



Published in final edited form as:

Mod Pathol. 2022 January ; 35(1): 52–59. doi:10.1038/s41379-021-00908-5.

Digital Validation of Breast Biomarkers (ER, PR, AR and HER2) in Cytology Specimens Using Three Different Scanners

Abeer M. Salama^{1,a}, Matthew G. Hanna^{1,a}, Dilip Giri¹, Brie Kezlarian¹, Marc-Henri Jean¹, Oscar Lin¹, Christina Vallejo¹, Edi Brogi¹, Marcia Edelweiss¹

¹Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

Abstract

Progression in digital pathology has yielded new opportunities for a remote work environment. We evaluated the utility of digital review of breast cancer immunohistochemical prognostic markers (IHC) using whole slide images (WSI) from formalin fixed paraffin embedded (FFPE) cytology cell block specimens (CB) using three different scanners.

CB from 20 patients with breast cancer diagnosis and available IHC were included. Glass slides including 20 Hematoxylin and eosin (H&E), 20 Estrogen Receptor (ER), 20 Progesterone Receptor (PR), 16 Androgen Receptor (AR), and 20 Human Epidermal Growth Factor Receptor 2 (HER2) were scanned on 3 different scanners. Four breast pathologists reviewed the WSI and recorded their semi-quantitative scoring for each marker. Kappa concordance was defined as complete agreement between glass/digital pairs. Discordances between microscopic and digital reads were classified as a major when a clinically relevant change was seen. Minor discordances were defined as differences in scoring percentages/staining pattern that would not have resulted in a clinical implication. Scanner precision was tabulated according to the success rate of each scan on all three scanners.

In total, we had 228 paired glass/digital IHC reads on all 3 scanners. There was strong concordance $\kappa \geq 0.85$ for all pathologists when comparing paired microscopic/digital reads. Strong concordance ($\kappa \geq 0.86$) was also seen when comparing reads between scanners.

Twenty-three percent of the WSI required rescanning due to barcode detection failures, 14% due to tissue detection failures, and 2% due to focus issues. Scanner 1 had the best average precision of 92%. HER2 IHC had the lowest intra-scanner precision (64%) among all stains.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: <https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms>

Corresponding author: Marcia Edelweiss, MD, Department of Pathology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York, 10065, edelweim@mskcc.org.

^aAuthors contributed equally.

Author contribution Statement: ME, MGH, ASA: Concept and design, data collection, data analysis/interpretation, manuscript drafting/revision. DG, BEK, CV, EB, OL: Data interpretation and manuscript revision, BEK: Data collection, manuscript revision, MJ: Data collection.

Conflict of Interest Statement: No competing financial interests exist for all contributory authors.

Unrelated: MGH is a consultant for PaigeAI and advisor for PathPresenter

Unrelated: MGH is a consultant for PaigeAI

Ethics Approval:

This study was performed with the approval of the Institutional Review Board of Memorial Sloan-Kettering Cancer Center (New York, NY)

This study is the first to address the utility of WSI in breast cancer IHC in CB and to validate its reporting using 3 different scanners. Digital images are reliable for breast IHC assessment in CB and offer similar reproducibility to microscope reads.

Keywords

ER; PR; HER2; Whole slide imaging; validation; cytology; digital cytology; digital imaging; digital pathology; interpretation; whole-slide imaging

Introduction:

Breast cancer biomarkers, namely Estrogen Receptor (ER), Progesterone Receptor (PR), Androgen Receptor (AR) and Human Epidermal Growth Factor Receptor 2 (HER2) are crucial components of pathology reporting and a known prognostic factor to determine therapy for patients both in the primary setting or with recurrent or metastatic disease. Studies have shown a substantial survival benefit from targeted therapy especially against hormone receptors such as ER and/or oncogenic proteins such as HER2¹. These prognostic markers in breast cancer are routinely tested using both histopathology specimens and cytology preparations including cell blocks². Digital pathology systems have been developed and validated for routine histopathology diagnoses using digitized glass slides as whole slide images (WSI). Digital pathology systems have also been demonstrated to assist in rapid diagnostic consultation in breast care clinics³.

Progression in the field of digital pathology has yielded new opportunities for digital reporting of such biomarkers. Over a decade, WSI has become popular in seeking a second opinion in teleconsultation⁴, education⁵, and most importantly for primary diagnosis, especially in surgical pathology⁶⁻¹⁹. There is limited literature on the use of WSI in cytopathology for primary diagnosis in comparison to surgical pathology mainly due to the cytological specimens' characteristics, longer scan times and the need for higher scanning resolution^{20,21}. More recently, the adoption of WSI for primary diagnosis has increasingly become a reality despite some barriers to implementation, such as cost and workflow considerations. This increased adoption of WSI has resulted in a relatively few publications regarding the validation of WSI for cytopathology diagnostic use²²⁻²⁵.

Although the College of American Pathologists (CAP) published a formal guideline on validating WSI for primary diagnosis in 2013¹⁵, there is limited data to support the use of WSI in reporting immunohistochemistry markers in cytology specimens. The FDA has also recently approved WSI for primary diagnosis purposes. There are currently two WSI scanner devices that has been cleared for use in FFPE hematoxylin and eosin stained tissue specimens but they do not make claims on the use for cytology specimens^{26,27}

Few studies have attempted to show that WSI diagnosis is equivalent to light microscopy interpretation in breast cytology specimens^{28,29}, and to the best of our knowledge, none have validated breast biomarkers reporting in cytology specimens. Among the few studies that used cytology specimens, only a single whole slide scanner was used, which

raises questions about inter-instrument as well as inter-observer variability and diagnostic accuracy.

