

# An Image Analysis Resource for Cancer Research: PIIP—Pathology Image Informatics Platform for Visualization, Analysis, and Management



Anne L. Martel<sup>1,2</sup>, Dan Hosseinzadeh<sup>3</sup>, Caglar Senaras<sup>4</sup>, Yu Zhou<sup>5</sup>, Azadeh Yazdanpanah<sup>1</sup>, Rushin Shojai<sup>1</sup>, Emily S. Patterson<sup>4</sup>, Anant Madabhushi<sup>5</sup>, and Metin N. Gurcan<sup>4</sup>

## Abstract

Pathology Image Informatics Platform (PIIP) is an NCI/NIH sponsored project intended for managing, annotating, sharing, and quantitatively analyzing digital pathology imaging data. It expands on an existing, freely available pathology image viewer, Sedeen. The goal of this project is to develop and embed some commonly used image analysis applications into the Sedeen viewer to create a freely available resource for the digital pathology and cancer research communities. Thus far, new plugins have been developed and incorporated into the platform for out of focus detection, region of interest transforma-

tion, and IHC slide analysis. Our biomarker quantification and nuclear segmentation algorithms, written in MATLAB, have also been integrated into the viewer. This article describes the viewing software and the mechanism to extend functionality by plugins, brief descriptions of which are provided as examples, to guide users who want to use this platform. PIIP project materials, including a video describing its usage and applications, and links for the Sedeen Viewer, plug-ins, and user manuals are freely available through the project web page: <http://pathiip.org>. *Cancer Res*; 77(21); e83–86. ©2017 AACR.

## Introduction

With the advent of whole slide digital scanners, histopathology slides can be digitized into very high-resolution digital images, realizing a new "big data" stream that can potentially rival "-omics data" in size and complexity (1–3). The first FDA approval for marketing a whole slide imaging system for primary diagnosis was granted in April 2017, and this will generate increased interest in using digital scanners by pathologists in hospital settings. Features extracted from whole slide images (WSI) using sophisticated image analysis methods have already been shown to improve diagnosis (4) and have prognostic power in a number of applications (5). There is also growing interest in combining morphological information extracted from WSIs with genomic data and *in vivo* medical images to discover new biomarkers (6, 7).

There is a need for user-friendly software tools that can support the display, annotation, and analysis of different sources of image data. For researchers, the ability to use analysis algorithms that have been contributed by other researchers is also a key requirement.

Although other image analysis platforms exist, most have not been designed to handle digital pathology WSIs. For example, Fiji

(fiji.sc) is an image processing package aimed at providing scientific image analysis tools for life sciences researchers and it is built on ImageJ (imagej.nih.gov), which is an extensible, open source, Java-based program developed by the NIH. Fiji is popular for cell image analysis but it was not designed to view and process large tiled WSIs.

It is necessary to crop images to a subregion on import and/or to limit their resolution, which means that the user has to process images tile by tile, losing the spatial context of each patch. This is a major limitation that makes Fiji incompatible with a clinical pathology workflow. OpenSlide (openslide.org) is a C library that allows application developers to parse various WSI formats. Although it provides a simple slide viewer, OpenSlide lacks support for annotations, analysis, and multimodality images. QuPath (qupath.github.io) is a new and interesting software application for the digital pathology community, which appears to provide good support for WSI viewing, annotations, and basic analysis. Extensions can be added via JavaScript and Groovy (a scripting language for Java); however, although these are widely used languages, they are less suited for image analysis in pathology because of the very high demands on memory and performance.

Sedeen is a pathology image viewer (<http://pathcore.com/downloads/>), which was initially developed with funding from the Ontario Cancer Research Institute. It was designed specifically to address the unmet need for a platform that could be used by the academic community to share both established and novel digital pathology visualization and analysis tools with pathologists and other imaging researchers (8). It is built in C++, which ensures good computational performance while also making it simple to integrate with common image analysis libraries such as ITK, VTK, and OpenCV.

The pathology image informatics platform (PIIP) is a multi-institutional project supported by the NCI/ITCR, which is leveraging the Sedeen viewer framework (see Fig. 1). It is a collaboration

<sup>1</sup>Sunnybrook Research Institute, Toronto, Canada. <sup>2</sup>Medical Biophysics, University of Toronto, Toronto, Canada. <sup>3</sup>Pathcore Inc., Toronto, Canada. <sup>4</sup>The Ohio State University, Columbus, Ohio. <sup>5</sup>Case Western Reserve University, Cleveland, Ohio.

