Leaders

A scoring system for immunohistochemical staining: consensus report of the task force for basic research of the EORTC-GCCG

Paul J van Diest, Peter van Dam, Sonja C Henzen-Logmans, Els Berns, Maria E L van der Burg, John Green, Ignace Vergote

Department of
Pathology of the Free
University Hospital,
Amsterdam, The
Netherlands
P J van Diest

Department of
Pathology, Daniel den
Hoed Cancer Institute,
Rotterdam, The
Netherlands,
S C Henzen-Logmans
E Berns

Department of
Gynaecologic
Oncology
M E L van der Burg

Department of
Gynaecologic
Oncology, University
of Antwerp, Belgium
P van Dam

Department of
Gynaecologic
Oncology, University
of Leuven, Belgium
I Vergote

Department of
Gynaecologic
Oncology,
Clatterbridge Centre
for Oncology,
Bebington, UK
J Green

EORTC-GCGG,
Gynaeological Cancer
Cooperative Group of the
European Organization for
Research and Treatment of
Cancer.

Correspondence to:
Dr van Diest, Department of
Pathology, Free University
Hospital, PO Box 7057,
1007 MB Amsterdam, The
Netherlands.

Accepted for publication
28 July 1997

Immunohistochemistry has become an important tool in surgical pathology, especially in the diagnosis of tumours. The demonstration of antigens more or less specific for different types of tumours has dramatically improved diagnostic accuracy. As well as being diagnostically important, the expression of several antigens as demonstrated by immunohistochemistry correlates with the prognosis of some cancers. Examples are steroid receptors in breast and gynaecological cancers,[1] oncogene products such as HER-2/neu in breast cancer,[2] apoptosis related gene products such as bcl-2 in breast cancer,[3] and cell cycle regulation proteins such as p53 (unpublished data), pRb,[5] and cyclin D1 in breast, and head and neck cancer.[6,7]

However, the initial promising results of studies on these markers could not always be confirmed. This may be because of different methodological problems. First, there may have been selection of patients and differences in treatment. Second, there are often differences in tissue processing from study to study, especially with regard to the type of antibody and the application of antigen retrieval. Third, and possibly most important, interpretation of staining and presentation of the results are not standardised resulting in low intraobserver and interobserver reproducibility. These problems are a very important obstruction for possible clinical applications. There is therefore a clear need for consensus on a protocol for scoring of immunohistochemical staining.

The aim of this article is to discuss the most important items that should be part of a protocol for the scoring of immunohistochemical staining (staining patterns, tumour area to be assessed, sampling of fields of vision, scoring method, decision thresholds) and to suggest some solutions for common problems to reach a more standardised interpretation and presentation of immunohistochemical staining results. Scoring of immunocytochemical staining will not specifically be addressed, although many of the issues discussed will be applicable to cytological material.

Tissue processing
There are several steps in tissue processing that may influence staining patterns and intensity. These include type and duration of fixation, section thickness, antigen retrieval procedures, type and concentrations of primary, second (and third) step antibodies, and way of staining development—for example, it has recently been reported that tumour marker immunostaining intensity may be lost on stored paraffin wax slides of breast cancer.[8,9] It is beyond the scope of this article to give an in depth overview of these factors, but these variables should be standardised as much as possible. Technically inadequate slides (uneven staining, uneven section thickness, heavy background) should be discarded.

Staining patterns
To arrive at reproducible interpretation, there must be agreement on the significance of nuclear, diffuse cytoplasmic or membrane staining for each specific antibody. Most antigens either localise (functionally) in the nucleus (such as pRb, p53, cyclin D1, steroid receptors), in the cytoplasm (Bax) or in the cellular membrane (HER-2/neu). Staining patterns other than the typical pattern for a given antibody (for example, cytoplasmic rather than nuclear staining for steroid receptors, or diffuse cytoplasmic instead of membrane staining for HER-2/neu) may have to be regarded as unfunctional and ignored, although for some antigens it will be useful to record these seemingly aberrant expression patterns until their biological role and clinical significance are better understood.[10]

Area to be assessed
Although a purely random approach may sometimes be optimal, it will for most applications be necessary to define in which part of the lesion the staining will be assessed. For instance, invasive areas may be chosen over in situ parts, for proliferation markers the periphery of tumours is preferred. For prognostic applications, often the most poorly differentiated area of the tumour will be selected as this yields best clinical correlations, but for

diagnostic applications the best differentiated areas are usually most useful because typical antigen patterns are best preserved. The tumour area to be selected should be clearly defined for each marker, and guidelines for the (minimal) size of this area (to be marked on the slide for future reference) must be given.