The advantages of WSI for surgical pathology can potentially be used in cytopathology specimens for interpretation of prognostic immunohistochemistry (IHC). Validation of WSI on multiple scanners is crucial to ensure that diagnostic performance based on digitized slides is non-inferior to that of glass slides and light microscopy. Among the few studies that used WSI in non-gyn cytology specimens²², none have validated the reporting of breast hormonal biomarkers in cytology specimens using three different scanners.

In this study, we aimed to evaluate the utility of the digital review of breast cancer biomarkers (ER, PR, AR, and HER2) and to validate their reporting in cytology specimens. We also compared concordance between digitally reviewed WSI to the conventional microscope results using three different whole slide scanners.

Material and Methods:

Case selection:

This study presents a validation of the digital pathology breast biomarkers in cytology specimens at a large academic tertiary cancer center in New York City after approval from the institutional review board. The validation encompasses digitization of fine needle aspiration and fluid specimens from patients with known breast cancer (recurrent or metastatic). Glass slides generated from formalin fixed paraffin embedded cell blocks included hematoxylin & eosin (H&E) stains and breast biomarker immunohistochemical stains (ER, PR, HER2 and AR). All methanol fixed specimens were excluded for the study.

Digitization of glass slides:

The whole slide imaging process included pre-analytic quality assurance of slide preparations, analytic process of glass slides digitization on each whole slide scanner, and post-analytic quality assurance of the generated WSI for digital artifacts (e.g. out of focus, stitching, banding). The whole slide scanning process included three different vendor whole slide scanners acquiring WSI from a glass slide at high resolution (~0.25 um/pixel). Each whole slide scanner uses an objective lens paired with an image acquisition sensor and stitches all captured images together to form a single digital file that can be navigated similarly to a glass slide on a microscope, in a whole slide image viewer. All glass slides were scanned in a single z-plane.

All glass slides used for validation ER, PR, AR and HER2 with their corresponding H&E cell block slides were scanned using three different scanners at x40 equivalent magnification (~0.25 um/pixel). The list of scanners used is indicated below:

- Leica Aperio GT450 (Leica Biosystems, Buffalo Grove, IL, USA)
- Panoramic 1000 (3DHistech, Budapest, Hungary)
- Ultra Fast Scanner (Philips Health, Amsterdam, Netherlands)

Whole slide scanner precision:

Each glass slide was scanned in triplicate across all scanners included in this study. This was conducted to evaluate the intra-scanner accuracy of digitization related to barcode detection, tissue detection, and image quality (e.g. blur, digital artifacts). A successful scan was defined as correct barcode decoding, complete capture of tissue on the glass slide, and digital slides free of image quality defects. Each glass slide was scanned in triplicate and tabulated according to the success rate of each scan for all three scanners.

Digital Slide Review and Scoring:

After whole slide scanning on each of the three different scanners at x40 equivalent resolution, slides were de-identified and were distributed to the study pathologists using an internally developed WSI viewer application³⁰.

A total of 96 glass slides were scanned; including 20 cell blocks H&E; 20, ER; 20, PR; 16, AR; and 20, HER2 stained immunohistochemistry slides three times on each scanner. All immunohistochemical stained slides had routine control tissue placement on each slide for pathologist reference. Cases were randomized and distributed to the study pathologists for digital cytology reporting on cases from the three different scanners. Quantification of ER, PR, AR and HER2 were assessed based on the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) guidelines^{31,32} by all four pathologists, blinded to the reported semi-quantitative IHC results; after a washout period of at least 6 months. All participating pathologists (n=4) were breast pathologists with cytology experience and at least 3-year experience in using WSI for secondary diagnostic use cases (tumor boards, reviewing archived scanned slides, etc.). Pathologist used HP Z24n 24-inch 1920×1200 resolution monitors. The monitors are not color calibrated by default and no color calibration adjustment was done. The review was conducted in 2 phases, first, each pathologist randomly reviewed a total of 15 digital cases with their corresponding markers, 5 cases from each scanner on 3 different occasions, at least 14 days apart.

The Second phase was conducted 6-months after completion of phase one. Two out of the four pathologists (Pathologist B and Pathologist C) reviewed all 20 digital cases with their corresponding IHC markers on each scanner to test inter-instrument concordance rate when comparing between multiple scanners and investigate inter-observer variability. Figure 1.

A referee pathologist not participating in reporting of the study reads verified all cases included in the study by ensuring pathologist case assignment, whole slide image slide quality, and respective rescanning for glass slides which failed scanning, noting the reason for each failure. The referee also annotated the readings of the biomarkers and included the percentage for the positive nuclear staining for ER, PR, AR, and membranous staining for HER2 staining for analysis.

Statistical Analysis:

Interobserver variability was calculated using unweighted Cohen's K, whereby a value of 0.01 to 0.20 indicated slight concordance, 0.21 to 0.40 fair, 0.41 to 0.60 moderate,

0.61 to 0.80 substantial, and 0.81 to 0.99 strong concordance. Statistical significance was established at $p < 0.05$.

The percentage agreement, 95% CI and the level of significance (using Fisher's exact test) was calculated using IBM SPSS Statistics for Windows, version 26 (IBM Corp., Armonk, N.Y., USA). This study was performed with the approval of the Institutional Review Board of Memorial Sloan-Kettering Cancer Center (New York, NY).