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

**Corresponding Author:** Anne L. Martel, Sunnybrook Research Institute, 2075 Bayview Avenue, M6-609, Toronto, Ontario M4N 3M5, Canada. Phone: 416-480-5702; Fax: 416-480-5714; E-mail: [Anne.martel@sri.utoronto.ca](mailto:Anne.martel@sri.utoronto.ca)

**doi:** 10.1158/0008-5472.CAN-17-0323

©2017 American Association for Cancer Research.

between five different participating academic institutions (University of Pennsylvania, University of Michigan, Ohio State University, Case Western Reserve University, University at Buffalo in the US and Sunnybrook Research Institute in Canada) and Pathcore, a for-profit company, which has been providing free access to Seden to the academic community since 2010. See Supplementary Video S1 for a brief overview of the PIIP project and a demonstration of the Seden viewer.

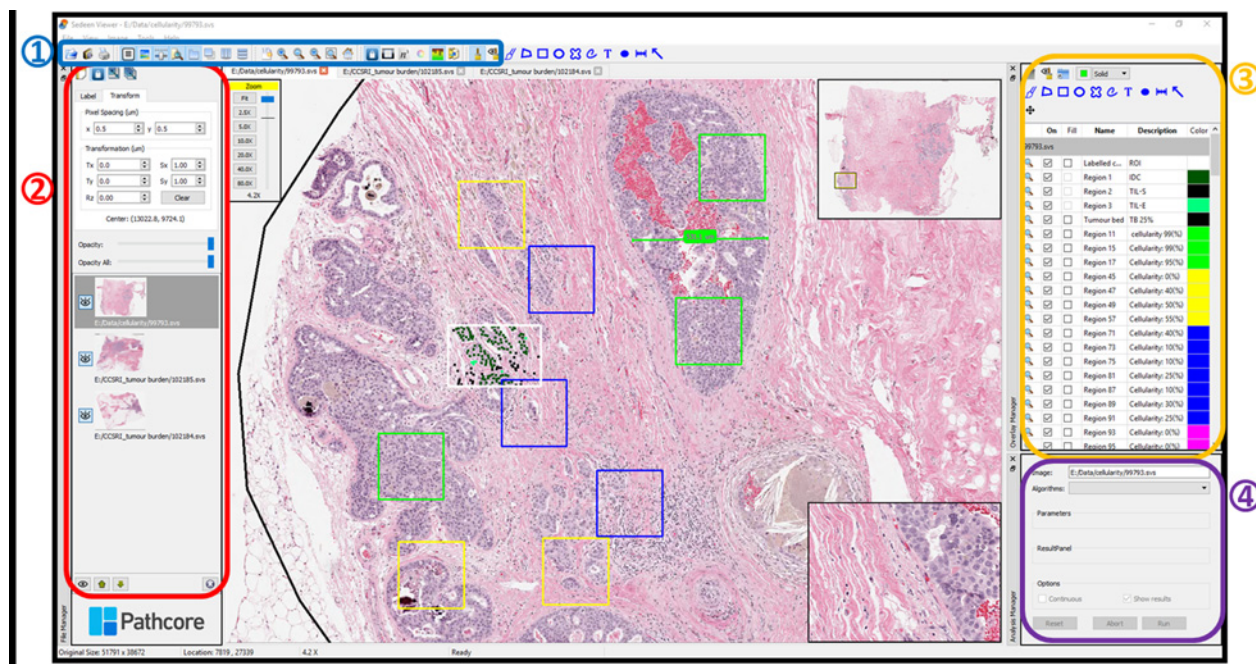
## Seden Viewer

Seden allows users to view images from multiple modalities (radiology and pathology in particular) and to perform visual overlays and registrations for biomarker validation (e.g., IHC and hematoxylin and eosin) and multimodality comparisons (hematoxylin and eosin and MRI). Among other capabilities, Seden provides rich annotation capabilities, supports a wide range of WSIs, and is extensible through the Pathcore Software Development Kit (SDK), allowing researchers to share and validate novel cancer informatics tools. At present, each slide scanner vendor has a proprietary WSI format. Typically, an image is saved as a series of tiles at multiple resolution levels; this allows subregions of the image to be retrieved without having to load the whole image into memory at once. An essential feature of the Seden viewer is the ability to import multiple file formats, including Aperio SVS, Leica SCN, Hamamatsu NDPI, Zeiss MRXS, VisionTek SVSlide, other TIFF-based formats, and JPEG-2000.

