Published in final edited form as:

Histopathology. 2011 July ; 59(1): 40–54. doi:10.1111/j.1365-2559.2011.03878.x.

Cell-based quantification of molecular biomarkers in histopathology specimens

Yousef Al-Kofahi¹, Wiem Lassoued^{2,3}, Kedar Grama⁴, Sumit K Nath¹, Jianliang Zhu², Ridha Oueslati³, Michael Feldman⁵, William M F Lee², and Badrinath Roysam^{1,4}

¹Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY

²Department of Medicine, Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA

³Unit of IMEC, University of Sciences Bizerte, Tunisia

⁴Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY

⁵Department of Pathology and Laboratory Medicine, Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA

Abstract

Aims—To investigate the use of a computer-assisted technology for objective, cell-based quantification of molecular biomarkers in specified cell types in histopathology specimens, with the aim of advancing current visual estimation or pixel-level (rather than cell-based) quantification methods.

Methods and results—Tissue specimens were multiplex-immunostained to reveal cell structures, cell type markers, and analytes, and imaged with multispectral microscopy. The image data were processed with novel software that automatically delineates and types each cell in the field, measures morphological features, and quantifies analytes in different subcellular compartments of specified cells. The methodology was validated with the use of cell blocks composed of differentially labelled cultured cells mixed in known proportions, and evaluated on human breast carcinoma specimens for quantifying human epidermal growth factor receptor 2, oestrogen receptor, progesterone receptor, Ki67, phospho-extracellular signal-related kinase, and phospho-S6. Automated cell-level analyses closely matched human assessments, but, predictably, differed from pixel-level analyses of the same images.

Conclusions—Our method reveals the type, distribution, morphology and biomarker state of each cell in the field, and allows multiple biomarkers to be quantified over specified cell types, regardless of abundance. It is ideal for studying specimens from patients in clinical trials of targeted therapeutic agents, for investigating minority stromal cell subpopulations, and for phenotypic characterization to personalize therapy and prognosis.

Keywords

automated image analysis; cell typing; digital histopathology; molecular biomarkers; multiplexed immunolabelling; multi-spectral imaging

Address for correspondence: Badrinath Roysam, JEC 7010, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA. roysam@ecse.rpi.edu. William M F Lee, Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA 19104, USA. leeemf@mail.upenn.edu.

[†]Y.A.-K. and W.L. contributed equally to this work.

Introduction

Histopathological evaluation of tissue samples is indispensable for cancer diagnosis, classification, and management,^{1,2} and is an important tool in animal-based research.^{3,4} Thin tissue sections are stained with haematoxylin, eosin and/or other chemical stains to reveal cell and tissue structures. Antibody staining to reveal specific molecular biomarkers is increasingly used to improve cancer diagnosis and classification, establish prognosis, and determine therapy. Even as molecular biomarkers are playing a growing role, the scoring of stained specimens remains largely a visual and subjective process: Cells are coarsely scored as positive or negative or graded for degree of antigen staining, the percentage of positive cells is estimated visually, and overall scores are arbitrarily binned/scaled. This process requires considerable expertise and is susceptible to inter-observer variability, despite standardization efforts.^{5–13} The use of rough composite score scales (e.g. 0, 1+, 2+, 3+) is a tacit acknowledgement of the inherent imprecision and subjectivity involved.

Recently, computer-automated methods have been developed to quantify antigen expression in tissue images,^{14–17} offering objectivity, reproducibility, and quantification on a continuous scale. Most operate by measuring the number of pixels stained for one or more antigens and quantifying co-localization of stains. They quantify at the level of individual pixels, groups of pixels, or image regions, however, and not at the level of individual cells, which are the fundamental units at which many biological processes occur. This is largely because of the lack of sufficiently reliable automated methods to segment (delineate) individual cells, identify subcellular compartments within cells, and quantify biomarkers within the subcellular regions. We set forth an approach that leverages recent advances in imaging, image analysis and pattern theory to enable biomarkers to be analysed and quantified on a cell-by-cell basis, providing additional data that cannot be obtained by pixel-level analysis and advancing prior efforts.^{18,19} Our segmentation algorithms are capable of delineating subcellular compartments by the use of image cues and geometric constraints. The subcellular compartment segmentations are consistently linked, enabling correct analysis in situations that challenge pixel-level analytical methods, e.g. multiple markers that are not co-localized but are present in the same cell. Importantly, our method explicitly identifies cell types, permitting selective measurement of biomarker expression in cell subpopulations regardless of their abundance

Materials and methods

Tissue Staining

Deparaffinized 5- μ m sections of formalin-fixed, paraffin-embedded human breast tissues were treated with citric acid (pH 6) for 15 min at 90°C prior to staining. Antibodies used for immunostaining included monoclonal mouse anti-human oestrogen receptor (ER), anti-human progesterone receptor (PR), anti-human Ki67, anti-epithelial membrane antigen (EMA), rabbit polyclonal anti-human epidermal growth factor receptor 2 (HER2) (Dako, Carpinteria, CA, USA), rabbit anti-phospho(p)-extracellular signal-related kinase (ERK), anti-p-S6 (Cell Signaling, Danvers, MA, USA), and mouse anti-multi-cytokeratin (CK) monoclonal antibodies (Vector Laboratories, Burlingame, CA, USA). ER, PR, Ki67, p-ERK and HER2 were detected by immunohistochemistry with biotinylated species-specific secondary antibodies, avidin-linked horseradish peroxidase (HRP) (ABC Kit) and 3,3'-diaminobenzidine or SG Blue (Vector Laboratories) HRP chromogen substrate. CK, EMA and p-S6 immunostaining were detected by fluorescence, with the use of Zenon Alexa Fluor-488 mouse IgG1 labelling (Invitrogen, Carlsbad, CA, USA), fluorescently labelled secondary antibodies (Invitrogen) or the ABC fluorescence detection kit. After immunostaining, slides were counterstained with haematoxylin.

