Cytometry
PART A
Journal of the International Society for Advancement of Cytometry

Detection and Classification of Thyroid Follicular Lesions Based on Nuclear Structure from Histopathology Images

Wei Wang,[1] John A. Ozolek,[2] Gustavo K. Rohde[1,3,4]*

[1]Center for Bioimage Informatics, Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

[2]Department of Pathology, Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania 15201

[3]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

[4]Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Received 25 September 2009; Revision Received 17 November 2009; Accepted 8 December 2009

Grant sponsor: NIH; Grant number: 5R21GM088816-02.

*Correspondence to: Gustavo K. Rohde, Ph.D., Assistant Professor of Biomedical Engineering, and Electrical and Computer Engineering (by courtesy) HHC 122, 5000 Forbes Avenue Pittsburgh, PA. 15232

Email: gustavor@cmu.edu

Published online 22 January 2010 in Wiley InterScience (www.interscience.wiley.com)

DOI: 10.1002/cyto.a.20853

Abstract

Follicular lesions of the thyroid are traditionally difficult and tedious challenges in diagnostic surgical pathology in part due to lack of obvious discriminatory cytological and microarchitectural features. We describe a computerized method to detect and classify follicular adenoma of the thyroid, follicular carcinoma of the thyroid, and normal thyroid based on the nuclear chromatin distribution from digital images of tissue obtained by routine histological methods. Our method is based on determining whether a set of nuclei, obtained from histological images using automated image segmentation, is most similar to sets of nuclei obtained from normal or diseased tissues. This comparison is performed utilizing numerical features, a support vector machine, and a simple voting strategy. We also describe novel methods to identify unique and defining chromatin patterns pertaining to each class. Unlike previous attempts in detecting and classifying these thyroid lesions using computational imaging, our results show that our method can automatically classify the data pertaining to 10 different human cases with 100% accuracy after blind cross validation using at most 43 nuclei randomly selected from each patient. We conclude that nuclear structure alone contains enough information to automatically classify the normal thyroid, follicular carcinoma, and follicular adenoma, as long as groups of nuclei (instead of individual ones) are used. We also conclude that the distribution of nuclear size and chromatin concentration (how tightly packed it is) seem to be discriminating features between nuclei of follicular adenoma, follicular carcinoma, and normal thyroid.

Key terms

follicular; thyroid; image analysis; nuclear structure; classification

THE distinction between follicular adenoma of the thyroid (FA; neoplastic, but not capable of metastases) and follicular carcinoma of the thyroid (FTC; neoplastic, capable of metastases) has largely been a tedious and difficult task for pathologists and until very recently pathologists have only been able to rely on the histopathology to distinguish them. The encompassing differential diagnosis of follicular lesions of the thyroid includes follicular adenoma, follicular carcinoma, follicular variant of papillary carcinoma, and nodular goiter. Follicular variant of papillary carcinoma is usually not as troublesome as the distinction between FA and FTC due to the presence of distinctive and unique nuclear features in this variant of papillary carcinoma. The microscopic presence of tumor invasion into the capsule (fibrous covering around the lesion) and/or invasion of the tumor into small vessels of the capsule (vascular invasion) are the distinguishing hallmarks of FTC. FA may have a fibrous capsule, but lacks capsular and/or vascular invasion. These two features are diagnostic because other visual features commonly used by pathologists including microarchitecture and cytological characteristics, notably nuclear and cytoplasmic morphology are virtually indistinguishable between FA and FTC (1). This similarity is illustrated in

ISAC
International Society for Advancement of Cytometry

Fig. 1. Specifically, a proper distinction between FA and FTC can be rendered only after the lesion has been removed surgically (thyroid lobectomy or total thyroidectomy) and the entire lesion has been sampled to determine if capsular and/or vascular invasion is present microscopically. This diagnostic algorithm has several limitations. The tissue processing is time consuming (fixation of the entire lesion, submission of the entire lesion in multiple tissue cassettes) and subject to sampling error (sampling the exact areas of capsular invasion). From the pathologists perspective, this requires a great deal of time to carefully examine every slide of the entirely submitted capsule/tumor particularly for evidence of capsular/vascular invasion. This may require multiple sections of each paraffin block to thoroughly sample the capsule to locate focal capsular and/or vascular invasion. Then, even after this intense scrutiny, pathologists may argue whether a focus seen under the microscope truly represents capsular and/or vascular invasion (minimally invasive FTC) (2).

As eluded earlier, more recent discoveries in the molecular pathogenesis of FA and FTC have yielded specific genes and pathways that may help to distinguish these lesions. Gene expression analysis of thyroid tumors shows differential expression of a number of genes between normal thyroid, follicular adenoma and follicular carcinoma (3–6). Interestingly, and not surprisingly, the study by Chevillard et al demonstrated only 43 genes with differential expression between FA and FTC, perhaps in some way paralleling their visual similarity (4). At our institution, a panel of PCR-based genes has recently been made available to help with this distinction. However, these tests are not widely available and only augment the total cost of health care. An image analysis tool that could separate and distinguish between FA and FTC, with perfect accuracy, would be extremely useful in the clinical diagnostic setting.

Approaches for Automated Image-Based Pathology

Automated image-based methods for detecting and classifying lesions have long been described in the scientific literature [see (7–11) for reviews]. They have been applied to studying cellular and tissue architecture in several types of cancers including prostate (12), cervix (13,14), thyroid (15–24), liver (25–29), breast (30), and several others. Most methods utilize an image processing pipeline that involves data preprocessing, feature extraction, and automated classification.

The first step, image preprocessing, consists of segmentation of subcellular structures, particularly nuclei because of the prominent role they have in the diagnostic process. Popular methods are based on thresholding (31), active contours (32), watershed (33), and others. Feature extraction is then employed so that different images/structures can be compared quantitatively. An extensive review of feature extraction methods is provided in (8,9). Among these features, popular classes include shape-related features and generic texture related features such as average intensity, entropy, Haralick features, Zernike polynomials, fractals, Wavelet, Fourier and related multiresolution/multiscale methods (9,34,35). In addition, features for measuring specific textures related to chromatin distribution have also been developed (36). In the third and final step, a wide variety of automated classification methods have been used to determine whether an image (at either tissue or cellular level) belongs to a diseased tissue or not, based on the features computed in the extraction step. Several multivariate machine learning-based classification methods have been used for this purpose. Some of these include neural networks (37), nearest neighbors (38), logistic regression (39), linear discriminant analysis (40), decision trees (41), support vector machines (42) and others. In addition, voting-based strategies for combining the output of several different classifiers have been designed (43–47).

