DOI: *10.4081/aiua.2026.15596*

ORIGINAL PAPER

Digital pathology for repeat prostate biopsy reassessment during active surveillance: An institutional experience

Ludovica Pepe [1, 2], Mara Curduman [3], Pietro Pepe [4], Vincenzo Fiorentino [1, 2]

[1] Anatomic Pathology Unit, Department of Human Pathology in Adult and Developmental Age "Gaetano Barresi", University of Messina, Messina, Italy;
[2] PhD Program in Translational Molecular Medicine and Surgery, Department of Biomedical Sciences, Dental Sciences and Morpho-functional Imaging, University of Messina, Messina, Italy;
[3] Pathology Unit, Cannizzaro Hospital, Catania, Italy;
[4] Urology Unit, Cannizzaro Hospital, Catania, Italy.

Summary ***Introduction: Repeat biopsy remains central to active surveillance (AS) because upgrading and tumor burden may change management.***
Methods: This retrospective paired workflow study included 38 men and 47 repeat biopsy sessions (2023-2025), comprising 846 biopsy cores and 871 H&E slides. Whole-slide imaging (WSI) and conventional microscopy were compared after a washout interval. QuPath was used only for pathologist-guided tumor-length and greatest percentage of cancer (GPC) documentation.
Results: On the first eligible repeat biopsy, upgrading to ISUP Grade Group (GG) 2 occurred in 7/38 men (18.4%; 95% CI, 7.7-34.3). No GG3, cribriform morphology, or intraductal carcinoma was identified. Protocol-linked reclassification occurred in 11/38 men (28.9%). Slide-level cancer-detection agreement was 856/871 (98.3%; kappa = 0.91), and raw patient-level agreement for upgrading was 36/38 (94.7%; kappa = 0.80). After adjudication, no upgraded patient remained missed. QuPath-supported measurement was interpretable in 79/101 positive core-level assessments (78.2%), with excellent agreement (ICC = 0.98) and shorter median measurement time (41 s vs 78 s).
Conclusions: WSI reproduced management-relevant conventional microscopy outputs in repeat AS biopsies. Its value was organizational and documentary, not autonomous cancer detection or grading.

KEY WORDS: *Prostate cancer; Active surveillance; Repeat biopsy; Confirmatory biopsy; Digital pathology; Whole-slide imaging; QuPath; Grade Group; Tumor burden.*

Submitted 25 April 2026; Accepted 3 May 2026

INTRODUCTION

Active surveillance (AS) is a standard management strategy for most men with low-risk prostate cancer and for selected men with favorable intermediate-risk disease. Its purpose is not to ignore cancer, but to reduce overtreatment while maintaining repeated clinical, radiological, and histological reassessment so that curative treatment can still be offered when the risk category changes (1-5). Repeat biopsy remains a clinically decisive component of AS. Although *magnetic resonance imaging* (MRI), *prostate-specific antigen* (PSA) kinetics, PSA density, PSMA PET/CT, and targeted biopsy strategies can refine risk stratification, histological re-evaluation still provides the most direct evidence of upgrading and tumor burden (4-16). Broader contemporary prostate-cancer assessment also increasingly incorporates molecular, genetic, immune, and quantitative pathology markers, which contextualize but do not replace standardized biopsy reassessment in AS (11, 12, 17-24). This is particularly important when a repeat biopsy changes a man from ISUP *Grade Group* (GG) 1 to GG2 or shows higher-volume persistent GG1 disease. For the urologist, the repeat-biopsy report must answer a limited number of practical questions: whether clinically significant upgrading is present, how many cores are positive, what the maximum linear extent of carcinoma is, what the *greatest percentage of cancer* (GPC) is in the most involved core, and whether cribriform morphology or intraductal carcinoma is present. These variables are embedded in contemporary prostate-cancer risk assessment and active-surveillance decision-making, and are more actionable than generic slide-level concordance alone (5, 15, 23, 25-28).

Whole-slide imaging (WSI) is increasingly implemented in routine pathology, but diagnostic use requires local validation and careful quality control (29). Prostate biopsy series have shown high concordance between digital and conventional workflows (30, 31). Nevertheless, AS biopsies represent a demanding setting because clinically relevant disease may be present as a minute focus of pattern 4, discontinuous low-volume carcinoma, or a small suspicious focus requiring ancillary immunohistochemistry (25-28, 32). Digital pathology should also be distinguished from autonomous *artificial intelligence* (AI). AI-based prostate-pathology tools are increasingly being studied, including in AS cohorts, but they require task-specific validation and should not be conflated with restricted pathologist-guided measurement tools (24, 33). The present study did not evaluate an AI screening or grading system. Instead, it evaluated a pathologist-led WSI workflow and the restricted use of QuPath, an open-source digital pathology platform, as an annotation and measurement aid (34).

Instead, it evaluated a pathologist-led WSI workflow and the restricted use of QuPath, an open-source digital pathology software platform that allows visualization of

whole-slide images, annotation of regions of interest, and calibrated measurement of tumor foci, as an annotation and measurement aid (34).

The aim of this study was therefore deliberately clinical. We asked whether a structured WSI workflow reproduced the same patient-level upgrading signal as conventional microscopy in repeat prostate biopsies performed during AS. Secondary aims were to characterize clinically relevant discordances, document the absence or presence of adverse architectural patterns, describe selective immunohistochemistry (IHC), and evaluate whether QuPath improved the reproducibility and practicality of tumor-burden reporting, including GPC.

Materials and methods

This retrospective workflow study evaluated repeat prostate biopsies from men managed with AS for biopsy-proven prostate adenocarcinoma. All biopsy specimens originated from *Azienda Ospedaliera Cannizzaro, Catania (Italy)*. WSI acquisition, digital review, conventional microscopy rereview, selective IHC, QuPath-supported measurement, and integrated adjudication were performed at *Pathology Unit of Policlinico Universitario G. Martino, Messina (Italy)*. Patients were identified from the institutional AS database between January 2023 and December 2025. Eligibility for AS at diagnostic biopsy required biopsy-proven ISUP GG1 carcinoma, clinical stage cT1c, PSA < 10 ng/ml, PSA density < 0.15 ng/ml/cc, absence of cribriform morphology or intraductal carcinoma, and low- volume diagnostic disease according to the local protocol, consistent with contemporary AS principles and local practice (1-7, 15). The final study dataset comprised 38 men, 47 repeat biopsy sessions, 846 biopsy cores, and 871 *hematoxylin and eosin* (H&E) slides. For patient-level analyses, only the first eligible repeat biopsy captured in the study dataset was used for each patient. These 38 first-repeat biopsies included 20 confirmatory biopsies and 18 later surveillance biopsies.

