RADIATION DOSE ESTIMATION BY AUTOMATED CYTOGENETIC BIODOSIMETRY

Peter K. Rogan[1,2,*], Yanxin Li[1], Ruth Wilkins[3], Farrah N. Flegal[4] and Joan H. M. Knoll[2,5]

[1]Department of Biochemistry, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario N6A 5C1, Canada

[2]Cytognomix Inc., 60 North Centre Rd., Box 27052, London, Ontario N5X 3X5, Canada

[3]Consumer and Clinical Radiation Protection Bureau, Health Canada, 775 Brookfield Rd., K1A 1C1, Ottawa, Ontario, Canada

[4]Canadian Nuclear Laboratories, Radiobiology & Health, Plant Road, Chalk River, Ontario, K0J 1J0, Canada

[5]Department of Pathology and Laboratory Medicine, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario N6A 5C1, Canada

*Corresponding author: progan@uwo.ca

The dose from ionizing radiation exposure can be interpolated from a calibration curve fit to the frequency of dicentric chromosomes (DCs) at multiple doses. As DC counts are manually determined, there is an acute need for accurate, fully automated biodosimetry calibration curve generation and analysis of exposed samples. Software, the Automated Dicentric Chromosome Identifier (ADCI), is presented which detects and discriminates DCs from monocentric chromosomes, computes biodosimetry calibration curves and estimates radiation dose. Images of metaphase cells from samples, exposed at 1.4–3.4 Gy, that had been manually scored by two reference laboratories were reanalyzed with ADCI. This resulted in estimated exposures within 0.4–1.1 Gy of the physical dose. Therefore, ADCI can determine radiation dose with accuracies comparable to standard triage biodosimetry. Calibration curves were generated from metaphase images in ~10 h, and dose estimations required ~0.8 h per 500 image sample. Running multiple instances of ADCI may be an effective response to a mass casualty radiation event.

INTRODUCTION

Biodosimetry is a useful tool for assessing the dose received by an individual when no reliable physical dosimetry is available. Traditionally, the dicentric chromosome (DC) assay is the method of choice for recent acute exposures to ionizing radiation. This cytogenetic method is based on measuring the frequency of DCs in metaphase cells and converting this frequency to dose through the use of *in vitro*-generated calibration curves[1–3]. Classical, microscope analysis of DCs is robust, allowing the estimation of doses in the range of 0.1–5 Gy. For dose estimates in the low end of this range, however, 1000 cells are typically scored making this method time-consuming and only feasible for small numbers of exposures. This manual approach lacks adequate throughput in a mass casualty event to estimate radiation exposures needed to triage individuals for diagnosis and treatment.

In response to the pressing demand to increase throughput in cytogenetic biodosimetry, capture of metaphase images and interpretation of DCs has been partially automated, with a concomitant reduction in the numbers of cells analyzed. Software (e.g. MSearch, DCScore [Metasystems]) has automated the scanning of microscope slides to locate metaphase cells and assisted review of DCs for triage biodosimetry[4]. This software has

also facilitated inter-laboratory collaboration and the assessment of partial-body exposures[5]. The adoption of triage scoring of 50 carefully selected cells greatly increases the throughput, while maintaining the ability to identify exposures of over 1 Gy[6] and reducing the time required by more than a factor of 5[7].

More recently, image analysis software designed to identify DCs (DCScore™) has been used to semi-automate biodosimetry[8–10]. However, it is still necessary to manually preprocess and supervise DC analyses performed with this software. After cells with abnormal chromosome counts and 'metaphases where the two chromatids are sticked or with twisted chromosomes, and metaphases where centromeric constrictions are not visible' are removed, the remaining images are analyzed with DCScore[10]. The operator then manually excludes images with 'twisted chromosomes, two aligned chromosomes, and other figures detected as dicentrics by the software'. False positive (FP) DCs will alter the estimated dose if these steps are not performed[10]. The high rate of FPs in raw data is not surprising in light of the known variation among chromosome morphologies. The detection of DCs, which are much less frequent than monocentric chromosomes (MCs), is also impacted by differences in sample processing procedures among laboratories[11].

The Automatic Dicentric Chromosome Identifier (ADCI) software described here addresses many of the issues of varying morphologies such that images captured on a metaphase finder can be analyzed without the need for manual verification. This unattended automated DC analysis would address a compelling demand for high throughput measurement of biological radiation overexposure (i.e. rapid response to industrial, clinical or terrorism incidents), without compromising accuracy.

MATERIALS AND METHODS

Samples and data used for automated biodosimetry

In this study, all images were obtained from samples prepared at two different biodosimetry reference laboratories (Health Canada [HC] and Canadian Nuclear Laboratories [CNL; formerly AECL or Atomic Energy of Canada Limited]). Two sets of images were supplied from each laboratory. The first set constituted images from X-ray calibration curve samples that had been exposed to doses between 0 and 4 Gy. The second set was blinded samples that had been exposed to various X-ray doses in the same dose range. All samples were irradiated at HC using an XRAD-320 (Precision X-ray, North Branford, CT) and processed at both of the laboratories.

Sample processing protocols were harmonized based on the general guidance provided by the International Atomic Energy Agency (IAEA)[12] and ISO (ISO 19238, 2014). Briefly, each laboratory cultured whole blood diluted with culture medium (RPMI 1640) containing 15% fetal bovine serum, with L-glutamine, penicillin, streptomycin and 15 mM BrdU and stimulated to cycle by the addition of 2% phytohemagglutinin. The cells were incubated at 37°C and 5% CO_2 for 48 h, with 1% colcemid at 10 mg/ml added at 44 h for mitotic arrest. The metaphase spreads were harvested after treatment in a hypotonic solution and fixation in 3:1 methanol: glacial acetic acid Carnoy's fixative. A temperature and humidity-controlled chamber was used to prepare metaphase cells on glass slides which were stained with fluorescence-plus Giemsa for chromosome aberration analysis by brightfield microscopy.

The slides were scanned at low magnification (×10) on an automated cytogenetic imaging platform (MetaSystems Inc., Newton, MA) to identify metaphase cells; images of individual metaphase spreads were then digitally captured at high magnification (×63). The metaphase cells were manually selected and scored for DCs at each of the laboratories; subsequently, images were analyzed with ADCI. The original centromere detection and MC–DC support vector machine (SVMs)[13] were derived using data sets from HC, which differ from data sets in the present study[14].