Previous Works in Automated Analysis of Thyroid Lesions

There have been several attempts to build automated systems for specifically classifying follicular thyroid lesions based on nuclear structure (17,18,20,23,24). The general idea employed is to utilize a set of features describing properties of sets of nuclei (e.g., average area of nuclei, coefficient of variation etc.) as the basis with which to compare several distributions. Classification accuracies for the several pathologies range from 50% to 99%. At the time of writing, we are not aware of a previously published system that considers three or more classes (or pathologies) that perfectly classifies more than one human case with 100% accuracy for all classes.

Our Contributions

In this article, we describe an automated approach capable of classifying real image data of FA, FTC, and normal thyroid (NT). We demonstrate that the information included in the chromatin distribution of nuclei from each of these classes contains enough information to classify the available test data with 100 % accuracy if considering groups of nuclei rather than individual ones. Our approach is based on the popular feature-based processing pipeline discussed above. In comparison with previous methods for classifying thyroid lesions, our approach utilizes a much larger number of features, and a different classification strategy. It combines a support vector machine classifier together with a simple voting strategy to determine if a set of nuclei emanating from some unknown tissue is most similar to FA, FTC, or NT. In addition to automated classification, we also describe newly developed methods for identifying unique and defining chromatin patterns pertaining to each FA, FTC, and NT. These calculations are accompanied by statistical significance assessments based on the *p*-value, and classification accuracies.

Finally, although this article demonstrates that these methods are successful at decoding and classifying nuclear structure information for thyroid lesions, none of the methods we describe are specific to these lesions. Therefore our methods could find applications in decoding nuclear structure information in several malignancies of interest to pathologists.

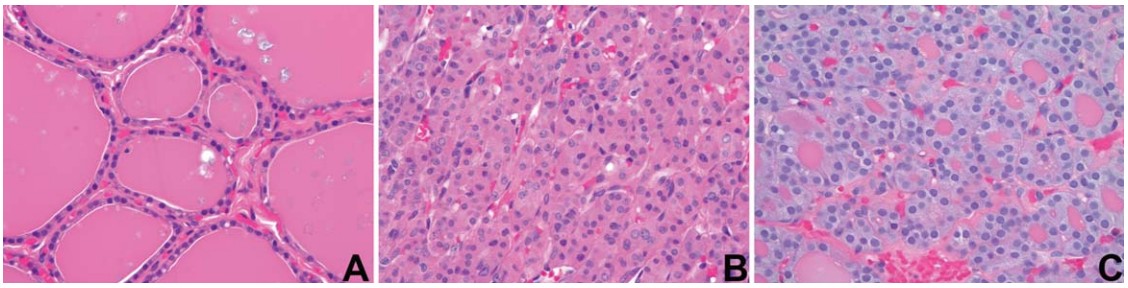

Figure 1. Examples randomly selected from hematoxylin and eosin stained sections of normal thyroid (**A**), follicular adenoma (**B**), and follicular carcinoma (**C**). By visual inspection, the follicular carcinoma appears to have larger nuclei overall than follicular adenoma and normal thyroid tissue, but the distinctions are subtle. (All micrographs taken at ×400). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Materials and Methods

In this section, we first illustrate how we acquire digital images from tissue blocks. After that, our methods for classifying different lesions and detecting unique characteristic chromatin patterns can be illustrated by flow charts shown in Fig. 2(A,B) respectively.

Data Acquisition and Preprocessing

Tissue procurement and processing. Tissue blocks were obtained from the archives of the University of Pittsburgh Medical Center (Institutional Review Board approval #PRO09020278). Cases for analysis included five resection specimens with the diagnosis of follicular adenoma of the thyroid (FA) and five cases of follicular carcinoma of the thyroid (FTC). For FA cases there are 4 females, 1 male with the mean age of 39.2 ± 20.3 years and mean lesion size of 4.8 ± 2.1 cm, and for FTC cases, there are 3 females, 2 males with mean age of 47.8 ± 9.2 years and mean lesion size of 2.8 ± 1.7 cm. Tissues were procured at the time of a surgical procedure retrospectively over a several year span. All tissues were fixed in 10% neutral buffered formalin and processed by routinely used methods on a conventional tissue processor using a series of graded alcohols and xylenes prior to paraffin-embedding. Tissue sections were cut at 5 micron thickness from the paraffin-embedded block and stained using the Feulgen technique which stains DNA only. This approach has been used in other morphometric studies to specifically isolate nuclei for computational analyses (48,49) and in our experience allows for more accurate segmentation of the nucleus compared to hematoxylin and eosin, hematoxylin alone, or periodic acid-Schiff stained sections. Briefly, the Feulgen technique involved rinsing of deparaffinized, and rehydrated slides in hydrochlo-

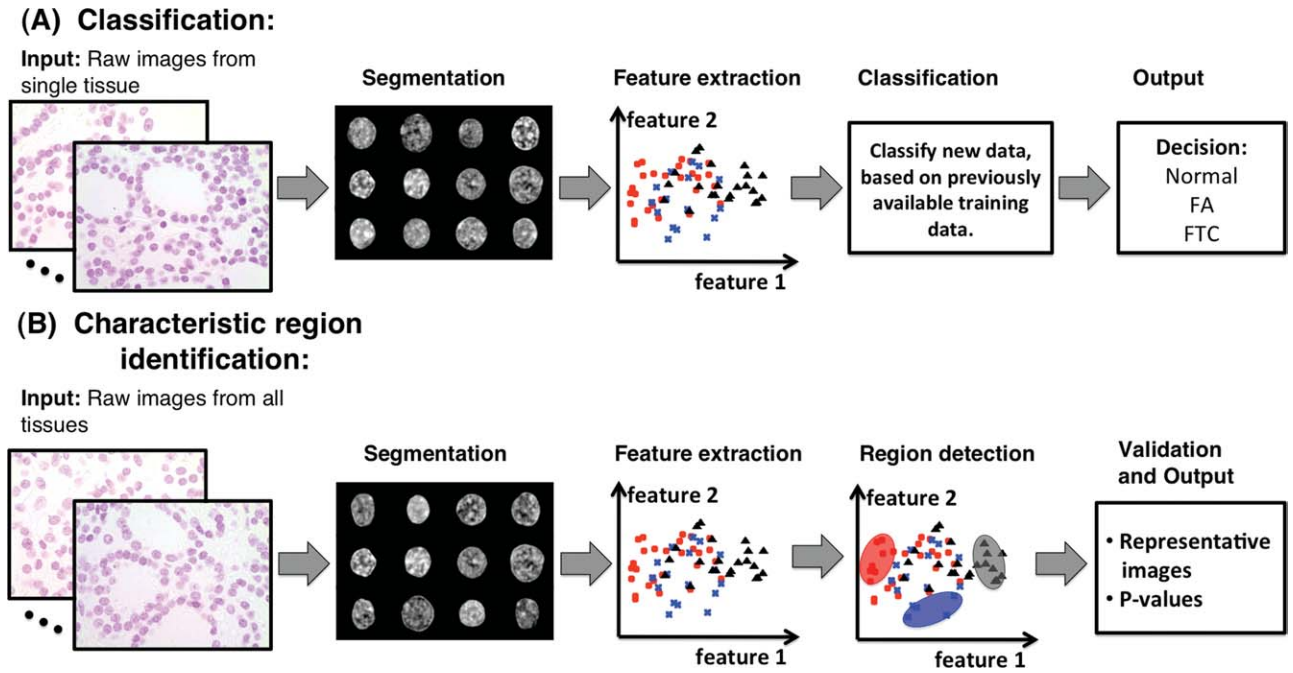

Figure 2. The flow charts of our method. **A** The flow chart of classifying different lesions as shown. **B**: The flow chart of detecting and validating unique characteristic chromatin patterns. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

