THE FIRST GAMMA-H2AX BIODOSIMETRY INTERCOMPARISON EXERCISE OF THE DEVELOPING EUROPEAN BIODOSIMETRY NETWORK RENEB

S. Barnard[1,*], E. A. Ainsbury[1], J. Al-hafidh[1], V. Hadjidekova[2], R. Hristova[2], C. Lindholm[3], O. Monteiro Gil[4], J. Moquet[1], M. Moreno[5], U. Rößler[6], H. Thierens[7], C. Vandevoorde[7], A. Vral[7], M. Wojewódzka[8] and K. Rothkamm[1]

[1]Public Health England, Centre for Radiation Chemical and Environmental Hazards, Chilton, UK
[2]National Centre of Radiobiology and Radiation Protection, Sofia, Bulgaria
[3]Radiation and Nuclear Safety Authority (STUK), Helsinki, Finland
[4]Instituto Superior Técnico, Universidade de Lisboa, CTN, Grupo de Protecção e Segurança Radiológica, Bobadela-LRS, Portugal
[5]Servicio Madrileño de Salud, Hospital General Universitario Gregorio Marañón, Madrid, Spain
[6]Bundesamt für Strahlenschutz, Oberschleissheim, Germany
[7]Department of Basic Medical Sciences, Ghent University, Ghent, Belgium
[8]Institute of Nuclear Chemistry and Technology, Center for Radiobiology and Biological Dosimetry, Warsaw, Poland

*Corresponding author: stephen.barnard@phe.gov.uk

In the event of a mass casualty radiation incident, the gamma-H2AX foci assay could be a useful tool to estimate radiation doses received by individuals. The rapid processing time of blood samples of just a few hours and the potential for batch processing, enabling high throughput, make the assay ideal for early triage categorisation to separate the 'worried well' from the low and critically exposed by quantifying radiation-induced foci in peripheral blood lymphocytes. Within the RENEB framework, 8 European laboratories have taken part in the first European gamma-H2AX biodosimetry exercise, which consisted of a telescoring comparison of 200 circulated foci images taken from 8 samples, and a comparison of 10 fresh blood lymphocyte samples that were shipped overnight to participating labs 4 or 24 h post-exposure. Despite large variations between laboratories in the dose–response relationship for foci induction, the obtained results indicate that the network should be able to use the gamma-H2AX assay for rapidly identifying the most severely exposed individuals within a cohort who could then be prioritised for accurate chromosome dosimetry.

INTRODUCTION

Following a major radiation incident, biological dosimetry would ideally assist in distinguishing the critically exposed from unexposed or worried-but-well individuals. This would enable a more efficient use of medical expertise and hospital facilities, as well as reassure those with little or no immediate health risks. The gamma-H2AX foci assay has become a well-established biomarker for radiation-induced DNA double-strand breaks since it was first reported over a decade ago[1]. The assay uses immunofluorescence staining techniques to 'tag' the phosphorylated histone variant gamma-H2AX, which is known to form at the site of DNA double-strand breaks. This enables the use of gamma-H2AX assay as a tool for sensitive detection and quantification of radiation-induced cellular damage in low-dose research[2], diagnostic radiology[3–5] as well as cancer research and therapy[6]. It can also potentially be used as a triage tool in the event of a radiation accident and may, in this setting, provide initial results faster (within hours) and with higher throughput than the gold standard chromosome dosimetry assays[7–13].

One aim of the EC-funded RENEB project (Realising the European Network of Biodosimetry)[14] is to further develop and harmonise the application of the gamma-H2AX assay as a biodosimetric triage tool across a network of participating laboratories within Europe, to increase the Europe-wide availability, total capacity and accuracy of this assay.

Here, results of the first intercomparison exercise performed by this network are presented. Fluorescence microscopy images of irradiated, isolated and gamma-H2AX-immunostained blood lymphocytes were captured and sent electronically to other laboratories within the network for scoring. The second task involved the shipment of 10 coded blood lymphocyte samples to the participating laboratories for processing, scoring and dose estimation using either existing or newly established calibration curves from included reference samples exposed to known doses.

MATERIALS AND METHODS

Induction

The two main technical hurdles for establishing the gamma-H2AX biodosimetry assay are sample processing, to achieve good quality staining, and consistent foci scoring criteria. The exercise was therefore preceded by a 2-d training course for those laboratories that had no or little previous experience with this assay, to provide a sound methodological basis for accurate analysis and future dose estimations. In addition, a 'picture book' defining foci scoring criteria was circulated among participants to harmonise scoring approaches between laboratories.