Individual slides were stained with combinations of the above antibodies to reveal antigens that reported on cell compartments, cell type and molecular analytes in each slide. Multiplex staining protocols were developed to minimize or avoid the opportunity for non-specific staining by secondary antibodies. Both chromogenic and fluorescent reporters were frequently used on the same slide, and only fluorochromes that could be resolved spectrally were used on the same slide.

Tissue Imaging

A Nuance[®] multispectral camera (CRI, Woburn, MA, USA) on a Leica DMRA2 epifluorescence microscope was used to record images at $\times 400$ magnification, 8 bits/pixel at 10-nm wavelength intervals from 420 to 720 nm in both brightfield and fluorescent modes. Nuance software was used to spectrally unmix the data into distinct channels representing haematoxylin and the individual chromogens and fluorochromes on the basis of the pure spectra.

Figure 1 shows a sample breast cancer specimen. The brightfield image (Figure 1 A) shows haematoxylin staining. Figure 1B shows the haematoxylin channel, unmixed using its spectral signature (Figure 1F), revealing cell nuclei. Such unmixed channels are ideal for automated segmentation, because they are monochrome and often contain only one type of object. Figure 1C shows the channel corresponding to CK fluorescent staining, which reveals the cytoplasmic domain of cells of epithelial origin. Figure 1D shows the channel corresponding to HER2 fluorescent staining, which reveals the plasma membrane of breast cancer cells expressing this biomarker. We use this image as a running example to illustrate the segmentation methods and process.

Image Analysis Overview

Our segmentation strategy focuses on cells whose nuclei are visible in the nuclear channel, because they mark individual cells—these are segmented first. Second, the cytosolic boundaries of cells whose nuclei are detected are segmented on the basis of markers and geometric constraints. The third step quantifies cell and nuclear morphologies, and measures biomarker expression over cellular compartments. Using these data, we identify cell types, classify cells as being positive/negative for antigens, and organize the measurements by cell type and subcellular compartment.

Automated segmentation of cell nuclei—We used our fully automated segmentation algorithm,²⁰ which is an improvement on the prior literature.^{21–35} Importantly, it is capable of automatic selection of parameter settings. It starts by binarizing the image, using the graph-cuts method, with automatic learning of foreground and background intensity profiles, using minimum error thresholding.^{36,37} Next, a multiscale Laplacian of Gaussian filter, with automatic and adaptive scale selection,²⁰ is used to identify nuclear centres. These points are used to generate an initial segmentation³⁸ that is refined using a multilabel graph-cuts algorithm with alpha-expansions³⁹ and graph-colouring.⁴⁰ Figure 2A shows sample automated segmentation results for the image in Figure 1 as red outlines overlaid on the nuclear channel displayed in greyscale. The green dots indicate nuclear centres whose locations and identifiers (IDs) are used in subsequent steps. Given the importance of this step, the user is provided with graphical tools to inspect the results and correct any errors before proceeding to the next step.

Automated delineation of cytoplasmic domains—This step generates the spatial mask for associating cytoplasmic markers with individual cells, using a mix of cues from cytoplasmic and membrane markers and geometric constraints. For example, CKs are found in the intracytoplasmic cytoskeleton of cells of epithelial origin (e.g. carcinoma cells in

Figure 1C), so they indicate cytoplasmic domains of a selected cell population. Cytoplasmic markers often highlight connected multicellular clusters that must be subdelineated into individual cells to permit cell-by-cell analysis. The cues for this subdelineation vary. Sometimes, it is possible to highlight cell boundaries by staining for a membrane-associated antigen, e.g. E-cadherin or EMA. Some analytes also can highlight membranes of cells, e.g. HER2 (Figure 2D). However, membrane labelling is often unreliable: HER2 is not always overexpressed, and E-cadherin expression can be lost in some cancers. Even when good cytoplasmic and membrane-bound markers are available, some ambiguities arise because histopathology slides are sections of three-dimensional specimens, and the sectioning plane cannot be planned accurately. For instance, the membranes of cells may be visible but not the nuclei, or the membrane signal can appear over a nucleus, appearing to cut across it. Finally, cells within a sample can show a variable degree of staining. Overall, cytoplasmic segmentation algorithms must be capable of coping with variable cues. Our strategy is to avoid direct segmentation of the cytoplasmic/membrane channels. Instead, we leverage the validated nuclear segmentations and build an adaptive algorithm that exploits cues in the cytoplasmic and/or membrane channels, when they are available, and that defaults to geometric constraints when they are inadequate. It automatically switches between two modes (defined below) on a cell-by-cell basis.

