510(k) SUBSTANTIAL EQUIVALENCE DETERMINATION
DECISION SUMMARY

I Background Information:

A 510(k) Number

K241232

B Applicant

Ibex Medical Analytics Ltd.

C Proprietary and Established Names

Galen™ Second Read™

D Regulatory Information

Product Code(s)	Classification	Regulation Section	Panel
QPN	Class II	21 CFR 864.3750 - Software algorithm device to assist users in digital pathology	PA - Pathology

II Submission/Device Overview:

A Purpose for Submission:

New Device

B Type of Test:

Software only device

III Intended Use/Indications for Use:

A Intended Use(s):

See Indications for Use below.

B Indication(s) for Use:

Galen™ Second Read™ is a software only device intended to analyze scanned histopathology whole slide images (WSIs) from prostate core needle biopsies (PCNB) prepared from hematoxylin & eosin (H&E) stained formalin-fixed paraffin embedded (FFPE) tissue. The device is intended to identify cases initially diagnosed as benign for further review by a pathologist. If Galen™ Second Read™ detects tissue morphology suspicious for prostate adenocarcinoma (AdC), it provides case- and slide-level alerts (flags) which includes a heatmap of tissue areas in the WSI that is likely to contain cancer.

Galen™ Second Read™ is intended to be used with slide images digitized with Philips Ultra Fast Scanner and visualized using the Galen™ Second Read™ user interface.

Galen™ Second Read™ outputs are not intended to be used on a standalone basis for diagnosis, to rule out prostatic AdC or to preclude pathological assessment of WSIs according to the standard of care.

C Special Conditions for Use Statement(s):

Rx - For Prescription Use Only

IV Device/System Characteristics:

A. Device Description:

The Galen Second Read (version 3.1-US) uses software algorithms, derived from deterministic deep convolutional neural networks. It analyzes WSIs of H&E-stained PCNB slides originating from FFPE tissue sections, that were originally diagnosed as benign by the pathologist. WSIs that are likely to contain prostatic AdC are flagged by providing heatmap of the relevant tissue areas for a second read by the pathologist. The final diagnosis is determined by the pathologist after review of the flagged findings.

The Galen Second Read is cloud-hosted and utilizes scanned WSIs generated from the Philips Ultra Fast scanner (UFS). For each input WSI, the Galen Second Read automatically analyzes the WSI and outputs the following:

- Binary classification of the likelihood (high/low) to contain AdC based on a predetermined threshold of the neural network output.
- For slides classified as high likelihood to contain AdC, slide-level findings are flagged and visualized (AdC score and heatmap) for additional review by a pathologist.
- For slides classified as low likelihood to contain AdC, no additional output is available.

Galen Second Read key functionalities include image upload and analysis, flagging slides with high likelihood to contain AdC and displaying WSIs uploaded to the system along with the analysis results. Flagged findings constitute a recommendation for additional review by a pathologist.

Figure 1 below presents a high-level view of the Galen Second Read design and workflow.

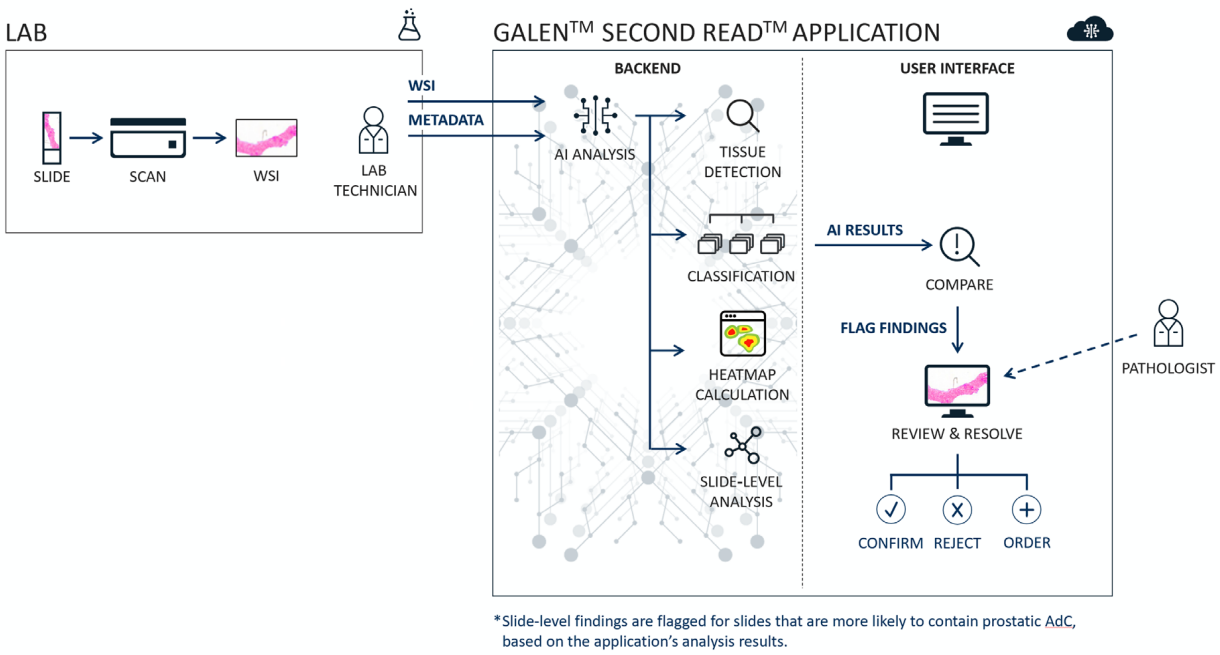

Figure 1. Galen Second Read High-Level System Design

Galen Second Read is operated as follows:

1. Scanned digital images of PCNB are acquired using the Philips IntelliSite Pathology Solution (PIPS) Ultra Fast Scanner (UFS). Image and other related quality control steps are performed per the scanner instructions for use and any additional user site specifications.
2. If a case is determined to be benign after initial review by the pathologist, then all WSIs from the case are uploaded along with its associated metadata (i.e., textual data describing the case and its WSIs such as Case ID, Slide ID, Tissue and Stain) to the cloud-hosted Galen Second Read. Once the WSI and its metadata are available, the data is automatically processed (analyzed) in the background by Galen Second Read.
3. For every slide (WSI), the device provides case- and slide-level alerts (flags). All flagged findings are available in the Galen Second Read user interface and the pathologist can select a patient case and open each flagged (pink square) WSI for additional review. The available information for review includes AdC score (where a higher AdC score indicates a high suspicion the slide contains AdC) and AdC heatmap marking the tissue region in the WSI that may contain cancer. The heatmap opacity can be controlled and toggled on/off to allow unobstructed review of the WSI. When a heatmap is displayed, a legend is shown, indicating the current active heatmap name and stating the color scale, e.g., blue for low likelihood, red for high likelihood.
4. After second review of the WSIs, the pathologist can either confirm the device result (revise the original diagnosis from benign to malignant), reject the device result (no additional action is required) or decide that more information (e.g., additional stains) is needed before making a diagnosis.
5. The final determination of diagnosis is made by the pathologist based on the histologic findings and/or additional tests. Pathologists should follow the standard of care (SoC) to obtain additional information, as needed, to render a final diagnosis.

Interoperable components intended for use with Galen Second Read and minimum system requirements are provided in Table 1 and Table 2, respectively.