Telescoring exercise

Blood lymphocytes were isolated, exposed to 0.5-, 1-, 2- and 4-Gy ^{137}Cs, incubated for 4 or 24 h, spotted onto silane-coated slides, fixed and immunostained as detailed below following an established protocol[11]. Maximum projection images of seven z-planes at 1-μm step size were captured for gamma-H2AX and DAPI staining using a $\times 40$ objective on a MetaSystems Metafer microscope system. Coded images were sent electronically to participants for foci scoring and ranking of exposure severity based on foci counts.

Blood sampling, irradiations and shipments

Following ethics approval (Berkshire REC 09/H0505/87) and informed consent, blood samples from two healthy volunteers were collected into lithium heparin tubes. Separate donors were required for each time point based on the quantity of whole blood needed for transport to the partner laboratories (ethics limit to 25 ml per donor). Based on existing studies[15, 16] and previous experience, variability between both donors was expected to be insignificant and make a negligible contribution to the overall uncertainty. Lymphocytes were isolated using Histopaque1077 (Sigma–Aldrich), suspended in a minimum essential medium (supplemented with 10 % foetal calf serum, 1 % L-glutamine and 1 % penicillin/streptomycin), exposed to ^{137}Cs gamma rays ($\sim$0.6 Gy min^{-1}) and incubated at 37°C for 4 or 24 h. For shipments, lymphocytes were suspended in foetal bovine serum, aliquoted into cryovials, wrapped and transported overnight on frozen cold packs, together with temperature loggers, to the participating laboratories. Calibration samples (0, 1, 2, 3 and 4 Gy) or, for laboratories with existing calibration curves, negative and positive controls (0- and 2-Gy samples, respectively) were included in the shipments, in addition to the coded intercomparison samples. These coded samples included 0-, 1- and 4-Gy uniformly irradiated lymphocytes and two non-uniformly exposed samples

consisting of 0 + 3 Gy mixed 1:1 to give a whole-body equivalent dose of 1.5 Gy and 2 + 3 Gy mixed 1:1 to give a whole-body equivalent dose of 2.5 Gy. Only samples from the same donor were mixed. Following 4- or 24-h incubation at 37°C, lymphocytes were shipped overnight to six participating laboratories. Temperature loggers included with the samples showed that temperatures always remained at 3–9°C for all shipments. Four participants received a full set of calibration samples for the 4-h repair time point, with two laboratories opting for reference samples only. All six participants received 4-h coded samples. Four of the participants also received coded 24-h samples.

Gamma-H2AX immunofluorescence staining and microscopic foci analysis

In brief, lymphocytes were washed and spotted or cytospun onto adhesive microscope slides, fixed in formaldehyde, permeabilised and extracted with Triton X-100, blocked with bovine serum albumin and immunostained for gamma-H2AX using fluorophore-conjugated secondary antibodies[11, 17]. Foci scoring was performed by fluorescence microscopy, either manually or automatically using MetaCyte software[17].

Data analysis

The Wilcoxon signed-rank test and the Mann–Whitney U test were used to test the ranking of samples based on foci counts by laboratories in the telescoring exercise. DoseEstimate software[18] was used to fit linear dose–response curves using iteratively reweighted least-squares fitting and calculate associated standard errors. Papworth's u-test was used to check for any deviation of foci frequencies from a random distribution described by the Poisson model[19]. General linear model analysis of variance (ANOVA) and post-hoc testing (Tukey's pairwise comparisons) were performed for reported foci counts and dose estimates using Minitab software with the following experimental factors: post-exposure time (4 or 24 h), method (automatic or manual scoring), laboratory (1–8) and number of cells scored (20 versus 50 for manual and 50 versus 200 for automated scoring).

