



Histological interpretation of differentiated vulvar intraepithelial neoplasia (dVIN) remains challenging—observations from a bi-national ring-study

Shatavisha Dasgupta¹ • Elf de Jonge² • Mieke R. Van Bockstal³ • Luthy S. M. Wong-Alcala⁴ • Suzanne Wilhelmus⁵ • Lex A. C. F. Makkus⁶ • Katrien Schelfout^{7,8} • Koen K. Van de Vijver^{9,10} • Sander Smits⁵ • Etienne Marbaix³ • Senada Koljenović¹ • Folkert J. van Kemenade¹ • Patricia C. Ewing-Graham¹

Received: 13 December 2020 / Revised: 22 February 2021 / Accepted: 24 February 2021
© The Author(s) 2021

Abstract

Differentiated vulvar intraepithelial neoplasia (dVIN) is a premalignant lesion that is known to progress rapidly to invasive carcinoma. Accurate histological diagnosis is therefore crucial to allow appropriate treatment. To identify reliable diagnostic features, we evaluated the inter-observer agreement in the histological assessment of dVIN, among a bi-national, multi-institutional group of pathologists. Two investigators from Erasmus MC selected 36 hematoxylin-eosin-stained glass slides of dVIN and no-dysplasia, and prepared a list of 15 histological features of dVIN. Nine participating pathologists (i) diagnosed each slide as dVIN or no-dysplasia, (ii) indicated which features they used for the diagnosis, and (iii) rated these features in terms of their diagnostic usefulness. Diagnoses rendered by > 50% participants were taken as the consensus (gold standard). p53-immunohistochemistry (IHC) was performed for all cases, and the expression patterns were correlated with the consensus diagnoses. Kappa (κ)-statistics were computed to measure inter-observer agreements, and concordance of the p53-IHC patterns with the consensus diagnoses. For the diagnosis of dVIN, overall agreement was moderate ($\kappa = 0.42$), and pair-wise agreements ranged from slight ($\kappa = 0.10$) to substantial ($\kappa = 0.73$). Based on the levels of agreement and ratings of usefulness, the most helpful diagnostic features were parakeratosis, cobblestone appearance, chromatin abnormality, angulated nuclei, atypia discernable under $\times 100$, and altered cellular alignment. p53-IHC patterns showed substantial concordance ($\kappa = 0.67$) with the consensus diagnoses. Histological interpretation of dVIN remains challenging with suboptimal inter-observer agreement. We identified the histological features that may facilitate the diagnosis of dVIN. For cases with a histological suspicion of dVIN, consensus-based pathological evaluation may improve the reliability of the diagnosis.

Elf de Jonge, Mieke R. Van Bockstal, Luthy S. M. Wong-Alcala, Suzanne Wilhelmus, Lex A. C. F. Makkus, Katrien Schelfout, Koen K. Van de Vijver, Sander Smits and Etienne Marbaix contributed equally to this work.

✉ Shatavisha Dasgupta
s.dasgupta@erasmusmc.nl

Elf de Jonge
elf.de.jonge@ghz.nl

Mieke R. Van Bockstal
mieke.vanbockstal@uclouvain.be

Luthy S. M. Wong-Alcala
LAlcala@amphia.nl

Suzanne Wilhelmus
S.Wilhelmus@pathan.nl

Lex A. C. F. Makkus
lmakkus@paldordrecht.nl

Katrien Schelfout
kschelfout@hotmail.com

Koen K. Van de Vijver
Koen.VandeVijver@uzgent.be

Sander Smits
S.Smits@pathan.nl

Etienne Marbaix
etienne.marbaix@uclouvain.be

Senada Koljenović
s.koljenovic@erasmusmc.nl

Folkert J. van Kemenade
f.vankemenade@erasmusmc.nl

Patricia C. Ewing-Graham
p.ewing@erasmusmc.nl

Extended author information available on the last page of the article

Keywords Lower genital tract · Carcinoma in situ · Vulvar diseases · Histology · Observer variation

Introduction

Differentiated vulvar intraepithelial neoplasia (dVIN) is the immediate precursor of human papillomavirus (HPV)–independent vulvar squamous cell carcinoma (VSCC), and is postulated to develop on the background of chronic dermatoses, driven by *TP53* mutations [1–4]. Recent literature suggests that dVIN has an accelerated rate of progression to VSCC (median interval: 41.4 months), and a high recurrence rate [5–7]. In view of this, current treatment guidelines [8, 9] recommend surgical excision of lesions that are histologically diagnosed as dVIN. Evidently, accurate histological diagnosis is crucial to allow appropriate patient management.

On histology, distinguishing dVIN from dermatoses, such as lichen sclerosus (LS), can present a challenge, as dVIN often exhibits subtle atypical features that mimic the reactive changes seen in chronic dermatoses [10–12]. The difficulty of diagnosing dVIN can give rise to diagnostic variability, which has the potential to critically affect treatment decisions [13, 14].

Although the diagnostic difficulty of dVIN has been acknowledged in literature [2–5], there is insufficient data on the inter-observer agreement in the histological assessment. In a previous study, we established the features that helped to reliably distinguish dVIN from LS, and could be interpreted with substantial agreement by pathologists at our center [15]. However, it remains to be determined whether similar level of agreement can be achieved between pathologists from different practice settings.

In the current study, therefore, we evaluated the inter-observer agreement for the diagnosis, and in the interpretation of histological features of dVIN, among a bi-national, multi-institutional group of pathologists. We also assessed the perception of the pathologists regarding the diagnostic usefulness of the histological features. Our aim was to thereby identify reliable diagnostic features that may facilitate the diagnosis of dVIN. In addition, we correlated the immunohistochemical expression patterns of p53 with the consensus histological diagnoses, as this marker is frequently used as an ancillary tool to support the histological diagnosis of dVIN.

Materials and methods

Study design

For the purpose of this study, two investigators (SDG and PCEG) identified all vulvar lesions from 2010 to 2013, from the electronic records of the Department of Pathology,

Erasmus MC. All of these lesions were from patients who underwent vulvar biopsies or excisions at Erasmus MC. Hematoxylin-eosin (HE)–stained slides of these lesions were retrieved from the archives, and the histology was reviewed by these investigators.

From this series, the investigators selected a set of 36 slides for inclusion in this study. The selection was enriched for lesions regarded as dVIN by the investigators on histology review, since the aim was to evaluate inter-observer agreement in dVIN. Furthermore, to provide a range of challenges to the participants, the selection was prepared in a way to include (i) lesions adjacent to VSCC, as well as standalone lesions, and (ii) lesions with classical histology, which were diagnostically straightforward, as well as lesions where the distinction between dVIN and no-dysplasia could be difficult. The selection did not comprise any slides with invasive carcinoma, as presence of VSCC in the adjacent epithelium can be considered by pathologists as a diagnostic clue for dVIN [14].