Concordance:

- Concordance analysis was performed in two steps:
 - Kappa concordance was initially calculated for all pathologist, defined as complete agreement between the first original signed out biomarkers IHC stains and the digital reads. (i.e., positive vs negative for ER, PR, AR, HER2 or equivocal for HER2 staining).
 - We also conducted a second analysis for ER, PR and AR to test within group concordance for pathologist B and C after dividing the IHC score readings into groups: Group 1=0%, Group 2=1%–10%, Group 3=11–75% and Groups 4 >75 % nuclear staining.

Discordance:

- Minor discordances were defined as differences that will not result in clinical or prognostic implications.
- Major discordances between microscopic and digital reads for all pathologists were classified when a clinically relevant change was seen (i.e. positive vs negative for ER, PR, AR, HER2 or equivocal versus negative for HER2 staining) between paired samples. Major discordances for pathologist B and C in the second phase of analysis were classified as differences in biomarkers scores that reflected different groups.

Final diagnosis for the discordant cases were made by two pathologists based on consensus (ME, OL).

Results:

Study Cohort:

Following random selection, the study cohort included 20 matched cell block H&E, ER, PR, and HER2 samples from 20 patients, with AR available in 16 cases. All samples were from metastatic disease (n=19) or were locally recurrent (n=1) in female patients. Among the study samples 70% (n=14), 35% (n=7), 75% (n=12), were reported positive for ER, PR, and AR, respectively by the original reported brightfield microscope interpretation. Fifteen cases were HER2 negative (0 or 1+), One case was reported HER2 positive (3+), and four cases as HER2 equivocal (2+) by immunohistochemistry. [supplementary table 1] FISH results were available on 3 out of the 4 equivocal HER2 cases of which one case showed amplification. Forty-five percent (9/20) of the study samples were pleural fluid specimens, 40% (8/20)

were from lymph nodes, 10% were from chest wall (2/9) and 5% (1/20) were from breast local recurrences. Clinical pathological characteristics are shown in Table 1.

The study set consisted of 96 glass slides scanned on the three scanners (total=288 scanned images). Quality assurance of WSIs by technicians or the referee pathologist prior to initiation of digital readings revealed 65 slides (23%) required rescanning due to barcode detection failures, 26 (9%) due to tissue detection failure, 13 (5%) due to partial tissue detection, and 5 (2%) due to out of focus areas in each scan. [Supplemental table 2]. Since Scanner 1 does not support manual focus plane depth adjustment, manual adjustments for out-of-focus and tissue failure were only done on Scanner 2 and Scanner 3. The slides for scanner 1 were rescanned without any adjustment and was at the scanner's ability for focus/tissue detection. After remediation of the glass slides and scanner adjustments, failures decreased to a total of 3 images (one tissue detection failure for PR on scanner 1 and two out-of-focus for HER-2 images on scanner 2 and 3).

The first-time successful scan rate and average number of rescans to successfully scan each slide are shown in Table 2.

We also separately tested for scanner precision results after scanning all slides from each case in triplicates on each scanner. Scanner 1 had the best average precision (92%) compared to scanner 2 and scanner 3 with 78% and 73%, respectively. When comparing precision between ER, PR, AR and HER2 staining, the HER2 immunohistochemical stains had the lowest intra-scanner precision at 64%.

Concordance:

All four pathologists successfully completed digital review of all WSI. Each pathologist randomly reviewed a total of 15 digital cases with their corresponding IHC biomarkers, 5 cases from each scanner on 3 different occasions. A total of 228 reads were performed using both WSI and glass slides.

There was strong concordance between all four pathologists' digital and the glass slide readings ($\kappa = 0.97, 0.85, 0.93$ and 0.90 for pathologist A, B, C and D; respectively (P -value < 0.0001), (Figure 3). Complete concordance between all study pathologists and the original sign-out diagnosis was achieved in 90% of ER, 80% of PR, 100% of AR and 95% of HER2 stains.

Inter-instrument concordance:

Six months after the initial scoring, two out of four pathologists (pathologist B and pathologist C) reviewed all digital cases on all three scanners. Concordance was similarly determined by comparing between multiple scanners to evaluate inter-observer variability on digital reads from different scanners.

There was a strong concordance ($\kappa = 0.97, 0.89,$ and 0.92 for pathologist B, and $\kappa = 0.92, 0.94,$ and 0.86 for Pathologist C) when comparing reads between scanners (scanner 1 vs scanner 2, scanner 1 vs scanner 3, and scanner 1 vs scanner 3; $P < 0.001$). (Figure 3)

Discordance:

Reassessment of the glass slide and the WSI for all discordant cases by 2 referee pathologists (ME, OL) revealed that overall discordance was seen in 2.3% (n=6) of glass/digital pairs; 1 for ER, 4 for PR, and 1 for HER2, of which 5 were of minor discordance without any clinical or prognostic implication. One ER and three PR stains from three different specimens were initially reported with rare nuclear stain positivity on the glass slide (5%, 1% and 1%; respectively) had a negative digital read. Another PR stain was reported negative on the initial glass read and was digitally reported positive (2%). All cases had very few tumor cells (<20 tumor cells) with weak nuclear staining shown in Figure 2.

One HER2 slide (0.3%) showed a major discordance. This HER2 stain that was initially reported equivocal (2+) on the glass read had a negative (1+) digital read. FISH results from this case performed at the time of the original reporting of IHC showed HER2 amplification. Table 3 details discordant cases.

