ImageJS: Personalized, Participated, Pervasive, and Reproducible Image Bioinformatics in the Web Browser

Jonas S. Almeida, James R. Hackney

Department of Pathology, Division of Informatics, University of Alabama at Birmingham, Birmingham, AL. E-mail: jalmeida@uab.edu

The ImageJS tool reported in the July 2012 issue of Pathology Informatics (J Pathology Informatics 3:25) describes a successful attempt to use a Web App ecosystem approach for the dissemination/deployment of Pathology Informatics applications. This approach, validated there with an image analysis application, only became a realistic proposition with the development of the modern web browser combined with the recent emergence of the third generation of Web technologies (Web 3.0). The key advantages of this approach derive from the browser’s reliance on the code migrating to the machine where the access to the data already exists. This decreases the exposure of sensitive data, which no longer leave the protected clinical informatics environment. It also facilitates having multiple partners, for example, computational statisticians, to develop individual modules which the user can then compose as needed for specific applications.

The basic module of ImageJS, which can be obtained by directing the browser to imagejs.org, includes the module orchestration methods that will respond to the URL of additional modules. This illustration can also be visualized in the two webcasts listed in that web page where multiple combinations of modules, for example to calculate cellular proliferation in KI67 labeled images, are assembled into a single link. As a result, a pathologist would only have to direct the browser to a tailored link to assemble an advanced image analysis application configured for a specific problem: with no download or installation. This gives substance to the argument that Web App ecosystems, such as ImageJS, have significant advantages over more conventional approaches, such as NIH’s ImageJ, as a vehicle to deploy Pathology Informatics applications.

CONTENT

This presentation describes and expands on the ImageJS tool reported in the July 2012 issue of Pathology Informatics (J Pathology Informatics 3:25). Modern web browsers now include code interpreters and support code distribution architectures that are advantageous as a computational environment to deliver informatics applications. In a nutshell, this presentation explores the use of web Apps to deliver image analysis directly (no downloads or installations) to where it is needed. We will argue, with examples such as determining proliferation from KI67 labeling images, that this approach has important advantages over conventional systems such as NIH’s ImageJ.

TECHNOLOGY

Only web technologies that are natively supported by the browser are used (HTML5 and JavaScript), combined with an open architecture that invites distributed application development.

DESIGN

A core module is delivered with the imagejs.org URL with analytical and user interface components being then loaded with their own script tags. The effect is that a particular image analysis solution is automatically configured from a string of URL’s, blurring the distinction between a “personalized” application and an analytical workflow.

RESULTS

A configurable Web App ecosystem for image analysis,
inviting module development by statisticians and human-machine interface researchers alike, is made available with open source and in the public domain. This result is detailed in a July 2012 Pathology Informatics report, and is also illustrated by a number of YouTube webcasts, including the two listed in http://imagejs.org. This ability of non-disruptively delivering advanced computational statistics applications to the point of care also caught the attention of the media which described it as the “angry birds” approach.

CONCLUSION

The emergence of the modern web browser coupled with the more open architectures of the third generation of Web technologies creates novel opportunities for pathology informatics. The ImageJS application to image analysis suggests that ecosystems of Web Apps have fundamental advantages as informatics solutions as concerns deployability, configurability, reproducibility, and protection of sensitive personal health information (PHI).

ClusterFASTQ: A Method for the Identification of Translocations in Clinical Next Generation Sequencing Data

Eric J. Duncavage¹, Haley Abel²

¹Department of Anatomic and Molecular Pathology, ²Division of Statistical Genetics, Washington University College of Medicine, St. Louis, MO. E-mail: eduncavage@path.wustl.edu

CONTENT

The identification of gene translocations has both diagnostic and prognostic significance in molecular oncology. Current clinical methods for translocation detection generally rely on low-resolution techniques such as fluorescence in-situ hybridization (FISH), however, recent work has demonstrated that translocations can be identified by targeted next generation sequencing (NGS). Existing methods for translocation detection in NGS data involve the identification of discordant paired-end sequencing reads, however, these methods are subject to a high false-positive rate that makes them unsuitable for clinical molecular diagnostics. ClusterFASTQ improves upon these methods by first identifying clusters of discordant paired reads, then validating the clusters by re-mapping the unmapped or incompletely mapped partners of nearby reads. Finally, it assembles all reads mapping to the vicinity of the cluster into a contig, typically 300-3000 bp long.

