A Clinically Motivated 2-Fold Framework for Quantifying and Classifying Immunohistochemically Stained Specimens

Bonnie Hall[1,2,3], Wenjin Chen[1,3], Michael Reiss[3,4], and David J. Foran[1,2,3]

[1] Center for Biomedical Imaging and Informatics
[2] Graduate School of the Biomedical Sciences
[3] The Cancer Institute of New Jersey, UMDNJ-Robert Wood Johnson Medical School
[4] Dept. of Internal Medicine and Dept. of Molecular Genetics, Microbiology, and Immunology
195 Little Albany St. New Brunswick, NJ 08903 USA
{huangbo, reissmi, djf}@umdnj.edu, wenjin.chen@gmail.com

Abstract. Motivated by the current limitations of automated quantitative image analysis in discriminating among intracellular immunohistochemical (IHC) staining patterns, this paper presents a two-fold approach for IHC characterization that utilizes both the protein stain information and the surrounding tissue architecture. Through the use of a color unmixing algorithm, stained tissue sections are automatically decomposed into the IHC stain, which visualizes the target protein, and the counterstain which provides an objective indication of the underlying histologic architecture. Feature measures are subsequently extracted from both staining planes. In order to characterize the IHC expression pattern, this approach exploits the use of a non-traditional feature based on textons. Novel biologically motivated filter banks are introduced in order to derive texture signatures for different IHC staining patterns. Systematic experiments using this approach were used to classify breast cancer tissue microarrays which had been previously prepared using immuno-targeted nuclear, cytoplasmic, and membrane stains.

Keywords: Quantitative IHC analysis, texture descriptors, expression signatures, automated classification, breast cancer.

1 Introduction

The capacity to quantify and characterize protein expression reliably is central to several key areas of investigative cancer research and discovery. Biomarkers can provide tremendous insight into the underlying mechanisms of disease progression and can have significant impact with regard to patient prognosis, treatment, and therapy planning. For instance, the presence of biomarkers such as the Her-2/neu (also called ErbB2) receptor in breast cancer indicates that a given patient may respond to treatment with Trastuzumab. Immunohistochemistry (IHC) is used to visualize these proteins by labeling them with a stain, such as diaminobenzidine (DAB). However, manual scoring of IHC stained pathology suffers from several

drawbacks. The most striking limitations include the inability to reproduce results, subjectivity of analyis, and intra- and inter-observer variability.

It has already been reported in the literature that quantitative IHC image analysis provides more consistent scoring than traditional manual scoring[1, 2] and better concordance with existing gold standards e.g. fluorescence in situ hybridization (FISH)[3, 4].   However, discrepancies between IHC scoring and actual clinical outcomes often arise irrespective of the means by which analysis is performed i.e. by eye or computer-assisted [3-5]. To address these ambiguities, this paper proposes a fresh approach for characterizing in-situ protein expression by utilizing texture-based feature metrics.

Estrogen Receptor   Anti-Keratin 5/8   Her-2/neu
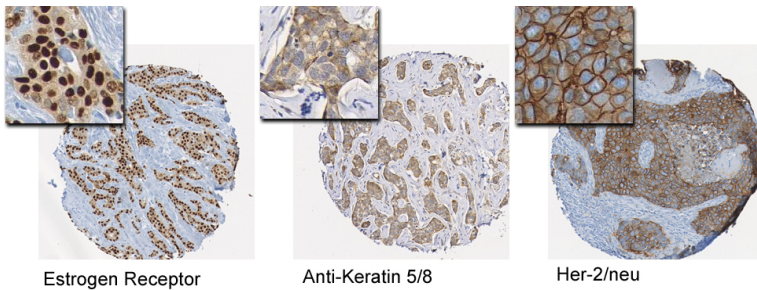
Fig. 1. Images taken at 20x resolution for breast tissue TMA cores stained with DAB *(brown regions)* and counter stained with hematoxylin *(blue regions)*. Inserts *(upper left)* show representative regions of staining from each of these cores at high resolution. Note that ER and PR are generally nuclear proteins, keratin is generally found in the cytoplasm, and Her-2/neu is active in disease processes when concentrated on the cytoplasmic membrane.

Constructing a model for quantifying in-situ protein expression required a close examination of the design specifications for our experiments. The first criteria was that the proposed model would reliably and reproducibly capture and quantify the underlying staining characteristics of the protein. Second, in order to be clinically meaningful, a means for detecting and objectively representing the contextual cellular and tissue-level architecture would need to be incorporated into the model since many proteins are only considered active or clinically relevant when they are localized in specific subcellular compartments[4, 6, 7]. Using these criteria we have developed a model that uses the texton signatures of the immunstains to characterize the protein expression pattern, and histological features to determine intracellular localization of target proteins. This information provides important insight into the in-vivo status of these proteins in tumors, such as indicating whether they are being overexpressed and specifying their functional location.

