Comparison of different anti-Ki67 antibody clones and hot-spot sizes for assessing proliferative index and grading in pancreatic neuroendocrine tumours using manual and image analysis

Roisin Owens,1,* Elaine Gilmore,2,* Victoria Bingham,2 Christopher Cardwell,3 Hilary McBride,1 Stephen McQuaid,2 Matthew Humphries2,* & Paul Kelly1,*
1Department of Cellular Pathology, Belfast Health and Social Care Trust, Belfast, 2Precision Medicine Centre of Excellence, Centre for Cancer Research and Cell Biology, Queen’s University Belfast, Belfast, and 3Centre for Public Health, Institute of Clinical Sciences, Queen’s University, Belfast, Northern Ireland, UK

Date of submission 21 April 2020
Accepted for publication 26 June 2020
Published online Article Accepted 2 July 2020

Owens R, Gilmore E, Bingham V, Cardwell C, McBride H, McQuaid S, Humphries M P & Kelly P. (2020) Histopathology 77, 646–658. https://doi.org/10.1111/his.14200

Comparison of different anti-Ki67 antibody clones and hot-spot sizes for assessing proliferative index and grading in pancreatic neuroendocrine tumours using manual and image analysis

Aims: Ki67 proliferative index (PI) is essential for grading gastroenteric and pancreatic neuroendocrine tumours (GEP NETs). Analytical and preanalytical variables can affect Ki67 PI. In contrast to counting methodology, until now little attention has focused on the question of clone equivalence and the effect of hot-spot size on Ki67 PI in GEP NETs. Using manual counting and image analysis, this study compared the Ki67 PI achieved using MM1, K2 and 30-9 to MIB1, a clone which has been validated for, and is referenced in, guidelines relating to assessment of Ki67 PI in GEP NETs.

Methods and results: Forty-two pancreatic NETs were each immunohistochemically stained for the anti-Ki67 clones MIB1, MM1, K2 and 30-9. Ki67 PI was calculated manually and by image analysis, the latter using three different hot-spot sizes. In manual comparisons using single hot-spot high-power fields, non-MIB1 clones overestimated Ki67 PI compared to MIB1, resulting in grading discordancess. Image analysis shows good agreement with manual Ki67 PI but a tendency to overestimate absolute Ki67 PI. Increasing the size of tumour hot-spot from 500 to 2000 cells resulted in a decrease in Ki67 PI.

Conclusion: Different anti-Ki67 clones do not produce equivalent PIs in GEP NETs, and clone selection may therefore affect patient care. Increasing the hot-spot size decreases the Ki67 PI. Greater standardisation in terms of antibody clone selection and hot-spot size is required for grading GEP NETs. Image analysis is an effective tool for assisting Ki67 assessment and allows easier standardisation of the size of the tumour hot-spot.

Keywords: Image Analysis, Ki67, Neuroendocrine Tumour, Pancreas, Proliferative Index, QuPath

Introduction

Grading of gastroenteric and pancreatic (GEP) neuroendocrine tumours (NETs), based on assessment of mitotic count and Ki67 proliferative index (PI), is essential for subclassifying NETs, determining their prognosis and informing therapeutic decisions.1–10

© 2020 The Authors. Histopathology published by John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
The Ki67 protein is associated with cellular proliferation and its expression is limited to active phases of the cell cycle; namely, G1, S, G2 and M phases. Ki67 protein can be detected using immunohistochemistry and the Ki67 PI is established by expressing, as a percentage, the ratio of positively labelled tumour nuclei to the total number of tumour nuclei in a selected area. Guidelines suggest that the Ki67 index for NETs should be assessed within areas of maximal immunolabelling, referred to as proliferative ‘hot-spots’, consisting of between 500 and 2000 tumour cells. These suggested hot-spot sizes were originally proposed on the basis of common sense rather than evidence-based data.

Analytical and pre-analytical issues exist with regard to Ki67 assessment. Multiple studies have sought to establish the best method for evaluating Ki67 PI and highlighted the need for higher quality standardised counting methods. As a result, visual approximation or ‘eyeballing’, which is prone to considerable interobserver variation, is discouraged in favour of more reproducible methods such as manual counts using printed digitally captured images or on a computer screen. Image analysis techniques, which overcome some of the intra- and interobserver variabilities in humans, have been shown to provide equivalent accuracy for Ki67 PI assessment compared to manual counting methods. Historically, the costs and technical requirements involved in establishing a digital infrastructure have been considered prohibitive to the adoption of digital pathology in routine practice. However, as implementation costs fall, digital deployments are becoming increasingly more attractive to clinical pathology laboratories.

Several pre-analytical variables may affect the assessment of Ki67 for which standardised practices do not exist. Ki67 immunohistochemistry produces variable staining intensity. In the past, different criteria have been applied in regard to the level of staining required to determine a positively labelled nucleus, resulting in inconsistencies in tumour grading. While some studies have only counted strong or dark brown nuclear staining as positive, others have accepted any degree of nuclear staining as positive. However, such variation in practice is important in view of the low threshold (≤3%) that separates grades 1 and 2 NETs, which can be influenced by very small differences in the number of cells determined to be positive. It is worth noting there is now a growing consensus among GEP NET experts that all nuclei that display positive staining (homogeneous or speckled) should be counted as positive in the setting of a high quality anti-Ki67 immunostain. This is logical, as the expression of Ki67 varies during the different phases of the cell cycle, beginning in the nucleolus before being redistributed to the rest of the cell. Differences in interlaboratory staining protocols can also result in different Ki67 PI results, even when the same antibody clone is used, ultimately leading to tumour grade shifts. Other pre-analytical factors, such as fixation, section thickness and the size of the proliferative hot-spot used, may also influence Ki67 PI.

