

1
2 **Strategies in genotoxicology: Acceptance of innovative scientific methods in a**
3 **regulatory context and from an industrial perspective.**
4

5 **Guy Steiblen^{1*}, Jan van Benthem², Georges Johnson³**

6 1. Solvay, Genas, France guy.steiblen@solvay.com + (33) 6 07 99 28 60

7 *to whom correspondence should be addressed

8 2. RIVM-National Institute for Public Health and the Environment, Bilthoven, The
9 Netherlands jan.van.benthem@rivm.nl, + (31) 6 5248 6578

10 3. Swansea University Medical School, Swansea University, UK
11 g.johnson@swansea.ac.uk, + (44) 1792 295158
12
13
14
15

16 **Abstract**
17

18 The tests used and the general principles behind test strategies are now often over 30
19 years old. It may be time by now, given that our knowledge of genetic toxicology has
20 improved and that we also technically are better able to investigate DNA damage
21 making use of modern molecular biological techniques, to start thinking on a new test
22 strategy. In the present paper, it is demonstrated that the time is there to consider a new
23 approach for genotoxicity assessment of substances. A fit for all test strategy was
24 discussed making use of the most recent technological methods and techniques.
25

26 It was also indicated that *in silico* tools should be more accepted by regulatory
27 institutes/bodies as supporting information to better conclude which tests should be
28 required for each separate substance to demonstrate its genotoxic potency. Next to that
29 there should be a good rationale for performing *in vivo* studies. Finally, the need for
30 germ cell genotoxicity testing, essential when classification and labeling of substances is
31 mandatory, was discussed. It was suggested to change the GHS for genotoxicity
32 classification and labelling from *in vivo* tests in germ cells into *in vivo* tests in somatic
33 cells.
34

35 Quantitative genotoxicology was also discussed. It appeared that we are currently at a
36 transition, where the science developing to justify carrying out human health risk
37 assessments based on genetic toxicology data sets supported by mechanistic data and
38 exposure data. However, implementation will take time, and acceptance will be
39 supported through the development of numerous case studies. Major remaining
40 questions are: is genetic damage a relevant endpoint in itself, or should the risk
41 assessment be carried out on the apical endpoint of cancer and which genotoxic
42 endpoint should be used to derive the point of departure (PoD) for the human exposure
43 limit?
44

45 **Keywords**

46 Strategy in genotoxicology; Adverse Outcome Pathway (AOP); Germ cells and Global
47 Harmonized System (GHS); Quantitative genotoxicity.

48
49

50 **Introduction**

51

52 This paper is a report from a workshop “Strategies in genotoxicology: Acceptance of
53 innovative scientific methods in a regulatory context: Strengths and Weaknesses. From
54 a regulatory point of view to an industrial perspective” at the 47th Annual Meeting of the
55 European Environmental Mutagenesis and Genomics Society, held in Rennes, France in
56 May 2019. Registrations of products always require the assessment of their genotoxic
57 potential in order to protect humans and the environment. During this workshop
58 strategies in genotoxicity were observed from a regulatory, an industrial and a
59 quantitative perspective. The views expressed in these presentations are always
60 exclusively those of the speakers.

61
62

63 **Discussion**

64

65 All new and existing chemicals, drugs, food ingredients, cosmetic ingredients, etc. are
66 tested before marketing to be sure that they do not pose a risk for public health. These
67 substances are tested on every toxicological endpoint, including genotoxicity. Testing for
68 genotoxicity has a dual-purpose: i) hazard assessment and ii) classification and labelling
69 according to the Global Harmonisation System (GHS).

70 Often, for instance for chemicals in Europe, genotoxicity has a special place in this
71 assessment since for genotoxic substances that are also carcinogenic a non-
72 thresholded risk assessment is mandatory.

73

74 Genotoxicity testing has some main principles. The maximum level of information should
75 be obtained with a minimum number of tests. The tests therefore have to be very
76 adequate and efficient. Next to that the extent of genotoxicity testing should be related to
77 the extent of (potential) human exposure: the more human exposure the more testing.
78 And finally the number or use of animals should be reduced. This means that *in vitro*
79 tests are strongly preferred.

80

81 Genotoxicity testing has three genotoxic endpoints gene mutations (mutagenicity),
82 structural chromosome aberrations (clastogenicity) and numerical chromosome
83 aberrations (aneuploidy). For a complete assessment of genotoxicity of substances all
84 endpoints have to be covered. Unfortunately, there is no genotoxicity test to cover all
85 endpoints, indicating that a combination of tests is necessary to cover genotoxicity. The
86 result is a strategy for genotoxicity testing.

87 For testing, a number of both *in vitro* and *in vivo* tests are available. Since 1982,
88 recommendations for the conduct of the commonly used tests are provided in the
89 Organization for Economic Co-operation and Development (OECD) test guidelines
90 describing the major steps of the tests. However, many tests are older and were already
91 developed (long) before the OECD started its test guidelines program. Since 1982 not
92 much has changed. A few tests were deleted in 2014 e.g. the Sister Chromatid

93 Exchange (SCE) test, the *Saccharomyces* and *Drosophila* tests as well as the *in vitro*
94 Unscheduled DNA Synthesis (UDS) test. The decision to delete some test guidelines
95 was made based on the observation that these tests are rarely used in the various
96 legislative jurisdictions, and on the availability of newer tests showing a better
97 performance for the same endpoint. The deleted test guidelines should not be used for
98 new testing and are no longer a part of the set of OECD recommended tests; however,
99 data previously generated from these deleted test guidelines can still be used in a risk
100 assessment or other types of regulatory decisions.

101 Only a few new tests were introduced: the *in vitro* mammalian micronucleus test (OECD
102 487 [1]), the transgenic gene mutation test (OECD 488 [2]) and the *in vivo* comet assay
103 (OECD 489 [3]). The former test guideline (OECD 476) was divided in two test
104 guidelines: one covering hprt/xprt mutations (revised as and a new guideline covering tk
105 mutations in recommended TK6 cells and L5171Y cells (OECD 490 [4]). All other tests,
106 were several times updated or revised. However, it may be clear that the possibilities to
107 keep updating or revising these tests are limited. There comes a point that a revision
108 becomes nit-picking.

109
110 At least within Europe, the genotoxicity testing strategy between regulatory bodies
111 (ECHA, EFSA, SCCS, EMA, COM) is rather similar. The strategy consists of a stepwise
112 approach starting with *in vitro* tests. In the event of negative *in vitro* results, it can be
113 concluded that the substance has no genotoxic potential. Positive *in vitro* results trigger
114 appropriate *in vivo* studies to assess whether the genotoxic potential observed *in vitro* is
115 expressed *in vivo*. A negative result in these confirming *in vivo* tests overrules the *in vitro*
116 positive result and the substance is considered not genotoxic. Positive *in vivo* results
117 make the substance tested a genotoxicant. For substances which are genotoxic in
118 somatic cells *in vivo*, the potential for germ cell mutagenicity should be considered. It is
119 recognised that standard reproduction studies do not cover all germ cell effects. Next to
120 that genotoxicity testing is more or less a pre-screen for carcinogenicity and as such
121 many legislations ask for a carcinogenicity test when a substance is a mutagen.

