Digital microscope test materials and test methods

Document version 3

Notes on the status of this document

- Areas where there is substantial uncertainty are shown in red test and many sections are as yet undefined and marked 'TBD' – volunteers to provide input for these sections would be welcomed.
- This is currently written as an ICC White Paper that provides recommendations. Is this appropriate or is there another direction that should be considered?
- The current approach suggested is on the whole a measurement-based approach and may be unnecessarily complex. An alternative visual assessment approach would also be possible but that approach may make it more difficult to define acceptance criteria.
- Measuring slides and displays is currently described as being out of scope for this white paper. Should we include pointers to 'best practice' documents? There are standards for the measurement of displays but I do not know of any standards for the measurement of microscope slides. In order to obtain representative measurements of slides, consideration must be given to the geometry of the viewing optics.

Contributing authors' declaration of interest

W Craig Revie	Employed by FFFEI Limited, a developer of imaging hardware and software. FFEI has developed a scanning hardware platform that is provided on an OEM basis to a number of manufacturers of whole slide imaging systems.
Peter Jackson	Retired Biomedical Scientist. Part time lecturer in Cellular Pathology, Bradford university. Honorary researcher University of Leeds Pathology and Tumour Biology.
Wei-Chung Cheng	[TBD]
Mike Shires	Employed by The University of Leeds and runs the Histology Service for the Leeds Institutes of Molecular Medicine.
Yukako Yagi	[TBD: Assistant Professor of Pathology at Harvard Medical School and Director of the Pathology Imaging and Communication Technology Center at MGH.]
David McDowell	[TBD: Retired imaging expert with many years of experience of standardising processes for imaging systems. Expert for Kodak in many ISO Committees for many years.]
Darren Treanor	Employed by Leeds Teaching Hospitals Trust and runs the Leeds digital pathology project at the University of Leeds (www.virtualpathology.leeds.ac.uk).

[TBD: is this necessary?]

Contents

1 Scope

This ICC white paper provides recommendations for the design of test materials, test methods for the assessment of colour performance of digital microscopes, in particular whole slide imaging (WSI) systems. This is limited to bright field RGB imaging of stained pathology slides and particular attention is given to the assessment of samples stained with variations of Haematoxylin and Eosin (H&E) as this is the predominant staining protocol used in most laboratories worldwide.

To some extent the assessment methods anticipate that ICC colour management methodology is used and in particular that an ICC profile for the image capture system is associated with the images produced. The test methods, however, do not rely on any particular calibration model and digital microscope vendors are free to use any calibration system they would like as long as they meet the overall colour aims.

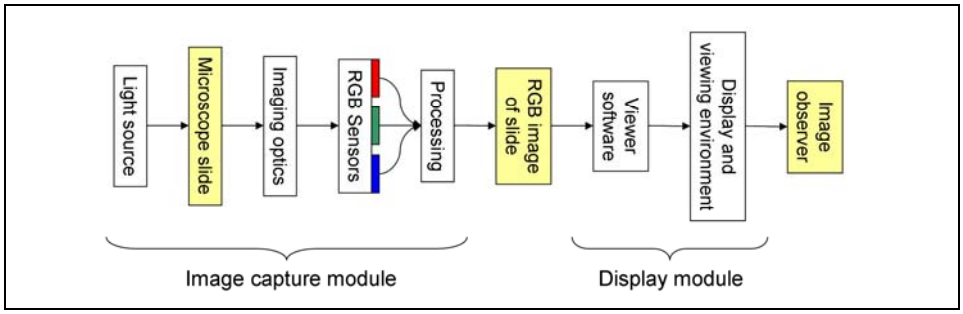
One source of substantial colour variation is the process of slide staining which has not been standardised and indeed staining to accommodate the preference of individual pathologists is a widespread practice. Although the staining of slides is out of the scope of this white paper, some understanding of the likely variations in colour due to different staining practice is necessary to ensure that the test materials includes a suitable set of colours.

The primary objective of these assessment methods is to ensure that digital microscopes are fit for purpose and to improve the reliability of diagnosis of disease using these systems.

2 Introduction

Digital microscopes, in particular whole slide imaging (WSI) systems provide an alternative to conventional optical microscopes as a means to assess stained tissue samples. WSI systems provide many potential benefits such as image analysis, remote diagnosis, education and easy case review. There is, however, a need for a framework that will allow regulators and users of these systems to check their colour consistency.

