Developing clinical decision support systems for cancer and COVID-19 precision medicine

Varun Suraj
Mentor: Dr. Gil Alterovitz
Motivations
Growth in Precision Medicine

“An emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person.”

- Precision Medicine Initiative, NIH
Availability of Electronic Patient Data

“EHRs are real-time, patient-centered records that make information available instantly and securely to authorized users.”

- HealthIT.gov

Sequencing DNA means determining the order of the four chemical building blocks - called "bases" - that make up the DNA molecule.

- National Human Genome Research Institute
<table>
<thead>
<tr>
<th>Gene</th>
<th>Variant</th>
<th>Clinical Trials</th>
<th>Associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRAF</td>
<td>D594G</td>
<td>673</td>
<td>chr7:g.140453154T&gt;C</td>
</tr>
</tbody>
</table>

**BRAF**  
B-Raf proto-oncogene, serine/threonine kinase

**Gene Description**

This gene encodes a protein belonging to the RAF family of serine/threonine protein kinases. This protein plays a role in regulating the MAP kinase/ERK signaling pathway, which affects cell division, differentiation, and secretion. Mutations in this gene, most commonly the V600E mutation, are the most frequently identified cancer-causing mutations in melanoma, and have been identified in various other cancers as well, including non-Hodgkin lymphoma, colorectal cancer, thyroid carcinoma, non-small cell lung carcinoma, hairy cell leukemia and adenocarcinoma of lung. Mutations in this gene are also associated with cardiofaciocutaneous, Noonan, and Costello syndromes, which exhibit overlapping phenotypes. A pseudogene of this gene has been identified on the X chromosome. [provided by RefSeq, Aug 2017].
Phase 1: Multiple EHRs

- Application
  - SMART Cancer Navigator
- EHR 1
- EHR 2
- Application
  - SMART COVID Navigator
OAuth2 Authorization Process

Abstract Protocol Flow

1. Authorization Request
2. Authorization Grant
3. Authorization Grant
4. Access Token
5. Access Token
6. Protected Resource

User (Resource Owner)

Authorization Server

Resource Server

Service API
Name: Gilberto712 Iglesias873 | Zip Code: 38004 | Age: 94 | Condition:

VA: 53741008 Coronary Heart Disease

✓ VA: J96.11 Chronic respiratory failure with hypoxia
VA: 53741008 Coronary Heart Disease
VA: G30.9 Alzheimer's disease, unspecified
VA: 22298006 Myocardial Infarction
VA: Z86.74 Personal history of sudden cardiac arrest

CMS: 70219 OTHER SBORHEIC KERATOSIS
CMS: 7099 SKIN DISORDER NOS
CMS: 7062 SEBACEOUS CYST
CMS: 7038 DISEASES OF NAIL NEC
Phase 2: Adding Genomic Data

VCF-to-FHIR Converter

Application

SMART Cancer Navigator

Variant APIs
Sample VCF File

```plaintext
##fileformat=VCFv4.3
##fileDate=20090805
##source=imputationProgramV3.1
##reference=file://seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126c9f8a6e0c7f379d618ff66be2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=HQ,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=qs50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">

CHROM POS ID REF ALT QUAL FILTER INFO          FORMAT       NA00001     NA00002     NA00003
20  14370  rs6054257  G  A  29  PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ  0|0:48:8:51,51 1|0:48:8:51,51 1/1:43:5:...
20  17330  .  T  A  3   q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ  0|0:49:3:58,50 0|1:3:5:65,3  0/0:41:3
20  1110696 rs6040355 A  G,T  67  PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ  1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
20  1230237  .  T  .  47  PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ  0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20  1234567  microsat1  G,GTCT  50  PASS NS=3;DP=9;AA=G GT:GQ:DP  0/1:35:4  0/2:17:2  1/1:40:3
```
Querying Gene-Variant Information
Phase 3: Extending Application to COVID-19

