Committee on CARCINOGENICITY

Committee on Carcinogenicity of Chemicals in Food, Consumer Products and the Environment (COC)

Statement COC/G07 - Version 1.1

Alternatives to the 2-year Bioassay

COC Secretariat
c/o Public Health England
Centre for Radiation, Chemical and Environmental Hazards
Chilton, Didcot, Oxfordshire OX11 0RQ

General introduction

1. Carcinogenicity studies of chemicals in laboratory animals usually employ higher doses of substances than those to which humans would normally be exposed and the response can be different to that of humans. Hence there are uncertainties in assessing carcinogenic risk in humans from the animal data, and steps are being taken to replace, where possible, the 2-year rodent bioassay with alternatives that more accurately predict carcinogenicity in humans. These alternatives should also limit the use of laboratory animals (in line with the '3Rs'[1] principle) and be more cost-effective.

2. This Guidance Statement provides an overview of approaches that have been proposed as alternatives, including some of the different types of animal and non-animal tests that may be used, and how it has been suggested these might be incorporated into an overall testing strategy. It is part of the Committee of Carcinogenicity (COC) guidance statement series which provides the Committee's views on all aspects of carcinogen risk assessment. It should be read in conjunction with G03 Hazard Identification and Characterisation: Conduct and Interpretation of Animal Carcinogenicity Studies. Guidance Statement G07 comprises 4 parts:

a. *in vivo* assays
b. cell transformation assays
c. omics, high-throughput screening technologies, and bioinformatics
d. alternative testing strategies for carcinogens incorporating results from short-term tests

3. The conduct of 2-year bioassays in 2 species, usually rat and mouse, has underpinned carcinogenicity risk assessment since the standard assay was developed in the 1960s (Cohen, 2010a,b). The objective of these long-term studies is to observe animals for the development of neoplastic lesions following exposure to a test substance for a major part of their life-span. The studies are usually designed to conform to closely defined test protocols and procedures (OECD TG 451 and 453, see Guidance Statement G03).

4. A significant body of data is available, particularly from the US National Toxicology Program (NTP), which has evaluated a large number of known carcinogens using the standard 2-year bioassay. Carcinogenicity testing strategies were developed taking into consideration the assumptions that, biologically, humans and animals are intrinsically similar and that carcinogenesis is a multistage process (Boobis et al., 2009). However, it has become evident that the conditions under which chemicals are tested are not necessarily relevant to human exposure, for example, the use of the maximum tolerated dose (MTD) and that some modes of carcinogenic action (MOA) are not relevant to human risk assessment. Furthermore, standard carcinogenicity study protocols involve the use of large numbers of animals (approximately 400-500 of each species) and, with increasing concern surrounding

[1] Replacement, refinement and reduction – see www.nc3rs.org.uk/the-3rs

unnecessary or poorly designed studies, efforts are being made to reduce animal use and to develop testing strategies that are more refined, in line with the principles of 3Rs.

5. The use of both rat and mouse 2-year bioassays in assessing the carcinogenic potential of chemicals has been subjected to close scrutiny. Several detailed evaluations of datasets have been undertaken with a view to assessing the utility of the mouse bioassay and the relevance of non-genotoxic, liver-only rodent carcinogens (Schach von Wittenau & Estes, 1983, cited by Alden et al., 1996; Huff et al., 1991, cited by Alden et al., 1996; Billington et al., 2010; Osimitz et al., 2013).

6. These investigations and analyses suggest that a single 2-year rodent assay is sufficient for cancer hazard identification. This view is endorsed by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), which indicates that bioassay data from only 1 species (eg the rat) is required for evaluation of carcinogenic potential, when supported by appropriate mutagenicity and pharmacokinetic studies and a study from a short-term *in vivo* assay, such as a transgenic mouse model (ICH, 1998).

7. For chemicals, some alternative strategies to the 2-year bioassay are being developed, which incorporate short-term endpoints (eg histopathology findings from toxicology studies) in carcinogenicity evaluations based on tiered and weight of evidence-based approaches, focusing on human-relevant modes of action. These methods vary depending on the type of compound being evaluated and the purpose of the evaluation, and it is not yet clear whether they will be suitable for risk assessment purposes.

8. ICH is now prospectively testing a strategy for evaluation of pharmaceuticals using a weight of evidence approach to define situations where a complete waiver of a 2-year bioassay would be justified (ICH, 2016a).

9. As well as alternative *in vivo* models, *in vitro* cell transformation assays have been developed as alternative methods to detect carcinogenic potential, in particular for use in testing scenarios where *in vivo* testing is not permitted (eg cosmetics testing). ‘Omics’ technologies, such as genomics, proteomics and metabolomics, enable detailed examination of chemically induced changes in the regulation of genes, proteins and metabolite profiles, respectively. These methods are considered useful in providing insight into the mode and mechanism of action of chemical carcinogens and as a prioritising and/or predictive tool for carcinogen identification. In parallel, high-throughput screening methods are being developed to screen large numbers of chemicals over a wide range of assay conditions. These newer methods produce large amounts of information that can be integrated and analysed using bioinformatics tools. They show promise for the future but are not yet sufficiently developed or validated to be used in the formal assessment of carcinogenic risk to humans from chemicals in the environment.

10. The following parts of this Guidance Document present the Committee's opinions and views on the approaches with the potential to be used as alternatives to the 2-year rodent bioassay in a carcinogenicity testing strategy.

Alternatives to the 2-year bioassay

COC/G07: Part a) *In vivo* assays

A1. A number of alternative animal models have been developed for the prediction of carcinogenesis. In a regulatory setting, ICH Guideline S1B (ICH, 1998) supports the use of certain alternative models instead of a second species (usually, but not exclusively, the mouse) in the carcinogenicity testing strategy for the evaluation of human pharmaceuticals. It states the following:

> "Several experimental methods are under investigation to assess their utility in carcinogenicity assessment. Generally, the methods should be based on mechanisms of carcinogenesis that are believed relevant to humans and applicable to human risk assessment. Such studies should supplement the long term carcinogenicity study and provide additional information that is not readily available from the long term assay. There should also be consideration given to animal numbers, welfare and the overall economy of the carcinogenic evaluation process. The following is a representative list of some approaches that may meet these criteria and is likely to be revised in the light of further information.
>
> a) The initiation-promotion model in rodents. One initiation-promotion model for the detection of hepatocarcinogens (and modifiers of hepatocarcinogenicity) employs an initiator, followed by several weeks of exposure to the test substance. Another multiorgan carcinogenesis model employs up to 5 initiators followed by several months of exposure to the test substance.
>
> b) Several transgenic mouse assays including the $p53^{+/-}$ deficient model, the Tg.AC model, the TgHras2 model, the XPA deficient model, etc.
>
> c) The neonatal in utero rodent tumorigenicity model"

These 3 models are considered below for the current COC guidance in an order considered to be predominance of use.

i) Transgenic (Tg) animal models

A2. A number of genetically modified mouse strains have been developed with the aim of providing models to facilitate the quick and accurate detection of chemical carcinogens. The mice develop tumours much more rapidly than wild-type mice as the transgenic modifications involve genes critical to the carcinogenic process. This underpins their utility in risk assessment strategies. The Health and Environmental Sciences Institute (HESI) (at the time part of The International Life Sciences Institute (ILSI)) co-ordinated a research and validation programme of work which evaluated the most commonly used models as part of the Alternative Cancer Test (ACT) programme: the $p53^{+/-}$ hemizygous knockout mouse, the rasH2 model, and the Tg:AC skin model. The COC evaluated this programme of work and other alternative

models for carcinogenicity testing (the $Xpa^{-/-}$ and $Xpa^{-/-}$ $p53^{+/-}$ transgenic mouse models and the neonatal mouse model) in 2002 (COC, 2002)

A3. The overall conclusion was:

> "The COC agreed an overall conclusion that none of the models used in the ILSI/HESI Alternative Cancer Test programme were suitable as a replacement for the mouse carcinogenicity bioassay (the primary purpose for the development of these models) and that further research should look to identify models with a greater relevance to mechanisms of carcinogenicity in humans. Of the animal models assessed there was evidence that $p53^{+/-}$ transgenic mouse model could identify some genotoxic carcinogens. There was insufficient data to suggest that the other animal models under consideration (RasH2, Tg.AC, Xpa, Xpa/$P53^{+/-}$ and $p53^{+/-}$) provide essentially similar results."

A4. Since the 2002 COC review, a number of studies and overviews evaluating the utility of the $p53^{+/-}$, RasH2 and Tg.AC models have been published and these have been considered for the current guidance. The Xpa/$P53^{+/-}$ model has not been considered as it is no longer commercially available.

$p53^{+/-}$ hemizygous knockout mouse

A5. $p53^{+/-}$ knockout mice are heterozygous for the tumour suppressor gene p53 - a point mutation in the remaining allele gives rise to a short latency period to tumour development. However, they have a low spontaneous tumour rate at 9 months thus making them sensitive to the detection of chemically-induced tumours, particularly those caused by genotoxic chemicals (French et al., 2001; Pritchard et al., 2003). The standard protocol involves daily oral dosing for 26 weeks, 5 dose groups including negative and positive controls, 25 mice/sex/group and extensive macroscopic and histopathological examination of tissues at the end of the study period. It is noted that it has been common practice to include a wild type control group at the high dose level and a wild type control group to establish whether any tumour response is related to inactivation of the p53 gene. The ILSI/HESI project on ACT examined early assay performance, spontaneous tumour incidence and results of commonly used positive controls (eg p-cresidine) and provides a comprehensive evaluation of the assay (Storer et al., 2001). Of the 16 genotoxic human and/or rodent carcinogens evaluated, 12 were positive (75%). Seven putative non-human carcinogens, which are rodent carcinogens, were examined. Two were equivocal (chloroform, DEHP) and the other 5 were negative (chlorpromazine, haloperidol, metaproterenol, sulfamethoxazole, WY-14643). The 3 non-genotoxic, non-rodent carcinogens were negative.

A6. The $p53^{+/-}$ knockout model has demonstrated the ability to identify hormonal carcinogenic mechanisms (diethylstilboestrol (DES), 17β-oestradiol) and immunosuppressive carcinogens (cyclosporine A), although the results are inconsistent (Storer et al., 2001; Alden et al., 2002). Some concerns have been

raised within the pharmaceutical industry with regards to assay performance during a review of the use of the assay. This includes some negative results in the $p53^{+/-}$ model following a positive *in vitro* clastogenicity response (Storer et al., 2010). It is noted that the genetic background (ie mouse strain) can have an influence on the biological outcome, for example, TSG $p53^{+/-}$ mice treated with the laxative phenolphthalein developed lymphoma, whereas $p53^{+/-}$ mice from the CBA and CIEA strains did not (Okamura et al., 2003). More recently, an evaluation of 52 NTP-tested chemicals (37 positives, 15 negatives) showed that the concordance of $p53^{+/-}$ mouse with NTP mouse carcinogens was 57% (Eastmond et al., 2013). It was noted that there is no biological reason why the $p53^{+/-}$ model would not be able to detect non-genotoxic carcinogens. However, the conclusion that the model is sensitive to genotoxic carcinogens but not non-genotoxic carcinogens remains following further evaluations of the assay data (Jacobsen-Kram et al., 2004; Storer et al., 2010).

A7. The Committee notes that the $p53^{+/-}$ knockout model has been shown to give some unexpected results and is not considered to be reliable for detecting non-genotoxic carcinogens. US and EU regulatory authorities do not consider it to be an acceptable model to replace the 2-year rodent bioassay

rasH2 model

A8. The rasH2 model is a hemizygous transgenic mouse which carries the human *c-Ha-ras* gene with a point mutation and its own promoter elements (Morton et al., 2002). These mice develop spontaneous and chemically induced tumours more rapidly than their non-transgenic counterparts and this enhanced sensitivity to neoplasia underpins the rationale for the utility of this model for carcinogenic risk assessment. The standard protocol is essentially the same as for the $p53^{+/-}$ model with 25 mice/sex/group (Nambiar and Morton, 2013). A positive control response can be elicited by a single dose of N-methyl-N-nitrosourea (MNU).

A9. Data from the ILSI/HESI ACT trial indicate the utility of the rasH2 model for detecting both genotoxic and non-genotoxic chemicals (Usui et al., 2001). Two of the 3 genotoxic human carcinogens tested were positive (cyclophosphamide, phenacetin) whilst melphalan generated equivocal results. DES and 17β-oestradiol were positive and negative respectively, and the immunosuppressive cyclosporine A was equivocal. Of the 11 non-genotoxic rodent carcinogens tested, 10 were negative; clofibrate gave equivocal and positive results in 2 separate studies. Analyses of 37 IARC classified chemicals indicated an 81% accuracy when assessing assay performance with regards to human carcinogenicity (Pritchard et al., 2003). More recent test results provide some evidence that the rasH2 assay also has the capacity to identify some non-genotoxic rodent carcinogens (namely clofibrate, DEHP and WY-14643, ethylenethiourea, ethylacrylate, 1,4-dioxane, troglitazone), though the majority of the assays of this class of chemicals were negative (Storer et al., 2010).

