Methods xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Digital pathology and image analysis in tissue biomarker research

Peter W. Hamilton *, Peter Bankhead, Yinhai Wang, Ryan Hutchinson, Declan Kieran, Darragh McArt, Jacqueline James, Manuel Salto-Tellez

Bioimaging and Informatics, Centre for Cancer Research & Cell Biology, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL, Northern Ireland, United Kingdom

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Digital pathology
Biomarker
Drug discovery
Personalized medicine
Tissue microarrays
Tumour analysis
Molecular pathology
Biobank

ABSTRACT

Digital pathology and the adoption of image analysis have grown rapidly in the last few years. This is largely due to the implementation of whole slide scanning, advances in software and computer processing capacity and the increasing importance of tissue-based research for biomarker discovery and stratified medicine. This review sets out the key application areas for digital pathology and image analysis, with a particular focus on research and biomarker discovery. A variety of image analysis applications are reviewed including nuclear morphometry and tissue architecture analysis, but with emphasis on immunohistochemistry and fluorescence analysis of tissue biomarkers. Digital pathology and image analysis have important roles across the drug/companion diagnostic development pipeline including biobanking, molecular pathology, tissue microarray analysis, molecular profiling of tissue and these important developments are reviewed. Underpinning all of these important developments is the need for high quality tissue samples and the impact of pre-analytical variables on tissue research is discussed. This requirement is combined with practical advice on setting up and running a digital pathology laboratory. Finally, we discuss the need to integrate digital image analysis data with epidemiological, clinical and genomic data in order to fully understand the relationship between genotype and phenotype and to drive discovery and the delivery of personalized medicine.

1. Introduction

While it is a common misconception that digital pathology and image analysis is new, research on the use of computers and software for analyzing and measuring cells or tissues in pathology date as far back as the 1960's and 70's [1–4]. That's over 40 years ago! Clearly, the hardware and software systems then were limited in their capability by comparison to today – but those studies were the first to demonstrate the value that computer-based imaging, cellular measurement and quantitation could play in pathological diagnosis and discovery.

As computer hardware advanced rapidly in the 1980's and 1990's, there was considerable promise that image analysis would be embraced as part of routine diagnosis in pathology. Some even posited that the technology would ultimately replace human pathologists. There was enormous investment in automated cytology screening based on IA, with the promise that this could be used to reduce cytology workload and improve diagnostic performance across laboratories worldwide. Clearly this did not happen on the scale predicted and even the most state of the art IA systems failed to significantly change practice in pathology. So the initial enthusiasm for digital IA technology in pathology waivered with the focus shifting to molecular pathology and the promise of diagnostic classification of tissue samples without the need for morphology. Three principle factors changed that: (1) the recognition that molecular pathology still relies heavily on tissue interpretation (2) the drive for targeted therapies based on the presence or absence of tissue-based markers and (3) digital scanning and whole slide imaging (WSI) of entire glass slides in pathology.

The last factor has been hugely instrumental in the recent upsurge in the adoption of image analysis again in both the research and diagnostic sector. Whole slide imaging (WSI), and associated viewing software, allows entire slides to be digitally scanned at high resolution, reviewed by an experienced morphologist, regions selected and image analysis applied to measure specific features. This potentially circumvents the need to use traditional microscopy, manual selection, restricted image capture using a CCD camera, transfer to an image analysis package and subsequent measurement of specific features. WSI can bring these processes together, making image analysis much more practicable and easy to adopt, while facilitating integration into existing workflows in both research and primary diagnostic laboratories.

* Corresponding author.
E-mail address: p.hamilton@qub.ac.uk (P.W. Hamilton).

http://dx.doi.org/10.1016/j.ymeth.2014.06.015
1046-2023/

Industrial image analysis systems have grown dramatically in recent years. This is likely to continue as the applications in discovery, preclinical and clinical research continue to demand quantitative methods, and as new diagnostic tools are translated into diagnostic practice.

This article aims to provide readers with a rapid overview of the current status of digital pathology and image analysis in biomarker research and diagnostic practice, including practical advice on adopting and developing these technologies.

2. Whole slide scanning and digital slides

2.1. Whole slide scanners

While the digital capture of individual images is still utilized widely in the research and tissue diagnostic community, whole slide scanning (WSI) is by far the most rapidly expanding means of digital image capture in pathology. WSI allows the digital capture of the entire tissue sample at high resolution and with appropriate software allows the viewing of the slide at any position and at any magnification. In this way it replicates what is achievable with standard microscopy, but provides a range of additional advantages – including facilitating image analysis.

