VOLUME 35 | SUPPLEMENT 2 | MARCH 2022

MODERN PATHOLOGY

ABSTRACTS (977–1017)

INFORMATICS

2022

USCAP 111TH ANNUAL MEETING

REAL INTELLIGENCE

MARCH 19–24, 2022 LOS ANGELES, CALIFORNIA

Published by
SPRINGER NATURE
www.ModernPathology.org

USCAP Creating a Better Pathologist
AN OFFICIAL JOURNAL OF THE UNITED STATES AND CANADIAN ACADEMY OF PATHOLOGY

EDUCATION COMMITTEE

Rhonda K. Yantiss
Chair

Kristin C. Jensen
Chair, CME Subcommittee

Laura C. Collins
Chair, Interactive Microscopy Subcommittee

Yuri Fedoriw
Short Course Coordinator

Ilan Weinreb
Chair, Subcommittee for Unique Live Course Offerings

Carla L. Ellis
Chair, DEI Subcommittee

Adebowale J. Adeniran
Kimberly H. Allison
Sarah M. Dry
William C. Faquin
Karen J. Fritchie
Jennifer B. Gordetsky
Levon Katsakhyan, Pathologist-in-Training
Melinda J. Lerwill
M. Beatriz S. Lopes
Julia R. Naso, Pathologist-in-Training
Liron Pantanowitz
Carlos Parra-Herran
Rajiv M. Patel
Charles "Matt" Quick
David F. Schaeffer
Lynette M. Sholl
Olga K. Weinberg
Maria Westerhoff

ABSTRACT REVIEW BOARD

Benjamin Adam
Oyedele Adeyi
Mariam Priya Alexander
Daniela Allende
Catalina Amador
Vijayalakshmi Ananthanarayanan
Tatjana Antic
Manju Aron
Roberto Barrios
Gregory R. Bean
Govind Bhagat
Luis Zabala Blanco
Michael Bonert
Alain C. Borczuk
Tamar C. Brandler
Eric Jason Burks
Kelly J. Butnor
Sarah M. Calkins
Weibiao Cao
Wenqing (Wendy) Cao
Barbara Ann Centeno
Joanna SY Chan
Kung-Chao Chang
Hao Chen
Wei Chen
Yunn-Yi Chen
Sarah Chiang
Soo-Jin Cho
Shefali Chopra
Nicole A. Cipriani
Cecilia Clement
Claudiu Cotta
Jennifer A. Cotter
Sonika M. Dahiya
Elizabeth G. Demicco
Katie Dennis
Jasreman Dhillon
Anand S. Dighe
Bojana Djordjevic
Michelle R. Downes
Charles G. Eberhart
Andrew G. Evans
Fang Fan
Julie C. Fanburg-Smith
Gelareh Farshid
Michael Feely
Susan A. Fineberg
Dennis J. Firchau
Gregory A. Fishbein
Agnes B. Fogo
Andrew L. Folpe
Danielle Fortuna
Billie Fyfe-Kirschner
Zeina Ghorab
Giovanna A. Giannico
Anthony J. Gill
Tamar A. Giorgadze
Alessio Giubellino
Carolyn Glass
Carmen R. Gomez-Fernandez
Shunyou Gong
Purva Gopal
Abha Goyal
Christopher C. Griffith
Ian S. Hagemann
Gillian Leigh Hale
Suntrea TG Hammer
Malini Harigopal
Kammi J. Henriksen
Jonas J. Heymann
Carlo Vincent Hojilla
Aaron R. Huber
Jabed Iqbal
Shilpa Jain
Vickie Y. Jo
Ivy John
Dan Jones
Ridas Juskevicius
Meghan E. Kapp
Nora Katabi
Francesca Khani
Joseph D. Khoury
Benjamin Kipp
Veronica E. Klepeis
Christian A. Kunder
Stefano La Rosa
Stephen M. Lagana
Keith K. Lai
Goo Lee
Michael Lee
Vasiliki Leventaki
Madelyn Lew
Faqian Li
Ying Li
Chieh-Yu Lin
Mikhail Lisovsky
Lesley C. Lomo
Fang-I Lu
aDeqin Ma
Varsha Manucha
Rachel Angelica Mariani
Brock Aaron Martin
David S. McClintock
Anne M. Mills
Richard N. Mitchell
Hiroshi Miyamoto
Kristen E. Muller
Priya Nagarajan
Navneet Narula
Michiya Nishino
Maura O'Neil
Scott Roland Owens
Burcin Pehlivanoglu
Deniz Peker Barclift
Avani Anil Pendse
Andre Pinto
Susan Prendeville
Carlos N. Prieto Granada
Peter Pytel
Stephen S. Raab
Emilian V. Racila
Stanley J. Radio
Santiago Ramon Y Cajal
Kaaren K Reichard
Jordan P. Reynolds
Lisa M. Rooper
Andrew Eric Rosenberg
Ozlen Saglam
Ankur R. Sangoi
Kurt B. Schaberg
Qiuying (Judy) Shi
Wonwoo Shon
Pratibha S. Shukla
Gabriel Sica
Alexa Siddon
Anthony Sisk
Kalliopi P. Siziopikou
Stephanie Lynn Skala
Maxwell L. Smith
Isaac H. Solomon
Wei Song
Simona Stolnicu
Adrian Suarez
Paul E. Swanson
Benjamin Jack Swanson
Sara Szabo
Gary H. Tozbikian
Gulisa Turashvili
Andrew T. Turk
Efsevia Vakiani
Paul VanderLaan
Hanlin L. Wang
Stephen C. Ward
Kevin M. Waters
Jaclyn C. Watkins
Shi Wei
Hannah Y. Wen
Kwun Wah Wen
Kristy Wolniak
Deyin Xing
Ya Xu
Shaofeng N. Yan
Zhaohai Yang
Yunshin Albert Yeh
Huina Zhang
Xuchen Zhang
Bihong Zhao
Lei Zhao

To cite abstracts in this publication, please use the following format: **Author A, Author B, Author C, et al. Abstract title (abs#). In "File Title."** ***Modern Pathology*** **2022; 35 (suppl 2): page#**

977 Deep Learning Model for the Prediction of EBV-associated Gastric Cancer

Sangjeong Ahn[1], Cristina Cho[2], Yeojin Jeong[3], Ji-Eon Kim[4], Jonghyun Lee[5], Namkug Kim[2], Jiyoon Jung[1], Ju Yeon Pyo[6], Jisun Song[1], Woon Yong Jung[7], Yoo Jin Lee[8], Kyoung Min Moon[9]

[1]International St. Mary's Hospital, Catholic Kwandong University College of Medicine, Incheon, South Korea, [2]Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea, [3]Genome & Health Big Data Laboratory, Graduate School of Public Health, Seoul National University, Seoul, South Korea, [4]Wonkwang University Medical Research Convergence Center, Wonkwang University Hospital, Iksan, South Korea, [5]Hanyang University, Seoul, South Korea, [6]Seogu, Incheon Metropolitan City, South Korea, [7]Hanyang University Guri Hospital, Hanyang University College of Medicine, Guri, South Korea, [8]Korea University Anam Hospital, Seoul, South Korea, [9]Gangneung Asan Hospital, Gangneung, South Korea

Disclosures: Sangjeong Ahn: None; Cristina Cho: None; Yeojin Jeong: None; Ji-Eon Kim: None; Jonghyun Lee: None; Namkug Kim: None; Jiyoon Jung: None; Ju Yeon Pyo: None; Jisun Song: *Grant or Research Support*, South Korea; Woon Yong Jung: None; Yoo Jin Lee: None; Kyoung Min Moon: None

Background: Detection of EBV status in gastric cancer is crucial for clinical decision making, as it identifies patient with different treatment response and prognosis. Despite its apparent importance, EBV status is not readily testable because of limited medical resources. Therefore, widely accessible and cost-effective tools for the testing is required. Here, we investigate the potential of a deep learning-based system for automated EBV prediction directly from hematoxylin and eosin (H&E)-stained whole-slide images (WSIs).

Design: The model was developed using 270 H&E-stained WSIs (24 with EBV positive and 246 with EBV negative), from The Cancer Genome Atlas, containing WSIs. Our deep learning model (EBVNet) consists of two sequential components: a tumor classifier and an EBV classifier (Figure 1). We visualized the features extracted by the tumor classifier (Figure 2A) and the EBV classifier (Figure 2B) using Uniform Manifold Approximation and Projection (UMAP). We internally and externally validated the model on patch-level dataset of TCGA and slide-level dataset of 61 H&E- stained WSIs (6 with EBV positive and 55 with EBV negative). Performance was primary evaluated using the sensitivity, specificity, precision, negative predictive value (NPV) and F1-score.

Results: The EBVNet model achieved a sensitivity of 92.4%, specificity of 99.5%, precision of 92.4%, NPV of 99.5%, and F1-score of 92.4% on the patch-level dataset. On the external dataset (slide-level dataset), our model achieved a sensitivity of 85.7%, specificity of 81.1%, precision of 37.5%, NPV of 97.7%, and F1-score of 52.2%.

Figure 1 - 977

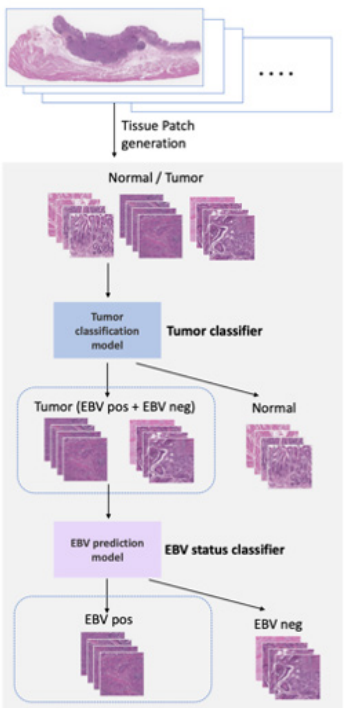
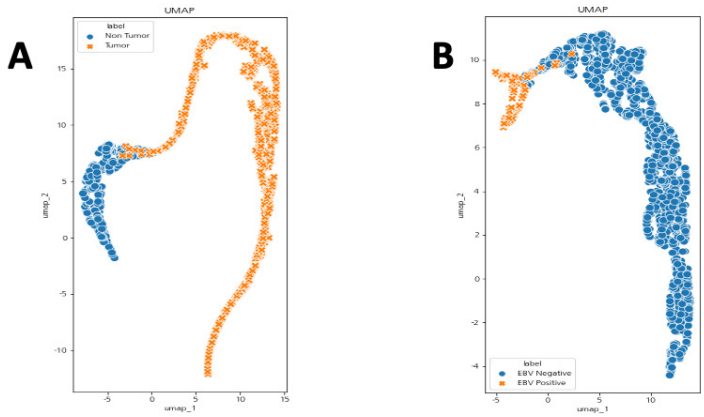


Figure 2 – 977



Conclusions: Within the current universal EBV testing paradigm, our model might contribute value as to prescreen patients before confirmatory testing, potentially reducing the number of tested patients, thereby resulting in substantial test-related labor and cost savings.

978 Automated Tumor Budding Detection by Computer Algorithm that Uses Knowledge Learned from Cytokeratin Immunohistochemistry

Vidya Arole[1], Thomas Tavolara[2], Muhammad Khalid Khan Niazi[3], Wendy Frankel[4], Metin Gurcan[2], Deborah Knight[4], Wei Chen[4]

[1]The Ohio State University, Columbus, OH, [2]Center for Biomedical Informatics at Wake Forest University, Winston-Salem, NC, [3]Wake Forest School of Medicine, Winston-Salem, NC, [4]The Ohio State University Wexner Medical Center, Columbus, OH

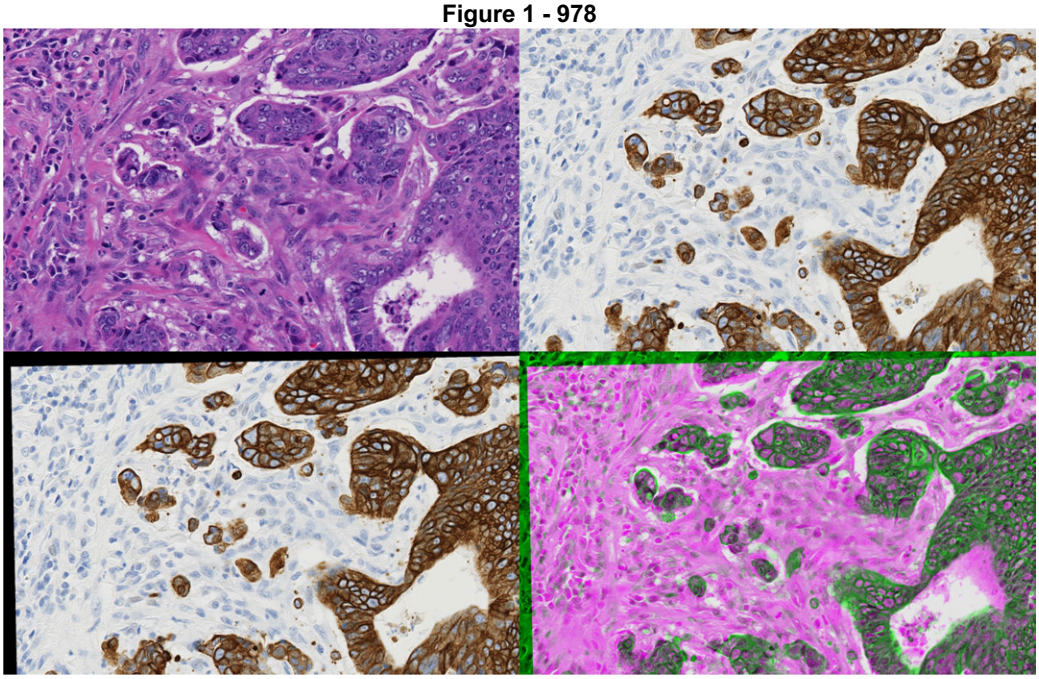
Disclosures: Vidya Arole: None; Thomas Tavolara: None; Muhammad Khalid Khan Niazi: None; Wendy Frankel: None; Metin Gurcan: None; Deborah Knight: None; Wei Chen: None

Background: Tumor budding (TB) is defined as a single tumor cell or a cell cluster of up to 4 cells at the tumor invasive front. It is an independent adverse prognostic factor in colorectal cancer (CRC). The detection of TB can be challenging when there is significant peritumoral inflammation and reactive stroma. For such cases, cytokeratin AE1/3 (CK) immunohistochemistry has been recommended to aid in TB identification. We aim to develop a computer algorithm for automated marking of TB on H&E images, which uses knowledge learned from CK images during training. No actual CK stain is needed when using the algorithm in clinical cases. Here we emphasize the challenges encountered during algorithm development.

Design: Adjacent CK and H&E images were registered to provide reliable ground truth for tumor regions and TB, and swin transformers were utilized to segment malignant tumor regions on H&E. Next, a series of automatic post-processing steps transferred small positive regions identified on CK (potential TB) onto registered H&E, where they were filtered by the number and size of nuclei. This yielded TB on H&E that benefit from registered CK. Notably, automatic non-rigid image registration techniques were utilized to deform CK images to maximize overlap with adjacent H&E images (Figure 1). We used Mattes mutual information as a metric, and a one-plus-one evolutionary algorithm as an optimizer. 60 paired H&E and CK images of CRC were analyzed by the algorithm and by consensus review of pathologists.

Results: In this initial phase of work, the concordance between the computer algorithm and a consensus reading (generated by fusing multiple readings by pathologists) was 60%. The discrepancy was mainly due to: 1) False positive TB that are TB mimickers, including fragmented glands embedded in peri-tumoral inflammation, degenerating tumor cell debris, and reactive myofibroblasts / macrophages expressing CK; 2) TB on CK image but not present on immediately adjacent H&E image. The algorithm accurately differentiated TB from several TB mimickers (reactive endothelial cells, ganglions, reactive stromal cells). The algorithm was helpful in a way similar to an actual CK stain.

Figure 1 - 978



Conclusions: A computer algorithm that uses knowledge learned from CK immunostain could aid pathologists in TB evaluation. TB mimickers on CK images are the main challenges that need to be overcome. Accurately annotated H&E-stained TB images (ground truth) is essential for training and refining the algorithm.

979 The ALTA Project: All Labs to All Albertans

David Beyer[1], Etienne Mahe[2], Leslee Phillips[3], Heather Sereda[3], Susan Nahirniak[4]
[1]University of Alberta Hospital, Edmonton, Canada, [2]University of Calgary, Calgary, Canada, [3]Alberta Health Services, Calgary, Canada, [4]Alberta Precision Laboratories, University of Alberta, Edmonton, Canada

Disclosures: David Beyer: None; Etienne Mahe: None; Leslee Phillips: None; Heather Sereda: None; Susan Nahirniak: None

Background: Alberta, Canada is home to approximately 4.4 million people, served by a publicly funded provincial healthcare system known as Alberta Health Services. Physicians are able to access lab results on any patient they are providing medical services for through an existing electronic health data repository called Alberta Netcare. In 2020, the provincial government mandated that patients have access to all of their own laboratory test data, including hematology, chemistry, microbiology, molecular genetics, and pathology. The goal of this project was to implement a user-friendly patient portal to allow patients to access their own lab data, no matter where in the province it was drawn or obtained.

Design: Currently, Alberta uses 5 major Laboratory Information Systems (LIS). Labs are pooled into a provincial data repository (see Figure 1). Our aim was to create a secure patient portal called MyHealthRecords (MHR) that pulled test result data from the provincial repositories and displayed it to patients accurately and securely. The MHR viewer references ALREP (a consolidated lab repository) which contains the pan-Canadian LOINC observation code database (pCLOCD/LOINC) name in the HL7 lab result message. The lab data elements are then displayed in MHR. For more complex tests (i.e. pathology reports) these results are generated from the Netcare data repository and rendered to the portal viewer, or sent as a PDF document directly from the source LIS.

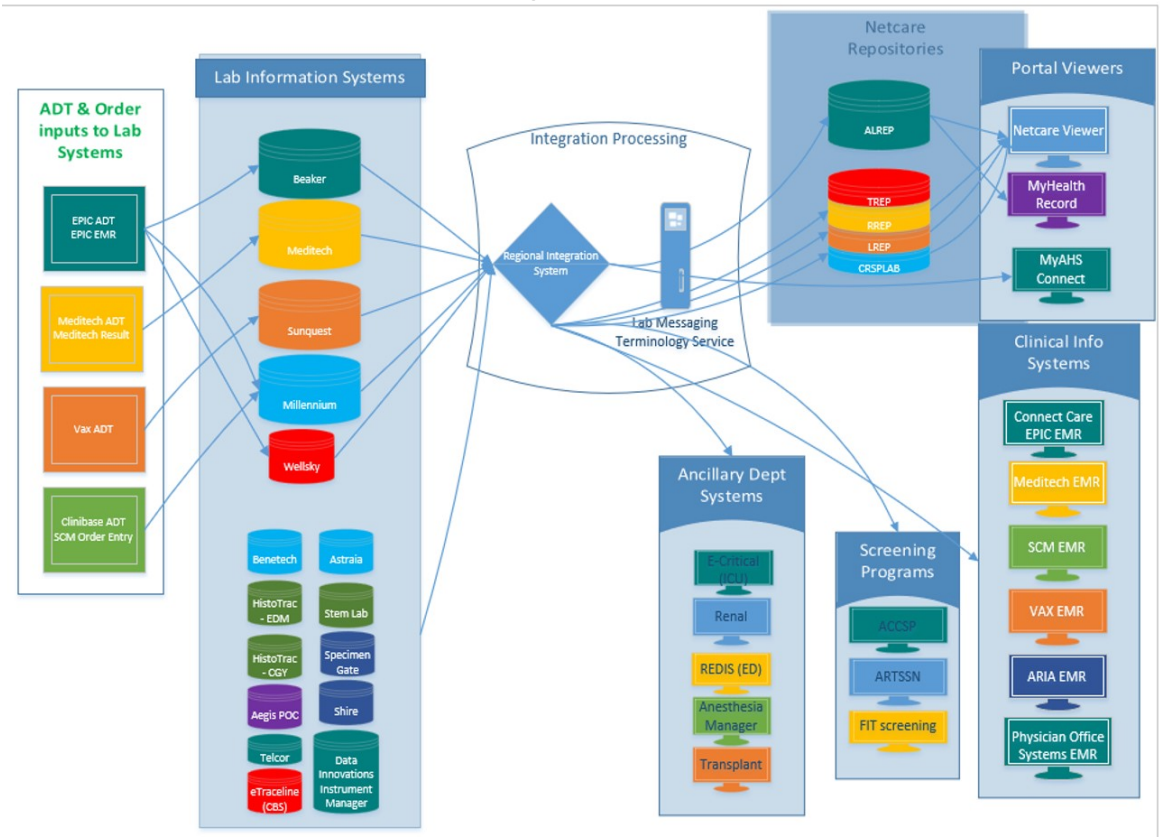
Results were validated by comparing the original LIS, Netcare, and the MHR viewer to ensure concordance. This validation process was completed for all tests that would be displayed on MHR, with physician sign off required.

Results: A limited test menu (accounting for 50% of labs by volume) was initially launched in 2017, followed by a 2nd launch in Feb. 2021, providing patients' access to 95% of laboratory tests by volume. The user's account was secured by using the already

existing My Alberta digital ID. The portal can be accessed online or via a secure app. In addition to test results being displayed, the portal also contains links to information about the various lab tests, what they mean, and how they are collected.

Fig. 1: Data flow through our 5 main LIS (Beaker, Meditech, Sunquest, Millennium, and Wellsky), as it moves to the patient accessed MHR.

Figure 1 - 979



Conclusions: As of August 2021, >99% of laboratory results are now available directly to patients, with over 1 million active users across the province. This has not only benefited patients, but also provided easily accessible vaccination records, of particular importance during the ongoing COVID-19 pandemic.

980 Best Practices of Dataset Generation for Clinical-Grade Deep Learning Image Analysis with Application to Mitosis Detection

Tom Bisson[1], Rasmus Kiehl[1], Rita Carvalho[1], Sebastian Lohmann[2], Tobias Lang[3], Sebastian Springenberg[3], Kai Saeger[4], Inti Zlobec, Peter Hufnagl[5], Norman Zerbe[1]
[1]Charité University Hospital Berlin, Berlin, Germany, [2]University of Applied Sciences Berlin, Berlin, Germany, [3]Mindpeak GmbH, Hamburg, Germany, [4]Berlin, Germany, [5]Charité University Medicine Berlin, Berlin, Germany

Disclosures: Tom Bisson: None; Rasmus Kiehl: None; Rita Carvalho: None; Sebastian Lohmann: None; Tobias Lang: *Employee*, Mindpeak GmbH; *Stock Ownership*, Mindpeak GmbH; Sebastian Springenberg: *Employee*, Mindpeak GmbH; Kai Saeger: *Employee*, VMscope GmbH; Inti Zlobec: None; Peter Hufnagl: None; Norman Zerbe: None

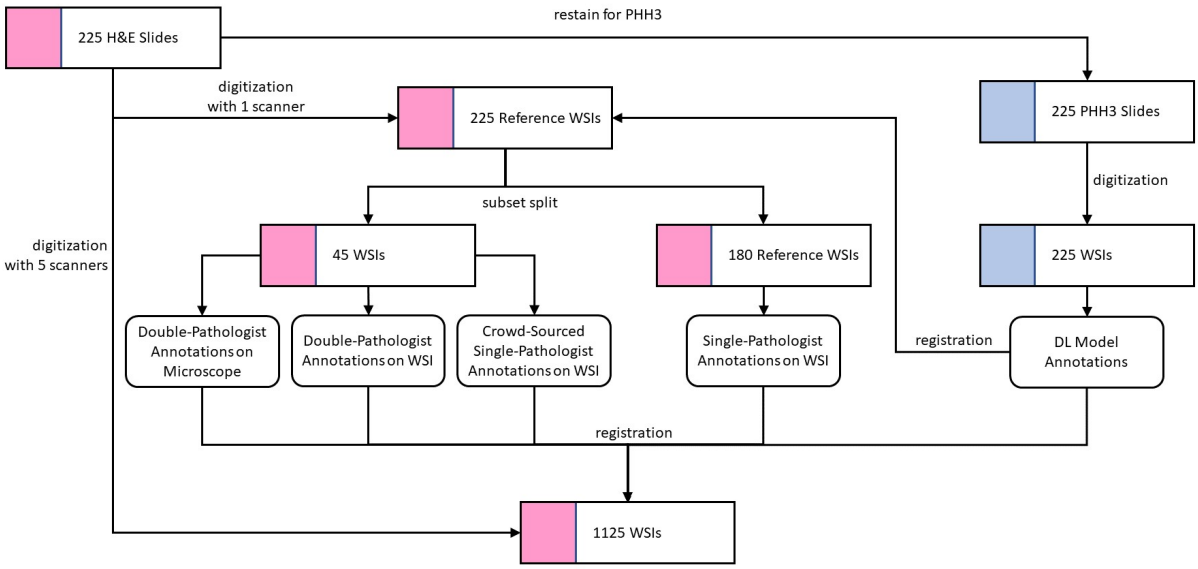
Background: Image analysis support using deep learning (DL) has gained significant attention in pathology, but adoption rates into clinical routine remain low. A barrier is the scarcity of appropriate generalized datasets for development and validation. Best practices for dataset construction are missing. A prime example is mitosis detection which although it was one of the first successful demonstrations of DL analysis in histopathology, has not been adopted in clinical routine due to missing generalizability

and insufficient methodic validation. We provide best practices to build datasets for clinical-grade DL, demonstrated using mitosis detection as an example.

