

**Computerized Calculation of Mitotic Count Distribution in Canine Cutaneous Mast Cell Tumor Sections: Mitotic Count is Area-Dependent**

Christof A. Bertram<sup>1\*</sup>, Marc Aubreville<sup>2\*</sup>, Corinne Gurtner<sup>1,6</sup>, Alexander Bartel<sup>3</sup>, Sarah M. Corner<sup>4</sup>, Martina Dettwiler<sup>6</sup>, Olivia Kershaw<sup>1</sup>, Erica L. Noland<sup>4,5</sup>, Anja Schmidt<sup>7</sup>, Dodd G. Sledge<sup>4</sup>, Rebecca C. Smedley<sup>4</sup>, Tuddow Thaiwong<sup>4</sup>, Matti Kiupel<sup>4,5</sup>, Andreas Maier<sup>2</sup>, Robert Klopfleisch<sup>1</sup>

<sup>1</sup> Institute of Veterinary Pathology, Freie Universität Berlin, Berlin, Germany,

<sup>2</sup> Pattern Recognition Lab, Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

<sup>3</sup> Institute for Veterinary Epidemiology and Biostatistics, Freie Universität Berlin, Berlin, Germany

<sup>4</sup> Michigan State University Veterinary Diagnostic Laboratory, Lansing, MI, USA

<sup>5</sup> Department of Pathobiology and Diagnostic Investigation, College of Veterinary Medicine, Michigan State University, East Lansing, MI, USA

<sup>6</sup> Institute of Animal Pathology, Department of Infectious Diseases and Pathobiology, Vetsuisse Faculty, University of Bern, Bern, Switzerland

<sup>7</sup> Vet Med Labor GmbH - Division of IDEXX Laboratories, Ludwigsburg, Germany

\* Authors contributed equally to the study

**Corresponding author:**

Robert Klopfleisch, Institute of Veterinary Pathology, Freie Universität Berlin, Robert-von-Ostertag-Straße 15, 14163 Berlin, Germany.

Email: robert.klopfleisch@fu-berlin.de

**Abstract:**

Mitotic count (MC) is an important element for grading canine cutaneous mast cell tumors (ccMCTs) and is determined in ten consecutive high-power fields with the highest mitotic activity. However, there is variability in area selection between pathologists. In this study, the MC distribution and the effect of area selection on the MC were analyzed in ccMCTs. Two pathologists independently annotated all mitotic figures in whole-slide images of 28 ccMCTs (ground truth). Automated image analysis was used to examine the ground truth distribution of the MC throughout the tumor section area, which was compared with the manual MCs of 11 pathologists. Computerized analysis demonstrated high variability of the MC within different tumor areas. There were 6 MCTs with consistently low MCs ( $MC < 7$  in all tumor areas), 13 cases with mostly high MCs ( $MC \geq 7$  in  $\geq 75\%$  of 10-hpf-areas) and 9 borderline cases with variable MCs around 7, which is a cut-off value for ccMCT grading. There was inconsistency among pathologists in identifying the areas with the highest density of mitotic figures throughout the three ccMCT groups; only 51.9% of the counts were consistent with the highest 25% of the ground truth MC distribution. Regardless, there was substantial agreement between pathologists in detecting tumors with  $MC \geq 7$ . Falsely low MCs below 7 mainly occurred in 4/9 borderline cases that had very few ground truth areas with  $MC \geq 7$ . The findings of this study highlight the need to further standardize how to select the region of the tumor in which to determine the MC.

**Keywords:** Area selection, high-power field, mitotic activity, mitotic figure distribution, tumor grading, tumor periphery.

Quantification of mitotic figures is one of the most common prognosticators of malignancy in veterinary and human tumor pathology.<sup>25,26</sup> It is well accepted that higher numbers of mitotic figures are commonly associated with a more aggressive behavior of neoplasms.<sup>14,15,18,34,38,47</sup> Accurate quantification of mitotic figures in standard histological sections stained with hematoxylin and eosin (HE) has been shown to have higher correlation to prognosis than a subjective impression by quick scanning in human breast cancer.<sup>39</sup> Also for malignant neoplasms in veterinary pathology, it is recommended to routinely quantify mitotic activity by means of the mitotic count (MC), which is defined as the number of mitotic figures in ten consecutive non-overlapping high-power fields (hpf, field visible at 400x magnification with a field number of the ocular of FN=22, size is 2.37mm<sup>2</sup>).<sup>25</sup> For some neoplasms, e.g., canine mast cell tumors (MCT),<sup>20,30,43</sup> canine and feline mammary tumors,<sup>29,31</sup> canine melanocytic tumors,<sup>42</sup> or canine soft tissue sarcomas,<sup>13,24</sup> the MC is also an important part of the tumor grading and cut-off values for the MC have been established that directly impact tumor grade. For example, the canine cutaneous MCT (ccMCT) grading developed by Kiupel *et al.*<sup>20</sup> utilizes a cut-off value of seven mitotic figures in ten hpf as one of the criteria to grade ccMCTs. Former grading systems have used the cut-off value of MC of 10.<sup>11</sup> Others have proposed prognostication of ccMCTs based entirely on the MC using a cut-off value of 5 or 7.<sup>15,32</sup>

The MC is very applicable for routine surgical pathology and can be determined on glass and digital slides without any specialized equipment. Previous studies have shown that there is strong agreement in MC as determined by either light or digital microscopy in canine round cell tumors and human breast cancer.<sup>1,7</sup> Nevertheless, the MC has some limitations in terms of inconsistency between observers. Several studies on human breast cancer<sup>1,9,12,27,28,44</sup>, canine mammary carcinomas<sup>36</sup> and canine round cell tumors<sup>7</sup> have shown that there is generally a high degree of inter-observer and

intra-observer variation of the MC. One of the most intensively discussed factors is the area unit 'hpf' which may vary up to 47% between different light microscopes at 400x magnification, depending on the field number (FN) of the respective ocular lens.<sup>6,25</sup> Surprisingly, many of the studies mentioned above did not define a 'hpf' beyond being the field of view at 400x magnification. A standard FN of 22mm with a subsequent area of a hpf at 400x magnification of 0.237mm<sup>2</sup> has been proposed for veterinary pathology by Meuten *et al.*<sup>25</sup>. Similarly, when digital whole-slide images are viewed on a computer screen, the displayed area depends significantly on the monitor's diagonal size.<sup>7,25</sup> Further attempts of standardization are to correct the MC by the proportion of neoplastic tissue within the hpf enumerated.<sup>12</sup> Indeed, standardization of the area size has been shown to somewhat improve concordance of the MC as has been demonstrated in canine round cell tumors with an increase of the concordance correlation coefficient by 0.044 (non-standardized: 0.648; size-standardized: 0.692) and human breast cancer with an increase of the coefficient of variance of 3.7% (56.0% in the conventional method and 52.3% in the volume corrected method).<sup>7,12</sup>

Another factor of influence is the difference between observers in their ability to identify individual mitotic figures and to distinguish them from other cell structures such as necrotic cells, especially in cases of prophase mitotic figures.<sup>23,27,44</sup> Few studies in human breast cancer reported only slight to moderate inter-observer agreement ( $\kappa$ = 0.13 - 0.57) or 6.9 – 54.5% disagreement on deciding whether an individual cell structure represents a mitotic figure or not.<sup>23,27,44</sup> Appropriate sample fixation and slide preparation, including section thickness, are likely to have impact on mitotic figure identification.<sup>6</sup> The influence of the viewing modality, i.e. light or digital microscopy, on identification of individual mitotic figures has not been fully assessed in veterinary or human pathology.