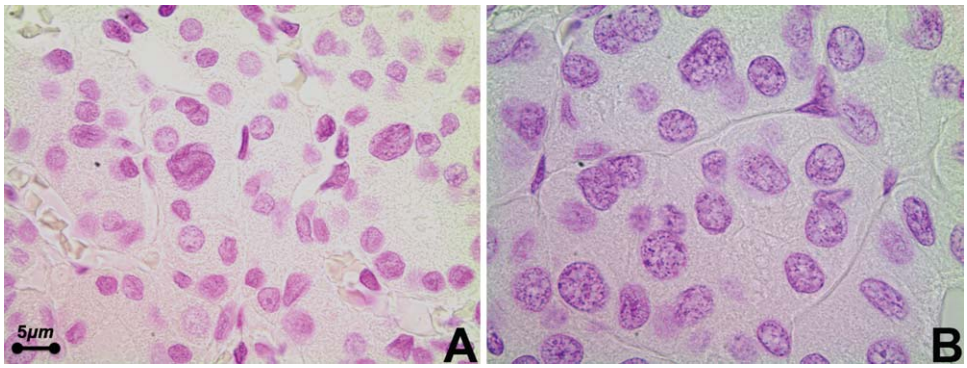

Figure 3. Examples randomly selected from Feulgen stained sections of follicular adenoma (**A**) and follicular carcinoma (**B**). This stain highlights DNA only and therefore stains only the nuclei. No cytoplasmic counterstaining was done. (All micrographs taken at ×1000). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

ric acid, followed by incubation in Schiffs reagent for one hour, followed by three successive two minute washes in bisulfate solution. Counterstaining was not performed to avoid possible interference from the cytoplasm with accurate isolation and segmentation of nuclear membrane boundaries. Only nuclei were stained with a deep magenta hue (Fig. 3).

Digital image acquisition. All images used for analysis in this study were acquired using an Olympus BX51 microscope equipped with a 100X UIS2 objective (Olympus America, Central Valley, PA) and 2 mega pixel SPOT Insight camera (Diagnostic Instruments, Sterling Heights, MI). Image specifications were 24bit RGB channels and 0.074 microns/pixel, 118 × 89 μm field of view. Slides were chosen by the pathologist (J.A.O.) that contained both lesion (FA or FTC) and adjacent normal appearing thyroid tissue (NT). In the majority of cases, we chose one slide that represented the lesion and adjacent normal thyroid tissue. If this was not possible, then a slide with normal thyroid from the case was chosen. The pathologist took 10 or 20 random images of lesion and normal to guarantee that at least 200 nuclei of both lesion and normal tissue per case were obtained. Nuclei were chosen by the pathologist and engineer (W.W.) for segmentation and analysis that demonstrated a complete and intact nuclear membrane within the focal plane.

Nuclear segmentation and preprocessing. We have implemented a semiautomatic method to segment large quantities of thyroid nuclei based on Feulgen stained histological images. It consists of three steps. First, a random field graph cut method (50) is utilized to search for the closed rough contours that best matches the intensity profile of nuclei. Briefly, we can formulate image segmentation as a pixel labeling problem and model the labeled image as a Markov random field. An energy function can be found to judge the quality of segmentation. This function is casted in a graph structure, and the min-cut of the graph corresponds to a globally optimal segmentation. Then an efficient level set active contour algorithm is used to refine the contours (51) to ensure best fit, without sharp cuts or edges. In the end, the pathologist (J.A.O.) reviews all the segmented nuclei, and removes those incorrectly segmented, or out of focus. As our goal in this study is to perform tasks related to digital pathology based on nuclear (chromatin) structure, the nuclei samples were converted to grayscale by selecting the green channel from the RGB images, and invert-

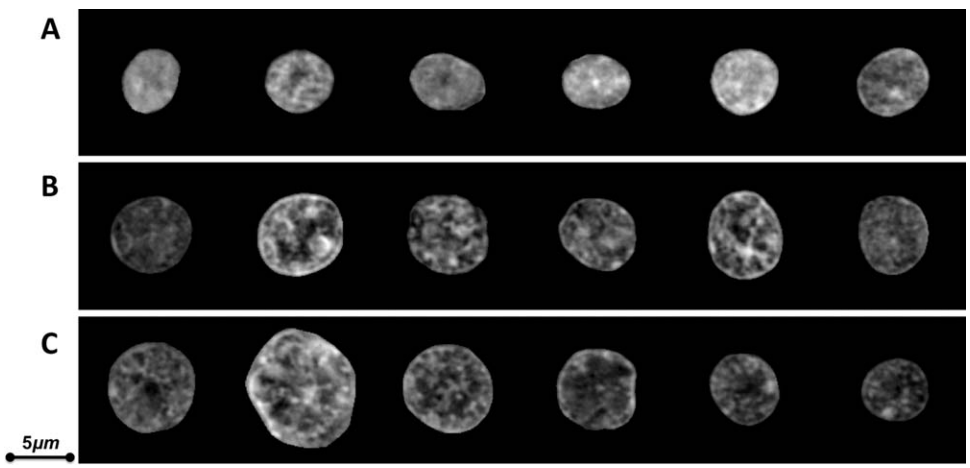

Figure 4. Random samples of preprocessed nuclei (after segmentation, gray level extraction, and intensity normalization).

Table 1. Results of classifying individual nuclei using the Mahalanobis nearest neighbor method

	NT (%)	FA (%)	FTC (%)
NT	40.13	33.25	26.62
FA	23.06	43.12	33.82
FTC	26.41	30.24	43.35

ing the intensity values such that a zero (color coded in black) corresponds to the relative minimum amount of chromatin in the nucleus. We note that selecting magenta channel in the CMYK color space gives the very similar results. All nuclei were normalized so that the sum of their intensity values is 1. This was done to guarantee that non-uniformities related to staining and image acquisition, from case to case, are not able to interfere with our method. In total, we extracted 871 normal thyroid nuclei, 489 follicular adenoma and 703 follicular carcinoma nuclei of thyroid from this dataset. A few sample nuclei randomly chosen for the entire data are displayed in Fig. 4.