The figure below shows a generic schematic for a digital microscope system. A microscope slide that has been stained and prepared for viewing is illuminated by a light source and focused by means of imaging optics on RGB sensors. The image captured by these sensors is then processed in some way to create an RGB image of the slide. Viewer software presents the RGB image of the slide on a display for viewing. The way in which the image is presented on the display along with the viewing environment determines the colour seen by the image observer (usually a pathologist).



The spectral responses of the RGB Sensors of digital microscopes do not in general match those of a typical human observer, such as that defined by the CIE standard colorimetric observer. Nor do the responses of different digital microscopes match each other. In addition the processing, viewer software, display response and viewing environment introduce further variables. Some form of colour management is required to reduce these errors so that the set of colours that are important to a pathologist are displayed consistently on the display.

There is therefore a need for an independent colour assessment method that uses a standard microscope slide and that will test that the set of colours on this slide are displayed correctly. This white paper makes recommendations for a reference microscope slide and defines a method for assessing images of the reference slide when presented on a display.

For some systems, the RGB image of the slide can be assessed independently of any display and so additional guidelines are provided for the assessment of these image files to allow the image capture module to be asses independently of other system components. Similarly, guidance is provided for assessment of the image display

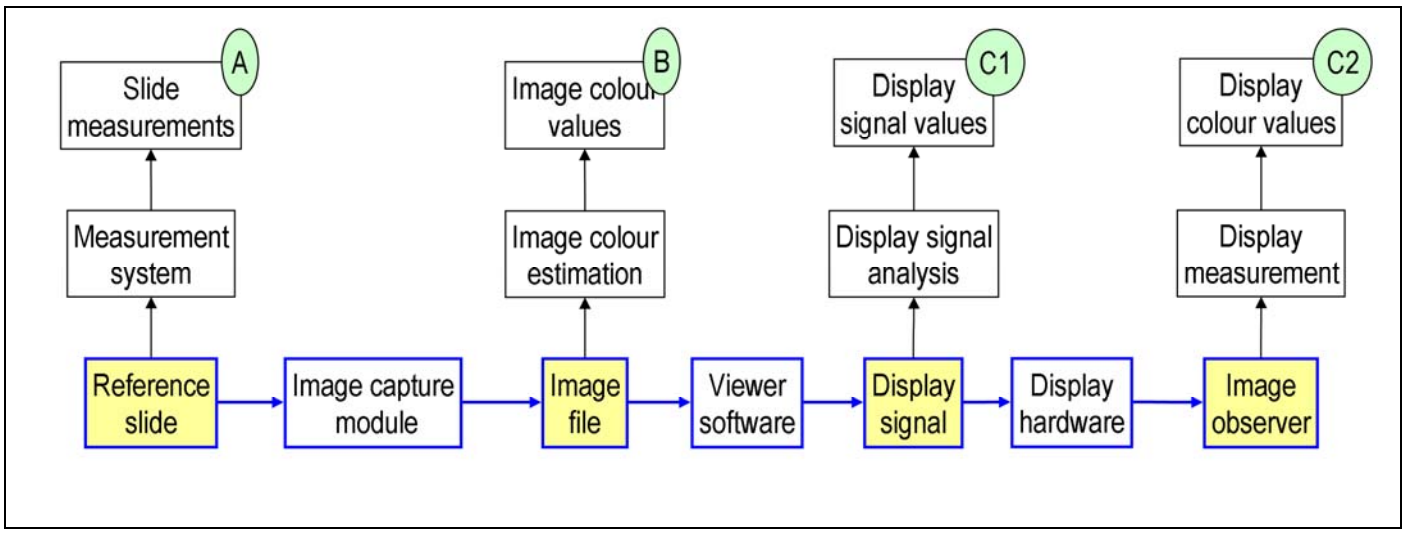
module comprising the viewer software, display hardware and viewing environment. In this way, these two major system components can be assessed independently.

Although this white paper is limited to colour evaluation we anticipate the need for the evaluation of other parameters such as dynamic range, resolution, focus and depth of field.

2.1 System overview

A simplified flow diagram for the test methods is shown below.