Mode 0: This applies to cells with detectable cytoplasmic and/or membrane marker. The cytoplasmic channel pixels $I_C(x,y)$ are automatically and adaptively binarized to separate the foreground and background, using the graph-cuts algorithm.^{36,37} Morphological opening and closing operators (radius = 3 pixels) are used to fill holes. If the membrane channel, $I_M(x,y)$, is available, the magnitude of its smoothed intensity gradient, $G_\sigma(x,y) = |X_\sigma I_M(x,y)|$, is computed by convolving $I_M(x,y)$ with the derivative of a Gaussian with $\sigma = 1.25$ pixels (fixed for a given magnification). If the membrane channel is unavailable, we compute $G_\sigma(x,y) = |X_\sigma I_C(x,y)|$ instead. The cues from the cytoplasmic and membrane channels are integrated with geometric distances by computing a gradient-enhanced distance map, $S(x,y)$, with respect to the segmented nuclei. This is used to compare the cue-adjusted proximity of each pixel to nuclei. If $d(i,j)$ denotes the Euclidean distance between neighbouring foreground pixels $i = (x_i, y_i)$ and $j = (x_j, y_j)$, the adjusted distance between them is $d(i,j) \times |G_\sigma(x_i, y_i) - G_\sigma(x_j, y_j)|$. The adjusted distance between non-neighbouring points $u_1 = (x_1, y_1)$ and $u_n = (x_n, y_n)$ is weighted by the length of the shortest path (with eight-neighbour connectivity) connecting them. The value at each cytoplasmic foreground point in $S(x,y)$ is set to the minimum of all of the adjusted distances from (x_F, y_F) to all of the nuclear boundary points that are connected by a path over foreground points. With the nuclei as the initial markers, a marker-controlled watershed transform⁴¹ is computed on $S(x,y)$. This produces a reliable segmentation of the cytoplasmic foreground into subregions, with one cytoplasmic region per segmented nucleus. Figure 2B shows sample cell segmentations of CK-positive cells using Mode 0 with the gradient information $G_\sigma(x,y)$ from the membrane channel. Figures 3, 5, 6 and S2 exemplify segmentations without the benefit of the membrane signal.

Mode 1: This is a geometric estimation that is invoked for cells for which cytoplasmic and membrane labels are unavailable (e.g. stromal cells that are CK-negative). The traditional geometric approach based on Voronoi diagrams^{42,43} produces unacceptably coarse polygonal approximations, so we use the Hamilton–Jacobi Generalized Voronoi Diagram (HJ-GVD),⁴⁴ which uses the Euclidean distance from segmented nuclear boundaries instead of their centroids, to produce more refined estimates. We impose a radius constraint $r_{\max}$ on the HJ-GVD to prevent unrealistically large cell domain estimates. Figure 2C shows sample results for the HER2 example, using $r_{\max} = 12$ pixels. The estimated cell boundaries are overlaid on the Euclidean distance map $D(x,y)$. Although these geometric estimates do not

reflect the cellular reality (the structures are unobservable), they are helpful for approximately associating extranuclear markers with cells when the limitations of immunostaining do not permit additional labels for cytoplasmic and membrane markers.

Morphological measurements of cells—From the nuclear and cytoplasmic segmentations, we compute cell features, including locations, areas, shape factors, boundary curvatures, convexity, eccentricity, radius variation, orientation, and various texture measures [average intensity, intensity variation, skew of intensity distribution, energy of intensity distribution, entropy of intensity distribution, interior gradient, and ratios of intensity values (e.g. max./min.)].⁴⁵ Not all features are needed for a given analysis, and the user can choose an appropriate subset. The cytoplasmic segmentation step produces one cytoplasmic domain per segmented nucleus, so the nuclear IDs are used for tabulating nuclear and cytoplasmic measurements.

Biomarker measurements of cells—Next, molecular biomarkers are quantified by measuring their distribution over cellular regions of interest (masks) defined by segmentation. Figure 2E shows a close-up view of these regions for an individual cell. The red outline shows the intranuclear compartment, the light blue contours delineate the intracytoplasmic compartment, and the orange contour runs parallel to the cell membrane outline (blue) separated by a fixed distance (five pixels).

Quantifying nuclear biomarkers: Directly summing the analyte signal over intranuclear compartments is naïve, as it does not correct for background fluorescence. Even when they appear dim, background pixels can add up to a significant sum over a region. To address this problem, we first perform an automatic two-level or three-level segmentation of the analyte channel.⁴⁶ When the contrast between the analyte-positive pixels and analyte-negative pixels is high, a two-level binarization separates the bright foreground from definite background pixels. When the analyte exhibits an intermediate background, a three-level binarization (e.g. Figure 4) segregates pixels into bright foreground, intermediate background, and dark background. Only the bright foreground pixels are used for analyte association. Figure S2 illustrates these steps for quantifying ER in a breast cancer specimen. Figure S2D shows the three-level binarization for background correction.

Quantifying cytoplasmic markers: Integration of markers over the cytoplasmic region proceeds as for nuclei—the background-corrected analyte signal is integrated over the cytoplasmic region of interest. In Figure 2E, the cytoplasmic region of integration is enclosed by the blue outlines, but excluding the intranuclear region.

Quantifying plasma membrane-bound markers: This computation must cope with the possibility of an unreliable membrane label that does not clearly and completely define the cytoplasmic domain of each cell. Happily, our cytoplasmic segmentation is designed to produce closed contours representing the best possible estimates of cell membrane locations based on available cues. When a user determines that the membrane signal is sufficiently reliable, membrane-bound analytes can be integrated within a narrow strip (typically five pixels wide) of the segmented membrane. When the locations of cytoplasm and plasma membrane markers are superimposed or extensively overlap, the integration is carried out over the entire cell domain, with background correction. The resulting biomarker measurements must be interpreted with care, as our images represent planar projections of subcellular compartments with finite thickness. When assigning analyte expression to subcellular compartments, one must acknowledge that these two compartments cannot be perfectly distinguished or separated in the images being analysed. Nevertheless, these

measurements are adequate from the standpoint of labelling cells as being positive/negative for membrane-bound antigens, and for statistical analysis.

Cell type identification—This step identifies whether a cell is of a specified type on the basis of its morphological and associative features. We use a supervised approach, where the user indicates a training set (containing examples of both classes from one or more images), from which a Bayesian classifier is constructed. Figure 2D illustrates cell classification results for the sample image shown in Figure 1, based on the CK signal. Yellow dots represent cells that are CK-positive and HER2-positive, and white dots represent other cells.

