

Immunohistochemical biomarker validation in highly selective needle biopsy microarrays derived from mpMRI-characterized prostates.

Jonathan Olivier^{1,2,3,*}, Vasilis Stavrinides^{1,3,*}, Jonathan Kay^{1,3}, Alex Freeman⁴, Hayley Pye^{1,3}, Zeba Ahmed^{1,3}, Lina Carmona^{1,3}, Susan Heavey^{1,3}, Lucy A. M. Simmons^{3,5}, Abi Kanthabalan^{3,5}, Manit Arya⁵, Tim Briggs^{5,6}, Dean Barratt⁷, Susan C. Charman^{8,9}, James Gelister⁶, David Hawkes⁷, Yipeng Hu⁷, Charles Jameson⁴, Neil McCartan^{3,5}, Shonit Punwani¹⁰, Jan van der Muelen^{4,8}, Caroline Moore^{3,5}, Mark Emberton^{3,5}, Hashim U. Ahmed^{3,11,12}, Hayley C. Whitaker^{1,3}.

¹Molecular Diagnostics and Therapeutics Group, University College London, UK, ²Department of Urology, Hospital Huriez, University Lille Nord de France, Lille, France, ³Division of Surgery and Interventional Science, Faculty of Medical Sciences, University College London, London, United Kingdom, ⁴Department of Pathology, UCLH NHS Foundation Trust, London, UK, ⁵Department of Urology, UCLH NHS Foundation Trust, London, UK, ⁶Department of Urology, The Royal Free London NHS Foundation Trust, London, UK, ⁷Centre for Medical Imaging and Computing, Department of Computer Science, University College London, London, UK, ⁸Clinical Effectiveness Unit, The Royal College of Surgeons of England, London, UK, ⁹Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK, ¹⁰ Department of Radiology, UCLH NHS Foundation Trust, London, UK, ¹¹Division of Surgery, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK, ¹²Imperial Urology, Charing Cross Hospital, Imperial College Healthcare NHS Trust, London, UK.

* these authors contributed equally to this manuscript.

Keywords: prostate cancer, tissue microarrays, MRI, immunohistochemistry

Address for correspondence:

Dr. Hayley Whitaker

Molecular Diagnostics and Therapeutics Group, Charles Bell House, Division of Surgery and Interventional Science, University College London, W1W 7TS.

E.mail: Hayley.Whitaker@ucl.ac.uk

The authors certify that they have no conflicts of interest.

ABSTRACT

Introduction: Diagnosing prostate cancer routinely involves tissue biopsy and increasingly image guided biopsy using multiparametric MRI (mpMRI). Excess tissue after diagnosis can be used for research to improve the diagnostic pathway and the vertical assembly of prostate needle biopsy cores into tissue microarrays (TMAs) allows the parallel immunohistochemical (IHC) validation of cancer biomarkers in routine diagnostic specimens. However, tissue within a biopsy core is often heterogeneous and cancer is not uniformly present, resulting in needle biopsy TMAs that suffer from highly variable cancer detection rates that complicate parallel biomarker validation.

Materials and methods: The prostate cores with the highest tumour burden (in terms of Gleason score and/or maximum cancer core length) were obtained from 249 patients in the PICTURE trial who underwent transperineal template prostate mapping (TPM) biopsy at 5 mm intervals preceded by mpMRI. From each core, 2 mm segments containing tumour or benign tissue (as assessed on H&E pathology) were selected, excised and embedded vertically into a new TMA block. TMA sections were then IHC-stained for the routinely used prostate cancer biomarkers PSA, PSMA, AMACR, p63 and MSMB and assessed using the h-score method. H-scores in patient matched malignant and benign tissue were correlated with the Gleason grade of the original core and the MRI Likert score for the sampled prostate area.

Results: A total of 2240 TMA cores were stained and IHC h-scores were assigned to 1790. There was a statistically significant difference in h-scores between patient matched malignant and adjacent benign tissue that is independent of Likert score. There was no association between the h-scores and Gleason grade or Likert score within each of the benign or malignant groups.

Conclusion: The construction of highly selective TMAs from prostate needle biopsy cores is possible. IHC data obtained through this method are highly reliable and can be correlated with imaging. IHC expression patterns for PSA, PSMA, AMACR, p63 and MSMB are distinct in malignant and adjacent benign tissue but did not correlate with mpMRI Likert score.

INTRODUCTION

Prostate cancer is one of the most common non-cutaneous cancer in males, with approximately 417,000 new cases diagnosed in 2012 in Europe alone¹. PSA testing has resulted in an increase in prostate cancer incidence and to a diagnostic migration towards smaller, low-grade disease with low metastatic potential and limited impact on mortality²⁻⁴. In the Western world, it is common practice to diagnose prostate cancer through transrectal, ultrasound-guided systematic needle biopsies (TRUS) in PSA-detected men. Clinico-pathological parameters obtained through this approach including serum PSA, Gleason grade and maximum cancer core length on biopsy are often used to stratify risk and guide patient management. In the last decade mpMRI has emerged as an important technique for characterising and targeting the biopsy of suspected prostate cancer, as it reduces the number of unnecessary biopsies and efficiently detects clinically significant targets without over-diagnosing insignificant disease⁵⁻⁸.

Incorporating tissue biomarkers in the patient stratification process could further refine the emerging imaging-based patient pathways, but selecting molecules for such purposes requires their parallel testing in very small amounts of diagnostic tissue. Tissue microarrays (TMAs) constructed from prostate needle biopsies are a promising tool for high-throughput biomarker development and validation^{9,10}. Numerous strategies have been proposed for maximising needle biopsy TMA performance, but a simple and productive approach is the vertical re-orientation of biopsy cores for the construction of high-density arrays^{11,12}.

