Received: 19 August 2022 | Revised: 30 December 2022 | Accepted: 6 January 2023

DOI: 10.1002/em.22527

Environmental and Molecular Mutagenesis | Environmental Mutagenesis and Genomics Society | WILEY

RESEARCH ARTICLE

Assessing testicular germ cell DNA damage in the comet assay; introduction of a proof-of-concept

Yvette Dirven[1,2] | Dag Markus Eide[1,2] | Erika Witasp Henriksson[3,4] | Rune Hjorth[5] | Anoop Kumar Sharma[6] | Anne Graupner[1,2] | Gunnar Brunborg[1,2] | Jarle Ballangby[1,2] | Anne Mette Zenner Boisen[6] | Stellan Swedmark[4] | Kristine Bjerve Gützkow[1,2] | Ann-Karin Olsen[1,2]

[1]Norwegian Institute of Public Health, Division of Climate and Environmental Health, Oslo, Norway

[2]Centre for Environmental Radioactivity (CERAD, Centre of Excellence of the Norwegian Research Council), Oslo, Norway

[3]Swedish Chemicals Agency, Department of Development of Legislation and Other Instruments, Unit of Proposals for Classification and Restriction, Sundbyberg, Sweden

[4]Swedish Chemicals Agency, Department of Development of Legislation and Other Instruments, Unit of Evaluation of Substances, Sundbyberg, Sweden

[5]The Danish Environmental Protection Agency, Odense, Denmark

[6]Technical University of Denmark, National Food Institute, Lyngby, Denmark

Correspondence
Ann-Karin Olsen, The Norwegian Institute of Public Health, Division of Climate and Environmental Health, P.O. Box 222, Oslo, Norway.
Email: ann-karinhardie.olsen@fhi.no

Funding information
Nordic Council of Ministers, Grant/Award Number: 2021-022,2021-101; Research Council of Norway through its Centers of Excellence funding scheme, Grant/Award Number: 223268/F50 CERAD; Swedish Chemicals Agency

Accepted by: F. Marchetti

Abstract

The *in vivo* comet assay is widely used to measure genotoxicity; however, the current OECD test guideline (TG 489) does not recommend using the assay to assess testicular germ cells, due to the presence of testicular somatic cells. An adapted approach to specifically assess testicular germ cells within the comet assay is certainly warranted, considering regulatory needs for germ cell-specific genotoxicity data in relation to the increasing global production of and exposure to potentially hazardous chemicals. Here, we provide a proof-of-concept to selectively analyze round spermatids and primary spermatocytes, distinguishing them from other cells of the testicle. Utilizing the comet assay recordings of DNA content (total fluorescence intensity) and DNA damage (% tail intensity) of individual comets, we developed a framework to distinguish testicular cell populations based on differences in DNA content/ploidy and appearance. Haploid round spermatid comets are identified through (1) visual inspection of DNA content distributions, (2) setting DNA content thresholds, and (3) modeling DNA content distributions using a normal mixture distribution function. We also describe an approach to distinguish primary spermatocytes during comet scoring, based on their high DNA content and large physical size. Our concept allows both somatic and germ cells to be analyzed in the same animal, adding a versatile, sensitive, rapid, and resource-efficient assay to the limited genotoxicity assessment toolbox for germ cells. An adaptation of TG 489 facilitates accumulation of valuable information regarding distribution of substances to germ cells and their potential for inducing germ cell gene mutations and structural chromosomal aberrations.

KEYWORDS
genotoxicity, OECD TG 489, primary spermatocyte, round spermatid, total fluorescence intensity

1 | INTRODUCTION

There is an increasingly scientific and regulatory demand for effective, rapid, and low-cost assays to assess genotoxicity in male germ cells with minimal use of test animals. This growing demand is related to the increased production and use of potentially hazardous chemicals (UN Environment Programme, 2019), and the associated human exposure. Moreover there is increased concern of potential harmful implications of preconception exposure of gametes to genotoxic substances (Brevik et al., 2012; Marcho et al., 2020; Zhang et al., 2020), as DNA damage is implicated in a plethora of adverse effects, of particular importance in germ cells. Exposure to a genotoxic agent can lead to mutations that can be transferred to the offspring, contributing to genetic disease, and impacting the health of coming generations (Dearfield et al., 2002; Griffiths et al., 2000; Marchetti et al., 2020; Yauk et al., 2015).

Today, few validated and accepted methods are available for the investigation of genotoxicity in male germ cells for chemical risk assessment from a regulatory perspective. The Organization for Economic Co-operation and Development (OECD) has published several internationally recognized guidelines for the testing of chemicals. Current *in vivo* methods, with approved OECD test guidelines (TG) for the detection of germline mutagenicity, include the rodent dominant lethal test (TG 478), mouse heritable translocation assay (TG 485), mammalian spermatogonial chromosomal aberration test (TG 483), and transgenic rodent mutation (TGR) assay (TG 488). These assays are generally labor-intensive, time-consuming, expensive, and hence rarely used. TG 478 and TG 485 both require large numbers of animals, which is not in accordance with the 3R's principle (replace, reduce, and refine). TG 483 and TG 488 require limited numbers of animals; however, TG 483 has low sensitivity. TG 488 has high sensitivity and specificity and was recently revised to obtain reliable data from epididymal spermatozoa, using a 28 + 28d exposure regimen to cover spermatogenesis (OECD, 2020). This recent revision will hopefully facilitate increased use of this assay as it has historically rarely been used as it is considered rather expensive, resource demanding, and requires transgenic animals available through few sources (ECHA, 2012).

The *in vivo* comet assay (TG 489) is cost-effective, easy to perform, requires the similar limited number of animals as the TG 488, is easily transferrable to new laboratories, has a high sensitivity for detecting DNA damage, and shows high specificity and few false-positive cases (Azqueta & Collins, 2013; EFSA, 2011; Langie et al., 2015; Singer & Yauk, 2010; Speit et al., 2015; Yauk et al., 2015). About a decade ago, the comet assay was reported to be more sensitive than both the TG 488 and rat liver unscheduled DNA synthesis (UDS) test (TG 486) for detecting rodent carcinogens that were negative in the conventional bone-marrow micronucleus assay (Kirkland & Speit, 2008). A recent comparison of TG 489 and TG 488 shows that the two assays generally identify the same compounds as genotoxic in liver, stomach, and colon (Kirkland, Levy, et al., 2019a).

The alkaline comet assay (single-cell gel electrophoresis) measures single-strand DNA breaks and alkali labile sites that are transformed into strand breaks under alkaline conditions (Singh et al., 1988; Cordelli et al., 2021). In short, cells are molded in agarose followed by lysis and deproteinization under alkaline conditions, leaving DNA nucleoid structures. The nucleoids are subjected to alkaline electrophoresis (pH > 13) to reveal DNA lesions prior to fluorescent staining of the DNA to visualize structures resembling comets. During alkaline electrophoresis, the broken or damaged DNA strands migrate from the head of the nucleoid structure creating a comet tail. The degree of DNA migration correlates with the level of DNA damage (Singh et al., 1988; Collins, 2004). DNA damage is most often recorded as the fraction of DNA in the tail (% tail intensity; % TI), as recommended in OECD TG 489.

Since its development in the late 1980s, the comet assay has been widely used to measure DNA damage and repair. The OECD test guideline (TG 489) was adopted in 2014; however, at this time the assay was not considered appropriate for specific measurements of the germ cells of the testis. The reasoning for this was that cell suspensions derived from the gonads contain a mixture of cells of both somatic and germ-cell origin, so testicular germ-cell-specific information could not be deduced (OECD, 2016). Moreover, the mature germ cells (spermatozoa) demonstrate variable background DNA damage levels when measured using the comet assay and may require specially adapted exposure regimens. The OECD suggests that protocol modifications and validation studies are necessary before TG 489 can be revised to include an analysis of testicular germ cells. Nevertheless, the comet assay has been used to measure DNA damage in gonadal cells in both research settings (Asare et al., 2016; Bjorge et al., 1995; 1996a; Brunborg et al., 2015; Olsen et al., 2001; 2003; Collins, 2004; Olsen et al., 2005; Olive & Banath, 2006; Speit & Hartmann, 2006; Dhawan et al., 2009; Hansen et al., 2010; 2014; Graupner et al., 2014; 2015; Gutzkow et al., 2016 Graupner et al., 2017; Sharma et al., 2018) and in a regulatory perspective (Brendler-Schwaab et al., 2005; Vasquez, 2012; Graupner et al., 2014; Frotschl, 2015; Koppen et al., 2017).

Here, we propose a methodological concept that enables specific analysis of genotoxicity in testicular germ cells, a framework that can be incorporated into the current TG 489. The concept is based on the unique DNA content (ploidy), shape, and size of different testicular germ cell nucleoids, enabling identification of specific germ cell populations in mixed testicular cell suspensions. Moreover, this approach allows both somatic and germ-cell genotoxicity data to be obtained from the same animal. Utilizing this concept will increase the availability of germ cell genotoxicity data used for identification of hazardous agents and extend the understanding of germ cell mutagens compared with somatic cell mutagens. The framework can also be used to understand genotoxic responses in specific germ cells, in a fundamental research perspective.

2 | MATERIALS AND METHODS

2.1 | Collection of historical testis comet data and comet assay protocol details

Data from a previously published in-house study were used to demonstrate the concept to identify and analyze specific germ cell comets

among a mixture of testicular cells. In this in-house study, the genotoxic effects of selenium deficiency were investigated (Graupner et al., 2015). Graupner and coworkers performed an in vivo comet assay with testicular cell suspensions from 8 to 11 weeks old Big Blue™ C57BL/six mice deprived of the antioxidant and essential element selenium (Se) through two generations. Testes from 14 animals were used in the study; eight mice were given a low selenium diet (0.01 mg Se/kg diet) and six mice were given a normal selenium diet (0.23 mg Se/kg diet). Selenium deprivation led to significantly increased overall DNA damage levels in mixed cell populations from the testes, as well as in lung, and blood compared to mice given a normal diet with adequate selenium. The complete protocol for tissue collection and the comet assay is described by Graupner et al., 2015, and a high throughput version of the alkaline comet assay was used (Gutzkow et al., 2013).

Briefly, testicular cell suspensions were generated by cutting the testis into small pieces in ice-cold Merchant's buffer (0.14 M NaCl, 1.47 mM K_2PO_4, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 10 mM Na_2EDTA). The testis pieces were transferred to a tube and mechanically squeezed through a stainless-steel screen (0.4 mm) using a plunger, followed by filtering through cotton gauze and nylon mesh (100 µm). The cell density of the testicular cell suspensions was adjusted to 1×10^6 cells/ml, followed by 1:10 dilution in 0.75% low melting point agarose at 37°C. Triplicates of cell suspension (4 µl) from each animal were molded on a cold GelBond© film. The cells were lysed and deproteinized overnight in a lysis solution (2.5 M NaCl, 0.1 M Na_2EDTA, 0.01 M Tris-base, 0.2 M NaOH, 0.034 M *N*-Lauroylsarcosine, 10% DMSO, 1% TritonX-100, pH 10). The films were transferred to an alkaline electrophoresis solution (0.3 M NaOH, 0.001 M Na_2EDTA, pH > 13) for 40 min to allow the DNA to unwind, followed by alkaline electrophoresis for 20 min at 8°C and 0.8 V/cm (25 V between electrodes) in fresh circulating electrophoresis solution. Following electrophoresis, the films were rinsed in water for 1 min followed by 2 × 5 min in neutralization solution (0.32 M Tris-HCl, 0.08 M Tris-base, pH 7.5). The films were fixed in ethanol and dried prior to DNA staining in TE-buffer with SYBR©Gold Nucleic Acid Gel Stain (10,000× dilution) for 20 min.

