

## EXPERT OPINION

# Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment II

C R Taylor & R M Levenson<sup>1</sup>

Department of Pathology, Keck School of Medicine of the University of Southern California, Los Angeles, CA and

<sup>1</sup>Cambridge Research and Instrumentation Inc., Woburn, MA, USA

---

Taylor C R & Levenson R M

(2006) *Histopathology* **49**, 411–424

## Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment II

Immunohistochemistry is entering its fourth decade of use on formalin-fixed paraffin-embedded tissues. Over this period the method has evolved to become a major part of the practice of diagnostic surgical pathology worldwide. From the beginning immunohistochemistry has been adapted to provide a range of markers of cell lineage and tissue type, with particular application to the diagnosis and classification of tumours. In this modality immunohistochemical methods were employed simply as 'special stains', the results of which were evaluated qualitatively by the pathologist, as for any other stain. More recently, attention has shifted to the demonstration of prognostic markers in tumour

cells, driven by the advent of molecular biology and the discovery of numerous regulatory molecules, coupled with manufacture of the corresponding specific antibodies. Immunohistochemistry has rapidly adapted to this new use, but in so doing the demand for quantification has become paramount; it is no longer enough that the 'stain' is there; rather it is a question of 'How much is there?'. This review explores the limitations of immunohistochemistry when employed in a semiquantitative mode, and explores the possibility of fulfilling the full potential of immunohistochemistry, as a true quantitative immunoassay applied in a tissue section environment.

**Keywords:** computer-assisted image analysis, immunohistochemistry, quantification, reference standards, spectral imaging, standardization

### Definitions

**Quantity** (noun): 1. a certain amount or number; 2. the property of something that is measurable in number, amount, size or weight; 3. a considerable number or amount (from Latin, *quantitas*, how much?).

**Quantitative** (alt. quantitive) (adjective): of, concerned with, or measured by, quantity.

(*Oxford Dictionary Compact Edition*, Oxford University Press, 2002)

---

## Introduction

The term 'semiquantitative' lacks clear definition, but would imply having some of the features of 'quantitative', as in 'semiprecious', or not quite precious.

While these definitions have some clarity in certain contexts, the use of the term 'quantitative' in anatom-

ical pathology is uncommon and inconsistent. By way of contrast, within the Clinical Laboratory many assays are quantitative and the characteristics that make up a quantitative assay can there be examined at leisure.

Anatomical pathology (surgical pathology, histopathology) *per se* is primarily observational, dependent upon pattern recognition in its broadest sense, without overt acknowledgement that within the context of pattern recognition there are elements that are quantitative. Biological stains, introduced in the mid-19th century (for review, see 1) lend tinctorial properties to

Address for correspondence: Dr Clive Taylor, Keck School of Medicine of the University of Southern California, Department of Pathology, HMR 204, 2011, Zonal Avenue, Los Angeles, CA 90033 (CRT), USA. e-mail: taylor@pathfinder.usc.edu

the tissue section. The interpretation of even the simple routine haematoxylin and eosin (H&E) stain does include elements of a quantitative assessment, albeit mostly at a subconscious level. Are the nuclei more or less blue (hyperchromatic)? Is the cytoplasm of the cardiac myocytes pinker than normal (hypereosinophilic), as in the early phases of myocardial infarction? What amount of atypia is present? These evaluations are made subjectively, with experience as the reference point, and formal quantitative methods are not usually employed, except for particular defined purposes.<sup>2</sup> Assessment of the degree of malignancy, formalized in some instances into grading criteria, again includes quantitative elements, such as the number of mitotic figures (sometimes going so far as to offer a count per high-power field), or the number of large cells versus small cells in a population, as in the grading of diagnosed follicular centre cell lymphomas of B-cell origin. Underlying these 'semiquantitative' approaches there is the subliminal concept of a covert reference standard, against which judgements, rather than 'measurements', can be made. Often this standard is crude, as in the use of a 'normal histiocyte' nucleus to separate large from small in the grading of follicular centre cell lymphomas, and the level of diagnostic agreement amongst different observers, including experts, is disturbingly poor (about 60% in this instance<sup>3</sup>).

Faced with the limited application of quantitative methods in day-to-day surgical pathology, a comparison with the quantitative methods in use in clinical pathology is of real value in determining how to improve the situation. Biological stains (including those based on aniline dyes) that are the basis of the usual histopathological stains are somewhat difficult to control in terms of intensity of colour (stain), from cell to cell and more so from section to section (different tissues on different days), although this may change with the advent of new generations of automated stainers. An immunohistochemical (IHC) reagent, by contrast, has the potential to provide quantitative data, for although we are not accustomed to thinking of it as such, it is in potential, if not in fact, an 'immunoassay' performed *in situ* on the tissue section. An IHC 'stain' is strictly analogous to an enzyme-linked immunosorbent assay (ELISA) test performed in the clinical lab and ELISA tests are widely recognized as being truly quantitative (if properly performed). Exactly the same reagents that are employed in an ELISA test on serum, e.g. an assay for insulin, may be employed to perform an IHC stain for insulin in a paraffin section. It is a curious oversight of scientists in general, and pathologists in particular, that the principles and reagents used in one environment are accepted as providing a strictly quantitative result

(ELISA-serum), but when applied to a tissue section (IHC), are addressed only as a 'stain'.

**Factors to be addressed in establishing quantitative IHC methods; towards an IHC assay as opposed to an IHC stain**

There have been several schools of thought as to the reason why IHC 'stains' are difficult to run in a manner that lends itself to quantitative analysis. If there is a consensus, it is that several reasons conspire together; these may conveniently be grouped into three general areas (Table 1).

Possibly the overriding factor in effecting significant change would be to transform the mindset of pathologists, at least of the next generation, so that the end result of an IHC protocol would come to be regarded not as just a stain, but rather as a precise immunoassay that is strictly quantifiable, if properly performed and controlled, similar to any other immunologically based assay of like principle (such as ELISA).

It would seem evident that in order to achieve a quantifiable result with an IHC stain, thereby converting it to a quantifiable immunoassay, the total assay (staining process) must itself first be standardized.<sup>6-10</sup> Those areas in assay performance that lead to signifi-

**Table 1.** The Total Test, adapted from the earlier proposal of the US Biologic Stain Commission<sup>4</sup> and modified from 5

|                                                         |
|---------------------------------------------------------|
| The Total Test                                          |
| Pre-analytical                                          |
| Specimen handling, from operating room to histology lab |
| Fixation: total fixation time and type of fixative      |
| Paraffin embedding, storage and sectioning              |
| Deparaffinization                                       |
| Analytical                                              |
| Antigen retrieval (exact method)                        |
| Assay (staining) method and protocol                    |
| Reagent validation                                      |
| Controls (Reference Standards)                          |
| Technologist and laboratory certification               |
| Proficiency testing and quality assurance               |
| Post-analytical                                         |
| Reading of result(s)/scoring/quantification             |
| Report                                                  |
| Turn-around time                                        |
| Outcomes analysis/economics/reimbursement               |

cant variability or errors, and are therefore targets for improvement, are reviewed below.

#### PRE-ANALYTIC ISSUES: TRANSPORTATION, FIXATION, SECTIONING

Pre-analytical issues fall under the broad rubric of 'sample preparation' (Table 1). This area is the least well controlled of all phases of the IHC staining process<sup>6,11</sup> and the least controllable, because of the ways in which tissues are obtained from diverse hospital and clinic settings. At long last the importance of good sample preparation in cancer diagnosis, or misdiagnosis, particularly with regard to measurement of prognostic and predictive markers, has reached the national consciousness in the USA, with issuance of requests for proposals from the National Cancer Institute (RFA-CA-07-003: *Innovations in Cancer Sample Preparation*, U.S. National Cancer Institute, 2006, [www.nih.gov](http://www.nih.gov)).