Across all 47 sessions, 20 were confirmatory and 27 were later surveillance biopsies. Thirty-three sessions used systematic sampling only, and 14 used combined systematic plus MRI-targeted sampling. Across all sessions, the number of biopsy cores per session had a mean of 18 and a median of 18 (IQR, 16-20); in sessions with suspicious MRI lesions, four targeted cores were added to the systematic sampling frame. The study flow, cohort structure, and analytic design are summarized in Figure 1.

The primary endpoint was patient-level clinically significant upgrading on the first eligible repeat biopsy, defined as ISUP GG ≥ 2. A supportive session-level endpoint was clinically significant disease across all 47 repeat biopsy sessions. An exploratory protocol-linked reclassification endpoint included clinically significant upgrading or higher-volume persistent GG1 disease according to the local AS pathway. Tumor burden was described using the number of positive cores, maximum linear cancer extent, and the greatest percentage of cancer in a single core (GPC); GPC was treated as a clinically practical descriptor for urologic decision-making, not as an autonomous endpoint detached from grade, core number, and morphology. Cribriform morphology and intraductal carcinoma were specifically assessed because of their relevance for AS suitability and management discussions. In the final study dataset, neither cribriform morphology nor intraductal carcinoma was identified.

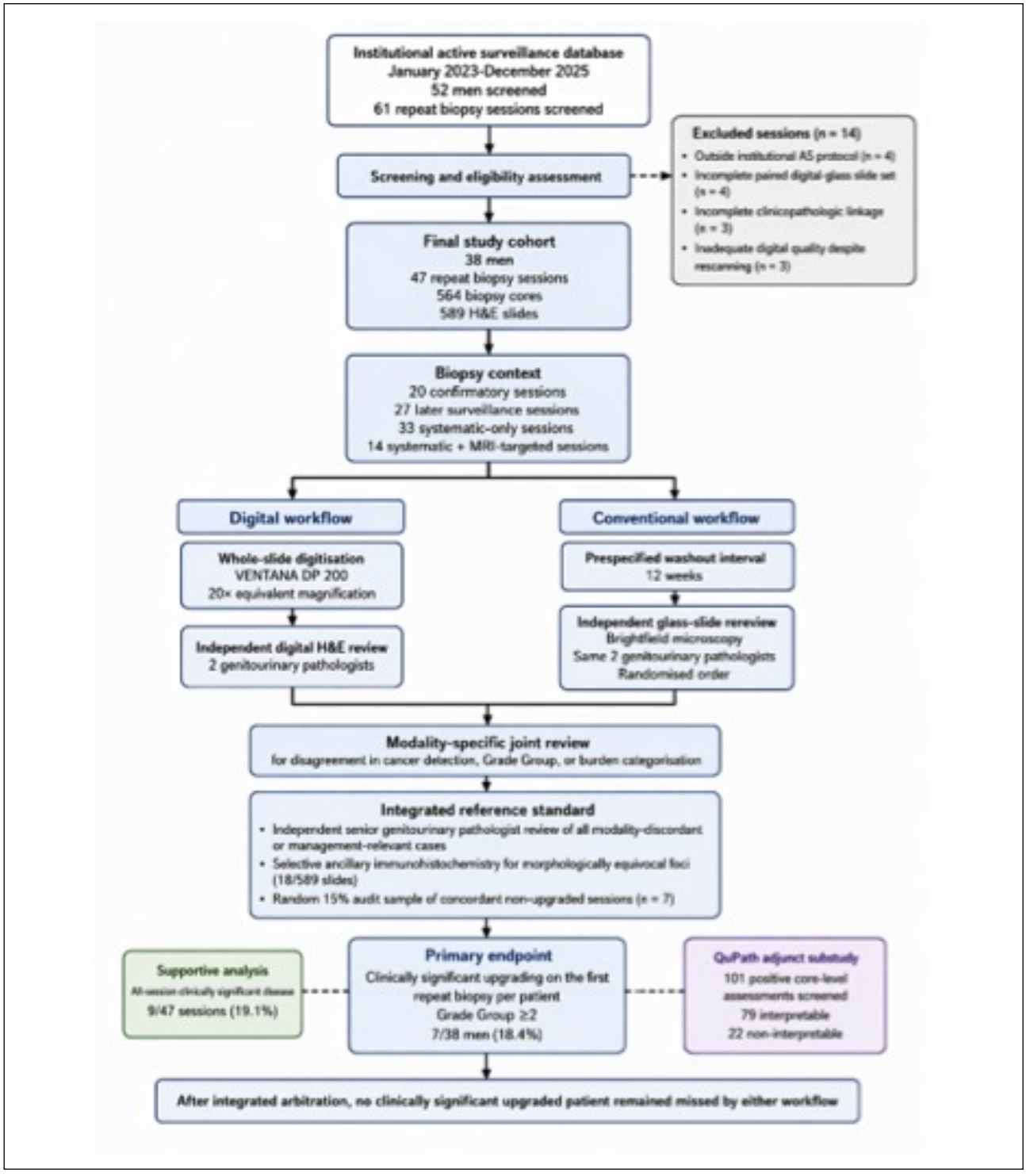
All material had been processed for routine diagnosis. Tissue was fixed in 10% buffered formalin, paraffin embedded, sectioned at 4-5 µm, and stained with H&E. All slides were digitized on a VENTANA DP 200 scanner at 20× equivalent magnification using a single focal plane. Tissue capture, focus, image completeness, and major artifacts were checked immediately after scanning. Slides with clinically relevant quality limitations were rescanned before inclusion. Digital review was performed on standardized workstations under controlled ambient lighting. Two genitourinary pathologists with 9 and 12 years of subspecialty experience independently reviewed all digitized H&E slides. Both had more than 18 months of routine digital sign-out experience and prior local WSI validation experience consistent with CAP-aligned guidance (29). After a 12-week washout, the complete slide set was rereviewed by conventional brightfield microscopy by the same readers in randomized order. Readers were blinded to their previous modality-specific classifications, to each other's assessments, to QuPath measurements, and to downstream management. The clinical context normally available at sign-out, including AS status, age, PSA trend, and biopsy type, remained available. For each modality, initial classifications recorded cancer presence, GG, number of positive cores, maximum cancer length, GPC, and the presence or absence of adverse architectural patterns. The operational workflow is illustrated in Figure 2.