122 Still, consideration should be given to whether specific features of the test substance
123 might require substitution of one or more of the recommended *in vitro* tests by other *in*
124 *vitro* or *in vivo* tests in the basic battery. The *in vivo* test to be performed is dependent
125 on the genotoxic endpoint that is positive *in vitro*, e.g. a positive *in vitro* micronucleus
126 test should be followed by an *in vivo* micronucleus test.

127
128 As an example, under the REACH regulation, the required *in vitro* assays for
129 genotoxicology are always the same. The Ames test (OECD 471) [5] is required for a
130 substance production or importation from 1 to 10 tons/year; the upper band tonnage
131 between 10-100 tons/year requires two additional tests (e.g. the *in vitro* micronucleus
132 test (OECD 487) [1] and *in vitro* mammalian cell gene mutation tests using the thymidine
133 kinase gene (OECD 490) [4]. If all these three assays are found to be negative, i.e. do
134 not generate any genotoxic response, then the substance is considered to be non-
135 genotoxic and no additional *in vitro* or *in vivo* assays are then required. Under REACH,
136 the use of animals for genotoxicity assessment is possible but needs to be justified.

137 Scientifically, this testing battery proposed by REACH regulation sounds acceptable
138 according to literature for assessing substance genotoxicity. In a review of Kirkland et al.
139 (2011) [6], the authors demonstrated that a battery of only two *in vitro* tests, the Ames

140 test (OECD 471) [5] and the *in vitro* micronucleus assay (OECD 487) [1], is enough as
141 both are covering all three endpoints of genotoxicology, i.e. gene mutations,
142 clastogenicity and aneuploidy. Nearly 100% (958 out of 962) of rodent carcinogens or *in*
143 *vivo* genotoxins were correctly detected with these two tests, which make this battery a
144 particular high sensitive combination.

145 Consequently, it may be justified to wonder whether *in vitro* tests alone are sufficient for
146 both single substance or mixture genotoxic potency assessment. Such philosophy would
147 be also in line with United States Environmental Protection Agency's (US-EPA) intention
148 which targets to ban animal use in toxicology studies by 2035 [7].

149 The tiered approach required by most regulations could then focus on *in vitro* tests and
150 on (Q)SAR in order to avoid unnecessary use of animals.

151
152

153 Table 1: predictivity of 3 common genotoxicity tests: Gene mutation test in
154 bacteria (Ames test); MLA: mouse lymphoma assay (gene mutation test in
155 mammalian cells); CA: chromosome aberration test. (modified or calculated from
156 Kirkland et al., 2005 [8])

	Ames	MLA	CA	Ames + CA	Ames + CA + MLA
Sensitivity (%)	58.8	73.1	65.6	82.2	84.7
Specificity (%)	73.9	39.0	44.9	33.1	22.9
Concordance (%)	62.5	62.9	59.8	71.1	64.8

157
158
159 However, the sensitivity (percentage of correctly predicted genotoxic substances) as
160 well as previously reported the specificity (percentage of correctly predicted non-
161 genotoxic substances) of the current test is rather poor. In table 1 the specificity and
162 sensitivity of the most commonly used *in vitro* genotoxicity tests are shown. It is clear
163 that particularly the specificity of some tests is much too low for tests that are used for
164 regulatory decisions [8], [9]). As the possibilities to update these tests are limited as was
165 discussed before, the chances to improve the specificity and sensitivity are limited as
166 well. That is why such a proposed standalone *in vitro* battery of tests can still be
167 challenged regarding its lack of specificity.

168 Indeed, there is a strong ethical pressure and policy against the use of experimental
169 animals. For example, there is already legislation for the testing of cosmetic ingredients
170 where the use of *in vivo* tests is prohibited. That leads to carcinogenicity assessment
171 being dependent on *in vitro* genotoxicity tests only, for which we just concluded that they
172 are not optimal.

173
174 Since 2005, several initiatives have been undertaken to improve the specificity of the *in*
175 *vitro* tests without compromising sensitivity, e.g. preference for p53-competent human
176 cells, cytotoxicity measures based on cell proliferation, carefully controlling the upper
177 limit of cytotoxicity or the highest non-toxic concentration, all of which have been
178 incorporated into the OECD guideline revisions [4].

179

180 When *in vivo* studies are required, registrants are often sensitive to the 3R principles.
181 This has contributed to an increase in the assessment of multiple genotoxic endpoints
182 from the same animals . As an example, it is common that data from assays which
183 combine a subacute toxicology study in the rodent, with micronucleus assessment in the
184 blood, as well as comet assay in the liver [10] (figure 1) are presented in submitted
185 dossiers. Combined studies are even recommended by regulations like the ICH S2 (R1)
186 (step 4 version, 2011) [11] for assessing the potential genotoxicity of a substance under
187 pharmaceutical regulation.

188 There is also be a good rationale to perform *in vivo* assays in rats rather than in mice for
189 the *in vivo* micronucleus test. The logic behind is that there is usually more kinetic data
190 available on rats compared to mice (e.g. absorption, distribution, metabolism, and
191 excretion (ADME) studies). Toxicokinetic (TK) information is required in most regulations
192 and is performed mainly on rats compared to mice, so when TK information is already
193 available from previous TK studies, it is not required to repeat the generation of TK data
194 during an *in vivo* micronucleus assay.

195 In addition, if the absorption has to be assessed during a micronucleus test to prove
196 animal's exposure, rats should be also prioritized, if possible, since it is possible to use
197 one animal for assessing the two different endpoints, i.e. bioanalysis and micronucleus
198 assessment. In mice, assessment of both parameters would require double the animals
199 because there is not enough blood available per animal to assess both endpoints.

