Artificial intelligence to improve cytology performances in bladder carcinoma detection: results of the VisioCyt test

Thierry Lebret1,2, Geraldine Pignot3, Marc Colombel4,5, Laurent Guy6, Xavier Rebillard7, Laurent Savareux8, Mathieu Roumigue9, Sebastien Nivet10, Monique Coutade Saidi11, Eric Piaton12 and Camelia Radulescu13

1Urology Department, Foch Hospital, Suresnes, 2UVSQ, Paris Saclay University, Versailles, 3Department of Surgical Oncology 2, Institut Paoli-Calmettes, Marseille, 4Urology Department, Hospices Civils de Lyon, 5Claude Bernard University, Lyon, 6Urology Department, CHU Clermont-Ferrand and Clermont Auvergne University, Clermont-Ferrand, 7Urology Department, Beausoleil Private Hospital, Montpellier, 8Urology Auvergne Centre, Private Hospital La Chataigneraie, Beaumont, 9Department of Urology, CHU Rangueil IUCT Oncopole Toulouse, 10VitaDX Company, Rennes, 11Anatomo and Cytopathology Department, IUCT Oncopole, Toulouse, 12Est Pathology Centre, Woman-Mother-Child Hospital, Bron, and 13Anatomo and Cytopathology Department, Foch Hospital, Suresnes, France

Objective
To explore the utility of artificial intelligence (AI) using the VisioCyt® test (VitaDX International, Rennes, France) to improve diagnosis of bladder carcinoma using voided urine cytology.

Patients and Methods
A national prospective multicentre trial (14 centres) was conducted on 1360 patients, divided in two groups. The first group included bladder carcinoma diagnosis with different histological grades and stages, and the second group included control patients based on negative cystoscopy and cytology results. The first step of this VISIOCYT1 trial focussed on algorithm development and the second step on validating this algorithm. A total of 598 patients were included in this first step, 449 patients with bladder tumours (219 high-grade and 230 low-grade) and 149 as negative controls. The VisioCyt test was compared to voided urine cytology performed by experienced uro-pathologists from each centre.

Results
Overall sensitivity was highly improved by the VisioCyt test compared to cytology (84.9% vs 43%). For high-grade tumours the VisioCyt test sensitivity was 92.6% vs 61.1% for the uro-pathologists. Regarding low-grade tumours, VisioCyt test sensitivity was 77% vs 26.3% for the uro-pathologists.

Conclusion
In comparison to routine cytology, the results of the first phase of the VISIOCYT1 trial show very clear progress in terms of sensitivity, which is particularly visible and interesting for low-grade tumours. If the validation cohort confirms these results, it could lead to the VisioCyt test being considered as a very useful aid for pathologists. Moreover, as this test is in fact software based on AI, it should become more and more efficient as more data are collected.

Keywords
bladder tumour, urine cytology, artificial intelligence, diagnosis, voided urine, #blcsm, #BladderCancer, #uroonc

Introduction
Urothelial carcinoma is the fourth most common cancer. Most tumours are diagnosed as non-muscle-invasive bladder carcinoma (NMIBC; 70–80%), including a large majority of low-grade tumours (LGTs). Nevertheless, up to 70% of patients will relapse during follow-up [1–3]. Moreover, one-third of high-grade NMIBC will progress to invasive diseases leading to a cystectomy. Therefore, early diagnosis and early detection of recurrence is a real challenge. In practice, urologists mainly use two key examinations for bladder cancer diagnosis: cystoscopy and cytology. The cystoscopy
examination must be repeated throughout the patient’s follow-up, according to their recurrence risk. However, repeating this invasive examination may lead to patients abandoning the process.

Conversely, voided cytology is very well accepted by patients, as urine collection is simple and non-invasive. Cytology is a routine examination based on the analysis of cell morphology, but it requires highly experienced pathologists, and suffers from low sensitivity when it comes to identifying urothelial carcinoma, especially in LGTs. Despite this, cytology is in some cases the only alternative when the urothelial tumour occurs in the upper urinary tract (5% of urothelial tumours) [4], and given that cystoscopy cannot be performed without general anaesthesia.