TECHNOLOGY

ClusterFASTQ is a command line utility implemented in Java. It accepts aligned sequence (BAM) files as input and outputs the chromosomal positions of both translocation partners as well as a contig spanning the breakpoint sequence.

DESIGN

Many existing methods for the detection of translocations in NGS data rely on the identification of discordant paired-end reads where paired reads map to different chromosomal regions. Although such methods are sensitive and computationally efficient, they produce a high number of false-positive results due to repetitive elements in the human genome. Other methods reduce the false-positive rate by looking for chimeric single-end reads, but the performance can be heavily dependent on the choice of aligner, and the methods generally do not produce large breakpoint-spanning contigs. ClusterFASTQ improves upon these methods by first identifying clusters of discordant paired reads, then validating the clusters by re-mapping the unmapped or incompletely mapped partners of nearby reads. Finally, it assembles all reads mapping to the vicinity of the cluster into a contig, typically 300-3000 bp long.

RESULTS

DNA sequences from eight formalin-fixed cancers with known ALK or MLL translocations were analyzed by both Breakdancer and ClusterFASTQ. Both tools correctly identified the translocations, however, Breakdancer produced an average of ~1200 putative translocations/case whereas ClusterFASTQ identified only one/case.

CONCLUSION

ClusterFASTQ allows for the rapid detection of translocations from targeted NGS data. Using a secondary single-end read re-alignment step, ClusterFASTQ greatly reduces the number of false-positive translocation results obtained, thereby permitting NGS-based translocation detection in the clinical laboratory.

Patients Accessing Laboratory Results via Patient Portal: What are the Risks?

Maggie L. Hopkins¹, Alexis B. Carter²

¹Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, ²Department of Pathology and Laboratory Medicine, Emory University School
CONTENT

Patient portals (PTPRs) are web portals typically offered by a healthcare entity to allow patients to view their entity-related medical information. The use of PTPRs is increasing, in part because of programs established by the US government including meaningful use (MU) of electronic health records. Several objectives of MU Stage 1 and recently finalized MU Stage 2 require that a percent of patients have access to a PTPR that includes lab results. However, controversy exists on whether direct release of medical information exposes patients to risks of misinterpretation or other harm. To answer this question, the authors undertook a review of the existing medical literature.

TECHNOLOGY

PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) and SurveyMonkey (http://www.surveymonkey.com) were used to review the literature and collect data, respectively.

DESIGN

Recent medical literature on PTPRs was retrieved from PubMed by searching for terms in quotes: “patient portal” and “patient portals.” A total of 98 articles were retrieved and filtered down to 34, based on whether the title was relevant to the topic. Subsequently, two independent reviewers reviewed each of the 34 articles. Articles were excluded if their focus was on manual entry of data by the patient (e.g., personal health record), electronic visits or if the full text could not be acquired. After exclusion, 10 articles remained which were categorized according to their focus.

RESULTS

The papers were reviewed for whether or not they covered the following items, and the percent which did address was recorded: Lab results available: 50% (5), ability to access portal: 70% (7), described patient demographics: 50% (5), reported access disparities: 40% (4), described usability: 60% (6), access to sensitive results: 20% (2), compliance with laws or standards: 10% (1), specific vendor platforms: 10% (1) [Table 1].

CONCLUSIONS

Peer-reviewed literature indicates that patients are most likely to use PTPLs to view laboratory results. Despite this, peer-reviewed articles examining how patients interpret and use laboratory results are lacking. Limitations of this study include lack of standard terminology and overlap with “personal health record(s),” thereby likely artificially reducing the number of articles for review. Regardless, much more work is needed to determine compliant standards by which laboratory data should be displayed in PTPLs for ease of patient comprehension.