However, a recurring challenge in needle biopsy TMAs is the significant variability in their cancer content. This is generally due to (i) the considerable heterogeneity of prostate cancer, (ii) random tissue sampling approaches that often result in disease misrepresentation (which is not necessarily the case with TMAs derived from prostatectomy specimens) and (iii) tissue loss during sampling, fixation, embedding or staining. These difficulties are further complicated by the scarcity of biopsy material, which is very precious and cannot be easily substituted if TMA quality is poor.

It follows that, for biomarker validation purposes, the ideal diagnostic needle biopsy TMA should (i) incorporate a large number of specimens, (ii) contain tissue from well-characterised prostate areas with clinically significant disease (iii) have a high cancer detection rate for maximum performance and (iv) produce results that can be correlated with imaging data. Here, we present the construction of a biopsy TMA from prostates thoroughly characterised using mpMRI and 5mm transperineal mapping (TPM) biopsies. We divided the biopsy cores in segments such that only either benign or malignant tissue was included in a specific array position. To test our tissue-selective TMAs, we performed IHC for routinely used prostate biomarkers and correlated IHC h-scores with imaging parameters and original pathology.

MATERIALS AND METHODS

Patient cohort

All prostate tissue was acquired during the PICTURE trial, a paired-cohort confirmatory study designed to assess the accuracy of mpMRI in detecting clinically significant cancer¹³. For this purpose, 249 men with a previous TRUS biopsy requiring a repeat evaluation underwent a 3T mpMRI followed by TPM biopsies of the entire prostate at 5 mm intervals. The likelihood of significant cancer by mpMRI was reported using the Likert scale, as previously defined¹⁴. In all MRI scans, the base, middle and apex of the prostate were divided in 4 quadrants resulting in Likert scores assigned to a total of 12 prostate areas for each patient. Ethical approval previously given for the study allowed the use of needle biopsy specimens for TMA construction as described below.

For each patient, the pathology report was reviewed and the biopsy cores with the highest Gleason score and/or the longest maximum cancer core length (MCCL) were identified and selected for retrieval from the UCL/UCLH Biobank for Studying Health & Disease. All cores were previously fixed and routinely processed to formalin fixed paraffin embedded (FFPE) blocks. Haematoxylin and eosin (H&E)-stained 4 µm sections were evaluated by an expert uropathologist (AF and CJ) and tumor foci were identified and graded according to the Gleason grading system. A total of 448 tissue blocks were retrieved. Clinical data (disease stage, age, PSA at study entry) and Likert scores for all prostate regions were collected for all patients. A summary of the pathology of all selected cores and the distribution of their corresponding Likert scores is presented in Table 1.

Microarray construction

An overview of the TMA construction process is shown in Figure 1. Diagnostic 4µm H&E-stained sections were obtained and scanned using a Hamamatsu scanner. The digital H&E images were inspected and 2 mm-long benign or tumour regions were identified and then

marked on the original FFPE block for cutting. Each cutting plan was mutually agreed by at least two investigators. FFPE blocks were incubated at 60°C for 20 minutes and all marked tumour or benign biopsy core tissue was dissected and excised with a microtome blade. In this way, wax chips containing 2 mm-long core segments containing exclusively benign or tumour tissue according to the H&E cutting plan were produced. Each chip was re-marked for orientation purposes and placed onto a new individual plastic cassette before being re-embedded vertically. Once all benign or malignant core segments were vertically re-embedded, an Estigen MTA-1 Manual Tissue Arrayer was used to extract 1.5 x 6 mm wax cores containing the vertical 2 mm core segments from each donor block and place them in the recipient wax block. Benign and tumour core segments were randomly positioned 0.7 mm apart in a 6 x 10 format. Liver tissue and blank positions were also used for orientation purposes. Each newly constructed TMA block was placed on a glass slide for 40 minutes at 60°C in an incubator and then cooled on a cold plate for tempering. Seven TMA blocks were produced and cut into 4 µm sections, with one slide every 50 retained for H&E staining and quality control.

Immunohistochemistry (IHC)

FFPE TMA sections were deparaffinised and rehydrated with successive 5-minute washes in xylene and alcohol (100% 90% and 70%). For all immunohistochemical stains, a Leica BOND-MAX Autostainer (Leica Biosystems, UK) was used. Heat-induced epitope retrieval was performed using either a pH 6.0 citrate-based or a pH 9.0 ethylene-diamine-tetra-acetic acid-based, ready-to-use solution (ER1 and ER2 respectively, Leica Biosystems, UK). All sections were incubated with the following primary antibodies under appropriately optimized conditions: PSA (rabbit polyclonal antibody, Dako A/S, Denmark; 1:9000 dilution, no retrieval), PSMA (mouse monoclonal antibody, clone 1D6, Leica Biosystems, UK; 1:50 dilution, ER1 for 20 min), p63 (mouse monoclonal antibody, clone 7JUL, Leica Biosystems, UK; 1:50 dilution, ER2 for 20 min), AMACR (rabbit monoclonal antibody, clone 13H4, Dako A/S, Denmark; 1:100 dilution, ER2 for 20 min) and MSMB (mouse monoclonal antibody, clone YPSP-1, Abcam, UK;

1:2500 dilution, enzymatic pre-digestion with Leica Biosystems Enzyme 1 for 15 min). Diaminobenzidine (DAB) was used as the chromogen and counterstaining was performed with haematoxylin for 1 minute. Following dehydration, the slides were cover-slipped using DPX (Leica Biosystems, UK).