Overview of ADCI data analysis

Our strategy for automating DC identification involved the development of algorithms capable of recognizing variable chromosome morphologies, i.e. bent or straight appearance, irregular or noisy boundaries, and differences in the degree of chromosome condensation within and between cells that result from asynchronous entry into metaphase. Chromosome image segmentation is also confounded by undetected overlapping chromosomes and sister chromatid separation (SCS). SCS, which is often present in biodosimetry laboratory data, affects the accuracy of computer automated image analysis. Many cells are required (typically ≥500) to accurately estimate radiation exposure; therefore, efficient and validated image analysis methods are required to detect DCs[3].

We previously described earlier versions of the ADCI software based on a set of novel image segmentation algorithms[15–18]. The current version processes Giemsa- (or DAPI-) stained human metaphase cell images and specifically recognizes DC chromosomes in the chromosome-segmented images. The algorithm performs image segmentation to identify chromosomes, localizes candidate centromeres and discriminates DCs from other objects in metaphase cells. The image segmentation module scans and separates foreground objects that include MCs or DCs in a metaphase image, overlapped chromosome clusters, nuclei and debris. Depending on the chromosome morphology, chromosome fragments may be split from complete chromosomes as a result of SCS, and this can adversely affect automated scoring of DCs. The number of segmented objects in an image is a pivotal indicator of the quality of the image. A method is implemented to preferentially avoid metaphase cells containing chromosome segmentation artefacts by thresholding object counts. DCs are detected in thresholded images with two sequential machine learning classifiers. First, a SVM, is used to identify centromere candidates[19], then these results are used to discriminate DCs from other objects[14]. Dose estimation is automated by building dose calibration curves and estimating exposures from DC frequencies in metaphase images of radiation exposed cells.

In many respects, the procedure to detect DCs in images in ADCI is analogous to manual scoring of metaphase cell images (Figure 1). The algorithm accounts for variations in chromosome morphology associated with differences in laboratory protocols. The software determines and reports the numbers of segmented and processed chromosome objects, and the number of positive DC detections for an image.
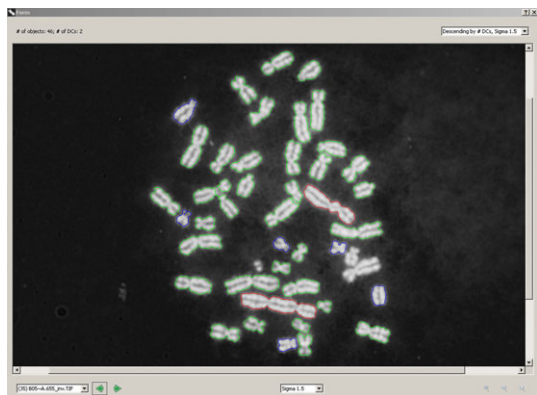
Figure 1. Representative processed metaphase image displayed by the microscopy image viewer in ADCI. The image in the display is B05–A.655 from the CNL-blinded sample INTOC0S05. The result is scored by SVM with $\sigma = 1.5$. DCs, MCs and unprocessed chromosomes have red, green and blue contours, respectively. Objects without highlighted contours are unresolvable overlapping or touching chromosomes or artifactual objects. Captured images are inverted prior to processing by ADCI.

Performance of the combined SVM algorithms is determined from the true positive (TP) detection rates and positive predictive values (PPV) compared to manual or ground truth scoring of DCs by experts.

MCs and DCs are discriminated with a radial basis kernel SVM (termed the MC–DC SVM). The tuning parameter, σ, can be changed to maximize either the sensitivity or specificity of DC detection[14]. The analysis of each metaphase image generates counts of detected objects for 11 different MC–DC SVMs, each derived for a different σ value. Larger values of σ increase TPs and decrease PPV of the SVM. Different values of σ are searched while training the SVM; the σ value is chosen that yields the most accurate classification results in multiple tests with different data sets. Selection of the optimal σ value may vary among different laboratories.

Description of ADCI software elements

ADCI identifies and separates chromosome objects, identifies centromere candidates and discriminates DCs from MCs. The core functionality was originally developed as a software library written in C++[14,18]. In ADCI, the library is integrated with functions to derive calibration curves and estimate sample exposure from the computed curves. The ADCI software application can be navigated with a central graphical user interface (GUI) suitable for analysis by cytogenetic and biosimetry laboratory personnel. The GUI is functionally organized by sectors and panels for entering and processing sample data, and displaying results

(Figure 2). The processing function attempts to remove non-chromosomal objects (e.g. nuclei, fragments and debris), thresholds and separates objects, and identifies DCs. The left sector contains the main work space for users to instruct the software to select samples and generate curves, and the right sector provides plot and text console panels to visualize results. An object count filter selects higher quality images with near normal chromosome complements for determination of DC frequencies. The display in the right panel shows the distribution of the number of segmented objects in a sample; however, it may also show calibration curve(s) and the estimated dose computed for different test sample(s) superimposed on the curve. Users create and process a sample by providing the path to the folder with the metaphase images for each sample and by selecting icons corresponding to the relevant function. The software scores DCs and records the contour for each object in a processed sample for each of the DC/MC SVMs. Multiple samples may be processed in the same run. Laboratory source information for new samples is specified at processing. Unprocessed samples can be added to the process queue in the work space. Users start this queue to process a selected sample or all samples in the queue. A progress bar indicates which samples have been processed, those yet to be processed, and the percent completion of the current sample. Users can abort a multi-sample processing procedure, while preserving results from samples that have completed analysis. Once the processing procedure completes, those samples are removed from the queue. Users can save a processed sample with all annotated metaphase images; conversely, a saved sample can be restored for further analysis.

After sample processing, the software outlines each processed chromosome in a metaphase cell in color. Red and green contours highlight DCs and MCs, respectively, while those with blue contours are unclassifiable, as a result of incomplete Intensity Integrated Laplacian image segmentation[17] (Figure 1). Objects without outlines represent overlapped chromosomes, nuclei and debris, which are not included in counts of objects in a cell. Individual cells may be sorted, automatically or manually filtered out with a built-in microscope image viewer.

