Received Date : 04-Dec-2015
Revised Date : 17-Mar-2016
Accepted Date : 27-Mar-2016
Article type : Original Article

TITLE PAGE

Title: Gleason Grade 4 Prostate Adenocarcinoma Patterns: An Inter-observer Agreement Study among Genitourinary Pathologists

Running title: Reproducibility of Gleason grade 4 patterns

List of authors:

Charlotte F. Kweldam[1*]

Daan Nieboer[2]

Ferran Algaba[3]

Mahul B. Amin[4]

Dan M. Berney[5]

Athanase Billis[6]

David G. Bostwick[7]

Lukas Bubendorf[8]

Liang Cheng[9]

Eva Compérat[10]

Brett Delahunt[11]

Lars Egevad[12]

Andrew J. Evans[13]

Donna E. Hansel[14]

Peter A. Humphrey[15]

This is the author's manuscript of the article published in final edited form as:
Kweldam, C. F., Nieboer, D., Algaba, F., Amin, M. B., Berney, D. M., Billis, A., ... & Delahunt, B. (2016). Gleason grade 4 prostate adenocarcinoma patterns: an interobserver agreement study among genitourinary pathologists. Histopathology, 69(3), 441-449. http://dx.doi.org/10.1111/his.12976

Glen Kristiansen[16]

Theodorus H. van der Kwast[13]

Cristina Magi-Galluzzi[17]

Rodolfo Montironi[18]

George J. Netto[19]

Hemamali Samaratunga[20]

John R. Srigley[21]

Puay H. Tan[22]

Murali Varma[23]

Ming Zhou[24]

Geert J.L.H. van Leenders[1]

[1] Department of Pathology, Erasmus Medical Center, Rotterdam, The Netherlands

[2] Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands

[3] Department of Pathology, Universitat Autónoma de Barcelona, Barcelona, Spain

[4] Department of Pathology & Laboratory Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[5] Department of Cellular Pathology, The Royal London Hospital, London, UK

[6] Department of Anatomic Pathology, School of Medical Sciences, State University of Campinas (Unicamp), Campinas, Brazil

[7] Bostwick Laboratories, Orlando, FL, USA

[8] Institute for Pathology, University Hospital Basel, Basel, Switzerland

[9] Department of Pathology & Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA

[10] Service d'Anatomie & Cytologie Pathologiques du Pr Capron, Hôpital de la Pitié-Salpêtrière, Paris, France

[11] Department of Pathology and Molecular Medicine, Wellington School of Medicine and Health Sciences, University of Otago, Wellington, New Zealand

[12] Department of Oncology and Pathology, Karolinska University Hospital, Solna, Stockholm, Sweden

[13] Department of Pathology & Laboratory Medicine, University Health Network, University of Toronto, Toronto, Ontario, Canada

[14] Department of Pathology, University of California San Diego, La Jolla, CA, USA

[15] Department of Pathology, Yale University School of Medicine, New Haven, CT, USA

[16] Institute of Pathology, University of Bonn, Bonn, Germany

[17] Robert J. Tomsich Pathology and Laboratory Medicine Institute, Cleveland Clinic, Cleveland, OH, USA

[18] Section of Pathological Anatomy, Department of Biomedical Sciences and Public Health, Polytechnic University of the Marche Region (Ancona), Ancona, Italy

[19] Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, USA

[20] Aquesta Pathology and University of Queensland, Brisbane, Qld, Australia

[21] Trillium Health Partners, Mississauga and McMaster University, Hamilton, ON, Canada

[22] Department of Pathology, Singapore General Hospital, Singapore

[23] Department of Medical Genetics, Haematology and Pathology, Cardiff University, Cardiff, UK

[24] Department of Pathology, NYU Langone Medical Center, New York, NY, USA

* **Corresponding author**. E-mail: c.kweldam@erasmusmc.nl, Tel. +31 10704384

Disclosure: None of the authors has a conflict of interest.

ABSTRACT

Aims: To assess the inter-observer reproducibility of individual Gleason grade 4 growth patterns.

