The *in vivo* erythrocyte *Pig-a* gene mutation assay

Part 1: Detailed review paper and performance assessment

Prepared by:

Robert H. Heflich*\(^1\), Stephen D. Dertinger\(^2\), Vasily N. Dobrovolsky\(^1\), Javed A. Bhalli\(^3\),

Michelle O. Kenyon\(^4\), Takafumi Kimoto\(^5\), Daishiro Miura\(^5\), Katsuyoshi Horibata\(^6\),

James T. MacGregor\(^7\), Daniel J. Roberts\(^8\), Bernd-Wolfgang Igl\(^9\), David P. Lovell\(^10\),

Roberta A. Mittelstaedt\(^1\), Jennifer M. Shemansky\(^1\), Rebecca A. Wynne\(^1\), L. Patrice McDaniel\(^1\),

University of Maryland CERSI, and HESI Genetic Toxicology Technical Committee

for the Organisation for Economic Cooperation and Development Working Group of the
National Coordinators of the Test Guidelines Programme

\(^1\)U.S. Food and Drug Administration, Jefferson, AR USA

\(^2\)Litron Laboratories, Rochester, NY USA

\(^3\)Covance, Greenfield, IN USA
4Pfizer, Groton, CT USA
5Teijin Pharma, Tokyo, Japan
6National Institute of Health Sciences, Kanagawa, Japan
7Toxicology Consulting Services, Bonita Springs, FL USA
8Charles River Laboratories, Skokie, IL USA
9Boehringer Ingelheim Pharma, Biberach, Germany
10St. George’s, University of London, UK

*Direct correspondence to:

Robert H. Heflich, Ph.D.
Division of Genetic and Molecular Toxicology
U.S. Food and Drug Administration
National Center for Toxicological Research
3900 NCTR Rd., Jefferson, AR USA
Ph: +1-870-543-7493, Fax: +1-870-543-7393; email: robert.heflich@fda.hhs.gov

The information and conclusions in this document are the sole responsibility of the authors and do not necessarily represent an official position of the U.S. Food and Drug Administration, the Health and Environmental Sciences Institute, or any of the other institutions and businesses associated with the authors.
Contents
Executive Summary........................................................................................................................................... 5
Introduction to Part 1 ........................................................................................................................................ 12

1. Principle of the Pig-a assay .................................................................................................................... 14
   a. Genesis of the Pig-a assay ...................................................................................................................... 14
      i) Glycosylphosphatidylinositol anchors and Pig-a.............................................................................. 14
      ii) PIG-A and paroxysmal nocturnal hemoglobinuria ........................................................................ 15
      iii) Analysis of Pig-a mutant cells as a reporter of gene mutation ...................................................... 16
   b. Considerations in assay design ............................................................................................................. 18
   c. Initial observations on the source of the mutants detected in the erythrocyte Pig-a assay .......... 20

2. Existing in vivo genetic toxicology assays ............................................................................................ 22
   a. TG 474: the mammalian erythrocyte micronucleus test .................................................................... 23
   b. TG 475: the mammalian bone marrow chromosomal aberration test ............................................... 24
   c. TG 489: the in vivo mammalian alkaline comet assay ....................................................................... 25
   d. TG 488: the transgenic somatic and germ cell gene mutation assays .............................................. 27

3. Strengths and weaknesses of the Pig-a assay ....................................................................................... 28

4. Assay protocol ...................................................................................................................................... 30
   a. Factors affecting Pig-a assay performance: treatment and sampling protocol and number of cells interrogated ...................................................................................................................... 34
   b. Option of preserving peripheral blood samples for later analysis ..................................................... 36
   c. Updated recommendations for conducting the Pig-a gene mutation assay .................................... 37

5. Recommendations for reporting data, statistical analysis, and interpretation of results .................. 39

6. Confirming that Pig-a mutation is responsible for the Pig-a mutant phenotype .................................. 43
   a. Evidence that GPI-marker-deficient bone marrow erythroid cells and granulocytes in rats and mice contain mutations in the Pig-a gene........................................................................... 43
      i) Mouse studies ................................................................................................................................... 43
      ii) Rat studies ...................................................................................................................................... 44
   b. Rat bone marrow cells with GPI-anchored protein markers contain wild-type Pig-a genes .......... 46
   c. GPI-deficient cells in PNH patients are almost always associated with PIG-A mutations .......... 47
   d. Analysis of Pig-a mRNA in mutant RETs from mice ...................................................................... 48
   e. Evidence that GPI-deficient rat T lymphocytes contain Pig-a mutations ...................................... 49
   f. Evidence that Pig-a mutations cause GPI deficiency in mammalian cells in vitro ................. 51
g. Evidence that Pig-a gene mutation responses are consistent with TGR and endogenous reporter gene mutation responses in the same animals .................................................................53

h. Evidence that animals with genomic instability have elevated Pig-a mutant frequencies ........54

i. Lack of evidence that loss of GPI anchors and GPI-anchored proteins is due to something other than mutation, e.g., epigenetic gene silencing .................................................................54

7. Evaluation of Pig-a assay results ..................................................................................57

   a. Identification and cataloging Pig-a testing data .......................................................57

   b. Analysis of Pig-a test responses .............................................................................58

      i) Criteria for evaluating Pig-a test responses .........................................................58

      ii) Pig-a test responses in rats and mice .................................................................61

      iii) Discussion of test responses .............................................................................64

8. Performance assessment of the Pig-a assay ................................................................69

   a. Approach employed ...............................................................................................69

   b. Establishing responses in reference assays ...........................................................71

   c. Performance assessment results .............................................................................72

      i) Performance assessment using Pig-a responses evaluated using the WG Criteria .....73

      ii) Performance assessment using Pig-a responses evaluated using the Extended Criteria....75

      iii) Performance assessment using only Pig-a responses from assays conducted with short-term treatment protocols .................................................................77

9. Positioning of the assay relative to existing in vivo genetic toxicology assays .................80

10. Potential for expanding the applicability of the assay ..................................................82

    a. Erythrocyte PIG-A assays conducted in humans ....................................................82

    b. Erythrocyte Pig-a assays conducted in nonhuman primates ..................................84

    c. In vivo Pig-a/PIG-A assays conducted in germ cells and other tissues ..................84

       i) Germ cells ............................................................................................................84

       ii) Other cell types .................................................................................................85

    d. In vitro Pig-a assays ..............................................................................................87

11. Animal welfare issues associated with the Pig-a assay ...............................................89

12. Summary and conclusions .........................................................................................91

Acknowledgements ........................................................................................................93

References .......................................................................................................................95
Executive Summary

Glycosylphosphatidylinositol (GPI) anchors tether >150 unique proteins to the outer surface of mammalian cells. The phosphatidylinositol glycan class A gene (Pig-a) is involved in an early step in the biosynthesis of these anchors and inactivating mutations in Pig-a result in loss of GPI anchors as well as loss of the proteins that GPI tethers to the cell surface. Along with Pig-a, approximately 30 genes are involved in the biosynthesis of GPI anchors; however, all the other genes are present in two copies, while the Pig-a gene is located on the X chromosome and present as one functional copy per cell. Thus, loss of GPI anchors, and GPI-anchored proteins, are almost always due to mutation of the Pig-a gene. This relationship between Pig-a mutation and loss of anchors and their associated proteins is the basis for the Pig-a gene mutation assay (commonly referred to as ‘the Pig-a assay’). Because the GPI biosynthesis pathway is highly conserved, the assay has been performed in several mammalian species, including humans. In addition, the assay has been conducted in vitro, using established mammalian cell lines. However, it is most commonly performed as an in vivo assay for gene mutation in rodents. In the rodent assay, flow cytometry is used to analyze erythrocytes from small samples of peripheral blood for the presence or absence of GPI-anchored proteins.

The mutants measured in the erythrocyte Pig-a assay are formed in nucleated erythroid progenitor cells that, in adult rodents, are mainly located in the bone marrow. Thus, when used as a test for mutagenicity, the test substance or its metabolites must be able to reach the bone marrow for an adequate evaluation of in vivo hazard. This requirement is shared with commonly used in vivo tests for clastogenicity and aneugenicity in somatic cells, the erythrocyte micronucleus (MN) test (described in Organisation for Economic Cooperation and Development [OECD] Test Guideline [TG] 474) and the bone marrow chromosomal aberration test (TG 475). Other commonly used in vivo somatic cell genotoxicity tests, the transgenic rodent (TGR) gene mutation assay (TG 488) and the in vivo comet assay (TG 489), detect in vivo genotoxicity in virtually all animal tissues. The comet assay, however, measures only short-lived DNA damage, not heritable gene mutation. Also, the TGR assay measures gene mutation, but in bacterial
transgenes rather than an endogenous mammalian gene, and the assay is resource intensive and it only can be performed with specific genetically engineered rodents. Thus, the *Pig-a* assay fills an unoccupied niche of measuring *in vivo* gene mutation rapidly, in virtually any animal model, and in a manner that complements the blood-based cytogenetic assays. Additionally, as it is a peripheral blood-based assay, the animals need not be sacrificed to generate *Pig-a* gene mutation data. This property facilitates making mutation measurements in the same animals over time (longitudinal data collection) and integration of the assay into longer-term studies, including subchronic and chronic toxicology studies.

Mutagenicity responses in the *in vivo Pig-a* assay generally accumulate with repeat dosing, at least with strong mutagenic test substances. Thus, a treatment protocol similar to the one recommended for the TGR assay (TG 488) is also recommended for the *Pig-a* assay: repeat dosing for 28 days and mutant cell analysis following the final dose. Although the data indicate that short-term treatment protocols (especially those employing a single treatment) do not detect a small number of weaker mutagens, other treatment and sampling schedules may be conducted, with justification. Ideally assays are conducted with three appropriately spaced dose groups and a vehicle control, with the highest dose being the maximum tolerated dose or the limit dose (1000 mg/kg/day for longer-term studies of ≥14 consecutive days of treatment and 2000 mg/kg/day for studies with shorter treatment schedules). Concurrent positive control animals typically are not used for the assay, although the mutant mimic control used to optimize flow cytometry conditions and frozen samples giving positive responses may be used to lend confidence that the assay is capable of detecting a positive response.

Mature erythrocytes are abundant in peripheral blood and thus very small blood volumes are sufficient to accurately measure mutant frequency. Newly formed reticulocytes (RETs) are only a few percent of total red blood cells (RBCs) in peripheral blood; however, it is recommended that both RETs and total RBCs be analyzed for *Pig-a* mutant frequency. RET *Pig-a* mutant frequencies increase rapidly, over the first one or two weeks after treatment with a mutagen, and thus are a leading indicator of mutation. Because RBCs are more abundant than RETs, measurement of mutation in total RBCs produces highly robust data. In addition, the kinetics of
mutant manifestation, RETs before RBCs, lends credence to the induction of weak responses.

The percent of RETs among erythrocytes (%RETs) in the blood samples should be monitored as an indication of toxicity to the erythropoietic system. Also, because blood sampling can be conducted in a minimally invasive manner, it is recommended (but not required) that assays be conducted prior to the initiation of treatment, during the treatment (e.g., at Day 15 of a 28-day dosing protocol), and, if feasible, at a later time point (e.g., Day 56 of a 28-day repeated dosing study). Although this protocol provides an opportunity to integrate the assay into standard 28-day or 90-day repeat dose general toxicology tests (TG 407), short-term dosing protocols (e.g., single acute treatment or treatments made on three consecutive days) for many known test substances have yielded results qualitatively similar to those generated by repeat-dosing, and can be considered. Whichever treatment and sampling protocol is employed, it is recommended that a minimum of $1 \times 10^6$ RETs and $1 \times 10^6$ total RBCs be assayed for Pig-a mutation. Assaying this many RETs normally requires enriching the blood samples for either RETs or mutants using immunomagnetic separation technologies.

The animal is the experimental unit for the statistical evaluation of in vivo Pig-a mutant frequency data. Transformations may be used to reduce any non-normality of data and the addition of a small value is often helpful to reduce complications associated with log-transforming ‘zero’ mutant frequencies (e.g., 1.0, 0.1 or 0.01 added to the number of mutant cells per million). Positive responses exhibit a statistically significant, dose-related increase relative to the negative control, have at least one treatment response that is significantly greater than the concurrent vehicle control, and have at least one treatment response that exceeds the distribution of historical negative control frequencies (e.g., vehicle control and pretreatment frequencies). Negative responses do not fulfil these criteria and have evidence of sufficient bone marrow exposure. There is no requirement for verification of a clear positive or negative response, but further experimentation may be necessary to resolve borderline or equivocal results. Application of expert judgement and computational methods that take the kinetics of RET and total RBC mutant manifestation into account can be helpful for interpreting Pig-a data.
Studies have been conducted to confirm that the *Pig-a* gene mutation assay measures gene mutation, and specifically mutation of the *Pig-a* gene. Analyses of *Pig-a* mutation have been conducted in erythroid and granulocyte precursor cells from the bone marrow of treated mice and rats. These analyses indicate that the *Pig-a* mutant phenotype in peripheral blood erythrocytes is due to mutation in the *Pig-a* gene of nucleated precursor cells. Compelling indirect evidence includes finding *Pig-a* mutations in mRNA from mutant RETs, and the consistent relationship between *Pig-a* mutation and the mutant phenotype in rat lymphocytes, in L5178Y/Tk<sup>+</sup> cells, and in hematopoietic cells from humans with paroxysmal nocturnal hemoglobinuria. Also, the positive relationships between *Pig-a* mutation and *TGr* mutation and *Pig-a* mutation and genomic instability also are consistent with the assay measuring gene mutation. There is no evidence that alternative mechanisms, *e.g.*, gene silencing by methylation, are responsible for the phenotype measured in the *Pig-a* assay. Overall, the accumulated evidence indicates that the *Pig-a* phenotype, most commonly measured as loss of GPI-anchored proteins, is almost always due to mutations in the *Pig-a* gene. The rare instances when this was not the case have only been associated with human diseases or in particular cell lines; there is no indication that anything other than *Pig-a* mutation is responsible for mutants in the rodent *Pig-a* assay.

Ionizing radiation and over 80 chemicals, chemical mixtures, and nanomaterials have been tested for *Pig-a* mutation in rats, mice, or both. Several test substances have been tested in multiple laboratories and using different protocols. A committee was formed to establish criteria for evaluating *Pig-a* assay responses and categorize the responses as positive (P), negative (N), equivocal (E), or inconclusive (I). The committee used three sets of criteria that employed data from experiments conducted with short-term treatment protocols to different degrees. Although different definitions can be found in the literature, for the purpose of this document, short-term treatments were treatments conducted on 13 or fewer consecutive days and longer-term treatments were treatments conducted on 14 or more consecutive days. Using these criteria, data on 45 to 53 test substances were classified as producing P or N calls in the assay, and thus were useful for evaluating the performance of the assay.
Among the Pig-a positives, all were anticipated to be positive in the Pig-a assay based on their structure and/or responses in other assays. Pig-a negatives fell into several categories: 1) compounds that are generally regarded as negative for genotoxicity; 2) substances that are primarily or exclusively genotoxic through clastogenicity, aneugenicity, or mitotic recombination; and 3) substances that are genotoxic in other assays but appear to have insufficient bone marrow exposure by the route of exposure employed in testing. The data indicate a mechanism-based complementarity between the types of genetic changes detected by the in vivo Pig-a and MN assays, where several genotoxic compounds are positive for Pig-a and negative for MN and several are positive for MN and negative for Pig-a. In addition, studies have been conducted that discount the effects of potential confounders for the assay and that demonstrate the ability of the assay to distinguish between genotoxic and nongenotoxic pairs of structurally related compounds. Confounders that have been considered include toxicity to the erythropoietic system, and the effects of sex and of gene silencing.

In addition, many Pig-a studies include a mutant frequency analysis before exposure to any test substance or vehicle control substance. These data were used as a baseline control mutant frequency to evaluate the effect of treatment with 28 consecutive daily doses of vehicle control substances using a longitudinal analysis of data. Data from 10 of these substances were sufficient to make a call of P or N; all were evaluated as N.

An independent committee evaluated the test substances and vehicle control substances with either N or P responses in the in vivo Pig-a assay for genotoxicity in conventional genotoxicity assays and the rodent cancer bioassay. Responses in the in vivo TGR assay for bone marrow mutagenicity and mutagenicity in any tissue were used to define the in vivo mutagenicity of test substances; these defined responses were used to test the performance of the Pig-a assay as a test of in vivo gene mutation. Although the relatively small data sets employed for the analysis compromised differentiating between the variables that were evaluated, in general, the Pig-a assay showed a near perfect degree of accuracy when TGR mutagenicity in bone marrow or cancer in hematopoietic tissue was used to define true P and N responses. When TGR mutagenicity in any tissue or in vivo MN induction was used to define true P and N responses,
the accuracy of the assay was slightly lower, with assay accuracy negatively impacted by the responses of test substances that were mutagenic in tissues other than bone marrow and/or mainly genotoxic by a mechanism that did not result in gene mutation. Lower accuracy when rodent cancer in any tissue was used to define P and N responses also was affected by several nongenotoxic carcinogens that were positive in the tumor bioassay but negative in the Pig-a assay. Although positive responses for a small number of test substances are known to require multiple treatment protocols, the effect of treatment protocol (short-term vs. longer-term) was not a major factor in the overall accuracy of the assay.

When used to address regulatory requirements, the Pig-a assay is generally regarded as a follow-up assay for evaluating the in vivo mutagenicity of substances testing positive in bacterial or in vitro mammalian cell gene mutation assays. In this regard, the most recent International Council on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use M7 guidance recommends the Pig-a gene mutation assay as a follow-up in vivo assay for drug impurities that are positive for bacterial mutagenicity as long as there is sufficient evidence for bone marrow exposure. The assay has additional applications in conducting weight-of-evidence genotoxicity assessments, quantitative measurements of in vivo mutation, extended-time monitoring of mutant frequencies, and as a routine complement to MN testing that affords more information on in vivo genotoxicity without using additional animals. Other possible applications include using the human PIG-A assay as a direct confirmation of observations from animal studies and for human biomonitoring studies in potentially exposed populations. The TGR assay is not useful for these later applications as it requires the use of genetically manipulated animals, and it cannot be conducted with humans. Pig-a gene mutation assays also have been used in a variety of hypothesis-driven human and rodent studies that benefit from measurement of in vivo mutation.

GPI anchors are found on all mammalian cells that have been examined, along with cells from lower eukaryotes. Thus, Pig-a gene mutation assays, in theory, can be conducted in virtually all animal species of toxicological interest, making the Pig-a assay and Pig-a data highly translatable. The Pig-a assay began with work on measuring mutation in humans, and several
recent studies have demonstrated the feasibility of a human erythrocyte PIG-A assay. Preliminary observations indicate that the assay can be conducted with erythrocytes from nonhuman primates; male rat germ cells; human granulocytes, lymphocytes, monocytes, and bone marrow progenitor cells; lymphocytes, granulocytes, and monocytes from rodents; and T lymphocytes from nonhuman primates. In addition to in vivo assays, assays measuring the Pig-a mutant phenotype (either loss of GPI anchors or GPI-anchored proteins) have been conducted in vitro with L5178Y/Tk<sup>+</sup> cells, TK6 human lymphoblastoid cells, and chicken DT-40 cells. There is a clear relationship between mutation in the Pig-a gene and loss of GPI-anchored proteins for L5178Y/Tk<sup>+</sup> cells. In DT-40 cells the mutant phenotype is due to mutation in the Pig-o gene (there are two copies of Pig-a and one copy of Pig-o in chicken cells), and the mutant phenotype in TK6 cells is caused by mutation of either PIG-A or PIG-L, both of which genes are present in one functional copy in these cells. None of these alternative assays has yet been developed to the extent of the rodent erythrocyte Pig-a gene mutation assay.

Sample collection in the rodent erythrocyte Pig-a assay is minimally invasive; samples can be collected over time from the same animal, and the assay lends itself to integration with other toxicology studies. Thus, even though it is an in vivo assay, these attributes make the assay consistent with the 3Rs principles, in that additional valuable data can be readily derived from animals already dedicated to toxicological testing. In addition, approaches have been described that use blood samples from one or a few animals to train personnel to conduct the assay and to establish laboratory proficiency in the assay as recommended by OECD test guidelines.

The Detailed Review Paper in Part 1 is accompanied by Part 2, which describes the state of Pig-a assay validation according to the principles set forth in OECD Guidance Document 34. Following review and acceptance of these documents by the OECD Working Group of the National Coordinators of the Test Guidelines Programme (WNT), a Test Guideline will be prepared for conducting the assay.
Introduction to Part 1

Lucio Luzzatto, David Araten and their colleagues were the first to suggest using the phosphatidylinositol glycan class A gene (PIG-A) as a reporter of in vivo mutation (Araten et al., 1999). As explained in Section 1, mutations in the PIG-A gene are known to disrupt the synthesis of glycosylphosphatidylinositol (GPI) anchors, resulting in the loss of cell-surface proteins. These losses can be identified by immunofluorescent staining and flow cytometric analysis of peripheral blood cells. The Luzzatto and Araten labs subsequently conducted a number of studies using PIG-A as a reporter of mutation in human cells (e.g., Peruzzi et al., 2010; Rondelli et al., 2010); others built upon the basic principles used for their human PIG-A assay to develop assays for measuring gene mutation in rodent erythrocytes (Bryce et al., 2008; Miura et al., 2008a, b; Phonethepswath et al., 2008; reviewed in Dobrovolsky et al., 2010b). Unless otherwise specified, the term ‘Pig-a assay’ (Pig-a is the rodent gene) is used in this report to refer to assays that measure gene mutation induced in the bone marrow erythroid cells of rats or mice. As explained below, these mutations are almost always detected by flow cytometric evaluation of the presence or absence of GPI-linked protein markers in peripheral blood erythrocytes.

The potential regulatory applications of the Pig-a assay were recognized early on by the Health and Environmental Sciences Institute (HESI) (until 2018, HESI was associated with the International Life Sciences Institute) (Schuler et al., 2011), and currently HESI sponsors a Pig-a workgroup as part of its Genetic Toxicology Technical Committee (GTTC). Additional support for development of the Pig-a assay came from a workgroup of the International Workshop on Genotoxicity Testing (IWGT) that met in 2013 in Foz do Iguaçu, Brazil. The review and recommendations of that workgroup (Gollapudi et al., 2015) form the basis for a major part of this report.