Table 1: WSI Scanner and Display

Manufacturer	Model
Philips Medical Systems Nederland B.V.	Philips IntelliSite Pathology Solution (PIPS) [Ultra-Fast Scanner (UFS)] Image Management System - Philips IMS
Display	Philips PS27QHDCR, Barco N.V. NV MDPC-8127

Table 2: Computer Environment/System Requirements

Workstation Component	Specifications
Computer System	RAM: 8.0 GB or higher CPU: 1 GHz or higher
Web Browser	Google Chrome v120 or later
Operating system	Windows v10 or Mac OS v11 or higher
Network	Internet access Minimum 20 Mbps download speed

B. Algorithm Development

Galen Second Read algorithm development was performed on training, tuning, and test datasets. Each dataset contained slides from unique patients ensuring that training, tuning, and test datasets do not have any slides, cases, or patients in common. Slides were labeled by board-certified pathologists as benign or cancer, including other cell types and structures typically found in a PCNB, as applicable. Datasets and their characteristics are provided in Table 3 and 4.

These datasets were completely independent from each other and the validation (pivotal clinical performance studies) datasets.

Table 3: Dataset Split for Training, Tuning and Test Sets

Algorithm Development		
Training Dataset	**Tuning Dataset**	**Test Datasets**
1,312 de-identified slides from PCNB Slides were scanned using Philips UFS, Hamamatsu NanoZoomer S210 scanner, Leica Aperio AT2 scanner, Ventana DP 200 scanner and 3DHistech P250 scanner.	1335 slides scanned using Philips UFS.	5312 slides scanned using Philips UFS and Leica Aperio AT2 scanner.

Table 4: Distribution of Slide Images by Geography in Algorithm Development

Race	Training Dataset	Tuning dataset	Test Datasets
EMEA	1011 (77%)	1335 (100%)	3685 (69.4%)

US	232 (17.7%)	NA	1627 (30.6%)
APAC	63 (4.8%)	NA	NA
OUS	6 (0.46%)	NA	NA
Total	**1312 (100%)**	**1335 (100%)**	**5312 (100%)**

EMEA: Europe, the Middle East and Africa (including Israel, UK, Germany, Netherlands, Austria, Spain and France sites); US - United States (including US labs); APAC - Asia–Pacific (including India, Australia, Japan and labs); OUS-Outside the US - other geographies (including Brazil lab)

C. Instrument Description Information:

6. Instrument Name:
   Galen Second Read

7. Specimen Identification:
   Galen Second Read uses WSIs of H&E-stained glass slides originating from PCNBs obtained from PIPS UFS to produce a WSI file. The WSI files, and their associated metadata, are uploaded to the Galen Second Read. When displayed in the user interface, a preliminary view of the WSI is available in the form of a slide thumbnail showing the image itself next to the slide ID/barcode as obtained by the scanner.

8. Specimen Sampling and Handling:
   Specimen sampling and handling are performed prior to and independent of the use of the Galen Second Read. Specimen sampling includes PCNB specimens which are processed using histology techniques. The FFPE tissue section is H&E stained. Digital images are then obtained from these glass slides using the PIPS UFS.

9. Calibration:
   Not applicable

10. Quality Control:
    Before reading pathology images using the Galen Second Read, pathologists should ensure that all scanned slide images from all slides from a case have been uploaded. It is the user's responsibility to apply appropriate process and quality assurance steps to ensure the quality of the images obtained and, when necessary, support the diagnosis by use of light microscopy.

V Substantial Equivalence Information:

A. Predicate Device Name(s):

Paige Prostate

B. Predicate Device Number(s):

DEN200080

C. Comparison with Predicate(s):

Device & Predicate Device(s):	K241232	DEN200080
Device Trade Name	Galen Second Read	Paige Prostate
General Device Characteristic Similarities		
Intended Use/ Indications For Use	Galen™ Second Read™ is a software only device intended to analyze scanned histopathology whole slide images (WSIs) from prostate core needle biopsies (PCNB) prepared from hematoxylin & eosin (H&E) stained formalin-fixed paraffin embedded (FFPE) tissue. The device is intended to identify cases initially diagnosed as benign for further review by a pathologist. If Galen™ Second Read™ detects tissue morphology suspicious for prostate adenocarcinoma (AdC), it provides case- and slide-level alerts (flags) which includes a heatmap of tissue areas in the WSI that is likely to contain cancer. Galen™ Second Read™ is intended to be used with slide images digitized with Philips Ultra Fast Scanner and visualized using the Galen™ Second Read™ user interface. Galen™ Second Read™ outputs are not intended to be used on a standalone basis for diagnosis, to rule out prostatic AdC or to preclude pathological assessment of WSIs according to the standard of care.	Paige Prostate is a software only device intended to assist pathologists in the detection of foci that are suspicious for cancer during the review of scanned whole slide images (WSI) from prostate needle biopsies prepared from hematoxylin & eosin (H&E) stained formalin fixed paraffin embedded (FFPE) tissue. After initial diagnostic review of the WSI by the pathologist, if Paige Prostate detects tissue morphology suspicious for cancer, it provides coordinates (X,Y) on a single location on the image with the highest likelihood of having cancer for further review by the pathologist. Paige Prostate is intended to be used with slide images digitized with Philips UFS and visualized with Paige FullFocus WSI viewing software. Paige Prostate is an adjunctive computer-assisted methodology and its output should not be used as the primary diagnosis. Pathologists should only use Paige Prostate in conjunction with their complete standard of care evaluation of the slide image.
Specimen Type	PCNBs prepared from H&E stained FFPE tissue	Same
Type of Test Performed	Software device intended to identify cases initially diagnosed as benign for further review by a pathologist. The Galen Second Read provides case-level alerts (flag) and slide-level alerts	Software device to identify digital histopathology images of PCNBs that are suspicious for cancer and to localize a focus with the highest probability for cancer

	(heatmap) for the WSIs, if prostate adenocarcinoma (AdC) is suspected by the device.	
Image file format	Philips UFS iSyntax File	Same
Type of Software Application	Internet browser-based applications	Same
Device interoperable components	PIPS UFS and applicable cleared Display	Same
	General Device Characteristic: Differences	
End User's Interface	Galen Second Read	FullFocus (K201005)
Image Manipulation Functions	Panning, zooming and measurements (distance)	Panning, zooming, color manipulation function, annotations, and measurements (distance & area)
Principle of Operation	After WSI images are successfully acquired using PIPS UFS and related quality control steps are performed, the scanned WSI digital images of the cases initially diagnosed as benign are sent to the Galen Second Read and immediately processed and analyzed. In case the slide analysis results indicate a high likelihood to contain AdC, the application flags for additional review by the pathologist. All flagged findings are available in the Galen Second Read and the pathologist can select a patient case and open each flagged WSI for an additional review. The available information for review includes AdC score (the likelihood to contain AdC) and the AdC heatmap marking the region suggestive to include cancer. Based on the device output, the pathologist can re-examine the slide and modify the original diagnosis of the cases to reflect the additional findings, if required.	After WSI images are successfully acquired using PIPS UFS and related quality control steps are performed, the scanned digital images are immediately processed by Paige Prostate. The pathologist selects a patient case and opens the WSI for review in the designated digital pathology viewing software, and after review and diagnosis done, Paige Prostate is activated and outputs binary classification (suspicious/not) for cancer based on predefined threshold by the neural network and if suspected, a single coordinate (X,Y) of the location with the highest probability of cancer on an image determined to be suspicious for cancer. Based on the device output, the pathologist can reexamine the slide and modify the original diagnosis to reflect the additional findings.