To improve and be more reproducible, cell morphology analysis based on cytology may benefit from artificial intelligence (AI) algorithm capabilities combined with image processing. The automation of this examination using AI allows for exhaustive morphological analysis of all the urothelial cells.

The VitaDX Company (VitaDX International, Rennes, France) developed a digital medical device (VisioCyt®) able to detect tumour cell aspect (LGT and high-grade tumour [HGT] cells) in voided urine samples.

A clinical trial (VISIOCYT1) was conducted to investigate this new method using AI. This research programme was built in two phases; the data collected during the first phase being used for developing the algorithm, and the data from the second phase being the validation cohort used to assess the performance of the VisioCyt test. The present study reports the results of the first phase of the VISIOCYT1 clinical trial.

**Patients and Methods**

**Patient Clinical Data Acquisition**

Biological, clinical and digital patient data were collected as part of the VISIOCYT1 clinical trial, which is a national, multicentric, prospective clinical study in collaboration with 14 French hospitals (10 University hospitals and four private hospitals), including 1360 patients and taking place in two phases: patients from the first phase were used for developing the algorithm, and the data from the second phase being the validation cohort used to assess test the performance of the VisioCyt test. The present study reports the results of the first phase of the VISIOCYT1 clinical trial.

- Group 1: patients with a bladder tumour, confirmed by positive histology showing urothelial carcinoma, whatever the grade. The histology was performed on tissues from the first transuretral resection (TUR).
- Group 2: control patients confirmed by a negative cytology and a negative endoscopy (patients without a history of urothelial tumours).

The enrolment of patients was competitive between each centre and continuous without interruption between the two phases. For this first phase presented here, 600 patients were enrolled; Group 1 comprised 450 patients with a bladder tumour and Group 2 comprised 150 patients without bladder tumours. The split between patients with and without bladder tumours was arbitrated according to the needs of the algorithm development. Thus, the enrolment of patients with HGTs or LGTs was natural.

The enrolment took place over 3 years, 2 years for the first phase in which the algorithm was developed and 1 year more for the enrolment of the 400 patients, making up the validation cohort. In addition, after the enrolments there were 12 months of follow-up. The follow-up is still ongoing and the algorithm is still being improved with additional data from the cohort of phase 1.

The inclusion criterion common to both groups of patients was the completion of a bladder endoscopy, either as part of bladder cancer suspicion (de novo, follow-up or relapse) or as part of a lower urinary tract exploration, excluding bladder or prostate cancer suspicion. All patients had to have a negative urinary analysis before the TUR or the screening flexible cystoscopy to avoid patients with urinary tract infection symptoms.

In order to only select patients with bladder urothelial carcinoma, the exclusion criteria were carcinoma of the upper urinary tract and non-urothelial carcinoma (i.e. adenocarcinoma and squamous cell carcinoma). All patients from Group 1 had abdominal CT with late image under furosemide prior to the cytology and the TUR or the screening flexible cystoscopy.

To limit false positives related to the presence of reactive cells, patients with urinary lithiasis, pelvic radiotherapy history and renal transplantation were also excluded.

The cytology examinations were performed on natural voided urine obtained before the TUR or to the screening flexible cystoscopy or to any scopy or catheterisation in order to avoid reactional cells.

The urine samples, fixed with ThinPrep® CytoLit, were split into two equal parts:

- One remained in the inclusion centre for the routine cytology test.
- The other was sent to the Suresnes hospital laboratory, where all urine samples were processed on the same equipment line.

Cytology was analysed in the centre by its referring expert pathologists in cytopathology, according to the Paris System.
2016 requirements. All pathologists were members of the French Society of Pathology and have >10 years of professional experience.

The pathologists analysed the routine cytology before the final result from tissue biopsy.

The urine cells sent to the Suresnes hospital laboratory were deposited on a glass slide using the ThinPrep® UroCyt technique and coloured using the Papanicolaou stain. Finally, the cytopathology slides were covered with a protective film (Tissue-Tek® Sakura).

The cytopathology slides were scanned with a Hamamatsu NanoZoomer S60 scanner at ×40, on three planes (z-stacks) spaced 3 µm apart, with two acquisition channels: white light and fluorescence light (Fig. 1).