The number of chromosome objects per cell can be specified as a range, eliminating images with either too many (comprising multiple cells, extreme levels of SCS or non-chromosomal debris) or too few objects (incomplete metaphase cells or highly overlapped sets of chromosomes). DC counts and frequencies are calculated and reported by the GUI. The DC frequency calculation is performed only on cells within the thresholded range. The overall distribution of processed object counts among all cells can be viewed as a histogram plot with the range of object thresholds highlighted. The expected Poisson distribution for an average DCs/cell at a given dose can also be plotted
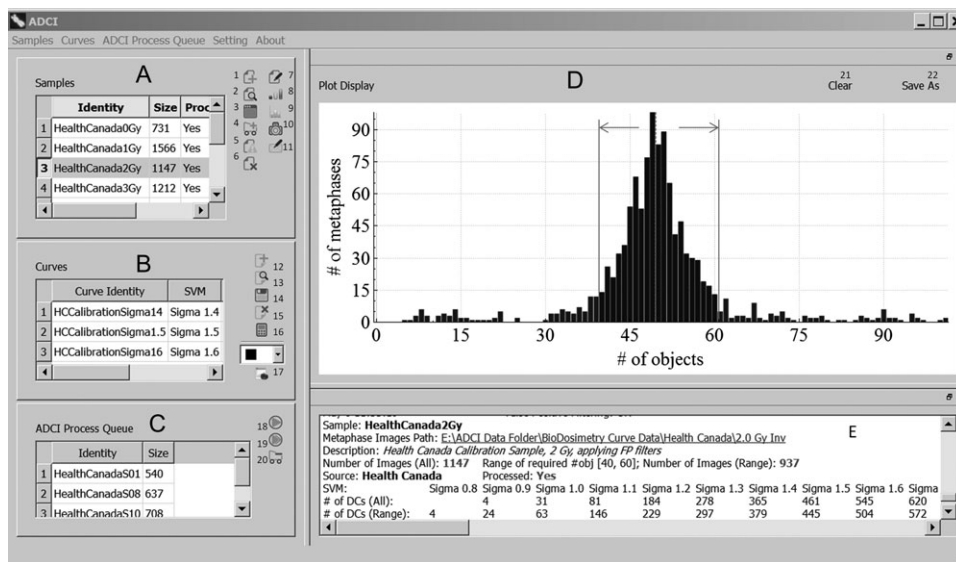
Figure 2. Main window of the GUI software with icons numbered for description purposes. Panel **A** is a list of sample operations. In Panel A, selecting icons performs: (1) add a new sample, (2) open a sample from file, (3) save a processed sample to file, (4) add an unprocessed sample to process queue, (5) discard existing result for a sample, (6) remove a sample from work space, (7) edit a sample, (8) show distribution of segmented objects in a sample, (9) show distribution of positive objects in a sample, (10) view ADCI result in metaphase images in a sample, (11) export ADCI result of a sample to a comma separated values file. Panel **B** is a list of curve operations. In Panel B, selecting icons performs: (12) create a new curve, (13) open a curve from file, (14) save a curve to file, (15) remove a curve from work space, (16) make dose estimation using a curve, (17) color selection for a curve in the plot panel. Panel **C** is the ADCI process queue. In Panel C, selecting icons performs: (18) process the selected sample, (19) process all samples in the process queue, (20) remove a sample from the process queue. Panel **D** is a plot display panel in the software (displaying distribution of segmented objects in a sample). In Panel D, selecting icons performs: (21) clear the plot, (22) save current plot to a graphical file (png, jpeg, etc). Panel **E** is the console display panel (displaying text information about a sample below the plot). The current plot display panel demonstrates the distribution of segmented objects in the HC 2-Gy calibration sample; however, it can also display the outputs shown in Figures 3–5. Here, the *X*- and *Y*-axes indicate the number of segmented objects in an image and the number of images in the sample, respectively. The distribution of segmented objects that fall within the user-specified limits (40–60 is shown; the interval is defined prior to metaphase cell image processing) are indicated by the bracketed interval of the histogram. This and the remaining figures are presented in gray scale; however, the software itself produces color graphics.

as a dotted line overlaying a bar graph of the actual distribution cells with different DC counts.

Generation of calibration curves

Despite the use of a common set of laboratory protocols, the dose calibration curves produced from each laboratory are unique and may produce different dose estimates[3]. To determine DC counts and ensure internally consistent dose estimates, data from the same laboratory for test samples and calibration curves are both processed using ADCI. To prepare dose–response curves, peripheral blood from a single individual was exposed to a 250-kVp radiation source, at 0.1, 0.25, 0.5, 0.75, 1, 2, 3, 4 and 5 Gy, and metaphase cells were prepared according to standard cytogenetic laboratory protocols. Along with unexposed control samples (0 Gy), the exposed samples were manually scored by HC and CNL and

each laboratory generated their own calibration curve. Metaphase images were provided from a subset of these samples (0, 0.5, 1, 2, 3 and 4 Gy) for analysis with ADCI. At HC, cells were selected by the Metafer slide scanning system (Metasystems), which removes some, but not all unsuitable cell images. CNL manually curated images of metaphase cells based on the spatial distribution, morphology and numbers of chromosomes detected. Calibration curves were derived both manually and with ADCI (both are shown in Figure 3). ADCI curves are generated for each MC–DC SVM (defined for σ values ranging from 0.8 to 1.8) for each set of dose-calibrated samples. These curves can be directly compared with the manual calibration curves by plotting the linear–quadratic formulae based on the respective polynomial coefficients.

