Tissue heterogeneity contributes to suboptimal precision of WHO 2010 scoring criteria for Ki67 labeling index in a subset of neuroendocrine neoplasms of the pancreas

Łukasz Liszka

Department of Pathomorphology, Medical University of Silesia, Katowice, Poland

Supplementary Information:

Methods

Histopathological diagnoses

Grossing of the specimens was performed by this author, and it was based on very extensive sampling of tumor tissue for histology. Many neuroendocrine neoplasms (NEN) were embedded in total. Specimens were fixed by immersion in 4% buffered formaldehyde for 48-72 hours at room temperature. Tissue processing, paraffin embedding, and hematoxylin-eosin staining were performed in a routine manner.

In general, histopathological diagnoses were based on a reference source [1]. Pathological data for the majority of the samples were gathered for diagnostic purposes by the author. Other cases were fully re-examined for this study.

Ki67 staining procedure

Ki67 stains were performed using pre-diluted mouse monoclonal MIB-1 antibody (Dako, Glostrup, Denmark) using automated machines (Autostainer Plus, Autostainer Link 48, Dako) following recommendations provided by the vendor. Freshly cut 4-μm-thick sections were placed on adhesive glass slides (Menzel Gläser, Thermo Fisher, Braunschweig, Germany). Heat-induced antigen retrieval was performed using Envision FLEX Target Retrieval Solution, Low pH (97°C, 20 minutes) (Dako) in the PT Link module (Dako). Slides were incubated with primary antibody at room temperature for 20 minutes. For signal detection, the Envision FLEX High pH polymer detection system with diaminobenzidine (Dako) was used. Slides were counterstained with hematoxylin (Dako). Tonsil served as a positive control. For the negative control, primary antibody was omitted.

Slide digitization

Ki67 slides and corresponding hematoxylin-eosin sections were digitized using a slide scanner (Hamamatsu Photonics, Hamamatsu, Japan) using the 40x mode (0.23 μm/pixel) and evaluated using a medically certified display (NEC Display Solutions, Tokyo, Japan) and dedicated software (NDP.view2, Hamamatsu).

Statistical analysis

For parametric tests, Ki67 LI values were transformed, as recommended [2, 3, 4]. Firstly, 0.1% was added for each recorded Ki67 LI, then obtained values were transformed with a natural logarithm. This resulted usually in normally distributed values [2, 3, 4], as checked using Kolmogorov-Smirnov tests. The concordance between Ki67 LI obtained in sets of different numbers of cells and between different hot spots (HS) and cold spots (CS) was examined using Lin's concordance correlation coefficients (CCC). CCC were interpreted using criteria by McBride [5]. Pearson's correlation coefficients were calculated for completeness. Concordance was also examined by inspection of Bland-Altman plots, which were interpreted as in [6]. For comparisons between independent (between spots) and dependent (within spots) indices, unpaired and paired t-tests, respectively, were used.

Results of Ki67 immunostains were also recorded as an ordinal variable, i.e. a Ki67-LI-based grade. In general, rules of grading provided by ENETS 2006 [7] and WHO 2010 [1] were followed. It is not known how to classify samples with Ki67 LI between 2% and 3%: some experts proposed that the index above 2% is sufficient to diagnose grade G2 [8, 9], while others proposed that samples with indices at least 2.5% [10, 11] or at least 3% [12, 13] should be used for establishing the G2 category. In this study, the latter approach (3% cut-off value) was used, in agreement with North American Neuroendocrine Tumor Society 2013 [14] and National Comprehensive Cancer Network 2016 [15] guidelines. As stated in the main text, cases with an index below 3% were recorded as G1, cases with an index between 3% and 20% were coded as G2, and cases with an index above 20% were coded as G3. As mentioned earlier, mitotic indices were not considered for grading. Concordance between grades obtained in sets of different numbers of cells or in different spots was described using weighted κ values with linear weights. κ values were interpreted following criteria by Landis and Koch [16]. McNemar's tests with continuity correction and Spearman rank correlation coefficients were used for comparisons between dependent ordinal variables.

Based on results of examination of Ki67 LI in HS-A, 5 subgroups of NET were distinguished: (1) cases with indices below 3% as measured in 500 cells and in 2000 cells (G1 subgroup), (2) cases with indices between 3% and 20% as measured in 500 cells, but below 3% as measured in 2000 cells (G1.5 subgroup), (3) cases with indices between 3% and 20% as measured in 500 cells and in 2000 cells (G2 subgroup), (4) cases with indices above 20% as measured in 500 cells, but between 3% and 20% as measured in 2000 cells (G2.5 subgroup), and (5) cases with indices above 20% as measured in 500 cells and in 2000 cells (G3 subgroup). Clinico-pathological characteristics between these subgroups were compared using the Mann-Whitney U test for continuous variables, and the χ^2 test and Fisher's test for comparisons of nominal or ordinal variables in $2 \times n$ and 2×2 contingency tables, respectively.

For documentation of utility of Ki67 LI as a predictor of regional lymph node metastasis, positive

and negative likelihood ratios, diagnostic odds ratios, Youden's statistics, and areas under receiver-operating characteristics curves were calculated. Likelihood ratios were interpreted as in [17]. Areas under receiver-operating characteristics curves were interpreted as cited in [18].

The number of samples included in this study was sufficient to detect the odds ratio of 10 and 20% percentage of disagreement in comparison of G1/G2 grading categories when counting 500 cells vs. 2000 cells in HS (McNemar's test, power 0.8). This required 57 pairs of observations in total and 9 pairs of disagreement. The large odds ratio value was justified by the assumption that the probability of identification of a tumor which would be diagnosed as G1 in 500 cells but as G2 in 2000 cells was low.

Statistical significance was set at an alpha value of 0.05 (two-sided). No adjustments for multiple testing were applied. Statistical analyses and figures were done using Statistica 12 (Dell Software, Tulsa, OK, USA), Winpepi [19], and Gene-E [20].

Working hypotheses

Working hypotheses were: (1) Grading of pancreatic NEN is consistent irrespective of number of counted cells in HS, at least within limits provided by the WHO 2010 guidelines. (2) Selection of suboptimal HS (i.e. not HS with the highest Ki67 LI in tissue section) for counting does not necessarily result in under-grading. (3) Counting of large number of cells for Ki67 LI *not* in HS (in this study: in the area with subjectively the lowest Ki67 LI, CS) usually still allows adequate (i.e. identical with Ki67-LI-based grade in HS) grading.