Encouraged by the IWGT recommendations, the GTTC Pig-a workgroup prepared an Organisation for Economic Cooperation and Development (OECD) Standard Project Submission Form (SPSF) proposing the development of a Test Guideline (TG) for the in vivo Pig-a assay (Test
Guideline for the *Pig-a* assay: an *in vivo* gene mutation assay promoting the 3Rs principles. Following review and revision, that proposal was accepted by the OECD Working Group of the National Coordinators of the Test Guidelines Programme (WNT) in April 2015 (Project 4.93), with the understanding that a Detailed Review Paper (DRP) and a Validation Document should first be approved by the WNT before TG development. This DRP has been prepared in response to this WNT request. In particular, the WNT review of the SPSF emphasized supplying convincing evidence that the assay measures true gene mutation. Reviewers expressed enthusiasm about the integration potential and animal welfare benefits of the assay but were concerned about sufficient data existing for performing a useful Retrospective Performance Assessment (RPA) on the assay. They also expressed interest in the possibility of developing a *Pig-a* germ cell assay and an *in vitro* version of the assay. This report pays special attention to all these issues.
1. Principle of the Pig-a assay

Inactivating mutations and mutations that otherwise significantly affect the structure of an encoded protein often can alter the phenotype of the mutant cell. These phenotypic changes can result from shutdown, attenuation or repurposing of an endogenous metabolic process, accumulation or depletion of a specific metabolite, or disturbances in the networks involved in signal transduction, intracellular trafficking, gene expression and/or proliferation. Thus, many genes could potentially serve as a reporter of mutation. In practice, however, the number of useful reporter genes is quite limited because most mutations cause relatively subtle phenotypic changes in affected cells that are difficult to detect using common laboratory equipment. The Pig-a gene is particularly useful for mutant detection because rare mutant phenotype cells are easily detected and quantified among a large number of wild-type cells after minimal and relatively inexpensive sample processing. Also, the mutant phenotype, in rodents at least, does not affect the growth, viability, and persistence of mutant cells relative to wild-type cells.

a. Genesis of the Pig-a assay

i) Glycosylphosphatidylinositol anchors and Pig-a

Most eukaryotic cells have evolutionally conserved mechanisms for displaying proteins on the exterior surface of their cytoplasmic membrane. A diverse subset of >150 of these proteins is maintained exoplasmically on the cytoplasmic membrane by the anchor molecule glycosylphosphatidylinositol (GPI) (Stevens, 1995; Brodsky and Hu, 2006; Brodsky, 2014; Kinoshita, 2014). The anchor is composed of a hydrophobic tail consisting of two or three fatty acid carbon chains and a charged three-mannose glycan core that covalently bridges to the protein via a phosphoethanolamine (Figure 1). The hydrophobic tail of the GPI molecule is embedded in the cytoplasmic membrane so that the attached protein does not require a lipophilic transmembrane domain to remain tethered to the cell membrane.
The syntheses of GPI and of the GPI-anchored protein are independent (Figure 2). The conjugation of the protein to the finished anchor occurs in the lumen of the endoplasmic reticulum at the ω-site of the peptide. The entire assembly is exported via COP-II-mediated endocytosis to the Golgi apparatus, and then endosomically to the surface of the cell, presenting the GPI-anchored protein on the exoplasmic leaflet (Figure 2). At the surface, GPI-anchored proteins congregate within restricted domains, termed lipid rafts, which differ in lipid and protein content from the rest of the plasma membrane. Partial or complete disruption of GPI biosynthesis that results in GPI anchor deficiency causes a partial or complete deficiency of GPI-anchored proteins at the cellular surface.

At least 31 genes are involved in the synthesis of the GPI anchor and the posttranslational modification of the anchor-protein complex (Figure 2) (Brodsky, 2014; Kinoshita, 2014). Inactivating mutations in several of these genes are known to result in a surface-protein-deficient phenotype, which can be detected by immunofluorescent labelling and flow cytometry (Figure 3). However, because Pig-a is the only one of these genes that is on the X chromosome in mammals (Takeda et al., 1993; Ware et al., 1994), it is the only gene involved in GPI biosynthesis that can reliably function as a reporter of mutation. A single copy of Pig-a is present in male cells and a single functional copy in female cells (the second copy of the Pig-a gene is transcriptionally silenced in females). All other genes in the pathway are autosomal and are likely to be present in two functional copies. The probability of inducing a detectable mutation-associated phenotype by mutating the single functional copy of an X-linked gene is much higher than that of detecting a phenotype by independently mutating the two functional copies of an autosomal gene (Brodsky and Hu, 2006). Thus, the vast majority of mutant cells that lack the GPI anchor and acquire a surface-protein-deficient phenotype are expected to result from mutation in the Pig-a gene. This association of GPI-anchored protein loss with Pig-a mutation is the basis for the Pig-a assay.

ii) PI(G-A and paroxysmal nocturnal hemoglobinuria

Much of what is known about Pig-a mutation is derived from studies of a rare, acquired human disease, paroxysmal nocturnal hemoglobinuria (PNH), a bone marrow disorder that is
characterized by hemolytic anemia, thrombosis, and peripheral blood cytopenia (Brodsky and Hu, 2006; Brodsky, 2014). The proximal cause of PNH is GPI deficiency, which is almost always caused by mutation of the PI@A gene. The rare exceptions (sometimes referred as causing ‘PNH-like’ disease) involve mutations in both alleles of autosomal GPI biosynthesis genes (e.g., PI@T [Krawitz et al., 2013; Mason et al., 2019] or PI@S [Nguyen et al., 2018]) or in genes for specific GPI-linked proteins (Yamashina et al., 1990; Nevo et al., 2013).

In PNH patients, the bone marrow is populated by one or a small number of expanded hematopoietic stem cell (HSC) clones that contain single PI@A mutations. These clones give rise to the hematopoietic progenitor cell lineages (e.g., erythroid, lymphoid and myeloid; see Figure 4 for general scheme of the hemopoietic system), all having those same mutations. The PI@A mutant progenitors, in turn, produce mature differentiated cells that transit into the peripheral circulation and that can result in specific cell populations that are nearly 100% deficient in GPI anchors and surface-tethered GPI-anchored proteins (Mortazavi et al., 2003; Schrezenmeier et al., 2014).

Deficiency of the GPI-anchored proteins CD55 and CD59 at the surface of peripheral RBCs results in a loss of complement regulation and triggers the lysis of mutant RBCs by the complement system (Wilcox et al., 1991). This accounts for the intravascular RBC lysis that is a primary clinical manifestation of PNH and the source of the pigmented urine (hemoglobinuria) that is often seen in PNH patients (Schrezenmeier et al., 2014).

**iii) Analysis of Pig-a mutant cells as a reporter of gene mutation**

A standard method for diagnosing PNH and for estimating the size of PI@A mutant clones involves labeling the cells of peripheral blood with fluorescent monoclonal antibodies against one or more GPI-anchored proteins and subsequent quantitation of marker-deficient (nonfluorescent) PI@A mutant cells by flow cytometry (Hall and Rosse, 1996; Madkaikar et al., 2009; Borowitz et al., 2010). Light scatter properties, sometimes in conjunction with additional antibodies against transmembrane proteins not anchored by GPI, can be used to identify the
specific cell lineage. Granulocytes are often used for the diagnosis of PNH using anti-CD59 antibodies.

Besides the large clonal expansions of PIG-A mutant clones associated with PNH, healthy, non-PNH subjects have a low frequency (approximately 0.1-2 x 10⁻⁵) of marker-deficient cells in the blood (Araten et al., 1999; Rawstron et al., 1999; Ware et al., 2001; Hu et al., 2005; Brodsky and Hu, 2006). This frequency is similar to the background mutant frequency of the HPRT gene in human peripheral blood lymphocytes (Robinson et al., 1994). Like PIG-A, HPRT is an X-linked gene, and the HPRT lymphocyte assay has been widely used to monitor mutant frequencies in humans. Based on their understanding of how PIG-A operates in GPI anchor synthesis, Araten et al. (1999) proposed that a flow cytometry-based PNH diagnostic protocol for detecting PIG-A mutant cells might form the basis of an assay for monitoring mutation in humans, specifically mutations induced in the bone marrow.

As the pathway for GPI synthesis is conserved in mammals (Kawagoe et al., 1994), similar flow cytometric methodology was designed for the quantitative detection of GPI-deficient and GPI-marker-deficient Pig-a mutant cells in rats, mice, and rhesus monkeys (Figure 3; Bryce et al., 2008; Miura et al., 2008a; Phonethepsworth et al., 2008; Dobrovolsky et al., 2009). The unifying characteristic of all current flow cytometric assays detecting Pig-a (or PIG-A) mutant cells is the use of fluorescent antibody(s) against GPI-anchored protein marker(s), or (infrequently) a fluorescent version of aerolysin (FLAER) to distinguish between GPI-proficient and GPI-deficient cells (Miura et al., 2008a). FLAER is a fluorescent, nontoxic derivative of the bacterial toxin aerolysin that binds the GPI anchor rather than the anchored protein marker (Brodsky et al., 2000). With either labeling approach, wild-type cells will be fluorescent, and GPI-deficient cells or cells lacking surface-displayed GPI-anchored proteins, will be non-fluorescent and presumed to harbor Pig-a mutations.

Finally, native aerolysin is toxic to cells with GPI anchors. Aerolysin binding results in disruption of the cytoplasmic membrane and cell death; thus, aerolysin can be used as a selection agent for identifying GPI-deficient (presumably Pig-a mutant) cells. Aerolysin (more commonly as the
protoxin, proaerolysin) has been used as a selection agent for limiting dilution cloning versions of the Pig-a assay, conducted both with rat and human T lymphocytes and with cell lines (e.g., Miura et al., 2008b; McDiarmid et al., 2011; Nakamura et al., 2012; Nicklas et al., 2015; Labash et al., 2015b; Y Wang et al., 2018). Although aerolysin selection is suitable for confirming the GPI-deficient phenotype and generating mutant clones for sequencing, it lacks the integration potential and convenience of the erythrocyte flow cytometry assay and thus is not considered useful as a routine regulatory assay for in vivo mutation.

b. Considerations in assay design

Commercially available antibodies against GPI-anchored protein markers are usually species-specific (e.g., anti-rat CD59 antibody will not bind GPI-anchored CD59 on the surface of mouse or human cells) while FLAER is not species-specific (i.e., it will bind the anchors present at the surface of mouse, human and rat cells). But, in many cases, monoclonal antibodies are superior to the bacterial toxin-based FLAER for discriminating rare populations of GPI-deficient cells (e.g., Miura et al., 2008a); thus, FLAER is rarely used for routine Pig-a mutation assays.

At least 150 GPI-anchored proteins have been identified; often their expression is variable in different cell types and at different stages of maturation, which is something that should be considered for designing a reliable GPI-anchor detection method (Brodsky et al., 2000; Hernández-Campo et al., 2006, 2007). Just as not every GPI-anchored protein marker is suitable for diagnosis of PNH, not every marker is suitable for designing a flow cytometry-based Pig-a mutation detection assay. For specific cell types, it is necessary to choose a marker that is expressed ubiquitously and at a high level at the cell surface. Experience suggests that rat CD59 is suitable for rat erythrocyte assays while mouse CD24 is suitable for mouse erythrocyte assays.

Peripheral blood is the preferred source of samples for flow cytometry-based Pig-a assays: peripheral blood is easy to collect, and the samples are naturally single-cell suspensions amenable for labeling and processing on a flow cytometer. Importantly, enzymatic dissociation
is typically used to generate individual cells from solid tissues. The need for this processing may limit the utility of the assay for mutation assessment in solid tissues due to the potential for damaging the surface-bound GPI-anchored markers employed in the antibody labeling (but see Section 10d). Although Pig-a assays have been conducted with other cell types (Section 10c), erythrocyte-based Pig-a assays have a major advantage owing to the extraordinarily high concentration of erythrocytes in peripheral blood. Microliter volumes of blood contain sufficient erythrocytes for accurately measuring mutant cell fractions. The requirement for only small blood samples translates into an ability to collect samples repeatedly, even from small laboratory animals, thus allowing the acquisition of longitudinal data sets based on assays conducted with individual animals.

In adult animals, most peripheral blood erythrocytes are derived from progenitors in the bone marrow (Sun et al., 2014; Figures 3 and 4). These progenitors generate a series of morphologically identifiable precursors that enucleate to form reticulocytes (RETs) and finally give rise to the mature red blood cells (RBCs) found in peripheral blood. As RETs mature into RBCs over a period of days, they progressively lose their non-GPI-anchored surface protein CD71 and their residual cytoplasmic RNA (Serke and Huhn, 1992), and so CD71 and residual RNA are convenient markers of the RET subpopulation of erythrocytes. The proportion of RETs among bone marrow cells is quite high (e.g., 40-80 %; Fiedler et al., 2010); however, after RETs migrate to the periphery, they are rapidly converted into mature RBCs and their proportion among total erythrocytes is reduced to a few percent or less, depending on species, age, and RET labeling technique (e.g., Serke and Huhn, 1992; Dertinger et al., 2015). RETs are important for the Pig-a assay because they are the earliest reporters of Pig-a mutations induced in erythroid precursors of the bone marrow that are accessible in peripheral blood. Since RETs are short-lived, the maximum frequency of mutant RETs after an acute treatment is achieved rapidly. After an acute treatment of rats, for example, the maximum mutant frequency in peripheral RETs usually occurs within two weeks, while it may take 6 to 8 weeks before the maximum frequency is reached in mature RBCs (which corresponds to the approximate lifetime of RBCs in the periphery) (Miura et al., 2009).
Another important reason for studying mutation using RETs is that in human PNH patients, the complement system lysed CD59-deficient $\textit{PIG-A}$ mutant RBCs (Wilcox et al., 1991). Observations by Dertinger et al. (2015) suggest that this factor affects the erythrocyte $\textit{PIG-A}$ assay in humans. If this were to occur in mutagen-treated animals, then the longer a mutant erythrocyte remains in circulation, the more likely is the chance that it can be destroyed by complement-mediated lysis. In this respect RET $\textit{Pig-a}$ assays not only convey information on mutagenicity in bone marrow earlier, data from RET assays also lend plausibility to weak responses measured in total RBCs. There is, however, accumulating evidence indicating that immune lysis does not affect $\textit{Pig-a}$ mutant frequencies in mutagen-treated laboratory animals as it appears to do in humans (see below: Miura et al., 2009; Phonethepswath et al., 2010).

c. Initial observations on the source of the mutants detected in the erythrocyte $\textit{Pig-a}$ assay

Relatively early observations about the rat erythrocyte $\textit{Pig-a}$ assay indicated that responses accumulate with repeated treatments and once an elevated mutant frequency is achieved, it persists for a prolonged period of time. Miura et al. (2009) found that four smaller doses of $\textit{N}$-ethyl-$\textit{N}$-nitrosourea (ENU), given at weekly intervals, produced virtually the same RBC mutant frequency as when a single dose was given that was equivalent to the sum of the four small doses. Also, after the maximum mutant frequency was achieved in total RBCs (6-8 weeks after the single treatment, or 6-8 weeks after the final weekly treatment), the response persisted until the experiment was terminated 16 to 20 weeks later. Phonethepswath et al. (2010) confirmed the persistence of elevated $\textit{Pig-a}$ mutant frequencies in an experiment that monitored ENU-induced mutant frequencies in rat total RBCs and RETs over a period of 6 months, and Dertinger et al. (2014a) observed a similar persistence of elevated $\textit{Pig-a}$ mutant frequencies in rats treated with cisplatin. Since the persistence of elevated mutant frequencies exceeded the lifetime of the differentiated peripheral blood erythrocytes, these observations suggested that a relatively stable, renewable cell population, e.g., long-lived erythroid
progenitors or perhaps even hematopoietic stem cells (HSCs), was the source of the induced mutants.

Bone marrow, the probable source of erythrocyte Pig-a mutants in the rodent Pig-a assay, contain a mixture of HSCs and multipotent and lineage-committed progenitor cells (Sun et al., 2014; Busch et al., 2015; Figure 4). Although all these cells are nucleated, and many are actively dividing, and thus could be the source of Pig-a mutations, it is unlikely that mutant HSCs make a major contribution to the mutant frequencies measured in the Pig-a assay. The number of ‘true’ HSCs in an animal is estimated to be 0.5-2 x 10^4 (Abkowitz et al., 2002; Busch et al., 2015); thus, assuming a neutral mutant phenotype, Pig-a mutant cell frequencies of less than 50 x 10^-6 are unlikely if Pig-a mutation is restricted to HSCs. In humans with PNH, where the PIG-A mutations are most likely derived from HSCs, the mutant frequencies in all hematopoietic lineages are quite high, and can approach 100% (Mortazavi et al., 2003; Schrezenmeier et al., 2014). These extraordinarily high mutant frequencies are almost always due to the expansion of one, or a very few, mutant HSC clones (Brodsky, 2014; Kinoshita, 2014). Spontaneous Pig-a mutant frequencies in rodents, on the other hand, are almost always <5 x 10^-6, and often <1 x 10^-6 (see Figure 8); also, only potent mutagens produce Pig-a mutant frequencies in rodents of >50 x 10^-6. Finally, Busch et al. (2015) indicate that the time necessary for a de novo HSC mutation to be expressed fully in the periphery of an adult mouse may exceed the mouse’s lifespan. Thus, it is unlikely that HSC mutation makes a major contribution to the Pig-a mutant phenotype measured in peripheral blood.

Bone marrow hematopoietic precursor cells are a more likely source of mutations in the assay. Recent studies indicate that much of steady-state erythropoiesis is driven not by true ‘long-term’ HSCs, but rather by relatively long-lived multipotent and lineage-specific progenitor cells in the bone marrow (Sun et al., 2014; Busch et al., 2015). The lifespans of these cells are consistent with the persistence of induced mutant frequencies seen in the assay. Although it is not clear how large these cellular compartments are, they are certainly larger than the HSC compartment, and thus minimum mutant frequencies resulting from mutations in these precursor cell populations could be of the order seen in the Pig-a gene mutation assay. The
analysis of Pig-a mutation in rat bone marrow erythroid cells and granulocytes described in Section 6 is consistent with the Pig-a assay detecting mutations induced in these hematopoietic precursor cells.

2. Existing in vivo genetic toxicology assays

Establishing the utility of the Pig-a assay for evaluating in vivo genotoxicity requires understanding the principles, methodologies, and applications of the current tests used for that purpose. In vivo tests are used for hazard identification, both as parts of test batteries with in vitro tests or to follow-up in vitro findings (Cimino, 2006). A number of in vivo genotoxicity tests exist, but each is focused on detecting a particular class of genotoxicity and/or detecting genotoxicity in a specific cell type. Thus, no one test is capable of reliably evaluating the genotoxicity of all test substances, which is the rationale for using batteries of tests (Eastmond et al., 2009). In addition, in vivo tests are used for risk characterization and dose-response analysis, as is the case when they are used for cancer mode-of-action studies or evaluating the potential health risks of low-dose exposures (Boverhof et al., 2011).

Of the many tests that have been developed over the years, eight tests for measuring in vivo genotoxicity currently have OECD TGs (OECD, 2013, 2016a,b,d,e,f). These tests have attained regulatory acceptance and have clear recommendations for their use and interpretation. Three of these in vivo assays are used only for somatic cell testing (TG 475 for detecting clastogenicity, TG 474 for detecting clastogenicity and aneugenicity, and TG 489, which detects DNA damage as single- and double-strand breaks), two are specifically for germ cell testing (TG 478 and TG 483, both of which mainly detect clastogenicity), and one can detect genotoxicity in both germ cells and somatic cells (TG 488, which detects gene mutation in transgenes). TGs for these assays are either new (TG 489, the in vivo alkaline comet assay) or recently revised (OECD, 2015). Two additional tests, a somatic cell test measuring DNA damage as unscheduled DNA synthesis (UDS) in rat liver (TG 486; OECD, 1997), and the heritable translocation assay for germ cell mutation in mice (TG 485; OECD, 1986b) are rarely used, and have not been updated
recently. In addition, the OECD TG for the Mouse Spot Test (TG 484; OECD, 1986a), an in vivo test for somatic cell mutation, was deleted in 2014 because it has become obsolete and is no longer recommended. An extensive analysis of in vitro and in vivo genotoxicity tests is available in the review and DRP prepared as part of TG 488 development (Lambert et al., 2005; OECD, 2009). The following, using language taken directly from the TGs, is intended to provide a general description of the tests that are most relevant to the Pig-a assay, the four most commonly used tests with OECD TGs that evaluate in vivo genotoxicity in somatic cells.

a. TG 474: the mammalian erythrocyte micronucleus test

The following description is taken verbatim from the Introduction to OECD TG 474 (OECD, 2016a).

“The mammalian in vivo micronucleus test is especially relevant for assessing genotoxicity because, although they may vary among species, factors of in vivo metabolism, pharmacokinetics and DNA repair are active and contribute to the responses. An in vivo assay is also useful for further investigation of genotoxicity detected by an in vitro system.”

“The mammalian in vivo micronucleus test is used for the detection of damage induced by the test chemical to the chromosomes or the mitotic apparatus of erythroblasts. The test evaluates micronucleus formation in erythrocytes sampled either in the bone marrow or peripheral blood cells of animals, usually rodents.”

“The purpose of the micronucleus test is to identify substances that cause cytogenetic damage which results in the formation of micronuclei containing either lagging chromosome fragments or whole chromosomes.”

“When a bone marrow erythroblast develops into an immature erythrocyte (sometimes also referred to as a polychromatic erythrocyte or reticulocyte), the main nucleus is extruded; any micronucleus that has been formed may remain behind in the cytoplasm.
Visualization or detection of micronuclei is facilitated in these cells because they lack a main nucleus. An increase in the frequency of micronucleated immature erythrocytes in treated animals is an indication of induced structural or numerical chromosomal aberrations.”

“Newly formed micronucleated erythrocytes are identified and quantitated by staining followed by either visual scoring using a microscope, or by automated analysis. Counting sufficient immature erythrocytes in the peripheral blood or bone marrow of adult animals is greatly facilitated by using an automated scoring platform.”

“Although not normally done as part of the test, chromosome fragments can be distinguished from whole chromosomes by a number of criteria. These include identification of the presence or absence of a kinetochore or centromeric DNA, both of which are characteristic of intact chromosomes. The absence of kinetochore or centromeric DNA indicates that the micronucleus contains only fragments of chromosomes, while the presence is indicative of chromosome loss.”

b. TG 475: the mammalian bone marrow chromosomal aberration test

The following description is taken verbatim (with modifications as noted) from the Introduction to OECD TG 475 (OECD, 2016b).