200
201
202
203
204



205
206 Figure 1: Combine multi-assay in *in vivo* rodent

207
208
209 It is generally accepted that standard genotoxicity tests are rather old. The test strategy
210 may use some refreshment. Most regulatory guidelines contain a list of tests which have
211 to be ticked off, and they do not cover new endpoints or new groups of chemicals. It may
212 also be time to think about a new approach for genotoxicity testing. A requisite for a new
213 strategy is that the three genotoxic endpoints, gene mutations, structural chromosome
214 aberrations and numerical chromosome aberrations are still covered. Moreover, there
215 should be room for classic as well as new tests, but this will required a paradigm shift in
216 genotoxicity thinking. A new harmonised and more advanced approach, should include
217 options to use mechanism-based assays, and additionally, both toxicokinetics and
218 exposure have to be included. Of course also the Adverse Outcome Pathway (AOP)

219 science will be a substantial part of it. However, the most dramatic change in thinking,
220 will be that the test substance and not the list of tests (the strategy) should be the main
221 consideration. On a substance by substance basis the real strategy for each specific
222 substance has to be determined. The idea is that there is a general framework of a
223 strategy that will be fine-tuned for each single substance, and chemical group.
224

225 One of the major changes will be that, based on the strong ethical pressure and policy
226 against the use of experimental animals, the framework predominantly or even
227 exclusively consists of *in vitro* genotoxicity tests. Fortunately, there is already experience
228 with what can be expected when *in vivo* tests are not allowed. Since the marketing ban
229 of March 11, 2013 [12], in Europe *in vivo* testing is prohibited for cosmetic products or
230 ingredients, irrespective of the availability of alternative non-animal tests. This major
231 change, lead to evaluations on genotoxicity of cosmetic ingredients being done
232 exclusively *in silico*, i.e. QSAR and read across, or on *in vitro* tests, i.e. the gene mutation
233 test in bacteria and the *in vitro* micronucleus test. A conclusion when one of the two
234 tests was positive (considered genotoxic) or negative (considered non-genotoxic) may
235 be clear but difficulties were expected when one or the 2 tests was positive. In the end
236 the problems were excellently tackled by collaboration between all stakeholders,
237 including industry and regulatory institutes. It resulted in a revised Notes of Guidance
238 [13] in which it is stated that the test should be critically evaluated for e.g. false positive
239 results. The latter may, for example, be the result of the quality of a study, excessive
240 cytotoxicity, positive results inside the historical control data or by the presence of
241 impurities. For specific tests like the gene mutation test in bacteria, the reason for a false
242 positive result may be bacterial toxicity or specific metabolism in bacteria.
243

244 If after checking for false negatives the positive result remains, then further testing may
245 be considered. Further tests can be classic genotoxicity tests like the mammalian
246 chromosomal aberration test or the gene mutation test in mammalian cells, but also
247 tests for which an OECD test guideline is not available like the 3D-skin micronucleus
248 test, the 3D-skin comet assay, γ H2AX assay, the hen's egg test (Het) micronucleus test,
249 indicator approaches like Multiflow or Toxtracker and even techniques which are very
250 new like array approaches, next generation sequencing or recombinant cell models. A
251 prerequisite is that a justification for the test is available and that the final conclusion is
252 made on weight of evidence. Until now, excessive problems in deciding whether a
253 substance is genotoxic or not did not occur. Industry screens their new substances also
254 for genotoxicity and uses sensitive new techniques including systems biology and high
255 throughput tests. Why is it not possible to use new approaches for decision making as
256 well, rather than sticking with the rather old fashioned tests low throughput and low
257 content assays as described before!
258

259 A common comment on why these new approaches are not suitable, is that they are not
260 validated and that an OECD test guideline for these tests is not available. Of course this
261 is true, and the lack of protocol standardisation comes with its own issues. On the other
262 hand, adequate scientific justification of why a specific test is used along with a clear
263 study design and test protocol, should be sufficient to overrule the comment that an
264 OECD test guideline is not available. Moreover, the tests mentioned are already known

265 in the genotoxic community as well as their characteristics. Concerning validation, that is
266 a more difficult story. The present validation project is a long, time consuming and
267 expensive course. Today, there is much discussion regarding the simplification of the
268 validation process [14, 15]. Hopefully this will lead in the near future to a shorter process
269 without losing the same level of quality.

270
271 It may be time to think about a new strategy for genotoxicity testing [16], or at least
272 introducing increased flexibility for inclusion of new approaches. When developing a new
273 strategy, it would be worthwhile considering an overall strategy for genotoxicity testing
274 that fits for all substances: chemicals, drugs, cosmetic ingredients, food etc. It is clear
275 that all strategies are similar, which is expected since they are all designed for detecting
276 hazard. Determination of genotoxic hazard can be performed identically according to
277 one single strategy for all groups of substances. However, divergences in regulatory
278 guidelines become important when the risk of the hazard is considered: e.g. no
279 genotoxic substance in food or a risk/benefit approach for drugs/chemicals.

280
281 Such a new framework is already described by the “clean sheet” workgroup of the
282 Genetic Toxicology Technical Committee (GTTC) of the Health and Environmental
283 Sciences Institute (HESI) [16]. In this framework the substance is leading. Figure 2
284 clearly shows the different steps to be taken in such a new approach. It is clear that next
285 to the planning and scoping, also exposure, knowledge base, test selection and
286 appropriate Point of Departure (PoD) selection play a very important role. In the paper
287 by Dearfield et al., 2017 [16] all steps are clearly explained. In the end it should lead to
288 estimated acceptable levels for endpoint of human exposure and risk characterization.
289 This “Clean Sheet framework” is now tested with (industrial) case studies.

