



Principles and approaches for reproducible scoring of tissue stains in research

David K. Meyerholz¹ · Amanda P. Beck²

Received: 25 January 2018 / Revised: 16 March 2018 / Accepted: 31 March 2018 / Published online: 30 May 2018
© United States & Canadian Academy of Pathology 2018

Abstract

Evaluation of tissues is a common and important aspect of translational research studies. Labeling techniques such as immunohistochemistry can stain cells/tissues to enhance identification of specific cell types, cellular activation states, and protein expression. While qualitative evaluation of labeled tissues has merit, use of semiquantitative and quantitative scoring approaches can greatly enhance the rigor of the tissue data. Adhering to key principles for reproducible scoring can enhance the quality and reproducibility of the tissue data so as to maximize its biological relevance and scientific impact.

Introduction

Tissue examination (histopathology) retains an important role in the diagnosis of clinical disease and evaluation of tissues in the research setting [1–5]. Examination and interpretation of tissue changes can be enhanced through labeling of cellular and tissue markers to identify features not observable by routine stains, such as specific cell types, activation states, and protein expression, to name a few. Common examples of cellular/tissue-labeling techniques include immunohistochemistry (IHC), immunofluorescence, in situ hybridization, and lectin histochemistry. These techniques can also provide a link between morphology and the in situ assessment of protein/marker expression to further corroborate other protein-based assays (such as enzyme-linked immunosorbent assay) [2, 3, 5–11]. While qualitative evaluation of labeled tissues has some

value, the use of semiquantitative and quantitative scoring can further clarify or validate morphologic interpretations, especially in the research setting.

Experimental pathologists are often experienced and well-versed in reproducible approaches for tissue scoring. These professional experts are invaluable resources and collaborators for experimental studies of tissues and can be useful in tissue/stain evaluations, peer-review, scoring, quality control, etc. In contrast, it is not uncommon for junior/trainee pathologists as well as biomedical personnel to inquire about resources to guide them in scoring tissue stains to produce effective and reproducible data. In this review, we present common principles and approaches to score labeled cells/tissue that can collectively enhance the rigor and reproducibility of the resulting tissue data. For simplicity in this review, we will primarily focus on examples using IHC “stains”, though these scoring principles and approaches can be readily applied to other labeling techniques as well.

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41374-018-0057-0>) contains supplementary material, which is available to authorized users.

✉ David K. Meyerholz
david-meyerholz@uiowa.edu

✉ Amanda P. Beck
amanda.beck@einstein.yu.edu

¹ Department of Pathology, University of Iowa Carver College of Medicine, Iowa City, IA, USA

² Department of Pathology, Albert Einstein College of Medicine, Bronx, NY, USA

Tissue factors

In order to have reproducible tissue staining assays, there is a necessity to develop clearly defined and standardized protocols to ensure consistent and valid results [7, 10, 12–14]. Prior to scoring tissues sections, it is important to consider pre-analytic, analytic, and post-analytic variables that may influence the reproducibility, quality, and extent of the staining procedure [7, 9, 10, 13–26]. One study reported that nearly a third of IHC slides evaluated by an external

Table 1 Tissue factors can influence immunostaining assessment

Tissue handling
Antemortem conditions (e.g., pyrexia)
Collection/transport (e.g., time postmortem, temperature)
Fixation (type, time, fixative:tissue ratio)
Trimming
Processing
Embedding
Sectioning
Drying (time, temperature)
Storage
Tissue staining
Deparaffinization
Staining method
Batch effects
Controls (positive and negative)
Validation/optimization
Antibody (clonality, source)
Antibody optimization (dilution, protocol, antigen retrieval)
Detection method (chromogen type, fluorophore)
Counterstaining

quality assessment program did not give a staining result that was satisfactory for analysis [5]. Reproducible immunostaining requires optimized and standardized tissue protocols to ensure consistent and reproducible results [13, 14, 27, 28]. A study by Engel and Moore [18] identified more than 60 variables in the pre-analytical stage alone, beginning with proper sample collection and handling and including multiple aspects of fixation, processing, embedding, slide drying, and storage. For simplicity, these tissue factors have been grouped into tissue handling and tissue staining variables (Table 1).

Tissue handling

Consistent sample preparation using validated protocols is critical to maintain both morphology and, in the case of IHC, antigenicity of target epitopes [7, 16, 17]. Tissue handling (Table 1) encompasses the steps from tissue collection at autopsy or biopsy until a sectioned tissue is ready to be stained. Fixation must be carefully considered, including consistency in time, volume, and type of fixative [3, 9, 13, 20, 21]. It is important to ensure there is no delay in sample fixation (known as “ischemic time”), as sub-optimal fixation will negatively affect later interpretation [9, 10, 12, 13, 16, 25]. Proper tissue preservation can prevent autolysis (“self-digestion”), a postmortem change that morphologically resembles necrosis and is characterized by degradation of cellular constituents (e.g., DNA, RNA, and protein). Autolytic processes increase degradation of

epitopes, incidence of nonspecific staining, and sloughing of epithelial cells that can readily confound scoring assessments [29, 30]. Higher temperature and prolonged postmortem interval (i.e., time from collection to fixation) are proportional to the extent of autolysis. Thus, keeping tissues cool and placing them into fixative in a timely fashion is important. However, even if the temperature and time are well controlled, tissues may still experience significant autolysis as a result of inadequate fixation protocols. The most common fixative used is 10% neutral buffered formalin, popular due to the fact it yields good morphology, is inexpensive, easily stored and readily available [14, 16]. However, other options are available, and a fixative that is most compatible with downstream techniques and end points commonly used in each laboratory are often selected [13, 14, 16, 21, 24]. In general, ~20 times volumes of fixative to tissue is recommended, and the total time in fixative may vary depending on the type of tissue being studied. The rate of formalin penetration into tissues is approximately 0.5 mm/h. We recommend gross-sectioning tissues thin enough (less than approximately 0.5 cm in at least one plane so that fixative can adequately penetrate. Placement of fixative and experimental tissues in conical tubes should be avoided as these containers inherently reduce the exposure of tissue surfaces to fixative. Tissues in fixative containers can be placed on a rotary table for gentle agitation and the fixative can be replenished after the first few hours of tissue exposure to further enhance tissue fixation [21]. Fixation time varies according to the size of tissue and chosen markers to be evaluated. In general, a minimum of 48–96 h of fixation and a maximum of 2 weeks is recommended as an initial starting point for most IHC validation and optimization procedures on tissues, but this may be modified depending on the target tissue and epitope. Once fixed, tissues can be trimmed and placed into cassettes, taking into consideration thickness, plane of sampling, and relevant orientation [12, 13, 16, 21].

The remaining steps of tissue handling include those from processing and paraffin embedding, through sectioning and storage (Table 1). Orientation of samples should remain consistent, and the relevant anatomy of the tissue for end point examination should be respected [13, 16, 21]. The final step in slide preparation for staining is sectioning. Tissue sections should be made of consistent thickness, such as 3–5 μm . Thicker sections tend to result in darker stains (as there is simply more antigen-containing tissue present), while thinner sections have lighter stains, but give greater visual resolution. At sectioning, artifacts, such as ripped, folded, or wrinkled tissue sections or uneven cuts (“chatter”), can constrain effective scoring evaluation even in well-stained tissues [14, 22].