Cases with disagreement in cancer detection, GG, or management-relevant tumor burden underwent joint modality-specific review. An integrated reference standard was then used for the final analyses. All modality- discordant cases, all cases with possible clinically significant upgrading, all cases in which tumor burden crossed or approached the local management trigger, and all cases triaged to IHC were reviewed by a senior genitourinary pathologist who was not one of the two primary readers and who was blinded to which modality generated the initial call. A random audit sample of otherwise concordant non-upgraded sessions was also reviewed to reduce latent verification bias. Selective ancillary IHC was not used as a screening test. It was reserved for morphologically equivocal foci that remained unresolved after independent review and joint reassessment, including atypical small acinar proliferation-like foci, crowded atypical glands in atrophic or inflamed backgrounds, and limited suspicious glands in fragmented tissue. The IHC panel included high-molecular-weight cytokeratin, p63, and alpha- methylacyl-CoA racemase (AMACR/P504S), interpreted strictly with morphology (32).

QuPath version 0.5.1 was used only as an annotation and measurement platform; specifically, it allowed pathologists to annotate tumor foci on digitized H&E slides and to perform calibrated tumor-length and core-length measurements for GPC calculation (34).

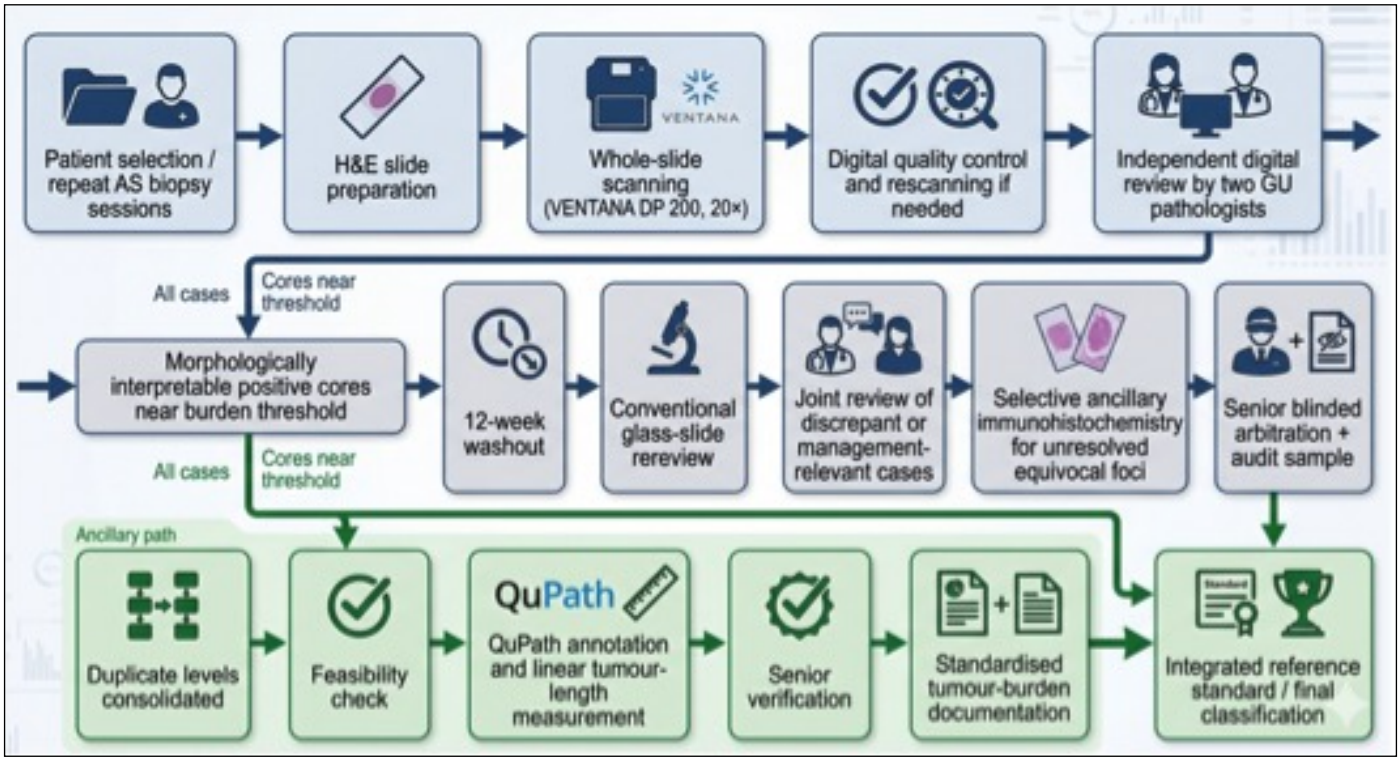
No autonomous AI classifier was used for cancer detection, grading, slide screening, or treatment recommendation. QuPath- supported measurement was attempted

Figure 1.
Study flow and analytic design. Flow diagram showing repeat-biopsy session selection, final study cohort, paired digital and conventional microscopy workflows, integrated reference standard, primary endpoint, supportive session-level analysis, and QuPath adjunct substudy. The primary endpoint was patient-level clinically significant upgrading on the first eligible repeat biopsy, defined as ISUP Grade Group ≥ 2. QuPath was evaluated only as a pathologist-guided measurement tool in selected positive cores and was not used for autonomous cancer detection, grading, or slide screening.



only in morphologically interpretable positive cores, particularly when low-grade tumor burden was close to the institutional management trigger. Maximum linear tumor extent was defined as the greatest continuous linear dimension of invasive carcinoma in an evaluable core profile. In discontinuous tumor foci separated by clearly

Figure 2.
Digital pathology and QuPath-assisted tumor-burden assessment workflow. Schematic representation of the operational digital pathology pathway used for repeat prostate biopsy reassessment during active surveillance. H&E slides were digitized, quality-controlled, reviewed independently by two genitourinary pathologists, and subsequently rereviewed by conventional microscopy after a prespecified washout interval. Discordant or management-relevant cases underwent joint review, selective ancillary immunohistochemistry when needed, senior arbitration, and integrated final classification. QuPath was used only for annotation and linear tumor-length/GPC documentation in selected morphologically interpretable positive cores.



benign tissue, only the largest continuous focus was entered into the primary burden analysis unless integrated review supported continuity.
QuPath-derived GPC was calculated from the measured tumor length and the corresponding evaluable core length using the same dataset rule applied for reference GPC.