Comets were scored using the Comet assay IV (Perceptive Instruments Ltd., Suffolk, UK). During scoring, at least 50 randomly selected comets were scored per gel, with three gels per animal (technical replicates), for a total of at least 150 comets scored per animal. Elongating/elongated spermatids and spermatozoa were not scored, based on their visually distinct nuclear shapes (see Figure 1d). Elongating spermatids have an elongated nucleus where the chromatin is increasingly condensed to one side of the nucleoid, like a crescent, whereas spermatozoa have a small nucleoid with a highly condensed appearance. Only comets with a round nucleus were scored; the roundness of each comet was judged visually during scoring in the original study (Graupner et al., 2015). It was not possible to quantify DNA damage levels in primary spermatocytes per se because we used a historical dataset. Primary spermatocytes must be identified during scoring based on their large nucleoid and high DNA content. The focus of this proof-of-concept is therefore on identification of haploid round spermatids.

2.2 | Identification and modeling of testicular cell populations

Frequency distribution histograms of the total DNA content of the comets (total intensity values) were created for each animal, using a class width of 10,000. Two approaches were used to identify the testicular cell populations based on the distributions of the total intensity values: visual inspection and modeling. Using visual inspection, as briefly discussed in our previous publications (Olsen et al., 2001; 2003; Brunborg et al., 2015) and (Zheng & Olive, 1997), the main peaks on the distribution histogram were defined as comet populations.

The data were modeled using a normal mixture distribution function. Nonrelevant comet registrations (data points with very high or low total intensities) were excluded using quantile-based flooring and capping. In short, several combinations of limits at the lower and/or upper quantiles were tested, starting at the outer quantiles and increasing the range, excluding more and more comets for each combination. For each combination, the model was refitted to the distribution histograms. Typically, the lower quantile was <0.5 and had to be set carefully to avoid excluding relevant comets. The upper quantile cut-off was set to 0.92. The upper cut-off is more adjustable because these comets never will be included in the 1N cell population. Once the model was consistently able to identify the same comet populations as observed using visual inspection in all animals, combined with low standard deviations (σ) of the identified populations, the comets in the excluded quantiles were marked as outliers in the dataset and excluded from further processing. Identical quantile limits were set for data from all animals. The number of comets that need to be excluded for the model to identify the most likely cell populations depends on the distribution of the data. The iterative process of removing comets from the outer quantiles of the total intensity distribution, reapplying the normal mixture distribution function and visually confirming that the model fits the observed cell populations should be performed on each unique dataset, until the most appropriate model describing the total intensity distributions of the three cell populations in all animals is identified.

2.3 | Transformation of total intensity values

To potentially improve the separation between populations the raw total intensity values were transformed using square root, cube root, and logarithmic transformations. Following transformation, distribution histograms were created for each animal and the cell populations were identified as described above.

2.4 | Standardization of total intensity values

The total intensity values were standardized to facilitate comparison between gels, animals, treatment groups, and experiments. Standardization was performed using a simple scaling factor or Z-score standardization. When using the scaling factor, the total intensity values

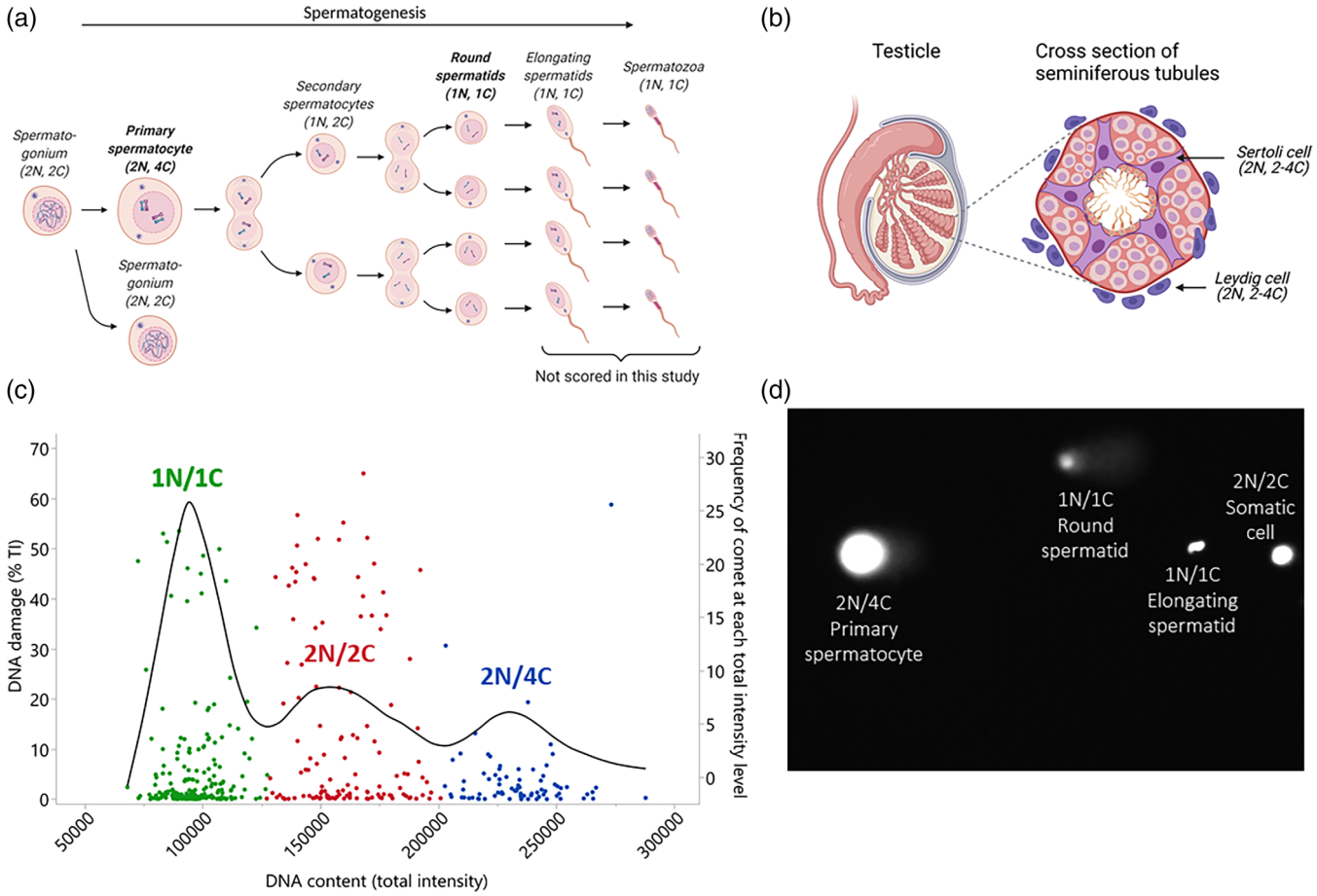

FIGURE 1 Spermatogenesis and DNA content. (a) During spermatogenesis primary spermatocytes (2N, 4C) undergo meiosis I to produce two secondary spermatocytes (2N, 2C) which subsequently divide during meiosis II to produce haploid round spermatids (1N, 1C) that further differentiate into elongating spermatids and ultimately spermatozoa. In the data presented here, round spermatids are selectively scored, since cells in later stages of spermiogenesis are excluded based on visual appearance. Created with BioRender). (b) The testicle contains seminiferous tubules where spermatozoa are produced through the process of spermatogenesis. The testicle contains germ cells at different stages of spermatogenesis, as well as somatic cells (2-4N, 2-4C), such as Sertoli and Leydig cells (Created with BioRender). (c) Cells in a testicular cell suspension exhibit different DNA content (measured as total intensity in the comet assay) which can be used to identify the 1N/1C cells (green; round spermatids), 2N/2C cells (red), and the 2-4N/4C cells (blue; primary spermatocytes and somatic cells). (d) Comet image illustrating the size and shape difference of the different testicular cell types.

from each animal were divided into quantiles. Afterward, the total intensity values were scaled so that the total intensity of a specific quantile was scaled to an arbitrary total intensity value. In our case, the 40% quantile was set to an arbitrary total intensity value of 100,000, and all the comet total intensities for each animal were scaled accordingly following the formula:

$$\text{total intensity value} * \left(\frac{\text{arbitraty total intensity value}}{\text{total intensity of quantile}}\right).$$

Z-score standardization was achieved based on the mean (μ) and standard deviation (σ) values of the leftmost population identified by the model, which also had the lowest total intensity, representing the haploid germ cell population (round spermatids). The total intensity values for each animal were standardized using the Z-score formula:

$$\frac{\text{total intensity value} - \text{mean}\,(\mu)}{\text{standard deviation}\,(\sigma)}.$$

2.5 | Identification of haploid germ cell comets

Three approaches were used to identify and differentiate comets belonging to the haploid germ cell population, hence distinguishing them from the other cell populations hereafter referred to as "other cells." For each approach, the % TI of comets were plotted along with their total fluorescence intensities, as separate plots for each animal.

I. The first approach for identifying the haploid germ cell population was based on the fitted normal mixture distribution function of Z-score standardized total intensity values. All comets with a Z-score standardized total intensity value lower than the mean + 1 standard deviation (SD) were considered haploid germ cells.

II. The second approach was based on identifying the haploid germ cells through visual inspection of the % TI (*y*-axis) against total intensity plots (*x*-axis). A line showing the frequency of comets (in small total intensity intervals) was added to the plots, and the comet population with the lowest total intensity was recognized as the haploid germ cells. Visual identification of the haploid germ cells was performed separately for each animal.
III. The third approach consisted of setting a conservative total intensity threshold, based on the expected distribution of comets in each of the testicular cell populations. Unexposed testicular cell suspensions consist of 60%–80% haploid germ cells (Bjorge et al., 1996b; Hansen et al., 2010; Rodriguez-Casuriaga et al., 2013; Brunborg et al., 2015). Therefore, a conservative cut-off at a lower level would identify haploid germ cells, with a very low likelihood of including comets from higher ploidy cell populations. The total intensity values for each animal were divided into quantiles, and all comets with a total intensity < 40% quantile were considered haploid.

2.6 | Statistical analyses

Data analysis was performed using the JMP Pro 15 software (JMP, Version 15.2.0 SAS Institute Inc., Cary, NC). % TI values were summarized as medians per gel, and the medians of the three gels were summarized as mean per animal (Bright et al., 2011; OECD, 2016). The total intensity data from each animal were modeled using a normal mixture distribution which fits the multimodal data and estimates a separate mean, standard deviation, and proportion for each population following exclusion of outliers, normalization, and standardization of data, through several iterations. Standardization of total intensity values between animals was tested using Tukey's test. Wilcoxon rank of sums was used to investigate differences in % TI. To obtain adjusted % TI values for the experimental animals, the mean % TI of the control animal group was subtracted from the mean % TI of each individual experimental animal. This was done separately for the comets identified as haploid cells or the other cells.

3 | RESULTS

3.1 | Concept for identifying round spermatids and primary spermatocytes in testicular cell suspensions

This proof-of-concept study relies on the assumption that the total intensity of a comet reflects the total amount of DNA in each cell. When performing the comet assay, both DNA damage (% TI) and DNA content (total intensity) of each individual cell is measured using software-based scoring techniques. This has the logical implication that comets with different DNA content can be identified based on their total intensity values. Additionally, the physical size and shape of each comet are easily observed during comet scoring. We exploit these two features to gather genotoxicity data from specific populations of testicular germ cells.

During spermatogenesis male germ cells exhibit different ploidy (*N*, number of complete sets of chromosomes) and DNA content (*C*, amount of DNA in the cell). Germ cells in the later stages of spermatogenesis (round spermatids, elongating/elongated spermatids, and spermatozoa) are haploid (1N) with only one copy of each chromosome (1C) (Figure 1a). Germ cells at earlier stages of spermatogenesis (2N; somatic cells in S phase or G2) have 2C or 4C DNA content (Figure 1b). Sorting all scored comets from the testicular cell suspension according to their total intensities thus results in distinct populations of cells with different ploidy as illustrated in Figure 1c. We, and others, have previously used a simplistic version of this principle to obtain information regarding testicular germ cells at specific stages of spermatogenesis, both in rodent (Bjorge et al., 1996b; Zheng & Olive, 1997 Jansen et al., 2001; Olsen et al., 2001; 2003; 2005) and human testicular cells (Olsen et al., 2001; 2003; Brunborg et al., 2015).