Results

FARSIGHT (www.farsight-toolkit.org) was written with the use of standard software tools (C++, ITK, VTK, QT), and allows a user to perform automated segmentation, view and edit the results, compute morphological and associative features, classify cells, and export the results to spreadsheets. It is both a free and an open source. Each row of the output corresponds to one numbered cell in the image. The software was validated in two ways. First, its results were compared with determinations made by a human expert. Another validation was based on cells cultured *in vitro*, labelled with different fluorochromes, and mixed in different ratios to create cell blocks from which slides were cut for fluorescence imaging and analysis. Specifically, cultured cells were labelled with the membrane dye PKH26, or with a combination of PKH26 and PKH67. The PKH26 cells and PKH26/PKH67 cells were mixed in different ratios (10:0, 9:1, 2:1, 1:1, 1:2, 1:9, and 0:10), fixed, and frozen in OCT embedding medium. Slides cut from these cell blocks were stained with 4',6-diamidino-2-phenylindole to reveal nuclei and membrane proteins PKH26 and PKH67. The details of the protocols and results are given in Doc. S1. Ten images (×400) were taken of slides from each block, and processed by **FARSIGHT** to segment cells, classify them as PKH67-negative or PKH67-positive, and compute the ratio of the two cell populations. The results were in concordance with the scoring of a human expert (Table A.1 in Doc. S1). The averages of cell proportions determined by **FARSIGHT** closely approximated the known truth (Figure A.1 in Doc. S1). We then proceeded to evaluate **FARSIGHT** for human breast histopathology samples.

Cell Membrane Analyte (HER2)

Figure 2 shows our analysis of the image in Figure 1. The histogram in Figure 2F shows the distribution of HER2 in the cells. The cut-off value was 12.6 greyscale units, at which 98.5% of the tumour cells (CK-positive cells) are HER2-positive. These data concord with an expert human reading of 99%. In some cases, HER2 staining was not also usable for cell boundary determination [e.g. HER2 staining overlays cell nuclei or is extremely dark and thick (Figure S1)], so the cell boundaries were estimated geometrically (Mode 1).

Nuclear Analytes (ER, PR, and Ki67)

We applied our methodology to specimens stained for three common nuclear-bound markers, ER, PR, and Ki67. Figure S2 shows the detailed steps for the ER case—the steps were identical for PR and Ki67. Figure 3 shows the results for breast cancer specimens stained for ER (Figure 3A,B), PR (Figure 3C,D), and Ki67 (Figure 3E,F), respectively. As a crosscheck, we computed the nuclear/cytoplasmic level ratios of the analytes for every cell. Histograms of these ratios (Figure 3B,D,F) show that these analytes were strongly nuclear-bound, as expected for antigens that are located in nuclei. The automatically determined percentages of ER-positive, PR-positive and Ki67-positive cells were 39%, 40% and 27% of the CK-positive cells, as compared with expert-determined percentages of 38%, 39% and 26%, respectively. For comparison, pixel-level analysis to determine the percentage of haematoxylin-positive pixels (the image area occupied by nuclei) that were also ER-

positive, PR-positive or Ki67-positive yielded 17.3%, 28.5% and 14.5%, respectively. Clearly, area measurements do not reflect cell numbers.

Figure 4 illustrates analysis of chosen subpopulations of cells. To measure cell proliferation and its relationship with activity of the Raf–MEK–ERK signaling pathway, a human breast carcinoma was immunostained for Ki67, p-ERK, and CK. CK staining revealed a cluster of carcinoma cells to the right, but these constituted a minority of the cells; the majority were lymphocytes within a reactive lymphoid nodule. Ki67 immunostaining showed that 34.6% of all cells were proliferating. For comparison, pixel-level analysis showed that 16.7% of haematoxylin-positive pixels were Ki67-positive. Only 2.1% (2/96) of carcinoma cells were Ki67-positive, whereas 37.8% (414/1094 stromal cells) were Ki67-positive. Thus, the total number or percentage of Ki67-positive cells did not accurately report tumour cell proliferative activity. This demonstrates that a cell-based method with the ability to type cells as tumour or stromal prior to analyte quantification is important for characterizing human tumours, where the cellular composition is always heterogeneous, and tumour cells may not predominate. Further analysis to examine the correlation between Raf–MEK–ERK signalling and proliferation showed a high coefficient ($R = 0.89$) between p-ERK and Ki67 expression in cells (Figure 4E). This suggests that ERK activation and proliferation may be linked events among the cells in this image. This is expected, as the majority of proliferating cells are lymphocytes, and ERK activation has been shown to accompany mitogenic activation of lymphocytes *in vitro*.⁴⁷ Because of the low frequency of Ki67 and p-ERK positivity among CK-positive cells in this image, little can be learned from it about the concurrence of ERK activation and proliferation in carcinoma cells (Figure 4F).

To examine the relationship between ERK activation and proliferation in breast cancer cells, another region of the same tumour (Figure 5A–C) and a region of a second, similarly stained tumour (Figure 5G–I) were analysed. In both fields, tumour cells were in the majority, and a significant fraction were Ki67-positive (10% for tumour 1; 7.5% for tumour 2). Scatter plots of p-ERK and Ki67 expression in individual cells revealed that the correlation between p-ERK and Ki67 staining was lower among the CK-positive carcinoma cells of tumour 1 ($R = 0.59$) and tumour 2 ($R = 0.29$) than among the reactive lymphocytes in tumour 1 (Figure 4, $R = 0.89$). On the basis of these images, the link between ERK activation and cell proliferation appears to be weaker in the tumour cells than in the reactive lymphocytes, illustrating the utility of specific cell-level analysis as a research tool.