In the 'routine' environment of diagnostic surgical pathology, specimens that ultimately may be subject to IHC analysis may be obtained by fine-needle aspiration, or from the operating room by biopsy, or by more extensive therapeutic surgical procedures. Following removal of the tissue from the body, autolysis generally is arrested by immersion in a fixative. By far the most commonly employed fixative is formalin (in common practice a 4% solution of formaldehyde).<sup>6,11,12</sup> Other fixatives have been employed, and others are being explored, in order more effectively to meet some of the current needs for performing molecular analyses of tissues or cells.<sup>13</sup> Formalin has many advantages, not least a long tradition of use and the fact that it yields good morphological detail; or rather, it yields the morphological detail we are accustomed to, which is deemed the equivalent of good. Formalin is also inexpensive, easily stored (with some reservations as to quality) and universally available. Formalin therefore is what we have, and what we must learn to work with for the immediate future.

Recognizing that the autolytic process begins immediately, the routine practice is to place the excised tissue directly in formalin, prior to leisurely transportation to the laboratory, with emphasis on 'leisurely'. Once in the surgical pathology suite ('grossing' room), the specimen is cut in, meaning that if not already sufficiently small it is cut into small blocks to facilitate rapid penetration by the fixative (formalin penetrates relatively slowly) and placed in fresh fixative for further processing. Whereas the ideal time for fixation of a 5 mm thick tissue block is perhaps 12–24 h (no uniform agreement here<sup>11,12</sup>), in practice, the total

time in fixative is very variable, due to differing transportation times to the laboratory and accumulation of specimens for batch processing. Fixation time in reality is almost entirely uncontrolled, varying anywhere from 6 to 24 h, or more. Add to this, questions as to whether the formalin is freshly prepared and adequately buffered, plus variability in the rate of penetration of formalin in different types of tissues and into differently sized blocks, and the result is a major impediment to standardization of an IHC stain and an obstacle to quantification.

As an aside, *in situ* hybridization (ISH) methods have a probe–target pairing that is not immunologically based, and thus strictly do not fall under the title of IHC. Nonetheless, the principles are closely analogous, particularly with reference to interpretation and scoring. For RNA analysis by ISH methods, there is a further complication, namely the rapid degradation of RNA by intrinsic enzymes, probably beginning as soon as the blood supply to the tissue is interrupted as part of excision. For useful results, and certainly for quantification, it is essential therefore to process such materials immediately, and control over transportation time becomes critical so as to minimize the time elapsed prior to complete fixation.

Following fixation, the process of embedding in paraffin, and subsequent de-paraffinization of the cut sections, also involves exposing the tissues (and therefore the analytes) to a series of chemicals and to heat. The end result is a formalin-fixed paraffin-embedded (FFPE) tissue section. While anecdotes exist, there are no good data as to the adverse effects of processing upon the various analytes that might be detected by IHC staining. This aspect therefore is usually ignored, but in the absence of data it appears sensible that these steps of the overall preparation of the tissue section are performed as consistently as possible.

The importance of consistent sectioning may also be overlooked. For routine staining a precision microtome is used to achieve a section thickness of about 5 µm. A nucleus that is 5 µm in diameter may thus be entirely within the plane of the section, or only partially included, with effects upon the apparent intensity of a nuclear IHC stain, all other things being equal. Thicker sections may manifest the same problem even for quite large nuclei, whereas generally thinner sections will minimize it. Uniform preparation of FFPE sections that are < 5 µm in thickness is not possible; plastic embedding media, or other special media, allow consistency in sectioning down to 1 µm, but do not lend themselves well to routine use, or to larger blocks. Even slight variations in thickness, over a 5-µm section, due to 'chatter' or unevenness of cut, may also produce

changes in intensity of the staining reaction that are inapparent to the naked eye, but are readily appreciable using quantitative imaging techniques.<sup>6</sup>

#### ANALYTICAL ISSUES: ANTIGEN RETRIEVAL, PROTOCOLS, REAGENTS, CONTROLS

Antigen retrieval (AR), considered here as part of the analytic process, has shown spectacular benefits in terms of the ability of all and sundry to achieve a positively stained FFPE section, but there have been some unexpected and unwanted consequences.<sup>5,14–16</sup> The fact that many antigens, which hitherto could be stained only with difficulty, are now readily demonstrable following AR has led to renewed laxity with regard to fixation and to diminished efforts in developing alternative and superior fixatives. The AR method itself is also open to great variation in practical performance, and this may affect the intensity of stain achieved, or even the number of cells that are perceived as demonstrating a positive staining reaction. The degree to which any particular antigen is 'retrieved' is also entirely unknown with reference to the absolute amount present postfixation (in the FFPE section) and the amount present postfixation is itself not known with reference to the amount present (per cell) when the tissue was first removed from the body (fresh, prior to sample preparation—'prefixation'). Some standardization may be achieved through the practice of testing the different variables in the retrieval process (method of heating, temperature, time, pH, etc.) to achieve the optimal AR protocol for each specific antigen using a defined set of reagents and staining methods.<sup>5,15</sup> This approach would seem to risk the possibility of uncovering significantly different AR protocols for many antigens, but in practice yields only three major variations of the basic AR method, one of which will generate excellent results for the great majority of clinically relevant antigens.<sup>5,15</sup>

Reagents and staining protocols, once seen as the primary impediment to qualitatively reproducible staining, are now regarded as perhaps the least of the difficulties, provided that certain procedures are followed, a tribute to the fine efforts of the Biological Stain Commission/Food and Drug Administration (FDA) working groups more than a decade ago.<sup>4,17</sup> A common error is to neglect to read the package insert for each new reagent carefully; at a minimum, perusal will provide performance characteristics (does it work on FFPE sections?) and expected patterns of staining. It should also provide a detailed staining protocol, with a judicious reminder that should a laboratory choose to depart from the protocol, then validation becomes the

entire responsibility of the performing laboratory. In any event, every new reagent introduced into the laboratory, whether a primary antibody, or a different labelled antibody system, must undergo an initial validation by the laboratory to establish the performance characteristics. So called positive-control tissues serve this purpose and properly should have been fixed and processed in a manner identical to the test specimens (same fixative, fixation time, etc.).<sup>4,5,8,18,19</sup> Tissue microarrays are useful in evaluating a new primary antibody, allowing a quick and efficient study of the pattern of staining on potentially hundreds of tumour or tissue types, in duplicate or triplicate, deposited on a single slide. These basic control sections serve to validate qualitatively the reagents and protocol, but as usually constituted cannot serve as absolute reference materials for calibration and quantification. This limitation is because the control materials themselves, while demonstrably positive in a qualitative sense, have been fixed and processed in ways that preclude knowing, in absolute terms, how much of the test analyte is present postfixation; it is merely that there is enough to detect a positive staining reaction with the reagents and protocol employed. Indeed, the amount of analyte present prefixation (when fresh) is also totally unknown.

From this brief review it is argued that the 'total test' must be standardized in order for any conceivable quantitative scoring method to achieve a useful degree of reliability. Even so, for all the reasons described, the best that can be achieved today is a 'semiquantitative' type of assay, given the lack of a defined reference standard.

It follows that a primary requirement should be to develop reference materials that can be used to establish the integrity of the sample, as well as to standardize the assay and to calibrate the results. The criteria for such a standard can be derived once more by extrapolation from clinical pathology (Table 2).