290
291
292

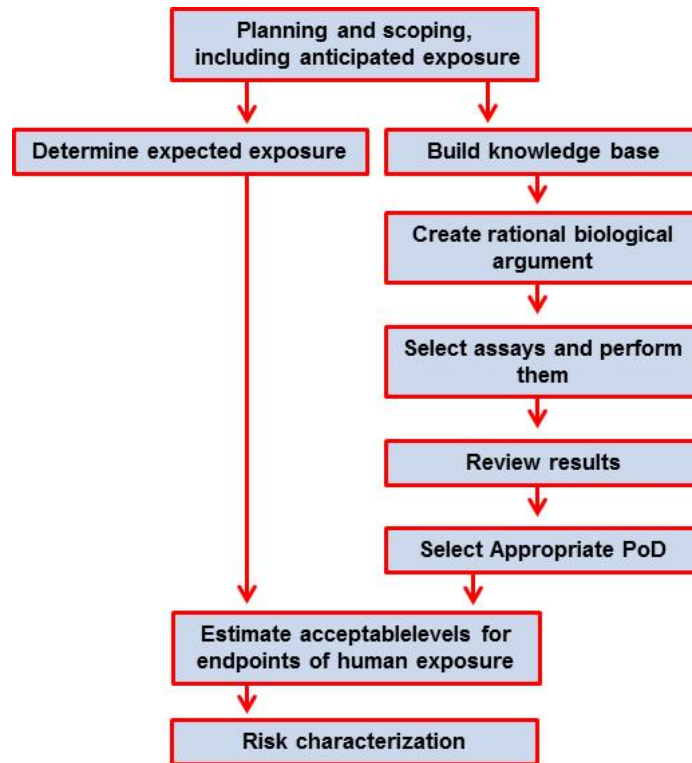
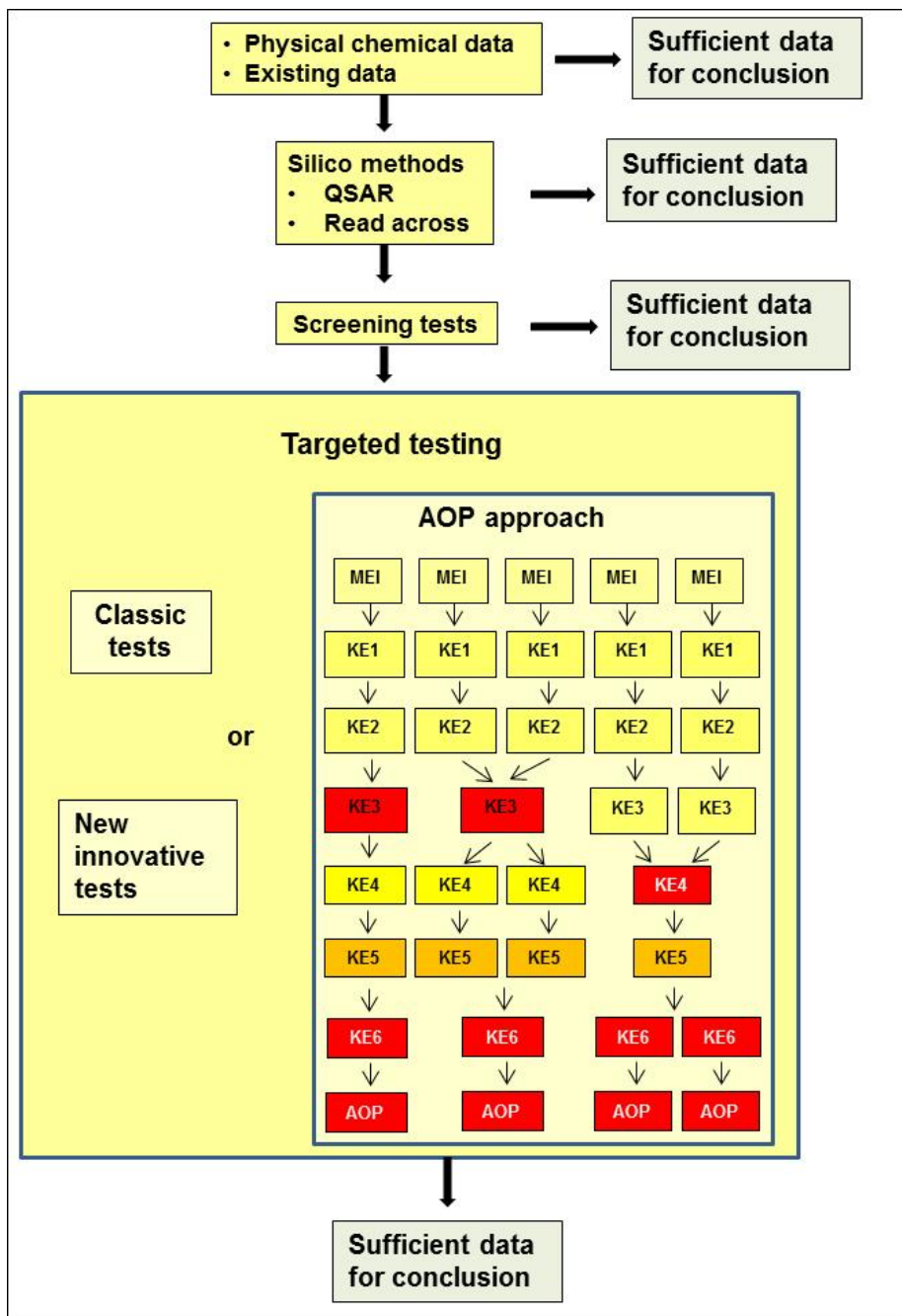


Figure 2: Strategy for examining genomic damage (from Dearfield et al., 2017 [16])

293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312

In figure 3, this theoretical framework is filled in in a more practical way. Still the substance is leading and it starts with a planning and scoping including which factors are important for a specific risk assessment of a specific substance. Relevant questions are, for example, how humans are exposed, are some population groups more sensitive for the substance, which legislation is applicable. The first step in a putative framework could deal with consideration of physical chemical data and/or existing data. Theoretically this may give enough information for a conclusion on genotoxicity of the substance, e.g. when there is no exposure, testing is redundant. Other follow up levels may be *in silico* methods, like quantitative structure activity relationship (QSAR) and read across as well as screening tests.



313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323

Figure 3: Putative strategy to determine the genotoxicity of substances. Following exposure with a substance, a Molecular Initiating Event (MEI) will be generated followed by Key Event(s) (KE) leading to the characterization of an AOP in a dedicated matrix (cell, protein-chips, etc.).

Most of EU regulations like REACH regulation EC [17], pesticide (PPP) [18], biocide (BPR) [19] and many others have a common philosophy in line with the above suggestion. The philosophy is to reduce the use of animals for experiments by using a

324 tiered approach, requesting at first alternative, assays like *in vitro* tests and/or *in silico*
325 tools. The cosmetic regulation (1223/2009) [12] is even more drastic, because animal
326 use is completely forbidden for experimental use.

327 For screening purposes, these tendencies or requirements have enhanced an
328 adaptation of industrials from the 2000's in terms of assay development or alternative
329 method developments. As an example *in silico* tools have been well developed and
330 applied in genotoxicology for many years, giving a wide range of (Q)SAR tools with a
331 powerful ability to predict genotoxic potency of a substance [20]. It is nevertheless not
332 possible to replace completely *in vitro* or *in vivo* assays by these *in silico* tools. In fact,
333 (Q)SAR are currently considered for providing alerts on hazardous substances.
334 However, they are also tools which can be used to replace some classic assays for
335 some regulations like REACH particularly for well studied groups of compounds, and this
336 option is possible for many endpoints required in REACH, i.e. from ecotoxicology to
337 toxicology assays (Annex XI of REACH regulation) [17]. Currently, it looks like
338 nevertheless that these *in silico* tools are essentially accepted as "supporting studies" by
339 regulators, so the "key study", i.e. the main assay (*in vitro* or *in vivo*) is still essential for a
340 substance submission under REACH. Thus far, the authors are unaware whether
341 genotoxicity data obtained by *in silico* tools only was allowed to address the requirement
342 for a substance submission under REACH.

343 Concerning screening tests, one may think of all kinds of high throughput tests that
344 already are commonly used for decision making purposes within the industrial setting.
345 Again the combined information from these different assays may be good enough to
346 conclude on genotoxicity. However, this conclusion can probably only be taken after
347 targeted testing and information on a putative mode of action becomes clear. For
348 targeted testing one should think of using an already available AOP in full in first
349 instance or a tiered number of classical or new innovative genotoxicity tests. Until now
350 six genotoxicity AOPs are available [21-23].