A10. A recent report reviews data from studies evaluating 10 chemicals in the rasH2 model in pharmaceutical laboratories and compared outcomes with the

conclusions from 2-year rat bioassays (Nambiar and Morton, 2013). All chemicals tested were negative in genotoxicity tests. Two of the 10 chemicals were positive in the rasH2 model. Both of these chemicals were also positive in rat 2-year bioassays at the same histological sites and were also associated with proliferative findings in the target organs. Non-genotoxic MOAs were assumed for these chemicals. A review of the spontaneous tumours and histology in rasH2 mice from 11 studies indicated little variation in the background incidence and consistent qualitative and quantitative responses with MNU as the positive control (Nambiar et al., 2012). These studies provide control data, which aid the interpretation of studies and support the use of this model as an alternative to the mouse 2-year assay. Another review of NTP chemicals tested in mice indicated an overall 82% concordance of the rasH2 assay with the mouse 2-year bioassay (16/20 positives, 7/8 negatives) (Eastmond et al., 2013).

A11. The Committee concludes that the rasH2 model performs adequately and can be considered as an alternative to the mouse bioassay.

Tg.AC skin model

A12. Tg.AC transgenic mice are hemizygous for mutant *v-Ha-ras* and can be considered as genetically 'initiated' due to the presence of this transgene. This model differs from the other 2 models as the most commonly used protocol involves topical application of the test chemical and the induction of squamous cell papillomas or carcinomas as the endpoint (Tennant et al., 2001). The protocol comprises topical application of the test chemical 3 times per week for 26 weeks. Evaluation of the assay in the ILSI/HESI ACT indicated that the Tg.AC model detects both genotoxic and non-genotoxic human carcinogens but only 9/14 chemicals positive in a standard 2-year bioassay with a variety of carcinogenic modes of action demonstrated to be active in the model. The number of false positives was low (1/14), therefore the model was not considered over-sensitive (Tennant et al., 2001). However, more recent experience has shown that irritant chemicals, and some commonly used vehicle formulations which are slightly irritant, can increase the yield and latency of skin tumours (Lynch et al., 2007), therefore a positive result can be complicated by confounding inflammation.

A13. In a separate evaluation, 27/35 (77%) chemicals were accurately predicted for carcinogenesis (23 carcinogens, 12 non carcinogens) (Pritchard et al., 2003). A recent review indicates a 61% concordance between the Tg.AC assay and NTP mouse carcinogens (12/23 positives, 8/10 negatives) (Eastmond et al., 2013). It is considered that the Tg.AC model is able to predict both genotoxic and non-genotoxic carcinogenesis when they are applied dermally. However, due to the persistent concerns with regards to tumourigenesis caused by inflammatory or irritant properties of chemicals the Tg.AC mice model is now generally considered unreliable for general use if interpretation of a positive result is complicated by confounding inflammation (Jacobs and Brown, 2015). Therefore its value in a carcinogenicity testing strategy is considered to be limited and accordingly it is not recommended.

Evaluation of the transgenic animal models

A14. A comprehensive overview and evaluation of all 3 assays ($p53^{+/-}$, Tg.AC and rasH2) used various combinations of the models, with and without consideration of rat 2-year bioassay data, to predict the carcinogenicity of 99 chemicals. It was concluded that correct identification of human carcinogens and non-carcinogens was 74-81% (whilst the similar evaluation of 2-year bioassay data was 69%). However some IARC Group 1 and Group 2A carcinogens were not identified and there were also a few false positives (Pritchard et al., 2003). A more recent evaluation of the 3 principal models suggests that, used alone, these assays would miss some probable human carcinogens (phenacetin, glycidol, 17β-oestradiol) (Storer et al., 2010). Furthermore, several issues of concern have been highlighted: methodological uncertainties, such as the effect of sample size on assay sensitivity and variability in spontaneous tumour frequencies, and reproducibility issues have been raised, together with questions on whether or how the dose-response data can be used for human risk assessment (Eastmond et al., 2013).

A15. A survey devised by the Carcinogenicity Alternative Mouse Models (CAMM) working group (Long et al., 2010) elicited 21 responses (90% of responses were from pharmaceutical organisations and 75% had used CAMM to support product development). The most commonly used model was the $p53^{+/-}$ mouse model with fewer laboratories using the rasH2 mouse model. There was only 1 example where the regulatory agency had rejected the submitted data. Feedback from agencies on study design was most often concerned with dose selection, in particular, whether the proposed high dose level was sufficiently close to the MTD to adequately ensure the sensitivity of the test. The most common positive controls used were p-cresidine for the $p53^{+/-}$ model, and urethane and MNU for the rasH2 model. However, it was considered by some respondents (5/15) that a positive control was not required if the model was well characterized within their laboratory. The tissues/organs which require pathological examination in the positive control animals are still under debate (ie all or only target organs). The importance of dose level selection was also discussed. Recommendations were proposed by the CAMM working group to improve study design and regulatory acceptance of transgenic animal studies.

Committee's evaluation of transgenic models

A16. The Committee's overall conclusion is that the rasH2 mouse model is the first choice model to replace the conventional mouse long-term bioassay as the assay has been shown to perform adequately for both genotoxic and non-genotoxic human carcinogens and is not overly sensitive. However, currently it is only supported when undertaken in addition to a rat 2-year bioassay (ICH, 1998). It is noted that, in a typical carcinogenic risk assessment strategy, chemicals with genotoxic properties will have been identified using the standard genotoxicity testing battery. Therefore the $p53^{+/-}$ assay is considered less useful than the rasH2 model as it is considered that it has an uncertain ability to predict chemicals with the potential to be carcinogenic in the absence of DNA reactivity. However, because of concerns about the insensitivity of C57BL $p53^{+/-}$ mice to detect non-genotoxic carcinogens, they have

not been used routinely to test compounds which have given negative results in genotoxicity assays.

A17. The Committee notes that transgenic assays can also provide insight into carcinogenic mode of action. For example, they may be useful for investigating chemicals where high dose cytotoxicity or cell proliferation leads to the development of age-related tumours or where the carcinogenic MOA is attributable to pharmacodynamics action. Attention is drawn to the need for rigorous optimization of protocols and validation of study designs, and it is recommended that attempts are made to improve the understanding of false positives and negatives.

ii) In utero/neonatal exposure models of carcinogenesis

A18. The Committee evaluated the rat neonatal model of carcinogenesis in 1998 as part of the ICH initiative and the conclusions are provided in a statement (COC, 1999). It was noted that there were very limited validation data and the Committee concluded that the available information showed tumour yields with genotoxic carcinogens were highly dependent on the strain of animal, age at start of treatment, and treatment protocol. There were no validation data regarding the use of short-term neonatal rodent bioassays for the identification of non-genotoxic carcinogens. Overall, the Committee concluded that there was no current evidence to support the use of the neonatal mouse or rat bioassays as part of the regulatory testing strategy for human medicines. Data retrieved since the 1998 review were limited to mechanistic studies; for example, investigating arsenic-induced murine carcinogenesis (Tokar et al., 2011; Waalkes et al., 2006a; Waalkes et al.,2006b; Ahlborn et al., 2009).

Committee's evaluation of the in utero/neonatal model

A19. The Committee considers that whilst ICH Guideline S1B (ICH, 1998) allows the use of the neonatal mouse model, there are limited data available. The majority of the studies are investigations designed on a case-by-case basis and as such there is no single protocol. The Committee concludes that the model is not relevant and not suitable as a general replacement for a 2-year bioassay.

iii) Initiation-Promotion models

A20. In the Solt Farber initiation-promotion model, rats are treated with a single dose of diethyl nitrosamine (DEN) as an initiator, followed by partial hepatectomy and repeated treatment with the test compound for several weeks to stimulate the formation of glutathione-S-transferase positive (GST+) foci which are considered to be pre-neoplastic lesions. The method was originally published in 1976 (Solt & Farber, 1976) and was developed and refined to become what is known as the Ito liver model (Ito et al., 1996; Ito et al., 2003). This is a medium-term treatment strategy and is based on the recognition that a large number of known carcinogens (genotoxic and non-genotoxic; >50% is quoted) are hepatocarcinogens in rodent bioassays and it is believed that the mode of action of many is mitogenic by stimulating hepatocyte proliferation. A multi-organ model based on the same

principles was subsequently developed with the goal of identifying the carcinogens not detected by the Ito liver model.

A21. A single published study evaluating the model concluded that, of the 159 compounds tested, 61 of 66 rodent liver carcinogens were identified as positive, 10 of 43 compounds which were carcinogens but not in the liver (non-hepatocarcinogens) showed a positive result and 1 of 50 non-carcinogens was positive in this assay (Tsuda et al., 2010).

A22. Two published studies using the multi-organ initiation-promotion model described the testing of 44 chemicals. All of the 12 rodent liver carcinogens, 10 of the 11 non-hepatocarcinogens and 0 of the 1 non-carcinogens were positive in this assay (Fukushima et al., 1991; Ito et al., 1996).

A23. These models of carcinogenesis were developed over 40 years ago, and the Committee notes that there are few studies using this methodology published in the literature other than those published by the originators of the protocol.

Committee's evaluation of the initiation-promotion models

A24. The Committee concludes that initiation-promotion models are not suitable for use in a carcinogenicity testing strategy, but may be useful to investigate the mode of action of certain carcinogens.

COC
January 2016

Alternatives to the 2-year bioassay

COC/G07: Part b) Cell transformation assays

B1. The Committee on Mutagenicity (COM) recently undertook a detailed review of the available cell transformation assays (CTAs). The assays considered were: SHE pH6.7 or pH7.0; BALB/c 3T3; C3H10T1/2; and Bhas 43. A statement was produced in which it was concluded that, to date, the CTAs are not suitable for use in a regulatory testing strategy for carcinogenicity. However, they may have value in predicting rodent carcinogenicity if used in the scenario where *in vitro* positive results were obtained for a cosmetic ingredient and no *in vivo* testing is allowed. It is noted that the OECD is pursuing the improvement and validation of the cell transformation assays and the COM and COC are actively involved in monitoring and contributing to their development (COM, 2012).

B2. The COC accepts the conclusions reached by the COM.

COC
January 2016

Alternatives to the 2-year bioassay

COC/G07: Part c) Omics, high-throughput screening technologies, and bioinformatics

1. Introduction

C1. Characterisation of the carcinogenic potential of the vast number of untested chemicals present in the human environment using conventional *in vivo* bioassays is not feasible and alternative methods are required. New approaches that are being developed include omics technologies and high-throughput screening (HTS) assays. The goal of these approaches is to develop predictive methods that are rapid, cheaper than current bioassays, and/or high throughput, based on human-relevant mechanisms of carcinogenesis.

2. Omics technologies

C2. The collective term ‘omics’ refers to the genomic (DNA sequence analysis) and post-genomic (eg transcriptomics, proteomics, metabolomics, epigenomics) technologies that are used for the characterisation and quantitation of pools of biological molecules (eg DNA, mRNAs, proteins, metabolites), and the exploration of their roles, relationships and actions within an organism (Ward & Daston, 2014). The term ‘toxicogenomics’ is sometimes used to describe the application of omics technologies to the study of adverse effects of toxicants or environmental stressors (Waters, 2016). The Committee will use the term ‘omics’ out of preference as it avoids the suggestion of being focussed on genomic techniques.

C3. Chemicals producing similar types and levels of toxicity are expected to share similar gene, protein or metabolite expression profiles, and such patterns of toxicant-induced molecular changes (‘fingerprints’ or ‘signatures’, sometimes referred to as biomarkers) can be used to assess toxicity. Omics methods may identify changes at much earlier time points than adverse effects observed at the tissue, organ or whole-organism level, and the post-genomic technologies can be used to follow toxicant-induced changes dynamically. Omics methods produce large amounts of biological information that can be integrated and analysed using bioinformatics tools.

C4. The aim of predictive omics in carcinogenicity evaluation is to create high-resolution profiles of biological responses, and to map causal events, processes and pathways that occur as a function of dose and time, reflecting carcinogenic modes of action (Waters, 2016). Omics methods are not yet suitable as high-throughput screening tests, but have shown great utility in determining mechanisms of action of chemical carcinogens, and as a prioritising and/or predictive tool for carcinogen identification. They are being developed to evaluate the effects of exposures to genotoxic and non-genotoxic carcinogens, *in vivo* and *in vitro*. Studies *in vivo* have been used mostly to identify mechanisms of carcinogenicity in rodents (eg Guyton et al., 2009; Fielden et al., 2011; Uehara et al., 2011) and for the classification and prioritisation of compounds for further evaluation (eg Ellinger-Ziegelbauer et al. 2008;

Thomas et al., 2009; Watanabe et al., 2012; Yamada et al. 2012; Melis et al., 2014).