RESULTS AND DISCUSSION

Eight laboratories took part in the telescoring exercise, with three participants opting to score automatically as well as manually, and two participants using more than one scorer. Results indicated considerable discrepancies between laboratories (less so between scorers from the same laboratory) regarding radiation-induced foci yields obtained by manual and

automated scorings (Table 1). However, samples could still be ranked in order of lowest to highest estimated radiation exposure based purely on mean foci/cell counts, and these could then be correlated for all participants (indicated by the consistent shading patterns in Table 1). The Wilcoxon signed-rank test revealed no significant difference between the median ranked dose and the results of each lab, for any of the samples (p all > 0.505). The Mann–Whitney U test was then used to compare the rankings of each laboratory for each sample, assigned in order of triage priority. There was no evidence of any significant difference between the labs' average rankings and the true triage rank ($p > 0.999$). Overall, the results suggest that even in the absence of a calibration curve, and—in the case of laboratories #2 and 3—with hardly any experience in the foci assay, all laboratories were able to distinguish critically high (2 and 4 Gy) and low-to-moderate (0.5 and 1 Gy) exposures from unirradiated samples, when processed 4 h post-exposure. Foci counts for the 1-Gy sample processed 24 h post-exposure were consistently lower than those at 4 h for the same dose, although still higher than the 0-Gy samples, in line with previous observations of biexponential foci loss post-exposure[11]. The reason for using incubation periods of 4 and 24 h within this study is based mainly on practicality: as discussed earlier, the assay shows much promise as a triage tool to initial screening of persons during a mass radiation accident. Realistically, 4–24 h is the most appropriate time window during which a sample is likely to be taken post-exposure.

Calibration coefficients in Table 2 indicate once again considerable discrepancy of foci yields between laboratories and lower foci yields at 24 h (mean of 0.9 foci per cell per gray) than at 4 h (mean of 2.1 foci per cell per gray), consistent with recent findings for shipments of ^{60}Co gamma-irradiated lymphocyte samples[19] and X-irradiated whole-blood samples[20]. Based on general linear model ANOVA, all doses except the 3 and 4 Gy points differed significantly from each other for all laboratories that produced new calibration curves ($p < 0.001$, individual data not shown). Significant differences were also found between laboratories and between 4- and 24-h incubation times ($p < 0.001$ for both). No significant difference between calibration curves was seen for manual versus automated scoring methods ($p = 0.115$). Post-hoc analysis confirmed that the significant difference between laboratories could be attributed mainly to one participant; excluding this data set would reduce significance greatly. Additionally, see Supplementary on-line Figure 1 for plotted calibration curves from all participating laboratories. This figure may give an easier understanding of the variation among calibration and a direct comparison of manual versus automated scoring methods in relevant cases.

Foci counts for the different doses differed significantly from each other for all but the non-uniform exposures of 2.5 and 1.5 Gy, which were statistically different from all other coded samples except from 2 and 1 Gy, respectively. Papworth's u-test was used to determine whether non-uniform exposure could be detected based on an overdispersed foci distribution. While 42 out of 59 analysed coded samples were technically overdispersed ($u > 1.96$) and 3 were underdispersed ($u < -1.96$), the highest u value was associated with the 1.5-Gy sample in 9 out of 12 data sets (5/6 manual 4 h; 2/3 manual 24 h; 2/3 auto), confirming previous reports that significant partial exposure (in this case, 3 Gy to 50 % of cells) may be detectable using foci distribution analysis[3, 11, 17]. As expected, u values for the only slightly non-uniformly exposed 2.5-Gy sample were indistinguishable from those for the uniformly exposed samples. The use of more sophisticated distribution analysis techniques[21] may help further improve the performance of the foci assay for detecting non-uniform exposures.

Each partner estimated the whole-body equivalent dose delivered to each coded sample based on a laboratory-specific calibration curve (Table 2) that had been created either independently or using lymphocytes sent together with the coded samples. ANOVA of dose estimates in Table 3 demonstrated a significant effect of dose ($p < 0.001$), with all dose levels statistically different from each other apart from the 1-Gy uniform and 1.5-Gy non-uniform samples and the 2.5-Gy non-uniform and 4-Gy uniform samples. There was no significant effect of number of cells or post-exposure time, but a significant difference was observed between manual and automatic scoring ($p = 0.001$) and between participating laboratories ($p < 0.001$). However, post-hoc testing demonstrated that the inter-laboratory variation was chiefly due to variation between two laboratories (3 and 4) and the others. Average dose estimates across all laboratories correlated very well with actual whole-body equivalent doses, at least for manual scoring, whereas automatic scoring tended to result in dose underestimation but not significantly ($p = 0.394$ general linear model ANOVA) (Figure 1).