Primary spermatocytes have a larger physical size than all other cells in a testicular cell suspension as well as a high DNA content (2N, 4C; corresponds to 2N duplicated chromosomes) (Johnson et al., 1999; Holstein et al., 2003). To retrieve primary spermatocytes' specific DNA damage measurements, we propose to selectively identify primary spermatocytes during scoring, based on these features (Figure 1d). As our study was developed using historical data, without possibility of physically examining each comet during scoring, we were unable to identify the primary spermatocytes in this dataset. Accordingly, the analyses described herein focus on identifying the haploid round spermatids in mixed testicular cell populations. Both primary spermatocytes and round spermatid have round nuclei, which can be visually distinguished from the more crescent-shaped nucleus of elongating spermatids and small, highly condensed nucleoid of rat spermatozoa.

3.2 | Identification and modeling of testicular cell populations

Raw comet total intensity data for each animal were plotted as distribution histograms. For most animals, the total intensity distribution histograms represented three populations of comets, easily identified by visual inspection. For some animals a few comet registrations had either very low or high total intensities, representing clear nonrelevant comet registrations, hampering the identification of the cell populations.

When modeling the three cell populations using the raw total intensity dataset from each animal, the normal mixture distribution function did not always fit the visually established total intensity distributions (Figure 2). For some animals, a few high-intensity data points were modeled as a separate population, with all remaining data points modeled as one separate, large population (Figure 2a). For other animals, data points with particularly low intensities were modeled as a distinct population, containing a limited number of data

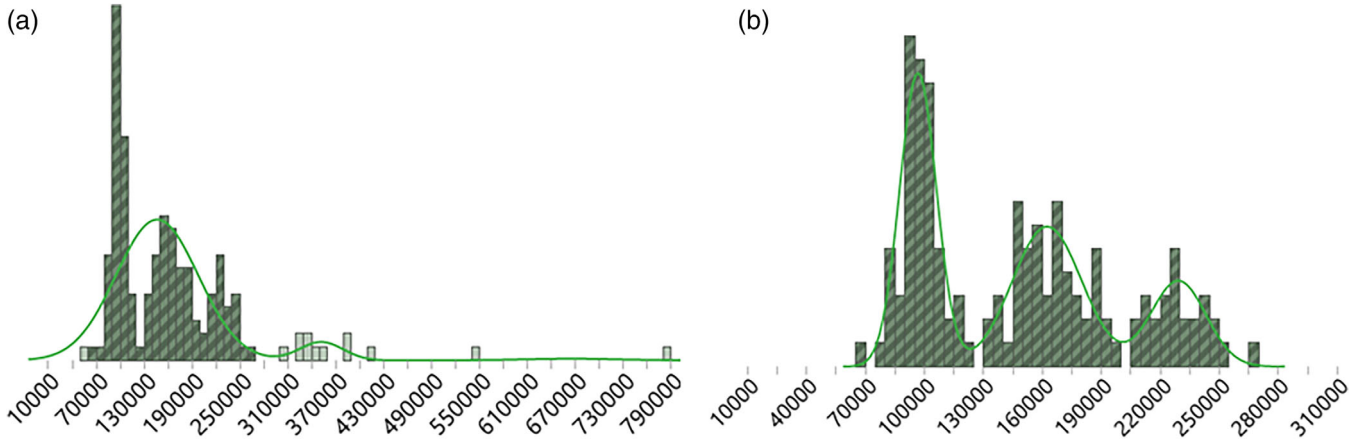

FIGURE 2 Distribution histograms of the normal mixture distribution before (left) and after (right) exclusion of nonrelevant data points. Data are shown for comet data from one representative animal. The x-axis shows the total intensities of the comets, the height of each column represents the frequency of comets within a specific total intensity interval; the green line shows the fitted three normal density distributions. In (a) shaded columns represent comets included for further analysis (seen in b)), while the light green columns represent nonrelevant data points. Distribution of total intensities and normal mixture distribution of the datasets (a) before removing nonrelevant datapoints, total # comets = 150 or (b) after exclusion of nonrelevant datapoints, total # comets = 137.

points. To overcome this, excluding the comet registrations in the 0.01 quantile, as well as comets with a total intensity higher than the 0.92 quantile, resulted in a normal mixture distribution model where the estimated mean (μ) and standard deviation (σ) for each population in the model fit well with the distribution histograms, without apparent exclusion of comets from the first (1N/1C) or third (2N, 2-4C) populations (Figure 2b).

3.3 | Transformation of total intensity values

The datasets were clearly left skewed-right tailed (Figure 3a, b). To improve separation between the cell populations, the data were transformed using square root, cube root, and logarithmic transformations. Generally, as illustrated by representative datasets in Figure 3, in this particular dataset transformation of the data of resulted in increased overlap between the first and second populations, due to larger peak dispersion. This was most pronounced for the cube root and logarithmic transformations that caused marked overlap in the experimental animals (Figure 3b). In general, there were no major differences between the untransformed and the square root transformed data; however, the dispersion of the haploid germ cell population was slightly larger for the square root transformed data, resulting in the conclusion that the untransformed data demonstrated the highest normality and resolution between populations and this was chosen for further analyses.

3.4 | Standardization of total intensities

Next, two methods were used for standardization of the total intensity data: quantile scaling and Z-score standardization. After quantile scaling, the mean total intensity values of all animals were no longer significantly different from each other ($p > .2$, Tukey's HSD test), whereas among the non-standardized total intensity data, 9 out of 91 animal pairs showed significantly different mean total intensities ($p > .05$, Tukey's HSD test).

When using Z-score normalization, the μ and σ values of the haploid germ cell population were used as parameters for standardization, meaning that only the total intensities of the haploid cells were truly standardized between animals. Our focus was on obtaining standardizing information regarding the haploid germ cell population, at the expense of the other populations. Using the μ and σ values of the haploid germ cells resulted in this population of comets being standardized to have a mean of 0 and a standard deviation of 1.

3.5 | Identification of haploid germ cell comets

Three approaches were attempted to select the haploid germ cells from the remaining cells in the testicular cell suspension (Figure 4). Depending on the nature of the comet data, and aim of study, prior standardization may not be necessary. Some of the approaches can use raw data directly (Figure 4a, b), while other approaches require prior data processing. Identification of the haploid germ cells based on visual inspection and the conservative cut-off can be performed on both standardized and non-standardized total intensity values. When basing the identification of the populations identified by the fitted normal mixture distribution function, Z-score standardized total intensity values should be used as this allows for faster processing, since the threshold for selecting the haploid population will be the same for all animals. For each approach, a cut-off/threshold is set to separate the haploid germ cell comets from the other comets in the testicular cell suspension.

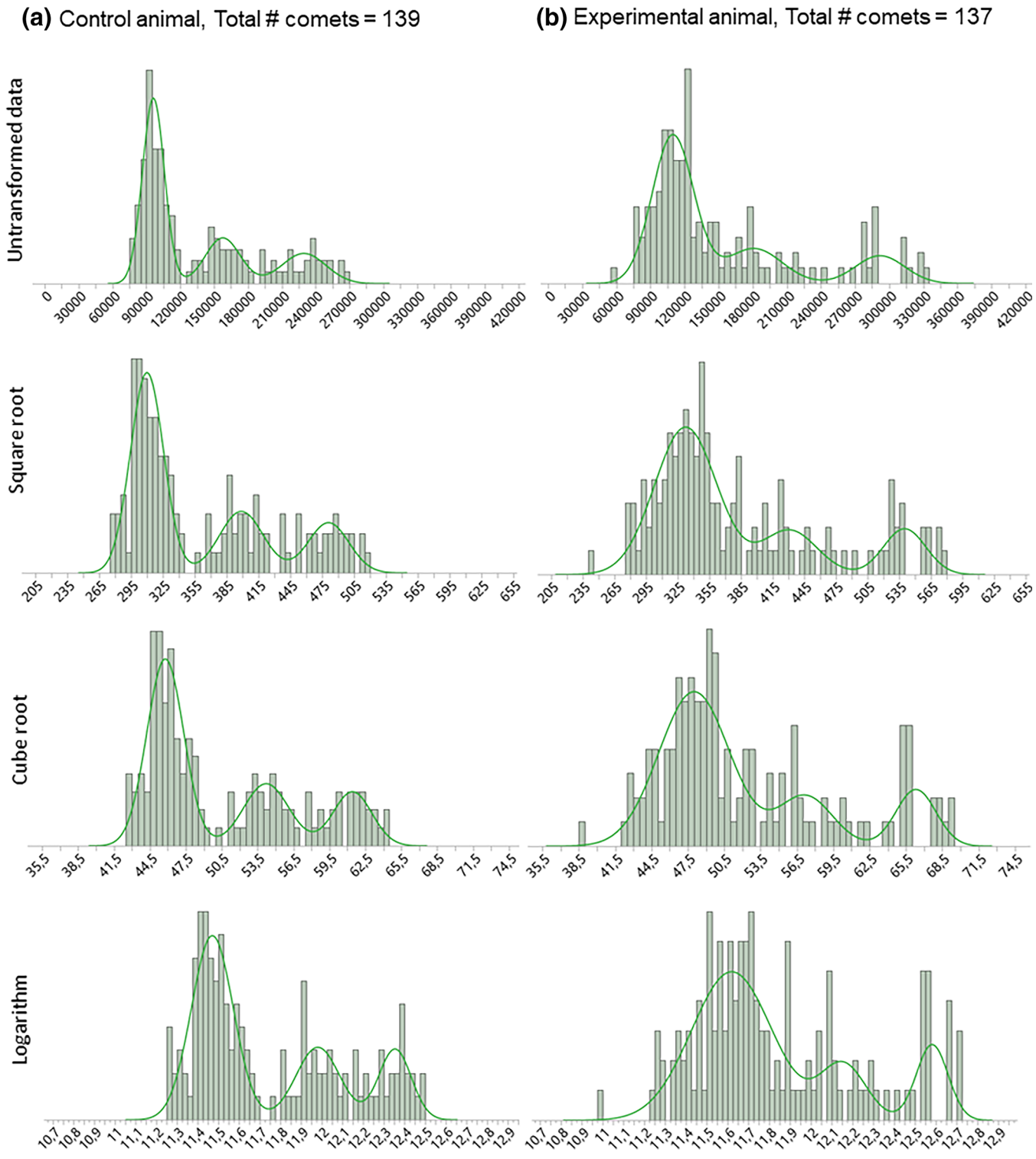

FIGURE 3 Distribution histograms of mixture distributions after different transformations of raw data. The *x*-axis shows the total intensities of the comets, and the height of the column represents the frequency of comets at specific total intensity intervals. The green line shows the fitted three normal density distribution, and the height of each column represents the number of comets in each column. The datasets are from a representative (a) control animal given a normal selenium diet and (b) experimental animal given a low selenium diet.

I. For the first approach the cut-off for haploid germ cells was set at mean + 1 SD, based on the μ and σ values obtained from the fitted distribution model. In other words, all comets with a total intensity <1 were considered haploid germ cells, whereas comets with total intensity >1 were labeled as "other cells" (Figure 4c, d).