Design: Constructing a clinical-grade dataset requires heterogeneous tissue samples from multiple labs digitized with scanners from various vendors (Fig.1). Annotations should be made on a reference subset and be transferred to other images by registration, ideally involving different annotation modalities to decrease interobserver variance. Re-staining using IHC allows confirmation of annotations and provides a complementary datasource by using an additional DL model to translate DAB-positive structures into annotations. Cascading models in dataset generation requires careful data stratification to avoid statistical bias. For validation, out-of-sample evaluation on data from an external source is needed.

Results: Using our best practices, we have built a clinical-grade dataset to develop DL models for mitosis detection. Our dataset is wide covering large-scale preanalytical heterogeneity and deep involving finely tuned annotations. It contains 1350 WSIs corresponding to 225 physical slides, involving a total of 2736 representative tumor regions and 67332 annotated mitotic figures. Systematic evaluation of annotation methods on a subset revealed interobserver variances on the order of 26% and intraobserver variances of 30% among pathologists depending on the annotation method. IHC staining allowed us to confirm manual annotations as well as extending annotations to outside the annotated regions.

Figure 1 - 980



Conclusions: Construction efficiency and data quality are fundamental to clinical-grade datasets. Both depend on the annotation method and the pathologists involved. To achieve a reliable and verified data quality and reduce interobserver variance, different methods should be combined on at least a subset of the data. Using mitosis detection we showed how best practices for dataset construction can be used for algorithm development and validation.

981 Detection of Lymph Node Metastasis of Breast Carcinomas Using Digital Imaging Analysis

Bindu Challa[1], David Kellough[2], Swati Satturwar[2], Giovanni Lujan[3], Wendy Frankel[1], Anil Parwani[3], Shaoli Sun[1], Zaibo Li[1]
[1]The Ohio State University Wexner Medical Center, Columbus, OH, [2]The Ohio State University Wexner Medical Center/James Cancer Hospital, Columbus, OH, [3]The Ohio State University, Columbus, OH

Disclosures: Bindu Challa: None; David Kellough: *Employee*, Inspirata, Inc; Swati Satturwar: *Grant or Research Support*, University Pittsburgh Medical Center Developmental Laboratory, University Pittsburgh Medical Center Developmental Laboratory; Giovanni Lujan: None; Wendy Frankel: None; Anil Parwani: None; Shaoli Sun: None; Zaibo Li: None

Background: Artificial intelligence (AI)-based digital imaging analysis (DIA) has been explored to detect metastatic tumor cells in lymph node with promising results. We have investigated the “Tumor Detection” DIA algorithm (Visiopharm) to assess hematoxylin and eosin (H&E) stained slides of lymph nodes from invasive breast carcinoma specimens.

Design: The study cohort consisted of 258 lymph nodes from 21 consecutive invasive breast ductal or lobular carcinoma specimens. H&E slides of lymph nodes were scanned using Philips IntelliSite Scanners and digital images were analyzed using Visiopharm Tumor Detection App. The 4 steps in DIA consisted of tissue detection, metastasis detection, post-processing and calculation of results. At the end of 4 steps, the image shows areas suspicious for metastasis in yellow (low suspicion), orange (intermediate suspicion) and red (high suspicion) annotations and also shows the size of the largest area of metastasis. DIA results were compared with the ground truth which was confirmed by pathologists' read on H&E with/without cytokeratin stains.

Results: DIA detected metastasis with an accuracy of 85.2%. The true positive rate was 100% (81/81), false negative rate was 0% (0/81), false positive rate was 21.5% (38/139) and true negative rate was 71.5% (139/177) (Table). In true positive cases, DIA detected the largest area of metastasis ranging from 0.08 mm to 15.33 mm (deviation of 0 to +/- 5.98 mm compared to the ground truth). In false positive cases, the largest area of metastasis was small in size and ranged from 0.01 mm to 1.1 mm (average size - 0.407 mm). Almost all false positive cases (37/38) were caused by presence of histiocytes in lymph nodes except one case, which was due to the presence of melanocytes in lymph node.

Figure 1: Digital image analysis algorithm detects (A) metastatic lobular carcinoma and (B) isolated tumor cells in lymph node.

Figure 2: False positive cases. (A) One case with histiocytes, which are negative for AE1/3. (B) One case with melanocytes, which are negative for AE1/3, but positive for S100 and SOX10.

		Ground truth		
		Positive	Negative	Total
Digital imaging analysis	Positive	81	38	119
	Negative	0	139	139
	Total	81	177	258

Figure 1 - 981

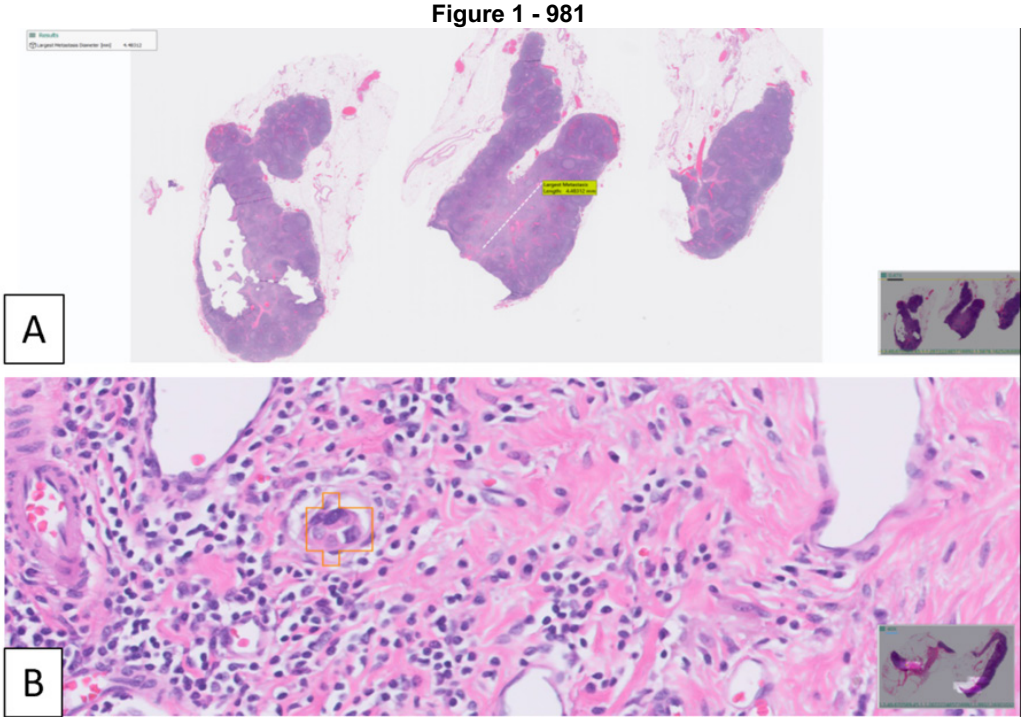
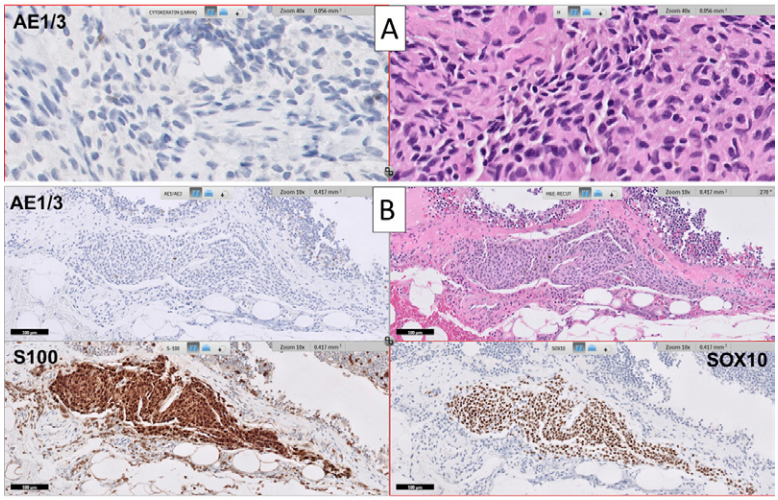


Figure 2 – 981



Conclusions: DIA demonstrates high sensitivity (100%) in detecting tumor metastasis but tends to overcall metastasis in order to avoid missing tumor detection. Given its high sensitivity, it can serve as a screening tool in identifying metastasis including micrometastasis and isolated tumor cells to save pathologists' time, and may reduce the usage of immunohistochemistry.

982 The Value of Performing Next Generation Sequencing in Solid Tumors

Zachary Chelsky[1], Lester Layfield[1], Katie Wilkinson[1], Richard Hammer[1], Yaswitha Jampani[1]
[1]*University of Missouri, Columbia, MO*

Disclosures: Zachary Chelsky: None; Lester Layfield: None; Katie Wilkinson: None; Richard Hammer: *Grant or Research Support*, Roche, GE; *Advisory Board Member*, Foundation Medicine, Caris Lifesciences; Yaswitha Jampani: None

Background: The value of next generation sequencing (NGS) is often debated, yet few studies have shown the frequency where NGS can identify therapeutically useful variants. The sequencing of a tumor by NGS may identify variants for which FDA-approved therapy exists for that tumor type, pathogenic variants for which an FDA-therapy is approved for a different tumor type, pathogenic variants which help select a therapeutic protocol, or variants of undetermined significance. We examined NGS results at our institution to determine frequency of therapeutically useful variants.

Design: We reviewed NGS reports of the 10 most common solid tumor types tested (143 cases). The tumor types examined were lung adenocarcinoma, lung squamous cell carcinoma, breast carcinoma, head and neck squamous cell carcinoma, prostate adenocarcinoma, colon adenocarcinoma, pancreas adenocarcinoma, ovary serous carcinoma, unknown primary adenocarcinoma, and unknown primary carcinoma (nos). Reports were reviewed for the presence of pathologic variants associated with an FDA-approved drug therapy for the patient's tumor type, a different tumor type, or selection of treatment protocol.

Results: Cases containing pathogenic variants with a matching treatment were identified. The total percentage was 63% for lung adenocarcinoma, 45% for lung squamous cell carcinoma, 86% for breast carcinoma, 57% for head and neck squamous cell carcinoma, 64% for prostate adenocarcinoma, 36% for colon adenocarcinoma, 86% for pancreas adenocarcinoma, 50% for ovary serous carcinoma, 43% for unknown primary adenocarcinoma, and 67% for unknown primary carcinoma (nos). The matching treatment was either FDA-approved for the patient's tumor type or for a different tumor type. Several pathogenic variants identified were not associated with any FDA-approved treatment, and patients may have had variants in multiple categories as detailed in Table 1. In 83% of cases a clinical trial was identified.

Tumor Type	Total Number of Cases	Number of Patients with a pathogenic variant associated with FDA-approved treatment in patient's tumor type	Patients having a pathogenic variant associated with FDA-approved treatment in tumor type different from patient's tumor type*	Patients having a pathogenic variant associated with no treatment type*
lung adenocarcinoma	46	12	25	24
lung squamous cell carcinoma (scc)	15	2	10	8
breast carcinoma (nos)	14	11	6	10
head and neck squamous cell carcinoma (hnscc)	14	8	8	8
prostate acinar adenocarcinoma	14	3	8	3
colon adenocarcinoma (crc)	11	6	6	4
pancreas ductal adenocarcinoma	8	0	3	7
ovary serous carcinoma	8	1	5	1
unknown primary adenocarcinoma	7	0	5	3
unknown primary carcinoma (cup) (nos)	6	0	3	4
*Patients may have a pathogenic variant associated with an FDA-approved treatment and other variants associated with a treatment in a different tumor type, or no treatment at all. As a result patients may be counted in more than one column or more than once in a column.				

Conclusions: NGS panels for solid tumors frequently identifies pathogenic variants associated with FDA treatments, however the treatment choices may not always be approved for a patients' tumor type. NGS provides valuable results for potential treatment in many tumors examined. Further analysis of the data is needed to determine the impact of NGS results and if the patient responded to that treatment. In addition, further clarification is needed to determine if the utility of NGS differs with cancer stage or treatment history.

983 Deep Learning-based Classification Algorithm of Colorectal Biopsy Using 19000 Whole Slide Images

Yosep Chong[1], Nishant Thakur[1], Hongjun Yoon[2], Gyoyeon Hwang[1], Gisu Hwang[2], Hyeongsub Kim[3]
[1]College of Medicine, The Catholic University of Korea, Seoul, South Korea, [2]Deepnoid, Seoul, South Korea, [3]Pohang, South Korea

Disclosures: Yosep Chong: None; Nishant Thakur: None; Hongjun Yoon: None; Gyoyeon Hwang: None; Gisu Hwang: None; Hyeongsub Kim: None

Background: Colorectal cancer (CRC) is one of the most leading causes of cancer mortality worldwide and colonoscopic biopsy is the most widely used screening exam. Recently, there have been many attempts to classify histologic images of CRCs with a help of artificial intelligence (AI). However, the limited number of available datasets with quality annotation by expert pathologists is still the biggest issue to develop robust AI models with desirable performance for assisting daily practice. In this study, we developed an AI model to classify the histologic images of colonoscopic biopsy into five histological classes using 19,000 whole slide images (WSIs).

Design: Hematoxylin eosin-stained WSIs from two university hospitals of the Catholic University of Korea (Yeouido and Uijeongbu St. Mary's Hospitals) were annotated with labels of five histologic classes including normal (ACI), hyperplastic polyp (HYP), adenoma (ADN), adenocarcinoma (ADC), and carcinoid (CAR) by 3 expert pathologists. Extracted patch images were pre-processed with cropping, filtering artifacts, and normalizing colors, and resized to 1,024×1,024 pixels. The dataset was split into a 7:2:1 ratio for training, validation, and test dataset. Pre-processed images were fed into the efficient classification network, EfficientNet-B0. Cross sampling using horizontal/vertical flips and rotations was used to balance the number of each class.

Results: A total of 92,996 image patches from 18,992 WSIs of 9,631 patients were used for training AI models. Our best model achieved 83.18% of accuracy for the entire WSI test dataset and 69.91%, 86.97%, 68.03%, 91.84%, and 87.50% of accuracy for ACI, HYP, ADN, ADC, and CAR, respectively.

Conclusions: With the help of the magnitude of a well-annotated dataset, our model showed the potential to classify the five histological subtypes very precisely and can also reduce the workload of pathologists in a clinical setting. In addition, this is the first study to classify histologic images of CRC with carcinoid class showing a decent performance.

984 Performance and Reproducibility of a Deep Learning-Based Approach at Diagnosing Celiac Disease

Kevin Côté[1], Yan Miao[1], Mina Farag[2], Farhad Maleki[1], Peter Savadjiev[1], Zu-Hua Gao[3]
[1]McGill University, Montreal, Canada, [2]McGill University Health Centre, Glen site, Montreal, Canada, [3]The University of British Columbia, Vancouver, Canada

Disclosures: Kevin Côté: None; Yan Miao: None; Mina Farag: None; Farhad Maleki: None; Peter Savadjiev: None; Zu-Hua Gao: None

Background: Rendering an accurate diagnosis of celiac disease (CD) can often be difficult, depending on identification of relevant semiological, serological, genetic, endoscopic, and histopathological evidence. Although it remains the diagnostic 'gold standard,' histopathological diagnosis of CD suffers from suboptimal interobserver variability. This study evaluated the performance and reproducibility of a deep learning-based approach at diagnosing CD.

Design: We assembled a total of 671 histological slides from a single tertiary referral hospital, representing 612 adult patients investigated for CD between 2016 and 2019. Whole glass slides were scanned at 40x using Aperio's ScanScope XT, reviewed by a pathologist-in-training (ground truth) blind to the original histopathological diagnoses, and graded according to a simplified Marsh-Oberhuber classification, where Marsh types 1 and 2 become type 1/2, and Marsh type 3 is not further divided between 3a, 3b, and 3c. Image patches of size 224x224 pixels were automatically extracted from each digitized slide at 3 different magnification levels: 20x, 10x and 5x. Three ResNet50 convolutional neural network (CNN) models, one for each magnification level, were trained to compute the probability of a patch belonging to one of the above histotypes. This probability was then aggregated across patches and magnification levels to obtain an overall classification label for the entire slide.

Results: Table 1 summarizes the diagnostic performance of this method. The mean per-class sensitivity, specificity, positive predictive value, and negative predictive value were 0.852, 0.905, 0.823, and 0.902, respectively. Also, Cohen's κ was computed – using the algorithm's predictions (see Table 1) as one rater and the ground truth as another – and showed good agreement between the two ($\kappa = 0.72$, $p < .001$).

Grading of biopsies		By pathologist (ground truth)			Total
		0	1/2	3	
By algorithm (predictions)	0	32	5	0	37
	1/2	10	27	0	37
	3	3	0	21	24
Total		45	32	21	98

Table 1. Agreement between the conclusions of the pathologist and the predictions of our deep learning-based algorithm concerning 98 digitized slides used for testing the latter's performance. Results are expressed following a simplified Marsh-Oberhuber classification.

Conclusions: Artificial intelligence (AI) represents a major paradigm shift in diagnostic medicine. Most of the previous studies have focused on the diagnosis of neoplastic lesions in the context of their tumor morphological and molecular classifications. Our study shows that AI could also be applied to diagnosing non-neoplastic lesions. In addition to achieving high precision, recall, and accuracy in differentiating CD biopsies from normal biopsies, we presented some evidence that our method may deliver more reproducible results than the standard histopathological diagnostic pipeline.

985 Cell-Specific Parameters Differentiate Benign from Malignant Soft Tissue Tumors

Jing Di[1], Caylin Hickey[1], Cody Bumgardner[1], Shadi Qasem[2]
[1]University of Kentucky College of Medicine, Lexington, KY, [2]University of Kentucky Healthcare, Lexington, KY

Disclosures: Jing Di: *Primary Investigator*, Janssen R&D, LLC; Caylin Hickey: None; Cody Bumgardner: None; Shadi Qasem: *Primary Investigator*, Janssen Pharmaceuticals

Background: Soft tissue pathology is a challenging area for most pathologists, and consultations are often needed from experts in the field. Molecular techniques, including next generation sequencing, have aided in the classification of some of these entities. One of the most challenging aspects of soft tissue pathology is differentiating benign and reactive processes from malignant ones. Current machine learning and artificial intelligence models are difficult to apply to soft tissue pathology given the large number of entities and the rarity of these tumors. In this study, we sought to identify cell-specific parameters that may be helpful in classifying these tumors into benign versus malignant.

Design: Ninety-five (95) soft tissue cases were retrieved from the files of one of the authors (SQ). The cases were scanned at 40X magnification using a high throughput scanner (Aperio AT2). At the highest layer of resolution slide images were divided into 768-pixel tiles. Each tile was evaluated for tissue percentage and color factors using pre-processing software original developed as part of the Deep HistoPath project. Cell detection was performed on each tile with at least 25% tissue using QuPath, providing 38 cell-specific parameters, which were averaged per tile. Whole Slide Images (WSI) were uploaded to an OMERO server for annotation. A pathologist marked regions of interest (ROI) on all slides. Metadata associated with tile and cell metrics was used to select the top 500 tiles per case within ROI regions. Cell-based metrics were averaged across the top tiles per case and were averaged on the case-level. Welch's *t* test was used with $p<0.05$ considered significant. **[Figure 1]**

Figure 1. Slide images were divided into 768-pixel tiles at the highest layer of resolution

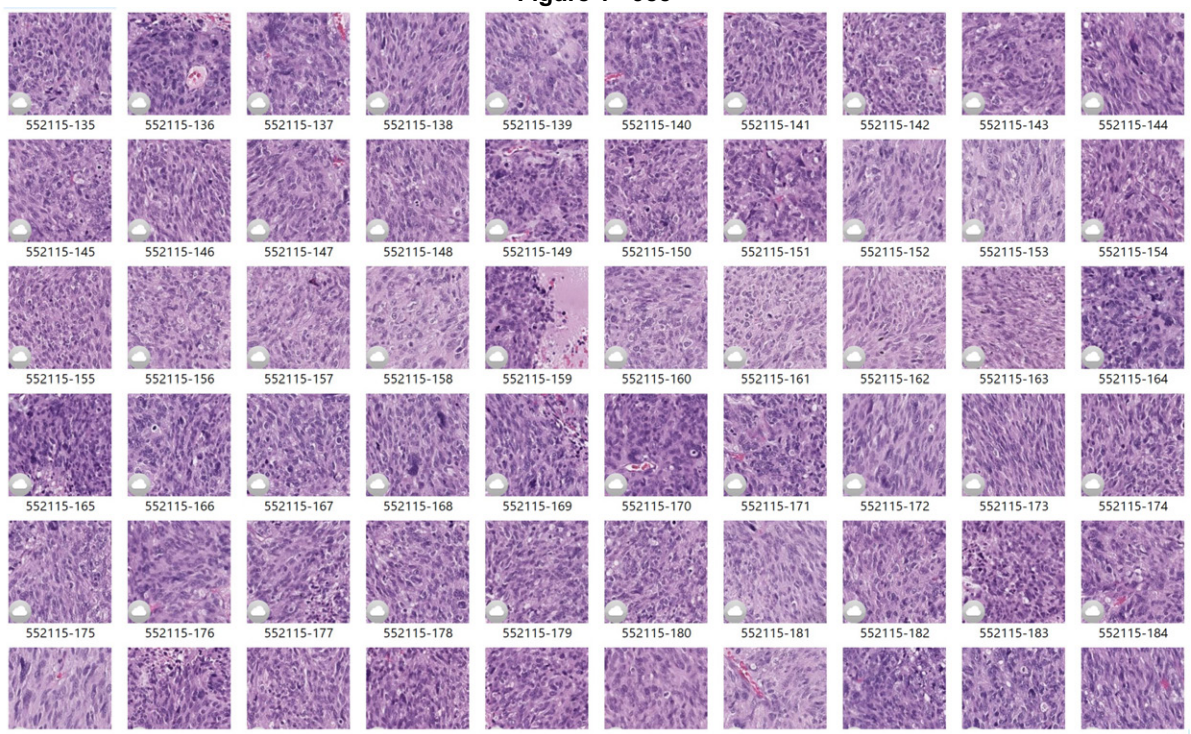
Results: The patients were 52 females and 43 males ranging in age from 9 months to 90 years (average 41, median 42). There were 60 benign and 35 malignant cases encompassing 68 distinct soft tissue entities from various locations in the body. Several cell-specific parameters were shown to be statistically significant ($p <0.05$) including cell count, nucleus/cell area ratio, nucleus hematoxylin OD* mean, cell max caliper, cell area, cell perimeter, cell circularity, cell min caliper. **[Table 1]**

*Optical density

Table 1: Cell-specific parameters and their P values

	Mean		Difference between means ±SEM	95% confidence interval	P value
	Benign	Malignant			
Cases (n)	60	35			
Nucleus/Cell area ratio	0.218	0.2731	0.05504 ± 0.008515	0.03804 to 0.07204	<0.0001
Cell count	171.9	278.9	107.0 ± 18.51	70.08 to 143.9	<0.0001
Cell: Max caliper	16.33	14.95	-1.381 ± 0.2853	-1.949 to -0.8130	<0.0001
Nucleus: Hematoxylin OD sum	36.71	47.15	10.44 ± 2.362	5.738 to 15.15	<0.0001
Cell: Area	135.1	110.7	-24.42 ± 5.009	-34.38 to -14.46	<0.0001
Cell: Perimeter	44.33	40.5	-3.831 ± 0.7991	-5.422 to -2.240	<0.0001
Cell: Circularity	0.8166	0.7968	-0.01988 ± 0.004069	-0.02796 to -0.01180	<0.0001
Cell: Min caliper	10.88	9.898	-0.9780 ± 0.2126	-1.401 to -0.5548	<0.0001

Figure 1 - 985



Conclusions: Cell-specific parameters such as cell count, nucleus/cell area, and nuclear and cytoplasmic staining characteristics can differentiate benign and malignant soft tissue tumors. Building algorithms and models using cell-specific parameters may be a helpful way to aid in classifying these tumors.