Although the viewer is currently distributed as a Windows application, it has been designed to be portable to other platforms, for example, MacOS and Linux, if required. Dr. Emily Patterson (Ohio State University, Columbus, OH), a human factors expert, has conducted a heuristic review of all the toolbars and menus on the Seden interface and will also provide input on the design of plugin interfaces (9). This will ensure that the software, including its plugins, is intuitive to users, which is important in collaborative research.

## Software Development Kit

The Pathcore SDK is a tool for integrating quantitative image analysis algorithms with the Seden viewer. The SDK is written in C++ and uses an open-source tool, *CMake* (cmake.org), to manage the build process in a way that is independent of both the operating system and the compiler used. The SDK application program interface (API) is fully documented and generated using *Doxygen*, which is a standard tool for extracting documentation from source file comments. It is distributed with a set of tutorials and exemplar algorithms, which can be selected from a pull-down menu at run time. Seden uses an on-the-fly image decoding strategy (directly from the disk), which is transparent to the clients and is very efficient. This allows client code to access arbitrary regions in an image without first loading the entire image into memory. The SDK simplifies access to WSI data by providing a set of format-independent data extraction routines, the ability to



**Figure 1.**

The user interface of Seden Viewer is shown. Certain key aspects of Seden Viewer's functionality are highlighted here, whereas further details of its use are described in the user manual. The toolbar (1) provides access to controls for image navigation and allows the display layout (stacked, vertical, and horizontal) to be customized. Screenshot, format conversion, and cropping tools are also available. The File Manager panel (2) allows the order of images and their visibility to be controlled. Opacity of images and transformation parameters can also be modified, which supports manual alignment. The Overlay Manager (3) provides tools for annotation and markup of images. Drawing tools include polygons, rulers, arrow, rectangles, ovals, and freehand shapes. The properties of overlays (color, name, and description) can be modified, and reports on their relevant characteristics (e.g., area, length) can be displayed. Annotations can be exported in a human readable XML format and can be imported into other programs and applications. The Analysis Manager (4) allows plugins generated with Pathcore SDK to be loaded into Seden. SDK generated plugins allow developers to collect parameters and display results of analysis in real-time.

create analysis pipelines, and the ability to easily visualize analysis results. Algorithm designers can therefore develop algorithms much more easily by leveraging the capabilities of Sedeen and the image analysis architecture provided by the Pathcore SDK.

The Pathcore SDK also provides numerous widgets that allow user parameters to be collected from the analysis manager GUI and transferred to the analysis algorithms. The output from the algorithm can then be rendered within the Sedeen viewer in the form of an overlay, new annotations, or as text.

## PIIP Plug-ins

With funding from the ITCR program, groups at Case Western University, The Ohio State University, and Sunnybrook Research Institute are implementing a variety of quantitative image analysis tools as plugins. Examples are described below and all of these plugins will be distributed with the next release of Sedeen. A github repository (<https://github.com/sedeen-piip-plugins>) contains the source code and user documentation for the plugins developed by the PIIP. Pathcore is providing support and is also working to expand the capabilities of the SDK to facilitate academic research, for example by developing methods for accessing MATLAB and Python code directly from the Sedeen Viewer.

### Biomarker quantification

The co-principal investigator (A. Madabhushi) and his team have developed a powerful tool for high-throughput biomarker quantification named Hierarchical Normalized Cuts (HNCuts; ref. 10). The method combines a frequency-weighted version of the mean shift (MS) algorithm with the normalized cuts scheme to segment all the image pixels into representative classes. Unlike supervised segmentation strategies, our scheme only requires specification of a small representative sample of the target class to rapidly identify similar objects in the image.