Statistical analysis and ethics

Continuous variables were summarized as medians and IQRs unless otherwise specified; categorical variables were summarized as counts and percentages. Exact two-sided 95% confidence intervals (CIs) were calculated for main proportions. Agreement for cancer detection was summarized by overall agreement and Cohen's kappa.
GG agreement among cancer-positive slides was summarized by quadratic weighted kappa because GG is ordinal. The QuPath substudy used biopsy core as the unit of analysis and compared QuPath-supported with reference tumor-length measurements using absolute-agreement *intraclass correlation coefficient* (ICC), Bland-Altman analysis, and measurement-time comparisons. Because clinically relevant discordant patient-level pairs were few, digital versus conventional microscopy differences were interpreted descriptively rather than as formal superiority or non-inferiority testing. This study was a retrospective observational analysis based on archived diagnostic material and de-identified clinicopathologic data. No additional procedure was performed for research purposes. According to local institutional policy for retrospective studies based exclusively on archived diagnostic material and anonymized data, formal ethics committee approval was waived. All patients had provided consent for diagnostic procedures and for use of anonymized clinicopathologic information according to institutional procedures. The study was conducted in accordance with the Declaration of Helsinki.

Results

Cohort, biopsy context, and digital quality

The final cohort included 38 men, 47 repeat biopsy sessions, 846 biopsy cores, and 871 H&E slides. Median patient age was 67 years (IQR, 64.3-72.0), median PSA was 6.1 ng/ml (IQR, 5.3-7.3), median prostate volume was 45.0 ml (IQR, 40.3-59.5), and median PSA density was 0.132 ng/ml/cc (IQR, 0.123-0.133). The median interval from diagnostic biopsy to the first eligible repeat biopsy captured in the dataset was 12.0 months (IQR, 11.0-14.8). Baseline patient characteristics, biopsy-session distribution, sampling context, and digital-slide

composition are summarized in Table 1. Across all sessions, the median number of biopsy cores per session was 18 (IQR, 16-20). Four targeted cores were included in each combined systematic plus MRI-targeted session performed for suspicious lesions. Pre- analytic digital quality control identified 24/871 slides (2.8%) requiring repeat scanning, mainly because of focal out-of-focus regions, incomplete tissue capture near the section edge, or stitching artifacts. After rescanning, all included slides were acceptable for review. Residual minor artifacts were recorded descriptively in 29/871 slides (3.3%) and did not preclude interpretation. Cancer was present on

Table 1.
Cohort characteristics, biopsy context, and clinical outcomes.

	Variable	Value
Cohort and biopsy context	Patients included, n	38
	First eligible repeat biopsies used for primary patient-level analysis, n	38
	Total repeat biopsy sessions, n	47
	Biopsy cores, n	846
	Digitized H&E slides, n	871
	Cancer-positive slides, n	96
	Positive core-level assessments after level consolidation, n	101
Clinical baseline	Median age, years (IQR)	67.0 (64.3-72.0)
	Median PSA, ng/ml (IQR)	6.1 (5.3-7.3)
	Median prostate volume, ml (IQR)	45.0 (40.3-59.5)
	Median PSA density, ng/ml/cc (IQR)	0.132 (0.123-0.133)
	Median interval from diagnostic biopsy to first eligible repeat biopsy, months (IQR)	12.0 (11.0-14.8)
Sampling context	Confirmatory first-repeat biopsies, n/N (%)	20/38 (52.6%)
	Later surveillance first-repeat biopsies, n/N (%)	18/38 (47.4%)
	Confirmatory sessions across all repeat biopsies, n/N (%)	20/47 (42.6%)
	Later surveillance sessions across all repeat biopsies, n/N (%)	27/47 (57.4%)
	Systematic-only sessions, n/N (%)	33/47 (70.2%)
	Combined systematic + MRI-targeted sessions, n/N (%)	14/47 (29.8%)
	Biopsy cores per session, median (IQR)	18 (16-20)
	Targeted cores per combined MRI-targeted session, n	4
	H&E slides per session, median (IQR)	18 (17-20)
Primary and protocol-linked outcomes	Clinically significant upgrading on first eligible repeat biopsy, n/N (%)	7/38 (18.4%)
	95% CI for clinically significant upgrading	7.7-34.3
	Upgraded cases classified as GG2, n	7
	GG3 cases, n	0
	Cribriform morphology, n	0
	Intraductal carcinoma, n	0
	Median maximum cancer core length in upgraded first-repeat biopsies, mm (IQR)	9.0 (8.0-10.6)
	Median positive cores in upgraded first-repeat biopsies, n (IQR)	5 (5-6)
	Median GPC in upgraded first-repeat biopsies, % (IQR)	60.0 (53.4-70.7)
	Median maximum cancer core length in positive non-upgraded first-repeat biopsies, mm (IQR)	6.0 (5.3-6.2)
	Median positive cores in positive non-upgraded first-repeat biopsies, n (IQR)	3 (3-3.8)
	Median GPC in positive non-upgraded first-repeat biopsies, % (IQR)	40.0 (35.5-41.2)
	Exploratory protocol-linked reclassification on first eligible repeat biopsy, n/N (%)	11/38 (28.9%)
	95% CI for exploratory protocol-linked reclassification	15.4-45.9
	Persistent higher-volume GG1 cases within the above category, n	4
Session-level and management context	Clinically significant disease across all sessions, n/N (%)	9/47 (19.1%)
	95% CI for clinically significant disease across all sessions	9.1-33.3
	Documented post-biopsy management escalation across all sessions, n/N (%)	13/47 (27.7%)
	95% CI for documented post-biopsy management escalation	15.6-42.6
Digital quality control	Slides requiring repeat scanning, n/N (%)	24/871 (2.8%)
	Residual minor artifacts not precluding interpretation, n/N (%)	29/871 (3.3%)

AS = active surveillance; CI = confidence interval; GG = ISUP Grade Group; GPC = greatest percentage of cancer in the most involved core; H&E = hematoxylin and eosin; IQR = interquartile range; MRI = magnetic resonance imaging; PSA = prostate-specific antigen. Protocol- linked reclassification used institution-specific criteria and should not be interpreted as a universal AS rule.