C5. Several groups have reported studies to predict the outcomes of 2-year rodent bioassays by applying omics methods to short-term (from single to 90-day exposures) studies *in vivo*. The majority of these studies have focussed on mRNA profiling in rat liver, but proteomics, microRNA profiling, and metabolomics methods have also been employed (eg Yamanaka et al., 2007; Waterman et al., 2010; Koufaris et al., 2012; Ament et al., 2013). Gene signatures have been identified to discriminate between direct- and indirect-acting genotoxic carcinogens, non-genotoxic carcinogens and non-carcinogens (reviewed by Waters et al., 2010; Auerbach, 2016). The identification of non-genotoxic carcinogens using omics biomarkers is complex due to the large variety of modes of action involved, many of which are tissue-specific. Some modes of non-genotoxic carcinogenicity, for example oxidative stress, may show very early signature gene expression changes after a single exposure and repeat-dose studies can then be useful to determine 'false positives'. Auerbach et al. (2010) reported that signatures for non-genotoxic hepatocarcinogenicity in rats were more predictive from 90-day than from shorter-term studies. From this, the authors proposed the concept of a 'shared cancer biology', whereby a common pre-cancerous biology may be identified by common gene expression markers that are to some degree independent of the specific exposure. 'Profiling to the phenotype' takes as the starting point a transcriptional profiling of tissue samples corresponding to cancer pathologies identified in 2-year bioassays, 'working backwards' to use these profiles as markers for earlier prediction based on the shared pre-cancer biology concept. Such data could be cross-referenced to archived human tissue samples to improve human relevance (Waters, 2016).

C6. Gene expression studies in cultured cells exposed to toxicants have also focussed mainly on liver, using either primary hepatocytes or cell lines. These studies have shown utility in identifying genotoxic carcinogens, for which the importance of using p53-competent cell types is emphasised. *In vitro* studies have proven less useful for discriminating non-genotoxic carcinogens, in large part due to the wide diversity of modes of action involved. Indeed, the feasibility of using *in vitro* models for predicting the development of cancer *in vivo* has been questioned, for reasons including the following: the carcinogenicity of a chemical may require the presence of and interaction between different cell and tissue types in an organism, the biotransformation of parent compounds into metabolites, and correlation of effective doses *in vitro* with corresponding concentrations in different tissues in an organism (Waters, 2016). Nevertheless, methods are considered to be useful in characterising toxicity pathways to elucidate modes of action (Doktorova et al., 2012; Luijten et al., 2016).

C7. A 'comparison approach' to the identification of non-genotoxic carcinogens using *in vitro* omics-based studies has been described, whereby a limited set of the most significantly up- and down-regulated genes is compared for overlap across different chemical exposures, to identify the best match for a chemical of interest. A test of this method to categorise chemicals by their mode of action using primary

mouse hepatocytes or mouse embryonic stem cells indicated the requirement to use a combination of different *in vitro* systems, and these studies are being extended to incorporate tests over chemical concentration ranges (Schaap et al., 2015, 2016).

C8. A large catalogue of (*in vivo* and *in vitro*) datasets is now available, based on a large set of compounds, consistent study designs and standardised experimental protocols. Databases contain dynamic gene expression data over multiple doses/concentrations plus companion data (eg compound pharmacology, toxicology, clinical chemistry and histopathology). This information can be used for 'phenotypic anchoring' – relating specific changes in gene-expression profiles to adverse effects observed in conventional toxicity tests, to allow the identification of gene-expression changes that are causally related to the development of the toxicity phenotype (Paules, 2003). Studies should now be extended to include targeting of organs/cell populations other than liver, the abstraction from individual signature genes to higher-order levels, such as pathway enrichments and molecular interactions, and the integration of expression data obtained across multiple omics platforms (DNA, mRNA, miRNA, protein, metabolites) (Römer et al., 2014).

C9. It has been suggested that the 'parallelogram approach' could be useful to compare early key events and toxicity pathways indicated by omics studies performed using sets of chemicals with well-established apical endpoints, to evaluate the likelihood of a similar mode of action in humans. This method, initially proposed by Sobels (1977) and further developed by Sutter (1995), can be used in the assessment of risk to humans by extrapolating findings from 2 different *in vitro* model systems, 1 of which should be human (eg rodent *in vitro* and human *in vitro*) and from *in vivo* studies in the non-human species (eg rodent *in vivo*). The parallelogram approach has been applied to studies in hepatotoxicity, integrating phenotypic and omics data from rodent studies *in vivo* with data obtained using rodent and human hepatocytes *in vitro*. The 'concordance model' extends this approach to include data from several animal species plus several *in vitro* (human) assays, which should lead to a greater level of confidence in the biological significance of the common toxicity pathways identified (Kienhuis et al., 2016). These approaches may also be applicable to carcinogenicity evaluations.

C10. Progress is being made to integrate omics data into quantitative cancer risk assessments. Dose-response assessments are applied to derive points of departure (PoDs) for omics-derived endpoints, usually benchmark doses (BMDs; often the lowest BMD), which can be compared with PoDs from conventional/apical endpoints. Case studies using transcriptomic biomarkers for several model compounds have been described, and the standardisation of study protocols as well as methods to derive BMD values have been discussed (reviewed by Thomas and Waters, 2016). A mode of action-based context is preferred in applying transcriptomic dose-response to the derivation of the BMD. Dose-response studies performed over time can relate BMD value changes with adverse responses to identify transcriptional changes that are progressive or resolve. Use of the 'most-sensitive BMD' derived from omics data may lead to an over-conservative risk assessment as the most sensitive changes in gene expression/ pathway alterations may represent adaptive

rather than toxicity endpoints. At present, this issue is addressed by phenotypic anchoring to traditional apical endpoints, with the intention that there will eventually be sufficient well-validated data that apical endpoints *in vivo* will no longer be required. The parallelogram approach and concordance model can be integrated to select pathways of human biological significance (Kienhuis et al., 2016).

C11. Thomas et al. (2013) outlined a framework for applying transcriptomic data to (non-cancer and cancer) risk assessment. The proposed weight of evidence analysis incorporates estimation of genotoxic potential and an extrapolation factor based on the PoD estimated from the lowest BMD determined from transcriptomic dose-response studies in 8 specified tissues at a single time point between 5 days and thirteen weeks in rats and mice. The assumption is that basing the PoD on the most sensitive pathway is generally protective until key adverse effect pathways are identified. This approach might be applicable to obtaining margins of exposure when cancer data are not available, but advice on relative risk is required. Thomas and Waters (2016) commented that although there may be issues of concern in using such an approach, pragmatically, a PoD based on such information may be preferable to no PoD, which is currently the case for the vast majority of chemicals.

C12. The advent of omics technologies has increased vastly the amount of data available, but the value of this data as predictive information has not matched expectations. In part this is because of the very real difficulty in assimilating data from different sources, for example genomic and metabonomic, and from it deriving a single ground truth. The widespread adoption of artificial intelligence, in particular deep learning where multi-scale integration and use of less-than-perfect data can occur to identify patterns, promises real progress in this area. Many data mining initiatives to date have relied on supervised analysis, looking for patterns that can predict a relatively general, pre-set outcome, such as cancer occurrence in animal model. Increased computing power, even larger datasets and new technologies allow truly unsupervised learning, where complex patterns can be identified that may stratify responses into groups, offering a degree of subtlety that is more appropriate for authentic studies in human populations. Coordinated work in this field across industry and academia is essential (Luechtefeld & Hartung, 2017).

3. High-throughput screening

C13. Individual omics-based assays can provide information about multiple changes (eg expression levels of large numbers of genes) in response to a chemical exposure, but they currently have limited applicability for use in high-throughput screening (HTS). Conversely, HTS methods, which evaluate only 1 or a small number of genes or processes per assay, are adapted to screen large numbers of chemicals over a wide range of assay conditions. They have the advantage of providing rapid, high-throughput, standardised testing of chemicals. A number of these methods were initially developed in the pharmaceutical industry for the rapid screening of libraries of candidate drugs or small molecules for specific types of biological activity or disease processes (Pereira and Williams, 2007) and are now being applied robotically to study chemical perturbations of biological pathways in

relation to toxicity.

C14. HTS assays comprise 2 general categories. Biochemical (cell-free) assays are usually homogenous reactions that measure effects on specific molecular targets and can be easily miniaturised. Cell-based assays can determine perturbations at different points in cellular pathways and are often run in multiwell formats (from Waters, 2016).

C15. HTS approaches are being developed with the aim to predict carcinogenicity *in vivo.* A wide range of doses can be tested in each individual assay allowing the description of dose-response curves at low (human-relevant) doses, which can be useful for comparison with low-dose omics and *in vivo* study data. HTS is of particular value for hazard identification and prioritisation for further testing, and can be run in parallel with structure-activity relationships (SARs) to predict potential targets prior to screening. A major challenge is how to incorporate the toxicokinetic and toxicodynamic parameters of *in vivo* studies, and some authors have questioned whether *in vitro* methods can actually be useful in risk assessment to support regulatory decision-making (see paragraph C6).

C16. The landmark report, 'Toxicity Testing in the 21st Century: A Vision and a Strategy' proposed a paradigm shift in toxicity testing from high-dose studies *in vivo* to an approach based on *in vitro* assays using human-relevant cells or tissues using a mode of action approach based on the evaluation of dynamic pathways underlying biological response (NRC, 2007; Bhattacharya et al., 2011). This concept has been generally labelled 'TT21C'. The aim stated is to test whether chemical compounds have the potential to disrupt processes in the human body that may lead to negative health effects. The 2 central aspects of the TT21C approach are the evaluation of innate cellular pathways that may be perturbed by chemicals and the determination of chemical concentration ranges in which these perturbations are likely to lead to adverse health effects. The TT21C approach is being evaluated in proof-of-concept studies using well-studied prototype compounds whose toxicity has already been examined with *in vivo* and *in vitro* assays.

C17. To date, the major initiatives applying the TT21C approach have been based in the US, in projects such as Tox21 and ToxCast. There are also various European projects moving to a toxicity pathway approach linked in with a reduction, replacement and refinement in the use of animals in toxicity testing. The AXLR8 consortium includes details of other EU funded research investigating these (http://axlr8.eu/, accessed 10/10/16).

C18. Tox21 (Toxicology in the 21st Century, www.epa.gov/chemical-research/toxicology-testing-21st-century-tox21, accessed 17/10/16)) is a collaboration in the US between partners at the National Institutes of Health (NIH), Environmental Protection Agency (EPA) and Food and Drug Administration (FDA) that began in 2008 in response to TT21C. The stated goals are to identify environmental chemicals that lead to biological responses and determine their mechanisms of action on biological systems, prioritize specific compounds for more

extensive toxicological evaluation, develop models that predict chemicals' negative health effects in humans, and annotate all human biochemical pathways and design assays (tests) that can measure these pathways' responses to chemicals. Tox21 utilises quantitative HTS *in vitro* assays and computational toxicology approaches to cover a range of cell responses and signalling pathways to rank and prioritise chemicals. The HTS assays target multiple genes, proteins, pathways and cancer-related processes. To date, over 10,000 chemicals have been screened in approximately 50 assays.

C19. The EPA ToxCast (Toxicity Forecaster) project is related to, but separate from, Tox21. The results from ToxCast form a contribution to Tox21. ToxCast uses a similar approach to Tox21, but includes a much wider range of assays and endpoints. To date, more than 1800 chemicals, including industrial and consumer products and food additives have been screened in the ToxCast program for over 700 endpoints. All of the resulting information is publicly available on a database, together with tools for visualising and analysing the data (www.epa.gov/chemical-research/toxicity-forecasting, accessed 29/09/16).

C20. In ToxCast Phase I, a set of around 300 chemicals with pre-existing toxicity data were run through >600 HTS assays. ToxCast HTS data relating to perturbation of carcinogenesis-related pathways were then used to develop a model for classifying carcinogens (mostly non-genotoxic) based on 2-year data in the EPA Toxicity Reference Database (ToxRefDB), comprising largely pesticides. This dataset was applied to an external test set of 33 pesticides. The model showed some (limited) capability to discriminate between possible/probable and negative/unlikely carcinogens, but several known carcinogens were identified as false negatives (Kleinstreuer et al., 2013). Further, independent analyses using this data set have found that assay design and coverage are not yet adequate and need development to improve the accuracy of prediction of rodent carcinogenicity and of the relevance of predictions to humans (Benigni, 2013; Cox et al., 2016). Problems faced in developing and improving the ToxCast assays are discussed in the review article by Benigni (2014), which concluded that the next phase should focus on including exogenous metabolic activation in the HTS assay systems and developing a set of well-characterised, standard carcinogens.

4. Summary and COC conclusions

C21. Use of the 2-year rodent bioassay to evaluate the carcinogenicity of the vast numbers of untested chemicals that are currently marketed is not feasible and alternative methods are required for this purpose. Newer approaches are being developed, such as omics technologies and high-throughput screening (HTS) assays. The goal of these approaches is to develop predictive methods that are rapid, cheaper than current bioassays, and/or high throughput, based on human-relevant mechanisms of carcinogenesis.

C22. Omics technologies may be useful as a part of new strategies (eg Cohen, 2004, 2010b and 2018) based on human-relevant modes of action. To date, most

studies have used transcriptomic methods, but newer approaches such as metabolomics show promise for the future. Omics approaches may be used to extrapolate between animal *in vivo* and *in vitro* experiments and human *in vitro* experiments to predict likely outcomes for humans *in vivo*. This requires the development of biomarkers, and while a lot of information has been generated in this area, a better understanding of the key markers is required before this can progress. This should include where assays evaluate genotype, understanding what effect this might have on phenotype.