VI Standards/Guidance Documents Referenced:

1. FDA Guidance “Content of Premarket Submissions for Device Software Functions"; June 14, 2023
2. FDA Guidance “Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions”; September 2023
3. FDA Guidance “Off-The-Shelf Software Use in Medical Devices”; August 2023
4. IEC 62304:2006+A1:2015, Medical Device Software – Software Life Cycle Processes
5. CLSI EP12-Ed3 “Evaluation of Qualitative, Binary Output Examination Performance” (2023)

VII Performance Characteristics (if/when applicable):

A. Analytical Performance:

1. Precision/Reproducibility

The objective of the studies was to assess the precision (repeatability and reproducibility) of Galen Second Read in identifying WSI of PCNBs suspicious for cancer, its localization accuracy and localization precision measured by its sensitivity and specificity. Precision studies included three separate studies: precision (repeatability and reproducibility) of the Galen Second Read Prostate slide-level outcome, localization accuracy of the produced heatmap within the cancer area and localization precision.

Slides used in precision studies were different from the slides used during the development (algorithm training and testing) of the Galen Second Read. The precision study was conducted at 4 sites - 2 US sites and 2 OUS sites. De-identified H&E slides were scanned with a PIPS scanner at 40x magnification. Reported original sign-out diagnosis was used as ground truth (GT) for the precision study. To evaluate precision at a slide level, unique consecutive cases diagnosed as cancer or benign originating from PCNBs were enrolled. One H&E slide from each case was selected by an enroller pathologist to be processed by the device. Cases diagnosed as prostatic AdC or other cancer and atypical small acinar proliferation (ASAP) were considered positive. The enroller selected the first slide that was compatible with the case category. For cancer cases, the first slide reported with the respective Gleason score identical to the Gleason score assigned to the case was selected. The study included 39 positive and 38 negative slides. The endpoints for the repeatability and reproducibility study were defined at the *level of slide*, as the binary status Positive or Negative.

Subject Characteristics

Tables 5−7 provide descriptive statistics for the slide analysis sets used for the precision, localization precision, and localization accuracy studies.

Table 5: Distribution of Slide Characteristics in Analytical Studies

Slide-Level Precision (Reproducibility & Repeatability)		
Characteristic	**Cancer**	**Benign**
	(N=39)	**(N=38)**

ASAP	0	0
Atrophy Present	Not Reported	35 (92.1%)
High-Grade PIN Present	Not Reported	0
Tumor size (per original report)		
≤ 0.5 mm	3 (7.7%)	
> 0.5 mm	10 (25.6%)	
Not reported	26 (66.7%)	
Gleason Grade		
Grade Group 1	31 (79.5%)	
Grade Group 2	2 (5.1%)	
Grade Group 3	3 (7.7%)	
Grade Group 4	2 (5.1%)	
Grade Group 5	1 (2.6%)	
Localization Precision		
Characteristic	**Cancer (N=14)**	**Benign (N=4)**
Original Diagnosis		
Cancer	14 (100.0%)	1 (25.0%)
ASAP	0	
Benign		3 (75.0%)
Tumor size (per original report)		
≤ 0.5 mm	3 (21.4%)	
> 0.5 mm	9 (64.3%)	1 (25.0%)
Not reported	2 (14.3%)	
Gleason Grade		
Grade Group 1	7 (50.0%)	1 (25.0%)
Grade Group 2	2 (14.3%)	
Grade Group 3	3 (21.4%)	
Grade Group 4	1 (7.1%)	
Grade Group 5	1 (7.1%)	
Localization Accuracy		
Characteristic	**Cancer (N=31)**	**Benign (N=9)**
Original Diagnosis		
Cancer	30 (96.8%)	3 (33.3%)
ASAP	1 (3.2%)	
Benign		6 (66.7%)
Tumor size (per original report)		
≤ 0.5 mm	2 (6.5%)	1 (11.1%)
> 0.5 mm	27 (87.1%)	2 (22.2%)
Not reported	2 (6.5%)	
Cancer: Tumor size (per original report)		

≤ 0.5 mm	2 (6.5%)	1 (11.1%)
> 0.5 mm	26 (83.9%)	2 (22.2%)
Not reported	2 (6.5%)	
ASAP: Tumor size (per original report)		
≤ 0.5 mm	0	
> 0.5 mm	1 (3.2%)	
Gleason Grade		
Grade Group 1	12 (38.7%)	3 (33.3%)
Grade Group 2	9 (29.0%)	
Grade Group 3	4 (12.9%)	
Grade Group 4	3 (9.7%)	
Grade Group 5	2 (6.5%)	
Not reported	1 (3.2%)	

Table 6: Descriptive Statistics of Subjects' Age at the Time of Biopsy

Analysis Set / Ground Truth		Subjects' Age at Time of Biopsy (Years)					
		Mean	SD	Min	Median	Max	N
Slide-Level Precision Analysis Set	**Positive**	70.4	6.9	54.0	72.0	82.0	39
	Negative	65.0	8.0	48.0	65.5	81.0	38
	All	67.7	7.9	48.0	70.0	82.0	77
Localization Precision Analysis Set	**Positive**	73.7	5.9	61.0	73.5	82.0	14
	Negative	60.3	8.2	48.0	64.0	65.0	4
	All	70.7	8.5	48.0	72.5	82.0	18
Localization Accuracy Analysis Set	**Positive**	71.2	6.9	57.0	72.0	86.0	31
	Negative	64.7	8.6	48.0	65.0	77.0	9
	All	69.7	7.7	48.0	71.0	86.0	40

Table 7: Distribution of Cases by Ground Truth

Analysis Set / Ground Truth		ASAP		Cancer		Benign		Total	
		N	%	N	%	N	%	N	%
Precision Analysis Set	**Positive**	0	0.0	39	100.0	0	0.0	39	100.0
	Negative	0	0.0	0	0.0	38	100.0	38	100.0
	Total	0	0.0	39	50.6	38	49.4	77	100.0
	Positive	0	0.0	14	100.0	0	0.0	14	100.0
	Negative	0	0.0	1	25.0	3	75.0	4	100.0

Localization Precision Analysis Set	**Total**	0	0.0	15	83.3	3	16.7	18	100.0
Localization Accuracy Analysis Set	**Positive**	1	3.2	30	96.8	0	0.0	31	100.0
	Negative	0	0.0	3	33.3	6	66.7	9	100.0
	Total	1	2.5	33	82.5	6	15.0	40	100.0

Slide-Level Precision Study

i. **Repeatability (Within Scanner/Operator):** Each slide was evaluated repeatedly 3 times using a single scanner by the same operator (Scanner 1 acquired by Operator 1). Table 8 presents the agreement rates between device output and ground truth for the three runs performed by Operator 1. In all three runs, the same 38 out of the 39 positive slides were correctly identified as true positives (97.4%). Therefore, the overall percent of correct calls for positive slides across all 3 runs was 97.4%. For negative slides, in the first and third run the same 33 out of 38 slides were correctly identified as true negatives (86.8%), and in the second run, 2 additional slides were correctly identified, for a total of 35 out of 38 slides that were identified as true negatives (92.1%). The overall percent correct calls for negative slides across all 3 runs was 88.6%.