There are numerous commercially available antibody clones which detect the Ki67 protein. Data available from an external quality assurance scheme in the United Kingdom (UK-NEQAS) revealed that more than 10 different clones were in use in registered clinical pathology laboratories in addition to a variety staining protocols. The MIB1 clone, which is specifically referenced in some international NET guidelines, and has been validated for NET grading, was the most commonly used clone (167 of 306 laboratories) in this scheme. The use of clones other than MIB1 may result from an assumption that all clones are equivalent, a lack of awareness of the specific reference to MIB1 in guidelines, or an unawareness that multiple clones exist at all. However, studies in breast cancer cases and in tonsillar tissue have indicated that different Ki67 antibody clones do not produce equivalent proliferative indices, and this may affect reproducibility and capacity to predict prognosis in breast cancer. Therefore, clone equivalence should not be assumed in GEP NET grading.

The primary aim of this study was to compare the absolute Ki67 PI and indicative tumour grade obtained using four different commercially available Ki67 antibodies under typical laboratory conditions applied to a series of well-differentiated (grades 1 and 2) pancreatic neuroendocrine tumours (PanNETs). In view of the increasing availability of whole slide imaging and image analysis software, we also sought to validate QuPath, an open source image analysis platform, for assessing Ki67 PI in PanNETs. Using digital image analysis it was also possible to provide a quantitative evaluation of the effect of hot-spot size (500, 1000 or 2000 cells) on Ki67 PI and tumour grade.

Materials and methods

Tissue selection

Forty-two well-differentiated grades 1 and 2 PanNETs surgically resected between 2011 and 2018 were
identified from the pathology archives of the Belfast Health and Social Care Trust, Northern Ireland. The most representative tumour blocks were identified and accessed through the Northern Ireland Biobank (NIB study number 18-0276); 5 × 4 μm sequential sections were cut from the selected formalin-fixed, paraffin-embedded (FFPE) blocks. Each selected case had a haematoxylin and eosin (H&E) section cut to review morphology and identify and annotate the tumour area. The remaining four sections were stained for four separate anti-Ki67 antibodies (Table 1). All the antibodies [MIB1-Dako (Dako, Stockport, UK), K2-Leica (Leica, Newcastle upon Tyne, UK), MM1-Leica, 30-9 Ventana (Ventana, Tucson, AZ, USA)] had already been validated and used previously in the laboratory.

MANUAL CLINICAL ASSESSMENTS

Each immunohistochemically (IHC)-stained section was reviewed on a Leica DM 4000B microscope at ×10 objective to identify and annotate the area of greatest Ki67 positivity, i.e. the proliferative hot-spot. A single high-power field, representing the maximal hot-spot within this area, was then identified at ×40 objective, photographed and subsequently printed on white paper using a colour printer. Ki67 positively and negatively stained tumour nuclei were separately marked on each printed image using two different colours, respectively, and the total tumour cell count and positively labelled tumour nuclei recorded. For the purpose of this study any Ki67 staining was considered positive, as outlined by Gentil Perret et al.29 Figure 1 shows specific examples of the expression of the four Ki67 clones in a single hot-spot within a single case in the cohort. Hot-spot identification, annotation and manual counting was performed by a pancreatic pathologist with expertise in neuroendocrine tumours (P.K.). The Ki67 PI was calculated as a percentage by dividing the total number of Ki67-positive tumour nuclei by the total number of tumour nuclei in the hot-spot multiplied by 100. The tumour grade suggested by the calculated Ki67 PI was then determined in line with the grading thresholds proposed in World Health Organisation (WHO) guidelines (<3% = grade 1; ≥3–20% = grade 2; >20% = grade 3).5,7 For this study the MIB1 clone was chosen as the gold standard antibody, as it is recognised as a validated anti-Ki67 clone in guidelines for grade assessment in NETs.1,2,4,13,29

STATISTICAL ANALYSIS

The mean cell count and mean Ki67 PI were compared between MIB1 and other anti-Ki67 clones using a paired-samples t-test. The proportion of samples where a non-MIB1 clone was higher than MIB1 was calculated, and the sign test was used to test if this was different from 50%. A Bland–Altman plot of the mean against the difference was generated for MIB1 and each anti-Ki67 clone. Pearson’s correlation coefficient was used to formally test for an association between the average and the difference. Limits of agreement were calculated for MIB1 and, where possible, each anti-Ki67 clone. Limits of agreement give range values which, for 95% of samples, capture the difference in Ki67 between MIB1 and each non-MIB1 clone. In addition, the proportion of values within 1% of MIB1 Ki67 PI with 95% confidence intervals (CIs) was determined. The grade concordance on Ki67 category (<3% or ≥3%) was determined by calculating the proportion agreement and kappa (chance corrected agreement) with 95% CIs. Statistical analysis was conducted in STATA version 14/GraphPad Prism version 5.03.

DIGITAL IMAGE ANALYSIS

Digital image analysis of IHC-stained Ki67 whole-slide images was performed using the open source image analysis programme QuPath version 0.1.34 The same hot-spot regions used for the manual Ki67 assessments were used for image analysis. Scanning of the

| Antibody | Dilution | Pretreatment conditions | Manufacturer | Staining platform |
|----------|----------|-------------------------|--------------|------------------|
| MIB1     | 1:75     | ER2, 20 min             | Dako         | Leica Bond III   |
| MM1      | RTU      | ER2, 20 min             | Leica        | Leica Bond III   |
| K2       | RTU      | ER2, 20 min             | Leica        | Leica Bond III   |
| 30-9     | RTU      | CC1, 32 min             | Ventana      | Ventana Ultra    |

RTU, ready-to-use.

