Original Article

Image microarrays derived from tissue microarrays (IMA-TMA): New resource for computer-aided diagnostic algorithm development

Jennifer A. Hipp, Jason D. Hipp, Megan Lim, Gaurav Sharma, Lauren B. Smith, Stephen M. Hewitt1, Ulysses G. J. Balis

Department of Pathology, University of Michigan, M4233A Medical Science I, 1301 Catherine Ann Arbor, Michigan 48109-0602, 1Laboratory of Pathology, Advanced Technology Center, National Institutes of Health, National Cancer Institute, 8717 Grosvenor Circle, Gaithersburg, MD 20877

E-mail: *Ulysses J. Balis - ulysses@med.umich.edu
*Corresponding author

Received: 26 January 12 Accepted: 01 May 12 Published: 12 July 2012

This article may be cited as:
Hipp JA, Hipp JD, Lim M, Sharma G, Smith LB, Hewitt SM, Balis UG. Image microarrays derived from tissue microarrays (IMA-TMA): New resource for computer-aided diagnostic algorithm development. J Pathol Inform 2012;3:24.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2012/3/1/24/98168

Abstract

Background: Conventional tissue microarrays (TMAs) consist of cores of tissue inserted into a recipient paraffin block such that a tissue section on a single glass slide can contain numerous patient samples in a spatially structured pattern. Scanning TMAs into digital slides for subsequent analysis by computer-aided diagnostic (CAD) algorithms offers the possibility of evaluating candidate algorithms against a near-complete repertoire of variable disease morphologies. This parallel interrogation approach simplifies the evaluation, validation, and comparison of such candidate algorithms. A recently developed pair of digital tools, digital core (dCORE), and image microarray maker (iMAM) enables the capture of uniformly sized and resolution-matched images, with these representing key morphologic features and fields of view, aggregated into a single monolithic digital image file in an array format, which we define as an image microarray (IMA). We further define the TMA-IMA construct as IMA-based images derived from whole slide images of TMAs themselves. Methods: Here we describe the first combined use of the previously described dCORE and iMAM tools, towards the goal of generating a higher-order image construct, with multiple TMA cores from multiple distinct conventional TMAs assembled as a single digital image montage. This image construct served as the basis of the carrying out of a massively parallel image analysis exercise, based on the use of the previously described spatially invariant vector quantization (SIVQ) algorithm. Results: Multicase, multifield TMA-IMAs of follicular lymphoma and follicular hyperplasia were separately rendered, using the aforementioned tools. Each of these two IMAs contained a distinct spectrum of morphologic heterogeneity with respect to both tingible body macrophage (TBM) appearance and apoptotic body morphology. SIVQ-based pattern matching, with ring vectors selected to screen for either tingible body macrophages or apoptotic bodies, was subsequently carried out on the differing TMA-IMAs, with attainment of excellent discriminant classification between the two diagnostic classes. Conclusion: The TMA-IMA construct enables and accelerates high-throughput multitcase, multifield based image feature discovery and classification, thus simplifying the development, validation, and comparison of CAD algorithms in settings where the heterogeneity of diagnostic feature morphologic is a significant factor.

Key words: TMA, IMA, SIVQ, dCORE, iMAM, WSI, CAD, image analysis
BACKGROUND

Tissue microarrays (TMAs) consist of cores of tissue from preexisting donor blocks that are inserted into a recipient paraffin block, and subsequently assembled into an array of such cores. This technology has facilitated high-throughput immunophenotypic analyses, where a single tissue section on a single glass slide that represents a plurality of patient samples can serve as a powerful tool for parallel discovery.[1] While the use of conventional tissue microarrays (TMA) is well established at both the basic science and translational investigative fronts, there is significantly less technical and operational experience with the use of TMAs in the digital domain in general and high-throughput production settings, in particular. Thus, the opportunity to combine assemblages of key fields-of-view or morphologic features from multiple conventional TMAs into a single and contiguous, reconstituted digital construct is a compelling prospect, with its potential availability leading to improved training, evaluation, comparison, and validation of computer aided diagnostic (CAD) algorithms.

Digital image feature detection, extraction, and aggregation tools (spatially invariant vector quantization (SIVQ), digital core (dCORE), and image microarray maker (iMAM), respectively) were recently developed, as previously reported[2-7] and formed the basis of creating the TMA-IMA construct. Briefly, SIVQ operates as a highly sensitive and specific pattern matching algorithm that can be used to match histological and cytological features across a range of length scales, including nuclear, cellular, and architectural features (i.e., capable of matching features of interest at high, medium, and low magnifications, respectively).[6] dCORE enables the systematic capture of images that contain the desired histopathologic feature(s) of interest from digital slides, while at the same time constraining size and digital image resolution.[7] Finally, iMAM completes the ensemble, with it tailored to create single digital image constructs from a cohort of source digital images, in an array-based format (with this effectively being a digital “TMA” of multiple actual TMAs). We term the overall construct TMA-IMA.