Sampling of fields of vision

Within the marked tumour area, it will usually neither be possible (considering the workload involved) nor necessary to assess all available fields of vision. Therefore, some form of sampling will have to be applied. It is a matter of experience[11][12] that this should not be left to the convenience of the observer as this will lead to interobserver variation and lack of reproducibility. Since some degree of clustering is usually present, the best way of sampling fields of vision is usually by systematically spreading them over the whole chosen tumour area (fig 1). This makes sure that cells over the whole chosen area are assessed and will lead to the most representative sample and highest reproducibility.[11][12]

Assessment of positivity of immunohistochemical staining

QUALITATIVE ASSESSMENT

Qualitative assessment of immunohistochemical staining is widely used to judge visually whether any positivity of staining (compared with the background) is present in the cells under study. This way of assessing positivity is often the usual practice for diagnostic reasons and is quite satisfactory for stains where the mere presence of antigens is informative irrespective of the degree of positivity or the fraction of positive cells, such as for keratins and CD31.

QUANTITATIVE ASSESSMENT

For certain antigens, especially those predictive of response to treatment or related to prognosis such as p53, cyclin D1, steroid receptors,

cathepsin D, and multidrug resistance related proteins, one needs to know the degree of positivity and the fraction of positive cells. A more sophisticated scoring procedure than qualitative assessment is needed. Such staining has semiquantitatively been scored as: −(negative), + (positive), ++ (strongly positive), or +++ (very strongly positive). Although this approach may be very intuitive, problems arise when there is heterogeneity in staining—for example, it is usually not clear what to do with heavy staining in just a small number of cells. Some have advocated to report the staining intensity and the percentage of positive cells. An even more refined method is the Histo-score,[13][14] where the fractions of negative (score 0), weakly positive (score 1), positive (score 2), strongly positive (score 3), and very strongly positive (score 4) cells are estimated, the fractions are multiplied with the scores and summed, the total being the H-score. This method is widely applied to score steroid receptor positivity in breast cancer. However, in practice, the fractions are only roughly estimated without counting, and assigning a score to an individual nucleus is highly subjective. Studies have shown that the scoring of no staining (score 0) and very strong staining (score 4) is quite reproducible, but that there are many interobserver discrepancies in scores 1–3.[14] Therefore, this method is not optimal.

Alternatively, image analysis has been applied. Theoretically, image analysis allows the assessment of the amount of staining per cell by measuring absorption. However, in practice, immunohistochemistry using diaminobenzidine (DAB) as chromogen appeared to be only stoichiometric (linear relation between the amount of antigen and staining intensity as measured by degree of absorption) at low levels of intensity and not at the usual intensities used in daily immunohistochemistry.[15] This approach has therefore lost some of its initial popularity. Using the alkaline phosphatase antialkaline phosphatase (APAAP) approach or antibodies directly labelled with a fluorescent dye seems more promising, but more experience is needed. Nevertheless, image analysis may well serve to assess objectively the number of positive cells,[16][17] although it remains a rather labourious and relatively expensive procedure.

Recently, an interactive approach for immunoscoring has been developed based on stereology,[4][18–20] which is very quick, simple, and reproducible. In a defined area of the lesion, fields of vision systematically spread over the whole area of interest are chosen, as illustrated in fig 1. As the first field is chosen at random, all fields are random, so this way of sampling is called systematic at random. This can be done manually, but is most efficiently done by means of a software guided automatic scanning stage. In these fields of vision, cells are selected using a point grid, and the positivity of these cells is scored. The principle here is to sample relatively few cells from relatively many fields of vision, or the "do more less well" principle.[21][22] The number of fields of vision and cells to be assessed will vary from application to application and is
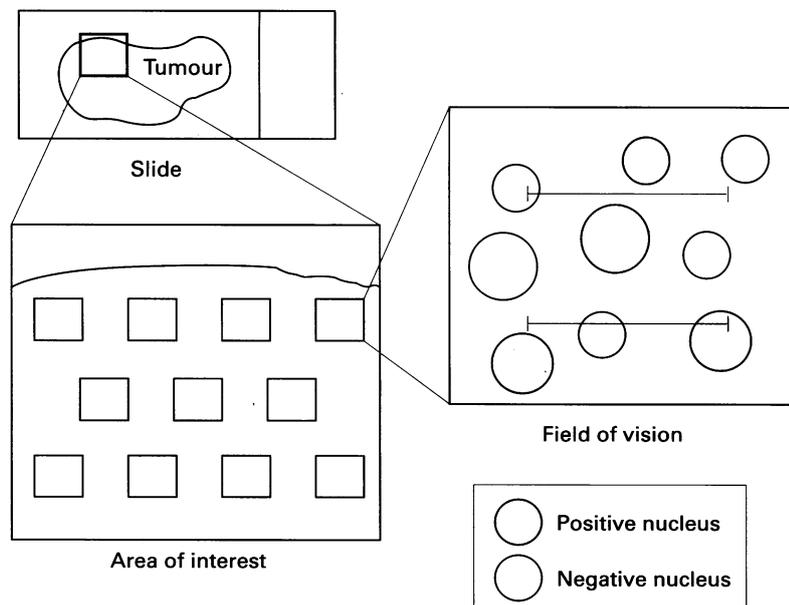

