Published in final edited form as:

Med Image Anal. 2014 July ; 18(5): 772–780. doi:10.1016/j.media.2014.04.004.

Accurate diagnosis of thyroid follicular lesions from nuclear morphology using supervised learning

John A. Ozolek^a, Akif Burak Tosun^{b,*}, Wei Wang^b, Cheng Chen^b, Soheil Kolouri^b, Saurav Basu^b, Hu Huang^b, and Gustavo K. Rohde^{b,c,d}

^a Department of Pathology, Children's Hospital of Pittsburgh, Pittsburgh, PA 15224, USA

^b Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

^c Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

^d Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract

Follicular lesions of the thyroid remain significant diagnostic challenges in surgical pathology and cytology. The diagnosis often requires considerable resources and ancillary tests including immunohistochemistry, molecular studies, and expert consultation. Visual analyses of nuclear morphological features, generally speaking, have not been helpful in distinguishing this group of lesions. Here we describe a method for distinguishing between follicular lesions of the thyroid based on nuclear morphology. The method utilizes an optimal transport-based linear embedding for segmented nuclei, together with an adaptation of existing classification methods. We show the method outputs assignments (classification results) which are near perfectly correlated with the clinical diagnosis of several lesion types' lesions utilizing a database of 94 patients in total. Experimental comparisons also show the new method can significantly outperform standard numerical feature-type methods in terms of agreement with the clinical diagnosis gold standard. In addition, the new method could potentially be used to derive insights into biologically meaningful nuclear morphology differences in these lesions. Our methods could be incorporated into a tool for pathologists to aid in distinguishing between follicular lesions of the thyroid. In addition, these results could potentially provide nuclear morphological correlates of biological behavior and reduce health care costs by decreasing histotechnician and pathologist time and obviating the need for ancillary testing.

Keywords

Nuclear morphology; Image analysis; Thyroid follicular lesions; Optimal transport

1. Introduction

Follicular lesions of the thyroid (to include adenomatous (hyperplastic) nodules, nodular goiter (NG), follicular adenoma (FA), follicular carcinoma (FC), and follicular variant of papillary carcinoma (FVPC)) remain significant diagnostic challenges in surgical pathology (Baloch and LiVolsi, 2002; LiVolsi and Baloch, 2004). From surgical resection specimens, parsing of these differential diagnoses often requires considerable time to carefully examine many tissue sections and occasionally, but with increasing frequency, the addition of immunohistochemistry, molecular studies, and at times expert consultation. However, some ancillary tests are not widely available or specific, and augment the total cost of health care. While FVPC has familiar nuclear morphological features that are helpful diagnostically, these features are not specific and not always present uniformly. In addition, concerning other follicular lesions of the thyroid, visual interpretation of nuclear features are not particularly helpful in diagnostic discrimination.

The nucleus represents a wealth of pixel data because of the myriad of molecular components and organizational structures contained within (e.g. DNA (deoxyribonucleic acid), RNA (ribonucleic acid), proteins, nucleosomes, nuclear matrix, nucleoli, chromatin complexes, nuclear lamina) (Stein et al., 2000; Zink et al., 2004). Many of these components can be stained and imaged in tissue sections. Changes in chromatin structure, constitution, and location can initiate changes in nuclear morphology (e.g. size, shape, contour) and conversely changes in nuclear morphology can alter gene expression in the course of cancer genesis and tumor progression (Stein et al., 2000; Zink et al., 2004). These changes can be seen in gross fashion visually under the microscope. Although rarely seen in follicular lesions of the thyroid (Nikiforov, 2008; Nikiforov et al., 2012), RET/PTC (rearranged in transformation/papillary thyroid carcinomas) transformation of primary cell cultures of normal human thyroid epithelial cells results in nuclear shape irregularities and heterochromatin dispersion (Zink et al., 2004; Fischer et al., 1998a, 2003). RAS (Rat sarcoma) mutations are more commonly found in the follicular variant of papillary carcinoma and activated HRAS (a type of RAS gene) expression results in chromatin coarsening and correlates with metastatic potential (Fischer et al., 1998b; Nikiforov, 2008; Nikiforov et al., 2012). Gene expression analyses of thyroid tumors shows differential expression of genes between normal thyroid and thyroid tumors (Chevallard et al., 2004; Finley et al., 2005; Nikiforov, 2008; Nikiforov et al., 2012; Zhao et al., 2008). Interestingly, Chevallard et al. (2004) demonstrated only 43 genes with differential expression between FA and FC, perhaps paralleling the visual similarity of their nuclei.

Previous designs of studies of nuclear morphometry by image analysis in thyroid follicular lesions have used a set of *a priori* selected numerical features for the necessary comparisons (e.g. nuclear size, area, perimeter, maximum/minimum diameter, major axis length, longest axis/shortest axis ratio) (Aiad et al., 2009; Frasoldati et al., 2001; Gupta et al., 2001; Karsho lu et al., 2005). A few studies have employed chromatin and texture features in addition to morphometric parameters (maximum approximately 30 features) and compared using higher levels of analysis (Karakitsos et al., 1996; Murata et al., 2002; Shapiro et al., 2007). The highest degrees of accuracy in classification (80–90%) have been achieved using higher levels of analysis for select diagnostic challenges (e.g. neural networks with training

data). We have previously shown with a very limited number of cases of FA, FC, and normal thyroid (NL) that by selecting many features (125 in total) the method was able discriminate between these three entities, in a total of 10 patients, with 100% accuracy when comparing groups of nuclei (not single nuclei) (Wang et al., 2010).

Here we describe a method for classifying tissue samples from different patients using nuclear morphology that is highly correlated with the available clinical diagnosis gold standard. The processing pipeline takes as input light microscopy images of stained tissue sections containing the representative lesion and outputs a label (class) for the lesion based on the comparison of the extracted nuclei to a ‘training’ set of labeled nuclei via a specially tailored supervised learning classification technique. Rather than utilizing the commonly used numerical feature approach described above, our method is based on comparing segmented nuclei using a linearized version of the well-known optimal transport between two distributions (Wang et al., 2013). Classification is performed by applying the K-nearest neighbor algorithm on a discriminant subspace extracted using a modified version of the Fisher linear discriminant analysis technique Wang et al. (2011a). We show this method can outperform traditional numerical feature-based approaches for comparing nuclei, and can achieve very high accuracy in a cohort of 94 patients extracted from the archives of the University of Pittsburgh Medical Center. In addition, we show the approach can be used to visualize interesting differences in nuclear morphology between different lesion types.