If tissue section slides will not be stained immediately, the date of tissue sectioning can be recorded, and slides

stored in a consistent, climate-controlled environment until needed. Antigen decay of unstained paraffin slides is a well-documented phenomenon during time spent in storage, though the exact mechanism of antigen degradation is unknown [31, 32]. Multiple factors have been suggested to affect immunoreactivity and antigen of unstained tissue sections, including length of storage time, temperature, humidity, light exposure (specifically UVA rays), oxidation, and the specific antigen being examined [31–35]. Loss of antigenicity in unstained tissue sections is proportional to duration of storage and this loss of antigenicity reduces the intensity and extent of staining, possibly resulting in false-negative results. Storage of unstained slides at room temperature is not recommended due to significant loss of antigenicity [31–33]. Some have advocated that unstained paraffin slides can be stored at 4 °C and others suggest –20 °C or even –80 °C may be preferable; however, regardless of the temperature there can still be some reduced immunoreactivity over time [34]. Exposure of tissues to exogenous (environmental humidity) or endogenous (inadequate fixation or processing of samples leading to water retention in tissue blocks) water can negatively affect immunoreactivity, with antigen loss observed after only several days with certain biomarkers. Decreased immunoreactivity has been reported in sections exposed to fluorescent light or sunlight, with antigenicity loss proportional to the length of exposure. Exposure of the unstained section to air leads to oxidation, which has been suggested to negatively impact immunoreactivity. To combat this, some laboratories have established procedures to coat slides with paraffin or parafilm to “seal” the tissues, though others have reported that this coating does not significantly protect against antigen loss [32, 33, 35]. Additional published methods to reduce antigen decay in stored tissue sections include vacuum packing with desiccant proteins, storage in a nitrogen chamber, and optimizing fixation and antigen retrieval methods. While unstained tissue sections are susceptible to antigen degradation, paraffin block on the other hand is resistant to antigen degradation and can be stored for years. Therefore, to avoid some of these issues associated with antigen degradation of unstained tissues on slides, many investigators and labs often wait to section tissues from paraffin blocks until immediately prior to immunostaining.

Tissue staining

Tissue staining is a common source of variability for scoring and encompasses the steps involved in taking an unstained tissue section to an immunostained slide ready for scoring (Table 1) [5, 13, 21, 26]. With IHC, molecular markers of interest in cells/tissues are specifically labeled using antibodies and chromogens (i.e., “stains”) [2, 3, 5–7, 9–11].

Development of an IHC protocol often requires optimization, a process of testing and adjusting various parameters to achieve the desired sensitivity and specificity. The parameters include, but are not limited to: antibody (Ab) type (e.g., monoclonal vs. polyclonal); method (direct, indirect, sandwich, polymer, etc.); reagent (peroxidase/antiperoxidase, avidin/biotin, etc.); origin (species, vendor, etc.); target (native protein vs. synthetic epitope); protocol (antigen unmasking, Ab dilution, incubation times, temperature, diluents, etc.); chromogen (3,3'-diaminobenzidine (DAB), etc.); and type of counterstain [5, 7, 10, 13, 14, 16, 20, 24, 25, 27, 28, 36]. An additional challenge is non-specific cross-reactivity, the binding of the antibody to an epitope different from the directed target (off-target structure) [37, 38]. This can be seen with both monoclonal and polyclonal antibodies. Also nonspecific staining (e.g., Van der Waals forces, etc.) can be a problem too. All antibodies cross-react to some extent, but the amount seen in a given assay depends on a variety of factors, including the specific antibody and tissue being tested, antibody concentrations, and testing parameters as mentioned above. While it is beyond the scope of this paper to go through each parameter, off-target staining issues are relevant considerations when developing, optimizing, or troubleshooting IHC staining.

For study validation, one can demonstrate that the antibody used is sensitive, specific, and reproducible in a given assay [38–40]. An important factor in staining validation is the usage of appropriate controls to help establish the integrity of the sample, success of the staining protocol, calibration of results, and standardization of the assay [5, 7, 14, 16, 36]. This is especially important when quantitatively assessing staining intensity [16]. Positive controls are tissues or samples containing the target molecule in a known anatomic that can be visualized by a stain. In quantitative analysis, the positive control tissue should ideally have areas with different levels of staining to compare to test samples, including structures with low-intensity staining to reduce the risk of false-negative results as well as regions of known high protein expression to ensure that staining strength does not interfere with diagnosis [10, 25]. The goal of a negative control is to check for nonspecific staining (false positive results). Appropriate negative controls include an isotype control and a negative tissue control [38–40]. Isotype controls are used in assay with monoclonal primary antibodies and involve substitution of pre-immune or isotype-specific sera at the same protein concentration as the primary antibody. Negative tissue controls are performed by staining of a cell line or tissue known to not express the protein of interest. Negative controls that omit the primary antibody only are inadequate and are better known as secondary antibody only controls. The use of this type of control ensures only that the secondary antibody

does not exhibit nonspecific binding but does not provide information regarding specificity of staining with the primary antibody. Finally, it is important to ensure that the antibody of interest is providing the appropriate staining pattern as per the subcellular location of the antigen of interest. This includes both extracellular and intracellular antigens, with intracellular antigens further categorized by their predominant cellular distribution as membranous, nuclear, and cytoplasmic. The methods of validation mentioned above are best performed with the assistance of a trained pathologist, as their expertise ensures quality interpretation.

Each step in the tissue staining process can potentially be a source of variation. For instance, manual IHC protocols tend to have more “opportunities” for minor variations compared to automated systems [21]. For studies that involve a medium to large volume of IHC slides, multiple runs (“batches”) may be required to complete the project. Each batch has the potential to produce differences in staining quality (“batch effects”) due to variations in solutions, incubation times, temperatures, and other factors. Randomization of the slides can prevent biasing the data between batches and the use of appropriate controls can assure the intended sensitivity and specificity of each IHC batch [20, 27]. As another example, counterstains are not often considered to be a major influencing factor in IHC; however, similar coloration (e.g., red) or localization (e.g., nucleus) between counterstains and expected IHC stains can potentially confound effective scoring during analysis.

Tissue scoring

Key principles and approaches for scoring tissues can be applied to a variety of tissue-labeling techniques, including histochemical, immunohistochemical, or immunofluorescence. Best practices put forward by the Society of Toxicologic Pathology state that scoring systems should be definable, reproducible, and meaningful [41]. This includes a thorough examination of all tissues with clearly articulated lesion parameters and scoring definitions. Transparency in the details of experimental design or ancillary data such as clinical chemistries, functional studies, or imaging may further aid in assessment [4, 19, 28].