Discussion:

Digital validation reporting of hormonal breast cancer markers has not been clearly established in the literature, especially in cytology specimens. In breast cancer patients, determination of prognosis and treatment strategies based on ER, PR, AR and HER2 status greatly depends on the accurate evaluation of overexpression by IHC and/or FISH.

Studies have attempted to use WSI to validate primary diagnosis in multiple surgical pathology specialty applications¹² including prostate³³, pediatric⁶, dermatopathology¹⁴, gastrointestinal^{16,19}, and gynecological pathology specimens³⁴. Krenacs et al. was among the first to address the potential use of digital imaging in breast cancer¹⁸. This was further followed by using WSI in primary breast cancer diagnosis^{17,28} and in reporting prognostic factors such as the Nottingham histology grading^{17,35}, PDL-1³⁶ and HER2 immunohistochemistry stains^{37,38} in histology specimens.

In surgical pathology, among the largest studies that used WSI to validate primary diagnosis in multiple organs; the concordance rate was 98% and 96% and the major discordance rate was 0.7% and 0.9% in 3017 and 1070 specimens, respectively^{39,40}. In breast pathology, the concordance and discordance rate followed a similar trend, 95–97% and 3%, respectively⁹.

However, comparing surgical pathology to cytology studies, the reported results were variable. In two recent meta-analysis studies, a surgical pathology, and a cytology study showed slightly different concordance rates between WSI and original diagnosis. Araujo et al in 2019 (included 13 surgical pathology studies) showed overall strong concordance of 87–98.3% with inter-observer κ coefficient 0.8–0.98. In contrast, Girolami et al. in 2020 (included 19 cytology studies), showed a very wide range of concordance and inter-observer κ coefficient, 14%–100% and 0.57–0.82 respectively between cytology smears only but not cell blocks. Even after correction for the difference in study size the mean percentage concordance and κ coefficient remained inferior to that reported in surgical pathology with 84% and 0.69, respectively²². This can be partially explained by the differences in the histological characteristics between surgical pathology and cytology specimen.

Focusing attention on the application of immunohistochemistry WSI for reporting prognostic breast markers, Campbell et al. used IHC (eg, AE1/AE3, P63, and E-cadherin) WSI to study the diagnostic reads in breast core needle biopsies⁹. Others have evaluated digitally reporting prognostic immunohistochemical factors such as PDL-1 and HER2 in breast cancers and showed similar concordance results. Two studies of digital HER2 and PDL-1 reporting showed substantially equivalent kappa co-efficient (0.72) and percent agreement ranging from 61% –92 in 180 and 79 cases; respectively^{36,41}, others have confirmed these findings and concluded there is non-inferiority for interpreting breast markers IHC by either glass slides or digital images³⁷. Not surprisingly, 90–97% concordance of glass/digital pairs for all pathologists seen in this study are comparable those published in the literature.

In our study, overall discordance was seen in 2.3% (n=6) of glass/digital pairs; of which one case (HER2) showed a major discordance. Among the 5 cases that had minor discordance, they all had very few tumor cells staining (1–5%). There are limited data on the overall benefit of endocrine therapies for patients with low level (1–10%) ER expression. The literature suggests tumors with such results are heterogeneous in both behavior and biology and may be more similar to ER-negative cancers³¹. For this reason, we considered those cases to be relatively concordant.

The only case that had a major discordance was a HER2, that was initially reported equivocal (2+) on the glass read and had a negative (1+) digital read. FISH studies performed at the time of the original IHC reporting showed HER2 amplification. Upon review of this case, the tumor cells had heterogenous HER2 membranous staining ranging from weak to strong, which might be subject to inter-observer variation regardless of the method used for scoring. HER2 scores on WSI were shown to be higher than glass slides in a previous study, possibly due to increased color contrast on WSI⁴², an issue that we did not face in this study since immunohistochemistry positive and negative controls were part of the scanned slide. Additionally, previous studies have addressed diagnostic concerns regarding color inconsistency of WSI between different scanners and within the same scanner on different occasions⁴³. These parameters may be important for future validation guidelines, since adjusting color or contrast/brightness might alter visibility and impact reporting of membranous or even nuclear staining.

In 2017, The FDA cleared the first device for using WSI for primary diagnosis²⁶. A systemic review showed that among the studies that mentioned the type of scanners used; Leica Aperio seemed to be the most used (37%) followed by Hamamatsu (21%), and Roche/Ventana (16%)²². Another study³⁹, used eight different scanners from five different manufactures and showed no significant discordances when rendering diagnosis using digital versus glass slides. As per laboratory accreditation guidelines⁴⁴, each digital pathology system requires its own studies for validation. In this single large center study, three different scanners from three different manufacturers were used, and indeed the concordance rate between scanners were analyzed after accounting for interobserver variability and showed excellent concordance between all three scanners.

Errors rates by first successful scan rate differed by scanner vendor. Each scanner has their own technical specifications, including tissue detection—where immunohistochemistry typically has lower contrast compared to hematoxylin and eosin-stained tissue. Scanners also require barcode decoding to ensure digital slides are viewable within the laboratory information system. Additionally, scanning glass slides may show digital artifacts such as out of focus regions on the digital slide that may necessitate rescan. Scanner 1 showed the highest performance related to first time successful scans, as well as intra-scanner precision. One PR stain persistently showed tissue detection failure on scanner 1; contributing reasons might be due to scant tissue and pale counterstain.

