# RADIATION DOSE ESTIMATION BY COMPLETELY AUTOMATED INTERPRETATION OF THE DICENTRIC CHROMOSOME ASSAY

Yanxin Li[1], Ben C. Shirley[1], Ruth C. Wilkins[2], Farrah Norton[3], Joan H.M. Knoll[1,4] and Peter K. Rogan[1,5,*]
[1]CytoGnomix, Inc., POB 27052, 60 N. Centre Rd, London, ON N5X 3X5, Canada
[2]Health Canada, Environmental and Radiation and Health Sciences Directorate, Ottawa, ON K1A 1C1, Canada
[3]Canadian Nuclear Laboratories, 286 Plant Rd, Chalk River, ON K0J 1J0, Canada
[4]Department of Pathology and Laboratory Medicine, University of Western Ontario, 1151 Richmond St., London, ON N6A 5C1, Canada
[5]Department of Biochemistry, University of Western Ontario, 1151 Richmond St., London, ON N6A 5C1, Canada

*Corresponding author: progan@uwo.ca

Accuracy of the automated dicentric chromosome (DC) assay relies on metaphase image selection. This study validates a software framework to find the best image selection models that mitigate inter-sample variability. Evaluation methods to determine model quality include the Poisson goodness-of-fit of DC distributions for each sample, residuals after calibration curve fitting and leave-one-out dose estimation errors. The process iteratively searches a pool of selection model candidates by modifying statistical and filter cut-offs to rank the best candidates according to their respective evaluation scores. Evaluation scores minimize the sum of squared errors relative to the actual radiation dose of the calibration samples. For one laboratory, the minimum score for the curve fit residual method was $0.0475\,\mathrm{Gy}^2$, compared to $1.1975\,\mathrm{Gy}^2$ without image selection. Application of optimal selection models using samples of unknown exposure produced estimated doses within 0.5 Gy of physical dose. Model optimization standardizes image selection among samples and provides relief from manual DC scoring, improving accuracy and consistency of dose estimation.

## INTRODUCTION

The Automated Dicentric Chromosome Identifier and Dose Estimator (ADCI) has been developed to automate the cytogenetic DC assay for dose estimation from radiation exposure[1]. The software uses image processing segmentation techniques to find chromosomes and machine learning (ML)-based algorithm to recognize centromeres and designate DCs[2]. The ML parameter, $\sigma$, balances sensitivity and specificity of DC detection (1.4 or 1.5 have highest accuracy for dose estimation). Nevertheless, the algorithm was originally less accurate at < 1 Gy, due to the failure to eliminate false positive (FP) DCs[3]. One goal of this study was to mitigate, and if possible, eliminate FPs. Sources of FPs in cell images included incorrectly classified monocentric chromosomes (MC), overlapped chromosomes, separated sister chromatids or other nonchromosomal objects. FP assignments were caused by limitations of the current ML algorithm which incorrectly interpreted certain chromosome morphologies or non-chromosomal objects.

We have implemented a computational approach which corrects FPs at the chromosome-level by analyzing individual positive detections and reduces FPs at the image-level by selecting optimal images in samples for analysis[4]. Previous approaches for metaphase cell selection have applied fixed thresholds to features extracted from objects present in these images for binary classification as either suitable or unanalyzable[5, 6]. Rule-based criteria for metaphase assessment have also extracted morphological features of objects enclosed by rectangles to partition chromosomes[7] or images[8] into small numbers of generally analyzable classes, with the exception of poor quality samples with high levels of debris or low mitotic indices[8]. User-scored features have been combined into quality measures which then classify metaphase cells into discrete groups[9]. Our approach integrates aspects of metaphase cell selection with radiation dose determination. We perform image segmentation with gradient vector flow snake and integrated intensity Laplacian algorithms to precisely define boundaries of all objects in an image[1]. From these objects, ADCI derives all features required for both image selection and DC detection. ADCI does not rely on fixed thresholds for image selection and removes only images lacking metaphase cells. Features used for image selection include object count and shape (used in ref.[5, 6]), contour finite difference, a novel measure of texture coarseness (used in ref.[10]), as well as likely centromeres[2]. Two different, novel quality measures have been developed for ranking metaphase images[4]. We enumerate a range of parameter values used in image ranking to

determine the best image selection criteria (termed models). These criteria generate optimal calibration curves based on either minimizing fit residuals, maximizing Poisson distribution of DCs, or estimating radiation dose of calibration samples whose exposures are blinded during curve generation.