2. Materials and methods

2.1. Tissue processing and imaging

Tissue blocks were obtained from the archives of the University of Pittsburgh Medical Center (approved as an exempt protocol by the Institutional Review Board of the University of Pittsburgh). Cases for analysis included resection specimens with the diagnosis of nodular goiter (NG, $n = 28$), follicular adenoma of the thyroid (FA, $n = 27$), follicular carcinoma of the thyroid (FC, $n = 20$), follicular variant of papillary carcinoma (FVPC, $n = 10$), and widely invasive follicular carcinoma (WIFC, $n = 9$). More information regarding the distribution of patient data is provided in Table 1. These groups were chosen since they represent the groups that usually require considerable resources and effort to distinguish. All cases were reviewed by at least two pathologist(s) who either specialize in thyroid pathology or head and neck pathology (at the time of diagnosis) and the study pathologist (J.A.O.) who selected appropriate representative blocks for staining and image acquisition.

Tissues were procured at the time of a surgical procedure retrospectively over an eleven year span (1999–2009). All tissues were fixed in 10% neutral buffered formalin and processed on a conventional tissue processor. Tissue sections were cut at 5 μm thickness from the paraffin-embedded block and placed on conventional Superfrost Plus microscope slides using Fisherbrand Superslip cover slips (0.17 mm thickness; Fisher Scientific, Thermo Fisher Scientific, Inc., Waltham, MA). All tissue sections for imaging were stained using the Feulgen technique. Counterstaining was omitted primarily to achieve better automated nuclear segmentation at 1000 $\times$ magnification. The Feulgen technique is a Schiff-based histochemical stain whereby the Schiff reagent binds exposed aldehyde groups thus staining DNA exclusively imparting nuclei with a magenta color (Bancroft and Gamble, 2008) (Fig.

1) by light microscopy. Feulgen staining has precedent in studies using computational image analysis of nuclear morphometry (Nielsen et al., 2012; Salmon et al., 1992a,b).

All images were acquired using an Olympus *BX51* microscope equipped with a 100× UIS2 UPlanFl oil immersion objective (numerical aperture 1.30; Olympus America, Central Valley, PA) and 2 megapixel SPOT Insight camera (Diagnostic Instruments, Sterling Heights, MI). Image specifications were 24 bit RGB channels and 0.074 $\mu\text{m}/\text{pixel}$, 118 μm , 89 μm field of view. Slides were chosen that contained both lesion and adjacent normal appearing thyroid tissue (NL) where possible. The pathologist (J.A.O.) took 10–30 random images from non-overlapping fields of view of lesion and/or normal thyroid to guarantee that the data for each patient contained adequate numbers of nuclei for subsequent analysis.

2.2. Nuclear segmentation

Nuclei were segmented using an existing supervised learning approach described in Chen et al. (2012, 2013). Briefly, a training set of nuclei selected from the dataset is first utilized to construct a model encompassing the variations in shape and texture expected in the dataset. Segmentation of any nucleus from images of the same kind was performed by finding the model parameters that best match the regions chosen by the pathologist for segmentation. The segmented nuclei are stored in a database of individual images, each containing one carefully segmented nucleus. The database was filtered to discard lesional nuclei which were smaller in area than the minimum area of a set of sample nuclei extracted from non-lesional (normal) nuclei from a similar set of patients. This is done to avoid potentially miss segmented nuclei given that, in practice, nuclei belonging to any thyroid lesion are not expected to be smaller than the smallest nucleus in normal tissue. Finally, the database was visually inspected by the pathologist to include only nuclei of lesional cells (or normal thyroid epithelial cells) that were accurately segmented and had sharp distinct nuclear membranes (in focus). In total, 2647 nuclei were selected for classification from a total of 94 patients for this study. The total numbers of extracted nuclei for each class were: 840 nuclei for NG, 756 nuclei for FA, 509 for FC, 300 for FVPC, and 242 for WIFC. Fig. 1 contains a few examples of segmented nuclei (second row), as well as the corresponding nuclei after our normalization (third row), which is described in more detail below.

2.3. Extracting a linear embedding

The images corresponding to each segmented nuclei were first normalized so as to remove the effects of random translation, and orientation as described in Rohde et al. (2008). In addition, the intensities of each nuclear images were also normalized as described in Wang et al. (2011b) so that the pixel brightness indicated the amount of locally concentrated chromatin (the brighter the pixel, the more chromatin), and so that the sum of all pixels' intensities in a given image was set to 1. This helps ensure that slight variations in tissue acquisition and staining normally encountered do not interfere with our ability to quantify the chromatin content of each segmented nucleus. After normalization, the chromatin content of each segmented nucleus was processed to obtain its linear optimal transport (LOT) embedding, as described in detail in Wang et al. (2013). Fig. 2 contains a graphical description of the idea. A linear (Euclidean) embedding for each nucleus using a tangent plane mapping of the Riemannian optimal transport manifold centered around a 'reference'

nucleus. To that end, the optimal transport plan between each nucleus and the reference is computed and, in conjunction with the ‘mass’ (collection of pixel intensity values) from the reference image, is used as the linear embedding for the corresponding nucleus. One of the major benefits of this technique is a dramatic reduction in computational complexity when computing pairwise transport distances between images in a dataset.