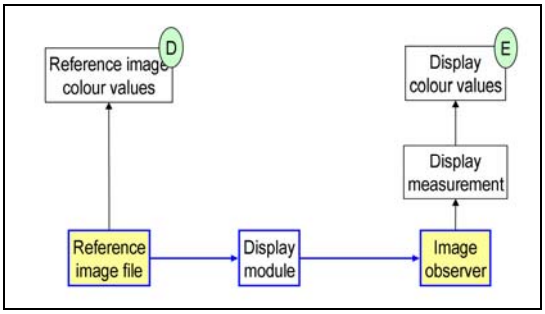
A reference slide is scanned by a digital microscope whose image capture module prepares an image file which is stored on the system. This image is prepared for on-screen viewing by a display module and the displayed image is viewed by an observer or measured by a measurement system.



The details of the image capture module vary from system to system but generally include imaging optics, RGB image sensors and image processing software. Similarly the display module includes viewer software which includes an image processing component, device driver software, display hardware and firmware and the viewing environment. The test methods have been designed to test the image capture module and display module without requiring any knowledge of the specific details of how these systems operate.

This white paper provides recommendations for the construction of a reference slide, a method for obtaining (A) spectral measurements of the reference slide, (B) a method for estimating colorimetric values of the slide image and (C1) a method for estimating intended colour by monitoring display drive values and (C2) a method for measuring colorimetric values as perceived by an image observer.

In some cases it is useful to be able to assess the display module independently of the image capture module, for example where third party viewer software is used, and this is shown schematically in the figure below. This white paper provides recommendations for a reference image file, (D) associated reference image colour values and (E) defines a method for measuring the display colour values [do we need E1 and E2 steps similar to C1 and C2?].



Methods for assessing colour accuracy for an end-to-end test (A compared to C2), for an image capture module test (A compared to B) and for the display module test (D compared to E) are described.

3 Measurement methods

The measurement of microscope slides and of displays is a complex topic which is out of scope of this white paper but best practice methods (where they exist) should be adopted when making measurements. In particular, the measurement geometry shall be such that the measurement data represents colour as viewed by a typical microscope. It is also important to specify measurements sufficiently well so that a third party could obtain a set of measurements that are close to the reference set.

Measurements of the reference slide shall be provided as spectral absorbance or spectral transmittance and shall cover the range 380-720nm and samples shall be spaced by no more than 10nm. Measurement uncertainty shall be provided.

The region of the slide over which the measurement was made shall be clearly specified and where measurements have been averaged over an area the method of averaging shall be indicated.

NOTE 1 colour measurements are highly dependent on the measurement geometry – specifically the light path through the material being measured. The difference between different measurement geometries can be as much as 20% in optical density.

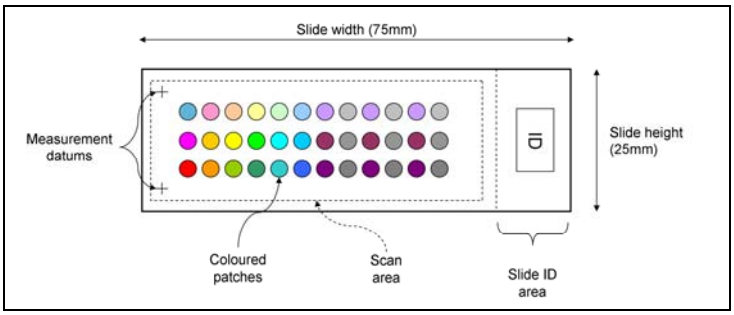
NOTE 2 providing only tristimulus values would limit the usefulness of the phantom and if spectral absorbance (or transmittance) is provided additional data about the phantom colour under different light sources is not necessary.

[TBD: Wei-Chung Cheng to provide details of his measurement method which will be included as an example.]

4 Test materials

4.1 Reference slide (colour)

The basic idea of a reference slide to be used to test colour is shown in the figure below. The slide comprises a number of coloured patches with known (measured) colours along with measurement datums and slide identification information. The content, size, shape and location of coloured patches, the shape and location of the measurement datums and details of slide identification may vary significantly from what has been shown here.



For a reference slide to be useful when testing the colour reproduction of a microscope system a number of aspects need to be considered:

- the set of colours on the slide should be representative of the colours encountered in day-to-day work,
- colour of patches should be measurable and it should be possible to repeat these measurements,
- colour of patches should be stable over time,
- the slide should be clearly identified,
- it should be possible to scan the slide easily on digital microscope systems.

