The TIES Cancer Research Network (TCRN): Computational Pathology Support for Precision Oncology

Michael J. Becich, Jonathan Silverstein, Michael Davis, Joyce Zeinis, Degan Hao, Melissa Schwenk, Faina Linkov, Nita Mahle, Roni Bollag, Ashish Sharma, Carmelo Gaudioso, Chakra Chennubhotla, Akif Burak Tosun, Tahsin Kurc, Joel Saltz, Jack London, Mathias Brochhausen, Lawrence Tarbox, Fred Prior, Michael Feldman and John Gilbertson


Abstract

Advances in cancer research and personalized medicine require significant new bridging infrastructures, including more robust biorepositories that link human tissue to clinical phenotypes and outcomes data. The Text Information Extraction System (TIES) ([http://ties.dbmi.pitt.edu Cancer Research Network (TCRN)]) is a novel platform developed at the Department of Biomedical Informatics, University of Pittsburgh that enables cancer researchers to mine the content of pathology reports and to share biospecimens across a federated network. Member sites can access pathology data that are de-identified and processed with the TIES natural language processing system (NLP), which creates a repository of rich phenotype data linked to clinical biospecimens. TIES incorporates multiple security and privacy best practices that, combined with legal agreements, network policies, and procedures, enable regulatory compliance. The institutional trust fabric TCRN has developed has been "road tested" by our network and has been used by several funded research projects.

Innovation

TCRN

A deployed imaging analysis pipelines implemented in QuIP and THRIVE to generate large sets of imaging features from whole slide images (WSI). Imaging features will be stored for each batch of images at the image level (i.e., WSI will be partitioned into patches, and imaging features and classification labels will be generated for each patch). The image-level features will be aggregated and summarized at the tumor level. QuIP is used to carry out image classifications and computation of features from segmentation of nuclear material in images. THRIVE is used to identify and quantitatively analyze individual ITH and infer interactions between different cell phenotypes and non-cellular constituents.

TCRN Framework and Trust Fabric

SUSTAINABLE OPEN SOURCE SOFTWARE ARCHITECTURE

Network Site

- University of Pittsburgh Cancer Region (TCRN)
- Stony Brook Cancer Center
- Abrams Research Cancer Center
- Roswell Park Cancer Institute
- Georgia Cancer Center
- SKCC-TSU
- To-Be-Named Site #1
- To-Be-Named Site #2
- Unfunded Additions

Network Site

- NLP
- QuIP
- THRIVE

Each node in the infrastructure includes sharing of whole slide images (WSI) linked to pathology report data (NLP), tissue specimens and outcomes data (Cancer Registry). Quantitative Imaging in Pathology (QuIP) and Tumor Heterogeneity Research Interactive Visualization Environment (THRIVE) modules will provide imaging features (e.g., cyto-epithelial/spatial heterogeneity, host immune response) to enhance deeper search capability for cohorts and specimens. Software service tiers:

- Discovery
  - Cohorts
  - Knowledge
  - Insights

- Tumor Microenvironment
- Tumor Immunology
- Tumor Heterogeneity
- Other Areas of Cancer Research

Figure 3 – TCRN Architecture

TCRN has established an innovative inter-institutional trust fabric, based on TCRN security and regulatory compliance best practices and a TIES adoption blueprint that allows federation of data and biospecimen resources, but these network agreements must be modified to allow for HIPAA-compliant open data sharing with the NCI community in partnership with TCI. An established Policy Working Group for TCRN composed of one member from each of the participating network sites who will establish a path to modifying and implementing this new network agreement and will further develop the existing Material Transfer and Data Use Agreements to allow researchers to efficiently access data and biospecimens. TCRN currently has nearly 7 million fully de-identified, annotated pathology reports linked to tissue specimens.