For IHC these requirements are exacting and have yet to be fully met in a practical sense. As discussed above, the usual positive-control tissue employed in laboratories meets only some of these requirements, as does the FDA-approved Her2-kit produced by Dako (HercepTest; Dako, Glostrup, Denmark, or [www.dakousa.com](http://www.dakousa.com)). In all instances the most important deficiency is the lack of data relating to the absolute amount of the analyte present in the control material prior to the first step of the total test (i.e. prior to specimen preparation/fixation). Efforts to meet the requirements set out in Table 2 have been few, but do show some promise in the use either of peptide deposits,<sup>20,21</sup> cell lines (including cell-line blocks)<sup>22,23</sup> or faux tissues (histoids) (M. Ingram and A. Imam,

**Table 2.** Summary of desirable characteristics of any reference standard that would provide a basis for accurate quantification of immunohistochemistry (or *in situ* hybridization)<sup>19</sup>

| Immunohistochemical Reference Standard: Requirements                                                   |
|--------------------------------------------------------------------------------------------------------|
| It must be subjected to the same rigours of sample preparation (fixation) as the 'test' tissue         |
| It must be integrated into all phases of the test (assay) protocol, including evaluation of the result |
| It should contain a known amount of the analyte(s) subject to assay                                    |
| It should be universally available                                                                     |
| It should be inexhaustible and inexpensive                                                             |

unpublished collaboration, 2005; see 5, p. 35, Figures 1–27). Another proposed approach is to use quantifiable internal reference standards, the characteristics of which have been measured by experimental observation under differing conditions of formalin fixation, paraffin embedment and antigen retrieval.<sup>19</sup> Such internal standards, once established in terms of absolute quantity of analyte per specific cell type, have the potential to serve as calibration points for test analytes demonstrated in adjacent cells by double-IHC stain methods, using multiplex-capable imaging techniques, which are described later.

Lacking quantifiable internal reference standards for calibration, all IHC stains at best can only be semiquantitative, comparing the intensity of stain, or the number of positive cells, or both, with the control, or with other cases, with results that are relative, not absolute.

#### POST ANALYTIC: RESULTS AND INTERPRETATION (SCORING)

One school of thought held that the lack of reliability of IHC methods for measurement of oestrogen receptor (ER) or progesterone receptor (PR) expression was attributable to the nature of the 'semiquantitative' scoring process and the intrinsic deficiencies of an observer-based, subjective manual method. The underlying belief was that, however clearly the criteria are set forth, the application of such criteria and the reporting of the outcome will vary from pathologist to pathologist, or even for the same pathologist from day to day. Computer-assisted image analysis was a touted solution to the scoring of IHC stains, where a quantifiable result was the desired outcome. Comparative studies<sup>7,9,24,25</sup> indeed do show that under controlled

circumstances image analysis is superior to manual methods as performed by most observers.

The problem of interpretation of an IHC stain should not be minimized. With basic lineage-related markers, the problem of consistent evaluation is real, even with reference to relatively simple questions: is the cell or tissue positive for  $\kappa$  chain or ER or CD30, or is it not? Is specific staining present or not, with reference to the controls? Where is the staining localized? How much staining is there (begging the questions as to whether the amount of staining correlates with the absolute amount of antigen)? What scoring system should be used and how reproducible is it? The general consensus is that IHC methods, applied as qualitative 'special' stains, if properly applied and interpreted, increase the accuracy of diagnosis in surgical pathology, as is well established by studies of lymphoma.<sup>3</sup> However, it is known, although not often publicly acknowledged, that the eyes and brains of different observers do not see and interpret the same H&E section the same way.<sup>18,26,27</sup> For IHC stains the variability of interpretation may be even greater, as is revealed in some of the proficiency-testing exercises carried out by the College of American Pathologists (Chicago, IL, USA) and UK National External Quality Assessment Immunocytochemistry. It turns out that the answers are dependent not only upon the experience and acuity of the eye of the beholder, but also upon the integrity of the staining process, as already emphasized.<sup>6,7,9,10,28–30</sup>

With respect to prognostic markers, the problem of interobserver consistency is much greater, requiring not just a decision as to whether or not there is specific positive staining, but some sort of scored or semiquantitative result. The inherent difficulties are well recognized for such commonly tested analytes as ER and PR,<sup>28</sup> where commercially available reference standards are not usually available and where both methodology and scoring vagaries contribute to error. The problem is arguably even greater for Her2.<sup>29,30</sup> The FDA-approved Dako kit contains a cell-line standard and includes instruction about how to read the result, and most published reports utilize some form of reference control. Even with these important provisions, scoring of the same cases for Her2, ER and PR by residents and pathologists shows clinically important variations and is short of the desired uniformity.<sup>28–30</sup>

Some investigators believe that the solution to the problem of interpretation, especially the quantitative or scoring aspects of interpretation, may be found in improved methods of image analysis.<sup>7,9,24,25</sup> Methods and instruments currently exist that yield improved results; many of these instruments are available commercially. At present the larger reference laborat-

ories are more likely to use such aids than smaller laboratories, or even academic centres. In part, this is a matter of economics; the instruments are expensive and hard to justify where volumes are insufficient, or where special expertise cannot be developed and committed to their operation. In part, it is a reflection of the fact that image analysis still requires interactive input by the pathologist, and that often leads to increased time requirements for reading the assay without conclusive evidence that the result is of more value clinically. Nonetheless, a visit to the exhibitor display at any of the major pathology meetings leaves little doubt as to which way the wind is blowing, as reviewed in the following paragraphs.

The last decade has seen enormous advances in the capabilities of image analysis systems applied to tissue sections, both in software and hardware, especially in digital cameras and in data management of the resulting large files. However, realization of the potential for increased accuracy in the post-analytic phase of the assay has served to focus renewed attention on the basic deficiencies of the IHC staining process as a whole, and its intrinsic lack of reproducibility, as discussed in the first part of this review. Even the most sophisticated image analysis hardware/software system cannot produce accurate results if the underlying stain (read immunoassay) itself suffers from non-reproducibility or significant non-linear behaviour. In this context, accuracy (and reproducibility) can be determined only if rigorous quantifiable reference standards<sup>19</sup> are available and are used to calibrate the system. The notion of accuracy should embrace not only the measurement of an analyte in a particular section, validated against a reference standard, but also the ability to repeat the result on the same case, day to day, in the same and in different laboratories, and the ability to measure the same (and ultimately different) analyte(s) in different specimens and cases, again reproducibly. Thus, standardization of the overall assay must proceed hand-in-hand with accurate and reliable reading (scoring) of the assay; both are essential for achievement of an IHC stain, which in practice could be, and should be, more than just a stain but rather a system of controlled and interlocked processes, analogous to immunoassays in the clinical laboratory.

Finally, expression-array-based research has emphasized that pathology and, in particular, cancer biology, reflects the simultaneous workings of multiple molecular pathways. For maximum relevance, these should be assessed on a per cell, rather than a per tissue slice basis, since ultimately cells are the units of behaviour and their individual phenotypes are the relevant metric. In a practical sense this implies multiplexed

molecular (IHC or ISH) assays in which more than one analyte is assessed on a tissue section at one time, in identifiable individual cells. As can be imagined, in addition to the imaging challenges this may pose, it also amplifies all the demands on controls and standards elaborated above.

## Image analysis; approaches and systems

While image analysis of molecular labels can include a number of applications, the following section will be limited to the discussion of the problem of estimating abundance of stains in histological tissue, with an emphasis on IHC as opposed to immunofluorescence. The previous section has addressed issues of sample preparation and provision of appropriate controls that can ensure that the IHC procedures have generated a valid signal for the imaging system to capture. The assumption is made that the signal on the slide is representative and in some way quantitatively related to the abundance of the antigens in the tissue section, which in turn is related, albeit in ways unknown, to the absolute amount of the analyte in the original tissue. The example used here will be estimation of nuclear antigens rather than membrane staining, since the latter may require additional considerations beyond simple intensity measurements, such as spatial patterns of expression that have their own subtleties. In addition, this review will not dwell on the well-documented subjectivity and intra- and interobserver variability of manual, visual-based semiquantitative estimation of intensity or even of percent positivity<sup>31,32</sup> and will simply postulate that properly designed automated imaging methods, because they are immune to the consequences of fatigue and subjectivity, can outperform human observers, certainly in terms of precision and quantitative reproducibility.

