Automated discrimination of dicentric and monocentric chromosomes by machine learning-based image processing

Yanxin Li¹, Joan H. Knoll²,³, Ruth C. Wilkins⁴, Farrah Flegal⁵, and Peter K. Rogan¹,³

¹ Schulich School of Medicine and Dentistry, Biochemistry Dept, Ontario, Canada
² Schulich School of Medicine and Dentistry, Pathology Dept, Ontario, Canada
³ Cytognomix, Inc., London, Ontario, Canada
⁴ Health Canada, Ottawa, Ontario, Canada
⁵ Canadian Nuclear Laboratories, Ontario, Canada

Study Goal: Exposure to radiation can be determined from dicentric chromosome (DC) frequencies in metaphase cells. We have fully automated DC detection and discrimination with image processing and Machine Learning (ML) algorithms that extract features and classify objects in Giemsa-stained microscopy images. The algorithms are implemented as desktop software. The goal is to achieve high detection accuracy across a range of radiation levels, while minimizing missed DCs and inclusion of false positive objects.

Abstract: The detection of DCs involves several steps, including: 1) image segmentation of objects in metaphase cells by intensity thresholding, 2) chromosome separation by watershed transformation, 3) filtering out inseparable chromosome clusters, fragments and staining debris with a morphological decision tree, 4) chromosome width determination by intensity-integrated Laplacian measurement, 5) ranking of centromere candidates by ML, and 6) distinction of DCs from monocentric chromosomes (MCs) by ML. Chromosome intervals with constrictions may either be correct or false positive centromeres due to twisting of chromosomes/chromatids or other cellular or imaging artifacts. A Support Vector Machine (SVM) generates the optimal centromere candidates for each chromosome from the minimum distances to a multidimensional hyperplane that is inferred from 14 image features. For DCs, these are true centromeres and for MCs, one is a true centromere and the other is false. Features (n=16) from selected candidates and local image regions are sequentially input to a Boosting classifier and a second SVM, which makes the final identification of DC or MC. The SVMs were trained with 292 DCs and 3135 MCs, and then tested with independent sets of cells exposed to low (1 Gy) and high (3-4 Gy) radiation levels. One expert scored training and test data as ground truth; however discrepancies of 10% were noted between all three experts in scoring true DCs in test data. Performance of the software was determined from the true positive rate (TPR) and positive predictive value (PPV) for different values of the SVM tuning parameter, sigma. At larger sigma, PPV decreases and TPR increases. At high dose (1.15 - 1.50 DCs per cell [DpC], 360 cells), for sigma = 1.3, the TPR = 0.52 and PPV = 0.83, while at sigma = 1.6, the TPR = 0.65 and PPV = 0.72 (14,428 chromosomes analyzed). At low dose (<0.2 DpC, 208 cells), a sigma value of 1.3 shows a TPR = 0.67 and PPV = 0.26 (8,041 chromosomes analyzed).

Conclusion: Our Automated Dicentric Chromosome Identifier software differentiates DCs from MCs, overlapped chromosomes and other objects in metaphase cells with minimal assistance. Images from automated microscopy systems of various qualities may be analyzed. The desired accuracy and error rate can be controlled through ML tuning parameters. Compared to expert scoring, the system is accurate at 3-4 Gy radiation doses, and exhibits high sensitivity and moderate specificity at low doses (1 Gy).
Acknowledgements: This work was supported by the University of Western Ontario, Canada Research Chairs Secretariat, and the Canadian Foundation for Innovation.