© 2020 The Authors. Histopathology published by John Wiley & Sons Ltd, Histopathology, 77, 646–658.
IHC-stained slides was performed using an Aperio AT2 scanner at ×40 magnification. Scans were imported into QuPath and cell detection was performed within the pre-annotated areas. Areas of necrosis, tissue folds and artefacts were excluded from analysis, as previously described. Each hot-spot was classified into tumour and stromal compartments using a detection classifier based on training regions. Following quality control, the tumour compartment was analysed for Ki67 staining using the single feature classifier, modified for a 0.15 minimum mean nuclear diaminobenzidine (DAB) optical density, to mitigate background staining. Therefore, all tumour cells with a mean nuclear DAB greater than or equal to 0.15 were classed as positive.

Digital image analysis readily allows for the accurate counting of tumour cells within a selected area on a digital slide. Taking advantage of this, three objects were drawn within the annotated hot-spot containing 500, 1000 and 2000 tumour cells, ensuring capture of the most proliferative areas. Positive cell detection data for each hot-spot size among all four Ki67 clones in the 42 PanNET cases were extracted from QuPath for analysis.

Results

MANUALLY ASSESSED Ki67 PROLIFERATIVE INDEX FOR MIB1 VERSUS MM1, K2 AND 30-9

Total cell counts for each antibody clone

Compared with MIB1, the mean total cell counts were similar for MM1 ($P = 0.74$) and K2 ($P = 0.185$), but lower for 30-9 ($P = 0.026$); see Table 2. For one case, the total cell count was below 500 cells for all four clones. Although some guidelines suggest that hot-spots used for Ki67 assessments

Table 2. Mean cell count and range of cell counts for each antibody clone

| Antibody Clone | Mean total cell count | SD  | Range       |
|----------------|-----------------------|-----|-------------|
| MIB1           | 925                   | 252 | 419–1543    |
| MM1            | 931                   | 247 | 437–1424    |
| K2             | 952                   | 261 | 374–1493    |
| 30-9           | 874                   | 233 | 327–1446    |

SD, standard deviation.
should be at least 500 cells, this particular case was still included in the final analysis.

**MANUAL KI67 PROLIFERATIVE INDEX**

The individual Ki67 PIs obtained for each case for MM1, K2 and 30-9 versus the gold standard MIB antibody clone are illustrated graphically in Figure 2A: parts (i) MM1, (ii) K2 and (iii) 30-9. Table 3 shows the range, mean and median for Ki67 PI for each antibody clone throughout the cohort. Compared with MIB1, the mean Ki67 PI was similar for MM1 ($P = 0.56$), but significantly higher for K2 ($P < 0.0001$) and 30-9 ($P < 0.0001$). MM1 overestimated Ki67 PI for 45% (19 of 42; sign test $P = 0.76$) of samples; K2 and 30-9 both significantly overestimated Ki67 PI compared to MIB1 [90% (38 of 42; sign test $P < 0.001$) and 88% (37 of 42; sign test $P < 0.001$)] of samples, respectively.

| Ki67 PI | Range | Mean | Median |
|---------|-------|------|--------|
| MIB1    | 0.2–6.9 | 2.1  | 1.4    |
| MM1     | 0.2–5.3 | 2.2  | 1.2    |
| K2      | 0.1–11.2 | 3.5  | 2.5    |
| 30-9    | 0.2–13.0 | 3.2  | 2.2    |

There was no correlation in the difference between MIB1 and MM1 values and the average Ki67 value ($r = 0.06, P = 0.71$) (Bland–Altman plots, Figure 2B). Calculating limits of agreement revealed that for 95% of samples the difference between the MM1 and MIB1 values were within the range +1.5 to -1.5.

**Figure 2.** Comparison of manual scores of the Ki67 proliferative index (PI) for the three different clones against the gold standard MIB1 (blue line) are shown in A, parts (i) MM1 (ii) K2 (iii) 30-9. Cases have been arranged in order of increasing MIB1-assessed Ki67 PI. A Ki67 PI of 3%, the clinical cut-off for grading, is highlighted by the black horizontal line. Accompanying Bland–Altman plots (B) display the difference (of clone positivity minus MIB1 positivity) against the average (of clone and MIB1 positivity).
In contrast, there was a strong correlation in the difference between the MIB1 and K2 values and the average Ki67 PI (of MIB1 and K2) ($r = 0.77$, $P \leq 0.001$), indicating that as Ki67 PI increased, K2 overestimated Ki67 PI compared to MIB1 by increasing amounts. Therefore, standard limits of agreement could not be calculated. Similarly, there was a strong correlation in the difference between the MIB1-determined Ki67 PI and 30-9 PI values and the average Ki67 value ($r = 0.86$, $P \leq 0.001$), also indicating an overestimation of 30-9 Ki67 PI when compared to MIB1.

When assessing an agreement within 1% PI between clones, 78% of MM1 values (33 of 42, 95% CI = 63%, 90%), 45% of K2 values (19 of 42, 95% CI = 30%, 61%) and 62% of 30-9 values (26 of 42, 95% CI = 46%, 76%) were within 1% of the MIB1-determined PI for the same case.

**Assessment of grade concordance between MIB1 versus MM1, K2 and 30-9**

Table 4 shows grade concordance between MIB1 and MM1 (a), K2 (b) and 30-9 (c). Generally, in discordant cases, non-MIB1 anti-Ki67 clones showed a strong tendency to indicate grade 2 status compared to grade 1 by MIB1 (11 of 13 discordant cases).

**Digital image analysis assessment of Ki67 index**

The original cohort of 42 cases, each stained with four different anti-Ki67 antibody clones, generated a total cohort of 168 (42 x 4) comparisons between manually assessed Ki67 PI and that obtained by digital image analysis. An example of QuPath masks on two cases is depicted in Figure 3. Examples of high Ki67 [A (ii)] and low Ki67 [B (ii)] are shown. In Figure 3C the yellow rectangles outline the hot-spot regions of interest of approximately 500, 1000 and 2000 tumour cells used for Ki67 assessment.