Factors that affect performance of the imaging system include the choice of camera and illumination source, the optical performance of the stains themselves, as well as the presence and degree of multiplexing. After image acquisition, it is then necessary to deploy appropriate mathematical techniques to extract quantitative intensity and area measurements from the imaging data.

### IMAGING HARDWARE: RGB VERSUS MULTISPECTRAL APPROACHES

There is a long history of the application of image processing to pathology samples.<sup>33,34</sup> While some early automated imaging systems employed greyscale cameras and filter wheels to collect images, most current

brightfield (transmitted light) pathology imaging systems rely on standard colour cameras similar in many respects to consumer digital cameras. These typically employ a Bayer-pattern colour mask over a charge coupled device or complementary metal oxide semiconductor detector, and use various algorithms to process the raw image data to generate colour images that can be presented to the pathologist, and that are also used in downstream automated analysis. Single-chip, Bayer-pattern red-green-blue (RGB) cameras that are often employed, especially in many 'home-grown' systems, can generate imaging artefacts, especially with respect to fine structures or edges, and have poorer spatial fidelity than more expensive three-chip systems in which separate pixel-registered cameras are used to acquire simultaneously red, green and blue images. While the simple acquisition of good-looking colour images is appealing, RGB detectors can introduce significant problems when one is trying to achieve quantification and inter-instrument precision. There are a number of ways that variation arises. For example, colour values can vary significantly with the colour temperature of the illumination source, different colour-correction routines in camera firmware can play a role in the exact colour values that are reported out, and different camera chips have differing spectral responsiveness. Some cameras employ automatic gain control or related circuitry designed to 'optimize' image quality, with unpredictable effects on resulting images.

Even if an RGB imaging system is working perfectly, there are intrinsic limitations to its ability to distinguish between similar chromogens and, even more challengingly, to be able to 'unmix' such signals if they overlap spatially. 'Unmix' in this sense means to isolate the optical signal from each chromogen so that each can be measured quantitatively, and separately. Signal processing theory suggests that at least  $n$  if not  $n + 1$  measurements are needed to unmix  $n$  signals. In theory, therefore, it is impossible to unmix more than three chromogens with an RGB sensor. In practice, while it is possible to do a good job unmixing diaminobenzidine (DAB) (brown) from haematoxylin (blue), it has proven extremely difficult to unmix brown from red from blue (a typical combination of stains for a double-labelled sample) using only RGB measurements, due to the colour overlap of the spectral profiles. To accomplish such tasks properly, true multispectral imaging approaches may be necessary.

#### SPECTRAL IMAGING

Spectral imaging microscopy represents a technological advance over visual or RGB camera-based ana-

lyses. By acquiring a stack of images at multiple wavelengths, spectral imaging allows the determination of precise optical spectra at every pixel location. With this spatially resolved spectral information in hand, it is possible to enhance the utility of IHC and ISH stains, and even the standard biological stains used in surgical pathology. There are a number of ways to perform spectral imaging (reviewed<sup>24,35</sup>). The focus in this review is on the commercially available liquid crystal tunable filter-based system (Nuance<sup>TM</sup>; CRi, Woburn, MA, USA), from which all examples here will be drawn; this is not to imply that the Nuance system is the best or only approach, merely that it is the model with which the authors have had most experience. This system is suitable for both brightfield and fluorescence imaging. Under automatic control, a series of images (from three to as many as  $\geq 20$ ) are taken from blue to the red (e.g. 420–700 nm) and the resulting image 'stack' or 'cube' is assembled in memory in such a way that a spectrum is associated with every pixel. The ability to sample the spectrum with many discrete wavelength regions spanning the visible wavelength range allows for accurate unmixing of multiple spatially colocalized chromogens, even if they are similar in colour and have largely overlapping absorption spectra. Thus, it becomes straightforward to separate dark reds from light browns, or even varieties of blue stains (haematoxylin versus nitroblue tetrazolium-5-bro-4-chloro-3-indolylphosphate).<sup>36,37</sup>

#### IMAGE PROCESSING AND UNMIXING

The key process, either with RGB images or multispectral datasets, is to partition the overall signal in a given pixel correctly into its component species. Linear unmixing algorithms (as described<sup>38–40</sup>) rely on the signals adding together linearly. This is true with fluorescent dyes (which emit light), but this is not the case with chromogens imaged in brightfield, since they absorb light. Fortunately, the Lambert–Beer (or simply Beer's) law relating concentrations to absorbance indicates that when the transmission data are converted to optical density (absorbance) units, linearity is restored and quantification and unmixing<sup>39</sup> can be successfully achieved. There are many benefits attendant on the conversion to optical density (OD), which is typically performed by taking the negative (base 10) log of the transmitted image divided by the illumination (usually a clear area on the microscope slide). First, absorbance values are an intrinsic property of the sample and do not depend on vagaries of illumination or camera responsiveness. This means that absorbance measurements of a

given specimen performed on any appropriate system should, in theory, be comparable. Second, in the process of creating an absorbance image, flat-fielding is automatically performed, which removes the effects of uneven illumination and minor flaws in the optical train. Conversion to OD can be performed on monochrome, RGB or multispectral images.

OD (absorbance) units are dimensionless and logarithmic: zero absorbance means all photons transmitted; an OD of 1.0 absorbs 90% of all photons, and an OD of 2.0 absorbs 99% of all potentially detected photons. IHC stains can individually generate signals of 1 OD. Accordingly, having two or more dense and overlapping stains can result in virtually black deposits from which little or no useful spectral or quantitative data can be recovered. This, plus the lesser dynamic range achievable with IHC versus fluorescence-based approaches may mean that immunofluorescence may be preferable or necessary for some applications.<sup>32</sup> Nevertheless, IHC has some practical advantages over immunofluorescence, including the fact that pathologists prefer it largely because it allows integration of 'phenotypic' features in the IHC stain with the traditional morphological features, long the 'gold standard' for diagnosis.

An important caveat is that the optical properties of the chromogens will affect the linearity and dynamic range of the assay. The Lambert–Beer law that underlies the unmixing approach applies only to pure absorbers. Some chromogens, most notably the popular brown DAB stain, exhibit scattering behaviour similar to that of melanosomes. In fact, it can be impossible to separate DAB from melanin pigmentation spectrally, since their spectra arise from the same optical properties. However, in practice, this does not seem to pose insuperable problems, since linearity and reasonable dynamic range can be achieved using DAB approaches.<sup>41</sup> Other chromogens, such as Vector Red, have been shown to have excellent linearity and dynamic range.<sup>42</sup>

In addition to the specific molecular labelling procedure, a counterstain is almost always applied. Thus the challenge for quantification begins with the unmixing of the chromogen (typically DAB) from the counterstain (typically haematoxylin). The latter pair can be successfully unmixed using simple RGB imagery if conversion to OD is performed,<sup>39</sup> but other pairs may not be so amenable. One of the challenges (see below) is the accurate determination of the spectra of the chromogens as input values into the unmixing procedure. Small variations in the spectra chosen can have quite dramatic effects on the calculated abundance values. While in many cases it

suffices to measure the spectrum of the isolated chromogens (single stain, no counterstain), we have found that it may be necessary to measure the spectrum of the chromogens in the actual sample, after all the staining procedures have been performed, since the spectra can be affected by the presence of other dyes and reagents.