The fluorescence acquisition is under patent and is process that the VitaDX Company cannot detail in the present article. The fluorescence observed is intrinsic due to the auto-fluorescence of the Papanicolaou in particular the Orange G6 and eosin contained in EA50 (Eosin Azure), which create a peripheral fluorescence cell. The haematoxylin contained in the Papanicolaou that colours the nucleus, reinforced the fluorescence contrast.

Cell Annotation Based on Digital Slides and Slide Labelling

Two levels of annotations are necessary to train machine learning methods: local annotations and global annotations. Local annotations are used for the detection, classification and segmentation of cell objects. Global annotations define the ground truth for the diagnosis algorithm.

For local annotations, >50 000 cells were selected and labelled according to five categories (by pathologists and cytotechnicians): urothelial cells, squamous cells, inflammatory cells, artefacts, and cell clusters. Then, urothelial cells were analysed and marked by four cytotechnicians specialised in urinary cytology. All these annotations were conducted according to the Paris System 2016 criteria, such as nucleus size, shape, and colour. The cytotechnicians worked on a web-based whole-slide labelling tool developed for this task by VitaDX (Fig. 2).

The global slide annotation process used to train the algorithm formed the patient’s clinical diagnosis. It is based on the results of the ‘gold-standard’ examinations, as follows:

- Negative slide: patient with a negative cytology (negative for high-grade urothelial carcinoma or atypical urothelial cells) and a negative endoscopy.
- Positive slide: patient with histology confirming urothelial neoplasia.

Automated Computed Processing

The VisioCyt algorithm consists in the following five steps (Fig. 3):

1. Object detection: all biological objects belonging to the foreground are detected and segmented at a ×20 zoom scale using a thresholding method.

Fig. 1 Processing from urine to AI diagnosis.
2. Object classification: each detected object is classified as one of the five previously defined types (urothelial cell, squamous cell, inflammatory cell, artefacts or cell clusters) using a ResNet-based neural network.

3. Local feature computation: the nuclei and cytoplasm of urothelial cells are segmented with a Unet-based neural network. Local features based on expert knowledge are computed on the bright-field image of the urothelial cells (nuclear cytoplasmic ratio, nucleus shape and colour, etc.). Confidential fluorescence features were used by the VitaDX Company, combining know-how and patents.

4. Global feature computation: global colour descriptors of the slide staining are computed on all the detected foreground components.

5. Slide classification: the local features from Step 3 are agglomerated and combined with the global features of Step 4 to produce a vector of features representing the slide. A linear classifier produces a diagnosis score between 0 and 1 for the slide. A positive diagnosis corresponds at a score superior to the threshold value of 0.5 and negative under 0.5.

6. Classification quality control: when the classification score is close to the decision threshold, the diagnosis is uncertain. To avoid making a prediction in these cases, a safety zone around 0.5 is established: scores between 0.45 and 0.55 are rejected. Only scores <0.45 or >0.55 are diagnosed respectively as negative and positive.

Performance Estimation Through Cross-Validation

The evaluation of machine learning algorithms relies on dataset split in two: the training data is used to build the algorithm and the test data for an independent performance evaluation [5]. For this first clinical phase, the algorithm was evaluated according to the leave-one-out cross-validation approach. This method consists in generating multiple patient splits (randomised), to create two datasets:

- An evaluation dataset with a unique patient used to measure performance.
- A training dataset with all the other patients.

In this clinical phase, 598 patients were included. For the leave-one-out evaluation, 598 different splits were created, meaning that each patient is only in the evaluation dataset once, and 597 times in the training dataset. The final estimated performance is the average performance for all 598 evaluation splits (Fig. 4).

Results

A total of 598 patients were included in this first study phase, including 149 negative control patients and 449 patients with bladder tumours (219 HGTs and 230 LGTs; Tables 1 and 2). The results below take into account only the patients from the first phase of this study.

Concerning the VisioCyt test specificity, the results cannot present the difference between HGT and LGT specificity given that the test does not differentiate HGTs from LGTs. The result of VisioCyt test is purely positive for bladder carcinoma whatever the grade.

Discussion

Since the first five-class scheme was reported by Papanicolaou and Marshall [6], much progress has been made in the diagnosis of bladder carcinoma using urine tests. Nevertheless, in everyday pathology practice, urinary cytology is a complicated examination when looking to discriminate...
**Fig. 3** Illustration of automated computed processing.