**Comparison of manual versus digital assessment of Ki67 proliferative index**

Figure 4A illustrates the manually determined Ki67 PI (blue line) and the digitally determined Ki67 PI using image analysis (red) for digital counts of 500 cells [4A(i)], 1000 cells [4A(ii)] and 2000 cells [4A (iii)]. Supporting information, Figure S1 illustrates manual Ki67 PI versus digital Ki67 PI broken down by clone. The Bland–Altman plots (Figure 4B) show a correlation between the difference between the manual and image analysis values and the average Ki67 value for 500 cells ($r = 0.81$, $P \leq 0.001$), 1000 cells ($r = 0.74$, $P \leq 0.001$) and 2000 cells ($r = 0.54$, $P \leq 0.001$), indicating that as Ki67 PI increased, image analysis overestimated compared to the manual assessment by a greater amount, therefore standard limits of agreements were not calculated.

Throughout the cohort, image analysis resulted in higher Ki67 PIs in 93% of cases (156 of 168; sign test $P < 0.001$) when 500 cell hot-spots were used, 83% of cases when 1000 cell hot-spots were used (139 of 168; sign test $P < 0.001$) and in 75% of cases when 2000 cell hot-spots were used (126 of 168; sign test $P < 0.001$). Of the 500 cell hot-spot digital Ki67, 42% PIs were within 1% of the manually determined value (73 of 168, CI = 35%, 51%); 61% of the 1000 cell hot-spots digital Ki67 PIs were within 1% of the manual Ki67 PI (102 of 168, CI = 53%, 68%) compared to 89% for the 2000 cell hot-spot Ki67 PIs (149 of 168, CI = 83%, 93%).

**Table 4. Grade concordance between manually assessed MIB1, Ki67, PI and Ki67 PI determined by alternative anti-Ki67 clones, in terms of number of cases (upper tier) and level of agreement (lower tier)**

|            | MM1         | K2          | 30-9        |
|------------|-------------|-------------|-------------|
|            | G1 | G2 | G1 | G2 | G1 | G2 |
| MIB1        |    |    |    |    |    |    |
| G1          | 26 | 3  | 24 | 5  | <3 | 3  |
| G2          | 1  | 12 | 1  | 12 | 0  | 13 |
| Percentage agreement with MIB1 (95% CI)  | 90% (80%, 91%) | 86% (71%, 95%) | 93% (81%, 99%) |
| Kappa for agreement with MIB1 (95% CI) | 0.79 (0.49, 1.00) | 0.69 (0.40, 1.00) | 0.84 (0.54, 1.00) |

CI, confidence interval; PI, proliferation index.

© 2020 The Authors. Histopathology published by John Wiley & Sons Ltd, Histopathology, 77, 646–658.
The tendency of image analysis to overestimate Ki67 PI compared to manually assessed Ki67 PI was also reflected when the mean Ki67 PI for image analysis was compared to the mean manually determined Ki67 PI ($P < 0.0001$, all hot-spot sizes) (Table 5). In addition, the mean Ki67 PI determined by image analysis decreased as the size of the tumour hot-spot increased. Using image analysis, MM1 was found to have a lower mean Ki67 PI when assessed by image analysis compared to digitally assessed MIB1 for all three hot-spot sizes.

**GRADE CONCORDANCE BETWEEN MANUALLY ASSESSED KI67 AND DIGITAL IMAGE ANALYSIS**

The proportion of concordant cases and kappa between manual and image analysis grades for the three hot-spot sizes is shown in Table 6. In the discordant cases, image analysis was more likely to result in a grade 2 designation compared to manual assessments accounting for 96, 88 and 72% of discordances when 500, 1000 and 2000 tumour cell hot-spots were used, respectively. Grade concordance breakdown by clone is shown in Supporting information, Table S1, with highest overall concordance between manual and image analysis observed for MM1.

**Discussion**

Assessment of Ki67 PI is essential for grading, classifying and treating GEP NETs. The MIB1 anti-Ki67 clone has been validated for grading GEP NETs and its use is specifically referred to in NET guidelines. However, alternative anti-Ki67 clones are routinely in use in clinical pathology laboratories, yet studies in breast cancer and tonsillar tissue have shown that different anti-Ki67 clones do not produce equivalent Ki67 PIs. Furthermore, the recommended size of the proliferative hot-spot to
be used to evaluate Ki67 is subject to variation in guidelines, but the impact of hot-spot size on Ki67 PI has received limited attention.

This study shows that there is considerable variation in the absolute Ki67 PI values determined by the different antibody clones directed against Ki67. In our cohort, all non-MIB1 clones showed a tendency to overestimate Ki67 compared to MIB1 when PI was assessed manually. MM1 showed closest agreement and no significant difference with MIB1 overall. In contrast, there were significant differences between MIB1 and both 30-9 and K2, with increasing