While a detailed explanation of the LOT approach is available in Wang et al. (2013), for completeness, here we offer the following summary. Briefly, a Euclidean average of intensities across the entire image dataset (after translation and rotation effects have been removed) is used as a reference template. That is, let the dataset be composed of N images

$I_1, I_2, \dots, I_n$. The reference image is computed as $I_0 = 1/N \sum_{i=1}^N I_i$. A particle approximation for each image (as well as the template or reference image) is computed as described in Wang et al. (2013). Briefly, the particle approximation is based on a weighted K-means approach to summarize the intensity values of each image. The purpose of the particle approximation is to reduce the overall computational complexity of the procedure. In this work, we have chosen $M = 800$ particles to be used for approximating each image. The output for each image is the position of each particle, as well as the weight (mass) of each particle in approximating the corresponding image. For example, the approximation for the reference image be written as $I_0 = \sum_{k=1}^M q_k \delta_{z_k}$ where δ_{z_k} corresponds to a discrete delta function placed at position z_k in that image, while q_k corresponds to the ‘mass’ at that image (computed as described in Wang et al. (2013)). Likewise, let $I_1 = \sum_{j=1}^M m_j \delta_{y_j}$ be the particle approximation of another image in the database. To obtain the linear embedding of image I_1 we solve for the optimal transport between the particle approximations of I_0 and I_1 by minimizing

$$d^2(I_0, I_1) = \min_{f \in \Pi(\mu, \nu)} \sum_{i=1}^M \sum_{j=1}^M |z_i - y_j|^2 f_{i,j} \quad (1)$$

subject to the constraints $\sum_{j=1}^M f_{i,j} = m_i$ for $i = 1, \dots, M$, and $\sum_{i=1}^M f_{i,j} = q_j$ for $j = 1, \dots, M$. This minimization is performed utilizing the linear programming approach described in Wang et al. (2013). The linear embedding of I_1 is obtained by applying the discrete transportation map between the reference I_0 and I_1 to the coordinates y_j via

$$\mathbf{x}_1 = \left(\sqrt{q_1} a_n^1 \cdots \sqrt{q_{N_\sigma}} a_n^M \right)^T \quad (2)$$

where a^k is the centroid of the particle $q_k \delta_{z_k}$ computed via

$$a_n^k = \sum_{j=1}^{N_{v_n}} f_{k,j} y_j / q_k \quad (3)$$

We denote $\mathbf{x}_1$ to be the linear optimal transport (LOT) embedding of I_1 . This embedding has dimensions $\mathbf{x}_1 \in \mathbb{R}^{M \times 2}$ for two dimensional images.

2.4. Image classification

The next step in our training procedure relies on computing a one dimensional linear discriminant subspace that can be used for classifying sets of nuclei. Here we opt towards applying the standard Fisher Linear Discriminant Analysis (LDA) technique (Bishop and Nasrabadi, 2006) on the LOT embedding of each labeled nucleus in the training set

computed as described above. That is, let $S_T = \sum_n (\mathbf{x}_n - \bar{\mathbf{x}}) (\mathbf{x}_n - \bar{\mathbf{x}})^T$ represent the ‘total scatter matrix’, $S_W = \sum_c \sum_{n \in c} (\mathbf{x}_n - \bar{\mathbf{x}}_c) (\mathbf{x}_n - \bar{\mathbf{x}}_c)^T$ be the ‘within class scatter matrix’, and $\bar{\mathbf{x}}_c$ be the center of class c . The LDA technique computes a projection line $\mathbf{w}_{LDA}^*$ as the solution of the following minimization problem:

$$\mathbf{w}_{LDA}^* = \arg \max_{\mathbf{w}} \frac{\mathbf{w}^T S_T \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}} \quad (4)$$

The solution of the minimization problem above is computed via the solution of an associated generalized eigenvalue problem

$$S_T \mathbf{w} = \lambda S_W \mathbf{w} \quad (5)$$

Given a set of LOT embeddings $\mathbf{x}_1, \dots, \mathbf{x}_N$ from a given class, the projection of each nucleus is computed via $\mathbf{x}_k^T \mathbf{w}_{LDA}^*$ and stored as training data. Given a set of LOT embeddings from nuclei from an unlabeled patient $\mathbf{y}_1, \dots, \mathbf{y}_K$, the projection of each nucleus $\mathbf{y}_k^T \mathbf{w}_{LDA}^*$ is similarly computed. The class of the unknown set is computed by classifying each $\mathbf{y}_1$ via a K-nearest neighbor algorithm (see Bishop and Nasrabadi, 2006 for details), and taking the majority ‘vote’ (the most common class assignment). The appropriate number of nearest neighbor K is computed with a blind cross validation procedure.

2.5. Cross validation

In order to test the ability of our method to classify patients whose nuclei were not used in the training process, we utilized the standard ‘leave one patient out’ cross validation strategy (Bishop and Nasrabadi, 2006). That is, in order to classify a certain patient, the nuclei belonging to that patient are removed from the training dataset and kept as testing data. A classifier is then trained using training data only. The trained classifier is used to compute the projection score of the held out patient's nuclei. In all classification accuracy tables shown in the results section below, each patient in the diagnostic challenge was diagnosed by using the LDA-based K-nearest neighbor voting classifier (computed as described above) trained without using nuclei from that patient (as detailed earlier). In each training and classification task, care was taken to ensure that the same number of nuclei (selected at random) for each class was included in computing the classifier. In summary, no testing data is used in the training process, while the LOT embedding procedure above is the same for every image, and does not utilize class labels. Classification accuracy tables were computed for the following diagnostic challenges of interest to practitioners in this area: FA vs. FC, FA vs. FVPC, FC vs. WIFC, FC vs. FVPC, FA vs. NG, and FC vs. NG. Using this

paradigm, we are able to compare benign vs. malignant as well as malignant vs. malignant lesions.

2.6. Visualizing differences in nuclear structure distributions

The LOT embedding not only allows for quantitative comparisons between nuclear images, but when combined with a penalized version of the LDA method described in Wang et al. (2011a), also enables visualization of the most relevant and statistically significant differences between two sets of images (Wang et al., 2013). For each diagnostic challenge, we also sought to display and learn what statistically significant differences exist between nuclei of the different classes (e.g. benign vs. malignant). For this, the LDA technique is applied to the linear embedding obtained from the LOT framework. As shown in Wang et al. (2011a), maximizing the LDA criterion is closely related to finding the projection that maximizes the Student's t statistic. Thus, the line that enables classification in LOT space can be used for visualizing differences between the classes. The only caveat is the LDA must be modified to ensure the differences displayed are adequately representative (as discussed in Wang et al. (2011a, 2013)). Nuclei generated by the LOT embedding (i.e. geodesic paths through the LOT manifold) are represented in the figures as the bottom rows below the histograms shown in Figs. 3–8. These nuclei represent what the nuclei would look like as they traverse along the path in geometric space that best discriminates one distribution of nuclei from another (see Fig. 2 of Wang et al. (2011b)). In these figures, we display this information using grayscale, as well as a color coding scheme aimed to mimic nuclei in histopathology images. The color images are obtained by taking the negative of the grayscale images shown on the top row, renormalizing their range to 255, and using a colormap for displaying them. The p -value for the Student's t -test reported in each figure was computed using ‘held out’ data, in a cross validation procedure using half the available data for training, and half for testing.