Dose estimates were then used to assign samples to triage categories of <1 Gy, $1–2$ Gy and >2 Gy (white, light grey and dark grey shading, respectively, in Table 3). Manual scoring of 4-h samples achieved the most accurate assignment of triage categories, with only one laboratory systematically underestimating doses significantly enough to affect triage categories. Dose estimates reported for 24-h samples and those based on automated scoring tended to show more deviation from the correct triage categories, potentially questioning the usefulness of this approach for triage. However, only three laboratories contributed to each of these data sets, and a larger-scale study

Table 1. Telescoring exercise.

Lab_scorer	D 0Gy+4h	F 0Gy+4h	H 0.5Gy+4h	C 1Gy+4h	G 1Gy+4h	B 2Gy+4h	A 4Gy+4h	E 1Gy+24h
Manual scoring								
1_1	0.02	0.06	0.44	1.12	1.42	1.78	2.90	0.42
1_2	0.00	0.16	1.06	1.50	1.70	2.26	2.40	0.52
1_3	0.00	0.10	1.42	1.56	1.52	2.52	3.30	0.48
1_4	0.02	0.58	1.24	1.88	2.02	2.04	2.50	0.88
2	0.00	0.86	2.72	3.14	3.20	4.44	3.70	0.72
3	0.10	0.64	2.34	2.74	3.46	3.16	3.80	1.24
4	0.04	0.52	2.56	4.88	4.62	7.18	7.90	2.28
5	0.00	0.46	2.20	3.32	3.94	4.82	4.10	1.30
6	0.06	1.48	6.08	7.00	6.18	11.00	7.08	3.46
7_1	0.00	0.66	1.96	3.68	3.16	5.03	3.32	1.22
7_2	0.00	0.76	2.80	4.12	3.80	5.02	3.60	1.74
8	0.02	0.58	1.60	2.16	2.54	3.38	3.22	1.20
Auto scoring								
6	0.34	1.00	5.40	6.25	4.11	11.05	7.11	3.10
5	0.01	0.08	0.92	2.10	3.34	3.85	3.45	0.36
4	0.01	0.06	0.38	0.91	2.20	3.57	1.54	0.30

Average foci counts per cell reported for eight coded samples following image-based scoring by 12 manual scorers and 3 automated analyses using MetaCyte®. Results are ranked (white to dark grey) for each scorer in order of estimated severity of exposure based on foci counts.

Table 2. Constant (C) and linear calibration coefficients (α) and associated standard errors used by the different labs to convert foci counts into dose estimates.

Lab	Analysis	$C \pm SE$	$\alpha \pm SE$
4 h post-exposure			
1	Manual	0.04 ± 0.02	2.15 ± 0.09
3	Manual	0.44 ± 0.16	1.90 ± 0.21
4	Manual	0.06 ± 0.03	2.24 ± 0.09
5	Manual	0.1 ± 0.09	1.47 ± 0.22
5	Automated	0.48 ± 0.09	1.49 ± 0.12
6	Manual	0.12 ± 0.08	3.57 ± 0.23
6	Automated	0.45 ± 0.28	3.52 ± 0.46
7	Manual	0.69 ± 0.06	0.72 ± 0.07
24 h post-exposure			
1	Manual	0.03 ± 0.04	0.79 ± 0.14
3	Manual	0.28 ± 0.12	0.92 ± 0.14
4	Manual	0.09 ± 0.03	1.35 ± 0.06
5	Automated	0.39 ± 0.08	0.65 ± 0.08

accuracy of dose estimations. Reducing the number of cells scored manually from 50 to 20 or automatically from 200 to 50 did not significantly affect dose estimates or assigned triage categories ($p = 0.855$) and could therefore be considered in a large-scale emergency, to increase the assay throughput.

It is recognised from this study that there are considerable variables to consider when drawing comparisons and conclusions of data collected from a range of laboratories. Each participant will have developed different predetermined criteria for scoring foci, resulting in different absolute yields scored for a given dose by the different laboratories. The authors attempted to control this factor with the above-mentioned training sessions and picture library, but inevitably some variation is unavoidable. One must also consider the practicalities of equipment. There will be variation in the microscopes used; spectral and brightness differences in the light sources and fluorophores, wavelength ranges between different fluorescence filters and the availability of antibody from suppliers in each country will alter. These issues have previously been discussed in-depth[7, 22]. Standardising these equipment requirements is an expensive and complicated process and would not equate to a cost-effective and rapid assay. Sending each laboratory calibration samples and encouraging the participants to produce their own curve coefficients seem the most appropriate solution for

would really be required to fully address this question. Furthermore, the post-hoc ANOVA results for dose, above, show that the triage categories used in this work were statistically distinguishable as only doses within triage categories were not significantly different from each other. Also, additional training and frequent calibration of the assay may help improve the

Table 3. Foci-based dose estimates reported for five coded blood lymphocyte samples shipped from lab 1 to the other labs following ^{137}Cs gamma irradiation and 4-/24-h incubation.