The PIIP provides algorithms needed to compute user-provided inputs (images, parameters, and annotations) and reports results through the Sedeen Viewer's GUI. In this case, users access the viewer's annotation tools to define the sample color swatch, which can be accessed by the HNCuts plugin. Once processing is carried out, the segmentation and quantification results are rendered within the Sedeen Viewer as a color/contour overlay. This plugin also demonstrates how algorithms implemented in Matlab can be integrated into a plugin.

### Out of focus detection

As whole slide imaging becomes more prevalent and many centers, particularly in Europe and Canada, adopt digital pathology, the demand for scanning whole slides will increase. Although many modern scanners are equipped with software to automatically detect areas that are out of focus, and adjust the scanning parameters accordingly, large numbers of digital slides contain out of focus areas. According to one study conducted at a Dutch academic pathology laboratory, 5% of the cases had problems with scanning artifacts such as out of focus areas (11). In a large-scale digitization effort, this percentage is expected to increase considerably, as the large volume of cases make it impractical to visually identify problematic cases and then rescan them if necessary.

The goal of this plugin is to detect out of focus areas and report them to the user. In its implementation, a plugin was created using

Pathcore SDK and integrated with the Sedeen Viewer. This plugin also demonstrates the use of the OpenCV API, a popular computer vision library, with the SDK. The plugin identifies regions that are out of focus and produces an overlay showing where these regions are located.

### Transforming and exporting annotations

One of the key differences between the Sedeen viewer and other platforms is that it provides tools for users to manually coregister images. This allows users to make direct comparisons between images acquired using different modalities or stained using different antigens. The ExportTransformedROI plugin extends the capabilities of Sedeen by allowing an annotation defined on a source image to be exported to a target image. If the source image and the target image have been aligned, then the transformations needed to achieve the alignment are also applied to the annotations.

### Use case example: analysis of multiple IHC slides

Serial sections cut from a tissue block and stained using different IHC markers can be used to compare the relative expression of different antigen receptors in tumors. In the Supplementary File, the workflow needed to spatially align two IHC images and then carry out a quantitative comparison of tissue staining over corresponding regions of interest is described. This demonstrates how Sedeen's built-in functionality can be combined with user-contributed plugins to achieve a specific analysis task.

## Discussion

The Sedeen Viewer has all of the advantages of a viewer built specifically for digital pathology, with a user interface that is intuitive for pathologists and biologists, while also providing a mechanism to expand its functionality through Pathcore SDK. The viewer is easy to download and install on Windows platforms and, in time, it will be made available for other platforms. We have thus far successfully demonstrated that image analytic algorithms from three different institutions (Ohio State University, Case Western Reserve University, and Sunnybrook Research Institute) can be integrated within the PIIP. Our future plans involve enabling the seamless integration of analytic tools developed by other users into the platform. We envision a PIIP that will be dynamically populated with validated image analytic "apps" developed by the community, and hence provide immense benefit to end users, including clinical and research centric pathologists and computational image analysis scientists. Pathcore has been providing free access to Sedeen since 2010 and is extending the capabilities of both the viewer and the Pathcore SDK in response to feedback from the PIIP project team. They are currently working on an improved mechanism to support the compilation and distribution of contributed plugins from the wider user community.

Other novel aspects of this project are also underway. Working with an expert on disease ontologies, Dr. Barry Smith (University of Buffalo, Buffalo NY), we have proposed a quantitative histopathology image ontology (QHIO), which will facilitate interoperability between histopathology datasets and methods used in pathological imaging and analysis. This ontology fosters data compatibility and provides a mechanism for linking algorithms with compatible datatypes. In ref. 12, we demonstrate how the

QHIO can be applied to the problem of hot spot detection in breast IHC WSIs. Human-computer interaction and user experience is an integral part of the PIIP. We are working with a human-computer interaction expert, Dr. Emily Patterson, to define user interface design requirements. On the basis of her suggestions, several design decisions have already been modified to make the interface friendlier to users with different experience levels.

### Disclosure of Potential Conflicts of Interest

A.L. Martel is a chief science officer and has ownership interest (including patents) in Pathcore.com. D. Hosseinzadeh has ownership interest (including patents) in Pathcore. A. Madabhushi reports receiving a commercial research grant from Philips and Inspirata Inc., has ownership interest (including patents) in Inspirata Inc. and Elucid Bioimaging, and is a consultant/advisory board member for Inspirata Inc. and Astrazeneca. M. Gurcan is a consultant/advisory board member for Inspirata, Inc. No potential conflicts of interest were disclosed by the other authors.

### Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the NIH.

### Authors' Contributions

**Conception and design:** A.L. Martel, D. Hosseinzadeh, E.S. Patterson, A. Madabhushi, M.N. Gurcan

### References

- Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: a review. *IEEE Rev Biomed Eng* 2009;2:147–71.
- Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: challenges and opportunities. *Med Image Anal* 2016;33:170–5.
- Bhargava R, Madabhushi A. Emerging themes in image informatics and molecular analysis for digital pathology. *Annu Rev Biomed Eng* 2016; 18:387–412.
- Fauzi MFA, Pennell M, Sahiner B, Chen W, Shana'ah A, Hemminger J, et al. Classification of follicular lymphoma: the effect of computer aid on pathologists grading. *BMC Med Inform Decis Mak* 2015;15:115.
- Lewis JS, Ali S, Luo J, Thorstad WL, Madabhushi A, Madabhushi A. A quantitative histomorphometric classifier (QuHbIC) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma. *Am J Surg Pathol* 2014;38:128–37.
- Singanamalli A, Rusu M, Sparks R, Shih N, Ziober A, Wang LP, et al. Identifying *in vivo* DCE MRI markers associated with microvessel architecture and gleason grades of prostate cancer—preliminary findings. *J Magn Reson Imaging* 2016;43:149–58.
- Lee G, Singanamalli A, Wang H, Feldman MD, Master SR, Shih NNC, et al. Supervised multi-view canonical correlation analysis (sMVCCA): integrat-

**Development of methodology:** A.L. Martel, C. Senaras, Y. Zhou, A. Yazdanpanah, R. Shojaii, A. Madabhushi, M.N. Gurcan

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** A. Madabhushi, M.N. Gurcan

**Writing, review, and/or revision of the manuscript:** A.L. Martel, D. Hosseinzadeh, C. Senaras, Y. Zhou, E.S. Patterson, A. Madabhushi, M.N. Gurcan

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** A.L. Martel, A. Madabhushi, M.N. Gurcan  
**Study supervision:** A.L. Martel, A. Madabhushi

### Acknowledgments

We are grateful to Craig Madho and Deyu Wang at Pathcore for their help in the coordination of the project and facilitating the writing process.

### Grant Support

The project described was supported in part by award number U24CA199374 (principal investigators: M.N. Gurcan, A. Madabhushi, A.L. Martel) from the National Cancer Institute. This project is also supported by Pathcore, a digital pathology software company. Pathcore has been supporting the research community since 2011 via the Seden Viewer and PIIP.

Received February 1, 2017; revised April 28, 2017; accepted August 9, 2017; published online November 1, 2017.

ing histologic and proteomic features for predicting recurrent prostate cancer. *IEEE Trans Med Imaging* 2015;34:284–97.

- Clarke GM, Peressotti C, Constantinou P, Hosseinzadeh D, Martel A, Yaffe MJ. Increasing specimen coverage using digital whole-mount breast pathology: implementation, clinical feasibility and application in research. *Comput Med Imaging Graph* 2011;35:531–41.
- Mount-Campbell A, Hosseinzadeh D, Gurcan M, Patterson E. Applying human factors engineering to improve usability and workflow in pathology informatics. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care* 2017;6:23–7.
- Janowczyk A, Chandran S, Singh R, Sasaroli D, Coukos G, Feldman MD, et al. High-throughput biomarker segmentation on ovarian cancer tissue microarrays via hierarchical normalized cuts. *IEEE Trans Biomed Eng* 2012;59:1240–52.
- Stathonikos N, Veta M, Huisman A, van Diest PJ. Going fully digital: perspective of a Dutch academic pathology lab. *J Pathol Inform* 2013;4:15.
- Gurcan MN, Tomaszewski J, Overton JA, Doyle S, Ruttenberg A, Smith B. Developing the Quantitative Histopathology Image Ontology (QHIO): a case study using the hot spot detection problem. *J Biomed Inform* 2017;66:129–35.