II. The second approach is based on a visual inspection of total intensity distributions. The haploid germ cell population was

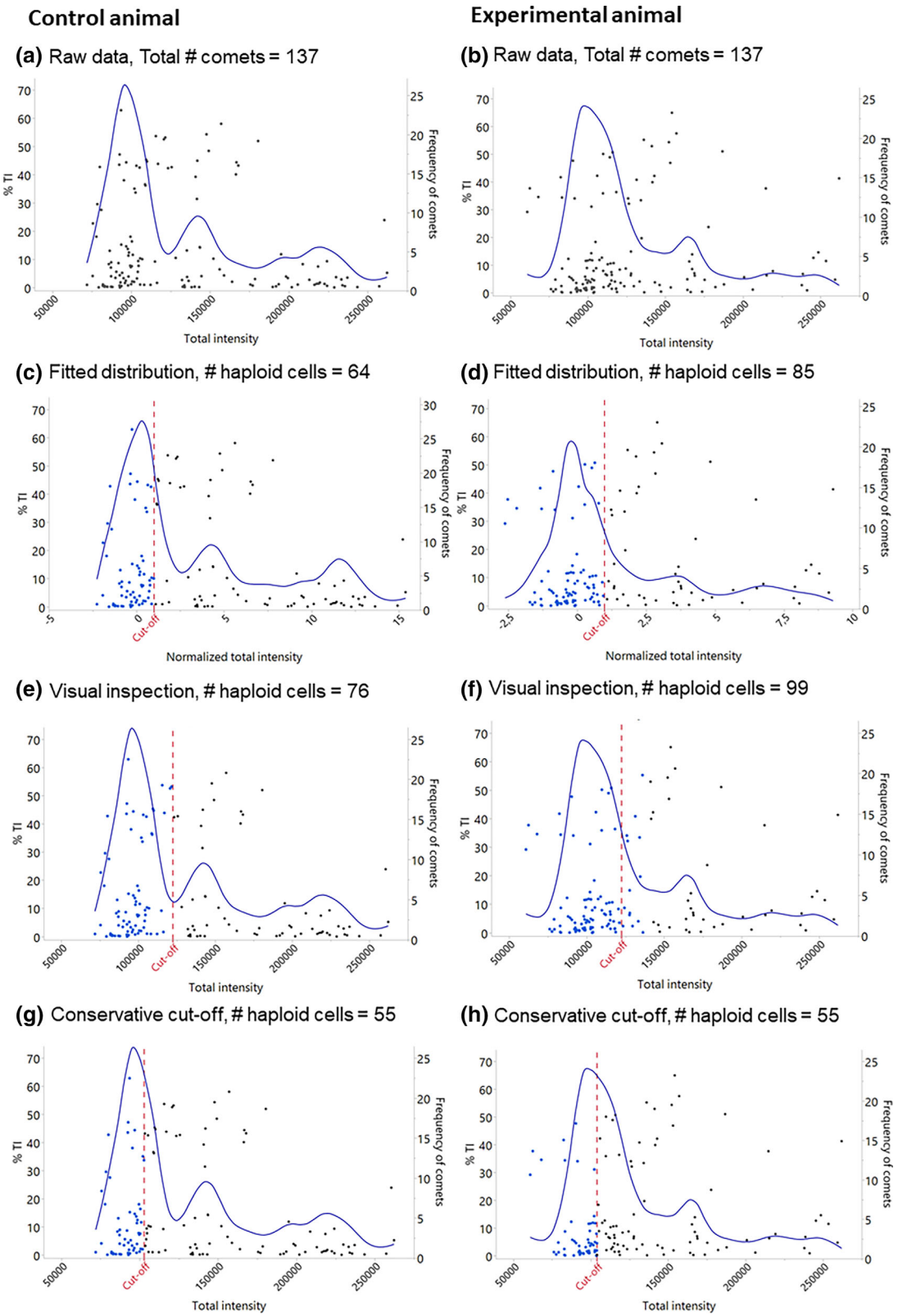

FIGURE 4 Identification of haploid germ cell populations based on different approaches. The total intensity (amount of DNA/comet) was plotted against % tail intensity (DNA damage level) for each comet. Data for one representative animal from the control group (left) and the experimental group (right) are presented. The vertical red dotted line represents the threshold (cut-off) between haploid germ cells with low total intensity and the 2C/4C cells, and the blue line represents the frequency of comets at each total intensity level. (a) and (b) represent raw data. (c) and (d) represent haploid germ cells identified based on the dispersion of the haploid germ cell population in the fitted normal distribution function after Z-score standardization. (e) and (f) represent identification based on visual inspection. (g) and (h) represent selection of the haploid germ cells based on a conservative cut-off where 40% of comets with the lowest total intensity were classified as haploid germ cells.

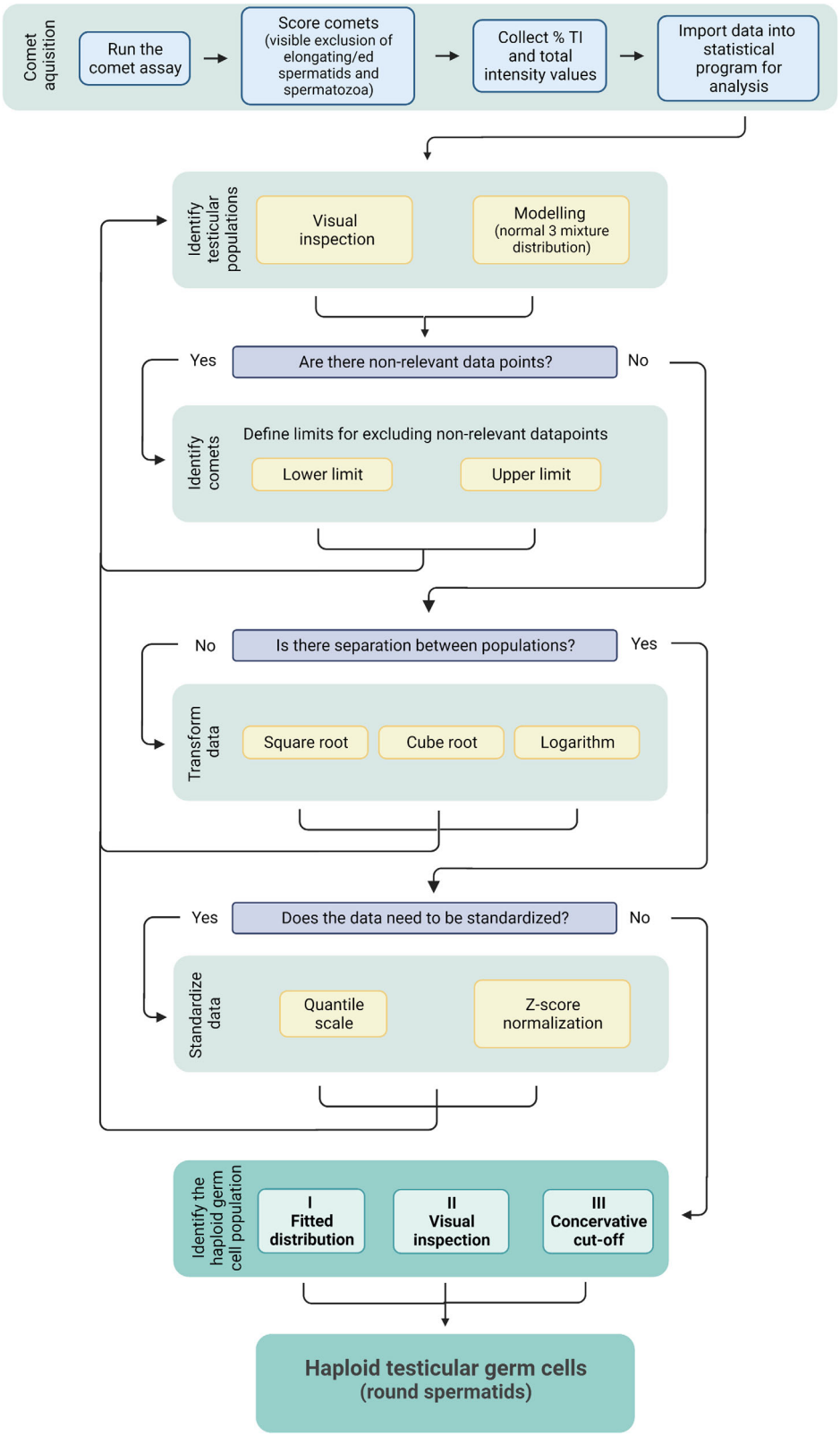

FIGURE 5 Flow diagrams for the identification of haploid germ cells. After acquiring and importing the raw data, germ cell populations are identified either through visual inspection or by modeling the data using a normal mixture distribution function. If there are nonrelevant data points that should not be included for further analysis, the lower and/or upper exclusion limits should be determined. In case of poor separation between raw data cell populations, transformation of data may increase resolution between populations and reduce skewness. Standardization of total intensity values between animals or groups can be achieved using two proposed methods. Lastly, the haploid germ cell population can be identified based on three approaches. Which path to follow depends on the raw data being analyzed (Created with BioRender).

identified evaluating a frequency line, showing the frequency of comets at total intensity intervals of 10,000. The observed comet population with the lowest total intensity was identified as the haploid germ cell population (Figure 4e, f).

III. The third approach was based on setting a conservative threshold, where all comets with a total intensity lower than the 40% quantile were identified as haploid germ cells (Figure 4g, h).

TABLE 1 The number of haploid germ cells and the mean tail intensity (% TI) for each animal (Big Blue™ C57BL/six mice) using three different approaches to identify the haploid germ cell population in the testicular cell suspension. Total # comets are the number of comets following exclusion of nonrelevant comet data points. % TI values are shown as the mean for each animal with standard deviation, where the mean was calculated from the median of each of the 3 gels per animal. Significance was tested using Wilcoxon rank of sums, ** $p < .005$, * $p < .05$.

		Fitted distribution			Conservative cut-off		Visual inspection	
Mouse number		Total # comets	# haploid cells	Tail intensity (%)	# haploid cells	Tail intensity (%)	# haploid cells	Tail intensity (%)
Control animals	1	138	68	2.4 ± 1.2	55	2.3 ± 2.0	82	3.0 ± 1.0
	2	137	50	2.7 ± 1.5	55	2.7 ± 1.8	60	2.9 ± 1.5
	3	137	64	6.8 ± 1.5	55	6.5 ± 2.2	76	7.4 ± 1.9
	4	137	52	3.3 ± 1.0	55	3.3 ± 0.6	71	3.1 ± 0.7
	5	137	64	2.4 ± 0.9	55	2.1 ± 1.3	85	3.5 ± 1.8
	6	139	65	0.3 ± 0.1	56	0.3 ± 0.1	82	0.3 ± 0.1
	Mean	137.50	68.0	3.0 ± 2.1	55.2	2.9 ± 2.1	77.7	3.4 ± 2.3
Experimental animals	1	148	68	3.9 ± 1.2	59	4.7 ± 2.1	92	6.7 ± 1.9
	2	152	97	8.2 ± 1.7	61	6.4 ± 1.1	116	9.6 ± 0.7
	3	138	71	8.8 ± 1.3	55	9.8 ± 0.6	82	10.3 ± 0.5
	4	139	77	8.5 ± 1.1	56	8.8 ± 1.4	85	8.9 ± 1.0
	5	137	85	4.7 ± 0.7	55	3.3 ± 1.2	99	5.1 ± 0.6
	6	137	67	7.8 ± 3.7	55	7.3 ± 3.5	83	9.5 ± 3.2
	7	137	76	7.3 ± 1.6	55	5.3 ± 0.0	95	8.7 ± 3.5
	8	137	76	5.4 ± 2.6	55	4.9 ± 2.6	92	9.5 ± 3.3
	Mean	140.63	77.1	6.9 ± 1.8	56.4	6.3 ± 2.2	93.0	8.6 ± 1.8
Difference between control and experimental			3.9 ± 1.1 **		3.4 ± 1.2 *		5.2 ± 1.1 **	