986 Clinical Validation of a Web Application to Create Nomenclature of FLT3 Internal Tandem Duplication from Next Generation Sequencing Result

Yi Ding[1], Geoffrey Smith[2], Kristin Deeb[2], Andrew Campbell[1], Linsheng Zhang[3]
[1]Geisinger Medical Center, Danville, PA, [2]Emory University School of Medicine, Atlanta, GA, [3]Emory University Hospital, Atlanta, GA

Disclosures: Yi Ding: None; Geoffrey Smith: None; Kristin Deeb: None; Andrew Campbell: None; Linsheng Zhang: None

Background: FMS-like tyrosine kinase 3 (*FLT3*) internal tandem duplication (ITD) in the juxtamembrane region is an activation mutation occurs in 20-25% of acute myeloid leukemia and some cases of precursor T-cell leukemia/lymphoma. The Archer® VariantPlex® Myeloid next generation sequencing panel and the coupled analysis pipeline are proven to have high sensitivity in identifying *FLT3*-ITD; however, the pipeline failed to generate appropriate nomenclature as recommended by the Human Genome Variation Society (HGVS), especially when the duplication/insertion is over 30 bp.

Design: We compiled a python script to create *FLT3*-ITD HGVS nomenclature automatically from the assembled consensus sequence generated by Archer® VariantPlex® Myeloid panel. The script is refactored into a web application (available for free at http://tinyurl.com/flt3itd). Assembled sequences of sixty-one *FLT3*-ITD cases identified from laboratory sequencing databases, and 10 artificial sequences of variable ITDs, are used for the clinical validation of this web application. Each sequencing result is inspected manually to confirm that the nomenclature is correct on cDNA, amino acid levels and matches accurate chromosome coordinates.

Results: Sixty cases and 10 artificial sequences created nomenclatures fully compliant with HGVS recommendations. The sizes of ITDs in 61 cases range from 21 to 195 (median 48) bp; 2 cases each contain 2 ITD variants of different sizes. Of the 63 real ITD sequences, only 35 (55.6%) are exact duplicate sequences; the other 28 (44.4%) are partial duplications containing nucleotide or amino acid variants and can only be named as insertion. One of the 63 sequences is not named correctly due to mutations and/or sequencing errors at the 3'end of the inserted sequence. Random character strings are reported as "not DNA sequence" and wild

type *FLT3* or unrelated DNA sequences are reported “no ITD in the sequence”. Cases with insertions containing intronic sequences may have duplicated DNA sequences, but not duplication of wild-type amino acids due to disruption of the splice sites. Determining the final amino acid sequence can be challenging due to our limited understanding of RNA splicing.

Conclusions: Our free web application can correctly create HGVS *FLT3*-ITD nomenclature from next generation sequencing results, as long as there is no error or mutation at the 3’ end of the sequence. RNA-based sequencing is required to definitively determine how the intronic sequence in ITD is translated into amino acids.

987 Histomorphometric Features of Prostate Cancer Identify Patients Who Biochemically Recur after Prostatectomy

Savannah Duenweg[1], Michael Brehler[1], Allison Lowman[2], Samuel Bobholz[2], John Sherman[1], Fitzgerald Kyereme[1], Kenneth Iczkowski[2], Peter Laviolette[2]
[1]Froedtert and the Medical College of Wisconsin, Wauwatosa, WI, [2]Medical College of Wisconsin, Milwaukee, WI

Disclosures: Savannah Duenweg: None; Michael Brehler: None; Allison Lowman: None; Samuel Bobholz: None; John Sherman: None; Fitzgerald Kyereme: None; Kenneth Iczkowski: None; Peter Laviolette: None

Background: Prostate cancer (PCa) is the most diagnosed non-cutaneous cancer in men, accounting for one in five new cancer diagnoses [PMID: 31912902]. If organ confined, removal of the prostate is considered curative. In some cases, distant metastases form resulting in poor patient prognosis. PCa is currently graded using Gleason Grade Groups [PMID: 26166626] to predict prognosis; however, subjective grading may vary between pathologists. This study seeks to determine whether quantitative pathomic features of PCa can discriminate patients who biochemically recur (BCR) following surgery.

Design: Whole mount prostate histology from 72 patients was analyzed for this IRB approved study. In total, 562 slides were H&E stained, digitized, and annotated by a GU fellowship-trained pathologist (KAI). High-resolution tiles were extracted from each annotation. Individual glands were identified using automated image processing algorithms. Eight pathomic features were calculated per gland (**Figure 1)** and averaged across each patient. Tiles were separated categorically into cancer and benign. Patients were followed with PSA testing according to standard of care practices (median 2.5 years, range 0.1-6.2 years). Logistic regressions were fit to assess the ability of the pathomic features, as well as tumor volume and cancer percentage to predict BCR (PSA ≥ 0.1 ng/ml post-surgery) per tissue type. Additionally, a model was fit to predict BCR from clinical information to compare accuracy.

Results: Logistic regression classified each of the tile groups at an accuracy > 85%, as well as receiver operating characteristic curve (ROC) area under the curve (AUC) of approximately 0.9 (**Figure 2**). The pathomic feature model outperformed clinical information which achieved an accuracy of 83% and an AUC of 0.87.

Figure 1 - 987

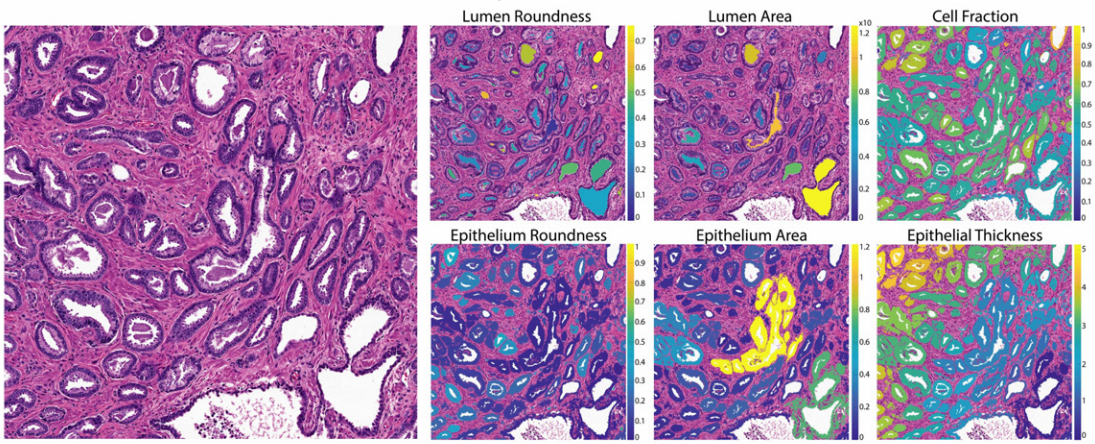


Figure 1 A Gleason 3 tile (left) was segmented to calculate pathomic features including (top) lumen roundness and area; cell fraction (percent of nuclei per gland area); (bottom) epithelium roundness, area, and thickness; and (not shown) overall tile stroma and epithelial area. Calculated values are overlaid on the respective gland. Units of area maps are in mm2, thickness in mm; roundness and cell fraction are unitless.

Figure 2 – 987

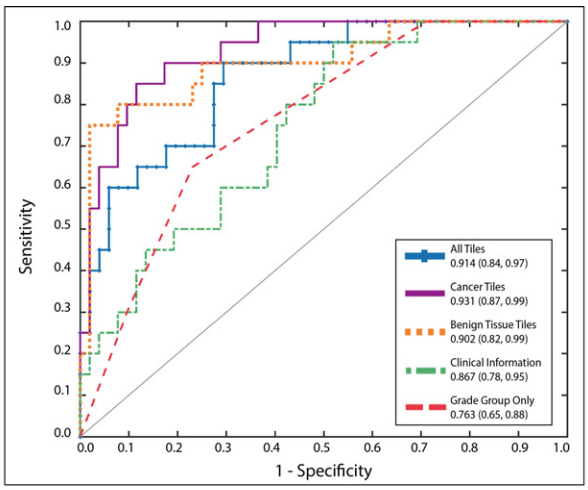


Figure 2 ROC curves to predict BCR stratified by tile group, clinical information (i.e., age, surgery stage, Grade Group, pre-surgery PSA, tumor ratio and volume), and Gleason Grade Group only. Calculated AUCs are shown with 95% confidence intervals.

Conclusions: Use of quantitative pathomic features calculated from digital histology of PCa as well as tumor size may provide clinicians with information beyond the traditional qualitative pathologic assessment. Specifically, we show that across benign tissue and all tiles, pathomic features are predictive of BCR. A prior study showed luminal features to be prognostic of BCR [PMID: 33941830]; however, we show that luminal, epithelial, and stromal features can all stratify patients by BCR risk. Further research is warranted to determine potential inclusion in treatment guidance.

988 HemeLabel: A Web-Based Platform for Rapid Bone Marrow Aspirate Image Annotation for Artificial Intelligence Applications in Hematopathology

Gregory Goldgof[1], Jacob Van Cleave[2], Laura Brown[1], Linlin Wang[1], Sonam Prakash[1], Atul Butte[1]
[1]University of California, San Francisco, San Francisco, CA, [2]San Francisco, CA

Disclosures: Gregory Goldgof: None; Jacob Van Cleave: None; Laura Brown: None; Linlin Wang: None; Sonam Prakash: None; Atul Butte: None

Background: Computational image classification methods are increasingly aiding pathologists in the diagnostic workup, ability to pick-up rare malignant cells, and improve case triage and efficiency. However, there has been limited development of tools for bone marrow aspirate smears (BMAs). Artificial intelligence (AI) systems could assist pathologists with cell counting, classification and abnormal cell detection on BMAs, improving both the efficiency and accuracy of diagnosis. However, a major bottleneck to the development of these algorithms is the lack of libraries of annotated images of cells from BMAs needed for training and testing, as well as tools for building these libraries.

Design: We built HemeLabel, a web-based application for the rapid labeling of hematopoietic cells from blood and BMAs. It is built with Django, a Python-based, open-source web-framework. Whole slide image (WSI) handling is performed using the OpenSeaDragon.js and user interactivity is implemented using JavaScript, AJAX, JQuery, and Fabric.js. Labeling speeds were then compared to Qupath, a general-purpose WSI annotation application.

Results: HemeLabel allows users to rapidly assign classifications to a spectrum of normal and abnormal cell types using custom key bindings based on lineage-based patterns (Figure 1). The platform allows multiple annotators to provide blinded classifications for studies of inter-pathologist variability, as well as manual cell segmentation (Fig 2a, 2b). Annotated cells are displayed in cell-type and lineage-based lists for rapid review (Fig 2c). Speed tests demonstrate an average classification speed of 1,282 ± 119 cells/hour for 20 marrow cell classes, 2.9x faster than QuPath ($p<0.001$) (Table 1).

We used HemeLabel to build a library of over 10,000 pathologist-labelled cells from hundreds of WSIs of BMAs and peripheral blood smears. This library includes a diverse range of hematopathology cases archived over almost 20 years at a quaternary care academic hospital system.

	Labeling Speed (Cells/Hour)	Time to Classify 1,000 Cells (Minutes)
HemeLabel	1,282 ± 119	47 ± 4.2
QuPath	439 ± 35.6	138 ± 10.8

Table 1: Speed test for benign cell segmentation and classification task (20 classes). HemeLabel performs 2.9x faster than QuPath ($p<0.001$).

Figure 1 - 988

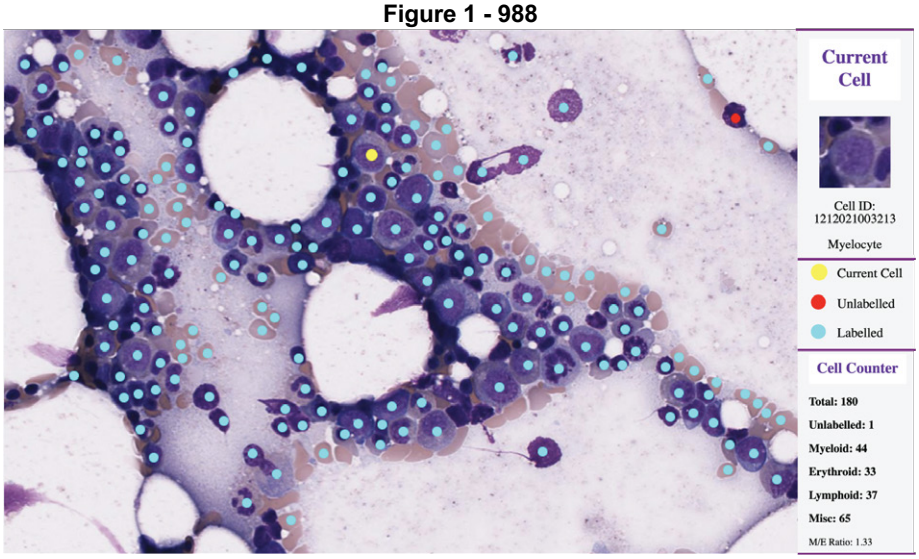
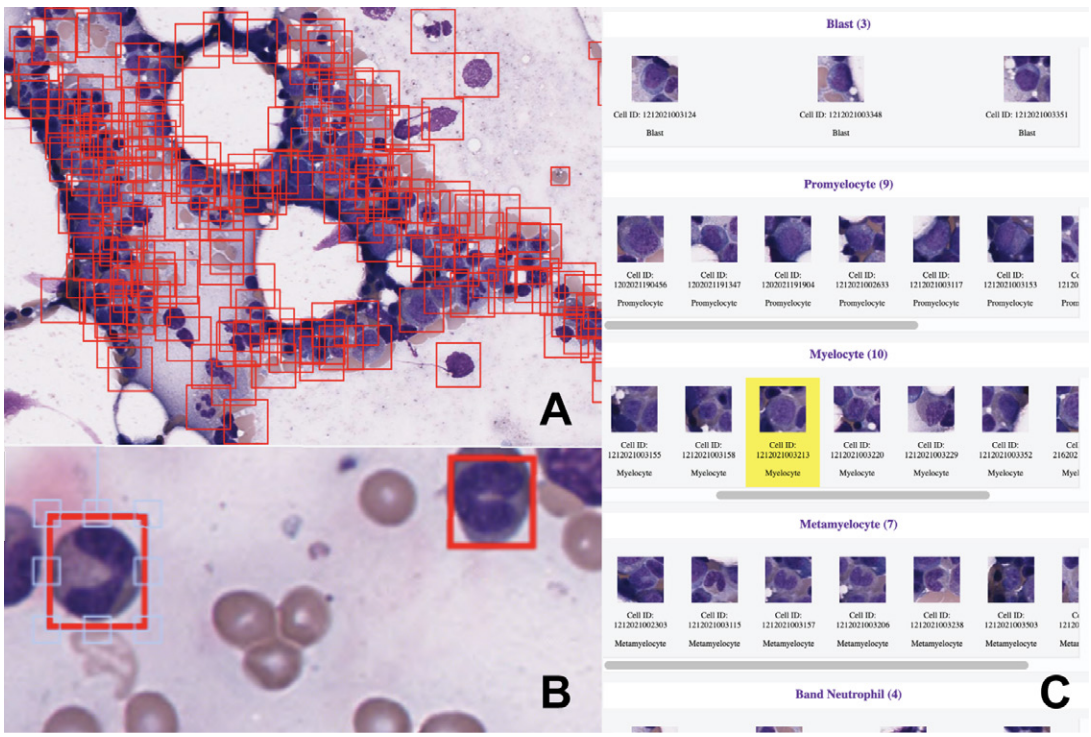


Figure 2 – 988



Conclusions: This library is fueling efforts to build AI tools for hematopathology digital workflow that integrates WSI systems for BMAs. The HemeLabel platform can be adapted to other image classification tasks, including alternative magnifications and different sources such as cytology or tissue patches. The online nature of HemeLabel enables collaboration across institutions to build diverse libraries that can be used to train AIs that are robust across institutions.

989 Diagnostic Value and Reliability of Digital Images in the Detection of Infectious Microorganisms

Mehrvash Haghighi[1], Clare Bryce[1], Shafinaz Hussein[1], John Paulsen Jr.[2], Christian Salib[2], Brandon Veremis[2], Roshanak Alialy[3], Mega Lahori[4], Yansheng Hao[5], Yuanxin Liang[6], Arnold Szporn[2], William Westra[2]
[1]Mount Sinai Hospital, New York, NY, [2]Icahn School of Medicine at Mount Sinai, New York, NY, [3]Mount Sinai St. Luke's Roosevelt Hospital, New York, NY, [4]Memorial Sloan Kettering Cancer Center, New York, NY, [5]University of Rochester Medical Center, Rochester, NY, [6]Yale School of Medicine, New Haven, CT

Disclosures: Mehrvash Haghighi: None; Clare Bryce: None; Shafinaz Hussein: None; John Paulsen Jr.: None; Christian Salib: None; Brandon Veremis: None; Roshanak Alialy: None; Mega Lahori: None; Yansheng Hao: None; Yuanxin Liang: None; Arnold Szporn: None; William Westra: None

Background: The reliable diagnosis of microorganisms in pathology samples is crucial to initiate and monitor treatment. As more institutions are adopting digital pathology for primary diagnosis, the evaluation of diagnostic value of digital images in detecting microorganisms seems crucial. This study investigates the visibility of different microorganism on the digitized special stain and immunostain slides.

Design: Original glass slides from all cases were randomized and de-identified. The positive and negative slides were placed in random order and divided based on the stain category as follows: Herpesvirus (immunostains) (n=100); Spirochete (immunostains) (n=80); Cryptococcus (mucin) (n=20); fungi (GMS) (n=100); H.pylori (immunostains) (n=100); H.pylori (Giemsa) (n=100); Cytomegalovirus (immunostains) (n=100); acid-fast bacilli (AFB) (n=20). All slides were scanned by Philips Ultra Fast Scanner at 40x. Of the 620 slides, 4 were excluded from the study due to lack of tissue detection by scanners and poor quality of scanned images after multiple rescanning. The diagnosis rendered by the original pathologist who signed out the glass slide was considered the reference standard. Six pathologists reviewed the digital images. The results were documented as: positive, negative and equivocal. Fleiss' kappa was run to determine the level of agreement among observers in each category. The equivocal values were excluded from the statistic measurement of diagnostic values.

Results: Fleiss' kappa showed good agreement among observers' interpretation results in the HSV, Spirochete, H.pylori, CMV immunostains and GMS stains. The AFB, mucin and Giemsa stains showed moderate agreement in the interpretation result among observers. The review of digital and glass slides of discrepancies revealed stain artifact in 2% and lack of focus on microorganisms in 1% of slides.

Table: Diagnostic test values of digital pathology for detection of microorganisms

	HSV (IHC)	Spirochete (IHC)	Cryptococcus (Mucin)	Fungi (GMS)	H.pylori (IHC)	H.pylori (Giemsa)	CMV (IHC)	Acid-fast bacilli (AFB)
Accuracy	93.25	93.17	75	92.46	98.60	87.12	94.6	67.82
Sensitivity (%)	95.66	87.55	64.28	89.93	97.8	84.82	91.56	54.05
Specificity (%)	88.65	99.09	100	95.88	100	90.58	98.76	92.68
PPV (%)	94.13	99.02	100	96.72	100	93.13	99.02	93.02
NPV (%)	91.48	88.30	54.54	87.59	96.33	79.84	89.55	52.77
Kappa coefficient	0.72 (Good)	0.78 (Good)	0.55 (Moderate)	0.67 (Good)	0.68 (Good)	0.54 (Moderate)	0.73 (Good)	0.51 (Moderate)

Conclusions: Our study demonstrates that diagnostic values are highest for immunostain H.pylori, HSV, CMV, Spirochete, and GMS stains. The examination of digital images of H.pylori immunostains provides better diagnostic value compared to Giemsa stain. The additional training for the interpreting AFB, Giemsa and mucin stains could improve the level of agreement among observers and the diagnostic values of digital images. Based on our findings, the interpretation of digital AFB and mucin stains is associated with low interrater reliability, low accuracy, sensitivity and NPV and it should not be used as an alternative to conventional glass slides.

990 Osteosarcoma Patient Stratification Based on Objective and Reproducible Post-Therapy Necrosis Assessment by Pixel-wise Deep Segmentation

David Ho[1], Narasimhan Agaram[1], Marc-Henri Jean[1], Chad Vanderbilt[1], John Healey[1], Paul Meyers[1], Thomas Fuchs[2], Meera Hameed[1]
[1]Memorial Sloan Kettering Cancer Center, New York, NY, [2]Icahn School of Medicine at Mount Sinai, New York, NY

Disclosures: David Ho: *Other*, Paige AI; Narasimhan Agaram: None; Marc-Henri Jean: None; Chad Vanderbilt: *Consultant*, Paige AI; John Healey: None; Paul Meyers: None; Thomas Fuchs: *Stock Ownership*, Paige.ai Inc.; *Employee*, Paige.ai Inc.; Meera Hameed: None

Background: Osteosarcoma is the most common primary bone tumor whose standard treatment includes pre-operative chemotherapy followed by resection and post-operative chemotherapy. Response to chemotherapy in the form of >90% tumor necrosis correlates with 5-year survival. Necrosis assessment is manually performed by pathologists' microscopic review of H&E-stained slides. This is a subjective measurement and is prone to inter-observer and intra-observer variabilities. In this study, we use a pixel-wise deep learning-based segmentation model to calculate case-level necrosis ratio in an objective and reproducible way and stratify patients to predict overall survival (OS) and progression-free survival (PFS).