The ability of our method to separate each cell into nuclear and extranuclear compartments is valuable. Figure 6 shows a breast tumour that was stained with antibodies to p-S6 (the activated form of ribosomal protein S6), CK and EMA, all by immunofluorescence, and counterstained with haematoxylin. Figure 6D shows cell segmentation and classification results with yellow contours outlining the cytoplasmic boundaries of CK-positive cells determined by use of the CK and EMA channels jointly. The subpopulation of CK-positive cells that were p-S6-positive was in the minority (11%) in this tumour (for comparison, pixel-based analysis showed that 8.9% of CK-positive pixels were p-S6-positive). Visual examination of the p-S6-positive cells shows that p-S6 staining, as expected, was predominantly cytoplasmic. This was confirmed by plotting a histogram of the extranuclear/nuclear ratio of p-S6 signal in cells that expressed this antigen (Figure 6F), which showed that only 10% of p-S6 signal was nuclear. This small amount of ‘nuclear’ p-S6 may be explained by the fact that the image represents a planar projection of a tumour section that is 5 μm thick; p-S6 staining in cell cytoplasm situated above or below nuclei in these sections would register as nuclear.

Discussion

The 'histocytometric' analyses performed by FARSIGHT on the images shown demonstrate the practicality and value of quantifying molecular analytes on a cellular scale with cell type and subcellular compartment specificity. Although these studies focused on breast cancer, our methodology and tools are applicable to other cancers and conditions. Our approach requires more extensive immunostaining and sophisticated imaging than traditional visual histopathology, but offers important benefits. It reveals the type, distribution, intrinsic characteristics and biomarker state of each cell in its tissue context. It allows multiple biomarkers to be quantified selectively over specified cell types, regardless of their abundance. Our efforts were focused on quantifying analytes in tumour cells, but stromal cells (endothelial cells, fibroblasts, lymphocytes, macrophages, etc.) are omnipresent in tumours and are gaining attention for their contributions to malignant progression and behaviour.^{48,49} The ability of histocytometry to specify the cell type for analysis makes it a sensitive and specific tool for investigating minority stromal cell subpopulations, whose attributes would otherwise be overshadowed by more abundant cell types.

Our cell-based method shares some advantages with pixel-level analysis, such as objectivity, reproducibility, and the ability to quantify on a continuous scale. However, by using the cell as the unit of analysis, it generates additional and potentially complementary measurements that are expressible in terms of cell counts and cell types. Such measurements are unaffected by the area occupied by cells and other tissue structures in the image. Although the two types of measurement can be correlated for some samples, they can differ greatly for others, as shown by our examples. For analysis of histopathology specimens, both methods are usable diagnostically, but we believe that event reporting by cell number or percentage is biologically more informative, as reflected in the fact that it is the preferred form of reporting for many *in-vitro* cellular studies. Our software system makes it possible to generate these reports.

Histocytometry correctly assigns analytes to appropriate subcellular locations within one cell (e.g. a nuclear analyte and a cytoplasmic analyte) to the same unit. Results so organized have obvious benefits, particularly when investigating tissues for biological processes and events that occur, or are regulated, at the level of individual cells but involve different subcellular compartments. This feature of FARSIGHT analysis also provides the ability to examine and quantify analytes in different compartments of cells. This is an advantage when studying analytes whose subcellular location, by itself, is informative about activity. For example, the transcription factor nuclear factor kappaB (NF- κ B) is kept transcriptionally inactive when it is constrained in the cytoplasm through binding to its inhibitor, I κ B. NF- κ B becomes active upon its translocation to the nucleus following stimuli that induce release from and degradation of I κ B.⁵⁰ An extension of this is the study of yet other analytes that produce different effects, depending on whether they are localized to the cytoplasm or nucleus. Finally, by providing analyte data for each cell in an image, rather than one result for the image as a whole, FARSIGHT analysis can reveal population characteristics, such as analyte range, distribution, and variance among cells, that can be additionally informative. Histocytometry can provide information similar to that provided by flow cytometry, with the added benefit of preserving tissue architecture, which allows concurrent examination of morphological features and quantification of spatial relationships and distributions that is not possible with the dissociated cells used for flow cytometry.

We developed our multiplex immunostaining protocols for the study of formalin-fixed, paraffin-embedded histopathology specimens. This allows histocytometric analysis to be performed on the tissue material most commonly available from cancer patients and most often stored in pathology archives. However, frozen and other forms of preserved tissues are