Reproducibility of tissue scoring can be maximized by following key principles and approaches to prevent bias. For instance, “masking” (or “blinding”) is a process that prevents the observer from having knowledge of treatment and/or group assignments, with the goal of limiting unintentional or subjective observation biases that may skew data interpretation [19, 42, 43]. Different types of masking approaches can be applied to investigational studies, though it is important for the observer to have sufficient

information or background regarding the study to reproducibly and accurately score the tissues. For example, complete masking (i.e., each individual sample is labeled distinctly as A, B, C, D, E, etc.) significantly restricts the scorer from any details of the study goals, treatments, or grouping. This approach may seem unbiased, but it actually hinders the observer (ideally a pathologist experienced with the disease and model) from identifying group-specific or unexpected changes that may be outside the purview of investigational expectations, thus creating the potential bias for false-negative results [41]. Another approach is grouped masking (i.e., A1, A2, A3 vs. B1, B2, B3), which allows for the observer to have sufficient information regarding the study groups (e.g., A vs. B) to make appropriate scientific interpretations and yet still limits observer knowledge of specific treatments for each group assignment [19]. In contrast, some situations may call for the observer to integrate the microscopic data with all other available sources of information to reach an appropriate interpretation. In such cases, an initial evaluation of the tissues and data is performed in an unblinded fashion to ensure that all lesion parameters are detected. This is followed by a masked review of randomized tissues (post-examination masking), at which time scores may be applied [9, 19].

Sampling is another issue that can greatly influence scoring systems. How are tissues sampled for harvest, for histopathology, or for image analysis? Is tissue collection at necropsy random, based on defined anatomic sites, or are only obvious lesions collected? At histopathology, does scoring represent a whole tissue section, or a defined area (random $\times 10$ objective microscopic field), or multiple areas (average score from five random $\times 10$ objective microscopic fields)? There is not often a uniform answer for sampling; rather it is project-dependent and influenced by lesion qualities (such as distribution, incidence, severity, etc.) within tissues. Increased sampling of the tissue often leads to better estimations of the true features of the lesion; however, increased sampling typically comes at a cost, both fiscally and in terms of time and labor. Importantly, the goal of tissue sampling is to best represent the true nature or quality of the tissue lesion for scoring analysis.

Consistency in tissue scoring is critical for reproducibility [19, 44]. While this may seem obvious, it can be difficult to maintain in certain situations. For example, consistency in semiquantitative (a.k.a. “grading”) and sometimes quantitative scoring can vary slightly when (1) one observer evaluates a large cohort of slides, (2) multiple pathologists evaluate different cohorts of slides, or (3) batches of slides are evaluated over periods of time; situations known as “diagnostic drift.” [45] Awareness of these issues along with quality control checks (by the same observer or an outside consultant) can help identify and mitigate diagnostic drift. As another example, data reporting

of methodology allows for transparency and clarity. Vague and subjective reporting can limit intra-observer and inter-observer reproducibility. In contrast, distinct, well-defined, and evaluable parameters can increase reproducibility and scientific rigor.

Semiquantitative assessment

Semiquantitative scoring systems are widely used methods to assess stained tissues and can often serve as a first-line or complementary approach to quantitative methods for statistical evaluation of groups. In semiquantitative scoring, the observer assigns a score (or “grade”) to tissue changes, to allow for subsequent statistical analysis [13, 19, 28]. As there are several types of semiquantitative approaches available, the type of scoring system should match the study design and questions to be addressed. Common approaches are described below.

Incidence method

Categorical data are defined by a group or qualitative trait to form a simple classification approach [10, 16, 19]. This type of data structure lacks hierarchical and progressive changes in extent or severity. Applying this concept to evaluation of tissue changes between two groups, a scoring system can divide tissues into two groups—e.g. “affected” (presence of a defined lesion) and “unaffected” (normal) groups using a predefined phenotype and clear definition of normal [44, 46]. For example, immunostaining for cellular markers can be qualitatively evaluated in tumors as present or absent, as defined by a certain threshold, to produce an case incidence (%) for each group [47]. Immunostaining for tumor protein 53 from two groups of benign ($n = 20$) and malignant ($n = 40$) tumors could be scored using a predefined “positive” staining threshold of $>50\%$ of cell immunostaining (Supplemental Table 1, Supplemental Fig. 1). In this mock example, if positive immunostaining was noted in three of the benign samples (17 negative) and 29 of the malignant tumors (11 negative), a Fisher’s exact test analysis yields a statistically significant difference of $P < 0.0001$. While the incidence method can be useful to detect the presence or absence of a phenotype, distinguishing ranges of a variable phenotype may be limited.

Rank method

The rank (“ordering”) method of scoring is a simple and quick approach in which each sample from a cohort is sorted from least to most severely affected for a given finding/lesion [19, 44]. For example, if someone was comparing two groups ($n = 6/\text{group}$), then these samples

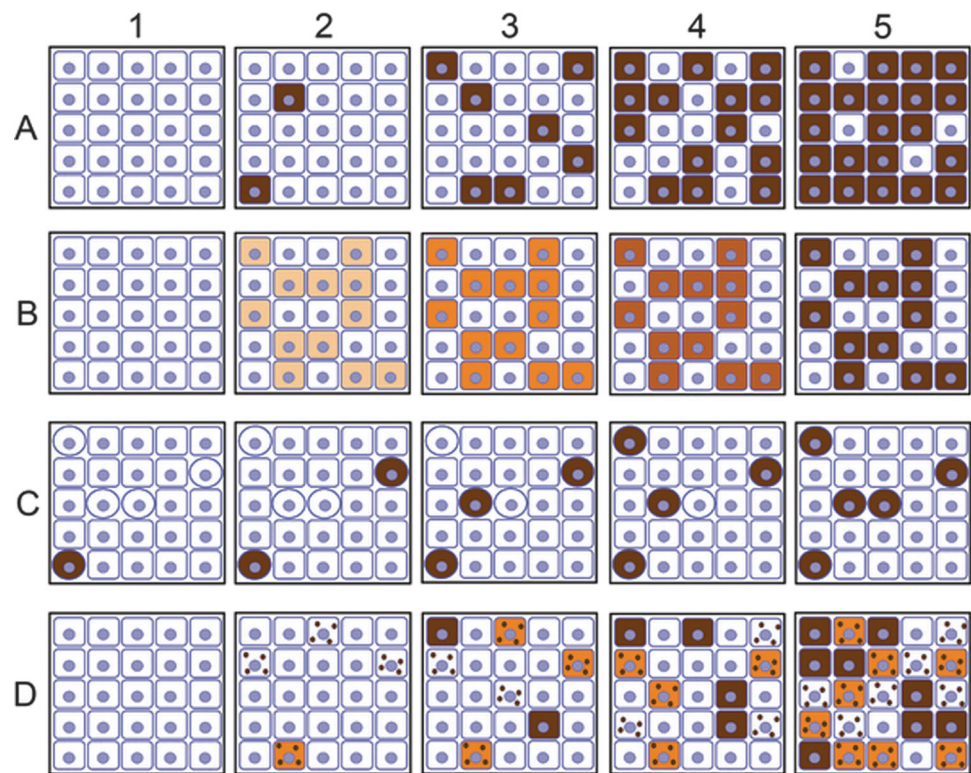
($n = 12$ total) could be masked, randomized, and then ranked according to a defined extent or severity of lesion. In this mock example, results for group A (ranked as 1, 2, 3, 5, 6, and 7) and B (ranked as 4, 8, 9, 10, 11, and 12) can then be analyzed by a nonparametric statistical test ($P = 0.0152$, Mann–Whitney test). This approach is simple to apply and can reduce the potential for “diagnostic drift” as tiered grades are not used [44].