Calibration curves can also be entered with physical dose paired with responses (i.e. DCs/cell) from
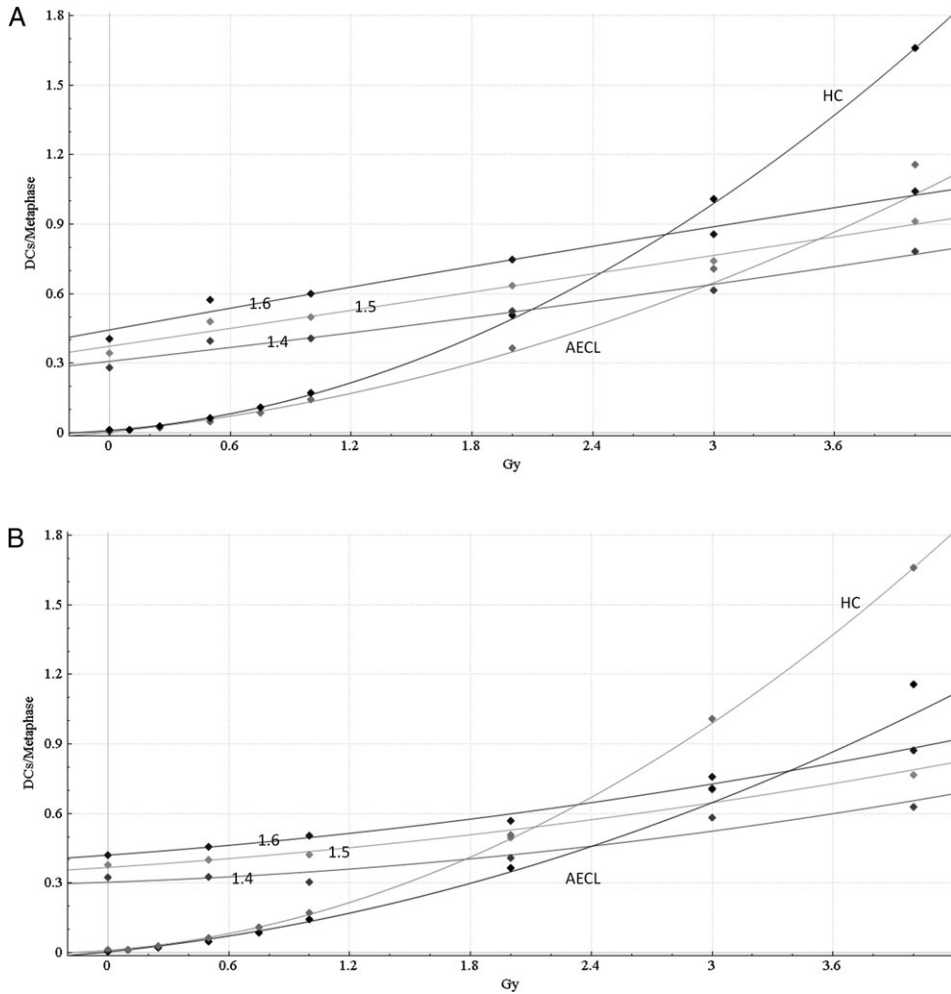
Figure 3. ADCI calibration curves for HC and CNL, in comparison with HC and CNL (labeled as AECL) manual cali-bration curves. The manual calibration curves contain additional values at 0.1, 0.25 and 0.75 Gy that were not analyzed with ADCI. *X*- and *Y*-axes indicate dose (Gy) and response (DCs/metaphase), respectively. Curves are plotted with the ori-ginal calibration dose–response points indicated as diamonds. Panels **A** and **B** show ADCI and manual calibration curves for HC and for CNL data, respectively, annotated with the corresponding σ values ($\sigma = 1.4, 1.5, 1.6$). In Panel A, the man-ual HC curve is denoted in black and corresponding CNL curve is shown in gray. In Panel B, the CNL manual curve is indicated in black and HC curve is shown in gray.

either manually scored DCs or ADCI-detected DCs. A linear–quadratic curve can also be created from the corresponding polynomial coefficients. Curve data can either be saved to a file or loaded from a saved file. The graphical interface displays the poly-nomial regression curve fit to the DC frequencies at different doses and the curve equation.

Blinded dose estimation of test samples

To estimate the dose–response (in Gy) for a sample of unknown exposure, the observed DC frequency in a processed sample is evaluated with the linear polynomial curve equation (Figure 4). The exposure can also be estimated for a test sample in ADCI by entering the DC count, i.e. the response, for a selected curve. The software interpolates the best estimate of the dose corresponding to the DC frequency by searching the applicable range of the curve in incre-ments of 0.05 Gy.

For R DCs in a given sample, the lower bound R_L and upper bound R_U is obtained according to Equation (1), with $\alpha = 0.05$. The exact confidence interval (CI) for a Poisson distribution of events R is
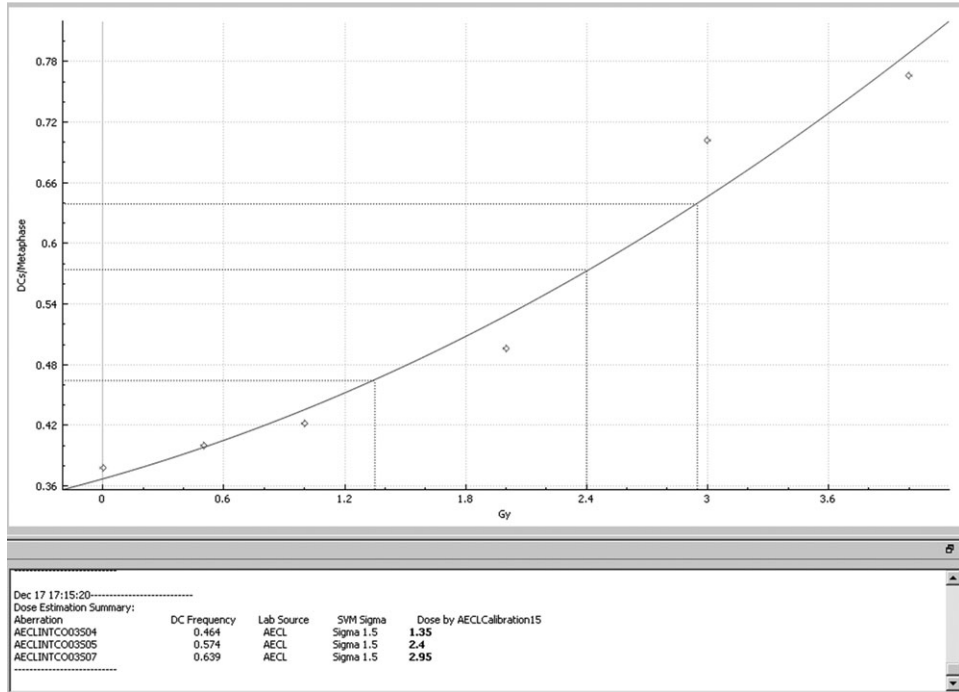
Figure 4. Dose estimation for blinded samples using an ADCI calibration curve. Doses of CNL samples AECLINTCO0S04, AECLINTCO0S05 and AECLINTCO0S07 (corresponding to INTC03S04, INTC03S05 and INTC03S07 in Table 3, respectively) are estimated using the ADCI calibration curve for CNL at $\sigma = 1.5$ (AECLCalibration15). The console panel describes the results in the table with estimated doses in **bold**. The plot panel shows the curve in gray, with projections of dose estimations as dotted lines. Confidence intervals for these estimates are indicated in Table 3.