Design: We scanned 92 cases with 2864 whole slide images (WSIs) by Aperio AT2 in 20x magnification. Seventy five WSIs from 15 cases were annotated to train the segmentation model and the remaining 77 cases with 2463 WSIs were used as an independent validation set. We trained 7-class Deep Multi-Magnification Network segmenting viable tumor (VT), necrosis with bone (NwB), necrosis without bone (NwoB), normal bone, normal tissue, cartilage, and blank. The number of pixels segmented as VT, NwB, and NwoB were counted to calculate the case-level necrosis ratio. The validation cases were stratified by necrosis ratio to predict OS and PFS. The log-rank test was used to analyze patient stratification based on necrosis ratio calculated by the segmentation model.

Results: Our deep segmentation model can successfully segment tissue subtypes including VT in red, NwB in blue, and NwoB in yellow, shown in Figure 1, and case-level necrosis ratio calculated by the model highly correlated with manual assessment. With the 90% threshold for OS stratification, we achieved p=0.0021 (Figure 2A). Furthermore, we found the optimal threshold for OS stratification is 80% for our segmentation model with p=0.000001 (Figure 2B). Using a subset of validation set with 66 cases excluding cases with metastases at the time of diagnosis, we stratified patients for PFS with p=0.018 at the optimal threshold of 70% (Figure 2C).

Figure 1 - 990

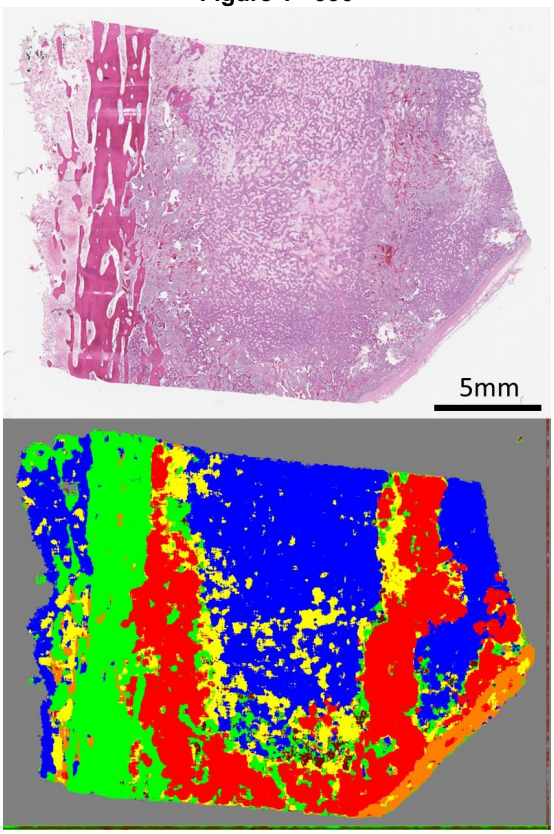
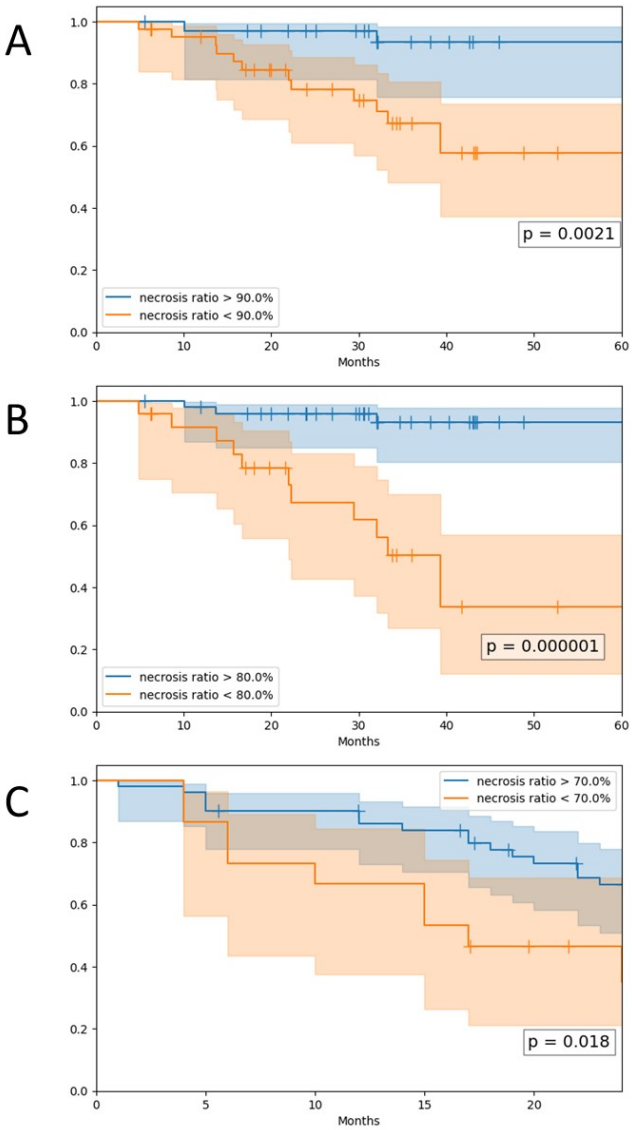


Figure 2 - 990



Conclusions: Our deep segmentation model was able to not only estimate necrosis ratio of post-therapy osteosarcoma cases, but also stratify patients for OS and PFS.

991 Effects of Color Calibration via ICC Profile on Inter-scanner Generalization of AI Models

Kshitij Ingale[1], Rohan Joshi[1], Irvin Ho[1], Aicha BenTaieb[1], Martin Stumpe[1]
[1]*Tempus Labs, Chicago, IL*

Disclosures: Kshitij Ingale: *Employee*, Tempus Labs; Rohan Joshi: *Employee*, Tempus Labs; Irvin Ho: None; Aicha BenTaieb: None; Martin Stumpe: None

Background: Generalizing deep learning models to whole slide images (WSIs) from different sources is a well-known challenge in digital pathology. Scanners used for digitization of slides usually yield images in a different color space than the standard RGB (sRGB) space, characterized by the scanner ICC profile. Some scanners do not apply color space transform to sRGB during WSI creation, limiting the utility of imaging algorithms. Data augmentation is a popular method to aid model generalization, however, it requires retraining a deployed model as well as additional experiments to tune the extent of color augmentation. In this study, we evaluate the relative effect of color augmentation and ICC profile transformation (which does not require re-training) by assessing the performance of a model trained to predict microsatellite instability from H&E slides with and without these procedures.

Design: For this study, a dataset of 1203 slides scanned with both Philips and Aperio scanners was curated and split into training (962 slides) and holdout-test sets (241 slides). Models were trained with or without color augmentation using only Philips WSIs, and evaluated on an independent Phillips and Aperio hold-out test set with (or without) ICC profile application. This process was repeated for 2 variants of attention based multiple instance learning model architectures (termed model 1 and model 2).

Results: Figure 1a-c demonstrates color differences in Philips and Aperio WSIs with and without ICC. Figure 1d highlights that the ICC sRGB transform partially corrects differences in saturation for the Aperio scanner and moves the mean closer to the Philips mean. Model performance on unseen scanner slides is lower without color augmentation or ICC profile transformation (Figure 2). This gap can be bridged by retraining models with color augmentation. Additionally, ICC profile transform either improved model generalization without requiring retraining (model 1, Figure 2a) or produced qualitatively similar results(model 2). ICC profile transform does not negatively impact performance when used with or without color augmentation.

Figure 1 - 991

Figure 1

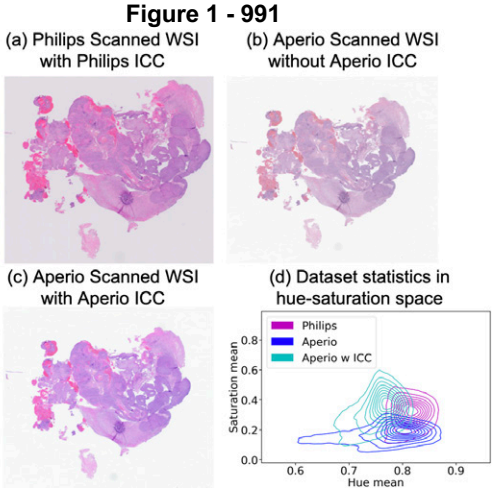
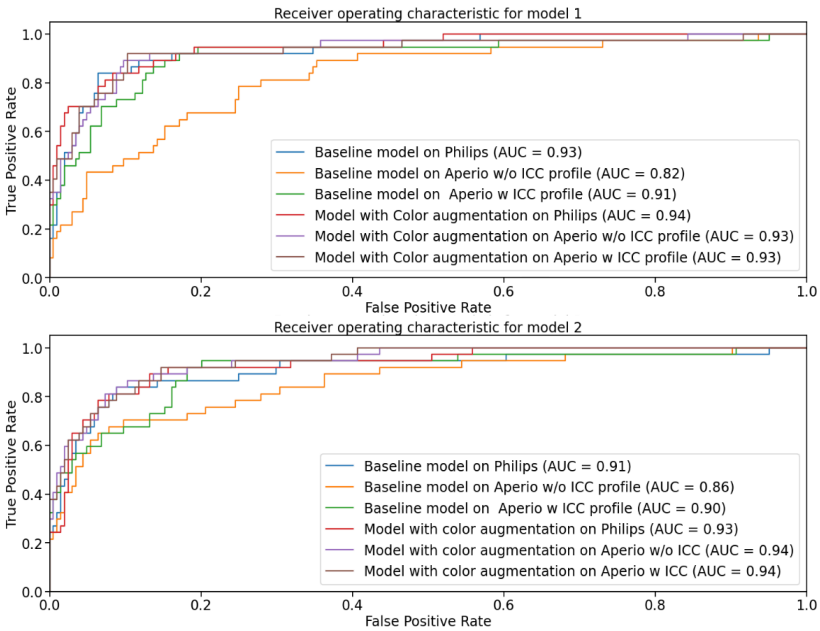


Figure 2 – 991

Figure 2



Conclusions: Using ICC profile transform can yield similar or better results on unseen scanner slides without model retraining. Color normalization via ICC profile could facilitate inter-scanner generalization when data is scarce. It can also be used as a part of the standardization workflow in digital pathology.

992 Deep Learning Identifies Microsatellite Instability in H&E Whole Slide Images from Prostate, Esophageal, and Gastric Cancers and Generalizes across Cancer Types

Rohan Joshi[1], Andrew Kruger[1], Elle Moore[1], Ryan Jones[1], Martin Stumpe[1]
[1]*Tempus Labs, Chicago, IL*

Disclosures: Rohan Joshi: *Employee*, Tempus Labs; Andrew Kruger: *Employee*, Tempus Labs; Elle Moore: *Employee*, Tempus Labs; *Employee*, Tempus Labs; Ryan Jones: *Employee*, Tempus Labs; Martin Stumpe: None

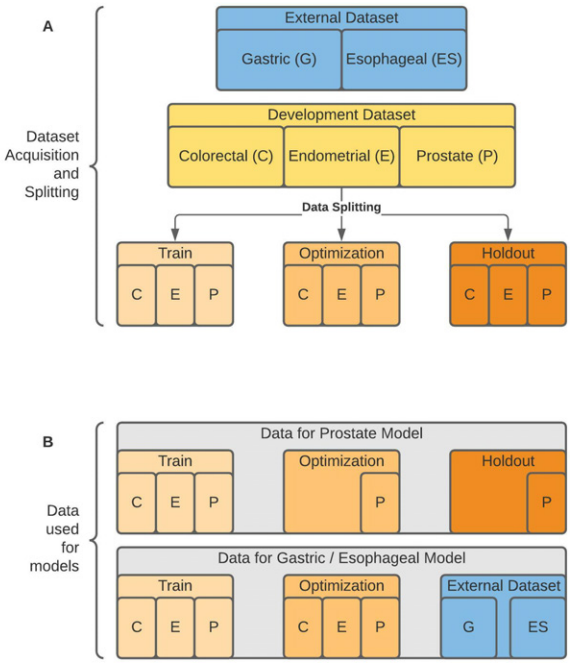
Background: Defective mismatch repair (dMMR) proteins and high microsatellite instability (MSI-H) are associated with a positive response to checkpoint inhibitor therapy in colorectal and non-colorectal cancers. Because of low prevalence in certain non-colorectal cancers, testing for dMMR and/or MSI-H is not routinely performed, particularly at the time of initial biopsy. Assaying would be more practical if a method existed to enrich for likely-positive patients. Here, we tested the ability to predict MSI status from H&E whole slide images (WSIs) in prostate cancer and to generalize MSI prediction to gastric and esophageal cancers.

Design: WSIs and MSI-H—or microsatellite stable (MSS)—labels (obtained by next-generation sequencing) were collected from primary or metastatic colorectal, endometrial, and prostate cancer specimens (N=3296, Figure 1). An attention-based convolutional neural network was trained to predict the MSI label from each WSI in the training set. Hyperparameters and operating points were selected using the optimization set of data, targeting high sensitivity for prostate cancer prediction (prostate cancer model) or simultaneous prostate, endometrial, and colorectal cancer prediction (gastric/esophageal cancer model). Results are reported on a fully-independent holdout set (prostate cancer model) or independently collected datasets (gastric and esophageal cancer model).

Results: We trained two different models and assessed accuracy across three datasets, finding comparable sensitivity/specificity between models trained and tested on the same cancer type vs different types (Table 1). Using estimated real-world prevalences of 5%, 2%, and 20% of MSI-H (in prostate, esophageal, and gastric cancers, respectively) and a positive MSI-H prediction from the models, we expect that 15%, 6%, and 32% (respectively) of patients would have detectable MSI-H status on follow-up testing after the model result.

Prostate Cancer	**True Status**		
	MSS	**MSI-H**	**Total**
Predicted MSS	193	0	193
Predicted MSI-H	81	8	89
Total	274	8	
Summary Metrics	Sensitivity = 100% (95% CI: 63–100%) Specificity = 70% (95% CI: 65–76%) Positive Likelihood Ratio = 3.4 (95% CI: 2.8–4.0)		
Esophageal Cancer	**True Status**		**Total**
	MSS	**MSI-H**	
Predicted MSS	423	3	426
Predicted MSI-H	152	12	164
Total	575	15	
Summary Metrics	Sensitivity = 80% (95% CI: 52–96%) Specificity = 74% (95% CI: 70–77%) Positive Likelihood Ratio = 3.0 (95% CI: 2.3–4.0)		
Gastric Cancer	**True Status**		
	MSS	**MSI-H**	**Total**
Predicted MSS	185	5	190
Predicted MSI-H	135	21	156
Total	320	26	
Summary Metrics	Sensitivity = 81% (95% CI: 61–93%) Specificity = 58% (95% CI: 52–63%) Positive Likelihood Ratio = 1.9 (95% CI: 1.5–2.4)		

Figure 1 - 992



Conclusions: H&E WSIs contain features that can predict the MSI-H phenotype in prostate, esophageal, and gastric cancers. A deep learning model showed discriminative ability on esophageal and gastric cancers despite being trained on other cancers, suggesting generalization across cancer types. Enrichment of patients likely to be MSI-H using this approach could make MSI or dMMR testing more feasible for routine use in these cancers. Future work includes expansion of the study to the institution's entire set of data and examination of potential model biases.

993 Drivers of Digital Pathology Adoption for Clinical Practice: An Experience from a Single Large Academic Institution

David Kellough[1], Saba Shafi[2], Nada Shaker[3], Rachel Scarl[3], Swati Satturwar[1], Giovanni Lujan[3], Wendy Frankel[2], Anil Parwani[3], Zaibo Li[2]

[1]The Ohio State University Wexner Medical Center/James Cancer Hospital, Columbus, OH, [2]The Ohio State University Wexner Medical Center, Columbus, OH, [3]The Ohio State University, Columbus, OH

Disclosures: David Kellough: *Employee*, Inspirata, Inc; Saba Shafi: None; Nada Shaker: None; Rachel Scarl: None; Swati Satturwar: None; Giovanni Lujan: None; Wendy Frankel: None; Anil Parwani: None; Zaibo Li: None

Background: Digital pathology has enormous potential to make routine pathology practice more efficient and accurate, however, full adoption has been slow. We aimed to identify driving factors that encourage pathologists to adopt digital pathology for their daily practice at our institution.

Design: We have collected data on four indicators of pathologist adoption since the implementation of digital pathology: (1) number of pathologists receiving training and certification for primary diagnosis using digital whole slide images (WSI); (2) average daily number of users logged in imaging managing system (IMS); (3) average daily number of primary diagnosis slides scanned; and (4) average daily number of slides scanned for immediate pathologist use (including consultation, urgent cases, etc.) Since adoption of digital workflow was voluntary and slides were only scanned for pathologists who have indicated to use WSIs for routine practice, these are accurate indicators of pathologists' transition from glass slides to digital workflow. These data were correlated with potential events during the study period.

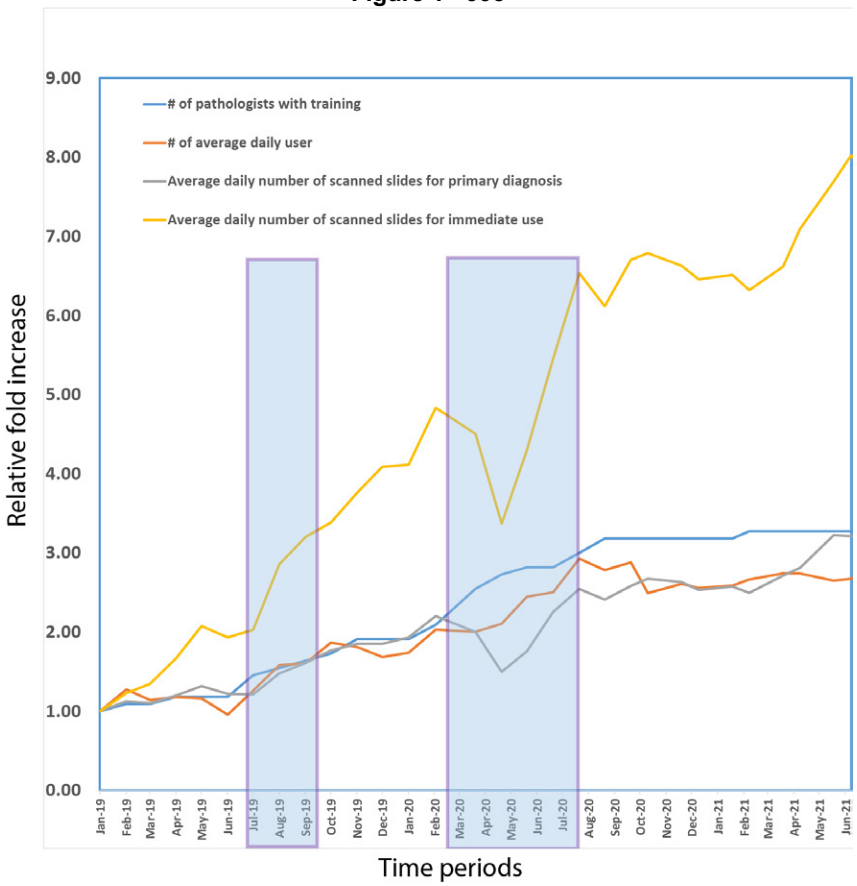
Results: The data of four indicators were summarized in the table. We observed two spikes: the first one was from July to September 2019 and the second was from March to May 2020 (Figure 1). The first spike correlates with our pathology laboratory

information system (LIS) transition from Sunquest Copath to Epic Beaker, which enables single-click access to WSIs in IMS from case working drafts. Previously, pathologists had to switch from pathology LIS to IMS and type in case numbers in order to access WSIs. The second spike correlates with the beginning of COVID-19 pandemic when many academic activities transitioned from live to remote using digital platforms. The need to work remotely, conduct education and consultation at distance, and minimize interaction with others appears to have driven many fence-sitting pathologists to adopt digital pathology. CMS waiver to loosen regulatory requirements during this pandemic has hastened pathologists' decision to switch to digital pathology for primary diagnosis.

- The suspension of elective surgeries due to COVID-19 in March and April 2020 resulted in a decrease of scanned slides before rapid increase after resumption of elective surgeries in May and June 2020.

Month	# of pathologists with training	# of average daily user	Average daily number of scanned slides for primary diagnosis	Average daily number of scanned slides for immediate use
Jan-19	11	11.9	813	258
Feb-19	12	15.2	915	317
Mar-19	12	13.6	895	347
Apr-19	13	14.1	974	431
May-19	13	13.8	1070	535
Jun-19	13	11.4	990	499
Jul-19	16	15.0	985	523
Aug-19	17	18.9	1201	737
Sep-19	18	19.1	1309	826
Oct-19	19	22.3	1437	873
Nov-19	21	21.6	1505	970
Dec-19	21	20.1	1505	1055
Jan-20	21	20.7	1570	1061
Feb-20	23	24.2	1791	1247
Mar-20	28	23.9	1627	1162
Apr-20	30	25.1	1217	870
May-20	31	29.2	1428	1109
Jun-20	31	29.8	1833	1407
Jul-20	33	34.9	2068	1686
Aug-20	35	33.2	1959	1578
Sep-20	35	34.3	2100	1729
Oct-20	35	29.7	2173	1751
Nov-20	35	31.1	2138	1710
Dec-20	35	30.5	2058	1666
Jan-21	35	30.8	2093	1680
Feb-21	36	31.8	2029	1631
Mar-21	36	32.7	2205	1707
Apr-21	36	32.6	2286	1831
May-21	36	31.6	2622	1985
Jun-21	36	31.9	2610	2068

Figure 1 - 993



Conclusions: Our data suggests that ease of use and the ability to work remotely are the most powerful drivers of digital pathology adoption.

994 Quantification of Ki67 in Breast Cancer Using Image Analysis Software: A Validation Study

Melissa Krystel-Whittemore[1], Hannah Wen[1], Dilip Giri[1], Maria Kuba[1], Edi Brogi[1], Matthew Hanna[1]
[1]Memorial Sloan Kettering Cancer Center, New York, NY

Disclosures: Melissa Krystel-Whittemore: None; Hannah Wen: None; Dilip Giri: None; Maria Kuba: None; Edi Brogi: None; Matthew Hanna: None

Background: Ki67 is a protein involved in mitosis and can be detected by immunohistochemistry (IHC) as a surrogate marker of proliferation. Ki67-index has been evaluated as a prognostic marker in ER+ breast cancer as well as a potential surrogate for gene panel testing. Recent neoadjuvant endocrine therapy trials incorporated Ki67 measurement to monitor response. Analytical challenges with Ki67 evaluation include intra/interobserver variability and are time-consuming. Image analysis (IA) has become available for automated quantification and may assist pathologists with Ki67 assessment. The aim of the study was to clinically validate IA analysis of Ki67 in breast cancer.

Design: Whole tissue sections of surgical resection from 57 consecutive invasive breast carcinomas were included in the study. Ki67 IHC was performed using two antibody clones (MIB1 and 30-9). A Ki67 hotspot 400x field of view (FOV) was captured for each case. One pathologist manually counted Ki67 positive/negative tumor cells in each FOV as ground truth. Tumor cells with nuclear staining at any intensity were considered positive based on International Ki67 Working Group Guidelines. Two other pathologists independently estimated Ki67 proliferation of the same FOV microscopically. Three nuclear IA tools were sequentially optimized for maximal performance across all slides and quantified Ki67 expression in the same FOV (Figure 1). Time to quantify each FOV was recorded.

Results: 57 invasive breast carcinomas (Nottingham G1=12, G2=37, G3=8) from our institution were stained with each Ki67 antibody (114 hotspot FOVs). Median time to score each FOV was 360 (Pathologist 1/direct count), 27 (Pathologist 2), 21 (Pathologist 3), 3 (IA1), 8 (IA2), and 5 (IA3) seconds. Using a ≥20% cut-off for Ki67 high versus low according to the MonarchE trial, interrater-Kappa agreement of Ki67 scores between all pathologists was 0.86 and 0.74, for MIB and 30-9, respectively. Compared to the reference standard IA1 showed highest K for MIB (0.856) and IA3 showed highest K for 30-9 (0.816). Kappa agreement of Ki67 scores across all pathologists and image analysis software was 0.79 (MIB) and 0.76 (30-9).