96/871 slides, corresponding to 101 positive core-level assessments after consolidation of duplicate levels from the same core.

Primary endpoint and urologist-facing pathology output

On the first eligible repeat biopsy per patient, clinically significant upgrading occurred in 7/38 men (18.4%; 95% CI, 7.7-34.3). All upgraded patients were GG2. No GG3 carcinoma was identified. No cribriform morphology or intraductal carcinoma was identified in upgraded or non-upgraded first-repeat biopsies. For urologic readability, 31/38 men did not show clinically significant upgrading, whereas 7/38 crossed the clinically significant threshold of GG ≥2. Among upgraded patients, the median maximum cancer core length was 9.0 mm (IQR, 8.0-10.6), the median number of positive cores was 5 (IQR, 5-6), and the median GPC was 60.0% (IQR, 53.4-70.7). Among positive but non-upgraded first-repeat biopsies, the corresponding values were 6.0 mm (IQR, 5.3-6.2), 3 positive cores (IQR, 3-3.8), and GPC 40.0% (IQR, 35.5-41.2). These data are summarized in Table 1.

Protocol-linked reclassification and post-biopsy management

Exploratory protocol-linked reclassification, defined as GG ≥ 2 or higher-volume GG1 according to the local AS pathway, occurred in 11/38 men (28.9%; 95% CI, 15.4-45.9). Four of these 11 men had higher-volume GG1 without GG ≥ 2 carcinoma. These cases explain why the protocol-linked reclassification count was higher than the clinically significant upgrading count: they were not upgraded cancers, but they had tumor-burden features, including clinically relevant GPC, that justified closer surveillance in the local protocol. Across all 47 repeat biopsy sessions, clinically significant disease (GG ≥ 2) was identified in 9/47 sessions (19.1%; 95% CI, 9.1-33.3). Documented post-biopsy management escalation followed 13/47 sessions (27.7%; 95% CI, 15.6-42.6). This variable was used only to provide clinical context; it was not interpreted as proof that one pathology workflow caused a treatment decision. Among the seven men upgraded on the first eligible repeat biopsy, available documentation showed escalation after all seven: radical prostatectomy in four, radiotherapy- based treatment in one, and intensified multidisciplinary surveillance in two.

Digital versus conventional microscopy and discordant cases

Raw slide-level agreement between digital review and conventional microscopy for cancer detection was 856/871 slides (98.3%), with Cohen's kappa = 0.91. At patient level, raw agreement for the primary endpoint before integrated arbitration was 36/38 patients (94.7%), with kappa = 0.80. Intermodality agreement, reproducibility, adjudication, selective IHC, and QuPath findings are summarized in Table 2. The two clinically relevant discordant patients before arbitration are important to interpret correctly. In one patient, the digital workflow identified a limited GG2 focus that the conventional pre-arbitration call did not identify. In the other patient, the conventional workflow identified a limited GG2 focus that the digital pre- arbitration call did not identify. The maximum cancer core lengths in these two cases were 7.3 mm and 7.2 mm, with five and four positive cores, respectively, and GPC values of 48.7% and 48.0%, respectively. Neither showed cribriform morphology or intraductal carcinoma. After integrated review, both were classified as GG2 and both proceeded to radical prostatectomy. Therefore, the discordance was not an AI error, because no AI diagnostic classifier was used. It represented one pre-arbitration false-negative interpretive call in each pathology workflow. The clinical implication is that borderline upgraded foci should enter a transparent adjudication pathway; after integrated adjudication, no clinically significant upgraded patient remained missed by either workflow. Within the digital workflow, reader disagreement on cancer detection before adjudication occurred in 27/871 slides (3.1%; kappa = 0.85). Within the conventional workflow, reader disagreement occurred in 30/871 slides (3.4%; kappa = 0.83). On the 96 reference-positive slides, weighted kappa for GG assignment against the integrated reference standard was 0.74 for the digital workflow and 0.80 for the conventional workflow. Discrepancies clustered in minute atypical foci, discontinuous carcinoma at the edge of a core, and limited suspected pattern 4 in scant tissue.

Selective IHC and QuPath-supported measurement

Selective IHC entered the adjudication pathway in 18/871 slides (2.1%). IHC resolved persistent uncertainty in 15/18 slides (83.3%) and changed the provisional diagnostic category in 6/18 slides (33.3%). Its role was to support morphology in equivocal foci, not to screen all biopsies. After consolidation of duplicate levels, 101 positive core-level assessments were screened for QuPath feasibility. QuPath-supported measurement was technically interpretable in 79/101 cores (78.2%) and non-interpretable in 22/101 cores (21.8%). Reasons for non-interpretability were discontinuous multifocal involvement, marked fragmentation, severe folding or section distortion, and extensive crush or edge artifact.

In interpretable cores, QuPath-supported measurements showed high agreement with reference manual tumor-length measurements, with an absolute-agreement intraclass correlation coefficient (ICC) of 0.98.

Bland-Altman analysis showed a mean bias of +0.12 mm, with 95% limits of agreement from -0.72 to +0.95 mm. For GPC, the mean bias was +0.8 percentage points, with 95% limits of agreement from -4.8 to +6.3 percentage points. In the independent second-observer subset of 40 cores, interobserver ICC was 0.98 and the coefficient of repeatability was 1.08 mm. Median measurement time in this subset was 41 s with QuPath and 78 s with manual on-screen measurement. In the prespecified subset of 22 borderline low-grade positive cores near the institutional trigger, QuPath changed the provisional burden category in 7/22 cores (31.8%; 95% CI, 13.9-54.9). A patient-level management-oriented change was flagged in three patients. These findings support a selective documentation role for QuPath in borderline GG1 disease, not a role in cancer detection or Gleason grading (Table 2).