Table 8: Within-Scanner Precision: Percent of Correct Calls

Operator / Scanner / Run	**Agreement with Ground Truth by Slide Type**			
	Positive Slides		**Negative Slides**	
	Percent of Correct Calls, % (n/N)	**95% CI**	**Percent of Correct Calls, % (n/N)**	**95% CI**
Run 1	97.4% (38/39)	(86.8%, 99.5%)	86.8% (33/38)	(72.7%, 94.2)
Run 2	97.4% (38/39)	(86.8%, 99.5%)	92.1% (35/38)	(79.2%, 97.3%)
Run 3	97.4% (38/39)	(86.8%, 99.5%)	86.8% (33/38)	(72.7%, 94.2%)
Overall Average*	97.4%	(92.1%; 99.1%)	88.6%	(81.5%; 93.2%)

* 95%CI are calculated using the Wilson score method (the 95%CIs can be overstated).

The within-slide repeatability across the 3 runs is summarized in Table 9. As described above, the same 38 out of 39 positive slides were correctly identified by the device in all 3 runs [(97.4%, 95%CI: (86.5%, 99.9%)].

Table 9: Within-Slide Repeatability: Percent Correct Calls for Scan Repetitions (Slide-Level Precision Analysis Set)

Type of Slide Images	**Total Number of Runs**	**Percent of Slide Images in Which All 3 Runs Yielded Correct Result**		**Percent of Runs with Correct Result Out of All Runs**	
		% (n/N)	**95% CI**	**% (n/N)**	**95% CI**
Positive	**117**	97.4% (38/39)	(86.8%; 99.5%)	97.4% (114/117)	(92.7%; 99.1%)

Negative	**114**	86.8% (33/38)	(72.7%; 94.2%)	88.6% (101/114)	(81.5%; 93.2%)

ii. Reproducibility (Between-Scanner/Operator): Each slide was scanned using 3 different scanners, where each scanner was operated by a different operator (one operator per scanner): total 3 repeated assessments (Scanner 1 acquired by Operator 1, Scanner 2 acquired by Operator 2, Scanner 3 acquired by Operator 3); 39 positive and 38 negative slides = 231 runs.

Table 10 presents the percentages of correct calls by the device vs. ground truth for the three scanner/operators, along with the overall percent of correct calls across the scanner/operators. For two scanner/operators, the same 38 out of 39 positive slides (97.4%) were correctly identified as true positives, and for one scanner/operator, 37 slides were correctly identified as true positives (94.9%). The overall percent of correct calls for positive slides across all 3 operators was 96.6%. For two scanner/operators, 33 out of 38 negative slides (86.8%) were correctly identified as true negatives, and for the third scanner/operator, 32 slides were correctly identified as true negatives (84%). The overall percent of correct calls for negative slides across all 3 operators was 86.0%.

Table 10: Reproducibility (Between-Scanner/Operator): Percent of Correct Calls

	Agreement with Ground Truth by Slide Type			
	Positive Slides		**Negative Slides**	
Operator / Scanner	**Percent of Correct Calls, % (n/N)**	**95% CI**	**Percent of Correct Calls, % (n/N)**	**95% CI**
Scanner 1/Operator 1	97.4% (38/39)	(86.8%; 99.5%)	86.8% (33/38)	(72.7%; 94.2%)
Scanner 2/Operator 2	94.9% (37/39)	(83.1%; 98.6%)	86.8% (33/38)	(72.7%; 94.2%)
Scanner 3/Operator 3	97.4% (38/39)	(86.8%; 99.5%)	84.2% (32/38)	(69.6%; 92.6%)
Overall Average*	96.6%	(92.7%; 99.1%)	86.0%	(78.4%; 91.2%)

* 95%CI are calculated using the Wilson score method (the 95%CI can be overstated)

The within-slide reproducibility across the 3 scanner/operators is summarized in Table 11. Out of a total of 117 repeated slide-runs, the device correctly identified 113 as true positive slides [96.6%, 95%CI: (91.5%, 98.7%)] and out of a total of 114 repeated slide-runs, the device correctly identified 98 as true negative slides [86.0%, 95% CI: (78.4%, 91.2%)].

Table 11: Reproducibility (Between-Scanner-Between-Operator): Percent of Correct Calls for Repetitions with 3 Different Scanner/Operators

Type of Slide Images	**Total Number of Runs**	**Percent of Slide Images in Which All 3 Runs Yielded Correct Result**		**Percent of Runs with Correct Result Out of All Runs**	
		% (n/N)	**95% CI**	**% (n/N)**	**95% CI**
Positive	117	94.9% (37/39)	(83.1%; 98.6%)	96.6% (113/117)	(91.5%; 98.7%)

Negative	114	81.6% (31/38)	(66.6%, 90.8%)	86.0% (98/114)	(78.4%, 91.2%)

2. **Localization Accuracy Study**

Galen Second Read generates the cancer heatmap in a two-step process: First, the tissue detection model detects the tissue areas in the slide. The model also detects out of focus (OOF)) areas. A tissue mask is created (without the OOF areas). Second, the classification algorithm (only on tissue areas) gives each pixel a probability for cancer (between 0 and 1). To visualize those probabilities, a heatmap is created as shown below in Figure 2.

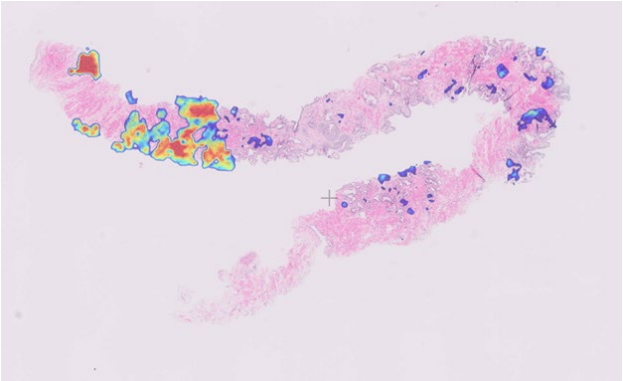

Figure 2. Cancer Heatmap

The lowest threshold for heatmap is 0.25 (pixel with a lower probability will not be colorized i.e., will be transparent). The rest of the spectrum (0.25 and above) will have one out of 16 colors from blue (lowest) to red (highest), the thresholds are shown below in Figure 3.

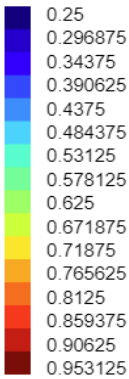

Figure 3. Heatmap Color Spectrum and its Matching Score Thresholds

The goal of the localization accuracy study was to demonstrate that a heatmap (area of concern) produced by the device is accurate and may provide the pathologist an area of concern that they can focus on and determine the correct diagnosis.

To evaluate localization accuracy 31 cancer cases and 9 benign cases were selected. All cases were de-identified, and one slide of each unique case was selected by an enroller pathologist as the representative slide of the case. For cancer cases, the first slide reported with the respective Gleason score identical to the Gleason score assigned to the case was selected. For each slide in the localization and localization precision analysis, one ROI (region of interest) was chosen. The enroller pathologist selected the ROI, that was defined as one entire biopsy level, most representative of the slide. In the ROI, the annotating pathologists marked

polygons (annotations) with a label “cancer” or a label “unsure”. All pixels inside the ROI, but outside any marked polygon are considered “benign”. GT was established for the pixels and defined as the majority of diagnoses reported by at least two out of the three pathologists not participating in the clinical validation studies. The study slides (31 cancer slides and 9 benign slides) were annotated by three pathologists blinded to each other’s annotation/diagnosis (and to the device results) as follows:

1. Annotate cancer at a whole slide image level selected by the Enroller pathologist (the selected ROI), representing the whole tissue in the specific level on the slide. The annotators were instructed to mark "tightly"' the cancer area margins.
2. Annotations must include only cancer foci. All tissue areas inside the ROI, but outside any marked polygon are considered benign.