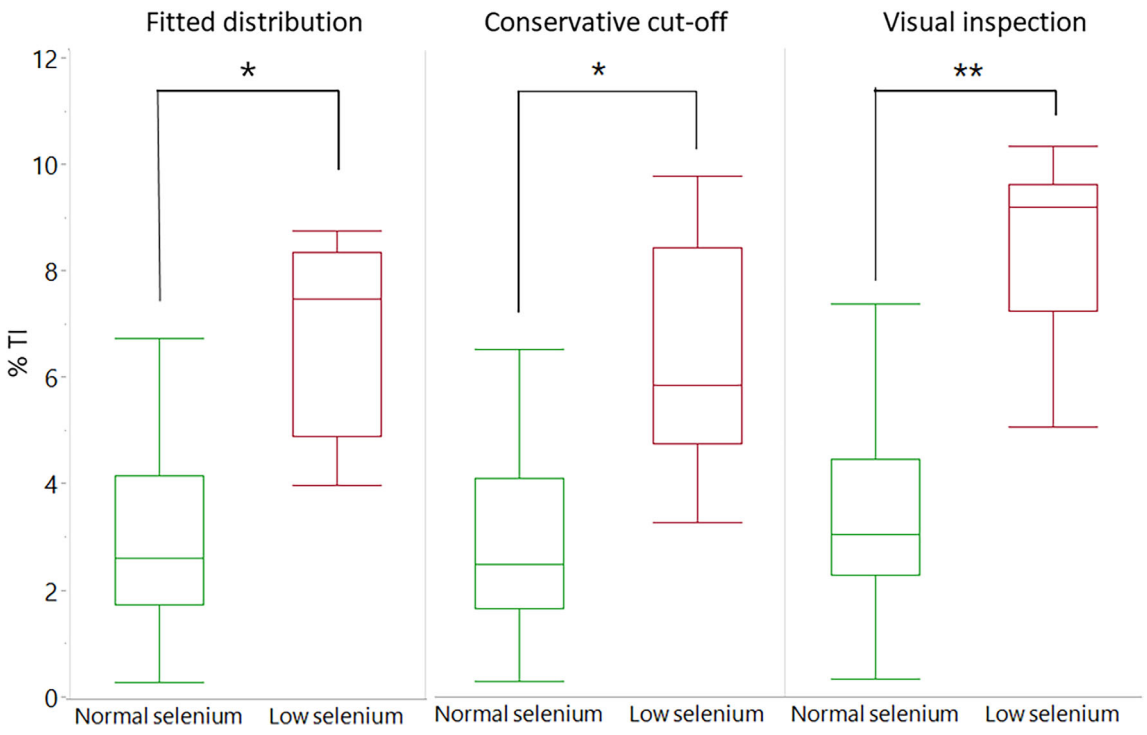

FIGURE 6 DNA damage levels (% TI) of haploid germ cells were identified using three different approaches in mice given a low or normal selenium diet. *Boxplot*: median with quartile. Whiskers indicate the lowest/highest value still within 1.5 IQR (interquartile range) of the 25th and 75th quartile. * $p < .05$, ** $p < .005$, Wilcoxon rank of sums.

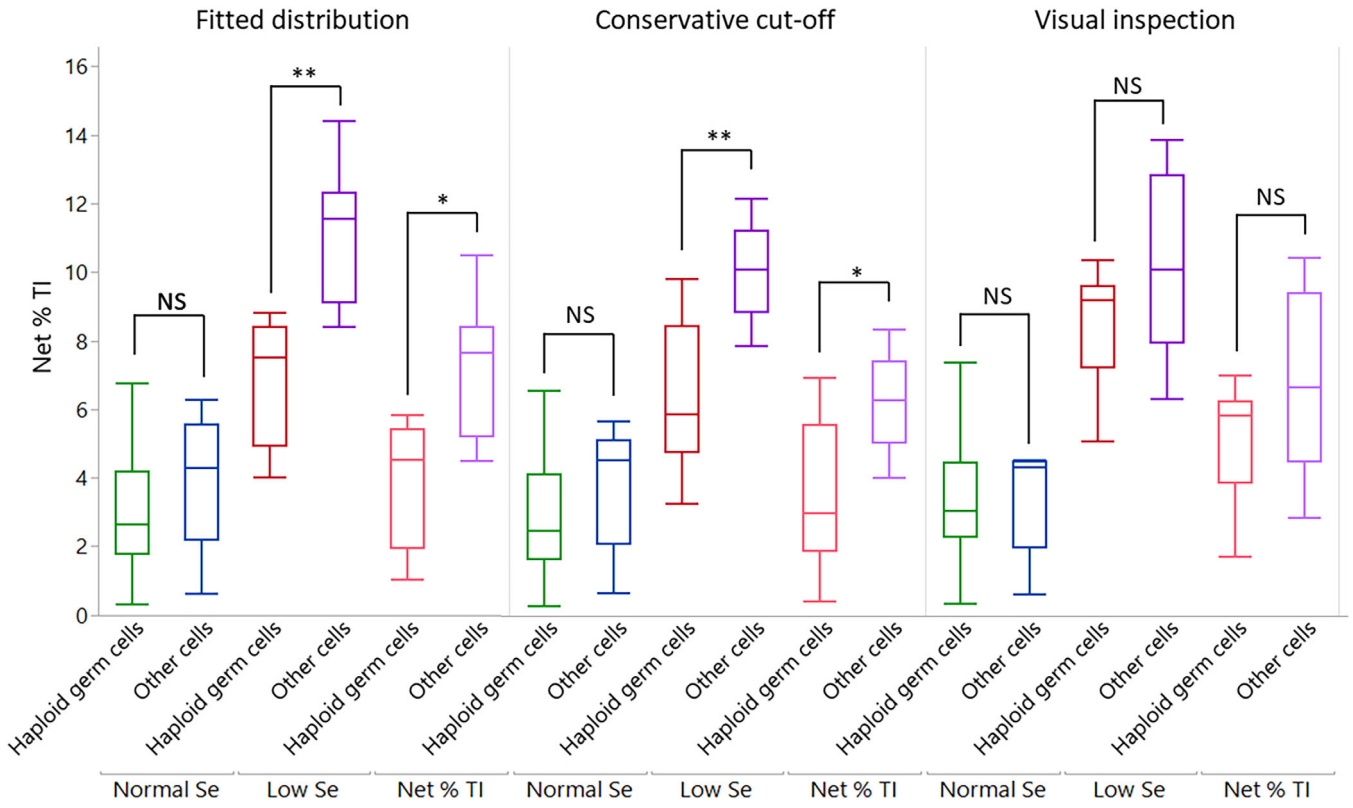

FIGURE 7 DNA damage levels (% TI) of haploid germ cells and the other cells identified using three different approaches in animals given a low or normal selenium diet. *Boxplot*: median with quartile. Whiskers indicate the lowest/highest value still within 1.5 IQR (interquartile range) of the 25th and 75th quartile. For the net low selenium boxes, the mean % TI of the control animals (green and blue box plots) were subtracted from the mean of medians for each experimental animal. The adjusted mean % TI of the experimental (low selenium) animals is shown in the boxes to the right. ** $p < .005$, * $p < .05$, NS = *not significant*, Wilcoxon rank of sums.

3.6 | Summary of the framework for analysis of haploid germ cells

The different steps of the developed protocol are summarized in Figure 5. As indicated, all steps may not be relevant for each experiment, and for some cases, several approaches can be considered appropriate to achieve the purpose of each step. It is imperative to examine each unique dataset separately when performing the analysis to decide which steps to follow, to ensure the best possible separation of the haploid germ cells from the other cell types of the testicle.

3.7 | DNA damage levels in haploid germ cells

Following the selection of haploid germ cells, DNA damage levels were compared between the datasets comprising control animals (normal selenium diet) and experimental animals (low selenium diet), to investigate whether selenium deprivation caused genotoxicity in haploid testicular germ cells.

Table 1 and Figure 6 summarize the mean % TI and standard deviation values of the haploid germ cells, based on the three different approaches, as shown in Figure 4. Regardless of the approach used to identify the haploid germ cells, significant differences were observed in the % TI between the control and experimental animals (Table 1 and Figure 6). The conservative cut-off resulted in the lowest number of comets being identified as haploid germ cells, along with the lowest average % TI, both in the control group and experimental group. Visual inspection resulted in the highest number of comets identified as haploid germ cells and the largest difference in % TI levels between the experimental group and the control group, and the highest % TI in general.

Next, we investigated whether there were differences in the % TI of the comets representing haploid germ cells versus the somatic cells and germ cells with higher ploidy, the so-called "other cells". For all three approaches, there were no significant differences in mean % TI between the haploid cells and the other cells in the control animals (normal selenium diet). However, for the experimental animals (low selenium diet) the other cells had a significantly higher mean % TI, compared to the haploid germ cells, using both the fitted distribution and conservative cut-off approaches (Figure 7). To investigate whether these differences were due to increased DNA damage in the experimental animals, the % TI was adjusted to account for the background levels of DNA damage observed in the control animals. The low selenium experimental animals showed that the net % TI was also significantly higher in the other cells, compared to the haploid germ cells. The differences in % TI were significant for two of the

approaches and were smallest for visual inspection and largest for the fitted distribution approaches.

4 | DISCUSSION

In the present study, our aim was to develop an approach to distinguish specific germ cells from other cells of the testicle, using data obtained from the in vivo comet assay, to provide germ cell-specific DNA damage level assessments. Using the proposed framework, we demonstrate that Se-deficiency through two generations in mice leads to germ cell genotoxicity with elevated DNA damage levels in round spermatids.

The proposed framework can be implemented in the current OECD test guideline "In Vivo Mammalian Alkaline Comet Assay" (TG 489), thereby increasing the availability of germ cell genotoxicity data for regulatory risk assessments. We describe a concept to identify haploid round spermatids (1N/1C) based on their low DNA content, whereas primary spermatocytes (2N/4C) can be identified through their particularly large size combined with high DNA content (Figure 1). Historical comet data from mice given a normal or low selenium diet through two generations (Graupner et al., 2015) were used to demonstrate the identification of round spermatids. We previously showed that selenium deprivation induces elevation of DNA damage in a mixed suspension of testicular cells (Graupner et al., 2015) and thus serves as a good model to study germ cell genotoxicity to address the question at hand.

We believe that round spermatids and primary spermatocytes cover relevant germ cell genotoxicity effects since they represent germ cells at early and late stages of spermatogenesis (Figure 1a). Moreover, they reside in different locations within the seminiferous tubules (Figure 1b). In the seminiferous epithelium a gradient of increasing protection from exogenous influences, known as the blood testis barrier, protects the germ cells in the apical compartment from some potential genotoxic substances (Mruk & Cheng, 2015). The primary spermatocytes are located close to the basal membrane, whereas the round spermatids are located closer to the lumen and may thus be more protected from genotoxic insults. The location and susceptibility of each cell type to genotoxic attack are important in context of their capacity for error-free elimination of induced DNA lesions. Round spermatids may be particularly sensitive to induction of DNA damage since they rely on error-prone nonhomologous end-joining for repair of double-strand breaks, as opposed to homologous recombination, as mechanism of DNA repair (Ahmed et al., 2010; Gouraud et al., 2013; Ahmed et al., 2015). In addition, we and others have demonstrated that male germ cells exhibit limited capacities for repair via several DNA repair pathways, including both nucleotides- and base-excision repair and that the repair capacity depends on both the spermatogenic cell stage and the type of DNA lesion induced (Jansen et al., 2001; Olsen et al., 2001; 2003; 2005). These characteristics of the different testicular germ cell have implications for their propensity to accumulate DNA mutations. Moreover, if DNA damage is observed in these cells, it indicates that the genotoxicant can reach even the most protected germ cell types (Cheng & Mruk, 2012; Su et al., 2011).

When using the raw dataset to model the different cell populations, using the normal mixture distribution function, we did not obtain a model that fit the three observable populations for all animals in the experiment. This was likely due to inclusion of nonrelevant data points, such as elongated spermatids or mature spermatozoa that exhibit more densely packed DNA, which may be refractory to DNA staining and results in data points with low total intensities (Figure 2a). Alternatively, scoring of cell doublets/overlapping cells results in data points with high total intensity. Including these data points in the dataset hampers the downstream analysis and causes increased skewness. We excluded these nonrelevant data points to facilitate identification of the most likely cell populations for each animal (Figure 2b). Additionally, establishing a well-fitted model is important when using Z-score normalization as the μ and σ values derived from the model are used to define the haploid round spermatid population. For these reasons, all further analysis steps were conducted using trimmed data.