“The mammalian in vivo bone marrow chromosomal aberration test is especially relevant for assessing genotoxicity because, although they may vary among species, factors of in vivo metabolism, pharmacokinetics and DNA-repair processes are active and contribute to the responses. An in vivo assay is also useful for further investigation of genotoxicity detected by an in vitro system.”

“The mammalian in vivo chromosomal aberration test is used for the detection of structural chromosome aberrations induced by test chemicals in bone marrow cells of animals, usually rodents......Structural chromosomal aberrations may be of two types,
chromosome or chromatid. While the majority of genotoxic chemical-induced
aberrations are of the chromatid-type, chromosome-type aberrations also occur.
Chromosomal damage and related events are the cause of many human genetic diseases
and there is substantial evidence that, when these lesions and related events cause
alterations in oncogenes and tumour suppressor genes, they are involved in cancer in
humans and experimental systems. Polyploidy (including endoreduplication) could arise
in chromosome aberration assays in vivo. However, an increase in polyploidy per se does
not indicate aneugenic potential and can simply indicate cell cycle perturbation or
cytotoxicity. This test is not designed to measure aneuploidy. An in vivo mammalian
erythrocyte micronucleus test (Test Guideline 474) or the in vitro mammalian cell
micronucleus test (Test Guideline 487) would be the in vivo and in vitro tests,
respectively, recommended for the detection of aneuploidy.”

c. TG 489: the in vivo mammalian alkaline comet assay

The following description is taken verbatim (with modifications as noted) from OECD TG 489
(OECD, 2016f).

“The purpose of the comet assay is to identify substances that cause DNA damage. Under
alkaline conditions (>pH 13), the comet assay can detect single and double stranded
breaks, resulting, for example, from direct interactions with DNA, alkali labile sites or as a
consequence of transient DNA stand breaks resulting from DNA excision repair. These
strand breaks may be repaired, resulting in no persistent effect, may be lethal to the cell,
or may be fixed into a mutation resulting in a permanent viable change. They may also
lead to chromosomal damage which is also associated with many human diseases
including cancer.”

“The in vivo alkaline comet assay is especially relevant to assess genotoxic hazard in that
the assay’s responses are dependent upon in vivo ADME (absorption, distribution,
metabolism and excretion), and also on DNA repair processes. These may vary among species, among tissues and among the types of DNA damage.”

To fulfill animal welfare requirements, in particular the reduction in animal usage (3Rs - Reduction, Refinement, Replacement - principles), this assay can also be integrated with other toxicological studies, e.g., repeated dose toxicity studies......or the endpoint can be combined with other genotoxicity endpoints such as the in vivo mammalian erythrocyte micronucleus assay......The comet assay is most often performed in rodents, although it has been applied to other mammalian and non-mammalian species.”

“The selection of route of exposure and tissue(s) to be studied should be determined based on all available/existing knowledge of the test chemicals e.g. intended/expected route of human exposure, metabolism and distribution, potential for site-of-contact effects, structural alerts, other genotoxicity or toxicity data, and the purpose of the study. Thus, where appropriate, the genotoxic potential of the test chemicals can be assayed in the target tissue(s) of carcinogenic and/or other toxic effects. The assay is also considered useful for further investigation of genotoxicity detected by an in vitro system. It is appropriate to perform an in vivo comet assay in a tissue of interest when it can be reasonably expected that the tissue of interest will be adequately exposed.”

“The technique is in principle applicable to any tissue from which analysable single cell/nuclei suspensions can be derived. Proprietary data from several laboratories demonstrate its successful application to many different tissues, and there are many publications showing the applicability of the technique to (many different) organs or tissues...”

“Cross-links cannot be reliably detected with the standard experimental conditions of the comet assay. Under certain modified experimental conditions, DNA-DNA and DNA-protein crosslinks, and other base modifications such as oxidized bases might be detected......But further work would be needed to adequately characterize the necessary
protocol modifications......The assay is not appropriate, even with modifications, for detecting aneugens.”

d. TG 488: the transgenic somatic and germ cell gene mutation assays

The following description is taken verbatim (with modifications as noted) from OECD TG 488 (OECD, 2013).

“(TGR gene mutation assays) use transgenic rats and mice that contain multiple copies of chromosomally integrated plasmid or phage shuttle vectors. The transgenes contain reporter genes for the detection of various types of mutations induced in vivo by test chemicals.”

“Mutagenesis in the TGR models is normally assessed as mutant frequency; if required, however, molecular analysis of the mutations can provide additional information.”

“These rodent in vivo gene mutation tests are especially relevant to assessing mutagenic hazard in that the assays’ responses are dependent upon in vivo metabolism, pharmacokinetics, DNA repair processes, and translesion DNA synthesis, although these may vary among species, among tissues and among the types of DNA damage. An in vivo assay for gene mutations is useful for further investigation of a mutagenic effect detected by an in vitro system, and for following up results of tests using other in vivo endpoints.....”

“If there is evidence that the test chemical, or a relevant metabolite, will not reach any of the tissues of interest, it is not appropriate to perform a TGR gene mutation assay.”

“In the assays described above, the target gene is bacterial or bacteriophage in origin, and the means of recovery from the rodent genomic DNA is by incorporation of the transgene into a λ bacteriophage or plasmid shuttle vector......and subsequent detection
of mutations in bacterial hosts under suitable conditions. The assays employ neutral transgenes that are readily recoverable from most tissues.”

“The basic TGR gene mutation experiment involves treatment of the rodent with a chemical over a period of time.....Administration is usually followed by a period of time, prior to sacrifice, during which the agent is not administered and during which unrepaired DNA lesions are fixed into stable mutations.......After the animal is sacrificed, genomic DNA is isolated from the tissue(s) of interest and purified.”

“The mutations scored in the lacI, lacZ, clI and gpt point mutation assays consist primarily of base pair substitution mutations, frameshift mutations and small insertions/deletions. The relative proportion of these mutation types among spontaneous mutations is similar to that seen in the endogenous Hprt gene. Large deletions are detected only with the Spi- selection and the lacZ plasmid assays.......”

“It is anticipated that in the future it may be possible to combine a TGR gene mutation assay with a repeat dose toxicity study (TG 407). However, data are required to ensure that the sensitivity of TGR gene mutation assays is unaffected by the shorter one day period of time between the end of the administration period and the sampling time, as used in the repeat dose toxicology study, compared to the three days used in the TGR gene mutation assays. Data are also required to indicate that the performance of the repeat dose assay is not adversely affected by using a transgenic rodent strain rather than traditional rodent strains.”

3. Strengths and weaknesses of the Pig-a assay

The Pig-a assay has several compelling strengths for measuring gene mutation in vivo. Perhaps most importantly, it can be performed in most mammalian species, mutant analysis is rapid and relatively low in cost, and the assay integrates well into general toxicity testing, making maximum use of animal resources already committed to in vivo testing. In addition, the same
endpoint that is monitored in rodents (i.e., *Pig-a* mutation), can be measured in humans and mammalian cells in culture, making possible mechanistic studies *in vitro* and direct assessment of human effects *in vivo*. These advantages stem from the following characteristics of the assay:

- The assay measures a mutant phenotype dependent upon the functioning of a pathway conserved in mammalian species: it does not depend on the properties of any specific strain of animal.

- The assay requires only microliters of peripheral blood, which facilitates relatively non-invasive collection of serial data and integration into studies without the need for animal sacrifice.

- The assay measures mutation in unattached hematopoietic cells (i.e., erythrocytes), which facilitates high throughput scoring via flow cytometric analysis without need for *ex vivo* culture or extensive manipulations.

These strengths are evident in comparison with the TGR assay, which is currently the only *in vivo* assay for gene mutation with an OECD Test Guideline (OECD, 2013). The costs associated with performing the TGR assay, in terms of time, labor, and materials, are often cited as a major barrier to its adoption for routine genotoxicity testing (Boverhof et al., 2011; Page et al., 2015). In addition, more than cost, the requirement for using specific strains of transgenic rodents and mutation sampling times that are incompatible with a 28-day repeat-dose general toxicology study reduce the potential for integration with other toxicology testing. Although efforts are on-going to address these problems (Akagi et al., 2015; Nohmi et al., 2017; Hori et al., 2019), these barriers to integrating TGR mutation analysis with the measurement of other toxicology endpoints currently remain. In addition, using a transgene imposes limitations on the types of mutations that can be measured. Finally, the relatively complex structure of endogenous mammalian genes is such that they can recover a greater variety of mutation types than can a
bacterial transgene, *e.g.*, megabase deletions (Tao et al., 1993; Chen et al., 1998; Walker et al., 1999).

The use of erythrocytes for conducting the *Pig-a* assay facilitates rapid analysis, minimally invasive sample collection, and integration of the assay into general toxicity testing, which are all major advantages. Measuring mutation using erythrocytes, however, also results in the two greatest weaknesses of the assay.

- The assay only detects mutations induced in erythropoietic tissue (*i.e.*, mainly the bone marrow in adult animals; see Section 1c), reducing its sensitivity for substances that are strongly mutagenic only in other tissues, like liver. This limitation of measuring genotoxicity in hematopoietic tissue is also true for the somatic cell cytogenetic assays described in OECD TGs 474 and 475 (see Sections 2a and 2b). Related to this, a major advantage of the TGR assay is that it can be used to detect mutations in virtually any tissue including germ cells, while germ cell *Pig-a* assays only are in the beginning stages of development (see Section 10c).

- Because mutations are detected in cells that do not have DNA, it is challenging to characterize the mutations measured in the *Pig-a* assay (see Section 6). By contrast, analysis of mutations in TGR assays can be conducted on any tissue with DNA, and mutation characterization is straightforward for most TGR systems, whose small bacterial transgene mutational targets are readily sequenced (Lambert et al., 2005), including by using newer high-throughput methods (Beal et al., 2015).

4. Assay protocol

Protocol recommendations for conducting the assay were given in the IWGT report (Gollapudi et al., 2015; summarized in Table I). The report indicated that the recommendations were based mainly on assays conducted using erythrocytes from male rats. More recently, however, considerable data have appeared on conducting the assay in mice (Olsen et al., 2017; see Table VII), and several studies have been published demonstrating similar responses to test
substances in male and female rats (Chikura et al., 2014; Labash et al., 2015a,c) (see Section 7b). Thus, the available data indicate that protocol recommendations can be generalized for use with male and female rats and with male mice.

A general scheme for performing the *in vivo* Pig-a assay in rodents is provided in Figure 5. The in-life portion of the assay follows the protocol recommendations in TG 488 for the TGR gene mutation assay (OECD, 2013). Approximately 28-day repeat-dose treatments are recommended both to take advantage of the accumulation of the Pig-a mutational response with the number of treatments and to better integrate the assay with 28-day repeat-dose toxicity studies (OECD, 1998; OECD, 2009). However, many Pig-a assays have been performed with acute treatment protocols (*i.e.*, single treatments) or short-term repeat-dose treatments, generally treatments on three consecutive days (see Tables II, III, VI, and VII). Shorter treatments can be economically attractive, or necessary when only limited amounts of the test substance are available, and the IWGT report indicates that acute or short-term treatments are permissible with justification. For example, short-term dosing may be justified if adequate dose levels can be achieved or if it is desirable to integrate the Pig-a assay with short-term genotoxicity (*e.g.*, micronucleus or comet), toxicity, or other *in vivo* tests (*e.g.*, single-dose pharmacokinetics study).

*Pig-a* mutant frequencies have been analyzed using limiting-dilution cloning of spleen lymphocytes and proaerolysin as a selection agent (*e.g.*, Miura et al., 2008b; see Section 10c), but for the reasons stated in Section 1, *Pig-a* assays are almost always conducted by flow cytometric evaluation of erythrocytes from peripheral blood. Prior to flow cytometry, blood processing involves methods to reduce interference by other cells in the blood, either by enriching the sample for erythrocytes using density-gradient centrifugation (*e.g.*, Phonethepswath et al., 2010), or specifically labeling erythroid cells using fluorescent HIS49 antibody for rats (Dobrovolsky et al., 2010a; Kimoto et al., 2011a) or TER-119 for mice (Kimoto et al., 2011b). RETs are distinguished from mature RBCs, either by staining with a fluorescent nucleic acid dye that labels the residual RNA in RETs (*e.g.*, SYTO 13), or by using a fluorescently labeled antibody against the CD71 transferrin receptor, which is expressed in RETs but not in mature RBCs. Finally, cells that are wild-type for GPI anchor expression, and presumably
descendant from cells having a wild-type Pig-a gene, are identified by labeling the blood sample with fluorescently labelled antibodies to a highly expressed GPI-anchored protein molecule, usually CD59 for rats or CD24 for mice. Samples are analyzed with a flow cytometer for light scatter as well as for the fluorescent signals produced by the various fluorochromes used for labeling. Mutants are typically distinguished from wild-type cells based on the fluorescence produced by the antibody specific for GPI-anchored protein (low fluorescence, mutant; high fluorescence, wild-type). Gates are usually set for mutant and wild-type cells based on the position of a mutant-mimic sample (erythrocytes not reacted with the GPI-anchored protein antibody; e.g., Phonethepswath et al., 2010), and the number of events in the mutant and wild-type gates counted to estimate a mutant frequency. Interlaboratory trials have demonstrated the transferability, reproducibility and relative sensitivity of the basic flow cytometric Pig-a assay (Dertinger and Heflich, 2011; Kimoto et al., 2013) (a description of interlaboratory trials is presented in Part 2, Section 7).

Once the labeling methodologies were sufficiently optimized, it became clear that spontaneous in vivo erythrocyte Pig-a mutant frequencies were relatively low, often 1 x 10^{-6} or less. To avoid samples having mutant frequencies of ‘zero’, and thus compromising the power of the test to detect positives, the IWGT established a minimum number of cells that should be detected. Directly detecting an adequate number of cells for mutation using a flow cytometer, especially in the case of RETs which are 20-100-fold less abundant in peripheral blood than mature RBCs, makes analysis times unacceptably long (e.g., ≥15 min/sample). In order to overcome this problem, immunomagnetic enrichment procedures have been developed (Dertinger et al., 2011a; Kimoto et al., 2011a) as a practical solution to increasing the number of total RBCs and/or RETs interrogated for mutation in the assay.

The two major immunomagnetic enrichment protocols used for the Pig-a assay take different approaches to increase the number of mutants that are analyzed by the assay: one by reducing the number of wild-type erythrocytes in the sample, the other by increasing the number of RETs analyzed. The process devised by Dertinger and colleagues, commonly performed in North America and Europe and commercially known as the In Vivo MutaFlow® or HT (high
throughput) method (hereafter referred to as the MutaFlow method), involves enrichment for erythrocytes followed by labeling with phycoerythrin (PE) conjugated anti-CD59 (rats) or PE-conjugated anti-CD24 (mice) (Figure 6). A PE-conjugated antibody against platelets (CD61) also is incorporated into these assays to prevent platelets from being misread as mutant cells. The samples are then treated with immunomagnetic beads that bind to the PE fluorochrome and passed through a column in the presence of a strong magnet, so that the eluted cells are highly enriched for non-PE-labeled mutants (Figure 6). The cells are then stained with SYTO 13 to distinguished between RETs and mature RBCs and counting beads are added to estimate the total number of ‘cell equivalents’ analyzed. These enriched samples allow flow cytometric analysis of approximately $3 \times 10^6$ RETs and $>100 \times 10^6$ total RBCs per sample in a flow cytometric run of approximately 5 min. The MutaFlow method has been extensively evaluated as described by Gollapudi et al. (2015) and Raschke et al. (2016).

The method developed by Kimoto and colleagues, and commonly used in Japan, is known as the PIGRET assay. In the PIGRET assay, blood is first labeled with PE-conjugated anti-CD71, reacted with immunomagnetic beads recognizing PE, and using a powerful magnet, a sample enriched for RETs is collected (Figure 7). This sample is then reacted with fluorescent antibodies for erythroid cells (HIS49 for rats, TER-119 for mice) and a highly expressed GPI-anchored protein (CD59 for rats, CD24 for mice), and processed by flow cytometry to evaluate the frequency of Pig-a mutants in RETs (CD71-positive, erythroid-marker-positive, GPI-anchored-protein-negative). In addition, fluorescent antibodies for erythroid and GPI-anchored protein markers are reacted with blood samples without prior immunoseparation, and processed by flow cytometry to estimate a Pig-a mutant frequency in total RBCs. As conducted in the most recent Japanese Environmental Mutagen Society/Mammalian Mutagenicity Study group (JEMS/MMS) interlaboratory trial (Kimoto et al., 2016; see Part 2, Section 7a), this approach analyzes at least $1 \times 10^6$ RETs and $1 \times 10^6$ total RBCs for estimating mutant frequency.

To summarize the differences between the two approaches: 1) the cell numbers analyzed by the PIGRET approach are less than with the Mutaflow method, but within the IWGT guidelines; 2) in contrast to the MutaFlow method, the PIGRET analysis is conducted directly on RETs and
total RBCs, without the need for counting beads to estimate the total number of cells analyzed; and 3), the PIGRET protocol requires two labeling reactions and two flow cytometry analyses—one to score mutant RETs and one to score total RBCs, while the MutaFlow method utilizes one labelling reaction and two flow cytometric analyses. For the MutaFlow method, one of the flow analyses estimates total RETs and total RBCs based on pre-immunomagnetic samples, and the second counts the number of mutant RETs and mutant RBCs in the post-immunomagnetic separation samples.

a. Factors affecting Pig-a assay performance: treatment and sampling protocol and number of cells interrogated

To a certain extent, the IWGT protocol recommendations were based on theory, with a reliance on limited data from a small number of prototypical chemical substances. Table II shows results with 10 test substances, where there is some evidence that employing a longer repeat-dose treatment protocol and/or employing an enrichment technique that increases the number of cells analyzed for mutation affected the response in the assay. For 2-AAF and azathioprine, assays conducted using a single treatment were negative, while assays conducted with a 3- or 28-day treatment protocol were positive. With aflatoxin B1, assays conducted using 3-day and 15-day treatment protocols were negative, while aflatoxin B1 was positive using a 29-day repeat-dose treatment protocol. As for the number of erythrocytes interrogated for mutation, diethylnitrosamine and cisplatin were negative using low-throughput analysis techniques that did not use immunomagnetic enrichment, but positive when using an immunomagnetic enrichment protocol. While protocols not using immunomagnetic enrichment were largely negative for cyclophosphamide mutagenicity, and occasionally negative for melphalan, MMS, and urethane mutagenicity, protocols using immunomagnetic enrichment were consistently positive.

Thus, the IWGT workgroup recommendations for using repeat-dose, approximately 28-day treatments, which generally result in the animal being treated with a greater amount of test
substance than can be delivered in a 1- or 3-day treatment protocol, and the analysis of the
greatest number of erythrocytes practical, will increase assay performance. With current
technologies, the latter means employing an immunomagnetic enrichment technique that
facilitates examining a large number of erythrocytes. As relatively robust, genotoxic
carcinogens, like 2-AAF and aflatoxin B1, can be missed by using low throughput and/or short-
term treatment protocols, establishing a negative response in the assay with confidence will
require following these more stringent methods. One possible exception to these
generalizations is described for a study on 4-nitroquinoline-1-oxide (4NQO) that was reported
in two papers. 4NQO produced quantitatively higher responses when given as a short-term
treatment than an equal total treatment spread over 28 daily treatments (Stankowski et al.,
2011a; Roberts et al., 2016). Both short-term and longer-term treatments were positive, but
the magnitude of the response may have been greater with the short-term treatment.

While the IWGT recommended a single sampling time at Day 28-31 for both short-term and 28-
day treatment protocols, it encouraged the use of additional, optional sampling at earlier and
later times as an aid to data interpretation. The desirability of a later time point is based on the
expectation that positive responses in total RBCs may be most fully manifested at a later time
point, especially with longer repeat-dose treatment protocols. This is because Pig-a mutant
frequencies accumulate with multiple doses and are generally persistent (e.g., Miura et al.,
2009; Dertinger et al., 2010; Phonethepswath et al., 2010). The desirability of an earlier time
point is based on the expectation that positive total RBC responses will be preceded, or at least
accompanied by, increases in RET mutant frequencies. In addition to conforming to
erthropoiesis theory (Dobrovolsky et al., 2010b), this expectation has been confirmed in
numerous assays, most recently in the JEMS/MMS interlaboratory trial comparing responses in
the PIGRET and total RBC Pig-a assay (Kimoto et al., 2016). As an example of using these data
for interpreting responses, Dobrovolsky et al. (2016) discounted positive RBC responses in
occasional rats treated with acrylamide, in part because these responses were not preceded or
accompanied by similar increases in RET Pig-a mutant frequencies.
In addition to establishing the plausibility of a positive response, collecting total RBC data, in addition to RET data, can be useful in detecting a weak positive response (e.g., Khanal et al., 2018). This may be because the MutaFlow analysis method, in particular, interrogates 30-fold or more total RBCs for mutation than it does RETs, potentially increasing the precision with which RBC mutant frequencies are measured.

b. Option of preserving peripheral blood samples for later analysis

A significant logistical challenge associated with erythrocyte-based Pig-a assays has been the inability to store samples for more than several days before processing and flow cytometric analysis becomes necessary. Recently, however, effective whole blood freezing and thawing reagents have been successfully tested at several facilities (Avlasevich et al., 2019). This technical advance addresses several logistical considerations such as: 1) holding blood samples until a flow cytometer and trained technical staff are available; 2) accumulating samples so that longitudinal data can be collected on the same day using identical labeling reagents and instrument settings; and 3) using frozen samples from a genotoxicant-treated animal as a positive control rather than treating positive control animals for each study.

Aside from these considerations, perhaps the most consequential implication of effective freeze/thaw protocols is deferred decision-making. For example, if a decision is made to forgo Pig-a analysis in the conduct of a 28-day toxicology or other repeat-dose study, it may be advantageous to freeze and store whole blood samples at appropriate times point(s), for instance at the termination of the treatment phase. Whereas mutation analysis may not have been indicated based on data collected up to the time of study initiation, there may be occasions when data are subsequently generated that suggest the merit of assessing in vivo gene mutation potential. For example, observations such as hyperplasia, preneoplastic lesions, or even frank tumors could raise new concerns about possible carcinogenicity due to genotoxic mechanism(s). In these cases, rather
than setting up a new study and exposing additional animals, it would be highly advantageous to thaw archived blood samples to investigate in vivo genotoxicity. In this 3Rs-friendly scheme, no additional animals or other additional costly resources would be necessary.

c. Updated recommendations for conducting the Pig-a gene mutation assay

Based on recent experience, and the analysis in Sections 4a and 4b, the following relatively minor additions, clarifications, and alterations can be made to the IWGT protocol recommendations of Gollapudi et al. (2015):

- Assays may be conducted in mice or rats. Both male and female rats may be used. For all treatment schedules, it may be advantageous to perform an assay prior to the first treatment (e.g., 1 to 5 days before the first treatment) in order to identify animals having mutant frequencies outside of historical control limits so that they can be removed from the main study.