The HER2 immunostain showed the lowest first-time successful scan performance in relation to the other immunohistochemical stains. After scanner adjustments, two out-of-focus HER-2 images persisted. Taking a closer look at the images, we believed that staining heterogeneity and faint tissue causing low contrast might have contributed to focus issues. Understanding these parameters are important in evaluating scanners as well as validations. Implementing WSI in primary reporting of immunohistochemistry will refine pathology practice for prognostication where limited tissue is available particularly in subspecialties like breast cytopathology.

Digital validation will help Integrate pathology images to the clinical information that will eventually permit easily comparing prior and prospective patient's breast hormonal status to better understand the progress and aid in treatment modification for recurrent or metastatic breast cancer cases.

When comparing pathologists, digital biomarker quantification to the original glass slide microscope, respective kappa values were consistent even when comparing digital reads on various scanners. However, some difficulties were encountered in reporting biomarker stains in cases with low cellularity and heterogenous weak nuclear staining even after manual adjustment for multiple scanning at different focus plane depths.

Excluding methanol fixed cases is one of this study's major limitations. Another limitation was the time needed for image exploration. Examining WSI was perceived to take more time than evaluation by conventional microscope (although no formal timing was conducted in this study). Additional research is required to document minimum number of tumor cells required in cytology specimens for optimal WSI performance and IHC digital reporting. Such data are important for validation guidelines or protocols dedicated to cytology specimens to enhance our pathology daily practice.

In conclusion this study is the first to address the feasibility of WSI of breast biomarkers in cytology specimens and validate primary reporting using inter-instrument comparisons between three different scanners. Digital scanning is an acceptable method to report ER, PR, AR and HER2 quantification assessment for clinical decision-making and offers similar reproducibility to routine microscope reads. More studies are needed in cytology specimens to better understand discordances and compare concordance between different scanners, resolutions, and specimen preparations. This will ultimately help better refine WSI guidelines and standards dedicated to cytology subspecialty.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

Research reported in this publication was supported in part by the Cancer Center Support Grant of the National Institutes of Health/National Cancer Institute under award number P30CA008748

Data Availability Statement:

The data that support the findings of this study are available from the corresponding author on reasonable request.

REFERENCES

1. Nofech-Mozes S, Vella ET, Dhesy-Thind S, et al. Systematic review on hormone receptor testing in breast cancer. *Appl Immunohistochem Mol Morphol* 2012;20:214–63. [PubMed: 22505008]
2. Gorman BK, Kosarac O, Chakraborty S, Schwartz MR, Mody DR. Comparison of breast carcinoma prognostic/predictive biomarkers on cell blocks obtained by various methods: Cellient, formalin and thrombin. *Acta Cytol* 2012;56:289–96. [PubMed: 22555532]
3. López AM, Graham AR, Barker GP, et al. Virtual slide telepathology enables an innovative telehealth rapid breast care clinic. 2009;40:1082–91.
4. Rocha R, Vassallo J, Soares F, Miller K, Gobbi H. Digital slides: Present status of a tool for consultation, teaching, and quality control in pathology. *Pathology - Research and Practice* 2009;205:735–41.
5. Romer DJ, Suster S. Use of virtual microscopy for didactic live-audience presentation in anatomic pathology. 2003;7:67–72.
6. Al-Janabi S, Huisman A, Nikkels PGJ, Ten Kate FJW, Van Diest PJ. Whole slide images for primary diagnostics of pediatric pathology specimens: a feasibility study. 2013;66:218–23.
7. Araujo ALD, Arboleda LPA, Palmier NR, et al. The performance of digital microscopy for primary diagnosis in human pathology: A systematic review. *Virchows Arch* 2019;474:269–87. [PubMed: 30685784]
8. Bauer TW, Schoenfield L, Slaw RJ, et al. Validation of whole slide imaging for primary diagnosis in surgical pathology. *Arch Pathol Lab Med* 2013;137:518–24. [PubMed: 23323732]
9. Campbell WS, Hinrichs SH, Lele SM, et al. Whole slide imaging diagnostic concordance with light microscopy for breast needle biopsies. 2014;45:1713–21.
10. Fónyad L, Krenács T, Nagy P, et al. Validation of diagnostic accuracy using digital slides in routine histopathology. *Diagn Pathol* 2012;7:35. [PubMed: 22463804]
11. Goacher E, Randell R, Williams B, Treanor D. The diagnostic concordance of whole slide imaging and light microscopy: A systematic review. *Arch Pathol Lab Med* 2017;141:151–61. [PubMed: 27399211]
12. Hanna MG, Reuter VE, Ardon O, et al. Validation of a digital pathology system including remote review during the COVID-19 pandemic. *Mod Pathol* 2020.
13. Jones NC, Nazarian RM, Duncan LM, et al. Interinstitutional whole slide imaging teleconsultation service development: Assessment using internal training and clinical consultation cases. 2015;139:627–35.
14. Kent MN, Olsen TG, Feeser TA, et al. Diagnostic accuracy of virtual pathology vs traditional microscopy in a large dermatopathology study. *JAMA Dermatology* 2017;153:1285. [PubMed: 29049424]
15. Pantanowitz L, Sinard JH, Henricks WH, et al. Validating whole slide imaging for diagnostic purposes in pathology: Guideline from the college of american pathologists pathology and laboratory quality enter. 2013;137:1710–22.