Therefore, of the 36 selected slides, 25 contained lesions adjacent to VSCCs, and 11 contained standalone lesions. The investigators had judged 26 (72%) slides as dVIN and 10 (28%) slides as no-dysplasia, comprising 6 lichen sclerosus and 4 non-specific reactive lesions. The investigators perceived 67% of the diagnoses as straightforward and 33% as difficult.

The original diagnoses of these slides, or the diagnoses rendered by the investigators on review were not used for the analyses. For each slide, the diagnosis rendered by > 50% of the participants was taken as the consensus diagnosis/gold standard.

For de-identification, all slides were re-labeled with opaque stickers bearing a random number. No serial sections were prepared. To ensure that all pathologists evaluated identical areas, the regions of interest were marked on the glass slides with red lines.

For all included slides, immunohistochemistry (IHC) was conducted with (i) p16 (E6H4-clone, Ventana), to confirm that the selection did not contain any HPV-related lesion, and with (ii) p53 (Ventana), to correlate with the consensus diagnosis. The IHC protocol is detailed in supplementary document 1. IHC slides were read only by the investigators and were not provided to the participants. IHC was scored and interpreted as described below:

- (i) p16-IHC patterns were scored as block-type or non-block-type (patchy), following the guidelines of The Lower Anogenital Squamous Terminology Standardization Project (LAST) [16]. Block-type p16-expression is considered to be indicative of a high-risk

HPV-infection [16]. This pattern was not present in any slide, confirming that the selection did not contain any HPV-related lesion.

- (ii) p53-IHC patterns were scored as p53-mutant or p53-wild-type, following recent literature [17, 18]. p53-mutant patterns include basal to para-basal/diffuse overexpression, basal overexpression, null-pattern, or cytoplasmic expression, and these have been reported to strongly correlate with the presence of *TP53* mutations [17–19]. Presence of any of these patterns, therefore, can be considered supportive of a histological diagnosis of dVIN.

p53-wild-type pattern, i.e., scattered, heterogeneous, basal/para-basal expression, is primarily seen in non-dysplastic lesions. However, this pattern has been also occasionally observed in dVIN [15, 20–23]. Hence, a p53-wild-type pattern does not preclude a histological diagnosis of dVIN. p53 patterns observed in our slides are presented in the “Results.”

Next, a list of histological features of dVIN was compiled from previously published literature [13–15, 24–26], and incorporated into an assessment form. These comprised

- A. Features of nuclear atypia: (i) atypia discernable under $\times 100$ magnification; (ii) angulated nuclei; (iii) macronucleoli, i.e., nucleoli visible under $\times 100$ magnification; (iv) chromatin abnormality (open or hyperchromatic pattern); (v) multinucleation; (vi) suprabasal mitoses; (vii) atypical mitoses; and (viii) mitotic count $>5/5$ mm
- B. Features of disturbed maturation/architecture: (i) individual cell keratinization; (ii) deep keratinization; (iii) deep squamous eddies, i.e., abortive pearls of keratin; (iv) cobblestone appearance, i.e., combination of premature keratinization and spongiosis; (v) elongated and/or anastomosing rete ridges; (vi) altered cellular alignment; and (vii) parakeratosis.

Participants

Pathologists who attend the gynecological-pathology working group of the Rotterdam region were invited to participate. HE-stained glass slides were circulated among the participants for histological assessment. Instructions and forms for the assessment (supplementary document 2) were sent to the participants electronically. Clinical information, original diagnoses, or IHC results were not provided. There was no consensus meeting prior to the assessment to determine any diagnostic criteria. To allow the participants to interpret the histological features in light of their own experience, detailed instructions regarding this were not provided. For measuring 5 mm to

assess the mitotic count, participants could use an eye-piece graticule, or the field-diameter of the eye-pieces of their microscopes. Since the measure of 5 mm was an arbitrarily chosen cut-off, a rough estimate of this measurement was considered sufficient. The participants were masked from each other's assessments. Information regarding the nature of practice (academic/non-academic), country of practice, and length of practicing experience was gathered from the participants.

Histological assessment

Participants were asked to independently examine the areas marked on the slides, and:

- (i) Provide a diagnosis as – dVIN or no-dysplasia
- (ii) Score the histological features (listed above) as – not present or present, and if present, indicate whether they were useful, or very useful for the diagnosis of dVIN
- (iii) Indicate whether the diagnosis was easy or difficult

Ethics statement

This study was conducted in accordance with the guidelines of the Dutch Federation of Biomedical Scientific Societies (www.federa.org/codes-conduct), which state that no separate ethical approval is required for the use of anonymized residual tissue procured during regular treatment.

Statistical analysis

Data were analyzed after all participants had completed their assessments, using R Core Team (2020) (Version 4.0.0, <https://www.R-project.org/>). Histological diagnoses and features were assessed categorically. Inter-observer agreement was measured by computing (i) percentages of agreement – to obtain an absolute measure, and (ii) kappa (κ) statistics – to obtain a relative measure. Fleiss' κ was computed to measure the overall agreement, i.e., agreement among all participants, using packages “irr” and “raters” [27, 28]. Cohen's κ was computed to measure the agreement between each participant pair; this resulted in 36 κ -values for the diagnoses, as well as for each of the 15 histological features. Cohen's κ was also used to measure the concordance of the p53-IHC patterns with the consensus diagnoses. Bootstrapping (10,000 runs) was performed to calculate the 95% confidence intervals (CI) of the κ -values using the package “boot” [29]. κ -values were interpreted as follows: < 0.20 = slight, 0.21 – 0.40 = fair, 0.41 – 0.60 = moderate, 0.61 – 0.80 = substantial, or 0.81 – 1.00 = near-perfect agreement. Correlation between categorical variables was measured with chi (χ^2)-squared test; two-sided p -value < 0.05 was considered statistically significant. Heat maps and bar charts were constructed to visualize the data.

Results

Participants

Nine pathologists participated in this study; 6 practice at 5 non-academic centers in the Netherlands, which handle a high diagnostic case load, and 3 practice at 2 academic centers in Belgium. Lengths of their practice experience ranged from less than 5 years ($n = 2$) to more than 15 years ($n = 3$). All participants routinely read vulvar pathology cases, including dVIN and VSCC, in their practice. The participants have been anonymized and are represented by acronyms (P1–P9), which do not correspond to their order in the author list.

Histological assessment

The participants diagnosed 28–81% (median = 58%) of the slides as dVIN, and perceived 21–80% (median = 58%) of these diagnoses to be difficult (Fig. 1). Nineteen to 72% (median = 42%) of the slides were diagnosed as no-dysplasia, of which 6–73% (median = 29%) were perceived as difficult.

Consensus diagnoses and diagnostic agreement for dVIN

The consensus diagnosis for 23 slides (64%) was dVIN (Fig. 2). For these slides, rates of diagnostic agreement ranged from 56 to 100% (median = 78%). Unanimous agreement (100%) was obtained for 5 slides.