Over recent years WSI instrumentation has become more widely accepted and affordable in pathology research laboratories and in primary diagnostic laboratories. However, given the pace of development, there are likely to be further systems available from new providers as the market continues to expand.

Most of the systems rely on two variants of image capture (1) line scanning and (2) tile scanning, both of which generate multiple high resolution images (in the form of lines or tiles) that are subsequently aligned or stitched together to create a complete, composite image of the original whole tissue section. Collecting image data by either method is achieved by passing the slide underneath the objective using a carefully controlled motorized scanning stage or objective assembly. The image data is rapidly recorded as the slide is traversed and image data stitched together in real time.

In most systems the magnification at which the slide is scanned can be adjusted. This is commonly either at 20× or 40× magnification. Other select systems can scan under oil at 63× to provide higher resolution systems. 20× scans are sufficient for most standard H&E remote viewing applications although some institutions prefer to scan at 40× to ensure higher resolution. Fig. 1A shows a whole slide scan of a pancreatic cancer, scanned at 40× magnification where the image can be viewed at any magnification (Fig. 1B) and where multiple slides can be viewed side by side for comparison at any location or any magnification (Fig. 1C). Image analysis can benefit from high resolution scans, particularly for applications that involve nuclear detection and analysis. Applications such as in situ hybridization (ISH) can be carried out at 40× with fluorescence but may benefit from high magnification scans in order to resolve individual spots with chromogenic ISH. Haematology applications may require 63× scanning (restricted to certain models of scanner) in order to better resolve morphology and cell types. There is however a storage premium to pay for high resolution scans.

Accurate focus across large areas of tissue during the scanning process is essential. In most instrumentation, this is achieved by mapping the topography variations that inherently exist across even a very thin tissue section, and rapidly adjusting the focus as the scan is being created [5]. The reliability of this process has improved dramatically over recently years and most systems can automatically scan large batches of slides with no human intervention at all.

Some WSI systems can also generate "multiplane" scans, which capture image data along the z-axis (Fig. 2) as numerous large images in a stack. With appropriate viewing software, this provides the ability navigate images in the z-plane, creating a digital focus effect. This is particularly effective for cytology preparations, where the ability to focus is extremely important.

Finally, many scanning systems now offer fluorescence WSI. This makes use of the benefits of fluorescence (see Section 4.6) while providing full slide scans, digitally capturing all relevant data for storage, remote review and image analysis. There are specific challenges associated with fluorescence WSI, not least of which is focus. Fluorescence images tend to contain less contextual background information than brightfield images, and so provide less data to support automated focus over large areas. However most systems provide the ability to select defined regions of interest for scanning, allowing large areas of slides to be successfully scanned under fluorescence.

2.2. Image size and compression

Whole slide digital images are large. Scanning a typical tissue section of 15 × 20 mm in size at 20× viewing magnification (0.24 μm per pixel) can generate images as large as 3.6 GB in size. Scanning at 40× will generate images as large as 14.5 GB. These can be compressed to more manageable sizes (approx. 25:1 compression or more), reducing the file size without impacting on the visual quality of the image. Studies on the compression of images in digital pathology [6] have shown that extensive image compression can be applied without experts being able to visually perceive differences in image quality. Even images with high compression ratios can still be interpretable visually.

An important consideration, however, is how image compression can affect quantitative image analysis. Commercial systems routinely apply different compression methods and levels as part of their standard configuration and so variation from one instrument to the next could be detrimental. Basic studies have shown that densitometric measurement (which is used routinely for quantitative IHC image analysis) is more sensitive to compression that morphological measurement (e.g., nuclear size). Different compression methods offered by different vendors can have very different effects on image analysis fidelity [7]. Kieran et al. [7] showed that with some methods of compression, 5% of the nuclei were segmented in error, with an error rate that steadily increased as compressed image quality decreased. Care therefore needs to be taken to assess the impact of compression artifacts on image analysis, and the impact of compression needs to be validated for each study depending on the features calculated.

2.3. Scanning speeds and automation

Most instruments can now scan slides in 1–3 min, some with the capability of automatically loading multiple slides without user intervention. Some of the larger scanning devices can accommodate in excess of 300 slides, making them ideal for high volume applications, including in busy clinical diagnostic laboratories or large scale tissue research facilities where large numbers of slides need to be scanned and archived daily. Smaller scanners are available, which can scan from 1 to 10 slides in a single action. These are ideal for specialist or incidental research requirements, for educational organizations that are scanning relatively small teaching collections, or for diagnostic labs that want to use digital pathology for infrequent second opinion or frozen section review.

2.4. Storage of digital slides

Given the size of digital slides and the numbers that are now being routinely scanned in many research and diagnostic laboratories, storage represents a significant element of the investment