C23. There is a need to continue to develop the appropriate statistical and informatics tools to analyse the large volumes of data generated by these technologies to ensure consistent and appropriate conclusions can be drawn.

C24. High-throughput screening (HTS) technologies using biochemical or cell-based assays that allow rapid screening of large numbers of chemicals over a wide range of concentrations may be useful for hazard identification and prioritisation, but are currently not useful for risk assessment.

C25. Overall, these newer methods show promise for the future but are not yet sufficiently developed or validated to be used in the formal assessment of carcinogenic risk to humans from chemicals in the environment. However, the Committee is aware of these developments and will keep progress in this area under review, and update or fully revise this statement as appropriate.

COC
November 2018

Alternatives to the 2-year bioassay

COC/G07: Part d) Alternative testing strategies for carcinogens incorporating results from short-term tests

1. Introduction

D1. The current paradigm for assessing the carcinogenicity of a chemical is a combination of genotoxicity testing and 2-year rodent bioassays. Genotoxicity/mutagenicity tests provide a useful screen to indicate positive carcinogenic potential for those compounds with a genotoxic mode of action, however they produce high numbers of false positive results, have little or no capacity to identify non-genotoxic carcinogens, and generally lack dose–response characterisation. Over the last few decades, the 2-year rodent bioassay has become the gold standard to assess carcinogenicity. Nevertheless there are limitations in extrapolating from the 2-year bioassay to human cancer risk; the primary disadvantage being that positive findings for carcinogenicity in rodents may be of limited or no relevance to human cancer risk due to issues of dose, species specificity and/or mode of action.

D2. New strategies are being developed to assess carcinogenicity in which the use of 2-year rodent bioassays is reduced or replaced with shorter term study data (from *in vivo*, *in vitro* and/or *in silico* tests). Some of these strategies have followed an approach of attempting to predict the outcome of 2-year rodent bioassays, with subsequent evaluation of the applicability of these findings to the human situation, while other approaches aim more directly at identifying and/or assessing the potential for carcinogenicity in humans. As these new strategies are currently in development, they have not been fully validated.

D3. The following section of this guidance statement is a review of developments to date and proposed strategies for future developments with relevance to this topic. The aim of the COC guidance will be to list some of the alternative approaches that are being developed, to discuss some of the advantages and disadvantages of using these approaches, and to consider the potential utility of such approaches for evaluating carcinogenic risks posed to the public from exposure to chemicals present at ambient levels in the environment.

2. Purpose of the Assessment

D4. Carcinogenicity studies are performed for a variety of reasons. These include hazard identification, hazard characterisation, and/or safety assessment of substances such as pharmaceutical products, industrial chemicals, food additives, cosmetics, and chemicals present in the general environment.

D5. The first step in a carcinogenicity assessment is normally a genotoxicity test battery. A lifetime rodent bioassay may then be required depending on the regulatory and legislative setting. For example, for small molecule pharmaceuticals intended for

continuous use or regular intermittent use, data from at least one 2-year rodent carcinogenicity bioassay are currently required by regulatory agencies. Conversely, the use of data from tests performed *in vivo* is not permitted for some products, such as the constituents of cosmetics intended for sale in the European Union. Regulatory frameworks for carcinogenicity testing of chemicals vary, but in many cases *in vivo* carcinogenicity bioassays are not performed and/or requested.

D6. For public health protection relating to chemicals present at ambient levels in the environment, the principal goals of carcinogenicity evaluations are the identification and risk assessment of human-relevant carcinogens. The aim is to decide whether exposure to a certain level of a particular chemical is acceptable in terms of the likelihood that it will cause cancer in humans, and to allow for management of this risk. The task is complex as the answer required is not a simple, binary 'yes' or 'no' but ideally a probabilistic evaluation of the risk effectively encountered by humans. It also depends on the cancer mode of action for the chemical.

D7. For application to the risk assessment of chemicals present in the environment, new systems for carcinogenicity evaluation would ideally have the potential to produce organ-specific, dose-dependent information relevant to humans.

3. History and Developments To Date

D8. The utility of short-term toxicological findings *in vivo* as an element to predict the outcomes of 2-year rodent carcinogenicity bioassays has been tested in several retrospective analyses of information in existing toxicological databases. Some studies have looked at the ability of short-term findings to predict tumourigenicity at the organ-specific level, whilst others have used a broader approach to evaluate whether the presence or absence of changes in short-term tests can predict tumourigenicity more generally at the whole-organism level.

3.1 Evaluations of the National Toxicology Program (NTP) database

D9. Allen et al. (2004) reviewed existing data in the NTP database with the aim to correlate specific hepatocellular pathology in pre-chronic studies (≤ 12 months) with carcinogenic endpoints in 2-year studies. Data were evaluated for mice (83 compounds) and rats (87 compounds). The pre-chronic endpoints evaluated were hepatocellular cytomegaly, hepatocellular necrosis, bile duct hyperplasia, hepatocellular hypertrophy, and hepatocellular degeneration (rats only). Increased liver weight was also included. Hepatocellular hypertrophy was the single most predictive feature (10/27 mouse carcinogens, 0 false positives; 5/11 rat carcinogens, 10 false positives). Three features as a group (hepatocellular necrosis, hypertrophy, and cytomegaly) correctly predicted carcinogenicity findings at 2 years for 17/27 (2 false positives) mouse and 7/11 (16 false positives) rat liver carcinogens. Inclusion of liver weight as a fourth criterion improved the sensitivity of the screen, but decreased the specificity (25/27 mouse carcinogens, 18 false positives; 11/11 rat carcinogens,

32 false positives). Genotoxicity results (*Salmonella* test and Micronucleus assay) did not correlate well with liver carcinogenesis outcomes in either mice or rats.

D10. The International Life Sciences Institute (ILSI) Health and Environmental Sciences Institute (HESI) conducted a retrospective analysis of the NTP database to test the hypothesis that the signals of importance for human cancer hazard identification can be detected in shorter term studies than the 2-year bioassay (Boobis et al. 2009). Sixteen chemicals were selected on the basis that they were positive in liver, kidney or lung in lifetime rodent (rat and/or mouse) carcinogenicity bioassays and that genotoxicity and short-term rodent study data were available. Thirteen-week study data for immuno-, liver, kidney and lung toxicity were reviewed for correlation against tumour outcomes in the corresponding tissues in 2-year bioassays.

D11. Genotoxicity data were obtained from the NTP database, summarised and evaluated. Data from at least 3 tests (Ames, *in vitro* chromosome aberrations, *in vivo* micronucleus) were available for 7 chemicals, while it was noted that a more recent NTP strategy had been to use the Ames test plus *in vivo* micronucleus assay. A literature review was carried out for 1 compound, for which no NTP genotoxicity data were available. An overall designation of genotoxic, not genotoxic, or equivocal was assigned based on all the available test data. Five chemicals were positive, 8 negative and 3 equivocal. The authors noted the requirement for a reliable battery of genotoxicity tests.

D12. Markers of *immune system* changes (downregulation, proliferation, or neoplasia) included haematology (total leukocyte, segmented neutrophil, lymphocyte, and monocyte counts), spleen and/or thymus weights, and histopathological findings in bone marrow, spleen, thymus and lymph nodes. None of the 16 chemicals caused direct immunosuppression in 13-week studies and there was no clear evidence of neoplasia in elements of the immune system. Several chemicals showed immune changes that were attributed to stress. The authors noted the requirement for further definition and evaluation of short-term tests for immunosuppressive effects, suggesting further work to include evaluations using a range of known positive and negative compounds.

D13. Liver findings examined for 13-week studies were organ weight, clinical pathology, and histopathology, including relative liver weight, hepatocellular hypertrophy, altered foci, hepatocyte necrosis, hepatocyte vacuolation, hepatocyte degeneration, bile duct hyperplasia, increased alanine transaminase levels, increased sorbitol dehydrogenase levels, and increased bile acid/bilirubin levels. Six chemicals were tumourigenic in the liver of rats, 9 in mice. In 13-week studies, liver weight was the best single predictor of tumour outcome (5/6 in rat, 6/9 in mouse). Grouping liver weight with other criteria increased the positive predictivity to 6/6 in rat and 8/9 in mouse. Considering the results collectively for rats and mice, there were no false positives, and 1 false negative (one chemical induced tumours in the mouse bioassay but no changes in rats or mice at 13 weeks). On this basis, the authors concluded that conventional liver endpoints currently identified in 13-week toxicity

studies were not adequate to identify all chemicals with carcinogenic potential and that additional endpoints may identify other key events that might more accurately predict carcinogenic potential in rats and mice. These would then be useful for defining modes of action to assess human carcinogenic potential and risk more effectively. Such endpoints include increases in cell proliferation (S-phase response) and induction/inhibition of apoptosis (measurement of labelling indices for both events), constitutive androstane receptor nuclear receptor activation (reporter assays), cytochrome P450 induction (direct biochemical measurement), and peroxisome proliferation (measurement of palmitoyl coenzyme A oxidase activity). The potential for –omics platforms to identify additional indicators was noted.

D14. Thirteen-week study criteria for *kidney* changes included hyaline droplets, inflammation, chronic progressive nephropathy, and absolute and relative kidney weights. Five compounds were tumourigenic in the kidney of rats, none in mice. All 5 chemicals were positive for 13-week changes. The best predictor of tumourigenicity at 13 weeks was increased kidney weight. When this parameter was combined with histologic findings, no false negatives were identified. The authors noted that 13-week study findings for kidney may give clues to carcinogenicity mode of action, which may help interpretation of human relevance (eg 4 chemicals that induced kidney tumours showed increased hyaline droplets, which indicates a rodent-specific mode of action that is not relevant to humans).

D15. For *lung*, diagnostic terms for histomorphologic alterations used by NTP to describe lung lesions in 13-week studies were: chronic active inflammation, inflammation NOS (not otherwise specified), alveolar epithelial hyperplasia, bronchiolar hyperplasia, proteinosis, fibrosis, histiocytic infiltration, and foreign body. In total, 11/16 correct predictions of lung carcinogenesis were made from short-term data. Seven chemicals induced tumours in rats and/or mice. Four of these showed inflammation and/or hyperplasia at 13 weeks and an additional 1 was genotoxic (giving 5 true positives, with the other 2 chemicals being false negatives). Two chemicals induced inflammation and/or hyperplasia at 13 weeks but did not show tumours in the 2-year study, ie were false positives, and there were 7 true negatives.

D16. Overall, the authors concluded that for most, but not all, of the chemicals producing tumours in 2-year studies, cellular changes indicative of a tumourigenic endpoint could be identified after 13 weeks using routine evaluations, but that such evaluations are not adequate to identify all non-genotoxic chemicals that will eventually produce tumours in rats and mice. Additional endpoints are needed to identify signals not detected with routine evaluation. Such endpoints might include BrdU labelling and a measure of apoptosis. Further efforts would be required to determine false-positive rates of this approach.

3.2 Pharmaceuticals

D17. Approaches are being developed to allow situations in which the regulatory evaluation of potential human cancer risks from pharmaceuticals may in some cases be made without the requirement for a 2-year rodent bioassay, based on the integration of other data using weight of evidence approaches.

Center for Drug Evaluation and Research (CDER)/FDA

D18. In 1998, the US FDA reviewed the use of 2-year rodent studies and alternative strategies for carcinogenesis testing and stated an aim to move away from reliance on the results of 1 test (the traditional lifetime bioassay in both sexes of 2 rodent species) towards a decision-making process based on a profile of data, using a weight of evidence approach that takes into account the increased knowledge of carcinogenic mechanisms that has been gained since the 2-year bioassay was adopted as a routine screen in the 1970s (Schwetz and Gaylor, 1998). A conceptual strategy was proposed, including a preliminary evaluation for genotoxicity to include data on physical–chemical properties, structure alert information, information from computer-based prediction systems and the results of a genetic toxicity screen, and subsequent tests to include transgenic mouse models and then possibly a 2-year rodent study. The inclusion of data relating to non-genotoxic mechanisms of carcinogenicity would be important, including the following mechanisms: hormone modulation, growth factor perturbation, cell proliferation (mitogenic, cytotoxic), inhibition of apoptosis, cell-to-cell communication, P450 induction, spindle fibre effects, altered methylation status, and specific mechanisms (ß-agonist, uterine tissue; H_2 antagonist, glandular stomach; peroxisome proliferation). It was proposed to evaluate these new test systems in parallel with the conduct of traditional 2-year bioassays.

D19. Jacobs (2005) compared the findings from short-term dose-ranging studies with the outcomes of 2-year rodent carcinogenicity studies for 60 pharmaceutical compounds in the CDER/FDA database. This evaluation considered liver, kidney, mammary, thyroid, adrenal, urinary bladder, lymph node/spleen, and lung. Contrary to the findings of Allen et al. (2004) (described in Section 3.1, above), short-term indicators such as hyperplasia, hypertrophy, increased organ weights, tissue degeneration or atrophy, and mineralisation were not reliable predictors of tumour outcome in the corresponding tissues in carcinogenicity bioassays. It was noted that some differences may be attributed to the different types of databases evaluated – many genotoxic and liver-toxic compounds are screened out in the pharmaceutical development process, there is greater variation in the rodent strains used for bioassays for pharmaceutical regulatory submissions than in NTP studies, and carcinogenicity bioassays for pharmaceuticals do not necessarily test the maximum tolerated dose.