Application
SMART COVID Navigator
CORD-19 Dataset
<table>
<thead>
<tr>
<th>Risk Factors</th>
<th>CSV Headers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Date</td>
</tr>
<tr>
<td>Asthma</td>
<td>Severe Adjusted</td>
</tr>
<tr>
<td>Autoimmune disorders</td>
<td>Severe Calculated</td>
</tr>
<tr>
<td>COPD</td>
<td>Study</td>
</tr>
<tr>
<td>Cancer</td>
<td>Severe Fatality</td>
</tr>
<tr>
<td>Cardio- and cerebrovascular disease</td>
<td>Severe Fatality lower bound</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>Severe Fatality upper bound</td>
</tr>
<tr>
<td>Chronic digestive disorders</td>
<td>Severe Fatality p-value</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>Severe Fatality Significant</td>
</tr>
<tr>
<td>Chronic liver disease</td>
<td>Severe Fatality Adjusted</td>
</tr>
<tr>
<td>Chronic respiratory diseases</td>
<td>Severe Fatality Calculated</td>
</tr>
<tr>
<td>Dementia</td>
<td>Study Link</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Journal</td>
</tr>
<tr>
<td>Drinking</td>
<td>Severe</td>
</tr>
<tr>
<td></td>
<td>Severe lower bound</td>
</tr>
<tr>
<td></td>
<td>Severe upper bound</td>
</tr>
<tr>
<td></td>
<td>Severe p-value</td>
</tr>
<tr>
<td></td>
<td>Severe Significant</td>
</tr>
<tr>
<td></td>
<td>Fatality</td>
</tr>
<tr>
<td></td>
<td>Fatality lower bound</td>
</tr>
<tr>
<td></td>
<td>Fatality upper bound</td>
</tr>
<tr>
<td></td>
<td>Fatality p-value</td>
</tr>
<tr>
<td></td>
<td>Fatality Significant</td>
</tr>
<tr>
<td></td>
<td>Fatality Adjusted</td>
</tr>
<tr>
<td></td>
<td>Fatality Calculated</td>
</tr>
<tr>
<td></td>
<td>Multivariate Adjustment</td>
</tr>
<tr>
<td></td>
<td>Study Type</td>
</tr>
<tr>
<td></td>
<td>Sample Size</td>
</tr>
<tr>
<td></td>
<td>Study Population</td>
</tr>
<tr>
<td></td>
<td>Added on</td>
</tr>
<tr>
<td></td>
<td>Critical only</td>
</tr>
<tr>
<td></td>
<td>Discharged vs. death?</td>
</tr>
</tbody>
</table>
VA: J96.11 Chronic respiratory failure with hypoxia
✓ VA: 53741008 Coronary Heart Disease
VA: G30.9 Alzheimer's disease, unspecified
VA: 22298006 Myocardial Infarction
VA: Z86.74 Personal history of sudden cardiac arrest
CMS: 70219 OTHER SBORHEIC KERATOSIS
CMS: 7099 SKIN DISORDER NOS
CMS: 7062 SEBACEOUS CYST
CMS: 7038 DISEASES OF NAIL NEC
### Age

- **Number of papers:** 33
- **Proportion of significant findings for severity:** 67%
- **Proportion of significant findings for fatality:** 84%

<table>
<thead>
<tr>
<th>Study Name</th>
<th>Date</th>
<th>Severity Significance</th>
<th>Fatality Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent of prior lung irradiation and mortality in COVID-19 patients with a cancer history</td>
<td>2020-05-20</td>
<td>Not Significant</td>
<td></td>
</tr>
<tr>
<td>Clinical Characteristics and Outcomes of Patients With Diabetes and COVID-19 in Association With Glucose-Lowering Medication</td>
<td>2020-05-14</td>
<td>Significant</td>
<td>Significant</td>
</tr>
<tr>
<td>Clinical Characteristics and Outcomes of Patients With Diabetes and COVID-19 in Association With Glucose-Lowering Medication</td>
<td>2020-05-14</td>
<td>Significant</td>
<td>Significant</td>
</tr>
<tr>
<td>Correlation of coagulation parameters with clinical outcomes in Coronavirus-19 affected minorities in United States: Observational cohort</td>
<td>2020-05-06</td>
<td>Significant</td>
<td></td>
</tr>
</tbody>
</table>
Future Work

- Add Reboot Rx COVID-Cancer Dataset
- Add AI/machine learning to cancer prediction using genomic data
- Add AI/machine learning to improve COVID-19 severity and mortality predictions
Acknowlegements

- MIT PRIMES
- Dr. Gil Alterovitz, Ning Xie, and Ling Teng
- Past PRIMES students who helped me understand prior work
- My family