A reduction of the observed left-skewed-right-tailed total intensity data were attempted by employing different data transformations. For the dataset in this study, the untransformed total intensity values generally resulted in the highest resolution between populations, and hence the highest likelihood of correctly identifying comets belonging to the haploid germ cell population. However, this was not the best approach for every animal. For some animals, the square root, cube root, or logarithmic transformations resulted in better-fitting models. Nevertheless, since data from all animals within an experiment must be subjected to similar processing, we chose to use the untransformed data, as this produced the better model for the highest fraction of animals in the experiment (Figure 3).

Variability in comet assay results between different experimental runs and laboratories has long been a challenge, also for the recorded total intensity values of comets. Several strategies for standardizing the protocol and reporting comet assay results have been proposed with regards to % tail intensity values (Zainol et al., 2009, Forchhammer et al., 2010, 2012, Collins et al., 2014, Brunborg et al., 2015, Moller et al., 2020). Very few efforts to standardize total intensity values have been conducted. Because a testicular cell population is heterogeneous and consists of different cell types, at different stages of maturation, with different DNA content and ploidy, finding a suitable method of standardization is inherently challenging. Not only does each cell population have its own distribution, but the proportion of cells in each population may also vary between animals. The comet data used in this study (Graupner et al., 2015) was generated in one lab by the same researcher over a short period of time; even so, variabilities in total intensity values were observed between comets scored on different days. Having a reliable method for standardization is thus important to facilitate comparison of data across and between experiments and laboratories.

An important assumption for using a scaling factor for standardization is that the total intensity distribution across animals is

comparable, which might not always be the case due to cytotoxicity or protocol variations during cell isolation. To circumvent the lack of comparability in total intensity distributions we employed Z-score standardization and concentrated on standardizing the haploid round spermatid population. The Z-score standardized total intensity values were based on μ and σ values from the distribution of the haploid germ cell population, obtained from the model. After Z-score standardization, the distributions of the haploid germ cell populations showed little to no difference across animals. If other testicular cell populations should be relevant, the same principle could easily be applied by using the σ and μ values of the population of interest.

To discriminate the haploid germ cells from other cells in the testicular cell suspension three different approaches were explored (Figure 4). The distributions of the comet total intensity values for each specific experiment define which of the approaches is considered most suitable, as each approach has different criteria for setting a cut-off, or threshold, separating the cell populations. Using Approach I (fitted distribution model), which was based on Z-score standardized total intensity values, the cut-off was similar for each animal. Therefore, the selection of haploid germ cells can be scripted and performed on data from all animals in each experiment simultaneously, facilitating rapid and easy identification of the desired cell population. In Approach II (visual inspection of the total intensity distributions) the haploid germ cell population is identified visually for each biological replicate, in this case per animal. The cut-off and number of identified haploid germ cells are different for each animal and this approach is therefore more subjective and labor-intensive than Approach I. At the same time, Approach II requires less raw data processing. In Approach III (conservative cut-off) the total intensity value at which the threshold is set, separating the haploid germ cells from the remaining cells, also varies between each biological replicate and must be set separately for each animal, unless quantile scaled data is used.

The fitted distribution model (Approach I) is most likely to reflect the true haploid cell population, as it considers both the total intensity mean as well as the dispersion (SD) of the cell populations. The fitted distribution model assigns comets to the haploid population based on a cut-off at mean + 1 standard deviation (SD). The SD range should be considered on a case-to-case basis. Generally, the cut-off should be determined using transformed data (roots, logarithms, or reciprocals) to improve the likelihood of normal distributions and increase separation between populations. In the historical dataset used herein deviations from normality were minor and did not improve with data transformation (Figure 3), we, therefore, continued with untransformed data.

Basing the selection of round spermatids on visual inspection can be challenging due to subjectiveness, wide dispersions, and potential overlap between the haploid and diploid comet populations. This makes it challenging to set the most appropriate threshold to separate the populations (Figure 4e, f), potentially causing an over- or underestimation of the number of haploid germ cells, which may influence the recorded DNA damage levels (% TI). In general, the distinction between populations was clearer in the control animals that had smaller dispersions and overlap, compared to the experimental animals. This was independent of whether the data were transformed or not (Figure 3), The same was true (subjectiveness and under-estimation) for the conservative cut-off method (Figure 4g, h), where the threshold is set to include fewer comets that are assumed to be present in the nonexposed haploid population. This approach is also sensitive to testicular germ cell cytotoxicity since the relative abundance of each cell type may change due to cell death.

To investigate whether there would be differences in the levels of DNA damage observed due to approaching used, we compared the % TI of the identified haploid cell populations in the control and experimental groups in each approach (Figure 6 and Table 1). We found no significant differences in % TI between the different approaches in the control animals ($p > .79$, Tukey's test). However, for the experimental animals, there was a significant difference in % TI for the haploid germ cells identified through the visual and conservative approaches ($p = .01$, Tukey's test). This was not surprising, as these two approaches had the largest difference in the number of comets identified as haploid and in the % TI. This indicates that the chosen approach can influence the results because including more or fewer comets in the haploid germ cell population might significantly alter the amount of DNA damage observed. Careful consideration should therefore be used to determine the most appropriate approach to use.

Our data indicate that being more restrictive in the selection of haploid cells, as opposed to including as many haploid comets as possible, might be beneficial, especially for the experimental group. As Figure 7 shows, for the fitted distribution and conservative cut-off approaches, the % TI was significantly lower for the haploid germ cells (round spermatids) compared to the other cells in the testicular cell suspension for the experimental animals on a low selenium diet. These approaches identified the lowest number of comets as haploid. The visual inspection approach might, therefore, be less accurate in distinguishing cell populations, leading to inclusion of non-haploid comets that may have different DNA damage levels (pending on stressor) than the haploid round spermatids. Figure 4 shows that the diploid population (2N/2C) in general has higher % TI than both the haploid and the tetraploid (2N/2-4C) populations. Using visual inspection and potentially identifying more highly damaged diploid cells as haploid, may explain the higher % TI of the haploid population. Specific differences in damage levels between cell populations may be lost. It should be noted that the haploid germ cells consistently showed a lower steady-state DNA damage level compared to the other cells in this experiment (Figure 7), which could be ascribed to mere comet dynamics of the haploid round spermatids or to factual lower DNA damage levels.

In addition to comparing the control and experimental animals, we also adjusted the % TI of the experimental animals by subtracting the steady state % TI of the control animal group. The adjusted % TI revealed that the haploid germ cells remained lower compared to the other cells, validating that the differences observed in % TI are not due to differences in % TI background levels. In addition, we observed little difference between the haploid cells of the control and experimental animals, after subtracting the steady-state DNA damage. This shows that similarly to the non-adjusted % TI, the other cells in general seem to exhibit higher levels of DNA damage due to the low selenium diet, compared to the round spermatids.

One important assumption for this analysis is that different cell types, at any DNA damage level (within the limits of the dynamic range of the comet assay), exhibit acceptably similar total intensity levels. If the total intensity of highly damaged comets is significantly changed, these comets may be ascribed to the erroneous population. Moreover, if the substance under investigation selectively causes high levels of cytotoxicity, cell cycle delays, or induces apoptosis in haploid germ cells, this cell population may consist fewer comets than anticipated. Based on our experience, DNA damage does not significantly affect the total intensity of comets, except at very high DNA damage levels. Other reports have shown that high levels of DNA damage can result in comets with similar (Olive & Banath, 2006) or lower total intensities (Olive et al., 1994; Brunborg et al., 2015; Moller et al., 2020). One explanation could be that small-sized DNA from heavily damaged comets migrate beyond what is considered and scored as the whole comet during electrophoresis. Moreover, factors such as choice of DNA stain, microscope/lamp physics, and robustness of the staining protocol may influence the total intensity values and would benefit from standardization.

In this study, selenium deprivation resulted in mean % TI levels between 0% and 15% (Figures 6, 7). We, therefore, concluded that these DNA damage levels are within acceptable limits for obtaining germ cell-specific genotoxicity data. We recommend that investigation of the effect of high DNA damage on total intensity levels should be pursued as part of validating the comet assay in testicular germ cells.

To obtain robust results, it is important to score comets to ensure sufficient power to detect minor effects. Currently, OECD TG 489 suggests scoring 150 comets per animal, from at least five animals per experimental condition (Smith et al., 2008; OECD, 2016). The adequate number of scored comets depends on the potency of the substance tested and the variability of the data. For this proof-of-concept, we focused on round spermatids and primary spermatocytes (identified post-scoring). To comply with TG 489, it is recommended to score > 150 comets to obtain a high enough number of scored round spermatids. Based on power calculations performed on this dataset, scoring 60 round spermatids was sufficient to detect the difference between the treatment groups with a power of 0.95. These results were obtained from 300 simulations with a combined sampling (permutation) of between 20 and 100 random comets with a random set of five mice in each treatment group. When the complete dataset was included in the power estimation (six control mice and eight experimental mice), fewer than 30 round spermatid comets per mouse were required. Depending on effect size, other comet numbers than the conventional 150 per biological sample may be appropriate and validation studies including weak mutagens or low exposure levels should be conducted to determine appropriate numbers of comets to score.

The comet assay gives information regarding DNA damage levels in the cells investigated and acts as an indicator test for mutagenicity and cancer (Kirkland & Speit, 2008; Kirkland, Uno, et al., 2019b). It should be noted that pre-mutagenic DNA lesions are measured, not mutations *per se*; however such DNA lesions predict mutagenicity. The DNA strand breaks measured in the comet assay originate from, and reflect, different DNA lesions with variable potential for inducing mutations (Akbari & Krokan, 2008). Single strand breaks may give rise to double-strand breaks that may bring about chromosomal aberrations, and base lesions and abasic (base-less) sites may lead to point mutations or small insertions/deletions. The comet assay has lower sensitivity for detecting some bulky adducts including those induced by polyaromatic hydrocarbons and heterocyclic amines (Kawaguchi et al., 2010; Graupner et al., 2014; Nixon et al., 2014; Ngo et al., 2020). A range of DNA lesions not readily detectable in the classical alkaline comet assay, including oxidized base lesions, can be detected using an enzyme-modified comet assay (Olsen et al., 2003 Hansen et al., 2010; Graupner et al., 2014; Graupner et al., 2015; Asare et al., 2016; Hansen et al., 2018; Muruzabal et al., 2021). In this way, the sensitivity may be enhanced, and we can obtain information regarding lesion specificity, which is also relevant for germ cell genotoxicity assessments.

The *in vivo* comet assay is one of the most commonly used tests to follow up positive results from in vitro mutagenicity tests (ECHA, 2017). Revising the *in vivo* comet assay test guideline (TG 489) to include a protocol for assessing DNA damage in germ cells will increase the availability of germ cell genotoxicity data, which in turn will aid hazard identification and classification of germ cell mutagens. Positive results in gonadal germ cells, obtained with the concept proposed herein, provide direct evidence that the substance interacts with the genetic material of germ cells. Such information is required for classification in germ cell mutagenicity category 1B according to GHS (GHS, 2009), rather than category 2. Classifying a substance in Muta category 1B rather than category 2 is challenging since the evidence needed for classification is rarely available, e.g., data from suitable, approved test guidelines, such as OECD TG 478, TG 483, TG 485, and TG 488. A revised TG 489 could provide such evidence, resulting in hazard identification and classification of potential germ cell mutagenic substances based on germ cell data. In cases of negative results from the in vivo comet assay in male gonadal cells, no definite conclusion can be made on mutagenicity in germ cells or distribution of substances to the testis. It should be noted that a revision of the GHS classification criteria for germ cell mutagenicity has been initiated in order to facilitate the use of other types of data as indirect evidence of interaction with germ cells and update the wording in Chapter 3.5 (UNECE, 2020; UNECE, 2021).