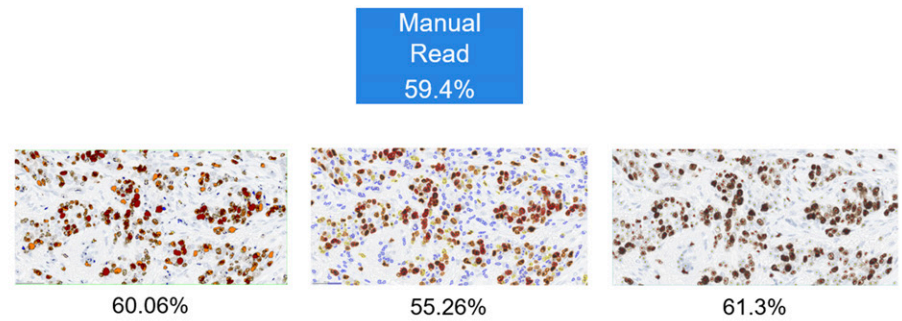
Table 1. Correlation between concordance of Ki67 quantification with 20% cutoff

		P2	P3	IA1	IA2	IA3
MIB	**P1**	0.89	0.89	0.86	0.79	0.82
	P2	X	0.81	0.71	0.86	0.85
	P3	X	x	0.79	0.76	0.71
	IA1	X	x	x	0.83	0.71
	IA2	X	x	x	X	0.66
30-9	**P1**	0.78	0.82	0.81	0.76	0.82
	P2	x	0.73	0.96	0.74	0.63
	P3	x	x	0.76	0.78	0.73
	IA1	x	x	x	0.80	0.66
	IA2	x	x	x	x	0.68

P, Pathologist; IA, Image analysis

Figure 1 - 994

Figure 1. Sample Image Output with Three IA Tools



Conclusions: There is substantial agreement between pathologists as well as between pathologists and image analysis tools for evaluating Ki67 IHC. IA offers an objective and reliable alternative to manual counting, decreases time to quantify Ki67 significantly, and could be a useful tool to implement Ki67 analysis in the clinical workflow.

995 Artificial Intelligence (AI)-Assisted Ki-67 Scoring Offers Improved Correlation with 5-Year Survival in Breast Cancer

Anna Laury[1], Nelli Sjoblom, Marja Kovala[2], Juuso Juhila[3]
[1]University of Helsinki, Helsinki, Finland, [2]Helsinki University Hospital, [3]Aiforia, Inc, Helsinki, Finland

Disclosures: Anna Laury: *Consultant*, Aiforia, Oyj; Nelli Sjoblom: None; Marja Kovala: *Consultant*, Aiforia Technologies; *Employee*, Helsinki University Hospital, Diagnostic center; *Grant or Research Support*, Helsinki University Hospital, Diagnostic center; Juuso Juhila: None

Background: The use of immunohistochemical stains in the prognostication of invasive breast cancer is well established. One stain that has received much attention is Ki-67; though widely used, the variability of interpretation and staining has been sufficient to preclude a recommendation for routine inclusion. Recently, the International Ki-67 Working Group recommended manual scoring of Ki-67 within limited patient groups. Multiple publications have shown that digital image analysis can produce accurate and reproducible results, however, the relationship to prognosis has been less well documented.

Design: Scanned whole slide images (WSI) of diagnostic H&E and MIB-1 stained slides from invasive breast cancer (ductal and lobular) from 131 patients were obtained from a biobank, along with associated clinicopathologic information. The slides were uploaded to Aiforia's cloud-based image analysis platform and independently reviewed by three pathologists. Each MIB-1 WSI was evaluated and the expression in invasive tumor was assessed on a scale from 0-100%, in 5% increments. Following a 2-week washout period the same slides were assessed by the same pathologists, this time with AI-assisted scoring. The AI-assistant provided a suggested score (% positive tumor cells) and visualization of the class assignment (positive or negative) for each cell; the pathologists reviewed the AI suggestion before assigning a score.

Results: The clinical performance of AI-assisted WSI assessment was very good as compared to manual scoring (sensitivity 100%, specificity >80% and accuracy >90%), showed excellent correlation with the manual scoring (>95%), and improved diagnostic concordance by decreasing diagnostic variation between the pathologists. Of particular interest, the AI-assisted score correlated better with the survival data, as compared to the manual score (F-values of 11.62 vs 8.04, respectively).

Conclusions: AI-assisted WSI Ki-67 scoring offers a tool to improve clinical prognostication in patients with invasive breast carcinoma.

996 Five Critical Genes Linked to the Risk of Hepatocellular Carcinoma: A Data Science Discovery

Yongjun Liu[1], Zhengjun Zhang[2]
[1]University of Wisconsin School of Medicine and Public Health, Madison, WI, [2]University of Wisconsin, Madison, WI

Disclosures: Yongjun Liu: None; Zhengjun Zhang: None

Background: Dysregulation of expression of functional genes plays critical roles in the etiology and progression of hepatocellular carcinoma (HCC). The published transcriptomic studies have shed light on various aspects of the molecular mechanisms of HCC. However, the number of the reported genes is large and their sensitivity/specificity in predicting the risk of HCC are unknown. Moreover, whether those genes have causal effects is controversial. The limitation is partially due to the inefficient detecting power of the existing analytical methods and the uninterpretable features of gene-gene interactions of the existing models.

Design: This study was aimed to apply competing linear factor analysis methods in differentially expressed gene detection to identify critical genes for HCC. Our approach takes advantage of the most recently developed machine learning methods of max-linear competing factor models, max-linear regression models, and max-linear logistic models. The dataset used was from The Cancer Genome Atlas (TCGA) with publicly available clinical and RNA-seq data of HCC samples (n=374) and controls (n=52) and the AJCC staging system. The goal was to select a sparse (single digit) number of genes with the highest performance. We started with three competing factors in max-linear logistic regression models. A Monte Carlo method with extensive computation was used to find the final model with the best performance of sensitivity and specificity and the smallest number of genes.

Results: Using a probability >50% as the threshold, we identified five critical differentially expressed genes: CCDC107, CXCL12, GIGYF1, GMNN, and IFFO1. These genes had 100% sensitivity and 100% specificity in classifying HCC and could classify the 374 HCC patients into 6 groups with different features of histologic grades and tumor staging (Figures 1 and 2). Interestingly, four of the five identified genes (except GIGYF1) showed evidence of replication (FDR <0.10) in our independent RNA-seq dataset which included 17 HCC and paired non-HCC samples. We also found GIGYF1 interacted with the combinations of GMNN/IFFO1, CCDC107/IFFO1, and CXCL12/GMNN, separately and respectively.

Figure 1 - 996

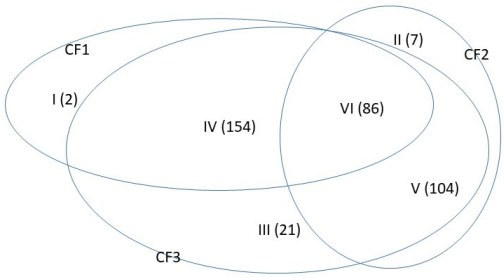
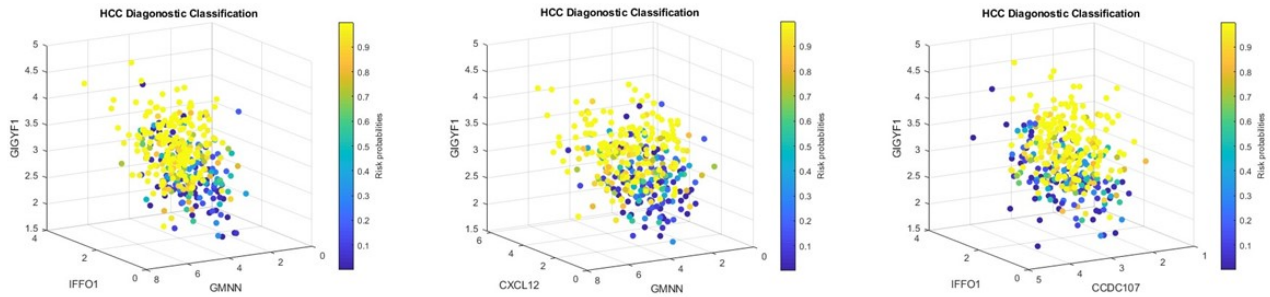


Figure 2 – 996



Conclusions: The inference/analysis approach used in this study is novel, and the five identified genes can be a starting point to conduct further gene network analysis, test other reported genes, find the causal directions of gene expression for HCC, and conduct prognostic and survival analysis.

997 The Use of Deep Machine Learning for Individual Cell Type Classification in Clinical Flow Cytometry Data

Joseph Lownik[1], Serhan Alkan[2], Qin Huang[1], Sumire Kitahara[1]
[1]Cedars-Sinai Medical Center, Los Angeles, CA, [2]Cedars-Sinai Medical Center, Beverly Hills, CA

Disclosures: Joseph Lownik: None; Serhan Alkan: None; Qin Huang: None; Sumire Kitahara: None

Background: While advances have been made in both supervised and unsupervised statistical methods for population clustering and gating, these methods rarely give informative cell classification for each cluster. Additionally, most unsupervised statistical methods for population clustering require new clustering iterations for each new sample acquired, often resulting in different and non-reproducible results, making their utility in clinical diagnostics difficult. In this study, we utilize deep machine learning to accurately classify cell populations using a training set with hematopathologist-determined cell classifications.

Design: 100 peripheral blood samples with 10-color flow cytometric data from 3 panels (B-cell, T-cell, myeloid) showing no evidence of abnormal cell populations were utilized for this study. Flow cytometry files were initially cleaned to exclude doublets and debris, then were bulk analyzed using unsupervised clustering methods to characterize every individual event, followed by cell type classification of each cluster based on immunophenotype by a hematopathologist. These classifications were then used in R statistical software to develop and train deep machine learning algorithms to predict cell types based on raw fluorescence data. Performance of each model was evaluated based on sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for each cell population.

Results: The different machine learning models tested varied in their performance for sensitivity, specificity, PPV and NPV. Overall, the Random Forest model had the greatest overall accuracy compared to basic decision trees, neural networks and k-nearest neighbor models. Our findings suggest that populations showing heterogenous antigen expression have slightly lower

sensitivity and specificity and are often classified as 2 or more individual populations. Overall accuracy for the Random Forest model was >99%.

Table 1: Machine learning model results for cell type prediction.

	Basic Decision Tree				Random Forest				Neural Net				k-Nearest Neighbors			
	Sens.	Spec.	PPV	NPV	Sens.	Spec.	PPV	NPV	Sens.	Spec.	PPV	NPV	Sens.	Spec.	PPV	NPV
NK Cells	0.952	0.997	0.917	0.998	0.981	1	1	0.999	0.952	1	1	0.998	0.981	1	1	0.999
CD8 T Cells	0.911	0.996	0.917	0.995	0.993	0.999	0.975	0.999	0.975	0.999	0.981	0.999	0.987	0.998	0.963	0.999
DN T Cells	0.923	0.999	0.923	0.999	0.923	1	1	0.999	0.846	1	1	0.999	0.845	0.999	0.917	0.999
CD4 T cells	0.938	0.994	0.931	0.994	0.981	0.999	0.981	0.999	0.962	0.998	0.973	0.996	0.984	0.997	0.969	0.998
CD7^{lo} CD4 T cells	0.718	0.997	0.836	0.994	0.922	0.997	0.868	0.998	0.953	0.996	0.835	0.996	0.938	0.997	0.882	0.999
Kappa^{+} B cells	0.877	0.998	0.893	0.998	1	1	1	1	0.93	0.998	0.883	0.999	1	0.999	0.95	1
Lambda^{+} B cells	0.8	0.999	0.842	0.998	1	0.999	0.952	1	0.85	0.999	0.944	0.999	0.95	0.999	0.95	0.999
Neutrophils	0.963	1	1	0.998	0.985	1	1	0.999	0.993	1	1	0.999	0.985	1	1	0.999
Immature Granulocytes	0.951	0.998	0.939	0.999	0.988	0.999	0.976	0.999	1	0.999	0.988	1	1	0.999	0.976	1
Classical Monocytes	0.982	0.999	0.991	0.999	1	0.999	0.996	1	0.987	0.999	0.995	0.998	0.996	0.998	0.978	0.999
Intermediate Monocytes	0.781	0.999	0.892	0.997	0.875	1	1	0.999	0.969	0.999	0.886	0.999	0.625	1	1	0.996

Figure 1 - 997

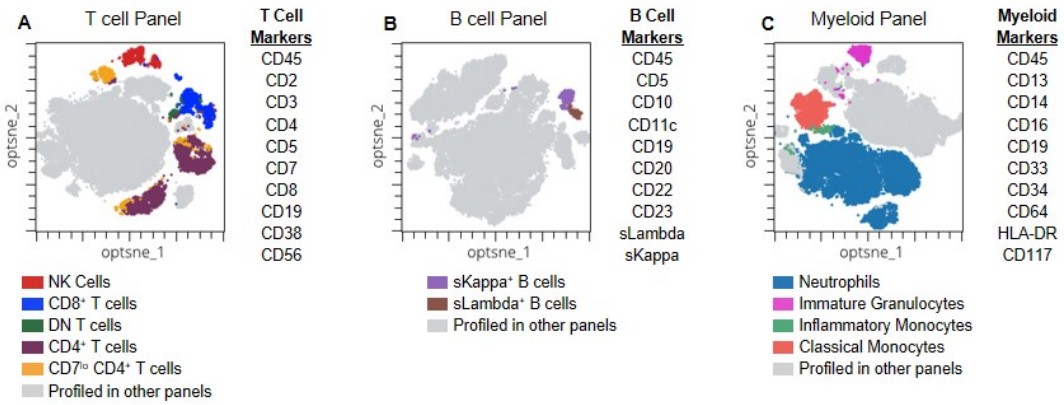


Figure 1: Cell population representations. Each indicated color represents a different cell population as clustered on a dimensionality reduction plot. (A) Populations identified using the indicated T cell panel. (B) Populations identified using the indicated B cell panel. (C) Populations identified using the indicated Myeloid panel. Note that all light grey populations on the dimensionality reduction plots indicated clustered cells which are not adequately characterized in that individual panel, but are examined in a different panel.

Figure 2 – 997

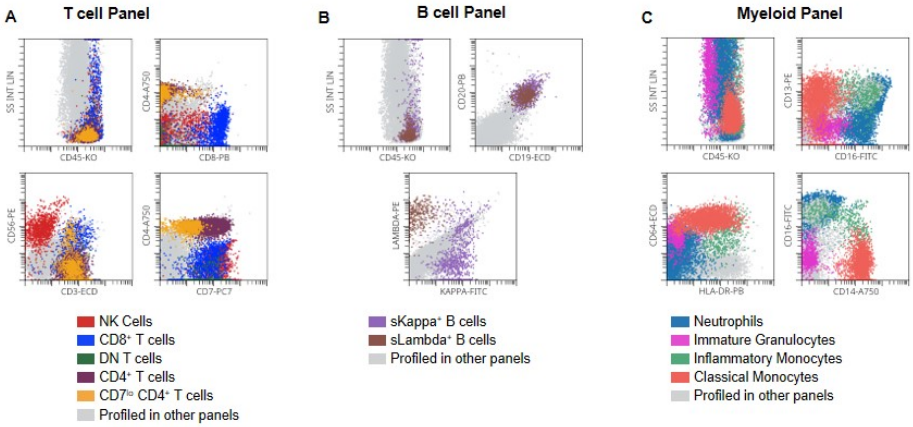


Figure 2: Cell population verification. Each indicated color represents a different cell population as clustered on a 2-Dimensional scatter plot. (A) Populations identified using the indicated T cell panel. (B) Populations identified using the indicated B cell panel. (C) Populations identified using the indicated Myeloid panel. Note that all light grey populations on the dimensionality reduction plots indicated clustered cells which are not adequately characterized in that individual panel, but are examined in a different panel.

Conclusions: Our study shows the feasibility of the application of machine learning prediction to cell type identification in flow cytometry studies. Each training set is lab-specific, however, the algorithm can be utilized in any laboratory. In addition to having an accuracy >99% in cell prediction for basic peripheral blood flow cytometry using a 3-tube, 10-color panel, prediction takes just a few seconds compared to the more intense time required by technicians or pathologists for gating.

998 Faster Than Glass: A Digital Pathology Workflow Unlocks Major Time-Savings

Giovanni Lujan[1], David Kellough[2], Wendy Frankel[3], Namrata Chhibber[4], Anil Parwani[1], Wei Chen[3], Zaibo Li[3], Audrey Lombardo[5], Jason Walsh[5], Mark Lloyd[4]

[1]The Ohio State University, Columbus, OH, [2]The Ohio State University Wexner Medical Center/James Cancer Hospital, Columbus, OH, [3]The Ohio State University Wexner Medical Center, Columbus, OH, [4]Inspirata, Inc., Tampa, FL, [5]The Ohio State University Medical Center/James Cancer Hospital, Columbus, OH

Disclosures: Giovanni Lujan: None; David Kellough: *Employee*, Inspirata, Inc; Wendy Frankel: None; Namrata Chhibber: None; Anil Parwani: None; Wei Chen: None; Zaibo Li: None; Audrey Lombardo: None; Jason Walsh: None; Mark Lloyd: *Employee*, Inspirata, Inc

Background: Speed is an issue for digital pathology adoption. Digital pathology affords advantages but scanning slides takes time and adds a layer between the slides and the pathologist. Is it worth it? As technicians and practitioners in the United States' most fully digital pathology practice, our experience suggests that-- even with scanning--pathologists are receiving images for diagnosis faster than in a traditional workflow. The elimination of paperwork, case assembly, and transport seems to more than offset scanning time. To test this, we measured and compared time elapsed from slide creation to receipt by pathologist with and without digital. We also examined time involved in acquisition of slides for prior cases.

Design: We explored time from slide creation to availability for pathologist review for 40 cases. Using timestamps on the slides as they moved to the pathologist's desk (or monitor), we noted slide creation time, whole slide image (WSI) creation time, and receipt by pathologist. Cases were selected from throughout the day to minimize bias from slides being processed at busy or slow times. Time for scanning was subtracted from the total to replicate a traditional workflow and left in for a digital workflow.

For priors, 25 total cases were selected from 2011 to 2020 to reflect storage locations of slides by age. Requests were made for these cases from our physical archive and cases were searched in our digital archive. Time from request to receipt was noted as well as completeness.

Results: The mean time for glass slides to get from slide foldering to pathologist desk was 1.99 hours. The mean time for current cases to get from slide foldering to WSIs ready for diagnosis was 0.94 hours, of which actual scanning was just 0.5 hours. It should

be noted that scanning time was calculated until the last slide finished but WSIs populate as soon as they scan so availability for diagnosis is actually even quicker.

Priors, on average, took 15.2 hours to reach the pathologist's desk from archives. Age of the case did not make a significant difference. We received 92% of the cases and 91% of the slides. WSIs, on the other hand, were instantly available in all cases and 100% of the cases and slides were present.

Table 1. Summary of Findings

	DIGITAL- Time to WSIs ready for pathologist review (hours)	TRADITIONAL- Time to slides on pathologist's desk, ready for review (hours)	Time Savings from Digital Workflow (hours)
Current cases	0.94	2.94	1.99
Prior cases	0.00	15.17	15.17

Conclusions: A digital pathology workflow saves time. Concerns about pre-diagnosis time added by scanning ignore the considerable time taken up by paperwork, case assembly, and transport. Digital enables pathologist access to diagnostic images within minutes of slide production.

999 Leveraging the Structural Information in Prostate Biopsies to Improve the Generalisation of Deep Learning Models across Datasets

Darragh Maguire[1], Jenny Fitzgerald[1], Mark Gregson[1], Donal O'Shea[1]
[1]Deciphex Ltd, Dublin, Ireland

Disclosures: Darragh Maguire: *Employee*, Deciphex Ltd; Jenny Fitzgerald: *Speaker*, Deciphex Ltd.; Mark Gregson: *Primary Investigator*, Deciphex Ltd.; Donal O'Shea: *Primary Investigator*, Deciphex Ltd.

Background: Classification of prostate cancer severity in needle biopsies is time-consuming and subject to inter-observer variability. Nonetheless, Gleason/ISUP grading is critical in guiding the clinical management of patients with prostate cancer. Artificial intelligence can be employed to extract information from tissue images in order to enhance pathologists' decision-making, improving on the objectivity and repeatability of clinical diagnoses. For example, deep learning methods have shown potential to grade prostate cancer accurately in whole slide images (WSI) of needle biopsies. Generalisation across biopsies from different labs, with variable staining and scanning techniques, is a major pitfall for many of these diagnostic models - and thus a primary barrier to their adoption. In this research, we present a technique to improve the robustness of models on variable datasets by providing additional structural information pertaining to the tissue inputs.

Design: Segmentation models were trained for pixel-level segmentation of Gleason patterns using prostate needle biopsy WSIs. The models were trained using slides from one dataset, and an additional independent dataset from different sources/scanners was used to assess the generalisability of the models. The models were trained using inputs of 512x512 pixel tissue tiles extracted from the WSIs. Morphological information predictions were combined to create greater context. Stratification of patient survival was correlated using biochemical recurrence as an endpoint from RNA sequencing data. Performance and generalisability of the models were evaluated both with and without this additional context.

Results: Models trained using the tissue input alone were unable to generalise to the data from unseen scanners, resulting in poor stratification across the cancer grades. However, a significant improvement was observed in cancer grading and prognostication when utilising the image-based assessment with molecular data and clinical data.

The additional molecular, clinical and structural context provided, alleviated the difficulty in classification of visually different tissue samples and allowed for prognostic determination. This novel multi-omic approach resulted in an increase in performance of the subsequent Gleason grading model and associated survival analysis prediction.

1000 Multilayer Whole Slide Imaging Improves AI-based Identification of Acid-fast Mycobacteria (AFB) in Ziehl-Neelsen Stained Slides

Karolina Nurzynska[1], Ann Walts[2], Arkadiusz Gertych[2]
[1]Silesian University of Technology, Gliwice, Poland, [2]Cedars-Sinai Medical Center, Los Angeles, CA

Disclosures: Karolina Nurzynska: None; Ann Walts: None; Arkadiusz Gertych: None

Background: Manual screening of Ziehl-Neelsen (ZN) stained slides that are negative or contain rare acid-fast bacteria (AFB) is labor-intensive and time-consuming requiring continuous refocusing at high magnification/oil immersion to visualize and optimally evaluate out-of-focus AFB candidates. New slide scanners that output multilayer whole slide images (WSIs) enable AI-based evaluation of multiple image planes, each focused automatically. We sought to determine if multilayer slide scanning improves upon single-layer CNN-based classification of low-burden AFB WSIs as AFB+ or AFB-.

Design: The P1000 scanner (3DHISTECH, Hungary) generated 3-layer WSIs (40X) of 15 low-burden ZN-stained AFB+ and 5 AFB- slides that had been manually screened for AFB by the pathologist with assessment confirmed by auramine-rhodamine stain and culture results. Each layer was divided into tiles for CNN-classification to produce a heatmap of AFB+ probabilities. Multilayer slide heatmaps were aggregated into one using the average (AVG) or the maximum (MAX) of AFB probability scores across the layers for each tile. The mid-layer heatmap from each WSI served as equivalent to the heatmap from a single-layer scan of that slide. Heatmap features representing AFB+ probability distributions were entered into a logistic regression (LR) model that had been trained to classify WSIs as AFB+ or AFB- using a separate set of 46 AFB+ and 88 AFB- single-layer WSIs. The CNN had been trained on 18,426 AFB+ and 28,913 AFB- tiles. A tile was considered AFB+ if its AFB probability score was >0.5. The balanced accuracy, sensitivity, and specificity were used to compare the classification performances of each multilayer WSI with that of its corresponding single-layer WSI.