351
352 Targeted testing may be repeated as one or more modes of action may be applicable.
353 When *in silico* and *in vitro* analysis does not give a reliable answer, *in vivo* testing may
354 be considered, but as a last resort. In the end it should lead to a final conclusion on
355 genotoxicity for this specific substance according to this specific strategy.

356
357
358 Most of the available results in genetic toxicology are focused on bacterial or somatic
359 cells. Currently, the Global Harmonized System (GHS) which is used in most of
360 substance/mixture evaluation regulations (i.e. pesticide, biocide, drug...) [24] contains a
361 germ cell mutagen classification. Therefore, there seems to be a gap between what is
362 most available, i.e. data from bacterial and somatic cells and the requirements to
363 conclude on substance classification, i.e. germ cells.

364
365 Today, most EU regulations try to avoid the use of *in vivo* tests in genotoxicology as
366 much as possible, only some still require such performance independently of *in vitro*
367 results. For example, pesticide regulation [18] required the systematic performance of an
368 *in vivo* rodent assay. The consequence is that most results generated for a substance in
369 genotoxicity assays are mainly available in *in vitro* tests and more seldom in *in vivo* tests
370 that, are predominantly performed on somatic cells. The impact is not neutral,

371 particularly for the substance classification under the Global Harmonization System
372 (GHS) [25], because GHS requires germ cell genotoxicity data for classification (Table
373 2). With data available mainly on somatic cells, it is therefore difficult to conclude on a
374 substance classification, because nearly no information is available on germ cells [26].
375 In the GHS, the category 1A or 1B will lead to a market ban of substance or mixtures.
376

377 Another problem is that currently, regulations recognize only animal assays to have the
378 ability to assess the genotoxic potential in germ cells. Among the *in vivo* assays
379 available to assess germ cell potent genotoxicity are:

- 380 - The transgenic mice test (OECD 488) [2] which has the ability to detect gene
381 mutations in the mouse or rat.
- 382 - The rodent dominant lethal test (OECD 478) [27], the mammalian spermatogonial
383 chromosomal aberration test (OCDE 483) [28] and the mouse heritable
384 translocation assay (OCDE 485) [29] these last three allow to detect
385 chromosomal aberration endpoint.

386 These tests are very demanding, use a lot of animals and are therefore far from the 3R
387 principles. In addition, these tests are not performed by a lot of laboratories which make
388 their use very difficult for most of industrials.
389

390
391
392 Moreover, there is no example of a substance positive only in germ cells and not in
393 somatic cells in genotoxicity. This statement is clearly indicated in the guidance on
394 genotoxicity testing and data interpretation for pharmaceuticals intended for human use
395 (ICH S2 (R1): 2.4 Germ cells paragraph) [11]. Based on the present knowledge and the
396 presently available tests, which are insensitive and uses large numbers of animals, germ
397 cell genotoxicity testing may be considered redundant. However, in a recent review
398 Marchetti et al. (2019, in press) [30] report on indications that many substances may
399 also be a risk for human germ cells, justifying germ cell testing.

400 To focus on the germ cells is not necessarily to claim that germ cells and somatic cells
401 are different, it can be more driven by the fact that the kinetics of substances are not the
402 same between substances. Consequently a substance could be positive in somatic
403 cells, whilst it is negative in germ cells, for instance because it is not able to reach the
404 germ cells. Marchetti et al. [30] indicate that if mutation is considered a toxicological
405 endpoint and mutation will be used in human risk assessment, the PoD for germ cells
406 surely will be different from PoD's in somatic cells which makes germ cell mutagenicity
407 testing essential. Therefore, the idea to focus in somatic cells only for the GHS
408 classification purpose instead of germ cells is not always applicable.
409

410
411

412

413

414

415 Table 2: Hazard categories for germ cells mutagens extracted from ECHA guidance (v.5,
416 2017)

Annex I: 3.5.2.2. For the purpose of classification for germ cell mutagenicity, substances are allocated to one of two categories as shown in Table 3.5.1.

Table 3.5.1
Hazard categories for germ cell mutagens

Categories	Criteria
CATEGORY 1:	<p>Substances known to induce heritable mutations or to be regarded as if they induce heritable mutations in the germ cells of humans.</p> <p>Substances known to induce heritable mutations in the germ cells of humans.</p>
Category 1A:	<p>The classification in Category 1A is based on positive evidence from human epidemiological studies.</p> <p>Substances to be regarded as if they induce heritable mutations in the germ cells of humans.</p>
Category 1B:	<p>The classification in Category 1B is based on:</p> <ul style="list-style-type: none"> – positive result(s) from in vivo heritable germ cell mutagenicity tests in mammals; or – positive result(s) from in vivo somatic cell mutagenicity tests in mammals, in combination with some evidence that the substance has potential to cause mutations to germ cells. It is possible to derive this supporting evidence from mutagenicity/genotoxicity tests in germ cells in vivo, or by demonstrating the ability of the substance or its metabolite(s) to interact with the genetic material of germ cells; or – positive results from tests showing mutagenic effects in the germ cells of humans, without demonstration of transmission to progeny; for example, an increase in the frequency of aneuploidy in sperm cells of exposed people.
CATEGORY 2:	<p>Substances which cause concern for humans owing to the possibility that they may induce heritable mutations in the germ cells of humans.</p> <p>The classification in Category 2 is based on:</p> <ul style="list-style-type: none"> – Positive evidence obtained from experiments in mammals and/or in some cases from in vitro experiments, obtained from: <ul style="list-style-type: none"> – Somatic cell mutagenicity tests in vivo, in mammals; or – Other in vivo somatic cell genotoxicity tests which are supported by positive results from in vitro mutagenicity assays. <p>Note: Substances which are positive in in vitro mammalian mutagenicity assays, and which also show chemical structure activity relationship to known germ cell mutagens, shall be considered for classification as Category 2 mutagens.</p>

417
418
419
420
421
422

423
424
425
426 Until now the genotoxicity of chemicals has been evaluated in a yes/no fashion.
427 Genotoxicity is hazard assessment. But there is already a paradigm shift ongoing from
428 this qualitative approach into a quantitative approach. Dose response modelling to
429 establish a PoD was introduced in genotoxicity testing. Assessing the potency and PoD
430 in addition to the positive/negative effect, leads to the use of mutations for human health
431 risk assessment. In fact there are already examples where this approach was
432 successfully used [31-33]. In this approach gene mutation has to be considered as a
433 toxicological endpoint. For more detail on this consideration see the review by Heflich et
434 al. (2019, in-press) [34].