Another asset of a revised TG 489 is the possibility of obtaining both somatic and germ cell data in one *in vivo* comet assay, hence reducing the number of test animals used. The comet assay is ideal for integrating different tissues into the same assay, meaning that germ cell and somatic data can be analyzed and compared within the same animal (Speit et al., 2009; Azqueta & Collins, 2013; Groff et al., 2021). Genotoxicity tests on somatic cells are still more frequently requested than tests on germ cells (Steiblen et al., 2020). Slides/films with gonadal comets can simply be stored and analyzed if additional *in vivo* germ cell mutagenicity tests are required to follow up positive results in somatic cells. This negates the need to perform additional *in vivo* germ cell mutagenicity tests, thereby reducing the use of animals and

contribution to the 3R framework. For pragmatic reasons, comet assay results from somatic cells have been used as indicators of the genotoxicity in germ cells (Sasaki et al., 2000; Eastmond et al., 2009 Speit et al., 2009; Yauk et al., 2015). Data obtained from testicular germ cells using the concept proposed herein, combined with data from metabolically active tissues, such as the liver, may shed light upon the predictability of, and correlation between, data from somatic cells and testicular germ cells.

Of the other germ cell genotoxicity assays with approved test guidelines (TG 478, TG 483, TG 485, and TG 488) the TGR assay (TG 488) is the only other *in vivo* test that can adress both somatic and germ cells. However, TG 488 is most often limited to detecting gene mutations in the form of base pair substitutions, frameshift mutations, and small deletions/insertions (OECD, 2020); depending on the transgenic rodent model used. The comet assay detects a range of pre-mutagenic lesions, such as DNA single and double strand breaks, alkali labile lesions, and DNA strand breaks resulting from DNA excision repair (OECD, 2016). The comet assay can thus be used to follow up both gene mutation as well as clastogenic effects (EFSA, 2011; ECHA, 2017).

Mutations measured with TG 488 depend on DNA replication to transition DNA damage into permanent mutations, thus requiring sufficient expression periods for mutations to be fixed. The only proliferating germ cells are the spermatogonia, which reside outside the blood-testis barrier and may thus be most susceptible to hazardous chemical exposure (Cheng & Mruk, 2012). The comet assay, on the other hand, detects pre-mutagenic DNA lesions in post-replicative testicular germ cells (primary spermatocytes and round spermatids) located within the blood testis barrier. The comet assay is independent of DNA replication, thus requiring a significantly shorter time from exposure to sampling (TG 489 recommends sampling 2–6 h after the last exposure). We believe that a revised TG 489 *in vivo* comet assay, with the inclusion of a guideline to investigate testicular male germ cells, will complement results obtained with the TG 488 TGR assay. The TGR assay detects mutations in germ cells whereas the comet assay contributes with information regarding potentially heritable DNA damage arising in testicular germ cells, that in some cases are repair deficient. Due to limited repair capacities in post-meiotic testicular germ cells, DNA damage will likely be present in the fertilizing spermatozoa, leading to potential mutation induction after fertilization that could manifest as health implications in offspring. Together, these two assays will contribute with valuable complementary genotoxicity/mutagenicity data for different germ cell types.

In conclusion, this proof-of-concept study provides a framework for obtaining testicular germ cell-specific genotoxicity data. It demonstrates that haploid round spermatids can be identified in a testicular cell suspension based on their low total intensity and round shape. We provide three different approaches to identifying round spermatids based on total intensity distributions of the comets: visual discrimination, setting conservative threshold values for comet fluorescence, or fitting a normal mixture distribution function to the total intensity values of the comets. We propose to use the latter approach, when possible. Additionally, we describe a method to identify primary spermatocytes, based on their high DNA content combined with their particularly large physical size. Revising OECD TG 489 to include analysis of male testicular germ cells provides an important contribution to the currently very limited and time-, and animal consuming set of tests available for assessing genotoxicity in germ cells, facilitating more correct hazard identification and classification of agents with respect to germ cell genotoxicity. The comet assay is suitable to follow up on both gene mutation and clastogenicity concerns. In addition, a revision of the TG 489 *in vivo* comet assay is easily implementable, versatile, sensitive, relatively fast, easy to perform, and provides both somatic and germ cell data from the same animals.

AUTHOR CONTRIBUTIONS

Ann-Karin Olsen and Gunnar Brunborg conceptualized and developed the main idea. Ann-Karin Olsen and Kristine Bjerve Gützkow supervised the project. Dag Markus Eide aided in developing and testing the data analysis model and interpreting the results. Jarle Ballangby collected data for a historical in-house database from which Anne Graupner provided data that was selected as being relevant for development of the analysis method. Yvette Dirven wrote the manuscript and created the figures. Erika Witasp Henriksson, Rune Hjorth, and Ann-Karin Olsen consulted on the writing of the manuscript. All authors, including Anne Mette Zenner Boisen and Stellan Swedmark, provided feedback and input on the manuscript. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

Thank you to all the members of the project steering group of the project for excellent input and feedback during the writing process. A special thanks to Nina Landvik, Sjur Andersen, and Linda Reierson from the Norwegian Environment Agency for their excellent input and feedback during the writing process and consultation regarding the final manuscript. Also, thanks to Frank Le Curieux and Jonas Nygren from ECHA for providing feedback and guidance on the paper. Funding was received from the Nordic Working Group for Chemicals, Environment, and Health, under the Nordic Council of Ministers (2021-022,2021-101). This publication was funded by the Nordic Council of Ministers; however, the content does not necessarily reflect the Nordic Council of Ministers' views, opinions, attitudes, or recommendations. Financial support was also received from Swedish Chemicals Agency and the Research Council of Norway through its Centers of Excellence funding scheme, project number 223268/F50 CERAD.

CONFLICT OF INTEREST

The authors of this work declare no competing interest.

REFERENCES

Ahmed, E.A., De Boer, P., Philippens, M.E., Kal, H.B. & De Rooij, D.G. (2010) Parp1-XRCC1 and the repair of DNA double strand breaks in mouse round spermatids. *Mutation Research*, 683, 84–90.

Ahmed, E.A., Scherthan, H. & De Rooij, D.G. (2015) DNA double Strand break response and limited repair capacity in mouse elongated spermatids. *International Journal of Molecular Sciences*, 16, 29923–29935.

Akbari, M. & Krokan, H.E. (2008) Cytotoxicity and mutagenicity of endogenous DNA base lesions as potential cause of human aging. *Mechanisms of Ageing and Development*, 129, 353–365.

Asare, N., Duale, N., Slagsvold, H.H., Lindeman, B., Olsen, A.K., Gromadzka-Ostrowska, J. et al. (2016) Genotoxicity and gene expression modulation of silver and titanium dioxide nanoparticles in mice. *Nanotoxicology*, 10, 312–321.

Azqueta, A. & Collins, A.R. (2013) The essential comet assay: a comprehensive guide to measuring DNA damage and repair. *Archives of Toxicology*, 87, 949–968.

Bjorge, C., Wiger, R., Holme, J.A., Brunborg, G., Andersen, R., Dybing, E. et al. (1995) In vitro toxicity of 1,2-dibromo-3-chloropropane (DBCP) in different testicular cell types from rats. *Reproductive Toxicology*, 9, 461–473.

Bjorge, C., Wiger, R., Holme, J.A., Brunborg, G., Scholz, T., Dybing, E. et al. (1996) DNA strand breaks in testicular cells from humans and rats following in vitro exposure to 1,2-dibromo-3-chloropropane (DBCP). *Reproductive Toxicology*, 10, 51–59.

Brendler-Schwaab, S., Hartmann, A., Pfuhler, S. & Speit, G. (2005) The in vivo comet assay: use and status in genotoxicity testing. *Mutagenesis*, 20, 245–254.

Brevik, A., Lindeman, B., Rusnakova, V., Olsen, A.K., Brunborg, G. & Duale, N. (2012) Paternal benzo[a]pyrene exposure affects gene expression in the early developing mouse embryo. *Toxicological Sciences*, 129, 157–165.

Bright, J., Aylott, M., Bate, S., Geys, H., Jarvis, P., Saul, J. et al. (2011) Recommendations on the statistical analysis of the comet assay. *Pharmaceutical Statistics*, 10, 485–493.

Brunborg, G., Collins, A., Graupner, A., Gutzkow, K.B. & Olsen, A.K. (2015) Reference cells and ploidy in the comet assay. *Frontiers in Genetics*, 6, 61.

Cheng, C.Y. & Mruk, D.D. (2012) The blood-testis barrier and its implications for male contraception. *Pharmacological Reviews*, 64, 16–64.

Collins, A.R. (2004) The comet assay for DNA damage and repair: principles, applications, and limitations. *Molecular Biotechnology*, 26, 249–261.

Collins, A.R., Yamani, E.N., Lorenzo, Y., Shaposhnikov, S., Brunborg, G. & Azqueta, A. (2014) Controlling variation in the comet assay. *Frontiers in Genetics*, 5, 359.

Cordelli, E., Bignami, M. & Pacchierotti, F. (2021) Comet assay: a versatile but complex tool in genotoxicity testing. *Toxicology Research*, 10, 68–78.

Dearfield, K.L., Cimino, M.C., Mccarroll, N.E., Mauer, I., Valcovic, L.R. & Agency, U.S.E.P. (2002) Genotoxicity risk assessment: a proposed classification strategy. *Mutation Research*, 521, 121–135.

Dhawan, A., Bajpayee, M. & Parmar, D. (2009) Comet assay: a reliable tool for the assessment of DNA damage in different models. *Cell Biology & Toxicology*, 25, 5–32.

Eastmond, D.A., Hartwig, A., Anderson, D., Anwar, W.A., Cimino, M.C., Dobrev, I. et al. (2009) Mutagenicity testing for chemical risk assessment: update of the WHO/IPCS harmonized scheme. *Mutagenesis*, 24, 341–349.

ECHA. (2012) ECHA Report on Survey of Worldwide Cros: Costs and Practicalities of Two New OECD Guidelines For Testing Chemical Substances OECD 443, Extended One-Generation Reproductive Toxicity Study, and OECD 488, Transgenic Rodent Somatic and Germ Cell Mutation Assay.

ECHA. (2017) Guidance on information requirements and chemical safety assessment, chapter R.7a: endpoint specific guidance. Version 6.0. AGENCY, E. C. (ed.).

EFSA, S. C. (2011) Scientific opinion on genotoxicity testing strategies applicable to food and feed safety assessment. *EFSA Journal*, 9, 2379.

Forchhammer, L., Ersson, C., Loft, S., Möller, L., Godschalk, R.W.L., Schooten, F.J.V.A.N. et al. (2012) Inter-laboratory variation in DNA damage using a standard comet assay protocol. *Mutagenesis*, 27, 665–672.

Forchhammer, L., Johansson, C., Loft, S., Moller, L., Godschalk, R.W., Langie, S.A. et al. (2010) Variation in the measurement of DNA damage by comet assay measured by the ECVAG inter-laboratory validation trial. *Mutagenesis*, 25, 113–123.

Frotschl, R. (2015) Experiences with the in vivo and in vitro comet assay in regulatory testing. *Mutagenesis*, 30, 51–57.

GHS 2009. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). vi, 559.

Gouraud, A., Brazeau, M.A., Gregoire, M.C., Simard, O., Massonneau, J., Arguin, M. et al. (2013) "Breaking news" from spermatids. *Basic and Clinical Andrology*, 23, 11.

Graupner, A., Eide, D.M., Brede, D.A., Ellender, M., Lindbo Hansen, E., Oughton, D.H. et al. (2017) Genotoxic effects of high dose rate X-ray and low dose rate gamma radiation in Apc(min/+) mice. *Environmental and Molecular Mutagenesis*, 58, 560–569.

Graupner, A., Instanes, C., Andersen, J.M., Brandt-Kjelsen, A., Dertinger, S.D., Salbu, B. et al. (2015) Genotoxic effects of two-generational selenium deficiency in mouse somatic and testicular cells. *Mutagenesis*, 30, 217–225.