It would be impossible to include and test every possible colour that can appear on stained slides and so a subset must be selected. This subset should include colours that cover the gamut of colours, should include intermediate tones and a set of neutral tones.

4.1.1 The set of colours

It is usual to stain slides using a small number of stains which means that for any given staining protocol the range of colours expressed on an individual slide is constrained to a small region of colour space. The most commonly used staining protocol uses Haematoxylin and Eosin stains (H&E) and the typical set of colours for an H&E stained sample is shown below.

[TBD: add plot of H&E slide colours]

Decision needed:

Option A: It would be possible to define a reference slide for each staining protocol, for example an 'H&E slide' would include only those colours that arise in H&E stained samples. This approach would mean that such a system can be qualified for a subset of staining protocols. This option would add complexity to the workflow for slide scanning and communication.

Option B: An alternative would be to qualify systems for a range of staining protocols. This option is potentially less precise.

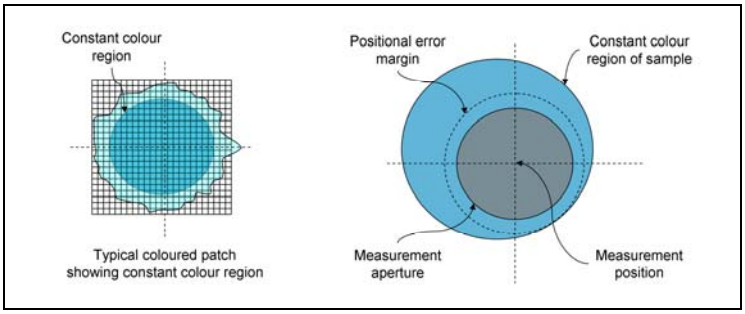
A set of candidate stains and staining protocols is provided in Annex A.

Substantiation required:

There is some evidence that, for most stains, the colours arising in stained tissue depend only on the amount of stain present and that stains combine in a simple linear way (i.e. there is no chemical interaction). The colour for most staining protocols depends only on the ratio of the two dyes present. This ratio changes depending on the tissue type.

4.1.2 Measurability

It is important to be able to make measurements of coloured patches in a way that can be repeated. To achieve this, patches should be uniform in colour when viewed at the imaging resolution to be qualified. The measurement centre, measurement aperture dimension and the required positional tolerance should be clearly specified.



The objective magnification must be included when discussing the physical size of the colour patches. Common magnification for digital microscopes is X20 and X40 and both of these magnifications should be considered when designing a suitable target.

The measurement location (in XY coordinate) should be included as part of the measurement data if the colour patches are not perfectly uniform and any averaging method defined.

[TBD: There may be significant variation in the ability of scanners to pinpoint locations over the entire imaging area based on a single datum. This can be solved by pattern recognition algorithms if the texture around the spot is unique or additional datums are added to the slide. A simple cross correlation coefficient method can be used to identify some "markers" in tissue samples on the glass slide, and then use these markers to construct the XY coordinate system.

We anticipate that a commercialized colour phantom will come with a long file describing the spectral reflectance of every patch. Technically it is not difficult to add two more columns to record the XY positions along with a description of the datums and any averaging method.]

4.1.3 Stability

Many coloured materials change in colour when exposed to light and the materials used to make the calibration reference slide coloured patches should be selected to ensure that the colours are stable over a suitable period (how long). The conditions in which the reference slide is stored may affect its stability.

The required storage conditions along with the period over which the slide is stable should be clearly identified on the reference slide.

4.1.4 Slide identification

Each calibration reference slide should have a unique identification code to allow it to be referenced by any assessment. This is particularly important in cases where the measurement data is stored separately from the reference slide.

4.1.5 Scanning requirements

It should be easy to scan the reference slide using the system configuration to be used for diagnostic work. The automated scanning of uniform colour areas presents some difficulties for digital microscopes as in many cases these systems need some kind of variation in order to focus.

[TBD: it may be necessary to add some marks to the slide to ensure that suitable focus can be achieved.]

4.2 Reference image file

[TBD: this image file will be equivalent to a scan of an idealised reference slide on a reference digital microscope and will include an ICC profile that specifies image colours. There are many details to be considered when defining this ideal image but the most significant factors are the illumination source and sensor sensitivities of the reference digital microscope. Candidates for these are CIE Illuminant E (spectrally flat) and the CIE colour matching functions (x_bar (λ), y_bar (λ), z_bar (λ)) which represent a standard observer (or perhaps more realistically a simple transform of these). Together along with spectral absorbance measurements of a reference slide these can be used to produce an idealised image of the slide.]