IHC scoring

Digital images of the IHC slides were obtained using a Hamamatsu scanner. Each individual TMA core was assessed for the presence of cancer and h-scored by at least two independent investigators (HW, VS, ZA) without prior knowledge of clinical data. A proportion representing the estimated percentage of positively stained epithelial cells (0-100%) and an intensity score (0=none; 1=weak; 2=intermediate; 3=strong) were assigned to each core. A final h-score ranging from 0 to 300 was calculated by multiplying the proportion score with the intensity score (h-score = % no staining x 0 + % weak staining x 1 + % moderate staining x 2 + % strong staining x 3). The designation of a core section as “benign” or “malignant” was reassigned in cases of discrepancy with the H&E appearance in the original tissue block. Only tumour tissue was scored in cores containing both tumour and benign tissue. Missing cores or purely stromal areas were excluded from the analysis.

Data analysis

All statistical analyses and visualisation were performed in the R programming environment (<http://www.R-project.org/>, version 3.4.1). Continuous data distributions (h-scores) were tested for normality using quantile-quantile plots and the Shapiro-Wilk test. Due to substantial non-normality of the h-score distributions, the paired Wilcoxon signed rank procedure was used to test for significant differences in h-scores between malignant and paired, adjacent benign tissue. Kruskal-Wallis analysis of variation was used for comparisons between multiple groups. All tests were two-sided and a statistical significance level of 0.01 was considered significant.

RESULTS

Performance measures

Although other methods of producing efficient biopsy TMAs have been described, data on their performance in terms of cancer detection rates are not comprehensive⁸⁻¹¹. For this TMA seven blocks were constructed containing 448 core segments in total. Slides were IHC stained against five biomarkers (PSA, PSMA, AMACR, p63 and MSMB), yielding a theoretical maximum of 2240 (5 x 448) stained core sections. A summary of these results is given in Table 2. Of these, 338 (15%) were either lost during the staining process or were not assessable on scoring due to poor tissue quality. In addition, 112 (5%) core sections contained stroma only. When missing, un-assessable or purely stromal tissue was excluded from the analysis, 1790 (83%) of cores remained. Of these 371 (82.8%) core sections were h-scored for PSA, 345 (77%) for PSMA, 343 (76.6%) for p63, 367 (81.9%) for AMACR and 364 (81.3%) for MSMB. During h-scoring, each core section was re-evaluated to confirm that it contained tumour or benign tissue as designated in the original H&E cutting plan before vertical re-embedding. For each TMA slide three separate levels, each 50 slides apart, were assessed for pathology on H&E appearances and demonstrated consistent tumour or benign content at two or more levels for 81.9% of cores (Figure 2A and Table 3). In total, IHC and H&E appearances at a single level agreed in 1670 out of 1790 cases, with concordance in 349 cores stained for PSA, 324 for PSMA, 325 for p63, 342 for AMACR and 330 for MSMB. Concordance rates (i.e. number of h-scored core sections with IHC-H&E concordance/total number of h-scored core sections) were 94%, 94%, 95%, 93% and 90% for each stain, respectively.

IHC correlations with pathology and Gleason grade

Gleason grade is routinely used to indicate the aggressiveness of prostate cancer and markers preferentially diagnosing clinically significant disease (often characterised by the presence of \geq Gleason 4 pathology) are increasingly sought. For data analysis only patient matched pairs of malignant tissue and paired, adjacent benign tissue (from the same tissue block) were considered. The number of h-scored malignant-benign pairs was 105, 92, 101, 103 and 99 for PSA, PSMA, p63, AMACR and MSMB, respectively. There was a statistically significant difference between the h-scores for malignant and paired benign tissue for PSA ($p < 0.001$), PSMA ($p < 0.00001$), p63 ($p < 0.00001$), AMACR ($p < 0.00001$) and MSMB ($p < 0.00001$) (Figures 2B and 2C). Overall, AMACR and PSMA h-scores were higher in tumour tissue compared to matched benign, whereas the opposite was true for p63, MSMB and PSA. These differences were also seen when visualising h-scores for all cores (including unmatched) (Supplementary Figure 2), although no statistical tests were performed as each group contained a mixture of paired and unpaired values.

As Gleason grade ≥ 4 is often associated with more aggressive disease each of the benign and malignant groups were analysed separately to investigate h-score differences between different Gleason grades at diagnosis (Figure 3). Non-parametric analysis of variation failed to demonstrate any significant h-score difference between both the benign and tumour tissues originating from cores with different Gleason grades indicating that protein expression of these markers was associated with tumorigenesis but not aggressiveness of disease.

IHC correlations with mpMRI

mpMRI has been shown to efficiently diagnose clinically significant prostate tumours and is rapidly becoming a mainstay of prostate cancer diagnosis. Despite this, very few routinely used biomarkers have been studied in conjunction with mpMRI data. In this study paired h-scores were compared for different mpMRI Likert scores (Figure 4). When each benign and malignant group was divided to two Likert subgroups ("lower" Likert ≤ 3 versus "higher" Likert ≥ 4), there was a significant difference between the paired h-scores for PSMA, p63, AMACR

and MSMB in both subgroups ($p < 0.01$). PSA was the exception: although there was a significant h-score difference in the “lower” Likert subgroup ($p = 0.0023$), there was no similar difference in the “higher” Likert subgroup ($p = 0.0945$).

Different Likert scores were considered for each group (benign or malignant) separately (Supplementary Figure 1). There was no significant h-score difference in benign tissue with different Likert scores assigned to the prostate area of origin for any biomarker ($p > 0.1$). This was also the case for malignant tissue, although AMACR reached the level of marginal statistical significance ($p = 0.03872$), suggesting that there could be a difference in AMACR h-scores between malignant tissues from prostatic areas with different Likert scores.