## METHODS

Data have been generated by biodosimetry laboratories at Health Canada (HC) and Canadian Nuclear Laboratories (CNL). Blood samples were irradiated by HC with an X-RAD-320 (Precision X-ray, North Branford, CT), at CNL with a $^{137}$Cs GammaCell40 (Atomic Energy of Canada Ltd, Ottawa, ON), and processed at both facilities using established protocols[11]. One set of metaphase images from CNL (referred as CNL-low, 1 Gy physical dose) and two sets from HC (referred as HC-low and HC-high, 1 Gy and 3.0–4.0 Gy physical dose, respectively) were used for algorithm development and testing. Images were captured using a Metafer slide scanning platform (Metasystems, Newton, MA). Calibration curves were prepared for samples irradiated at known exposures (0, 0.25, 0.5, 0.75, 1, 2, 3, 4 and 5 Gy). Samples from previous exercises (designated HS## and CS## in Table 1) were analyzed with these curves before versus after application of

matched image selection models. After completing training on ADCI[12], authors from HC and CNL performed these analyses independently.

## Image and DC Filtering

False positive DCs (FPs) and true positive DCs (TPs) were discriminated based on their quantitative morphological characteristics such as contours and width profiles using a set of quantitative filters to delineate FPs[3]. All TPs were retained after processing of FPs (i.e. 100% specificity). On average, $55 \pm 9.6\%$ of FPs were removed among all sets; individually the filters eliminated 52% of FPs from the CNL set, which was comparable to the HC sets (66% and 48% for low and high dose image sets, respectively).

Additionally, ADCI automatically selects images by filtering and scoring all objects detected in an image, only a small fraction of which are DCs. Image segmentation filters measure morphological features and use empirically determined criteria to eliminate poor images. Filtering calculates and eliminates images exceeding thresholds for:

(I)  average length–width ratio of objects
(II)  average centromere candidate density of objects
(III)  average contour finite difference (measuring object concavity) of objects

**Table 1. Actual versus estimated exposures (in Gy) of Health Canada (HS) and Canadian Nuclear Laboratories (CS) Exercise Samples before and after image selection.**

*Selection model*

| Sample ID | HS01 | HS08 | HS10 | HS04 | HS05 | HS07 | | | | RMSE (Gy) |
|---|---|---|---|---|---|---|---|---|---|---|
| Actual dose | 3.10 | 2.30 | 1.40 | 1.80 | 2.80 | 3.40 | | | | |
| Unselected | 2.40 | 0.60 | 0.00 | 2.85 | 1.80 | 3.90✓ | | | | 1.13 |
| *A_C*[1] | 3.15✓ | 1.80✓ | 1.20✓ | 3.90 | 2.70✓ | 3.90✓ | | | | 0.91 |
| *A_D*[2] | 2.85✓ | 1.50 | 1.25✓ | 3.45 | 2.45✓ | 3.90✓ | | | | 0.80 |
| ***A_B***[3] | 3.25✓ | 2.45✓ | 1.00✓ | 1.75✓ | 3.25✓ | 2.90✓ | | | | 0.33 |
| *Automated,178981*[4] | 3.60✓ | 2.15✓ | 1.25✓ | 3.90 | 2.75✓ | 3.90✓ | | | | 0.91 |