5 Display and viewing conditions

Many of the details of display requirements are outside of the scope of this white paper but care should be taken to ensure that the display white point, colour gamut, dynamic range, uniformity and viewing angle are suitable for its use in diagnostic pathology.

[TBD: we should refer to other best practice documents for display measurement and perhaps to provide recommendations for these criteria.]

5.1 Presentation of slide images

[TBD: this section will specify how to present images of the reference slide on screen ensuring that measurements of the display correspond with measurements of the reference slide.]

5.2 Viewing conditions

It is anticipated that displays will be used in a variety of different viewing conditions and that the image will be appropriately adjusted to ensure consistent appearance across viewing environments. Steps should be taken to control the viewing environment to ensure that the ambient lighting remains well controlled and that there are no highly coloured elements visible (directly or reflected on screen).

[TBD: more details of minimum and maximum recommended illumination levels, ambient illumination colour and anticipated display compensation, perhaps based on DICOM GSDF.]

[TBD: what kind of colour match is required? This is perhaps the biggest point for discussion.

It seems clear from discussions with pathologists they can work effectively in a wide variety of viewing conditions and that the important aspect seems to be to ensure that the 'colour context' is preserved. I think it would be impossible to agree what is meant by a perceptual match as this is inevitably to some extent subjective but we could probably agree requirements for a relative colorimetric match and this may be the simplest approach. If we were to adopt this approach the viewing environment would be included by making measurements with a telespectroradiometer (for Product Qualification tests) and would be modelled by measuring the ambient using a contact spectrophotometer with an ability to measure ambient.

So perhaps achieving a Relative colorimetric match is a reasonable objective.]

6 Test methods

[TBD: is the following set of tests appropriate?]

6.1 Capture-to-display test using telespectroradiometer

[TBD: describe how to conduct A-C test using teleradiometer. Scan slide and present image on screen and measure image. Will need some way to measure the average colour of a region presented on screen, also need to consider magnification.

One alternative to measuring the average colour of a certain region, which may be an odd shape would be to first profile the display (i.e., to build a display model, or obtain the mapping from DVI to CIEXYZ) and then predict the CIEXYZ of every pixel in this region. An FPGA could be used to retrieve the DVI signals.

This may be more complex overall and may become quite complex as it would seem to require a profile of every point of the display to provide a reasonable measure of its characteristics and is only practical when the display is sRGB.]

6.2 Capture-to-display test using contact spectrophotometer

[TBD: describe how to conduct A-C1 and A-C2 tests using contact instrument with radiant mode to perform a simple check of the viewing environment. Details will be similar to the above.]

6.3 Capture module test

[TBD: describe how to conduct A-B test. This is basically a matter of scanning the reference slide and performing an assessment of the image created (including the associated ICC profile). This should include some description of how to calculate the average and standard deviation of colour of a region of an image and on how to compare two sets of colours.]

[If the ICC profile is included in the A-B test, what colour manager do we use? The factory software or our own?

There may be scope for a new for a new product in this area to automate the process but the choice of colour management module is unlikely to introduce significant variation compared to other variables although with 8-bit data precision, different CMMs always obtain different results due to the errors accumulated in the calculation process. This is worse for display devices because speed precedes accuracy so the differences are often in the range of several delta-E. Such colour errors seem minor when comparing with the other variables in a WSI system.

We do need to consider the effect on the dynamic range reduction and colour quantisation introduced by such a process. Concern about this issue has been expressed by some pathologists who consider this to be a more serious problem than the colour accuracy issue.]

6.4 Display module test

[TBD: describe how to conduct D-E test. Will be similar to the capture-to-display testing but will use the reference image rather than a scanned image.]

6.4.1 Test for displayed digital values

[Shall we add another checkpoint where the digital RGB values for examining the effects of the image viewer, colour manager, and the display system hardware (graphics card) and software (display driver)? An FPGA board could be used to monitor the DVI (and presumably equivalent) signals.