435
436 The idea of quantitative genetic toxicology is to advance beyond the standard hazard
437 assessment of substances, to a risk based approach. This changes the assessment
438 from whether substances are genotoxic or not genotoxic to calculations of exposure
439 levels that are of negligible concern to the exposed population. Advancements have
440 been made in defining PoD metrics from genetic toxicity data, and this has progressed
441 the consideration of these data for risk assessment purposes. The standard genetic
442 toxicity models of the *in vivo* micronucleus assay and the transgenic gene mutation tests
443 are suitable for dose response analysis and derivation of PoD. These *in vivo* assays can
444 be assessed using Benchmark Dose (BMD) analysis. The BMD approach uses
445 statistical models to define a small but measurable increased above the background,
446 and it is a more advanced statistical approach than pairwise testing which is used to
447 derive No Observed Effect Levels (NOEL).

448
449 Some recent developments with the BMD approach have led to increase precision, with
450 the major ones being the covariate BMD approach put forward by Slob and Setzer [35]
451 which uses conserved shape parameters to increase precision in the analyses. Another
452 improved part of BMD analysis comes from adjusting the Critical Effect Sizes (CES)
453 from the default of 5% or 10% to a higher part of the dose response. Slob [36] and Zeller
454 et al. [37], used extensive data sets to calculate that for genetic toxicity data sets, a CES
455 of 50% should be used. Both modifications to the BMD analysis, lead to decreased
456 range in BMD confidence interval and increased precision in the BMD.

457
458 There are numerous different approaches for calculating human exposure limits using
459 genetic toxicity PoD metrics. They are the same as those derived using PoD from other
460 toxicological endpoints, or even from the cancer bioassay. The calculations involve an
461 extrapolation to a human comparable dose, and then division by a number of
462 uncertainty/adjustment factors to cover the diversity of the human population, the study
463 duration, severity of endpoint, animal model being used, and which metric was being

464 used. This provides a number such as a Permissible Daily Exposure (PDE), Acceptable
465 Daily Intake (ADI) or other comparable approaches [38]. There is also the less complex
466 method, where the PoD is divided by the human exposure level, to calculate the Margin
467 of Exposure (MOE). If the MOE is over 10,000, then it is considered to be of negligible
468 concern, but a cause for concern if the MOE is below this value.

469
470 The HESI GTTC has been tasked with publishing case studies, where example
471 substances are assessed using these different approaches. In the most case examples,
472 PDEs were calculated and compared based on genetic toxicity data and cancer
473 bioassay data for certain alkylnitrosamines. These have recently been identified as
474 impurities in some pharmaceutical products, and the products were recalled. It is
475 therefore important to determine whether the exposed population has an increased risk
476 of mutation and cancer due to this exposure. The PDE derived from the cancer bioassay
477 data were excellent due to the high power and high-quality cancer bioassay study for the
478 alkylnitrosamines N-nitrosodiethylamine (NDEA) and N-nitrosodimethylamine (NDMA).
479 Data from the transgenic gene mutation studies were also good, but the studies included
480 short term dosing, so the adjustment factors were larger and potentially required further
481 adjustment. However, in this case, the PDEs were comparable and the mutation PDEs
482 supported the cancer bioassay derived PDEs. For N-Nitrosodibenzylamine (NMBA), no
483 suitable cancer bioassay data were available, however the gene mutation data were
484 relevant and usable. These case studies show that PDE can be derived, but there are
485 still ongoing discussions about when they would be suitable for use in a risk
486 assessment. The major considerations are that the dosing strategy and study design
487 should be suitable, and that multiple data sets are available to derive the PoD. Extensive
488 mechanistic understanding for a defined non-linear or threshold dose response are
489 required for the PDE or MOE type approach.

490
491 Benzene has also become an interesting case study. In 2018, ECHA carried out a risk
492 assessment for benzene, used the *in vivo* micronucleus data to define Occupational
493 Exposure Levels (OELs), and these were compared to those from OELs using human
494 exposure data. They were very comparable, and when the *in vivo* MN data were re-
495 analysed using the new best practice of covariate BMD and a CES of 50%, the data
496 were of great interest and provided a higher utility than those derived from the previous
497 analysis [39].

498
499 We are currently at a transition, where the science used to justify carrying out human
500 health risk assessments based on genetic toxicology data sets is suitable. However,
501 there needs to be a development of case studies in order to overcome the numerous
502 hurdles that are appearing. This will take time, and acceptance will be supported through
503 the development of numerous case studies in the different areas from pharmaceutical
504 impurities, food contaminants, industrial chemicals, cosmetics etc.

505
506 Resistance to using genetic toxicology data for human health risk assessments is
507 common within the field. The main arguments are;

- 508 1. Is genetic damage a relevant endpoint in itself, or should the risk assessment be
509 carried out on the apical endpoint of cancer?
- 510 2. Which endpoint should be used to derive the PoD for the human exposure limit?
- 511 3. Which study design, including acute, sub-chronic or chronic dosing, number of
512 days, replicate number, and more.
- 513 4. The adjustment factors are also a huge area. Issues include which factors to use
514 for the exposure time, 1-10 or even higher? Severity of endpoints, is mutation as
515 severe as cancer?
- 516 5. How much mechanistic data is required to support the use of a PDE through the
517 'threshold mechanism' type approach, outlined in the ICH M7 framework?

518
519
520

521 Future perspectives for quantitative genetic toxicology

522
523 Once the issues have been overcome, the result will be that genetic toxicity data can be
524 used for risk assessment purposes. *In vivo* data from relevant endpoints of mutation
525 and/or chromosome damage will be used to derive BMDL metrics from which Health
526 Based Guidance Values (HBGV) will be calculated. These will be compared with those
527 derived from other toxicological endpoints, and the most relevant and most conservative
528 metric will be used as the final value from which the risk assessment is carried out. This
529 is based on the recognition that mutation is a relevant endpoint and that the human
530 population should be protected from increased risk of mutation [34]. In addition to this,
531 mutation is linked with cancer, and the development of case studies will potentially show
532 that genetic toxicity based HBGV will be comparable to those from the cancer bioassay,
533 obviously ensuring exposure, tissue, study design, adjustment factors are suitable. Such
534 a major advancement will lead to a reduction in animal testing through less cancer
535 bioassay studies being required, and additionally through the increased use of combined
536 genetic toxicity assays.

537
538 Additional advances that could improve risk assessment based on genetic toxicity data,
539 would be the use of next generation sequencing in place of transgenic gene mutation
540 data. This would lead to mutations in cancer genes and other disease related genes
541 forming the assessment, which would increase relevance of the mutation endpoint for
542 use in risk assessment.

543
544

545

546

547 **Summary and perspectives**

548

549 The field of genetic toxicity testing is strongly in motion. There are many new
550 developments and approaches, including the development of high throughput and
551 screening tests. *In silico* approaches like QSAR and read across are generally accepted.
552 The idea for quantitative genotoxicity, in which genotoxicity is a toxicological endpoint,
553 gains more and more ground. Taking into account the quality of the current classical
554 tests as compared to the new developments, it may be clear that genotoxicity testing is
555 ready for an update.