- When exposure occurs daily for several weeks, e.g., using a 28-day repeat-dose protocol,
  
  o Collect blood for assay within day(s) of cessation of the treatment (e.g., Day 28-31 of 28-day repeat-dose protocol). Data collected from testing diverse genotoxicants suggest this is sufficient time for adequate manifestation of mutant RET and mutant RBC responses.

  o Optionally, and when logistically feasible, there may be merit to conducting Pig-a analyses on blood samples collected from a 28-day repeat dose protocol at an additional, later time point. For instance, some experiments include satellite “recovery groups” to evaluate whether toxic effects are diminished or increased upon discontinuation of exposure. Such blood samples, typically collected between 2 to 4 weeks after cessation of exposure, represent an
opportunity to evaluate mutant RBC frequencies when a mutagen-induced effect is more fully manifested in the total RBC population.

- Optionally, and when logistically feasible, there may be merit to conducting an earlier blood sampling, for instance at Day 14 or 15 during a 28-day repeat-dose protocol. RET data from such samples, in particular, may be useful in supporting responses detected at the 28-31-day sampling point.

- When dosing occurs for more than 28 days (e.g., 90 days), there may be merit in conducting assays at approximately 28 days (e.g., at Day 28-31) in addition to at the cessation of treatment in order to mitigate the possible effect of clonal expansion on mutant frequency at extended sampling times (e.g., Mittelstaedt et al., 2019).

- When test substance exposure occurs for 1-3 consecutive days,

  - Sampling at two collection times is recommended:
    - Collect blood and perform assays approximately 1-2 weeks after cessation of treatment. Data collected from testing on diverse genotoxicants suggest that this is sufficient time for adequate manifestation of mutant RET responses, and in the case of a very few potent mutagens, for mutant RBC responses.
    - It is important to include a second blood collection time, one at approximately Day 30. This provides time for the peripheral blood pool to more completely turn over, which enhances the sensitivity of the mutant RBC endpoint. Based on the data accumulated to date and differences between mutant RET and mutant RBC manifestation times, scientific justification would be required to support a negative result from a short-term-treatment study based on only one post-treatment blood collection time-point.
Other treatment and harvest schedules may be scientifically justified, but as above, they should take into consideration the different manifestation times of mutant RET and mutant RBC responses.

5. Recommendations for reporting data, statistical analysis, and interpretation of results

The recommendations given in this section are based on those made in the IWGT Pig-a report (Gollapudi et al., 2015). In addition, they are supported by Pig-a data currently available and are aligned with the OECD Mammalian Erythrocyte Micronucleus Test Guideline (TG 474; OECD, 2016a). Although not all Pig-a data in the literature have been reported and analyzed in this way, the following recommendations are provided based on a current understanding of how Pig-a gene mutation data are generated and used.

Individual animal data should be reported or at least made available upon request. These data should include: the number of cells evaluated, the number of mutant cells observed, and the frequency of mutant cells (typically reported as number per million). Furthermore, group means for like-treated animals should be reported, and covariates such as blood sampling time, animal sex, etc. should be clearly indicated. A measure of statistical significance (e.g., p-values) and within group variation (e.g., standard deviation, standard error of the mean, and/or confidence intervals) should be reported. These can be helpful for interpreting the biological relevance of a statistical finding.

Data from Pig-a experiments are generally evaluated with the intention of determining whether or not a test substance affects the observed proportions of mutant RETs among all RETs and mutant RBCs among all RBCs. Such analyses should be accomplished by appropriate statistical methods, using the animal as the experimental unit. The statistical method(s) employed should be indicated. Providing that all the study acceptability criteria are fulfilled (Table I and Section 4; see also Section 7b), a test substance is considered clearly positive if:
A) At least one of the treatment groups exhibits a statistically significant increase in the frequency of Pig-a mutant phenotype cells compared with the concurrent negative control; 

B) This increase is dose-related at least at one sampling time when evaluated with an appropriate trend test; and 

C) Mutant frequencies exceed the distribution of the historical negative control data.

For Criterion A, appropriate statistical methods should be used to compare mutant frequencies in treated groups with the concurrent vehicle control group. It is important to ensure that, whether a parametric or non-parametric statistical method is used, the underlying assumptions for the method are reviewed. In particular, it is important to consider the residual distribution and variance-homogeneity.

An appropriate transformation of the mutant frequencies may be used to reduce non-normality and/or heterogeneous variances. One example is a logarithmic transformation, using $\log_{10}$ or the natural log ($\ln$). The complication of mutant frequencies of zero being observed occasionally can be overcome in the case of log transformation by adding a small value (such as 1.0, 0.1 or 0.01) to each mutant cell frequency (expressed as mutants x $10^{-6}$).

When the objective of the analysis is to declare a comparison significant, the tests should be performed with a predefined alpha (α) value, and should generally be carried out as one-sided tests, with the expectation of detecting a treatment-related increase in mutant cells. If pairwise comparisons of treatment group means are carried out using methods which correct for multiple comparisons, then the p-values and confidence intervals should be adjusted appropriately. The method of multiple comparison must be clearly stated to allow this to be considered in any assessment of the results.

In the case of Criterion B, at least three treated dose groups should be analyzed to test for a dose-related increase. In general, a trend test can be applied to such data. These often test for a linear trend and care must be taken in interpreting the test when the highest dose level of a mutagen/genotoxicant does not result in the highest mutant cell frequency (e.g., 2AAF;
Dertinger et al., 2012). Care is therefore needed in the interpretation of results when the dose-response is non-linear, particularly when there is a downturn in mutant frequency at a high dose or a non-monotonic response.

Criterion C is used for assurance that a statistically significant increase in mutant cell frequency has biological relevance. The experimental results are compared with an appropriate historical negative control distribution (*e.g.*, 95% Tolerance Interval), assuming that the laboratory performing the test has a good historical database (see Figure 8 for an example). The emphasis is on the distribution of the historical negative control data rather than the range (as determined by maximum and minimum values observed in the historical control data). The historical control distribution of mutant cell frequencies allows a comparison of data from the current experiment with metrics derived from the historical negative control database such as tolerance intervals, control limits and related values. This approach helps understand the likelihood of the observation of high mutant cell frequency value(s) instead of employing simple and less informative within- or out-of-range comparisons.

In some studies, background or baseline measures may be available, where blood samples were taken before the compound was applied. It may be possible to include this covariate data into the analysis of the experiment.

Experimental designs that include multiple time points may be used for conducting longitudinal analysis. As there are expectations regarding the manifestation of mutant RET and mutant total RBC frequencies with time, such analysis may aid in establishing the biological relevance as well as statistical significance of a response. These experimental designs may involve dose groups consisting of sub-groups of animals being examined at different time points or, alternatively, the animals in the dose groups providing blood samples on multiple days. Only a few studies analyzing longitudinal data have appeared thus far in the literature (*e.g.*, Dertinger et al., 2010; Dobrovolsky et al., 2016; Igl et al., 2018) so that consensus recommendations cannot be made at this time. However, designs with both dose and time as factors need to be analyzed by considering whether the blood samples are from independent animals or are measures from
the same animal on different days. As is the case with all Pig-a studies, these studies need careful design to ensure that pre-specified effect sizes considered biologically relevant can be detected, if they exist, with high probability (i.e., the design has sufficient power). Analysis of longitudinal data was used in the evaluation of vehicle control mutant frequencies as explained in Section 7b.

Pig-a assays collect data on both mutant RET and mutant total RBC frequencies and these endpoints have usually been analyzed separately in the literature. Statistical approaches capable of considering them simultaneously exist and these approaches may be useful in certain instances.

A Pig-a experiment that does not meet any of the criteria (A, B and C) is a clearly negative test result. Negative results indicate that, under the test conditions, the test substance does not cause gene mutation to hematopoietic cells of the test species. However, a negative result is only considered relevant when evidence is provided that the bone marrow has been exposed to the test substance. With toxic compounds, direct evidence of bone marrow exposure is provided by a depression of the immature to mature erythrocyte ratio; in the absence of toxicity to the erythropoietic system, indirect evidence (measurement of the plasma or blood levels of the test substance and its metabolites) can be useful for documenting exposure (a discussion of this topic is provided in EFSA [2017]). Alternatively, ADME (absorption, distribution, metabolism, and excretion) data, obtained in an independent study using the same route and same species can be used to demonstrate bone marrow exposure. It is currently unresolved, however, as to what plasma/blood levels are sufficient to support a negative result, especially when the negative in vivo test was performed as a follow-up to an in vitro positive. In the case of intravenous administration, evidence of exposure is not needed.

There is no requirement for verification of a clear positive or clear negative response. In experiments where the results are neither clearly negative nor positive, and to assist in establishing the biological relevance of the result (e.g., a weak or borderline increase), the data
should be evaluated by expert judgment and/or further investigations of the existing experiments. In some cases, analyzing more cells or performing a repeat experiment using modified experimental conditions could be useful. In rare cases, even after further investigations, the data may preclude making a conclusion that the test substance produces either a clear positive or negative result, and the experimental result will be concluded to be equivocal.

6. Confirming that Pig-a mutation is responsible for the Pig-a mutant phenotype

There are several lines of evidence indicating that the phenotype measured in the rodent Pig-a assay, loss of GPI-anchored proteins, is due to Pig-a mutation. While there are rare instances associated with human diseases and assays conducted with in vitro cell cultures where mutations in another GPI-biosynthesis pathway gene caused GPI-anchor deficiency, there are no data indicating that this occurs in the rodent Pig-a gene mutation assay. As indicated in Section 1c, the mutations that are responsible for the phenotype measured in the erythrocyte Pig-a assay are induced primarily in nucleated erythroid precursor cells in the bone marrow. Thus, the most direct way of demonstrating that Pig-a mutations are being detected by the Pig-a assay is to evaluate GPI-marker-deficient bone marrow erythroid cells for Pig-a mutations. Other lines of evidence that relate to the mutational basis for the Pig-a phenotype are presented in approximate order of relevance.

a. Evidence that GPI-marker-deficient bone marrow erythroid cells and granulocytes in rats and mice contain mutations in the Pig-a gene

i) Mouse studies

Kimoto et al. (2011b) treated mice with a single dose of 100 mg/kg ENU and measured Pig-a mutant frequencies in bone marrow erythroid cells and total RBCs from peripheral blood over a period of 4 weeks. Consistent with mutations being induced in the nucleated bone marrow erythroid cells and these cells differentiating and transiting, with time, into cells found in
peripheral blood, mutant frequencies increased more rapidly in the bone marrow cells than in the peripheral blood RBCs. Pools of mutant bone marrow erythroids were prepared from 7 treated mice by flow cytometric sorting. Pig-a cDNAs were synthesized from RNA recovered from each pool; the cDNAs were amplified by PCR and cloned into plasmid vectors. 18 of 25 cloned cDNA pools produced recombinant clones that were successfully analyzed by Sanger sequencing. All recombinants contained Pig-a mutations, with the same mutation detected multiple times in clones from four of the mice. The mutations were either basepair substitutions (70% of which were at A:T, with the mutated T on the nontranscribed DNA strand), or exon deletions that could have been produced by basepair substitutions. The findings demonstrated that all the Pig-a sequences that could be recovered from phenotypically mutant erythroid cells contained Pig-a mutations. In addition, the types of mutations that were detected were consistent with those expected to be induced by ENU in a transcribed mammalian gene.

ii) Rat studies

A series of studies have been conducted to evaluate mutation induction in the Pig-a gene of bone marrow erythroid cells and granulocytes of rats. Rats were treated with either vehicle or three doses of 40 mg/kg ENU given every other day and assayed for Pig-a mutation in nucleated bone marrow erythroid cells 10 or 12 days following the last dose (on Day 17 or 19; Revollo et al., 2018). Bone marrow was stained with Hoechst33342, to identify DNA-containing cells, allophycocyanin (APC) conjugated anti-CD71, to identify early erythroid cells, and PE anti-CD59, to identify cells with GPI-anchored protein markers. The average erythroid cell CD59-deficient mutant frequency in 6 vehicle control rats was $17 \times 10^{-6}$, while the mutant frequency increased to an average of $406 \times 10^{-6}$ in 6 ENU-treated rats. Phenotypically mutant cells were sorted from each of the ENU-treated rats and a next generation sequencing (NGS) method called MAML (Mutation Analysis with Multiplexed Libraries; Revollo et al., 2017a) was used to identify Pig-a mutations in pools of either 5 cells or 10 genome equivalents. A total of 116 different Pig-a mutations were detected in the mutant erythroid cells, with one mutation found in three different rats and 16 mutations found in two different rats; the remaining 99 mutations
were found in only one rat. The spectrum of mutations was typical of that produced by ENU in a mammalian gene (e.g., Mittelstaedt et al., 1995): 73% of mutations occurred at A:T basepairs, with the mutated T located on the nontranscribed DNA strand of the gene; T→A transversion was the most frequent mutation followed by T→C transition.

*Pig-a* mutant frequencies also were measured in bone marrow granulocytes. These studies were performed 1) to provide additional evidence that *Pig-a* mutation was responsible for the *Pig-a* mutant phenotype using a cell type closely related to erythroid cells; and 2) to gain insight into the timing of mutation induction in hematopoietic cells, *i.e.*, in multipotent or myeloid precursor cells before commitment to the erythroid and granulocyte lineages and/or in lineage-committed cells (see Figure 4). Rats were treated with ENU and bone marrow harvested as described for the erythroid studies outlined above, with some rats being evaluated both for bone marrow erythroid and granulocyte mutation (Dad et al., 2018). Mutant granulocytes were quantified by flow cytometry using a combination of APC anti-CD11b, for identifying granulocytes, and PE anti-CD48, for identifying cells with GPI-anchored protein markers. CD48 is a GPI-linked marker that is highly expressed in granulocytes and is much more useful than CD59 for identifying GPI-deficient granulocytes. CD48-deficient mutant frequencies increased from an average of $8.4 \times 10^{-6}$ in 8 vehicle control rats to $567.1 \times 10^{-6}$ in 8 ENU-treated rats. Mutants were sequenced by MAML in pools of 25 cells; a total of 133 mutations were identified in cells from four of the ENU-treated rats (granulocytes from the other four treated rats were not sequenced). As was the case with the erythroid ENU-induced mutation spectrum, most of the mutations were at A:T basepairs, with the mutated T located on the nontranscribed DNA strand. The most common mutations were T→A transversion, followed by T→C transition.

Both granulocyte and erythroid mutants were analyzed for *Pig-a* mutations in bone marrow harvested from three of the 8 ENU-treated rats, yielding a total of 61 granulocyte mutations and 69 erythroid cell mutations. Although the overall mutation spectra for granulocytes and erythroid cells were very similar, only three mutations were common to the two cell populations from a single rat.
iii.) Summary of bone marrow Pig-a studies

The observations made in these studies indicate that Pig-a mutations can be found in GPI-anchored-marker-deficient erythroid cells isolated from rat and mouse bone marrow, thus confirming that the erythrocyte Pig-a assay is measuring Pig-a mutation. The rat observations also shed light on the source of the mutations. Early after treatment, a variety of mutations can be detected in the erythroid cells and granulocytes of mutagen-treated rats, indicating that the mutant phenotype cells were caused by many independent mutations. In addition, 10-12 days following ENU treatment there was little overlap between the mutations detected in bone marrow erythroid cells and granulocytes from the same rat. This suggests that a small number of Pig-a mutations may have been induced in the bone marrow precursors that produce both erythroid and granulocyte cells, but that many of the mutations detected early after treatment are induced in cells already committed to be either erythroid cells or granulocytes.

Section 4b recommends that peripheral blood be assayed for Pig-a mutation at least twice for 1-3-day treatments, at Days 7-14 and at approximately Day 30. For 28-day repeat dose treatments, it is recommended that mutants be evaluated at least once, at approximately Day 30. The findings from these bone marrow mutation studies imply that, at these sampling times, the mutant phenotype measured in peripheral blood erythrocytes will be due to a variety of mutations that were induced independently, primarily in bone marrow progenitor cells committed to the erythroid developmental pathway.

b. Rat bone marrow cells with GPI-anchored protein markers contain wild-type Pig-a genes

In the experiment performed by Revollo et al. (2018) phenotypic wild-type bone marrow erythroid cells (i.e., cells positive for the CD59 GPI-anchored protein) also were sorted for Pig-a sequence analysis by MAML. The wild-type cells that were examined contained no Pig-a mutations at all. The same result was found for pools of wild-type granulocytes in the studies conducted by Dad and colleagues (2018): no Pig-a mutations were found in phenotypically
wild-type bone marrow granulocytes. These results serve to confirm that analysis of GPI-anchored protein status (i.e., performing the flow cytometric erythrocyte PIG-a assay) distinguishes between cells having mutant and wild-type PIG-a genes.

c. GPI-deficient cells in PNH patients are almost always associated with PIG-A mutations

As described in Section 1, classical hemolytic PNH almost always is associated with the expansion of a mutant HSC clone or occasionally a limited number of mutant HSC clones, each clone containing a PIG-A mutation (Nafa et al., 1995, 1998; Rosse, 1997; Brodsky and Hu, 2006; Brodsky, 2014). The mutations typically result in GPI deficiency in a substantial fraction of all hematopoietic cell lineages descendent from HSCs, including erythroid cells (Brodsky and Hu, 2006). PIG-A mutant clones also can be found in patients that develop PNH following aplastic anemia, although the development of GPI-deficient erythrocytes sometimes is not as apparent in these patients (Mortazavi et al., 2003).

The PIG-A mutations in PNH may have originated as spontaneous mutations, and although several possibilities exist accounting for their expansion (e.g., Luzzatto and Besseler, 1996; O’Keefe et al., 2011; Sugimori et al., 2012; Luzzatto, 2016), why particular mutations develop into expanded PNH clones is not totally clear (Brodsky and Hu, 2006). However, the important observation is that PNH, with rare exception, is due to GPI-deficient hematopoietic cells that harbor mutations in the PIG-A gene. The rare exceptions, described in Section 1, include patients with other mutations that either result in anchor deficiency or in a deficiency in an anchored protein (Krawitz et al., 2013; Brodsky, 2014). Thus, the PNH literature indicates that GPI-deficiency and GPI-anchored-protein deficiency is due to mutation, almost always to mutation in the PIG-A gene (Brodsky and Hu, 2006). It should be noted that even though hemolysis is a hallmark of the classical disease, the association of PNH with PIG-A mutation is almost entirely based on mutations detected in granulocytes and lymphocytes, and rarely with bone marrow erythroid cells (e.g., Nafa et al., 1995, 1998; Mortazavi et al., 2003; Hu et al., 2005).
d. Analysis of Pig-a mRNA in mutant RETs from mice

Since RETs contain RNA, they may contain Pig-a mRNA descendant from mutated nucleated erythroid precursors that can be analyzed for mutations. Even though mRNA analysis doesn’t evaluate the Pig-a gene directly, mRNA sequencing potentially can be conducted with a cell type, RETs, that is used for detecting mutants in the assay. Unfortunately, efforts to amplify Pig-a mRNA by RT-PCR of RNA from the mutant RETs of rats and mice largely have been unsuccessful (D.J. Roberts, V.N. Dobrovolsky, J.R. Revollo, R.A. Mittelstaedt personal communications).

A single report (Byrne et al., 2014), however, describes the detection of Pig-a mutations in GPI-deficient RETs from mice. This study used the PIGRET assay to evaluate genomic instability in a transgenic myelodysplastic syndrome (MDS) mouse model. The results indicated higher Pig-a mutant frequencies in peripheral blood RETs from MDS mice than from wild-type mice. Mutations were analyzed by sorting 2000 mutant RETs from each of three MDS mice (analyses were not done with mutants from wild-type mice), extracting RNA from the three mutant pools, and amplifying Pig-a cDNAs by RT-PCR. Amplified cDNAs were cloned into a plasmid vector and four recombinant clones from each mouse sequenced. All 12 recombinant clones contained Pig-a mutations, with all but one of the clones containing multiple mutations—as many as 7 mutations in a single clone. For one mouse, the one to four mutations in each of the four sequenced clones were all different from one another; another mouse had the same basepair substitution in three of the four clones (along with several unique mutations); while in the third mouse, all clones contained four identical mutations. Although mutant frequencies in the transgenic mice were quite high, approaching $1 \times 10^{-2}$, it is difficult to rationalize the existence of multiple mutations in 11 of the 12 clones that were sequenced.
e. Evidence that GPI-deficient rat T lymphocytes contain Pig-a mutations

Each mammal contains several thousand HSCs from which the various cell types in peripheral blood are derived (Abkowitz et al., 2002; see Figure 4). Thus, PNH, a disease that is caused by PIG-A mutation in HSCs, is characterized by large mutant clones in the differentiated cell types formed by HSCs, including lymphocytes, granulocytes, and erythrocytes (e.g., Mortazavi et al., 2003). As indicated in Section 6c, even though hemolytic lysis is a major pathogenic manifestation of classical PNH, the association between PIG-A mutation and the disease largely rests on mutation analyses conducted with nucleated hematopoietic cells, most usually granulocytes. Mutation induction in the Pig-a assay may not be perfectly analogous to PIG-A mutation in PNH; however, the persistence of the mutant phenotype and the bone marrow sequencing data described in Section 6a indicate that mutations in early precursor cells likely are involved in the responses. Thus, evaluating mutation induction in nucleated myeloid and lymphoid cells has relevance to mutation induction in the erythrocyte Pig-a assay.