CONCLUSION

Despite the large number of emerging genomic models the current predictive models for stratifying prostate cancer patients for treatment remain based on clinico-pathological variables such as age, serum PSA levels, disease stage and Gleason grade¹⁵. Using refined classification schemes which utilise biomarkers provides a route to increased predictive ability and personalised patient management¹⁶. Using additional immunohistochemical markers at diagnosis is a simple and cheap approach that does not require any additional infrastructure, allowing rapid implementation in a pathology laboratory. However, all novel markers require validation on large numbers of representative patient samples before widespread use and TMAs are a useful way of examining expression in large numbers of tissue samples simultaneously. However, the majority of TMAs are derived purely from radical prostatectomy tissue that are more likely to be of a lower stage and Grade and are not sampled in an unbiased manner. As a result most TMAs do not accurately represent the tissue biopsies used for routine diagnosis and bias any subsequent biomarker validation studies.

Utilising archival biopsy tissue for routine biomarker validation we have constructed tissue-selective microarrays from vertically re-arranged prostate needle biopsy samples for the purposes of parallel IHC and radiological biomarker validation (Figure 1). Numerous strategies have been proposed for maximising needle biopsy TMA performance, but a simple and productive approach is the vertical re-orientation of biopsy cores for the construction of high-density arrays^{9,10}. Although this approach has previously worked well in prostate tissue, the inherent tumour heterogeneity and low tumour content in TRUS biopsy samples results in TMAs where tumour content is often low or missing. In addition, true TMA performance is not routinely reported, at least meticulously^{11,12}.

The TMA we describe here contains only patient-matched MRI-characterized tissue cores containing the highest disease burden from 5 mm TPM biopsies or adjacent benign tissue. By using 2 mm core segments rather than entire needle biopsy cores we were able to ensure a significant degree of homogeneity and produce high-quality TMA measures (Table

2). Excluding missing/not assessable cores or cores containing only stroma, the IHC and H&E concordance rate was greater than 90% for all stains. This suggests that, in instances where a core section is present for h-scoring and contains epithelial tissue, the scorer can be fairly confident that it contains benign or tumour tissue as originally intended in the H&E cutting plan. The numbers of missing, un-assessable, stroma-containing, concordant or discordant cores were very similar between the five stains suggesting inter-slide reproducibility. We also demonstrated that tissue consistency was, on the whole, preserved along the entire TMA block, with concordance at two or more levels reaching almost 82%. TMA performance metrics are generally under-reported and our method of measuring performance could be widely adopted to facilitate comparisons between different needle biopsy TMA construction methods.

To demonstrate the tissue within the TMA is suitable for IHC it was used to assess expression of five widely used prostate biomarkers PSA, PSMA, p63, AMACR and MSMB¹⁷⁻²⁰. The differential expression of PSA, PSMA, p63, AMACR and MSMB (as represented by h-scores) differed considerably between malignant and paired, neighbouring benign tissue (Figure 2C) although h-scores within each of the benign or tumour groups were not associated with either Likert score or Gleason grade (Figure 3 and Supplementary Figure 2). However, when Likert score was grouped into lower risk (3 or lower) and higher risk (Likert 4/5) significant differences were seen for PSMA, p63, AMACR and MSMB, but not PSA in both risk groups (Figure 4).

Although the tissue was obtained through extensive TPM biopsies outside the standard of care we have demonstrated the feasibility, reproducibility and effectiveness of this TMA construction method and propose that it is possible to reproduce similar results with standard TURP or image-guided biopsies. This TMA represents a unique paired tissue, high quality, resource with clinical and radiological data which will allow validation of novel biomarkers correlated with imaging using a large number of biologically relevant patient samples.

Acknowledgements

We acknowledge the support of University College London in completing this study. We are grateful to Prostate Cancer UK for their support via the Prostate Cancer Centre of Excellence that supports the work in our laboratory and funding for HW and JK for the time spent working on this manuscript. We are also grateful to the European Association of Urology and Professor Arnauld Villers for supporting the work of JO. We would like to thank NIHR UCH/UCL Biomedical Research Centre who provide core funding for ME and the Wellcome Trust who provide funding to HA. We would also like to acknowledge the UCL/UCLH Biobank for Studying Health & Disease for assistance in accessing samples. Finally, we are very grateful to the patients who gave their samples for this research.

REFERENCES

1. Ferlay, J. *et al.* Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur. J. Cancer Oxf. Engl. 1990* **49**, 1374–1403 (2013).
2. Welch, H. G. & Albertsen, P. C. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005. *J. Natl. Cancer Inst.* **101**, 1325–1329 (2009).
3. Ross, H. M. *et al.* Do adenocarcinomas of the prostate with Gleason score (GS) ≤ 6 have the potential to metastasize to lymph nodes? *Am. J. Surg. Pathol.* **36**, 1346–1352 (2012).
4. Eggener, S. E. *et al.* Predicting 15-Year Prostate Cancer Specific Mortality After Radical Prostatectomy. *J. Urol.* **185**, 869–875 (2011).
5. Moore, C. M. *et al.* Image-Guided Prostate Biopsy Using Magnetic Resonance Imaging–Derived Targets: A Systematic Review. *Eur. Urol.* **63**, 125–140 (2013).
6. Schoots, I. G. *et al.* Magnetic resonance imaging-targeted biopsy may enhance the diagnostic accuracy of significant prostate cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis. *Eur. Urol.* **68**, 438–450 (2015).
7. Ahmed, H. U. *et al.* Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *The Lancet* **389**, 815–822 (2017).