$$(\chi^2_{2R,\alpha/2}, \chi^2_{2(R+1),1-(\alpha/2)}). \tag{1}$$

This method uses the chi-square distribution to calculate the CI for Poisson distribution (Equation 1). Using these two values with a selected ADCI calibration curve in the same way estimated dose D is calculated for R, we obtain the corresponding lower bound D_L and upper bound D_U for dose estimation. This CI measures the uncertainty in the dose caused by the uncertainty of chromosome aberrations in a sample. It does not reflect the uncertainty of the calibration curve, but it is acceptable for well-established curves[3].

For ADCI-generated curves, the SVM with the σ value that results in the maximum combined TP rate and PPV is selected for testing of samples of unknown exposure. After the sample is processed to identify DCs, the DC frequencies for each of these SVMs are determined. When chromosome aberration frequencies are compared between samples in ADCI, the σ parameter for each sample should be the same and consistent with the σ value used to derive the corresponding calibration curve. The software does not

extrapolate doses for exposures beyond the range of the calibration data; in those cases, it indicates the dose to be below the minimum dose or above the maximum dose of the applicable range.

RESULTS

The automated biodosimetry method estimates the radiation dose in samples constituting a set of metaphase cell images from an individual exposed to low-linear energy transfer (LET) radiation. Instead of manually counting DCs in metaphase images to determine the frequency of this aberration, the automated method detects and accrues DCs in a sample. A sufficient number of images (typically ≥500) are required to accurately estimate the dose received.

Sample quality assessment and DC analysis

The ADCI algorithm has optimal performance for detection of DCs in metaphase images with a near complete complement of well-separated, linear chromosomes. However, the underlying methods can accommodate a range of chromosome structures,

6

counts and spacing. Nevertheless, variation in mitotic index, cell culture protocols, slide preparation and chromosome length affect DC detection due to irregular chromosome morphology, staining or distribution, such as overlapping or incomplete sets of chromosomes and extreme SCS. The computed DC frequency can be skewed by incomplete metaphase cells, multiple cells in an image, or excessive cellular debris or non-metaphase objects, which can affect the estimated dose (see discussion of INTC03S10 below). An image with a paucity of objects (<30) is more likely to contain an incomplete set of chromosomes or multiple overlapped chromosomes. Extreme SCS can split chromosomes, falsely elevating the number of 'chromosome objects' detected. Experts can compensate for these complications by discarding these images with a counting filter that defines the number of allowable objects in an image at the time that image processing is initiated.

The ADCI software preprocesses images to assess metaphase cell quality and indirectly, chromosome morphology. ADCI retains images based on the number of segmented objects (Figure 2). A range of 40 and 60 objects per image has been defined empirically to limit overlapped chromosomes and incomplete metaphase cells. ADCI then processes the objects with detectable centromeres. The Integrated Intensity Laplacian module requires that each object has a contour that can be evenly divided in order for it to be processed[17]. Unprocessed objects are infrequent and tend to exhibit irregular shapes, such as overlapped or short chromosomes or fragments, or cellular debris. Only metaphase cells in which ≥20% of all objects can be processed are retained for DC identification and dose estimation.

The software was validated with image data from radiation exposed samples of known and masked radiation doses. For each laboratory, we processed the calibration samples and test samples separately in ADCI, built ADCI calibration curves, compared these curves with manual calibration curves, and estimated radiation doses for the blinded test samples using the calibration curves. During sample processing, the software calls the ADCI library to process the selected sample or all samples in the queue, in parallel with multiple CPU threads. For a sample containing 500–1000 images, the process requires 0.8–1.6 h with a laptop computer (equipped with an Intel I7-M620 2.67 GHz CPU and 8 Gb RAM).

Generation of calibration curves

The results of processing calibration samples are indicated in Table 1. ADCI detects a monotonic, linear increase of DCs/cell with dose in both HC and CNL data sets. However, the distribution of DCs detected by ADCI shows that the expected Poisson distribution is somewhat overdispersed in the HC data set above 1 Gy ($\sigma = 1.4$), and in the CNL data set at 0.4 Gy ($\sigma = 1.4$) and 2 Gy ($\sigma = 1.5$).

A set of linear–quadratic calibration curves for σ values from 0.8 to 1.8, in increments of 0.1, were constructed separately for data from HC and CNL. The coefficients for these functions are summarized in Table 2. We found that SVMs based on $1.4 \leq \sigma \leq 1.6$ maximize the combined TP rate and PPV. The algorithm is accurate for 2–4 Gy high-dose data and currently has satisfactory, but less accurate, performance at lower dose (1 Gy).

The ADCI-derived calibration curves at $\sigma = 1.4$, 1.5 and 1.6 for the HC data are shown in Figure 3A and for CNL in Figure 3B, respectively, as well as the calibration curves derived of the same images by manual scoring by these laboratories.

Table 1. Detection of DCs by ADCI at different doses in images from HC and CNL reference laboratories.