3. Results

Pertinent pathologic and clinical characteristics of the thyroid follicular lesions are represented in Table 1. Pathological and clinical data were compared using one-way unstacked ANOVA using Tukey method of comparison and performed using MiniTab 16 statistical software (Minitab, Inc., State College, PA). Not unexpectedly, WIFC lesions were significantly larger than FA, FC, or FVPC ($p = 0.001$) while no significant differences in lesion size were noted between FA, FC, or FVPC. Patients with NG and WIFC were significantly older than patients with FC ($p = 0.006$), but not significantly older than patients with FA or FVPC.

Table 2 contains a summary of the classification results obtained for each diagnostic challenge. Results utilizing the LOT-based approach are shown in bold face. For comparison purposes, here we have also included results obtained utilizing a numerical feature-based approach similar to the approach described earlier in Wang et al. (2010). In total, 256 numerical features, which included morphological and texture features, were extracted per nucleus. Since nuclear morphological features are widely used in discriminating cancer cells from normal ones in image analysis in digital pathology (Mahfouz et al., 2012), we extracted

six of the most popular features in our experiment: area, convexity, circularity, perimeter, eccentricity and equivalent diameter. Additionally, we computed three intensity-derived features (average intensity, standard deviation, and entropy), Haralick features and Gabor features as described in Haralick et al. (1973), Boland and Murphy, 2001, and Wang et al., 2010. Using these approaches we computed 220 texture features in total for each nuclear image. Finally, 30 wavelet-based features were computed as described in Wang et al. (2010) and added to the set. Features were normalized so that the variance of each feature was set to one. The same 'leave one patient out' cross validation strategy is applied to train the classifier. The training process consisted of applying the stepwise discriminant analysis (SDA) (Jennrich and Sampson, 1977) technique, and subsequently applying the same LDA-based classification procedure described above. In the comparisons below, the training and testing procedures utilized identical data for both LOT and numerical feature-based approaches.

All cases of FA were distinguished from cases of NG, FC and FVPC on the basis of nuclear morphology alone with 100% accuracy in pairwise comparisons utilizing the LOT-based approach. The feature-based approach, on the other hand, performs well, though not with the same accuracy. Likewise, the approach we described above is able to perfectly distinguish FC from NG and WIFC on this data, while the numerical feature-based approach makes 7 erroneous assignments in a total of 29 patients for FC vs. WIFC and 9 erroneous assignments in a total of 48 patients for FC vs. NG. Finally, differentiating FC from FVPC turned out to be the most difficult challenge. Here the LOT approach classified 87% of the cases correctly, while the numerical feature approach classified 90% of the cases correctly. Table 3 contains the statistical analysis (Kappa statistic, p -values) for each of the classification challenges. For assessing the significance of the classification accuracies reported we have used Cohen's Kappa (Cohen, 1960). This statistic measures the agreement between two classifiers (in our case the gold standard and the output of our method) and takes into account the agreement occurring by chance. According to our results, we can conclude that the LOT approach was able to perfectly agree with the pathologist's ground truth labels except for FVPC vs. FC challenge, for which the agreement is still substantial. In all cases involving data analysis using the LOT approach, the statistical significance of the result (assessed through the p value) is significant at the $p = 0.005$.

As described above, we also utilized the LOT-embedding in combination with a penalized version of the LDA classifier in order to visualize meaningful differences between the nuclei in several diagnostic challenges. Comparing FA with FC, nuclei of FA showed more uniform and denser chromatin distribution (Fig. 3, thin arrows). FC nuclei tended to have chromatin concentrated around the nuclear membrane with punctate dense concentrations within the central region of the nucleus (thick arrows). FC nuclei were significantly larger, had more random distribution of chromatin, and greater amounts of chromatin at the periphery compared to FA.

Not unexpectedly, the distribution histograms were more widely separated between FA and FVPC compared to FA/NG and FA/FC distributions. As noted, greater numbers of FA nuclei had more uniform chromatin distribution with centralized areas of fine to coarse stippling (Fig. 4, thin arrows). Greater numbers of FVPC nuclei had central clearing and

peripheralization of chromatin compared to FA (thick arrows). The tails of the FA and FVPC distributions clearly show that more cancer nuclei had very open chromatin patterns with occasional central stippling and a thin rim of peripheral chromatin, while FA nuclei maintained the more concentrated central chromatin with peripheral chromatin present but not as sharply defined as in FVPC. Nuclei of FA were significantly smaller with less random chromatin distribution but more peripheral chromatin than FVPC.

In similar fashion, and not unexpectedly, all cases of FC were distinguished from cases of WIFC with 100% accuracy in pairwise comparisons (Fig. 5). A wide separation of distributions was seen between the means of FC and WIFC. Greater numbers of FC nuclei had a relatively homogeneous appearance with diffuse chromatin distribution and only slight central clearing (Fig. 5, thick arrows). Nuclei of WIFC had a more heterogeneous appearance with gradual central clearing, coarse clumped chromatin, and chromatin peripheralization (thin arrows). WIFC nuclei, as expected based on visual inspection, were significantly larger with greater random chromatin distribution and greater peripheral chromatin than FC.

Results for the FC vs. FVPC challenge are shown in Fig. 6. The distribution histograms had significant p -values, although the means appear close together. The generated nuclei have similar appearances and chromatin distributions even at the tail of both distributions (Fig. 6). Greater numbers of FVPC nuclei (thin arrows) appear to show more discrete coarse chromatin located centrally compared to FC which had very concentrated areas of chromatin peripherally and centrally (thick arrows). FVPC nuclei were significantly larger, had greater random chromatin distribution, but less peripheral chromatin than FC.

The geodesics that distinguished FA best from NG are shown in Fig. 7 below the histogram bars. Greater numbers of FA nuclei compared to NG nuclei had appearances that ranged from more diffusely coarse chromatin to loss of chromatin at the membrane border with a few central chromatin concentrations (thin arrows). Greater numbers of NG nuclei had a more open central chromatin with peripheral distinction (thick arrows). FA had significantly greater nuclear area, greater random distribution of chromatin, and greater peripheral distribution of chromatin than NG.