Ordinal method

In clinical and preclinical research, ordinal scoring is a common method for semiquantitative scoring. In this approach, tissue changes are segregated into tiered scores (or “grades”) of progressive severity that best reflect the magnitude or distribution of tissue involvement [13, 23, 28, 46, 48, 49]. The number of score categories in an ordinal scoring system typically ranges from about 3 to 5, but has rarely extended upwards of 50 [13, 19, 23, 24, 44]. Fewer score categories may reduce the sensitivity of the system, while increased categories tend to reduce reproducibility as there is less obvious distinction between each one. It has been suggested that 4–5 score categories may be the optimal number to maximize detection and repeatability [19, 23, 50].

Ordinal scores (e.g., 0, 1, 2, 3, and 4) are generally reflective of cellular immunostaining frequency or intensity (see Fig. 1a, b, respectively). Using frequency (Fig. 1a), the observer/pathologist can estimate the cellular staining incidence (%) in each tissue and these will define the respective score (e.g. “1”: none, “2”: 1–25%, “3”: 26–50%, “4”: 51–75%, and “5”: 76–100% cells stained). With well-defined parameters, this approach has moderate to good reproducibility. However, if the scope of the ordinal scoring system does not closely match the scope of the score data, evaluation of the samples for group-specific differences may prove difficult. For instance, if the frequency scoring from Fig. 1a was applied to the incidence of immunostaining seen in Fig. 1c, determination of group-specific differences would be difficult as all samples would be scored as low incidence (e.g., “1” or “2”). In situations where a scoring system does not match the range of the tissues being evaluated, one can either fine-tune the definition of each ordinal grade to fit the full range of tissue changes (for all groups being evaluated) or change to another type of scoring system (e.g., quantitative). This brings up a point of “normalization”. For example, when evaluating all the cells in this section (Fig. 1c) as the denominator, the variation in immunostaining frequency is small (frequency range of 0–25%). But if one can identify specific and biologically relevant cell types by morphology (e.g., Fig. 1c, round cells only) or special stains, then the dynamic range of immunostaining variation can dramatically change and in this case, the frequency can increase to

Fig. 1 Types of cellular staining patterns (brown coloration) with differences in frequency (a), intensity (b), cell type (c), or mixed staining (d)



~20–100% staining of target cells. This approach can allow for more biologically relevant and testable evaluation of IHC in cells/tissues. While the previous mock examples (Fig. 1a–c) have straightforward immunostaining for scoring purposes, in practice, IHC tissue samples often have a more diverse appearance (Fig. 1d). In these situations, one may need to define “positive” staining by clear and reproducible thresholds as deemed relevant for the project. For example, if “positive” immunostaining is defined by moderate to strong and diffuse cytoplasmic staining (i.e., Fig. 1b, columns 3–5), then applying these parameters to Fig. 1d yields a range of immunostaining frequency of 0%, 4%, 20%, 32%, and 64%, respectively.

Scores can also be assigned from multiple parameters (e.g., frequency and intensity) to form a “composite” score [51, 52]. The immunoreactivity score (IRS) is a commonly utilized composite score in both the clinical setting and translational research. The IRS is the sum of the ordinal scores for distribution and intensity of immunostaining (Supplemental Table 2) [53]. A clinical example of this type of composite score is the Allred score, originally developed for assessment of estrogen receptor immunostaining [54–57]. A variation to the IRS is the *H*-score (Supplemental Table 3), which also assigns an ordinal score to the immunostaining intensity and multiplies this by an estimate of the percentage of immunostained tissue for each intensity grade, yielding total scores between 0 and 300 [20, 54].

Semiquantitative scoring methods have several advantages, including ease of use and requiring nominal to no specialized expertise or equipment (e.g., software packages). As such, it is also a cost-effective approach as there are seldom any associated input costs. Most often, semiquantitative scoring is used in research for determination of group-specific differences. These initial scores can be used as stand-alone data, to corroborate clinical data on the project or to better target quantitative assessments.

Semiquantitative systems are not without limitations. As these are based on manual and subjective visual assessment, they are susceptible to some level of observer bias and variability that could contribute to reproducibility issues [2, 3, 7, 15–17, 19, 23, 46, 58, 59]. These reproducibility issues can be constrained through proper masking of observers and well-defined scoring systems that clearly define each grade. In addition, ordinal scoring systems may not always represent a true linear relationship in grades [23, 44, 46]. That is to say that the differences (intervals) in points on the ordinal scale are not always equivalent (i.e., the difference between a score of “1” and “3” is not necessarily of the same magnitude as the difference between a score of “2” and “4”). This consideration may be most relevant when trying to correlate semiquantitative scoring to quantitative biology data.

Analysis of semiquantitative scores is important once scoring is completed and can be broadly segregated into two

general approaches [43]. The first approach of analysis is that of validation, which often occurs when first establishing a scoring system. Validation of repeatability is a way to evaluate whether the scoring results on stained tissues can be consistently repeated at different times [19, 60, 61]. This repeated examination and scoring in a masked fashion can be performed by the same person (intra-observer) or by other observers (inter-observer), but the use of multiple observers in demonstrating reproducibility is arguably a more robust approach. The overarching goal is to show a significant correlation between the observers scores from the same stained tissues, thus, demonstrating reproducibility. Validation of pathobiology is also important consideration [19, 48, 62, 63]. Similar to validation of reproducibility, the overarching goal in validation of pathobiology is to show a significant correlation between tissue scores and relevant parameters of biological disease severity. For example, if the semiquantitative scores for a putative tissue marker of lung disease does not correlate with clinical and ancillary parameters of disease severity, then this raises serious questions about the utility and feasibility of the scoring system.

The second approach is that of evaluating tissue scores between treatment groups by statistical analyses and these types of tests have been discussed elsewhere in more detail [23, 43, 64]. Investigators who analyze semiquantitative data should try to avoid common pitfalls. Most semiquantitative data are nonparametric in nature and when evaluating for group-specific differences, nonparametric statistical test should be used. For instance, if comparing two groups of semiquantitative data, nonparametric tests that can be applied in many situations include Mann–Whitney *U*-test or the Kolmogorov–Smirnov test. Paired or unpaired *T*-tests are parametric tests that if used on semiquantitative (nonparametric) data, could result in interpretations that are prone to error and should be avoided. Alternatively if multiple groups or parameters are studied, then analysis of variance approaches can be used in the analysis of the semiquantitative data. These are just a few common examples for how to analyze semiquantitative data. It is not the intention for this paper to give comprehensive advice on statistical evaluation of semiquantitative data as there are multiple permutations and considerations that can influence such a decision. Importantly, inclusion of professional expertise in statistical analysis (e.g., statistician) from the start of a project is highly recommended as it can greatly enhance the rigor and impact of scientific studies [65].

Quantitative assessment

While semiquantitative approaches can be useful to detect group differences in many situations, quantification of

tissue staining may be warranted to provide increased robustness to the dataset. The use of image analysis of tissue labeling has increased in recent years for numerous applications in the clinical setting, including, diagnostic and prognostic determinations, as well as in the research setting for evaluating protein expression and correlating with other quantitative assays, such as real-time PCR [6, 66, 67]. Compared to semiquantitative approaches, quantitative assessment of tissue staining has the potential to produce data that are more rigorous and on a continuous scale that allow for more precise correlations to clinical or biological data [14, 19, 24, 68]. Several recent papers have discussed the approaches, advantages, and limitations of quantitative scoring of tissues and the reader is encouraged to examine these for more specific information [69, 70].