Supplementary Tables

Table S1. Guidelines for assessment of Ki67 LI in neuroendocrine neoplasms of the pancreas

European Neuroendocrine Tumor Society 2006 guidelines [7]:
"The Ki67 index should be assessed in 2,000 tumor cells in areas where the highest nuclear labeling is observed (often but not exclusively at the tumor periphery)".
European Neuroendocrine Tumor Society 2009 guidelines [21]:
"To determine Ki67 (MIB1) labeling index, 100 tumor cells have to be assessed in a hot-spot area."; "In case the Ki67 positivity is unevenly distributed, several tumor areas should be evaluated".
World Health Organization 2010 guidelines [1]:
"The grading requires mitotic count (…) and Ki67 index using the MIB antibody as a percentage of 500-2000 cells counted in areas of strongest nuclear labeling ("hot spots")".
"Multidisciplinary team of physicians interested in NETs" 2010 guidelines [22]:
"Eyeballed estimate of the labeling percentage was agreed to be the only method that could be strongly advocated at present. However, there was a recognition of many shortcomings of this approach"; "The group recommended to count the most densely staining regions ("hot spots") and to count a variety of areas with the tumor; it was specifically noted that counting of random areas or single regions is inadequate".
North American Neuroendocrine Tumor Society 2013 guidelines [14]:
threshold Ki67 LI values for grading were given but without description of counting methodology
National Comprehensive Cancer Network 2016 guidelines [15]:
"Ki67 index is reported as the percentage of positive tumor cells in the area of highest nuclear labeling. Although recommendation have been to count 2000 tumor cells in order to determine the Ki67 index, this is not practical in routine clinical practice. It is therefore currently acceptable to estimate the labeling index, despite the recognition that estimation is subject to limitations in reproducibility".
College of American Pathologists 2016 guidelines [23]:
"Ki67 index is reported as percent positive tumor cells in area of highest nuclear labeling, although the precise method of assessment has not been standardized. It has been recommended that 500 to 2000 tumor cells be counted to determine the Ki67 index".
European Neuroendocrine Tumor Society 2016 guidelines [24]:
"P-NETs should be classified and graded using the current WHO 2010 classification and grading system".

Table S2. Clinico-pathological data of the study cases

	Neuroendocrine tumors (n = 71)	Neuroendocrine carcinomas (n = 6)
Age (median, range)	59 (19-78)	63 (38-68)
Sex (Female : Male)	35 : 36	4 : 2
Specimen:		
1) Resection specimen	70	4
Enucleation	13	0
Partial pancreatectomy	55	4
Total pancreatectomy	2	0
2) Incisional biopsy of primary tumor	1	2
Tumor localization:		
Head	28	5
Body	7	1
Tail	32	0
Entire pancreas (multiple tumors)	1	0
Not known	3	0
Histopathological subtype:	–	
Small cell		1
Large cell		5*
pM stage:		
cM0	59	3
pM1	11	3
cM1	1	0
Tumor diameter (median, range, in mm):**	28 (6-140)	48 (32-50)
ENETS pT stage [7]:**		
pT1	26	0
pT2	16	0
pT3	25	4
pT4	3	0
pN stage:**		
pN0	30	2
pN1	24	2
pNx	16	0
Non-ischemic tumor necrosis	10	6
Lymph-vascular invasion**	34	4
Perineural invasion**	28	4
Chromogranin A expression	71	4
Synaptophysin expression	71	6

** two cases were diagnosed as mixed adenoneuroendocrine carcinomas*
*** in resected cases (n = 74)*

Table S3. Concordance between Ki-67 LI in hot spots and in cold spots in neuroendocrine carcinomas[1]

		HS-A (100 cells)	HS-A (500 cells)	HS-A (1000 cells)	HS-A (2500 cells)	HS-B (2000 cells)	HS-C (2000 cells)	CS (2000 cells)	HS-(A + B + C) – 500 cells in each (1500 cells in total)
HS-A (2000 cells)	t-test (p-value)	0.275[2]	0.165[2]	0.177[2]	0.085[2]	0.387[3]	0.472[3]	0.112[3]	0.264[2]
	Pearson's R	0.437 (p = 0.343)	0.981 (p = 0.000)	0.978 (p = 0.000)	0.999 (p = 0.000)	0.887 (p = 0.019)	0.962 (0.002)	0.896 (p = 0.016)	0.959 (p = 0.002)
	Lin's CCC	0.373 (95% CI: from –0.36 to 0.82)*	0.968 (95% CI: from 0.82 to 1)***	0.918 (95% CI: from 0.71 to 0.98)**	0.997 (95% CI: from 0.98 to 1)****	0.391 (95% CI: from 0.08 to 0.63)*	0.856 (95% CI: 0.47-0.97)*	0.388 (95% CI: from –0.003 to 0.68)*	0.882 (95% CI: from 0.60 to 0.97)*

[1] *calculations made on transformed data*
[2] *paired t-test*
[3] *unpaired t-test*
* *poor agreement*
** *moderate agreement*
*** *substantial agreement*
**** *almost perfect agreement*
CCC – concordance correlation coefficient; CI – confidence interval; CS – cold spot; HS – hot spot

Table S4. Concordance of Ki-67-LI-based grade in neuroendocrine tumors (in 2000 cells) between hot spots and cold spots

		HS-B (2000 cells)			HS-C (2000 cells)			CS (2000 cells)			HS - (A+B+C) – 500 cells in each HS (1500 cells in total)		
		G1	G2	G3	G1	G2	G3	G1	G2	G3	G1	G2	G3
HS-A (2000 cells)	G1	30	0	0	30	0	0	30	0	0	22	8	0
	G2	10	26	0	12	24	0	30	6	0	1	35	0
	G3	0	3	2	0	4	1	4	1	0	0	0	5
McNemar's test		p = 0.001			p = 0.000			p = 0.000			p = 0.046		
Percentage of agreement		58/71 (81.7%)			55/71 (77.5%)			36/71 (50.7%)			62/71 (87.3%)		
Weighted κ		κ = 0.70 (95% CI: 0.55-0.84)* (p = 0.000)			κ = 0.62 (95% CI: 0.47-0.77)* (p = 0.000)			κ = 0.13 (95% CI: 0.04-0.23)** (p = 0.009)			κ = 0.79 (95% CI: 0.66-0.92)* (p = 0.000)		
Weighted κ significantly above 0.6		p = 0.000			NS			NS			p = 0.003		
Spearman's rho		R = 0.777 (p = 0.000)			R = 0.737 (p = 0.000)			R = 0.279 (p = 0.018)			R = 0.804 (p = 0.000)		

* *substantial agreement*
** *slight agreement*
CI – confidence interval; CS – cold spot; HS – hot spot; NS – not significant

Table S5. Subgroups of resected cases of neuroendocrine tumors distinguished based on Ki67 LI (hot spot A; 2000 cells, n = 70).