also suitable for this type of analysis; their study only requires development of appropriate immunostaining protocols. These protocols have involved immunostaining for four or more antigens on the same slide to study a single analyte. This level of complexity stems from the need to stain for cell type, subcellular compartments and analyte antigens on the same slide. Some of this complexity may be reduced by algorithms for direct multispectral identification of tumour cells and tumour areas in slides stained only with haematoxylin and eosin (H&E). For tumour cell analysis, computer-generated ‘tumour masks’ may eliminate the need to immunostain for cell type and compartment antigens. By combination of the use of tumour masks with cell segmentation based on geometric algorithms, histocytometric analysis may be performed on slides stained only for analyte and H&E, such as breast cancer specimens stained for ER, PR and HER2 in hospital pathology laboratories. Although the utility of developing methods for histocytometric analysis of simply stained slides is primarily clinical, expanding the current limits of immunostain multiplexing will make histocytometry an even more potent instrument for biological research. FARSIGHT can also be applied to H&E-stained sections, but the caveat rests with the fluorescence of eosin, which must be properly accounted for in the spectral unmixing. This will allow study of numerous analytes on the same slide. Accompanied by FARSIGHT cell-based quantification of their expression, this will enable examination of complex patterns of signalling pathway activity and other molecular events in cells in an authentic tissue context. Although our examples did not show analysis of multiple cell types, the system itself is capable of such analysis, and we expect to report validation of this capability in subsequent articles. As part of our effort to hasten the development and advancement of this histopathology analysis platform, FARSIGHT has been made available as a free and open source software system (www.farsight-toolkit.org). In the future, we expect this system to be adapted to automated analysis of larger batches of specimens, which may be multiplex-stained by automated systems, and whole-slide scanning.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Various portions of this work were supported by S-IDEA grant W81XWH-07-1-0325 from the US Army Breast Cancer Research Program, NIH grant R01 EB005157, NIH grant R01 CA135509, and NSF grant EEC-9986821. Portions of this project were funded by a grant from the Pennsylvania Department of Health. The Department specifically disclaims responsibility for any analysis, interpretations, or conclusions. The authors thank Dr Cliff Hoyt (CRi, Woburn, MA, USA) for helpful discussions, and the staff at CRi Inc. for technical assistance.

References

1. Cheng J, Qiu S, Raju U, Wolman SR, Worsham MJ. Benign breast disease heterogeneity: association with histopathology, age, and ethnicity. *Breast Cancer Res Treat.* 2008; 111:289–296. [PubMed: 17917807]
2. Hammerschmied CG, Walter B, Hartmann A. Renal cell carcinoma 2008: histopathology, molecular genetics and new therapeutic options. *Pathologie.* 2008; 29:354–363. [PubMed: 18592242]
3. Crocker, J.; Murray, P. *Molecular biology in cellular pathology.* Chichester: Wiley; 2003.
4. Research Signpost (Trivandrum India). Recent research developments in histopathology. Trivandrum, India: Research Signpost; 2002. v
5. Menard S, Balsari A, Tagliabue E, et al. Biology, prognosis and response to therapy of breast carcinomas according to HER2 score. *Ann Oncol.* 2008; 19:1706–1712. [PubMed: 18544559]
6. Chivukula M, Bhargava R, Brufsky A, Surti U, Dabbs DJ. Clinical importance of HER2 immunohistologic heterogeneous expression in core-needle biopsies vs resection specimens for

- equivocal (immunohistochemical score 2+) cases. *Mod Pathol*. 2008; 21:363–368. [PubMed: 18246053]
7. Montironi R, Mazzucchelli R, Barbisan F, et al. HER2 expression and gene amplification in pT2a Gleason score 6 prostate cancer incidentally detected in cystoprostatectomies: comparison with clinically detected androgen-dependent and androgen-independent cancer. *Hum Pathol*. 2006; 37:1137–1144. [PubMed: 16938518]
 8. Bowles DW, Rabinovitch R, Borges V, et al. A young woman with a small ER-positive breast cancer, a micrometastatic axillary lymph node, and an intermediate oncotype DX recurrence score. *Oncology (Williston Park)*. 2007; 21:1212–1217. [PubMed: 17926799]
 9. Remmele W, Stegner HE. Recommendation for uniform definition of an immunoreactive score (IRS) for immunohistochemical estrogen receptor detection (ER-ICA) in breast cancer tissue. *Pathologie*. 1987; 8:138–140. [PubMed: 3303008]
 10. Neudert M, Fischer C, Krempien B, Seibel MJ, Bauss F. A rapid histological score for the semiquantitative assessment of bone metastases in experimental models of breast cancer. *Onkologie*. 2008; 31:521–527. [PubMed: 18854651]
 11. Egyed Z, Jaray B, Kulka J, Pentek Z. Triple test score for the evaluation of invasive ductal and lobular breast cancer. *Pathol Oncol Res*. 2008
 12. Kurosumi M. Immunohistochemical assessment of hormone receptor status using a new scoring system (J-Score) in breast cancer. *Breast Cancer*. 2007; 14:189–193. [PubMed: 17485905]
 13. Remmele W, Schickelanz KH. Immunohistochemical determination of estrogen and progesterone receptor content in human breast cancer Computer-assisted image analysis (QIC score) vs subjective grading (IRS). *Pathol Res Pract*. 1993; 189:862–866. [PubMed: 8302707]
 14. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nat Med*. 2002; 8:1323–1327. [PubMed: 12389040]
 15. Mulrane L, Rexhepaj E, Penney S, Callanan JJ, Gallagher WM. Automated image analysis in histopathology: a valuable tool in medical diagnostics. *Expert Rev Mol Diagn*. 2008; 8:707–725. [PubMed: 18999923]
 16. Taylor CR, Levenson RM. Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment II. *Histopathology*. 2006; 49:411–424. [PubMed: 16978205]
 17. Tamai S. Expert systems and automatic diagnostic systems in histopathology—a review. *Rinsho Byori*. 1999; 47:126–131. [PubMed: 10097629]
 18. Peterson RA, Krull DL, Butler L. Applications of laser scanning cytometry in immunohistochemistry and routine histopathology. *Toxicol Pathol*. 2008; 36:117–132. [PubMed: 18337230]
 19. Davis DW, Takamori R, Raut CP, et al. Pharmacodynamic analysis of target inhibition and endothelial cell death in tumors treated with the vascular endothelial growth factor receptor antagonists SU5416 or SU6668. *Clin Cancer Res*. 2005; 11(2 Pt 1):678–689. [PubMed: 15701856]
 20. Al-Kofahi, Y.; Lassoued, W.; Lee, W.; Roysam, B. IEEE Trans Biomed Eng. 2009. Improved automatic detection & segmentation of cell nuclei in histopathology images. in review
 21. Chang H, Yang Q, Parvin B. Segmentation of heterogeneous blob objects through voting and level set formulation. *Pattern Recognition Lett*. 2007; 28:1781–1787.
 22. Parvin B, Yang Q, Han J, et al. Iterative voting for inference of structural saliency and characterization of subcellular events. *IEEE Trans Image Processing*. 2007; 16:615–623.
 23. Wahlby C, Sintorn IM, Erlandsson F, Borgefors G, Bengtsson E. Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. *J Microsc*. 2004; 215:67–76. [PubMed: 15230877]
 24. Ortiz de Solorzano C, Garcia Rodriguez E, Jones A, et al. Segmentation of confocal microscope images of cell nuclei in thick tissue sections. *J Microsc*. 1999; 193(Pt 3):212–226. [PubMed: 10199001]
 25. Lin G, Chawla MK, Olson K, et al. A multi-model approach to simultaneous segmentation and classification of heterogeneous populations of cell nuclei in 3D confocal microscope images. *Cytometry A*. 2007; 71:724–736. [PubMed: 17654650]

