Building Cognitive Computing for Healthcare

Patrick McNeillie M.D.
Clinical Lead and Senior Architect on Watson for Genomics
Agenda

• Building Cognitive Computing
  - Defining the Importance of Natural Language Processing (NLP)

• Applying Cognitive to a Complex Problem

• Surfacing Valuable Insights in Big Data
Moving Beyond Jeopardy: Bringing Cognitive Computing to Healthcare
Define Attributes

Clinical Expert

What is the best treatment for a given patient?

Data Set

Lung Cancer
Squamous Cell
PD-L1 Receptor
Stage IIIA

Attribute Extraction

Compare to Prior Knowledge

Applying Cognitive Computing to Complex Problems

Boston University Digital Health Initiative Round Table
Define Attributes

Attribute Extraction

Compare to Prior Knowledge

Clinical Expert

What is the best treatment for a given patient?

Data Set

Natural Language Processing

Stage IIIA

Squamous Cell

Lung Cancer

PD-L1 Receptor

Chieff Complaint/Identifying Information:
The member is 50-60 year old man who has been diagnosed with non-small cell lung cancer clinical stage IIIA, squamous cell. He presented with a 10X increase in his serum level of succinate dehydrogenase. He saw the primary care physician with complaints of a dry cough and neuter grade fever (Temp: 38.3 C). It was his view to several over the counter cold relief. He saw his primary care physician again and was referred to a specialist for further workup and treatment. He was given a prednisone 4 week course of a prednisone. His symptoms improved but returned once he was tapered off the prednisone.
What is the best treatment for a given patient?

Define Attributes
- Stage IIIA
- Squamous Cell
- Lung Cancer
- PD-L1 Receptor

Attribute Extraction
- Natural Language Processing

Compare to Prior Knowledge
- Lung Cancer
- Squamous Cell
- PD-L1 Receptor
- Treatment
- Atezolizumab
- Docetaxel
- Survival: 13.8 Months
- Survival: 9.6 Months

Clinical Expert

Data Set

Chief Complaint/Identifying Information:

The patient is a 50-60 year old man who has been diagnosed with squamous cell lung cancer. His medical history includes a diagnosis of stage IIIA lung cancer, and he has been treated with Atezolizumab, a PD-L1 inhibitor, for several months. He presents with intermittent cough and shortness of breath.

History of Present Illness:

Based on the medical records provided, the patient has a history of smoking. He has recently noted increased shortness of breath and a persistent cough. Physical examination revealed a 2 cm mass in the right lung field. Laboratory tests showed a slight elevation in LDH levels. The patient has been asymptomatic and has not had any prior treatment for lung cancer.

Treatment:
- Atezolizumab
- Docetaxel

Survival:
- Atezolizumab: 13.8 Months
- Docetaxel: 9.6 Months

Boston University Digital Health Initiative Round Table

Applying Cognitive Computing to Complex Problems
What is the best treatment for a given patient?

Chief Complaint/Identifying Information:
The member is 50-60 year old man who has been diagnosed with non-small cell lung cancer. His primary clinical stage is Stage IIIA with right adenopathy, bilaterally. He is positive for PD-L1 on imaging consultation conducted.

History of Present Illness:
Based on the medical records provided, the member is a 50-60 year old male with a 35-pack year smoking history. In 2018, his primary care physician noticed a dry cough and a mild fever that lasted for several days. He was treated with over-the-counter cold relief. He saw his primary care physician again in 2019 and was referred for imaging. He demonstrated a small right lower lobe nodule concerning for pneumonia. He was treated with a course of antibiotics. His symptoms continued and he was treated with a second course of antibiotics. A repeat chest x-ray on XXXXX showed some improvement in the right lung infiltrate. However, as of late April, his symptoms continued. He also went on to develop shortness of breath, dyspnea on exertion, chest pain, and fatigue. He was given a 2-week course of a prednisone. His symptoms improved but returned once he was tapered off of the prednisone.
Differences in attribute extraction between Machine Learning and Rules-Based NLP

**Machine Learning NLP**

- Disease: steroid resistant nephrotic syndrome
  - Subtype: glomerulosclerosis
- Therapy: rxdx-101

**Rules-Based NLP**

- Disease: nephrotic syndrome
  - Subtype: glomerulosclerosis
- Therapy: cortisol

**Rule:** connect disease to therapy if within 10 words of each other
Applying Cognitive to Genomics
Impact on Medical Genomics
### What’s the best treatment given a tumor sequence?