	G1 subgroup (n = 19)	G1.5 subgroup (n = 11)	G2 subgroup (n = 33)	G2.5 subgroup (n = 2)	G3 subgroup (n = 5)	G1.5 vs. G1 p-value	G1.5 vs. G2 p-value
Grade based on 500 cells in hot spot A	1	2	2	3	3		
Grade based on 2000 cells in hot spot A	1	1	2	2	3		
Age (median, range)	59 (33-78)	64 (41-70)	59 (19-77)	38 and 68	46 (35-65)	0.898	0.244
Sex (Female : Male)	8 : 11	8 : 3	16 : 17	0 : 2	3 : 2	0.142	0.294
Tumor localization:						0.863	0.517
Head	9 (47.4%)	6 (54.%)	10 (30.3%)	1 (50%)	2 (40%)		
Body	2 (10.5%)	1 (9.1%)	4 (12.1%)	0	0		
Tail	6 (31.6%)	4 (36.4%)	18 (54.5%)	1 (50%)	3 (60%)		
Entire pancreas (multiple tumors)	1 (5.3%)	0	0	0	0		
Not known	1 (5.3%)	0	1 (3.0%)	0	0		
pM stage:						0.126	1
cM0	19 (100%)	9 (81.8%)	26 (78.8%)	1 (50%)	3 (60%)		
pM1	0	2 (18.2%)	7 (21.2%)	1 (50%)	1 (20%)		
cM1	0	0	0	0	1 (20%)		
Tumor diameter (median, range, in mm)	14 (6-60)	22 (6-70)	38 (6-90)	12 and 55	35 (25-140)	0.426	0.110
ENETS pT stage [7]:						0.316 *1	0.241 *0.081
pT1	13 (68.4%)	5 (45.4%)	7 (21.2%)	1 (50%)	0		
pT2	2 (10.5%)	4 (36.4%)	9 (27.3%)	0	1 (20%)		
pT3	3 (15.8%)	2 (18.2%)	16 (48.5%)	1 (50%)	3 (60%)		
pT4	1 (5.3%)	0	1 (3.0%)	0	1 (20%)		
pN stage:						0.629	0.383
pN0	12 (63.2%)	5 (45.4%)	13 (39.4%)	0	0		
pN1	2 (10.5%)	2 (18.2%)	13 (39.4%)	2 (100%)	5 (100%)		
pNx	5 (26.3%)	4 (36.4%)	7 (21.2%)	0	0		
Non-ischemic tumor necrosis	0	0	6 (18.2%)	0	4 (80%)	NA	0.075
Lymph-vascular invasion	6 (31.6%)	4 (36.4%)	17 (51.5%)	2 (100%)	5 (100%)	1.0	0.494
Perineural invasion	6 (31.6%)	5 (45.4%)	11 (33.3%)	2 (100%)	4 (80%)	0.696	0.492

** pT1 + pT2 vs. pT3 + pT4 (Fisher's exact tests)*
NA – cannot be calculated

Table S6. Diagnostic performance of Ki-67-LI-based grade in neuroendocrine tumors as predictor of regional lymph node metastasis[1, 2]

	HS-A (100 cells)	HS-A (500 cells)	HS-A (1000 cells)	HS-A (2000 cells)	HS-A (2500 cells)	HS-B (2000 cells)	HS-C (2000 cells)	CS (2000 cells)	HS-(A+B+C) (1500 cells)
Positive likelihood ratio	1.036	1.443	1.636	1.716	2.028	2.788	2.868	6.692	1.575
Negative likelihood ratio	0	0.372	0.319	0.418	0.372	0.319	0.406	0.797	0.186
Diagnostic odds ratio	NA	3.879	5.128	4.105	5.452	8.740	7.064	8.394	8.468
Youden's statistics	0.034	0.260	0.329	0.321	0.390	0.493	0.451	0.196	0.337

[1] This evaluation was limited to patients with neuroendocrine tumors, who were treated with pancreatic resection, have at least 1 regional lymph node detected in the pancreatectomy specimen or submitted in a separate container during the procedure, and did not have synchronous distant metastases (n = 42).
[2] G2 or G3 status interpreted as positive test result, G1 status interpreted as negative test result, reference value: pN status – pN0 vs. pN1
CS – cold spot, HS – hot spot, NA – cannot be calculated

Table S7. Area under receiver-operating characteristics (AUROC) curves: Ki67 LI as predictor of regional lymph node metastasis[1, 2]

	AUROC	P[3]
HS-A (100 cells)	0.764 (95% CI: 0.58-0.95)	0.005
HS-A (500 cells)	0.751 (95% CI: 0.58-0.92)	0.005
HS-A (1000 cells)	0.768 (95% CI: 0.60-0.94)	0.002
HS-A (2000 cells)	0.771 (95% CI: 0.60-0.94)	0.002
HS-A (2500 cells)	0.777 (95% CI: 0.61-0.94)	0.001
HS-B (2000 cells)	0.763 (95% CI: 0.59-0.63)	0.002
HS-C (2000 cells)	0.775 (95% CI: 0.61-0.94)	0.001
CS (2000 cells)	0.635 (95% CI: 0.44-0.84)	0.185
HS (A+B+C) (1500 cells)	0.768 (95% CI: 0.60-0.94)	0.002

[1] This evaluation was limited to patients with neuroendocrine tumors, who were treated with pancreatic resection, have at least 1 regional lymph node detected in the pancreatectomy specimen or submitted in a separate container during the procedure, and did not have synchronous distant metastases (n = 42).
[2] G2 or G3 status interpreted as positive test result, G1 status interpreted as negative test result, reference value: pN status – pN0 vs. pN1
[3] these p values describe statistically significant differences between calculated AUROC and AUROC = 0.5. The latter value describes a diagnostic value of a coin toss.
CI – confidence interval; CS – cold spot; HS – hot spot

Table S8. Concordance of Ki67-LI-based grade in neuroendocrine tumors in hot spot A related to the number of counted cells – 5% Ki67 labeling index as a cut-off value for G1/G2 distinction

		HS-A (100 cells)			HS-A (500 cells)			HS-A (1000 cells)			HS-A (2500 cells)		
		G1	G2	G3	G1	G2	G3	G1	G2	G3	G1	G2	G3
HS-A (2000 cells)	G1	19	24	0	31	12	0	39	4	0	42	1	0
	G2	0	14	9	1	20	2	1	21	1	0	23	0
	G3	0	0	5	0	0	5	0	0	5	0	0	5
McNemar's test		p = 0.000			p = 0.002			p = 0.221			p = 1		
Percentage of agreement		38/71 (53.5%)			56/71 (78.9%)			65/71 (91.5%)			70/71 (98.6%)		
Weighted κ		κ = 0.41 (95% CI: 0.27-0.55)* (p = 0.000)			κ = 0.68 (95% CI: 0.54-0.83)** (p = 0.000)			κ = 0.87 (95% CI: 0.76-0.97)*** (p = 0.000)			κ = 0.98 (95% CI: 0.93-1.00)*** (p = 0.000)		
Weighted κ significantly above 0.6		NS			p = 0.136			p = 0.000			p = 0.000		
Spearman's rho		R = 0.701 (p = 0.000)			R = 0.742 (p = 0.000)			R = 0.881 (p = 0.000)			R = 0.976 (p = 0.000)		

** moderate agreement*
*** substantial agreement*
**** almost perfect agreement*
CI – confidence interval; HS – hot spot; NS – not significant

Table S9. Concordance of Ki-67-LI-based grade in neuroendocrine tumors (in 2000 cells) between hot spots and cold spots – 5% Ki67 labeling index as a cut-off value for G1/G2 distinction

		HS-B (2000 cells)			HS-C (2000 cells)			CS (2000 cells)			HS-(A+B+C) – 500 cells in each (1500 cells in total)		
		G1	G2	G3	G1	G2	G3	G1	G2	G3	G1	G2	G3
HS-A (2000 cells)	G1	43	0	0	43	0	0	43	0	0	37	6	0
	G2	6	17	0	8	15	0	22	1	0	3	20	0
	G3	0	3	2	0	4	1	5	0	0	0	0	5
McNemar's test		p = 0.001			p = 0.001			p = 0.000			p = 0.505		
Percentage of agreement		62/71 (87.3%)			59/71 (83.1%)			44/71 (62.0%)			62/71 (87.3%)		
Weighted κ		κ = 0.77 (95% CI: from 0.64 to 0.90)* (p = 0.000)			κ = 0.68 (95% CI: from 0.48 to 0.81)* (p = 0.000)			κ = 0.04 (95% CI: from –0.03 to 0.11)** (p = 0.106)			κ = 0.79 (95% CI: from 0.66 to 0.93)* (p = 0.000)		
Weighted κ significantly above 0.6		p = 0.005			NS			NS			p = 0.002		
Spearman's rho		R = 0.848 (p = 0.000)			R = 0.800 (p = 0.000)			R = 0.128 (p = 0.286)			R = 0.787 (p = 0.000)		