Importantly, selection of the evaluated hpf has been poorly standardized with vague selection criteria.<sup>4,44</sup> Therefore, we hypothesized that variable area selection may be another factor influencing the MC regardless of the tumor type. Most grading systems as well as general recommendations request that mitotic figures should be counted in a single area comprised of 10 contiguous, non-overlapping hpf located in the tumor region with the highest mitotic activity.<sup>6,13,14,20,25,29,31,33,42,45,46</sup> A tumor section with an area of 4.0 cm<sup>2</sup> contains more than 1600 non-overlapping, standard-sized hpf. Therefore, scanning the tumor section and selecting the one area with the highest MC might be problematic.<sup>4,44</sup> Some studies circumvented this problem by counting mitotic figures in randomly selected hpf.<sup>19,22,41</sup> Other chose the most cellular or most anaplastic part of the tumor as well as excluding areas with necrosis, hemorrhage and other artifacts.<sup>6,13,24,25</sup> It is also recommended to perform the MC in the periphery of the tumor as this is the invasive front and the most optimally fixed material.<sup>6,25,26</sup> Although there is no systematic study in veterinary pathology to date, some authors have hypothesized that the potentially more aggressive tumor cells with possibly more proliferation are located at the invasive front of the tumor.<sup>25,37</sup>

In this study, we hypothesized that mitotic figures are not randomly distributed in ccMCTs and that an arbitrary selection of the area enumerated will influence the MC. To test this hypothesis, two pathologists annotated all mitotic figures in whole-slide images of HE-stained tissue sections of 28 different ccMCTs. Consecutively, a self-developed automated image analysis software was used to systematically analyze the distribution of mitotic figures throughout entire tumor sections (the software determined how many mitotic figures, which were labeled by two pathologists, were located in every possible 10-hpf-area of the tumor; we did not use artificial intelligence to identify mitotic figures). MCs performed manually by 11 anatomic pathologists were compared with the computerized ground truth MC distribution in order to determine how

successfully pathologists determine highest MCs. In addition, this study also tested the hypothesis that the density of mitotic figures is higher in the tumor periphery in comparison to the tumor center.

## **Material and methods**

### *Specimens*

Thirty-two ccMCTs, including 16 low-grade and 16 high-grade ccMCT (two-tier grading system according to original diagnosis; all high-grade cases had a MC  $\geq 7$ ),<sup>20</sup> with acceptable tissue quality (determined by sufficient histological perceptibility of cellular details) were randomly selected from the authors' archive. Subsequently, we excluded 4 cases with tumor section areas of  $<12\text{mm}^2$  from systematic analysis. These 4 ccMCT contained very few mitotic figures (1 –5 mitotic figures in the entire tumor section with a highest possible MC of 1 or 2) and, therefore, were considered inappropriate for calculating of MC distribution (exemplary heatmaps in Supplemental Figs. S1 and S2). MCTs had been formalin-fixed, sectioned along their largest and orthogonal diameter, and paraffin-embedded during routine diagnostic service. For this study we produced new tissue sections of a single representative tissue block per case, which were subsequently stained with hematoxylin and eosin (HE) by a tissue stainer (ST5010 Autostainer XL, Leica, Germany). All slides were digitalized with a linear scanner (ScanScope CS2, Leica, Germany) in one focal plane by default settings at a scanning magnification of 400x, resulting in an image resolution of 0.25  $\mu\text{m}$  per pixel. Largest histopathological tumor diameters of the 28 cases were measured in the whole-slide image by the ruler tool of the Aperio ImageScope© (Leica Biosystem Imaging) software and ranged from 5 mm to 28 mm (mean: 17.3 mm; Supplemental Table S1). Tumor diameters had to be determined from digitalized slides as gross measurements were not available for all cases. Statistical difference in largest tumor diameter between the later-defined three groups was calculated by a two-sided T-test by use of the SPSS Statistics 25 for Windows (IBM, New York, United States).

### *Tumor area and mitosis annotation*

The open source software package SlideRunner<sup>2</sup> was used for two annotation tasks (Fig. 1A). First, mitotic figures were labeled within the entire whole-slide image by two pathologists [CB, RK]. Initially, the complete slide was screened twice for all mitotic figures within the tumor section by the first pathologist [CB]. The screening was performed in a semi-automated manner, where the software would propose partly overlapping, consecutive image sections at a magnification of 400x to ensure that the entire whole-slide image was viewed thoroughly. As part of our annotation process, many of the non-mitotic cells (i.e. mast cells, eosinophils, and other cell structures such as apoptotic cells) were annotated as a different annotation class. We considered this an important approach in order to reduce confirmation bias for the second annotation task. The second pathologist [RK] was then asked to assign a second class, i.e. mitotic figure or non-mitotic cell, to all initially annotated structures, which were blinded to the cell class given by the first pathologist. This second opinion annotation was also performed in a semi-automated manner, where the software showed initial annotations that obscured the class previously assigned until all annotations on the slide had been given a second opinion. Both pathologists were instructed to only label convincing mitotic figures. Telophase figures with cytokinesis were labeled as a single mitotic figure. In total, 52,295 cell structures were annotated as mitotic figures by at least one of two pathologists in the 28 whole-slide images. Of those, 43,421 cell structures (83.0 %) were agreed to be mitotic figures by both pathologists and were included in the ground truth data set for further analysis. Ground truth is a term for labeled data sets in image analysis that describes image information determined by the defined gold standard method, in this case the image location (coordinate) and annotation class (i.e. mitotic figure) of manual annotations agreed on by both pathologists. For the remaining 8874 mitosis-annotations (17.0 %), the two pathologists disagreed on whether or not the cell structure was a mitotic figure and these annotations were disregarded for the



following study. Furthermore, a total of 152,866 annotations were labelled as non-mitotic cell structures by consensus

Second, the tumor section area within the whole tissue section was encircled by a single pathologist [CB] using a polygon annotation tool. This tumor area map included the entire tumor area and excluded necrotic areas, non-tumor-tissue within the tumor area larger than one hpf, such as a large blood vessel or entrapped residual tissue (sebaceous glands, hair follicles) and, in one case, a blurry image section (scanning artifact). In three cases (case 7, 16, and 22) there were areas within the tumor section area with variable cellularity.

#### *Generation of the mitotic count map*

Computerized image analysis was used to determine distribution of the manually labeled mitotic figures within the valid tumor area. For this purpose, different image maps were created from ground truth mitotic figure and tumor area annotations. First, a size-reduced digital map of the ground truth mitotic figure annotations was created using a downsampling factor of 32 (Fig. 1B). In this binary mitotic figure map, the center of each mitotic figure was assigned the value "1" and other coordinates the value "0". The result was a map that included all single mitotic events at their respective downsampled coordinates with reduced computational complexity over the typically very large original whole-slide image dimensions.

In a second step, we used a sliding window summation to assess the number of mitotic figures within a defined area in the slide. This tumor area was specified as a continuous region of adjacent ten standard-sized hpf with each spanning  $0.237\text{mm}^2$  as recommended by Meuten *et al.*<sup>25</sup> The resulting window covered a total area of  $2.37\text{mm}^2$  with an aspect ratio of 4:3 centered around the respective coordinates and a width/height of 1777.6/1333.2 microns, or approximately 7110/5333 pixels at original

scanning resolution, respectively. Downsampled as described above, the window was thus reduced to a size of 222 by 167 pixels. We decided to use the shape of a 4:3 rectangle for the 10-hpf-area, as the viewing area of digital microscopy (computer screen) is rectangular, not round as for light microscopy. For each pixel position of the downsampled mitotic figure map, a frame of that defined size was centered on this pixel position and the sum of all mitotic figures within that frame was calculated. This sliding window summation yielded the mitotic count map that contained every possible mitotic count given an arbitrary center coordinate of the 10-hpf-area (Fig. 1C). In other words, a pixel of the mitotic count map represents the MC within the 10-hpf-area of which it is the center coordinate. Those steps described above have been replicated for defined areas of 20 hpf (4.74mm<sup>2</sup>), and 50hpf (11.85mm<sup>2</sup>).