In this regard, several studies have been conducted to evaluate Pig-a mutation in GPI-deficient and GPI-anchored-protein-deficient lymphocytes from rats. Miura et al. (2008b; 2011) treated rats acutely with ENU, and 4 weeks later, assayed for GPI-deficient splenic lymphocytes using proaerolysin selection. Pig-a cDNA prepared from the mutant clones was used for sequence analysis. Including putative splicing mutations, approximately 70% of the mutant clones contained clear alterations of Pig-a cDNA sequence, and about 80% of these changes were clearly independent. A total of 76 Pig-a mutations were identified in 66 clones analyzed. Most mutations were basepair substitutions at A:T, with the mutated T on the nontranscribed DNA strand (36 of 38 mutations at A:T). The spectrum was consistent with the types of mutations produced by an ethylating agent in a transcribed mammalian cell gene, and indeed was very similar to a published spectrum for ENU-induced mutation in the rat lymphocyte Hprt gene (Mittelstaedt et al., 1995).

The Pig-a assay, however, is normally conducted by identifying the mutant phenotype using flow cytometric evaluation of GPI-anchored proteins, not by proaerolysin selection of mutants.
Flow cytometric analysis also has the potential to speed up molecular analysis as it rapidly identifies mutants for analysis. Revollo and colleagues (2015) flow sorted rat spleen lymphocytes deficient in the GPI-anchored marker CD48 from ENU-treated rats, expanded the sorted cells into clones, and then performed Pig-a cDNA sequencing on RNA extracted from the mutant clones. In this study, the rats were treated with ENU, and analyzed for mutant lymphocytes in small groups (to accommodate flow sorting) four to 8 weeks later. Approximately 88% of the sorted mutant clones contained Pig-a mutations, and at least 87% of the mutations were induced independently. The mutation spectrum was characteristic of that described for ENU above: a predominance of basepair substitutions at A:T, with the mutated T on the nontranscribed strand of DNA. This spectrum was very similar to the ENU-induced Pig-a mutant spectrum induced in proaerolysin-selected lymphocyte mutants (Miura et al., 2011) and in rat lymphocyte Hprt mutants (Mittelstaedt et al., 1995). To increase the efficiency of the analysis, Revollo et al. also used massively parallel DNA sequencing to analyze a pool of cDNAs from 64 of the clones derived from the flow-sorted mutants. The distribution (spectrum) of mutations using this approach was virtually identical to that derived by clone-by-clone Sanger sequencing analysis.

Similar mutation analysis, clone-by-clone analysis using Sanger sequencing and NGS sequencing on a pool of mutant cDNAs, produced a distinct Pig-a mutation spectrum for flow-sorted lymphocyte mutants from 7,12-dimethylbenz[a]anthracene (DMBA) treated rats. Rats were treated with DMBA and assayed for lymphocyte mutants 6 weeks later. In this case, approximately 78% of mutants had Pig-a mutations, at least 60% of which contained independent basepair substitutions (other mutants contained large deletions and indels). The spectrum was dominated by basepair substitutions at A:T with the mutated A on the nontranscribed DNA strand (Dobrovolsky et al., 2015; Revollo et al., 2016). The Pig-a mutation spectrum from DMBA-treated rats also was very similar to Hprt lymphocyte mutation spectra for DMBA-treated rats that were reported previously (Heflich et al., 1996; Mittelstaedt et al., 1998). Analysis of flow-sorted pools of Pig-a mutant T cells from procarbazine-treated rats
indicated that at least 75% had *Pig-a* mutations, with the mutations caused by adducts with dA (Revollo et al., 2017a).

Like the erythroid cell and granulocyte mutation spectra that were detected in bone marrow soon after ENU treatment (Section 6a), the lymphocyte spectra consisted of many independent (different) mutations. For instance, of the 249 mutant lymphocyte clones from DMBA-treated rats that were analyzed by Dobrovolsky et al. (2015), 197 contained mutations in the *Pig-a* gene, and 117 of the basepair substitution mutations were unique. Unlike the bone marrow mutants, the lymphocyte mutations were analyzed in the peripheral circulation (spleen), and in a long-lived cell (T lymphocytes), whose mutations may have reflected both the few mutations that potentially were induced by the treatment in early precursor cells plus mutations induced in the intermediate/progenitor cells that had transited from the bone marrow (see Section 1c).

f. Evidence that *Pig-a* mutations cause GPI deficiency in mammalian cells *in vitro*

As indicated in Section 10d, several *in vitro* gene mutation assays have been devised that measure deficiency in GPI or GPI-anchored protein markers as an endpoint. Sequence analysis of the mutants detected in these assays has shown that mutation in a GPI-biosynthesis gene is the cause of the phenotype. However, in some biological systems mutations in genes other than *Pig-a* can contribute to the GPI-deficient phenotype. The reasons why this is the case generally support the model for how the *Pig-a* assay detects mutation *in vivo*.

An assay for gene mutation in chicken DT-40 cells identifies mutants by clonal selection in the presence of proaerolysin (Nakamura et al., 2012). Analysis of 18 MMS-induced DT-40 mutants found that all had mutations in the *Pig-o* gene. In chickens, *Pig-a* is an autosomal gene, while chickens have a single copy of *Pig-o* on the Z chromosome. Thus, the same logic that is used to indicate that mutations in the *Pig-a* gene are the most likely cause of GPI-deficiency in mammals can be used to argue that mutations in *Pig-o* are the likely cause of GPI deficiency in DT-40 cells.
Assays for detecting GPI-deficient mutants in human lymphobastoid TK6 cells have been described by several researchers. Nicklas et al. (2015) were the first to report that PIG-L was heterozygous in these cells, so that a ‘PIG-A assay’ based on detecting GPI-anchor deficiency actually detects mutants that are mutated in either PIG-A or PIG-L. Sequence analysis indicated that almost all spontaneous proaerolysin-resistant mutants (59 out of 60 analyzed) contained mutations in the PIG-L gene, and none in the PIG-A gene, while mutants induced by EMS were almost equally divided between those with PIG-A and PIG-L mutations. These general observations were subsequently confirmed in sequencing studies conducted by Krüger et al. (2016).

Although the genome of mouse L5178Y/Tk\textsuperscript{+/−} cells has literally millions of sequence alterations, both large and small, relative to the mouse genome (Sawyer et al., 2006; McKinzie and Revollo, 2017), Pig-a mutants, determined by both proaerolysin resistance and flow cytometric identification of cells deficient in GPI-anchored proteins, almost always contain mutations in the Pig-a gene. Wang et al. (2018) found that 38 of 41 spontaneous mutants, 18 of 18 ENU-induced mutants, and 27 of 27 B[a]P-induced mutants contained mutations in the Pig-a gene. In subsequent studies, Revollo et al. (2017b) found that all 81 of the B[a]P-induced mutants that they evaluated contained mutations in the Pig-a gene, while Bemis et al. (2018) found all 38 of their EMS-induced Pig-a mutants contained Pig-a mutations. The induced mutants contained mainly Pig-a basepair substitution mutations reflecting the mutational specificity of the test substance, while mutants from negative control cultures contained a high proportion of frameshifts and large deletions.

Taken together, these studies indicate that the Pig-a assay phenotype, loss of GPI-anchored proteins, is caused by mutation. In mammalian cells where Pig-a is the only single-copy GPI-biosynthesis gene (\textit{i.e.}, in L5178Y/Tk\textsuperscript{+/−} cells), the phenotype is caused by mutation of the Pig-a gene.
g. Evidence that Pig-a gene mutation responses are consistent with TGR and endogenous reporter gene mutation responses in the same animals

Pig-a mutant frequencies have been assayed in conjunction with transgene mutant frequencies in studies conducted with TGRs (Table IV). The rationale for this combination of assays is that the Pig-a assay serves as a rapid and sensitive approach for determining systemic (or bone marrow) mutation responses while the transgenes can be used to evaluate tissue-specific gene mutation. In most cases, the Pig-a and tissue-specific transgene responses were similar: EMS, BaP, ENU, and procarbazine were positive in the Pig-a and transgenic assays, while carbon nanotubes, 3-monochloropropane-1,2-diol and various ester conjugates of 3-monochloropropane-1,2-diol, dichloromethane, dichloropropane, and TiO₂ nanoparticles were negative in both types of assays. The discordant responses are for a mixture of dichloromethane and dichloropropane and for 4NQO where these compounds tested negative for Pig-a mutation and positive in the liver of gpt-delta mice (Horibata et al., 2013; Suzuki et al., 2014). These ‘discordant’ responses, however, are only superficially discordant, and consistent with the known tissue specificity of the Pig-a assay. In the Horibata et al. study, 4NQO was also tested in the bone marrow gpt assay, where it was negative. Thus, these data are consistent with both assays detecting a similar response, i.e., gene mutation, with the Pig-a assay being limited to mutations induced in the bone marrow. Note that many of these negative Pig-a responses were generated using protocols that were judged by the team reviewing Pig-a data for the RPA as insufficient to support a negative call (see Section 7b), and were classified as Inconclusive (I) (see Tables VI, VII, and VIII).

Hprt lymphocyte mutation assays also have been paired with the Pig-a assay in in vivo gene mutation studies. Both Pig-a and Hprt are endogenous genes residing on the X chromosome, and assays for mutation in both genes are conducted using hematopoietic cells (i.e., erythrocytes and T lymphocytes), so it is anticipated that mutational responses in the two assays will be similar. That has proven to be the case in a limited number of studies conducted with ENU and BaP (Bhalli et al., 2011b; Cammerer et al., 2011). As with the case of the TGR-Pig-
a data, these results are consistent with the Pig-a and Hprt assays both measuring gene mutation.

h. Evidence that animals with genomic instability have elevated Pig-a mutant frequencies

If the Pig-a assay is responding to mutation, then the expectation is that mutant frequencies will increase in mice with a mutator phenotype or genomic instability. Byrne et al. (2014) detected increased erythrocyte (as well as T cell and monocyte) Pig-a mutant frequencies in a mouse model for MDS, a HSC/progenitor cell disease associated with increased levels of genomic instability. Pig-a mutant frequencies in RETs increased with the age of the mice and were approximately 50-fold greater in MDS than wild-type mice. RT-PCR analysis of the mutant RETs revealed multiple Pig-a mutations (see Section 6d). MDS mouse Pig-a mutant frequencies were also greater in T cells and monocytes, but the differences with wild-type mice were not nearly as great as with erythrocytes.

Graupner et al. (2014, 2015, 2016) reported on Pig-a assays conducted in Ogg knockout, heterozygous, and wild-type mice. Ogg is involved in the repair of oxidative lesions and the expectation is that decreased lesion repair will result in increased levels of DNA damage and mutation. Higher levels of DNA damage were detected in nucleated blood cells from Ogg mice; however, higher Pig-a mutant frequencies were not detected. These studies, however, were compromised by the use of an early mouse Pig-a mutant analysis protocol that is now known to result in the lysis (and thus loss) of GPI-deficient RETs. These results, therefore, should be treated with caution.

i. Lack of evidence that loss of GPI anchors and GPI-anchored proteins is due to something other than mutation, e.g., epigenetic gene silencing.

An early observation in patients treated with CAMPATH-1H for lymphoid malignancies was the emergence of large clones of GPI-anchor-deficient cells (Hertenstein et al., 1995; Taylor et al.,
CAMPATH-1H is a humanized monoclonal antibody against CD52, a GPI-anchored protein that is expressed in differentiated lymphoid cells, but not in stem cells; importantly, CAMPATH-1H is unlikely to be genotoxic. The expanded cells sometimes contained PIG-A mutations, but sometimes no mutation could be found, an observation that planted the idea that Pig-a mutant cells could, under the right conditions, be caused by non-genotoxic mechanisms. Further investigation with more sensitive mutation analysis tools, however, indicated that the CAMPATH-1H treatment most likely results in in vivo selection of the progeny of pre-existing PIG-A mutant HSCs (Rawstron et al., 1999), producing large increases in PIG-A mutant frequency in peripheral blood.

As Pig-a is an endogenous gene on the X chromosome, and proper gene expression is necessary for the assay to correctly discriminate between mutant and wild-type cells, aberrant gene silencing has been cited as possible mechanism for generating GPI-deficient cells. This possibility was discussed in the IWGT Pig-a report (Gollapudi et al., 2015), but without any evidence to support such a mechanism. Since then, studies have been conducted comparing Pig-a mutant frequencies in male and female rats (Chikura et al., 2014; Labash et al., 2015a,c). Because one of the X chromosomes in females is inactivated during development (i.e., by lyonization), it might be hypothesized that inappropriate inactivation (methylation) might occasionally inactivate both Pig-a alleles, which might elevate mutant frequencies. Even though differences in metabolism, weight and %RETs can occur between males and females, there is no evidence that sex affects induced Pig-a mutant frequencies (Chikura et al., 2014; Labash et al, 2015a,c). Inappropriate X-chromosome inactivation might affect spontaneous mutant frequencies with higher mutant frequencies occurring in females than in males. However, there also is no evidence that male and female spontaneous mutant frequencies differ (Labash et al., 2015a).

Aberrant gene silencing may also affect Pig-a responses when testing substances known to modulate endogenous methylation. This may be seen most clearly with tests conducted with substances not expected to induce gene mutation but that are known to hypermethylate DNA. Hydroxyurea is one of several compounds known to hypermethylate DNA (Townsend et al.,
20 May 2019

2017), a process that might be hypothesized to silence the *Pig-a* gene and increase *Pig-a* mutant frequencies. However, in independent studies conducted in two laboratories, using the most sensitive immunomagnetic enrichment technologies and up to 28-day treatments with up to MTD doses, rats exposed to hydroxyurea had no increase in *Pig-a* mutant frequency, although hydroxyurea treatment did increase the frequency of micronuclei (Dertinger et al., 2012; Adachi et al., 2016). There is also evidence that furan, phenobarbital, and caffeic acid alter endogenous methylation (both hypomethylation and hypermethylation) (Watson and Goodman, 2002; Lee and Zhu, 2006; Phillips and Goodman, 2008; de Conti et al., 2014, 2016). All these compounds have tested negative (or negative but inconclusive) in the *Pig-a* assay (see Tables VI, VII and VIII), although caffeic acid was positive for peripheral blood micronucleus induction (Bhalli et al., 2016).

Dietary folate deficiency might lead to uracil incorporation and hypomethylation of DNA. Hypomethylation could decrease *Pig-a* mutant frequencies by causing expression of silenced genes; the opposite is also conceivable, where hypomethylation could lead to an increase in *Pig-a* mutant frequency. RET and total RBC *Pig-a* mutant frequencies did not differ in mice fed a folate-deficient diet and mice fed a control diet; however, total RBC mutant frequencies for folate-deficient mice were greater than for folate-supplemented mice (MacFarlane et al., 2015). Thus there was an effect but the differences in mutant frequency were quite small, the data in some groups quite variable, and similar differences were not seen for RET mutant frequencies.

In addition to DNA methylation, various chemical marks (e.g., methylation and acetylation) on histones bound at gene regulatory regions are important for controlling transcriptional status. Thus, aberrant gene silencing might also be mediated by chemicals that modulate histone modifications. The only chemical known to exert effects on histone methylation or acetylation evaluated in the *Pig-a* assay is caffeic acid, which as mentioned above, tested negative for *Pig-a* mutation.

Thus, to date, there is no convincing experimental evidence for epigenetic silencing as a mechanism for producing GPI-anchored-protein-deficient cells.
7. Evaluation of Pig-a assay results

a. Identification and cataloging Pig-a testing data

Pig-a assay data were identified in two ways: 1) in studies reported in the published, peer-reviewed literature, and 2) data from a variety of less formal routes, e.g., posters, platform presentations, and personal communications. In all cases efforts were made to secure the original data by directly contacting the scientists that were involved with the studies. The responsible scientists were asked to populate a spread sheet (Excel file) that was developed by Stephen D. Dertinger (Litron Laboratories), Takafumi Kimoto (Teijin Pharma), and Vasily N. Dobrovolsky (USFDA/NCTR), with important help from Paul White (Health Canada), George Johnson (Swansea University, UK), and Wout Slob (RIVM, The Netherlands). Completed spread sheets contain animal-by-animal and sampling-point-by-sampling-point data derived from the study, and include details about the test article, vehicle, treatment protocol and analysis methods, the results, observations of associated toxicity, other assays conducted as part of the study, conclusions about responses, and any reports generated that describe the findings. These spread sheets were collected and curated by a group at the USFDA/NCTR that included L. Patrice McDaniel, Roberta A. Mittelstaedt, and Jennifer M. Shemansky. As part of this process, statistical evaluations were performed on the data, as necessary, for confirming or establishing the significance of test substance responses and the mutagenicity of vehicle controls in the assay (see Section 7b). The curated data spread sheets then were transferred to a group at the University of Maryland, School of Pharmacy (Baltimore USA) that included Christopher Klimas and James Polli. This group developed the design and search functions of a public website on which the data files were posted. This website can be accessed through the following link:

http://www.pharmacy.umaryland.edu/centers/cersi-files/

A second website was established for data that are not available for public distribution (mainly data that are intended for inclusion in future publications), but that could be useful for evaluating the performance of the assay. The sources of the data used for evaluating each of
the test substance responses in the assay (publication citation and/or website postings) are indicated in Tables VI and VII.

b. Analysis of Pig-a test responses

i) Criteria for evaluating Pig-a test responses

(a) Test substances

A subcommittee of the HESI GTTC Pig-a Assay Workgroup was formed to provide advice on criteria for Pig-a data analysis and to form consensus where necessary on assay responses. The subcommittee consisted of Robert H. Heflich (USFDA/NCTR), Javed A. Bhalli (Covance and BioReliance), B. Bhaskar Gollapudi (Exponent), and Daniel J. Roberts (IUVO Biosciences and Charles River Laboratories).

The subcommittee defined the following response categories:

- Positive (P): overall positive call
- Negative (N): overall negative call
- Inconclusive (I): weakness in study design or other considerations makes it impossible to interpret results with a reasonable degree of certainty
- Equivocal (E): results from well-designed studies are insufficient for supporting a P or N call---often involves multiple studies having differing conclusions or single studies having data points supporting differing conclusions

The Workgroup subcommittee also established a set for rules for analyzing Pig-a assay data and arriving at a response call referred to as the ‘Workgroup (or WG) Criteria’ in Table V. The subcommittee designed the criteria in a manner to produce the greatest possible certainty in the positive and negative Pig-a response calls, at the expense of evaluating many studies as inconclusive. Key elements in these criteria included the use of a repeat-dose, longer-term (typically 28-day) treatment protocol; dosing up to the MTD or limit dose; assaying both RETs
and total RBCs for *Pig-a* mutation; and assaying at least $1 \times 10^6$ of both RETs and RBCs (a requirement that typically involved using an immunomagnetic enrichment protocol). For the purpose of the data analysis, ‘short-term’ was defined as single (or acute) treatments plus repeated treatments of up to 13 days, with treatments most often conducted on Days 1-3. All longer treatment protocols were considered ‘longer-term’ treatments, which were, most commonly, treatments on 28 consecutive days.

A great deal of latitude was applied to responses that were clearly positive. For example, clearly positive responses that were generated from short-term treatment protocols, or by assaying only total RBCs, typically were judged to be positive. The WG Criteria resulted in many nominally negative responses being evaluated as inconclusive because of various protocol deficiencies.

When preliminary results using the WG Criteria were reviewed by the *Pig-a* Expert Group (EG) formed by the OECD, it was strongly suggested that the criteria be modified to include all data from studies using short-term treatment protocols that otherwise conformed with the WG Criteria. It was recognized that a few test substances are more readily detected as positive using longer-term rather than short-term dosing protocols; however, the number of test substances falling into this category were relatively few, and mainly involve single treatments (see Tables II and III). It was felt that the practical advantages of conducting a short-term dosing study made it advisable to at least document the effect of accepting data from short-term and longer-term dosing protocols as equivalent. Thus, the evaluations summarized in Tables VI, VII and XIII were conducted both using the WG Criteria and the ‘Extended Criteria’ described in Table V.

(b) Vehicle controls

A practical consideration in analyzing the performance of the *in vivo* *Pig-a* assay was the availability of sufficient negative data in comparison to positive responses in the assay. This was a problem not only for the *Pig-a* assay, where much more test data are available on presumed
genotoxicants than data on presumed nongenotoxicants, but also for the reference assays used to perform an RPA. For the validation of the TGR assay, which led to establishing TG 488, this problem was partially overcome by using vehicle controls as negative substances (OECD, 2009).

The RPA for the TGR assay assumed that commonly used vehicles are negative in both the TGR assay and reference assays without testing that assumption (OECD, 2009). In the case of the Pig-a assay, this assumption was verified by analyzing vehicle controls for their mutagenicity using a modification of the analysis method commonly used for test substance data. Many Pig-a studies use a pretreatment sampling point (e.g., on Day -1) to establish the baseline mutant frequencies of the animals (and perhaps eliminate outliers; see Gollapudi et al., 2015). As this data point was generated before the animals were exposed to the vehicle, it can serve as a negative control for assays conducted subsequently on vehicle control animals (see Figure 8 for examples of pretreatment mutant frequency data in comparison with mutant frequencies from animals treated for 28 days with different vehicles).

Several deviations from the approaches used for analyzing data on test substances were necessary for evaluating the mutagenicity of vehicle controls. The dose was generally limited by the volume administered (e.g., 10-20 ml/kg/rat for aqueous vehicles), rather than by the mass or toxicity of the test substance, and only one test volume was generally used. These volumes are governed by established guidelines on the volumes that can be tolerated by animals during experimentation (e.g., Gad et al., 2016; IQ 3Rs Leadership Group, 2016). It is also the case that the same limit is generally applicable both to repeat dose studies and single dose studies, so that the animal experiences amounts of the vehicle control in direct proportion to the number of treatments conducted.

A review of appropriate test data in the Pig-a database was conducted by the data identification and cataloging subcommittee. All data for this analysis were required to be available in the on-line Pig-a database (see above). A preliminary screen identified studies that evaluated pretreatment samples as well as performed Pig-a assays after various short-term and longer-term treatments with vehicle controls (Annex 1). For the analysis conducted here, only
studies were considered that were conducted with immunomagnetic enrichment, where 28 or more daily treatments with the vehicle control were conducted, and where Pig-a assays were conducted within one week of the final dose. This design provides the maximum exposure to the vehicle substance, uses the most sensitive methods for detecting Pig-a mutant frequencies, and, from a practical standpoint, eliminated from consideration only three of the vehicles shown in Annex I (corn oil, designated etoposide solvent, and 1.25% HPMC, 0.18% methylparaben, 0.02% polyparaben, 0.1% docusate sodium in water).