Methods: Twenty-three genitourinary pathologists participated in the evaluation of 60 selected high-magnification photographs. The selection included 10 cases of Gleason grade 3, 40 Gleason grade 4 (10 per growth pattern) and 10 Gleason grade 5. Participants were asked to select a single predominant Gleason grade per case (3, 4, 5) and to indicate the predominant Gleason grade 4 growth pattern if present. "Consensus" was defined as at least 80% agreement and "favoured" as 60-80% agreement.

Results: Consensus on Gleason grading was reached in 47/60 (78%) cases, 35 of which were assigned a grade 4. In the 13 non-consensus cases, ill-formed (6/13, 46%) and fused (7/13, 54%) patterns were involved in disagreement. Among the 20 cases where at least 1 pathologist assigned the ill-formed growth pattern none (0%, 0/20) reached consensus. Consensus for fused, cribriform and glomeruloid glands was reached in 2%, 23%, and 38%, respectively. In 9/35 (26%) consensus Gleason grade 4 cases, participants disagreed on growth pattern. Six of these were characterized by large epithelial proliferations with delicate intervening fibro-vascular cores, which were alternately given fused or cribriform growth pattern ("complex fused").

Conclusions: Consensus on Gleason grade 4 growth pattern was predominantly reached on cribriform and glomeruloid patterns, but rarely on ill-formed and fused glands. The complex fused glands seem to be a borderline pattern of unknown prognostic significance on which a consensus could not be reached.

Key words: Gleason grading; inter-observer variability; prostate cancer

INTRODUCTION

The Gleason grading system is one of the most important predictors of prostate cancer progression. In 2005, the International Society of Urological Pathology (ISUP) organized a consensus conference on Gleason grading of prostate cancer.[1] The goal of this meeting was to achieve consensus among leading genitourinary pathologists in specific areas of Gleason grading. At this meeting, large cribriform and ill-formed glands, classically described as Gleason grade 3, were re-defined as Gleason grade 4.[1] Subsequently, during the 2014 ISUP consensus conference on prostate cancer grading it was decided that small cribriform and glomeruloid glands should also be considered a Gleason grade 4 pattern.[2] As a result, contemporary Gleason grade 4 prostate cancer represents a heterogeneous group of various growth patterns consisting of ill-formed, fused, cribriform and glomeruloid glands. While Gleason grade 4 is not sub-classified in daily clinical practice, recent studies have suggested that among Gleason grade 4 growth patterns, cribriform growth is associated with a worse clinical outcome.[3-6]

One of the major limitations of the Gleason grading system is its considerable inter-observer variability.[7-19] Fused glands or small glands without lumina may, for instance, be interpreted as tangentially sectioned Gleason grade 3 or as a focal Gleason grade 4 prostate cancer.[15] Despite this inter-observer variability, accurate identification of Gleason grade 4 prostate cancer is important for subsequent therapeutic decision-making.[7-19] For instance, patients with Gleason score 3+3=6 prostate cancer are often eligible for active surveillance, while patients with 3+4=7 are frequently excluded.[20] To support future studies on Gleason grade 4 architectural sub-classification, it is important that individual growth patterns are well-defined and reproducible. We therefore undertook an inter-observer variability study among genitourinary

pathologists to assess the inter-observer agreement of individual Gleason grade 4 growth patterns.

MATERIALS AND METHODS

Case selection

Case selection (n=60) was undertaken by two investigators (C.K., G.v.L.) and included a representative collection of various Gleason grade 4 growth patterns.[1] The selection aimed at including 10 Gleason grade 3, 40 Gleason grade 4 (10 per growth pattern) and 10 Gleason grade 5 cases, several of which deliberately showed ambiguous Gleason grade 4 patterns. In this study, we did not include ductal, intraductal or small cell prostate cancer. Digital images of hematoxylin and eosin–stained slides from selected prostate cancer cases were obtained using a NanoZoomer digital slide scanner (Hamamatsu Photonics K.K., Hamamatsu City, Japan). To ascertain that all participants evaluated the same tumour structures, tumour regions of interest were delineated in each case with a yellow line.