The consensus diagnosis for 13 (36%) slides was no-dysplasia (Fig. 2). For these slides, rates of diagnostic agreement ranged from 67 to 100% (median = 89%). Unanimous agreement was obtained for 4 slides.

The overall agreement for the diagnosis of dVIN was moderate ($\kappa = 0.42$), and the pair-wise agreements ranged from slight ($\kappa = 0.10$) to substantial ($\kappa = 0.73$) (Table 1). Substantial agreement was obtained between 19%, and moderate agreement between 39% of the participant pairs (Fig. 2).

Fig. 1 Bar chart depicting the proportions of slides diagnosed as dVIN or no-dysplasia per participant (P1–P9), along with the proportion of slides perceived as diagnostically difficult; the numbers depict the percentages of slides perceived as difficult

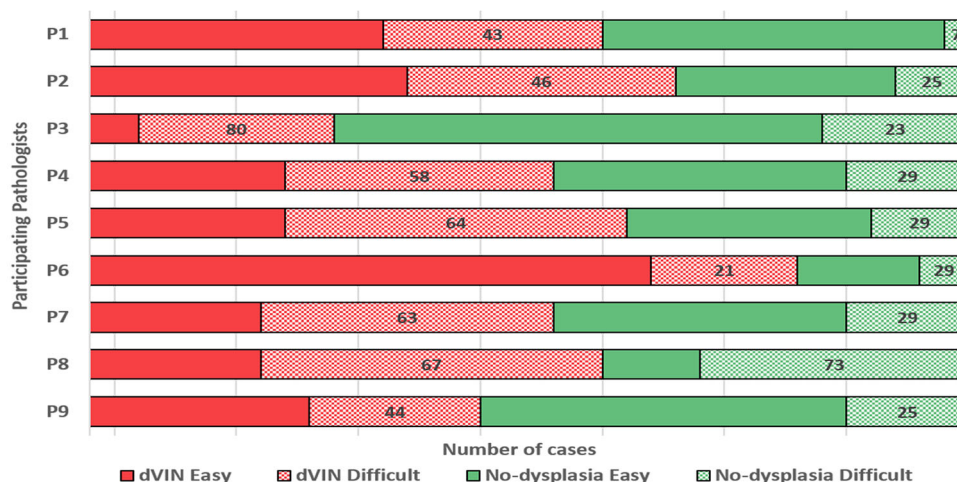


Fig. 2 Left: Heat map depicting the types of specimens, diagnoses rendered by the participants (P1–P9), p53-IHC results, and the consensus diagnoses; *, ‡, § slides were from the same specimen. Right: Heat map depicting the levels of agreement between the participant pairs for the diagnosis; color-coding corresponds to the levels of agreement

Pair-wise κ -values with 95% CI are provided in Table S1. The diagnosis of dVIN was more frequently perceived to be difficult than the diagnosis of no-dysplasia ($p = 0.02$). For all slides (dVIN or no-dysplasia), diagnostic difficulty perceived by the participants correlated significantly with lower percentages of agreement ($p = 0.001$).

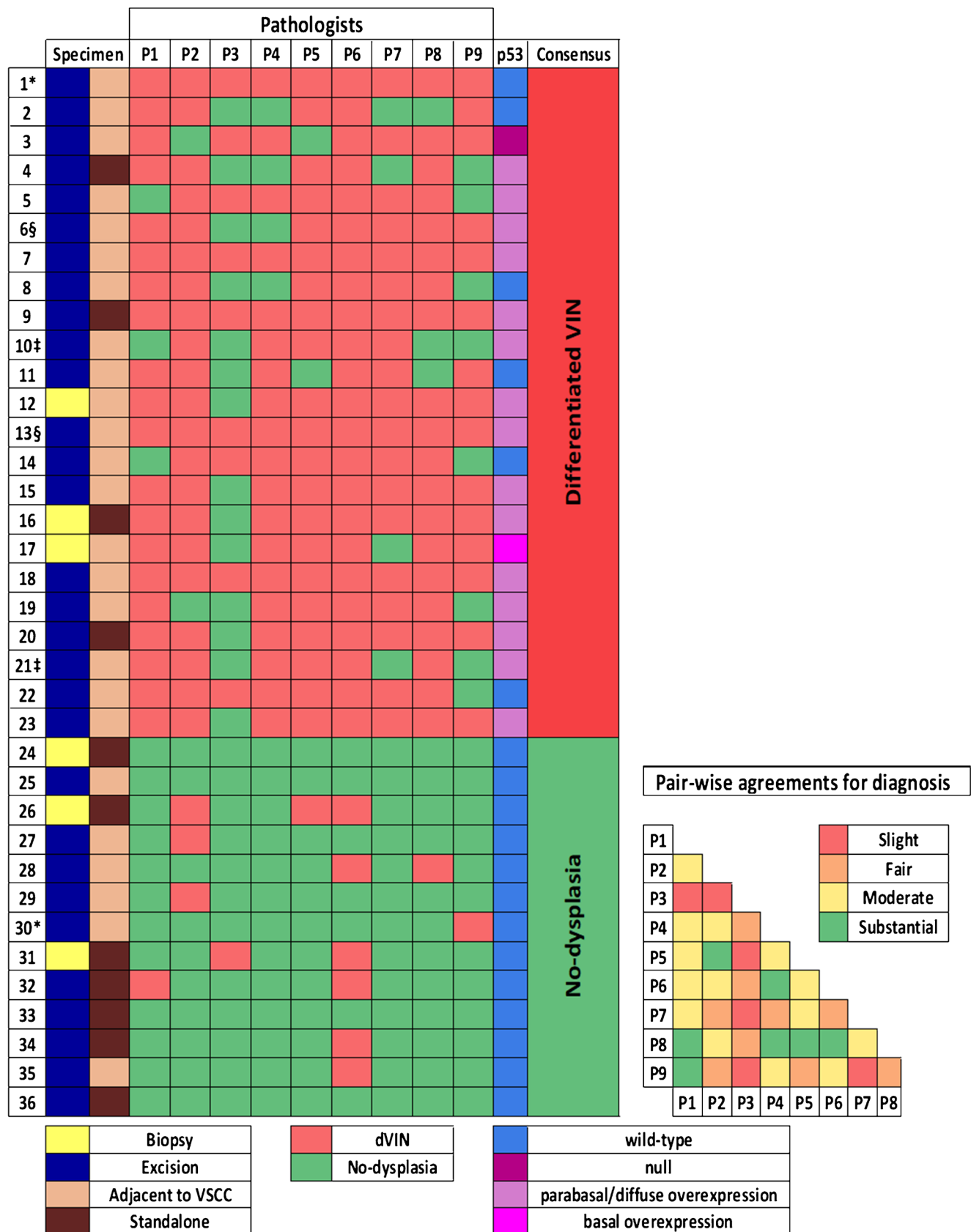
Correlation of the consensus diagnoses with p53-IHC patterns

Of the slides with a consensus diagnosis of dVIN, 17 (74%) showed p53-mutant patterns, which were basal to para-basal/diffuse overexpression in 15 slides, basal overexpression in 1 slide, and null-pattern in 1 slide. Six slides (26%) showed p53-wild-type pattern, i.e., scattered, heterogeneous, basal/para-basal expression (Fig. 2).