26. Lin G, Chawla MK, Olson K, et al. Hierarchical, model-based merging of multiple fragments for improved three-dimensional segmentation of nuclei. *Cytometry A*. 2005; 63:20–33. [PubMed: 15584021]
27. Mackin RW Jr, Newton LM, Turner JN, Holmes TJ, Roysam B. Accuracy of nuclear classification in cervical smear images Quantitative impact of computational deconvolution and 3-D feature computation. *Anal Quant Cytol Histol*. 1998; 20:77–91. [PubMed: 9569965]
28. Mackin RW Jr, Newton LM, Turner JN, Roysam B. Advances in high-speed, three-dimensional imaging and automated segmentation algorithms for thick and overlapped clusters in cytologic preparations Application to cervical smears. *Anal Quant Cytol Histol*. 1998; 20:105–121. [PubMed: 9569968]
29. Ancin H, Roysam B, Dufresne TE, et al. Advances in automated 3-D image analyses of cell populations imaged by confocal microscopy. *Cytometry*. 1996; 25:221–234. [PubMed: 8914819]
30. Byun JY, Verardo MR, Sumengen B, et al. Automated tool for the detection of cell nuclei in digital microscopic images: application to retinal images. *Mol Vision*. 2006; 12:949–960.
31. Chawla MK, Lin G, Olson K, et al. 3D-catFISH: a system for automated quantitative three-dimensional compartmental analysis of temporal gene transcription activity imaged by fluorescence in situ hybridization. *J Neurosci Methods*. 2004; 139:13–24. [PubMed: 15351517]
32. De Solorzano CO, Malladi R, Lelievre SA, Lockett SJ. Segmentation of nuclei and cells using membrane related protein markers. *J Microsc*. 2001; 201:404–415. [PubMed: 11240857]
33. Gudla PR, Nandy K, Collins J, et al. A high-throughput system for segmenting nuclei using multiscale techniques. *Cytometry A*. 2008; 73:451–466. [PubMed: 18338778]
34. Li G, Liu TM, Nie JX, et al. Detection of blob objects in microscopic zebrafish images based on gradient vector diffusion. *Cytometry Part A*. 2007; 71A:835–845.
35. Lin G, Adiga U, Olson K, et al. A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks. *Cytometry A*. 2003; 56:23–36. [PubMed: 14566936]
36. Kittler J, Illingworth J. Minimum error thresholding. *Pattern Recognition*. 1986; 19:41–47.
37. Pal NR, Pal SK. Image model, Poisson distribution and object extraction. *Int J Pattern Recognition Artificial Intelligence*. 1991; 5:25.
38. Wu XW, Chen YD, Brooks BR, Su YA. The local maximum clustering method and its application in microarray gene expression data analysis. *Eurasip J Appl Signal Processing*. 2004; 2004:53–63.
39. Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. *IEEE Trans Pattern Analysis Machine Intelligence*. 2001; 23:1222–1239.
40. Nath SK, Bunyak F, Palaniappan K. Robust tracking of migrating cells using four-color level set segmentation. *Lect Notes Comput Sci*. 2006; 4179(LNCS):920–932. [PubMed: 18650960]
41. Vincent L, Soille P. Watersheds in digital spaces—an efficient algorithm based on immersion simulations. *IEEE Trans Pattern Analysis Machine Intelligence*. 1991; 13:583–598.
42. Bertin E, Parazza F, Chassery JM. Segmentation and measurement based on 3D Voronoi diagram: application to confocal microscopy. *Comput Med Imaging Graph*. 1993; 17:175–182. [PubMed: 8402524]
43. Sanchez-Marin FJ. A simple procedure for simulating samples of tissue using Voronoi diagrams. *Anal Quant Cytol Histol*. 2005; 27:225–231. [PubMed: 16220834]
44. Nath SK, Palaniappan K, Bunyak F. Accurate spatial neighborhood relationships for arbitrarily-shaped objects using Hamilton–Jacobi GVD. *Lect Notes Comput Sci*. 2007; 4522(LNCS):421–431. [PubMed: 18820751]
45. Bjornsson CS, Lin G, Al-Kofahi Y, et al. Associative image analysis: a method for automated quantification of 3D multi-parameter images of brain tissue. *J Neurosci Methods*. 2008; 170:165–178. [PubMed: 18294697]
46. Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved automatic detection & segmentation of cell nuclei in histopathology images. *IEEE Trans Biomed Eng*. 2009
47. Cantrell D. T cell antigen receptor signal transduction pathways. *Annu Rev Immunol*. 1996; 14:259–274. [PubMed: 8717515]
48. Kerbel RS. Tumor angiogenesis. *N Engl J Med*. 2008; 358:2039–2049. [PubMed: 18463380]