In this paper, we report the design, development, and evaluation of a two-fold model for IHC characterization that utilizes both the information provided by the targeted protein (visualized by the DAB stain), and the histological context (visualized by the hematoxylin stain). The model is evaluated on a breast cancer tissue microarray (TMA) stained for four proteins: keratin (indicating regions of carcinoma) and three clinically used markers: estrogen receptor (ER), progesterone receptor (PR), and Her-2/neu. Proficient characterization of IHC with this model is

demonstrated by its ability to classify ER and PR, keratin, and Her-2/neu into three respective classes: nuclear, cytoplasmic, and membrane staining. Highlights of this paper include the use of texton-based features to characterize protein expression, the design and systematic investigation of a new set of filter banks, and the use of a two-fold model which combines information derived from both protein staining pattern and underlying tissue architecture. Throughout the course of these experiments a stringent 1/2 cross-validation analysis was carried out to ensure the reliablility and robustness of the results.

2   Two-Fold Modeling of In-situ Protein Expression

The two-fold modeling approach we present (Fig. 2) utilizes color decomposition[8] to separate the immunostained tissue disc into its constituent staining maps: the diaminobenzidine (DAB) staining plane which indicates level and distribution of the targeted protein, and the hematoxylin plane, which reveals cellular and tissue architecture.
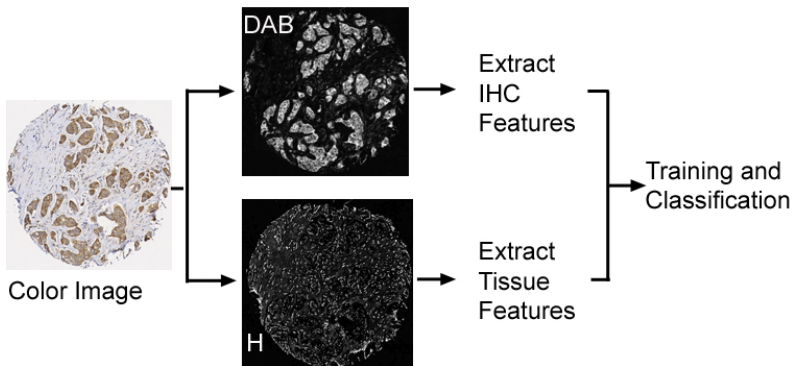
Fig. 2. The color image is decomposed into the targeted protein (DAB) and the hematoxylin (H) staining maps. Features are extracted from each map and combined for classification.

The DAB staining map is characterized using textons, a texture feature composed of a vocabulary of consensus filter responses derived through cluster analysis [9, 10]. The vocabulary is learned directly from the image by convolving a filter bank with image patches from the immunohistochemically stained map and clustering the response vectors in order to obtain IHC (in this case, DAB) textons. IHC textons are derived for each class of staining pattern and combined to form the vocabulary. An IHC texture histogram is derived by convolving the segmented immunohistochemically stained regions with the same filter bank and matching the resulting filter responses to the closest IHC texton. A normalized histogram of textons detected in the stained area is then generated.

In the tissue architecture map, regions corresponding to the segmented DAB stain are analyzed with various features such as an intensity histogram, mean, standard deviation, median intensity, and Haralick's 2nd order features[11]. The intensity

histogram is a normalized tally of intensities (0-255) where 2 intensities are placed per bin for smoothness. Since the counterstain usually does not stain very strongly, our preliminary results show that using intensities 0 through 31 (a 16 dimensional vector) was sufficient to capture the discriminating information present in the intensity histogram.

$$D = D^{DAB} + D^{H} \ . \tag{1}$$

The DAB stain feature and the hematoxylin (H) feature distances were computed between training and test images and summed according to Eq.1 in order to arrive at the total combined feature distance, D. The χ^2 distance metric was used for D^{DAB} and D^{H} . A weighted K-Nearest Neighbor classifier was implemented to perform this operation.