All slides with a consensus diagnosis of no-dysplasia showed p53-wild-type pattern, i.e., scattered, heterogeneous, basal/para-basal expression (Fig. 2). Concordance of the p53-IHC patterns with the consensus diagnoses was substantial ($\kappa = 0.67$; $p < 0.001$).

Agreements in the interpretation of histological features and ratings of their usefulness

Overall agreement was moderate in the interpretation of parakeratosis, mitotic count $> 5/5$ mm, and atypia discernable under $\times 100$ magnification. Fair agreement was obtained for multinucleation, angulated nuclei, chromatin abnormality, suprabasal mitoses, deep squamous eddies, elongated and/or



anastomosing rete ridges, altered cellular alignment, individual cell keratinization, and cobblestone appearance (Table 1).

Pair-wise agreements in the interpretation of the histological features ranged from slight ($\kappa = 0.01$) to near-perfect ($\kappa = 0.94$) (Table 1). The highest proportion of substantial/near-perfect agreement between participant pairs was obtained for parakeratosis (39%), and cobblestone appearance was rated most frequently (24%) as “very useful” for the diagnosis of dVIN (Table 2). Taking into consideration the levels of pair-wise agreements and the ratings of usefulness, the most helpful features were parakeratosis, cobblestone appearance, chromatin abnormality, angulated nuclei, atypia discernable under $\times 100$, and altered cellular alignment.

For each histological feature, the levels of pair-wise agreements are depicted in Figures S1 and S2, and the pair-wise κ -values with 95% CI are provided in Tables S2–S16. The ratings of usefulness are depicted in Fig. 3, and the histological features are demonstrated in Figs. 4 and 5.

Discussion

To the best of our knowledge, this is the first bi-national, multi-institutional, ring-study to assess the inter-observer agreement in the histological assessment of dVIN. Agreement on the diagnosis between nine participating

pathologists was moderate, while that between the participant pairs varied from slight to substantial. These results were similar to that of the only previous study on inter-observer agreement in dVIN [13], and indicate that the diagnostic agreement for dVIN remains suboptimal.

As histological diagnoses guide treatment decisions, variability in the diagnoses can result in treatment disparities [31]. Therefore, to improve the diagnostic reliability and to assure a similar standard of care, we suggest consensus evaluation of dVIN cases with a panel of pathologists experienced in vulvar neoplasia. Regular inter-disciplinary communication between gynecologists/dermatologists and pathologists can also enhance relevant knowledge and expertise.

An essential step to ensure a reliable histological diagnosis is to identify representative features which can be reproducibly interpreted by pathologists. We identified the most helpful features as parakeratosis, cobblestone appearance, chromatin abnormality, angulated nuclei, atypia discernable under $\times 100$, and altered cellular alignment, based on the proportions of substantial/near-perfect agreement between the participant pairs, and the ratings of diagnostic usefulness. We observed that the participants recorded parakeratosis and cobblestone appearance as very useful for diagnosing dVIN, particularly where the nuclear atypia could not be discerned under $\times 100$.

Previously, van den Einden et al. proposed that the presence of atypical mitoses in the basal layer, basal cellular

Table 1 Kappa values for the histological diagnoses and features of dVIN

	Overall agreement		Pair-wise agreements Range of κ -values
	κ -values (95% CI)	Level of agreement	
I. Histological diagnosis	0.42 (0.40–0.49)	Moderate	0.10–0.73
II. Histological features			
A. Nuclear atypia			
i. Atypia discernable under $\times 100$ magnification	0.42 (0.33–0.46)	Moderate	0.06–0.70
ii. Angulated nuclei	0.32 (0.29–0.35)	Fair	0.01–0.65
iii. Macronucleoli	0.10 (0.04–0.13)	Slight	0.01–0.68
iv. Chromatin abnormality	0.32 (0.27–0.37)	Fair	0.08–0.67
v. Multinucleation	0.34 (0.32–0.41)	Fair	0.01–0.87
vi. Suprabasal mitoses	0.28 (0.22–0.31)	Fair	0.01–0.61
vii. Atypical mitoses	0.11 (0.04–0.14)	Slight	0.01–0.91
viii. Mitotic count $>5/5$ mm	0.45 (0.38–0.51)	Moderate	0.01–0.94
B. Disturbed maturation and architecture			
i. Individual cell keratinization	0.21 (0.17–0.24)	Fair	0.01–0.46
ii. Deep keratinization	0.19 (0.13–0.21)	Slight	0.01–0.48
iii. Deep squamous eddies	0.31 (0.23–0.42)	Fair	0.06–0.60
iv. Cobblestone appearance	0.22 (0.19–0.27)	Fair	0.10–0.61
v. Elongated and/or anastomosing rete ridges	0.30 (0.22–0.35)	Fair	0.04–0.69
vi. Altered cellular alignment	0.23 (0.18–0.29)	Fair	0.01–0.73
vii. Parakeratosis	0.57 (0.49–0.61)	Moderate	0.17–0.82

Table 2 Histological features of dVIN, in descending order of the proportions of substantial/almost-perfect agreement, and ratings as “very useful” for diagnosis