Table 5. Mean Ki67 proliferation index (PI) for each antibody clone determined by image analysis

|       | Manual Mean | 500 Mean | SD | 500 SD | P manual versus 500 | 1000 Mean | SD | 1000 SD | P manual versus 1000 | 2000 Mean | SD | 2000 SD | P manual versus 2000 |
|-------|-------------|----------|----|--------|---------------------|-----------|----|---------|---------------------|-----------|----|---------|---------------------|
| MIB1  | 2.1         | 1.88     | 3.5 | 2.7     | P < 0.0001          | 2.9       | 2.4 | P < 0.0001 | 2.6                 | 2.0       | P = 0.0001 |
| MM1   | 2.2         | 1.9      | 3.4 | 2.9     | P < 0.0001          | 2.8       | 2.4 | P = 0.0002 | 2.4                 | 2.1       | P = 0.22   |
| K2    | 3.5         | 3.1      | 5.8 | 5.0     | P < 0.0001          | 5.2       | 4.7 | P < 0.0001 | 4.7                 | 4.2       | P < 0.0001 |
| 30-9  | 3.2         | 3.1      | 5.3 | 4.9     | P < 0.0001          | 4.4       | 4.1 | P < 0.0001 | 4.0                 | 3.7       | P = 0.0005 |
| All clones | 2.7         | 2.6      | 4.5 | 4.1     | P < 0.0001          | 3.9       | 3.7 | P < 0.0001 | 3.4                 | 3.3       | P < 0.0001 |

CI, confidence interval; SD, standard deviation.
disparity as the Ki67 PI increased. One issue noted was that the total cell count for the 30-9 clone was, on average, 50 cells less than that used for the other clones. It could be argued that the smaller total cell count contributed to a concentration effect on Ki67 PI for 30-9 compared to MIB1. However, when the same clone-to-clone comparisons were duplicated using image analysis, which ensured that equivalent total cell counts were applied for all assessments, the same trend was noted. This suggests that the differences between manually assessed MIB1 and 30-9 PIs were independent of different mean tumour cell counts.

The grading threshold that exists between grades 1 and 2 is low (3%) and can be influenced by relatively small differences in the number of positively labelled cells, as described in a study of the impact of different interlaboratory staining practices using the MIB1 clone.20 Our data show that although grade concordance or agreement between MIB1 and non-MIB1 clones is moderate to substantial the grade was influenced by the Ki67 clone selected, as non-MIB1 clones produced more grade 2/1 than grade 1/2 discordances compared to MIB1 (11 versus 2). That the differences in absolute Ki67 values did not manifest with more grade discordances is almost certainly a reflection of the wide range of Ki67 PI values captured in the grade 2 category (3–20%). None of the cases used in our study had a Ki67 PI at the grade 2/3 (20%) boundary. However, it would be anticipated that the use of certain clones, such as 30-9 and K2, will lead to the designation of more PanNETs as grade 3. Prior to the most recent WHO NET grading system, such tumours would have been classified as grade 3 neuroendocrine carcinomas (NECs).

Absolute Ki67 PI is important to clinicians. Although current guidelines now widely adopt the 3% threshold for grade 1 versus 2, some studies have suggested that a 5% cut-off may be more clinically relevant in PanNETs.40-43 Ki67 PI, rather than grade, is increasingly used to inform treatment decisions in NETs. For example, a Ki67 PI cut-off of 10% has been used to select patients suitable for liver transplantation and is comparable with the inclusion criteria of the CLARINET study, and of some oncologists when selecting systemic therapy.40,44,45 ENETS guidelines for PanNETs refer to ‘low’ and ‘high’ grade 2 tumours, implying that treatments may be adapted according to grade 2 subdivisions.46 The response of GEP NETs to chemotherapy has been reported to improve with increasing Ki67 PI.47 With regard to higher-grade NENs, although NECs are no longer formally graded according to the current WHO classification scheme a Ki67 PI threshold of 55% appears to continue to have prognostic relevance.5,7,48-50 Based on our data, it is likely that throughout large patient cohorts the overall relationship between outcomes, such as response to treatment and Ki67 PI, would be observed regardless of the clone selected because, in general, an increase in Ki67 PI was mirrored across all clones. However, for the individual patient the situation is different, and the choice of antibody clone could have a significant impact on treatment.

In the past, use of image analysis for assessing Ki67 PI has been compared to manual counting methods and shown to be comparable, but cost has been seen as prohibitive for routine use.17,18,22 The technological improvements of the past decade have seen increased adoption of digital pathology in clinical practice, creating more potential for the use of image analysis and artificial intelligence. Previous work in our laboratory in breast cancer samples demonstrated high concordance between image analysis using QuPath and manual scoring of Ki67 (MM1

| Table 6. Grade concordance between manually assessed and image analysis using three different hot-spot sizes (500, 1000, 2000 cells), in terms of number of cases (upper tier) and levels of agreement (lower tier) |
|---------------------------------|---|---|---|---|---|
| MIB1                            | 500| 1000| 2000|
| G1 G2                          | 85 | 93  | 93  |
| Percentage agreement            | 90% (80%, 91%) | 90% (85%, 94%) | 89% (84%, 94%) |
| with manual (95% CI)            | 0.72 (0.58, 0.81) | 0.80 (0.65, 0.95) | 0.77 (0.62, 0.93) |
| Kappa for agreement             | 0.72 (0.58, 0.81) | 0.80 (0.65, 0.95) | 0.77 (0.62, 0.93) |
| with manual (95% CI)            | 0.72 (0.58, 0.81) | 0.80 (0.65, 0.95) | 0.77 (0.62, 0.93) |
| G1 G2                          | 1  | 2  | 5  |
| Percentage agreement            | 90% (80%, 91%) | 90% (85%, 94%) | 89% (84%, 94%) |
| with manual (95% CI)            | 0.72 (0.58, 0.81) | 0.80 (0.65, 0.95) | 0.77 (0.62, 0.93) |
| Kappa for agreement             | 0.72 (0.58, 0.81) | 0.80 (0.65, 0.95) | 0.77 (0.62, 0.93) |
| with manual (95% CI)            | 0.72 (0.58, 0.81) | 0.80 (0.65, 0.95) | 0.77 (0.62, 0.93) |

© 2020 The Authors. *Histopathology* published by John Wiley & Sons Ltd, *Histopathology*, 77, 646–658.
clone) and also suggested a dependence of Ki67 PI on the specific antibody used.\textsuperscript{31}