The first step (Step#1) in localization accuracy was to demonstrate that the entire area of a heatmap covers all positive regions. The entire heatmap area (from the warmest up to the coldest color) can be treated as “rule-out” area. In other words, pixels outside the marked area can be safely considered as “negative”, and pathologists review can mainly focus on the heatmap area. Table 12 presents sensitivity and NPV results obtained for Step#1. Note that each parameter (sensitivity and NPV) was first calculated at the level of slide, and then summarized over all slides. For that reason, N=31 for sensitivity estimate, and N=40 for NPV estimate.

Table 12: Descriptive Statistics for Localization Sensitivity and NPV for the Entire Heatmap Area

Parameter	Mean	SD	Min	Median	Max	N
Sensitivity (%)	98.7	2.2	90.0	99.7	100.0	31
NPV (%)	99.8	0.5	96.9	100.0	100.0	40
Specificity (%)	91.0	10.1	53.1	94.7	99.7	40
PPV (%)	31.4	27.4	0.0	30.3	87.0	40

The next step (Step#2) was to demonstrate that the warmest sub-area of a heatmap contains mainly positive pixels. The warmest sub-area can be treated as “rule-in” area. In other words, pixels within the warmest area should be positive with high probability, and very low likelihood for False Positive. Table 13 presents specificity and PPV results obtained in the study for Step#2. Note that each parameter (specificity and PPV) was first calculated at the level of slide, and then summarized over all slides. For some slides, warmest area was not produced as part of the heatmap, therefore for PPV estimation N=27.

Table 13: Descriptive Statistics for Localization Specificity and PPV for the Warmest Heatmap Area

Parameter	Mean	SD	Min	Median	Max	N
Specificity (%)	100.0	0.0	99.8	100.0	100.0	40
PPV (%)	99.6	1.0	95.7	100.0	100.0	27
Sensitivity (%)	25.0	21.8	0.0	18.7	79.6	31
NPV (%)	92.8	11.8	59.2	98.7	100.0	40

Step#3 was to demonstrate that all accuracy parameters (sensitivity, specificity, NPV and PPV) associated with the intermediate areas represented by different colors (from warmest to coldest) have monotonic pattern, i.e., the change of a color in the heatmap from coldest to warmest is indicative of decreased likelihood of having FP; and change of a color in the heatmap from warmest to coldest is indicative of decreased likelihood of having FN. Table 14 presents accuracy parameters for the intermediate heatmap areas. Based on the data, a monotonic pattern is present for each accuracy parameter indicating that the color breaks of the heatmap is associated with increased/decreased likelihood of FN and FP.

Table 14: Descriptive Statistics for Localization Accuracy Parameters for All Heatmap Colors

Threshold	Sensitivity (%)			NPV (%)			Specificity (%)			PPV (%)		
	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N
0.25	98.7	2.2	31	99.8	0.5	40	91	10.1	40	31.4	27.4	40
0.296875	97.7	4.1	31	99.7	0.8	40	92.8	9.0	40	35.3	28.8	40
0.34375	97	4.7	31	99.6	1.2	40	94.1	7.8	40	38.9	30.2	40
0.390625	95.6	6.8	31	99.4	1.5	40	95.2	6.8	40	42.7	31.5	40
0.4375	94.2	9.0	31	99.3	1.9	40	96.1	5.8	40	46.4	32.6	40
0.484375	92.3	11.3	31	99.1	2.4	40	96.8	4.9	40	51.6	33.3	39
0.53125	90.2	13.4	31	98.9	2.8	40	97.4	4.1	40	55.4	34.4	39
0.578125	88.4	14.6	31	98.6	3.4	40	98	3.4	40	59.4	35.3	39
0.625	85.7	15.8	31	98.3	3.9	40	98.4	2.7	40	66.4	34.1	37
0.671875	82.6	17.0	31	98.0	4.5	40	98.9	2.1	40	69.8	34.9	37
0.71875	78.9	19.2	31	97.6	5.2	40	99.2	1.6	40	73.1	35.9	37
0.765625	74.4	20.9	31	97.0	5.9	40	99.5	1.1	40	82.1	29.8	34
0.8125	67.1	22.1	31	96.4	6.8	40	99.7	0.7	40	88.1	26.8	33
0.859375	56.9	23.9	31	95.5	8.1	40	99.8	0.4	40	94.9	18.0	31
0.90625	43.3	25	31	94.4	9.7	40	99.9	0.2	40	99.1	2.3	29
0.953125	25.0	21.8	31	92.8	11.8	40	100	0	40	99.6	1.0	27

3. Localization Precision

For localization precision analysis, sensitivity, specificity, PPV and NPV were calculated for each slide and run (defined by a combination of Run / Scanner / Operator). Repeatability and reproducibility of the localization were assessed via accuracy parameters obtained through different runs (Run/Scanner/Operator): sensitivity, specificity, NPV and PPV.

To evaluate localization precision 14 cancer slides and 4 benign slides (18 in total) were selected from the slides annotated for the localization accuracy study. Each of the 18 slides was repeatedly scanned, 5 times in total: 3 repeated assessments on Scanner 1 by Operator 1, one assessment on Scanner 2 by Operator 2 and one assessment on Scanner 3 by Operator 3. Localization precision was assessed through sensitivity and specificity obtained through repeated runs for each one of the 16 color-thresholds versus GT, as defined above.
The endpoints for the localization study are defined at the *level of pixel* within a slide.

(i) Repeatability Analysis: Table 15 presents Repeatability accuracy parameters – sensitivity, and NPV for each run for the entire heatmap area (determined by the lowest threshold of 0.25) and specificity and NPV for warmest heatmap area (determined by threshold >0.95).

Table 15: Repeatability: Within-Scanner Localization Precision

Analysis	Scanner/Operator, Run	Sensitivity	Negative Predictive Value
Entire Heatmap Area	Scanner 1/Operator 1 Run 1	99.1%	99.8%
	Scanner 1/Operator 1 Run 2	99.1%	99.7%
	Scanner 1/Operator 1 Run 3	98.6%	99.7%
	Overall Average	98.9%	99.7%
Warmest Heatmap Area	**Scanner/Operator, Run**	**Specificity**	**Positive Predictive Value**
	Scanner 1/Operator 1 Run 1	100%	99.7%
	Scanner 1/Operator 1 Run 2	100%	99.2%
	Scanner 1/Operator 1Run 3	100%	99.6%
	Overall Average	100%	99.5%

(ii) Reproducibility Analysis: Table 16 presents Reproducibility accuracy parameters for each Operator/Scanner – sensitivity, and NPV for the entire heatmap area (determined by the lowest threshold of 0.25) and specificity and NPV for the warmest heatmap area (determined by threshold >0.95).

Table 16: Reproducibility: Between-Scanner/Operator Localization Precision

Analysis	Scanner/Operator, Run	Sensitivity	Negative Predictive Value
Entire Heatmap Area	Scanner1/Operator 1 Run 1	99.1%	99.8%
	Scanner 2/Operator 2 Run 1	98.6%	99.7%
	Scanner 3/Operator 3 Run 1	98.9%	99.7%
	Overall Average	98.9%	99.7%
Warmest Heatmap Area	**Scanner/Operator, Run**	**Specificity**	**Positive Predictive Value**
	Scanner 1/Operator 1 Run 1	100%	99.7%
	Scanner 2/Operator 2 Run 1	100%	99.5%
	Scanner 3/Operator 3 Run 1	100%	99.6%
	Overall Average	100%	99.6%

B. Clinical Studies

Two clinical studies were conducted to assess the performance of the Galen Second Read as follows:

(i) Clinical Performance Study of The Galen™ Prostate AI-Powered Solution in Identifying Missed Cancers in Prostate Biopsies Previously Diagnosed as Benign (AIDER-2)

The study objective was to assess the performance of the Galen Second Read in identifying prostatic adenocarcinoma cases (subjects) initially missed by SoC method in PCNBs. This study was performed with retrospectively collected samples and conducted at three sites: two sites in the US and one OUS site. Three hundred forty-seven (347) cases were enrolled in the study. Characteristics of the study cohort are presented in Table 17.