Finally, the geodesics that distinguished FC from NG showed that greater numbers of NG nuclei had central clumped chromatin without a distinctive rim of peripheral chromatin (Fig. 8, thin arrows). Greater numbers of FC nuclei demonstrated a more heterogeneous appearance ranging from more evenly distributed chromatin with some stippling to very open central chromatin with a distinct border (thick arrows). FC nuclei were significantly larger, had more random chromatin distribution, but similar peripheral chromatin distribution compared to NG.

4. Summary and conclusions

We have described a novel image analysis technique for analyzing nuclear morphology to classify thyroid follicular lesions from routinely processed tissue and images obtained using a suitably equipped clinical brightfield microscope (standard equipment in the offices of practicing pathologists). The technique we described combines a transport-based linear

embedding for each image together with simple majority voting strategy based on a K-nearest neighbor classifier obtained utilizing a linear discriminant subspace. Overall, the approach was shown to achieve correlation with the known clinical diagnosis in a total of 94 patients. In addition, the approach is also well suited for visual interpretation of nuclear morphology differences in different diagnostic challenges.

We demonstrated the ability to separate several important thyroid follicular lesions including FA/FC, FA/FVPC, FC/WIFC, FA/NG, and FC/NG with 100% accuracy (according to the known clinical diagnosis gold standard) based on digital images of groups of nuclei from 94 patients in total. The distinction of FA from FC with 100% accuracy based on nuclear morphology alone was encouraging as this remains a diagnosis that requires more extensive histology and pathology resources to render because visual determination of differences in nuclear morphology cannot be done accurately. These results demonstrate that the nuclear morphology is significantly different between diagnostic groups to allow their distinction. The selection of thyroid follicular lesions to test this methodology stems not only from the previously noted challenges for pathologists, but also that one group of lesions (FVPC), in most cases, shows well-documented visible nuclear features associated with papillary carcinoma. We presumed, while recognizing the visible nuclear morphological heterogeneity that can occur in FVPC, that we would have a ‘positive control’ for the methodology assuming that this group of lesions (and their nuclei) would more readily segregate from the other follicular lesions. This was true for the FA/FVPC comparison in which these two entities were segregated with 100% accuracy and had widely separated means for their distributions.

However, we note that, FVPC had cases misclassified as FC for both feature and LOT approaches. Upon reviewing the misclassified cases of FVPC as FC, it is noteworthy that several of the cases lacked all of the distinctive nuclear features associated with papillary carcinoma and the nuclear features were not seen in all cells throughout the lesion. The FVPC tumors selected for analysis had visual evidence of all characteristic and documented nuclear features of papillary carcinoma of the thyroid that include central nuclear clearing, nuclear overlap, nuclear grooves, nuclear enlargement, and nuclear contour irregularities. This misclassification of a few cases is not without a plausible biological corollary since FVPC does not consistently and uniformly exhibit all diagnostic nuclear features of papillary carcinoma (Baloch and LiVolsi, 2002; Nikiforov et al., 2012). Some experts propose nuclear enlargement and nuclear contour irregularities are the minimum criteria necessary to consider FVPC. Some believe that FVPC progresses from a benign to malignant phenotype in stages with ‘mature’ lesions showing the greatest number of nuclear characteristics associated with papillary carcinoma. Thus, the nuclear phenotype may depend on tumor age (tumor size) at the time of removal (Nikiforov et al., 2012). A limitation of these approaches in general may occur with lesions that have both inter- and intranuclear morphological heterogeneity. This study did not specifically address the impact of lesion size on classification accuracy.

We also note that while we showed experimentally that the LOT-based approach can yield high classification accuracies, theoretical justification for this improvement is the subject of current investigation. We note, however, that the LOT framework, in theory, can be

considered as an invertible nonlinear transform (Wang et al., 2013). That is, given enough particles to approximate each image, the process of embedding that image in LOT domain is reversible. The embedding is computed by ‘matching’ each image in a database to a chosen reference image, through the optimal transport framework. This ‘matching’ operation is possible even without the specification (or existence) of corresponding anatomical landmarks in the images (as in the case of the nuclei images used here). In contrast, the process of numerical feature extraction, generally speaking, is not reversible and thus inherently involves information loss which could explain, in part, the increase in classification accuracy demonstrated above.

As important as the high classification accuracies obtained using nuclear morphology alone, we are able to generate actual representations of nuclei that represent the distributions compared. Having this information allows one to examine entire distributions and choose elements or features that can then be compared statistically. Another important discrimination seen by this visual representation is that many lesions had most of their nuclei having similar appearance (overlapping means spanning two or three geodesics) while other lesions had significant tails on their distributions representing up to 5 geodesics. This seemed to be present in several of the benign/malignant comparisons with the malignant lesions having seemingly greater numbers of nuclei extending out in the tails of the distributions (FA/FVPC, and FA/FC). This may reflect molecular genetic changes that produce significantly different morphological changes in the nucleus.

In addition, we have also studied how many nuclei per patient were necessary to achieve the results shown above. Comparisons that yielded 100% accuracy classifications required only 10–15 nuclei to achieve this classification accuracy. For the comparison with less than 100% classification accuracy, 20 nuclei achieved the classification accuracy shown. For all comparisons, no further increase in accuracy was achieved with increasing numbers of nuclei beyond these minimum numbers.

This study has several limitations. The manner in which tissue was processed (one stain) and imaged (conventional CCD-charge-coupled device camera) almost certainly excluded additional pixel data that may have proved useful. However, we chose routine methods of staining and image acquisition with the idea that any methodology developed proving reliable and accurate could be readily translated into a practical tool for the histotechnician and pathologist. In addition, by using the Feulgen technique, DNA alone was targeted for staining. Theoretically, we could have eliminated some pixel information by not using a standard hematoxylin and eosin (HE) stain which stains all nucleic acids and protein. Future work will involve performing the same analysis on HE stained slides. Of note, the Feulgen stain does not require special laboratory reagents and can be performed in most histology laboratories that perform periodic acid Schiff staining. We also note that in this study we chose to utilize 1000× magnification for image acquisition. As a result, relatively few nuclei are present per field of view, hence helping limit the number of nuclei that were extracted per patient. In future work, we plan to study the effect that different magnification would have on the resulting accuracies. Finally, we note that the segmentation methods here were not fully automated in that the pathologist ultimately reviewed the nuclei in the segmented database and only ‘accepted’ nuclei of cells relevant to each category, and that were

appropriately segmented. This was done in order to limit the amount of ‘noise’ in the system. It is also our future goal to develop more robust methods for image segmentation.