** substantial agreement*
*** slight agreement*
CI – confidence interval; CS – cold spot; HS – hot spot; NS – not significant

Table S10. Diagnostic performance of Li-67-LI-based grade in neuroendocrine tumors as a predictor of regional lymph node metastasis – 5% Ki67 labeling index as a cut-off value for G1/G2 distinction[1, 2]

	HS-A (100 cells)	HS-A (500 cells)	HS-A (1000 cells)	HS-A (2000 cells)	HS-A (2500 cells)	HS-B (2000 cells)	HS-C (2000 cells)	CS (2000 cells)	HS-(A+B+C) (1500 cells)
Positive likelihood ratio	1.174	1.673	2.231	3.187	2.788	2.974	3.569	0	2.510
Negative likelihood ratio	0.669	0.525	0.446	0.304	0.319	0.485	0.465	1.036	0.425
Diagnostic odds ratio	1.755	3.187	5.002	10.484	8.740	6.132	7.675	0	5.906
Youden's statistics	0.114	0.279	0.382	0.528	0.493	0.408	0.443	–0.034	0.416

[1] *This evaluation was limited to patients with neuroendocrine tumors, who were treated with pancreatic resection, have at least 1 regional lymph node detected in the pancreatectomy specimen or submitted in a separate container during the procedure, and did not have synchronous distant metastases (n = 42)*
[2] *G2 or G3 status interpreted as positive test result, G1 status interpreted as negative test result, reference value: pN status – pN0 vs. pN1*
CS – cold spot; HS – hot spot

Table S11. Heterogeneity of Ki-67-LI-based grade in neuroendocrine tumors across hot spots

Cases with the same grade as counted in 100, 500 and 2000 cells:				
Grade	in hot spot A	in hot spot B	in hot spot C	in all hot spots
G1	2	8	7	1
G2	27	20	22	12
G3	5	2	1	1
Total	34/71 (47.9%)	30/71 (42.2%)	30/71 (42.2%)	14/71 (19.7%)
Cases with different grade as counted in 100, 500 and 2000 cells:				
–	37/71 (52.1%)	41/71 (51.8%)	41/71 (51.8%)	57/71 (80.3%)