| Sample ID | CS01 | CS02 | CS03 | CS04 | CS05 | CS06 | CS07 | CS08 | CS09 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Actual dose | 1.20 | 3.10 | 0.00 | 0.90 | 4.00 | 2.80 | 1.70 | 3.60 | 2.30 | |
| Unselected | 1.50✓ | 3.95 | 1.25 | 1.35✓ | 4.65 | 3.20✓ | 2.55 | 4.35 | 2.35✓ | 0.70 |
| *C_B*[5] | 1.30✓ | 3.50✓ | 0.20✓ | 1.40✓ | 5.00 | 3.00✓ | 2.70 | 3.90✓ | 3.20 | 0.61 |
| *Automated 295*[6] | 0.40 | 2.70✓ | 0.10✓ | 0.00 | 3.70✓ | 2.20 | 2.30✓ | 3.50✓ | 2.10✓ | 0.52 |
| ***Automated 869***[7] | 1.05✓ | 3.00✓ | 0.80 | 0.60✓ | 3.95✓ | 2.30✓ | 2.30✓ | 3.80✓ | 2.60✓ | 0.41 |

✓: fulfills IAEA triage criteria for dose estimation; RMSE: Root Mean Square Error. Descriptions of *Selection Models* evaluated: Unselected: includes all metaphase images [1]*A_C* selects top 250 images ranked by combining Z-scores (based on the standard deviations from mean values of each Filter for all images in the same sample) of Filters I–VI weighted according to [5,2,4,3,4,1]; [2]*A_D* applies standard deviation thresholded image Filters I–VI; [3]*A_B* selects top 250 images scored by Group Bin Distance; [4]Leave-one-out (LOO)-optimized *Automated Model 178981* selects top 300 images ranked by combining Z-scores of Filters I–VI weighted according to [5,4,1,4,5,3]; [5]*C_B* applies standard deviation thresholded image Filters I–III and selects top 250 images ranked by Group Bin Distance; [6]LOO-optimized *Automated Model 295* applies thresholded Filters I–VI (Standard deviation range <1–1.5); [7]Poisson-optimized *Automated Model 869* applies Filters II–IV to all images. The most accurate dose estimates are indicated by **bolding** of the Selection Models with the lowest RMSE for data from each laboratory.

(IV)  total number of objects
 (V)  number of segmented objects
(VI)  ratio of classified objects (di- + monocentric) to segmented objects (di- + mono- + unclassifiable).

Ranking sorts images either by combining $Z$-scores of features in Filters I–VI (or a subset of these) to provide an assessment of image quality (i.e. combined $Z$-score), or binning chromosomes into three groups based on their areas which approximate numbers of base-pairs in each (i.e. Group Bin Distance). The weights of each of the features are specified prior to generating the combined $Z$-score. The Group Bin Distance uses the Euclidean distance between chromosome length bins and the standard karyotype[13] to score images. Regardless of which ranking method is used, we select ≥250 top-scoring images, based on our finding that ≥200 images are needed to stabilize DC frequencies[4]. Image selection models of empirically or automatically-derived parameter combinations were used to select images for calibration curves and estimating exposures.

### Automated Image Selection Model Generation

Aside from the pre-computed image selection models provided with ADCI[3], models can be computed and derived prospectively and then applied to calibration sample data. These models select images at a particular $\sigma$ value by an exhaustive search of parameters consisting of: combinations of Filters I–VI at different $Z$-score thresholds and/or top ranked images ordered either by combined $Z$-scores of Filters I–VI or by group bin distance to normal karyotypes (Figure 1). Thousands of potential image selection models can be derived by changing these parameters. Since optimal model configurations can differ between laboratories due to their particular sample preparation procedures, ADCI provides automated search functionality to compute optimal models for sets of samples. These automated searches involve both generation of a pool of possible models and a comparative evaluation of the performance of each model in the pool with all of the others. Each filter can either be enabled at a user specified threshold value or be disabled altogether. Images can be scored using the combined $Z$-score method (contents of an image selection model heading) or with the group bin method. The combined $Z$-score method apportions adjustable contributions of each feature using a weight vector. The number of selected top images, after images are scored and ranked, is also adjustable. Optimized image selection models are evaluated by determining overall scores, $S$, based on either the collated $P$-values of Poisson fit for every sample, or calibration curve fit residuals (requires ≥ three samples), or LOO cross-validation of dose estimation (requires ≥4 calibration samples).