#### *Generation of the valid tumor area map*

In a next step, a map of the valid tumor area in the tissue section was created as the region on the whole-slide image in which the MC derived from the mitotic count map had to be determined. We took the annotated area of tumor tissue on each slide and removed all exclusion image parts representing non-tumor tissue, necrotic tumor tissue, and blurry image regions. This resulted in a tumor area map, in which the entire tumor tissue was represented by the value “1”, and all non-tumor tissue by “0” (Fig. 1D). The size of the tumor section area was determined from this map. The selected tumor sections had a mean tumor section area of 167.86 mm<sup>2</sup> (range: 14.86 – 346.10 mm<sup>2</sup>; standard deviation: 95.01 mm<sup>2</sup>; see Supplemental Table S1 for tumor size of individual sections).

Subsequently, a derived valid tumor map was created that comprised all coordinates in which the center coordinate of the 10hpf, 20hpf, and 50hpf area was placed while covering tumor tissue (tumor area map) by at least 95% of the hpf-area (Fig. 1E). To

achieve this, we used a moving window averaging approach with a subsequent threshold of 0.95, effectively yielding a valid tumor map with reduced area over the initial tumor area map.

#### *Computerized analysis of the ground truth MC distribution*

For statistical analysis, all entries of the mitotic count map that were valid as per valid tumor area map were considered. Depending on the size of the valid tumor area, the total amount of 2.37mm<sup>2</sup> area windows evaluated varied from 57,665 to 4,185,973 between the 28 cases (Supplemental Table S1). These numbers included all possible pixel-positions of the 2.37mm<sup>2</sup>-area window within the down-sampled valid tumor area that was determined by computerized calculation and used for the ground truth MC distribution. First, we computed the probability to arbitrarily select a high MC ( $MC \geq 7$ ) 10-hpf-area; i.e. the number of the 10-hpf-areas that contained equal to or more than 7 mitotic figures divided by all possible 10-hpf-areas within the valid tumor area (without exclusion tissue). The likelihood to select low MC areas ( $MC < 7$ ) is the reciprocal of the indicated values. Based on that determined probability of the ground truth MC being above the cut-off value of 7, we defined 3 groups of ccMCTs (Fig. 2): 1) Group 1 (low MCs, n=6) has a 0% ground truth probability of a  $MC \geq 7$ , i.e. all potential contiguous 10-hpf-areas have MCs  $< 7$ ; 2). Group 2 (borderline cases, n=9) has a 0.01 – 74.99% probability of selecting an area with a MC above the cut-off value of 7, i.e. at least contains one 10-hpf-area with  $MC \geq 7$ ; and 3) Group 3 (mostly high MCs, n=13) has a  $\geq 75\%$  ground truth chance of arbitrarily selecting an area that has a  $MC \geq 7$ , i.e. mostly comprises of 10-hpf-areas with  $MC \geq 7$ . Additionally, we similarly evaluated two further published cut-off values of 5, and 10.<sup>11,20,32</sup> Secondly, we determined the regional variability of the mitotic count of each case by heat maps, box-and-whisker plots (see Supplemental Table S2 for minimum, 1<sup>st</sup> quartile, median, 3<sup>rd</sup> quartile and maximum

value), and histograms. Heat maps were generated by adding a semi-transparent green overlay to the microscopic image with the maximum value (as indicated by the color bar) having no transparency and the minimum value having full transparency (transparency is relative to the maximum MC). Box-and-whisker plots and histograms were generated with the python matplotlib 2.2.2 software. Effects of cellular density were determined subjectively based on the heat maps.

#### *Manual MC by pathologists*

Manual mitotic counts were obtained according to current guidelines by 11 board-certified anatomic pathologists [AS, CG, DS, EN, MD, MK, OK, RK, RS, SC, TT; order of participants has been randomized for analysis] from four different diagnostic laboratories by examining the same digital images that were used for the ground truth data set. Pathologists were instructed to perform the MCs in the area with the highest mitotic density as requested by the current grading system to simulate a diagnostic setting.<sup>20</sup> According to current recommendations, the number of required hpf was calculated for each pathologist in order to examine a total area of 2.37 mm<sup>2</sup>, as the size of a hpf varies between different computer monitors.<sup>25</sup> Presentation order of the 28 cases was randomized. Pathologists were blinded to the tumor grade and the MCs of other pathologists. A total of 308 MC were available as a result of 11 pathologists examining all 28 cases (Supplemental Table S3). The MCs of the pathologists were categorized using the quartiles of the ground truth MC distribution. We chose a classification in favor of the pathologists by defining the categories as greater or equal than the lower bound quartile. Following this rule, the categories were defined as: Category 1 (MCs < 1st quartile), Category 2 (1st Quartile ≤ MCs < Median), Category 3 (Median ≤ MCs < 3rd Quartile), Category 4 (3rd Quartile ≤ MCs). Low MC (group 1)

cases had a very narrow ground truth distribution with close to equal quartile cut-off values (discrete counts) and therefore category 4 may include significantly more than 25% of the values. Categories 3 and 4 combined represent the highest 50% of all possible ground truth MCs. Additionally, we categorized the MCs into below and above the cut-off value for ccMCT grading/prognostication.<sup>11,15,20,32</sup> 95% Confidence intervals (CI) were determined by the Wilson method with a continuity correction. Agreement for the high MC classification ( $MC \geq 7$ ) was evaluated by Light's kappa ( $\kappa$ ) for fully crossed designs and interrater agreement for the raw MC values of all pathologists was evaluated by intra-class correlation coefficient (ICC; two-way agreement, single-measures, random). The  $\kappa$  and ICC values are measures of the level of agreement corrected by chance between different participants; however, they do not indicate degree of accuracy to the ground truth. We evaluated kappa as poor = 0, slight = 0.01–0.20, fair = 0.21–0.40; moderate = 0.41–0.60, substantial = 0.61–0.80, or almost perfect = 0.81–1.00.<sup>17</sup> ICC was evaluated as poor = 0 – 0.39, fair = 0.40 – 0.59, good = 0.6 – 0.74 and excellent = 0.75 – 1.00.<sup>17</sup> Calculations were performed using R version 3.52 (R Foundation, Vienna) using the package irr version 0.84.1.<sup>16</sup>

#### *Differences between tumor periphery and center*

Counting mitotic figures at the periphery of the tumor section has been recommended due to better fixation and suspected higher proliferation of tumor cells at the invasion front.<sup>25</sup> For this reason, the ground truth mitotic activity index (MAI) was compared between the tumor periphery and center. MAI is defined as the number of mitotic figures within a tumor area, in this case tumor periphery or center, divided by the size (in  $\text{mm}^2$ ) of that respective area.<sup>25</sup> As there is currently no consensus on the proportional and absolute size of the periphery in ccMCTs, we evaluated three different ratios of the tumor periphery containing 25%, 50%, and 75% of the total tumor area,

respectively. The tumor periphery was defined as the area between the tumor border margins, which was annotated by a single pathologist (see above), subtracted by the area of the tumor center, which was determined by step-wise reduction of the tumor area from the outside (erosion operator with uniform 3x3 filter kernel) until the target percentage of 75%, 50% and 25% (standard deviation: <0.15%) was reached. Tissue sections that contained wide-spread necrosis and ulceration, non-tumor tissue and scanning artifacts were excluded from calculation as described above. If the whole-slide image comprised several tissue sections, area proportions were determined for the joint area of the sections and calculations of tumor size and MAI were added. Differences of the MAI between periphery and center for each defined proportion were tested for significance by a two-sided t-test with the python scipy version 1.1.0 stats-package software.