Table 2.
Digital/conventional agreement, adjudication, selective IHC, and QuPath adjunct measurement.

	Variable	Value
Intermodality agreement	Slides compared across modalities, n	871
	Raw agreement for cancer detection, n/N (%)	856/871 (98.3%)
	Cohen's kappa for cancer detection	0.91
	Cancer-positive slides for GG comparison, n	96
	Weighted kappa for GG assignment versus integrated reference	Digital 0.74; conventional 0.80
	Raw patient-level agreement for primary endpoint before integrated arbitration, n/N (%)	36/38 (94.7%)
	Patient-level kappa before integrated arbitration	0.80
Discordance and adjudication	Clinically relevant discordant patient pairs before arbitration, n	2
	Digital pre-arbitration false-negative upgraded patient, n	1
	Conventional pre-arbitration false-negative upgraded patient, n	1
	Clinically significant upgraded patients captured after integrated arbitration, digital workflow	7/7
	Clinically significant upgraded patients captured after integrated arbitration, conventional workflow	7/7
	Audit sample of concordant non-upgraded sessions reviewed independently, n	7
	Additional clinically significant cases found in audit sample, n	0
Reader reproducibility and IHC	Digital reader disagreement for cancer detection, n/N (%)	27/871 (3.1%)
	Digital interobserver kappa for cancer detection	0.85
	Conventional microscopy reader disagreement for cancer detection, n/N (%)	30/871 (3.4%)
	Conventional microscopy interobserver kappa for cancer detection	0.83
	Slides requiring joint review in digital workflow, n/N (%)	43/871 (4.9%)
	Slides requiring joint review in conventional microscopy workflow, n/N (%)	40/871 (4.6%)
	Slides requiring third-reader arbitration in digital workflow, n/N (%)	9/871 (1.0%)
	Slides requiring third-reader arbitration in conventional microscopy workflow, n/N (%)	8/871 (0.9%)
	Slides undergoing selective ancillary IHC, n/N (%)	18/871 (2.1%)
	IHC resolved uncertainty, n/N (%)	15/18 (83.3%)
	IHC changed provisional diagnostic category, n/N (%)	6/18 (33.3%)
QuPath feasibility	Positive core-level assessments screened, n	101
	Technically interpretable for QuPath-supported measurement, n/N (%)	79/101 (78.2%)
	Non-interpretable, n/N (%)	22/101 (21.8%)
	Reasons for non-interpretability, n	Discontinuous multifocal involvement, 9; marked fragmentation, 6; severe folding or section distortion, 4; extensive crush or edge artifact, 3
QuPath measurement	Absolute-agreement ICC versus reference tumor- length measurement	0.98
	Mean bias, QuPath minus reference, mm	+0.12
	95% limits of agreement, mm	-0.72 to +0.95
	Mean GPC bias, QuPath minus reference, percentage points	+0.8
	95% limits of agreement for GPC, percentage points	-4.8 to +6.3
	Independent second-observer subset, n	40
	Interobserver ICC	0.98
	Coefficient of repeatability, mm	1.08
	Median measurement time, QuPath vs manual on- screen	41 s vs 78 s
Borderline-core impact	Borderline low-grade positive cores near institutional trigger, n	22
	Cores in which QuPath changed provisional burden category, n/N (%)	7/22 (31.8%)
	Patients with management-oriented change flagged by QuPath, n	3

GG = ISUP Grade Group; GPC = greatest percentage of cancer; IHC = immunohistochemistry; ICC = intraclass correlation coefficient; WSI = whole-slide imaging.
The integrated reference standard was an adjudicated pathology reference, not an external histological gold standard. QuPath was not used for cancer detection, grading, slide screening, or treatment.

Discussion

This study addressed a practical question for AS: when repeat prostate biopsies are reviewed digitally and by conventional microscopy, does the digital workflow reproduce the pathology information that may change management? In this cohort, it did. Digital review closely matched conventional microscopy for patient-level clinically significant upgrading, while clinically relevant discordance before arbitration was uncommon and balanced between workflows. The analysis was deliberately cen-

tered on the information that matters most to urologists managing AS: whether repeat biopsy shows clinically significant upgrading, how many cores are positive, the extent of tumor involvement, GPC in the most involved core, the presence or absence of cribriform morphology or intraductal carcinoma, and whether the overall findings support continued AS, intensified monitoring, or active treatment. In this setting, generic slide-level concordance alone is not sufficient, because a single small upgraded focus may be more clinically relevant than a large number of concordant benign slides. A clinically important point is that all upgraded patients in the final dataset were GG2. No GG3 carcinoma was identified, and no cribriform morphology or intraductal carcinoma was recorded. This makes the cohort coherent with a real-world AS reassessment setting in which the main management threshold is not high-grade disease, but the recognition of limited GG2 carcinoma or higher-volume GG1 disease that may justify escalation or closer surveillance (5, 15, 25-28). The two discordant patients clarify the meaning of the raw discrepancy. One limited GG2 focus was missed pre- arbitration by the digital workflow and one limited GG2 focus was missed pre-arbitration by conventional microscopy. These were not AI failures and not clinical-management errors; they were difficult borderline pathology calls, resolved by integrated review. This balanced pattern argues against a systematic failure of either modality, but it also argues against claiming that WSI is superior to conventional microscopy for detecting upgrading. The main value of the digital workflow was therefore operational and documentary rather than autonomous diagnosis. WSI supported easier slide retrieval, review of cancer-positive slides, side-by-side discussion in multidisciplinary settings, and more standardized annotation of small cancer foci. This is especially relevant because prostate-biopsy tumor extent and biopsy/tumor-volume relationships are clinically informative but may vary among pathologists, potentially affecting AS eligibility, reclassification, or treatment discussions (10, 23, 28). GPC deserves specific emphasis. In AS, the urologist is rarely interested only in whether a single microscopic focus is present; the proportion of an involved core occupied by carcinoma is often part of the management conversation. By deriving GPC consistently from the same measurement framework used for maximum cancer length, the present workflow provides a clinically familiar variable in a reproducible and auditable format. This is a more realistic use of digital pathology than attempting to replace the pathologist's morphologic interpretation. The QuPath substudy supports this restricted role. In interpretable positive cores, QuPath showed excellent agreement with reference tumor-length measurement, acceptable agreement for GPC, and shorter measurement time. It also changed the provisional burden category in a subset of borderline low-grade cores. This does not mean that QuPath detected cancer or assigned Gleason grade. Rather, it suggests that once the pathologist has identified and graded a focus, digital annotation may make the reporting of length and GPC more reproducible. The data do not show that digital review improves Gleason score interpretation. Weighted kappa for GG assignment was high for both workflows and numerically higher for conventional microscopy. This should be stated clearly: final diagnosis and grading remain the responsibility of the pathologist. Digital tools may make measurement and documentation of a morphologically recognized focus more reproducible, but they should not be presented as tools that autonomously detect prostate cancer, select treatment, or resolve limited pattern 4 interpretation. This distinction is important because dedicated AI studies in AS biopsies have shown promising cancer-detection and workload-triage performance, but within task-specific algorithmic frameworks that differ from the restricted QuPath measurement approach evaluated here (33). Selective IHC remained important. In AS biopsies, avoiding overdiagnosis is as important as avoiding underdiagnosis, because an equivocal minute focus may trigger anxiety, intensified follow-up, or treatment discussion. The use of basal-cell markers and AMACR/P504S only after morphological triage was therefore consistent with standard prostate pathology practice and prevented overcalling in several equivocal foci (32). This study has some limitations. It is retrospective, modest in size, and conducted within a single institutional network. The same two primary readers participated in both workflows, although washout and blinding