#### MULTIPLEXING

Typically, only a single IHC–chromogen–antigen combination is used per slide; if more than one antigen is to be analysed, serial sections are made and a different antibody is applied to each. This procedure benefits from simplified protocols and quality control regimens compared with multicolour techniques, but generates more slides and possibly more preparation steps than if the reagents are 'multiplexed' on a single slide. Moreover, multiple molecular events cannot be evaluated on a per-cell basis when parallel sections are employed, and this capability is very important in establishing the phenotype of individual tumour cells (e.g. lymphoma cells) distributed in a mixed cell population. Multicolour immunohistochemistry is thus an important goal, but is challenging to achieve. The prerequisite to quantitative accuracy in a multiple labelled section is lack of interference between the labels. Not only can one label physically block the successful labelling of the next antigen due to steric hindrance, but the various labelling procedures can be chemically incompatible. Suffice it to say that the performance of multiple labellings on a single specimen increases the demands for appropriate controls.<sup>43</sup> Assuming that the labelling procedures have been performed satisfactorily, unmixing of three or more chromogens is entirely feasible<sup>38,44</sup> (Levenson, submitted). In addition, multiple chromogenic ISH signals can be combined with IHC.<sup>45,46</sup>

#### Examples of spectral unmixing and multiplexing

Figures 1 and 2 illustrate the application of spectral imaging to a determination of Ki67 levels in lymph node cells. The Ki67 antigen was visualized using DAB and the sample counterstained with haematoxylin (H). Figure 1A shows the visual appearance of the sample, which, like all the subsequent examples, was spectrally imaged using a Nuance multispectral imaging system. The unmixed DAB and haematoxylin channels are shown in Figure 1B,C. Note that the haematoxylin staining accurately recapitulates the dense staining of

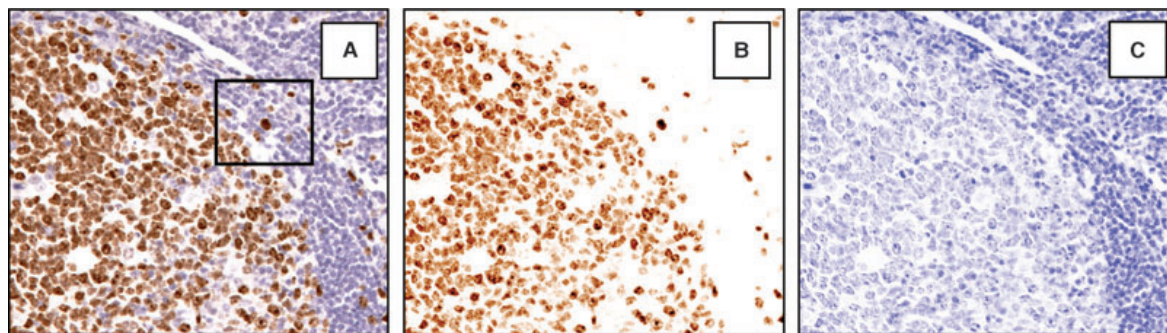


**Figure 1.** Unmixing of diaminobenzidine (DAB) from haematoxylin: Ki67 in a lymph node germinal centre imaged with a Nuance multispectral imaging system. A, Visual (RGB) appearance of the sample. B, Unmixed DAB signal. C, Unmixed haematoxylin signal, which accurately recapitulates the dense staining of the mantle cells and the paler staining of the germinal centre. The small box indicates the region highlighted in Figure 2.

the mantle cells and the paler staining of the germinal centre. The small box indicates the detail region highlighted in Figure 2, which addresses the importance of accurately estimating the 'pure' spectrum of the DAB for use in the unmixing procedure. Three different spectra for the DAB component were used as inputs into the unmixing procedure. If one simply captures the spectrum of a DAB-labelled nucleus (top row), unmixes and examines the haematoxylin channel, it can be seen that all of the absorbance (due to DAB plus haematoxylin) ends up in the DAB channel, and a white 'hole' is seen in the DAB+ regions in the H channel. The integrated intensity of the DAB-labelled nucleus is indicated. If one attempts to calculate the 'pure' spectrum of the DAB by removing the H component, a variety of curves can be generated, depending on the nature of the algorithm used. The second row shows what happens if overcompensation occurs—in this case, some of the DAB signal remains in the H channel, leading to an overly intense H signal and an underestimation of DAB intensity. Finally, if the DAB spectrum is correctly estimated, unmixing generates a clean partition of DAB and H signals, in which the H intensity of the labelled nucleus is essentially indistinguishable from that of its neighbours. The integrated intensities of the DAB label in the circled nucleus varied by more than twofold, depending on the spectra chosen, illustrating the quantitative importance of correct unmixing. Of course, the importance of using appropriate spectra for the unmixing process only increases with the number of chromogens being considered simultaneously.

Figure 3 is intended to demonstrate that three-colour unmixing is feasible, using three strips of coloured plastic arranged so that all possible combinations of single, double and triple mixtures are captured.

The spectra of the individual strips are shown, as are the unmixed images for each strip separately (pseudo-coloured according to the colour of the spectral library curves in Figure 3B), along with intensity profiles along each strip. As can be seen, calculated absorbance values of each strip are unaffected by the presence of the other absorbers.

Finally, Figure 4 illustrates the application of unmixing to a histological section of formalin-fixed, paraffin-embedded breast tissue containing both non-malignant and invasive breast epithelial cells, stained for ER and PR and counterstained with haematoxylin. This example has considerable current relevance because the detection and evaluation of nuclear positivity of breast cancer steroid hormone receptors can affect the choice of treatment and is useful in predicting patient outcomes.<sup>7,47</sup> Receptor levels are currently evaluated manually, typically using a 0–3+ grading system and/or a simple visual estimate of the number of positive nuclei in a relevant cellular population. In this example, ER and PR antigens were visualized with DAB and Vulcan Red chromogens and counterstained with haematoxylin (H). The six panels illustrate the original visual appearance and after unmixing the H channel (which can be used to identify the nuclear compartment for quantitative purposes) and separate channels for ER and PR (green and red, respectively). The dotted oval identifies a region of presumptively normal epithelium and the red oval a region of invasive ductal carcinoma. The bottom panels show an overlay of the green and red channels, and finally, a depiction of the original image with ER–PR double-positive cells is indicated using a yellow mask. It is striking that the normal and the malignant regions exhibit different colocalization patterns (normal, ~ 5%; malignant, ~ 55%, on a pixel-wise basis).