## Results

### *Probability of assigning high MC in arbitrary area selection*

The ground truth chances to arbitrarily select a tumor area that contains MCs above the cut-off value of  $MC \geq 7$  ranged from 0% to 99.97% (Supplemental Table S4). Group 1 (always low MC, 0% chance) contained 6 ccMCTs (cases 1 – 6) and had the significantly ( $p \leq 0.003$ ) smallest histopathological tumor diameters, ranging from 5 to 16 mm (Median: 9.5mm; Fig. 3). Group 2 (borderline cases) included 9 ccMCTs with a very small (0.20 – 7.43%; case 7 – 10; subgroup 1) to moderate (26.48% – 73.97%; case 11 – 15; subgroup 2) ground truth chance of arbitrarily selecting a high MC area. In other words, these cases included few (subgroup 1; third quartile below 7) to numerous (subgroup 2, third quartile equal or greater 7) tumor areas with a  $MC \geq 7$  and they included many tumor areas with a  $MC < 7$ . In group 2, largest histopathological tumor diameters ranged from 9 to 28 mm (median: 21 mm) and were

significantly ( $p= 0.003$ ) larger than for group 1 MCTs, while two cases (case 9 and 10) had smaller tumor diameters than the maximum tumor diameter of group 1 MCTs (Fig. 3). Group 3 (mostly high MC,  $\geq 75\%$  chance) comprised 13 ccMCTs (cases 16 – 28). In group 3, the largest tumor diameters of histopathological sections ranged from 7 to 24 mm and were overall significantly ( $p= 0.002$ ) larger than tumor diameters of group 1 MCTs. However, three cases (cases 19, 21 and 25) overlapped with the tumor diameter range of group 1 MCTs (Fig. 3). Although in the majority of the cases the MC range correlates with the tumor diameter, cases 19, 21 and 25 were noteworthy exceptions.

#### *Ground truth MC distribution*

Ground truth mitotic figures were irregularly distributed throughout the tumor section area for all groups, i.e. there were different MC depending on the tumor region assessed (Supplemental Figs. 1 and 2). Always-low MC cases (group 1) had MC ranges between 0 – 5 (Fig. 4 and 5). Therefore, area selection seems to be of minor importance in this group. Borderline cases (group 2; Fig. 6) had lower MC ranges in subgroup 1 between 7 – 13 (, Fig. 7) compared to subgroup 2 with MC ranges between 30 – 108 (; Supplemental Fig. 3). In contrast to group 1, area selection seems to be essential in group 2. Mostly-high MC cases (group 3) had the highest ground truth MC between 28 and 335 (Fig. 8 and 9). In this group, random area selection would rarely lead to MCs below the cut-off value. Case 19 had an exceptionally narrow distribution and this tumor also had the smallest size. The present study included three ccMCTs with distinctly variable cellular density throughout the tumor section. Whereas in one case, the highest mitotic counts were in areas with high cellular density (case 18; Supplemental Fig. S4), another case had the highest activity in a region with low

cellular density (case 7; Supplemental Fig. S5). In the third case, cell-sparse regions had medium mitotic activity (case 22; Supplemental Fig. S6).

By increasing the size of the area examined (10 hpf to 20 hpf and 50 hpf), the range of MC distribution (maximum and minimum) was generally mildly to moderately reduced (Supplemental Fig. 7-9), and the highest MC value was lowered to beneath the cut-off value in cases 7 to 10 (all cases of subgroup 1 from the borderline group). However, the median values largely remained unchanged.

#### *Pathologist's Manual MC*

Overall inter-observer agreement of the manual MCs of eleven board-certified pathologists was good (ICC = 0.715; CI: 0.592 – 0.830). In comparison, ICC of group 3 was smaller (ICC = 0.603; CI: 0.402 – 0.816) indicating that agreement on the highest values becomes more difficult with higher MC ranges. ICC of other groups are not indicated since the range of possible values was quite small and relative deviation was therefore overinflated (ICC values were quite small and should not be interpreted as such).

Compared to the ground truth distribution, the manual MCs ranged from low to high values (Fig. 4, 6, 8). Of all manual counts of all pathologists, 74.4% (CI: 68.2 – 80.5%) were within the highest  $\geq 50\%$  (Category 3 and 4) of all ground truth MCs and 51.9% (CI: 42.6 – 61.6%) were within the highest  $\geq 25\%$  (Category 4) (Table 1, Supplemental Fig. S10). Manual MC of MCTs in group 3 were less often within the highest half (56.6%; CI: 44.8 – 68.9%) and quarter (36.4%; CI: 22.8 – 48.8%) of all possible ground truth MCs in comparison to group 1 (within highest half in 90.9%, CI: 81.2 – 95.9%; within highest quarter in 66.7%, CI: 53.2 – 79.0%) and group 2 (within highest half in 88.9%, CI: 79.7 – 97.9%; within highest quarter in 64.6%, CI: 53.0 – 76.0%). When comparing the individual pathologists, there was high variability between MCs being



within the ground truth highest half ranging from 60.7 to 92.9%; (Fig. 10 and Supplemental Table S5) and highest quarter ranging from 32.1 to 82.1% (Fig. 11 and Supplemental Table S5).

Additionally, we tested whether the manual MCs were above or below the grading cut-off of 7 (Supplemental Fig. 11). Overall, there was substantial agreement between pathologists to detect high MCs above or equal 7 ( $\kappa= 0.865$ ) for all cases combined. When comparing the three tumor groups, the MCs resulted consistently in a good prognosis and poor prognosis designation (based on the MC alone) for ccMCTs in groups 1 and group 3, respectively. However, for the borderline group, the MC was variable around the cut-off value, which would potentially impact the assigned grade. In low MC cases (group 1), the manual MCs were below 7 in 98.5% (CI: 91.6 – 99.8%; 1.5% false positive) with only one overestimated count (case 4). In group 3 ccMCTs, pathologists determined MCs  $\geq 7$  with a detection rate of 96.5% (CI: 91.9 – 98.6%; mean ground truth chance in random area selection is 92.9%) with disagreement only in case 19, which had a very narrow distribution. Manual MC of borderline cases, which contained at least one area with MC  $\geq 7$  according to the ground truth distribution, were above the cut-off only in 53.5% (95.5-CI: 43.5 – 63.3%; mean ground truth chance in random area selection is 34.2%). In this group, underestimation of the MCs mainly occurred in subgroup 1, which had the cut-off value of 7 in the upper quartile of the ground truth distribution. Detection rate for MC  $\geq 7$  was 4.5% (CI: 1.1 – 16.4%; mean ground truth chance in random area selection is 2.9%). In contrast, pathologists had a detection rate of 92.7% (CI: 51.0 – 99.9%) for MC  $\geq 7$  for subgroup 2 cases (cases 11 – 15; ground truth chance above 25%; mean is 59.2%); i.e. consistently reached cut-off values.

#### *Usage of different cut-off values*

We tested the influence of various published cut-off values, i.e.  $MC \geq 5$ ,  $MC \geq 7$  and  $MC \geq 10$ .<sup>11,15,20,32</sup> Compared to the commonly used cut-off value of 7, we found considerable differences in the probability to assign a prognosis with cut-off values of 5<sup>32</sup> leading to a higher chance and the cut-off values 10<sup>11</sup> leading to a lower chance of assigning a poor prognosis/outcome (Supplemental Fig. S11, S12 and Table S4). The influence was most significant in the borderline cases (group 2) with the greatest difference in chance between the cut-off values  $MC \geq 5$  and  $MC \geq 10$  being 42.56% in the most extreme case (tumor case 14; mean difference: 20.5%). In group 1, there was a mean difference of 0.04% (range: 0.00% - 0.18%) and in group 3, a mean difference of 9.09% (range: 0.28% - 38.66%). Similar effects were observed with the manual MCs performed by the 11 pathologists (Supplemental Table S4). Based on the manual MCs, borderline cases would have been assigned a high grade in 59.6%, 53.5%, and 33.3% (a difference of 26.3%) by the cut-off value of 5, 7, and 10, respectively. Therefore, the usage of different cut-off values resulted in relevant inter-observer variation for borderline cases. On the other hand, group 1 cases had a difference of only 1.5% (3.0%, 1.5%, and 1.5% probability, respectively) and group 3 cases had a difference of only 4.6% (97.9%, 96.5%, and 93.3% probability, respectively). As such, the cut-off value applied had minimal impact on inter-observer variation.