**Fig. 4** Illustration of the leave-one-out cross-validation process used to learn and estimate the performance of the ‘mean model’.

This process is repeated N times in order to see the N patients.
benign from malignant cells, including atypia, especially in the case of the LGTs. It is not that rare for the pathologist not to have enough or all the atypia cells that could confirm urothelial carcinoma diagnoses. The major challenge lies in increasing voided urine cell discrimination to support urothelial carcinoma diagnoses for bladder cancer, considered as one of the most prevalent malignancies. The purpose of the present study was to find out if AI could be an aid for the pathologist to detect tumour cells in voided urine. As the urine cytology specimen contains a wide variety of normal cellular elements (i.e. urothelial cell, squamous cell, umbrella and inflammatory cells, bacterial colonies, proteinaceous casts, corpora amylacea), AI needs to recognise specific parameters to improve diagnoses.

This first study step for the VisioCyt procedure, as reported in the present study, shows that progressively built algorithms provide major support to cytology interpretation. Standard cytology sensitivity is globally reported to be between 40% and 50%, and specificity 80–90% [7]. With AI and using the VisioCyt test, sensitivity increased to 84.9% in the present series. The most impressive results concern LGTs. Of the 230 patients with LGTs, sensitivity increased from 26.3% to 77.0%, respectively for standard cytology and the VisioCyt technique. Increasing sensitivity should have an important impact in two clinical situations: primary detection of bladder tumours in high-risk populations (smokers or professional exposure), and mostly during follow-up after primary LGT resection. During follow-up after LGT resection, urethro-bladder cystoscopy and cytology are both key examinations repeated every 3, 6 or 12 months (depending on multifocality, size and aggressiveness parameters), in line with European Association of Urology (EAU) guidelines [8].

The VisioCyt test should be available in different European centralised laboratories before the end of 2021 with an expected price ranging from €100 to €150 (Euros) for the final payer. The results of the analysis should be returned within 48 h. Depending on the needs of the laboratories and the laboratory management system, the test should be carrying out through software available on a specific cloud or on a computer in the locality.

Currently, cytology’s very poor sensitivity for LGTs is acknowledged in the literature [9–11] and iterative cystoscopies are considered by patients as an unpleasant approach, often leading to cessation of follow-up [12–14] that sometimes required for 10 years. Moreover, cystoscopy is a relatively expensive examination, potentially accompanied by several adverse effects. Therefore, early and accurate help aiming to improve diagnosis of LGTs has become crucial. With sensitivity >75% for the VisioCyt technique, the clinician’s point of view should be to discuss the possibility of reducing the number of cystoscopies during follow-up after NMIBC, especially for cases with LGTs. Therefore, with AI, new perspectives could improve patient compliance with the monitoring protocol, aiming to reduce the number of repeated cystoscopies.

Otherwise, as it is a simple procedure, it could be used more often in order to reduce the delay in detecting recurrence. A

### Table 1

<table>
<thead>
<tr>
<th>Test, n (%)</th>
<th>Global sensitivity, n (%)</th>
<th>Sensitivity HGT, n (%)</th>
<th>Sensitivity LGT, n (%)</th>
<th>Specificity</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>VisioCyt test</td>
<td>381 (84.9)</td>
<td>204 (92.6)</td>
<td>177 (77.0)</td>
<td>121 (81.2)</td>
<td>121 (64.0)</td>
</tr>
<tr>
<td>Pathologists</td>
<td>193 (43.0)</td>
<td>133 (61.1)</td>
<td>60 (26.3)</td>
<td>149 (100.0)*</td>
<td>149 (36.7)</td>
</tr>
<tr>
<td>Total</td>
<td>449 (100.0)</td>
<td>219 (100.0)</td>
<td>230 (100.0)</td>
<td>149 (100.0)</td>
<td>149 (100.0)</td>
</tr>
</tbody>
</table>

*The pathologists’ specificity is 100% due to clinical trial construction, as the inclusion criteria for patients in the negative control group was based on two conditions: a negative cytological examination result and a negative bladder endoscopy. In the event of positive or uninterpretable cytology, the patient was excluded from the negative control group. †The negative predictive value (NPV) presented here is not representative of the NPV that can be routinely obtained because the prevalence is forced by clinical inclusion and is not representative of natural prevalence.