Images were incorporated into a digital questionnaire. An international group of 26 genitourinary pathologists was invited to participate in the study. Investigators involved in the case selection (C.K., G.v.L.) were not included in this group. Participants were asked to select a single predominant Gleason grade per case (3, 4, 5). If a case was assigned a Gleason grade 4, participants were asked to indicate a single predominant growth pattern (ill-formed, fused, cribriform or glomeruloid). Participants had the opportunity to add comments by free-text. Finally, we collected basic demographic information from participants. No consensus training preceded the study. The questionnaire including all cases and instructions is shown in *Supplemental Appendix* 1.

Analysis

Consensus was defined as at least 80% agreement on Gleason grade or Gleason grade 4 sub-classification. When 60% to 80% of the participants agreed, the classification was considered as "favoured". The latter thresholds were chosen arbitrarily. Although participants were asked to select only one predominant grade and/or pattern per case, a combination of grades or patterns, e.g. "grade 3 & 4" or "fused & ill-formed", was assigned in a few cases. We considered these combinations as separate categories in the analysis. This also applied to cases that were assigned by free-text as Gleason grade 2, which was not a selection option. In addition, some cases had been assigned "Gleason grade 3 cribriform" (n=4) or "Gleason grade 3 glomeruloid" (n=1). The latter cases were considered as Gleason grade 3 for the analysis. Their growth patterns, however, were included in Gleason grade 4 pattern analysis.

RESULTS

Participants

Replies were received from 23 genitourinary pathologists (88% response rate) residing in 13 countries: United States (n=8), Canada (n=3), United Kingdom (n=2), Australia (n=1), Brazil (n=1) France (n=1), Germany (n=1), Italy (n=1), Singapore (n=1), Spain (n=1), Sweden (n=1), Switzerland (n=1), and New Zealand (n=1). Twenty participants were in academic practice (87%), 2 in private practice (8.7%), and 1 in academic, community and private practice (4%). The number of diagnostic prostate biopsies analysed per year was as follows: 100-250 (n=1; 4%), 250-500 (n=7; 30%), 500-1000 (n=7; 30%), and more than 1000 (n=8; 35%).

Gleason grading

Consensus on Gleason grade was reached in 47 (78%) cases: 8 had Gleason grade 3, 35 Gleason grade 4, and 4 Gleason grade 5. In 9 cases (15%) a Gleason grade was favoured: 4 cases were favoured as Gleason grade 3, 1 Gleason grade 4, and 4 Gleason grade 5. In 4 cases (7%) there was less than 60% agreement on Gleason grading (Figure 1). If less than 80% of participants agreed on Gleason grade, ill-formed (6/13, 46%) and fused (7/13, 54%) were the involved growth patterns . The frequencies of the Gleason grades and grade 4 patterns per case are listed in *Supplemental table 1*.

Gleason grade 4 patterns

The number of cases in which a pattern 4 was assigned by at least 1 pathologist was as follows: fused in n=41 (68%), cribriform n=30 (50%), ill-formed n=20 (33%) and glomeruloid n=13 (22%). Consensus on Gleason grade 4 pattern was reached in 13 (37%) out of 35 consensus Gleason grade 4 cases, 7 of which were on cribriform (54%), 5 on glomeruloid (38%), and 1 on fused (8%). None of the cases reached consensus for ill-formed glands.

Among the 13 cases where at least 1 pathologist assigned a glomeruloid growth pattern, consensus on glomeruloid was present in 5 (38%) cases (Case 12, 18, 25, 29 and 31). The glomeruloid pattern was favoured in 2 (15%) cases (Case 15 and 38); in both cases a considerable number of participants preferred cribriform, or a combination of both. Consensus on glomeruloid pattern was found in cases showing relatively small glomerulations, while the favoured cases all comprised large glomeruloid formations in which cribriform pattern was the preferred alternative (Figure 2).

Among the 30 cases where at least 1 participant assigned cribriform pattern, consensus on cribriform pattern was reached in 7 (23%) cases (Case 4, 9, 17, 27, 30, 43 and 48), 2 of which (Case 4 and 48) reached 100% agreement. All 7 consensus cases had large cribriform glands with multiple punched-out lumina (Figure 3A-D). When cribriform pattern was favoured (n=2; 7%), fused glands were also considered by other participants in these cases (Case 16 and 59). Case 16 represented a relatively small gland containing several small mucin-filled spaces and delicate intervening stroma containing capillaries (Figure 3E). Case 59 showed a large glandular proliferation with elongated and round lumina and discrete fibro-vascular cores (Figure 3F).