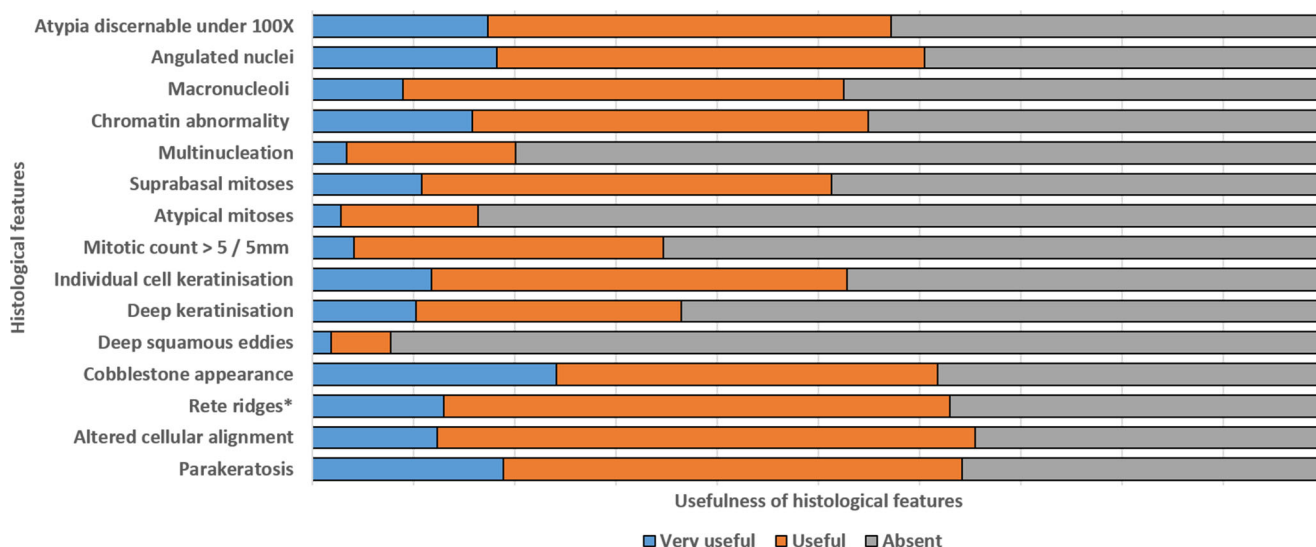
Proportion of substantial/near-perfect agreement	Very useful for the diagnosis of dVIN
1. Parakeratosis (39%)	1. Cobblestone appearance (24%)
2. Mitotic count > 5/5 mm (19%)	2. Parakeratosis (19%)
3. Deep squamous eddies (14%)	3. Angulated nuclei (18%)
4. Multinucleation (11%)	4. Atypia discernable under × 100 (16%)
5. Chromatin abnormality (8%)	5. Chromatin abnormality (16%)
6. Atypical mitoses (8%)	6. Elongated and/or anastomosing rete ridges (13%)
7. Atypia discernable under × 100 (6%)	7. Altered cellular alignment (12%)
8. Angulated nuclei (6%)	8. Individual cell keratinization (12%)
9. Macronucleoli (6%)	9. Suprabasal mitoses (11%)
10. Cobblestone appearance (6%)	10. Deep keratinization (10%)
11. Altered cellular alignment (6%)	11. Macronucleoli (9%)
12. Elongated and/or anastomosing rete ridges (5%)	12. Multinucleation (3%)
13. Suprabasal mitoses (3%)	13. Mitotic count > 5/5 mm (3%)
14. Individual cell keratinization (0%)	14. Atypical mitoses (3%)
15. Deep keratinization (0%)	15. Deep squamous eddies (2%)

atypia, dyskeratosis, prominent nucleoli, and elongated and anastomosing rete ridges were the most predictive features of dVIN [13]. In a subsequent survey among vulva pathology experts, only basal layer atypia was judged by consensus as an “essential” diagnostic feature [14]. However, neither of these studies assessed the agreement in the interpretation of these features. In our previous study, we obtained substantial agreement in the interpretation of macronucleoli, angulated nuclei, individual cell keratinization, deep keratinization, and deep squamous eddies, between pathologists at our center [15]. In the current study, however, similar level of agreement for these features was not observed. We speculate that our previous results may have been influenced by the similar standard

of histological interpretation among participants who work in close collaboration at the same center.

In this study, we also correlated the histological consensus diagnoses with the immunohistochemical expression of p53, as this marker is commonly used to aid the diagnosis of dVIN. p53-mutant patterns have been reported to accurately reflect underlying *TP53* mutations, which characterize dVIN [19, 20, 32]. Substantial concordance of p53-IHC patterns with the histological consensus diagnoses was recorded, which confirms that routine use of this marker can improve the diagnostic accuracy for dVIN.

However, 6 (26%) of the slides in this study that were diagnosed as dVIN by consensus, showed wild-type p53-

**Fig. 3** Bar charts representing the proportions of ratings of usefulness for each histological feature; *elongated and/or anastomosing rete ridges

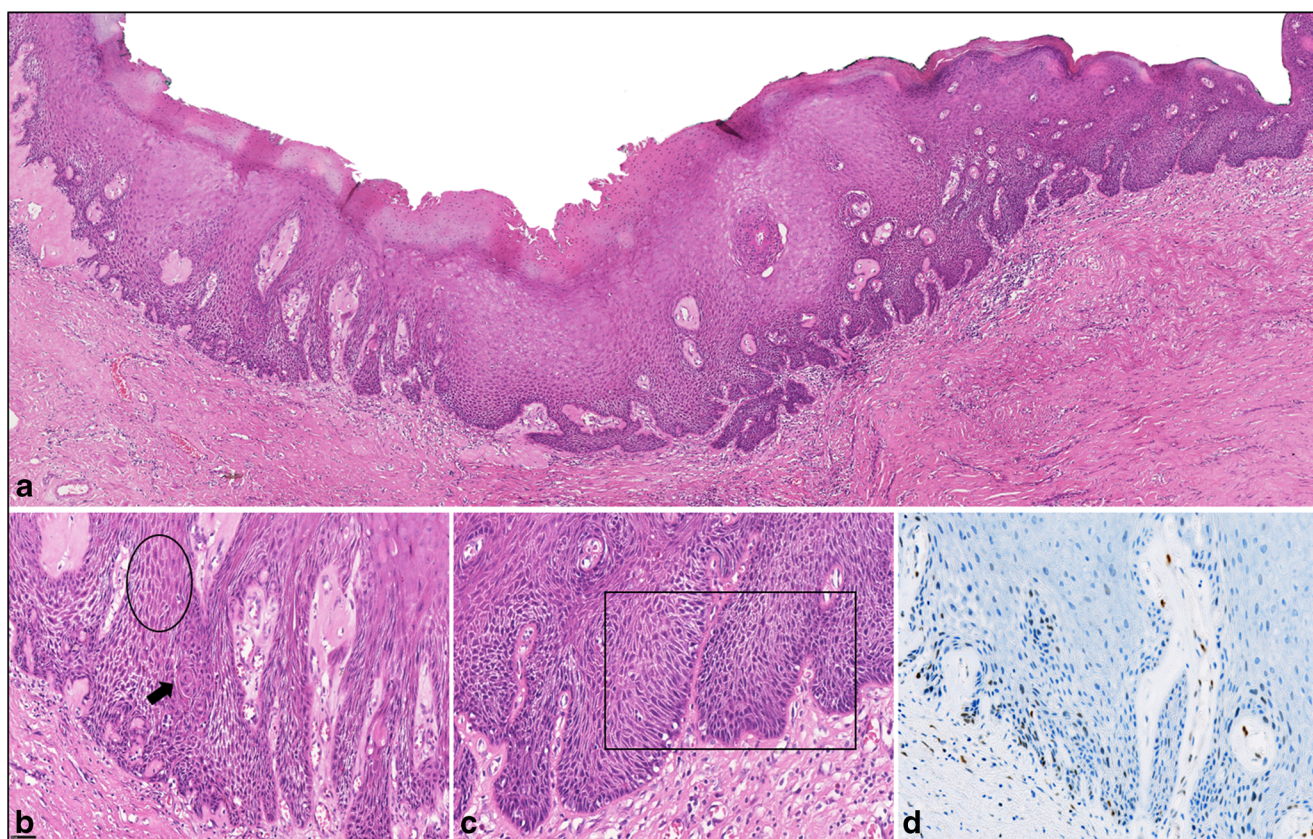


Fig. 4 Low- (A) and high-magnification (B and C) images of one of the slides (HE-stain) that was diagnosed by all participants as dVIN, with corresponding p53-IHC (D). (A) Epithelial acanthosis, hyperkeratosis, parakeratosis, and an eosinophilic appearance can be appreciated under low magnification (original magnification $\times 5$). (B) Cobblestone appearance (circled area), deep squamous eddies (arrow), elongated rete

ridges, and (C) altered cellular alignment and angulated nuclei (squared area) were rated by the participants as “very useful” features for the diagnosis for this slide (original magnification $\times 200$); (D) p53-IHC shows a wild-type pattern, i.e., scattered nuclear p53 staining of heterogeneous intensity in the basal and the para-basal layers