Our study of 42 cases and four Ki67 antibody clones facilitated the creation of 168 manual versus digital comparisons, one of the largest validation cohorts reported to date in PanNETs. We found that mean Ki67 PI generated using image analysis was higher than the mean of manually assessed Ki67 PI and demonstrated that tumour hot-spot size has an inverse effect on Ki67 PI, with a decrease in mean Ki67 as the hot-spot size increases. This inverse relationship between hot-spot size and Ki67 PI has also been described by Volynskaya et al.\textsuperscript{22} The highest correlations between image analysis and manual Ki67 were seen using a 2000 tumour cell hot-spot for image analysis Ki67 assessments. The average hot-spot size used for manual assessment was closer to 1000 cells, so the better correlation at 2000 cells is probably due to a dilution effect counteracting the apparent overestimation of Ki67 PI by image analysis. Despite the differences in absolute Ki67 PI, grading concordance between manual and digital assessments was substantial for all hot-spot sizes. The overestimation of Ki67 PI by image analysis manifested in more grade 2 versus 1 discordant cases between digital versus manual, stressing the importance of quality control of image analysis results by pathologists.

Using image analysis, we also replicated the manual MIB1 versus non-MIB1 clone comparison study. As described above, similar to the observations in the manual study, the K2 and 30-9 clones showed overestimation of Ki67 PI compared to MIB1. Given the unbiased nature of image analysis, this supports the original observation that K2 and 30-9 are more sensitive under the conditions used. In a review of selected cases, to ascertain the reasons for discrepancy between digital and manual assessments, it was noted that instances of misclassification of non-tumour cells as tumour cell was quite rare. During initial quality control of the image analysis software, it was noted that there was a tendency to ascribe positivity to tumour cells that were interpreted as negative on review of the digital image by a pathologist. The main reason for this phenomenon was often the presence of pre-analytical staining artefacts, such as increased section thickness, nuclear crowding or excess background staining. A loss of contrast was also noted when digital images were printed for manual counting, compared to the high-resolution digitally scanned images viewed on screen. This led to occasional under-recognition of faintly positive staining in print images compounding the differences in individual comparisons. Ultimately, this limited discrepancy review highlights the need for high technical quality of tissue sectioning and immunostaining, particularly when image analysis is being utilised. Pathologist supervision of image analysis software is therefore paramount if deployed in clinical practice.

This study did not attempt to determine if any single clone was superior in terms of clinical outcomes, as has been conducted in breast cancer,\textsuperscript{30} but we suggest that such studies are warranted. Furthermore, we used a different antibody detection platform for 30-9 compared to the other clones, which introduces another variable known to have potential to influence the determination of PI when a single clone is used.\textsuperscript{20} The MIB1 clone was the only clone in the study that was not ready-to-use (RTU). Unlike Blank et al.,\textsuperscript{20} we did not alter the dilution, retrieval and incubation schedules for MIB1 to assess the impact of these factors on Ki67 staining. Similarly, we did not assess the impact of altering the retrieval and incubation schedules for the non-MIB1 clones. The work by Blank et al. suggests that it may be possible to achieve greater Ki67 PI agreement between the various clones by altering their respective staining protocols; however, this was not considered feasible for this particular study. These points could be seen as criticisms and limitations of our study design, although it is important to stress that that study was conducted in such a way as to replicate typical and variable interlaboratory practices in clinical pathology departments. Therefore, all antibodies were optimised independently, and the most appropriate detection platforms used rather than a single staining platform for all antibodies. The time taken to perform manual Ki67 and image analysis PI assessments was not formally recorded in this study. Manual counts have been reported to take 6 min on average, although individual cases may take 55 min.\textsuperscript{17,22} In contrast, image analysis software can perform this task on selected hot-spots in seconds.\textsuperscript{22}

**Conclusion**

In a series of 42 PanNETs, we established that MIB1 and non-MIB1 anti Ki67 antibody clones do not always produce equivalent Ki67 PI results, which may impact upon patient care. MM1 correlated most closely with MIB1 in both manual and image analysis comparisons, suggesting that it may be the best MIB1 alternative in clinical practice. Nevertheless, validation of individual non-MIB1 clones may be required.
for grading of NETs, and further studies should assess if any particular clones are better at predicting the clinical outcomes of NETs in large patient cohorts. We also show that the size of the tumour cell hot-spot used to assess Ki67 has an inverse relationship on the PI, stressing the need for greater standardisation in guidelines and clinical practice. Our results also question the effectiveness of quality assurance schemes which do not incorporate a quantitative assessment of Ki67 antibody clone performance. Our findings support the continuing validation and use of image analysis as a rapid method assessing Ki67 PI in clinical practice, but stress the importance of high technical quality with regard to immunostaining and the need for overall pathologist quality assurance of results generated through automated platforms.

Author Contributions
Roisin Owens – antibody validations, case retrievals, preliminary data collection, manuscript preparation. Elaine Gilmore – QuPath validation and data acquisition. Victoria Bingham – data acquisition, analysis, manuscript preparation. Christopher Cardwell – statistical analysis, manuscript preparation. Hilary McBride – antibody validation. Stephen McQuaid – data acquisition, analysis, manuscript preparation, study design. Matthew P. Humphries – data acquisition, analysis, manuscript preparation, study design. Paul Kelly – study design, manual assessments, manuscript preparation.

Acknowledgements
Funding for this study was from training funds in the Department of Cellular Pathology, BHSCT. The samples used in this study were received from the Northern Ireland Biobank, which has ethical approval to use de-identified tissue samples from the Belfast Health and Social Care Tissue Pathology archive (REC:11/NI/0013. NIB15-0168). Informed consent was not required for this study.