Automated Classification

We utilized the popular numerical feature space approach for automatically classifying sets of nuclei. The general approach follows the steps: 1) select a portion of the available data for training, 2) train a classifier to determine to which class an unknown nucleus belongs. Then, in step 3, a group of nuclei can be assigned to a single class by testing each individual nucleus, and choosing the majority "vote" as the assignment for that group of nucleus. In all cases training and testing data never overlap.

In our implementation we utilized a 125-dimensional feature space to compare nuclei. In this case, each nucleus i is represented by a vector $x_i \in R^{125}$. These features are described in detail in Appendix. Briefly, they included shape features, Haralick texture features, wavelet and Gabor-based features, and others. Most features chosen are (some approximately) invariant to translations and rotations. However, we also normalized all of the nuclei data to account for rotations and translation according to a method we previously described in (52,53). The individual features were normalized by subtracting their mean and dividing by the standard deviation. The mean and standard deviation of each feature were computed from the training set of data.

We tested a total of four different classifiers: Mahalanobis distance nearest neighbors (MNN) based on the centroid of the class, and support vector machines (SVM) with linear, quadratic, and Gaussian radial basis function (RBF) kernels. The basic idea in MNN is to assign an individual nucleus x_i by finding the centroid of the class that is closest to it, as measured by the Mahalanobis distance (54) (akin to a multidimensional z-score). SVM on the other hand, tries to find a set of hyperplanes, in the feature space R^{125}, that have the largest distance to the nearest training datapoints of other classes (maximum margin) to obtain a lower generalization error of the classifier. By transformations that map the original features into a higher dimensional space [the famous kernel approach (55)], we can also find the maximum-margin hyperplanes in transformed feature space R^m, $m \geq 125$, which is equivalent to finding a nonlinear classifier in the original feature space R^{125}. The kernel functions we used to map the features are linear $K_L(x_i,x_j) = x_i^T x_j$, quadratic $K_L(x_i,x_j) = ({x_i}^T x_j + r)^2$ and radial basis function $K_L(x_i,x_j) = e^{-\gamma\|x_i - x_j\|^2}$, $\gamma > 0$. The optimal r and γ parameters were selected by performing 20 folds of the cross validation procedure in the training set. Finally, we used one-versus-all multiclass SVM to classify this 3-class problem, which reduced the multiclass problem into multiple binary problems and used a max-wins voting strategy to classify the testing instance. For more detailed descriptions of these classification methods, we refer the reader to a standard text in machine learning (54).

The classifiers were implemented in the MATLAB (Mathworks, Natick, MA) programming language using the Bioinformatics and Statistics toolboxes and LIBSVM (56). We have tested the classification procedures using both normalized and un-normalized data.

Identifying Unique Chromatin Patterns

Unique chromatin patterns pertaining to each NT, FTC, and FA class can be defined as regions in the feature space that are well populated by one class, but not the others. To that end, we analyzed the entire dataset in feature space in the following manner. The pairwise distance between all nuclei (from all subjects, from all classes) was computed. For each nucleus, the average distance to its W nearest neighbors was computed. The nuclei for which this value fell within the top 5% of all distance pairs were removed from analysis to eliminate outliers (nuclear structures extremely isolated in feature space).

For each class, unique patterns were chosen as those that were most distant (in the average of W nearest neighbor sense) to the two other classes remaining. Since the distribution of average distances for the entire dataset can be approximated as described earlier, an estimate for the p-value of each unique pattern identified can be computed. In our application, we

Table 2. Results of classifying individual nuclei using a support vector machine classifier with a quadratic kernel

	NT (%)	FA (%)	FTC (%)
NT	51.10	23.52	25.38
FA	25.6	48.7	22.7
FTC	26.41	30.24	43.35

Table 3. Results of classifying individual nuclei using a support vector machine classifier with a radial basis function kernel

	NT (%)	FA (%)	FTC (%)
NT	90.22	5.48	4.30
FA	7.5%	79.8	12.7
FTC	3.49	14.70	81.81

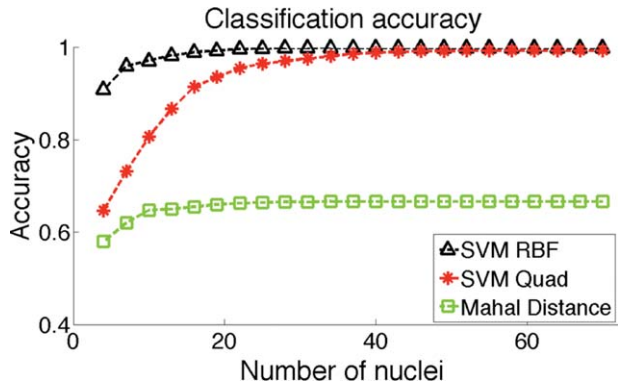

Figure 5. Average classication accuracy groups of nuclei with voting strategy. The SVM based on RBF can reach 100% accuracy when 28 nuclei are used. The SVM based on quadratic kernel also shows a good classification accuracy, but does not reach 100% accuracy until 70 nuclei are used. The MNN classifier does not show a good tendency to reach 100% accuracy until 70 nuclei are used. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

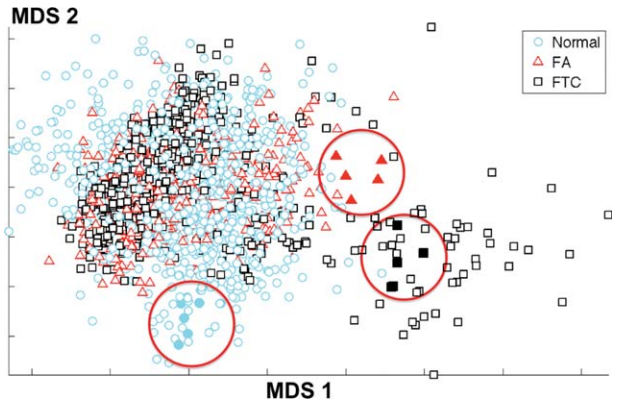

Figure 6. Two-dimensional representation nuclear population (NT, FTC, and FA), with axis corresponding to directions computed by multidimensional scaling technique excluding the top 5% outliers (see text). The solid shapes indicate the three unique regions corresponding to NT, FTC, and FA that have been identified by our algorithm. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

performed our experiments with $W = 5,\ldots,10$ (results not shown). The results show that, after removing outliers, the detected unique chromatin patterns (nuclei) do not change very much in general (the *P*-values are roughly the same), although few nuclei maybe different.