Results: The classification accuracy of 3-layer WSIs was 30% higher than that of the single-layer WSIs using the AVG-based method (Table 1). This method increased the specificity by 80% but decreased the sensitivity by 20%. The MAX-based method increased sensitivity, specificity and accuracy by 13%, 20% and 17%, respectively.

Table 1. Performances of the LR model in classifying 3-layer and single (middle) layer in digitized ZN-stained slides as AFB+ or AFB-. Performance was measured for the maximum and the average AFB probability score aggregation method. The highest metrics are bolded.

Number of WSI layers used in AI-based classification	AFB probability score aggregation method across layers	Sensitivity [%]	Specificity [%]	Balanced accuracy [%]
single (middle) layer	n/a*	73.3	20.0	46.7
multilayer (3 layers)	MAX	**86.7**	40.0	63.3
multilayer (3 layers)	AVG	53.3	**100.0**	**76.7**

* n/a = not applicable

Conclusions: AI-based analysis of slides scanned with the multilayer mode yielded higher accuracy and specificity than analysis of the same slides when represented by a single (the middle) layer. Improvement in the classification is attributed to the increased number of layers capturing more in-focus tissue for analysis than the single-layer. Our study illustrates the potential of multilayer slide scanning combined with AI-based analysis.

1001 AI-Based Analysis and Three-Dimensional Visualization Software for Whole-Block Image Toward Clinical Application

Takashi Ohnishi[1], Alexei Teplov[1], Kareem Ibrahim[1], Masao Yoshida[2], Emine Cesmecioglu[1], Yukako Yagi[1]
[1]Memorial Sloan Kettering Cancer Center, New York, NY, [2]Shizuoka Cancer Center, Nagaizumi, Japan

Disclosures: Takashi Ohnishi: None; Alexei Teplov: None; Kareem Ibrahim: None; Masao Yoshida: None; Emine Cesmecioglu: None; Yukako Yagi: None

Background: It is known that reviewing tumors with their surroundings is crucial for pathological evaluation. With the emerging imaging technologies, three-dimensional (3D) visualization and analysis will lead to a histopathological review in the spatial context. According to previous studies, whole-block image (WBI) scanned by micro-computed tomography (micro-CT) allows us to provide a 3D morphology of a neoplasm non-invasively. Segmenting pathologically significant structures like blood vessels from the whole

tissue would be a step forward to using these technologies. To assist 3D analysis on WBI, we have been developing several software. This study demonstrates two of them.

Design: Figure 1 details a schematic of our analytical procedure. In this study, two developed software were implemented as plug-ins for Dragonfly (Ver. 2021.1, Object Research Systems, Quebec, Canada), a commercial 3D visualization and analysis application. First software automatically adjusts window level (WL) of the WBI by histogram analysis and digitally stains like H&E image. Second software performs an AI-based blood vessel segmentation on WBI processed by the first software. For test of the developed software, we randomly selected five WBIs of formalin-fixed paraffin-embedded tissue sample by the endoscopic submucosal dissection from around 1000 WBIs of various tissue samples. All WBIs were scanned by a custom-build micro-CT (Nikon Metrology NV, Leuven, Belgium).

Results: Figure 2 shows three example results. The first software successfully applied H&E-like appearance on the WBI automatically. By using the second software, blood vessels were visualized and superimposed with the digitally stained WBI. If needed, operator can easily adjust WL, color and other visualization parameters and set a region of interest for the blood vessel segmentation by default tools of the Dragonfly. Digital stain and blood vessel segmentation completed in about 5 seconds and 10 minutes using GeForce RTX 2080 Ti, respectively.

Figure 1 - 1001

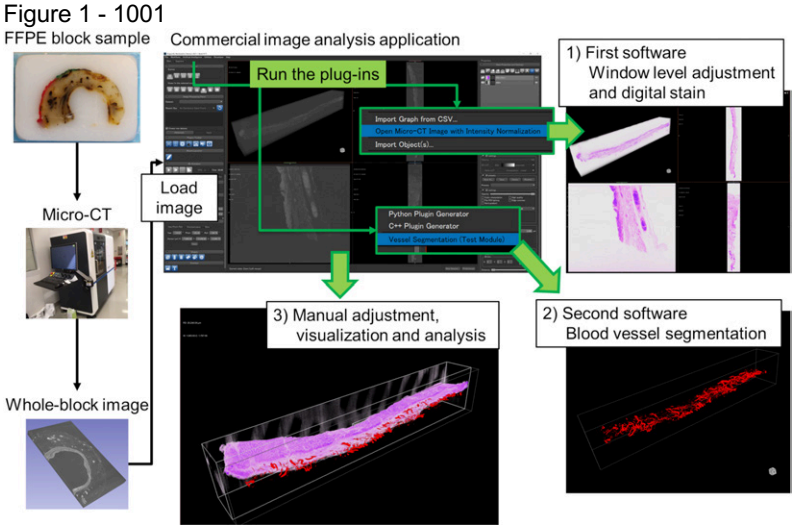
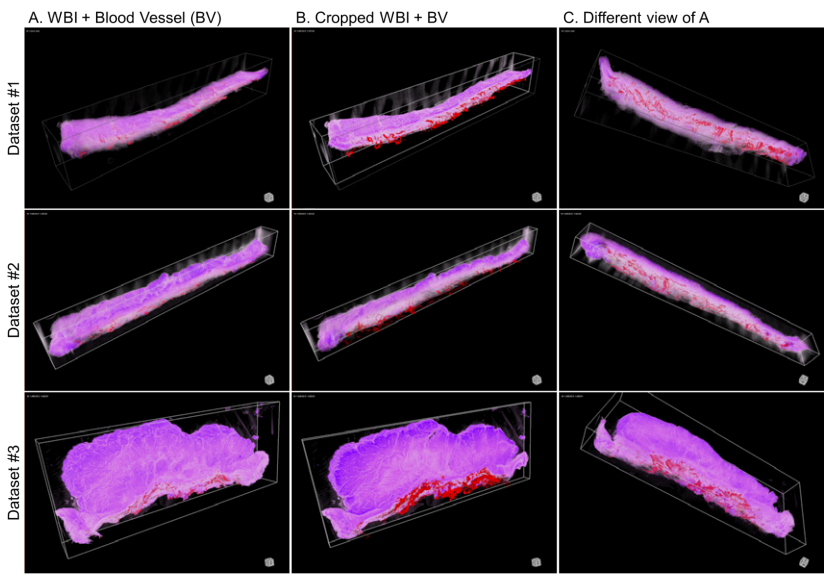


Figure 2 – 1001



*) Red indicates segmented blood vessels, H&E like appearance is applied on WBI.

Conclusions: We have developed image visualization and analysis software for whole-block image. Our software will facilitate three-dimensional reviewing the pathology samples by providing an H&E-like appearance and segmenting the blood vessels. The next step is to expand the segmentation subjects such as lymph node and tumor lesion.

1002 Benchmarking Bioinformatics Approaches for Tumor Mutational Burden Evaluation From a Large Cancer Panel Against Whole Exome Sequencing

Jiuhong Pang[1], Mahesh Mansukhani[2], Helen Fernandes[3], Susan Hsiao[4]
[1]Columbia University Irving Medical Center, New York, NY, [2]Columbia University Medical Center, New York, NY, [3]Columbia University, New York, NY, [4]New York-Presbyterian/Columbia University Medical Center, New York, NY

Disclosures: Jiuhong Pang: *Employee*, Columbia University Medical Center; Mahesh Mansukhani: None; Helen Fernandes: None; Susan Hsiao: *Grant or Research Support*, Bristol Myers Squibb; *Speaker*, Medscape, Illumina; *Consultant*, Opentrons Labworks, Inc; *Advisory Board Member*, Loxo Oncology

Background: Tumor mutational burden (TMB), typically defined as the number of mutations per megabase, is a biomarker used to predict response to immune checkpoint inhibitors, which target the immune checkpoints that regulate T cell mediated immunity. A high TMB is thought to result in a larger number of neoantigens that can be presented and recognized by T cells as “foreign”, thereby increasing the efficacy of immune checkpoint therapy. Although different groups have proposed different calculations for TMB determination, a clear consensus of variant types for inclusion has not yet emerged. In this study we sought to explore approaches to TMB calculation using a custom, in-house developed 586 gene cancer panel benchmarked to the TMB as measured using whole exome sequencing (WES) as the orthogonal method.

Design: Libraries on 30 tumor cases (melanoma, n=3; pancreas, n=4; lung, n=4; renal, n=4; bladder, n=5; biliary tract, n=4; other, n=6) with paired normal samples were prepared using IDT xGen Exome v2 (WES) reagents or custom probes to 586 cancer genes covering 1.57 Mb of coding regions and sequenced on an Illumina NextSeq2000 instrument (average depth 1000X). Data was analyzed using Genome Analysis Toolkit (GATK) Best Practices. TMB as measured by WES was compared to the TMB as calculated by the tumor-only cancer panel (after filtering with a panel of normals and the population database gnomAD). Variables commonly considered for TMB calculation inclusion were evaluated and included nonsynonymous variants, synonymous variants, indels, hotspot variants, and variants in flanking (+/- 10bp) intronic sequence.

Results: TMB as measured by WES ranged from 0.43 to 72.92 mutations/Mb, with 7 samples having >10 mutations / MB. The correlation along with analytical sensitivity and specificity at 10 mutations/Mb between WES TMB and 586 gene panel TMB was determined for each method of TMB calculation, and results are shown in the table below. The Spearman correlation rho values ranged from 0.707 to 0.783, sensitivity ranged from 85.7-100%, and specificity ranged from 87-95.7%, with each method of TMB calculation.

	Method 1 Non-synonymous	Method 2 Non-synonymous Synonymous Exclude hotspot	Method 3 Non-synonymous Indel	Method 4 Non-synonymous Intronic	Method 5 Non-synonymous Indel Intronic
Spearman correlation (rho)	0.787	0.707	0.783	0.776	0.783
Sensitivity, (95% CI)	85.7% (42.1-99.6%)	100% (59.0-100%)	100% (59.0-100%)	100% (59.0-100%)	100% (59.0-100%)
Specificity, (95% CI)	95.7% (78.1-99.9%)	87.0% (66.4-97.2%)	87.0% (66.4-97.2%)	95.7% (78.1-99.9%)	91.3% (72.0-98.9%)

Conclusions: Strong correlation between WES TMB and 586 gene panel TMB was seen regardless of method used to calculate panel TMB, suggesting that this panel is representative of WES TMB. All methods of TMB calculation evaluated showed good sensitivity and specificity using 10 mutations/Mb as a cutoff, suggesting that multiple TMB calculation approaches may yield comparable results.

1003 Comparison and Availability of Artificial Intelligence-Based Commercial Software Solutions for Pathologists

Ankush Patel[1], Nada Shaker[2], Bindu Challa[1], Saba Shafi[1], Anil Parwani[2]
[1]The Ohio State University Wexner Medical Center, Columbus, OH, [2]The Ohio State University, Columbus, OH

Disclosures: Ankush Patel: None; Nada Shaker: None; Bindu Challa: None; Saba Shafi: None; Anil Parwani: None

Background: Commercial software vendors have leveraged artificial intelligence (AI) techniques and advanced algorithms to develop solutions for pathologists who navigate a progressively digitized departmental landscape. Such tools are primarily purposed for applications in image analysis, image viewing, and workflow. Though they have demonstrated faculty as adjunctive devices for enhanced diagnostic utility, implementation has been stymied due to many factors including standardization, validation, regulatory, and data transparency, e.g. "black box" concerns arising from convolutional neural networks forged from deep and machine learning techniques. We seek to discern the efforts commercial vendors have implemented towards addressing such challenges, the utilities these solutions offer, and their efficacy as aids for the modern pathologist.

Design: 35 commercial software vendors offering platforms designed to assist pathologists in workflow, image analysis, and image viewing applications were identified. A qualitative literature review was conducted using multiple databases, e.g. PubMed, Medline, employing search terms to identify artificial intelligence techniques, e.g. machine learning; deep learning; neural networks, utilized in commercial software solutions. Text analysis was conducted on relevant literature for application of software solutions within academic, clinical, and research forums in pathology.

Results: Results demonstrated an improved approach to diagnostic challenges within computational pathology, e.g. cytological applications and analysis of regions presenting complex histology and diverse morphology, e.g. detection, grading, and quantification of tissue areas suspicious for prostate carcinoma in whole slide images (WSI)s of needle biopsy tissue sections. Movement towards user-friendly compatibility with existing digital pathology laboratory implementations, e.g. WSI scanning devices, monitors, LIS) was demonstrated, as well as increased efforts towards achieving generalizability. Some solutions have now been granted 510(k) premarket FDA approval, with many still retaining RUO status in the US. Validation studies have increased markedly from the three year period spanning 2019-2021, with results demonstrating improvement in sensitivity and specificity of disease detection throughout multiple subspecialties.

Conclusions: Though many improvements remain upon the horizon, commercial developers have began to effectively address the needs of pathologists.

1004 Live Telemicroscopy is Substantially Equivalent to In-person Intraoperative Frozen Section Diagnosis

Joseph Rohr, University of Nebraska Medical Center, Omaha, NE

Disclosures: Joseph Rohr: None

Background: Intraoperative diagnosis by frozen section is a mainstay of surgical pathologic practice. Access to surgical pathology with an appropriate turn-around time (TAT) has been a limiting factor for small or remote surgical centers with negative impacts on cost and patient care. As a major regional referral center for pathology services, my institution has utilized remote telemicroscopy to provide intraoperative diagnoses to satellite hospitals. A pathology assistant or high-level technician cut the tissue and loaded the slides into the telemicroscope, and the pathologist at the main site provided the diagnoses. I sought to compare this mechanism to our standard intraoperative diagnoses.

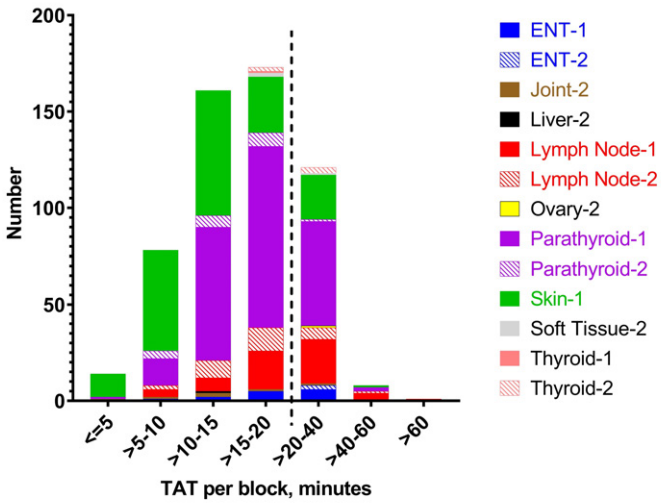
Design: All frozen section diagnoses in a four-year period were queried in our laboratory information system for surgical center, anatomic site of frozen section, intraoperative organ site, TAT, reading pathologist, and concordance with final (paraffin) diagnosis. Intraoperative diagnoses performed by telemicroscopy were compared to intraoperative diagnoses on glass slides. Surgeries performed at one institution and read elsewhere, and intraoperative reads for adequacy only, were excluded. Comparisons were performed by ANOVA and Student's t-tests as appropriate.

Results: A total of 609 cases comprised of 1060 parts and 1496 cryoblocks had intraoperative diagnoses by telemicroscopy at two satellite surgical centers. Ten major anatomic sites were represented (Table 1), the majority of which were parathyroid (261 [42.9%]) and skin (214 [35.1%]). TAT was available for 556 cases (91.3%; Figure 1). The average TAT per block was 17m03s±8m03s. Soft tissue sites had the highest average TAT per block (27m20s). In two cases (0.33%), computer technical

issues prevented diagnosis. Nineteen cases (3.13%) were discordant, including three Category C (15.8%), distributed among 23 reading pathologists with various subspecialties. In this same period, the average TAT per block for in-person reads was 15m25s±8m08s, and the total discordance rate was 2.94%, of which 12.2% were Category C. None of these differences were significant (p>>0.05).

	Location 1			Location 2			Combined		
Anatomic site	**Number**	***n* with TAT**	**TAT**	**Number**	***n* with TAT**	**TAT**	**Number**	***n* with TAT**	**TAT**
Parathyroid	243	234	18:01	18	18	15:07	261	252	17:49
Skin	213	182	13:08	1	0	N/A	214	182	13:08
Lymph node	66	59	23:14	31	30	19:15	97	89	21:53
ENT	15	14	20:04	2	2	24:30	17	16	20:37
Thyroid	1	1	18:00	6	5	22:00	7	6	21:20
Joint	0	0	N/A	6	6	13:02	6	6	13:02
Soft tissue	0	0	N/A	4	3	27:20	4	3	27:20
Liver	0	0	N/A	1	1	13:00	1	1	13:00
Brain	0	0	N/A	1	0	N/A	1	0	N/A
Ovary	0	0	N/A	1	1	23:00	1	1	23:00
sum/average	538	490	16:53	71	66	18:15	609	556	17:03

Figure 1 - 1004



Conclusions: Diagnoses by telemicroscopy did not significantly differ from in-person diagnoses on TAT and concordance as performed by varied pathology staff. Rare cases of technological issues prevented telemicroscopic diagnosis. This provides further justification for continued use and expansion of telemicroscopic services in primary intraoperative diagnoses.

1005 Using a 3D Printed Microscope For Assessing Liquid Based Cervical Cytology Specimens

Daniel Rosen[1], Farinaz Arbab[2], Nisha Ramani[3]
[1]Michael E. DeBakey VA Medical Center, Houston, TX, [2]Baylor College of Medicine/University of Houston, Houston, TX, [3]The University of Texas MD Anderson Cancer Center, Houston, TX

Disclosures: Daniel Rosen: None; Farinaz Arbab: None; Nisha Ramani: None

Background: Currently, there are several different digital pathology platforms available including whole slide image scanners, cloud services, and even mobile cellular devices. However, equipment cost, operating costs, IT infrastructure, and network limitations are some barriers that impede telepathology collaboration. The purpose of this study is to evaluate the diagnostic concordance for assessing cervical cytology smears using the low-cost automated high-resolution OpenFlexure 3D printed microscope.

Design: A total of 60 liquid based cervical cytology glass slide smears were reviewed by 3 board certified pathologists and scored as NILM, ASCUS, LSIL or HSIL. Whole slide images were generated using the 3D printed low-cost automated high-resolution

OpenFlexure 3D microscope (figure 1). Images were captured with an 8-megapixel raspberry pi camera on the OpenFlexure microscope and stitched to generate a whole slide image. After a washout period of at least 2 weeks the 3 observers reviewed the digital images blinded to each other and the previous glass slide readings. Cohen's Kappa coefficient was used to calculate intraobserver agreement. In addition concordance was investigated by grouping the cases as NILM or >ASCUS (including ASCUS, LSIL and HSIL).

Results: The maximum scanned area was up to 60% of the sample (figure 2). The reported concordance for each observer for NILM, ASCUS, LSIL and HSIL was 73%, 73% and 67% (Kappa: 0.50, 0.51, 0.44) respectively (table 1). When cases were group ed as NILM versus >ASCUS concordance was 82%, 78% and 82% (Kappa: 0.60, 0.54 and 0.63) respectively. Significant discrepancies between glass slide and digital image interpretation were identified in 8, 12 and 7 cases by each pathologist corresponding to: 2 LSIL cases and 6 ASCUS by pathologist 1, 1 LSIL and 11 ASCUS by observer 2, and 7 ASCUS by observer 3. No discrepancies between HSIL or NILM cases were reported by any of the observers.

DIGITAL SLIDE	GLASS SLIDE			
OBS1	NILM	ASCUS	LSIL	HSIL
NILM	34	6	2	0
ASCUS	2	3	0	1
LSIL	1	2	2	2
HSIL	0	0	0	5
OBS2				
NILM	33	11	1	0
ASCUS	0	2	3	1
LSIL	0	0	2	0
HSIL	0	0	0	7
OBS3				
NILM	28	7	0	0
ASCUS	4	7	6	2
LSIL	0	1	1	0
HSIL	0	0	0	4

Figure 1 - 1005

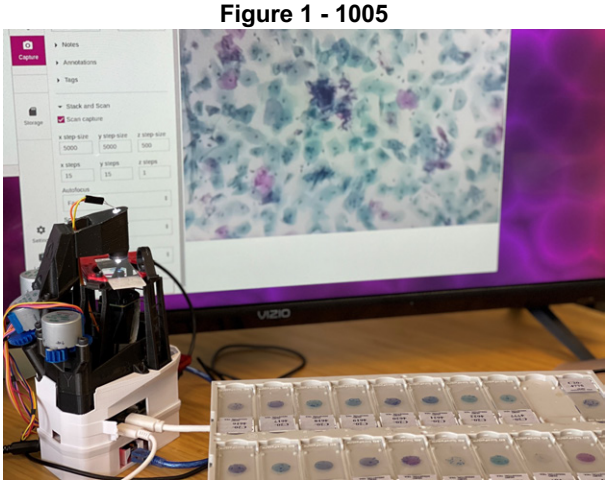
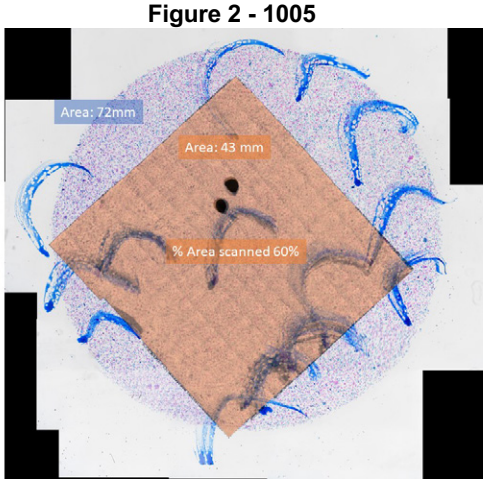


Figure 2 - 1005



Conclusions: The major limitation of this study was the reduced scanned area produced by the microscope precluding examination of the whole specimen. The OpenFlexure microscope offers a viable platform for microscopy imaging providing high quality images for the examination of liquid based cervical cytology specimens. Further modifications to the microscope may be required (such as improving the optics and increasing the sample scan area) to improve image interpretation.

1006 Implementation of an Integrated AI Solution for Breast Cancer Diagnosis and Reporting in Clinical Routine

Judith Sandbank[1], Alona Nudelman[2], Yuval Raz[3], Raz Ziv[3], Ira Krasnitsky[3], Yuval Globerson[3], Nir Benor[3], Gev Decktor[3], Tal Patalon[4], Manuela Vecsler[3], Chaim Linhart[3]
[1]Maccabi Health Services, Kiriat Ono, Israel, [2]Rehovot, Israel, [3]Ibex Medical Analytics, Tel Aviv, Israel, [4]Maccabi Health Services, Tel Aviv, Israel

Disclosures: Judith Sandbank: *Primary Investigator*, Ibex Medical Analytics; Alona Nudelman: None; Yuval Raz: None; Raz Ziv: *Employee*, Ibex Medical Analytics; Ira Krasnitsky: None; Yuval Globerson: *Employee*, Ibex Medical Analytics; Nir Benor: *Employee*, Ibex Medical Analytics; Gev Decktor: *Employee*, Ibex Medical Analytics; Tal Patalon: *Primary Investigator*, KSM, Maccabi Research and Innovation Center, Maccabi Healthcare Services; Manuela Vecsler: *Employee*, Ibex Medical Analytics; Chaim Linhart: *Employee*, Ibex Medical Analytics

Background: There is high demand to develop clinically useful computer-assisted diagnostic solutions to bring about significant efficiency improvements for pathologists, reduce turnaround times, decrease error rates, and provide objective, reproducible, and detailed diagnoses. Successful implementations of digital pathology in a variety of scenarios have started to demonstrate benefits for both research and routine diagnosis, encouraging adoption, however implementation of AI as clinical support solution during diagnosis reporting is yet to be reported.