Figure 1 Schematic explanation of the suggested protocol for interactive stereological immunoscoring based on systematic random sampling.

dependent on the desired accuracy. In practice, 200 cells sampled from 50–100 fields of vision will allow a coefficient of variation < 5%.[19 21] However, on occasion (for example, in case of values very close to the decision threshold) one may have to count more cells to reach an even higher accuracy. In general, it is not necessary to count 1000 cells as has been suggested.[23] As fields spread over the whole area of interest are assessed (which copes best with clustering that is often observed) and selection of cells is un-biased, this approach leads to the highest reproducibility.[4 19] Good clinical correlations with this method have been found in Barrett's oesophagus,[20] breast cancer,[4] and lymphomas.[18]

This approach is especially useful to assess the percentage of positive cells irrespective of the staining intensity, but obviously does not overcome the problems involved in the subjective assessment of degree of positivity as previously mentioned.

Choice of cut off points

After having assessed the percentage of positive cells, the question is obviously how to choose the cut off value for decision making on "positivity". To this end, different approaches may be followed. First, staining percentages may be compared with biochemical findings. Second, a cut off point may be tuned to molecular biological results—for example, presence of DNA amplification or the amount of mRNA. Third, one may derive a cut off point from clinical studies taking clinical end points such as recurrence or survival, or response to a certain treatment. Preferably, such clinical cut off points should be established in a learning set of patients and confirmed in an independent patient group. Also, cut off points should be standardised from one study to another. This is essential for semiquantitative scoring.

Reporting

Results of immunoquantitation should preferably not just be reported as positive or negative, but the quantification method should be provided as well as exact data from the quantitation and the decision threshold. This can be done by merely providing the data, but graphically presenting the data may be more illuminating and easy to understand.

Summary of suggested protocol

In summary, when you are interested not just in mere presence of antigen but the fraction of positive cells is of interest irrespective of the staining intensity, we suggest an interactive stereological approach in a carefully defined part of the lesion. This consists of a systematic random sampling of fields of vision, preferentially guided by an automatic scanning stage, and assessment of positivity of cells in these fields selected with a point grid. Technically adequate slides are a prerequisite. An example of a protocol for such an approach is presented below:

● define the relevant staining pattern (nuclear, cytoplasmic or membrane) and consider only this pattern

● define the area of the lesion to be assessed (periphery *v* central parts), relevant tissue parts (in situ *v* invasive), and cells of interest
● in this area, apply a systematic random sampling approach for selection of a defined number of fields of vision
● in the fields of vision, assess positivity of cells selected with a point grid
● define the number of cells to be assessed, using the principle that few cells should be sampled from many fields of vision (for example, 2–4 cells from 50–100 fields of vision yielding a total of 200 counts)
● define cut off value
● define how the results should be reported.

Such methods are fast, simple and highly reproducible. Application of such protocols will yield more objective data on the distribution of antigens in different lesions, which will be easier to compare from study to study, and will speed the process towards clinically applicable tests.