Vehicle control data were analyzed by comparing pretreatment mutant frequencies with posttreatment mutant frequencies from individual studies. This involved employing a modified longitudinal analysis of Pig-a mutant frequency data as discussed in Section 5. As also described in Section 5, a small value was added to each mutant frequency, the data were log
10 transformed, and the frequencies were tested for increases using either a one-tailed Student’s t-test or Dunn’s test, as appropriate. Rare exposure-associated decreases were considered irrelevant, and where significant increases were detected, they were compared with historical controls established for the individual laboratories. With the exception of Litron Laboratories, which have published historical controls (e.g., Avlasevich et al., 2018), laboratory historical controls were generated ad hoc when sufficient pretreatment data were available.

Vehicle control treatments with significant posttreatment increases relative to the pretreatment frequency and that exceeded the historical control for the laboratory were considered positive. Vehicle control treatments that did not result in significantly increased mutant frequencies or did not exceed the range of historical negative control data for the laboratory were considered negative. This approach to evaluating vehicle control responses was reached by consensus agreement of the Pig-a data analysis subcommittee.

ii) Pig-a test responses in rats and mice

Ionizing radiation and over 80 chemicals, chemical mixtures, and nanomaterials have been tested for Pig-a mutation in either rats or mice. Seventy-two have been assayed in rats, 23 in
mice, and 13 test substances have been assayed in both rodent species. Responses for the test substances evaluated in rats are listed in Table VI, substances evaluated in mice are listed in Table VII, and a summary of the Pig-a responses in mice and rats and overall consensus calls are given in Table VIII. Tables VI and VII provide details about most of the tests that have been conducted (not all tests in rats involving ENU are listed), the sources of the information used for making response calls, and the reasons for the rat and mouse response E and I calls.

The 13 test substances that were tested in both mice and rats produced qualitatively similar Pig-a assay results. Thus, for most of the following discussion and analysis, data from mice and rats were combined, and the overall responses shown in the three right columns of Table VIII were used for analyzing assay performance.

Several test substances have been tested in multiple laboratories and using different protocols (details given in Tables VI and VII; see also Table II). In addition to its use as a positive control, ENU is the first compound most laboratories use when establishing the assay. Including assays conducted as part of interlaboratory trials, ENU has been tested in more than 20 laboratories, and only a fraction of these tests are listed in the tables (but many additional data sets can be found on the web-based databases). Other compounds tested for in vivo Pig-a mutation in three or more laboratories include 1,3-propane sultone: 7 labs; DMBA and 4-nitroquinoline-1-oxide (4NQO): 5 labs; acrylamide, benzo[a]pyrene (BaP), and ethyl methanesulfonate (EMS): 4 labs; 2-acetylaminofluorene (2-AAF), aristolochic acids, chlorambucil, cisplatin, cyclophosphamide, N-propyl-nitrosourea, ionizing radiation, melphalan, methyl methanesulfonate (MMS), mitomycin C, N-nitroso-N-methyl urea (MNU), and ethyl carbamate/urethane: 3 labs.

As described above, all response calls were made using two sets of criteria. In general, strict application of the Extended Criteria had the effect of rendering some (i.e., four) of the P responses using the WG Criteria, E calls (Table VIII, Summary of calls), as there were several test substances that tested P only with repeat-dose treatments (see Tables II and III and analysis in Section 4a). In addition, the Extended Criteria evaluated five compounds that tested negative,
but only used short-term treatment protocols, as N, while with the WG Criteria, they were evaluated as I (Table VIII). Thus, the Extended Criteria resulted in fewer P and I calls, and a greater number of N and E calls. As only P and N calls were used for the performance assessment, the WG Criteria produced 52 useful data calls (34 P and 18 N), while the Extended Criteria produced 53 (30 P and 23 N).

Table VIII also lists overall calls for data from the studies conducted using short-term treatment protocols (i.e., 13 or fewer daily treatments). Altogether, 56 test substances were assayed using short-term treatment protocols. The summary data in Table VIII indicate that results from studies with these shorter treatment protocols had a lower fraction of I calls (11% as opposed to approximately 30% when all studies were considered), with the result that the fraction of P and N calls was increased from approximately 60% when all Pig-a data were considered to 80% when only studies with short-term treatment protocols were considered. The significance of this observation is not clear and the changes in the relative number of P and N calls may reflect the relatively small number of test substances in the Pig-a database.

The 14 vehicle controls that were evaluated for their mutagenicity are listed in Annex I. Eleven of these vehicle controls were used in studies that had pretreatment data, 28 or more daily treatments, immunomagnetic separation protocols, and assays conducted at Day 28 to Day 35 of the study (Table IX). Expert data curation indicated that there were sufficient primary data to evaluate Pig-a mutant frequency responses for 10 of these vehicle controls, including phosphate buffered saline, water, and vegetable oil. All these substances were evaluated as negative in the Pig-a assay. There were insufficient data available to establish a Pig-a mutant frequency response for 0.5% methylcellulose (Table IX). For the purpose of the performance assessment, vehicle controls that were negative after 28-days of treatment were assumed to be negative with short-term treatments.
iii) Discussion of test responses

(a) Testing relevant to establishing the chemical space and applicability domain of the assay

Testing has been conducted on substances from a wide range of chemical classes, although in many cases the number of substances from each group is not large. Of the most commonly evaluated chemical classes, data have been collected on 10 alkylating agents, 6 aromatic nitro/amino compounds, 6 nanomaterials, 4 polycyclic aromatic hydrocarbons (PAHs), and 3 mycotoxins (Tables VI, VII and VIII).

Other than vehicles, 18 (WG Criteria) or 23 (Extended Criteria) test substances tested negative in the assay (Table VIII). These test substances can be divided into several categories in order to account for the negative responses (Table X). Note that the data underlying this categorization was not the result of a formal rigorous literature review; such a review, however, can be found for select conventional assays used for the performance assessment described in Section 8, below.

First, there are 9 substances that are generally regarded as nongenotoxic based on their structure and previous testing and are expected to be negative in any genotoxicity assay (see analysis of test substances in conventional genetic toxicity assays in Annex II). Second, there are six genotoxic compounds that were negative in the Pig-a assay, but the primary (or only) known mechanism of genotoxicity for these compounds is clastogenicity, aneugenicity, or recombination. These compounds are expected to be difficult to detect using a mutation reporter gene located on the X chromosome (DeMarini et al., 1989).

Third, there are 8 compounds known to be positive in other in vitro and in vivo gene mutation and genotoxicity assays but tested negative in the in vivo Pig-a assay. For these compounds, it is possible that adequate amounts of the reactive metabolites that caused them to be mutagenic in other assays did not reach the bone marrow of the test animals. It is not always clear whether this lack of adequate exposure is due to a specific problem with bone marrow
exposure, or due to a more generalized negative in vivo response (e.g., due to a difference in metabolite profile). The distinction is important since the former reflects a limitation of the assay (the assay detects only mutations induced in the bone marrow), while if the latter is true, the result would indicate that the test substance possesses little genotoxic risk in vivo.

In the third group, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butone (NNK) has induced gene mutations in vivo using other assays (von Pressentin et al., 1999, 2001) and was positive for Pig-a mutation in an in vitro assay (Table XV; Mittelstaedt et al., 2019). In the in vivo Pig-a assay, NNK was delivered by inhalation, and the limit dose by inhalation (determined by nasal tissue toxicity) was considerably lower than doses producing positive responses in the mouse in vivo MN assay where the compound was administered by intraperitoneal injection (Padma et al., 1989). Thus, it is likely that insufficient amounts of the test substance reached the bone marrow to produce a positive response in vivo (Mittelstaedt et al., 2019).

Insufficient exposure of the bone marrow also could be the reason for the negative Pig-a response for CeO₂. Although the best way of testing the genotoxicity of nanomaterials is currently uncertain (Elespuru et al., 2018), the negative Pig-a response, like the NNK response, was generated in an inhalation experiment, where relatively low doses of the test substance were tolerated, and where exposure of bone marrow is often a question (Cordelli et al., 2017). Relatively higher doses used for gavage exposures resulted in positive bone marrow MN responses (Kumari et al., 2014a,b). In contrast, the negative response produced by 1-(3-dimethylamino-propyl)-3-ethylcarbo-diimide hydrochloride (EDAC) has been attributed to a specific detoxification pathway or process present in vivo (Custer, 2012), and the negative produced by 4-chloro-1,2-diaminobenzene may be due to a sex-specific effect (see discussion in Section 8).

(b) Evidence for complementarily between the Pig-a assay and other in vivo genetic toxicity tests
Many of the studies summarized in Tables VI and VII evaluated other in vivo genotoxicity endpoints in the same animals used to conduct the Pig-a gene mutation assay (see notes in the tables). Comparison of the responses in the different assays is informative for determining if combining assays could give a more complete evaluation of the genotoxicity of a test substance. For example, the second group of Pig-a negatives listed in Table X lists seven compounds that have tested positive for blood or bone marrow MN induction in the same animals used for the Pig-a assay. It can be concluded that the genotoxicity that these test substances produce is more readily detected with assays measuring aneugenicity and/or clastogenicity (like the MN assay) than gene mutation (like the Pig-a assay).

Among other compounds with differential responses in different genetic toxicology assays, acrylamide was negative for Pig-a in mice but positive for MN induction (Hobbs et al., 2016). In addition, acrylamide was equivocal for Pig-a and negative for MN induction in rats, but positive for liver comet (Dobrovolsky et al., 2016; Hobbs et al., 2016). In separate studies, NNK was positive for TGR gene mutation in tissues other than blood or bone marrow (e.g., oral tissue; von Pressentin et al., 1999, 2001); presumably similar positive responses might occur if these assays were conducted in the animals used for the Pig-a assay.

Among Pig-a positives, several have been negative in the in vivo MN or comet assays conducted with the same animals. For example, aristolochic acids have tested positive for Pig-a gene mutation in rat studies conducted independently in three labs (Bhalli et al., 2013b; Elhajouji et al., 2014; Koyama et al., 2016). In the Bhalli et al. study, which used both 3-day and 28-day treatment protocols, aristolochic acids were consistently positive for the Pig-a endpoint, but at best produced an equivocal response for peripheral blood MN and were weakly positive with no dose response for liver comet. Other compounds that appear to fall into this category of a positive Pig-a response and weak or negative responses in blood or bone marrow-based assays for clastogenicity and DNA strand breakage include diethylnitrosamine (DEN, using immunomagnetic enrichment; Avlasevich et al., 2014; Kanal et al., 2018), a proprietary compound tested by BioReliance (Compound X-A, Table VI; Dutta, 2017), and 5-(2-chloroethyl)-2’-deoxyuridine (CEDU) (Table VI; Elhajouji et al., 2018), with the latter two compounds being
nucleoside analogues. A more formal analysis of the apparent complementarity between the blood-based *Pig-a* and MN assays can be found in the performance assessment described in Section 8.

Some of the sensitivity of *Pig-a* mutation relative to other genotoxicity assays may be due to its ability to accumulate mutations with repeat treatments. While 4NQO is positive for *Pig-a* gene mutation using both short-term and longer-term dosing protocols (Table VI), it reliably positive for the MN and comet endpoints only using relatively high doses in short-term treatment protocols, and not using doses that can be tolerated for a 28-day repeat-dose study (Stankowski et al., 2011a; Roberts et al., 2016). With regards to this latter observation, because of the large dynamic range for *Pig-a* mutation and its extremely low background frequency, it is often the case that greater fold-increases are noted for *Pig-a* mutation than for other genotoxicity endpoints, even when both endpoints are considered positive (e.g., Dertinger et al., 2010, 2012). This suggests that the *Pig-a* endpoint may have an intrinsic advantage in sensitivity over other *in vivo* genotoxicity endpoints, at least for gene mutagens.

(c) Studies of genotoxicant/nongenotoxicant pairs

In studies intended to test the specificity of the assay, *Pig-a* assays have been conducted with structurally related genotoxicant/nongenotoxicant pairs. The genotoxicant BaP and the nongenotoxicant pyrene have been tested in both rats and mice, and only BaP has induced *Pig-a* mutation (Tables VI, VII, and VIII; Torous et al., 2012; Labash et al., 2016). A similar distinction can be made between the genotoxic carcinogen ethyl carbamate (urethane) and the nongenotoxic carcinogen methyl carbamate. These two test substances have tested positive and negative, respectively, in both mice and rats (Tables VI, VII and VIII; Bemis et al., 2015; Labash et al., 2016; Stankowski et al., 2015). Although regarded as a genotoxic carcinogen, the genotoxicity of ethyl carbamate is difficult to detect *in vitro*, since it requires CYP2E1 whose activity is not present at sufficient levels in many *in vitro* genotoxicity assays (Tweets et al., 2007).
(d) Assays conducted to evaluate confounders and limitations of the assay

Additional testing has been conducted with protocols, and test substances, designed to discover limits to how the assay can be performed and interpreted. Cytotoxicity is regarded as a major confounder for genotoxicity assays and can produce nonrelevant responses. This is especially clear for in vitro assays that measure chromosome breakage (Greenwood et al., 2004; Fellows et al., 2008; Galloway et al., 2011) and it is a suspected confounder of the in vivo comet assay (e.g., Rothfuss et al., 2011). However, most genotoxicants are cytotoxic and cytotoxicity is also a valuable indicator of exposure, so there is always a question of how much cytotoxicity is sufficient to demonstrate exposure without producing an artefactual response. Recommendations on dose selection and cytotoxicity in the Pig-a assay (Table I; Gollapudi et al., 2015) are consistent with existing recommendations for conducting other in vivo genotoxicity assays (i.e., TG 474; TG 475; TG 488; TG 489). Two studies have been conducted in rats to stress the hematopoietic system and induce a wave of compensatory erythropoiesis. However, no effect was seen on the Pig-a mutant frequencies (Kenyon et al., 2015; Nicolette et al., 2017).

Another possible complication for the assay is its reliance on an endogenous gene located on the X-chromosome as a reporter of mutation. One of the two copies of Pig-a in female animals is silenced during development, resulting in one functional copy of the gene in both males and females. Inappropriate silencing of the ‘expressed allele’ in females (or the single allele in males) could increase Pig-a mutant frequencies in untreated animals. Also, if a treatment had the effect of altering methylation patterns or inducing gene silencing in the animals used for a Pig-a assay, increases in Pig-a mutant frequency might result that do not represent increases in Pig-a mutation. Testing conducted to explore the possibility that epigenetic effects confound the Pig-a assay is discussed in Section 6i.
(e) Additional studies

The Pig-a assay also has been used as a mutational reporter in studies evaluating the effect of nutritional factors and genotype on genomic stability. The primary objective of these studies was to answer basic science questions and not to test substances for in vivo mutagenicity; thus, many are not included in Tables VI or VII. Elevated Pig-a mutant frequencies have been detected in a mouse model of leukemia (Byrne et al., 2014); as a result of dietary protein deficiency (Pacheco-Martinez et al., 2016), dietary selenium deficiency (Grauptner et al. 2014, 2015, 2016), diet-induced obesity (Wickliffe et al., 2016); and in BaP-treated P450 reductase knockout mice (L Wang et al., 2017). A study that varied the concentration of dietary folate given to mice found the well-known effect of folate deficiency increasing MN frequency, but its effects on Pig-a mutant frequency were questionable (MacFarlane et al., 2015). Finally, a study using knock-out mice for DNA glycosylases NEIL1, NEIL2, and NEIL3 found no effect of the genotype on either cancer predisposition or spontaneous Pig-a mutant frequency (Rolseth et al., 2017).

8. Performance assessment of the Pig-a assay

a. Approach employed

OECD Guidance Document 34 recommends evaluating the performance of a new assay primarily by describing the ‘….. accuracy (e.g., sensitivity, specificity, positive and negative predictivity, false positive and false negative rates) of the proposed test method’, indicating that ‘….it should be compared to that obtained for the reference test method currently accepted by regulatory agencies…….’ (OECD, 2005).

As described in Sections 1 and 6, the Pig-a assay measures in vivo gene mutation, specifically mutations induced in the Pig-a gene of rat and mouse bone marrow cells. Thus, the primary
approach chosen to evaluate the performance of the Pig-a assay was to compare its accuracy in evaluating in vivo gene mutation relative to the mutagenicity of test substances whose mutagenicity responses are defined by the TGR assay. The TGR assay is the only in vivo gene mutation assay widely accepted by regulatory agencies and the only in vivo gene mutation assay currently with an OECD TG (TG 488). As described in Section 2d, the TGR assay is capable of measuring mutation in the bone marrow cells of rats and mice, as well as in virtually any other tissue of transgenic rodents. As indicated in Section 7b above, Pig-a assay results meeting test criteria have been established for over 50 test substances. The accuracy of the Pig-a assay was evaluated by comparing these Pig-a responses with responses in the TGR assay for these same test substances. Separate comparisons were made between Pig-a responses and TGR responses in rodent bone marrow and in any tissue of rats and mice. In order to make these comparisons, responses in rats and mice were merged into single data sets for each endpoint. As is evident from Tables VI and VII, most Pig-a data are from experiments conducted with rats, while most TGR data are from studies with mice (Lambert et al., 2005).

In addition, the most widely used in vivo genotoxicity assay is the in vivo erythrocyte MN assay as described in OECD TG 474 (see Section 2a). Section 7b describes data indicating that, even though they measure genotoxicity in the same basic target tissue, the in vivo Pig-a assay may compliment the in vivo MN assay due to their measuring mechanistically different genotoxic responses. In order to determine the extent to which this mechanistic specificity affects the predictive value of these assays for evaluating genotoxicity, responses in the in vivo Pig-a assay were compared with responses defined by the in vivo MN assay.

Finally, cancer is a primary health concern addressed by in vivo gene mutation data. Gene mutation provides regulatory agencies with a practical, relatively short-term method for identifying potential genotoxic carcinogens. Therefore, in addition to evaluating the accuracy of the Pig-a assay as a test for in vivo gene mutation, we evaluated its accuracy for identifying rodent carcinogens.
b. Establishing responses in reference assays

A separate subcommittee, that included members of the HESI GTTC Pig-a Assay Workgroup as well as other genetic toxicology experts, was formed to provide a curated set of responses in conventional genotoxicity tests and rodent cancer bioassays for the test substances evaluated in the Pig-a gene mutation assay. Besides establishing mutation responses in the TGR assay for bone marrow and all tissues, and responses in the rodent cancer bioassay (separating cancer in hematopoietic tissue and cancer in all tissues), responses also were determined for the in vitro MN and chromosome aberration assays, the in vivo MN assay, and the Ames’ bacterial gene mutation assay. This subcommittee, which operated independently of the Pig-a assay subcommittee described above, consisted of Michelle O. Kenyon (Pfizer), Stephen D. Dertinger (Litron Laboratories), Leon F. Stankowski (Charles River Laboratories), Martha M. Moore (Rambol), Katsuyoshi Horibata (NIHS, Japan), and Takafumi Kimoto (Teijin Pharma). The same P, N, E, and I categories used for describing Pig-a responses were used to describe responses in the conventional assays.

First, genotoxicity and carcinogenicity data for the test agents were obtained from the Vitic Database (Lhasa, Ltd). Where all results for a test substance in an assay were in agreement, the result was considered to be accurate. When there were different test results in an assay for a given test substance, the results and citations were provided to an independent reviewer for assessment. Additional databases were consulted as needed. The reviewer provided a suggested call for discussion by the subcommittee. If a consensus call could not be made, that test substance was documented as Equivocal (E) in the assay.

Annex II provides the consensus calls developed by the subcommittee for the complete set of conventional genetic toxicology assays. Note that the analysis of in vivo MN responses was not limited to in vivo erythrocyte assays. In at least one case (that of DEN), data from liver MN assays drove the overall response call for a test substance.
c. Performance assessment results

Table XI compares Pig-a assay responses generated by applying both the WG and Extended Criteria with responses defined by TGR responses in bone marrow and in all tissues, by responses in the in vivo MN assay, and by rodent cancer bioassay responses in the hematopoietic system and in all tissues. Only test substances producing either P or N overall responses in the Pig-a assay were used for this analysis.

The performance evaluation of the Pig-a assay was conducted by Robert H. Heflich (USFDA/NCTR) in consultation with David P. Lovell (St. George's University, London). Pig-a assay performance was evaluated by defining true positive and negative responses as the responses in the reference assay and calculating the following metrics:

- Sensitivity: the percent of all the positives in the reference assay detected as positives in the Pig-a assay
- Positive predictivity: the percent of positives in the Pig-a assay that are positive in the reference assay
- Specificity: the percent of all negatives in the reference assay detected as negatives in the Pig-a assay
- Negative predictivity: the percent of negatives in the Pig-a assay that are negatives in the reference assay
- Concordance: the number of true negatives and positives (defined by their responses in the reference assay) correctly identified in the Pig-a assay divided by the total number of test substances x 100.

The calculation of these metrics from data arrayed in two-by-two tables is shown in Annex III.

In addition, performance was evaluated by the application of two tests of agreement between two different measures:

- Cohen’s kappa test: an estimate of the inter-assay agreement (i.e., the agreement between the Pig-a assay and a reference assay) for qualitative (categorical) items. The
kappa statistic ranges from -1 (perfect disagreement) to 0 (no agreement) to +1 (perfect agreement); among positive agreements, 0.01-0.20 is considered slight agreement, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial, and 0.81 to 1.00 almost perfect.

- PABAK (prevalence and bias adjusted kappa): since Cohen’s kappa is affected by prevalence (e.g., unbalanced data sets having more positives than negatives), tests like the PABAK have been devised that adjust the statistic for prevalence and distribution of the marginal totals. However, even with adjustment, none of the association tests is perfect, and must be view in context with other quantitative measures.