NEG CARC

D20. Reddy et al. (2010), tested a 'whole animal negative predictivity' strategy, finding, in agreement with Jacobs (2005), that histopathological changes indicative of hyperplasia, cellular hypertrophy, and atypical cell foci were not reliable predictors of tumour outcome in the corresponding tissues. However, the complete absence of histopathological evidence of pre-neoplasia in all tissues in short-term toxicity studies was a reliable indicator for negative tumour outcome in a 2-year bioassay. In this study, 2-year rat bioassay data for 80 pharmaceuticals from commercial and Merck databases (30 carcinogens and 50 non-carcinogens) were compared with findings from corresponding 6- or 12-month toxicity studies.

D21. The 'whole animal negative' model specified the presence of pre-neoplasia (hyperplasia, cellular hypertrophy, and atypical cellular foci) in any single tissue (25 of the 30 carcinogens) as positive, and the absence of pre-neoplasia in all tissues (35 of the 50 non-carcinogens) as negative (sensitivity 83%, specificity 70%, negative predictive value 88%, positive predictive value 63%[2]). The 5 false negatives (ie negative from analysis of 6-month data but positive for tumours in 2-year rat bioassays) were all negative in genotoxicity assays and 2-year mouse carcinogenicity bioassays, and all produced tumours in rats based on proliferative or hormonal effects. The authors considered that the positive 2-year rat bioassay results for these 5 compounds were of questionable relevance to carcinogenicity in humans. They were all approved compounds currently marketed for non-life-threatening specifications and tumourigenicity was considered to be associated with rat-specific mechanisms.

D22. A larger project incorporating data from 13 companies was set up to further test the whole animal negative predictivity strategy proposed by Reddy and colleagues, using an expanded database maintained by the Pharmaceutical Research and Manufacturers of America (PhRMA) and including 182 pharmaceutical compounds (66 positive and 116 negative in 2-year rat carcinogenicity studies) (Sistare et al., 2011). In this study, negative outcome was specified as the absence of all of 3 criteria:

- genotoxicity
- any knowledge or significant evidence of hormonal perturbation activity
- evidence of histopathologic risk factors of rat neoplasia in all tissues examined in the corresponding chronic rat toxicity study conducted at similarly matching doses to those used in 2-year carcinogenicity studies.

[2] Sensitivity=TP/(TP+FN)X100, Specificity=TN/(TN+FP)X100, positive predictive value=TP/(TP+FP)X100, negative predictive value=TN/(TN+FN)X100 (TP=true positive, TN=true negative, FP=false positive, FN=false negative)

D23. This approach was termed ‘NEG CARC’ (Negative for Endocrine, Genotoxicity, and Chronic Study Associated Histopathological Risk Factors for Carcinogenicity).

D24. Immunosuppression was not included as a criterion on the basis that results in rat carcinogenicity tests do not reliably reflect human risk for this effect (Bugelski et al., 2010). It was noted there are likely to be significant differences between broad-based immunosuppressants and selective immune modulatory compounds that would be important to understand in helping to provide perspective for human risk assessment.

D25. *Genotoxicity* was assessed as any clear, single, positive genetic toxicology result in the good laboratory practice-compliant standard battery of assays that was not otherwise explained as an irrelevant finding.

D26. *Hormonal perturbation.* A weight of evidence approach was used, which included evidence of treatment-related microscopic and/or macroscopic changes in multiple endocrine tissues within a sex, measurements of changes in hormone levels, and knowledge of pharmacological mechanism of action (hormone receptor binding, alteration of hormone levels, alteration of activity of endogenous hormones).

D27. *Histopathology.* Positive findings were considered to be treatment-related hyperplasia, cellular hypertrophy, atypical cellular foci, or neoplasia in chronic studies (including multinucleated cells, basophilia, basophilic foci, cellular enlargement, cytomegaly, cellular swelling, cellular alteration, dysplasia, eosinophilic foci, karyomegaly, or tumour; excluding vaginal metaplasia and myocardial hypertrophy).

D28. The NEG CARC strategy identified 52 compounds as true positives (7 genotoxicity, 42 histopathology, 26 hormonal perturbation), 54 false positives (17 genotoxicity, 38 histopathology, 15 hormonal perturbation), 62 true negatives and 14 false negatives (sensitivity 79%, specificity 53%, negative predictive value 82%, positive predictive value 49% to predict rat carcinogenicity). Sensitivity was similar when considering endpoints at 6 or 12 months. As observed by Reddy et al. (2010), the sensitivity of microscopic findings to predict neoplasia in the 2-year rat study on an organ-by-organ basis was lower than on a whole-animal basis: for 9/42 true positives identified by histopathology, the tumour site in the carcinogenicity study did not match any of the positive tissues in the repeat-dose toxicity study (4 of these were considered to have hormonally linked mechanisms, 1 to be related to site of initial high exposure, and for 4 cases the mechanism was unknown).

D29. Eleven tissues (liver, thyroid, adrenals, ovaries, mammary gland, bone, pituitary, urinary bladder, kidneys, skin, stomach) served as sentinels in the 6/12-month studies for 90% of tumours in the 2-year studies. (The spectrum of positive tissues for the histopathology false positives was noted to be similar). Nine sites accounted for over 80% of tumours (liver, thyroid, ovaries, testes, urinary bladder, skin, mammary gland, kidneys, adrenals). The authors suggested that tissues with

the highest expected exposure after dosing or with high sensitivity to hormonal perturbations are thus most likely to be predictive of tumour risk to the rat. Many of the true positives were identified by early hormonal perturbation (often hormonal agents designed for this purpose). These were associated with ovarian granulosa cell, bone, mammary, testicular, pancreatic and/or thyroid tumours and all had earlier documented effects on hormones or hormonally regulated tissues in the rat in tissues related to the tumours seen in the lifetime bioassay. Development of such tumours in rats at sites distal to the primary drug target tissue was noted to be often due to rodent-specific mechanisms associated with chronic trophic hormonal stimulation at the target site for tumourigenesis that may or may not translate to humans.

D30. The human health relevance of positive 2-year rat bioassays for the 14 false-negative compounds was considered to be questionable and is discussed on a case-by-case basis. The overall conclusions were that the tumour signals were marginal, inconsistent across sexes, inconsistent across species and with a tendency to occur only at high doses. Ten of these compounds were marketed, 2 were not marketed for reasons unrelated to the rat carcinogenicity findings and 2 were still in development at the time of publication despite the positive rat carcinogenicity findings.

D31. An evaluation of data for 78 IARC Group 1 and 2A chemicals + 8 pharmaceuticals that had been withdrawn for cancer concerns was similarly carried out. Most of these (72) were positive for genotoxicity. Of the 14 non-genotoxic compounds, 10 would have been triggered for 2-year carcinogenicity testing by sub-chronic/chronic histopathology and/or known hormonal perturbation using the NEG CARC approach. Of the remaining 4 compounds, 3 were not carcinogenic in rats at doses that could be tolerated in 2-year studies (Group 1 – ethanol; Group 2A – 4-chloro-ortho-toluidine and tetrachloroethylene). Thus for ethanol (IARC Group I) the method would fail to predict the need to conduct a rat carcinogenicity study to identify a known human carcinogen, nevertheless the negative outcome of the rat study would have been correctly predicted. The other NEG CARC-negative compound (Group 1 – cyclosporine) was an immunosuppressant that would be expected to be negative in a 2-year rat assay but tumourigenic in humans.

D32. On the basis of this retrospective study, Sistare and colleagues proposed that a 2-year rat study is not necessary for compounds that are negative by the NEG CARC paradigm, and that human cancer risk assessment for such compounds can be carried out on the basis of a 6-month rat study + transgenic mouse study.

D33. Van der Laan et al. (2016) proposed that it would also be important to include the category ‘pharmacological evidence’ as part of the NEG CARC approach. This hypothesis was tested in a detailed evaluation of a primary dataset of 298 pharmaceuticals, including 191 compounds from the ‘PhRMA’ database evaluated by Sistare and colleagues, 44 compounds from the CDER/FDA database, and 63 compounds from the Japanese Pharmaceutical Manufacturers Association (JPMA) database. Excluding 43 compounds that did not have a primary mammalian

pharmacologic target (ie antivirals/antimicrobials), 255 compounds were categorised into 6 pharmaco-therapeutic areas (CNS, cardiovascular, respiratory, metabolic, hormonal, anti-inflammatory and immunomodulatory) plus 'remaining' compounds. Within these categories, 172 of the 255 compounds were sub-categorised into 45 pharmacological classes according to the primary drug target. Classification of tumourigenicity, based on the 'NEG CARC' criteria (genotoxicity + short-term histopathology + hormone perturbation) and on findings in 2-year rat bioassays was then correlated with pharmacological class.

D34. The aim was to identify pharmacological classes with a high proportion of positive class members. Ten of the 45 classes were 'positive' (contained > 50% compounds identified as rat carcinogens, see Table 9 from van der Laan et al., 2016, Annex 1), 17 classes were 'negative', and 18 classes had 'mixed' results. Not all compounds in each positive class were carcinogens, perhaps related to pharmacological, exposure, and replicability issues. Some compounds induced tumours considered to be unrelated to their pharmacology (eg induction of liver and/or thyroid tumours via induction of drug metabolising enzymes).

D35. Findings based on pharmacological class were discussed in the context of the NEG CARC prediction system for rat carcinogenicity, with particular reference to the 15 'false negatives' present in the database investigated (11 from the PhRMA database evaluated by Sistare et al. (2011), 1 from the CDER/FDA database, and 3 from the JPMA database). For several compounds the inclusion of pharmacological class effects would have designated a true positive instead of false negative result, indicating that this could be a valuable additional criterion in a weight of evidence evaluation in cases where histopathology is negative at 6 months. The NEG CARC category 'evidence for hormonal effects' was proposed as too broad and better replaced with 'oestrogenic/progestenic effects'. The spectrum associated with immunosuppressants was noted to be complex, with this class placed in the 'mixed' group (2 positive and 2 negative compounds in 2-year bioassay). The positives were an anti-TNFα compound that induced mammary gland tumours and systemic malignant lymphoma, and an immunosuppressant associated with granulocytic leukaemia in bone and interstitial cell tumours in testis. At least 1 of the 2-year bioassay negatives was positive in repeat dose toxicity studies (decreased thymus weight).

D36. Luijten et al. (2016) stated that the NEG CARC approach has also been tested in a retrospective analysis (Woutersen et al., 2016) of around 200 'environmentally relevant chemicals' using data from sub-chronic 90-day studies in rats, with findings in agreement with those of Sistare et al. (2011) that the absence of pre-neoplastic histological changes can accurately predict the lack of carcinogenicity of a non-genotoxic chemical.

Revision of ICH Guideline S1

D37. The International Council on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), in 1998, in its Guideline S1,

proposed that carcinogenicity testing of small molecule pharmaceuticals for regulatory purposes be based on a 2-year test in 1 (rather than, historically, 2) rodent species, supplemented with other data (a short- or medium-term *in vivo* rodent test or a second long-term carcinogenicity test in another rodent species) (ICH, 1998). Approaches using transgenic mice have subsequently been adopted (reviewed in G07 Part A), while the utility of other short-term study data is currently being evaluated.

D38. Ongoing revision of ICH S1 now aims to define situations where complete waiver of a 2-year bioassay would be justified (ICH, 2016a). A recent review of this process noted that the various available datasets that have been evaluated retrospectively have indicated that sufficient information should be available from pharmacology, genotoxicity and chronic toxicity data to conclude that a given pharmaceutical in certain cases presents a negligible risk or, conversely, a likely risk of human carcinogenicity without conducting a 2-year rat carcinogenicity study. Compounds could thus be listed in 1 of 3 main categories:

- Category 1 - highly likely to be tumourigenic in humans such that a product would be labelled accordingly and 2-year rat, 2-year mouse, or transgenic mouse carcinogenicity studies would not add value

- Category 2 - the available sets of pharmacologic and toxicologic data indicate that tumourigenic potential for humans is uncertain and rodent carcinogenicity studies are likely to add value to human risk assessment

- Category 3a - highly likely to be tumourigenic in rats but not in humans through prior established and well recognised mechanisms known to be human irrelevant, such that a 2-year rat study would not add value, or Category 3b - highly likely not to be tumourigenic in both rats and humans such that no 2-year rat study is needed. A 2-year or transgenic mouse study would be needed in most cases

D39. A set of weight of evidence criteria has been developed to assign candidate compounds to these categories, including: knowledge of intended drug target and pathway pharmacology, secondary pharmacology, and drug target distribution in rats and humans; genetic toxicology study results; histopathologic evaluation of repeated dose rat toxicology studies; exposure margins in chronic rat toxicology studies; metabolic profile; evidence of hormonal perturbation; immune suppression; special studies and endpoints (eg emerging technologies, new biomarkers..); results of non-rodent carcinogenicity study; transgenic mouse study.