2.1 Filter Bank Construction and Development

A range of filter banks was systematically applied in order to derive IHC textons that optimize classification. Based on its superior classification performance when used to assess a natural image database, the first filter bank chosen was the MR8 Filter Bank[10]. Subsequently, a set of novel filter banks were introduced based on the observation that cellular features tended to be more isotropic and less like edges in nature: such as blob-like nuclei and circular membranes. Adapted from the MR8 filter bank, the first biologically based filter bank 1 (BF1) also contains isotropic filters developed by Schmid[12] as well as two sizes of Gaussian and Laplacian of Gaussian (LOG) filters. This filter bank produces an 8 dimensional response vector (Fig. 3).
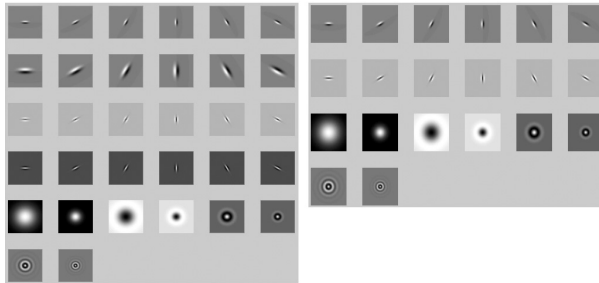
Fig. 3. The BF1 filter bank *(left)*, includes edge filters at 2 scales and 6 rotations (σ_x, σ_y) = {(1,3), (2,6)}, a black and white-centered bar filter at one scale (1,3) and 6 rotations, 2 sizes of isotropic Gaussian and LOG at 2 scales σ = {5,10}, and 4 Schmid filters where (σ, τ) = {(4,1), (4,2), (6,1), (6,2)}. In order to achieve rotational and size invariance, the six rotations for bar and edge filters and the two scales for the Gaussian and LOG were collapsed down to their maximum response, resulting in an 8 dimensional response vector. Similarly, the BF2 filter bank *(right)* produces a 10 dimensional response vector through rotation invariance of the smallest BF1 bar and edge filters, but retaining responses from each isotropic filter.

Similarly, the second biologically based filter bank (BF2) was derived through principle component analysis of a subset of IHC texton histograms generated with

BF1. Based on this analysis, the number of bar and edge type filters were reduced, while isotropic filters would retain size differences. The result of BF2 is a 10 dimensional response vector (Fig.3). The filters are L_1 normalized so that outputs are of similar range.

2.2   Breast TMA Data Set

Tissue microarray datasets comprised of 186 breast carcinoma cases, with two tissue cores extracted from each case. Consecutive sections 34 and 35 were stained with anti-keratin 5/8. Sections 36, 38, 39 were stained with anti-estrogen receptor antibody, Her-2/neu, and progesterone receptor, respectively. These specimens were all counter-stained with hematoxylin.

Images were digitized utilizing a Trestle MedMicro (Irvine, CA) automatic whole slide scanner system using a 40x volume scan setting. The TMA image was automatically registered, parsed into separate tissue cores at 20x equivalence, and decomposed into DAB and hematoxylin maps [8].

In order to treat stained regions as texture, background pixels near DAB pixels were segmented using a stain mask derived by blurring the DAB region with an 11x11 pixel averaging filter and applying Otsu's method[13] for thresholding. 212 cytoplasmic-stained images were obtained from tissues that stained specifically for anti-keratin 5/8 and 212 nuclear-stained images from ER and PR positive tissues. Two image patches (150x150 pixels) taken from twenty (~10%) randomly chosen images per class were dedicated to vocabulary formation, leaving 192 per class for training and classification. 50 Her-2/neu images with positive membrane staining were obtained. Similarly, 2 image patches taken from five randomly chosen images were used for vocabulary formation, and reused solely as training.
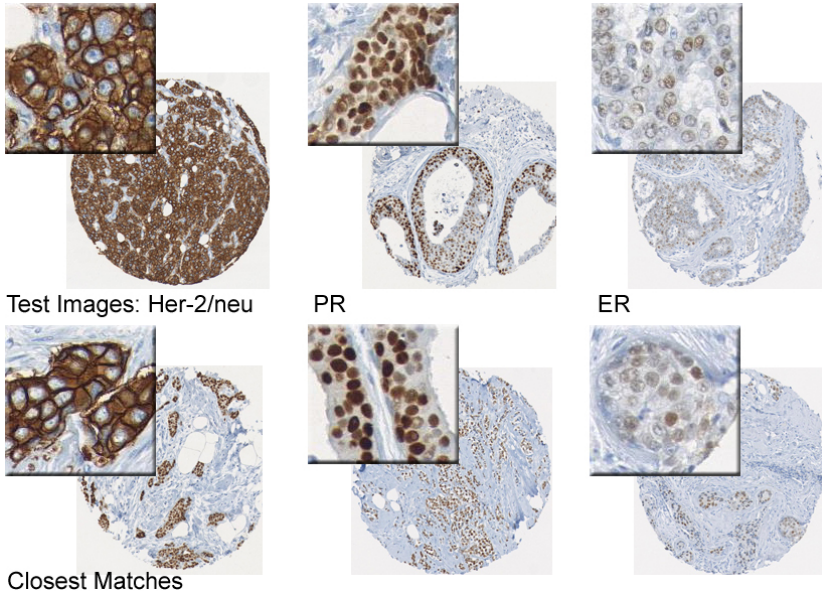
3   Results