No consensus was reached on the ill-formed pattern in any of the cases, although this pattern was favoured in 5 (25%) (Case 2, 5, 7, 39 and 45; Figure 4). In all of these cases, the fused pattern was also considered by other participants. Consensus on the fused pattern was reached in only 1 (2%) case (Case 46), while the fused pattern was favoured in 4 (10%) (Case 19, 32, 54 and 57; Figure 5). In the latter cases, cribriform was also given by others. Here, the tumour glands had a complex architecture demonstrating anastomosing glands with elongated lumina and delicate intervening stroma and capillaries.

There was less than 80% agreement on the pattern sub-classification in 22/35 consensus Gleason grade 4 cases. In these cases, pattern disagreement occurred mostly on fused *versus* ill-formed and fused *versus* cribriform. Cases in which no agreement could be reached on fused or ill-formed pattern showed glands that were relatively small with an irregular border (Case 3 and 36; Figure 6 A-B). The other type of disagreement was characterized by complex epithelial anastomosing glands with delicate intervening fibro-vascular cores ("complex fused"), which were alternately assigned fused or cribriform pattern (Case 20, 23, 37, 40, 42 and 53, Figure 6C-H). The

distribution of the Gleason grade 4 patterns related to disagreement on Gleason grading or grade 4 sub-classification are summarized in Figure 7.

DISCUSSION

To facilitate future studies on Gleason grade 4 sub-classification, it is important that individual growth patterns are well-defined and reproducible. We therefore initiated an inter-observer study among genitourinary pathologists to assess the inter-observer agreement of individual Gleason grade 4 growth patterns. We found that consensus on Gleason grade 4 growth pattern was predominantly reached on cribriform and glomeruloid, but hardly on ill-formed and fused glands. Although most participants agreed on classifying large epithelial proliferations with multiple punched-out lumina lacking intervening stroma as the cribriform pattern, there was considerable inter-observer variation in the sub-classification of large complex epithelial proliferations with subtle intervening stroma; these were alternately regarded as cribriform or fused. To our knowledge, this histological category "complex fused" has not been described previously. Also, glomerulations with large intra-luminal cribriform proliferations were alternately regarded as cribriform or glomeruloid.

Fused and ill-formed glands were the two patterns mostly related to Gleason grade variability (3, 4 or 5). Our findings are in line with two previous studies by Egevad et al. who looked at the inter-observer reproducibility of Gleason grade 4 patterns among 15 and 337 pathologists respectively.[21, 22] In one of these studies, the percentage of fused and ill-formed glands was inversely correlated with agreement among pathologists, while the cribriform pattern had no significant correlation with inter-observer variability.[22] Zhou et al. recently suggested that adjacent tumour glands play an important role in decision-making in cases showing ambiguous ill-formed

patterns.[19] They recommend that >10 poorly formed glands not immediately adjacent to other well-formed glands, should be considered ill-formed Gleason pattern 4. In contrast, poorly formed glands that are intermixed with well-formed glands, or ≤5 poorly formed glands regardless of their location should be diagnostic features against Gleason pattern 4. Because our study specifically focused on growth pattern sub-classification and lacked the presence of sufficient adjacent tumour glands, we were unable to make such observations.