8. Kasivisvanathan, V. *et al.* MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N. Engl. J. Med.* (2018). doi:10.1056/NEJMoa1801993
9. Jhavar, S. *et al.* Construction of tissue microarrays from prostate needle biopsy specimens. *Br. J. Cancer* **93**, 478–482 (2005).
10. Singh, S. S. *et al.* Feasibility of constructing tissue microarrays from diagnostic prostate biopsies. *The Prostate* **67**, 1011–1018 (2007).
11. McCarthy, F. *et al.* High-density tissue microarrays from prostate needle biopsies. *J. Clin. Pathol.* **64**, 88–90 (2011).
12. McCarthy, F. *et al.* An improved method for constructing tissue microarrays from prostate needle biopsy specimens. *J. Clin. Pathol.* **62**, 694–698 (2009).
13. Simmons, L. A. M. *et al.* The PICTURE study: diagnostic accuracy of multiparametric MRI in men requiring a repeat prostate biopsy. *Br. J. Cancer* **116**, 1159–1165 (2017).
14. Dickinson, L. *et al.* Magnetic Resonance Imaging for the Detection, Localisation, and Characterisation of Prostate Cancer: Recommendations from a European Consensus Meeting. *Eur. Urol.* **59**, 477–494 (2011).
15. D'Amico, A. V. *et al.* Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* **280**, 969–974 (1998).
16. Abdollah, F., Dalela, D., Haffner, M. C., Culig, Z. & Schalken, J. The Role of Biomarkers and Genetics in the Diagnosis of Prostate Cancer. *Eur. Urol. Focus* **1**, 99–108 (2015).
17. Epstein, J. I., Egevad, L., Humphrey, P. A. & Montironi, R. Best Practices Recommendations in the Application of Immunohistochemistry in the Prostate: Report From the International Society of Urologic Pathology Consensus Conference. *Am. J. Surg. Pathol.* **38**, e6–e19 (2014).
18. Bergström, S. H., Järemo, H., Nilsson, M., Adamo, H. H. & Bergh, A. Prostate tumors downregulate microseminoprotein-beta (MSMB) in the surrounding benign prostate epithelium and this response is associated with tumor aggressiveness. *The Prostate* (2017). doi:10.1002/pros.23466
19. Giannico, G. A., Arnold, S. A., Gellert, L. L. & Hameed, O. New and Emerging Diagnostic and Prognostic Immunohistochemical Biomarkers in Prostate Pathology. *Adv. Anat. Pathol.* **24**, 35–44 (2017).
20. Sjöblom, L. *et al.* Microseminoprotein-Beta Expression in Different Stages of Prostate Cancer. *PLoS One* **11**, e0150241 (2016).

Figure legends

Figure 1. TMA construction: Benign and malignant areas of 2 mm were identified within a biopsy core on H&E and selected for inclusion in the TMA (A). Each segment of core was divided according to the H&E cutting plan with a microtome blade, in order to obtain wax chips that contain either 2mm of malignant or benign tissue core (B). The wax chips were marked on their edges for orientation, re-positioned vertically and then embedded in a new paraffin donor block (C). All vertically re-embedded core segments are introduced into the final TMA block before tempering at 37°C for smoothing (D). In total, seven TMA slides were constructed and sectioned in their entirety, yielding 200-300 slides per TMA. The first slide of every 50 was stained with H&E for quality control while the rest were stored for immunohistochemistry after dipping in wax.

Figure 2. IHC for common prostate cancer biomarkers on paired samples: The seven TMA blocks were cut in their entirety and slides were wax dipped to prevent from oxidation. Slides at three levels (typically slide 50, 100, and 150) from each of the seven TMA slides were H&E stained and each core assessed for tumour or benign content. Concordance was measured as slides that had the same pathology at all three levels (dark green), two levels (pale green), one level (orange) and no levels (red). All IHC was performed on the BondMax Autostainer with staining shown in brown and nuclei are shown in blue. Representative images are shown for PSA, PSMA, p63, AMACR and MSMB (20x magnification) (B). IHC for paired samples was analysed using h-score method that takes into account staining intensity and the number of positively stained cells. Data for all samples (paired and unpaired) is shown in Supplementary Figure 1. H-scores in tumour tissue (black) were compared to paired h-scores in benign tissue (grey) from the same biopsy block using a paired Wilcoxon signed rank test. The p-values and number of pairs are separately shown for each stain (C).

Figure 3. IHC h-scores versus Gleason grade: All IHC was analysed using h-score method that takes into account staining intensity and the number of positively stained cells. H-scores for paired samples only in tumour tissue (black) and benign tissue (grey) are shown. All

associations were tested within either benign or tumour groups using Kruskal-Wallis analysis of variation. The p-values and number of pairs are separately shown for each stain.

Figure 4. TMA IHC h-scores versus appearance on prostate mpMRI: All IHC was analysed using h-score method that takes into account staining intensity and the number of positively stained cells. H-scores for paired samples only in tumour tissue (black) and benign tissue (grey) are shown. mpMRIs were graded using Likert score, a 5-point ordinal scale where Likert scores 1-2, Likert 3 and Likert 4-5 reflect a low, equivocal and high probability of underlying clinically significant disease, respectively. For this analysis non-visible mpMRI areas were defined as Likert 1-3 and visible lesions as Likert 4/5. A detailed breakdown of paired and unpaired h-scores combined against all Likert scores is provided in Supplementary Figure 2. H-scores were compared using a Wilcoxon test.