Supplementary Figures

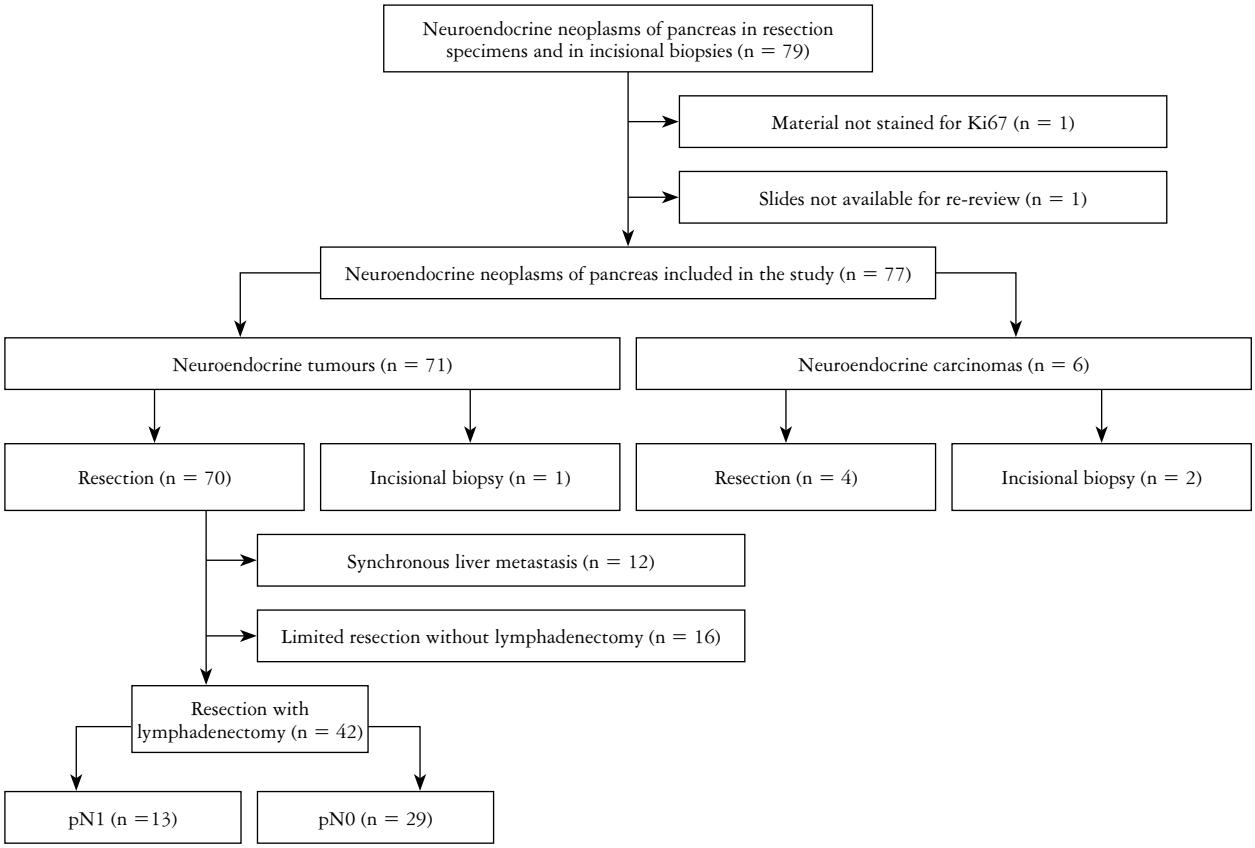


Fig. S1. Flow chart describing study population

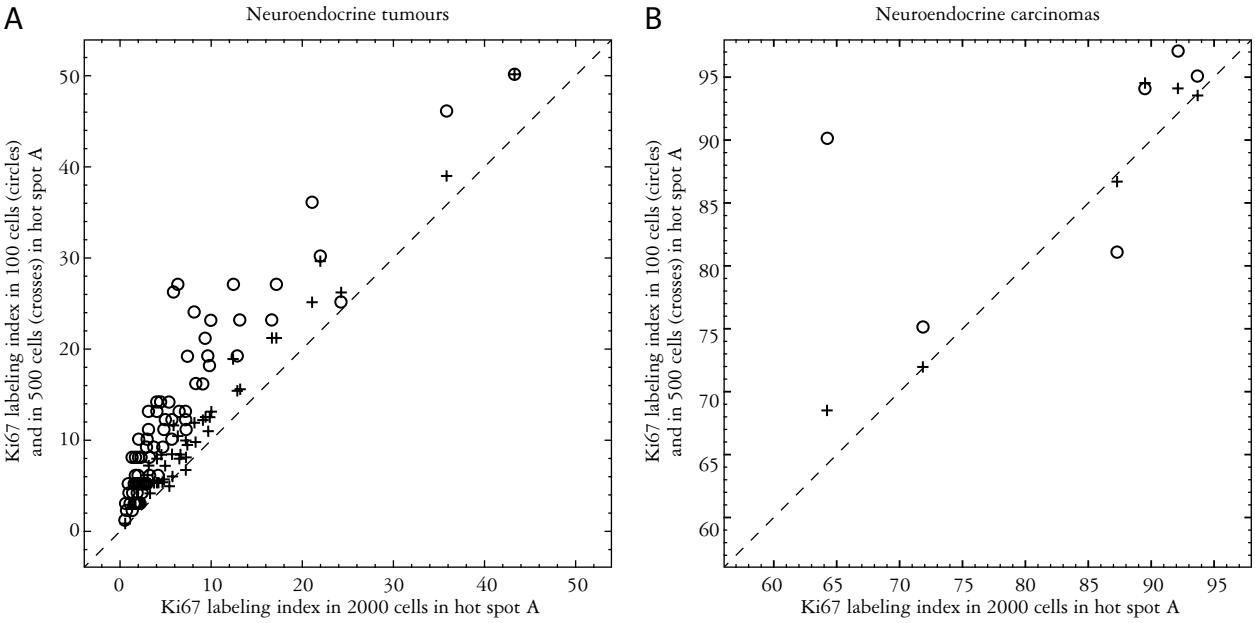
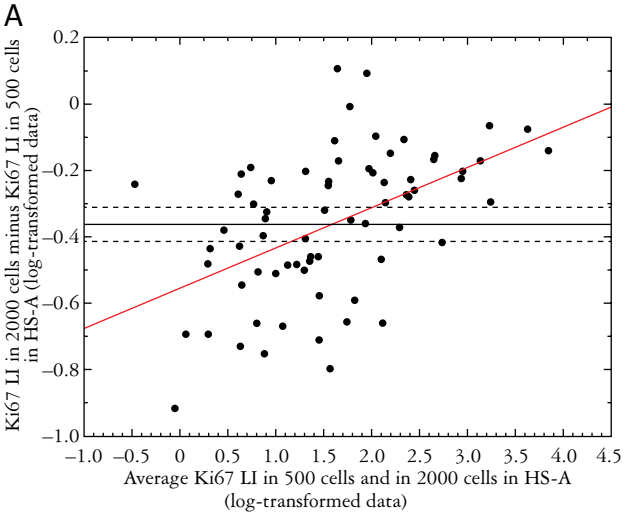
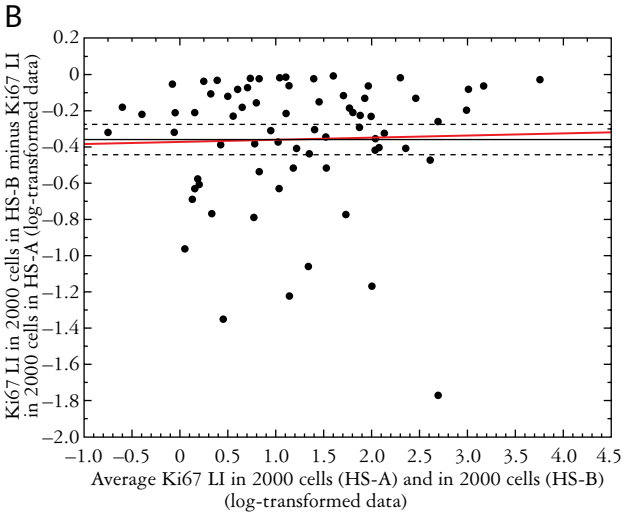


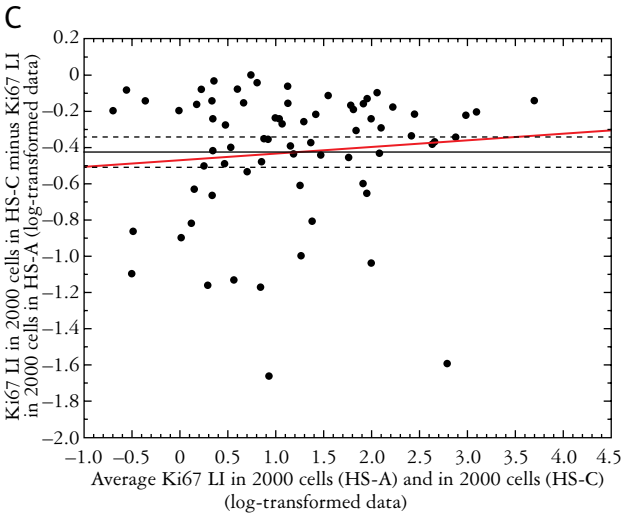
Fig. S2a, S2b. The relationship between raw Ki-67 LI scored in 100 cells and in 500 cells vs. 2000 cells in hot spot A in neuroendocrine tumors (Fig. S2a) and in neuroendocrine carcinomas (Fig. S2b)



Bland-Altman plot showing Ki-67 LI in 500 cells vs. 2000 cells (hot spot A) – transformed data. The mean difference between transformed Ki67 LI in 2000 and in 500 cells was –0.362 (solid black line). This corresponded to a geometric mean of the ratios (Ki67 LI in 2000 cells/Ki67 LI in 500 cells) of 0.696. The 95% CI for the mean (dotted lines) was: from –0.31 to –0.41. The confidence interval did not include 0, suggesting fixed bias. Pearson's correlation coefficient (0.50) (red solid line) was significantly different from 0 (p = 0.000), suggesting proportional bias



Bland-Altman plot showing Ki67 LI in 2000 cells in hot spot A vs. 2000 cells in hot spot B – transformed data.The mean difference between transformed Ki67 LI in 2000 in hot spot A and in 2000 cells in hot spot B was -0.356 (solid black line). This corresponded to a geometric mean of the ratios (Ki67 LI in hot spot B/Ki67 LI in hot spot A) of 0.700. The 95% CI for the mean (dotted lines) was: from –0.27 to –0.44. The confidence interval did not include 0, suggesting fixed bias. Pearson's correlation coefficient (0.03) (red solid line) was not significantly different from 0 (p = 0.791), suggesting no proportional bias



Bland-Altman plot showing Ki67 LI in 2000 cells in hot spot A vs. 2000 cells in hot spot C – transformed data. The mean difference between transformed Ki67 LI in 2000 in hot spot A and in 2000 cells in hot spot B was –0.425 (solid black line). This corresponded to a geometric mean of the ratios (Ki67 LI in hot spot C/Ki67 LI in hot spot A) of 0.654. The 95% CI for the mean (dotted lines) was: from –0.34 to –0.51. The confidence interval did not include 0, suggesting fixed bias. Pearson's correlation coefficient (0.10) (red solid line) was not significantly different from 0 (p = 0.422), suggesting no proportional bias