Although Gleason grading is subject to inter-observer variability, accurate identification of Gleason grade 4 prostate cancer is important for subsequent therapeutic decision-making.[7-19] In many institutes, patients with Gleason score 6 prostate cancer are candidates for active surveillance, while patients with Gleason score 7 generally undergo therapeutic intervention.[20] It has been estimated that up to 13% of patients would have been recommended different treatment solely based on pathologic re-evaluation of diagnostic biopsies.[13, 18]. Although the presence of a Gleason pattern 4 is a clinical threshold for active treatment, Gleason grade 4 comprises a heterogeneous tumour group, covering at least 4 morphologically distinct patterns. Recent studies have suggested that individual Gleason grade 4 growth patterns are associated with clinical outcome.[3-6] Dong et al. found in a consecutive series of 214 patients that the presence of cribriform growth was associated with a shorter time to biochemical recurrence and metastasis after radical prostatectomy, while fused and ill-formed glands were not.[5] In a cohort of 161 Gleason score 7 prostate cancer patients, our group also found that the presence of cribriform growth was associated with a worse metastasis-free survival and disease-specific survival, while fused, glomeruloid and ill-formed glands were not.[6] In a screen-detected cohort of 1031 men, we recently reported similar findings for cribriform growth in diagnostic biopsies.[23] We therefore recommend reporting

presence of cribriform tumour glands in pathology reports, as this pattern appears to confer a less favourable outcome.

The present study has several limitations. Cases were selected by two investigators, and intentionally included both classic and ambiguous cases. For instance, large epithelial proliferations with punched-out lumina and subtle scattered fibro-vascular cores, which are difficult to classify as either cribriform or fused, are relatively uncommon. It is therefore difficult to draw conclusions on the general agreement of Gleason grade 4 growth patterns in daily clinical practice. Secondly, participants were asked to score delineated tissue areas to ensure categorisation discordances were not due to evaluation of different tumour areas. However, the participants were therefore not able to interpret tumour growth patterns in a larger context.

The present study did not make use of a statistical measure of agreement, such as Cohen's κ or Krippendorff's α.[23, 24] Due to study design, in which a Gleason grade 4 sub-classification only had to be indicated when Gleason grade 4 was present, we had missing values for grade 4 sub-classification when participants assigned a case a Gleason grade 3 or 5. Cohen's κ was therefore not a suitable statistical measure in this context, as it is known to be highly influenced by missing values.[25] Although Cohen's κ is commonly used in the medical literature, its value, however, may not be sufficiently informative because it relates the proportion of observed agreement to variation in a sample (i.e. relative measure), while the clinician's question of observer variation in individual cases calls for an absolute measure, i.e. percentage agreement.[26]

In conclusion, consensus on Gleason grade 4 growth pattern was predominantly reached on glomeruloid and cribriform patterns, and rarely on ill-formed and fused patterns. These data indicate that Gleason grade 4 cribriform morphology, which has

been associated with a worse clinical outcome, is a reasonably reproducible pattern in Gleason grade 4 prostate cancer. The complex fused glands seem to be a borderline pattern of unknown prognostic significance on which a consensus could not be reached.

LIST OF ONLINE SUPPORTING INFORMATION

Supplemental Appendix 1. The questionnaire including all cases and instructions.

Supplemental Table 1. The frequencies of the Gleason grades and grade 4 patterns per case

REFERENCES

1. Epstein JI, C AW, Jr., Amin MB, Egevad LL, Committee IG. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am J Surg Pathol* 2005 Sep;**29**(9);1228-1242.

2. Epstein JI, Egevad L, Amin MB *et al.* The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol* 2016 Feb;**40**(2):244-252.

3. Iczkowski KA, Torkko KC, Kotnis GR *et al.* Digital quantification of five high-grade prostate cancer patterns, including the cribriform pattern, and their association with adverse outcome. *Am J Clin Pathol* 2011 Jul;**136**(1);98-107.

4. Kryvenko ON, Gupta NS, Virani N *et al.* Gleason score 7 adenocarcinoma of the prostate with lymph node metastases: analysis of 184 radical prostatectomy specimens. *Arch Pathol Lab Med* 2013 May;**137**(5);610-617.

5. Dong F, Yang P, Wang C *et al.* Architectural heterogeneity and cribriform pattern predict adverse clinical outcome for Gleason grade 4 prostatic adenocarcinoma. *Am J Surg Pathol* 2013 Dec;**37**(12);1855-1861.

6. Kweldam CF, Wildhagen MF, Steyerberg EW, Bangma CH, van der Kwast TH, van Leenders GJ. Cribriform growth is highly predictive for postoperative metastasis and disease-specific death in Gleason score 7 prostate cancer. *Mod Pathol* 2015 Mar;**28**(3);457-464.