556 There is no need to immediately go to a clean sheet approach but why not start with
557 considering the chemical as the lead and not the tests, and think of, next to the classical
558 strategy, an alternative new approach and framework when preparing a legislation
559 dossier. Hopefully this will trigger a new framework in the near future.

560

561 Databases using a significant number of substances, indicate that *in vitro* test batteries
562 using the Ames test as well as the micronucleus assay would have a sensitivity of nearly
563 100% making the use of animals test not necessary. The drawback is that there is a lack
564 of specificity of such tests that could make their use, as standalone, critical.

565 Currently, although many data are available on somatic cells, on substances or
566 mixtures, they will not allow to classify a substance according to the GHS which is
567 mainly based on germ cells results. As there is no unique germ cell mutagen, it would
568 therefore be more appropriate to change the GHS for the genotoxicity classification and
569 propose instead *in vivo* assays in somatic cells instead of germ cells.

570

571 We are currently at a transition, where the science is developing in order to justify
572 carrying out human health risk assessments based on genetic toxicology data sets,
573 which is supported by mechanistic data and exposure data. With this in mind, *in vitro*
574 genetic toxicity data should be used for potency ranking and mechanistic understanding
575 which can support the *in vivo* derived BMDL and HBGV. Future advances in this area, in
576 line with the efforts from EPA [40] will lead to *in vitro* to *in vivo* comparisons to allow
577 calculation of usable BMDL from *in vitro* studies. However, there are numerous issues
578 that need to be overcome in order to address the shortcomings of the current
579 approaches in this area. *In silico* analysis combined with potency ranking, chemical
580 grouping and chemical categorization also offer opportunities to enhance the application
581 of *in vitro* data.

582

583

584

585

586

587 **Bibliographical review**

588

589 [1] OECD, OECD Guideline for the Testing of Chemicals. In Vitro Mammalian Cell Micronucleus Test. 487,
590 In Vitro Micronucleus Test 2010.

591 [2] OECD, OECD Guidelines for Testing of Chemicals. Section 4: Health Effects: Test No.488: Transgenic
592 rodent somatic and germ cell mutation assays, 2011.

593 [3] OECD Guidelines for the Testing of Chemicals, Section 4: Test No. 489: In Vivo Mammalian Alkaline
594 Comet Assay, 2016.

595 [4] OECD Guidelines for Testing of Chemicals. Section 4: Health Effects: Test No.490: In vitro mammalian
596 cell gene mutation test using the Thymidine Kinase gene, 2015.

597 [5] OECD, OECD Guideline for the Testing of Chemicals. Bacterial Reverse Mutation Test. 471, in: OECD
598 (Ed.)1997.

599 [6] Kirkland D, Reeve L, Gatehouse D, V. P., A core in vitro genotoxicity battery comprising the Ames test
600 plus the in vitro micronucleus test is sufficient to detect rodent carcinogens and in vivo genotoxins. ,
601 Mutagenesis, 721 (2011) 27-73.

602 [7] United States Environmental Protection Agency: Directive to prioritize efforts to reduce animal
603 testing, 2019.

604 [8] D. Kirkland, M. Aardema, L. Henderson, L. Muller, Evaluation of the ability of a battery of three in
605 vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity
606 and relative predictivity (vol 584, pg 1, 2005), Mutat. Res. Genet. Toxicol. Environ. Mutagen., 588 (2005)
607 70-70.

608 [9] D. Kirkland, E. Zeiger, F. Madia, R. Corvi, Can in vitro mammalian cell genotoxicity test results be used
609 to complement positive results in the Ames test and help predict carcinogenic or in vivo genotoxic
610 activity? II. Construction and analysis of a consolidated database, Mutat. Res. Genet. Toxicol. Environ.
611 Mutagen., 775 (2014) 69-80.

612 [10] S. Khanal, P. Singh, S. Avlasevich, D.K. Torous, J. Bemis, D. SD, Integration of liver and blood
613 micronucleus and Pig-a gene mutation endpoints into rat 28-day repeat-treatment studies: Proof-of-
614 principle with diethylnitrosamine., Mutat. Res., (2018) 30-35.

615 [11] ICH, ICH Topic S2 (R1). Guidance on Genotoxicity Testing and Data Interpretation for
616 Pharmaceuticals Intended for Human Use, in: CPMP/ICH (Ed.)2008, pp. 1-13.

617 [12] EC, Regulation (EC) No 1223/2009 of the European parliament and of the Council of 30 November
618 2009 on cosmetic products (2009).

619 [13] Scientific Committee on Consumer Safety 2018. The SCCS notes of guidance for the testing of
620 cosmetic ingredients and their safety evaluation, 10th revision. , 2018.

621 [14] A.H. Piersma, T. Burgdorf, K. Louekari, B. Desprez, R. Taalman, R. Landsiedel, J. Barroso, V. Rogiers,
622 C. Eskes, M. Oelgeschlager, M. Whelan, A. Braeuning, A.M. Vinggaard, A. Kienhuis, J. van Benthem, J.
623 Ezendam, Workshop on acceleration of the validation and regulatory acceptance of alternative methods
624 and implementation of testing strategies, Toxicology in Vitro, 50 (2018) 62-74.

625 [15] T. Burgdorf, A.H. Piersma, R. Landsiedel, R. Clewell, N. Kleinstreuer, M. Oelgeschlager, B. Desprez, A.
626 Kienhuis, P. Bos, R. de Vries, L. de Wit, T. Seidle, J. Scheel, G. Schonfelder, J. van Benthem, A.M.