Calculations were made with the DAG Stat statistical suite for Microsoft Excel, available at: https://www.biostats.com.au/DAG_Stat/. The two-by-two tables and performance statistics, including 95% confidence intervals for some of the statistics and a list of nonconcordant test results, are shown in Annex III. Note that the performance metrics that were calculated for the assay have large 95% confidence intervals, a function of the relatively small data sets (n = 25-53) that were used for the calculations. Thus, any perceived differences in the performance of the assay must be tempered by the level of confidence in the estimates.

i) Performance assessment using Pig-a responses evaluated using the WG Criteria

Table XII summarizes the results of the performance analysis conducted on Pig-a assay responses determined using the WG Criteria. The responses with the Pig-a assay displayed a high degree of accuracy when the bone marrow TGR assay defined P and N test substances, with only DEN producing discordant results. DEN, a genotoxic carcinogen whose genotoxicity is difficult to detect in blood-based in vivo assays (Kanal et al., 2018), was positive in the Pig-a assay using immunomagnetic separation and negative in the bone marrow TGR assay, suggesting that, at least for bone marrow mutation, the Pig-a assay was more sensitive to the mutagenicity of this compound than the TGR assay.
When TGR responses in any tissue were used as the reference assay, the accuracy of the Pig-a mutation assay was only slightly lower. Discordant results were found for three test substances, hydroxyurea, NNK, and 4-chloro-1,2-diaminobenzene, which were negative in the Pig-a assay but were mutagenic in the TGR assay in tissues other than bone marrow. In addition, as discussed in Section 7b, hydroxyurea is known mainly as a clastogen (Ames-test positive only at concentrations that exceed the current recommendation of 5000 µg/plate), and the Pig-a study in which NNK was evaluated used inhalation treatment, which produced limited exposure of bone marrow (Mittelstaedt et al., 2019). Finally, when 4-chloro-1,2-diaminobenzene, an Ames’ positive mouse liver and rat bladder carcinogen, was tested in the mouse TGR assay, it was only positive for liver mutation induction in female mice, and it was negative in male mice (Suter et al., 1996). Since the Pig-a data were generated in male rats, it is possible that the discordance between the positive TGR response and negative Pig-a response for 4-chloro-1,2-diaminobenzene not only could be due to its tissue specificity, but also its species and sex specificity. Note that DEN, which produced discordant responses when Pig-a and TGR responses were evaluated in bone marrow, was positive for other tissues in the TGR assay so that DEN was positive in both the Pig-a and TGR assays when all tissues were considered for the TGR assay.

When the in vivo MN assay was used as the reference assay, the accuracy of the Pig-a assay was still good (‘substantial’ using the Kappa test interpretation guidelines described above), but lower that when the TGR assay was used to define P and N responses. Seven discordant compounds were detected: acetaminophen, aristolochic acids, CEDU, 4-chloro-1,2-diaminobenzene, CeO2, hydroxyurea, and vinblastine. As noted in Section 7b acetaminophen, hydroxyurea, and vinblastine mainly are genotoxic through genotoxic mechanisms resulting in clastogenicity or aneugenicity, and thus were positive in the MN assay and negative for Pig-a mutation. On the other hand, aristolochic acids and CEDU, which are known to be genotoxic mainly through the induction of gene mutation, were positive in the Pig-a assay and negative for MN induction. Thus, the mechanistic complementarity of the MN and Pig-a assays noted in Section 7b can be detected in this quantitative assessment.
When cancer in hematopoietic tissues was used to define positive and negative responses, responses in the Pig-a assay had a very high degree of accuracy (‘near perfect’ in the terminology described above). Only two test substances produced discordant results: acetaminophen, whose genotoxicity is most readily observed as clastogenicity and not gene mutation, and glycidyl methacrylate, which is positive in many conventional gene mutation and clastogenicity assays (see Annex II) and may be an example of a genotoxic noncarcinogen. Note that this high degree of precision was based on observations with very few compounds (25), and that the precision benefited greatly from using vehicle control data for the analysis.

Defining positive and negative responses by the ability of a test substance to induce cancer in any rodent tissue markedly reduced the accuracy of the Pig-a assay. This was mainly reflected in a large reduction in the negative predictivity of the assay, i.e., there were several compounds negative in the Pig-a assay that were rodent carcinogens. In all, there were 8 discordant responses, including responses from test substances that are often classified as nongenotoxic carcinogens (methyl carbamate, 2-butoxyethanol), compounds that are mainly clastogens (acetaminophen, hydroxyurea) and compounds that may have poor bone marrow exposure (NNK) (see Table X and associated discussion). As indicated above, 4-chloro-1,2-diaminobenzene may exhibit several possible specificities in its mutagenic response, and it is possible that this influences its carcinogenicity. In addition, glycidyl methacrylate was positive in the Pig-a assay, while it was negative for rodent cancer. As mentioned above, this compound may be an example of a genotoxic noncarcinogen.

Table XIII summarizes the results of the performance analysis conducted with Pig-a assay responses determined using the Extended Criteria. Using these criteria, there was perfect agreement between the responses in the Pig-a assay and the bone marrow TGR assay. Note that in strictly applying the Extended Criteria, DEN (which produced the only discordant results using the WG Criteria) was not used for the analysis since it was scored as equivocal (two well
conducted assays producing different results, discussed further in Section 4a). Also, two compounds, azathioprine and 2-acetylaminofluorene, were scored as equivocal since studies with well-conducted single-treatment protocols were negative while these test substances were positive when animals were treated with three or 28 daily exposures (see discussion in Section 4a). Thus, results with these two substances were not used for the performance evaluation. All other responses were the same using both the WG and Extended Criteria.

The accuracy of the Pig-a mutation assay was slightly lower (but still substantial, applying the Kappa test interpretation guidelines) when TGR responses in any tissue were used as the reference assay to define positives and negatives. Discordant results were found for the same three test substances that were discordant using the WG Criteria: hydroxyurea, NNK, and 4-chloro-1,2-diaminobenzene. All were negative in the Pig-a assay, regardless of including or excluding short-term treatment studies, but were mutagenic in the TGR assay in tissues other than bone marrow. Note that DEN, which produced discordant responses when Pig-a and TGR responses were evaluated in bone marrow and Pig-a responses were evaluated using the WG Criteria, was scored as equivocal in the Pig-a assay using the Extended Criteria since a well-conducted Pig-a study with a single treatment was negative and Pig-a studies having 28-daily treatment protocols were positive.

When the in vivo MN assay was used as the reference assay, the performance of the Pig-a assay fell to ‘moderate’, lower than when the TGR assay was used to define P and N responses. When analyzed using the Extended Criteria, 11 discordant compounds were detected: acetaminophen, aristolochic acids, CEDU, 4-chloro-1,2-diaminobenzene, CeO₂, hydroxyurea and vinblastine, as with the WG Criteria, plus caffeic acid, 1,2-dimethylhydrazine, 5-fluorouracil, and AZT. As noted in Section 7b these later four compounds are considered to operate primarily through mechanisms that produce clastogenic/aneugenic responses. As was the case when Pig-a responses were evaluated using the WG Criteria, the mechanistic complementarity of the MN and Pig-a assays noted in Section 7b is evident in this assessment.
When cancer in hematopoietic tissues was used to define positive and negative responses, responses in the Pig-a assay, analyzed with the Extended Criteria, displayed a very high degree of accuracy. As was the case when the WG Criteria were used to evaluate Pig-a responses, only two test substances produced discordant results, acetaminophen and glycidyl methacrylate. Again, this high degree of accuracy was based on observations with very few compounds (26 in this case) and benefited greatly from using vehicle control data for the analysis.

Defining positive and negative responses by the ability of a test substance to induce cancer in any rodent tissue reduced the accuracy of the Pig-a assay to ‘moderate’. This was mainly reflected in a large reduction in the negative predictivity of the assay, from 94% when only hematopoietic cancer data were used to 56% when cancer in any tissue was used. There were 13 discordant responses among 50 test substances with relevant data. Along with the 8 discordant test substances identified with the WG Criteria, discordant test substances included four powerful clastogens that are weak or negative in gene mutation assays (caffeic acid, AZT, 1,2-dimethylhydrazine, and 5-fluorouracil) and melamine, which is believed to be a nongenotoxic carcinogen (see Table X and associated discussion).

**ii) Performance assessment using only Pig-a responses from assays conducted with short-term treatment protocols**

Table XIV summarizes the results of the performance analysis conducted with Pig-a assay responses determined using data only from studies conducted using short-term treatment protocols and analyzed using the Extended Criteria. Under these conditions, there was perfect agreement between the responses in the Pig-a assay and the bone marrow TGR assay. Note that a study with a single treatment with DEN produced a negative response, the same as the response in the bone marrow TGR assay. Also, two compounds, azathioprine and 2-acetylaminofluorene, were scored as equivocal using short-term protocols since well-conducted studies using a single treatment were negative while these test substances were positive when
treated with three daily exposures (see discussion in Section 4a). Thus, data from these two substances were not used for the performance evaluation.

The accuracy of the Pig-a mutation assay using only short-term treatment protocols was slightly lower (but still substantial) when TGR responses in any tissue were used as the reference assay to define test substances with positive and negative responses. Discordant results were found for four test substances, hydroxyurea, NNK, 1,2-dimethylhydrazine, and DEN, all of which were negative in the Pig-a assay and positive in the TGR assay in a tissue other than bone marrow. Reasons why hydroxyurea and NNK might be negative for the Pig-a assay are discussed above for Pig-a data evaluated by both the WG and Extended Criteria. Although it wasn’t evaluated in the bone marrow, the mutagenicity of 1,2-dimethylhydrazine in the TGR assay was strongly associated with the colon (Newell and Heddle, 2004). DEN was discordant since a Pig-a mutation study employing a single treatment was negative (it hasn’t been tested with other short-term testing protocols) while DEN is positive for TGR mutation in a tissue other than bone marrow.

When the in vivo MN assay was used as the reference assay, the performance of the Pig-a assay was still good, but lower than when the TGR assay was used to define P and N responses (falling into the moderate range). The lower accuracy can be seen in the lower negative predictivity value, meaning that a number of negatives in the Pig-a assay were positive in the MN assay. When data from short-term Pig-a studies were analyzed using the Extended Criteria, 10 discordant compounds were detected: acetaminophen, aristolochic acids, CEDU, hydroxyurea, aflatoxin B1, caffeic acid, 1,2-dimethylhydrazine, 5-fluorouracil, and AZT. As noted above, several of these compounds are considered primarily to be aneugens/clastogens (acetaminophen, hydroxyurea, caffeic acid, 1,2-dimethylhydrazine, 5-fluorouracil, AZT), while aristolochic acids and CEDU operate mainly through a pathway resulting in gene mutation. As was the case when Pig-a responses were evaluated using the WG or Extended Criteria, the mechanistic complementarity of the MN and Pig-a assays noted in Section 7b can be detected in this assessment. The negative response with aflatoxin B1 with short-term treatments is notable since it is one of the few compounds whose mutagenicity is detected only with longer-
term treatment protocols (see Table III); this limitation in the gene mutation response is not true for the MN assay. Finally, DEN was classified as positive for in vivo MN by the conventional assays subcommittee, but this call was based on strong responses in the liver MN assay. DEN was found to be negative in several erythrocyte MN studies that were conducted in conjunction with Pig-a gene mutation assays (Shi et al., 2011; Avlasevich et al., 2014; Kanal et al., 2018).

When cancer in hematopoietic tissues was used to define positive and negative responses, responses in the Pig-a assay, analyzed with the Extended Criteria applied to responses in assays conducted with short-term dosing protocols, displayed a very high degree of accuracy. In this case, only acetaminophen produced discordant results, and, as explained above, this may be because the genotoxicity of acetaminophen is characterized by clastogenicity and not gene mutation. Again, this high degree of accuracy was based on observations with very few compounds (24, the smallest data set used in this analysis), and the precision benefited greatly from using vehicle control data for the analysis.

As was the case when the results from longer-term dosing studies were included in the analysis, defining positive and negative responses by the ability of a test substance to induce cancer in any rodent tissue markedly reduced the precision of the Pig-a assay (into the moderate range). This was mainly reflected in a large reduction in the negative predictivity of the assay, from 93% when only hematopoietic cancer data were used to 56% when cancer in any tissue was used. This analysis resulted in 11 discordant responses among 46 test substances with relevant data, and in each case the discordant finding involved a negative Pig-a response for a rodent carcinogen. Ten of the discordant compounds were also discordant using the WG or Extended data analysis criteria on the full set of Pig-a data responses. The situation with aflatoxin B1 is discussed above; aflatoxin B1 was only positive in the Pig-a assay when a 29-day daily dosing protocol was used.
9. Positioning of the assay relative to existing *in vivo* genetic toxicology assays

The IWGT report described placement of the assay relative to existing genotoxicity testing strategies (Gollapudi et al., 2015), and those recommendations are essentially unchanged in this document. The major use currently envisioned for the *Pig-a* assay is as a follow-up for evaluating test substances that are positive for gene mutation *in vitro* using either bacterial or mammalian cell tests. The IWGT report pointed out that the assay is a gene mutation assay and therefore may not detect substances that are primarily clastogens or aneugens. It also noted that negative tests should only be accepted if there was an indication of bone marrow exposure. Test substance exposure could be demonstrated by a reduction in the frequency of RETs or by demonstrating adequate plasma levels of the test substance or its metabolites. Table X contains examples confirming the wisdom of these recommendations: genotoxic compounds like hydroxyurea that are primarily clastogens, and gene mutagens like NNK that have poor bone marrow exposure, have produced negative results in the *in vivo Pig-a* assay.

The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) M7 guidance (ICH, 2014) recommends the *Pig-a* assay as a follow-up *in vivo* assay for drug impurities positive for bacterial mutagenicity. The ICH and IWGT recommendations are quite similar for this follow-up application, although the ICH recommendations do not consider clastogens/aneugens at all, and suggest demonstration of adequate exposure to metabolite(s) only for indirect (S9-mediated) bacterial mutagens. To our knowledge, the assay has been used in follow-up studies for at least two types of Ames-positive drug impurities. In addition, the *Pig-a* assay has been used as a follow-up test for a nucleoside analog Active Pharmaceutical Ingredient producing large colony mutants in the MLA. These examples, however, come from meeting presentations or personal communications and are not in the published literature (Custer, 2012; Dutta, 2017; Joel Bercu, personal communication).

The IWGT report also indicated that the *Pig-a* assay has value in developing weight-of-evidence arguments for the genotoxicity of a substance. The quantitative assessment of genotoxicity data has shown potential for estimating human risk (Johnson et al., 2014, 2015), and the
properties of the Pig-a assay may be useful for this application. Its large dynamic range, its ability to respond to repeated dosing in a cumulative manner, its ability to repeatedly sample test animals over time, and its potential to analyze the same endpoint in multiple species, including humans, may be beneficial to generating quantitative gene mutation data for human risk assessment.

In addition, the IWGT report suggested on mainly theoretical grounds that the Pig-a gene mutation assay can serve as a complement to the in vivo peripheral blood MN assay for many routine in vivo genotoxicity testing applications. While both measure genotoxicity in bone marrow cells, each measures a different spectrum of genotoxicity responses (gene mutation vs. clastogenicity/aneugenicity). Especially when used in conjunction with repeat dose studies, both could be integrated readily into standard toxicity protocols without impacting other testing. The analyses in Sections 7b and 8 provide evidence of the complementarity of the two tests for detecting the genotoxicity of several compounds. This includes clastogens like hydroxyurea, aneugens like vinblastine, and gene mutagens like CEDU and aristolochic acids, where one test was positive and the other negative for these known genotoxic substances.

Another potential regulatory application, as the human PIG-A assay becomes more widely used, is direct confirmation of findings in animal studies with data from humans. This ability is unique to assays like the Pig-a/PIG-A assay that measure genotoxicity in endogenous genes and thus provide the potential for translation between laboratory animals and humans. This degree of translation is not possible with TGR assays that employ genetically engineered rodents and exogenous transgenes.

The Pig-a assay also has applications in studies that go beyond genotoxicity testing. As outlined in Section 7b, the assay has been used to evaluate the effect of diet and genotype on genomic stability in rodents and the effect of these factors on responses to mutagens (Byrne et al., 2014; Graupner et al., 2014, 2015; Pacheco-Martinez et al., 2016; Wickliffe et al., 2016; Rolseth et al., 2017; L Wang et al., 2017). Because the assay can be performed with human cells, it has been used to evaluate mutant frequencies and mutation rates in humans and in cells cultured from
humans with diseases associated with genomic instability (Olsen et al., 2017; see Sections 10a and 10c). For example, Araten et al. (2010) have used the PIG-A assay to demonstrate genomic instability in cells derived from patients with T cell acute lymphoblastic leukemia, mantle-cell lymphoma, follicular lymphoma, and in some plasma cell neoplasms.

10. Potential for expanding the applicability of the assay

The potential for translation is one of the strongest features of the Pig-a assay. Given the conservation of the Pig-a gene, GPI anchors, and the function of GPI anchors across species and cell types, the assay can potentially be adopted to other species of toxicological relevance as well as to different cell types in vivo and to in vitro mammalian cell systems.

a. Erythrocyte PIG-A assays conducted in humans

The most important species for evaluating human risk is humans, and the genesis of the rodent Pig-a assay began with studies using the PIG-A gene as a reporter of somatic cell mutation in humans (Araten et al., 1999; see Section 1a). Several investigators have followed up on the initial observations that reported a ‘background’ frequency of $8 \times 10^{-6}$ CD59 plus CD55 doubly-deficient total RBCs in one normal donor (Araten et al., 1999). Dobrovolsky et al. (2011) greatly expanded the number of healthy volunteers assayed for CD59-deficient erythrocyte PIG-A mutant frequency and found a frequency of $5.1 \pm 4.9 \times 10^{-6}$ for total RBCs from 95 of 97 volunteers; there were no significant relationships between mutant frequency and age or smoking habits. Two individuals had markedly increased mutant frequencies of approximately 100 and $300 \times 10^{-6}$, with the frequencies remaining elevated upon resampling. In addition, 10 cancer patients receiving cytotoxic/mutagenic chemotherapy were monitored for total RBC mutant frequency before and for approximately 6 months after beginning therapy. Only one patient, treated with a combination therapy that included cisplatin, displayed an increase in PIG-A mutant frequency with time.
More recently Cao et al. (2016) used the total RBC \textit{PIG-A} assay to measure mutant frequencies in 217 Chinese volunteers. The mean mutant frequency, $5.25 \pm 3.6 \times 10^{-6}$, was remarkably similar to that detected by Dobrovolsky et al. (2011) in a U.S. population, and again, no relationship between age and mutant frequency was found. In this larger population, however, a positive relationship was found between mutant frequency and smoking expressed as cigarette-pack-years. In a subsequent pilot study, Cao et al. (2017) reported increased erythrocyte \textit{PIG-A} mutant frequencies in some (but not all) of 20 inflammatory bowel disease patients treated with azathioprine and in 17 barbeque restaurant workers who potentially are exposed to elevated levels of PAHs.

In addition, Horibata et al. (2016) monitored total RBC \textit{PIG-A} mutant frequencies in 10 healthy volunteers and 27 patients undergoing cancer chemotherapy. Although the study did not include pre-chemotherapy mutant frequency determinations, and the times of sampling were often too early to expect treatment-related increases in mutant frequency, two of the cancer patients had \textit{PIG-A} mutant frequencies significantly greater than the healthy controls.

These foregoing studies provide evidence that an erythrocyte \textit{PIG-A} assay can be deployed in humans. There has been a question, however, about the extent to which complement-mediated lysis of GPI-deficient cells may affect the sensitivity of the assay (Peruzzi et al., 2010). Complement-mediated lysis of \textit{PIG-A} mutants is a well-established pathology for PNH, and the sizes of PNH clones can be considerably greater for RETs than for mature RBCs (Iwamoto et al., 1995; Ware et al., 1995). In addition, none of the earlier studies measuring erythrocyte \textit{PIG-A} mutation in humans took advantage of immunomagnetic enrichment to increase the number of cells evaluated. With this in mind, Dertinger et al. (2015) have adopted the immunomagnetic enrichment techniques used for the rodent \textit{Pig-a} assay for measuring mutant frequencies in human RETs and total RBCs from 52 healthy volunteers. As anticipated, the frequency of CD59 plus CD55 mutants was approximately two-fold greater in RETs than in total RBCs (mean frequencies, $6.0 \times 10^{-6}$ vs. $2.9 \times 10^{-6}$, respectively), indicating a modest selection pressure in humans against mutant erythrocytes. The relatively large inter-individual variability in RET mutant frequencies (30-fold) and the consistency in RET mutant frequency with repeat
sampling was similar to that noted with previous studies measuring PIG-A mutant frequencies in total RBCs. These studies using immunomagnetic enrichment, however, detected age-related increases in both RET and total RBC mutant frequencies.

Methods for measuring RET PIG-A mutant frequencies have not been applied widely to populations at risk for elevated genotoxicity. However, a preliminary report presented as a meeting poster provided evidence that combination therapies that included cisplatin increased peripheral blood PIG-A and/or micronucleated RET frequencies (n=3; Dertinger et al., 2016). Importantly, these chemotherapy patients provided pretreatment blood samples to accommodate the fact that humans exhibit wide baseline variations in PIG-A mutant frequency.

In addition, human PIG-A assays have been conducted with cells other than erythrocytes, where complement-mediated selection against PIG-A mutants is expected to be less of an issue. These studies are described in Section 10c, below.

b. Erythrocyte Pig-a assays conducted in nonhuman primates

A single study by Dobrovolsky et al. (2009) evaluated PIG-A (CD59-deficient) mutant frequencies in total RBCs in 10 controls and in one male rhesus monkey (Macaca mulatta) treated twice with 50 mg/kg ENU. The ENU treatment was given by i.p. injection at approximately two years and 3.5 years of age. Total RBC mutant frequencies measured 6 months after the last ENU treatment indicated a mutant frequency of 46.5 x 10^{-6} vs a mean frequency of 7.8 x 10^{-6} in the controls.

c. In vivo Pig-a/PIG-A assays conducted in germ cells and other tissues

i) Germ cells

Mutations transmitted though germ cells are of concern since they will occur in every cell of the resulting offspring and can have a profound impact on the phenotype. Thus, the analysis of
germ cell genotoxicity and gene mutation is sometimes recommended under particular circumstances, e.g., when there is evidence of germ cell exposure whether or not there is evidence of genotoxicity in somatic cells (Yauk et al., 2015). In addition to cells of the bone marrow, male germ cells are actively dividing in adults, and can be made into single cell suspensions, making them potentially amenable for the detection of gene mutation using the Pig-a assay. TG 488 describes the use of the TGR assay for measuring gene mutation in male germ cells (OECD, 2013; note that this TG is currently being updated for measuring germ cell mutation). While there may be advantages to a Pig-a assay using male germ cells (e.g., it could be conducted in non-transgenic animals), it may not be practical to repeatedly sample animals for germ cell mutation as can be done for the erythrocyte Pig-a assay.