Fig. S3. Bland-Altman plot showing Ki67 LI in 500 cells vs. 2000 cells (transformed data) – hot spot A (Fig. S3A), Ki67 LI in 2000 cells (transformed data) – hot spot A vs. hot spot B (Fig. S3B), Ki67 LI in 2000 cells (transformed data) - hot spot A vs. hot spot C (Fig. S3C)

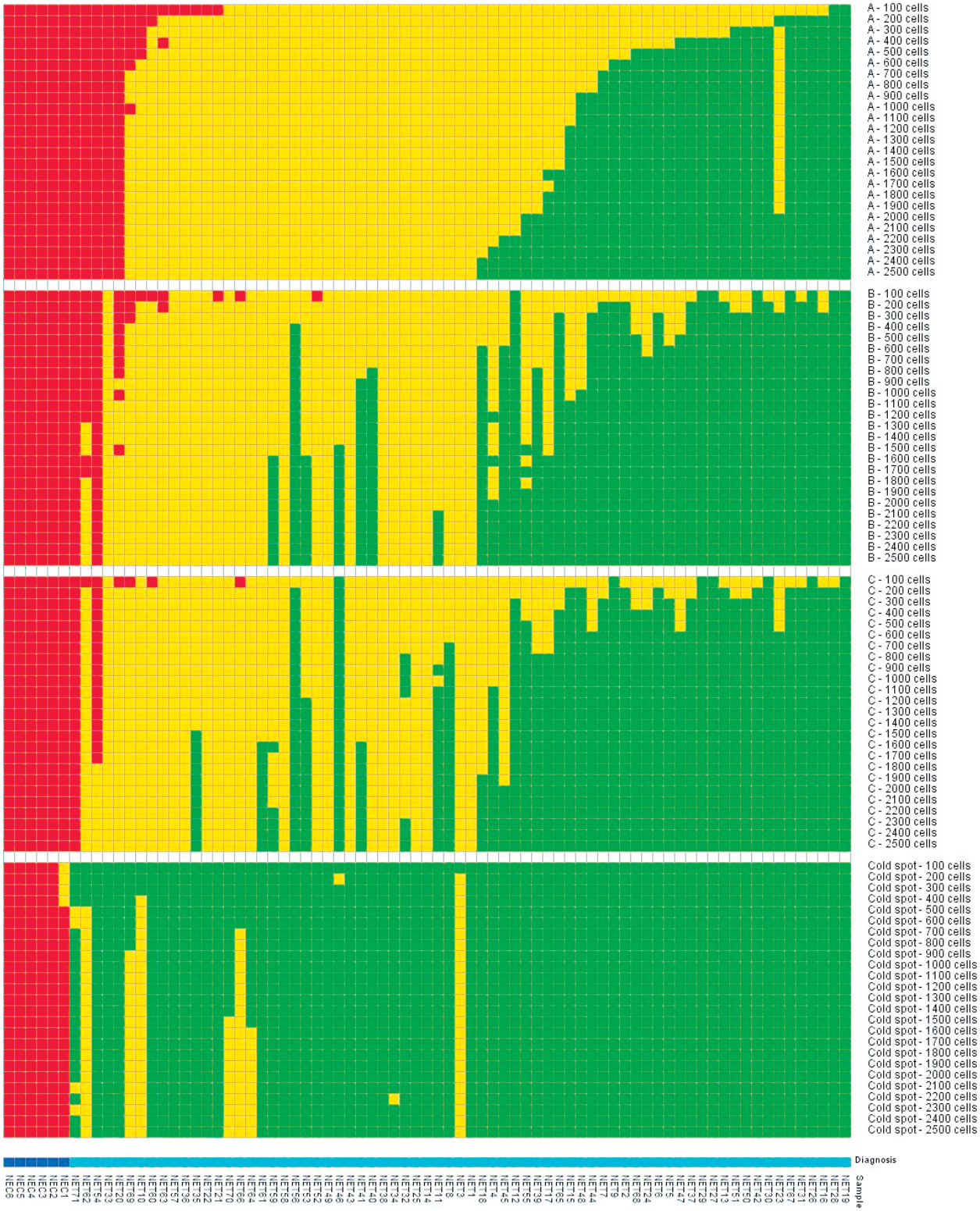
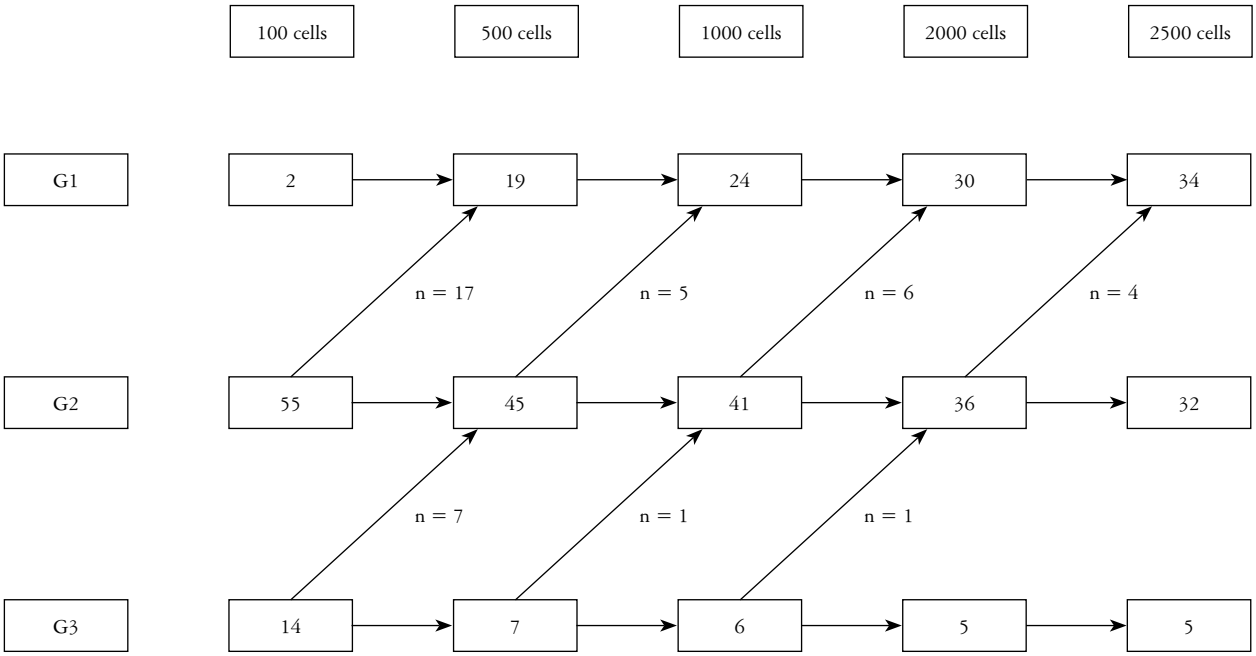


Fig. S4. Relationship between Ki67-LI-based grade proportions and number of examined cells in hot spots and in cold spots in the entire study population. A, B, C – hot spots. G1, G2, and G3 cases are presented in green, yellow, and red, respectively

A

Relationship between grade proportions and number of examined cells in hot spot A in neuroendocrine tumors



B

Relationship between grade proportions and number of examined cells in hot spot B in neuroendocrine tumors