D40. The ICH S1 revision process has been reviewed in the publication by Morton et al. (2014).

D41. The ICH is currently evaluating prospectively the reliability of this less-than-lifetime strategy through data generated by companies and will base their guidance on the outcome of this exercise (ICH, 2016a). Carcinogenicity assessment

documents submitted by sponsors based on the weight of evidence factors will be evaluated before completion of 2-year bioassays, allowing regulatory agencies to assess how well the weight of evidence predicts the 2-year rat carcinogenicity study results. The 'prospective evaluation period' for this work began in 2013 and is currently expected to have gathered sufficient data to assess the viability of the weight of evidence approach by the end of 2017, with the final study report expected to be submitted at the end of 2019 (ICH, 2016b). The goal is to evaluate carcinogenicity assessment documents plus 2-year data for 50 compounds, at least 20 of which are in Category 3.

4 Integrated Approaches For The Identification And Risk Assessment Of Human-Relevant Carcinogens

D42. The studies reviewed in Section 3 used data from rodent carcinogenicity bioassays as the comparator, ie the approach taken has generally been to evaluate the effectiveness of short-term tests to predict the results of carcinogenicity assays in rodents rather than directly addressing the likelihood of carcinogenicity in exposed humans. In addition, many strategies focus on carcinogen hazard identification and may support labelling requirements, but do not address the potential spectrum of risk over a range of exposure levels. They are therefore less well suited to the assessment of levels of carcinogenic risk posed by chemical exposures at ambient levels present in the human environment.

D43. A key issue that is re-iterated by many commentators is the need to move to a strategy based on the identification of human-relevant carcinogens (Meek et al., 2003). For the incorporation of short-term tests into such a strategy, it is necessary to establish which short-term data are required to achieve this. This should be informed by consideration of the key events and modes of action of carcinogenicity (see COC discussion paper CC/2016/08). Rodent-specific modes of action would be excluded from the strategy as the aim is to identify and evaluate human-relevant carcinogens. A combined *in vitro* and *in vivo* approach may be developed, with an initial evaluation for *in vitro* signals that might indicate carcinogenic potential (eg genotoxicity tests, high-throughput screening) and subsequent confirmation of the relevance or otherwise in short-term *in vivo* tests. Some generic key events (eg cell proliferation, immunosuppression) may be evaluated as short-term endpoints *in vivo*. Toxicogenomic techniques (eg transcriptomics) may also be applied to the evaluation of additional endpoints/biomarkers incorporated into sub-chronic toxicity studies, which may be able to highlight carcinogen class-specific signatures (*reviewed in* Doktorova et al., 2012). Toxicogenomic and high-throughput screening approaches are addressed in G07, part c and are not discussed in detail here.

4.1 Tiered and weight of evidence-based strategies to predict human carcinogenicity that incorporate parameters measured in sub-chronic toxicity studies

D44. Strategies and paradigms have been proposed that incorporate findings from short-term *in vivo* endpoints into human carcinogenicity risk assessments based on tiered and/or weight of evidence approaches.

D45. Cohen (2004, 2010a,b) has argued that the 2-year rodent bioassay is no longer necessary or appropriate for the evaluation of possible carcinogenic risk of chemicals to humans and that its use should be discontinued. An alternative model is presented that is based on shorter term tests, with an emphasis on mode of action and interpretation of the relevance to humans of findings in rodents. The premise is that increased carcinogenic risk occurs via: 1. increased net rate of DNA damage per cell division, occurring in pluripotential cell populations, and 2. increased number of DNA replications – ie increased cell proliferation (either by direct mitogenesis involving hormones or growth factors, or by cytotoxicity and regenerative proliferation) or decreased cell loss (by inhibition of apoptosis or cell differentiation). The model is represented as a tiered approach, incorporating a short-term screen for genotoxicity, immunosuppressive and oestrogenic activity using *in vitro* and *in vivo* tests, and the conduct of a 13-week assay using multiple doses to evaluate endpoints indicating toxicity/cell proliferation.

D46. The key events in this testing schedule involve precursor changes that can be identified in 13-week studies in rats and mice. The screening proposed has 2 phases: a general screen for any potential activity in any target tissue, and then a more detailed evaluation of the specific tissues identified as potential positives. The aim of this second stage is a careful mechanistic evaluation to identify the basis of the positive result, to determine whether the mode of action is relevant to humans and to define dose–response curves. This stage may eventually incorporate -omics methods.

D47. In this approach, the potential for DNA reactivity would be based primarily on Ames assay and structure activity relationships. The author proposed that other genotoxicity assays, such as micronuclei formation, chromosomal aberrations, clastogenicity, DNA repair, or effects on the mitotic apparatus would not be recommended for this screening process as they are highly influenced by cytotoxicity. *In vivo* tests may be carried outif doubts remained after *in vitro* tests. Positives could then be evaluated for dose–response for DNA reactivity and for cell proliferative effects (which may occur at higher doses), to aid in extrapolation of the assessment to humans.

D48. *Oestrogenic activity* would be detected by *in vitro* assays and/or histologic assessment of typical oestrogen-affected tissues (eg breast, endometrium, cervix).

D49. *Immunosuppression* could be assessed by *in vitro* assays and/or in 13-week studies by histopathologic evaluation of immunologic (eg thymus, lymph nodes, spleen) tissues.

D50. *Toxicity and/or increased cell proliferation* could be demonstrated on the basis of histopathological examination, and possibly also using screens for DNA synthesis such as BrdU, PCNA or Ki-67 labelling index assays. Clinical chemistry and organ weight data from 13-week studies may be helpful.

D51. The question of how to evaluate the different tissues is noted to be a subject of debate. For example, examination of rodent tissues that do not have human counterparts (eg forestomach, Zymbal's gland, Harderian gland) may be of uncertain relevance, and species-specific tumours in rodents that have no analogue in humans (eg splenic mononuclear cell leukaemia in rats, mouse submucosal mesenchymal lesion of the urinary bladder) may have little predictive value for human tumourigenicity. The evaluation of rodent endocrine tissues for carcinogenic activity is proposed to be of limited predictive value for human cancer risk, except for the evaluation of oestrogenic activity, due to differences in kinetics, metabolism and dynamics of these tissues and feedback mechanisms between humans and rodents. Many of these tumours occur at high rates spontaneously in rodents, such as the thyroid, pituitary, and testicular Leydig cell tumours in rats. Rodents are resistant to some tumours at sites that are common in humans, such as glandular stomach, colon, prostate and pancreas. Conversely, liver, kidney, lower urinary tract and, to some extent, lung, tumours show some correlation between humans and rodents. The correlation is strongest for carcinogenesis induced by DNA-reactive compounds. Various modes of action have been identified, some of which are considered to be relevant and some irrelevant to humans.

D52. Luijten et al. (2016) also proposed a tiered test strategy for cancer hazard identification, incorporating existing knowledge, genotoxicity data and data from sub-chronic rat studies. This would include:

- Tier 1. Review of existing data (physico-chemical, toxicokinetic/dynamic, intended use, (quantitative) structure activity relationships
- Tier II. Genotoxicity tests *in vitro*
- Tier III. Genotoxicity tests *in vivo*
- Tier IV. Carcinogenicity

A weight of evidence approach focussing on sub-chronic, repeat-dose toxicity data: histopathology (pre-neoplastic, proliferative or toxic lesions), organ weights, blood and urine chemistries and immunohistochemistry (eg Ki-67 as cell proliferation marker), plus pharmacological mode of action in the case pharmaceuticals. The authors noted that this strategy was developed to allow rapid implementation and does not fully address existing needs for mode of action information.

4.2 Proposal for an IATA for non-genotoxic carcinogens (OECD)

D53. Jacobs et al. (2016) (for the OECD) proposed the development of an IATA (Integrated Approach to Testing and Assessment)[3] to evaluate the carcinogenic potential of chemicals that are negative in genotoxicity screens, noting that the potential for carcinogenicity via non-genotoxic mechanisms is often not evaluated, due to the testing approach recommended under many regulatory frameworks.

D54. The selection of elements in an IATA can be based on an adverse outcome pathway concept incorporating biological changes, or key events, at the cellular, tissue, organ and organism levels that occur in response to molecular initiating events and leading to an adverse outcome. The relationships between molecular initiating events, key events and adverse outcomes are described in key event relationships. An IATA can also be developed empirically, containing elements other than those informed by the adverse outcome pathway, such as intended use and exposure, toxicokinetics and toxicodynamics.

D55. The proposed IATA comprises a structured information level framework with 5 levels of test information:

- **Level 0** incorporates available literature and *in silico* mode of action review information.

- **Level 1** (sub-cellular) and **Level 2** (cellular) tests evaluate mode of action groups *in vitro*, looking for molecular initiating events and early/initial key events. It is noted that widely accepted Level 1/2 test methods currently exist only for endocrine molecular initiating events (eg oestrogen receptor binding and transactivation, steroidogenesis). The ToxCast programme is cited as potentially useful for Level 1/2 tests, and also toxicogenomic approaches using *in vitro* test systems that group chemicals according to specific modes of action. A wide range of modes of action should be tested. Quantitative information such as dose–response relationships and points of departure will be required in order to be able to predict whether 1 key event would trigger the next key event.

- **Level 3** (multicellular tissue/organ) aims to identify cytoskeletal, tissue and organ changes and angiogenesis. It includes *in vitro* tests such as cell transformation assays and 3D cell models, *ex vivo* organ studies, *in vivo* data such as histopathology from repeat dose studies, and ‘organ-on-a-chip’ technologies. Level 3 *in vivo* information may not be needed if Level 2 and

[3] The OECD working definition of an IATA is: ‘a structured approach used for hazard identification (potential), hazard characterisation (potency) and/or safety assessment (potential/potency and exposure) of a chemical or group of chemicals, which strategically integrates and weights all relevant data to inform regulatory decision regarding potential hazard and/or risk and/or the need for further targeted testing and therefore optimising and potentially reducing the number of tests that need to be conducted’ (OECD, 2015).

Level 3 *in vitro* data are sufficient to meet regulatory requirements based on molecular initiating event and weight of evidence information.

- **Level 4** (organism) includes transgenic rodent assays, 2-year rodent carcinogenicity bioassays and chronic toxicity studies. The aim is to minimise the necessity for Level 4 data, in line with the principles of 'Toxicity Testing in the 21st Century' (described in G07, part c). This information may be required to gain insight into adverse effect levels, dose–response curves and tumour types/species affected.

D56. Assays or diagnostic tools may overlap 2 levels (an example given is cell transformation assays, which may belong to Levels 2 and 3). Quantitative and qualitative adverse outcome pathway/mode of action elements are required, based on the steps of the carcinogenic process. All modes of action should be tested for (as blocks of tests), and negative results for 1 block should not exclude all other modes of action. It should also be noted that mechanisms are not always related to adversity, with early molecular initiating events/key events not always leading to downstream adverse outcomes. IATA-based decisions may be made when several interconnected mechanisms are affected adversely (for example, all of the 3 hallmarks – oxidative stress, cell death, immune system evasion). Level 1, 2 and 3 assays require validation such that definitive decisions including the derivation of acceptable exposure levels can be made.

5 Summary

D57. For several decades, the standard method used to evaluate the carcinogenic potential of chemicals has been a genotoxicity test battery plus extrapolation from the results of high-dose 2-year rodent bioassay tests to low-dose exposures in humans. Key drawbacks of this approach are the high number of false positive results obtained and the question of relevance to human cancer risk, due to issues of species specificity, mode of action, and dose.

D58. Alternative strategies to the 2-year rodent bioassay are being developed that incorporate short-term data into carcinogenicity evaluations, based on tiered approaches and/or weight of evidence evaluations. Some of these approaches are likely to be feasible in the short term whilst others are more exploratory and it is not yet clear whether they will be feasible for risk assessment purposes. They vary depending on the type of compound being evaluated and the purpose of the evaluation. For use in risk assessment of chemicals present in the environment, new systems for carcinogenicity evaluation would ideally have the potential to produce organ-specific, dose-dependent information relevant to humans.

D59. Retrospective evaluations of existing databases have shown some utility of short-term *in vivo* test data to predict the outcomes of 2-year rodent bioassays, but with the development of further short-term endpoints necessary. A negative-predictive approach (the absence of short-term histopathologic risk factors in multiple tissues) has shown utility for screening out non-carcinogens, particularly in the

evaluation of pharmaceuticals. A strategy for evaluation of pharmaceuticals using a weight of evidence approach is being tested prospectively by the ICH to define situations where a waiver of the requirement for a 2-year rodent carcinogenicity bioassay can be granted.

D60. New tiered/integrated strategies are being developed using a mode of action-based approach incorporating modes of action that are of relevance to humans but not those that are rodent specific. The question of which key events/modes of action should be evaluated is a developing area. A combined *in vitro*/*in vivo* approach may be developed, looking for any *in vitro* signals that might indicate carcinogenic potential (eg in high-throughput screening) and then confirming relevance or otherwise in short-term *in vivo* tests.

D61. An IATA for the evaluation of carcinogenic risks posed to humans by non-genotoxic chemicals is in development (OECD). The goal is for a strategy without animal testing, based on tests for key events and key events relationships, as this knowledge base expands.