The LOT and statistical approaches described above (and in more detail in Wang et al. (2013, 2011a)) are aimed at characterizing differences between nuclear images directly from pixel intensities without first reducing each nuclear image to a set of numerical features. Given enough computation time, it is capable of utilizing all pixel information present in each image. The LOT framework does have the ability to quantify and visualize differences. However, these benefits come with a significant increase in computational complexity and time in comparison to more standard approaches (Wang et al., 2010). Because of computational time considerations, we first constructed an approximate version of each image by selecting its N most important pixels (800 pixels). By doing so, we may have eliminated pixel information that was potentially discriminative in certain applications. Additionally, in similar fashion to statistical methods utilized in the field of image analysis, the framework we describe is expected to be most accurate in two class discrimination problems. As the number of classes increases, such methods tend to have lower classification accuracies given that the complexity of the problem increases significantly.

We point out that while here we showed results using a total of 94 patients, in the future we intend to test the method with a larger number of patients with other malignancies as well. For that to become possible, certain issues may need to be addressed. One important issue to always keep in mind is the careful definition of “ground truth.” While in certain cases (e.g. when follow up data exists) this may not be an issue, the true confirmation of a given diagnosis can be hard to obtain unequivocally. In addition, performing accurate segmentation of vast numbers of nuclei in a large number of patients is an error prone process. In our experience, most existing nuclear segmentation methods have the potential for producing inaccurate segmentation contours at a certain failure rate. For larger studies, it is not clear how classification accuracy would be affected by any increase in the number of inaccurate segmented nuclei. Finally, it is also possible that in a larger study patient data from numerous institutions would need to be combined. As such, standardization in slide preparation and staining may become an issue as well.

In summary, in this study we described a novel, data-driven, computational image analysis methodology can separate a significant number of cases of thyroid follicular tumors by nuclear morphology alone with high accuracies and in most comparisons with 100% accuracy surpassing other traditional feature-based approaches. These methods use relatively few images of nuclei derived from routine histological tissue preparation and image acquisition methods. In addition, our approach is superior to feature-based methods, by enabling visualization and comparison of visual differences between distributions of nuclei. The ability to visualize distributions of nuclei for specific lesions might lead to nuclear morphological correlates of biological behaviors (prognosis) and molecular genetics signatures. Discovering phenotypic differences in other subcellular structures and organelles are clear advantages to this methodology. Other subcellular organelles and structures (e.g. mitochondria, lysosomes, and golgi proteins) would be alternative structures which could be analyzed using the same computational methods, provided they can be properly segmented. The next obvious application of these methods will be in the interpretation of thyroid

cytology and other diagnostic challenges in surgical pathology. A tool could be developed to assist pathologists with diagnostic challenges and potentially provide clinicians with important prognostic information. Additionally, it may obviate the need in some cases for additional tissue sectioning, immunohistochemistry, and molecular testing; elements that add to the overall cost of healthcare.

Acknowledgments

We gratefully acknowledge Clifford Smay, H.T., A.S.C.P. for all histology performed in this study. This work was supported in part by National Institutes of Health grants GM090033, GM088816, and Pennsylvania State Health Department award 4100059192.

Biography

John A. Ozolek graduated from the University of Pittsburgh School of Medicine with the M.D. degree in 1989. He completed residency in Pediatrics and Fellowship in Neonatal-Perinatal Medicine at the University of Pittsburgh before practicing in the United States Air Force and private practice. He returned to train in Anatomic Pathology and completed a Fellowship in Pediatric Pathology at Children's Hospital of Pittsburgh where he remains on staff. He has extensive collaborations with the Center for Bioimage Informatics and Department of Mathematics at Carnegie Mellon University to develop algorithms and computational approaches for automating tissue recognition in research and clinical applications.

Akif Burak Tosun received his B.S. degree in computer engineering from Hacettepe University, Turkey, in 2006, and is his Ph.D. degree in computer engineering at Bilkent University, Turkey, in 2012. He's currently a post doctoral research scholar and his work includes the development of novel approaches for cancer diagnosis using image analysis techniques and machine learning algorithms at Carnegie Mellon University, USA.

Wei Wang earned his B.S. and M.S. degrees in Electrical Engineering from Beijing Jiaotong University in 2004 and 2007, respectively, where he specialized in image processing. He completed his Ph.D. in Biomedical Engineering from Carnegie Mellon University in 2012. His research was focused in biomedical image analysis and machine learning.

Cheng Chen received the B.S. degree and the M.S. degree in electrical engineering from University of Science and Technology of China, in 2005 and 2008 respectively. He got Ph.D. degree in biomedical engineering from Carnegie Mellon University, PA, in 2013. His research interests include: biomedical image processing and analysis, machine learning.

Soheil Kolouri received his M.S. degree in electrical engineering from Colorado State University, Colorado, US in 2012, with specialization in signal/image processing. From 2010 to 2012 he was a research assistant in Signal/Image Processing Laboratory, Colorado State University (CSU), Fort Collins, Colorado. He is currently a Ph.D. candidate in Biomedical Engineering Department at Carnegie Mellon University, Pittsburgh, Pennsylvania. He is a member of the Center for Bioimage Informatics (CBI), where he is working under Dr. Gustavo K. Rohde on generic methods for biomedical image analysis.

Saurav Basu is currently working as a Research Staff Member in the IBM India Research Laboratory at New Delhi, India. He is involved in developing algorithms pertaining to inverse problems in tomographic imaging, signal processing and partial differential equations. Till July 2013, he was engaged in developing bioimage informatics retrieval systems in Center for Bioimage Informatics, Carnegie Mellon University, USA. He was conferred a Ph.D. in Image Analysis by the University of Virginia, USA, in 2011. His interests include application of differential geometry, machine learning, signal processing, optimization and PDEs to problems in imaging, computer vision, inverse problems and control.