16. Saco A, Diaz A, Hernandez M, et al. Validation of whole-slide imaging in the primary diagnosis of liver biopsies in a university hospital. *Dig Liver Dis* 2017;49:1240–6. [PubMed: 28780052]
17. Rakha EA, Aleskandarani M, Toss MS, et al. Breast cancer histologic grading using digital microscopy: concordance and outcome association. *J Clin Pathol* 2018;71:680–6. [PubMed: 29535212]
18. Krenacs T, Zsakovics I, Diczhazi C, et al. The potential of digital microscopy in breast pathology. *Pathology & Oncology Research* 2009;15:55–8. [PubMed: 18752051]
19. Al-Janabi S, Huisman A, Vink A, et al. Whole slide images for primary diagnostics of gastrointestinal tract pathology: a feasibility study. 2012;43:702–7.
20. Eccher A, Girolami I. Current state of whole slide imaging use in cytopathology: Pros and pitfalls. *Cytopathology* 2020;31:372–8. [PubMed: 32020667]
21. Hanna MG, Pantanowitz L. Why is digital pathology in cytopathology lagging behind surgical pathology? *Cancer Cytopathol* 2017;125:519–20. [PubMed: 28475271]
22. Girolami I, Pantanowitz L, Marletta S, et al. Diagnostic concordance between whole slide imaging and conventional light microscopy in cytopathology: A systematic review. *Cancer Cytopathol* 2020;128:17–28. [PubMed: 31600035]
23. Groen R, Abe K, Yoon H-S, et al. Application of microscope-based scanning software (Panoptiq) for the interpretation of cervicovaginal cytology specimens. *Cancer Cytopathol* 2017;125:918–25. [PubMed: 28960858]
24. Hanna MG, Monaco SE, Cuda J, et al. Comparison of glass slides and various digital-slide modalities for cytopathology screening and interpretation. *Cancer Cytopathol* 2017;125:701–9. [PubMed: 28558124]
25. ElGabry EA, Parwani A, Pantanowitz L. Whole-slide imaging: Widening the scope of cytopathology. *Diagn Histopathol* 2014;20:456–61.
26. U.S. Food and Drug Administration. FDA News Release: FDA allows marketing of first whole slide imaging system for digital pathology. <https://www.fda.gov/drugs/informationondrugs/approveddrugs/ucm553358.htm>. 2017. Accessed: 12 April 2017.
27. U.S. Food and Drug Administration. Leica biosystems receives FDA clearance for Aperio AT2 DX digital pathology system. https://www.accessdata.fda.gov/cdrh_docs/pdf19/K190332.pdf. 2019.
28. Williams BJ, Hanby A, Millican-Slater R, et al. Digital pathology for the primary diagnosis of breast histopathological specimens: an innovative validation and concordance study on digital pathology validation and training. *Histopathology* 2018;72:662–71. [PubMed: 28940580]
29. Ross J, Greaves J, Earls P, Shulruf B, Van Es SL. Digital vs traditional: Are diagnostic accuracy rates similar for glass slides vs whole slide images in a non-gynaecological external quality assurance setting? *Cytopathology* 2018;29:326–34. [PubMed: 29665117]
30. Hanna MG, Reuter VE, Hameed MR, et al. Whole slide imaging equivalency and efficiency study: experience at a large academic center. *Mod Pathol* 2019;32:916–28. [PubMed: 30778169]
31. Allison KH, Hammond MEH, Dowsett M, et al. Estrogen and progesterone receptor testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists guideline update. *Arch Pathol Lab Med* 2020;144:545–63. [PubMed: 31928354]
32. Wolff AC, Hammond MEH, Allison KH, et al. Human Epidermal Growth Factor Receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline focused update. *Arch Pathol Lab Med* 2018;142:1364–82. [PubMed: 29846104]
33. Camparo P, Egevad L, Algaba F, et al. Utility of whole slide imaging and virtual microscopy in prostate pathology. *APMIS* 2012;120:298–304. [PubMed: 22429212]
34. Ordi J, Castillo P, Saco A, et al. Validation of whole slide imaging in the primary diagnosis of gynecological pathology in a university hospital. 2015;68:33–9.
35. Davidson T, Rendi M, Frederick P, et al. Breast cancer prognostic factors in the digital era: Comparison of Nottingham grade using whole slide images and glass slides. *J Pathol Inform* 2019;10:11–. [PubMed: 31057980]
36. Hoda RS, Brogi E, D'Alfonso TM, et al. Interobserver variation of PD-L1 SP142 immunohistochemistry interpretation in breast carcinoma: A study of 79 cases using whole slide imaging. *Arch Pathol Lab Med* 2021.

37. Wilbur D, Brachtel E, Gilbertson J, et al. Whole slide imaging for human epidermal growth factor receptor 2 immunohistochemistry interpretation: Accuracy, Precision, and reproducibility studies for digital manual and paired glass slide manual interpretation. *J Pathol Inform* 2015;6:22–. [PubMed: 26110090]
38. Turashvili G, Leung S, Turbin D, et al. Inter-observer reproducibility of HER2 immunohistochemical assessment and concordance with fluorescent in situ hybridization (FISH): pathologist assessment compared to quantitative image analysis. *BMC Cancer* 2009;9:165. [PubMed: 19476653]
39. Tabata K, Mori I, Sasaki T, et al. Whole-slide imaging at primary pathological diagnosis: Validation of whole-slide imaging-based primary pathological diagnosis at twelve Japanese academic institutes. *Pathol Int* 2017;67:547–54. [PubMed: 28980740]
40. Snead DRJ, Tsang Y-W, Meskiri A, et al. Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology* 2016;68:1063–72. [PubMed: 26409165]
41. Nassar A, Cohen C, Albitar M, et al. Reading immunohistochemical slides on a computer monitor—A multisite performance study using 180 HER2-stained breast carcinomas. *Appl Immunohistochem Mol Morphol* 2011;19:212–7. [PubMed: 21475038]
42. Kondo Y, Iijima T, Noguchi M. Evaluation of immunohistochemical staining using whole-slide imaging for HER2 scoring of breast cancer in comparison with real glass slides. *2012;62:592–9.*
43. Clarke EL, Revie C, Brett D, et al. Development of a novel tissue-mimicking color calibration slide for digital microscopy. *Color Research & Application* 2018;43:184–97.
44. Pantanowitz L, Sinard JH, Henricks WH, et al. Validating whole slide imaging for diagnostic purposes in pathology: Guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *2013;137:1710–22.*