In this brief technical report, we demonstrate the utility of applying these digital tools toward conventional TMAs, to create a single resultant TMA-IMAs, with the assemblages of the many fields-of-view representing cohorts of highly specific morphologic features from multiple subjects (representing a spectrum of morphologic heterogeneity). These images constructs served as the basis for SIVQ-based feature classification.

METHODS

Conventional TMAs (TMA tissue sections on glass slides) of follicular lymphoma (FL) and of follicular hyperplasia (FH) were scanned with a 40× objective (0.25 μm/pixel) using the Aperio XT. Five cores corresponding to five subjects with FH and FL were used to generate the image cores of tingible body macrophages (TBMs) with apoptotic bodies in germinal centers using dCORE.[7] Resultant digital images were assembled into an IMA with iMAM with the resultant digital slide representing an IMA-TMA. This construct was then analyzed by use of SIVQ, as previously described by Hipp and Cheng et al.[2,3,5,6,8]

RESULTS

From the perspective of classical morphology, FH differs from FL by the presence in the former of TBMs and apoptotic bodies in germinal centers that can serve as a discriminant for use with CAD algorithms.[9] To assemble the full spectrum of morphologically heterogeneous TBMs that are normally present in FH, numerous image cores were created with the dCORE tool, from conventional de-identified, multisubject TMAs, which were already available as several digital slides [Figure 1a]. iMAM concatenated these separate images into a single TMA-IMA construct. Five FH image cores were of germinal centers containing TBMs and apoptotic bodies while the five FL cores were of neoplastic follicles. Multiple SIVQ-
three vectors elevated both sensitivity and specificity for TBM and apoptotic body detection, with this observation confirmed by a visual estimate which similarly confirmed the increased density of spatially colocated events, across all three sets of vector event maps.

**DISCUSSION**

The TMA-IMA construct facilitates the creation of highly specific IMAs that contain only the desired cell(s) or based ring vectors were identified such that three were selected for archetypal features of TBMs including large whitish-pink cytoplasm (vector 1), small apoptotic bodies with a blue and white interface (vector 2), and nucleoli of macrophages (vector 3). Subsequent analysis by SIVQ resulted in the generation of three image heatmaps, with resultant false-color output correlating to overall quality of matches [Figures 1b-1d]. While vectors 1, 2, and 3 individually identified 21, 21, and 28 out of a total possible 40 TBMs, respectively, the synergistic effect of using all three vectors elevated both sensitivity and specificity for TBM and apoptotic body detection, with this observation confirmed by a visual estimate which similarly confirmed the increased density of spatially colocated events, across all three sets of vector event maps.
morphic features of interest, as determined by the pathologist. Their availability, in tandem with constitutive diagnostic features of importance being available with sufficient heterogeneity across their natural morphological spectrum, served to accelerate the optimization of a high-throughput, multicase, and multifield feature detection CAD algorithm.

Creating TMA-IMAs constructs that represented the broadest morphological variability intrinsic to diagnostic entities facilitated the discovery of optimal feature classifier vectors that exhibited suitably high sensitivity and specificity. In addition, as the resultant TMA-IMA constructs were considerably reduced in file size, compared to their parent images (owing to their possessing essentially only optimized fields of view), significant improvement in computational speed and efficiency was realized, when carrying out SIVQ analysis.

Advantages of Using the TMA-IMA Construct

dCORE use, targeted toward WSI datasets derived from conventional TMAs, allowed for the creation of image arrays that captured highly specific features of interest with suitable capture of adequate events representing their full morphologic variability. With the ability to render these image libraries as monolithic TMA-IMA constructs, it became possible to easily create large and diverse cohorts of reference images (for both normal and diseased states), making CAD development and validation a significantly easier process. In addition, digital monolithic slides rendered as IMA-TMA constructs were several log smaller than the combined data size represented by the cohort of native TMA images. Significant reductions in file size, as encountered here, can be reasonably expected to contribute to reduced local data storage requirements and expedited transmission of images over networks; both of these improvements positively contributing toward simplified collaborative use of such digital media for research and education.

We anticipate that the use of TMA-IMA-like constructs will improve the efficiency of translational studies, which have curated collections of cores distributed over multiple slides (e.g., our studies have used up to 13 arrays for one cohort) and multiple TMAs that have replicate stains of the same antibody. The improvement will be a direct consequence of image-based analyses being able to be condensed down into computational queries on as few as even a single TMA-IMA construct. Lastly, using TMA-IMA-like constructs to screen for CAD algorithms will have a significant advantage over using TMAs alone to screen for molecular markers, given that the former does exhaust the specimen (tissue block).

Challenges of Working with Conventional TMAs

The creation of TMAs can result in image tiles that lack the desired histo- and cyto-morphologic features of interest; this is inherent in the stochastics of sub-

sampling large data sets.\textsuperscript{[10,11]} Because of the potential for imprecise tissue selection (tissue coring), multiple samples are often taken to ensure adequate and accurate sampling.\textsuperscript{[12]} By comparison, use of the TMA-IMA construct ensures that the most important morphologic predicate features are consistently present, and moreover, that they are present in suitably broad morphologic diversity. This assertion is validated in Figure 1a, which is representative of the typical construct made possible by the tools described herein. Although commercial image software packages have enabled users to create montage images of TMA cores for display purposes, they have not as of yet been optimized to address the compounded challenge of manipulating multiple WSI source data files to render a new WSI file. Finally, these applications have yet to simplify the execution of tasks such as image analysis, feature detection, feature quantification or CAD on very large file formats. Clearly, such capabilities are de rigueur prior to the specialty realizing maximum utility with respect to WSI data.