Declarations

Ethical approval and consent for participate: *This study was conducted as a retrospective observational analysis based on archived diagnostic material and de- identified clinicopathological data. According to local institutional policy for retrospective studies using archived diagnostic material and anonymized data, formal ethics committee approval was waived. All patients had provided consent for diagnostic procedures and for the use of anonymized clinicopathological information according to institutional procedures. The study was conducted in accordance with the Declaration of Helsinki.*

Consent for publication: *Not applicable. The manuscript does not contain identifiable individual patient data.*

Availability of data and material: *The data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to institutional regulations and applicable privacy restrictions.*

Competing interests: *The authors declare that they have no conflicts of interest related to this work.*

Funding: *This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.*

Authors' contributions: *Ludovica Pepe: Conceptualization; Methodology; Investigation; Data curation; Writing - original draft; Writing - review and editing. Vincenzo Fiorentino: Conceptualization; Methodology; Investigation; Validation; Data curation; Writing - original draft; Writing - review and editing. Mara Curduman: Investigation; Validation; Data curation; Writing - review and editing. Pietro Pepe: Conceptualization; Resources; Supervision; Project administration; Writing - review and editing. All authors read and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.*

Acknowledgments: *Not applicable.*

were used to reduce recall bias. The integrated reference standard strengthened clinically relevant classification but did not constitute an external histologic gold standard. The higher-volume GG1 trigger was institution-specific and should not be interpreted as a universal AS rule. Management escalation was interpreted cautiously because discontinuation or intensification of AS is influenced by clinical, radiological, pathological, and patient-related factors rather than by pathology workflow alone (35). Within these limits, the study supports a realistic implementation message. A validated WSI workflow can reproduce conventional microscopy for repeat-biopsy reassessment in AS, provided that pathologist-led review, selective IHC, transparent adjudication, and restricted use of digital measurement tools remain central to the workflow.

Conclusions

In this retrospective head-to-head workflow study, structured WSI review reproduced the patient-level clinically significant upgrading signal obtained by conventional microscopy in repeat prostate biopsies from men on AS. The result supports clinical alignment between workflows, not technological superiority. For the pathologist and the urologist, the practical advantages were clearer workflow organization, easier reassessment of cancer-positive slides, faster measurement in selected positive cores, and more reproducible documentation of tumor extent and GPC. QuPath was useful only as a restricted measurement aid in morphologically interpretable cores. Gleason score interpretation, including the recognition of limited pattern 4, remained pathologist-led.

References

1. Eastham JA, Auffenberg GB, Barocas DA, et al. Clinically Localized Prostate Cancer: AUA/ASTRO Guideline, Part I: Introduction, Risk Assessment, Staging, and Risk-Based Management. J Urol. 2022; 208:10-18.

2. Cornford P, van den Bergh RCN, Briers E, et EAU-EANM-ESTRO-ESUR-ISUP-SIOG Guidelines on Prostate Cancer-2024 Update. Part I: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur Urol. 2024; 86:148-163.

3. Willemse PM, Davis NF, Grivas N, et al. Systematic Review of Active Surveillance for Clinically Localised Prostate Cancer to Develop Recommendations Regarding Inclusion of Intermediate-risk Disease, Biopsy Characteristics at Inclusion and Monitoring, and Surveillance Repeat Biopsy Strategy. Eur Urol. 2022; 81:337-346.

4. Pepe P, Cimino S, Garufi A, et al. Confirmatory biopsy of men under active surveillance: extended versus saturation versus multiparametric magnetic resonance imaging/transrectal ultrasound fusion prostate biopsy. Scand J Urol. 2017; 51:260-263.

5. Pepe P, Pepe L, Pennisi M, Fraggetta F. Oncological Outcomes in Men With Favorable Intermediate Risk Prostate Cancer Enrolled in Active Surveillance. In Vivo. 2024; 38:1300-1305.

6. Luiting HB, Remmers S, Boevé ER, et al. A Multivariable Approach Using Magnetic Resonance Imaging to Avoid a Protocol-based Prostate Biopsy in Men on Active Surveillance for Prostate Cancer-Data from the International Multicenter Prospective PRIAS Study. Eur Urol Oncol. 2022; 5:651-658.

7. Osses DF, Drost FH, Verbeek JFM, et al. Prostate cancer upgrading with serial prostate magnetic resonance imaging and repeat biopsy in men on active surveillance: are confirmatory biopsies still necessary? BJU Int. 2020; 126:124-132.

8. Pepe P, Pepe L, Pepe G, Fiorentino V. PSMA PET/CT in diagnosing and staging high grade prostate cancer: accuracy and cost-effectiveness. Arch Ital Urol Androl. 2026; 98:14797.

9. Pepe P, Pepe L, Tamburo M, et al. 68Ga-PSMA PET/CT evaluation in men enrolled in prostate cancer Active Surveillance. Arch Ital Urol Androl. 2023; 95(:11322.

