PROGNOSIS OF STAGE I LUNG CANCER PATIENTS THROUGH QUANTITATIVE ANALYSIS OF CENTROSOMAL FEATURES

Dansheng Song[1], Tatyana A. Zhukov[1], Olga Markov[2], Wei Qian[3], Melvyn S. Tockman[1]
Dansheng.Song@moffitt.org, Tatyana.Zhukov@gmail.com, olgareunova@gmail.com, wqian@utep.edu, Melvyn.Tockman@Moffitt.org

[1] Molecular Epidemiology Program, Department of Cancer Epidemiology and Genetics, H. Lee Moffitt Cancer Center & Research Institute
[2] Life Sciences Advanced Technologies Inc.
[3] College of Engineering, University of Texas at El Paso

ABSTRACT

Centrosome amplification leads to the loss of regulated chromosome segregation, aneuploidy, and chromosome instability and has the possibility to be a biomarker of cancer prognosis.

To explore this feasibility, resected, stage I non-small cell lung cancer (NSCLC) tissues from six survivor and six fatal cases were immunostained and scanned. Regions of interest were selected to include one cell and its centrosomes. After segmentation, feature abstraction, and optimization, six non-redundant features were used for statistical analysis and classification.

Two analytic methods showed that for each feature, centrosomes from survivors differed from centrosomes of fatalities, indicating sampling from different populations. The data were classified using linear discriminant analysis (LDA) and support vector machines (SVM) with 10-fold cross-validation. Classification accuracy was 74% by LDA and 79% by SVM, respectively, and further improved to 85% with bagging.

Centrosome can be a biomarker for stage I NSCLC prognosis and has potential for clinical utility.

Keywords:
Biomarker, Centrosome, Features, Lung cancer, Prognosis.

INTRODUCTION

Non-small cell lung cancer (NSCLC) histopathology and staging have limited ability to predict individual lung cancer progression and outcome. Using gene expression technology, investigators have recently reported the value of expression profiles to forecast survival and outcome of NSCLC patients [1-5]. Centrosome protein expression is a component of many of these profiles. Given its central role in the biogenesis and prognosis of malignancy, the development of an objective, quantitative, and reproducible measurement of centrosome features, which could be available immediately to pathologists at the time of diagnosis, would be of immense clinical value.

We have developed a method to objectively and reproducibly quantify the centrosome features in lung cancer [6]. In this report, we have applied this approach to confocal images of resected, immunostained stage I NSCLC to measure centrosome numeral and morphological features for statistical analysis and classification. The result revealed the feasibility of predicting the outcome of stage I NSCLC patients by using centrosome features as prognostic biomarker.

MATERIALS AND METHODS

Specimen preparation and confocal scanning. After approval by the Institutional Review Board, archived blocks of lung tumor resected from stage I lung cancer patients were provided by the Tissue Core of the H. Lee Moffitt Cancer Center and Research Institute. Centrosomal images were acquired and processed as previously described [6]. Standard fluorescent microscopy detects in-focus and out-of-focus photons. Laser scanning confocal microscopy utilizes pinhole apertures in order to reject out-of-focus light typically detected by standard fluorescent microscopy. While disadvantages of thin optical sectioning parameters include the potential loss of out-of-focus information, the centrosomal features of interest in this project are extremely specific and, therefore, require precise detection optimized by limiting out-of-focus light detection. After comparison, a confocal microscope was used to image the tissue sections (see Fig. 1. a).

From confocal images taken of each section, regions of interest (ROI) are selected manually. Every ROI consists of one cell with its centrosomes (see Fig. 1b, 1c). 235 ROIs yielded 594 centrosomes from six survivors (survived >9 yr), and 211 ROIs yielded 309 centrosomes from six fatalities (survived < 4 yr) for this experiment. Among the six survivors were four cases of stage IA and two cases of stage IB lung cancer. Of the six fatalities were three cases of stage IA and three cases of stage IB lung cancer.

Image processing. Resolution enhancement was performed to improve the measurement of centrosome features, such as shape, boundary, and structure. Two-dimensional first-degree Lagrange polynomial interpolation [7] was performed. This linear interpolation technique uses information given only by the two adjacent pixels and leads to unbiased image approximation.