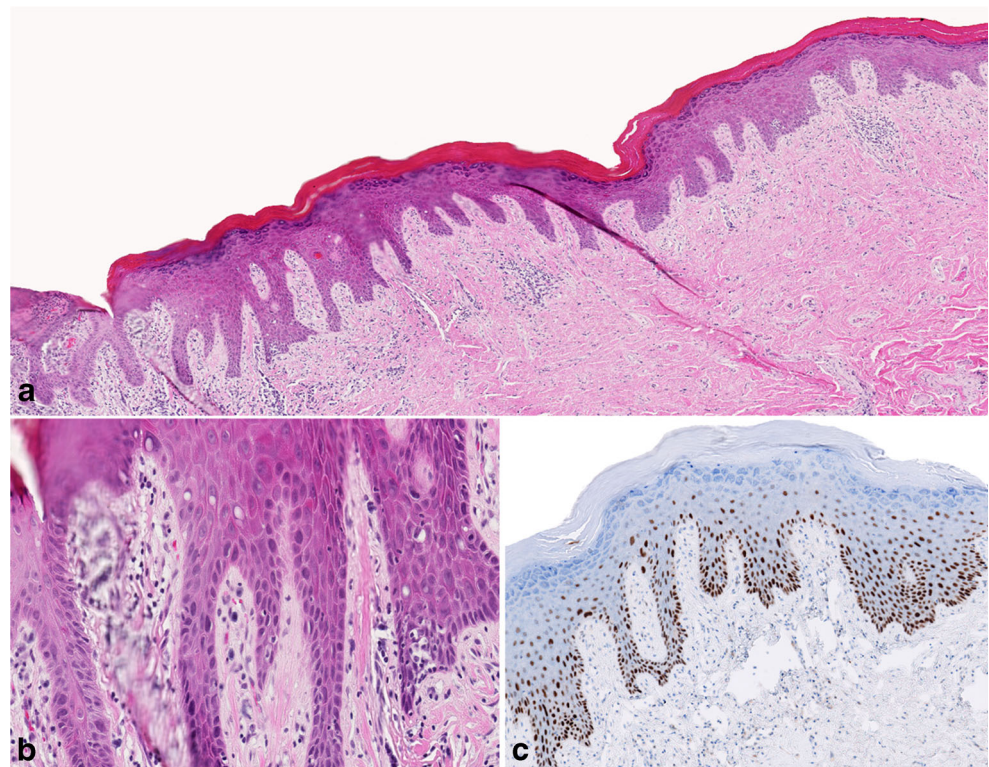
expression. This is in line with recent literature, which states that 17–42% cases of dVIN can show wild-type p53-expression [4], and implies that p53-IHC may not effectively inform the diagnosis in every case of dVIN. Furthermore, p53-IHC patterns in VSCC and the adjacent dVIN may not show perfect concordance [22]. A recent study reported that while dVIN adjacent to p53-wild-type VSCC always shows wild-type p53-expression, dVIN adjacent to p53-mutant VSCC can show wild-type p53-expression in 31.4% of cases [22]. In our study, all of the lesions judged as dVIN by consensus and showing wild-type p53-expression were present adjacent to VSCC. Similarly to the previous study [22], we observed that 67% (4/6) of these VSCCs showed wild-type p53-expression, while 33% (2/6) showed p53-mutant patterns (results not presented). This limitation of p53-IHC should be borne in mind particularly when using this marker to confirm the presence of dVIN in resection margins of VSCC. For dVINs that show wild-type p53-expression, the diagnosis defers to histological assessment, which, as our study indicates, may be fraught with variability. In view of this, we believe that ancillary biomarkers (immunohistochemical/molecular) need to be

established to aid the diagnosis of the p53-wild-type subcategory of dVIN.

Through this study, we intended to estimate the diagnostic variability of dVIN in the real world. To ensure an accurate representation of this variability, (i) pathologists with varying levels of experience and from academic and non-academic centers were included, (ii) diagnostic criteria were not predetermined to allow the participants to interpret the histology in light of their own experience, and (iii) assessments of outlier participants were not excluded.

Nevertheless, there are several limitations of this study. We used the majority (consensus) diagnosis of each slide to determine the diagnostic gold standard. It could be argued whether the consensus represents another diagnostic opinion rather than a standard of truth. dVIN is known to originate in a background of chronic dermatoses, and there is no clear, universally accepted threshold for identifying atypia/dysplasia. This threshold is often influenced by the pathologists’ training and/or practice experience. Unless a reliable IHC marker is established, every method to ascertain a gold-standard diagnosis will have some bias.

Fig. 5 Low- (A) and high-magnification (B) images of one of the slides (HE-stain) that was diagnosed by 5 participants as dVIN and 4 participants as no-dysplasia, with corresponding p53-IHC (C). (A) Elongated rete ridges and hyperkeratosis are appreciable under low magnification (original magnification $\times 5$). (B) Participants who diagnosed this slide as dVIN rated angulated nuclei, chromatin abnormality, cobblestone appearance, elongated rete ridges, and altered cellular alignment as “very useful” features for the diagnosis (original magnification $\times 300$); (C) p53-IHC shows mutant pattern, i.e., diffuse, strong, nuclear p53 staining in the basal and para-basal layers



There is also little consensus on the ideal method for measuring observer agreement in pathology diagnosis. It has been suggested that both percentages of agreement and κ -statistics do not take into account the prevalence of a particular diagnosis in a set of cases, or completely rule out concordances due to chance [33, 34]. Validity of the cut-offs that are used to interpret levels of agreement from κ -values has also been challenged [30, 35].

It could also be argued whether our study overestimated the diagnostic variability. Unlike in routine practice, participants diagnosed the slides without clinical information, serial sections, or IHC. The selection contained a higher proportion of dVIN than no-dysplasia slides, which may not reflect routine practice. We lacked statistical power to evaluate the influence of level of experience or practice setting on the diagnostic variability. Furthermore, the inter-observer agreement in the interpretation of p53-IHC was not assessed. To gain further insights on these contexts, we have set up a larger study among geographically disparate group of pathologists, which includes the assessment of p53-IHC.