Hu Huang received the B.S. and M.S. degrees in Biomedical Engineering from Jiangsu University (China), in 2007 and 2011, respectively. His research interests were signal processing, pattern recognition and their application on biomedical problems. He obtained the M.S. degree in Biomedical Engineering in 2013 from Carnegie Mellon University (Pittsburgh, PA), where his study mainly focused on machine learning and medical image analysis. He is currently pursuing his Ph.D. degree in Biomedical Informatics and Computational Biology at the University of Minnesota (Twin Cities, MN). His main research interests are in the areas of machine learning and metagenomics.

Gustavo K. Rohde earned B.S. degrees in physics and mathematics in 1999, and the M.S. degree in electrical engineering in 2001 from Vanderbilt University. He received a doctorate in applied mathematics and scientific computation in 2005 from the University of Maryland. He is currently an associate professor of Biomedical Engineering, and Electrical and Computer Engineering at Carnegie Mellon University.

References

- Aiad H, Abdou A, Bashandy M, Said A, Ezz-Elarab S, Zahran A. Computerized nuclear morphometry in the diagnosis of thyroid lesions with predominant follicular pattern. *Ecanermedicalsience*. 2009; 3
- Baloch ZW, LiVolsi VA. Follicular-patterned lesions of the thyroid the bane of the pathologist. *Am. J. Clin. Pathol.* 2002; 117:143–150. [PubMed: 11789719]
- Bancroft JD, Gamble M. *Theory and Practice of Histological Techniques*. Elsevier Health Sciences. 2008
- Bishop, CM.; Nasrabadi, NM. *Pattern Recognition and Machine Learning*. Vol. 1. Springer; New York: 2006.
- Boland M, Murphy R. 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. [PubMed: 11751230]
- Chen A, Wang W, Ozolek JA, Rohde G. A flexible and robust approach for segmenting cell nuclei from 2d microscopy images using supervised learning and template matching. *Cytometry A*. 2013; 83:495–507. [PubMed: 23568787]
- Chen, C.; Wang, W.; Ozolek, JA.; Lages, N.; Altschuler, SJ.; Wu, LF.; Rohde, GK. Biomedical Imaging (ISBI). *IEEE*; 2012. 2012. A template matching approach for segmenting microscopy images.; p. 768-771.
- 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. [PubMed: 15475448]
- Cohen J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 1960; 20:37–46.

- Finley DJ, Lubitz CC, Wei C, Zhu B, Fahey TJ III. Advancing the molecular diagnosis of thyroid nodules: defining benign lesions by molecular profiling. *Thyroid*. 2005; 15:562–568. [PubMed: 16029122]
- Fischer AH, Bond JA, Taysavang P, Battles OE, Wynford-Thomas D. Papillary thyroid carcinoma oncogene (RET/PTC) alters the nuclear envelope and chromatin structure. *Am. J. Pathol.* 1998a; 153:1443–1450. [PubMed: 9811335]
- Fischer AH, Chadee DN, Wright JA, Gansler TS, Davie JR. Ras-associated nuclear structural change appears functionally significant and independent of the mitotic signaling pathway. *J. Cell Biochem.* 1998b; 70:130–140. [PubMed: 9632114]
- Fischer AH, Taysavang P, Jhiang SM. Nuclear envelope irregularity is induced by RET/PTC during interphase. *Am. J. Pathol.* 2003; 163:1091–1100. [PubMed: 12937150]
- Frasoldati A, Flora M, Pesenti M, Caroggio A, Valcavi R. Computer-assisted cell morphometry and ploidy analysis in the assessment of thyroid follicular neoplasms. *Thyroid*. 2001; 11:941–946. [PubMed: 11716041]
- 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. [PubMed: 11349221]
- Haralick R, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* 1973:610–621.
- Jennrich RI, Sampson P. Stepwise discriminant analysis. *Stat. Methods Digit Comput.* 1977; 3:77–95.
- Karakitsos P, Cochand-Priollet B, Guillausseau PJ, Pouliakis A. Potential of the back propagation neural network in the morphologic examination of thyroid lesions. *Anal. Quant. Cytol.* 1996; 18:494.
- Karsho lu Y, Celasun B, Günhan Ö. Contribution of morphometry in the differential diagnosis of fine-needle thyroid aspirates. *Cytometry B Clin. Cytom.* 2005; 65:22–28. [PubMed: 15779051]
- LiVolsi VA, Baloch ZW. Follicular neoplasms of the thyroid: view, biases, and experiences. *Adv. Anat. Pathol.* 2004; 11:279–287. [PubMed: 15505528]
- Mahfouz S, El-Sharkawy S, Sharaf W, Hussein H, El-Nemr R. Image cytometry of fine needle aspiration of thyroid epithelial lesions. *Appl. Immunohisto. Mol. Morphol.* 2012; 20:25.
- Murata, S.i.; 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. [PubMed: 12442308]
- Nielsen B, Albrechtsen F, Danielsen HE. Automatic segmentation of cell nuclei in feulgen-stained histological sections of prostate cancer and quantitative evaluation of segmentation results. *Cytometry A*. 2012; 81:588–601. [PubMed: 22605528]
- Nikiforov YE. Thyroid carcinoma: molecular pathways and therapeutic targets. *Modern Pathol.* 2008; 21:S37–S43.
- Nikiforov YE, Biddinger PW, Thompson LD. *Diagnostic Pathology and Molecular Genetics of the Thyroid: A Comprehensive Guide for Practicing Thyroid Pathology*. Wolters Kluwer Health (second ed.). 2012
- Rohde GK, Ribeiro AJ, Dahl KN, Murphy RF. Deformation-based nuclear morphometry: capturing nuclear shape variation in HeLa cells. *Cytometry A*. 2008; 73:341–350. [PubMed: 18163487]
- Salmon I, Gasperin P, Pasteels J, 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.* 1992a; 97:776. [PubMed: 1375804]
- Salmon I, Kiss R, Franc B, Gasperin P, Heimann R, Pasteels J, Verhest A. Comparison of morphonuclear features in normal, benign and neoplastic thyroid tissue by digital cell image analysis. *Anal. Quant. Cytol.* 1992b; 14:47–54.
- Shapiro NA, Poloz TL, Shkurupij VA, Tarkov MS, Poloz VV, Demin AV. Application of artificial neural network for classification of thyroid follicular tumors. *Anal. Quant. Cytol.* 2007; 29:87.