The FPGA approach helps the calibration check process in the following ways:

- *Capture check: By isolating the display, the capture stage (A-B) can be examined independently. The readout of the FPGA (DVI in sRGB) should be compared against the measurement (the calculated truth) of the color phantom. If a WSI system passes such a test (even when a telespectroradiometer is not available), then at least its image files are sound.*
- *Display calibration: The FPGA can directly inject pixel data (in sRGB) into the display for testing or profiling. The display can evaluated quickly without using the WSI scanner hardware or software. In comparison, a color calibration kit relies on system software to control the display, and the output display signals may or may not be tampered by the display driver or graphics card.*
- *Cost: The FPGA hardware (by putting together off-the-shelf components) costs less than $350, which is cheaper than a decent telespectroradiometer.*
- *Speed: The FPGA measures pixel values in milliseconds. The telespectroradiometer takes, say, 5 seconds to obtain one reading.*

- *Accuracy: The FPGA measures digital data with 100% accuracy. In contrary, optical measurement is sensitive to the instrument calibration, measuring geometry, ambient light and temperature, etc.*

This model will be fine for sRGB but will be difficult to use in the case where the display is not an sRGB display (or is being used in sRGB mode). I think that sRGB would be OK as a choice for this and should be supported by the framework as an option but display technology has moved on since the sRGB standard was developed and it may be that we should also support forward-looking models such as that of mRGB where the full gamut of displays can be used to display images.

Should we consider this just another way to check the image data for the case where an sRGB workflow is used? We could add the FPGA / DVI check as an alternative to the image check in our model. These are not exactly equivalent for non sRGB models as the image presented to the video interface will need to have colour management applied.

Since some of the stain colours in use today are outside of the sRGB gamut adding support for wide gamut displays would seem to provide a flexible model but this is one of the first decisions that we should make.

How would we deal with the case where the display includes a lookup table?]

7 Test scenarios

[TBD: this version includes four test scenarios. Is this too many? Do we need to bother with test scenarios at all or should this be the responsibility of some other group? My expectation is that the measurement equipment and level of scrutiny (tolerances) required for each scenario will be different.]

7.1 Product qualification

[TBD: this is expected to be a set of tests conducted to qualify a product range and the set of tests described will demonstrate that the product range is capable of meeting the system requirements for colour. Tests will be conducted on a 'representative' product from the range. We should define what constitutes a product range and perhaps identify colour-critical system components.]

7.2 Installation qualification

[TBD: this is expected to be a test or set of tests conducted when each system is installed. Since product qualification has already been done the set of testing for this step should demonstrate that the system has been configured appropriately, should include measurement of the viewing environment and a method for checking the system setup and for controlling the viewing environment should be clearly identified.]

7.3 Periodic qualification

[TBD: this is expected to be a simple test that can be conducted by the pathologist or technician. Testing will demonstrate (in a way that can be recorded) that the system remains within calibration limits and that the viewing environment is appropriate.]

7.4 Visual assessment

[TBD: this is test intended for simple visual assessment of the system and will provide feedback to the pathologist of the colour performance of the system. This may not be suitable for primary diagnosis.]

8 System requirements

8.1 Identification of operating mode

In some cases digital microscope systems process slide images to enhance colour in some way, for example to increase colour discrimination. In this mode of operation, colours will no longer be calibrated. Digital microscope manufacturers should ensure that users can see clearly whether calibrated colour or enhanced colour is being presented.

9 Annex A: stains and staining protocols

[TBD: The following is a list of stains and staining protocols we may wish to include. The list is ordered in approximate frequency of usage with the most commonly used staining protocols listed first. The list has been compiled with help from Leeds Teaching Hospitals NHS Trust and the stains mentioned are the most widely used for histological diagnosis and are the stains often requested by UKNEQAS for assessment. Input from other regions is requested. Many of these staining protocols produce spectrally similar colours and if would be a good idea to identify a subset of the most commonly used staining protocols that covers the range of spectral colours on microscope slides.]

9.1 Haematoxylin and Eosin (H&E)

Haematoxylin and eosin staining protocol is used frequently in histology to examine thin sections of tissue. Haematoxylin stains cell nuclei blue, while eosin stains cytoplasm, connective tissue and other extracellular substances pink or red. Eosin is strongly absorbed by red blood cells, colouring them bright red. In a skilfully made H & E preparation the red blood cells are almost orange, and collagen and cytoplasm (especially muscle) acquire different shades of pink. Hematoxylin stains the cell nucleus and other acidic structures (such as RNA-rich portions of the cytoplasm and the matrix of hyaline cartilage) blue. In contrast, eosin stains the cytoplasm and collagen pink.