Results

The results of classifying individual nuclei using the MNN, SVM-Quadratic, and SVM-RBF methods are contained in Tables 1–3, respectively. We omit result of the SVM-linear method since these also performed poorly. These were obtained by randomly selecting 600 nuclei from the entire dataset (evenly distributed between human cases and classes) for training, and testing on the remaining set of nuclei. It is clear that the SVM-RBF approach yielded the best results in classification accuracy. The accuracy of classification, however, did not reach 100% using individual nuclei.

Figure 5 shows the average classification accuracy (average of the diagonal of the classification tables described above) when groups of nuclei are used instead to determine if an unknown set belongs to the class NT, FTC, or FA. Training was performed as described above. The average classification accuracy (calculated with 1,000 random drawings with replacement) is plotted as a function of the number of nuclei used in the calculation. In this calculation, the voting strategy utilizing the SVM-RBF classifier described above achieved 100% accuracy (no miss classification) when 28 nuclei were used. The SVM-quadratic classifier was also able to reach 100% classification accuracy, but required many more nuclei. The MNN and SVM-linear classifiers were not able to reach 100% classification accuracy, no matter how many nuclei were used.

Table 4. Results of classifying individual human cases using a leave one out cross validation strategy using 43 nuclei (each time)

	NT	FA	FTC
NT	10	0	0
FA	0	5	0
FTC	0	0	5

For this result, the support vector machine classifier (with a radial basis function kernel) was used.

The aforementioned results were obtained by pooling all data from all subjects together. We have also performed classification of individual cases using a "leave-one-out" cross validation strategy where the data from 9 cases are used for training, and testing is performed on the remaining case (a case-by-case strategy without mixing between patients). Table 4 shows that perfect classification accuracy can be achieved (for all cases) whenever 43 nuclei are used (some cases required fewer nuclei). Here we have also repeated the experiment 1,000 times, each time drawing 43 nuclei randomly.

In addition to automated classification, we have also identified the unique chromatin patterns pertaining to NT, FTC, and FA using the algorithm defined earlier. A two-dimensional representation [obtained using the multi dimensional scaling technique (57)] is shown in Fig. 6. In this figure each point represents a nucleus in feature space. We note that the top 5% outliers were removed, as described earlier. The solid shapes (indicated by large circles) indicate the regions in this space that have been identified as unique regions according to the algorithm described above. The nuclear structures pertaining to each identified pattern are shown in Fig. 7 (*P*-values indicated in the figure's legend).

By observing the patterns shown in Fig. 7 two immediate observations with regards to the differences between the patterns can be made. First there is an obvious difference in the area of the cells. This is further confirmed by the histogram of areas shown in Fig. 8(A). This figure shows that normal cells have nuclei that are smaller on average, and that neither FTC or FA nuclei have cells that are as small as those displayed in

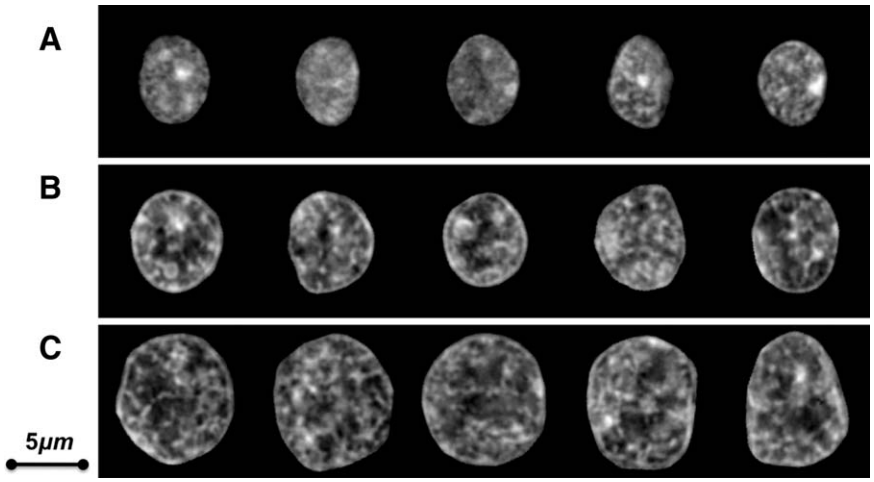

Figure 7. Nuclei samples of the unique regions from Fig. 6 are shown in rows **A**, **B**, and **C** corresponding to NT, FA, and FTC, respectively. The *p*-values of NT nuclei shown in (**A**) from left to right are 0.04, 0.05, 0.05, 0.06, 0.07. The *p*-values of FA nuclei shown in (**B**) from left to right are 0.02, 0.03, 0.04, 0.04, 0.05; The *p*-values of FTC nuclei shown in (**C**) from left to right are 0.05, 0.06, 0.07, 0.08, 0.09, approximately.

the first row of Fig. 7. Next, it is also clear that both FTC and FA have chromatin that are more tightly packed in certain areas, than do normal thyroid nuclei. A similar frequency plot of the entropy (58) of the individual nuclei image histograms is shown in Fig. 8(B). We note that there exist algorithms for determining which features are most significant in distinguishing the several classes. When stepwise discriminant analysis (54) was used on this dataset, it confirmed that size and texture information (including entropy) were amongst the best ranked features (1st and 3rd, respectively). Finally, we note this information (nuclear area and entropy of intensity values) is statistically significant in that it alone can be used to classify the data with 100% accuracy using a simple nearest neighbor strategy where the histogram (for each feature) of a group of unknown nuclei is compared (in the sum of difference squared sense) to the histogram of the training data. For the area feature, 110 nuclei are required for 100% classification accuracy, and 120 for the entropy.

Discussion

The distinction of FA and FTC is the paramount issue since no ancillary molecular or immunohistochemical techniques are readily available and current diagnostic modalities rely on tedious routine histopathological techniques. This is not a trivial problem since in the last year alone at only one hospital within our local health system ~100 thyroidectomies were performed with the diagnosis of follicular adenoma or follicular carcinoma. An algorithm or image analysis technique that requires only a few fields of view containing a few dozen nuclei from routinely processed tissue and that can distinguish diagnostic entities would save considerable time (pathologists, pathologist assistants, histotechnicians) and money (salaries for aforementioned personnel, technical and professional fees billed to medical insurance, personnel involved in billing, etc.).