After image enhancement, the centrosomes were isolated (segmented) from the background. After comparing thresholding methods, Kapur's maximum entropy-based thresholding was applied [8]. This method considers the foreground (centrosomes) and the background (nuclei, cytoplasm, etc.) within an ROI as separate signal sources and finds the threshold, which maximizes the sum of the entropies of the two classes. The entropy threshold resulted in consistently accurate segmentation (see Fig. 1. d, e).

Some features, such as fractal dimension, are highly sensitive to image processing. For example, using a morphological operation would change the roughness of boundary, which would affect the classification result. To reduce this affect, we were careful to use a morphological

operation as little as possible during centrosome segmentation. For further improving the classification performance, we'll totally avoid using of the morphological operation.

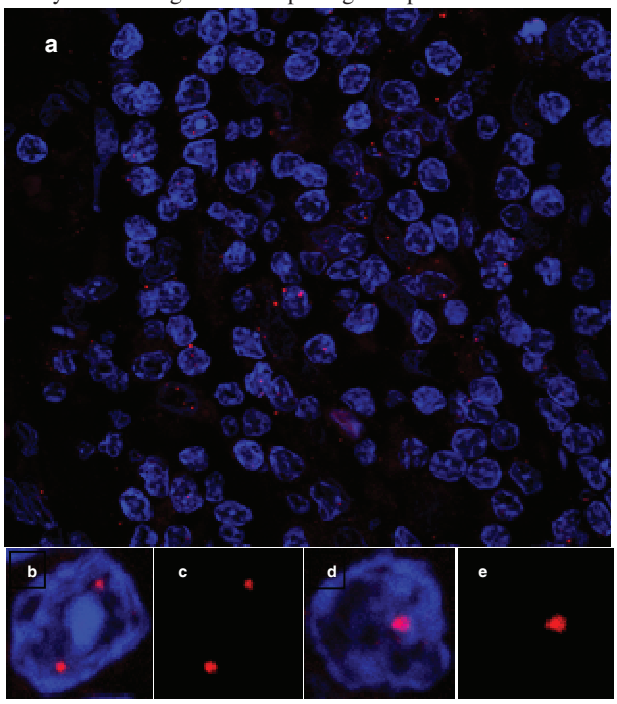


Fig. 1. (a) Z-stack confocal image in which centrosomes are shown as red spots. (b) and (c) are two ROIs selected from (a) with 2X interpolation. (d) and (e) are segmentation images of (b) and (c), which centrosomes are isolated from background and ready for feature extraction.

After centrosome isolation, 12 previously defined centrosomal image features were extracted for statistical analysis and classification [6]. These 12 features can be classified into 6 categories: centrosome number, size, shape, boundary, structure, and intensity.

Feature selection was then performed to improve computation efficiency and improve classification accuracy. We use the feature selection algorithm called Minimum Redundancy-Maximum Relevance (MRMR) [9], which selects features by testing whether certain preset conditions about features and target classes are satisfied. MRMR yields low classification errors and high generalization of selected features for different classifiers with low costs of computation.

Statistical analysis. Matlab Statistics Toolbox is used for statistical analysis. This toolbox is more directed towards numerical application. It provides a comprehensive set of tools to assess and understand data. The two-sample t-test is performed to verify weather the two samples have the same mean (the null hypothesis) and can be distinguished by these features. The test is carried out under the assumption that two samples are independent and normally distributed [10]. Our sample sizes of n = 309 and 594, respectively, are sufficient to approximate to normal distribution.

The Kolmogorov-Smirnov (KS) test determines whether two samples are drawn from the same distribution (the null hypothesis) and is sensitive to differences in both location and shape of the empirical cumulative distribution functions (CDF) of the two samples. KS test makes no assumption about distribution normality. In addition to h (hypothesis test result) and p (*p* value of the test), the KS test also returns the test statistic k, which quantifies the difference between distribution shapes of the two samples and can be written as: $k = \max(|F_1(x) - F_2(x)|)$, where $F_1(x)$ and $F_2(x)$ are empirical cumulative distribution functions of samples 1 and 2, respectively [11].

The two-sample t-test and KS test delineate the data from different aspects and expose more information of the data.

Classification and Evaluation. Centrosome features were entered into two types of classifiers: Linear discriminant analysis (LDA) and Support Vector Machines (SVM). Matlab Statistics Toolbox and Informatics Toolbox were used for implementation of LDA and SVM, respectively.

LDA projects the data points into a new space of lower dimension, which maximizes the between-class variability and minimizes their within-class variability [12, 13].