#### *Differences between the tumor periphery and center*

Computerized analysis was used to assess whether the ground truth mitotic density determined by the mitotic activity index (MAI, count of mitotic figures within the entire periphery or center area divided by the size of the area) varied between the periphery and center of the tumor area in the included ccMCTs. We defined the tumor periphery at three different sizes containing 25%, 50% or 75% of the entire tumor section area (Fig. 12). For a 25% periphery area ratio, the median MAI of the periphery was 2.65

mitotic figures/mm<sup>2</sup> and of the center was 4.55 mitotic figures/mm<sup>2</sup> (Fig. 13; for values of individual cases see Supplemental Table S6). In the case of the 50/50 area split, the median MAI was 3.60 mitotic figures/mm<sup>2</sup> and 4.68 mitotic figures/mm<sup>2</sup> for the periphery and center, respectively (Fig. 13; for values of individual cases see Supplemental Table S7). When the periphery contained 75% of the total valid tumor area, the median MAI of the periphery was 3.98 mitotic figures/mm<sup>2</sup> and of the center was 4.55 mitotic figures/mm<sup>2</sup> (Fig. 13; for values of individual cases see Supplemental Table S8). A two-sided t-test did not find a significant difference between the two groups for any of the three area proportions ( $p= 0.18$ ,  $p= 0.48$ , and  $p=0.87$ , respectively). Results were similar between the three tumor groups.

## Discussion

In order to standardize histopathological prognostication, several grading systems have been developed in veterinary pathology, all of which use the MC as one relevant parameter. An example is the two-tier grading system for ccMCT developed by Kiupel *et al.*<sup>20</sup>. This grading system uses the MC as one of four criteria to differentiate between aggressive high-grade and less aggressive low-grade MCT. Although the use of the MC, which is a seemingly objective absolute number, has increased the reproducibility of tumor grading systems, inter-observer inconsistency of the MC can potentially lead to variable prognostication. The area, in which the MC is performed, can only be hardly standardized potentially leading to inconsistency in prior studies and in routine diagnostics. In the present study, we however determined that the MC varies throughout ccMCT sections, which can potentially be explained by tumor heterogeneity. The MC varied among different regions of the tumor for those cases that had a high overall range of MCs (absolute variation). However, a relative regional MC variation was even higher for those cases with a small overall range of MCs,

although the absolute difference is small and likely of inferior clinical significance. This paradox needs to be considered when comparing agreement level (such as by the intra-class correlation coefficient) and this is why we decided to omit values for group 1 and 2. Regardless of the high variability of the MC throughout the study, we found that prognostication of all 28 cases solely based on manual MCs by pathologists was very consistent when a cut-off of 7 was applied.<sup>15,20</sup> High kappa values for grading indicated that all participating pathologists were able to perform in consensus with other pathologists, i.e. MCs were reproducible among the 11 board-certified pathologists. Only in a small subset of four borderline cases with very few ground truth areas having  $MC \geq 7$ , the manual MCs commonly did not reach the cut-off values (i.e. false negative; these tumors were often incorrectly identified as having  $MC < 7$  based on manual counts by pathologists). Most manual MCs performed in the cases belonging to group 1, subgroup 2 of group 2 as well as group 3 were consistently below respectively above the cut-off value in agreement with the ground truth. Future studies are required to determine whether those 'false negatives' are truly low or high grade tumors. Borderline cases (group 2) were problematic and might potentially represent those cases where histopathological prognostication does not fit to the clinical outcome.

The manual MC values of all pathologists had high variation being consistent with low to high mitotically active areas of the ground truth MC distribution. However, this high variability rarely impacted the prognostication based on current cut-off values. The upper quartile of the MC distribution (i.e. the MC in the mitotically most active tumor regions) was only reached in half of the cases; that is, pathologists' manual counts tended to underestimate the highest MC based on the ground truth data. For group 3 with the highest range of MCs, this was only true for one third of the manual MCs with

a 95% confidence interval that still included 25% (random chance). In addition, agreement of the 11 pathologists on the highest manual MCs was smaller for this group (group 3) in comparison to the overall agreement. Both findings indicate the difficulty in accurately and reproducibly performing the MC in cases with a high range of MCs and a potential large size of the histologic section.

Results of the present study prove that manual MCs often do not represent the most mitotically active tumor regions and suggest that area selection likely has a significant influence on the MC. In addition, the high degree of variability of the MC could also be impacted by the degree of inter-observer variation in identifying mitotic figures. Such variation has been estimated to range from 6.9% to 54.5% by previous studies while the present study had a disagreement of 17%.<sup>23,27,44</sup> Regardless, our study demonstrated that accurate and reproducible quantification of the most mitotically active region is hampered by marked inter-observer variation in digital slides.

A limitation of the present study is that disease-free intervals and survival times were not available. Therefore, the direct influence of variable area selection and MCs on tumor prognosis could not be determined. However, it is the current scientific consensus that the most mitotically active region (and not the region with mean mitotic activity) correlates best with prognosis.<sup>26</sup> Due to the moderate sensitivity of the current method in finding the most mitotically active area, the extent of correlation of the MC with disease outcome may have potentially been underestimated in previous studies. Further studies with large case numbers are required to verify whether the highest MC correlates best with clinical outcome.

There is currently no appropriate recommendation to reliably find the mitotically most active region in ccMCTs. It has been proposed that the periphery of a tumor might be the mitotically more active region due to the location of the most aggressive tumor cells at the invasive front of the tumor.<sup>25,26</sup> However, the present study found a minor but

statistically insignificant tendency towards a higher MC in the tumor center for all cases combined. However, individual cases were highly variable in whether the periphery or center of the tumor had higher MCs. These findings imply that the approach to count mitotic figures primarily in the tumor periphery would not improve standardization of the MC in ccMCTs. Larger studies are essential to further evaluate whether the invasive front or the center of the tumor is a more reliable area than the periphery to determine the highest MC. It has also been reported that the MC should be counted in the most cellular regions in order to standardize the MC.<sup>6,13,25,46</sup> In the present study, we included three cases with variable cellular density and in one case, the mitotically most active region was in an area of low cellularity. Although it seems conceivable that areas with lower cellularity may generally have a lower mitotic count, this may not be true for every case, and sometimes would have led to underestimation of the MC. Also, all current grading systems in veterinary pathology pay little attention to cellular density in terms of how the MC (a parameter independent of cellular density) is determined. In human breast cancer grading, a volume-corrected MC (MC divided by the estimated proportion of the hpf covered by tumor tissue) has been proposed;<sup>12</sup> however, we speculate that proportion estimates are more difficult in round cell tumors with solitary tumor cells and often minimal tumor stroma. The number of mitotically active cancer cells can also be determined as the mitotic index (MI), which is the number of neoplastic cells with mitotic figures divided by the total number of neoplastic cells within a defined area.<sup>25</sup> Although this parameter has been shown to be more reproducible than the mitotic count in some canine and feline tumors,<sup>37</sup> it would be too laborious in a diagnostic setting to count all neoplastic cells in a 10 hpf area (in the authors' experience there are often between 7,000 and 13,000 neoplastic cells in 10 hpf in a ccMCT). Future research will show if the MI can be implemented for routine diagnosis by the aid of computerized image analysis.