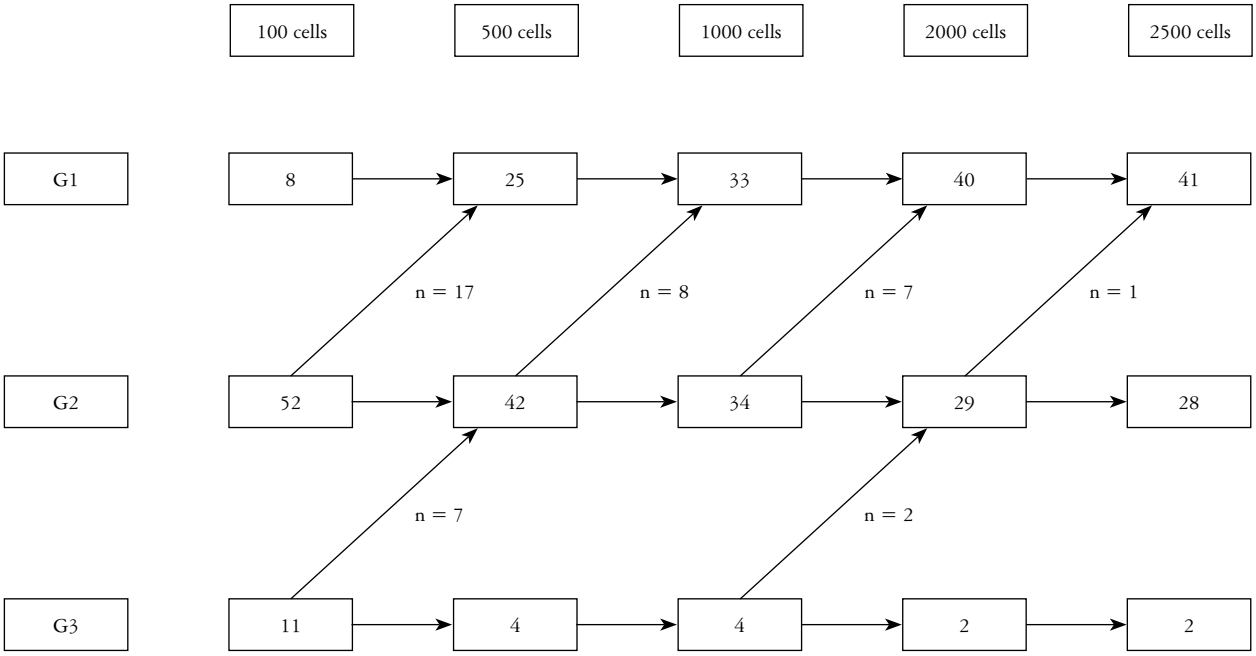
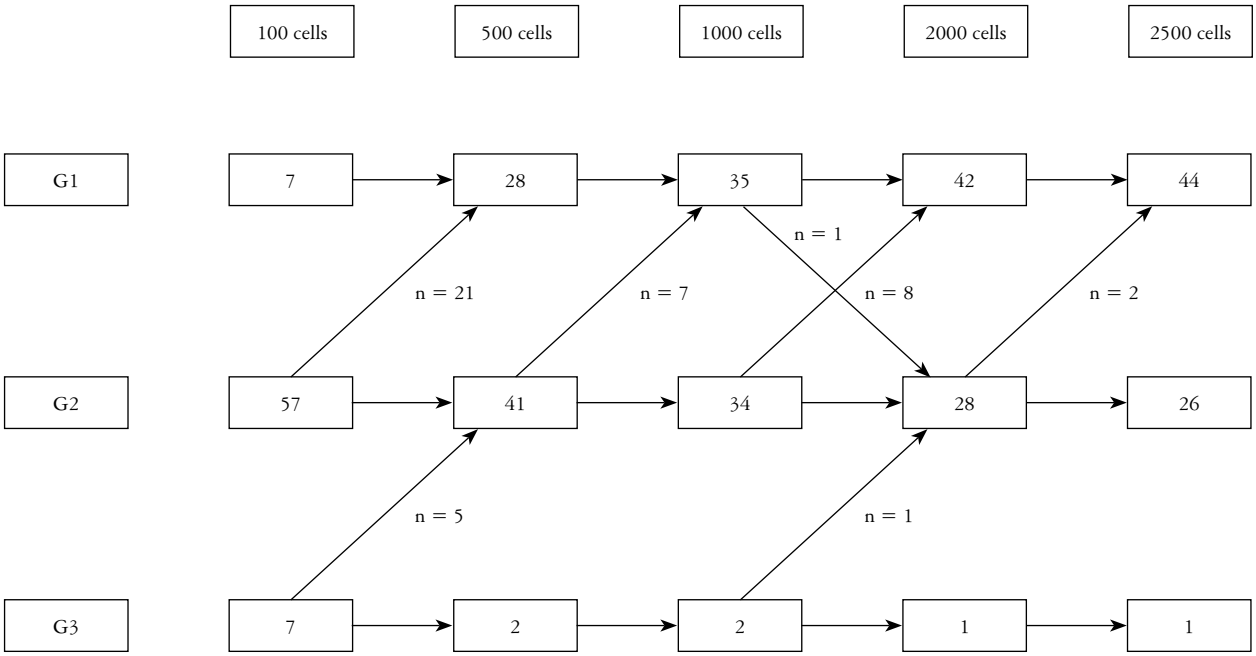


Fig. S5. Relationship between Ki67-LI-based grade proportions and number of examined cells in neuroendocrine tumors (flow diagrams) in hot spot A (A), in hot spot B (B), in hot spot C (C), and in cold spot (D)

C

Relationship between grade proportions and number of examined cells in hot spot C in neuroendocrine tumors



D

Relationship between grade proportions and number of examined cells in cold spot in neuroendocrine tumors

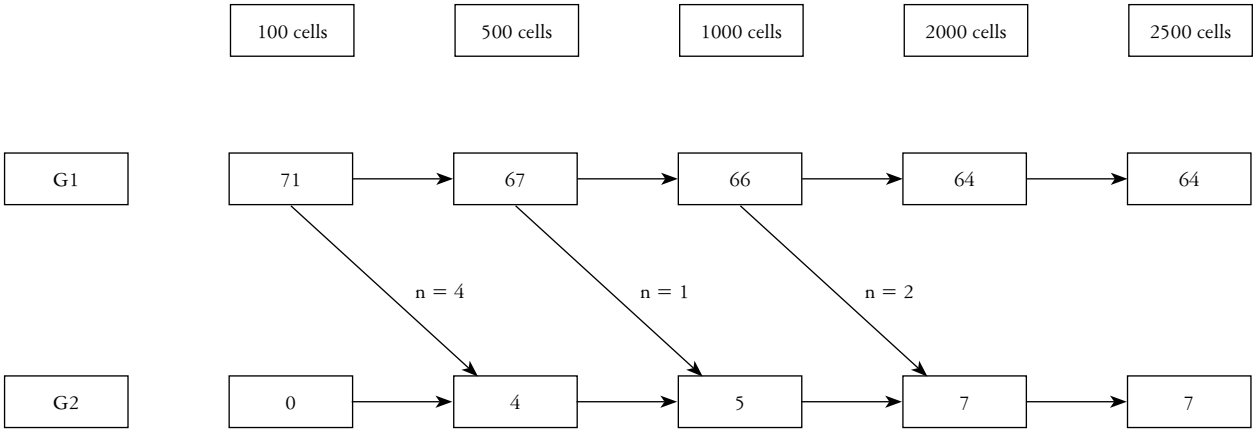


Fig. S5. Cont.

