Longitudinal Visual Analytics for Unpacking the Cancer Journey

Zhou Yuan[1], Sean Finan[2], Jeremy Warner[3], Guergana Savova[2], Harry Hochheiser[1]
[1]University of Pittsburgh, Pittsburgh, PA, USA; [2]Boston Children's Hospital, Boston, MA, USA; [3]Vanderbilt University, Nashville, TN, USA

Abstract

Retrospective cancer research requires identification of patients matching both categorical and temporal inclusion criteria, often based on factors exclusively available in clinical notes. Although natural language processing approaches for inferring higher-level concepts have shown promise for bringing structure to clinical texts, interpreting results is often challenging, involving the need to move between abstracted representations and constituent text elements. We discuss qualitative inquiry into user tasks and goals, data elements and models resulting in an innovative natural language processing pipeline and a visual analytics tool designed to facilitate interpretation of patient summaries and identification of cohorts for retrospective research.

Introduction

The complexities of cancer care create significant challenges for the extraction of information for retrospective research. Review of extensive collections of clinical notes notes can be a laborious interpretive challenge: Although ad hoc solutions such as the "oncologic history" have spontaneously developed as information collection devices, they are not necessarily universal, accurate, or complete[1].

The Cancer Deep Phenotype Extraction (DeepPhe) project is developing informatics solutions to overcome these inefficiencies. DeepPhe uses a combination of classic and state-of-the-art NLP techniques, inference rules, and a rich information model[2], to summarize diagnoses, treatments, responses and temporal relationships needed to support retrospective research[3]. We present a preliminary visualization tool focused on bridging the gap between textual mentions in individual and higher-level concepts of interest to clinical researchers.

Results

Our visualization provides a variety of views ranging from overall patient summaries to specific textual mentions. *Patient details* (Figure 1A) are shown in a series of panels, beginning with patient demographics and summary cancer attributes, treatments, cancer stage, cell line, and TNM values[4]. *Tumor details* (1B) are shown as expandable lists of attributes colored to indicate classes of information (body site, diagnosis, biomarker values, etc.). A toggle above the tumor summary pane supports switching to tabular views when desired. Tumor and cancer details can be selected to reveal individual text spans contributing to the summary element. The *Clinical note timeline* (1C) arranges notes color-coded by episode on a zoomable timeline with one lane for each type of note (progress, radiology, and surgical pathology). Episode labels above the timeline can be clicked to zoom the timeline to documents contained in the specified episode.

Below the timeline, the *explanation panel* (1D) bridges the gap between the inferred attributes of the cancer and tumor summaries (1A and 1B) and the text of the clinical note (1F). The *mention pane* (1E) provides a summary mentions extracted from the selected document. Each mention can be clicked to highlight the appropriate scan in the *note view* (1F).

Navigation through multiple levels of abstraction is illustrated in Figure 1: The selection of tumor summary item "ER+" (1B) led to the display of the "Estrogen Receptor Status" in the explanation pane (1D), and the display of relevant mentions from Report 22 (E). Clicking on the "positive" mention leads to text confirming the ER-positive observation (1F).

The DeepPhe cohort viewer is in early stages of development. Preliminary features include selectable histograms displaying frequencies of cancer stages and biomarker status, and box plots for distribution of age
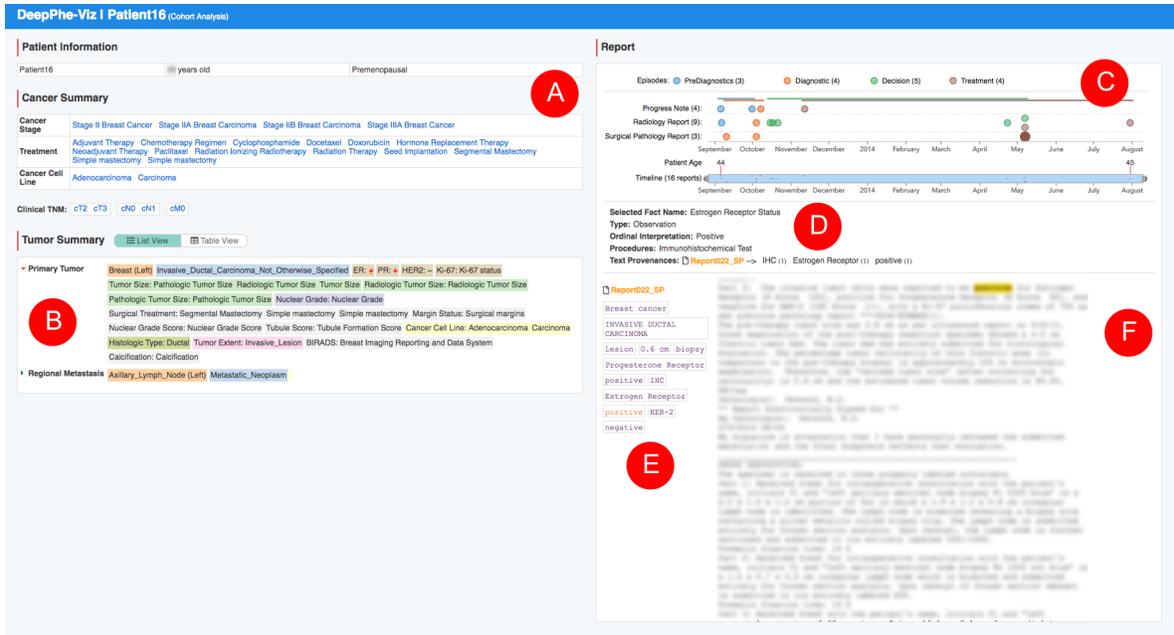
Figure 1: The patient visualization view: (A) Patient demographics and cancer summary. (B) Tumor summary. (C) Interactive timeline displaying documents by type and episode. (D) Inference provenance display. (E) Document-level mentions. (F) Clinical note. Text obscured to protect patient privacy.

at first encounter across stages, as well as distributions of stage and diagnosis information.

Discussion & Conclusion

The DeepPhe visualization tools represent a first step toward a visualization detailed cancer histories that bridges the semantic distance between individual concept mentions in notes and higher-level aggregations. Unlike may previous text analytics tools the DeepPhe tools attempt to combine NLP results with an analytics interface, thus forming a Extension of proposed prototype designs to handle richer data, particularly involving temporal spans, will set the stage for deployment with clinical researchers and subsequent evaluation studies.

Acknowledgments

References

[1] Warner JL, Anick P, Hong P, Xue N. Natural Language Processing and the Oncologic History: Is There a Match? Journal of Oncology Practice. 2011;7(4):e15–e19. PMID: 22043196.

[2] Hochheiser H, Castine M, Harris D, Savova G, Jacobson RS. An information model for computable cancer phenotypes. BMC Medical Informatics and Decision Making. 2016 Sep;16(1):121.

[3] Savova GK, Tseytlin E, Finan S, Castine M, Miller T, Medvedeva O, et al. DeepPhe: A Natural Language Processing System for Extracting Cancer Phenotypes from Clinical Records. Cancer Research. 2017;77(21):e115–e118.

[4] Gospodarowicz MK, Brierley JD, Wittekind C. TNM Classification of Malignant Tumours. John Wiley & Sons; 2017.