In conclusion, the suboptimal level of diagnostic agreement for dVIN observed in this study affirms the difficulty of the diagnosis. We identified parakeratosis, cobblestone appearance, chromatin abnormality, angulated nuclei, atypia discernable under $\times 100$, and altered cellular alignment as helpful

diagnostic features of dVIN. For cases with a histological suspicion of dVIN, we suggest consensus-based pathological evaluation to improve diagnostic reliability.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00428-021-03070-0>.

Acknowledgements The authors are grateful to Mr. Hongchao Qi, Statistical Consultant, Department of Bio-statistics, Erasmus MC, for his guidance on the data analyses, and to Dr. Thierry P.P. van den Bosch, Dr. Remko Hersmus, and Ms. Linda Coelers, from the Department of Pathology, Erasmus MC, for their excellent technical assistance.

Author contribution Conceptualization: S. Dasgupta, P.C. Ewing-Graham, S. Koljenović, F.J. van Kemenade; methodology: S. Dasgupta, P.C. Ewing-Graham, F.J. van Kemenade; software, formal analysis, visualization: S. Dasgupta; data curation: S. Dasgupta, P.C. Ewing-Graham, E. de Jonge, M.R. Van Bockstal, L.S.M. Wong-Alcala, S. Wilhelmus, L.A.C.F. Makkus, K. Schelfout, K.K. Van de Vijver, S. Smits, E. Marbaix; writing – original draft preparation: S. Dasgupta; writing – review and editing: all authors; supervision: P.C. Ewing-Graham, S. Koljenović, F.J. van Kemenade

Data Availability Whole slide images of the cases included in this study are available for sharing with physicians and researchers for educational and research purposes. Upon reasonable request, images will be shared in a secure manner, obeying our hospital guidelines. Requests can be directed to the corresponding author.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Williams A, Syed S, Velangi S, Ganesan R (2019) New directions in vulvar cancer pathology. *Curr Oncol Rep* 21(10):88. <https://doi.org/10.1007/s11912-019-0833-z>
- Dasgupta S, Ewing-Graham PC, Swagemakers SMA, van der Spek PJ, van Doorn HC, Noordhoek Hegt V, Koljenović S, van Kemenade FJ (2020) Precursor lesions of vulvar squamous cell carcinoma - histology and biomarkers: a systematic review. *Crit Rev Oncol Hematol* 147:102866. <https://doi.org/10.1016/j.critrevonc.2020.102866>
- Jin C, Liang S (2019) Differentiated vulvar intraepithelial neoplasia: a brief review of clinicopathologic features. *Arch Pathol Lab Med* 143:768–771. <https://doi.org/10.5858/arpa.2018-0019-RS>
- Heller DS, Day T, Allbritton JI, Scurry J, Radici G, Welch K, Preti M, for the ISSVD Difficult Pathologic Diagnoses Committee (2020) Diagnostic criteria for differentiated vulvar intraepithelial neoplasia and vulvar aberrant maturation. *J Low Genit Tract Dis* 25:57–70. <https://doi.org/10.1097/LGT.0000000000000572>
- Singh N, Gilks CB (2020) Vulval squamous cell carcinoma and its precursors. *Histopathology* 76:128–138. <https://doi.org/10.1111/his.13989>
- Goyal A, Zhang G, Yang B (2018) Differential expression patterns of GATA3 in usual and differentiated types of vulvar intraepithelial neoplasia: potential diagnostic implications. *Mod Pathol* 31:1131–1140. <https://doi.org/10.1038/s41379-018-0021-y>
- Griesinger LM, Walline H, Wang GY, Lorenzatti Hiles G, Welch KC, Haefner HK, Lieberman RW, Skala SL (2020) Expanding the morphologic, immunohistochemical, and HPV genotypic features of high-grade squamous intraepithelial lesions of the vulva with morphology mimicking differentiated vulvar intraepithelial neoplasia and/or lichen sclerosus. *Int J Gynecol Pathol*. <https://doi.org/10.1097/PGP.0000000000000708>
- NCCN Clinical Practice Guidelines in Oncology Version 3.2020 (2020) - Vulvar Cancer (Squamous Cell Carcinoma). https://www.nccn.org/store/login/login.aspx?ReturnURL=https://www.nccn.org/professionals/physician_gls/pdf/vulvar.pdf. Last accessed Nov 2020
- Morrison J, Baldwin P, Buckley L, Cogswell L, Edey K, Faruqi A, Ganesan R, Hall M, Hillaby K, Reed N, Rolland P, Fotopoulou C (2020) British Gynaecological Cancer Society (BGCS) vulvar cancer guidelines: recommendations for practice. *Eur J Obstet Gynecol Reprod Biol* 252:502–525. <https://doi.org/10.1016/j.ejogrb.2020.05.054>
- Watkins JC (2019) Human papillomavirus-independent squamous lesions of the vulva. *Surg Pathol Clin* 12:249–261. <https://doi.org/10.1016/j.path.2019.01.001>
- Day T, Wilkinson E, Rowan D, Scurry J, Committee* IDPD (2020) Clinicopathologic diagnostic criteria for vulvar lichen planus. *J Low Genit Tract Dis* 24:317–329. <https://doi.org/10.1097/LGT.0000000000000532>
- Day T, Marzol A, Pagano R, Jaaback K, Scurry J (2020) Clinicopathologic diagnosis of differentiated vulvar intraepithelial neoplasia and vulvar aberrant maturation. *J Low Genit Tract Dis* 24:392–398. <https://doi.org/10.1097/LGT.0000000000000569>
- van den Einden LC, de Hullu JA, Massuger LF et al (2013) Interobserver variability and the effect of education in the histopathologic diagnosis of differentiated vulvar intraepithelial neoplasia. *Mod Pathol* 26:874–880. <https://doi.org/10.1038/modpathol.2012.235>
- Reutter JC, Walters RA, Selim MA (2016) Differentiated vulvar intraepithelial neoplasia: what criteria do we use in practice? *J Low Genit Tract Dis* 20:261–266. <https://doi.org/10.1097/LGT.0000000000000211>
- Dasgupta S, Ewing-Graham PC, van Kemenade FJ, van Doorn HC, Noordhoek Hegt V, Koljenović S (2018) Differentiated vulvar intraepithelial neoplasia (dVIN): the most helpful histological features and the utility of cytokeratins 13 and 17. *Virchows Arch* 473: 739–747. <https://doi.org/10.1007/s00428-018-2436-8>
- Darragh TM, Colgan TJ, Cox JT et al (2012) The lower anogenital squamous terminology standardization project for HPV-associated lesions: background and consensus recommendations from the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology. *Arch Pathol Lab Med* 136: 1266–1297. <https://doi.org/10.1097/PGP.0b013e31826916c7>
- Kortekaas KE, Solleveld-Westerink N, Tessier-Cloutier B, Rutten TA, Poelgeest MIE, Gilks CB, Hoang LN, Bosse T (2020) Performance of the pattern based interpretation of p53 immunohistochemistry as a surrogate for TP53 mutations in vulvar squamous cell carcinoma. *Histopathology* 77:92–99. <https://doi.org/10.1111/his.14109>
- Tessier-Cloutier B, Kortekaas KE, Thompson E, Pors J, Chen J, Ho J, Prentice LM, McConechy MK, Chow C, Proctor L, McAlpine JN, Huntsman DG, Gilks CB, Bosse T, Hoang LN (2020) Major p53 immunohistochemical patterns in situ and invasive squamous cell carcinomas of the vulva and correlation with TP53 mutation status. *Mod Pathol* 33:1595–1605. <https://doi.org/10.1038/s41379-020-0524-1>
- Tessier-Cloutier B, Pors J, Thompson E, Ho J, Prentice L, McConechy M, Aguirre-Hernandez R, Miller R, Leung S, Proctor L, McAlpine JN, Huntsman DG, Gilks CB, Hoang LN (2021) Molecular characterization of invasive and in situ squamous neoplasia of the vulva and implications for morphologic diagnosis and outcome. *Mod Pathol* 34:508–518. <https://doi.org/10.1038/s41379-020-00651-3>
- Nooij LS, Ter Haar NT, Ruano D et al (2017) Genomic characterization of vulvar (pre)cancers identifies distinct molecular subtypes with prognostic significance. *Clin Cancer Res* 23:6781–6789. <https://doi.org/10.1158/1078-0432.CCR-17-1302>
- Liu YA, Ji JX, Almadani N, Crawford RI, Gilks CB, Kinloch M, Hoang L (2020) Comparison of p53 immunohistochemical staining in differentiated vulvar intraepithelial neoplasia (dVIN) with that in inflammatory dermatoses and benign squamous lesions in the vulva. *Histopathology* 78:424–433. <https://doi.org/10.1111/his.14238>
- Rakislova N, Alemany L, Clavero O, Saco A, Torné A, del Pino M, Munmany M, Rodrigo-Calvo M, Guerrero J, Marimon L, Vega N, Quirós B, Lloveras B, Ribera-Cortada I, Alejo M, Pawlita M, Quint W, de Sanjose S, Ordi J, VVAP Study Group (2020) p53 immunohistochemical patterns in HPV-independent squamous cell