<table>
<thead>
<tr>
<th>Gene</th>
<th>Alteration</th>
<th>Alteration Type</th>
<th>Functional Significance</th>
<th>Annotation</th>
<th>Actionable Gene</th>
<th>Actionable Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDGFRA</td>
<td>Y288C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTEN</td>
<td>K163*</td>
<td>Truncation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGFR2</td>
<td>K292M</td>
<td>SNV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ingest & Learn

Clinical Partners (currently 16 centers)

Genomic Databases

Un-structured Text

Best Practices
Guidelines
Patient Data

Reference Genomes
Molecular
Protein Pathways

Literature Articles
Clinical Trials
Conference Reports

Process Patient Data

Personalized Driver Alterations and Targeted Therapies

Output Report
Our Approach to NLP

Focused Annotation Specialist Team (FAST)

Label Documents

Standardize Ontologies and Dictionaries

Train Machine Learning

Measure Performance

Enable Integration & Reuse

Reduce Bias
Annotating mentions

Epidermal growth factor receptor (EGFR) gene mutations (G719X, exon 19 deletions/insertions, L858R, and L861Q) predict favorable responses to EGFR tyrosine kinase inhibitors (TKIs) in advanced non-small cell lung cancer (NSCLC).

However, EGFR exon 20 insertion mutations (~10% of all EGFR mutations) are generally associated with insensitivity to available TKIs (gefitinib, erlotinib, and afatinib).

The basis of this primary resistance is poorly understood.

We studied a broad subset of exon 20 insertion mutations, comparing in vitro TKI sensitivity with responses to gefitinib and erlotinib in NSCLC patients, and found that most are resistant to EGFR TKIs.

The crystal structure of a representative TKI-insensitive mutant (D770_N771insNPG) reveals an unaltered adenosine triphosphate-binding pocket, and the inserted residues form a wedge at the end of the C helix that promotes the active kinase conformation. Unlike EGFR-L858R, D770_N771insNPG activates EGFR without increasing its affinity for EGFR TKIs.

Unexpectedly, we find that EGFR-A763_Y764insFQEA is highly sensitive to EGFR TKIs in vitro, and patients whose NSCLCs harbor this mutation respond to erlotinib.

Analysis of the A763_Y764insFQEA mutant indicates that the inserted residues shift the register of the C helix in the N-terminal direction, altering the structure in the region that is also affected by the TKI-sensitive EGFR-L858R
Annotating relationships

Unlike EGFR \textit{L858R}, D770\_N771insNPG activates EGFR without increasing its affinity for EGFR TKIs.

Unexpectedly, we find that EGFR\_A763\_Y764insFQEA is highly sensitive to EGFR TKIs in vitro, and patients whose NSCLCs harbor this mutation respond to erlotinib.
Ensuring Agreement: Measuring Kappa Scores

Time

Agreement

100%

Domain Experts
Annotators
Concepts Detection: Compare to a standard and iterate

Model Performance for Labeling Concepts

When a plateau is reached, development teams need to assess if the plateau is secondary to bias or variance to efficiently apply resources and iterate.

F1 - Precision - Recall
Relationship Detection: Compare to a standard and iterate

When a plateau is reached, development teams need to assess if the plateau is secondary to bias or variance to efficiently apply resources and iterate.

Model Performance for Labeling Relationships

- F1
- Precision
- Recall
Unexpectedly, we find that EGFR A763_Y764insFQEA is highly sensitive to EGFR TKIs in vitro, and patients whose NSCLCs harbor this mutation respond to erlotinib.
Clinical Knowledge: Extracted Data from Corpus of Documents
Clinical Knowledge: Extracted Data from Corpus of Documents
### Output Sample from Knowledge Graph Query