Quantification of Epidermal Growth Factor Receptor Expression in Colorectal Cancer using Image Analysis

Ryan A. J. Hutchinson1, Jacqueline A. James1, Richard Wilson1, Richard A. Adams2, B. Jasani2, Manuel Salto-Téllez1, Peter W. Hamilton1

1Centre for Cancer Research and Cell Biology, School of Medicine, Dentistry and Biomedical Science, Queen’s University Belfast, Belfast, United Kingdom; 2Institute of Cancer and Genetics, Section of Oncology and Palliative Medicine, Cardiff University School of Medicine, Cardiff, United Kingdom. E-mail: rhutchinson09@qub.ac.uk

CONTENT

The COIN trial was a major multinational study examining the comparison of continuous chemotherapy with and without cetuximab to intermittent chemotherapy as a first-line therapy in patients with previously untreated, advanced colorectal cancer. Original findings from the COIN trial suggested that pathology based visual scoring of epidermal growth factor receptor (EGFR) expression using immunohistochemistry (IHC) in advanced colorectal cancer tissue microarrays did not indicate any predictive value for treatment with...
cetuximab and chemotherapy in first line therapy. The aim of our study was to examine whether automated IHC image analysis has the potential to provide new evidence as to the utility of EGFR IHC as a predictive biomarker.

TECHNOLOGY

The complete set (27 slides) of IHC-stained, advanced colorectal cancer tissue microarray slides was digitized using an Aperio ScanScope CS2 scanner. Representative regions were randomly marked up across 23 tissue microarray cores by the original COIN trial pathologist. A total of 70 regions were annotated which were representative of a range of staining intensities. The digitized, IHC-stained slides were imported into Definiens Tissue Studio for EGFR quantification. This cohort of marked-up tissue microarray cores was used as a training set to develop the algorithm. A region recognition method applied to EGFR IHC images identified histological regions and a novel image analysis solution was created to identify positive membrane staining using subsets of identified regions within the EGFR IHC images.

RESULTS

The defined algorithm was capable of accurately segmenting and measuring positive cell membrane EGFR expression across the range of colorectal cancer samples. This allowed for the successful selection of epithelial tumour cells and exclusion of positivity in stromal/non-tumour regions. A direct comparison between visual IHC scores and computerized image analysis-derived IHC scores showed a strong correlation ($r = 0.96$, $P < 0.0001$). Some results suggested that visually marked up cores with a lower histological score were scored higher in some instances using image analysis.

CONCLUSION

The application of image analysis facilitates the automated quantification of the EGFR IHC expression, providing accurate membrane segmentation and measurement. Although visual and automated IHC evaluation showed strong concordance in this sample set, the clear advantage of image analysis is its reproducibility and consistency across large numbers of samples. Having now established a reliable image analysis method for quantification of EGFR expression in colorectal cancer, future work will examine the role of EGFR image analysis in predicting response to cetuximab in the COIN trial patient cohort.

A Digital Imaging and Communication in Medicine Prototype in XML with Relationships

Robert C. Leif, Stephanie H. Leif

Newport Instruments, San Diego, CA. E-mail: rleif@rleif.com

CONTENT

CytometryML consists of XML schemas that describe the specimen processing, acquisition, and results of cytometric measurements. Presently, these measurements can be performed with a flow cytometer or digital microscope. Because of the considerable overlap between cytometry and digital microscopy, data-types have been incorporated from both the International Society for Advancement of Cytometry (ISAC) and the Digital Imaging and Communication in Medicine (DICOM) standards. The DICOM separation of studies and instances has been included. The architecture of CytometryML now includes the capacity to describe relationships in a manner, which is similar to that of the resource description framework (RDF) and the ZIP file container for both the meta- and binary data and is based on the popular EPUB standard.