SVM works oppositely from LDA, it transforms nonlinear input data into higher dimensional space, where the data are linearly separable by an optimal hyperplane. The classification task is implemented by using kernel functions. Via appropriate selection of kernel and optimized parameters, SVM is capable of adapting complicated classification tasks [14]. In our case, Gaussian Radial Basis Function kernel, $k(\mathbf{x_i}, \mathbf{x_j}) = \exp(-\gamma \| \mathbf{x_i} - \mathbf{x_j} \|^2)$, $\gamma = 2$ is selected after the optimization.

In K-fold cross-validation [15], after randomly partitioning the original sample into K subsamples, one is retained as the validation datum for testing the classifier, and the remaining K − 1 subsamples are used as training data. With cross-validation repeated K times, using all observations for both training and validation, each observation is used for validation exactly once. In this experiment, 10-fold cross-validation has been performed 3 times. Since the samples are selected randomly, there may be fluctuation between different validations. The average of three validations has been taken as the result.

Bagging is an acronym for bootstrap aggregating. It is a technique that improves classification accuracy by combining classifications from randomly generated training sets, sampling repeatedly from a data set with replacement. On average, a bootstrap sample contains approximately 63% of original training data, because each element has a probability $1 - (1 - 1/N)^N$ of being selected in each sample. If N is sufficiently large, this probability converges to $1 - 1/e \approx 0.632$. Overall accuracy is computed by combining the accuracies of each bootstrap sample (ε_i) with the accuracy computed from a training set that contains all the labeled examples in the original data: $acc_{boot} = \frac{1}{b}\sum_{i=1}^{b}(0.632 \times \varepsilon_i + 0.368 \times acc_s)$ where acc_s is the overall accuracy of the classifier; b is the number of bootstrap samples [16]. In this experiment, 100 bootstrap samples (b = 100) were taken.

Estimating a confidence interval (CI) for accuracy will verify the reliability of a classification result. Our

classification task is a binomial experiment. Therefore, it has a binomial distribution: $P(X=x) = {}_N C_x \, p^x (1-p)^{N-x}$ Where N is the number of centrosomes, X is the number that was correctly predicted by a classifier and p is the accurate rate of the classifier. When N is sufficiently large (Np >5 & N(1-p) > 5), it can be well approximated by a normal distribution. Based on the normal distribution, the CI of accuracy can be derived: $P\left(-Z_{\alpha/2} \le (acc-p)/(\sqrt{p(1-p)/N}) \le Z_{1-\alpha/2}\right) = 1-\alpha$ Where acc is accuracy, $Z_{\alpha/2}$ and $Z_{1-\alpha/2}$ are the upper and lower bounds obtained from a standard normal distribution at confidence level $(1-\alpha)$. Since standard normal distribution is symmetric around Z = 0, it follows $Z_{\alpha/2} = Z_{1-\alpha/2}$. Rearranging the inequality achieves the CI for p [16]:

$$p = \left(2\times N\times acc + Z_{\alpha/2}^2 \pm Z_{\alpha/2}\sqrt{Z_{\alpha/2}^2 + 4\times N\times acc - 4\times N\times acc^2}\right)/\left(2(N+Z_{\alpha/2}^2)\right)$$

RESULTS

Feature set optimization. Following MRMR optimization, the 12 features were ranked. We tested the ability of LDA to classify centrosomes by increasing the number of features from 1 to 12 selected according to the optimized order. After performing 10-fold cross-validation 3 times, we averaged 3 validation results as the overall result. The classification error rate is plotted against the number of features in the LDA model in Fig. 2.

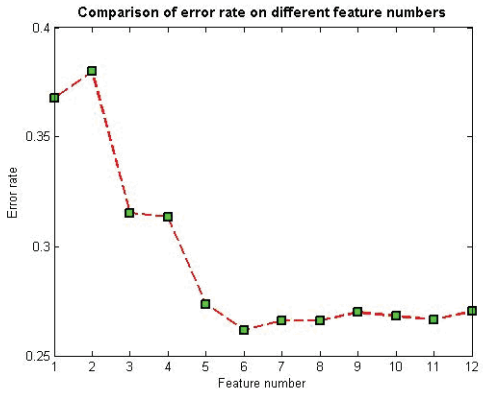


Fig. 2. The distribution of error rates reveals the variation of error rates by number of features analyzed. The order of features is optimized by MRMR. This plot shows the error rate reaches the minimum when the number of features is six.