Smaller scores, $S$, indicate better selection models. For Poisson evaluation, $P$-values for each of the filter scores applied in the model, $i$, are combined by Fisher's method:

$$S = -2 * \sum \ln p_i$$

For fit residual evaluation, least square residual differences between response, i.e. DC frequency, and the calibration curve estimate for each sample, j, are summed:

$$S = \sum (\text{Dose}_j - f(\text{Response}_j))^2$$

For LOO evaluation, one sample is treated as a test sample, and the remaining samples are treated as calibration samples. A curve is fit and the dose estimation error of the test sample calculated. This process is iterated over all samples. $S$ is the sum of the dose estimation errors over all samples:

$$S = \sum 2 * (\text{Dose}_j - f_j(\text{Response}_j))^2$$

## RESULTS

### Application of Image Selection Filters

Use of image selection models eliminates poor quality images, which reduces differences between samples based on features extracted from chromosome images. Heat maps of statistical comparisons of these feature distributions for different samples from the same laboratory indicates the results of applying image selection models to calibration and exercise samples (Figure 2). The differences between chromosome morphologies become less significant after the image selection model is applied (Panel B) which removes suboptimal images with extreme object counts from both types of samples. Nevertheless, several pairwise comparisons between the object count distributions in the calibration versus exercise sample image sets remain significant despite filtering, consistent with the possibility that there are batch effects that distinguish these groups of samples.

### Estimating Radiation Dose with Image Selection Models

Sample exposures were estimated and compared with corresponding actual radiation doses. The same image selection models used for calibration curve generation were also applied to exercise samples. The best image selection method for HC data scored images by Group Bin Distance and selected top ranked images in each sample to create a calibration curve. Accurate doses were obtained with CNL data
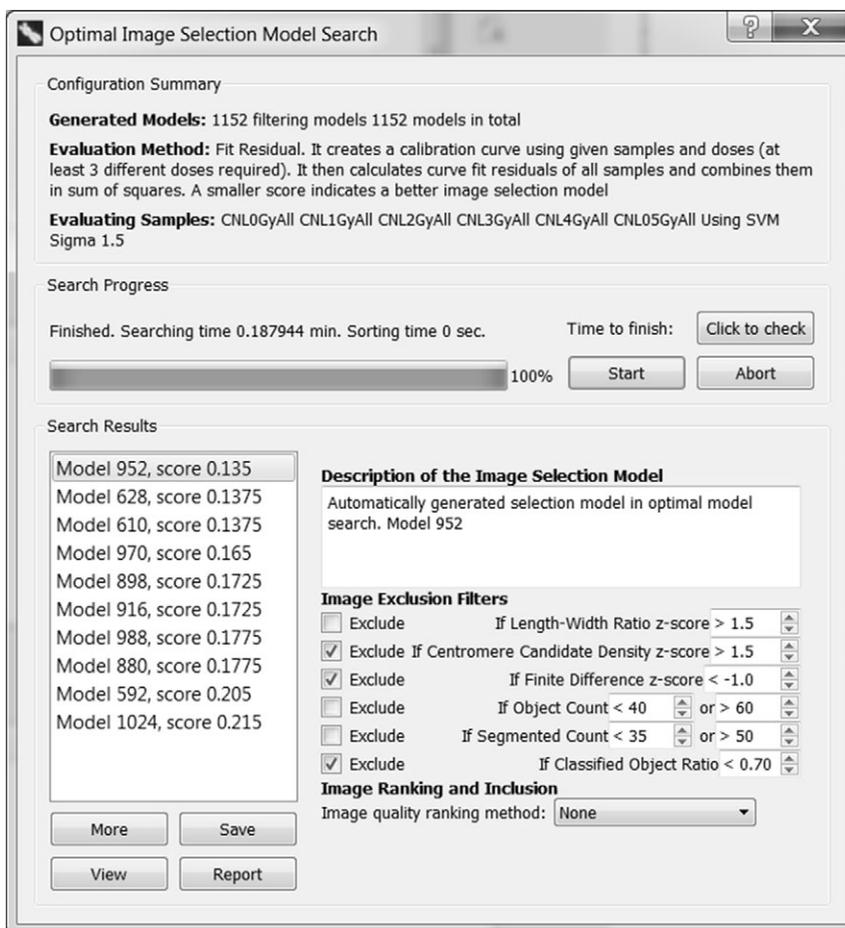
Figure 1. Panel showing results dialog after automated image selection model generation, according to criteria specified in the Wizard and Settings menus of ADCI. CNL calibration samples are fit to a curve with $\sigma = 1.5$. Image feature parameters, and optional image ranking and inclusion criteria are iterated across all selected combinations to generate 1152 different models. In this example, each model is evaluated by determining residuals upon fitting calibration curves. Scores indicate squared sums of residuals. Up to 50 Search Results can be displayed. Models are sorted by score; in this example, the optimal combination with smallest overall residuals is Model 952 (the first model highlighted). The Image Exclusion Filter elements shown correspond to the highlighted model. Individual or multiple models may be Saved, Viewed, or summarized as Reports. The same algorithm is applied with the other available evaluation methods, which can sometimes result in improved scores.