In general, there are two major approaches to quantitative evaluation of tissue staining; manual and automated image analysis with some minor overlap of both approaches. Historically, quantification of stained regions relied upon labor intensive manual methods, including point counting of images projected onto grids [71], using microscope-based micrometers [72], or evaluating black and white color micrographs [73]. While microscope-based quantification methods such as counting the frequency or percent area of stained cells are still used today (Table 2) [74–76], there have been significant strides in digital pathology applications [70, 77]. The remainder of this paper will focus on principles useful to quantitative scoring that is generally independent of the approach, methodology, or software application.

There are important analytical factors to initially consider when using quantitative analysis, including the choice of label (e.g., chromogen or fluorescent dye) for detection. Chromogens are frequently used as detection agents in anatomic pathology applications and are available in a range of colors, with the most commonly used compound being

Table 2 Examples of common quantitative scoring methods and parameters used to evaluate labeled tissues

Approach	Method	Parameters
Quantitative	% Stained	Cells [90]
		Nuclei [75]
		Area [75]
		Pixels [2]
Counts	Counts	Membrane [91, 92]
		Number (cell, nuclei, etc.) per field/area [76, 93]
Automated ^a	Automated ^a	Circularity [94, 95]
		Optical density [11, 96]
Stereology	Stereology	Stereology [83, 97]

^aSeveral options are available based for each software platform, but only a selected few examples of automated outputs are reported here

DAB, which results in brown staining [78]. Though the use of chromogens has advantages, such as ease of interpretation in morphological context and simple equipment requirements, there are limitations for their use with quantitative methods [10, 78]. Assessment of staining amount is based on measuring absorption, and the optimal absorbance for DAB is 1–2 units (meaning that up to 99% of the light is blocked by the substrate, leaving only 1% of the total signal available for analysis) [24, 66]. This limits the ability of chromogens to be used in multiplexing, and also contributes to difficulties in maximizing the dynamic range of the assay, which is the total range of values that can be obtained from a particular assay. When utilizing DAB, the dynamic range of IHC is about one to two logs; however, protein expression *in vivo* usually spans at least two logs and can vary up to four logs in cases of gene amplification. Thus, the limited dynamic range of a chromogen-based IHC assay may allow for only half of the information to be gathered, requiring more than one antibody concentration to be assayed in order to cover the entire dynamic range of protein expression [10, 66, 78].

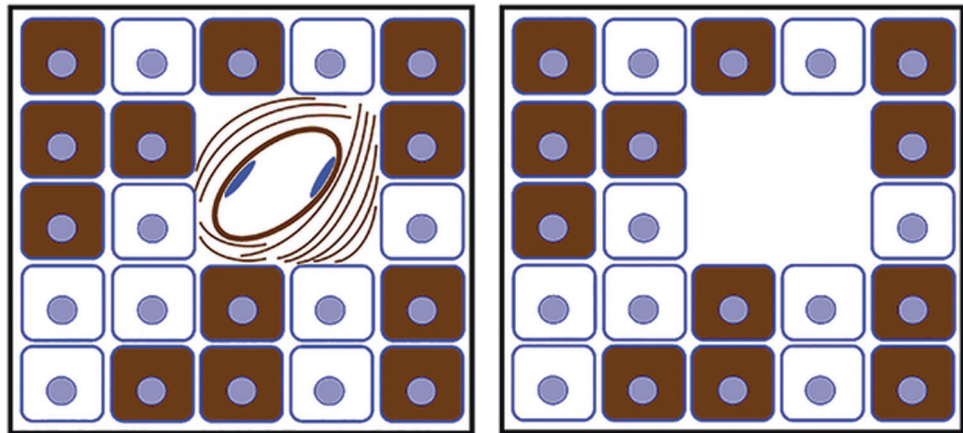
Using antibodies directly labeled with a fluorescent dye as a detection method in image analysis has been shown to have increased sensitivity and reproducibility as compared to the use of chromogens [3, 9, 24, 25, 42, 78]. Fluorescent probes have a broader dynamic range (approximately 2–3 logs) than chromogens. Additionally, as opposed to chromogens that absorb light, fluorescent dyes actually emit light, so they are more amenable to multiplexing, as the number of markers utilized is not limited by light absorption. Drawbacks of fluorescent staining include fading (limited half-life) of stains, presence of autofluorescence in tissues, increased expense, and limited morphologic assessment [79].

For quantitative digital analysis, a whole-slide imaging system can convert the glass slide into a high-resolution, high-contrast image [22, 56, 67]. Red–green–blue (RGB) images are produced by conventional imaging technology, in which each pixel contains information related to the extent of red, green, or blue color channels, each with a 1–256 (28 or 8 bit) intensity range [6, 16]. Technical limitations can result in imaging artifacts, such as suboptimal image contrast, sharpness and resolution, varied chromogen staining intensities, or overlapping chromogens in multiplexed stains, which directly affect data accuracy and reproducibility [6, 14, 58]. Multispectral imaging (MSI) systems have been developed to overcome the limits of RGB imaging and improve quantitation in both bright-field and fluorescent microscopy. By acquiring a stack of images at multiple wavelengths, MSI can obtain color spectral information at each pixel of an image that is not limited to only three channels. In addition, images are of increased resolution and contrast and MSI is able to separate overlapping and/or multiple chromogens [6, 58].

The resulting digital image is then analyzed by commercial and/or freely available software to provide quantitative information [11]. The software uses complex mathematical algorithms to process and separate the image into regions with similar characteristics (such as color, intensity, or texture) or to calculate relevant tissue parameters such as cell or staining per area (Table 2) [14, 16, 80]. All digital images consist of pixels, which are each composed of numerical values that define the color, allowing for patterns to be mathematically analyzed by computerized pixel profiling [11, 16]. After calibration of the system via control and reference samples to ensure accuracy and reproducibility, the next critical step for quantification and consistent data analysis is the determination of threshold values (or “cutoffs”) for intensity, as a means to define the limits of staining intensity for inclusion of cells as “positive” or “negative” in the scoring [2, 11]. Because thresholding may be subjective, this is preferably performed with the aid of statistical analysis, as artificially low threshold values lead to a percent area stained of up to 100% (all pixels identified as stained) and overly high threshold values lead to percent area stained values close to zero [7, 25]. A challenge that can commonly be seen in tissues is that of mixed staining (intensity and cellular distribution, see one example in Fig. 1d). Establishing threshold limits is an important step to maximize inclusion appropriate staining (e.g., cellular, pixels, etc.) while limiting the inclusion of potential nonspecific staining. The reader would be encouraged to read these reviews for more specific details about thresholding approaches quantitative analysis [16, 81, 82].

When reporting quantitative data, the assumption is made that the signal on the slide is representative and quantitatively related to the amount of antigen in the sampled section of tissue, which is in turn related to the absolute amount of the antigen in the tissue as a whole [14]. Data may be presented as a ratio, such as the amount of antigen expression (as assessed immunohistochemically) relative to the area in which the target of interest is expressed [16]. There has been debate surrounding the method of quantification, regarding its basis in either two-dimensional or three-dimensional counting methods [13]. Two-dimensional model methods are based on counting cell profiles in one (or a few) two-dimensional planar sections used to represent the three-dimensional tissue. As this may lead to biased results, there have been attempts to instead count cells in three-dimensional space. Stereology describes the mathematical methods used to obtain spatial information for three-dimensional structures (like tissues) from two-dimensional projections [83, 84]. Briefly, random fields of vision spread over a defined area of interest are sampled and in each field, cells are selected using a point grid imposed on the images and the positivity of the cell is scored [8, 24]. Three-

Fig. 2 Tissues for quantitative scoring can be examined and nontargeted tissues (left panel, interstitial tissues/vessels) can be removed to leave only targeted tissues/stains (right panel)



dimensional (or volumetric) approaches are based on statistical principles of sampling relatively few cells from relatively many fields; it is both laborious and complex as compared to two-dimensional analysis and often reserved for specialized labs.