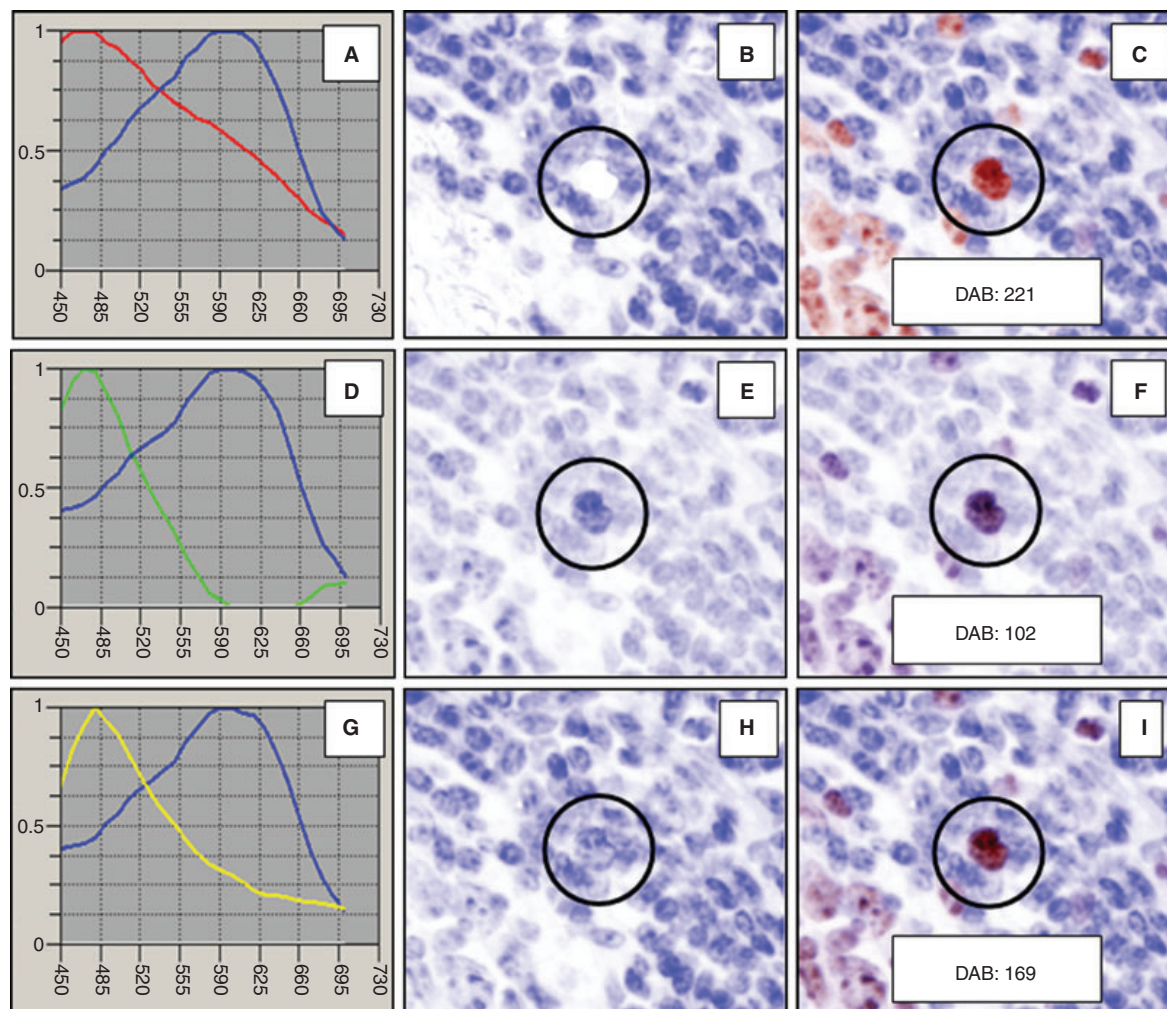


**Figure 2.** Unmixing of diaminobenzidine (DAB) from haematoxylin: choice of DAB spectrum affects quantitative results. Differing spectra for the DAB (along with a constant haematoxylin spectrum) are shown in A,D,G, and the respective unmixing results are shown in the corresponding rows. The unmixed haematoxylin channels are shown in the second column and the combined DAB plus haematoxylin result is shown in the third column. The numeric values shown represent the integrated optical density of the DAB signal from the circled nucleus. The third row represents the best DAB spectral estimate, with haematoxylin values for Ki67(+) and (–) nuclei displaying similar intensities. See text for additional discussion.

The biological significance of this and other patterns of markers revealed quantitatively on a per cell basis is currently unknown. What is important is that now there are tools to explore molecular interrelationships in individual cells using multicolour IHC-based techniques, with the potential for quantifiable results, prerequisites for the beginnings of 'Molecular Morphology'.<sup>48</sup>

In conclusion, quantitative immunohistochemistry is not a distant mirage, but is within our grasp. It will

require careful attention to the preimaging components, including provision of quantitative standards,<sup>19</sup> to be included in the entire sample processing pathway, and attention to all parameters of sample acquisition, fixation and staining, with good quality control procedures in place for each probe singly and in combination. For multiplexing, the interaction of one antibody–label combination with all the others must be understood and controlled, and choice of chromogen and counterstains will affect both the visual and quantitative results.

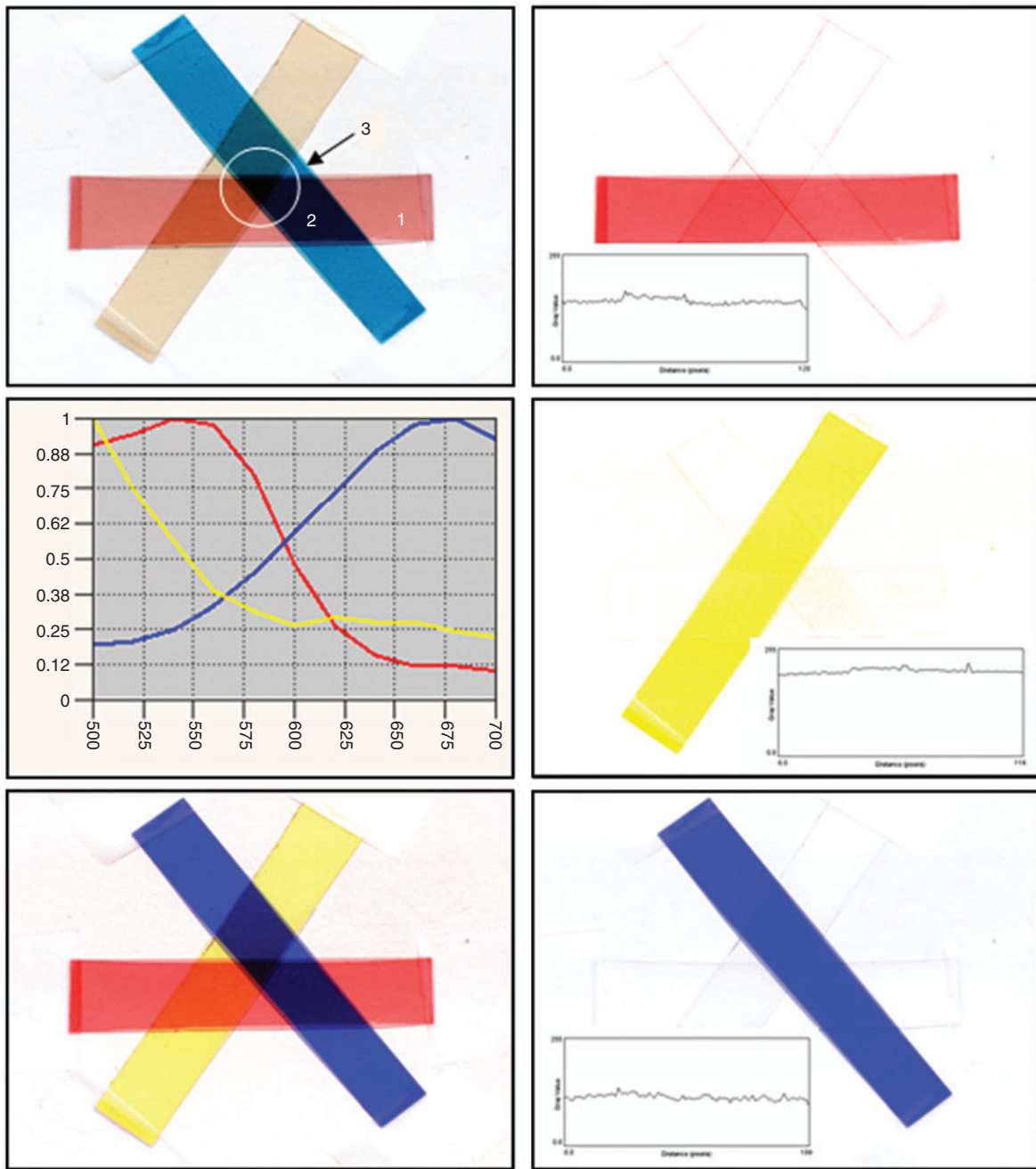


**Figure 3.** Three-colour unmixing of plastic films with spectra similar to brown and red immunohistochemical chromogens and haematoxylin. The strips were arranged so that single, double and triple overlapping regions were present (representative regions are indicated by numbers in A). A spectral dataset was acquired; spectra corresponding to the individual plastic strip species are shown in B. Using these spectra, the image cube was unmixed to create individual images of each coloured strip by itself (coloured in the pseudocolours of the spectral library used for unmixing). Intensity profiles are shown for each strip, indicating that quantitative unmixing could be achieved even when two or three absorbing species spatially overlapped.