Conflicts of interest
The authors declare no competing interests.

References
1. Rindi G, Klimstra DS, Abedi-Ardekani B et al. A common classification framework for neuroendocrine neoplasms: an International Agency for Research on Cancer (IARC) and World Health Organization (WHO) expert consensus proposal. Mod. Pathol. 2018; 31: 1770–1786.
2. Rindi G, Kloppe G, Alhman H et al. TNM staging of foregut (neuro)endocrine tumors: a consensus proposal including a grading system. Virchows Arch. 2006; 449: 395–401.
3. Inzani F, Petrone G, Rindi G. The new World Health Organization classification for pancreatic neuroendocrine neoplasia. Endocr. Relat. Cancer. 2018; 25: 463–470.
4. Kloppe G, Couvelard A, Perren A et al. ENETS consensus guidelines for the standards of care in neuroendocrine tumors: towards a standardized approach to the diagnosis of gastroenteropancreatic neuroendocrine tumors and their prognostic stratification. Neuroendocrinology 2009; 90: 162–166.
5. World Health Organization (WHO) Classification of Tumours of the Digestive System. 5th ed. Lyon: International Agency for Research on Cancer, 2019.
6. Bosman FT et al. Hruban RH, Theise ND eds. WHO classification of tumours of the digestive system. 4th ed. Lyon: International Agency for Research on Cancer, 2010.
7. Lloyd RV, Osamura RY, Kloppe G, Rosai J. WHO classification of tumours of endocrine organs. 4th ed. Lyon: International Agency for Cancer Research, 2017.
8. Vilar E, Salazar R, Perez-Garcia J, Cortes J, Oberg K, Tabernero J. Chemotherapy and role of the proliferation marker Ki-67 in digestive neuroendocrine tumors. Endocr. Relat. Cancer. 2007; 14: 221–232.
9. Oberg K, Akerstrom G, Rindi G, Jelic S. ESMO Guidelines Working Group. Neuroendocrine gastroenteropancreatic tumours: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann. Oncol. 2010; 21(Suppl. 5): v223–v237.
10. Boninsegni L, Panzuto F, Partelli S et al. Malignant pancreatic neuroendocrine tumour: lymph node ratio and Ki67 are predictors of recurrence after curative resections. Eur. J. Cancer 2012; 48: 1608–1615.
11. Gerdes J, Lenke H, Baisch H, Wacker HH, Schwaab U, Stein H. Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. J. Immunol. 1984; 133: 1710–1715.
12. Scholzen T, Gerdes J. The Ki-67 protein: from the known and the unknown. J. Cell. Physiol. 2000; 182; 311–322.
13. Luong TV, Watkins J, Chakrabarty B, Wang LM. Dataset for histopathological reporting of neuroendocrine neoplasms of the gastroenteropancreatic tract. London: Royal College of Pathologists, 2019.
14. Stephenson TJ, Cross SS, Chetty R. Dataset for neuroendocrine tumours of the gastrointestinal tract including pancreas. 3rd ed. London: Royal College of Pathologists, 2012.
15. Cavalcanti MS, Gonen M, Klimstra DS. The ENETS/WHO grading system for neuroendocrine neoplasms of the gastroenteropancreatic system: a review of the current state, limitations and proposals for modifications. Int. J. Endocr. Oncol. 2016; 3: 203–219.
16. Tang LH, Gonen M, Hedvat C, Modlin IM, Klimstra DS. Objective quantification of the Ki67 proliferative index in neuroendocrine tumors of the gastroenteropancreatic system: a comparison of digital image analysis with manual methods. Am. J. Surg. Pathol. 2012; 36: 1761–1770.
17. 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: 686–694.
Ki67 clones and hot-spot size impact assessment

18. Young HT, Carr NJ, Green B, Tilley C, Bhargava V, Pearce N. Accuracy of visual assessments of proliferation indices in gastroenteropancreatic neuroendocrine tumours. J. Clin. Pathol. 2013; 66: 700–704.

19. Kroneman TN, Voss JS, Lohse CM, Wu TT, Smyrk TC, Zhang L. Comparison of three Ki-67 index quantification methods and clinical significance in pancreatic neuroendocrine tumors. Endocr. Pathol. 2015; 26: 255–262.

20. Blank A, Weiheck L, Marioni I et al. Interlaboratory variability of MIB1 staining in well-differentiated pancreatic neuroendocrine tumors. Virchows Arch. 2015; 467: 543–550.

21. Remes SM, Tuominen VJ, Helin H, Isola J, Arola J. Grading of neuroendocrine tumors with Ki-67 requires high-quality assessment practices. Am. J. Surg. Pathol. 2012; 36: 1359–1363.

22. Volynskaia Z, Mote O, Pakbaz S, Al-Ghamdi D, Asa SL. Ki67 quantitative interpretation: insights using image analysis. J. Pathol. Inform. 2019; 10: 8.

23. Williams BJ, Bottoms D, Treanor D. Future-proofing pathology: the case for clinical adoption of digital pathology. J. Clin. Pathol. 2017; 70: 1010–1018.

24. Polley MY, Leung SC, Gao D et al. Comparison of 5 Ki-67 antibodies and their reproducibility of immunohistochemical assessment of the Ki-67 labelling index in a large multi-centre trial. J. Pathol. 2002; 198: 292–299.

25. Polley MY, Leung SC, Gao D et al. An international study to increase concordance in Ki67 scoring. Mod. Pathol. 2015; 28: 778–786.

26. Kim JY, Hong SM. Recent updates on neuroendocrine tumors from the gastrointestinal and pancreaticobiliary tracts. Arch. Pathol. Lab. Med. 2016; 140: 437–448.