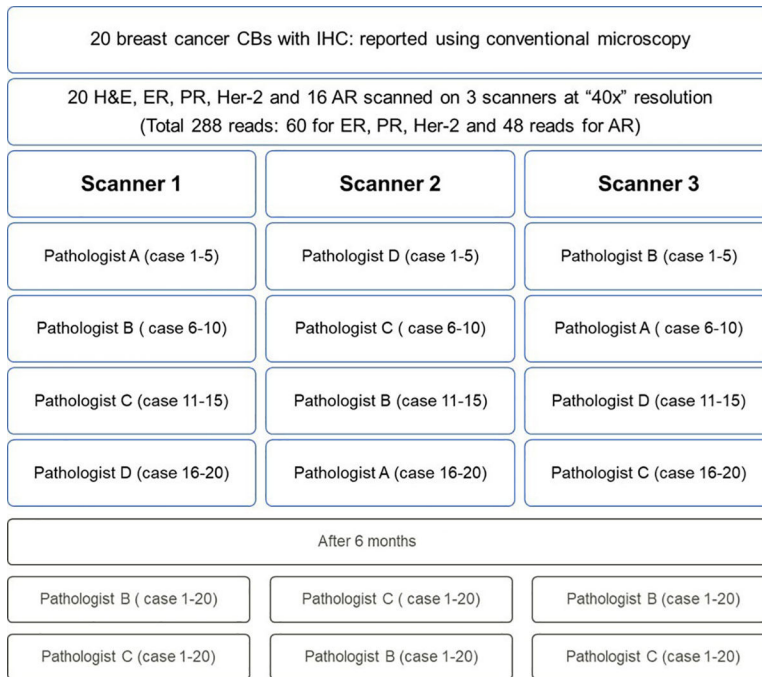


Figure 1.
Study design

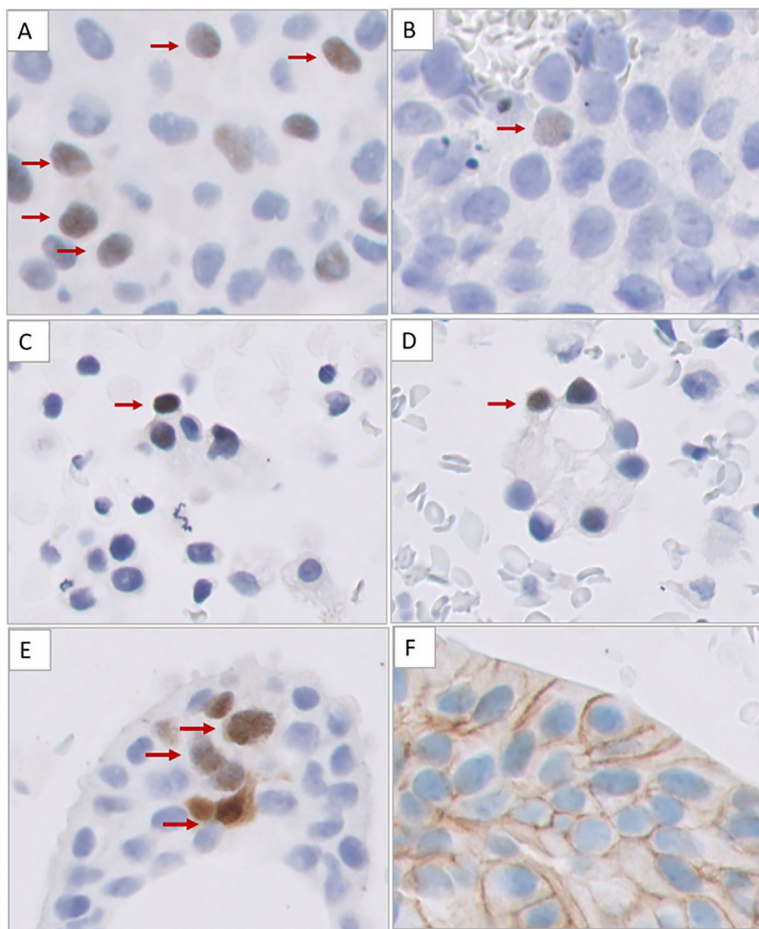


Figure 2. Five discordant cases IHC stained slides

A) ER stain: reported 5% on the glass slide and had a negative read B) PR stain: initially reported negative on the glass slide and had a 2% digital read, C-E) PR stains: initially reported 1% on the glass slide and had a negative digital read. F) HER2 stain: Initially reported equivocal (2+) on the glass read had a negative (1+) digital read. IHC: Immunohistochemistry, ER: Estrogen Receptor, PR: Progesterone Receptor, AR: Androgen Receptor, HER2: Human Epidermal Growth Receptor 2

