

## Emerging Tools for Computer-Aided Diagnosis and Prognostication

Scott Ritter\* and Kenneth B Margulies

Department of Medicine, Division of Cardiovascular Medicine, Cardiovascular Institute, Perelman School of Medicine, University of Pennsylvania

### Emerging Tools for Computer-aided Diagnosis and Prognostication

The ability to more accurately predict and prevent disease has the potential to transform clinical practice by improving response to specific treatment regimens and decreasing morbidity and mortality. Part of what limits the accuracy to which we can predict and prevent disease results from our limited understanding of the relationship between clinical presentation and disease progression [1].

Although vast amounts of data are collected at clinical presentation, ranging from macro-scale Magnetic Resonance Imaging (MRI) scans, to micro-scale pathology slides, to nano-scale proteins and genes, there are challenges associated with analyzing, combining, and correlating these data to make diagnostic, prognostic, and theranostic predictions [2-4]. Computerized image analysis and data integration methods have the potential to improve our understanding of the relationship between these heterogeneous multi-format, multi-scale data to better predict disease outcomes and treatment responses.

### Computer-based Image Analysis

Advances in imaging hardware and computational processing have catalyzed the growth of digital imaging and computer-based image analysis in pathology. Digitization of entire glass slides (whole-slide imaging) has increased the amount of morphologic data that can be obtained from tissue [3]. Whole-slide imaging has also aided pathologists with automated field selection and has begun to allow pathologists to supplement steps in image analysis (i.e., feature extraction, feature selection, dimensionality reduction, and classification) with automated machine-learning algorithms to minimize subjectivity and augment quality assurance [3,5,6].

One such tool, developed, evaluated, and applied by Beck et al., is an unbiased image analysis system called C-Path [7]. C-Path has been used to identify feature sets in tissue microarrays to predict 5-year survival of patients with breast carcinoma. Using a machine-learning algorithm and thousands of morphologic descriptors, the C-Path prognostic model accurately predicted good and poor prognosis patients and identified clinically significant morphologic features, some of which were not previously recognizable using traditional quantitative pathology techniques. Although the molecular basis for the prognositically significant morphologic phenotypes has yet to be elucidated, and the effectiveness of computer-aided pathological interpretation has yet to be established on whole-slide images and tested on a diverse set of images, this approach shows great potential because it has predicted survival outcomes with a high degree of statistical significance and has the potential for further refinement. This example illustrates the potential for using automated, unbiased image analysis and machine-learning systems for producing standardized, objective, reproducible results that could eventually support clinical practice [8].

### Heterogeneous Data Integration

Advances in computational processing have enabled quantitative integration of heterogeneous, multi-format, multi-scale data—particularly imaging and genomic data [2,9-12].

In one of the first applications to combine imaging and non-imaging (protein expression) data, Lee and Madabhushi developed a Generalized Fusion Framework (GFF) to integrate the micro-scale morphological features obtained from digital histopathology slides with nano-scale protein expression measurements from mass spectrometry [13]. This GFF was created to observe whether quantitative integration of image-based signatures from digital histopathology slides with corresponding peptide measurements from mass spectrometry could be used to differentiate prostate cancer progressors with prostate cancer non-progressors. The challenge of integrating this multi-scale, multi-modal, multi-protocol data was overcome by combining the 3 data modalities (architectural histopathology features, morphological histopathology features, and m/z mass spectrometry features in 51, 100, and 570 dimensions, respectively) into a common low-dimensional meta-space projection with 3 dimensions using principal component analysis. This projection was then normalized, concatenated, and reduced a second time with principal component analysis to yield the low-dimensional integration product of the original high-dimensional data. Results reflected the suitability of using this GFF to integrate heterogeneous multi-format, multi-scale data for differentiating between patients with different disease profiles.

Later applications by Madabhushi et al., have explored additional methods for combining data modalities beyond principal component analysis (e.g., non-linear dimensionality reduction methods) and correlations between disease and markers in digital pathology [10], gene and protein expression [11], spectroscopy [12,14], ultrasound [15], and MRI [9,14,16].

### Future Directions

While computer-based image analysis, heterogeneous data integration methods, and computer-aided prognostics are currently demonstrating their efficacy in the pre-operative or pre-therapeutic cancer population, they will inevitably have applicability in other fields.

In cardiovascular medicine, for instance, large amounts of macro-scale heart morphology and phenotype data (from MRI, hemodynamics, and echocardiograms), micro-scale whole-slide imaging data (from biopsies, donors, explants, and device placements), and nano-scale gene expression and transcriptome data are being collected at several institutions for clinical and research purposes [17]. Because typical cardiac pathology scoring systems are rather

\*Corresponding author: Scott Ritter, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, USA, Tel: (215) 573-2999; Fax: (215) 746-7415; E-mail: [sritter@mail.med.upenn.edu](mailto:sritter@mail.med.upenn.edu)

Received February 21, 2014; Accepted February 22, 2014; Published February 24, 2014

Citation: Ritter S, Margulies KB (2014) Emerging Tools for Computer-Aided Diagnosis and Prognostication. J Clin Trials 4: e117. doi:10.4172/2167-0870.1000e117

Copyright: © 2014 Ritter S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

rudimentary, such as the Dallas criteria for myocarditis [18] and the International Society for Heart and Lung Transplantation scoring of rejection in cardiac allografts [19], there is rich opportunity for computer-aided interpretation and multi-modality integration to provide new insights into myocardial disease mechanisms, severity and prognosis. As with the oncology applications described above, a key step in these myocardial applications will be correlation with clinical outcomes and current clinical reference standards. As heterogeneous data integration tools become increasingly sophisticated and validated, they could provide a rational basis for the identification of interpatient distinctions necessary for greater individualization of therapeutics.