Laboratory/dose	No. cell images	$\sigma = 1.4$		$\sigma = 1.5$	
		DC/cell	*u*-test	DC/cell	*u*-test
HC/0 Gy	731	0.282	1.54	0.345	1.96
HC/0.5 Gy	586	0.397	1.00	0.48	0.58
HC/1 Gy	1566	0.408	3.89	0.5	4.87
HC/2 Gy	1147	0.538	8.12	0.651	7.24
HC/3 Gy	1212	0.618	2.59	0.746	2.60
HC/4 Gy	909	0.785	2.84	0.915	2.99
CNL/0 Gy	500	0.324	0.38	0.378	0.92
CNL/0.5 Gy	500	0.327	2.44	0.401	2.24
CNL/1 Gy	500	0.304	1.26	0.422	1.76
CNL/2 Gy	500	0.408	1.49	0.496	3.29
CNL/3 Gy	500	0.583	1.9	0.703	2.6
CNL/4 Gy	500	0.628	1.28	0.766	1.3

Dose estimation of test samples

HC and CNL each provided three test samples at different, blinded radiation doses from an international biodosimetry exercise. Physical radiation exposures were not revealed until doses were first estimated by ADCI and by manual cytogenetic review. Images from samples INTC03S01, INTC03S08 and INTC03S10 were obtained from HC, and INTC03S04, INTC03S05, INTC03S07 came from CNL. HC used the Metafer system itself to perform preliminary image selection, while CNL manually selected metaphase images from the blinded samples. Metafer does not eliminate all images without metaphase chromosomes nor does it remove densely overlapped or incomplete metaphase cells. Blinded samples were processed by the ADCI software and

had doses estimated according to ADCI calibration curves. Cytogenetic specialists in the corresponding labs also scored these samples. Five experts at HC each scored 50 metaphase images for each sample. At CNL, four experts each scored 50 metaphase images for each sample. The frequency of DCs was calculated for each sample from which the radiation doses were then estimated based on the manual calibration curves in each laboratory.

The results and comparisons of traditional manual biodosimetry and automated biodosimetry by ADCI are shown in Table 3. With respect to physical doses, differences between expert-scored samples ranged from 0 to 0.3 Gy, whereas ADCI-scored samples differed by 0.3–1.1 Gy. Exposure was underestimated in sample INTC03S10, which exhibited the largest

Table 2. Coefficients for ADCI-generated, linear–quadratic calibration curves.

Curve	Coefficients			Statistics	
Lab, SVM parameter	Intercept	Linear	Quadratic	R^2	SSE
HC, $\sigma = 1.4$	0.308	0.097	0.004	0.981	0.003
HC, $\sigma = 1.5$	0.373	0.128	0.001	0.984	0.003
HC, $\sigma = 1.6$	0.444	0.157	−0.003	0.979	0.0055
CNL, $\sigma = 1.4$	0.303	0.029	0.014	0.934	0.0067
CNL, $\sigma = 1.5$	0.367	0.056	0.012	0.964	0.005
CNL, $\sigma = 1.6$	0.420	0.062	0.013	0.987	0.002

σ, kernel parameter for radial basis function in SVM; R^2, square of the correlation between the response values and the predicted response values; SSE, total deviation of the response values from the fit to the response values, i.e. sum of squares due to error.

Table 3. Estimation of radiation exposure of blinded test samples by ADCI.

Laboratory	Sample identifier		
HC	INTC03S01	INTC03S08	INTC03S10
No. of metaphases processed[a]	540	637	708
Physical dose, Gy	3.1	2.3	1.4
Expert inference, average Gy	3.4	2.5	1.4
ADCI estimation, Gy, $\sigma = 1.5$	3.9	1.65	0.3
95% Confidence interval [D_L, D_R], $\sigma = 1.5$	[3.3, 4]	[1.2, 2.1]	[0, 0.65]
ADCI u-test, $\sigma = 1.5$	9.45	6.46	4.23
CNL	INTC03S04	INTC03S05	INTC03S07
No. of metaphases processed[a]	448	500	385
Physical dose, Gy	1.8	2.8	3.4
Expert inference, average Gy	1.7	2.7	3.1
ADCI estimation, Gy, $\sigma = 1.5$	1.35	2.4	2.95
95% Confidence interval [D_L, D_R], $\sigma = 1.5$	[0.55, 2.0]	[1.8, 2.95]	[2.3, 3.55]
ADCI u-test, $\sigma = 1.5$	2.74	1.21	2.8

[a]Metaphase cells analyzed after filtering using thresholded number of objects in image; see Materials and Methods.

deviation from the actual dose, due to Metafer-generated images lacking metaphase cells. ADCI scoring was more accurate for CNL samples, in which the images were manually preselected.

DISCUSSION

The ADCI algorithm and software implement a fully automated approach to radiation biodosimetry using inferred DC frequencies as biological endpoints. Manual involvement is limited to data input and operational decision-making for DC detection; derivation of calibration curves and dose estimations are performed by the software. This relieves the workload of biodosimetry interpretation in cytogenetic biodosimetry personnel due to its relatively fast processing speed and by eliminating required interaction with the user during processing. Parallel versions of some ADCI modules have been successfully deployed on various distributed computing systems, which further accelerate the ADCI procedure[18]. The dose estimation performed by the software is fairly accurate, with an average difference of 0.65 Gy between the estimated dose and the physical dose in the tested samples.

The quality of the metaphase images in a sample significantly affects accuracy of the automated DC analysis and consequently DC frequency measurement and dose estimation for the sample in ADCI. Notable differences in image quality are evident between samples from HC and CNL data. CNL samples exhibit higher quality cell images with well-spread, and fewer overlapped chromosomes, as well as decreased SCS relative to the HC samples analyzed. However, CNL performed manual image selection that was more stringent than the automatic cell selection performed by HC with the Metafer system. The *u*-test values of HC calibration samples and blinded samples are larger than those of CNL samples. Over-dispersion is primarily attributed to FP DC detection in the ADCI algorithm. FP DC detection is more common in the HC samples than in CNL samples. Furthermore, the distribution of the number of segmented objects also differs between the blinded HC and CNL samples. A manual review of metaphases in samples from both HC and CNL confirmed these quality differences, and explains the differences in accuracy of dose estimation in the blinded samples from HC.

Calibration curves in ADCI

A minimum of three dose-calibrated samples at different exposures are required to fit a dose estimation curve from DC frequencies. The curve fit to these data is defined by a linear–quadratic formula defined by a set of coefficients by a maximum likelihood method[3]. The maximum dose and the minimum dose in the sample set define the applicable range for estimating dose of samples with undetermined test sample exposures. In practice, this range spans 0 to 4–5 Gy, which is approximately the maximum tolerable whole body dose.

For whole body irradiation, these chromosome aberrations occur randomly and follow a Poisson distribution. The distribution of chromosome aberrations in cells is related to the type of radiation[3]. In ADCI, a linear–quadratic curve is plotted from these data. The curve fits a linear–quadratic function in low-LET radiation scenarios, where two separate radiation events are thought to be required to create two double-stranded breaks needed for DC formation. For high-LET radiation, a linear relationship is expected. The probability of a DC being detected depends on stringency of the SVM and effectiveness of the feature combination that defines the SVM. While the distribution of TPs is expected to fit a Poisson at low LET, FPs are not independent from each other in ADCI. Chromosomes with SCS, overlapped and fragmented chromosomes are major sources of FPs. Cells with more FPs, which are more prevalent in lower quality metaphase cells, bias the calibration curve toward the linear component.