Numerous studies have indicated a high degree of correlation between digital image analysis and pathologist visual scoring, but there are several additional advantages to digital analyses [85].

Accuracy

Automated IHC measurements result in a greater degree of objectivity and reproducibility in the assessment of morphological features as compared to manual evaluation and are more suited to high-throughput sample processing [3, 7–9, 22, 56, 59, 84, 86]. The human visual system is highly skilled at pattern recognition for morphologic changes, but has limited ability to detect subtle changes in tissue, particularly in relation to spatial and density assessments [13, 58, 84]. The eye is also inaccurate at detecting differences at low intensity (weak staining), which are the conditions at which IHC staining is most linearly related to antigen concentration [22, 87]. Conversely, automated IHC measurements are precise in these staining/intensity ranges.

Speed

Digital systems can often quantitate staining data to a greater degree and with greater speed than the human eye. Current systems of analysis may require just seconds to complete for each tissue, whereas it may take several hours or even days for manual analysis by a pathologist.

Multiplex evaluations

Finally, digital image analysis methods allow for IHC to be multiplexed to assess the relationship of two or more targets

simultaneously [7, 22]. In situ hybridization and IHC techniques can even be combined to gain information about a given target at both the protein and DNA/mRNA level [7].

Despite all the advantages to digital image analysis, there are still limitations. As described above, there are a multitude of tissue variables that need to be controlled to create a stained slide of appropriately high quality for analysis. Inconsistencies such as uneven fixation, varying sectioning thickness, and irregular chromogen precipitation may cause inaccurate staining intensity measurements [2]. Cellular density in some tissues is often lower near the edges of a section (as opposed to the middle), so counting areas near the margins may result in artificially low results [88]. To combat this concern, it is recommended to use “guard zones” at the edge of tissues, regions in which quantitative is not performed.

Additionally, the use of computer-based analysis may still have limited usefulness in routine clinical examinations that have traditionally relied on manual evaluation by a pathologist [23]. Often, the limiting factor in such analysis is the quality of the image to be quantified. This can be camera-dependent, in that the pixel count of the image varies based on the camera’s resolution [11, 16]. Computer-based errors in hardware (such as inconstant illumination or insufficiently broad dynamic range for the camera) or software problems related to the analysis algorithms may also contribute [59, 87]. Most software programs also require manual involvement, at least during initial set-up. This includes the way in which the areas of interest (and conversely, areas not of interest), background staining, and foci of positive staining are identified and defined [13, 22]. Trained pathologists are often needed to help set threshold values and distinguish the area of interest for analysis from the tissue as a whole (Fig. 2) [2, 13, 89]. This is most important in heterogeneous samples (Fig. 1d) or in tissue sections containing both neoplastic and non-neoplastic elements. While these activities may eventually be automated (following a period of programming or “training” of the

software), there is an initial period of increased time and reduced efficiency of analysis. Also, there may be significant inter-laboratory variability in threshold values, as they can be set arbitrarily by the pathologist or observer [2, 7, 11, 25].

Summary

Semiquantitative and quantitative scoring of labeled tissues are useful ways to expand the scope, depth, and rigor of research studies. Following key principles of scoring can increase reproducibility and confidence in the resulting conclusions.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Adissu HA, Estabel J, Sunter D, et al. Histopathology reveals correlative and unique phenotypes in a high-throughput mouse phenotyping screen. *Dis Model Mech*. 2014;7:515–24.
- Choudhury KR, Yagle KJ, Swanson PE, et al. A robust automated measure of average antibody staining in immunohistochemistry images. *J Histochem Cytochem*. 2010;58:95–107.
- Cregger M, Berger AJ, Rimm DL. Immunohistochemistry and quantitative analysis of protein expression. *Arch Pathol Lab Med*. 2006;130:1026–30.
- Ladiges W, Ikeno Y, Liggitt D, et al. Pathology is a critical aspect of preclinical aging studies. *Pathobiol Aging Age Relat Dis*. 2013;3. <https://doi.org/10.3402/pba.v3i0.22451>.
- Nielsen S. External quality assessment for immunohistochemistry: experiences from NordiQC. *Biotech Histochem*. 2015;90:331–40.
- Liu WL, Wang LW, Chen JM, et al. Application of multispectral imaging in quantitative immunohistochemistry study of breast cancer: a comparative study. *Tumour Biol*. 2016;37:5013–24.
- Okoye JO, Nnatuanya IN. Immunohistochemistry: a revolutionary technique in laboratory medicine. *Clin Med Diagn*. 2015;5:60–9.
- Prasad K, Prabhu GK. Image analysis tools for evaluation of microscopic views of immunohistochemically stained specimen in medical research—a review. *J Med Syst*. 2012;36:2621–31.
- Rojo MG, Bueno G, Slodkowska J. Review of imaging solutions for integrated quantitative immunohistochemistry in the pathology daily practice. *Folia Histochem Cytobiol*. 2009;47:349–54.
- Rudbeck L. Adding quality to your qualitative IHC. *MLO Med Lab Obs*. 2015;47:21.
- Varghese F, Bukhari AB, Malhotra R, et al. IHC Profiler: an open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples. *PLoS ONE*. 2014;9:e96801.
- Lin F, Chen Z. Standardization of diagnostic immunohistochemistry: literature review and geisinger experience. *Arch Pathol Lab Med*. 2014;138:1564–77.
- Scudamore CL. Practical approaches to reviewing and recording pathology data. A practical guide to the histology of the mouse. John Wiley & Sons, Ltd; Chennai, India, 2014. p. 25–41.
- Taylor CR, Levenson RM. Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment II. *Histopathology*. 2006;49:411–24.
- Axelrod DE, Miller N, Chapman JA. Avoiding pitfalls in the statistical analysis of heterogeneous tumors. *Biomed Inform Insights*. 2009;2:11–8.
- Dunstan RW, Wharton KA Jr, Quigley C, et al. The use of immunohistochemistry for biomarker assessment—can it compete with other technologies? *Toxicol Pathol*. 2011;39:988–1002.
- Eaton KA, Danon SJ, Krakowka S, et al. A reproducible scoring system for quantification of histologic lesions of inflammatory disease in mouse gastric epithelium. *Comp Med*. 2007;57:57–65.
- Engel KB, Moore HM. Effects of preanalytical variables on the detection of proteins by immunohistochemistry in formalin-fixed, paraffin-embedded tissue. *Arch Pathol Lab Med*. 2011;135:537–43.
- Gibson-Corley KN, Olivier AK, Meyerholz DK. Principles for valid histopathologic scoring in research. *Vet Pathol*. 2013;50:1007–15.
- Kraus JA, Dabbs DJ, Beriwal S, et al. Semi-quantitative immunohistochemical assay versus oncotype DX((R)) qRT-PCR assay for estrogen and progesterone receptors: an independent quality assurance study. *Mod Pathol*. 2012;25:869–76.
- Olivier AK, Naumann P, Goeken A, et al. Genetically modified species in research: opportunities and challenges for the histology core laboratory. *J Histotechnol*. 2012;35:63–7.
- Rizzardi AE, Johnson AT, Vogel RI, et al. Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring. *Diagn Pathol*. 2012;7:42.
- Shackelford C, Long G, Wolf J, et al. Qualitative and quantitative analysis of nonneoplastic lesions in toxicology studies. *Toxicol Pathol*. 2002;30:93–6.
- van Diest PJ, van Dam P, Henzen-Logmans SC, et al. A scoring system for immunohistochemical staining: consensus report of the task force for basic research of the EORTC-GCCG. European Organization for Research and Treatment of Cancer-Gynaecological Cancer Cooperative Group. *J Clin Pathol*. 1997;50:801–4.
- Walker RA. Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment I. *Histopathology*. 2006;49:406–10.
- Scudamore CL, Soilleux EJ, Karp NA, et al. Recommendations for minimum information for publication of experimental pathology data: MINPEPA guidelines. *J Pathol*. 2016;238:359–67.
- Lin PL, Rodgers M, Smith L, et al. Quantitative comparison of active and latent tuberculosis in the *Cynomolgus* Macaque model. *Infect Immun*. 2009;77:4631–42.
- Scudamore CL. Integrating pathology into human disease modelling—how to eat the elephant. *Dis Model Mech*. 2014;7:495–7.
- Dettmeyer RB. *Forensic histopathology: fundamentals and perspectives*. New York: Springer; 2011.
- Williams JM, Duckworth CA, Burkitt MD, et al. Epithelial cell shedding and barrier function: a matter of life and death at the small intestinal villus tip. *Vet Pathol*. 2015;52:445–55.
- Ramos-Vara JA, Webster JD, DuSold D, et al. Immunohistochemical evaluation of the effects of paraffin section storage on biomarker stability. *Vet Pathol*. 2014;51:102–9.
- Grillo F, Pigozzi S, Ceriolo P, et al. Factors affecting immunoreactivity in long-term storage of formalin-fixed paraffin-embedded tissue sections. *Histochem Cell Biol*. 2015;144:93–9.
- Economou M, Schoni L, Hammer C, et al. Proper paraffin slide storage is crucial for translational research projects involving immunohistochemistry stains. *Clin Transl Med*. 2014;3:4.
- Cheung CC, Banerjee D, Barnes PJ, et al. Canadian Association of Pathologists-Association canadienne des pathologistes National Standards Committee for High Complexity Testing/Immunohistochemistry: guidelines for the preparation, release, and storage of

