Title

A response to: Campbell WS, Talmon GA, Foster KW, Lele SM, Kozel JA, West WW. Sixty-five thousand shades of gray: importance of color in surgical pathology diagnoses. Human Pathol 2015; 6:1945-50

Authors

Emily L Clarke[1] MBBS PG Cert(HR)

Claudia Mello-Thoms[3] MS PhD

Derek Magee[2] BSc PhD

Darren Treanor[1,2] MB BSc(Computing) PhD FRCPath

Institutions

1. Department of Histopathology, Leeds Teaching Hospitals NHS Trust, Leeds, UK
2. University of Leeds, Leeds, UK
3. Faculty of Health Sciences, University of Sydney, Australia

Corresponding Author

Dr Emily Clarke

Department of Histopathology

St James University Hospital

Beckett Street

Leeds

West Yorkshire,

UK

LS9 7TF

E-mail: e.l.clarke@leeds.ac.uk

Tel: +44(0)113 3438509

Keywords

Color calibration; digital pathology; WSI; grayscale; diagnostic accuracy

Main Text

Dear Sirs,

We congratulate Campbell et al for their interesting and thought-provoking paper 'Sixty-five thousand shades of gray: importance of color in surgical pathology diagnoses'[1]. We agree that there is insufficient research into this important area.

The need for further work alongside the necessity for color standardisation in digital pathology is highlighted in guidance by the US Food and Drug Administration[2]. Our group has conducted work in this area including; measurement of staining variation, inaccuracies in scanner color reproduction and the development of a test object for color calibration of digital pathology[3]. Our pilot research evaluating the effectiveness of color calibration in the clinical setting[4] indicates that end-to-end color calibration increases pathologist's diagnostic confidence as compared to monitor calibration alone. Therefore, we have a few thoughts regarding this study and some suggestions for future work.

We were surprised by the closing statement that *'color calibration does not appear to impact WSI diagnostic accuracy'*. The study did not assess the effect of color calibration, rather it compared diagnostic concordance between grayscale WSIs and the light microscope. So asserting that the absence of color calibration (i.e. not controlling color in a digital imaging system) does not impact diagnostic accuracy is unfounded. It is also at odds with guidance from the US Food and Drug Administration[2].

We support the author's statement that it is incorrect to conclude that color is unimportant in diagnosis, for many reasons. Firstly, the study is not appropriately powered to answer that question. Even when minimising the potential effect of color by assuming only 1% drop in concordance when using grayscale images (from 100% to 99%), 1556 cases would be needed for 80% power and a significance level of 0.05[5]. Secondly, as the authors acknowledge, restricting cases to breast biopsy specimens reduces the generalizability. A

large degree of low-magnification morphological assessment is used when diagnosing breast specimens and therefore color may be less important in this subspeciality. Additionally, the degree of diagnostic difficulty within the included cases is not clear; more difficult diagnoses (e.g. subtle neoplasia) might depend on color for accurate assessment. This notion of the importance of color varying between cases is supported by a study by Rubin et al 2009[6], who concluded that the usefulness of grayscale images as compared to color images varied depending on the case and were, at times, a hindrance despite participants having reduced color perception. Perhaps future experiments could involve cases representing a spectrum of diagnostic difficulty as well as a range of different tissue types and stains. Finally, we note that other important parameters concerning diagnosis (e.g. user fatigue and diagnostic speed) were not evaluated in this study, yet could be affected by color differences. We therefore suggest that this should be considered as an outcome in the future.

It is well-known that much of the information in a digital image is preserved when grayscale images are created, depending upon the colorspace (RGB or CIELAB) and the type of color-to-grayscale conversion used[7]. For example, at low levels the L (luminance) channel relates in a linear fashion (relating to concentration) with haematoxylin. Therefore, if the grayscale conversion was based upon low L, then much of the diagnostic information would be preserved (i.e. nuclear detail). We were therefore unsurprised that broadly similar diagnoses could be reached from grayscale images. Unfortunately, specific details regarding the conversion algorithm used was not provided in the paper, beyond the use of the software application NegativeScreen (ArcaneSanctum.Net, France).

Overall, this study is a welcome addition to a field with sparse evidence but a lot of debate. A common argument against the importance of color is the presence of color-blind pathologists. Whilst this seems a logical argument, a study by Poole et al 1997[8] indicated that color blind pathologists had a lower mean score (94% vs. 99%) as compared to their

colleagues with normal color vision when trying to identify pathological features. This finding is supported – amusingly - by a recent paper demonstrating that pigeons (Columba livia) had a reduced accuracy for detecting breast cancer in using monochrome images as compared to full color[9].

Although this study provides some interesting insights, we have concerns that the conclusions may be over-simplified and generalized by others to imply that color does not matter in whole slide imaging. It is likely that there is varying need for color depending on the specifics of the case, with grayscale and color providing differing and complementary functions – more research is needed to investigate this in greater detail.

References

[1] Campbell WS, Talmon GA, Foster KW, Lele SM, Kozel JA, West WW. Sixty-five thousand shades of gray: importance of color in surgical pathology diagnoses. Human Pathol 2015; 6:1945-50

[2] Technical Performance Assessment of Digital Pathology Whole Slide Imaging Devices: Draft Guidance for Industry and Food and Drug Administration Staff. Food and Drug Administration, http://www.fda.gov/ucm/groups/fdagov-public/@fdagov-meddev-gen/documents/document/ucm435355.pdf [accessed 16.10.15]

[3] Revie WC, Clarke EL, Shires M, Jackson P, Brettle D, Cochrane R, Treanor D. Color calibration for digital pathology: Development of a novel color calibration assessment target for digital microscopy (submitted)

[4] Clarke EL, Revie CW, Brettle D, Wilson R, Mello-Thoms C, Treanor D. Color calibration in Digital Pathology: The Clinical Impact of a Novel Test Object. (submitted)

[5] Sealed Envelope Power (Sample Size) Calculator, https://www.sealedenvelope.com/power/binary-superiority/ [accessed 22.3.16]

[6] Rubin LR, Lackey WL, Kennedy FA, Stephenson RB. Using Color and Grayscale Images to Teach Histology to Color-deficient Medical Students. Anat Sci Ed 2009; 2:84–8.

[7] Kanan C, Cottrell G. Color-to-Grayscale: Does the Method Matter in Image Recognition? Plos One 2012;7:1. DOI: 10.1371/journal.pone.0029740

[8] Poole CJM, Hill DJ, Christie JL, Birch J. Deficient colour vision and interpretation of histopathology slides: Cross sectional study. BMJ 1997; 315:1279– 1281

[9] Levenson RM, Krupinski EA, Navarro VM, Wasserman EA. Pigeons (Columba livia) as Trainable Observers of Pathology and Radiology Breast Cancer Images. PLoS ONE 2015; 10(11). doi:10.1371/journal.pone.0141357