### Table 2

<table>
<thead>
<tr>
<th>Threshold variation</th>
<th>Rejected slides, n (%)</th>
<th>Global sensitivity, n (%)</th>
<th>Sensitivity HGT, n (%)</th>
<th>Sensitivity LGT, n (%)</th>
<th>Specificity, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5-0.5</td>
<td>0 (0)</td>
<td>373 (83.1)</td>
<td>197 (90.0)</td>
<td>176 (76.5)</td>
<td>115 (77.2)</td>
</tr>
<tr>
<td>0.475-0.525</td>
<td>31 (5.2)</td>
<td>377 (83.9)</td>
<td>202 (92.3)</td>
<td>175 (76.0)</td>
<td>119 (79.7)</td>
</tr>
<tr>
<td>0.45-0.55</td>
<td>49 (8.2)</td>
<td>381 (84.9)</td>
<td>204 (92.6)</td>
<td>177 (77.0)</td>
<td>121 (81.2)</td>
</tr>
<tr>
<td>0.425-0.575</td>
<td>69 (11.5)</td>
<td>387 (86.3)</td>
<td>205 (93.6)</td>
<td>182 (79.2)</td>
<td>122 (81.7)</td>
</tr>
<tr>
<td>0.4-0.6</td>
<td>100 (16.7)</td>
<td>396 (88.1)</td>
<td>206 (93.9)</td>
<td>189 (82.1)</td>
<td>121 (81.2)</td>
</tr>
<tr>
<td>Total</td>
<td>598 (100.0)</td>
<td>449 (100.0)</td>
<td>219 (100.0)</td>
<td>230 (100.0)</td>
<td>149 (100.0)</td>
</tr>
</tbody>
</table>

This drop zone 0.45-0.55 was kept, as it represents the best ratio performance (sensitivity/specificity) vs the percentage of unanalysable VisioCyt cytology tests.
lot of tests, e.g. the bladder tumour-associated antigen (BTA) TRAK assay, nuclear matrix protein 22 (NMP22) and uCyt1 tests, have been approved by the United States Food and Drug Administration (FDA) but their utility is still very limited [15]. None of them have been incorporated in guidelines [16–18] and suffer from low sensitivity, especially for LGTs [19]. To improve cytology performance, new techniques have been also tested for harvesting intact urinary-exfoliated tumour cells, but the economic impact is high and the gain in term of sensitivity remains poor in cases with LGTs [20]. Liquid biopsies using DNA- and RNA-based markers in urine are promising, but are still being assessed [21]. One must consider VisioCyt as a new cost-effective technical approach, it is not a ‘test’, but instead a new concept using AI to help pathologists.

As mentioned, the main objective was to improve test sensitivity, especially for cases with LGTs. The variation of the threshold decision (positive or negative) could improve sensitivity, but this increases the number of cases where the test may not be contributory (percentage of rejected slides). One of the interests in using AI on digital cytology is that the system allows us to vary the threshold to improve sensitivity and specificity, but to the detriment of the number of rejected cytologies. As we can see in Table 2, the larger the threshold, the greater the sensitivity, whereas specificity remains at ~81%, compared to the number of rejected slides that increases faster (∼2 between 0.45–0.55 and 0.4–0.6). Finally, the most acceptable threshold is between 0.45–0.55 and 0.4–0.6, when sensitivity increases by 4% (84.6%/88.1%), while the percentage of patients without a VisioCyt diagnostic increases from 8.2% to 16.7%. In our case, a threshold between 0.45 and 0.55 seems to be an acceptable ratio between performance and the number of patients without a VisioCyt result.

Concerning specificity, the present study demonstrates that VisioCyt does not bring a huge benefit compared to pathologist diagnosis. Firstly, it was not the main objective and secondly, by using new morphometric parameters not included is this first algorithm (i.e. chromasia level, chromatin aspect, cytoplasmic consistence, nuclear border analysis, etc.), combined with more urines samples, the algorithm could be improved, and thus be more relevant. Inflammation is admitted to be a cause of false-positive diagnosis [22], but with new algorithms and progressively added parameters, AI could probably avoid this disruptive element. In practice, certain cytomorphological features associated with LGT could probably be included in new algorithms [23,24]. Nevertheless, definitively, it will probably not be that valuable in daily practice, firstly because cytology has suitable specificity and secondly because positive cytology always needs to be confirmed by histological samples taken from an endoscopic resection.