Computers are becoming increasingly ready to supplement and enhance imaging (MRI, ultrasound), morphologic information (tissue), and molecular classification (whole-genome sequencing, expression profiling, proteomics, and metabolomics) with diagnostic, prognostic, and therapeutic predictions [8]. These computer-based tools for heterogeneous data integration have begun to demonstrate their effectiveness in large retrospective studies and will soon be ready for prospective, multi-institutional validation studies as the next step before adoption into clinical practice.

## Funding

This work was supported by the Myocardial Applied Genomics Network (MAGNet) National Institutes of Health grant R01HL105993.

## References

1. Madabhushi A, Doyle S, Lee G, Basavanahally A, Monaco J, et al. (2010) Integrated diagnostics: a conceptual framework with examples. *Clin Chem Lab Med* 48: 989-998.
2. Madabhushi A, Agner S, Basavanahally A, Doyle S, Lee G (2011) Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data. *Comput Med Imaging Graph* 35: 506-514.
3. Ghaznavi F, Evans A, Madabhushi A, Feldman M (2013) Digital imaging in pathology: whole-slide imaging and beyond. *Annu Rev Pathol* 8: 331-359.
4. Cappola TP, Margulies KB (2011) Functional genomics applied to cardiovascular medicine. *Circulation* 124: 87-94.
5. Schnitt SJ, Connolly JL, Tavassoli FA, Fechner RE, Kempson RL, et al. (1992) Interobserver reproducibility in the diagnosis of ductal proliferative breast lesions using standardized criteria. *Am J Surg Pathol* 16: 1133-1143.
6. Wei BR, Simpson RM (2013) Digital pathology and image analysis augment biospecimen annotation and biobank quality assurance harmonization. *Clin Biochem*.
7. Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, et al. (2011) Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med* 3: 108ra113.
8. Rimm DL (2011) C-path: a Watson-like visit to the pathology lab. *Sci Transl Med* 3: 108fs8.
9. Litjens G, Toth R, van de Ven W, Hoeks C, Kerkstra S, et al. (2014) Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge. *Med Image Anal* 18: 359-373.
10. Lewis JS Jr, Ali S, Luo J, Thorstad WL, Madabhushi A (2014) A quantitative histomorphometric classifier (QuHbIC) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma. *Am J Surg Pathol* 38: 128-137.
11. Lee G, Rodriguez C, Madabhushi A (2008) Investigating the efficacy of nonlinear dimensionality reduction schemes in classifying gene and protein expression studies. *IEEE/ACM Trans Comput Biol Bioinform* 5: 368-384.
12. Tiwari P, Rosen M, Reed G, Kurhanewicz J, Madabhushi A (2009) Spectral embedding based probabilistic boosting tree (ScEPTre): classifying high dimensional heterogeneous biomedical data. *Med Image Comput Comput Assist Interv* 12: 844-851.
13. Lee G, Doyle S, Monaco J, Madabhushi A, Feldman MD, et al. (2009) A knowledge representation framework for integration, classification of multi-scale imaging and non-imaging data: Preliminary results in predicting prostate cancer recurrence by fusing mass spectrometry and histology. *Proceedings of the International Symposium on Biomedical Imaging*.
14. Tiwari P, Kurhanewicz J, Rosen M, Madabhushi A (2010) Semi supervised multi kernel (SeSMiK) graph embedding: identifying aggressive prostate cancer via magnetic resonance imaging and spectroscopy. *Med Image Comput Comput Assist Interv* 13: 666-673.
15. Sparks R, Bloch BN, Feleppa E, Barratt D, Madabhushi A (2013) Fully automated prostate magnetic resonance imaging and transrectal ultrasound fusion via a probabilistic registration metric. *Proc Soc Photo Opt Instrum Eng*.
16. Tiwari P, Viswanath S, Kurhanewicz J, Sridhar A, Madabhushi A (2012) Multimodal wavelet embedding representation for data combination (MaWERiC): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. *NMR Biomed* 25: 607-619.
17. Myocardial Applied Genomics Network (MAGNet) (2011).
18. Baughman KL (2006) Diagnosis of myocarditis: death of Dallas criteria. *Circulation* 113: 593-595.
19. Stewart S, Winters GL, Fishbein MC, Tazelaar HD, Kobashigawa J, et al. (2005) Revision of the 1990 working formulation for the standardization of nomenclature in the diagnosis of heart rejection. *J Heart Lung Transplant* 24: 1710-1720.