Selecting “crucial areas” in borderline cases seems very difficult. As discussed above, scanning the whole-slide image for the area with the highest mitotic density is quite ineffective especially in cases with a large range of MCs. One would assume that larger tumor sections may have higher a variability of the MC than smaller tumor sections simply based on the likelihood of such variations occurring more commonly in larger samples. In the present study, all tumor sections with a tumor diameter above 16 mm included areas with MC  $\geq 7$ . Although MCTs with low MC in all regions (group 1) had a maximum largest tumor diameter of 16 mm, borderline MCTs (group 2) and MCTs in which most regions had high MCs (group 3) had maximum tumor diameters as small as 9 mm and 7 mm, respectively. Therefore, tumor diameters were independent of MC at least in some cases and probably are inaccurate for use as a solitary distinguishing feature. In some feline and canine neoplasms (feline cutaneous MCT;<sup>33</sup> feline and canine mammary carcinoma;<sup>29,31</sup> canine melanocytic neoplasia<sup>42</sup>), tumor size has indeed been determined to be a prognostic indicator. To the authors’ knowledge, similar systematic investigations have not been performed for canine cutaneous MCTs. Depending on the cut-off value used for tumor grading, the overall chance to assign a high grade based on the MC alone differed considerably in borderline cases. The three cut-off values applied (7, 5, and 10) have been used either as part of grading systems or as solitary prognostic parameters for ccMCT in previous studies.<sup>11,15,20,32</sup> In contrast to previous grading approaches for ccMCT,<sup>11,30</sup> the current grading system developed by Kiupel *et al.*<sup>20</sup> comprises only two grades.<sup>30</sup> In the authors’ experience, some pathologists are more comfortable using a three-tier system that includes an intermediate grade regardless of the high prognostic and therapeutic uncertainty of the intermediate grade.<sup>35</sup> However, as we define biological behavior by the presence or absence of invasion and metastatic spread, an intermediate grade has not yet been proven to give additional prognostic or therapeutic information in ccMCT.<sup>20,35</sup>

Borderline cases of the present study should not be classified as an intermediate grade, but rather as difficult cases that are prone to inconsistent MC determination around the cut-off value. There are always borderline cases between two distinct grades regardless of the number of different cut-offs applied. In fact, a three-tiered system has borderline cases between the 1<sup>st</sup> and 2<sup>nd</sup> grade as well as between the 2<sup>nd</sup> and 3<sup>rd</sup> grade, which might add further variability in grading in comparison to a two-tiered system. The current results highlight the need to very carefully select the MC cut-off values when establishing a grading system as selection will have a great impact on borderline cases. In addition, it is likely that the MC cut-off value will have to be reevaluated with the development of methods that have higher sensitivity to find the most mitotically active region.

A limitation of this study is that there is some inter-observer variation in histologically identifying individual mitotic figures. We tried to reduce this as much as possible for the ground truth data set by blinded labeling of two independent pathologists. While the inter-observer variability (17%) was relatively small in comparison to previous studies (27 - 37%), the variability is only based on two pathologists establishing the baseline data.<sup>23,27,48</sup> The morphology of mitotic figures is defined as containing condensed chromosomes with hairy extension while lacking a nuclear membrane.<sup>26,46</sup> However, there are considerable differences in the morphology of mitotic figures depending on the phase of division as well as atypical morphologies and, therefore, discordance seems to be unavoidable to a certain degree. Even for well-trained pathologists it is difficult to differentiate mitotic figures from some apoptotic and pyknotic cells or overstained nuclei.<sup>23,27,44</sup> In the present study, mitotic figures were annotated in the entire tissue section; however, large areas with necrosis were excluded in a second step. We also excluded labels that were not in concordance by



two pathologists as it has been recommended to only count unambiguous mitotic figures.<sup>25</sup>

The examination of whole-slide images scanned at a single focal plane may make it more difficult to identify mitotic figures because of the lack of fine focus and the limited image resolution.<sup>8</sup> Although scanning at multiple focal planes (z-scanning) is technically available and allows the whole-slide image to be viewed at different levels comparable to the fine-focus of light microscopy, it is currently not used for routine diagnosis due to the extended scanning time and the huge file size of the whole-slide image.<sup>8</sup> To reduce misidentification of mitotic figures, we also excluded all slides with inappropriate image quality, as mitotic figures are increasingly difficult to distinguish from artifacts such as overstained nuclei and necrosis throughout the entire slide.<sup>6</sup> Lastly, the specific tumor type, i.e. MCTs, may pose an additional challenge for identifying mitotic figures, as a large numbers of mast cell granules may mask nuclear details. Limited image resolution and lack of fine focus may have also affected area selection by pathologists for manual MCs as many participants noted additional difficulty in assessing the digital scans for mitoses in comparison to light microscopy.

The present results emphasize the need for further standardization for area selection in order to achieve higher reproducibility in MC. As discussed above, variations in focal density of mitotic figures throughout the tumor section may pose a severe problem for area selection and subsequently for accurate counting. Improvement options include repeated counts by the same or different pathologists with the purpose of uniformity or to choose the highest count.<sup>13,25,26,29</sup> Regardless of the increased time required, even various repeats would have only slightly increase the chance to find very rare events in the present cases. Alternatively, it has been proposed to increase the size of the area enumerated (for example to 20 hpf),<sup>10,27</sup> which would also increase time

investment and therefore would be rather problematic in a diagnostic setting. Although the latter approach has been suspected to improve reproducibility of the MC by a computer model of human breast cancer,<sup>10</sup> it would also change the MC towards the median value as demonstrated by the current study and, therefore, is inadequate to quantify the most mitotically active region. A new approach for standardization of the MC could be automated image analysis based on deep learning methods. Computerized analysis of a whole-slide image has the potential to reduce laborious tasks while minimizing inter-observer variability and maximizing reproducibility.<sup>8</sup> Several robust deep learning based algorithms have been developed to detect mitotic figures automatically in human breast cancer tissue,<sup>23,48,49</sup> and similar systems are being investigated for ccMCT.<sup>3,4</sup> However, fully automated mitotic activity estimations still have limitations for clinical applications as current algorithms do not have a high enough sensitivity and specificity (F1-score: 0.66 – 0.89).<sup>5,21</sup> Therefore, algorithms are not necessarily better at identifying individual mitotic figures in comparison to pathologists; however, they may improve histopathological quantification of mitotic figures by investigating every pixel within the whole-slide images and therefore reduce area-selection bias. Subsequently, a potential way to standardize MCs by computerization is a fully automated preselection of the 10-hpf-area with the highest mitotic activity within the entire tumor section (field of interest) as has been shown recently.<sup>3,4</sup> Future research has to determine to which degree a computer-assisted diagnosis system could support the pathologist's task in tumor grading in a diagnostic situation. As the present study proposes that poor reproducibility of the MC can be explained to some degree by inconsistent area selection, we propose that determining the MC could significantly benefit from further attempts to standardize area selection. Nevertheless we acknowledge that representative MCs with high prognostic value are furthermore dependent on many other factors such as rapid tissue fixation.<sup>6</sup> It is also

emphasized that further markers of cell proliferation (such as AgNOR and Ki-67) have excellent prognostic value and should be considered as supplemental methods for prognostication, especially in borderline cases.<sup>40</sup>

In conclusion, the present study showed that in ccMCTs the MC may vary tremendously between different tumor areas. Pathologists were often unable to determine the area of the highest MC in whole-slide images, which is especially true for MCTs that had a high range of MCs throughout the tumor. Regardless, detection rate for current cut-off values of manual MCs were generally very reproducible between all pathologists. Inaccuracies in determining the highest MC was only problematic for prognostication in a subset of ccMCTs that contained few areas with ground truth MCs above the cut-off value. Our findings highlight the need for further standardization when quantifying mitotic figures in scanned slides of ccMCT. Computer-assisted diagnosis systems using automated image analyses and deep learning approaches might be a promising solution to accurately and reproducibly identify the area with the highest mitotic activity within large tumor sections.