In order to demonstrate robustness of our method, 1/2 cross-validation was used throughout the experiments, wherein 50% of our images from each class were used for training and 50% were used for testing. Due to the relatively limited number of specimens from the membrane staining class (only 15-30% of breast cancer patients overexpress Her-2/neu[14]) two sets of experiments were carried out. The two-class classification experiment contains 212 nuclear-staining images and 212 cytoplasmic-staining images. The three-class classification experiments used a balanced set of nuclear-staining, cytoplasmic-staining, and membrane-staining images of 50 each. Three filter banks, MR8, BF1, and BF2, were used to derive texton libraries of a range of sizes based on DAB information. Results are reported for performance with DAB information alone and in combination with hematoxylin stained tissue information (a 16 dimensional truncated intensity histogram).

All three filter banks classified the two-class case with >90% accuracy (Figure 4, *left*). When using the BF1 filter bank, the texton feature from DAB stain alone allowed us to classify the two class case with 94% correct classification. When tissue information was added to the experiment, the result modestly increased to about 96%.
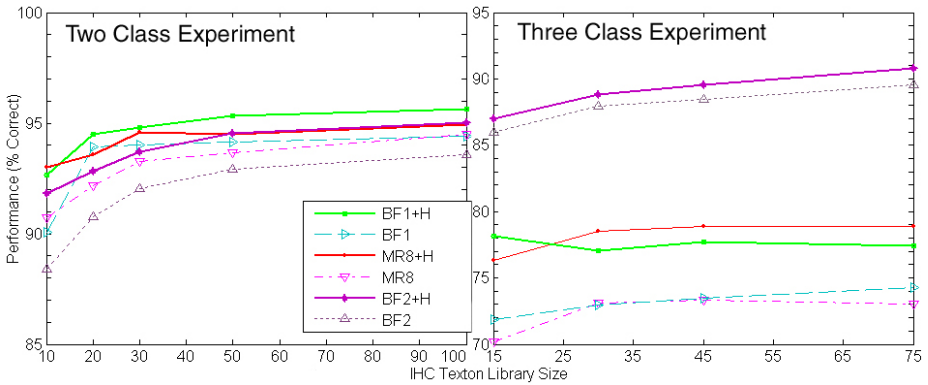
Fig. 4. Nuclear and cytoplasmic staining were classified by DAB textons with and without the addition of information from hematoxylin stained tissue (H)*(left).*The membrane staining class was added in the three class experiment *(right).* The same legend is used for both figures.

Fig. 5. Three sets of test images *(top row)* from the three class experiment and their closest matches*(bottom row)*. High resolution regions *(upper left inserts)* are shown.

Figure 4 *(right)* shows the results from the three class (nuclear, cytoplasmic, and membrane staining) experiments. The results for DAB textons derived from the BF1 filter bank are similar to those of MR8 filter bank, while those from BF2 clearly provide better classification performance. Significant improvement is noted in the three class case with the addition of tissue information to DAB textons derived from the MR8 and BF1 filter banks.

Experiments were also conducted on a range of features extracted from the hematoxylin map including 1st order features: mean, standard deviation, and median, as well as 2nd order texture features: Entropy, Angular Second Moment, Contrast, and Correlation. The combination of the first five features resulted in about 70% correct classification for the 2-class case. However, when combined with the DAB features according to Eq.1, there was no significant improvement in performance.

Figure 5 shows performance results from the three-class experiment utilizing 45 DAB textons (BF2, 15 textons per class) and hematoxylin information. Please note that since all cores were stained and imaged under identical conditions, the resulting texton library carries normalized staining intensity information. As a result, our model also captures the staining intensity level (Fig. 5, right panel).

4 Discussion and Conclusion

In this paper, a system for extracting biologically distinct feature sets was developed in order to classify and aid assist in characterizing immunohistochemically stained tissue specimens. The basic framework of this algorithm decomposes an IHC stained image into two clinically motivated parts: the protein of interest and its tissue context. A range of different feature sets are extracted from these images at both the expression and contextual levels and subsequently combined to improve classification performance. This approach allows the quantification of two key components for reliably interpreting immunohistochemistry, namely the protein of interest and its histological context.

