Cancer Deep Phenotype Extraction from Electronic Medical Records

Guergana Savova
Boston Childrens’ Hospital/Harvard Medical School

Harry Hochheiser
University of Pittsburgh

Jeremy Warner
Vanderbilt University

U24 CA248010
The Cancer Journey

Risk Assessment
- Early stage, curable (<50%)

Early Diagnosis

Treatment
- Surveillance and Survivorship

Advanced or relapsed, incurable (>50%)
- 1st line
- 2nd line
- 3rd
- 4th

Prevention & Diagnosis
- Treatment
- Palliative Care
- Hospice/EOL

2/4/21
Phenotype Extraction from Clinical Notes

Manual chart review

Automated chart review

Classifier

EHR Notes

HPI: 48 woman with right breast mammogram.
Final diagnosis: Breast, Left, Needle Biopsy:
Invasive Duct Carcinoma.
Patient currently on neoadjuvant therapy with Taxol. Due to cardiomyopathy, patient not candidate for Transtuzumab.

Phenotype Label
• positive/negative
DeepPhe Phase 1: 2014-2020

CANCER CONTAINER
- Body location
- Laterality
- Temporality
- Stage
  - cT, cN, cM
  - pT, pN

TUMOR CONTAINER
- General Attributes
  - Body location
  - Laterality
  - Diagnosis
  - Tumor type (primary, met, distant, regionalMet)
  - Histologic type (ductal, lobular, etc.)
  - Cancer type (carcinoma, adenocarcinoma, etc.)
  - Tumor extent (invasive, in-situ)
  - Calcifications
  - Grade

TUMOR CONTAINER
- Specific to BrCa
- Clockface position
- Quadrant
- ER/PR/HER2 interpretation

TUMOR CONTAINER
- Specific to Melanoma
- Clarks level
- Breslow depth

TUMOR CONTAINER
- Specific to OvCa
- CA-125 value

Extensions Specific to PrCa
- Gleason
- PSA

28 Cancer and Tumor attributes
Invasive Ductal Carcinoma. 4.4 cm
Tumor is ER-, PR-, Her2-.

Patient has triple negative breast cancer.

58 yo F presents to the ER with slurred speech.

2017 DOI: 10.1158/0008-5472.CAN-17-0615
DeepPhe-Viz

Cohort View

Patient View

Zhou, et al.JCO Clinical Cancer Informatics
2020 DOI: 10.1200/CCI.19.00115
SA1: Develop methods for comprehensive extraction of tumor and cancer characteristics based on 5 types of solid and hematologic malignancy

**Phase 1:**
- Breast
- Ovarian
- Melanoma

**Guidelines Annotation**

**Phase 2:**
- Diffuse Large B-Cell Lymphoma (DLBCL)
- Gastrointestinal malignancy

**Methods Development**
- Unsupervised Domain Adaptation

**Gold Standards**

**Evaluation**

**Synoptic Reports**
SA2: Extract longitudinal and context-specific treatment regimen and episode information for each patient’s cancer journey, with normalization to NCIt, RxNorm, and HemOnc
Protocols

Preceding treatment

Breast cancer surgery
Carboplatin & Paclitaxel
Docetaxel monotherapy
…6 more

Current treatment

AC

Subsequent treatment

Breast cancer surgery
CMF
Docetaxel monotherapy
…15 more
SA3: Extract detailed and comprehensive cancer clinical genomics information

“...Final diagnosis: KRAS codon 12 mutation IDENTIFIED (p.Gly12val)...”

Container:
- gene: KRAS
- mutation: codon 12 mutation
- codon: 12
- interpretation: mutation identified
- syntaxP: p.Gly12val
Repeat to determine resistance mechanisms at time of progression or at predefined intervals.

SA4: Create highly usable, interactive tools to help cancer researchers use data extracted by DeepPhe

- Index and search of extracted details?
- Temporal search?
  - Treatment before or after surgery?
- Patient Similarity?
- Temporal alignment with respect to sentinel events?
- “Debiasing” views?

- Discussions with key informants underway
- Input welcome!
SA5: Conduct predefined case studies to validate DeepPhe while answering important questions in cancer care; solicit input from the community for additional case studies.

<table>
<thead>
<tr>
<th>Capability</th>
<th>Breast</th>
<th>Ovarian</th>
<th>Melanoma</th>
<th>Colorectal</th>
<th>DLBCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic tumor facts</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Biomarkers</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Systemic treatments</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal ordering of treatment</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contextual placement of treatment</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Comorbidities</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toxicities</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SA6: Grow DeepPhe’s user community through technical support and user engagement

- **Easing the path to deployment, adoption, and participation**
  - Installation/configuration, containers, APIS, Web presence

- **Understanding and meeting user needs**
  - User groups and meetings

- **Demonstrating possibilities**
  - Success stories/testimonials – based on SA5 projects

- **Reaching out to key constituencies**
  - Sessions at meetings, “office hours”, webinars

- **Engaging with other ITCR-funded projects**
  - Discussion welcome
Conclusions

• Companion Project - DeepPhe-CR
  • Natural Language Processing Platform for Cancer Surveillance (UG3 CA243120), with Kentucky Cancer Registry

• Team
  • Boston Childrens’: **MPI Savova**, Tim Miller, Sean Finan, David Harris, Chen Lin
  • Pittsburgh: **MPI Hochheiser**, John Levander
  • Vanderbilt: **MPI Warner**, Alicia Beeghley-Fadiel, Doug Johnson
  • Dana-Farber: Elizabeth Buchbinder, Danielle Bitterman
  • Mayo: Piet De Groen