The AI diagnostic performance should increase with more data and added morphological parameters. With appropriate ethical control, the VisioCyt test based on deep learning using improving knowledge will probably over time also appear as an effective aid, and will become more and more relevant, in order to be efficient in everyday urocytopathology practice. VisioCyt could be an important modality both in screening for new carcinomas and for surveying cancer recurrence.

Acknowledgements

VitaDX (Rennes, France) was involved in this trial as promoter and co-sponsor associated with public funds. As a promoter, VitaDX has selected the Clinical Research Organisation, the statistical analysis company and co-selected the investigators with the principal investigator. VitaDX has co-defined the design of this study, in order to be sure that the algorithm constraints would be taken into account. VitaDX was also involved in the preparation of this article by the description of the algorithm material and method development, and also, the discussion about the limits of the technology.

This research behind it would not have been possible without the participation and support of all the persons who in one way or another have contributed in making this research project possible. We would like to express our deep and sincere gratitude to the following urologists, pathologists and cytotechnicians: Prof. L. Guy (Urologist, 63000 Clermont-Ferrand); Prof. J.L. Kemeny (Pathologist, 63000 Clermont-Ferrand); Dr C. Desplechain (Pathologist, 63000 Clermont-Ferrand); Dr R. Borges-Reis (Pathologist, 34000 Montpellier); Dr J. Thomassin-Piana (Pathologist, 13009 Marseille); Dr G. Verhoest (Urologist, 35000 Rennes); Prof. N. Rioux-Leclercq (Pathologist, 35000 Rennes); Prof. O. Bouchot (Urologist, 44000 Nantes); Dr C. Kandel-Aznar (Pathologist, 44000 Nantes); Dr K. Autain-Renaudin (Pathologist, 44000 Nantes); Prof. S. Droupy (Pathologist, 30009 Nantes); Dr S. Gonzalez (Pathologist, 30009 Nantes); Dr F. Kleinclauss (Urologist, 25000 Besançon); Dr B. Cochand-Priollet (Pathologist, 75014 Paris); Dr Y. Rouach (Urologist, 92160 Antony); Dr A. Balaton (Pathologist, 91940 Les Ulis); Dr P. Gros (Pathologist, 91940 Les Ulis); Prof. F. Kleinclauss (Urologist, 25000 Besançon); Prof. J. Hubert (Urologist, 54511 Vandoeuvre-les-nancy); Dr M. Antoine (Pathologist, 54511 Vandoeuvre-les-nancy); Mr H. Debaque (Cytotechnician, 91300 Massy); Mr J.M. Desnoyer (Cytotechnician, 89000 Auxerre); Mrs C. Selle-Patoz (Cytotechnician, 3555 Dudelange-Luxembourg), and Mrs D. Hmaidi (Cytotechnician, 06200 Nice).

Disclosure of Interests

Thierry Lebret: MSD, Astza Zeneca, Roche, Ferring, Janssen, Ipsen, BMS, vitaDx. Géraldine Pignot: consulting or advisory roles: Janssen, Astellas, Roche, Pifzer, Arquer, Astra-Zeneca, MSD, BMS, Bouchara-Recordati, Ipsen; travel expenses:

References
12 Gezginici E, Iyigun E, Kibar Y, Bedir S. Three distraction methods for pain reduction during cystoscopy: a randomized controlled trial evaluating effects on pain, anxiety, and satisfaction. J Endourology 2018; 32: 1078–84

Correspondence: Thierry Lebret, Urology Department, Foch Hospital, 40 rue Worth, Suresnes 92150, France.

e-mail: t.lebret@hopital-foch.org

Abbreviations: AI, artificial intelligence; HGT, high-grade tumour; LGT, low-grade tumour; NMIBC, non-muscle-invasive bladder carcinoma; TUR, transurethral resection.