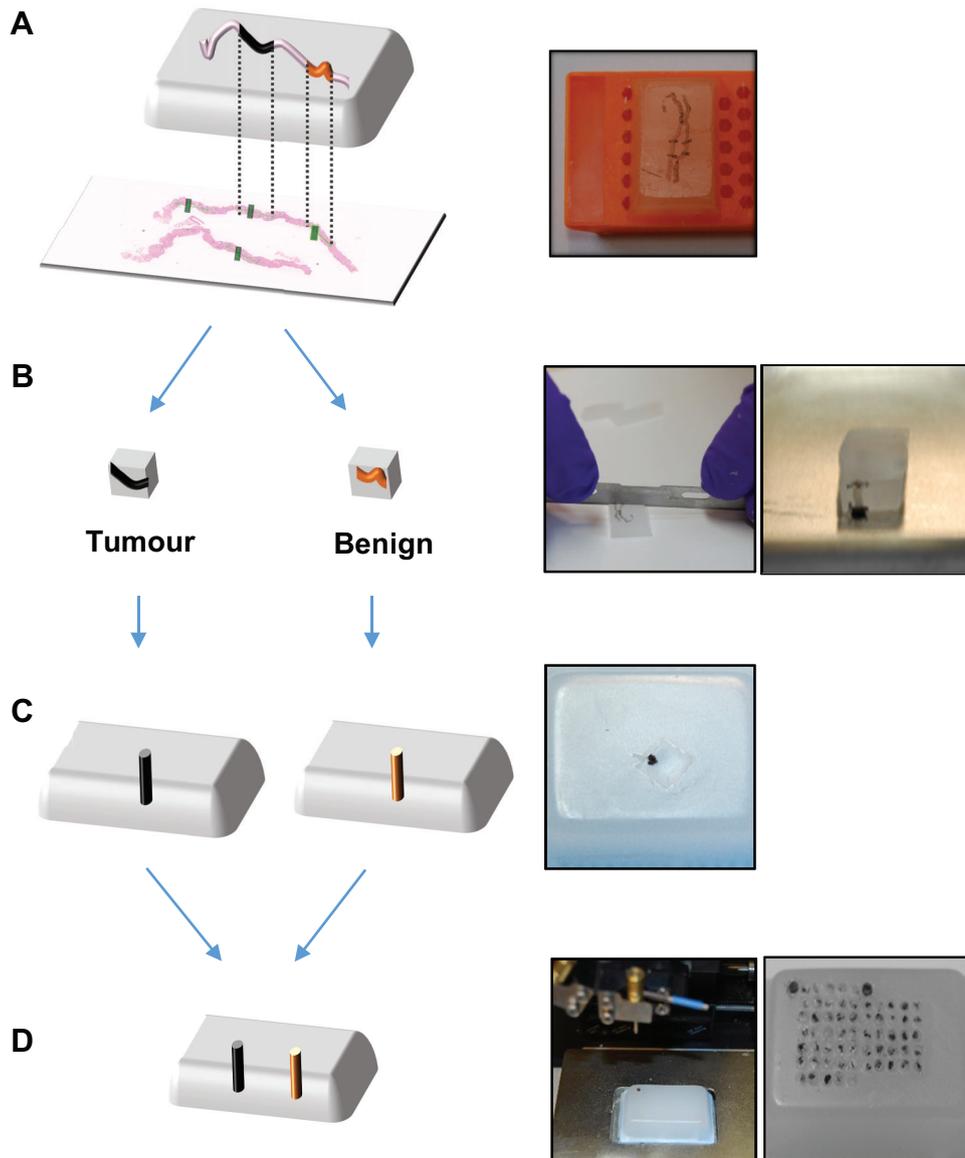
Table 1: Pathological and radiological characteristics of the cohort: TPM biopsy reports were scrutinised and cores with the highest Gleason grade and/or maximum cancer core length identified. True benign biopsies or biopsies containing only prostatic intraepithelial neoplasia (PIN) from patients without any cancer were also included in the TMA. MRI images of the base, middle and apex of each prostate were divided in quadrants and each quadrant assessed using a 5-point Likert scale for the likelihood of underlying clinically significant cancer (where scores of 4 or 5 denote a higher likelihood). For each TMA core, the corresponding Likert score for the sampled prostate quadrant was available.

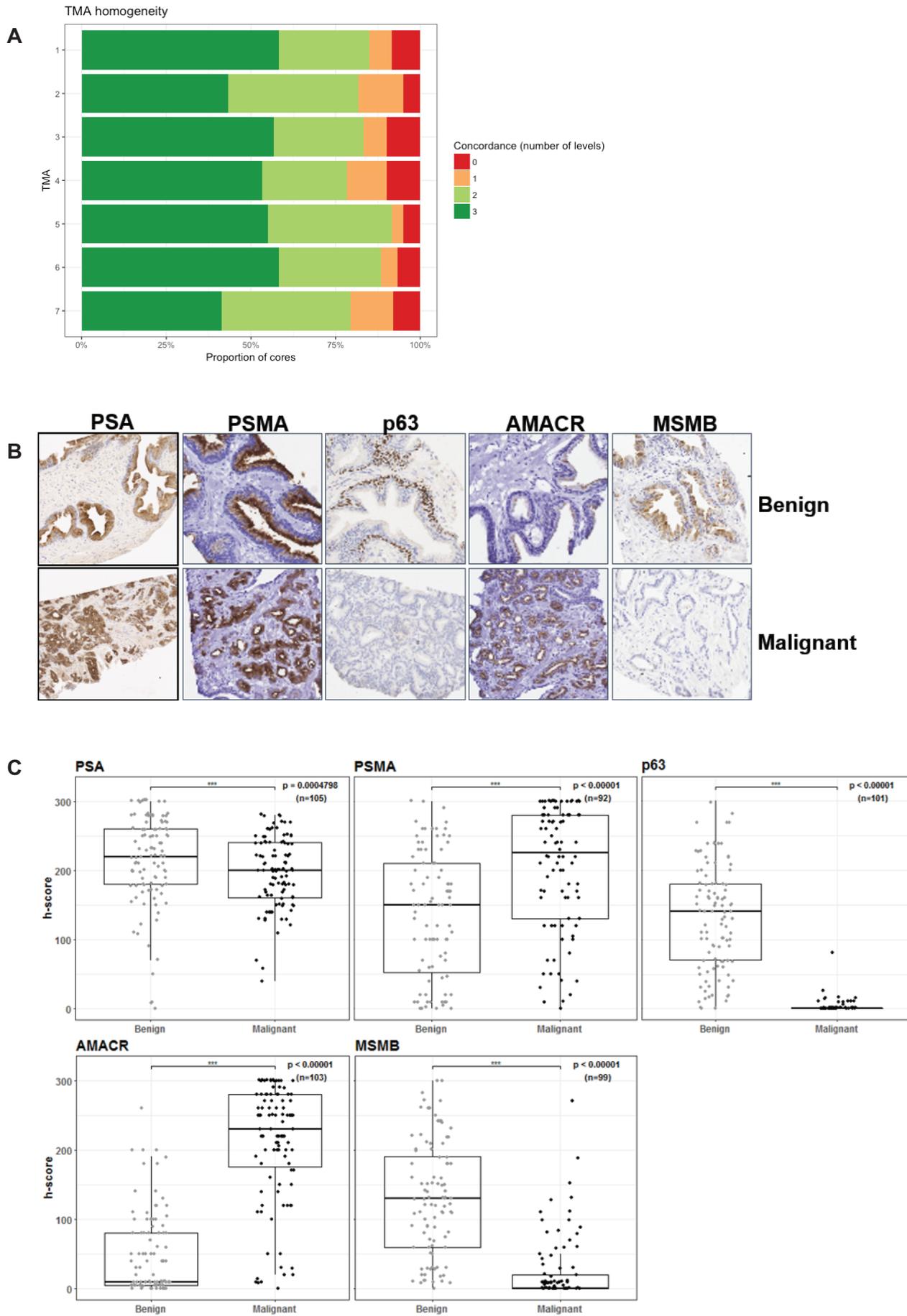
Table 2: TMA quality assessment: All IHC was performed on the BondMax Autostainer for PSA, PSMA, p63, AMACR and MSMB. Digital images of the IHC slides were obtained using a Hamamatsu scanner. Each individual TMA core was assessed for the presence of cancer and h-scored by at least two independent investigators (HW, VS, ZA) without prior knowledge of clinical data. The designation of a core section as “benign” or “malignant” was reassigned in cases of discrepancy with the H&E appearance in the original tissue block.