Graupner, A., Instanes, C., Dertinger, S.D., Andersen, J.M., Lindeman, B., Rongved, T.D. et al. (2014) Single cell gel electrophoresis (SCGE) and pig-a mutation assay in vivo-tools for genotoxicity testing from a regulatory perspective: a study of benzo[a]pyrene in Ogg1(−/−) mice. *Mutation Research, Genetic Toxicology and Environmental Mutagenesis*, 772, 34–41.

Griffiths, A., Miller, J., Suzuki, D.T. et al. (2000) *An introduction to genetic analysis*. New York, NY: W. H. Freeman.

Groff, K., Evans, S.J., Doak, S.H., Pfuhler, S., Corvi, R., Saunders, S. et al. (2021) In vitro and integrated in vivo strategies to reduce animal use in genotoxicity testing. *Mutagenesis*, 36, 389–400.

Gutzkow, K.B., Duale, N., Danielsen, T., Stedingk, V.O.N.H., Shahzadi, S., Instanes, C. et al. (2016) Enhanced susceptibility of obese mice to glycidamide-induced sperm chromatin damage without increased oxidative stress. *Andrology*, 4, 1102–1114.

Gutzkow, K.B., Langleite, T.M., Meier, S., Graupner, A., Collins, A.R. & Brunborg, G. (2013) High-throughput comet assay using 96 minigels. *Mutagenesis*, 28, 333–340.

Hansen, M.K., Sharma, A.K., Dybdahl, M., Boberg, J. & Kulahci, M. (2014) In vivo comet assay—statistical analysis and power calculations of mice testicular cells. *Mutation Research, Genetic Toxicology and Environmental Mutagenesis*, 774, 29–40.

Hansen, S.H., Olsen, A.K., Soderlund, E.J. & Brunborg, G. (2010) In vitro investigations of glycidamide-induced DNA lesions in mouse male germ cells and in mouse and human lymphocytes. *Mutation Research*, 696, 55–61.

Hansen, S.H., Pawlowicz, A.J., Kronberg, L., Gutzkow, K.B., Olsen, A.K. & Brunborg, G. (2018) Using the comet assay and lysis conditions to characterize DNA lesions from the acrylamide metabolite glycidamide. *Mutagenesis*, 33, 31–39.

Holstein, A.F., Schulze, W. & Davidoff, M. (2003) Understanding spermatogenesis is a prerequisite for treatment. *Reproductive Biology and Endocrinology*, 1, 107.

Jansen, J., Olsen, A.K., Wiger, R., Naegeli, H., Boer, P, D.E., Van Der Hoeven, F. et al. (2001) Nucleotide excision repair in rat male germ cells: low level of repair in intact cells contrasts with high dual incision activity in vitro. *Nucleic Acids Research*, 29, 1791–1800.

Johnson, L., Neaves, W.B., Barnard, J.J., Keillor, G.E., Brown, S.W. & Yanagimachi, R. (1999) A comparative morphological study of human germ cells in vitro or in situ within seminiferous tubules. *Biology of Reproduction*, 61, 927–934.

Kawaguchi, S., Nakamura, T., Yamamoto, A., Honda, G. & Sasaki, Y.F. (2010) Is the comet assay a sensitive procedure for detecting genotoxicity? *Journal of Nucleic Acids*, 2010, 541050.

Kirkland, D., Levy, D.D., Lebaron, M.J., Aardema, M.J., Beevers, C., Bhalli, J. et al. (2019a) A comparison of transgenic rodent mutation and in vivo comet assay responses for 91 chemicals. *Mutation Research, Genetic Toxicology and Environmental Mutagenesis*, 839, 21–35.

Kirkland, D. & Speit, G. (2008) Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens III. Appropriate follow-up testing in vivo. *Mutation Research*, 654, 114–132.

Kirkland, D., Uno, Y., Luijten, M., Beevers, C., Benthem, V.A.N.J., Burlinson, B. et al. (2019b) In vivo genotoxicity testing strategies: report from the 7th international workshop on genotoxicity testing (IWGT). *Mutation Research, Genetic Toxicology and Environmental Mutagenesis*, 847, 403035.

Koppen, G., Azqueta, A., Pourrut, B., Brunborg, G., Collins, A.R. & Langie, S. A.S. (2017) The next three decades of the comet assay: a report of the 11th international comet assay workshop. *Mutagenesis*, 32, 397–408.

Langie, S.A., Azqueta, A. & Collins, A.R. (2015) The comet assay: past, present, and future. *Frontiers in Genetics*, 6, 266.

Marchetti, F., Douglas, G.R. & Yauk, C.L. (2020) A return to the origin of the EMGS: rejuvenating the quest for human germ cell mutagens and determining the risk to future generations. *Environmental and Molecular Mutagenesis*, 61, 42–54.

Marcho, C., Oluwayiose, O.A. & Pilsner, J.R. (2020) The preconception environment and sperm epigenetics. *Andrology*, 8, 924–942.

Moller, P., Azqueta, A., Boutet-Robinet, E., Koppen, G., Bonassi, S., Milic, M. et al. (2020) Minimum information for reporting on the comet assay (MIRCA): recommendations for describing comet assay procedures and results. *Nature Protocols*, 15, 3817–3826.

Mruk, D.D. & Cheng, C.Y. (2015) The mammalian blood-testis barrier: its biology and regulation. *Endocrine Reviews*, 36, 564–591.

Muruzabal, D., Collins, A. & Azqueta, A. (2021) The enzyme-modified comet assay: past, present and future. *Food and Chemical Toxicology*, 147, 111865.

Ngo, L.P., Owiti, N.A., Swartz, C., Winters, J., Su, Y., Ge, J. et al. (2020) Sensitive CometChip assay for screening potentially carcinogenic DNA adducts by trapping DNA repair intermediates. *Nucleic Acids Research*, 48, e13.

Nixon, B.J., Katen, A.L., Stanger, S.J., Schjenken, J.E., Nixon, B. & Roman, S. D. (2014) Mouse spermatocytes express CYP2E1 and respond to acrylamide exposure. *PLoS One*, 9, e94904.

OECD. (2016) *Test No. 489: In Vivo Mammalian Alkaline Comet Assay*.

OECD. (2020) *Test No. 488: Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays*.

Olive, P.L. & Banath, J.P. (2006) The comet assay: a method to measure DNA damage in individual cells. *Nature Protocols*, 1, 23–29.

Olive, P.L., Banath, J.P. & Fjell, C.D. (1994) DNA strand breakage and DNA structure influence staining with propidium iodide using the alkaline comet assay. *Cytometry*, 16, 305–312.

Olsen, A.K., Bjortuft, H., Wiger, R., Holme, J., Seeberg, E., Bjoras, M. et al. (2001) Highly efficient base excision repair (BER) in human and rat male germ cells. *Nucleic Acids Research*, 29, 1781–1790.

Olsen, A.K., Duale, N., Bjoras, M., Larsen, C.T., Wiger, R., Holme, J.A. et al. (2003) Limited repair of 8-hydroxy-7,8-dihydroguanine residues in human testicular cells. *Nucleic Acids Research*, 31, 1351–1363.

Olsen, A.K., Lindeman, B., Wiger, R., Duale, N. & Brunborg, G. (2005) How do male germ cells handle DNA damage? *Toxicology and Applied Pharmacology*, 207, 521–531.

Rodriguez-Casuriaga, R., Folle, G.A., Santinaque, F., Lopez-Carro, B. & Geisinger, A. (2013) Simple and efficient technique for the preparation of testicular cell suspensions. *Journal of Visualized Experiments*, 78, 50102.

Sasaki, Y.F., Sekihashi, K., Izumiyama, F., Nishidate, E., Saga, A., Ishida, K. et al. (2000) The comet assay with multiple mouse organs: comparison of comet assay results and carcinogenicity with 208 chemicals selected from the IARC monographs and U.S. NTP Carcinogenicity Database. *Critical Reviews in Toxicology*, 30, 629–799.

Sharma, A.K., Boberg, J. & Dybdahl, M. (2018) DNA damage in mouse organs and in human sperm cells by bisphenol a. *Toxicological and Environmental Chemistry*, 100, 465–478.

Singer, T.M. & Yauk, C.L. (2010) Germ cell mutagens: risk assessment challenges in the 21st century. *Environmental and Molecular Mutagenesis*, 51, 919–928.

Singh, N.P., Mccoy, M.T., Tice, R.R. & Schneider, E.L. (1988) A simple technique for quantitation of low levels of DNA damage in individual cells. *Experimental Cell Research*, 175, 184–191.

Smith, C.C., Adkins, D.J., Martin, E.A. & O'donovan, M.R. (2008) Recommendations for design of the rat comet assay. *Mutagenesis*, 23, 233–240.

Speit, G. & Hartmann, A. (2006) The comet assay. In: Henderson, D.S. (Ed.) *DNA repair protocols: mammalian systems*. Totowa, NJ: Humana Press.

Speit, G., Kojima, H., Burlinson, B., Collins, A.R., Kasper, P., Plappert-Helbig, U. et al. (2015) Critical issues with the in vivo comet assay: a report of the comet assay working group in the 6th international workshop on genotoxicity testing (IWGT). *Mutation Research, Genetic Toxicology and Environmental Mutagenesis*, 783, 6–12.

Speit, G., Vasquez, M. & Hartmann, A. (2009) The comet assay as an indicator test for germ cell genotoxicity. *Mutation Research*, 681, 3–12.

Steiblen, G., Benthem, J.V. & Johnson, G. (2020) Strategies in genotoxicology: acceptance of innovative scientific methods in a regulatory context and from an industrial perspective. *Mutation Research, Genetic Toxicology and Environmental Mutagenesis*, 853, 503171.

Su, L., Mruk, D.D. & Cheng, C.Y. (2011) Drug transporters, the blood-testis barrier, and spermatogenesis. *The Journal of Endocrinology*, 208, 207–223.

UN ENVIRONMENT PROGRAMME. (2019) UN Environment Global Chemicals Outlook II-From Legacies to Innovative Solutions: Implementing the 2030 Agenda for Sustainable Development.

UNECE. (2020) Proposal to create an informal working group on clarification of the criteria for classification for germ cell mutagenicity in category 1B and set up its terms of reference. Committee of Experts on the Transport of Dangerous Goods and on the Globally Harmonized System of Classification and Labelling of Chemicals.

UNECE. (2021) Status of the work of the informal working group on clarification of the criteria for classification for germ cell mutagenicity. Committee of Experts on the Transport of Dangerous Goods and on the Globally Harmonized System of Classification and Labelling of Chemicals.

Vasquez, M.Z. (2012) Recommendations for safety testing with the in vivo comet assay. *Mutation Research*, 747, 142–156.

Yauk, C.L., Aardema, M.J., Benthem, J., Bishop, J.B., Dearfield, K.L., Demarini, D.M. et al. (2015) Approaches for identifying germ cell mutagens: report of the 2013 IWGT workshop on germ cell assays(). *Mutation Research, Genetic Toxicology and Environmental Mutagenesis*, 783, 36–54.

Zainol, M., Stoute, J., Almeida, G.M., Rapp, A., Bowman, K.J., Jones, G.D. et al. (2009) Introducing a true internal standard for the comet assay to minimize intra- and inter-experiment variability in measures of DNA damage and repair. *Nucleic Acids Research*, 37, e150.

Zhang, Y., Mustieles, V., Yland, J., Braun, J.M., Williams, P.L., Attaman, J.A. et al. (2020) Association of Parental Preconception Exposure to phthalates and phthalate substitutes with preterm birth. *JAMA Network Open*, 3, e202159.

Zheng, H. & Olive, P.L. (1997) Influence of oxygen on radiation-induced DNA damage in testicular cells of C3H mice. *International Journal of Radiation Biology*, 71, 275–282.

How to cite this article: Dirven, Y., Eide, D.M., Henriksson, E.W., Hjorth, R., Sharma, A.K., Graupner, A. et al. (2023) Assessing testicular germ cell DNA damage in the comet assay; introduction of a proof-of-concept. *Environmental and Molecular Mutagenesis*, 64(2), 88–104. Available from: https://doi.org/10.1002/em.22527