9.2 Diaminobenzidine (DAB) with Haematoxylin counter stain

Diaminobenzidine is oxidized by hydrogen peroxide in the presence of horseradish peroxidase to produce a dark-brown colour at the antigenic site. This is commonly used with a Haematoxylin counter stain.

9.3 Papanicolaou (PAP)

Papanicolaou staining, or Pap staining, is a frequently used method for examining cell samples from various bodily secretions. It is frequently used to stain cervical smears It uses a combination of haematoxylin, Orange G, eosin Y, Light Green SF yellowish, and sometimes Bismarck Brown Y.

9.4 Perls' Prussian blue

Perls' Prussian blue uses solution of potassium ferrocyanide and acid to stain tissue. Ferric iron deposits in tissue (present mostly as ferric iron within the storage protein ferritin) then react with the soluble ferrocyanide in the stain, to form insoluble Prussian blue dye in situ. They are then visualizable microscopically as blue deposits, within cells.

9.5 Periodic acid-Schiff (PAS)

Periodic acid-Schiff staining is used to mark carbohydrates (glycogen, glycoprotein, proteoglycans). It is used to distinguish different types of glycogen storage diseases.

9.6 Reticulin

In pathology, the reticulin stain is a popular staining method in histology. It is used to visualize reticular fiber and used extensively in liver histopathology.

9.7 Millers elastic Van Gieson

This method is used for identifying elastic fibers in tissues such as skin, aorta, etc. on formalin-fixed, paraffin-embedded sections, and may be used for frozen sections as well. The elastic fibers will be stained blue-black collagen fibres red, red blood cells and muscle will be stained yellow.

9.8 Shikata

Used to demonstrate both copper and Australia antigen (hepatitis B surface antigen). The method relies on permanganate oxidizing sulphur-containing proteins to sulphonate residues that react with orcein.

9.9 Giemsa stain

A Romanowsky type stain consisting of polychrome methylene blue and eosin. A very useful stain in histopathology for the demonstration of haemopoetic cells in bone marrow, protozoa and spirochaetes.

9.10 Ziehl Neelsen technique

Tubercle bacilli have a lipid rich cell wall which is capable of taking up carbol fuchsin (basic fuchsin in a phenolic solution) in such a way that they retain the dye upon subsequent differentiation in acid or alcohol i.e. they are acid and alcohol fast. Most other organisms lose the dye and take up the counterstain (methylene blue).

9.11 Grocott

Method of choice for demonstrating all fungi. The rationale is that chromic acid formed aldehydes from certain structures reduce a hexamine-silver mixture at an alkaline pH, and are thus selectively blackened.

9.12 Alcian blue PAS

Alcian blue used at pH 2.5 stains all acidic mucins. This is followed by the PAS technique which selectively stains neutral mucins. Thus acidic mucins are stained blue, neutral mucins are stained red and mixed mucins are stained with both dyes to give a purple colour.

9.13 Jones methenamine silver

Periodate formed aldehydes from certain carbohydrate containing material (in this case basement membranes) will selectively reduce an alkaline hexamine (methenamine) silver salt mixture to give a visible silver deposit on the basement membrane.

9.14 Gram

Blue cationic dyes (in this case crystal violet) are used to stain the nucleic acids of the organisms (and background tissue), following this the sections are treated with iodine and then differentiated in a suitable dye solvent such as acetone or alcohol. The tissue background and certain types of organisms lose their blue staining but take up the cationic dye of contrasting colour (Neutral red) subsequently applied. The blue stained organisms are termed "Gram positive" and those stained with the counterstain "Gram negative".

9.15 Congo red stain for amyloid

Congo red is an acidic diazo dye which binds to substances such as amyloid by hydrogen bonding. An important feature of the Congo red staining of amyloid is the red to green (dichroic) birefringence which is given due to the parallel alignment of the dye molecules on the amyloid fibrils.

9.16 Masson trichrome

Trichrome stains were originally developed to distinguish muscle from collagen. The method relies on the on the degree of differentiation of acid fuchsin by phosphomolybdic acid. It is important to control the differentiation until the connective tissue is unstained. The connective tissue is then stained with methyl blue or light green.