Perhaps the most significant contribution of this article is that it provides evidence that methods currently available from the machine learning and data analysis community, after suitable adaptation, can be used to detect and classify thyroid lesions based on nuclear morphometry (chromatin distribution patterns as depicted images produced with the Feulgen staining technique). We tested our method on 10 cases (5 FA, 5 FTC, each also containing normal tissue). The results show that nuclear structure alone contains enough diagnostic infor-

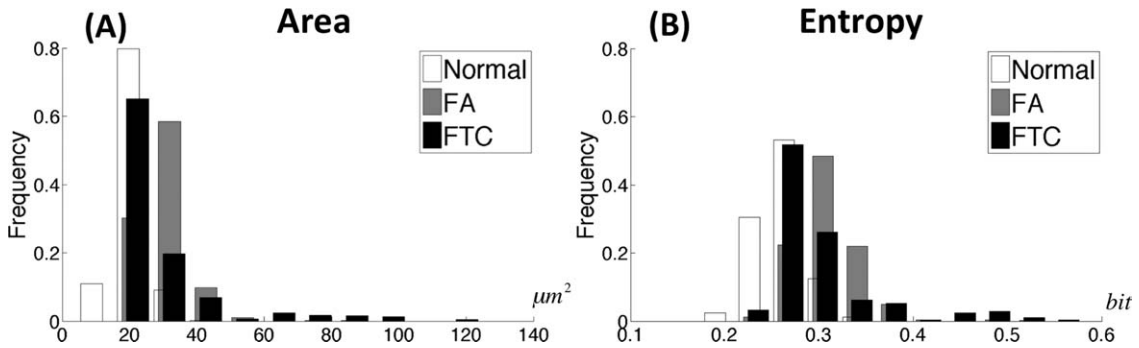

Figure 8. Histogram of individual features. **A**: The histogram of the area feature with units of μm^2. **B**: The histogram of entropy feature with the units of bit.

mation to classify the three thyroid tissue classes, we investigated, with 100% accuracy. Classification was performed both on a case-by-case basis, as well as with random samples drawn from the entire data population (all cases).

However, even though a high classification accuracy can be obtained, a fair amount of skepticism is useful here, since our study had a number of limitations. First, we believe that 10 human cases is not nearly enough to establish the validity of the technology in a clinical sense. It will be interesting to determine if our methods can maintain the high classification accuracy as the number of cases analyzed increases significantly. Another important limitation of our study is that we have focused exclusively on images obtained through the Feulgen staining technique. We did so because it greatly facilitated our automated segmentation approaches. Variation in staining procedure were accounted partially by applying a mathematical operation so that the sum of image intensity values in each segmented nucleus was 1. It will also be interesting to see whether the same methods we describe here can be applied to images obtained with other staining techniques (e.g., hematoxylin and eosin) as well.

From a methodological point of view, we highlight the following contributions. In comparison to previous attempts to classify thyroid lesions based on nuclear structure, we have used a much larger number of features. The specific classification strategy (SVM together with a voting method) is also new, within this context. Support vector machines with a RBF kernel have several well-documented theoretical advantages. Briefly, the algorithms available for estimating the solutions associated with support vector machines are able to estimate more adaptive and nonlinear separation borders between classes that are dictated by the data (55). Within the larger context of computational imaging-based pathology, the method we describe for identifying unique chromatin patterns, to our knowledge, is also new. We note that these methods can also be applied to other lesions/tissues, as long as individual nuclei (or other separate structures) can be accurately segmented.

One of the major keys to our approach and its potential ability to achieve 100% classification accuracy is that it considers groups of nuclei, as opposed to individual ones. Not surprisingly, as long as the classifier for individual nuclei has a reasonable accuracy, the classification based on group of data should reach higher accuracy than individual classification. Our results indicate, however, that relatively few nuclei (on the order of 50) may be needed to obtain very good automated classification results using nuclear structure alone. This corresponds to one or two 400x microscopic fields of view or $\sim$0.39 mm^2 of tissue area for these particular lesions. This is to be expected since cell populations in tissue are heterogeneous. We have illustrated this here showing considerable overlap in size and chromatin distribution between nuclei from all three groups. This heterogeneity can also be visualized through the microscope if enough time and careful observation is used. However, in the case of the pathologist trying to distinguish between a follicular adenoma and a follicular carcinoma, these distinctions of nuclear size and chromatin distribution may not be recognized over large groups of cells, or may be recognized but would not be enough to be absolutely confident in rendering a diagnosis of carcinoma which has far reaching implications for the patient; hence the current algorithm for processing and diagnosing these lesions as outlined in the introduction.

In addition, using these approaches, we have been able to show that size and chromatin concentration are the features that most significantly distinguish between NT, FTC, and FA (according to the limited set of test data available to us). Normal thyroid has overall smaller nuclear size and less variation (entropy) between nuclei as noted by their very symmetrical and normal distributions while FA and FTC have more skewed distributions with more overlap compared to normal thyroid with FTC showing the most skewed distributions of size and entropy. Interestingly, within feature space, nuclei of FTC appear to segregate into two major camps; those more resembling nuclei of NT and FA and those more distant from both FA and NT. This may reflect more nuclear size/chromatin heterogeneity within the population of nuclei in FTC. Alternatively, the subset of nuclei in FTC with more nuclear size/chromatin heterogeneity may belong to cases that have or may exhibit more aggressive histological appearance and/or clinical courses. Testing with more cases and correlation with clinical course and pathological appearance will further validate these findings.

Acknowledgments

The authors thank Dr. Dejan Slepcev from the Department of Mathematics and Dr. Ann B. Lee, of the Department of Statistics of Carnegie Mellon University for important discussions related to the geometric interpretation of data. They also thank Clifford Smay, HT from the histology core facility at the Magee Womens Research Institute, University of Pittsburgh, for slide preparation. Finally, the authors also thank Cheng Chen, from the Department of Biomedical Engineering of Carnegie Mellon University for assistance in implementing the nuclear segmentation routine used.