## Acknowledgements

None

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

## Literature

- 1 Al-Janabi S, van Slooten H-J, Visser M, Van Der Ploeg T, Van Diest PJ, Jiwa M. Evaluation of mitotic activity index in breast cancer using whole slide digital images. *PLoS One*. 2013;8:e82576.
- 2 Aubreville M, Bertram C, Klopfleisch R, Maier A. SlideRunner. In: Bildverarbeitung für die Medizin 2018, eds. Maier A, Deserno T, Handels H, Maier-Hain K, Palm C, Tolxdorff T, Springer Vieweg, Berlin, Heidelberg, 2018:309-314.
- 3 Aubreville M, Bertram CA, Klopfleisch R, Maier A. Field of interest proposal for augmented mitotic cell count: Comparison of two convolutional networks. Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: BIOIMAGING 2019;30-37
- 4 Aubreville M, Bertram CA, Marzahl C, et al. Field of interest prediction for computer-aided mitotic count. *arXiv preprint arXiv:190205414*. 2019;
- 5 Aubreville M, Krappmann M, Bertram C, Klopfleisch R, Maier A. A guided spatial transformer network for histology cell differentiation. Presented at the Eurographics Workshop on Visual Computing for Biology and Medicine 2017;21-25
- 6 Baak JP, Gudlaugsson E, Skaland I, et al. Proliferation is the strongest prognosticator in node-negative breast cancer: significance, error sources, alternatives and comparison with molecular prognostic markers. *Breast Cancer Res Treat*. 2009;115:241-254.
- 7 Bertram CA, Gurtner C, Dettwiler M, et al. Validation of digital microscopy compared with light microscopy for the diagnosis of canine cutaneous tumors. *Vet Pathol*. 2018;55:490-500.
- 8 Bertram CA, Klopfleisch R. The pathologist 2.0: an update on digital pathology in veterinary medicine. *Vet Pathol*. 2017;54:756-766.
- 9 Boiesen P, Bendahl P-O, Anagnostaki L, et al. Histologic grading in breast cancer: reproducibility between seven pathologic departments. *Acta Oncol*. 2000;39:41-45.
- 10 Bonert M, Tate AJ. Mitotic counts in breast cancer should be standardized with a uniform sample area. *BioMed Eng OnLine*. 2017;16:16-28.
- 11 Bostock D. The prognosis following surgical removal of mastocytomas in dogs. *J Small Anim Pract*. 1973;14:27-40.

Formatiert: Deutsch (Deutschland)

Formatiert: Deutsch (Deutschland)

Formatiert: Deutsch (Deutschland)

- 12 Collan Y, Kuopio T, Baak J, et al. Standardized mitotic counts in breast cancer evaluation of the method. *Pathol Res Pract.* 1996;192:931-941.
- 13 Dennis M, McSporran K, Bacon N, Schulman F, Foster R, Powers B. Prognostic factors for cutaneous and subcutaneous soft tissue sarcomas in dogs. *Vet Pathol.* 2011;48:73-84.
- 14 Edmondson E, Hess A, Powers B. Prognostic significance of histologic features in canine renal cell carcinomas: 70 nephrectomies. *Vet Pathol.* 2015;52:260-268.
- 15 Elston LB, Sueiro FA, Cavalcanti JN, Metzke K. Letter to the editor: the importance of the mitotic index as a prognostic factor for survival of canine cutaneous mast cell tumors: a validation study. *Vet Pathol.* 2009;46:362-364.
- 16 Gamer M, Lemon J, Fellows I, Singh P: Package 'irr': Various coefficients of interrater reliability and agreement. 2019
- 17 Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol.* 2012;8:23.
- 18 Horta RS, Lavallo GE, Monteiro LN, Souza MC, Cassali GD, Araújo RB. Assessment of canine mast cell tumor mortality risk based on clinical, histologic, immunohistochemical, and molecular features. *Vet Pathol.* 2018;55:212-223.
- 19 Kirpensteijn J, Kik M, Rutteman G, Teske E. Prognostic significance of a new histologic grading system for canine osteosarcoma. *Vet Pathol.* 2002;39:240-246.
- 20 Kiupel M, Webster J, Bailey K, et al. Proposal of a 2-tier histologic grading system for canine cutaneous mast cell tumors to more accurately predict biological behavior. *Vet Pathol.* 2011;48:147-155.
- 21 Li C, Wang X, Liu W, Latecki LJ. DeepMitosis: Mitosis detection via deep detection, verification and segmentation networks. *Med Image Anal.* 2018;45:121-133.
- 22 Loukopoulos P, Robinson W. Clinicopathological relevance of tumour grading in canine osteosarcoma. *J Comp Pathol.* 2007;136:65-73.
- 23 Malon C, Brachtel E, Cosatto E, et al. Mitotic figure recognition: agreement among pathologists and computerized detector. *Anal Cell Pathol* 2012;35:97-100.
- 24 McSporran K. Histologic grade predicts recurrence for marginally excised canine subcutaneous soft tissue sarcomas. *Vet Pathol.* 2009;46:928-933.
- 25 Meuten D, Moore F, George J. Mitotic count and the field of view area: time to standardize. *Vet Pathol.* 2016;53:7-9.
- 26 Meuten DJ. Appendix: Diagnostic schemes and algorithms. In: Tumors in domestic animals, ed. Meuten D, John Wiley & Sons, 2016:942 - 978.
- 27 Meyer JS, Alvarez C, Milikowski C, et al. Breast carcinoma malignancy grading by Bloom–Richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index. *Mod Pathol.* 2005;18:1067.
- 28 Meyer JS, Cosatto E, Graf HP. Mitotic index of invasive breast carcinoma: achieving clinically meaningful precision and evaluating tertial cutoffs. *Arch Pathol Lab Med.* 2009;133:1826-1833.
- 29 Mills S, Musil K, Davies J, et al. Prognostic value of histologic grading for feline mammary carcinoma: a retrospective survival analysis. *Vet Pathol.* 2015;52:238-249.
- 30 Patnaik A, Ehler W, MacEwen E. Canine cutaneous mast cell tumor: morphologic grading and survival time in 83 dogs. *Vet Pathol.* 1984;21:469-474.
- 31 Peña L, Andrés PD, Clemente M, Cuesta P, Perez-Alenza M. Prognostic value of histological grading in noninflammatory canine mammary carcinomas in a prospective study with two-year follow-up: relationship with clinical and histological characteristics. *Vet Pathol.* 2012;50:94-105.
- 32 Romansik E, Reilly C, Kass P, Moore P, London CA. Mitotic index is predictive for survival for canine cutaneous mast cell tumors. *Vet Pathol.* 2007;44:335-341.
- 33 Sabattini S, Bettini G. Grading cutaneous mast cell tumors in cats. *Vet Pathol.* 2019;56:43-49.
- 34 Sabattini S, Bettini G. Prognostic value of histologic and immunohistochemical features in feline cutaneous mast cell tumors. *Vet Pathol.* 2010;47:643-653.