This plot displays the sensitivity of the classification error rate to the addition of features. The error rate reached the minimum when the first six features were selected. The other six features were redundant and minimally relevant to the target, which had little or no contribution to the classification accuracy. We used these six features as our feature set for statistical analysis and classification. They were Fractal Dimension, Roundness, Area, Num/cell, Perimeter Ratio, and Mean Intensity. Correlation coefficients of these six features were calculated. Some features are strongly correlated, such as Perimeter Ratio with roundness (-0.91), and Perimeter Ratio with Fractal Dimension (-0.86). We tried to remove Perimeter Ratio, and tested on SVM. The accuracy fell from 0.786 to 0.755, which means this feature has a small contribution to classification accuracy. Therefore, we kept all six features.

Statistical analysis. The two-sample t-tests resulted in Table 1, h = 1 with the *p* values less than 0.001 for all six features, indicating rejection of the null hypothesis, which means at 99.9% confidence level the two samples have different means for all the six features. The confidence intervals (ci) on the mean differences of all six features did not contain zero, which means that there were significant mean differences for all six features from the two samples; the centrosomes represent two separate populations and are distinguishable on the six features.

The KS test results in Table 1, h = 1 with $p \le 0.0015$ for all six features rejected the null hypothesis, which verified, at 99.85% confidence level, the two samples were from the different distributions (locations and shapes). The statistic k values indicated distinct distributions based on the distances between the two CDFs. Four features' k values were > 40% and the smallest value of k was 12.2%, indicating that the distances between the two CDFs of the centrosome features for survivors and fatalities were large enough to be distinguished.

The feature data of survivors and fatalities were separable according to the statistical analysis results.

Table 1. Statistical analysis results

		Num /cell	Area	Round-ness	Intensity	Perim Ratio	Fractal Dimen
t-test	h	1	1	1	1	1	1
	p	3.04 e-07	1.02 e-13	5.32 e-22	7.47 e-4	2.22 e-33	1.686 e-36
	c i	-1.465 -0.661	168.8 287.2	0.554 0.827	-10.63 -2.826	-0.107 -0.078	0.070 0.094
KS test	h	1	1	1	1	1	1
	p	0.0015	5.87 e-31	8.48 e-34	4.20 e-03	1.85 e-30	1.87 e-39
	k	0.178	0.412	0.431	0.122	0.409	0.466

Classification. The results of 10-fold cross-validation for LDA and SVM are shown in Table 2. Comparing the results, the accuracy rate of SVM (79%) was somewhat better than LDA (74%). The fluctuations between different test results of 10-fold cross-validation were very small (< 0.0035 for LDA accuracy, < 0.0005 for SVM accuracy), demonstrating stability of the 10-fold cross-validation results.

Bagging with SVM significantly improved the accuracy of SVM from 79% to 85%, which means there is room to improve the accuracy.

The CI of accuracies on 95% confidence level can also be found in Table 2. These CIs confirmed that accuracies were reliable and acceptable.

Table 2. 10-fold cross-validation and 100 bagging results

	Accuracy rate	Sensitivity	Specificity	CI of Accuracy
LDA (cross-validation)	0.738	0.808	0.702	0.708 0.766
SVM (cross-validation)	0.786	0.838	0.683	0.758 0.811
SVM (bagging)	0.855	0.798	0.884	0.831 0.876

DISCUSSION

Our experiment provides an objective and quantitative assessment of centrosomal numeral and morphological abnormalities and shows the magnitude of these differences. Based on the centrosome classification, future statistical analysis and classification will be applied to a larger database. At present, there is no clinical method to predict the portion of patients whose resected, localized stage I NSCLC recurs with distant metastases. Building upon evidence that centrosomal abnormalities cause chromosomal instability (CIN) [17-20], a new quantitative biomarker of CIN, indicating subsequent lung tumor behavior would have great clinical significance. Such a marker could individualize the application of existing lung cancer therapies. For example, by identifying aggressive tumors in stage I for whom adjuvant treatment may be beneficial [21], as well as direct the development of new drugs against centrosome targets.

ACKNOWLEDGEMENTS

The authors thank Inna Fedorenko, Ph. D. student of H Lee Moffitt Cancer Center, for helping process of tissues and acquisition of high quality images. The authors also thank Mark Lloyd, supervisor of The Analytic Microscopy Core Facility at H. Lee Moffitt Cancer Center, for his tireless technical support with confocal imaging and other requirements. This project was supported by grant # 07KN-14-12307 (to TAZ) from State of Florida James & Esther King Biomedical Research Program and grant #30-15900-01-01 (to MST) from: State of Florida/Bankhead Coley.