49. Orimo A, Gupta PB, Sgroi DC, et al. Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell*. 2005; 121:335–348. [PubMed: 15882617]
50. Baeuerle PA, Baltimore D. I kappa B: a specific inhibitor of the NF-kappa B transcription factor. *Science*. 1988; 242:540–546. [PubMed: 3140380]

Abbreviations

CK	cytokeratin
EMA	epithelial membrane antigen
ER	oestrogen receptor
ERK	extracellular signal-related kinase
H&E	haematoxylin and eosin
HJ-GVD	Hamilton–Jacobi Generalized Voronoi Diagram
HRE2	human epidermal growth factor receptor
2HRP	horseradish peroxidase
ID	identifier
NF-κB	nuclear factor kappaB
PR	progesterone receptor

Figure 1.

Multiplex-stained human breast cancer specimen. A human breast cancer was stained for HER2 by immunofluorescence with Texas Red and for cytokeratin by immunofluorescence with Alexa-488, and counterstained with haematoxylin. The slide was imaged multispectrally in absorption and fluorescence modes, and the results were unmixed to yield non-overlapping channels. **A**, Brightfield image showing haematoxylin staining. **B**, Unmixed channel containing only cell nuclei, corresponding to the haematoxylin spectral signature. **C**, Unmixed channel for fluorescently stained cytokeratin. **D**, Unmixed channel corresponding to fluorescently stained HER2. **E**, Composite three-colour image with nuclei

(red), cytokeratin (green), and HER2 (blue). **F**, Spectral signatures used for the unmixing computations, displayed using blue for haematoxylin (nuclei), green for Alexa-488 (cytokeratin), and red for Texas Red (HER2).

Figure 2.

Automated image analysis steps for the specimen in Figure 1. **A**, Automatic nuclear segmentation (red outlines) of the nuclear channel. **B**, Estimated cytoplasmic domains for cytokeratin-positive cells for the boxed region in **D** overlaid on the gradient-enhanced distance map (Mode 0). **C**, Geometrically estimated cytoplasmic domains for stromal cells in the same region overlaid on the underlying dominance map (Mode 1). **D**, Composite cell segmentation and classification results, with yellow dots indicating cells that are cytokeratin-positive and HER2-positive, and white dots indicating other cells. **E**, Close-up

illustrating regions of interest used to quantify HER2. **F**, Histogram summary showing the cut-off point for declaring cells HER2-positive.

Figure 3.

Examples showing analysis of breast cancer specimens stained for three nuclear-bound biomarkers. Breast cancer slides were immunostained for oestrogen receptor (**A,B**), progesterone receptor (**C,D**), or Ki67 (**E,F**) plus cytokeratin, and counterstained with haematoxylin. Images were captured, and nuclear and whole-cell segmentation was performed, with yellow dots indicating the nuclei positive for the respective analytes (**A,C,E**). Analyte was quantified in the nuclear and extranuclear compartments of each cell, and histograms of the nuclear/extranuclear analyte level ratios in all positive cells are shown (**B,D,F**).

Figure 4.

Duplex analysis of phospho-extracellular signal-related kinase (p-ERK) and Ki67 immunostaining in lymphoid cells in a human breast carcinoma. A section of a breast tumour was stained sequentially with anti-p-ERK (SG Blue), anti-Ki67 [3,3'-diaminobenzidine (DAB)] and anti-CK (Alexa-488) antibodies, and this was followed by haematoxylin staining, multispectral imaging ($\times 400$), and cytometric analysis. The brightfield image of a lymphoid nodule in the tumour is shown (A), along with the unmixed channels for DAB (Ki67) (B), SG Blue (p-ERK) (C), and Alexa-488 (cytokeratin) (D).

Scatter plots of p-ERK (*x*-axis) and Ki67 (*y*-axis) staining intensity are shown for cells in the lymphoid nodule (**E**) and for tumour cells (**F**), with each dot representing one cell.

Figure 5.

Duplex analysis of phospho-extracellular signal-related kinase (p-ERK) and Ki67 immunostaining in human breast carcinoma cells. Sections of two different breast tumours were stained and analysed as described for Figure 5. Brightfield images of the two different tumours are shown (A,G), with the unmixed channels for 3,3-diaminobenzidine (Ki67) (B,H) and SG Blue (p-ERK) (C,I). Composite images showing whole-cell segmentation of the tumour (cytokeratin-positive) cells are shown (D,J). Scatter plots of p-ERK (x-axis) and Ki67 (y-axis) staining intensity are shown for tumour cells (E,K) and for non-tumour (stromal) cells (F,L), with each dot representing one cell.

Figure 6.

Analysis of phospho-S6 immunostaining in a human breast cancer. A section of a breast tumour was stained with anti-p-S6 (Alexa-488), anti-epithelial membrane antigen (EMA) (Alexa-594) and anti-cytokeratin (CK) (Alexa-555), and this was followed by haematoxylin staining, multispectral imaging ($\times 400$), and cytometric analysis. The brightfield image is shown (A), along with unmixed channels for Alexa-555 (CK) (B) and Alexa-594 (EMA) (C). Composite images of p-S6 analyte staining along with segmented whole tumour cells are shown (D; E shows an enlargement of the boxed area in D). In each cell, analyte in the

nuclear and extranuclear compartment was quantified. Extranuclear/nuclear analyte ratios were calculated for each positive tumour cell, and their distribution is shown (**F**).