Cells scored	Whole-body dose equivalent (Gy)									
	0		1		1.5		2.5		4	
	20	50	20	50	20	50	20	50	20	50
Manual scoring										
1–4 h	0	0	1.2	1.1	1.1	1.6	2.5	2.4	2.9	2.8
3–4 h	0.1	0.1	0.3	0.4	0.8	0.7	1.2	1.6	1.6	1.5
4–4 h	0.2	0.1	1	1.1	1.6	1.7	4.1	3.9	3.1	3.1
5–4 h	0.1	0	0.7	0.8	2	1.9	3.2	2.8	6.1	6.1
6–4 h	0	0	1.7	1.8	2.3	2	3.4	3	3.9	4
7–4 h	0.3	0.5	1.8	1.9	1.2	1.4	3.2	2.7	8.2	5.8
1–24 h	a	a	1.5	1.9	2.8	1.8	3.3	3.5	4.2	3.6
3–24 h	0	0	0.8	1.2	2.4	2.1	4.6	3.7	0.1	0.3
4–24 h	0.2	0.1	0.7	0.9	0.8	0.3	0.3	0.6	2.4	2.3
Cells scored	50	200	50	200	50	200	50	200	50	200
Auto scoring										
5–4 h	0	0	0.3	0.6	1.1	1	1.7	2.3	1.1	1.4
6–4 h	0	0.1	1.2	1	0.2	0.3	1.6	0.9	1.6	1.7
5–24 h	0	0	1.1	1	1.2	1.6	2.1	1.5	3.8	3.3

Lymphocytes were fixed, immunostained and scored manually and/or automatically with MetaCyte®. Shading of dose estimates reflects triage categories of <1 Gy (white), 1–2 Gy (light grey) and >2 Gy (dark grey). Dose estimates are shown based on manual scoring of 20 or 50 cells or automated scoring of 50 or 200 cells.
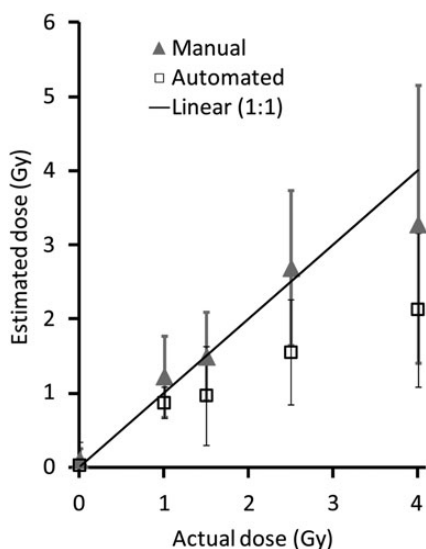aNot enough cells obtained.

Figure 1. Average gamma-H2AX-based estimates of whole-body equivalent doses versus true whole-body equivalent doses for manually or automatically analysed samples. Error bars show the standard deviation between the 8–9 (manual) and 3 (automated) measurements for each sample taken from Table 3. The line is not a fit but indicates the ideal 1:1 relationship.

these technical hurdles to establish a functional gamma-H2AX biodosimetry network across Europe.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Rogakou, E. P., Boon, C., Redon, C. and Bonner, W. M. *Megabase chromatin domains involved in DNA double-strand breaks in vivo.* J. Cell Biol. **146**, 905–916 (1999).
2. Rothkamm, K. and Löbrich, M. *Evidence for a lack of DNA double-strand break repair in human cells exposed*

to very low X-ray doses. Proc. Natl. Acad. Sci. USA. **100**, 5057–5062 (2003).

3. Rothkamm, K., Balroop, S., Shekhdar, J., Fernie, P. and Goh, V. *Leukocyte DNA damage following multi-detector row CT: a quantitative biomarker of low level radiation exposure*. Radiology **242**, 244–251 (2007).