Table 17: Distribution of Cases Characteristics

	Number of cases	Number of slides	Age of subjects (years): Mean (SD)
Site 1	100	861	66.6 (7.8)
Site 2	99	3686	64.5 (6.4)
Site 3	148	1395	66.8 (8.4)
Combined	347	5942	66.1 (7.7)

The device analyzed scanned WSIs of H&E from 347 cases who were initially diagnosed as benign based on the PCNBs. The slides were scanned with a Philips UFS scanner at 40x magnification and WSIs were then processed by Galen Second Read to provide a "Flag" (Positive) or "No Flag" (Negative). In the study, GT determination for a slide and a case (subject) was performed as follows: GT determination for a slide was performed by two independent expert pathologists; for slides where the pathologists disagreed, a third independent expert pathologist reviewed the slide and the majority rule determined the GT for the slide. Slides with prostatic AdC or other cancer and ASAP were considered GT positive slides.

The case was considered as "Flag" by the device if at least one slide from the case has "Flag" (Positive) and the case is considered as "No Flag" (Negative) by the device if all slides from the case has "No Flag".

Out of the total of 347 cases (5,942 slides) that were initially diagnosed as benign, the device Flagged 202 cases (573 slides) and 145 cases as No Flagged. All slides with Flag by the device from 202 cases were sent to GT determination. Out of 202 cases, 46 cases were confirmed as positive GT (ASAP or cancer) after review by expert pathologists.

For estimation of sensitivity and specificity of the device on a slide-level and on a case-level, the following randomly selected cases were sent to GT determination:
14 cases out of 46 Flagged cases with positive GT determination for the Flagged slides;
46 cases out of 156 Flagged cases with negative GT determination for the Flagged slides;
71 cases out of 145 No Flagged cases.
If the case was selected for GT determination, all slides from the case were sent to GT determination.

Device performance (sensitivity and specificity) is provided at the level of a slide and at the level of a case.

For slide-level analysis, the definitions of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) are as follows:

	Definition
TP slide	Slide is GT positive and Flagged by device
FP slide	Slide is GT negative and Flagged by device
FN slide	Slide is GT positive and Not Flagged by device
TN slide	Slide is GT negative and Not Flagged by device

For case-level analysis, the definitions of TP, FP, TN, and FN are as follows:

	Definition
TP case	Case has only one GT positive slide and this slide is Flagged by the device
	Case has more than one GT positive slide and at least one of these GT positive slides is Flagged by device
FP case	Case has all slides GT Negative and at least one slide is Flagged by the device
FN case	Case has one or multiple GT positive slides and all these GT positive slides are Not Flagged (missed) by the device
TN case	Case has all slides GT negative and all slides are not Flagged by the device

After GT determination of the slides for the randomly selected cases (subjects), an estimation of sensitivity and specificity was adjusted for a verification bias. In addition, two-sided 95% confidence intervals were calculated.

Slide-Level Performance Results: The analysis included 5,941 slides, of which 2,555 were verified. Table 18 presents the distribution of the Flag and No Flag slides versus GT status.

Table 18: Distribution of Slide Results by Galen Second Read and Ground Truth

	GT Positive	GT Negative	Not Verified	Total
Flag	81	492	0	573
No Flag	7	1975	3386	5368
Total	88	2467	3386	5941*

*1 slide did not have GT determination

Sensitivity and specificity on the slide-level adjusted for verification bias is presented in Table 19 below:

Table 19: Slide-Level Device Performance

Parameter	Estimate	95% CI
Sensitivity	81.0%	(69.2%; 92.9%)
Specificity	91.6%	(90.9%; 92.3%)

Positive and Negative Predictive values were as follows:
PPV=14.1%, NPV=99.6% at 1.7% prevalence of GT positive slides.

Case-Level Performance Results: In the analysis, the device provided Flag for 202 cases. For 46 cases of 202, positive slide(s) flagged by the device was confirmed by GT. Thus, 46 cases

were classified as TP. In the verification analysis, 46 randomly selected cases of 156 were verified for GT. Of these cases, 2 cases were found positive by GT. However, the GT positive slides were not the same as the slides flagged by the device; these two subjects had both FP, FN and TN slides. Per the definitions above, these subjects were considered in the analysis as FN. The unverified 110 cases out of 156 Flagged by the device were considered as "unverified". Out of 71 subjects that were not Flagged by the device and sent for verification, 2 cases were found to be GT positive. Table 20 presents the distribution of the Flag and No Flag cases versus GT status.

Table 20: Distribution of Case Results by Galen Second Read and Ground Truth

	GT Positive	GT Negative	Not Verified	Total
Flag	46	44	110	200
No Flag	4	69	74	147
Total	50	113	184	347

Sensitivity and specificity on the case-level adjusted for verification bias is presented in Table 21 below:

Table 21: Case-Level Device Performance

Parameter	Estimate	95% CI
Sensitivity	80.8%	(74.1%; 87.6%)
Specificity	46.9%	(39.5%; 54.3%)

Positive and Negative Predictive values were also calculated which are as follows: PPV=23.0%, NPV=92.6% at 16.4% prevalence of GT positive cases.

- The study demonstrated that sensitivity of the Galen Second Read is higher than the sensitivity of the SoC read (sensitivity of SoC was 0% because all the slides were diagnosed initially as benign by SoC), both at the slide-level and case-level.
- In the study there was a decrease in specificity of the Galen Second Read compared to the specificity of SoC (specificity of SoC was 100% because all the slides were diagnosed initially as benign by SoC). However, this can be managed by mitigation measures such as use of additional stains to confirm if the slide/case is positive.

(ii) Reader Study Comparing Pathologists' Diagnoses When Supported by the Galen™ Prostate AI-Powered Solution Versus the Standard of Care of Prostate Core Needle Biopsies (AIDER-1)

The objective of the study was to evaluate a difference in performances of a pathologist supported by the Galen Second Read vs GT result and a pathologist Standard of Care vs GT result in a set of slides from a population of subjects who have undergone a prostate core needle biopsy.

The study included retrospectively collected samples from subjects who had undergone a prostate core needle biopsy. The study dataset was composed of slides enrolled at 3 US clinical pathology laboratories (2 academic medical centers and 1 hospital), and 1 OUS

clinical pathology reference laboratory. The cancer slides included all the Gleason grade groups and the distribution were as follows: GG1-39.9%, GG2-19.9%, GG3-11.6, GG4-5.8%, and GG5-7.6%. The study cohort was enriched with challenging tumors as follows: ASAP - 4.2%, very small tumors <0.5 mm – 23.5%, and tumors 0.5-3 mm - 16.1%, slides with challenging histological appearance (i.e., atrophic pattern adenocarcinoma, pseudo hyperplastic adenocarcinoma, foamy gland adenocarcinoma and rare tumors (e.g., neuroendocrine tumors/carcinomas, urothelial carcinomas, squamous cell carcinomas) - 5.3%. Table 22 presents the baseline characteristics of the 396 cancer subjects and 376 benign subjects.