The Poisson fit is one measure that can be used to evaluate the overall performance of ADCI. An excess of FPs will produce Poisson distributions that underrepresent cells without any DCs or a single DC. ADCI computes the Poisson distribution for the average number of positive DCs per cell (the λ parameter), which is compared with the observed distribution of cells with different DC frequencies detected (Figure 5). The dispersion index, or u value[20] can be used to assess conformity of the ADCI-derived DC distribution to the Poisson. A value of $\mu > 1.96$ indicates over-dispersion due to non-uniform exposure, while values ≤ 1.96 represent under-dispersion. For the minimally curated HC test samples, the combination of unrecognized FP DCs and images lacking metaphase cells distort this statistic. By contrast, the curated images in the CNL test samples exhibited μ values within or close to these thresholds. Automated preprocessing and filtering of images and FP DCs will limit these sources of error, and should facilitate analyses partial-body exposures using ADCI.

The linear component of the fitted curve dominates the ADCI calibration curves relative to the corresponding manual curves. DC responses of ADCI curves increase at a slower rate due to the smaller quadratic coefficients, especially at high dose. The derivatives of ADCI curves are smaller than the ones of the manual curve for a certain dose, which is reflected as more constant rates of change in responses for ADCI curves. As seen in Figure 3, ADCI curves demonstrate stronger responses at low dose (1–2 Gy) and weaker responses at high dose (~4 Gy) than the manual curves. In the 3–4 Gy
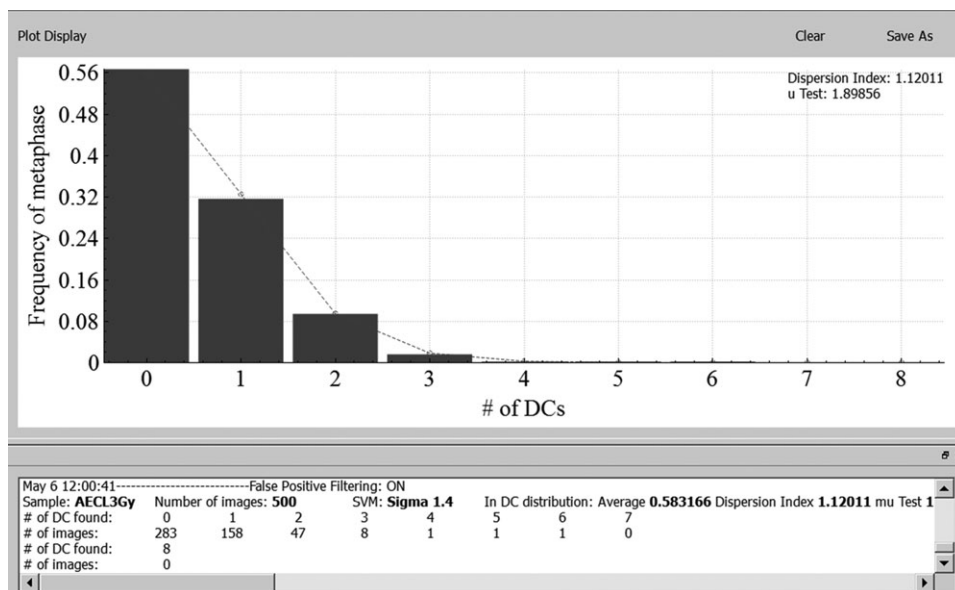
9

Figure 5. Poisson distribution of DCs and statistical analysis of a sample in ADCI. Distribution of positive detections in the CNL 3 Gy calibration sample. In the plot panel (top), the *X*-axis is the number of positive detections in a metaphase image, and the *Y*-axis represents the frequency of each metaphase image type in the sample. The bar graph shows the actual positive detection distribution and the dotted lines represent the corresponding Poisson distribution. The distribution of DCs is also summarized in the console display panel.

range, ADCI and manual curves are most similar. The detection of FPs by ADCI at low doses is primarily related to the adventitious association of short arms of acrocentric chromosomes and SCS chromosomes that are not eliminated by other image filters. SVMs created at low σ values are particularly susceptible to these artefacts. However, the frequency of these FPs appears to be related to image quality, and is invariant to increased radiation level. FPs have a proportionately larger impact at low doses, causing a stronger response in ADCI curves. At higher levels of radiation exposure, FPs tend to be balanced out by an increase in FNs, i.e. DCs missed by the SVM. Beyond the level at which the number of FNs exceed the number of FPs, the ADCI calibration curves exhibit a weaker response than manually derived curves. The differences between ADCI-derived calibration curves at different σ values largely affect the intercept (DCs/cell) rather than the slope. The intercept increases with higher values of σ. In fact, the manual curves prepared by HC and CNL also differ in the magnitude and growth rate of response. Both curves fit to linear–quadratic functions with high precision and have similar responses at low dose. However, the HC manual curve has a stronger response than the corresponding CNL curve at high dose.

While differences exist between ADCI calibration curves and manual curves and the ADCI algorithm is not free of errors in DC detection, it is still possible to infer radiation exposure doses from test samples with reasonable accuracy. However, accuracy is highly dependent on the quality of images in the samples. Manually selected images more accurately estimate physical radiation exposures, despite the introduction of automated preprocessing and filtering procedures in ADCI to eliminate segmentation and SVM classification errors.

DCScore™ software requires manual preprocessing and postprocessing review of DCs[4,9], especially at low radiation doses. By contrast, ADCI that substantially limits the need to manually review results. Images can, nevertheless, be manually curated in ADCI to eliminate those without metaphase cells or containing FP DCs. It is notable that dose estimates from the curated CNL data produced results of comparable accuracy to manual DC scoring without any image post-processing. These assay results may be of particular value for discriminating high from low dose exposures.