4. Beels, L., Bacher, K., Smeets, P., Verstraete, K., Vral, A. and Thierens, H. *Dose-length product of scanners correlates with DNA damage in patients undergoing contrast CT*. Eur. J. Radiol. **81**, 1495–1499 (2012).

5. May, M. S. *et al*. *Induction and repair of DNA double-strand breaks in blood lymphocytes of patients undergoing ^{18}F-FDG PET/CT examinations*. Eur. J. Nucl. Med. Mol. Imaging **39**, 1712–1719 (2012).

6. Ivashkevich, A., Redon, C. E., Nakamura, A. J., Martin, R. F. and Martin, O. A. *Use of the gamma-H2AX assay to monitor DNA damage and repair in translational cancer research*. Cancer Lett. **327**, 123–133 (2012).

7. Rothkamm, K. and Horn, S. *Gamma-H2AX as protein biomarker for radiation exposure*. Ann. Ist. Super. Sanita. **45**, 265–271 (2009).

8. Ainsbury, E. A. *et al*. *Review of retrospective dosimetry techniques for external ionising radiation exposures*. Radiat. Prot. Dosim. **147**, 573–592 (2011).

9. Redon, C. E., Nakamura, A. J., Gouliaeva, K., Rahman, A., Blakely, W. F. and Bonner, W. M. *The use of gamma-H2AX as a biodosimeter for total-body radiation exposure in non-human primates*. PLoS One **5**, e15544 (2010).

10. Horn, S. and Rothkamm, K. *Candidate protein biomarkers for radiation exposure*. Radiat. Meas. **46**, 903–906 (2011).

11. Horn, S., Barnard, S. and Rothkamm, K. *Gamma-H2AX-based dose estimation for whole and partial body radiation exposure*. PLoS One **6**, e25113 (2011).

12. Turner, H. C. *et al*. *Adapting the gamma-H2AX assay for automated processing in human lymphocytes. 1. technological aspects*. Radiat. Res. **175**, 282–290 (2011).

13. Ghardi, M., Moreels, M., Chatelain, B., Chatelain, C. and Baatout, S. *Radiation-induced double strand breaks and subsequent apoptotic DNA fragmentation in human peripheral blood mononuclear cells*. Int. J. Mol. Med. **29**, 769–780 (2012).

14. Kulka, U. *et al*. *Realising the European Network of Biodosimetry (RENEB)*. Radiat. Prot. Dosim. **151**, 621–625 (2012).

15. Roch-lefevre, S., Mandina, T., Voisin, P., Gaetan, G., Mesa, J. E., Valente, M., Bonnesoeur, P., Garcia, O. and Roy, L. *Quantification of gamma-H2AX foci in human lymphocytes: a method for biological dosimetry after ionizing radiation exposure*. Radiat. Res. **174**(2), 185–194 (2010).

16. Chua, M. L. *et al*. *Inter-individual and inter-cell type cariation in residual DNA damage after in vivo irradiation of human skin*. Radiother. Oncol. **99**(2), 225–230 (2011).

17. Rothkamm, K. *et al*. *Manual versus automated gamma-H2AX foci analysis across five European laboratories: can this assay be used for rapid biodosimetry in a large scale radiation accident?* Mutat. Res. **756**(1–2), 170–173 (2013).

18. Ainsbury, E. A. and Lloyd, D. C. *Dose estimation software for radiation biodosimetry*. Health Phys. **99**, 290–295 (2010).

19. International Atomic Energy Agency. *Cytogenetic dosimetry: applications in preparedness for and response to radiation emergencies*. EPR-Biodosimetry 2011. IAEA (2011).

20. Rothkamm, K. *et al*. *NATO biodosimetry study laboratory inter-comparison on the gamma-H2AX foci assay*. Radiat. Res. **180**(2), 149–155 (2013).

21. Ainsbury, E. A., Vinnikov, V. A., Maznyk, N. A., Lloyd, D. C. and Rothkamm, K. *A comparison of six statistical distributions for analysis of chromosome aberration data for radiation biodosimetry*. Radiat. Prot. Dosim. **155**(3), 253–267 (2013).

22. Barnard, S., Bouffler, S. and Rothkamm, K. *The shape of the radiation dose response for DNA double-strand break induction and repair*. Genome Integr. **4**(1), 1 (2013). doi: 10.1186/2041-9414-4-1.