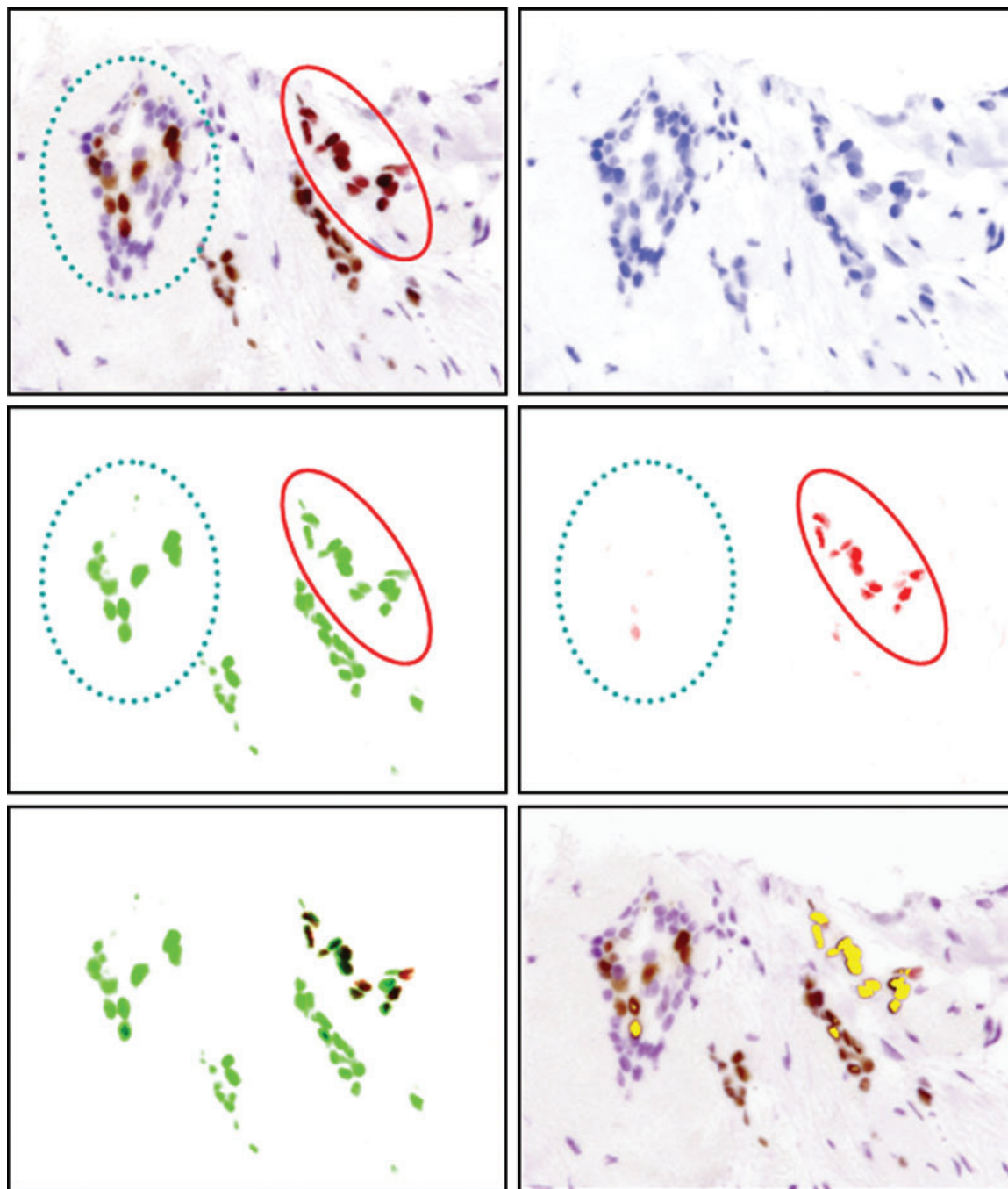


**Figure 4.** Detection and unmixing of oestrogen receptor (ER)– (diaminobenzidine) and progesterone receptor (PR)– (Vulcan Red) signals from a breast tissue specimen counterstained with haematoxylin. The six panels illustrate the original visual appearance (A), and after unmixing, the H channel (B, which can be used to identify the nuclear compartment for quantitative purposes) and separate channels for ER (C) and PR (D). The dotted oval identifies a region of presumptively normal epithelium and the red oval a region of invasive ductal carcinoma. The bottom panels show an overlay of the green and red channels (E) and finally, a depiction of the original image with ER–PR double-positive cells indicated using a yellow mask (F).

Finally, the imaging component has to be carefully performed, with appropriate sensors, exemplified by multispectral, reliable and validated unmixing algorithms. In addition, and not discussed above, it will be essential to incorporate appropriate downstream image analysis and quantification approaches that accurately report molecular events on a per pixel, per cell, or per 'relevant tissue component' basis, as appropriate. Ultimately, especially for clinical applications, this task becomes a systems problem, in which the entire process, from sample acquisition to reporting and interpretation, needs to be integrated, standardized<sup>11,19,49,50</sup> and, to the greatest extent possible, automated.

## Acknowledgements

We thank Dr C. Kerfoot (Mosaic Laboratories) for providing the breast cancer specimen and the spectral dataset. This work was supported by NIH grants nos R33 CA103455-01 and R44 CA88684.