627 Vinggaard, C. Eskes, J. Ezendam, Workshop on the validation and regulatory acceptance of innovative 3R
628 approaches in regulatory toxicology - Evolution versus revolution, *Toxicology in Vitro*, 59 (2019) 1-11.
629 [16] K.L. Dearfield, B.B. Gollapudi, J.C. Bemis, R.D. Benz, G.R. Douglas, R.K. Elespuru, G.E. Johnson, D.J.
630 Kirkland, M.J. LeBaron, A.P. Li, F. Marchetti, L.H. Pottenger, E. Rorije, J.Y. Tanir, V. Thybaud, J. van
631 Benthem, C.L. Yauk, E. Zeiger, M. Luijten, Next Generation Testing Strategy for Assessment of Genomic
632 Damage: A Conceptual Framework and Considerations, *Environmental and Molecular Mutagenesis*, 58
633 (2017) 264-283.
634 [17] EC, No. 1907/29006 concerning the Registration, Evaluation, Authorisation and restriction of
635 Chemicals (REACH), (2006).
636 [18] EC, Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October
637 2009 concerning the placing of plant protection products on the market and repealing Council directives
638 79/117/EEC and 91/414/EEC, 2009.
639 [19] EC, Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012
640 concerning the making available on the market and use of biocidal products 2012.
641 [20] N.F. Cariello, J.D. Wilson, B.H. Britt, D.J. Wedd, B. Burlinson, V. Gombar, Comparison of the
642 computer programs DEREK and TOPKAT to predict bacterial mutagenicity, *Mutagenesis*, 17 (2002) 321-
643 329.
644 [21] C.L. Yauk, J. Bishop, K.L. Dearfield, G.R. Douglas, B.F. Hales, M. Luijten, J.M. O'Brien, B. Robaire, R.
645 Sram, J. van Benthem, M.G. Wade, P.A. White, F. Marchetti, The development of adverse outcome
646 pathways for mutagenic effects for the organization for economic co-operation and development,
647 *Environmental and Molecular Mutagenesis*, 54 (2013) 79-81.
648 [22] F. Marchetti, A. Massarotti, C.L. Yauk, F. Pacchierotti, A. Russo, The adverse outcome pathway (AOP)
649 for chemical binding to tubulin in oocytes leading to aneuploid offspring, *Environmental and Molecular*
650 *Mutagenesis*, 57 (2016) 87-113.
651 [23] A.A. Sasaki JC, Bryce SM, Custer L, Dearfield K, Dietz Y, Elhajouji A, Escobar P, Fornace Jr AJ, Froetschl
652 R, Galloway S, Hemmann U, Hendriks G, Li H-H, Luijten M, Ouedraogo G, Peel L, Pfuhlker S, Roberts D,
653 Thybaud V, van Benthem J, Yaul CL, Schuler M., Understanding genotoxic mode of action to inform risk
654 characterization: application of the adverse outcome pathway framework. In preparation.
655 [24] Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008
656 on classification and labelling and packaging of substances and mixtures, 2008.
657 [25] EC, Regulation (EC). No 1272/2008 of the European Parliament and of the Council of 16 December
658 2008 on classification and labelling and packaging of substances and mixtures., 2008.
659 [26] M. Woutersen, M. Beekman, M.E.J. Pronk, A. Muller, J.A. De Knecht, B.C. Hakkert, Does REACH
660 provide sufficient information to regulate mutagenic and carcinogenic substances? , *Human and*
661 *Ecological Risk Assessment*, 0 (2018) 1-20.
662 [27] OECD, OECD Guidelines for Testing of Chemicals. Section 4: Health Effects: Test No.478: Rodent
663 Dominant Lethal test, (1984).
664 [28] OECD Guidelines for Testing of Chemicals. Section 4: Health Effects: Test No.478: Rodent Dominant
665 Lethal test, 2016.
666 [29] OECD, OECD Guidelines for Testing of Chemicals. Section 4: Health Effects: Test No.485: Mouse
667 heritable translocation assay, (1986).
668 [30] D.G. Marchetti F, Yauk CL. 2019., A return to the origin of the EMGS: rejuvenating the quest for
669 human germ cell mutagens and determining the risk to future generations., *Environ Mol Mutagen*. 2019
670 Aug 31.
671 [31] G.E. Johnson, L.G. Soeteman-Hernandez, B.B. Gollapudi, O.G. Bodger, K.L. Dearfield, R.H. Heflich,
672 J.G. Hixon, D.P. Lovell, J.T. MacGregor, L.H. Pottenger, C.M. Thompson, L. Abraham, V. Thybaud, J.Y.
673 Tanir, E. Zeiger, J. van Benthem, P.A. White, Derivation of Point of Departure (PoD) Estimates in Genetic

674 Toxicology Studies and Their Potential Applications in Risk Assessment, *Environmental and Molecular*
675 *Mutagenesis*, 55 (2014) 609-623.

676 [32] P.A. White, G.E. Johnson, Genetic toxicology at the crossroads-from qualitative hazard evaluation to
677 quantitative risk assessment, *Mutagenesis*, 31 (2016) 233-237.

678 [33] J.W. Wills, G.E. Johnson, S.H. Doak, L.G. Soeteman-Hernandez, W. Slob, P.A. White, Empirical
679 analysis of BMD metrics in genetic toxicology part I: in vitro analyses to provide robust potency rankings
680 and support MOA determinations, *Mutagenesis*, 31 (2016) 255-263.

681 [34] J.G. Heflich RH, Zeller A, Marchetti F, Douglas GR, Witt KL, Gollapudi BB, White PA. , Mutation as a
682 toxicological endpoint for regulatory decision-making, *Environ Mol Mutagen*. 2019 Aug 312019.

683 [35] W. Slob, R.W. Setzer, Shape and steepness of toxicological dose-response relationships of
684 continuous endpoints, *Crit Rev Toxicol*, 44 (2014) 270-297.

685 [36] W. Slob, A general theory of effect size, and its consequences for defining the benchmark response
686 (BMR) for continuous endpoints, *Crit Rev Toxicol*, (2016) 1-10.

687 [37] A. Zeller, G. Duran-Pacheco, M. Guerard, An appraisal of critical effect sizes for the benchmark dose
688 approach to assess dose-response relationships in genetic toxicology, *Arch Toxicol*, 91 (2017) 3799-3807.

689 [38] J.P. Bercu, S.M. Galloway, P. Parris, A. Teasdale, M. Masuda-Herrera, K. Dobo, P. Heard, M. Kenyon,
690 J. Nicolette, E. Vock, W. Ku, J. Harvey, A. White, S. Glowienke, E.A. Martin, L. Custer, R.A. Jolly, V.
691 Thybaud, Potential impurities in drug substances: Compound-specific toxicology limits for 20 synthetic
692 reagents and by-products, and a class-specific toxicology limit for alkyl bromides, *Regul Toxicol*
693 *Pharmacol*, 94 (2018) 172-182.

694 [39] M. Luitjen, N.S. Ball, K.L. Dearfield, B.B. Gollapudi, G.E. Johnson, F. Madia, F. Peel, S. Pfuhler, R.S.
695 Settivari, W. ter Burg, J. van Benthem, Utility of a Next Generation Framework for Assessment of
696 Genomic Damage: A Case Study Using the Industrial Chemical Benzene, *Environ Mol Mutagen*, (2019).

697 [40] R.S. Thomas, M.A. Philbert, S.S. Auerbach, B.A. Wetmore, M.J. Devito, I. Cote, J.C. Rowlands, M.P.
698 Whelan, S.M. Hays, M.E. Andersen, M.E. Meek, L.W. Reiter, J.C. Lambert, H.J. Clewell, 3rd, M.L.
699 Stephens, Q.J. Zhao, S.C. Wesselkamper, L. Flowers, E.W. Carney, T.P. Pastoor, D.D. Petersen, C.L. Yauk,
700 A. Nong, Incorporating new technologies into toxicity testing and risk assessment: moving from 21st
701 century vision to a data-driven framework, *Toxicological sciences : an official journal of the Society of*
702 *Toxicology*, 136 (2013) 4-18.

703