References

[1] Beer DG, Kardia SL, Huang CC, et al. "Gene-expression profiles predict survival of patients with lung adenocarcinoma," Nat Med. vol. 8 pp. 816-824, 2002.

[2] Wigle DA, Jurisica I, Radulovich N, et al. "Molecular profiling of non-small cell lung cancer and correlation with disease-free survival," Cancer Res. vol. 62, pp. 3005-3008, 2002.

[3] Chen H-Y, Yu S-L, Chen C-H, et al. "A five-gene signature and clinical outcome in non-small-cell lung cancer," N Engl J Med. vol. 356, pp. 11-20, 2007.

[4] Lu Y, Lemon W, Liu PY, et al. "A gene expression signature predicts survival of patients with stage I non-small cell lung cancer," PLoS medicine, vol. 3, pp. 2229-2243, 2006.

[5] Kadara H, Behrens C, Yuan P, et al. "A five-gene and corresponding protein signature for stage-I lung adenocarcinoma prognosis," Clin Cancer Res. vol. 17, pp. 1490-1501, 2011.

[6] Song D, Fedorenko I, Pensky M, et al. "Quantificational and statistical analysis of the differences in centrosomal features of untreated lung cancer cells and normal cells," Anal Quant Cytol Histol. vol. 32, pp. 280-290, 2010.

[7] Lehmann TM, Gonner C, Spitzer K. "Survey: Interpolation methods in medical image processing," IEEE Trans on Med Imaging, vol. 18, pp. 1049-1075, 1999.

[8] Yin P-Y. "Maximum entropy-based optimal threshold selection using deterministic reinforcement learning with controlled randomization," Signal Processing, vol. 82 pp. 993-1006, 2002.

[9] Peng HC, Long FH, Ding C. "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," IEEE Trans on Pattern Analysis and Machine Intelligence, vol. 27 pp. 1226-1238, 2005.

[10] Glass GV, Hopkins KD. "Statistical methods in education and psychology," Boston, MA: Allyn and Bacon; 1996.

[11] Kozmann GG, L.S. Lux, R.L. "Nonparametric identification of discriminative information in body surface maps," IEEE Trans on Biomedical Engineering, vol. 38, 1061-1068, 1991.

[12] Qiao Z, Zhou, L., Huang, J. Sparse, "Linear discriminant analysis with application to high dimension low sample size data," IAENG International Journal of Applied Mathematics, vol. 39, pp. 48-60. 2009.

[13] Kumar K, Bhattacharya, S. "Artificial neural network vs linear discriminant analysis in credit ratings forecast: A comparative study of prediction performances," Review of Accounting and Finance, vol. 5, pp. 216-227, 2006.

[14] Gokcen I, Peng J. "Comparing linear discriminant analysis and support vector machines," Advances in information systems, Book Series: Lecture notes in computer science, vol. 2457, pp. 104-113. 2002.

[15] Kohavi R. A, "Study of cross-validation and bootstrap for accuracy estimation and model selection," Proceedings of the 14th international joint conference on Artificial intelligence, vol. 2, pp. 1137-1143, 1995.

[16] Tan P-NS, M. Kumar, V. "Introduction to data mining," Boston, MA: Pearson & Addison Welsey;: pp. 145-198, 2006.

[17] Fukasawa K, Choi T, Kuriyama R, Rulong S, Vande Woude GF. "Abnormal centrosome amplification in the absence of p53," Science, vol. 271, pp. 1744-1747, 1996.

[18] Metzler M. "Mutations in a novel cilia–centrosome protein cause a cystic kidney disease associated with retinal degeneration," Clinical Genetics, vol. 79, pp. 222-224, 2011.

[19] Zhao X, Jin S, Song Y, et al. "Cdc2/cyclin b1 regulates centrosomal nlp proteolysis and subcellular localization," Cancer Biol Ther, vol. 10 pp. 945-952, 2010.

[20] Fukasawa, K. "Centrosome amplification, chromosome instability and cancer development," Cancer Lett.; vol. 230, pp. 6-19, 2005.

[21] Boutros R, Ducommun B. "Asymmetric localization of the cdc25b phosphatase to the mother centrosome during interphase," Cell cycle, vol. 7, pp. 401-406, 2008.