7. Allsbrook WC, Jr., Mangold KA, Johnson MH, Lane RB, Lane CG, Epstein JI. Interobserver reproducibility of Gleason grading of prostatic carcinoma: general pathologist. *Hum Pathol* 2001 Jan;**32**(1);81-88.

8. Allsbrook WC, Jr., Mangold KA, Johnson MH *et al.* Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. *Hum Pathol* 2001 Jan;**32**(1);74-80.

9. Egevad L. Reproducibility of Gleason grading of prostate cancer can be improved by the use of reference images. *Urology* 2001 Feb;**57**(2);291-295.

10. Glaessgen A, Hamberg H, Pihl CG, Sundelin B, Nilsson B, Egevad L. Interobserver reproducibility of modified Gleason score in radical prostatectomy specimens. *Virchows Arch* 2004 Jul;**445**(1);17-21.

11. Griffiths DF, Melia J, McWilliam LJ *et al.* A study of Gleason score interpretation in different groups of UK pathologists; techniques for improving reproducibility. *Histopathology* 2006 May;**48**(6);655-662.

12. Melia J, Moseley R, Ball RY *et al.* A UK-based investigation of inter- and intra-observer reproducibility of Gleason grading of prostatic biopsies. *Histopathology* 2006 May;**48**(6);644-654.

13. Berg KD, Toft BG, Roder MA, Brasso K, Vainer B, Iversen P. Prostate needle biopsies: interobserver variation and clinical consequences of histopathological re-evaluation. *APMIS* 2011 Apr;**119**(4-5);239-246.

14. Rodriguez-Urrego PA, Cronin AM, Al-Ahmadie HA *et al.* Interobserver and intraobserver reproducibility in digital and routine microscopic assessment of prostate needle biopsies. *Hum Pathol* 2011 Jan;**42**(1);68-74.

15. McKenney JK, Simko J, Bonham M *et al.* The potential impact of reproducibility of Gleason grading in men with early stage prostate cancer managed by active surveillance: a multi-institutional study. *J Urol* 2011 Aug;**186**(2);465-469.

16. Helpap B, Kristiansen G, Beer M *et al.* Improving the reproducibility of the Gleason scores in small foci of prostate cancer--suggestion of diagnostic criteria for glandular fusion. *Pathol Oncol Res* 2012 Jul;**18**(3);615-621.

17. Berney DM, Algaba F, Camparo P *et al.* The reasons behind variation in Gleason grading of prostatic biopsies: areas of agreement and misconception among 266 European pathologists. *Histopathology* 2014 Feb;**64**(3);405-411.

18. Thomsen FB, Marcussen N, Berg KD *et al.* Repeated biopsies in patients with prostate cancer on active surveillance: clinical implications of interobserver variation in histopathological assessment. *BJU Int* 2015 Apr;**115**(4);599-605.

19. Zhou M, Li J, Cheng L *et al.* Diagnosis of "Poorly Formed Glands" Gleason Pattern 4 Prostatic Adenocarcinoma on Needle Biopsy: An Interobserver Reproducibility Study Among Urologic Pathologists With Recommendations. *Am J Surg Pathol* 2015 Oct;**39**(10);1331-1339.

20. Dall'Era MA, Albertsen PC, Bangma C *et al.* Active surveillance for prostate cancer: a systematic review of the literature. *Eur Urol* 2012 Dec;**62**(6);976-983.

21. Egevad L, Algaba F, Berney DM *et al.* Interactive digital slides with heat maps: a novel method to improve the reproducibility of Gleason grading. *Virchows Arch* 2011 Aug;**459**(2);175-182.

22. Egevad L, Ahmad AS, Algaba F *et al.* Standardization of Gleason grading among 337 European pathologists. *Histopathology* 2013 Jan;**62**(2);247-256.

23. Kweldam CF, Kümmerlin IP, Nieboer D *et al.* Disease-specific survival of patients with invasive cribriform and intraductal prostate cancer at diagnostic biopsy. *Mod Pathol* 2016 Mar 4.