Initial work has been conducted on a male germ cell Pig-a assay in two laboratories. In a study conducted by Bhalli and coworkers and presented only in a poster (Bhalli et al., 2017), male rats were treated with procarbazine. Assays conducted with epidydimal sperm approximately 45 days after the treatment detected a dose-related increase in the frequency of CD59-negative sperm, with a relatively high background frequency that the authors attributed to the preliminary stage of methods development. Another study, using a similar approach, found positive responses with a germ cell mutagen (ENU), negative responses for a nonmutagen (clofibrate), and a high CD59-negative sperm frequency in samples from vehicle controls (84 and 88 x 10^6) (Ji and LeBaron, 2017). Although these data are quite preliminary, they suggest that, when the current problems and uncertainties are resolved, a modified version of the Pig-a assay could be useful for evaluating male germ cell mutagens.

**ii) Other cell types**

Several human PIG-A studies have used blood cells, other than erythrocytes, to measure PIG-A mutant frequencies. These studies have used granulocytes, lymphocytes, monocytes, and bone marrow progenitor cells, but many of these studies have been in the context of evaluating PNH or making other disease diagnoses rather than measuring small increases in mutant frequency
(e.g., Brodsky et al., 2000; Mukhina et al., 2001; Mortazavi et al., 2003). As reviewed in Olsen et al. (2017), however, low-level mutant frequencies have been measured in humans in several contexts: understanding the significance of low-level spontaneous PIG-A mutant frequencies and their connection to PNH and other diseases (Araten et al., 1999; Ware et al., 2001; Hu et al., 2005; Pu et al., 2012); using PIG-A mutant frequency as a reporter of genomic instability (Araten and Luzzatto, 2006; Araten et al., 2005, 2010, 2012, 2013; Grasso et al., 2014; Rondelli et al., 2013); using PIG-A as a biomarker of human exposure to potential genotoxins (McDiarmid et al., 2011); and using PIG-A for establishing mutational rates in humans. The mutation rate studies can be viewed as a hybrid in vivo/in vitro analysis, as B lymphocytes were taken from humans, transformed into lymphoblastoid cultures, cleansed of existing PIG-A mutants by flow sorting, and then mutant frequencies measured over a period of time to establish a mutation rate (Araten et al., 2005; Peruzzi et al., 2010).

McDiarmid et al. (2011) described using limiting-dilution cloning of proaerolysin-resistant peripheral blood T lymphocyte mutants for evaluating mutations in Gulf War veterans exposed to munitions containing depleted uranium. Since human blood samples don’t have the volume limitations of rodent samples, the use of T lymphocytes for evaluating PIG-A mutant frequencies may have some advantages for biomonitoring over the use of erythrocytes. T Lymphocytes from humans can be analyzed for PIG-A mutation either by flow cytometry or by limiting dilution cloning (e.g., Rawstron et al., 1999; Ware et al., 2001), lymphocytes expand readily and contain DNA so the mutations can be analyzed directly (e.g., Revollo et al., 2015; Dobrovolsky et al., 2017), and the use of lymphocytes avoids the problem of immune lysis that may complicate human erythrocyte assays (Section 10a). Cao et al. (2016) reported on the feasibility of using FLAER to monitor the frequency of GPI-negative white blood cells from a small number (5) of human volunteers.

Several studies also evaluated Pig-a mutant frequencies in rodents using blood cells other than erythrocytes. Pig-a mutant frequencies have been measured in peripheral blood T lymphocytes and monocytes from mice (Byrne et al., 2014) and in peripheral blood T cells from rats (Miura et al., 2008a,b, 2011; Bhalli et al., 2011b; Cammerer et al., 2011; Dobrovolsky et al., 2015;
Revollo et al., 2015, 2016). The mouse study measured the effects of a presumed mutator phenotype on mutant frequency, while the rat studies detected the in vivo mutagenicity of known mutagens (i.e., ENU, DMBA, and BaP). In addition, several studies have evaluated Pig-a mutation in nucleated rat and mouse bone marrow erythrocyte and granulocyte precursor cells (Kimoto et al., 2011a,b; Revollo et al., 2018; Dobrovolsky et al., 2017; Dad et al., 2018; Kimoto and Miura, 2017), with the primary objective of directly demonstrating that Pig-a mutation is responsible for the mutant phenotype measured in the assay (i.e., loss of GPI-anchored protein markers; see Section 6).

A single study evaluated Pig-a mutation in T lymphocytes from Macaca mulatta (rhesus monkeys) (Dobrovolsky et al., 2009). Increases in proaerolysin-resistant T lymphocytes were detected in a single monkey given two doses of ENU more than a year apart and monitored over a period of 21 months.

d. In vitro Pig-a assays

Flow analysis of Pig-a mutation may benefit in vitro analysis of gene mutation from a throughput standpoint and in developing high-density dose response data (Johnson et al., 2015). An in vitro version of an in vivo assay also may be envisioned as having value in testing hypotheses about results in the in vivo assay and in prescreening compounds for in vivo testing (Bemis and Heflich, 2019). Most of the work on developing in vitro versions of the Pig-a assay has been conducted using TK6 human lymphoblastoid cells and L5178Y/Tk+/− mouse lymphoma cells. These cell lines are grown in suspension, lending themselves to flow cytometry analysis, and they already are used for regulatory in vitro gene mutation assays (OECD, 2016c; OECD 2016g). Therefore, the characteristics of these cells are relatively well known, and standardized cultures suitable for genetic toxicology evaluations are available through cell repositories (Lorge et al., 2016).

TK6 cells were used in one of the first papers describing PIG-A as a reporter of mutation (Chen et al., 2001). Although there have been problems with extraordinarily high spontaneous mutant
frequencies in these cells (>1,000 x 10⁻⁶), methods have been devised to obtain lower, more stable background mutant frequencies (e.g., Rees et al., 2017). Using cleansed cultures, Krüger and colleagues (Krüger et al., 2015, 2016; Piberger et al., 2017) detected strong mutational responses for EMS, 4-NQO, BaP diol epoxide, and UVC (Table XV). It should be noted that the TK6 cell ‘PIG-A’ mutation assay detects mutants that have mutations in either PIG-A or PIG-L, as PIG-L is heterozygous in these cells (Nicklas et al, 2015; Krüger et al., 2016; see Section 6f for more information). Thus, although the TK6 cell assay may not be totally analogous to the in vivo Pig-a assay, it has been commented that the sensitivity of the TK6 cell assay may benefit from having both an X-linked and autosomal reporter of mutation in terms of increasing the types of mutations that the assay can detect (Nicklas et al., 2015).

Recent reports indicate that the Pig-a assay in L5178Y/Tk⁺⁻ cells is sensitive to standard mutagens and appropriately discriminates between mutagens and nonmutagens (Table XV; Bemis et al., 2018; David et al., 2018; Y Wang et al., 2018). Also, Bemis et al. (2018), Y Wang et al. (2018), and Revollo et al. (2017b) found that mutants detected in the assay almost always contain Pig-a mutations (see Section 6f, above), making it unlikely that any gene, other than Pig-a, is involved in the GPI-deficient mutational response detected in this cell line.

A recent study described a flow cytometric PIG-A assay conducted with HepG2 cells, an attached human liver cell line that retains some activity for promutagen activation (Kopp et al., 2018). If the genotype of mutants established as containing PIG-A mutations, this observation suggests that the long-held presumption that the Pig-a assay was only suitable for unattached cells (see Section 1b) may not be totally correct.

An additional Pig-a-type in vitro assay has been described by Nakamura et al. (2012) (Table XV). Like Pig-a assays, the assay measures GPI-anchor deficiency as the mutant phenotype. Because, however, the assay is conducted in DT-40 cells, which are chicken cells, the GPI biosynthesis gene that is present as a single copy on the sex chromosome (Z in chickens) is Pig-o and Pig-a is on an autosome. This means that Pig-o rather than Pig-a is the primary reporter for mutation in this assay. The assay also is conducted by limiting dilution cloning and proaerolysin selection.
rather than flow cytometry. MMS treatment of DT-40 cells resulted in an increase in proaerolysin-resistant mutants, and these mutants were shown to contain mutations in the Pig-a gene.

11. Animal welfare issues associated with the Pig-a assay

A compelling argument for developing an OECD TG for the erythrocyte Pig-a assay is its ability to make better use of existing animal resources devoted to toxicology testing, thus conforming with the principles commonly referred to as the 3Rs (replacement, reduction, refinement). The 3Rs principles recommend adopting non-animal methods (replacement), methods to obtain comparable information using fewer animals or gain more information from the same number of animals (reduction), and using methods to alleviate or minimize potential pain, suffering or distress (refinement) (Russell and Burch, 1959). Invasive tissue sampling or animal sacrifice is not necessary to conduct the erythrocyte Pig-a assay, consistent with refinement, and repeated sampling of the same animals greatly reduces the number of animals necessary for establishing the temporality of a response, as might occur when evaluating the cancer mode of action of a test substance (Moore et al., 2008), consistent with reduction. Section 3 describes how the characteristics of the assay make it easier to integrate the Pig-a assay into existing general toxicology studies than most other genotoxicity assays, another application that results in reduction. As an example, the U.S. National Toxicology Program has recently decided to incorporate the Pig-a assay along with the peripheral blood MN assay into their routine rodent toxicity tests to make better use of the animals (K Witt, personal communication).

Data generated with the Pig-a assay indicate that, with proper training, the reproducibility and transferability of the Pig-a assay are impressively high (Gollapudi et al., 2015; a discussion of interlaboratory trials is in Part 2). Training and a subsequent demonstration of laboratory proficiency are important elements to implementing any assay, and the OECD has established guidelines for how that should be done for various genetic toxicology assays, including in vivo assays (OECD, 2015). This can be especially challenging when a new assay is introduced into a
laboratory since adopting a new assay typically requires several experimental trials to learn the methodology, educate laboratory staff, and demonstrate that the performance of the assay is adequate for generating acceptable data. In case of *in vivo* tests, this learning and ‘laboratory validation’ period is ordinarily at direct odds with the 3Rs (Russell and Burch, 1959).

An in-house validation procedure that involved only 10 rats was described by Godin-Ethier et al. (2015). Raschke and colleagues (2016) described reconstruction experiments (also known as spiking experiments) that enable laboratories to measure mutant phenotype cells of varying frequencies, and to compare the observed results to expected frequencies. These experiments minimize the focus on steps that are not of high importance for laboratory proficiency (*e.g.*, ability to successfully treat animals with test substances), and focus on those elements that are considered critical to conducting the assay — sample handling/labeling, and flow cytometric analysis.

The Raschke study described two approaches that were employed by four laboratories — (a) CD59 masking of blood from an untreated animal, and (b) using blood from a mutagen-treated animal. The CD59 masking technique was developed and utilized by investigators at Bayer Pharma and did not require prior treatment of animals; thus, there was no need to wait for mutant manifestation. The whole experimental procedure could be performed in one day, and performed repeatedly, providing flexibility for training purposes. Mutant-mimic samples were generated by blocking the CD59 epitope using non-fluorescently labelled anti-CD59 antibody (CD59 masking), thereby preventing any CD59-associated fluorescence during flow cytometry. Varying amounts of the fully CD59-masked sample were spiked into aliquots of unmasked blood (peripheral blood from an untreated animal) to achieve very low to moderately elevated erythrocyte mutant frequencies. A second approach, which used two animals, was utilized by GSK, Covance, and Litron Laboratories. In this approach, single rats were administered 0 or 20 mg ENU/kg/day via oral gavage for three consecutive days. Approximately 1.0 mL blood was collected from each animal approximately 4 weeks after the treatment and the blood samples were mixed in different ratios to create samples with minimum to moderately elevated erythrocyte mutant frequencies.
These reconstruction experiments were conducted independently in four laboratories and showed good overall precision (correlation coefficients >0.99) and accuracy (estimated slope: 0.71–1.09) of mutant cell scoring measured using the Bland and Altman method (Bland and Altman, 1999). These data strongly support the use of reconstruction experiments for training purposes and demonstrating laboratory proficiency with very few animals.

Finally, although the current recommended protocol described in Section 4 does not include running a positive control, positive controls are often included with genetic toxicology assays. The method for preserving samples described in Section 4b allows including a positive control in addition to a mutant mimic standard in the flow cytometric analysis of Pig-a assay samples without including additional positive control animals in a study. Preserving samples also has the potential advantage of generating Pig-a data from previous studies where a measure of mutation was not considered important at the time the study was conducted.

12. Summary and conclusions

1. An in vivo gene mutation assay has been developed that is based on measuring the loss of cell surface proteins caused by mutation in the endogenous, X-linked Pig-a gene.

2. In theory, the assay can be conducted in any animal of toxicological interest. To date, it has been conducted in mice, rats, rhesus monkeys and humans; it is most fully developed in rodents. The ability to conduct the assay in humans, the species of greatest toxicological interest, is a major advantage. The assay evaluates hazard, but together with exposure data, it may be applicable for risk evaluation and assessment.

3. Most Pig-a assay data are for peripheral blood erythrocytes. The use of erythrocytes for conducting the assay imparts advantages in terms of speed, economy, and the efficient use of animals.

4. The gene mutations measured in the assay are induced in nucleated bone marrow erythroid cells. Because many of these cells are long-lived and have extensive replication
potential, mutations induced in the assay result in elevated frequencies of mutant peripheral blood erythrocytes that persist over many months.

5. The mutations detected in the rodent assay appear to have a neutral phenotype; mutant frequencies accumulate with repeated dosing.

6. Consensus protocols have been developed and evaluated in several interlaboratory trials.

7. More than 90 test substances and vehicle controls have been evaluated in rats and mice using the assay. Negative responses were generated either by known nongenotoxicants or are consistent with the properties of the assay. All positive responses were induced by test substances known to be gene mutagens. The assay detected the mutagenicity of several genotoxic substances that are difficult to detect in blood-based MN assays (DEN, CEDU, aristoclochic acids), and several genotoxicants that are primarily known as clastogens (cisplatin, ionizing radiation, cyclophosphamide).

8. Test data support consensus protocol recommendations, including using a repeat dose treatment protocol, assaying both total RBCs and RETs, and assaying as many RETs and RBCs as practical. Alternative treatment protocols have been used, and in most cases performed well; therefore, short-term (including acute) treatment protocols can be used with justification.

9. The test data also indicate that the assay is complementary to the erythrocyte in vivo MN assay, a combination that results in a more comprehensive analysis of in vivo genotoxicity than can be done using either test by itself.

10. A retrospective performance assessment conducted with curated Pig-a and conventional test response data indicated that the Pig-a assay is highly accurate when responses in the bone marrow TGR assay and rodent cancer in hematopoietic tissues are used to establish true positive and negative responses. The accuracy of the assay was lower, most clearly seen as lower for negative predictivity, when in vivo MN responses, TGR responses in all tissues, and cancer responses in all tissues were used to define true positive and negative responses. In general, within the limits of the available data, the assay performed well as an assay for detecting in vivo gene mutation.
11. The results of a recent evaluation of curated TGR and comet assay data contrasted with the findings of the current performance analysis for Pig-a gene mutation. While Pig-a mutation was highly accurate for identifying compounds that were positive and negative in the bone marrow TGR assay, the comet assay was relatively poor (Kirkland et al., 2019). Although the amount of data available for evaluating comet assay performance was even lower than for evaluating Pig-a assay performance (15 compounds for comet, 26-28 for Pig-a), the comparison suggests the value of using a gene mutation assay for identifying mutagenic substances.

12. The current regulatory applications of the test are: 1) a test for evaluating the in vivo risks associated with positive responses in in vitro gene mutation screening assays, and 2) an in vivo gene mutation test that can be integrated with the in vivo MN test into repeat-dose toxicology studies to provide information about gene mutation in combination with chromosomal damage and loss. Current guidances and other documents anticipate use of Pig-a data for regulatory decision-making.

13. In vitro Pig-a assays have been developed that should be useful for evaluating substances for in vivo testing and addressing hypotheses that are difficult or unethical to test in vivo.

14. It is envisioned that, as the human PIG-A assay becomes more widely used, laboratory studies that interface with human studies using this same gene as a reporter of mutation will become increasingly important for regulatory decision-making.

15. The characteristics of the erythrocyte assay lend itself to integration into in vivo general toxicology tests, making the assay consistent with the 3Rs principles of reduction and refinement of animal use. Responses in the Pig-a assay are dependent on the cumulative dose of the test substance, rather than the daily dose, making integration into repeat-dose studies (e.g., 28- or 90-day general toxicology studies) a particularly attractive approach to maximizing the sensitivity of measuring gene mutation in vivo.

Acknowledgements
In addition to the authors listed on the title page, all of whom contributed to the drafting of this manuscript and the gathering and analysis of data, this document benefited from suggestions and discussion from IP Pogribny, GR Douglas, E Zeiger, L Stankowski, A Zeller, PA White, AM Lynch, and BB Gollapudi. We also thank Yang Luan and Yiyi Cao (Shanghai Jiao Tong University, China), A Zeller (F. Hoffman-La Roche, Switzerland), Y-S Chung (Hoseo University, South Korea), F Nesslany (Pasteur Institute-Lille, France), J Kenny (GlaxoSmithKline, UK), L Custer and C Gleason (Bristol-Myers Squibb, USA), B-j Van der Leede (Janssen, Belgium), A Elhajouji (Novartis, Switzerland), F Marchetti (Health Canada), M LeBaron and Z Ji (Dow and Bristol-Myers Squibb, USA), and A Dad and JR Revollo (U.S. FDA/NCTR) for allowing us to refer to their unpublished data.
References

Abkowitz JL, SN Catlin, MT McCallie, P Guttorp (2002) Evidence that the number of hematopoietic stem cells per animal is conserved in mammals, Blood 100:2665-2667.


Boverhof DR, MP Chamberlain, CR Elcombe, FJ Gonzalez, RH Heflich, LG Hernandez, AC Jacobs, D Jacobson-Kram, M Luijten, A Maggi, MG Manjanatha, J van Benthem, BB


Cordelli E, J Keller, P Eleuteri, P Villani, L Ma-Hock, M Schulz, R Landsiedel, F Pacchierotti (2017) No genotoxicity in rat blood cells upon 3- or 6-month inhalation exposure to CeO₂ or BaSO₄ nanomaterials, Mutagenesis 32:13-22. DOI: 1093/mutage/gew005.


EFSA (European Food Safety Authority) (2011) Scientific opinion on genotoxicity testing strategies applicable to food and feed safety assessment, EFSA Journal 9:2379.


Gad SC, CB Spainhour, C Shoemake, DR Stackhouse Pallman, A Stricker-Krongrad3, PA Downing, RE Seals, LA Eagle, K Polhamus, J Daly (2016) Tolerable levels of nonclinical
vehicles and formulations used in studies by multiple routes in multiple species with notes on methods to improve utility. Inter J Toxicol 35:95-178.


IQ 3Rs Leadership Group (2016) Recommended Dose Volumes for Common Laboratory Animals. Available at: https://iqconsortium.org/images/LG-3Rs/IQ-CRO_Recommended_Dose_Volumes_for_Common_Laboratory_Animals_June_2016_(2).pdf


Kenyon MO, LS Coffing, JI Ackerman, WC Gunther, SD Dertinger, K Criswell, KL Dobo (2015) Compensatory erythropoiesis has no impact on the outcome of the in vivo Pig-a mutation assay in rats following treatment with the haemolytic agent 2-butoxyethanol, Mutagenesis 30:325-334.


Kimoto T, K Suzuki, Xm Kobayashi, VN Dobrovolsky, RH Heflich, D Miura, Y Kasahara (2011b) Manifestation of Pig-a mutant bone marrow erythroids and peripheral blood erythrocytes in mice treated with N-ethyl-N-nitrosourea: Direct sequencing of Pig-a cDNA from bone marrow cells negative for GPI-anchored protein expression, Mutat Res 723:36-42.


Kimoto T, K Horibata, D Miura, S Chikura, Y Okada, A Ukai, S Itoh, S Nakayama, H Sanada, N Koyama, S Muto, Y Uno, M Yamamoto, Y Suzuki, T Fukuda, K Goto, K Wada, T Kyoya,

Kinoshita T (2014) Biosynthesis and deficiencies of glycosylphosphatidylinositol, Proc Jpn Acad, Sec B 90:130-143.


Kulkarni R (2016) Regulatory acceptance of Pig-a assay for an API, Presentation at EMGS Pig-a Interest Group meeting, Kansas City, MO.


Labash C, K Carlson, SL Avlasevich, A Berg, JC Bemis, JT MacGregor, SD Dertinger (2015c) Induction of Pig-a mutant erythrocytes in male and female rats exposed to 1,3-propane sultone, ethyl carbamate, or thiotepa, Mutat Res 782:24-29.


OECD (Organisation for Economic Cooperation and Development) (2016c) OECD Guideline for the testing of chemicals: In vitro mammalian cell gene mutation tests using the
20 May 2019


Piberger AL, CT Krüger, BM Strauch, B Schneider, A Hartwig (2017) BPDE-induced genotoxicity: relationship between DNA adducts, mutagenicity in the in vitro PIG-A assay, and the transcriptional response to DNA damage in TK6 cells, Arch Toxicol. DOI: 10.1007/s00204-017-203-0.


Recio L, K Shepard, J Davis, M Friedman, D Marroni, C Hobbs (2016) Assessment of genotoxic and mutagenic potential of acrylamide in rats and mice by use of MN and Pig-a assays, Toxicologist


Stankowski LF Jr, MJ Aardema, TE Lawlor, K Pant, S Roy, Y Xu, R Elbekai (2015) Integration of Pig-\(a\), micronucleus, chromosome aberration and comet assay endpoints in a 28-day rodent toxicity study with urethane, Mutagenesis 30:335-342.


Paroxysmal nocturnal hemoglobinuria and concurrent JAK2\textsuperscript{V617F} mutation, Blood Cancer J 2:e63.


Wang L, W Xu, L Ma, S Zhang, K Zhang, P Te, G Xing, X Zhang, Y Cao, J Si, J Gu, Y Luan (2017) Detoxification of benzo[a]pyrene primarily depends on cytochrome P450, while bioactivation involves additional oxidoreductases including 5-lipoxygenase, cyclooxygenase, and aldo-keto reductase in the liver, J Biochem Mol Toxicol e21902.


Ware RE, TA Howard, T Kamitani, H-M Chang, ETH Yeh, MF Seldin (1994) Chromosomal assignment of genes involved in glycosylphosphatidylinositol anchor biosynthesis:
impli
cations for the pathogenesis of Paroxysmal Nocturnal Hemoglobinuria, Blood 83:3753-3757.


Yauk CL, MJ Aardema, J van Benthem, JB Bishop, KL Dearfield, DM DeMarini, UE Dubrova, M Honma, JR Lupski, F Marchetti, ML Meistrich, F Pacchierotti, J Stewart, MD Waters,