- unstained archived diagnostic tissue sections for immunohistochemistry. *Am J Clin Pathol.* 2014;142:629–33.
35. Xie R, Chung JY, Ylaya K, et al. Factors influencing the degradation of archival formalin-fixed paraffin-embedded tissue sections. *J Histochem Cytochem.* 2011;59:356–65.
 36. Torlakovic EE, Nielsen S, Vyberg M, et al. Getting controls under control: the time is now for immunohistochemistry. *J Clin Pathol.* 2015;68:879–82.
 37. Leach MW, Halpern WG, Johnson CW, et al. Use of tissue cross-reactivity studies in the development of antibody-based biopharmaceuticals: history, experience, methodology, and future directions. *Toxicol Pathol.* 2010;38:1138–66.
 38. Schonbrunn A. Editorial: antibody can get it right: confronting problems of antibody specificity and irreproducibility. *Mol Endocrinol.* 2014;28:1403–7.
 39. Bordeaux J, Welsh A, Agarwal S, et al. Antibody validation. *Biotechniques.* 2010;48:197–209.
 40. Hewitt SM, Baskin DG, Frevert CW, et al. Controls for immunohistochemistry: the Histochemical Society's standards of practice for validation of immunohistochemical assays. *J Histochem Cytochem.* 2014;62:693–7.
 41. Crissman JW, Goodman DG, Hildebrandt PK, et al. Best practices guideline: toxicologic histopathology. *Toxicol Pathol.* 2004;32:126–31.
 42. Cohen DA, Dabbs DJ, Cooper KL, et al. Interobserver agreement among pathologists for semiquantitative hormone receptor scoring in breast carcinoma. *Am J Clin Pathol.* 2012;138:796–802.
 43. Meyerholz DK, Sieren JC, Beck AP, et al. Approaches to evaluate lung inflammation in translational research. *Vet Pathol.* 2018;55:42–52.
 44. Holland T, Holland C. Analysis of unbiased histopathology data from rodent toxicity studies (or, are these groups different enough to ascribe it to treatment?). *Toxicol Pathol.* 2011;39:569–75.
 45. Cross SS. Grading and scoring in histopathology. *Histopathology.* 1998;33:99–106.
 46. Holland T. The comparative power of the discriminant methods used in toxicological pathology. *Toxicol Pathol.* 2005;33:490–4.
 47. Shalin SC, Sakharpe A, Lyle S, et al. p53 staining correlates with tumor type and location in sebaceous neoplasms. *Am J Dermatopathol.* 2012;34:129–35 (quiz 136–8).
 48. Klopffleisch R. Multiparametric and semiquantitative scoring systems for the evaluation of mouse model histopathology—a systematic review. *BMC Vet Res.* 2013;9:123.
 49. Fedchenko N, Reifemrath J. Different approaches for interpretation and reporting of immunohistochemistry analysis results in the bone tissue—a review. *Diagn Pathol.* 2014;9:221.
 50. Schafer KA, Eighmy J, Fikes JD, et al. Use of severity grades to characterize histopathologic changes. *Toxicol Pathol.* 2018;46:256–65.
 51. Knudson CJ, Hartwig SM, Meyerholz DK, et al. RSV vaccine-enhanced disease is orchestrated by the combined actions of distinct CD4 T cell subsets. *PLoS Pathog.* 2015;11:e1004757.
 52. Snyder JM, Ward JM, Treuting PM. Cause-of-death analysis in rodent aging studies. *Vet Pathol.* 2016;53:233–43.
 53. 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;8:138–40.
 54. Budwit-Novotny DA, McCarty KS, Cox EB, et al. Immunohistochemical analyses of estrogen receptor in endometrial adenocarcinoma using a monoclonal antibody. *Cancer Res.* 1986;46:5419–25.
 55. Ellis MJ, Coop A, Singh B, et al. Letrozole is more effective neoadjuvant endocrine therapy than tamoxifen for ErbB-1-and/or ErbB-2-positive, estrogen receptor-positive primary breast cancer: evidence from a phase III randomized trial. *J Clin Oncol.* 2001;19:3808–16.
 56. Engelberg JA, Giberson RT, Young LJ, et al. The use of mouse models of breast cancer and quantitative image analysis to evaluate hormone receptor antigenicity after microwave-assisted formalin fixation. *J Histochem Cytochem.* 2014;62:319–34.
 57. Allred DC, Harvey JM, Berardo M, et al. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol.* 1998;11:155–68.
 58. Levenson RM, Hoyt CC. Spectral imaging and microscopy. *Am Lab.* 2000;32:26.
 59. True LD. Quantitative immunohistochemistry: a new tool for surgical pathology? *Am J Clin Pathol.* 1988;90:324–5.
 60. Cross SS. Kappa statistics as indicators of quality assurance in histopathology and cytopathology. *J Clin Pathol.* 1996;49:597–9.
 61. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–74.
 62. Bleich A, Mahler M, Most C, et al. Refined histopathologic scoring system improves power to detect colitis QTL in mice. *Mamm Genome.* 2004;15:865–71.
 63. Scheinin T, Butler DM, Salway F, et al. Validation of the interleukin-10 knockout mouse model of colitis: antitumor necrosis factor-antibodies suppress the progression of colitis. *Clin Exp Immunol.* 2003;133:38–43.
 64. Festing MF. The design and statistical analysis of animal experiments. *ILAR J.* 2002;43:191–3.
 65. Zeiss CJ, Ward JM, Allore HG. Designing phenotyping studies for genetically engineered mice. *Vet Pathol.* 2012;49:24–31.
 66. Rimm DL. What brown cannot do for you. *Nat Biotechnol.* 2006;24:914–6.
 67. Zehntner SP, Chakravarty MM, Bolovan RJ, et al. Synergistic tissue counterstaining and image segmentation techniques for accurate, quantitative immunohistochemistry. *J Histochem Cytochem.* 2008;56:873–80.
 68. Bankhead P, Fernandez JA, McArt DG, et al. Integrated tumor identification and automated scoring minimizes pathologist involvement and provides new insights to key biomarkers in breast cancer. *Lab Invest.* 2018;98:15–26.
 69. Hamilton PW, Bankhead P, Wang Y, et al. Digital pathology and image analysis in tissue biomarker research. *Methods.* 2014;70:59–73.
 70. Aeffner F, Wilson K, Martin NT, et al. The gold standard paradox in digital image analysis: manual versus automated scoring as ground truth. *Arch Pathol Lab Med.* 2017;141:1267–75.
 71. Blyth DI, Pedrick MS, Savage TJ, et al. Lung inflammation and epithelial changes in a murine model of atopic asthma. *Am J Respir Cell Mol Biol.* 1996;14:425–38.
 72. Elwood CM, Hamblin AS, Batt RM. Quantitative and qualitative immunohistochemistry of T cell subsets and MHC class II expression in the canine small intestine. *Vet Immunol Immunopathol.* 1997;58:195–207.
 73. Kumagai K, Ohno I, Imai K, et al. The involvement of matrix metalloproteinases in basement membrane injury in a murine model of acute allergic airway inflammation. *Clin Exp Allergy.* 2002;32:1527–34.
 74. Reid MD, Bagci P, Ohike N, et al. Calculation of the Ki67 index in pancreatic neuroendocrine tumors: a comparative analysis of four counting methodologies. *Mod Pathol.* 2015;28:686–94.
 75. Meyerholz DK, Stoltz DA, Namati E, et al. Loss of cystic fibrosis transmembrane conductance regulator function produces abnormalities in tracheal development in neonatal pigs and young children. *Am J Respir Crit Care Med.* 2010;182:1251–61.
 76. Barr JY, Wang X, Meyerholz DK, et al. CD8 T cells contribute to lacrimal gland pathology in the nonobese diabetic mouse model of Sjogren syndrome. *Immunol Cell Biol.* 2017;95:684–94.