10. Pepe P, Fraggetta F, Galia A, et al. Is quantitative histologic examination useful to predict nonorgan-confined prostate cancer when saturation biopsy is performed? Urology. 2008; 72:1198-202.

11. Pepe P, Pepe L, Fiorentino V, et al. Transcutaneous perineal-ultrasound guided prostate biopsy in men with rectal amputation. Arch Ital Urol Androl. 2025; 97:14047.

12. Fiorentino V, Pepe L, Zuccalà V, et al. Gleason score down and upgrading at radical prostatectomy in targeted vs. systematic prostate biopsy: Findings from an institutional cohort. Pathol Res Pract. 2025; 271:156040.

13. Pepe P, Pepe L, Fiorentino V, et al. PSMA PET/CT Accuracy in Diagnosing Prostate Cancer Nodes Metastases. In Vivo. 2024; 38:2880-2885.

14. Pepe P, Pepe L, Fiorentino V, et al. Multiparametric MRI targeted prostate biopsy: When omit systematic biopsy? Arch Ital Urol Androl. 2024; 96:12992.

15. Pepe P, Pepe L, Pennisi M, Fraggetta FL. Which Prostate Biopsy in Men Enrolled in Active Surveillance? Experience in 110 Men Submitted to Scheduled Three-Years Transperineal Saturation Biopsy Combined With Fusion Targeted Cores. Clin Genitourin Cancer. 2021; 19:305-308.

16. Pepe P, Aragona F. Incidence of insignificant prostate cancer using free/total PSA: results of a case-finding protocol on 14,453 patients. Prostate Cancer Prostatic Dis. 2010; 13:316-9.

17. Pepe L, Pizzimenti C, Tralongo P, et al. PD-L1 Expression in Prostate Cancer: Anatomopathological Features, Methodological Pitfalls, and Therapeutic Potential. Int J Mol Sci. 2026; 27:1797.

18. Vatrano S, Pepe P, Pepe L, et al. BRCA mutations and prostate cancer: should urologist improve daily clinical practice? Arch Ital Urol Androl. 2025; 97:13635.

19. Pepe P, Fraggetta F, Galia A, et al. PCA3 score and prostate cancer diagnosis at repeated saturation biopsy. Which cut-off: 20 or 35? Int Braz J Urol. 2012; 38:489-95.

20. Fiorentino V, Germanà E, Ricciardi G, et al: SPOP and MMR/MSI alterations in prostate cancer: relationship with PD-L1, TILs and AR expression. Pathologica. 2025; 117:338-347.

21. Fiorentino V, Pepe L, Pizzimenti C, et al. PD-L1 expression in prostate cancer and Gleason Grade Group: Is there any relationship? Findings from a multi-institutional cohort. Pathol Res Pract. 2025; 269:155916.

22. Pecci V, Troisi F, Aiello A, et al: Targeting of H19/cell adhesion molecules circuitry by GSK-J4 epidrug inhibits metastatic progression in prostate cancer. Cancer Cell Int. 2024; 24:56.

23. Fiorentino V, Martini M, Dell'Aquila M, et al. Histopathological Ratios to Predict Gleason Score Agreement between Biopsy and Radical Prostatectomy. Diagnostics (Basel). 2020; 11:10.

24. Fiorentino V, Pizzimenti C, Franchina M, et al. The minefield of indeterminate thyroid nodules: Could artificial intelligence be a suitable diagnostic tool? Diagn. Histopathol. 2023; 29:396-401.

25. van Leenders GJLH, van der Kwast TH, Grignon DJ, et ISUP Grading Workshop Panel Members. The 2019 International Society of Urological Pathology (ISUP) Consensus Conference on Grading of Prostatic Carcinoma. Am J Surg Pathol. 2020; 44:e87-e99.

26. Hassan O, Matoso A: Clinical significance of subtypes of Gleason pattern 4 prostate cancer. Transl Androl Urol. 2018; 7(Suppl 4):S477-S483.

27. Perera M, Jibara G, Tin AL, et al. Outcomes of Grade Group 2 and 3 Prostate Cancer on Initial Versus Confirmatory Biopsy: Implications for Active Surveillance. Eur Urol Focus. 2023; 9:662-668.

28. Bernhardt M, Weinhold L, Bremmer F, et al. Unexpectedly high variability in determining tumour extent in prostatic biopsies: implications for active surveillance. Histopathology. 2025; 86:627-639.

29. Evans AJ, Brown RW, Bui MM, et al. Validating Whole Slide Imaging Systems for Diagnostic Purposes in Pathology. Arch Pathol Lab Med. 2022 ; 146:440-450.

30. Rao V, Subramanian P, Sali AP, et al. Validation of Whole Slide Imaging for primary surgical pathology diagnosis of prostate biopsies. Indian J Pathol Microbiol. 2021; 64:78-83.

31. Torresani E, Gentilini MA, Grassi S, et al. Diagnostic concordance between traditional and digital workflows. A study on 1427 prostate biopsies. Pathologica. 2023; 115:221-226.

32. Molinié V, Fromont G, Sibony M, et al. Diagnostic utility of a p63/alpha-methyl-CoA-racemase (p504s) cocktail in atypical foci in the prostate. Mod Pathol. 2004; 17:1180-90.

33. Arvidsson I, Svanemur E, Marginean F, et al. Artificial intelligence for detection of prostate cancer in biopsies during active surveillance. BJU Int. 2024; 134:1001-1009.

34. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep. 2017; 7:16878.

35. Van Hemelrijck M, Ji X, Helleman J, Roobol MJ, et Members of the Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance GAP3 consortium; Members of the Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance GAP3 consortium. Reasons for Discontinuing Active Surveillance: Assessment of 21 Centres in 12 Countries in the Movember GAP3 Consortium. Eur Urol. 2019; 75:523-531.

Correspondence

Ludovica Pepe

Vincenzo Fiorentino

Anatomic Pathology Unit, Department of Human Pathology in Adult and Developmental Age "Gaetano Barresi", University of Messina, 98125 Messina, Italy

Mara Curduman

Pathology Unit, Cannizzaro Hospital, 95126 Catania, Italy

Pietro Pepe (Corresponding Author)

piepepe@hotmail.com

Urology Unit, Cannizzaro Hospital, 95126 Catania, Italy