Initially, we anticipate that the ADCI software will be an adjunct to conventional manual scoring of DCs. Once laboratories gain experience with the ADCI software, it could become a routine tool for radiation biodosimetry. Practical applications of

ADCI will include i) triaging samples in large scale events freeing up biodosimetry personnel to accept and process more samples, ii) screening of populations exposed to radiation to study variation in radiation response between individuals and, iii) to create composite calibration curves based on multiple individuals or radiation sources. Furthermore, it may be useful for the identification of radiation exposed individuals with occupational, military or clinical overexposures, and may be relevant for assessing long term effects of radiation therapy in cancer patients.

REFERENCES

1. Bauchinger, M. *Cytogenetic Effects in Human Lymphocytes as a Dosimetry System*. In: Biological Dosimetry: Cytometric Approaches to Mammalian Systems. Eisert, W. S. and Mendelsohn, M. L. Berlin: Springer-Verlag), 15–24 (1984).
2. Lloyd, D. C., Edwards, A. A. and Prosser, J. S. *Chromosome aberrations induced in human lymphocytes by in vitro acute X and Gamma radiation*. Rad. Prot. Biodosim. **15**, 83–88 (1986).
3. International Atomic Energy Agency (IAEA). Cytogenetic analysis for radiation dose assessment. Technical Report Series No. 405, Vienna (2001).
4. Schunck, C., Johannes, T., Varga, D., Lorch, T. and Plesch, A. *New developments in automated cytogenetic imaging: unattended scoring of dicentric chromosomes, micronuclei, single cell gel electrophoresis, and fluorescence signals*. Cytogenet. Genome Res. **104**, 383–389 (2004).
5. Ainsbury, E. A., Livingston, G. K., Abbott, M. G., Moquet, J. E., Hone, P. A., Jenkins, M. S., Christensen, D. M., Lloyd, D. C. and Rothkamm, K. *Interlaboratory variation in scoring dicentric chromosomes in a case of partial-body x-ray exposure: implications for biodosimetry networking and cytogenetic 'triage mode' scoring*. Radiat. Res. **172**, 746–752 (2009).
6. Lloyd, D. C., Edwards, A. A., Moquet, J. E. and Guerrero-Carbajal, Y. C. *The role of cytogenetics in early triage of radiation casualties*. Appl. Radiat. Isot. **52**, 1107–1112 (2000).
7. Flegal, F. N., Devantier, Y., McNamee, J. P. and Wilkins, R. C. *Quickscan dicentric chromosome analysis for radiation biodosimetry*. Health Phys. **98**, 276–281 (2010).
8. Vaurijoux, A., Gruel, G., Pouzoulet, F., Grégoire, E., Martin, C., Roch-Lefèvre, S., Voisin, P., Voisin, P. and Roy, L. *Strategy for population triage based on dicentric analysis*. Radiat. Res. **171**, 541–548 (2009).
9. Vaurijoux, A., Gruel, G., Gregoire, E., Roch-Lefevre, S., Voisin, P., Martin, C., Voisin, P., Roy, L. and Barquinero, J.-F. *Automatic dicentric scoring a real option to be used in biological dosimetry*. Radiat. Emerg. Med. **4**, 16–21 (2015).
10. Gruel, G., Grégoire, E., Lecas, S., Martin, C., Roch-Lefevre, S., Vaurijoux, A., Voisin, P., Voisin, P. and Barquinero, J.-F. *Biological dosimetry by automated dicentric scoring in a simulated emergency*. Radiat. Res. **179**, 557–569 (2013).
11. Wilkins, R. C., Romm, H., Kao, T. C., Awa, A. A., Yoshida, M. A., Livingston, G. K., Jenkins, M. S., Oestreicher, U., Pellmar, T. C. and Prasanna, P. G. *Interlaboratory comparison of the dicentric chromosome assay for radiation biodosimetry in mass casualty events*. Radiat. Res. **169**, 551–560 (2008).
12. International Atomic Energy Agency. Cytogenetic dosimetry: applications in preparedness for and response to radiation emergencies. IAEA-EPR, Vienna (2011).
13. Hsu, C.-W., Chang, C.-C. and Lin, C.-J. "A practical guide to support vector classification." pp. 1–16 (2003). https://www.cs.sfu.ca/people/Faculty/teaching/726/spring11/svmguide.pdf
14. Li, Y., Knoll, J. H. M., Wilkins, R., Flegal, F. N. and Rogan, P. K. *Automated discrimination of dicentric and monocentric chromosomes by machine learning-based image processing*. Microsc. Res. Tech. **79**, 393–402 (2016).
15. Subasinghe, A., Samarabandu, J., Knoll, J., Khan, W. and Rogan, P. K. An image processing algorithm for accurate extraction of the centreline from human metaphase chromosomes. 2010 IEEE International Conference on Image Processing, pp. 3613–3616 (2010). DOI: 10.1109/ICIP.2010.5652017
16. Li, Y., Wickramasinghe, A., Subasinghe, A., Samarabandu, J., Knoll, J., Wilkins, R., Flegal, F. and Rogan, P. K. Towards large scale automated interpretation of cytogenetic biodosimetry data. paper #1569626685. IEEE 6th Annual International Conference on Automation for Sustainability. pp. 30–35 (2012). DOI: 10.1109/ICIAFS.2012.6420039
17. Subasinghe, A. A., Samarabandu, J., Knoll, J. and Rogan, P. K. *Intensity integrated laplacian based thickness measurement for detecting human metaphase chromosome centromere location*. IEEE Trans. Biomed. Eng. **60**, 2005–2013 (2013).
18. Rogan, P. K., Li, Y., Wickramasinghe, A., Subasinghe, A., Caminsky, N., Khan, W., Samarabandu, J., Wilkins, R., Flegal, F. and Knoll, J. K. *Automating dicentric chromosome detection from cytogenetic biodosimetry data*. Rad. Prot. Biodosim. **159**, 95–104 (2014).
19. Subasinghe, A., Samarabandu, J., Li, Y., Flegal, F., Wilkins, R., Knoll, J. H. M. and Rogan, P. K. Centromere detection of human metaphase chromosome images using a candidate based method, F1000Research, http://dx.doi.org/10.12688/f1000research.9075.1 [bioRxiv: http://dx.doi.org/10.1101/032110] (2016).
20. Papworth, D. *Curve fitting by maximum likelihood. Appendix in savage JH. Radiation induced chromosome aberrations in the plant: Tradescantia: dose response curves*. Radiation Botany **15**, 87–140 (1975).