Table 22: Subject Baseline Characteristics

	Cancer	**Benign**
Characteristic	**(N=396)**	**(N=376)**
Age (years), Mean (SD)	67.6 (8.3)	66.1 (7.2)
Case Category (Original Diagnosis)		
Cancer[a]	353 (89.1%)	10 (2.7%)
Cancer Tumor size		
≤ 0.5 mm	84 (21.2%)	5 (1.3%)
> 0.5 mm	268 (67.7%)	5 (1.3%)
Not reported	1 (0.3%)	0
ASAP	11 (2.8%)	5 (1.3%)
ASAP Tumor size		
≤ 0.5 mm	11 (2.8%)	5 (1.3%)
> 0.5 mm	0	0
Benign	32 (8.1%)	361 (96.0%)
High-Grade PIN Present[b]	120 (30.3%)	39 (10.4%)
Race[c]		
Asian-Far East/Indian Subcontinent	4 (1.0%)	5 (1.3%)
Black or African American	41 (10.4%)	21 (5.6%)
Native Hawaiian or Pacific Islander	2 (0.5%)	0
White	229 (57.8%)	232 (61.7%)
Unknown	121 (30.6%)	118 (31.4%)
Gleason Grade[d]		
Grade Group 1	158 (39.9%)	1 (0.3%)
Grade Group 2	79 (19.9%)	0
Grade Group 3	46 (11.6%)	0
Grade Group 4	23 (5.8%)	0
Grade Group 5	30 (7.6%)	0
Not reported	60 (15.2%)	375 (99.7%)

[a] Cancer cases included Acinar adenocarcinoma (AdC), either alone or with another subtype, mainly atrophic pattern or pseudohyperplastic pattern AdC and others (foamy gland AdC, AdC with intraductal carcinoma, signet ring pattern, oncocytic pattern, small cell carcinoma, focal mucinous features AdC, neuroendocrine carcinoma).

[b] According to GT, where at least one ground truth pathologists diagnosed HG-PIN present.

[c] For the subset of "Cancer" subjects, all numbers reported for Race sum up to 397, and not 396, due to one subject reporting two races - Black or African American and White.
[d] According to GT, where when Gleason grade group differed between the ground truth pathologists, it was defined based on the most experienced pathologist.

AIDER-1 study included 4 sites with 3 eligible pathologists in each site, for a total of 12 pathologists (10 general anatomic surgical pathologists and 2 genitourinary (GU) subspecialists). Nine pathologists were US-board certified anatomic/surgical pathologists, while two (2) of these were GU sub-specialized; three (3) pathologists were board certified in OUS. The study included 3 eligible pathologists in each site, for a total of 12 eligible pathologists, as summarized in Table 23 below.

Table 23: Summary of Pathologists Characteristics

Characteristic		N	%
Age Group	< 35 years	2	16.7
	35 - 45 years	6	50.0
	45 - 55 years	1	8.3
	> 55 years	3	25.0
	All	12	100.0
Years of Experience	< 5 years	4	33.3
	5 - 10 years	3	25.0
	10 - 20 years	2	16.7
	20 - 30 years	1	8.3
	> 30 years	2	16.7
	All	12	100.0
Number of Cases Diagnosed per Day	< 10	0	0.0
	10 - 20	3	25.0
	> 20	9	75.0
	All	12	100.0
Number of Diagnostic Hours per Day	< 3 hours	1	8.3
	3 - 6 hours	3	25.0
	> 6 hours	8	66.7
	All	12	100.0
Professional Setting	Academic Medical Centre lab	6	50.0
	Private lab	4	33.3
	Hospital lab	2	16.7
	All	12	100.0

The study consisted of two arms:

- **Arm A:** a standard of care (SoC) arm, in which slides were read digitally using the routine lab practice, i.e. on Philips Image Management System and Viewer and associated computer monitors [PP27QHD (Philips) or MDPC-8127 (Barco NV)].
- **Arm B:** a Galen Second Read workflow arm, in which after reading slides digitally, the pathologist reviewed images flagged by the Galen Second Read and determined the final diagnosis.

All cases and associated slides were read in both Arms of the study (Arm A and Arm B). The cases were assigned in random order to each of the study pathologists. Each pathologist read all

of their cases/slides, i.e. all the site cases, under both Arms, with a washout period of two weeks between Arms to minimize recall bias.

To minimize order bias, each pathologist reviewed half of their cases first in Arm A, and subsequently (after washout period) in Arm B, and the other half first in Arm B, and subsequently (after washout period) in Arm A. This study design enabled providing a SoC read and a read using the Galen Prostate workflow by the same pathologist for each slide. During the second session the pathologists were blinded to their results provided during the first session. For both arms, the pathologists only reviewed H&E slides and reached a positive or negative diagnosis. Study pathologists had the option to specify if they would have deferred the diagnosis to immunohistochemistry (IHC) stains or 2nd opinion for the respective slide.

Ground Truth Pathologists Characteristics: Two (2) pathologists, who were US-board certified anatomic/surgical pathology and GU sub-specialized, reviewed all the study slides, and determined each WSI as either positive or negative. The GT pathologists first reviewed only H&E slides, if requested and available, they were provided with IHCs. Discrepant slides were reviewed by a third US-board certified anatomic/surgical pathology GU sub-specialized pathologist.

Accountability: Overall, there were 798 enrolled cases for the study, out of which 26 cases were fully excluded, because 25 slides was found as out of focus (OOF) / not readable by the pathologists/GTs or because 1 slide was erroneously read together with corresponding IHC. Thus, 772 cases/slides were included in both analysis sets: 376 negative cases and 396 positive cases. In each arm, each case was read 3 times, once per pathologist. Therefore, the total number of negative reads was anticipated to be 376×3=1128 and the number of positive reads was anticipated to be 396×3=1188. One negative case in site 3 was out-of-focus for only pathologists 1 (per the pathologist professional opinion), therefore the total number of negative reads was 1127. One positive case in site 3 was erroneously read with IHC staining for pathologist 3, therefore the total number of positive reads was 1187. Thus, a total of 2314 reads were included in both analysis sets.

Each pathologist's result per slide in comparison with the GT (positive or benign) is presented in Table 24 and Table 25 below.

Table 24: Pathologist Results: GT Positive

GT=Positive						
		Without Galen				
Site	**Pathologist**			Positive	Benign	Total
1 (98 cases)	**1**	With Galen	Positive	93	2	95
			Benign	0	3	3
	2	With Galen	Positive	91	4	95
			Benign	0	3	3
	3	With Galen	Positive	86	2	88
			Benign	0	10	10
2	**4**	With Galen	Positive	101	0	101
			Benign	0	6	6
	5	With Galen	Positive	102	0	102

(107 cases)			Benign	0	5	5
	6	With Galen	Positive	90	6	96
			Benign	0	11	11
3 (86 cases)	**7**	With Galen	Positive	73	10	83
			Benign	0	3	3
	8	With Galen	Positive	73	5	78
			Benign	0	8	8
	9	With Galen	Positive	72	1	73
			Bening	0	12	12
4 (105 cases)	**10**	With Galen	Positive	101	2	103
			Benign	0	2	2
	11	With Galen	Positive	96	1	97
			Benign	0	8	8
	12	With Galen	Positive	96	8	104
			Benign	0	1	1
Combined	**1-12**	With Galen	Positive	1074	41	1115
			Benign	0	72	72