1 Gaskell DJ, Hawkins RA, Sangsterl K, Chetty U, Forrest APM. Relation between immunocytochemical estimation of oestrogen receptor in elderly patients with primary breast cancer and response to tamoxifen. *Lancet* 1989;i:1044–6.
2 Baak JPA, Chin D, Diest PJ van, Matze-Cok P, Bacus SS. Comparative long term prognostic value of quantitative Her2/neu protein expression, DNA ploidy, morphometric and clinical features in paraffin-embedded invasive breast cancer. *Lab Invest* 1991;64:215–22.
3 Visscher DW, Sarkar F, Tabaczka P, Crissman J. Clinico-pathologic analysis of bcl-2 immunostaining in breast carcinoma. *Mod Pathol* 1996;9:642–6.
4 Withdrawn.
5 Anderson JJ, Tiniakos DG, McIntosh GG, Autzen P, Henry JA, Thomas MD, *et al*. Retinoblastoma protein in human breast carcinoma: immunohistochemical study using a new monoclonal antibody effective on routinely processed tissues. *J Pathol* 1996;180:65–70.
6 Diest PJ van, Michalides RJAM, Jannink I, Valk P van der, Peterse JL, Jong JS de, *et al*. Cyclin D1 expression in invasive breast cancer: correlations and prognostic value. *Am J Pathol* 1997;150:705–11.
7 Michalides R, Veelen N van, Hart A, Loftus B, Wientjens E, Balm A. Overexpression of cyclin D1 correlates with recurrence in a group of forty-seven operable squamous cell carcinomas of the head and neck. *Cancer Res* 1995;55:975–8.
8 Henson DE. Loss of p53-immunostaining intensity in breast cancer [editorial]. *J Natl Cancer Inst* 1996;88:1015–16.
9 Jacobs TW, Prioleau JE, Stillman IE, Schnitt SJ. Loss of tumor marker-immunostaining intensity on stored paraffin slides of breast cancer. *J Natl Cancer Inst* 1996;88:1054–9.
10 Kay EW, Walsh CJ, Cassidy M, Curran B, Leader M. C-erbB-2 immunostaining: problems with interpretation. *J Clin Pathol* 1994;47:816–22.
11 Fleege JC, Diest PJ van, Baak JPA. Computer assisted efficiency-testing of different sampling methods for selective nuclear graphic tablet morphometry. *Lab Invest* 1990;63:270–5.
12 Fleege JC, Diest PJ van, Baak JPA. Systematic random sampling for selective interactive nuclear morphometry in breast cancer sections. Refinement and multiobserver evaluation. *Anal Quant Cytol Histol* 1993;15:281–9.
13 Bosman FT, Goeij AFPM de, Rousch M. Quality control in immunocytochemistry: experiences with the oestrogen receptor assay. *J Clin Pathol* 1992;45:120–4.
14 Diest PJ van, Weger AR, Lindholm J. Reproducibility of subjective immunoscoring of steroid receptors in breast cancer. *Analyt Quant Cytol Histol* 1996;18:351–4.
15 Fritz P, Wu X, Tuczek H, Multhaupt H, Schwarzmann P. Quantitation in immunohistochemistry. A research method or a diagnostic tool in surgical pathology? *Pathologica* 1995; 87:300–9.
16 Makkink-Nombrado SV, Baak JPA, Diest PJ van, Broeck-aert M, Meijer GA, Matze-Cok E, *et al*. Prognosis in premenopausal breast cancer without systemic adjuvant treatment: prospective evaluation of classical, morphometric, flow and image cytometric, and immunohistochemical cell- and tissue molecular biological markers. In: *Quantitative pathological, cell- and molecular biological markers in breast*

cancer. SV Makkink-Nombrado, Thesis. Free University Amsterdam, 1995.

17 Makkink-Nombrado SV, Baak JPA, Diest PJ van, Broeck-aert M, Schuurmans L, Brugghe J, et al. Comparison of classic prognostic factors with quantitative pathologic, cell- and molecular biologic features in premenopausal breast cancer. In: Quantitative pathological, cell- and molecular biological markers in breast cancer. SV Makkink-Nombrado, Thesis. Free University Amsterdam, 1995.

18 Oudejans JJ, Jiwa NM, Kummer JA, Ossenkoppele GJ, Heerde P van, Baars JW, et al. Activated cytotoxic T-cells as prognostic marker in Hodgkin's disease. Blood 1997;83: 1376–82.

19 Polkowski W, Meijer GA, Baak JPA, Kate FJW ten, Obertop H, Offerhaus GJA, et al. Reproducibility of p53 and Ki67 immuno-quantitation in Barrett's oesophagus. Analyt Quant Cytol Histol 1997;19:246–54.

20 Polkowski W, Baak JPA, Lanschot JJB van, Meijer GA, Schuurmans LT, Kate FJW ten, et al. Clinical decision making in Barrett's oesophagus can be supported by com-puterized immunoquantitation and morphometry of fea-tures associated with proliferation and differentiation. J Pathol. [In press.]

21 Brugghe J, Baak JPA, Meijer GA, Diest PJ van, Broeckaert MAM, Schuurmans LT. Volume percentage of epithelium in ovarian tumours: a methodological study [abstract]. J Pathol 1995;176:49A.

22 Gundersen HJG, Osterby R. Optimizing sampling efficiency of stereological studies in biology: or "do more less well!". J Microsc 1981;121:65–73.

23 Hall PA, Richards MA, Gregory WM, d'Ardenne AJ, Lister TA, Stansfeld AG. The prognostic value of Ki67 immunos-taining in non-Hodgkin's lymphoma. J Pathol 1988;154: 223–35.