Literature Cited

1. Rosai J, Ackerman LV. Rosai and Ackerman's Surgical Pathology, 9th ed., Edinburgh: Mosby, 2004.
2. Heffess CS, Thompson LD. Minimally invasive follicular thyroid carcinoma. Endocr Pathol 2001;12:417–422.
3. Barden CB, Shister KW, Zhu B, Guiter G, Greenblatt DY, Zeiger MA, Fahey T Jr. Classification of follicular thyroid tumors by molecular signature: results of gene profiling. Clin Cancer Res 2003;9:1792–1800.
4. Chevillard S, Ugolin N, Vielh P, Ory K, Levalois C, Elliott D, Clayman GL, El-Naggar AK. Gene expression profiling of differentiated thyroid neoplasms: diagnostic and clinical implications. Clin Cancer Res 2004;10:6586–6597.
5. Nikiforov YE. Thyroid carcinoma: Molecular pathways and therapeutic targets. Mod Pathol 2008;21 (Suppl 2):S37–S43.
6. Zhao J, Leonard C, Gemsenjager E, Heitz PU, Moch H, Odermatt B. Differentiation of human follicular thyroid adenomas from carcinomas by gene expression profiling. Oncol Rep 2008;19:329–337.
7. Mulrane L, Rexhepaj E, Penney S, Callanan J, Gallagher W. Automated image analysis in histopathology: A valuable tool in medical diagnostics. Expert Rev Mol Diagn 2008;8:707–725.
8. Demir C, Yener B. Automated cancer diagnosis based on histopathological images: a systematic survey. Tech Rep TR-05–09, Rensselaer, NY: Rensselaer Polytechnic Institute, 2005.
9. Rodenacker K, Bengtsson E. A feature set for cytometry on digitized microscopy images. Anal Cell Pathol 2003;25:1–36.
10. Gil J, Wu HS. Application of image analysis to anatomic pathology: realities and promises. Cancer Invest 2003;21:950–959.

11. Bartels PH, Gahm T, Thompson D. Automated microscopy in diagnostic histopathology: From image processing to automated reasoning. Ist J Imaging Syst Technol 1998;8:214–223.
12. Singh SS, Kim D, Mohler JL. Java web start based software for automated quantitative nuclear analysis of prostate cancer and benign prostate hyperplasia. Biomed Eng Online 2005;4:1:31.
13. Bengtsson E. Fifty years of attempts to automate screening for cervical cancer. Med Imaging Tech 1999;17:203–210.
14. Abulafia O, Sherer DM. Automated cervical cytology: meta-analysis of the performance of the PAPNET system. Obstet Gynecol Surv 1999;54:253–264.
15. Burger G, Ploem JS, Goerttler K. Clinical Cytometry and Histometry. London: Academic Press, 1987.
16. Murata SI, Mochizuki K, Nakazawa T, Kondo T, Nakamura N, Yamashita H, Urata Y, Ashihara T, Katoh R. Morphological abstraction of thyroid tumor cell nuclei using morphometry with factor analysis. Microsc Res Tech 2003;61:457–462.
17. Karslioglu Y, Celasun B, Gunhan O. Contribution of morphometry in the differential diagnosis of fine-needle thyroid aspirates. Cytometry Part B 2005;65B:22–28.
18. Gupta N, Sarkar C, Singh R, Karak AK. Evaluation of diagnostic efficiency of computerized image analysis based quantitative nuclear parameters in papillary and follicular thyroid tumors using paraffin-embedded tissue sections. Pathol Oncol Res 2001;7:46–55.
19. Vasko VV, Gaudart J, Allasia C, Savchenko V, Di Cristofaro J, Saji M, Ringel MD, De Micco C. Thyroid follicular adenomas may display features of follicular carcinoma and follicular variant of papillary carcinoma. Eur J Endocrinol 2004;151:779–786.
20. Frasoldati A, Flora M, Pesenti M, Caroggio A, Valcavil R. Computer-assisted cell morphometry and ploidy analysis in the assessment of thyroid follicular neoplasms. Thyroid 2001;11:941–946.
21. Murata SI, Mochizuki K, Nakazawa T, Kondo T, Nakamura N, Yamashita H, Urata Y, Ashihara T, Katoh R. Detection of underlying characteristics of nuclear chromatin patterns of thyroid tumor cells using texture and factor analyses. Cytometry 2002;49:91–95.
22. Tsai TH, Chang TC, Chiang CP. Nuclear area measurements of parathyroid adenoma, parathyroid hyperplasia and thyroid follicular adenoma, a comparison. Anal Quant Cytol Histol 1997;19:45–48.
23. Kayser K, Hoshang SA, Metze K, Goldmann T, Vollmer E, Radziszowski D, Kosjerina Z, Mireskandari M, Kayser G. Texture- and object-related automated information analysis in histological still images of various organs. Anal Quant Cytol Histol 2008;30:323–335.
24. Tsantis S, Dimitropoulos N, Cavouras D, Nikiforidis G. Morphological and wavelet features towards sonographic thyroid nodules evaluation. Comput Med Imaging Graph 2009;33:91–99.
25. Albregtsen F, Nielsen B, Danielsen H. Adaptive gray level run length features from class distance matrices. Int Conf Pattern Recognit 2000;3:3746–3749.
26. Baheerathan S, Albregtsen F, Danielsen H. New texture features based on complexity curve. Pattern Recognit 1999;32:605–618.
27. Kerenji A, Bozovic Z, Tasic M, Budimlija Z, Klem I, Polzovic A. Fractal dimension of hepatocytes' nuclei in normal liver vs hepatocellular carcinoma (hcc) in human subjects - preliminary results. Arch Oncol 2000;8:47–50.
28. Tsai HW, Tsai HH, Kuo FY, Chang KC. Computerized analyses of morphology and proliferative activity differentiate hepatoblastoma from paediatric hepatocellular carcinoma. Histopathology 2009;54:328–336.
29. Ikeguchi M, Sato N, Hirooka Y, Kaibara N. Computerized nuclear morphometry of hepatocellular carcinoma and its relation to proliferative activity. J Surg Oncol 1998;68:225–230.
30. Pereira RR, Marques PMA, Honda MO, Kinoshita S, Engelmann R, Muramatsu C, Doi K. Usefulness of texture analysis for computerized classification of breast lesions on mammograms. J Digit Imaging 2007;20:248–255.
31. Borst H, Abmayr W, Gais P. A thresholding method for automatic cell image segmentation. J Histochem Cytochem 1979;27:180–187.
32. Lee KM, Street WN. A fast and robust approach for automated segmentation of breast cancer nuclei. In: Proceedings of the 2nd IASTED International Conference on Computer Graphics and Imaging.; 1999. pp 42–47.
33. Adiga U, Malladi R, Ferndandez-Gonzalez R, de Solorzano CO. High-throughput analyzis of multi-spectral images of breast cancer tissue. IEEE Trans Imaging Proc 2006;15:2259–2268.
34. Cross SC. Fractals in pathology. J Pathol 1997;182:1–8.
35. Roula MA, Bouridane A, Miller P. Multispectral analysis of chromatin texture for automatic grading of prostatic neoplasia. In: Proceedings of Medical Image Understanding and Analysis, 2002; Portsmouth, UK.
36. Young IT, Verbeek PW, Mayall BH. Characterization of chromatin distribution in cell nuclei. Cytometry 1986;7:467–474.
37. Choi HK, Jarkrans T, Bengtsson E, Vasko J, Wester K, Malmström PU, Busch C. Image analysis based grading of bladder carcinoma, Comparison of object, texture and graph based methods and their reproducibility. Anal Cell Pathol 1997; 15:1–18.
38. Esgiar AN, Naguib RNG, Sharif BS, Bennett MK, Murray A. Fractal analysis in the detection of colonic cancer images. IEEE Trans Inf Technol Biomed 2002;6:54–58.
39. Wolberg W, Street W, Heisey D, Mangasarian O. Computer-derived nuclear features distinguish malignant from benign breast cytology. Hum Pathol 1995;26:792–796.
40. Smolle J. Computer recognition of skin structures using discriminant and cluster analysis. Skin Res Technol 2000;6:58–63.
41. Wiltgen M, Gerger A, Smolle J. Tissue counter analysis of benign common nevi and malignant melanoma. Int J Med lnform 2003;69:17–28.
42. Yang L, Tuzel O, Chen W, Meer P, Salaru G, Goodell LA, Foran DJ. Pathminer: A web-based tool for computer-assisted diagnostics in pathology. IEEE Trans Inf Technol Biomed 2009;13:291–299.
43. Huang PW, Lee CH. Automatic classification for pathological prostate images based on fractal analysis. IEEE Trans Med Imaging 2009;28:1037–1050.
44. Kong J, Sertel O, Shimada H, BOyer KL, Saltz JH, Gurcan MN. Computer-aided evaluation of neuroblastoma on whole slide histology images: classifying grade of neuroblastic differentiation. Pattern Recognit 2009;42:1080–1092.
45. Tahir MA, Bouridane A. Novel round-robin tabu search algorithm for prostate cancer classification and diagnosis using multispectral imagery. IEEE Trans Inf Technol Biomed 2006;10:782–793.
46. Yang L, Chen W, Meer P, Salaru G, Goodell L, Berstis V, Foran D. Virtual microscopy and grid-enabled decision support for large scale analysis of imaged pathology specimens. IEEE Trans Inf Technol Biomed (in press).
47. Bauer E, Kohavi R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Mach Learn l999;36:105–139.
48. Salmon I, Kiss R, Franc B, Gasperin P, Heimann R, Pasteels JL, Verhest A. Comparison of mor-phonuclear features in normal, benign and neoplastic thyroid tissue by digital cell image analysis. Anal Quant Cytol Histol 1992;14:47–54.
49. Salmon I, Gasperin P, Pasteels JL, Heimann R, Kiss R. Relationship between histopathologic typing and morphonuclear assessments of 238 thyroid lesions. Digital cell image analysis performed on feulgen-stained nuclei from formalin-fixed, paraffin-embedded materials. Am J Clin Pathol 1992;97:776–786.
50. Boyokov Y, Funka-Lea G. Graph cuts and efficient n-d image segmentation. Int J Comp Vis 2006;70:109–131.
51. Li C, Huang R, Ding Z, Gatenby C, Metaxas D, Gore J. A variational level set approach to segmentation and bias correction of images with intensity inhomogeneity. Med Image Comput Comput Assist Interv Int Conf 2008;11(Part 2):1083–1091.
52. Rohde GK, Wang W, Peng T, Murphy RF. Deformation-based nonlinear dimension reduction: Applications to nuclear morphometry. In: Proceedings of IEEE International Symposium Biomedical Imaging; 2008. pp 500–503.
53. Rohde G, Ribeiro AJS, Dahl KN, Murphy RF. Deformation-based nuclear morphometry: Capturing nuclear shape variation in HeLa cells. Cytometry Part A 2008;73A:341–350.
54. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning; data mining, inference, and prediction. New York: Springer; 2001.
55. Bishop CM. Pattern Recognition and Machine Learning (Information Science and Statistics). New York: Springer; 2006.
56. Chang CC, Lin CJ. Libsvm : A library for support vector machines. Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001.
57. Cox T, Cox M. Multidimensional Scaling. London: Chapman and Hall, 1994.
58. Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948;27:379–423.
59. Chen YW, Hu KH, Li FZ, Li SY, Su WF, Huang ZY, Hu YX. Studies on quantitative analysis and automatic recognition of cell types of lung cancertudy on quantitative analysis and automatic recognition of cell types of lung cancer. Biomed Mater Eng 2006;16:119–128.
60. Padfield D, Chen B, Roysam H, Cline C. Lin G, Seel M. Cancer tissue classification using nuclear feature measurements from dapi stained images. In: Proceedings of 1st Workshop on Microscopic Image Analysis with Applications in Biology, 2006. pp 86–92.
61. Haralick R, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybernet 1973; SMC-3:610–621.
62. Boland MV, Murphy RF. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of hela cells. Bioinformatics 2001;17:1213–1223.
63. Huang K, Murphy RF. Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. BMC Bioinformatics 2004;5:78.
64. Manjunath BS, Ma WY. Textural features for browsing and retrieval of image data. IEEE Trans Pattern Anal Mach Intell 1996;8:837–842.