FIGURE LEGENDS

Figure 1. Cases with less than 60% agreement on Gleason grade. A) **Case 22**. Grade 3 n=11, grade 4 n=12 (fused n=12). B) **Case 24**. Grade 4 n=11, grade 5 n=11, grade 4 & 5 n=1 (ill-formed n=7, fused n=2, cribriform n=1, fused & ill-formed n=2). C) **Case 51**. Grade 4 n=13, grade 5 n=8, grade 4 & 5 n=2 (fused n=1, ill-formed n=11, fused & ill-formed n=2). D) **Case 58**. Grade 3 n=12, grade 4 n=11 (fused n=7, cribriform n=4, cribriform grade 3 n=1).

Figure 2. Agreed (A, B) and favoured (C, D) cases with glomeruloid pattern. A) **Case 18**. Glomeruloid n=20, cribriform n=1, cribriform & glomeruloid n=1. B) **Case 29**. Glomeruloid n=19, cribriform n=3, fused n=1. C) **Case 15**. Glomeruloid n=16, cribriform n=5, cribriform & glomeruloid n=2. D) **Case 38**. Glomeruloid n=14, cribriform n=4, cribriform & glomeruloid n=5.

Figure 3. Agreed (A-D) and favoured (E, F) cases with cribriform pattern. A) **Case 4**. Cribriform n=23. B) **Case 9**. Cribriform n=21, fused n=2. C) **Case 27**. Cribriform n=20, glomeruloid n=2, cribriform grade 3 n=1. D) **Case 48**. Cribriform n=23. E) **Case 16.** Cribriform n=14, fused n=6, glomeruloid n=1, cribriform & glomeruloid n=1, cribriform & fused n=1. F) **Case 59**. Cribriform n=15, fused n=7, cribriform & fused n=1.

Figure 4. Cases with favoured ill-formed pattern. A) **Case 2**. Ill-formed n=14, fused n=7, fused & ill-formed n=1. B) **Case 5**. Ill-formed n=18, fused n=2, fused & ill-formed n=1. C) **Case 39**. Ill-formed n=16, fused n=2, fused & ill-formed n=2. D) **Case 45**. Ill-formed n=18, fused n=3, fused & ill-formed n=1.

Figure 5. Cases with favoured fused pattern. A) **Case 19**. Fused n=16, cribriform n=2, fused & ill-formed n=1. B) **Case 32**. Fused n=17, cribriform n=5, ill-formed n=1. C) **Case 54**. Fused n=17, cribriform n=2, cribriform & fused n=1, cribriform & ill-formed n=1. D) **Case 57**. Fused n=14, cribriform n=4, ill-formed n=1, fused & ill-formed n=1, cribriform & fused n=1.

Figure 6. Cases with less than 80% agreement on Gleason grade 4 sub-classification. A) **Case 3**. Ill-formed n=13, fused n=4, fused & ill-formed n=3. B) **Case 36**. Ill-formed n=13, fused n=4, fused & ill-formed=3. C) **Case 23**. Fused n=12, cribriform n=7, cribriform & fused n=2, fused & ill-formed n=1. D) **Case 20**. Fused n=13, cribriform n=6, cribriform & fused n=1, cribriform & fused & ill-formed =1. E) **Case 37**. Cribriform n=11, fused n=6, glomeruloid n=1, fused & glomeruloid n=1, cribriform & glomeruloid n=1, fused & ill-formed & glomeruloid n=1. F) **Case 40**. Cribriform n=13, fused n=8, fused & ill-formed

n=1. G) **Case 42**. Cribriform n=13, fused n=7, cribriform & fused n=1. H) **Case 53**. Fused n=11, cribriform n=10, fused & ill-formed n=1, cribriform & ill-formed n=1.

Figure 7. Schematic overview showing types of disagreement in cases with less than 80% consensus on Gleason grading (n=13/60, blue arrows) and grade 4 sub-classification (n=22/35, red arrows). The percentages represent the proportion of cases involved in a particular type of disagreement. The ellipse represents Gleason grade 3, the trapezium Gleason grade 5, and the rectangles Gleason grade 4 patterns.