6 COC conclusions on alternative testing strategies for carcinogens incorporating results from short-term tests

D62. Use of the 2-year rodent bioassay to evaluate the carcinogenicity of the vast numbers of untested chemicals that are currently marketed is not feasible and alternative methods are required for this purpose. Genotoxicity tests can detect many, but not all, genotoxic carcinogens, and cannot detect non-genotoxic carcinogens. Genotoxicity tests should be carried out according to COM guidelines.

D63. The development of alternative approaches for the identification and characterisation of chemical carcinogens is a rapidly evolving field. Currently available data do not give a clear indication of the direction of progress in replacement. Some of the approaches that have been used have conceptual problems and there are currently no methods that are generally accepted in replacement of animal carcinogenicity studies.

D64. In developing alternative strategies, it is important to determine the aspects of risk assessment which the strategy can address. Carcinogenicity studies are performed for a various aspects of risk assessment. The requirement for a lifetime rodent bioassay may depend on the regulatory and legislative setting.

D65. One approach that is being developed as an alternative to performing a 2-year bioassay is to use experimental data from shorter-term tests, which may be incorporated into evaluations based on tiered approaches or weight of evidence evaluations. Retrospective studies have indicated some utility of short-term (eg 3- or 6-month) *in vivo* test data to predict the outcomes of 2-year rodent carcinogenicity bioassays but further short-term endpoints are required.

D66. For public health protection relating to chemicals present at ambient levels in the environment, the principal goal is the identification and risk assessment of

carcinogenicity in humans, not rodents. There is a limit to how far animal tests can be refined to predict human cancer risk. The emphasis should now be moved away from the development of further rodent studies. Future approaches should take into account human-relevant modes of action and alternative strategies should focus on predicting potential human carcinogenicity rather than rodent carcinogenicity.

D67. While short-term studies *in vivo* may be used as part of the weight of evidence to provide an indication that a chemical is potentially carcinogenic, at the current time they do not provide a basis for estimation of tumour risk.

D68. In some cases the positive results observed in short-term tests *in vivo* have little biological plausibility. The signals indicating potential carcinogenicity are sometimes identified in different tissues to those in which tumours are identified in longer term studies. In addition, the signals in the short-term assays may be hypertrophy or hyperplasia, which are not of themselves pre-neoplastic characteristics. This question of biological plausibility adds weight to the view that these short-term assays cannot be used as a replacement for the 2-year bioassay for the identification of rodent carcinogens.

D69. Negative predictive approaches, which incorporate negative outcomes for genotoxicity and short-term histopathologic risk factors in multiple tissues, are of interest but have associated problems. Human metabolism may not be suitably accounted for in the genotoxicity tests included. These approaches often have very high false-positives rates.

D70. It is increasingly recognised that rodent carcinogenicity studies which use the maximum tolerated dose (MTD) of a test chemical, may not be relevant for the identification of human-relevant hazard or, therefore, useful for risk assessment. This is particularly so when tumours are only detected in rodents at the highest dose level tested, because such dose levels (>200mg/kg/day) can overwhelm biological homeostasis, by saturating metabolic detoxification and excretory processes. Moreover, maximally tolerated dose levels are frequently associated with stress responses, such as thymic involution - an indicator of immunosuppression. Immunosuppression in both humans and animals increases the risk of neoplasia.

D71. For these reasons, carcinogenicity studies of potential new medicines in conventional rats and mice, normally set the high dose level using a 25-fold multiple of the maximum human exposure level, rather than an MTD. However, some regulatory authorities have requested that carcinogenicity studies using transgenic mice should use the MTD level. This is because some human-relevant carcinogens (IARC Class1) are only positive in such transgenic mouse studies, when tested at or close to, the MTD. However, for biologically consistency, the concerns about the human relevance of carcinogenicity data from conventional mice using the MTD, must apply equally to studies using transgenic mice.

D72. Differences in approach between testing of pharmaceuticals compared to other chemicals are noted, with pharmaceuticals generally being tested in animals at

lower doses than other chemicals. Pharmaceuticals are also in themselves associated with a pharmacological effect in humans, whereas other chemicals either have only a technical purpose in the media they are in (eg food additives), or are tested to ensure they do not show adverse effects in non-target species (eg pesticides). It is important to recognise that different approaches may be required and that alternatives may not address the requirements of all the different sectors. Nevertheless, it is important to maintain the collaborative approach across the sectors.

D73. Overall, it is important for alternative means of assessing health risks from chemicals to be developed and for good lines of communication and sharing of data between different sectors to continue. The challenge for such alternative strategies is to avoid missing crucial adverse effects while not over-predicting issues of concern.

COC
March 2017

Abbreviations

BMD	Benchmark dose
EPA	Environmental Protection Agency
FDA	Food and Drug Administration
HTS	High-throughput screening
IARC	International Agency for Research on Cancer
IATA	integrated approach to testing and assessment
ICH	International Council on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
JPMA	Japanese Pharmaceutical Manufacturers Association
NEG CARC	Negative for Endocrine, Genotoxicity, and Chronic Study Associated Histopathological Risk Factors for Carcinogenicity
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
PhRMA	Pharmaceutical Research and Manufacturers of America
PoD	Point of departure
Tox21	Toxicology in the 21st Century
ToxCast	Toxicity Forecaster
TT21C	Toxicity Testing in the 21st Century

References

Adeleye Y, Andersen M, Clewell R et al. (2015). Implementing Toxicity Testing in the 21st Century (TT21C): Making safety decisions using toxicity pathways, and progress in a prototype risk assessment. Toxicology, 332, 102-111.

Ahlborn, G.J., Nelson, G.M., Grindstaff, R.D., Waalkes, M.P., Diwan, B.A., Allen, J.W., Kitchin, K.T., Preston, R.J., Hernandez-Zavala, A., Adair, B., et al. (2009). Impact of life stage and duration of exposure on arsenic-induced proliferative lesions and neoplasia in C3H mice. Toxicology, 262, 106-113.

Alden CL, Lynn A, Bourdeau A et al. (2011). A critical review of the effectiveness of rodent pharmaceutical carcinogenesis testing in predicting for human risk. Veterinary Pathology, 48, 772-284.

Alden, C., Smith, P., Morton, D. (2002). Application of genetically altered models as replacement for the lifetime mouse bioassay in pharmaceutical development. Toxicologic Pathology 30, 135-138.

Alden, C.L., Smith, P.F., Piper, C.E., Brej, L. (1996). A critical appraisal of the value of the mouse cancer bioassay in safety assessment. Toxicologic Pathology 24, 722-725.

Allen DG, Pearse G, Haseman JK and Maronpot RR (2004). Prediction of rodent carcinogenesis: an evaluation of prechronic liver lesions as forecasters of liver tumors in NTP carcinogenicity studies. Toxicologic Pathology, 32, 393-401.

Ament Z, Waterman, CL, West JA et al (2013). A metabolomics investigation of non genotoxic carcinogenicity in the rat. Journal of Proteome Research, 12, 5775-5790.

Auerbach SS (2016). *In vivo* signatures of genotoxic and non-genotoxic chemicals. In: Thomas RS, Waters MD (Eds), Toxicogenomics in Predictive Carcinogenicity. The Royal Society of Chemistry, pp 113-153. DOI:10.1039/9781782624059-00113.

Auerbach SS, Shah RR, Mav D et al (2010). Predicting the hepatocarcinogenic potential of alkenylbenzene flavoring agents using toxicogenomics and machine learning. Toxicology and Applied Pharmacology, 243, 300-314.

Benigni R. Evaluation of the toxicity forecasting capability of EPA's ToxCast Phase I data: can ToxCast in vitro assays predict carcinogenicity? Journal of Environmental Science and Health Part C Environmental Carcinogenesis & Ecotoxicology Reviews. 2013;31(3):201-12

Benigni R. Predicting the carcinogenicity of chemicals with alternative approaches: recent advances. Expert Opinion and Drug Metabolism & Toxicology. 2014 Sep;10(9):1199-208.

Bhattacharya S, Zhang Q, Carmichael PL et al. (2011). Toxicity Testing in the 21st Century: Defining New risk Assessment Approaches Based on Perturbation of Intracellular Toxicity Pathways. PLoS One, 6, e20887.

Billington, R., Lewis, R.W., Mehta, J.M., Dewhurst, I. (2010). The mouse carcinogenicity study is no longer a scientifically justifiable core data requirement for the safety assessment of pesticides. Critical Reviews in Toxicology, 40, 35-49.

Boobis AR, Cohen SM, Dellarco V et al. (2006). IPCS framework for analyzing the relevance of a cancer mode of action for humans. Critical Reviews in Toxicology, 36, 781-792.

Boobis, A.R., Cohen, S.M., Doerrer, N.G., Galloway, S.M., Haley, P.J., Hard, G.C., Hess, F.G., Macdonald, J.S., Thibault, S., Wolf, D.C., et al. (2009). A data-based assessment of alternative strategies for identification of potential human cancer hazards. Toxicologic Pathology 37, 714-732.

Bugelski PJ, Volk A, Walker MR et al. (2010). Critical review of preclinical approaches to evaluate the potential of immunosuppressive drugs to influence human neoplasia. International Journal of Toxicology, 29, 435-466.

COC (1999) COC Consideration of neonatal rodent bioassay COC/1999/C1. Available here: http://webarchive.nationalarchives.gov.uk/20131103234051/http://www.iacoc.org.uk/statements/neonatalrodentbioassaycoc99c1may1998.htm

COC (2002) COC statement on ILSI/HESI research programme on alternative cancer models COC/2002/S3. Available here: http://webarchive.nationalarchives.gov.uk/20140506122041/http://www.iacoc.org.uk/statements/ILSIHESIresearchprogrammeonalternativecancermodelscoc02s3april2002.htm

Cohen SM (2004). Human carcinogenic risk evaluation: an alternative approach to the two-year rodent bioassay. Toxicological Sciences, 80, 225-229.

Cohen, S.M. (2010a). An enhanced 13-week bioassay: an alternative to the 2-year bioassay to screen for human carcinogenesis. Experimental and Toxicologic Pathology 62, 497-502.

Cohen, S.M. (2010b). Evaluation of possible carcinogenic risk to humans based on liver tumors in rodent assays: the two-year bioassay is no longer necessary. Toxicologic Pathology 38, 487-501.

Cohen (2018). Screening for human urinary bladder carcinogens: two-year bioassay is unnecessary. Toxicology Research, 7, 565-575.

COM (2012) Cell Transformation Assays. COM/12/S4. Available here: http://webarchive.nationalarchives.gov.uk/20140506144845/http://www.iacom.org.uk/

statements/documents/COM12S4CellTransformationAssayStatementfinalforinternet_000.pdf

Cox LA, Popken DA, Kaplan AM et al. (2016). How well can *in vitro* data predict *in vivo* effects of chemicals? Rodent carcinogenicity as a case study. Regulatory Toxicology and Pharmacology, 77, 54-64.

Doktorova TY, Pauwels M, Vinken M et al. (2012). Opportunities for an alternative integrating testing strategy for carcinogen hazard identification. Critical Reviews in Toxicology, 42, 91-106.

Eastmond, D.A., Vulimiri, S.V., French, J.E., Sopnawane, B. (2013) The use of genetically modified mice in cancer risk assessment: challenges and limitations. Critical Reviews in Toxicology 43, 611-631.

Ellinger-Ziegelbauer H, Gmuender H, Bandenburg A and Ahr HJ (2008). Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. Mutation Research, 637, 23-39.

Fielden MR, Adai A, Dunn RT, 2nd et al (2011). Development and evaluation of a genomic signature for the prediction and mechanistic assessment of nongenotoxic hepatocarcinogens in the rat. Toxicological Sciences. 124, 54-74.

French, J., Storer R.D., Donehower, L.A. (2001) The nature of the heterozygous Trp53 knockout model for identification of mutagenic carcinogens. Toxicologic Pathology 229, Suppl, 24-9.

Fukushima, S., Hagiwara, A., Hirose, M., Yamaguchi, S., Tiwawech, D., Ito, N. (1991). Modifying effects of various chemicals on preneoplastic and neoplastic lesion development in a wide-spectrum organ carcinogenesis model using F344 rats. Japanese journal of cancer research: Gann 82, 642-649.

Guyton KZ, Kyle AD, Aubrecht J et al. (2009). Improving prediction of chemical carcinogenicity by considering multiple mechanisms and applying toxicogenomic approaches. Mutation Research. 681, 230-240.

Hanahan D and Weinberg RA (2000). The hallmarks of cancer. Cell, 100, 57-70.

Hanahan D and Weinberg RA (2011). Hallmarks of cancer: The next generation. Cell, 144, 646-674.

ICH (1998) S1B: Note for guidance on carcinogenicity: testing for carcinogenicity of pharmaceuticals (CPMP/ICH/299/95). Available here: www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002735.pdf

ICH (2016a). ICH Guideline S1. Regulatory notice on changes to core guideline on rodent carcinogenicity testing of pharmaceuticals. 29 February 2016. Accessed at:

www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2012/12/WC500136405.pdf, 07/06/16.