The texton library-based texture feature was chosen in our studies to characterize the staining pattern of the protein of interest because of its ability to adapt to complex tissue textures, which have repetitive structures at multiple scales. We introduced a set of new biologically oriented filter banks into the algorithm that proved to reliably capture the subtle morphological differences among various staining patterns and im-prove classification performance. We have also demonstrated that by incorporating the underlying tissue architecture into the formulation that performance is further enhanced.

This paper presents a proof-of-concept study that demonstrates the feasibility of discriminating among IHC staining patterns based on the simultaneous consideration of texture and tissue architecture. In its present form the model and associated algorithms may be used for investigating cases where there exists discrepencies between what is expected based on standard IHC analysis and response to therapy or other measures of protein amplification. Cases where discrepencies exist can potentially lead to the withholding of targeted therapies or the unwarranted exposure to drug toxicity. The textural information underlying the protein expression pattern may be an additional factor in elucidating these clinical inconsistencies and thus provide patients with the proper treatment protocol. In the next phase of our studies we plan to investigate the usefulness of this approach in analyzing a broader range of immunostains and disease states.

References

1. Hilbe, W., Gachter, A., Duba, H.C., Dirnhofer, S., Eisterer, W., Schmid, T., Mildner, A., Bodner, J., Woll, E.: Comparison of automated cellular imaging system and manual microscopy for immunohistochemically stained cryostat sections of lung cancer specimens applying, p. 53, ki-67 and p. 120. Oncol Rep 10, 15–20 (2003)
2. Camp, R.L., Chung, G.G., Rimm, D.L.: Automated subcellular localization and quantification of protein expression in tissue microarrays. Nat. Med. 8, 1323–1327 (2002)
3. Ellis, C.M., Dyson, M.J., Stephenson, T.J., Maltby, E.L.: HER2 amplification status in breast cancer: a comparison between immunohistochemical staining and fluorescence in situ hybridisation using manual and automated quantitative image analysis scoring techniques. J. Clin. Pathol. 58, 710–714 (2005)
4. Wang, S., Saboorian, M.H., Frenkel, E.P., Haley, B.B., Siddiqui, M.T., Gokaslan, S., Wians, F.H., Hynan Jr, L., Ashfaq, R.: Assessment of HER-2/neu status in breast cancer. Automated Cellular Imaging System (ACIS)-assisted quantitation of immunohistochemical assay achieves high accuracy in comparison with fluorescence in situ hybridization assay as the standard. Am. J. Clin. Pathol. 116, 495–503 (2001)
5. Chung, K.Y., Shia, J., Kemeny, N.E., et al.: Cetuximab shows activity in colorectal cancer patients with tumors that do not express the epidermal growth factor receptor by immunohistochemistry. J. Clin. Oncol. 23, 1803–1810 (2005)
6. Liang, J., Zubovitz, J., Petrocelli, T., et al.: PKB/Akt phosphorylates p. 27, impairs nuclear import of p. 27 and opposes p27-mediated G1 arrest. Nat. Med. 8, 1153–1160 (2002)
7. Rosen, D.G., Yang, G., Cai, K.Q., Bast Jr, R.C., Gershenson, D.M., Silva, E.G., Liu, J.: Subcellular localization of p. 27 kip 1 expression predicts poor prognosis in human ovarian cancer. Clin. Cancer Res. 11, 632–637 (2005)
8. Chen, W., Reiss, M., Foran, D.J.: A prototype for unsupervised analysis of tissue microarrays for cancer research and diagnostics. IEEE Trans. Inf. Technol. Biomed. 8, 89–96 (2004)
9. Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. Int. J. Comput Vision 43, 29–44 (2001)
10. Varma, M., Zisserman, A.: A statistical approach to texture classification from single images. Int. J. Comput Vision 62, 61–81 (2005)
11. Haralick, R.: Texture features for image classification. IEEE Trans. Sys. Man. Cyber. SMC-3, 610–621 (1973)
12. Schmid, C.: Constructing models for content-based image retrieval. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 39–45 (2001)
13. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Sys. Man. Cyber. 9, 62–66 (1979)
14. Slamon, D.J., Godolphin, W., Jones, L.A., Holt, J.A., Wong, S.G., Keith, D.E., Levin, W.J., Stuart, S.G., Udove, J., Ullrich, A., et al.: Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 244, 707–712 (1989)