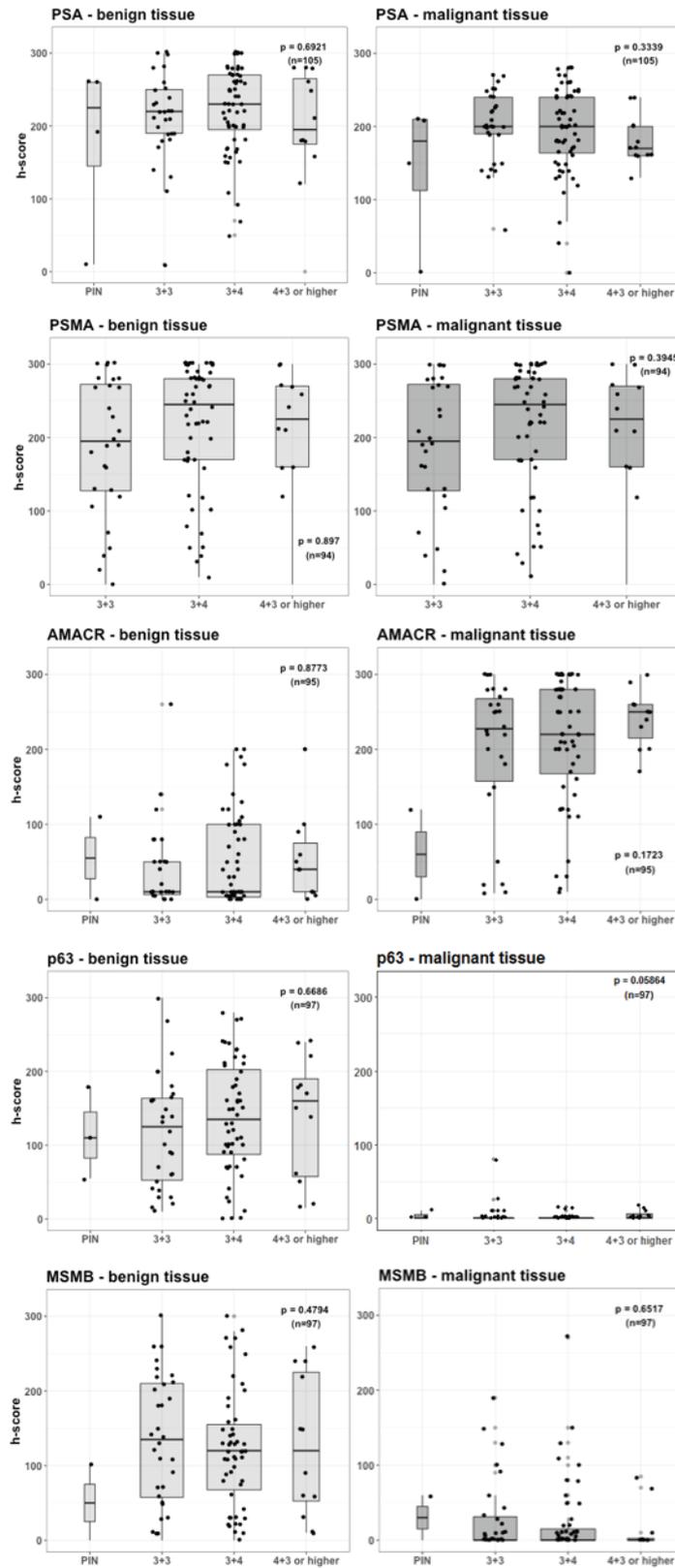
Table 3: Pathology concordance across the TMA. H&E slides were assessed every 50 slides over 150 slides of each TMA slide and assessed for tumour or benign pathology. Concordance was judged as partial when it agreed at two levels or full when all three levels exhibited the same pathology.