27. Mengel M, von Wasielewski R, Wiese B, Rudiger T, Muller-Hermelink HK, Kreipe H. Inter-laboratory and inter-observer reproducibility of immunohistochemical assessment of the Ki-67 labelling index in a large multi-centre trial. J. Pathol. 2002; 198: 292–299.

28. NRQAS U. Run 117/46. 2017. Available at: https://ukneqasicc.org/wp-content/uploads/2019/07/run_117_journal_final.pdf.

29. Gent Perret A, Mosnier JF, Helin H, Isola J, Arola J. Grading of neuroendocrine tumors with Ki-67 requires high-quality assessment practices. Am. J. Surg. Pathol. 2012; 36: 1359–1363.

30. Acs B, Kulkja K, Kovacs KA et al. Comparison of 5 Ki-67 antibodies regarding reproducibility and capacity to predict prognosis in breast cancer: does the antibody matter? Hum. Pathol. 2017; 65: 31–40.

31. 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.

32. Lindboe CF, Torp SH. Comparison of Ki-67 equivalent antibodies. J. Clin. Pathol. 2002; 55: 467–471.

33. Lewis C, McQuaid S, Hamilton PW, Salt-Nelles M, McArt D, James JA. Building a ‘Repository of Science’: The importance of integrating biobanks within molecular pathology programmes. Eur. J. Cancer 2016; 67: 191–199.

34. Bankhead P, Loughrey MB, Fernandez JA et al. QuPath: open source software for digital pathology image analysis. Sci. Rep. 2017; 7: 16878.

35. Humphries MP, Hynes S, Bingham V et al. Automated tumour recognition and digital pathology scoring unravels new role for PD-L1 in predicting good outcome in ER-/HER2+ breast cancer. J. Oncol. 2018; 2018: 2937012.

36. Humphries MP, McQuaid S, Craig SG et al. Critical appraisal of programmed death ligand 1 reflex diagnostic testing: current standards and future opportunities. J. Thorac. Oncol. 2019; 14: 45–53.

37. Loughrey MB, Bankhead P, Coleman HG et al. Validation of the systematic scoring of immunohistochemically stained tumour tissue microarrays using QuPath digital image analysis. Histopathology 2018; 73: 327–338.

38. Viudez A, Carvalho FL, Maleki Z et al. A new immunohistochemistry prognostic score (IPS) for recurrence and survival in resected pancreatic neuroendocrine tumors (PanNET). Oncotarget 2016; 7: 24950–24961.

39. Yamaguchi T, Fujimori T, Tomita S et al. Clinical validation of the gastrointestinal NET grading system: Ki67 index criteria of the WHO 2010 classification is appropriate to predict metastasis or recurrence. Diagn. Pathol. 2013; 8: 65.

40. Genc CG, Falconi M, Partelli S et al. Recurrence of pancreatic neuroendocrine tumors and survival predicted by Ki67. Ann. Surg. Oncol. 2018; 25: 2467–2474.

41. Scarpa A, Mantovani W, Capelli P et al. Pancreatic endocrine tumors: improved TNM staging and histopathological grading permit a clinically efficient prognostic stratification of patients. Mod. Pathol. 2010; 23: 824–833.

42. Pelosi G, Bresaloi E, Bognia G et al. Endocrine tumors of the pancreas: Ki-67 immunoreactivity on paraffin sections is an independent predictor for malignancy: a comparative study with proliferating-cell nuclear antigen and progesterone receptor protein immunostaining, mitotic index, and other clinicopathologic variables. Hum. Pathol. 1996; 27: 1124–1134.

43. Rindi G, Falconi M, Klersy C et al. TNM staging of neoplasms of the endocrine pancreas: results from a large international cohort study. J. Natl Cancer Inst. 2012; 104: 764–777.

44. Mazzaferrro V, Pulvirenti A, Coppa J. Neuroendocrine tumors metastatic to the liver: how to select patients for liver transplantation? J. Hepatol. 2007; 47: 469–466.

45. Caplin ME, Pavel M, Cwiklak JB et al. Lanreotide in metastatic enteropancreatic neuroendocrine tumors. N. Engl. J. Med. 2014; 371: 224–231.

46. Falconi M, Eriksson B, Kaltssen G et al. ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. Neuroendocrinology 2016; 103: 153–171.

47. Childs A, Kirkwood A, Edelcine J et al. Ki-67 index and response to chemotherapy in patients with neuroendocrine tumours. Endocr. Relat. Cancer 2016; 23: 563–570.

48. Milione M, Maisonneuve P, Spada F et al. The clinicopathologic heterogeneity of grade 3 gastroenteropancreatic neuroendocrine neoplasms: morphological differentiation and proliferation identify different prognostic categories. Neuroendocrinology 2017; 104: 85–93.

49. Sorbye H, Welin S, Langer SW et al. Predictive and prognostic factors for treatment and survival in 305 patients with advanced gastrointestinal neuroendocrine carcinoma (WHO G3): the NORDIC NEC study. Ann. Oncol. 2013; 24: 152–160.

50. Bousturk O, Yang Z, Tang LH et al. The high-grade (WHO G3) pancreatic neuroendocrine tumor category is morphologically and biologically heterogeneous and includes both well differentiated and poorly differentiated neoplasms. Am. J. Surg. Pathol. 2015; 39: 683–690.
Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Line graphs demonstrating the agreement between the manual Ki67 PI (blue line) versus digital Ki67 PI for 500 cells (yellow line), 1000 cells (green line) and 2000 cells (red line) for the individual clones in the 42 PNET cases. Clinical cut-off (Black line), M1B1 (i), MM1 (ii), K2 (iii) and 30-9 (iv).

Table S1. Percentage of cases showing grade concordance between individual manually assessed Ki67 clones and image analysis using 500, 1000 and 2000 total tumour cell hotspots.