- Stein GS, Montecino M, Van Wijnen AJ, Stein JL, Lian JB. Nuclear structure-gene expression interrelationships: implications for aberrant gene expression in cancer. *Cancer Res.* 2000; 60:2067–2076. [PubMed: 10786661]
- Wang W, Mo Y, Ozolek JA, Rohde GK. Penalized fisher discriminant analysis and its application to image-based morphometry. *Pattern Recogn. Lett.* 2011a; 32:2128–2135.
- Wang W, Ozolek JA, Rohde GK. Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images. *Cytometry A.* 2010; 77:485–494. [PubMed: 20099247]
- Wang W, Ozolek JA, Slepcev D, Lee AB, Chen C, Rohde GK. An optimal transportation approach for nuclear structure-based pathology. *IEEE Trans. Med. Imag.* 2011b; 30:621–631.
- Wang W, Slepcev D, Basu S, Ozolek JA, Rohde GK. A linear optimal transportation framework for quantifying and visualizing variations in sets of images. *Int. J. Comput. Vision.* 2013; 101:254–269.
- Zhao J, Leonard C, Genssenjäger 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. [PubMed: 18202778]
- Zink D, Fischer AH, Nickerson JA. Nuclear structure in cancer cells. *Nat. Rev. Cancer.* 2004; 4:677–687. [PubMed: 15343274]

Fig. 1.

Columns (top to bottom) show representative HE (hematoxylin and eosin) images of normal thyroid and benign and malignant thyroid follicular neoplasms, segmented nuclei from Feulgen stained images, and segmented nuclei normalized, rotated, and converted to grayscale (used for analysis): Column 1 (A, G, M)-normal thyroid, column 2 (B, H, N)-nodular goiter, column 3 (C, I, O)-follicular adenoma, column 4 (D, J, P)-follicular carcinoma, column 5 (E, K, Q)-widely invasive follicular carcinoma, column 6 (F, L, R)-follicular variant of papillary carcinoma (all HE images captured at 400 $\times$; Feulgen stained images captured at 1000 $\times$).

Fig. 2. Graphical representation of the linear optimal transport framework (left-hand side), and linear subspace extraction method (right-hand side). See text for details.

Fig. 3.

Nuclei distribution histograms and accuracy table for follicular adenoma (FA) and follicular carcinoma (FC) comparison. The two rows of images of nuclei beneath the histogram bins are the normalized grayscale (upper row) and colorized (lower row) visual representations of nuclei along the optimal transportation pathway (geodesic) that best discriminate between these two groups of nuclei. Actual nuclei in these groups most similar to the generated nuclei below were then counted in the corresponding bin to generate the bars of the histogram. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4.
Nuclei distribution histograms and accuracy table for comparison of follicular adenoma (FA) and follicular variant of papillary carcinoma (FVPC) comparison.

Fig. 5. Nuclei distribution histograms and accuracy table for comparison of follicular carcinoma (FC) and widely invasive follicular carcinoma (WIFC) comparison.

Fig. 6. Nuclei distribution histograms and accuracy table for comparison of follicular variant of papillary carcinoma (FVPC) and follicular carcinoma (FC) comparison.

Fig. 7. Nuclei distribution histograms and accuracy table for comparison of follicular adenoma (FA) and nodular goiter (NG) comparison.

Fig. 8.
Nuclei distribution histograms and accuracy table for comparison of follicular carcinoma (FC) and nodular goiter (NG) comparison.

Table 1

Pathologic and clinical characteristics of thyroid follicular lesions. NG-Nodular goiter, FA-Follicular adenoma, FC-Follicular carcinoma, FVPC-Follicular variant of papillary carcinoma, WIFC-Widely invasive follicular carcinoma. Range presented in parentheses.

Lesion	N	Size (cm)	Age (years)
NG	28	59.1 ± 38.3 ^a	55.8 ± 12.7(31 – 79)
FA	27	2.6 ± 1.4	47.5 ± 12.1(23 – 69)
FC	20	3.3 ± 1.6	42.9 ± 16.5(14 – 75)
FVPC	10	2.2 ± 1.5	46.1 ± 13.8(27 – 72)
WIFC	9	5.2 ± 2.9	60.6 ± 16.8(33 – 83)

^a(Weight (gr) for NG only).

Table 2

Confusion matrices for the challenges of FA vs. FC, FA vs. FVPC, FC vs. WIFC, FC vs. FVPC, FA vs. NG, and FC vs. NG. Tables are showing results of both LOT based and feature based approaches together in each table entry (LOT Predictions/Feature Predictions).

FA vs. FC				FA vs. FVPC			
		True diagnosis				True diagnosis	
		FA	FC			FA	FVPC
Test	FA	27/24	0/3	Test	FA	27/26	0/1
	FC	0/3	20/17		FVPC	0/0	10/10

FC vs. WIFC				FC vs. FVPC			
		True diagnosis				True diagnosis	
		FC	WIFC			FC	FVPC
Test	FC	20/16	0/4	Test	FC	20/18	0/2
	WIFC	0/3	9/6		FVPC	4/1	6/9

FA vs. NG				FC vs. NG			
		True diagnosis				True diagnosis	
		FA	NG			FC	NG
Test	FA	27/21	0/6	Test	FC	20/13	0/7
	NG	0/3	28/25		NG	0/2	28/26

Table 3

Summary of classification results on four different comparisons.

Comparisons	LOT		Feature	
	Accuracy (%)	Cohen's Kappa	Accuracy (%)	Cohen's Kappa
FA vs. FC	100	1.00 ($p = 0$)	87	0.74 ($p = 0$)
FA vs. FVPC	100	1.00 ($p = 0$)	97	0.93 ($p = 0$)
FC vs. WIFC	100	1.00 ($p = 0$)	76	0.45 ($p = .0157$)
FVPC vs. FC	87	0.67 ($p = .002$)	90	0.78 ($p = 0$)
FA vs. NG	100	1.00 ($p = 0$)	84	0.67 ($p = 0$)
FC vs. NG	100	1.00 ($p = 0$)	81	0.60 ($p = .0001$)