Aims: To introduce an AI tool to support the pathologists in their review and reporting of breast biopsies in a centralized pathology laboratory.

Design: A two-arm study in which the standard of care arm (using a microscope) was compared with an arm in which pathologist conducted the reporting using an AI solution workflow. Ten pathologists participated in the study and reported on 100 breast biopsies (10 cases/ pathologist). Each case was reported twice, both with microscope and with the AI solution, randomized between pathologists. To assess the effect on accuracy of reporting, discrepant reports were adjudicated and reviewed by a breast specialist pathologist. The study endpoints included accuracy of the AI algorithm on invasive and in situ carcinoma detection and pathologists' satisfaction.

Results: The AI algorithm demonstrated high performance with an AUC as high as 0.99 for the detection of invasive carcinoma, with specificity and sensitivity of 93.8% and 96.4%, respectively as well as high accuracy for the detection of DCIS/ADH with AUC of 0.98 and specificity and sensitivity of 94.3% and 94.9%, respectively. The AI also differentiated between subtypes/grades of invasive and in-situ cancers with an AUC of 0.97 for IDC vs. ILC and AUC of 0.94 for DCIS high grade vs. low grade/ADH, respectively. Moreover, the overall user experience, as reported by pathologists in a user survey, was markedly better with the AI solution compared to a microscope.

Conclusions: This is the first report of a successful implementation of a multi-feature AI solution that automatically reports clinically relevant diagnostic parameters regarding invasive and in situ breast carcinoma, offering an important tool for computer-aided diagnosis in routine pathology practice.

1007 Assessing the Ability of AI To Score Ki-67 in Sarcomas

Sahil Saraf[1], Aahan Singh[2], Logaswari M[3], Jan Sauer[2], Li Yan Khor[4], Sathiyamoorthy Selvarajan[4], Kiat Hon Lim[4], KV Santosh[5], Vani Ravikumar[5], Priyanka Somwanshi[6], Rajasa Jialdasani[2], Kaveh Taghipour[2], Aneesh Sathe[2]
[1]Qritive Pte. Ltd and Singapore General Hospital, Singapore, Singapore, [2]Qritive Pte. Ltd., Singapore, Singapore, [3]Singapore General Hospital, Singapore, Singapore, [4]Singapore General Hospital and Duke-NUS Medical School, Singapore, Singapore, [5]RV Metropolis Lab, Bangalore, India, [6]Medirad Diagnostics, Bangalore, India

Disclosures: Sahil Saraf: *Employee*, Qritive Pte Ltd; Aahan Singh: *Employee*, Qritive; Logaswari M: None; Jan Sauer: *Employee*, Qritive Pte. Ltd.; Li Yan Khor: None; Sathiyamoorthy Selvarajan: None; Kiat Hon Lim: None; KV Santosh: None; Vani Ravikumar: None; Priyanka Somwanshi: None; Rajasa Jialdasani: *Employee*, Qritive Pte. Ltd.; Kaveh Taghipour: *Employee*, Qritive Pte. Ltd.; Aneesh Sathe: *Employee*, Qritive

Background: Quantitative immunohistochemistry (IHC) is an integral part of histopathology reporting, providing vital information such as tumour characterisation and prognosis. Ki-67 proliferation index has traditionally been manually estimated on light microscopy over several fields of view, a laborious task given to interobserver variation. AI can potentially overcome this. We chose to evaluate the performance of an AI assistance tool for grading of sarcomas.

Design: The system design makes use of traditional computer vision algorithms which, unlike machine learning methods, do not require training. The algorithm identifies nuclei in a given field of view and categorizes them as positive or negative based on the intensity of the IHC staining. In this study, the algorithm was evaluated by 3 pathologists independently on 440 regions of interest (ROIs) from 88 Ki-67 stained sarcoma slides. To compare the output of the algorithm, each pathologist was first asked to score every ROI as per their traditional workflow. Next, the pathologist was shown the outputs of the AI (both cell segmentation and final Ki-67 score) and asked to score each ROI again. The discordance between the pathologist's initial impression and the final score after viewing the AI outputs was calculated.

Results: The metrics calculated show that, on average, the pathologists agree with 93.5% of the ROIs scored by the AI (number of ROIs accepted / total number of ROIs in the study). In the cases where the pathologist disagreed with the outputs of the AI, a majority of the differences lie between 0 and 10 percentage points (Figure 1). To quantify inter-pathologist differences, we compared their Ki67 scores before and after they were assisted by the AI outputs. We observed that the root mean squared error (RMSE) between the pathologists themselves was 14.08% on average before they viewed the AI outputs. In contrast, the average RMSE between the pathologists after AI assistance was 2.52 (Table 1). These results show that after AI assistance, the pathologists generally provide more accurate estimates of Ki-67 and discordance was reduced by 82.10%.

Root Mean Square Error

	P1	P2	P3	AI	P1-AI	P2-AI	P3-AI
P1	0.00	18.10	12.01	17.13	17.13	17.50	16.89
P2	18.10	0.00	12.12	5.07	4.90	4.00	4.99
P3	12.01	12.12	0.00	11.70	11.54	11.90	11.43
AI	17.13	5.07	11.70	0.00	2.18	2.69	3.05
P1-AI	17.13	4.90	11.54	2.18	0.00	2.53	2.10
P2-AI	17.50	4.00	11.90	2.69	2.53	0.00	2.93
P3-AI	16.89	4.99	11.43	3.05	2.10	2.93	0.00

Table 1: The discordance between the pathologists after AI assistance. P1, P2, P3 are Pathologists' scores before AI assistance, P1-AI, P2-AI, P3-AI are Pathologists' scores after AI assistance, and AI denotes the Ki-67 scores given by the AI system.

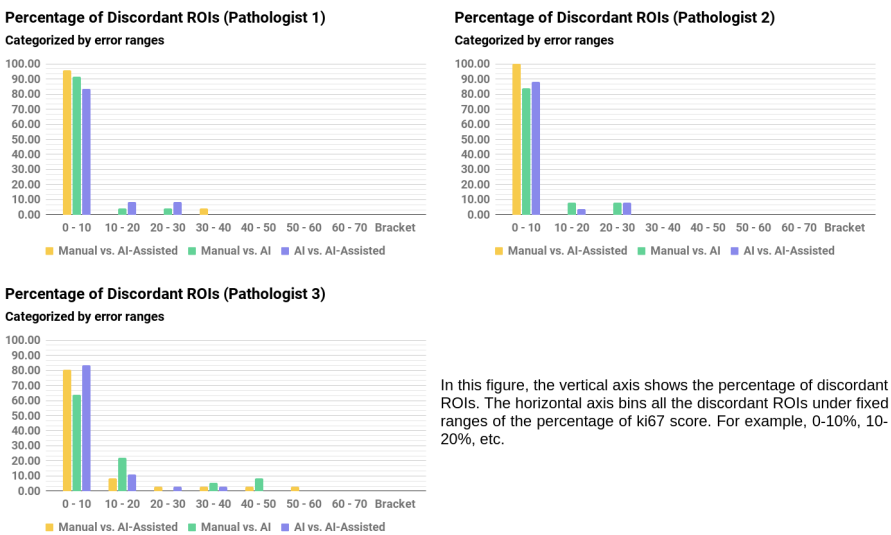
Each cell represents discordance measured by root mean square error of pairs of variables, defined

$$\sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N}}$$

where *x*, *y*, and *N* denote the row names, column names and the total ROIs used.

Figure 1 - 1007

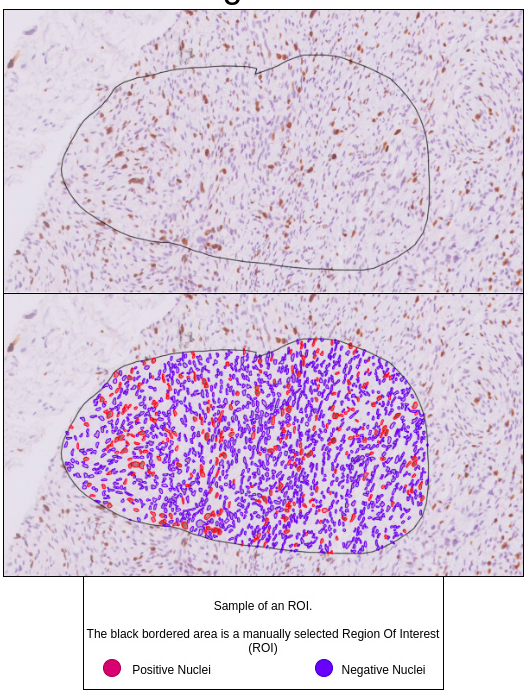
Figure 1.



In this figure, the vertical axis shows the percentage of discordant ROIs. The horizontal axis bins all the discordant ROIs under fixed ranges of the percentage of ki67 score. For example, 0-10%, 10-20%, etc.

Figure 2 – 1007

Figure 2.



Conclusions: We show that AI, correctly used in quantitative immunohistochemistry, can increase the accuracy of the output and reduced the discordance between the pathologists by 82.10%. This adds objective specificity to the scoring.

1008 Validation of Automated Digital Imaging Analysis of Estrogen Receptor Immunohistochemistry in the Clinical Setting

Saba Shafi[1], David Kellough[2], Giovanni Lujan[3], Swati Satturwar[2], Anil Parwani[3], Zaibo Li[1]
[1]The Ohio State University Wexner Medical Center, Columbus, OH, [2]The Ohio State University Wexner Medical Center/James Cancer Hospital, Columbus, OH, [3]The Ohio State University, Columbus, OH

Disclosures: Saba Shafi: None; David Kellough: *Employee*, Inspirata, Inc; Giovanni Lujan: None; Swati Satturwar: None; Anil Parwani: None; Zaibo Li: None

Background: The advent of whole-slide digital scanning and the use of deep learning (DL)-based neural networks has generated enormous interest in artificial intelligence (AI)-based digital pathology technologies. Objective assessment of various diagnostic and therapeutic cancer biomarkers can be done using various deep neural network-based approaches, which are fast, robust and reproducible and show good concordance with pathologists' manual estimation. This has resulted in the need for validation and evaluation of performance of these approaches in the clinical setting.

Design: We carried out an analysis of estrogen receptor (ER) immunohistochemistry (IHC) using Visiopharm (VIS, Hoersholm, Denmark) automated digital image analysis (DIA) platform to quantify ER positivity on ER IHC whole slide images (WSI) of breast cancer cases streamed directly from our clinical Image Management System (IMS). After setting the optimal color deconvolution, DAB optical density mean of DAB positive cells were detected using positive cell detection with varying thresholds to detect varying intensities of ER positive cells in VIS platform. The performance of the calibrated VIS platform was tested against two pathologists manual scoring of ER IHC.

Results: Out of the 73 ER-positive cases, VIS DIA categorized 70 (95.9%) as ER-positive. Twenty-one (87.5%) of 24 ER-negative cases were classified as ER-negative by VIS DIA. (Table 1) The overall concordance between pathologist reads and VIS reads

was excellent (93.8%) (Figure 1). Discordance between pathologists and VIS was seen in six cases (3 ER-negative and 3 ER-positive) with a Pearson Correlation Coefficient of 0.72. All 3 false negative cases had very weak ER staining and no more than 10% positivity. One false positive case was caused by intermixed benign glands which showed positive staining and were not excluded by VIS DIA. The second false positive case was due to inclusion of ER positive ductal carcinoma in-situ (DCIS). The third false positive case was due to nonspecific staining on folded tissue area (Figure 2).

Table 1. Comparison between pathologists' reads of ER with VIS

		Original reads		Total (n)	Concordant (%)
		ER-positive (%)	ER-negative (%)		
Total cases		73 (75.3)	24 (24.7)	97	
VIS reads	ER-positive	70 (95.9)	3 (12.5)	73	93.8
	ER-negative	3 (4.1)	21 (87.5)	24	
Pathologist reads - 1	ER-positive	69 (94.5)	3 (12.5)	72	92.8
	ER-negative	4 (5.5)	21 (87.5)	25	
Pathologist reads - 2	ER-positive	73 (100.0)	0 (0.0)	73	100.0
	ER-negative	0 (0.0)	24 (100.0)	24	

Figure 1. Correlation between original pathologist reads and VIS DIA reads.

Figure 2. Discordant cases. **A**. False positive case with intermixed benign glands. **B**. False positive case with staining in DCIS. **C**. False positive case due to tissue folds. **D**. False negative due to weak nuclear staining.

Figure 1 - 1008

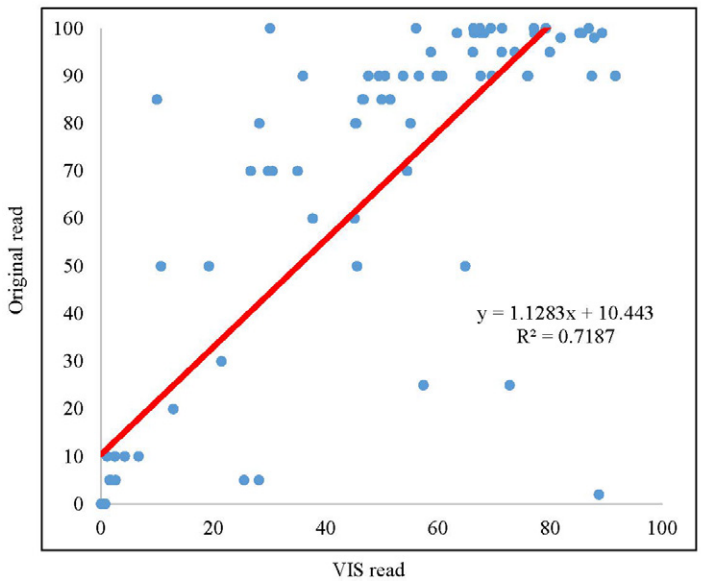
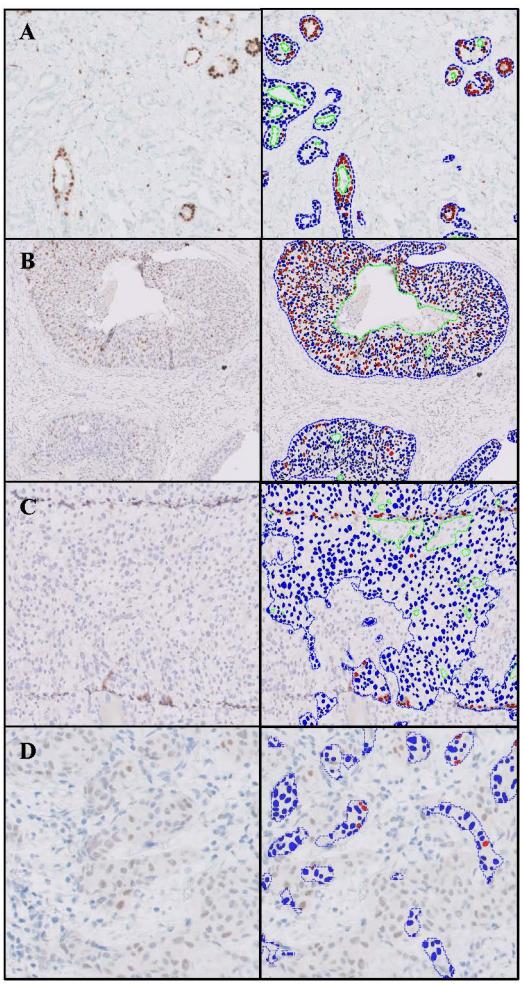


Figure 2 - 1008



Conclusions: In summary, our study provides real-world clinical data to support the robustness of this IMS integrated, VIS automated platform for ER quantification without downloading/uploading. The overall concordance between pathologist reads and VIS reads was excellent which is quite promising for its application in the clinical setting in the future.

1009 Comparison of Four Different Displays for Identification of Select Pathologic Features Extracted From Whole Slide Images of Surgical Pathology Cases

Nada Shaker[1], Konstantin Shilo[1], Ankush Patel[2], Ashwini Esnakula[3], Saba Shafi[2], Bindu Challa[2], David Kellough[3], Scott Hammond[1], Sehrish Javaid[1], Swati Satturwar[3], Martha Yearsley[2], Zaibo Li[2], Abberly Lott Limbach[2], Giovanni Lujan[1], Anil Parwani[1]

[1]The Ohio State University, Columbus, OH, [2]The Ohio State University Wexner Medical Center, Columbus, OH, [3]The Ohio State University Wexner Medical Center/James Cancer Hospital, Columbus, OH

Disclosures: Nada Shaker: None; Konstantin Shilo: None; Ankush Patel: None; Ashwini Esnakula: None; Saba Shafi: None; Bindu Challa: None; David Kellough: *Employee*, Inspirata, Inc; Scott Hammond: None; Sehrish Javaid: None; Swati Satturwar: None; Martha Yearsley: None; Zaibo Li: None; Abberly Lott Limbach: None; Giovanni Lujan: None; Anil Parwani: None

Background: Pathologists have to make decisions about choosing the best display for digital reads of images for clinical applications. Some of these displays may be part of a “closed” vendor system. No minimal standards for displays for the appropriate use in digital pathology have been defined. Pathologists are using home/remote monitors for clinical diagnostics, especially during the pandemic. Many factors such as the spatial resolution, contrast, hue, saturation, and need to brightness are to be considered. Our study is aimed to assess the performance of three fixed work displays and one remote personal display with varying resolutions to determine their effectiveness for the identification of ten features in digital pathology images.

Design: The study subjects included seven pathologists (minimum of 5 years of board-certified experience and at least two years of digital sign-out experience) and three residents. The subjects reviewed ninety images using the LG, Dell, and the Samsung displays. The subjects also reviewed the same image set using a fourth remote/home/personal monitor (including Philips, Dell, personal laptop, etc.) to assess the safety of cases diagnosis using a consumer monitor. All results recorded by the subjects were compared with features recorded as gold standard on a Philips display. Different features were randomly selected and annotated on whole slide images to ensure a wide range of different cases.

Results: All participants preferred the larger displays (>30”). The accuracy of the identification of ten features on different types of displays did not show significant difference between the three types of “fixed” workplace displays. Most of features were identified by the pathologists using the three types of monitors. The highest accuracy was recorded with identification of neutrophils, eosinophils, plasma cells, granuloma, and mucin. However, the identification of crystals, mitoses, necrosis, hemosiderin, and nucleoli showed less accuracy.

Table 1: The accuracy of the identification of ten features on different types of displays

	Monitor 1(LG)	Monitor 2 (Samsung)	Monitor 3 (Dell)	Monitor 4 (Personal)
Pathologist-1 (15 years’ experience)	92.1%	89.5%	92.6%	92% (Dell)
Pathologist-2 (15 years’ experience)	NA	91.1%	93.5%	94.3% (Philips)
Pathologist-3(20 years’ experience)	83.5%	85.6%	83.7%	82.3% (Dell laptop)
Pathologist -4(5 years’ experience)	82.7%	85.2%	82.7%	84.4% (Philips)
Pathologist-5 (15 years’ experience)	NA	NA	NA	84.2% (Philips)
Pathologist-6 (15 years’ experience)	NA	NA	NA	85.3%
Pathologist-7(10 years’ experience)	NA	NA	NA	82.7%(HP desktop)
Trainee-1 (PGY2)	81.6%	83.7%	80.6%	80.2% (HP laptop)
Trainee-2 (PGY2)	79.4%	82.6%	80.7%	80.4%(MAC book pro 16 inch)
Trainee-3 (PGY1)	78.3%	80%	79.4%	79.5%(Dell)

Conclusions: We conclude that most features could be identified using any display. However, there were some features which were more challenging to be identified across the three fixed display types. In addition, the remote personal monitors showed similar accuracy in feature identification. This study supports use of standard “unlocked” vendor agnostic displays for clinical digital pathology workflow rather than buying “locked”, more expensive displays that are part of a digital pathology system.

1010 Types and Frequency of Whole Slide Imaging Scan Failures in a Clinical High Through-Put Digital Pathology Scanning Laboratory

Nada Shaker[1], David Kellough[2], Savannah Erck[2], Erin Palermini[3], Zaibo Li[4], Giovanni Lujan[1], Anil Parwani[1], Swati Satturwar[2]

[1]The Ohio State University, Columbus, OH, [2]The Ohio State University Wexner Medical Center/James Cancer Hospital, Columbus, OH, [3]The Ohio State University Medical Center, OH, [4]The Ohio State University Wexner Medical Center, Columbus, OH

Disclosures: Nada Shaker: None; David Kellough: *Employee*, Inspirata, Inc; Savannah Erck: None; Erin Palermini: None; Zaibo Li: None; Giovanni Lujan: None; Anil Parwani: None; Swati Satturwar: None

Background: Digital pathology adoption for clinical diagnostics continues to increase due to favorable regulatory environment and need for remote diagnosis during COVID-19 pandemic. Whole slide imaging (WSI) scan failure is fairly uncommon. However, there is limited literature about the true incidence of WSI scan failure rates and the impact on the daily operations in a setting of complete digital workflow. Our digital pathology scanning facility is one of the largest clinical digital pathology operating the world. Our facility routinely monitors scan failure data as a part of quality control and quality assurance. This study was undertaken to address the issues related to scan failure and to assess impact on turn-around times (TATs). This data would be beneficial to health care providers considering transition to a complete digital work-flow.

Design: In 2017, we transitioned from scanning of archival slides to mostly new slides for primary diagnosis. We have operated 13 different Philips UFS scanners and scanned 2,289,266 slides representing nearly 233,864 cases.

Scan failure data was collected from 3 resources (1) Errors detected by machine, (2) Retrospective quality control review and (3) Errors reported by pathologists. Every slide image is appraised by the scanner for defects including failed region of interest (ROI) detections, slides skipped, slides dropped, tissue not detected, and other faults. Each image is also checked by scan technician to determine if the ROI was correctly captured or not. Routinely 1.5% of the daily scans are inspected by senior staff for quality assurance. Slides are scored on a scale of 1 to 10 using different parameters and scans scoring <8 are designated as failed scans and slides typically get rescan. Total scan failure rates, re-scan (since 2019) rates were recorded and monitored.

Results: Table 1 summarizes WSI scan failure data at our facility. Overall scan failure rate was just 1.19% with majority of the failures were attributable to machine error followed by failures due to slide preparation features. Most common machine error was failed ROI followed by skipped tissue error.

Table 1 Summary of whole slide imaging scan failure types and frequency

	Archival Slides	Primary Diagnosis Slides	Total Life-of-Project
Number of Tissue Skipped Errors	4,392	1,841	6,233
Number of Region of Interest Errors	8,030	9,512	17,542
Number of Slides Dropped	1,233	271	1,504
Number of Other Errors (Out of focus, too faint staining, tissue too thick or too small, broken slides etc)	894	670	1,564
Issues Found on Quality Control Review	283	203	486
Total Errors	14,832	12, 497	27, 329
Total Slides Scanned	1,244,763	1,044,503	2, 289, 266
% Total Errors	1.19%	1.20%	1.19%
Number of re-scan requests	Not available	327	327

Conclusions: WSI scan failure is extremely uncommon (1.19%) in a facility with experienced slide scanning staff and optimal slide preparations. Re-scanning was requested only for 1.19% cases and was feasible in 100% cases. Scanning of archival versus newly prepared slides did not have an impact on scan failure rates. Scan failure is not frequent enough to impact TATs and therefore need not be a concern for institutions considering transitioning to digital workflow.