TECHNOLOGY

The schemas are written in the XML Schema Definition (XSD1.1) language and validated to demonstrate adherence to XSD1.1. Their content was tested by translating specific XSD elements into XML and filling in the values of the objects contained therein. The use of an element-based implementation of relationships permits bidirectional and multiple relationships between two objects to be expressed.

DESIGN

Modularity of the design was enhanced by basing the schemas on objects. The DICOM hierarchical separation of a metadata-containing series and separate instances, which also include binary data, was maintained. Because of compatibility with draft Supplement 161, CytometryML should be able to read XML data that is stored on the picture archiving system, PACS while minimally loading the server, as well as being compatible with future efforts to write XML data on the PACS.

RESULTS

An XML-based system that incorporates data-types from existing standards and provides enhanced, but simple-to-
understand relationships has been created. Preliminary data indicate that these XML data-types can be used with XHTML5, which would permit the creation of a medical informatics system that has access to the full power of the Internet.

CONCLUSION

CytometryML can be considered as a basis for a collaborative effort between pathologists and cytometrists for development of a continuum of complimentary interoperable standards and a prototype of a future Internet-based version of DICOM, DICOM 4.

Best Block Designation in Surgical Pathology: A Help or a Hindrance to Subsequent Molecular Studies?

Andrew M. Quinn, Frank Kuo
Department of Pathology, Brigham and Women’s Hospital, Boston, MA. E-mail: andrew.m.quinn@gmail.com

CONTENT

Selecting blocks with tumor suitable for testing is a time-consuming step in molecular diagnostics workflow. In order to improve turnaround time, our department instituted a quality improvement initiative in 2010 by which surgical pathologists select a “best block” (BB) for molecular testing and record its tumor content during case finalization. The purpose of this study was to compare pre-analytical and analytical processing intervals between molecular cases for which a BB was designated and those for which it was not.

TECHNOLOGY

Two partially automated, real-time tracking systems were employed: an in-house, web-based laboratory information management system for following pre-analytical and analytical molecular case processing time points, and an in-house enhancement to Sunquest PowerPath (Sunquest Information Systems, Tucson, AZ, USA) for BB designation and characterization. Time points were logged via electronic verification within the web-based laboratory information management system. Data analyses were performed via GraphPad Prism (GraphPad Software, Inc., San Diego, CA, USA).

RESULTS

A summary of the results is presented in Table 1.

CONCLUSIONS

“BB” designation at the time of surgical pathology case finalization results in significant improvement in the overall turnaround time for subsequent molecular analyses. The improvement is best realized in the pre-analytical phase of reviewing a case, but may reflect selection bias. Surprisingly, BB not used cases have significantly shorter pre-analytical intervals than BB used cases, suggesting a confounding variable. The variability of numerous pre-analytical intervals (data not shown) indicates that further refinements and subsequent review of the tracking process is needed for optimization.

Table 1: Effect of prior BB designation on molecular case processing intervals*

| Group      | Report review to first block selecteda | Report review to material arrival in molecular labb | Material arrival in molecular lab to molecular report finalizationc | Total turnaround timeb |
|------------|---------------------------------------|---------------------------------------------------|-----------------------------------------------------------------|------------------------|
| BB used    | 21 (7-80)                             | 40 (25-72)                                        | 64 (55-72)                                                      | 112 (91-144)           |
| BB not used| 21 (7-120)                            | 31 (4-77)                                         | 59 (49-68)                                                      | 88 (59-119)            |
| No BB      | 214 (19-816)                           | 132 (48-399)                                      | 63 (55-70)                                                      | 172 (106-390)          |

*aData shown as median interval number of days with interquartile range in parentheses. *Non-parametric analysis of variance and all post-hoc comparison P values significant, except for comparison between BB used and BB not used. *Non-parametric analysis of variance and all post-hoc comparison P values significant. *Non-parametric analysis of variance and all post-hoc comparison P values significant, except for comparison between BB used and no BB