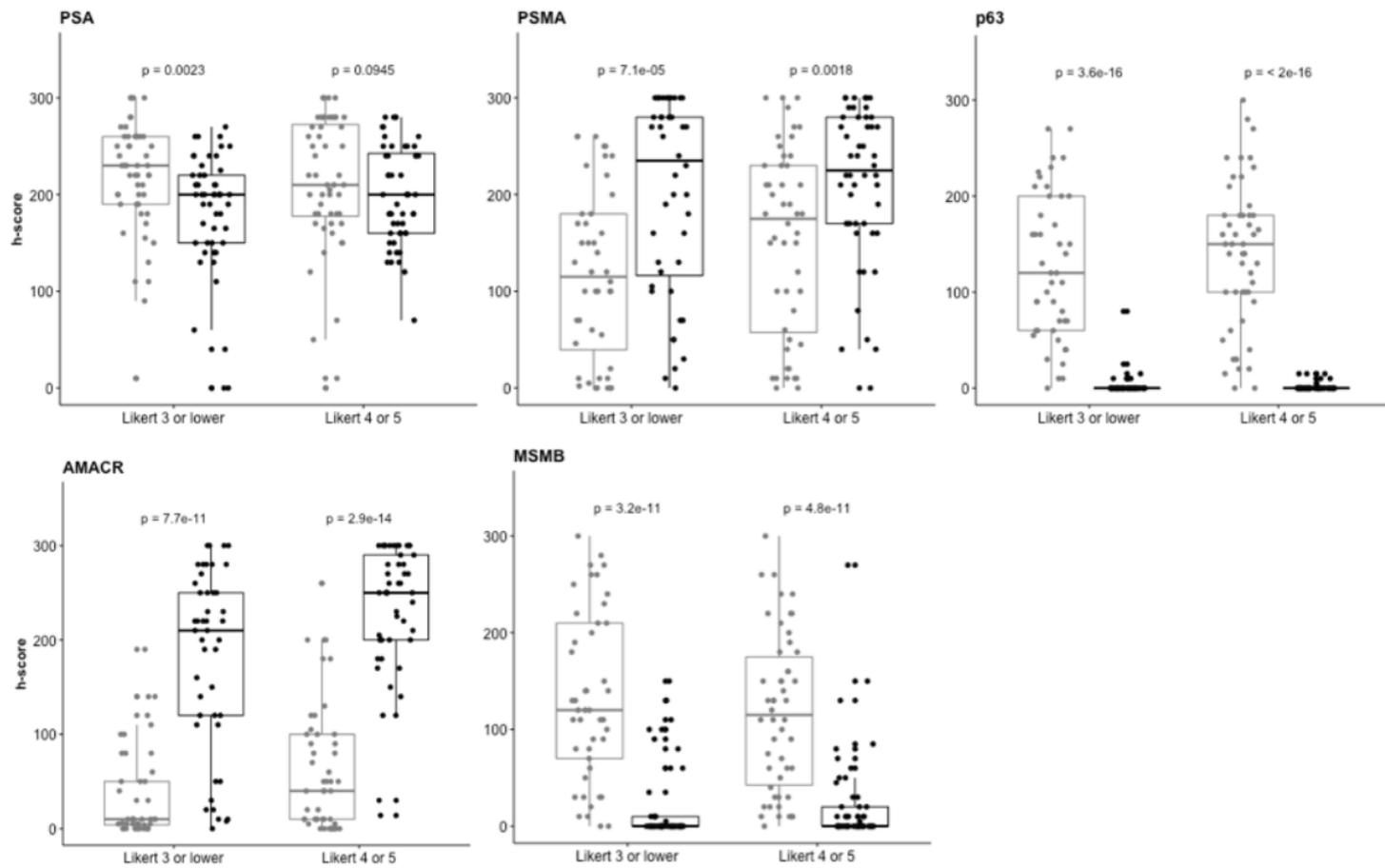
Supplementary Figure 1: IHC for common prostate cancer biomarkers on all samples (paired and unpaired). IHC for all samples was analysed using h-score method that takes into account staining intensity and the number of positively stained cells. The distributions are similar to those observed for paired-only data, but formal statistical tests were not performed due to non-independence of observations within each group.

Supplementary Figure 2. TMA IHC h-scores versus appearance on prostate mpMRI by Likert score: IHC was analysed using h-score method that takes into account staining intensity and the number of positively stained cells. H-scores for paired samples only in tumour tissue (dark grey) and benign tissue (light grey) are shown. mpMRI visible lesions were graded by Likert score where Likert 1/2 are unlikely to be tumour, Likert 3 is equivocal and Likert 4/5 have a high probability of being tumour. All associations were tested within either benign or tumour groups using Kruskal-Wallis analysis of variation. The p-values and number of pairs are separately shown for each stain.









TMA core characteristic	Number of cores
Pathology	
Benign	26
PIN	16
3+3	64
3+4	118
3+5	1
4+3	20
4+4	3
5+4	1
Likert score for sampled quadrant	
2	73
3	59
4	39
5	77

Table 1: Pathological and radiological characteristics associated with the TMA biopsy cores.

	PSA	PSMA	p63	AMACR	MSMB	Total
Missing or unassessable	42	87	95	68	46	338
Stroma only	35	16	10	13	38	112
Concordant benign tumour	349 167 182	324 155 169	325 158 167	342 170 172	330 160 170	1670 810 860
Re-assigned benign → tumour	22 9	21 7	18 4	25 10	34 10	120 40
tumour → benign	13	14	14	15	24	80
Total	448	448	448	448	448	2240

Table 2: TMA IHC outcomes.

Concordance	Partial (2 levels)	Full (3 levels)
Tumour	64	122
Benign	74	107
Total	138/448 (30.80%)	229/448 (51.12%)

Table 3: H&E slides were assessed every 50 slides over 150 slides of each TMA slide and assessed for tumour or benign pathology. Concordance was judged as partial when it agreed at two levels or full when all three levels exhibited the same pathology.