<table>
<thead>
<tr>
<th>Gene</th>
<th>Variant</th>
<th>Variant_Type</th>
<th>Functional_Impact</th>
<th>Cancer_Type</th>
<th>NCI Code</th>
<th>Therapy</th>
<th>Response</th>
<th>PMID</th>
<th>Evidence_Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIT</td>
<td>exon 9</td>
<td>p.449-514 mutation</td>
<td>gain-of-function</td>
<td>Gastrointestinal Stromal Tumor</td>
<td>C3668</td>
<td>sorafenib</td>
<td>sensitivity</td>
<td>22270258</td>
<td>Level 2b</td>
</tr>
<tr>
<td>KIT</td>
<td>exon 11</td>
<td>p.550-592 mutation</td>
<td>gain-of-function</td>
<td>Gastrointestinal Stromal Tumor</td>
<td>C3668</td>
<td>sorafenib</td>
<td>response</td>
<td>22270258</td>
<td>Level 2B</td>
</tr>
<tr>
<td>KIT</td>
<td>exon 17</td>
<td>p.788-828 mutation</td>
<td>gain-of-function</td>
<td>Gastrointestinal Stromal Tumor</td>
<td>C3668</td>
<td>nilotinib</td>
<td>sensitivity</td>
<td>1119758, 21456C</td>
<td>Level 2B</td>
</tr>
<tr>
<td>ABL1</td>
<td>T315I</td>
<td>missense mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>AURK inhibitors + BCR-ABL1 inh</td>
<td>response</td>
<td>22772060</td>
<td>Level 2</td>
</tr>
<tr>
<td>ABL1</td>
<td>T315I</td>
<td>missense mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>axitinib</td>
<td>response</td>
<td>25686603</td>
<td>Level 2</td>
</tr>
<tr>
<td>ALK</td>
<td>F856S</td>
<td>missense mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>crizotinib</td>
<td>sensitivity</td>
<td>26302424</td>
<td>Level 3</td>
</tr>
<tr>
<td>ALK</td>
<td>A348D</td>
<td>missense mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>crizotinib</td>
<td>sensitivity</td>
<td>26302424</td>
<td>Level 3</td>
</tr>
<tr>
<td>BCR-JAK2</td>
<td>rearrangement</td>
<td>fusion gene</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>ruvortinib</td>
<td>sensitivity</td>
<td>22987947</td>
<td>Level 3</td>
</tr>
<tr>
<td>CRLF2</td>
<td>rearrangement</td>
<td>fusion gene</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>BET inhibitors</td>
<td>sensitivity</td>
<td>22904298</td>
<td>Level 3</td>
</tr>
<tr>
<td>CRLF2</td>
<td>rearrangement</td>
<td>fusion gene</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>mTOR inhibitors</td>
<td>sensitivity</td>
<td>22955920</td>
<td>Level 3</td>
</tr>
<tr>
<td>FBXW7</td>
<td>any</td>
<td>mutation</td>
<td>loss-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>steroids in early setting</td>
<td>sensitivity</td>
<td>20861099</td>
<td>Level 2A</td>
</tr>
<tr>
<td>IL7R</td>
<td>237-255</td>
<td>indel or S185C</td>
<td>mutation</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>ruvortinib (in SH2 domain)</td>
<td>sensitivity</td>
<td>22904298</td>
<td>Level 3</td>
</tr>
<tr>
<td>IL7R</td>
<td>237-255</td>
<td>indel or S185C</td>
<td>mutation</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>mTOR inhibitors (in SH2 domain)</td>
<td>sensitivity</td>
<td>22955920</td>
<td>Level 3</td>
</tr>
<tr>
<td>JAK1</td>
<td>S646F</td>
<td>missense mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>ruvortinib</td>
<td>sensitivity</td>
<td>22955920</td>
<td>Level 3</td>
</tr>
<tr>
<td>JAK2</td>
<td>rearrangement</td>
<td>fusion gene</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>ruvortinib</td>
<td>sensitivity</td>
<td>875623, 228994</td>
<td>Level 3</td>
</tr>
<tr>
<td>KRAS</td>
<td>codon 12</td>
<td>missense mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>MEK inhibitors</td>
<td>sensitivity</td>
<td>18701506</td>
<td>Level 3</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>rearrangement</td>
<td>fusion gene</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>Gamma secretase inhibitor</td>
<td>sensitivity</td>
<td>6882243, 230335</td>
<td>Level 3</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>PEST domain</td>
<td>p.2245-2534 mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>Gamma secretase inhibitor</td>
<td>response</td>
<td>22006 (abs 68)</td>
<td>Level 2B</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>PEST domain</td>
<td>p.2245-2534 mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>Gamma secretase inhibitor</td>
<td>sensitivity</td>
<td>2778647, 225108</td>
<td>Level 3</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>PEST domain</td>
<td>p.2245-2534 mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>Gamma secretase inhibitors + mTOR</td>
<td>sensitivity</td>
<td>19246502</td>
<td>Level 3</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>PEST domain</td>
<td>p.2245-2534 mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>Gamma secretase inhibitors + CDK4A</td>
<td>sensitivity</td>
<td>19318552</td>
<td>Level 3</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>PEST domain</td>
<td>p.2245-2534 mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>steroids early setting</td>
<td>response</td>
<td>20861920, 24</td>
<td>Level 2A</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>HD</td>
<td>p.1529-1723 mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>Gamma secretase inhibitor</td>
<td>response</td>
<td>20861920, 24</td>
<td>Level 2B</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>HD</td>
<td>p.1529-1723 mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>Gamma secretase inhibitor</td>
<td>sensitivity</td>
<td>778842, 225108</td>
<td>Level 3</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>HD</td>
<td>p.1529-1723 mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>Gamma secretase inhibitors + mTOR</td>
<td>sensitivity</td>
<td>19246502</td>
<td>Level 3</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>HD</td>
<td>p.1529-1723 mutation</td>
<td>gain-of-function</td>
<td>Acute Lymphoblastic Leukemia</td>
<td>C3167</td>
<td>Gamma secretase inhibitors + CDK4A</td>
<td>sensitivity</td>
<td>19318552</td>
<td>Level 3</td>
</tr>
</tbody>
</table>
Leveraging Literature Articles to Support Clinical Care