ICH (2016b). The ICHS1 Regulatory Testing Paradigm of Carcinogenicity in rats – Status Report. 2 March 2016. Accessed at: www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S1/Final_Status_Report_ICH_S1_Jacksonville_2March2016.pdf, 07/06/16.

Ito, N., Hasegawa, R., Imaida, K., Hirose, M., Shirai, T. (1996). Medium-term liver and multi-organ carcinogenesis bioassays for carcinogens and chemopreventive agents. Experimental and Toxicologic Pathology 48, 113-119.

Ito, N., Tamano, S., Shirai, T. (2003). A medium-term rat liver bioassay for rapid in vivo detection of carcinogenic potential of chemicals. Cancer Science 94, 3-8.

Jacobs A (2005). Predication of 2-year carcinogenicity study results for pharmaceutical products: how are we doing? Toxicological Sciences, 88, 18-23.

Jacobs, A.C., Brown, P.C. (2015) Regulatory Forum Opinion Piece*: Transgenic/Alternative Carcinogenicity Assays: A Retrospective Review of Studies Submitted to CDER/FDA 1997-2014. Toxicologic Pathology, 43, 605-10.

Jacobs MN, Colacci A, Louekari K et al. (2016). International regulatory needs for development of an IATA for non-genotoxic carcinogenic chemical substances. ALTEX – Alternatives to animal experimentation, 33, 359-392.

Jacobson-Kram, D., Sistare, F.D., Jacobs, A.C. (2004). Use of transgenic mice in carcinogenicity hazard assessment. Toxicologic Pathology 32, Suppl 1, 49-52.

Julien E, Boobis AR, Olin SS (2009). The Key Events Dose-Response Framework: a cross-disciplinary mode-of-action based approach to examining dose-response and thresholds. Critical Reviews in Food Science and Nutrition, 49, 718-728.

Kienhuis AS, Mennes WC, Driessen M et al (2016). The parallelogram approach to assess human relevance of toxicogenomics-derived toxicity pathways in human health risk assessment. In: Thomas RS, Waters MD (Eds), Toxicogenomics in Predictive Carcinogenicity. The Royal Society of Chemistry, pp 423-441. DOI:10.1039/9781782624059-00423.

Kleinstreuer NC, Dix DJ, Houck KA et al. (2013). *In Vitro* Perturbations of Targets in Cancer Hallmark Processes Predict Rodent Chemical Carcinogenesis. Toxicological Sciences, 131, 40-55.

Koufaris C, Wright J, Currie RA, Gooderham NJ (2012). Hepatic microRNA profiles offer predictive and mechanistic insights after exposure to genotoxic and epigenetic hepatocarcinogens. Toxicological Sciences, 128, 532-543.

Long, GG (2010). Commentary on “Evaluation of possible carcinogenic risk to humans based on liver tumors in rodent assays: the two-year bioassay I no longer necessary.” Toxicologic Pathology, 38, 502-505.

Long, G.G., Morton, D., Peters, T., Short, B., Skydsgaard, M. (2010). Alternative mouse models for carcinogenicity assessment: industry use and issues with pathology interpretation. Toxicologic Pathology 38, 43-50.

Luechtefeld T & Hartung T (2017) Computational Approaches to Chemical Hazard Assessment. ALTEX – Alternatives to animal experimentation, 34, 459–478.

Luijten M, Olthof ED, Hakkert B et al (2016). An integrative test strategy for cancer hazard identification. Critical Reviews in Toxicology, 46, 615-639.

Lynch, D., Svoboda, J., Putta, S., Hofland, H.E., Chern, W.H., Hansen, L.A. (2007) Mouse skin models for carcinogenic hazard identification: utilities and challenges. Toxicologic Pathology, 35, 853-64.

Meek ME, Bucher JR, Cohen SM et al. (2003). A framework for human relevance analysis of information on carcinogenic modes of action. Critical Reviews in Toxicology, 33, 591-653.

Melis JP, Derks KW, Pronk TE et al (2014). In vivo murine hepatic microRNA and mRNA expression signatures predicting the (non-)genotoxic carcinogenic potential of chemicals. Archives of Toxicology, 88, 1023-1034.

Morton, D., Alden, C.L., Roth, A.J., Usui, T. (2002) The Tg rasH2 mouse in cancer hazard identification. Toxicologic Pathology, 30, 139-146.

Morton D, Sistare FD, Nambiar PR et al. (2014). Regulatory Forum commentary: alternative mouse models for future cancer risk assessment. Toxicologic Pathology, 42, 799-806.

Nambiar, P.R., Morton, D. (2013) The rasH2 mouse model for assessing carcinogenic potential of pharmaceuticals. Toxicologic Pathology, 41, 1058-67.

Nambiar, P.R., Turnquist, S.E., Morton, D. (2012) Spontaneous tumor incidence in rasH2 mice: review of internal data and published literature. Toxicologic Pathology, 40, 614-623.

NRC (2007). Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press. doi:10.17226/11970.

OECD (2015). Report of the workshop on a framework for the development and use of integrated approaches to testing and assessment. Series on Testing and Assessment No. 215. Accessed at: www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2015)22&doclanguage=en (07/06/16).

Okamura, M., Kashida, Y., Watanabe, T., Yasuhara, K., Onodera, H., Masao, H., Usui, T., Tamaoki, N., Mitsumori, K. (2003) Lack of susceptibility of heterozygous p53-knockout CBA and CIEA mice to phenolphthalein in a 6-month carcinogenicity study. Toxicology, 185, 17-22.

Osimitz, T.G., Droege, W., Boobis, A.R., Lake, B.G. (2013) Evaluation of the utility of the lifetime mouse bioassay in the identification of cancer hazards for humans. Food and Chemical Toxicology, 60, 550-562.

Paules R (2003). Phenotypic anchoring: linking cause and effect. Environmental Health Perspectives, 111, A338-339.

Pereira DA and Williams JA (2007). Origin and evoluation of high throughput screening. British Journal of Pharmacology, 152, 53-61.

Pritchard, J.B., French, J.E., Davis, B.J., Haseman, J.K. (2003) The role of transgenic mouse models in carcinogen identification. Environmental Health Perspectives, 111, 444-454.

Reddy MV, Sistare FD, Christensen JS et al. (2010). An evaluation of chronic 6- and 12-month rat toxicology studies as predictors of 2-year tumor outcome. Veterinary Pathology, 47, 614-629.

Römer M, Eichner J, Metzger U et al (2014). Cross-platform toxicogenomics for the prediction of non-genotoxic hepatocarcinogenesis in rat. PLoS ONE, 9, e97640.

Schaap MM, Wackers PFK, Zwart EP et al (2015). A novel toxicogenomics-based approach to categorize (non-) genotoxic carcinogens. Archives of Toxicology, 89, 2413-2427.

Schaap MM, van Benthem J, Jacobs MN et al (2016). Dissecting modes of action of non-genotoxic carcinogens. In: Thomas RS, Waters MD (Eds), Toxicogenomics in Predictive Carcinogenicity. The Royal Society of Chemistry, pp 209-235. DOI:10.1039/9781782624059-00209.

Schwetz B and Gaylor D (1998). Alternative tests: carcinogenesis as an example. Environmental Health Perspectives, 106, 467-471.

Sistare FD, Morton D, Alden C et al. (2011). An analysis of pharmaceutical experience with decades of rat carcinogenicity testing: support for a proposal to modify current regulatory guidelines. Toxicologic Pathology, 39, 716-744.

Sobels FH (1977). Some problems associated with the testing for environmental mutagens and a perspective for studies in ‘‘comparative mutagenesis’’. Mutation Research, 46, 245-260.

Solt, D., Farber, E. (1976) New principles for the analysis of chemical carcinogenesis Nature, 263, 701-703.

Storer, R.D., French, J.E., Haseman, J., Hajian, G., LeGrand, E.K., Long, G.G., Mixson, L.A., Ochoa, R. Sagartz, J.E., Soper, K.A. (2001) p53+/- hemizygous knockout mouse: overview of available data. Toxicologic Pathology, 29, 30-50.

Storer, R.D., Sistare, F.D., Reddy, M.V., DeGeorge, J.J. (2010) An industry perspective on the utility of short-term carcinogenicity testing in transgenic mice in pharmaceutical development. Toxicologic Pathology, 38, 51-61.

Sutter TR (1995). Molecular and cellular approaches to extrapolation for risk assessment. Environmental Health Perspectives, 103, 386-389.

Tennant, R.W., Stasiewicz, S., Eastin, W.C., Mennear, J.H., Spalding J.W. (2001). The Tg.AC (v-Ha-ras) transgenic mouse: nature of the model. Toxicologic Pathology, 29, 51-59.

Thomas RS, Bao W, Chu TM et al (2009). Use of short-term transcriptional profiles to assess the long-term cancer-related safety of environmental and industrial chemicals. Toxicological Sciences, 112, 311-321.

Thomas RS, Wesselkamper SC, Wang NC et al (2013). Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. Toxicological Sciences, 134, 180-194.

Thomas RS and Waters MD (2016). Transcriptomic dose-response analysis for mode of action and risk assessment. In: Thomas RS, Waters MD (Eds), Toxicogenomics in Predictive Carcinogenicity. The Royal Society of Chemistry,pp 154-184. DOI:10.1039/9781782624059-00154.

Tokar, E.J., Diwan, B.A., Ward, J.M., Delker, D.A., Waalkes, M.P. (2011). Carcinogenic effects of "whole-life" exposure to inorganic arsenic in CD1 mice. Toxicological Sciences, 119, 73-83.

Tsuda, H., Futakuchi, M., Fukamachi, K., Shirai, T., Imaida, K., Fukushima, S., Tatematsu, M., Furukawa, F., Tamano, S., Ito, N. (2010) A medium-term, rapid rat bioassay model for the detection of carcinogenic potential of chemicals. Toxicologic Pathology, 38, 182-187.

U.S. EPA (2013). A retrospective analysis of the immunotoxicity study (OCSPP test guideline No. 870.7800). February, 2013. Accessed at: www.epa.gov/pesticide-registration/retrospective-analysis-immunotoxicity-study (07/06/2016)

Uehara T, Minowa Y, Morikawa Y et al (2011). Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database. Toxicology and Applied Pharmacology, 255, 297-306.

Usui, T., Mutai, M., Hisada, S., Takoaka, M., Soper, K.A., McCullough, B., Alden, C. (2001). CB6F1-rasH2 mouse: overview of available data. Toxicologic Pathology, 29, Suppl, 90-108.

van der Laan JW, Kasper P, Silva Lima B et al. (2016). Critical analysis of carcinogenicity study outcomes. Relationship with pharmacological properties. Critical Reviews in Toxicology, 46, 587-614.

Waalkes, M.P., Liu, J., Diwan, B.A. (2007). Transplacental arsenic carcinogenesis in mice. Toxicology and Applied Pharmacology, 222, 271-280.

Waalkes, M.P., Liu, J., Ward, J.M., Diwan, B.A. (2006a). Enhanced urinary bladder and liver carcinogenesis in male CD1 mice exposed to transplacental inorganic arsenic and postnatal diethylstilbestrol or tamoxifen. Toxicology and Applied Pharmacology, 215, 295-305.

Waalkes, M.P., Liu, J., Ward, J.M., Powell, D.A., Diwan, B.A. (2006b). Urogenital carcinogenesis in female CD1 mice induced by in utero arsenic exposure is exacerbated by postnatal diethylstilbestrol treatment. Cancer Research, 66, 1337-1345.

Ward SL & Daston G (2014). 'Omics', bioinformatics, computational biology. http://alttox.org/mapp/emerging-technologies/omics-bioinformatics- computational-biology/ accessed 08/09/16.

Watanabe T, Suzuki T, Natsume M et al (2012). Discrimination of genotoxic and non-genotoxic hepatocarcinogens by statistical analysis based on gene expression profiling in the mouse liver as determined by quantitative real-time PCR. Mutation Research, 747, 164-175.

Waterman CL, Currie RA, Cottrell LA et al (2010). An integrated functional genomic study of acute phenobarbital exposure in the rat. BMC Genomics, 11, 9.

Waters MD (2016). Introduction to Predictive Toxicogenomics for Carcinogenicity. In: Thomas RS, Waters MD (Eds), Toxicogenomics in Predictive Carcinogenicity. The Royal Society of Chemistry, pp 1-38. DOI:10.1039/9781782624059.

Waters MD, Jackson M, Lea I (2010). Characterizing and predicting carcinogenicity and mode of action using conventional and toxicogenomics methods. Mutation Research, 705, 184-200.

Woutersen FA, Soffers AEMF, Kroese ED, Krul CAM, van der Laan JW, van Benthem J, Luijten M (2016). Prediction of carcinogenic potential of chemicals using repeated-dose (13-week) toxicity data. Regulatory Toxicology and Pharmacology, 81, 242-249.

Yamada F, Sumida K, Uehara T et al (2012). Toxicogenomics discrimination of potential hepatocarcinogenicity of non-genotoxic compounds in rat liver. Journal of Applied Toxicology, 33, 1284-1293.

Yamanaka H, Yakabe Y, Saito K et al (2007). Quantitative proteomic analysis of rat liver for carcinogenicity prediction in a 28-day repeated dose study. Proteomics, 7,

781-795.