with Filters II–IV. Dose estimates for exercise samples after image model selection were compared with unselected images from the same samples (Table 1). Models with the highest accuracies for these international exercise samples (smallest RMSE) for HC were A_B (automated Model 178981 also performed well for all except one sample), and automated Models 869 and 295 for CNL.

## DISCUSSION

Validation of optimal image selection models results in accurate radiation dose estimates on calibration and exercise samples. Both pre-computed and automated optimized models generated dose estimates that met IAEA criteria for triage biodosimetry[1]. Interestingly, the models that performed well to select metaphase image data produced by one laboratory differed from models chosen for the other laboratory. However, after the best images were selected, dose estimates between laboratories were consistent. This was also apparent using biodosimetry samples irradiated at the same exposure that were independently processed by each laboratory[4]. Differences between calibration curves from individual laboratories are well known, yet the estimated
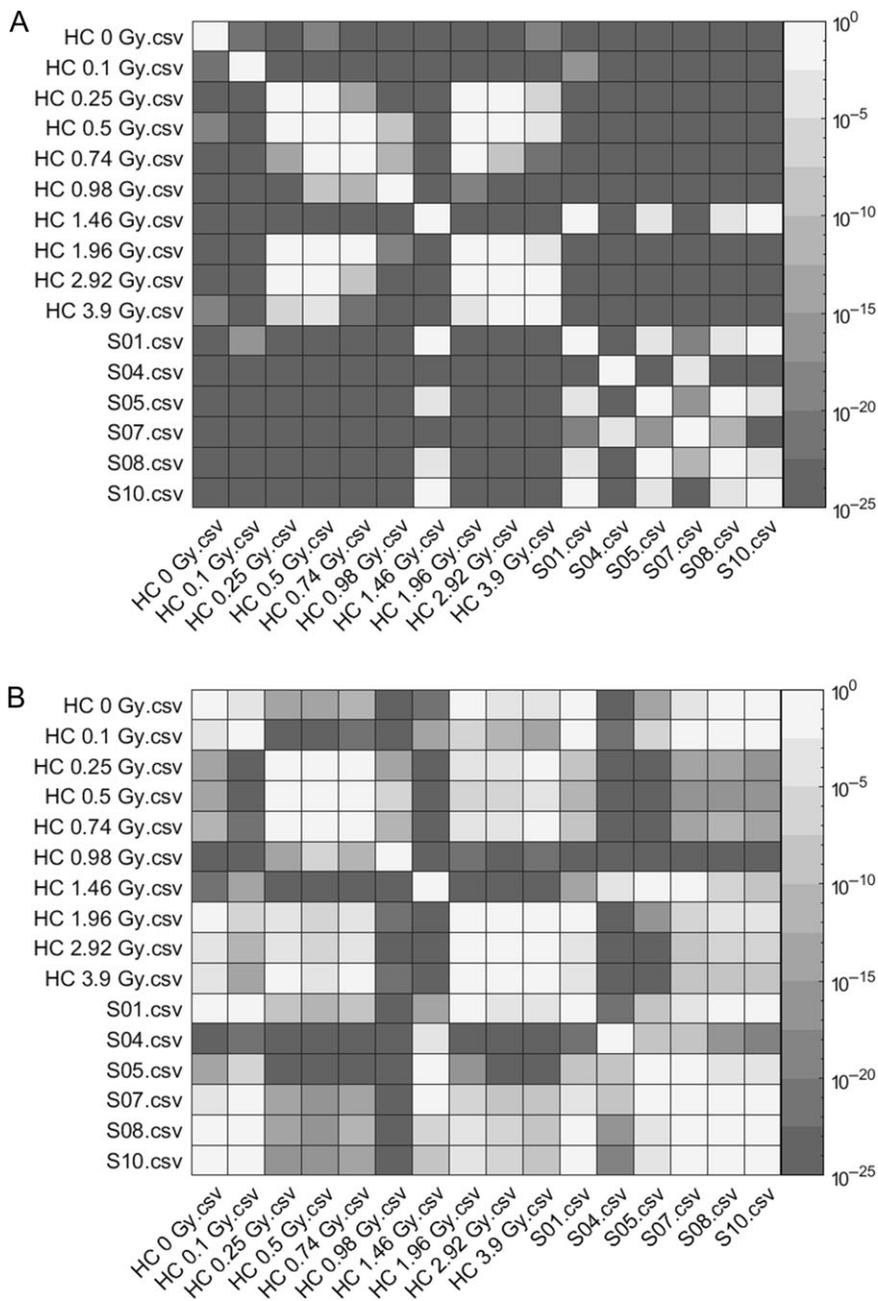
Figure 2. (**A**) Unfiltered samples. (**B**) Model B_C filtered samples. Representative heat maps of chromosome object count distributions for Health Canada calibration (HC ##Gy.csv) and exercise (S##.csv) samples. Panels A and B, respectively, show significance of pairwise comparisons of object counts in unselected versus Model B_C filter[3] selected images from each sample. The legends (rightmost column) display probability thresholds of Wilcoxon signed rank tests of chromosome object counts between each pair of samples from the HC laboratory. Similar results were obtained for other chromosomal features.

5

exposures of commonly sourced exercise samples were consistent to within 0.5 Gy (not shown). Since ADCI uses the same algorithms to call DCs for different samples, differences in dose estimates may largely be attributable to variations in metaphase image selection criteria.

Batch effects that differentiate calibration from exercise sample categories were also noted during image filtering. These effects may arise from inter-category differences in chromosome preparation methods, time in transit, or sample sources. Sample handling protocols, sources of culture media or reagents may differ between calibration and exercise sample processing between and within laboratories. Image filtering mitigated many of these batch effects. It is conceivable that uncorrected batch effects seen in unselected or poorly curated metaphase images could explain dose estimation errors in some exercise samples.