Appendix: Numerical Features Used

We used a total of 125 popular numerical features in our attempts to quantify nuclear structure:

- Six shape features (59,60): These shape features are widely used in distinguishing between normal and cancer cells in digital pathology. In our study we included the following: area, convexity, circularity, perimeter, eccentricity, and equivalent diameter. They can be computed based on the binarized object image.
- Three additional texture features (59,60): These additional textures are related to the pixel intensities of each images. We included average intensity, standard deviation, and entropy.

- 26 Haralick features (9): Based on vertical directions (+90, +270) and horizontal directions (+0, +180) with 1 pixel offset, 4 gray-level co-occurrence matrices were computed for each images. Then 13 statistics were computed based on each gray-level co-occurrence matrix. They include Haralick angular second moment, contrast, entropy, sum entropy, information measure of correlation 1 and 2, correlation, sum variance, difference variance, sum average, sum of squares, difference entropy and Haralick inverse difference moment. Therefore, 52 statistics were computed for each images. The 26 Haralick features were computed by averaging statistics for vertical directions and horizontal directions. More detailed explanations can be found in (61,62).
- 30 Daubechies 4 wavelet features (63): These features were calculated from a 10 level 10th wavelet decomposition using both scaling and wavelet functions of the Daubechies 4 wavelet transformation. We decomposed an image column-wise and row-wise sequentially, then obtained four small images, three of which captured the high-frequency information. The average energy of these three high-frequency components were taken as features at each level. Therefore, there were 3×10 features computed by decomposing an image 10 times.
- 60 Gabor features (64): Given a Gabor filter with scale m and orientation n, we can convolve an image with this filter and compute the mean and standard deviation of the resulting image as texture features. We used Gabor texture feature extractor according to (64) and used the default parameters to construct a Gabor wavelet bank with five different scales and six different orientations. Then 60 Gabor features were computed for each images.