

Computational Pathology: Unlocking Tissue Content in Precision Medicine

Chukka Srinivas, Scientific Fellow, Ventana Medical Systems Inc. (Roche Tissue Diagnostics), Mountain View, CA 94043.

Traditionally, pathologists detect, stage and grade cancer in patient tissue biopsies based on a microscopic manual evaluation of small set of cellular and morphological features in Hematoxylin and Eosin (H&E) stained tissue slides. This information along with an assessment of protein overexpression in the tumor, semi-quantitatively scored from multiple immunohistochemical (IHC) stained slides, is used to subtype the cancer for prognostic and precise treatment selection and planning.

Digital pathology refers to the technologies for digitization of tissue whole slides and image analysis algorithms for automated slide interpretation and precise quantification. So far, the image analysis algorithm development has been focused on generating consistent interpretation and reproducible slide scores, accounting for the inherent challenges of wide biological and staining variability in a clinical setting, but primarily limited to mimicking the manual interpretation process of the pathologists.

There is ever increasing medical evidence that while there is potentially a large amount of prognostic information for a given patient in the tissue, today this information is being analyzed separately from a clinical standpoint without its holistic integration into a single comparative prognostic dataset. Computational pathology is a data-driven pipeline, based on statistical and machine learning methods, for systematic extraction of multi-dimensional information in the digitized tissue slides at multiple scales, statistically combine these features and directly correlating against patient outcome, which is given in terms of patient survival or a response to a drug treatment, to discover discriminant features for prognostication and prediction to a drug response.

In this talk, we propose a computational pathology framework and showcase an end to end application to a specific example of prediction of risk of recurrence in early stage breast cancer patients. Supervised learning based fully automated image analysis algorithms are used to analyze H&E and multiple IHC whole slides and extract an exhaustive set of image features. Based on extracted image features and clinical outcome for a patient cohort, a L1-regularized logistic regression based prognostic model is constructed. We show here there are morphological, relational and co-expression image features which are significantly associated with patient overall survival and could therefore be used to improve prognosis and guide follow up treatment.