77. Dee FR, Meyerholz DK. Teaching medical pathology in the twenty-first century: virtual microscopy applications. *J Vet Med Educ.* 2007;34:431–6.
78. Carvajal-Hausdorf DE, Schalper KA, Neumeister VM, et al. Quantitative measurement of cancer tissue biomarkers in the lab and in the clinic. *Lab Invest.* 2015;95:385–96.
79. Gu J, Hacker GW. *Modern methods in analytical morphology.* USA: Springer; 2012.
80. Uc A, Olivier AK, Griffin MA, et al. Glycaemic regulation and insulin secretion are abnormal in cystic fibrosis pigs despite sparing of islet cell mass. *Clin Sci (Lond).* 2015;128:131–42.
81. Johnson SJ, Walker FR. Strategies to improve quantitative assessment of immunohistochemical and immunofluorescent labelling. *Sci Rep.* 2015;5:10607.
82. Jensen EC. Quantitative analysis of histological staining and fluorescence using ImageJ. *Anat Rec (Hoboken).* 2013;296:378–81.
83. Brown DL. Practical stereology applications for the pathologist. *Vet Pathol.* 2017;54:358–68.
84. Wanke R. Stereology—benefits and pitfalls. *Exp Toxicol Pathol.* 2002;54:163–4.
85. Rizzardi AE, Zhang X, Vogel RI, et al. Quantitative comparison and reproducibility of pathologist scoring and digital image analysis of estrogen receptor beta2 immunohistochemistry in prostate cancer. *Diagn Pathol.* 2016;11:63.
86. Cardiff RD, Hubbard NE, Engelberg JA, et al. Quantitation of fixative-induced morphologic and antigenic variation in mouse and human breast cancers. *Lab Invest.* 2013;93:480–97.
87. Rimm DL, Giltman JM, Moeder C, et al. Bimodal population or pathologist artifact? *J Clin Oncol.* 2007;25:2487–8.
88. Hou J, Riise J, Pakkenberg B. Application of immunohistochemistry in stereology for quantitative assessment of neural cell populations illustrated in the Gottingen minipig. *PLoS ONE.* 2012;7:e43556.
89. Fiore C, Bailey D, Conlon N, et al. Utility of multispectral imaging in automated quantitative scoring of immunohistochemistry. *J Clin Pathol.* 2012;65:496–502.
90. Elias S, McGuire JR, Yu H, et al. Huntingtin is required for epithelial polarity through RAB11A-mediated apical trafficking of PAR3-aPKC. *PLoS Biol.* 2015;13:e1002142.
91. Borthwick LA, Botha P, Verdon B, et al. Is CFTR-deltaF508 really absent from the apical membrane of the airway epithelium? *PLoS One.* 2011;6:e23226.
92. Meyerholz DK, Stoltz DA, Gansemer ND, et al. Lack of cystic fibrosis transmembrane conductance regulator disrupts fetal airway development in pigs. *Lab Invest.* 2018. <https://doi.org/10.1038/s41374-018-0026-7>
93. Yuri P, Hendri AZ, Danarto R. Association between tumor-associated macrophages and microvessel density on prostate cancer progression. *Pros Int.* 2015;3:93–98.
94. Schochlin M, Weissinger SE, Brandes AR, et al. A nuclear circularity-based classifier for diagnostic distinction of desmoplastic from spindle cell melanoma in digitized histological images. *J Pathol Inform.* 2014;5(1):40.
95. Agle CC, Velloso CP, Lazarus NR, et al. An image analysis method for the precise selection and quantitation of fluorescently labeled cellular constituents: application to the measurement of human muscle cells in culture. *J Histochem Cytochem.* 2012;60:428–438.
96. Zhu C, Wang J, Cheng T, et al. The potential role of increasing the release of mouse beta-defensin-14 in the treatment of osteomyelitis in mice: a primary study. *PLoS One.* 2014;9:e86874.
97. Adam RJ, Michalski AS, Bauer C, et al. Air trapping and airflow obstruction in newborn cystic fibrosis piglets. *Am J Resp Crit Care Med.* 2013;188(12):1434–1441.