Future of Conventional TMAs

Significant time and financial investments have been made in creating TMA blocks for molecular studies. Scanning and digitizing these powerful resources for future TMA-IMA studies is a byproduct of initial TMA creation, and the CAD algorithm community should consider piggybacking off the efforts of the TMA community. In addition, TMA-IMA-like constructs should be made publicly available, with the previously described digital slide repository (www.WSIRepository.org) resource perhaps being an effective method for doing so.\textsuperscript{[14]}

CONCLUSION

In summary, the TMA-IMA construct, in concert with a supporting suite of high-throughput image manipulation tools, holds significant promise to accelerate the pace of discovery for algorithm development, algorithm validation, and algorithm comparison, in settings where significant numbers of fields-of-view and additional morphologically disparate cases are required to generate suitably high statistical power.

dCORE was used to extract image cores consisting of representative fields of germinal centers from FH and FCL and these images were subsequently aggregated into an IMA using the IMAM tool (A). Three ring vectors were then used to analyze the image with SIVQ and generate heatmaps, where the paint corresponds to the quality of matches. Ring vector 1 was selected to identify features with large cytoplasm with a white and pink interphase (1B). Ring vector 2 was selected to identify and search for apoptotic bodies with a blue-white interphase (1C). Ring vector 3 was selected to identify and search for nuclei of macrophages (1D). When these three heatmaps were taken in aggregate, the visual estimate of all three
sets of vector events confirmed a synergistic effect for both elevating sensitivity and specificity for TBM and apoptotic body detection.

REFERENCES

1. Avninder S,Ylaya K, Hewitt SM. Tissue microarray: a simple technology that has revolutionized research in pathology. J Postgrad Med 2008;54:158-62.
2. Hipp J, Cheng J,Daignault S, Sica J, Dugan MC, Lucas D, et al. Automated area calculation of histopathologic features using SIVQ. Anal Cell Pathol (Amst). 2011. [In press]
3. Hipp J, Cheng J, Hanson JC,Yan W,Taylor P, Hu N, et al. SIVQ-aided laser capture microdissection: A tool for high-throughput expression profiling. J Pathol Inform 2011;2:19.
4. Hipp J, Monaco J, Kunju LP, Cheng J, Yagi Y, Rodriguez-Canales J, et al. Integration of architectural and cytologic driven image algorithms for prostate adenocarcinoma identification. Anal Cell Pathol (Amst) 2012. [In press].
5. Hipp J, Smith SC, Cheng J, Tomlins SA, Monaco J, Madabhushi A, et al. Optimization of complex cancer morphology detection using the SIVQ pattern recognition algorithm. Anal Cell Pathol 2012;35:41-50.
6. Hipp JD, Cheng J, Toner M, Tompkins RG, Balis UJ. Spatially Invariant Vector Quantization: A pattern matching algorithm for multiple classes of image subject matter including pathology. J Pathol Inform 2011;2:13.
7. Hipp J, Cheng J, Pantanowitz L, Hewitt S, Yagi Y, Monaco J, et al. Image Microarrays (IMA): Digital pathology’s missing tool. J Pathol Inform 2011;2:47.
8. Cheng J, Hipp J, Monaco J, Lucas DR, Madabhushi A, Balis UJ. Automated vector selection of SIVQ and parallel computing integration MATLAB™: Innovations supporting large-scale and high-throughput image analysis studies. J Pathol Inform 2011;2:37.
9. Ramzy I. Clinical cytopathology and aspiration biopsy. New York: McGraw-Hill; 2001.
10. Hewitt SM. The application of tissue microarrays in the validation of microarray results. Methods Enzymol 2006;410:400-15.
11. Shou JZ, Hu N, Takikita M, Roth MJ, Johnson LL, Giffen C, et al. Overexpression of CDC25B and LAMC2 mRNA and protein in esophageal squamous cell carcinomas and premalignant lesions in subjects from a high-risk population in China. Cancer Epidemiol Biomarkers Prev 2008;17:1424-35.
12. Fergenbaum JH, Garcia-Closas M, Hewitt SM, Lissowska J, Sakoda LC, Sherman ME. Loss of antigenicity in stored sections of breast cancer tissue microarrays. Cancer Epidemiol Biomarkers Prev 2004;13:667-72.
13. Hipp JD, Lucas DR, Emmert-Buck MR, Compton CC, Balis UJ. Digital slide repositories for publications: lessons learned from the microarray community. Am J Surg Pathol 2011;35:783-6.
14. Hipp JD, Sica J, McKenna B, Monaco J, Madabhushi A, Cheng J, et al. The need for the pathology community to sponsor a whole slide imaging repository with technical guidance from the pathology informatics community. J Pathol Inform 2011;2:21.