Table 25: Pathologist Results: GT Benign

GT=Benign						
			Without Galen			
Site	**Pathologist**			Positive	Benign	Total
1 (98 cases)	**1**	With Galen	Positive	8	3	11
			Benign	0	87	87
	2	With Galen	Positive	23	4	27
			Benign	0	71	71
	3	With Galen	Positive	8	2	10
			Benign	0	88	88
2 (93 cases)	**4**	With Galen	Positive	10	0	10
			Benign	0	83	83
	5	With Galen	Positive	8	1	9
			Benign	0	84	84
	6	With Galen	Positive	0	0	0
			Benign	0	93	93
3 (92 cases)	**7**	With Galen	Positive	2	8	10
			Benign	0	81	81
	8	With Galen	Positive	10	3	13
			Benign	0	79	79
	9	With Galen	Positive	3	1	4
			Benign	0	88	88
4 (93 cases)	**10**	With Galen	Positive	13	3	16
			Benignn	0	77	77
	11	With Galen	Positive	8	0	8

			Benign	0	85	85
	12	With Galen	Positive	7	11	18
			Benign	0	75	75
Combined	**1-12**	With Galen	Positive	100	36	136
			Benign	0	991	991

Sensitivity and specificity for each pathologist were estimated and results are presented in Table 26 and Figure 4 below.

Table 26. Sensitivity and Specificity Presented by Pathologist

	Sensitivity			Specificity		
Pathologist	**With Galen**	**SoC**	**Difference**	**With Galen**	**SoC**	**Difference**
1	96.9%	94.9%	2.0%	88.8%	91.8%	-3.1%
2	96.9%	92.9%	4.1%	72.4%	76.5%	-4.1%
3	89.8%	87.8%	2.0%	89.8%	91.8%	-2.0%
4	94.4%	94.4%	0.0%	89.2%	89.2%	0.0%
5	95.3%	95.3%	0.0%	90.3%	91.4%	-1.1%
6	89.7%	84.1%	5.6%	100%	100%	0.0%
7	96.5%	84.9%	11.6%	89.0%	97.8%	-8.8%
8	90.7%	84.9%	5.8%	85.9%	89.1%	-3.3%
9	85.9%	84.7%	1.2%	95.7%	96.7%	-1.1%
10	98.1%	96.2%	1.9%	82.8%	86.0%	-3.2%
11	92.4%	91.4%	1.0%	91.4%	91.4%	0.0%
12	99.0%	91.4%	7.6%	80.6%	92.5%	-11.8%
Combined	93.9%	90.5%	3.5%	87.9%	91.1%	-3.2%

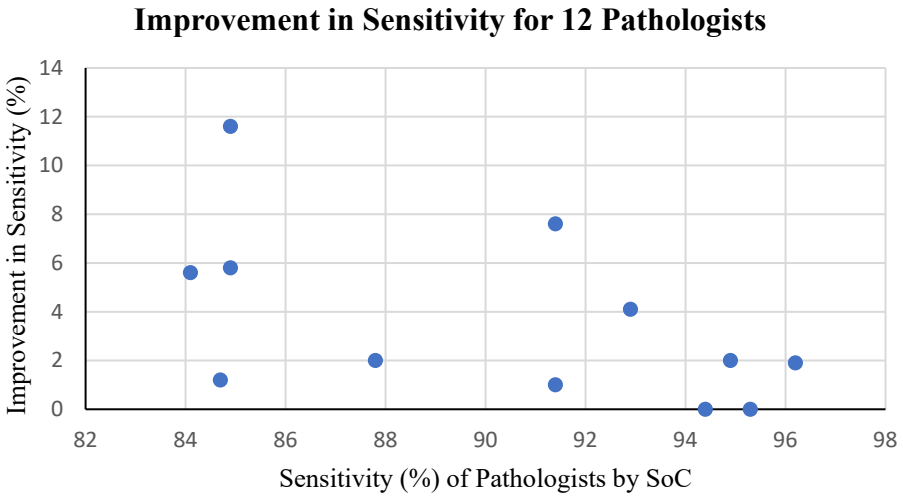

Figure 4. Improvement in Sensitivity for Pathologists

Sensitivity and specificity for the combined data were calculated and presented in Table 27.

Table 27: Sensitivity and Specificity of Pathologists with Galen Device vs SoC

	Sensitivity		
	Estimate	(n/N)	95%CI*
With Galen	93.9%	(1115/1187)	(92.2%; 95.8%)
SoC	90.5%	(1074/1187)	(88.5%; 92.6%)
Difference	3.5%	(41/1187)	(2.3%; 4.5%)
	Specificity		
With Galen	87.9%	(991/1127)	(85.8%; 90.4%)
SoC	91.1%	(1027/1127)	(89.3%; 93.2%)
Difference	-3.2%	(-36/1127)	(-4.3%; -1.9%)

* Confidence intervals are calculated by bootstrap

The AIDER-1 clinical study demonstrated a statistically significant improvement in sensitivity of 3.5% with 95%CI: (2.3%; 4.5%) and statistically significant decrease in specificity of –3.2% with 95%CI: (-4.3%; -1.9%).

Sensitivity and specificity for the slides originally assessed by a pathologist as benign (the intended use population of the device) are also calculated and presented in the Table 28 below.

Table 28: Sensitivity and Specificity for the Slides Originally Assessed as Benign vs GT

	Sensitivity		Specificity	
Pathologist	**With Galen**	**(n/N)**	**With Galen**	**(n/N)**
1	40.0%	(2/5)	96.7%	(87/90)
2	57.1%	(4/7)	94.7%	(71/75)
3	16.7%	(2/12)	97.8%	(88/90)
4	0.0%	(0/6)	100%	(83/83)
5	0.0%	(0/5)	98.8%	(84/85)
6	35.3%	(6/17)	100%	(93/93)
7	76.9%	(10/13)	91.0%	(81/89)
8	38.5%	(5/13)	96.3%	(79/82)
9	7.7%	(1/13)	98.9%	(88/89)
10	50.0%	(2/4)	96.3%	(77/80)
11	11.1%	(1/9)	100%	(85/85)
12	88.9%	(8/9)	87.2%	(75/86)
Combined	36.3% 95%CI*: (28.0%; 45.5%)	(41/113)	96.5% 95%CI*: (95.2%; 97.5%)	(991/1027)

* Confidence intervals are calculated by the Wilson score method (CLSI EP12-Ed3)

- The AIDER-1 study demonstrated that sensitivity of the pathologists using the Galen device for the cases/slides which were initially diagnosed as benign was 36.3% with 95% CI: (28.0%; 45.5%) (sensitivity of SoC was 0% because all the slides were diagnosed initially as benign by SoC).
- Specificity of the pathologist using the Galen device for the cases/slides which were initially diagnosed as benign was 96.5% with 95% CI: (95.2%; 97.5%) (specificity of

SoC was 100%). The decrease in specificity can be managed by mitigation measures such as use of additional stains to confirm if the slide/case is positive.

2. Linearity:
   Not applicable

3. Analytical Specificity/Interference:
   Not applicable

4. Accuracy (Instrument):
   See clinical study section above.

5. Carry-Over:
   Not applicable

B Other Supportive Instrument Performance Characteristics Data:

N/A

VIII Proposed Labeling:

The labeling is sufficient, and it satisfies the requirements of 21 CFR Parts 801 and 809, as applicable, and the special controls for this device type under 21 CFR 864.3750.

IX Conclusion:

The submitted information in this premarket notification is complete and supports a substantial equivalence decision.