- carcinomas of the vulva and the associated skin lesions: a study of 779 cases. *Int J Mol Sci* 21:8091. <https://doi.org/10.3390/ijms21218091>
23. Dasgupta S, Koljenović S, van Doorn HC, Noordhoek Hegt V, van Kemenade FJ, Ewing-Graham PC (2019) Immunohistochemistry with CK17 and SOX-2, in addition to p53, can be of diagnostic value for differentiated vulvar intraepithelial neoplasia (dVIN). *Virchows Arch* 475:1–436. <https://doi.org/10.1007/s00428-019-02631-8>
 24. WHO (2014) Classification of tumours of female reproductive organs, 4th edn. International Agency for Research on Cancer, Lyon
 25. Hoang LN, Park KJ, Soslow RA, Murali R (2016) Squamous precursor lesions of the vulva: current classification and diagnostic challenges. *Pathology* 48:291–302. <https://doi.org/10.1016/j.pathol.2016.02.015>
 26. Hart WR (2001) Vulvar intraepithelial neoplasia: historical aspects and current status. *Int J Gynecol Pathol* 20:16–30. <https://doi.org/10.1097/00004347-200101000-00003>
 27. Quatto P, Ripamonti E (2014) raters: a modification of Fleiss' Kappa in case of nominal and ordinal variables. R package version 2.0.1
 28. Gamer M, Lemon J, Fellows I, Singh P <https://CRAN.R-project.org/package=irr>. Accessed Dec 2020
 29. Davison AC, Hinkley DV (1997) Bootstrap methods and their applications. Cambridge University Press, Cambridge
 30. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
 31. Jonker LW, Dasgupta S, Ewing-Graham PC, van Doorn HC (2020) Can radical surgical treatment of the vulva be justified in the absence of a conclusive diagnosis of squamous cell carcinoma on biopsy? A retrospective 10-year cohort study. *Eur J Obstet Gynecol Reprod Biol* 248:238–244. <https://doi.org/10.1016/j.ejogrb.2020.03.027>
 32. Zieba S, Pouwer AW, Kowalik A et al (2020) Somatic mutation profiling in premalignant lesions of vulvar squamous cell carcinoma. *Int J Mol Sci* 21:4880. <https://doi.org/10.3390/ijms21144880>
 33. Marchevsky AM, Walts AE, Lissenberg-Witte BI, Thunnissen E (2020) Pathologists should probably forget about kappa. Percent agreement, diagnostic specificity and related metrics provide more clinically applicable measures of interobserver variability. *Ann Diagn Pathol* 47:151561. <https://doi.org/10.1016/j.anndiagpath.2020.151561>
 34. de Vet HCW, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL (2013) Clinicians are right not to like Cohen's κ . *BMJ: Br Med J* 346:f2125. <https://doi.org/10.1136/bmj.f2125>
 35. Banerjee M, Capozzoli M, McSweeney L, Sinha D (1999) Beyond kappa: a review of interrater agreement measures. *Can J Stat* 27:3–23. <https://doi.org/10.2307/3315487>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Shatavisha Dasgupta¹  • Elf de Jonge² • Mieke R. Van Bockstal³  • Luthy S. M. Wong-Alcala⁴ • Suzanne Wilhelmus⁵  • Lex A. C. F. Makkus⁶ • Katrien Schelfout^{7,8} • Koen K. Van de Vijver^{9,10}  • Sander Smits⁵ • Etienne Marbaix³  • Senada Koljenović¹  • Folkert J. van Kemenade¹  • Patricia C. Ewing-Graham¹ 

¹ Department of Pathology, Erasmus MC, University Medical Centre, Postbus 2040, Be-building, 3000CA, Rotterdam, The Netherlands

² Department of Pathology, Groene Hart Ziekenhuis, Gouda, The Netherlands

³ Department of Pathology, Cliniques Universitaires Saint-Luc Bruxelles, Brussels, Belgium

⁴ Department of Pathology, Pathologisch en Cytologisch laboratorium, Breda, The Netherlands

⁵ Department of Pathology, Pathan B.V., Laboratory for Pathology, Rotterdam, The Netherlands

⁶ Laboratory for Pathology, PAL Dordrecht, Dordrecht, The Netherlands

⁷ Department of Pathology, Bravis Ziekenhuis, Bergen op Zoom, The Netherlands

⁸ Department of Pathology, Ziekenhuis Geel, Geel, Belgium

⁹ Department of Pathology, Cancer Research Institute Ghent, Ghent University Hospital, Ghent, Belgium

¹⁰ Department of Pathology, Antwerp University, Antwerp, Belgium