## References

- Horobin RW, Kiernan JA eds. *Conn's biological stains*, 10th edn. Oxford: Biological Stain Commission, Bios Publishers 2002.
- Becker RL. Standardization and quality control of quantitative microscopy in pathology. *J. Cell. Biochem.* 1993; **17G**: 199–204.
- The Non-Hodgkin's-Lymphoma Classification Project. Clinical evaluation of the International Lymphoma Group Classification of Non-Hodgkin's Lymphoma. *Blood* 1997; **89**: 3909–3918.
- Taylor CR. Quality assurance and standardization in immunohistochemistry. A proposal for the annual meeting of the Biological Stain Commission, June, 1991. *Biotechnic. Histochem.* 1992; **67**: 110–117.
- Taylor CR, Cote RJ. *Immunomicroscopy; a diagnostic tool for the surgical pathologist*, 3rd edn. Edinburgh: Elsevier 2005.
- Grube D. Constants and variables in immunohistochemistry. *Arch. Histol. Cytol.* 2004; **67**: 115–134.
- Diaz LK, Sneige N. Estrogen receptor analysis for breast cancer: current issues and keys to increasing testing accuracy. *Adv. Anat. Pathol.* 2005; **12**: 10–19.
- Seidal T, Balaton AJ, Battifora H. Interpretation and quantification of immunostains. *Am. J. Surg. Pathol.* 2001; **25**: 1204–1207.
- Umemura S, Itoh H, Serizawa A *et al.* Immunohistochemical evaluation of hormone receptors in breast cancer: which scoring system is suitable for highly sensitive procedures. *Appl. Immunohistochem. Mol. Morphol.* 2004; **12**: 8–13.
- Umemura S, Osamura RY. Utility of immunohistochemistry in breast cancer practice. *Breast Cancer* 2004; **11**: 334–338.
- Leong AS-Y. Quantitation in immunohistology. Fact or fiction? A discussion of variables that influence results. *Appl. Immunohistochem. Mol. Morphol.* 2004; **12**: 1–7.
- Dapson RW, Feldman AT, Wolfe D. *Lessons In immunohistochemistry*, Battle Creek, MI: Anatech 2005.
- Nadji M, Nassiri M, Vincek V *et al.* Immunohistochemistry of tissue prepared by a molecular friendly fixation and processing system. *Appl. Immunohistochem. Mol. Morphol.* 2006; **115**: 814–822.
- Shi SR, Cote RJ, Young LL, Taylor CR. Antigen retrieval immunohistochemistry: practice and development. *J. Histochem. Technol.* 1997; **20**: 145–154.
- Shi SR, Cote RJ, Chaiwun B *et al.* Standardization of immunohistochemistry based on antigen retrieval technique for routine formalin-fixed tissue sections. *Appl. Immunohistochem. Mol. Morphol.* 1998; **6**: 89–96.
- Shi SR, Gu J, Taylor CR. Antigen retrieval techniques. *Immunohistochemistry and Molecular Morphology*. Eaton Publishing, Natick, MA 2000.
- Taylor CR. Report from the Biological Stain Commission. FDA issues final rule for classification/reclassification of immunohistochemistry (IHC) reagents and kits. *Biotechnic. Histochem.* 1998; **73**: 175–177.
- Taylor CR. An exaltation of experts; concerted efforts in the standardization of immunohistochemistry. *Appl. Immunohistochem.* 1993; **1**: 232–243.
- Taylor CR. Standardization in immunohistochemistry: the role of antigen retrieval in molecular morphology. *Biotechnic. Histochem.* 2006; **81**: 3–12.
- Sompuram SR, Kodola V, Zhang J *et al.* A novel quality control slide for quantitative immunohistochemistry testing. *J. Histochem. Cytochem.* 2002; **50**: 1425–1433.
- Shi SR, Liu C, Perez J *et al.* Protein-embedding technique: a potential approach to standardization of immunohistochemistry. *J. Histochem. Cytochem.* 2005; **53**: 1167–1170.
- Rhodes A, Jasani B, Couturier J *et al.* A formalin-fixed, paraffin-processed cell line standard for quality control of immunohistochemical assay of HER-2/neu expression in breast cancer. *Am. J. Clin. Pathol.* 2002; **117**: 81–89.
- Shi SR, Cote RJ, Liu C *et al.* A modified reduced temperature antigen retrieval protocol effective for use with polyclonal antibody to Cox2 (PG27). *Appl. Immunohistochem. Mol. Morphol.* 2002; **10**: 368–373.
- Cross SS. Observer accuracy in estimating proportions in images: implications for the semiquantitative assessment of staining reactions and a proposal for a new system. *J. Clin. Pathol.* 2001; **54**: 385–390.
- Lehr HA, Jacobs TW, Yaziji H *et al.* Quantitative evaluation of HER-2/neu status in breast cancer by fluorescence *in situ* hybridization and by immunohistochemistry with image analysis. *Am. J. Clin. Pathol.* 2001; **115**: 814–822.
- Foucar E, Foucar MK. Error in anatomic pathology. In Foucar MK ed. *Bone marrow pathology*, 2nd edn. Chicago: ASCP Press 2000.
- Sirota RL. Error and error reduction in surgical pathology. *Arch. Pathol. Lab. Med.* 2005; **129**: 1228–1233.
- Rhodes A, Jasani B, Balaton AJ, Miller KD. Immunohistochemical demonstration of estrogen and progesterone receptors: correlation of standards achieved on in house tumours with that achieved on external quality assessment material in over 150 laboratories from 26 countries. *J. Clin. Pathol.* 2000; **53**: 292–301.
- Hatanaka Y, Hashizume K, Kamihara Y *et al.* Quantitative immunohistochemical evaluation of HER2/neu expression with HercepTest™ in breast carcinoma by image analysis. *Pathol. Int.* 2001; **51**: 33–36.
- Rhodes A, Jasani B, Anderson E *et al.* Evaluation of Her2/neu immunohistochemical assays sensitivity and scoring on formalin-fixed and paraffin-processed cell lines and breast tumours: a comparative study involving results from laboratories in 21 countries. *Am. J. Clin. Pathol.* 2002; **118**: 408–417.

31. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nat. Med.* 2002; **8**: 1323–1327.
32. Rubin MA, Zerkowski MP, Camp RL *et al.* Quantitative determination of expression of the prostate cancer protein alpha-methylacyl-CoA racemase using automated quantitative analysis (AQUA): a novel paradigm for automated and continuous biomarker measurements. *Am. J. Pathol.* 2004; **164**: 831–840.
33. Bengtsson E. *Computerized cell image analysis: past, present, and future. Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29–July 2 2003. Proceedings Vol. 2749/2003, Lecture Notes in Computer Science.* Berlin/Heidelberg: Springer 2003; 395–407.
34. Bearman G, Levenson R. Biological imaging spectroscopy. In Vo-Dinh T ed. *Biomedical photonics handbook*. Boca Raton: CRC Press 2003; 818–826.
35. Barber PR, Vojnovic B, Atkin G *et al.* Applications of cost-effective spectral imaging microscopy in cancer research. *J. Phys. D. Appl. Phys.* 2003; **36**: 1729–1738.
36. Levenson RM. Spectral imaging perspective on cytomics. *Cytometry* 2006; Part A. **69A**: 592–600.
37. Mansfield JR, Levenson RM. Spectral imaging in biology and medicine: slices of life. *Cytometry* 2006; Part A. (in press).
38. Zhou R, Hammond EH, Parker DL. A multiple wavelength algorithm in color image analysis and its applications in stain decomposition in microscopy images. *Med. Phys.* 1997; **23**: 1977–1986.
39. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* 2001; **23**: 291–299.
40. Farkas DL, Du C, Fisher GW *et al.* Non-invasive image acquisition and advanced processing in optical bioimaging. *Comput Med. Imaging Graph.* 1998; **22**: 89–102.
41. Matkowskyj KA, Cox R, Jensen RT *et al.* Quantitative immunohistochemistry by measuring cumulative signal strength accurately measures receptor number. *J. Histochem. Cytochem.* 2003; **51**: 205–214.
42. Ermert L, Hocke AC, Duncker HR *et al.* Comparison of different detection methods in quantitative microdensitometry. *Am. J. Pathol.* 2001; **158**: 407–417.
43. Szeszel MK, Crisman CL, Crow L *et al.* Quantifying estrogen and progesterone receptor expression in breast cancer by digital imaging. *J. Histochem. Cytochem.* 2005; **53**: 753–762.
44. Levenson R, Cronin PJ, Pankratov KP. Spectral imaging for brightfield microscopy. *Proc. SPIE* 2003; **4959**: 27–33.
45. Speel EJ, Jansen MP, Ramaekers FC *et al.* A novel triple-color detection procedure for brightfield microscopy, combining *in situ* hybridization with immunocytochemistry. *J. Histochem. Cytochem.* 1994; **42**: 1299–1307.
46. Macville MV, Van Der Laak JA, Speel EJ *et al.* Spectral imaging of multicolor chromogenic dyes in pathological specimens. *Anal. Cell Pathol.* 2001; **22**: 133–142.
47. Press MF, Sauter G, Bernstein L *et al.* Diagnostic evaluation of HER-2 as a molecular target: an assessment of accuracy and reproducibility of laboratory testing in large, prospective, randomized clinical trials. *Clin. Cancer. Res.* 2005; **11**: 6598–6607.
48. Taylor CR. Immunohistochemistry for the age of molecular morphology. *Appl. Immunohistochem. Mol. Morphol.* 2001; **9**: 1–2.
49. O'Leary TJ. Standardization in immunohistochemistry. *Appl. Immunohistochem. Mol. Morphol.* 2001; **9**: 3–8.
50. Taylor CR. Quantifiable internal reference standards for immunohistochemistry. The measurement of quantity by weight. *Appl. Immunohistochem. Mol. Morphol.* 2006; **14**: 253–259.