<table>
<thead>
<tr>
<th>Gene</th>
<th>Alteration</th>
<th>Alteration Type</th>
<th>Functional Significance</th>
<th>Annotation</th>
<th>Actionable Gene</th>
<th>Actionable Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDGFRA</td>
<td>Y288C</td>
<td>SNV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTEN</td>
<td>K163*</td>
<td>Truncation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGFR2</td>
<td>K292M</td>
<td>SNV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Leveraging Literature Articles to Support Clinical Care

<table>
<thead>
<tr>
<th>Gene</th>
<th>Alteration</th>
<th>Alteration Type</th>
<th>Functional Significance</th>
<th>Annotation</th>
<th>Actionable Gene</th>
<th>Actionable Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDGFRA</td>
<td>Y288C</td>
<td>SNV</td>
<td>Activating</td>
<td><strong>Activating Mutation</strong>: Exogenous expression of the Y288C mutant in BaF3 cells increased the number of viable cells measured an average of 5.76 fold more compared with cells expressing wildtype PDGFRA.</td>
<td>Yes</td>
<td>Yes-Functional Genomics</td>
</tr>
<tr>
<td>PTEN</td>
<td>K163*</td>
<td>Truncation</td>
<td>Unknown</td>
<td><strong>Likely loss of function</strong>: This mutation likely leads to loss of function due to premature stop codon in the phosphatase domain. Missense mutations in K163 resulted in an ≈80-fold reduction in its membrane affinity (Das et al., 2003)</td>
<td>Yes</td>
<td>Yes – Inferred</td>
</tr>
<tr>
<td>FGFR2</td>
<td>K292M</td>
<td>SNV</td>
<td>Unknown</td>
<td><strong>Potentially activating</strong>: The functional significance of this mutation is unknown. However, it is located within the Ig-like domain type 3, and other mutations in this domain can create an autocrine, feed-forward FGF signaling loop resulting in FGFR2 activation.</td>
<td>Yes</td>
<td>Potentially</td>
</tr>
</tbody>
</table>
Summary
Empowering Healthcare Professionals with Cognitive Computing

The Process of Building
- Work Closely with domain experts
- Leverage tools (like NLP) to extract unstructured concepts and their associations
- Combine unstructured and structured data to create knowledge
- Validate and iterate with domain experts

Value of Applying to Healthcare
- Increased efficiency
- Reduced bias
- Scaling expert knowledge
- Innovating across a complex domain

With Cognitive computing, the extraction of information from clinical data and published literature can empower healthcare professionals to stay up-to-date, focus on the complex problems, and spend more time with caring for patients.
Questions?
(remember, I’m not Watson..)