1011 Evaluation of SpliceAI in Predicting the Impact of Synonymous Variants on Splicing Alterations

Reiri Sono[1], Liying Zhang[2]
[1]David Geffen School of Medicine at UCLA, Los Angeles, CA, [2]University of California, Los Angeles, Los Angeles, CA

Disclosures: Reiri Sono: None; Liying Zhang: *Other*, Decipher Medicine

Background: Clinical whole exome or genome sequencing must keep abreast of fast growth of molecular evidence. Splicing alteration is gaining attention as a possible pathogenic mechanism that may push more variants of uncertain significance to a definite pathogenicity status. Many splice prediction tools with machine learning have been developed. While they are advantageous in their freedom from wet lab setup, their accuracies are dependent on the volume and diversity of organically verified variants on which they are trained. SpliceAI is one of the newest and most accessible such application, which was trained on human reference sequence and validated with canonical splice site variants in the GTEx rare disease cohort. However, many more splice-altering variants have been described before its conception. Synonymous substitutions are of prime interest due to their inconsistent handling across institutions. Here, we aimed to measure the accuracy of SpliceAI predictions using synonymous pathogenic or likely pathogenic variants in the ClinVar database.

Design: A list of all variants and their evidence documentation were downloaded from the ClinVar FTP repository. The HGVS notations of synonymous substitutions were identified and left outer joined with all available evidence documents that contain any relevant keywords including "splice", "splicing", and "mRNA". The evidence was then manually adjudicated to weed out negative clauses, conflicts, or ambiguities. These variants were uploaded to the Ensembl cloud engine to generate SpliceAI outputs. The variant was called positive for splice alteration if the highest score out of any number of score quads (splice acceptor/donor gain/loss) was ≥ 0.1, as per our clinical protocol. The percentage of positive variants were compared between variants with or without in vitro evidence.

Results: From those variants with in vitro evidence of splice site alteration, 80/102 (78.4 %) were predicted positive. From those without, 2183/18145 (12.0 %) were predicted positive (p = 0.00).

Pathogenic or Likely Pathogenic Synonymous substitution variants on ClinVar	has in vitro mRNA evidence of splice alteration	lacks in vitro mRNA evidence of splice alteration	total
predicted to alter splice site	80	2183	2263
predicted NOT to alter splice site	22	15962	15984
total	102	18145	18247

Conclusions: SpliceAI has a moderate to high sensitivity and high specificity in predicting splice site changes on synonymous substitution variants deemed pathogenic or likely pathogenic (P/LP) in ClinVar when compared to the manually adjudicated, conservative evidence of in-vitro mRNA assay positivity. The false positives may represent the next focus of in-vitro assays where current evidence is insufficient. The false negatives may serve as re-training inputs to the algorithm.

1012 Detection of Y90 Microspheres by Super-Resolution Micro-CT of Liver Tissue Specimens Obtained Through 90Y PET/CT Guided Biopsies after Trans-Arterial Y90 Radioembolization of Liver Tumors

Alexei Teplov[1], Nicola Naydenov[1], Marissa Iraca[1], Kareem Ibrahim[1], Mahdi Zirakchian Zadeh[1], S Ruan[1], Takashi Ohnishi[1], Adam Kesner[1], Pat Zanzonico[1], John Humm[1], Assen Kirov[1], Yukako Yagi[1]
[1]Memorial Sloan Kettering Cancer Center, New York, NY

Disclosures: Alexei Teplov: None; Nicola Naydenov: None; Marissa Iraca: None; Kareem Ibrahim: None; Mahdi Zirakchian Zadeh: None; S Ruan: None; Takashi Ohnishi: None; Adam Kesner: None; Pat Zanzonico: None; John Humm: None; Assen Kirov: None; Yukako Yagi: None

Background: Since the mean activity of the microspheres used in trans-arterial ^{90}Y radioembolization is known, counting them in situ in specimens obtained through core needle biopsy provides an opportunity to estimate the local dose with higher resolution

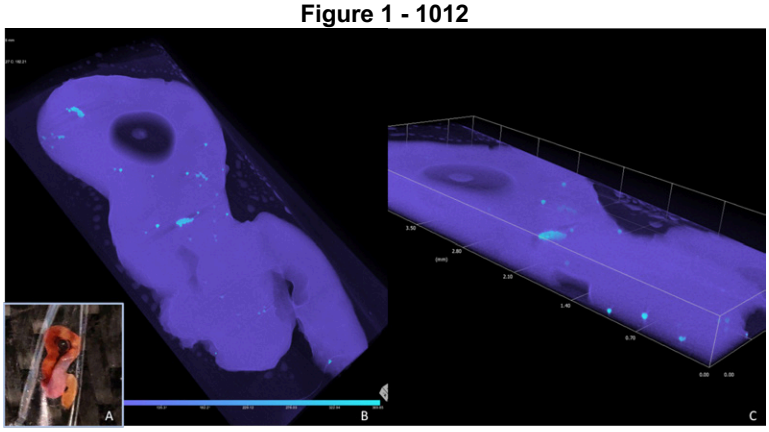
than what is possible with ^{90}Y PET images. This method could enable the investigation of the biological effects from ^{90}Y radiation at the microscopic level. We present a technique to visualize the microspheres using a super-resolution Micro-CT.

Design: Liver specimens were obtained under ^{90}Y PET/CT guided 18G core needle biopsies after the intrarterial hepatic injection of resin (SIR-SpheresR) microspheres in 3 patients, and glass microspheres (TheraSphereTM) in one patient were scanned along with 2 surrogate specimens made of gelatin-glycerol-water mixture which was pre-loaded with known activity glass microspheres. To minimize CT artifacts the specimens (1 to 3 mm3) were placed on carbon fiber plates which were then wrapped with 3-micron-thick mylar to prevent the specimens from drying. The samples were scanned using a custom-built micro-CT system (Nikon) at a voxel size 3-5 microns and the data were reconstructed. The number of spheres in each sample was counted manually by two observers, while the 3 samples with resin microspheres and the 4 samples with glass microsphere were counted using a commercial machine learning tool. For the resin samples, the in-house software which evaluates a sphericity with 3D Gaussian distribution automatically then counts microsphere using evaluated value.

Results: The technique described above allowed to visualize and count resin and glass microspheres in both clinical and surrogate gel specimens (Table). The number of spheres varied from 109 to 214 (standard deviation (SD) range 16 to 17) in the 3 clinical specimens loaded with resin microspheres, was 78 (SD 4) and 144 (SD 8) in the clinical specimens loaded with glass microspheres, and was 21 (SD 1) and 42 (SD 2) in the surrogate specimens loaded with glass microspheres (Table). The glass (compared to resin) microspheres are more clearly seen due to higher contrast and were easier to count in both clinical and surrogate samples.

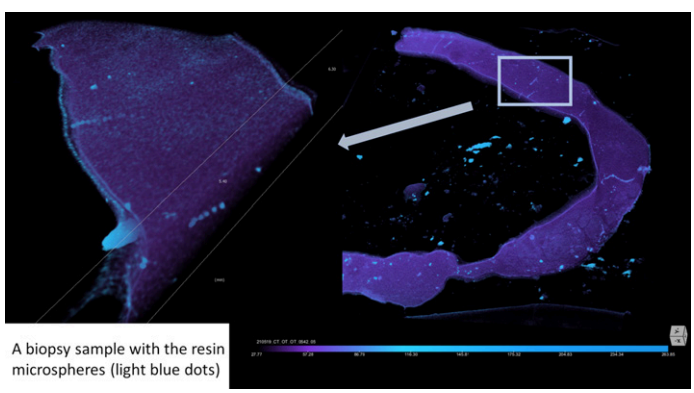
		Count	In-house	1	2				Volume (mm^3)
Study ID	Y90	Software1	Software	Manual	Manual	Mean	STD	%STD	Software
Case 1	Resin	215	187	210	225	**209**	16	7.7	
Case 2	Resin	139	177	149	139	**151**	18	11.9	
Case 3	Resin	125	110	125	97	**114**	14	11.8	
Case 4-1	glass	136		151	146	**144**	7.6	5.3	1.76
Case 4-2	glass	79		82	74	**78.3**	4	5.2	2.91
Case 5	glass	22		22	20	**21.3**	1.2	5.4	1.46
Case6	glass	42		42	41	**41.7**	0.6	1.4	1.05

Figure 1 - 1012



A: Biopsy tissue sample, B and C: 3D reconstructed model of sample in A. Light blue dots are glass microspheres

Figure 2 – 1012



A biopsy sample with the resin microspheres (light blue dots)

Conclusions: The super resolution micro-CT can be a promising method to count microspheres at their location in specimens obtained through core needle biopsy to estimate the local dose. The in-house application will be further developed so that it can be used for different types of spheres.

1013 MMR GINIE: A Novel Excel-based Natural Language Processing Tool Identifies Mismatch Repair Gene Status from Pathology Reports

Boris Virine[1], Aaron Pollett[2]
[1]Laboratory Medicine Program, Departments of Anatomical Pathology, University Health Network and University of Toronto, Toronto, Canada, [2]Mount Sinai Hospital, Toronto, Canada

Disclosures: Boris Virine: None; Aaron Pollett: None

Background: Natural Language Processing (NLP) is a computational tool used to categorize unstructured data, including pathology reports. Despite increasing use of synoptic reports for colorectal carcinoma, mismatch repair (MMR) gene status is often excluded. To improve the efficiency of data analysis NLP tools offer the potential to simultaneously analyze numerous pathology reports and provide categorized MMR status information. While high quality NLP software is available, these systems can be complicated to use and are not optimized for use in analyzing text data in spreadsheet form. We aimed to develop and validate an easy-to-use NLP tool that is integrated into Microsoft Excel, focused on pulling MMR status information from pathology reports.

Design: All colorectal cancer pathology reports from 2000-2001 and 2010-2016 containing MMR gene data at our institution were exported to an Excel spreadsheet. A Visual Basic (VBA) based NLP program was developed in-house, called MMR Gene Identification using NLP Integrated into Excel (MMR GINIE) to process the data. The software was trained using two year's datasets. Validation was performed by analyzing results from a representative sample of cases from datasets between 2011 and 2016.

Results: MMR GINIE was able to successfully extract MMR gene status from 88.8-99.4% of cases. A detailed error analysis revealed that 89.5% of errors involved mis-categorization of individual genes, 10.5% of errors involved misidentification of MSI status, and 5.3% of errors were associated with the failure to detect genes. The accuracy of the system was dependent on the report type, with conventional reports of IHC and MSI testing being correctly interpreted in 97.6% of cases, compared to 85.6% of molecular reports.

Conclusions: Here we demonstrate that an Excel-based NLP tool can be used to accurately extract MMR status data from unstructured or semi-structured pathology reports.

1014 Refined HOX Codes for the Identification of the Origins of Cancer Cells

Yunshin Yeh[1], Siyuan Cheng[2], Shu Yang[2], Yingli Shi[3], Xiuping Yu[2]
[1]Overton Brooks VA Medical Center, Shreveport, LA, [2]Louisiana State University Health Sciences Center, Shreveport, LA, [3]LSU Health Shreveport, Shreveport, LA

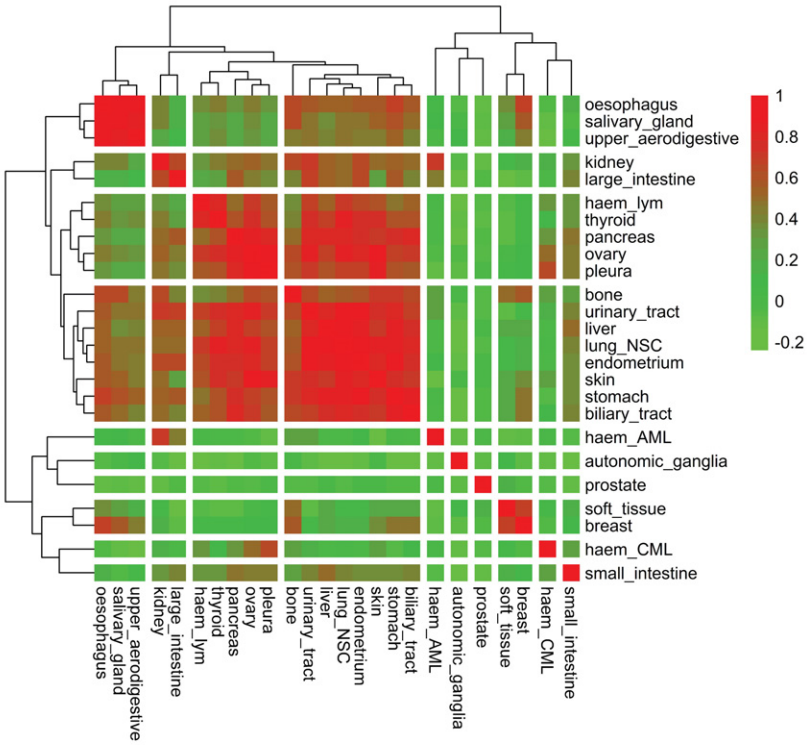
Disclosures: Yunshin Yeh: None; Siyuan Cheng: None; Shu Yang: None; Yingli Shi: None; Xiuping Yu: None

Background: The 39 human homeobox genes (HOXs) are located in 4 HOX clusters (HOXA-D). HOXs are highly conserved and are essential for human development. The 3' and 5' order of each gene cluster (1-13) reflects their temporal and spatial expression pattern during embryonic morphogenesis. In previous studies, we have established 24 HOXs expression patterns (HOX codes) that correspond to 24 anatomical sites. In this study, we have further refined the HOX codes and have generated 42 HOX codes for different cell lineages instead of the anatomical origins of cells.

Design: The 39 human HOX gene mRNA expression data were extracted from the Cancer Cell Line Encyclopedia (CCLE) dataset. A total of 1019 cell lines were re-grouped based on the information of refined cancer type, histology, histological subtype, and primary site in the CCLE annotation file. The median of each HOX gene expression level was calculated for each group to establish the HOX code for that specific tissue type. Pearson correlation was calculated between each individual HOX code and its corresponding cancer cell lines. The results of 173 HOX patterns were filtered or combined based on the number and similarity of the samples as well as the information retrieved from the literature review. The Cancer Genome Atlas (TCGA) datasets were used to validate the refined 42 HOX codes.

Results: All of the 42 HOX codes correspond to their own cell lineages. The majority of the HOX codes demonstrate a strong positive correlation with the cell lines from which they are derived and show a weak or negative correlation with different types of cell lines. Unsupervised clustering analysis indicated that the HOX codes originate from the same germ layer or closer anatomical location mimicked more than their counterparts. We also validated the HOX codes using RNASeq data collected by TCGA and found a consistent result that the majority of the HOX codes correlate well (Pearson correlation) with the types of cancer from which the cells originate.

Figure 1 - 1014



Conclusions: We established 42 HOX codes for 42 cell lineages. We found that the majority of the HOX codes are distinct gene groups and represent their corresponding cell lineages. These HOX codes may be used to identify the sites of tissue origin for metastatic cancers.

1015 STMN1 Expression is Elevated in Neuroendocrine Small Cell Carcinoma of the Prostate

Yunshin Yeh[1], Yingli Shi[2], Siyuan Cheng[3], Shu Yang[3], Xiuping Yu[3]
[1]Overton Brooks VA Medical Center, Shreveport, LA, [2]LSU Health Shreveport, Shreveport, LA, [3]Louisiana State University Health Sciences Center, Shreveport, LA

Disclosures: Yunshin Yeh: None; Yingli Shi: None; Siyuan Cheng: None; Shu Yang: None; Xiuping Yu: None

Background: Prostate cancer (PCa) is the second leading cause of cancer-related death in men in the United States. It has been shown that most of the PCa patients treated with androgen deprivation therapy eventually develop castration-resistant PCa and approximately 20-30% of the castration-resistant PCa acquire neuroendocrine (NE) features. To identify the mechanism involved in NE differentiation of PCa cells, we have conducted bioinformatics analysis and have identified the differentially expressed genes of NEPCa and prostatic adenocarcinoma (AdPCa).

Design: To explore the expression of STMN1 (stathmin1) in the prostate, we performed the bioinformatics analysis using the single and bulk RNA-Seq data obtained from the single-cell portal, cBioportal, the Cancer Genome Atlas, and the Cancer Cell Line Encyclopedia websites. To verify the expression pattern of STMN1, we performed Western blot and immunohistochemical staining on PCa cell lines and tissue samples of human PCa and mouse models.

Results: The bioinformatics transcriptomic analysis showed that the expression of STMN1 and STMN3 are both significantly upregulated in NEPCa compared with that of AdPCa samples. These findings were supported by the results that STMN1 and STMN3 expression levels are significantly higher in NEPCa cell line NCI-H660 than AdPCa cell lines including PC3, 22RV1, VCaP. and DU145. In AdPCa, overexpression of STMN1 was positively correlated with Gleason score and was associated with lower disease-free survival in PCa patients. STMN1 overexpression was further demonstrated by the positive staining of STMN1 immunomarker in human tissue samples and mouse models of NEPCa. In addition, increased STMN1 expression was also observed at 2 and 3 days after readministration of androgen in prostate regeneration mouse models and proliferating cancer cells indicated by the diffusely positive staining of Ki67 immunomarker.

Conclusions: The expression of STMN1 was increased in NEPCa and proliferating cells, that suggests a potential mechanism for the emergence and progression of NEPCa. STMN1 may serve as a prognostic marker for advanced PCa and a potential therapeutic target for the treatment of both NEPCa and advanced PCa.

1016 Comparison of Digital Analysis and Microscopic Review of Ki-67 Scoring in Follicular Lymphoma Disease Prognostication

Hira Yousaf[1], Aqsa Nasir[2], Ryan Shanley[1], Michael Linden[3], Sophia Yohe[1]
[1]University of Minnesota, Minneapolis, MN, [2]University of Florida College of Medicine, Gainesville, FL, [3]University of Minnesota Medical Center, Minneapolis, MN

Disclosures: Hira Yousaf: None; Aqsa Nasir: None; Ryan Shanley: None; Michael Linden: *Other*, Cell Signaling Technology; Sophia Yohe: None

Background: A recent study (unpublished) at our institute showed that a high interfollicular area (IFA) Ki-67 proliferation index (PI) is associated with lower progression free survival (PFS) in patients with follicular lymphoma, based on conventional manual microscopic review. While digital analysis has been approved in clinical practice for many immunohistochemical markers. It has not been well studied for Ki-67 PI, especially for follicular lymphoma cases. We performed digital analysis of these cases to verify our previous findings and to study the correlation with manual review.

Design: Grade 1 to 3A follicular lymphoma cases from 2003 to 2016 were stained for Ki-67 and scored by manual review by 2 pathologists on high-power (40x). 10 consecutive follicular areas (FA) and IFA were averaged separately when there was differential staining, while a single count was used in cases with no difference between areas. For digital analysis, Ki-67 stained slides were scanned with Aperio E slide manager and analyzed using Aperio image analysis Nuclear Version 9. Each case was annotated and analyzed in the best representative area. In cases with sufficient tissue, either a FA and IFA or a single area of analysis of 1.50+/-0.3mm^2 was set; while for cases with small tissue, the maximum available area was analyzed, giving an average area of 1.38 for FA and 1.45 for IFA. Intensity of positive nuclear staining was quantified as 0=no staining, 1+=weak, 2+=moderate and 3+=strong staining. Digital analysis generated data for total number and percentage of positive nuclei; average percentage and number of positive nuclei and area of analysis. Cox regression analysis was used to measure overall survival (OS) and progression free survival (PFS).

Results: The Ki-67 PI in some cases was markedly different between the FA and IFA. The median FA, Ki-67 was 25 (interquartile range (IQR) 15 to 40) and the median IFA Ki-67 was 10 (IQR 5 to 15). Digital analysis and Microscopic review were only moderately correlated for Ki-67 (Spearman's correlation 0.55 for FA, and 0.44 for IFA). However, associations with OS and PFS were similar for both methods, showing a similar trend with a significant correlation of higher IFA Ki-67 PI and lower PFS (Table 1).

	Manual Review	Digital analysis
Total Cases	70	61
Overall survival for follicular area	Hazard ratio for a 10 point increase in Ki-67 PI: 0.92 (95% CI 0.71 to 1.21; p= 0.56)	Hazard ratio for a 10 point increase in Ki-67 PI:0.79 (95% CI 0.47 to 1.34; p= 0.39)
Overall survival interfollicular area	Hazard ratio for a 10 point increase in Ki-67 PI: 0.85 (95% CI 0.46 to 1.58; p= 0.61)	Hazard ratio for a 10 point increase in Ki-67 PI: 1.08 (95% CI 0.59 to 1.99; p= 0.80)
Progression free survival for follicular area	Hazard ratio for a 10 point increase in Ki-67 PI :0.94 (95% CI 0.82 to 1.07; p= 0.34)	Hazard ratio for a 10 point increase in Ki-67 PI: 1.10 (95% CI 0.87 to 1.40; p= 0.42)
Progression free survival for interfollicular area	Hazard ratio for a 10 point increase in Ki-67 PI: 1.32 (95% CI 1.08 to 1.63;p= 0.01)	Hazard ratio for a 10 point increase in Ki-67 PI:1.38 (95% CI 1.06 to 1.80; p= 0.02)

Conclusions: Ki-67 scoring by digital analysis and manual review are moderately correlated and both are prognostic to predict PFS in patients with follicular lymphoma. Marked differences in FA and IFA Ki-67 PI can be seen and IFA PI was associated with PFS in our study.

1017 Predicting Breast Cancer Metastasis in Sentinel Lymph Node with Deep Learning Technology

Lan Zheng[1], Nghia Nguyen[2], Xiaohong Iris Wang[3], Songlin Zhang[2], Jianmin Ding[2], Hongxia Sun[2]
[1]McGovern Medical School at UTHealth, The University of Texas Health Science Center at Houston, Houston, TX, [2]The University of Texas Health Science Center at Houston, Houston, TX, [3]The University of Texas Health Science Center at Houston McGovern Medical School, Houston, TX

Disclosures: Lan Zheng: None; Nghia Nguyen: None; Xiaohong Iris Wang: None; Songlin Zhang: None; Jianmin Ding: None; Hongxia Sun: None

Background: Sentinel lymph node (SLN) biopsy has been widely used to assess axillary lymph nodes in patients with early breast cancer (BC). There is still a challenge on accurately and quickly determine the SLN status by histology screen alone. The objective of this study was to determine if deep learning could help.

Design: From a retrospective cohort of invasive ductal carcinoma patients who received upfront surgical management, H&E-stained SLN excisional biopsies slides were selected from 34 patients, 18 with positive findings (including macrometastasis and micrometastasis), and 16 with negative findings (including isolated tumor cells and negative). A positive slide and a negative slide (from a negative node), or two negative slides (for a negative case) were selected for each case to obtain a total of 68 digital whole-slide images (WSIs). Convolutional neural network algorithm was used to develop a breast cancer screening model. Forty representative images, 100x100 pixels in dimensions, were taken for each WSI in lymphocyte-rich area without tumor. A total of 2720 images were obtained from which 2160 (79%) were used for training the model, 240 (9%) for validation, and 320 (12%) for testing. For each test case, the predicted screen result is combined from the prediction for 5 images (at least 3 or more must agree), a process known as majority voting.

Results: In this study cohort, there is no statistical difference between positive and negative groups regarding to patient age, tumor size, histologic grade, tumor infiltrating lymphocytes, ER, PR and HER2 status. Using this screening model, a 92% accuracy with 84.4% sensitivity, 100% specificity, 100% positive predictive value and 86.5% negative predictive value were achieved in separating patient to positive or negative group. Note that images from either positive or negative node in patients with metastatic carcinoma has similar results and both can predict the metastatic status.

Conclusions: We developed a novel, digital pathology diagnostic model for BC screening in SLN with excellent results. Our findings of both positive and negative lymph node in patients with metastatic carcinoma having similar value in predicting metastatic status indicates the importance of tumor microenvironment in determining BC metastasis. This study provided a proof of concept for incorporating automated screening using digital microscopic images into pathology workflow to augment the pathologists' productivity.