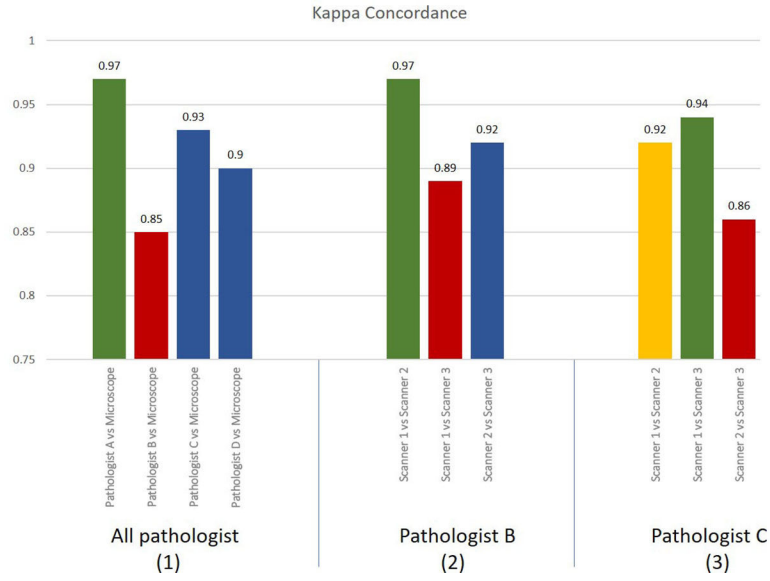


Figure 3. Kappa concordance

(1) Kappa concordance of ER, PR, AR and HER2 scores between all pathologist digital reads vs microscope. (2) Kappa concordance of ER, PR and AR scores divided into groups between all scanners for pathologist B. (3) Kappa concordance of ER, PR and AR scores divided into groups between all scanners for pathologist C. The p-value corresponds to a two-sided hypothesis test comparing reader-averaged accuracy with each scanner to the microscope and between scanners (0.41 to 0.60 moderate, 0.61 to 0.80 substantial, and 0.81 to 0.99 strong concordance). ER: Estrogen Receptor, PR: Progesterone Receptor, AR: Androgen Receptor, HER2: Human Epidermal Growth Receptor 2

Table 1:

Summary of Histology, Specimen Type, Site of 20 breast cancer samples

Age, median (range)	54 (35– 76)
Site type	n (%)
Locally recurrent	1 (5)
Metastatic	19 (95)
Specimen Type	
Chest wall	2 (10)
Lymph node	8 (40)
Pleural fluid	9 (45)
Breast	1 (5)
Tumor type	
IDC	85 (17)
IDC and ILC	5 (1)
Unknown	10 (2)

IDC: Invasive ductal carcinoma, ILC: Invasive lobular carcinoma

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

First-time successful scan rate, average number of rescans to successfully scan for each slide, Scanner precision by stain and by scanner

First-time successful scan rate (average)	CB	ER	PR	AR	HER2
Positive Negative	NA	14 6	7 13	12 4	1 15
Scanner 1 (81%)	85%	80%	80%	85%	80%
Scanner 2 (68%)	75%	85%	85%	81%	15%
Scanner 3 (43%)	85%	35%	35%	38%	20%
Average number of rescans to scan each slide	CB	ER	PR	AR	HER2
Scanner 1	1.3	1.5	1.5	1.5	1.3
Scanner 2	2.6	1.3	1.7	1.7	2.1
Scanner 3	1	2	1	1	1.5
Average Scanners precision (x3 times)	Average by scanner	ER	PR	AR	HER2
Scanner 1	92%	92%	90%	94%	92%
Scanner 2	78%	92%	92%	90%	40%
Scanner 3	73%	77%	77%	77%	60%
Average by stain		87%	86%	87%	64%

ER: Estrogen Receptor, PR: Progesterone Receptor, AR: Androgen Receptor, HER2: Human Epidermal Growth Receptor 2, CB: Cell block

Table 3:

Discordant diagnoses and reasons: 5 discordant cases showing glass reads, digital reads, reviewers reads, scanners used and reasons for discordance

Stain	ER	PR	PR	PR	HER2
Glass read	5	0	1	1	2+
Digital read	0	2	0	0	1+
Reviewer 1	2	2	1	1	1-2+
Reviewer 2	1	1	1	1	2+
Scanner	Scanner 2	Scanner 2	Scanner 2	Scanner 2 and 3	Scanner 2
Reason	Rare cells	Rare cells	Rare cells	Rare cells	Heterogenous weak staining

ER: Estrogen Receptor, PR: Progesterone Receptor, AR: Androgen Receptor, HER2: Human Epidermal Growth Receptor 2

Table 4:

Technical data: First-time failed scan for Negative versus positive IHC due to barcode or other reasons

First-time failed scan for Negative IHC:	ER % (n=6)	PR % (n=13)	AR % (n=4)	HER2 % (n=15)
Failed Total negative scans	6 18	14 39	4 12	29 45 (64%)
Due to barcode	50 (3)	64 (9)	100 (4)	79 (23)
Due to other reasons	50 (3)	36 (5)	0 (0)	21(6)
First-time failed scan for Positive IHC:	ER % (n=14)	PR % (n=7)	AR % (n=12)	HER2 % (n=1)
Failed Total positive scans	14 42	6 21	12 36	1 3 (33%)
Due to barcode	43 (6)	17 (1)	33 (4)	0 (0)
Due to other reasons	57 (8)	83 (5)	67 (8)	100 (1)

ER: Estrogen Receptor, PR: Progesterone Receptor, AR: Androgen Receptor, HER2: Human Epidermal Growth Receptor 2, IHC: immunohistochemistry

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript