Introduction

Tumors often contain varying levels of infiltrating lymphocytes, depending on the strength of the host’s immune response. Indeed, tumor-infiltrating lymphocyte (TIL) density is a strong predictor of clinical outcome, since it reflects whether or not the immune system is attacking the cancer. Potentially, tests that characterize TIL phenotype (CD3, CD4, CD8, FOXP3, etc.), location, and density in tumor and surrounding stroma could aid the selection of immunotherapies and define targets for individualized treatment. TIL assessment today is tedious and time consuming, and since it is based on human visual perception, is prone to inter- and intra-observer variability, due to complicated tissue architectures, ambiguous histology revealed through hematoxylin counterstain, and other human factors. In this study, performed in collaboration with Dr. Michael Feldman and Dr. Ian Hagemann at University of Pennsylvania, we investigate computer-aided histologic event counting, using as our sample use-case the counting of lymphocytes in serous ovarian carcinoma specimens. Samples were prepared as a tissue microarray (TMA) consisting of 618 cases, with clinical follow-up. PerkinElmer’s Vectra® multispectral slide analysis system was used to acquire images from each core and inForm® Tissue Finder™ image analysis software to automatically detect regions of tumor, and counted CD3+TILs in tumor regions, thus determining T cell density. Automated results were compared with visual assessments. Kaplan-Meier survival curves were generated for both automated and visual scores.
Methods

Sample
• Tissue microarray (TMA) containing 618 ovarian cancer samples.
• The TMA samples were stained with anti-CD3 antibody to identify T-lymphocytes.
• The samples were further stained with an epithelial cell marker (cytokeratin) to assist in automated segmentation of tumor and stroma (Figure 1).

Automated Scoring
• The TMA slides were scanned using PerkinElmer’s Vectra multispectral imaging system.
• The scanned images were processed using PerkinElmer’s inForm Tissue Finder. A machine-learning algorithm was trained to segment tumor from stroma and identify CD3 cells labeled with DAB, indicating T-lymphocytes.
• The density of T cells within the tumor areas was calculated.

Visual Scoring
• A pathologist rated lymphocyte density.
• A semi-quantitative scale of 0 to 3 was assigned to each core.

Analysis
• Automated counts were compared to visual scores.
• Manual and automated scoring were compared with survival.

Results
The machine-learning algorithm determined tumor T cell density for 70% of cores and stromal T cell density for 42% (Table 1). With triplicate representation of each tumor on the array, 93% of tumors had at least one core informative for intratumoral T cells, and 71% had at least one core informative for stromal T cells. There was a significant strong positive correlation between total visual and machine counts ($r = 0.6704, p < 0.0001$ by Spearman’s nonparametric test) (Figure 2). Kaplan-Meier analysis shows equivalent and significant P values (~0.03) for visual and automated scoring approaches (Figure 3). Although correlation between visual and automated scoring was high, automated scoring consistently determined larger numbers than visual. Larger numbers were due primarily to oversplitting of segmented lymphocytes, the inclusion of lymphocytes lacking nuclear counterstain which are ignored during visual scores, and inclusion of lymphocytes at periphery of tumor areas. Additionally, we found that a pathologist’s involvement was essential, to review segmentation results and assure data quality by rejecting data from areas improperly stained, out of focus, or folded, or otherwise inaccurately segmented.

Table 1. Automated segmentation results showing that 436 of 618 TMA cores successfully segmented. Errors were due to issues with the tissue, over- or under-segmentation of tissue, and over- or under-segmentation of lymphocytes.
Conclusions

- Preliminary results indicate machine scoring can meaningfully capture TIL status of tumors and yield quantitative, normalized feature count using consistent rules.

- The prognostic power of the test can be extended by adding labels for lymphocyte phenotyping (e.g., CD4+, CD8+, CD25, FOXP3+, etc), enabled by Vectra’s multispectral capability.

- TMAs, although useful for research investigations, do not support routine clinical work. Future investigations will involve whole biopsy sections.

- These results demonstrate the feasibility of a practical and viable clinical workflow, in which TIL counting is automated by computer and results are reviewed by pathologists to assure data quality.

Acknowledgement

Study performed in collaboration with I.S. Hagemann & M.D. Feldman, Hospital of the University of Pennsylvania, Philadelphia, PA; University of Pennsylvania Health System, Philadelphia, PA, USA

Figure 2. Correlation of visual scoring and automated scoring yielded an r-value of 0.6704.

Figure 3. Kaplan-Meier curves for visual and semi-automated scoring compared with survival. These curves have essentially equivalent P-values.