Study limitations

There were several limitations of the present study: (1) Ki67 scoring was performed by a single observer, so the inter-rater variability was not examined. (2) Ki67 scoring was performed manually rather than using digital image analysis. HS and CS were detected subjectively. However, manual counting of Ki67-positive cells in a printed image is a reference method for Ki67 LI assessment in NEN of the pancreas [25]. (3) Although consecutive NEN samples were included in this study, referral bias cannot be excluded. (4) The number of studied NEC cases was small, so conclusions on NEC are of limited reliability. Pancreatic NEC is a very rare disease. Ki67 LI in examined NEC cases was high – NEC with relatively lower Ki67 LI exist, but are even rarer [26]. (5) A single tissue block was examined for Ki67 LI in each case, but this is possibly enough [27]. (6) Data on functional status of NEN were not included, since they were missing for some earlier cases. However, functionality may be less important than previously thought [28]. (7) Follow-up data were not included. Many cases were relatively recent, so survival analysis would not be informative. (8) The immunohistochemistry protocol for this study included antigen retrieval in low pH buffer, as recommended by the antibody manufacturer. It was recognized by the author that according to standardization initiatives [29] the use of high pH buffer may give better results [30]. (9) Automated tools for comprehensive assessment of stain heterogeneity were developed [31, 32], but they were not available for the study.

References for Supplementary Information

1. Rindi G, Arnold R, Bosman FT, et al. Nomenclature and classification of neuroendocrine neoplasms of the digestive system. In: WHO classification of tumours of the digestive system. Bosman FT, Carneiro F, Hruban RH, Theise ND (eds). 4th ed. International Agency for Research on Cancer, Lyon 2010; 13-14.
2. Polley MY, Leung SC, Gao D, et al. An international study to increase concordance in Ki67 scoring. Mod Pathol 2015; 28: 778-786.
3. Dowsett M, Nielsen TO, A'Hern R, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. J Natl Cancer Inst 2011; 103: 1656-1664.
4. Polley MY, Leung SC, McShane LM, et al. An international Ki67 reproducibility study. J Natl Cancer Inst 2013; 105: 1897-1906.
5. McBride GB. A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. NIWA Client Report 2005: HAM2005-062.
6. Oshiro H, Czerniak BA, Sakamaki K, et al. Comparison between whole mount tissue preparations and virtual tissue microarray samples for measuring Ki-67 and apoptosis indices in human bladder cancer: A cross-sectional study. Medicine (Baltimore) 2016; 95: e4500.
7. Rindi G, Klöppel 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.
8. Rindi G. The ENETS guidelines: the new TNM classification system. Tumori 2010; 96: 806-809.
9. Rindi G, Klöppel G, Couvelard A, et al. TNM staging of midgut and hindgut (neuro) endocrine tumors: a consensus proposal including a grading system. Virchows Arch 2007; 451: 757-762.
10. Rindi G, Petrone G, Inzani F. The 2010 WHO classification of digestive neuroendocrine neoplasms: a critical appraisal four years after its introduction. Endocr Pathol 2014; 25: 186-192.
11. McCall CM, Shi C, Cornish TC, et al. Grading of well-differentiated pancreatic neuroendocrine tumors is improved by the inclusion of both Ki67 proliferative index and mitotic rate. Am J Surg Pathol 2013; 37: 1671-1677.
12. Klimstra DS. Pathology reporting of neuroendocrine tumors: essential elements for accurate diagnosis, classification, and staging. Semin Oncol 2013; 40: 23-36.
13. Reid MD, Balci S, Saka B, et al. Neuroendocrine tumors of the pancreas: current concepts and controversies. Endocr Pathol 2014; 25: 65-79.
14. Kunz PL, Reidy-Lagunes D, Anthony LB, et al. Consensus guidelines for the management and treatment of neuroendocrine tumors. Pancreas 2013; 42: 557-577.
15. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology. Neuroendocrine Tumors. Version 2.2016. https://www.nccn.org/professionals/physician_gls/pdf/neuroendocrine.pdf (Assessed: 24th July 2016).
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-174.
17. Søreide K, Kørner H, Søreide JA. Diagnostic accuracy and receiver-operating characteristics curve analysis in surgical research and decision making. Ann Surg 2011; 253: 27-34.
18. Carter JV, Pan J, Rai SN, Galandiuk S. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. Surgery 2016; 159: 1638-1645.
19. Abramson JH. WINPEPI updated: computer programs for epidemiologists, and their teaching potential. Epidemiol Perspect Innov 2011; 8: 1.
20. Gene-E. http://www.broadinstitute.org/cancer/software/GENE-E/index.html (Assessed: 7th August 2016).
21. Klöppel 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.
22. Klimstra DS, Modlin IR, Adsay NV, et al. Pathology reporting of neuroendocrine tumors: application of the Delphic consensus process to the development of a minimum pathology data set. Am J Surg Pathol 2010; 34: 300-313.
23. College of American Pathologists. Protocol for the Examination of Specimens From Patients With Tumors of the Endocrine Pancreas. Pancreas Endocrine 3.3.0.0. January 2016 Accessed: http://www.cap.org/ShowProperty?nodePath=/UCMCon/Contribution Folders/WebContent/pdf/cp-pancreasendo-16protocol-3300.pdf (30th July 2016).
24. Falconi M, Eriksson B, Kaltsas 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.
25. 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.
26. Basturk O, Tang L, Hruban RH, et al. Poorly differentiated neuroendocrine carcinomas of the pancreas: a clinicopathologic analysis of 44 cases. Am J Surg Pathol 2014; 38: 437-447.
27. Grillo F, Albertelli M, Brisigotti MP, et al. Grade increases in gastroenteropancreatic neuroendocrine tumor metastases compared to the primary tumor. Neuroendocrinology 2016; 103: 452-459.
28. Modlin IM, Moss SF, Gustafsson BI, et al. The archaic distinction between functioning and nonfunctioning neuroendocrine neoplasms is no longer clinically relevant. Langenbecks Arch Surg 2011; 396: 1145-1156.
29. Vyberg M, Nielsen S. Proficiency testing in immunohistochemistry – experiences from Nordic Immunohistochemical Quality Control (NordiQC). Virchows Arch 2016; 468: 19-29.
30. Nordic immunohistochemical Quality Control 2016. http://www.nordiqc.org/protocol_view.php?id=360 (assessed 30th July 2016).
31. Valous NA, Lahrmann B, Halama N, et al. Spatial intratumoral heterogeneity of proliferation in immunohistochemical images of solid tumors. Med Phys 2016; 43: 2936.
32. Plancoulaine B, Laurinaviciene A, Herlin P, et al. A methodology for comprehensive breast cancer Ki67 labeling index with intra-tumor heterogeneity appraisal based on hexagonal tiling of digital image analysis data. Virchows Arch 2016, in press.