Once a calibration curve and an optimal image selection model have been established, the exposure level of any subsequent sample can be determined by applying the model after processing the corresponding images for DC detection. After culturing and slide preparation, sample analysis is constrained only by the time needed for microscope image acquisition and to a lesser extent, for image segmentation by ADCI. The subsequent steps are rapid: (a) the optimal image selection model is determined previously as part of calibration curve generation, (b) DC identification is performed as part of image segmentation, (c) determination of radiation exposure follows immediately using a built-in software wizard. Unattended analysis of multiple samples may be performed in ADCI. We have carried out single runs with up to 750 duplicated samples, sourced from multiple exercise samples consisting of 500–800 images each, using the Windows version of the program on computers with 16 Gb RAM. The actual number of images used to estimate exposure is flexible in order to allow for analysis of samples with low mitotic indices. Each 500 image sample required ~5 min to estimate exposures with a computer accelerated with an Nvidia® GTX960M graphics processing unit (GPU). With two GPU-equipped workstations, this level of throughput is equivalent to 576 samples per day. This should be adequate to perform uninterrupted dose estimation of a moderate-sized population of potentially exposed individuals, for example, in the case of a local radiation accident.

## ACKNOWLEDGEMENTS

## REFERENCES

1. Rogan, P. K., Li, Y., Wilkins, R. C., Flegal, F. N. and Knoll, J. H. M. *Radiation dose estimation by automated cytogenetic biodosimetry*. Radiat. Prot. Dosim. **172**, 207–217 (2016).
2. Li, Y., Knoll, J. H. M., Wilkins, R. C., Flegal, F. N. and Rogan, P. K. *Automated discrimination of dicentric and monocentric chromosomes by machine learning-based image processing*. Microsc. Res. Tech **79**, 383–402 (2016).
3. Liu, J., Li, Y., Wilkins, R., Flegal, F., Knoll, J. H. M. and Rogan, P. K. *Accurate cytogenetic biodosimetry through automated dicentric chromosome curation and metaphase cell selection*. F1000Res **6**, 1396 (2017).
4. Rogan, P. K., Li, Y., Wickramasinghe, A., Subasinghe, A., Caminsky, N., Khan, W., Samarabandu, J., Wilkins, R., Flegal, F. and Knoll, J. H. *Automating dicentric chromosome detection from cytogenetic biodosimetry data*. Radiat. Prot. Dosim. **159**, 95–104 (2014).
5. Wang, X., Li, S., Liu, H., Wood, M., Chen, W. R. and Zheng, B. *Automated identification of analyzable metaphase chromosomes depicted on microscopy digital images*. J. Biomed. Inform. **41**, 264–271 (2008).
6. Qiu, Y., Song, J., Lu, X., Zheng, B., Li, S. and Liu, H. *Feature selection for the automated detection of metaphase chromosomes: performance comparison using a receiver operating characteristic method*. Anal. Cell. Pathol (Amst) **2014**, ID 565392, 9 pages (2014).
7. Uttamatanin, R., Yuvapoositanon, P., Intarapanich, A., Kaewkamnerd, S., Phuksaritanon, R., Assawamakin, A. and Tongsima, S. *MetaSel: a metaphase selection tool using a Gaussian-based classification technique*. BMC Bioinformatics **14**(Suppl 16), S13 (2013).
8. Finnon, P., Lloyd, D. C. and Edwards, A. A. *An assessment of the metaphase finding capability of the Cytoscan 110*. Mutat. Res. Gen. Tox. En. **164**, 101–108 (1986).
9. Van den Berg, H. T. C. M., De France, H. F., Habbema, J. D. F. and Raatgever, J. W. *Automated selection of metaphase cells by quality*. Cytometry **1**, 363–372 (1981).
10. Vega-Alvarado, L., Márquez, J. and Corkidi, G. *Inter-chromosome texture as a feature for automatic identification of metaphase spreads*. Med. Biol. Eng. Comput. **40**, 479–484 (2002).
11. International Atomic Energy Agency. *Cytogenetic Dosimetry: Applications in Preparedness for and Response to Radiation Emergencies* (Vienna: IAEA) (2011).
12. Shirley, B., Li, Y., Knoll, J. H. M. and Rogan, P. K. *Expedited radiation biodosimetry by automated dicentric chromosome identification and dose estimation*. J. Vis. Exp. **127**, e56245 (2017).
13. ISCN 2016, International Standing Committee on Human Cytogenetic Nomenclature. McGowan-Jordan, J., Simons, A. and Schmid, M., Eds (Karger) (2016).