Formatiert: Deutsch (Deutschland)

- 35 Sabattini S, Scarpa F, Berlato D, Bettini G. Histologic grading of canine mast cell tumor: is 2 better than 3? *Vet Pathol.* 2015;52:70-73.
- 36 Santos M, Correia-Gomes C, Santos A, de Matos A, Dias-Pereira P, Lopes C. Interobserver reproducibility of histological grading of canine simple mammary carcinomas. *J Comp Path.* 2015;153:22-27.
- 37 Sarli G, Benazzi C, Preziosi R, Salda LD, Bettini G, Marcato P. Evaluating mitotic activity in canine and feline solid tumors: standardizing the parameter. *Biotech Histochem.* 1999;74:64-76.
- 38 Schott CR, Tatierysky LJ, Foster RA, Wood GA. Histologic grade does not predict outcome in dogs with appendicular osteosarcoma receiving the standard of care. *Vet Pathol.* 2018;55:202-211.
- 39 Skaland I, van Diest PJ, Janssen EA, Gudlaugsson E, Baak JP. Prognostic differences of world health organization–assessed mitotic activity index and mitotic impression by quick scanning in invasive ductal breast cancer patients younger than 55 years. *Hum Pathol* 2008;39:584-590.
- 40 Sledge DG, Webster J, Kiupel M. Canine cutaneous mast cell tumors: A combined clinical and pathologic approach to diagnosis, prognosis, and treatment selection. *Vet J.* 2016;215:43-54.
- 41 Spangler W, Culbertson M, Kass P. Primary mesenchymal (nonangiomatous/nonlymphomatous) neoplasms occurring in the canine spleen: anatomic classification, immunohistochemistry, and mitotic activity correlated with patient survival. *Vet Pathol.* 1994;31:37-47.
- 42 Spangler W, Kass P. The histologic and epidemiologic bases for prognostic considerations in canine melanocytic neoplasia. *Vet pathol.* 2006;43:136-149.
- 43 Thompson J, Pearl D, Yager J, Best S, Coomber B, Foste R. Canine subcutaneous mast cell tumor: characterization and prognostic indices. *Vet Pathol.* 2011;48:156 - 168.
- 44 Tsuda H, Akiyama F, Kurosumi M, et al. Evaluation of the interobserver agreement in the number of mitotic figures breast carcinoma as simulation of quality monitoring in the japan national surgical adjuvant study of breast cancer (NSAS-BC) protocol. *Jpn J Cancer Res.* 2000;91:451-457.
- 45 Valli V, Kass P, Myint MS, Scott F. Canine lymphomas: association of classification type, disease stage, tumor subtype, mitotic rate, and treatment with survival. *Vet Pathol.* 2013;50:738-748.
- 46 van Diest PJ, Baak JP, Matze-Cok P, et al. Reproducibility of mitosis counting in 2,469 breast cancer specimens: results from the multicenter morphometric mammary carcinoma project. *Human Pathol.* 1992;23:603-607.
- 47 Vascellari M, Giantin M, Capello K, et al. Expression of Ki67, BCL-2, and COX-2 in canine cutaneous mast cell tumors: association with grading and prognosis. *Vet Pathol.* 2013;50:110-121.
- 48 Veta M, Van Diest PJ, Jiwa M, Al-Janabi S, Pluim JP. Mitosis counting in breast cancer: Object-level interobserver agreement and comparison to an automatic method. *PLoS one.* 2016;11:e0161286.
- 49 Veta M, Van Diest PJ, Willems SM, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med Image Anal.* 2015;20:237-248.

## Figure legends

**Figure 1.** Overview of the processing chain to derive on-slide, ground truth mitotic count distribution from annotation data. A) Original whole-slide image of a canine cutaneous mast cell tumor (hematoxylin and eosin [HE], case 20, group 3) with mitotic

figure annotations (green dots) and tumor boundary annotations (solid red line). B) Downsampled (reduction of the resolution) mitotic figure map derived from the mitotic figure annotations. C) Mitotic count map representing the mitotic count determined within a defined area of 2.37mm<sup>2</sup> (i.e., 10 hpf) for every possible center coordinate within the whole-slide image. Brighter colors show higher mitotic counts. D) Downsampled, binary tumor area map derived from the tumor boundary annotations. E) Valid tumor area map derived from the tumor area map excluding tissue in which the center coordinate of the 10-hpf-area window would be covered less than 95% by tumor tissue. Original tumor boundary is depicted as a dashed red line. F) Combined visualization of the whole-slide image (HE), the mitotic count map within the original tumor area (green overlay; higher opacity equals a higher mitotic count), and the valid tumor area indicated by a black line. The center of each 10-hpf-area window was required to be within the black line. In this example the mitotic count ranges from 0 (blue 10-hpf-area window) to 105 (red 10-hpf-area window).

**Figure 2 and 3.** Comparison of the three canine cutaneous mast cell tumor groups which were defined as: group 1 (uniformly low mitotic count [MC]), group 2 (borderline, with subgroups 1 and 2), and group 3 (mostly high MC). **Figure 2.** Ground truth percentage of 10-hpf-areas with a MC  $\geq 7$  relative to the total number of all possible 10-hpf-areas within the section. The three black broken vertical lines separate the defined groups and subgroups. The percentage of 10-hpf-areas with MC < 7 is the reciprocal of the indicated values. **Figure 3.** Comparison of the largest histopathological tumor diameter among the three tumor groups. The boxes show the first quartile, second quartile (median) and third quartile; the whiskers show the range (minimum and

maximum) of the data. There is a significant difference ( $p < 0.05$ ) between group 1 and 2 as well as group 1 and 3.

**Figure 4-9.** Distribution of ground truth mitotic counts (MC) among 28 histopathological sections of canine cutaneous mast cell tumors. **Figure 4, 6, and 8.** Box-and-whisker plots (boxes show the first, second and third quartile; whiskers show the minimum and maximum of the data) of the ground truth MCs distribution of individual cases, divided into group 1 (uniformly low MCs; Figure 4), group 2 (borderline MCs; Figure 6) and group 3 (mostly high MCs; Figure 8). The cut-off value for tumor prognostication ( $MC \geq 7$ ) is indicated by the red dashed line. Absolute numbers of the manual MC performed by eleven pathologists are indicated by the different symbols in the individual boxplots. **Figure 5, 7, and 9.** Exemplary cases of the three tumor groups including case 5 (Figure 5; low MC), 9 (Figure 7, borderline) and 28 (Figure 9, mostly high MC). The images show the whole-slide image (hematoxylin and eosin) of the respective case with tumor border annotations (red lines) and the ground truth heat map of the MC as a green overlay with a MC range indicated by different opacity (scale on the right side).

**Figure 10 and 11.** Forest plots of the manual mitotic count (MC) being above or equal to the upper half or upper 25%, respectively, of ground truth MC distribution comparing the eleven pathologists. Of each graph, the central symbol represents the result of the respective participant and the attached line represents the confidence interval (Wilson method with continuity correction). The left dotted lines demarcate the chance if the 10-hpf-areas were selected randomly (50% and 25%, respectively) and the right line indicates the ideal achievement (100%). **Figure 10.** Forest-plot of the MC being above



or equal the ground truth median. Chance in fully random area selection would be 50%.

**Figure 11.** Forest-plot of the MC being above or equal the ground truth third quartile. Chance in fully random area selection would be 25%.

**Figure 12 and 13.** Comparison of the mitotic density between the tumor periphery and tumor center in 28 canine cutaneous mast cell tumors. **Figure 12.** Example of the three different definitions of the border between the tumor periphery and center in which the mitotic density was determined. Whole-slide image (hematoxylin and eosin) of case 22 (mostly high MC) with the manually annotated tumor margins (red line). The blue line indicates the border circumscribing the periphery containing 25% and the center containing 75% of the total tumor area. The green line separates the tumor area in half. The cyan line marks the border between the 75% periphery area and the remaining 25% center area. The green overlay superimposed on the whole-slide image indicates the mitotic count distribution with higher opacities representing higher mitotic counts (scale on the right side). **Figure 13.** Box-and-whisker plots (0-100% interval) of the mitotic density of the tumor center and periphery for all cases combined with three different size definitions. Mitotic density was determined by mitotic activity index, which was defined as the number of mitotic figures in a specific tumor area (i.e. tumor periphery and center) divided by the size of that area. Two-sided t-tests did not show significant differences between the two tumor areas for any of the three size definitions.