Precision Cancer Medicine: A SMART Web App

Makiah Bennett & Jack Flahive
"I gave it a healthy dose of denial, but it didn't help."
Motivations

Why bother?
First Motivation: Vast Data

<table>
<thead>
<tr>
<th>Reference allele</th>
<th>POPFREQ for rs12121571</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1050456</td>
<td>0.50041054545469</td>
</tr>
<tr>
<td>rs1050450</td>
<td>0.64646464646464</td>
</tr>
<tr>
<td>rs1050451</td>
<td>0.43434343434343</td>
</tr>
<tr>
<td>rs1050452</td>
<td>0.59595959595959</td>
</tr>
</tbody>
</table>

**Start of Data Table for this file**

**Format Statement for Population Frequency**

- Reference allele
- POPFREQ for rs12121571

**Start of VCF Header**

- **First Motivation:** Vast Data
- **Start of VCF Header:**
  - **CHROM:** 11794553
  - **POS:** 12121577
  - **QUAL:** FILTER INFO

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 50.041054545469
- **FILTER:** INFO
- **INFO:** AF=0.50041054545469;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 64.646464646464
- **FILTER:** INFO
- **INFO:** AF=0.64646464646464;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 43.434343434343
- **FILTER:** INFO
- **INFO:** AF=0.43434343434343;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 59.595959595959
- **FILTER:** INFO
- **INFO:** AF=0.59595959595959;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 95.959595959596
- **FILTER:** INFO
- **INFO:** AF=0.95959595959596;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 55.555555555556
- **FILTER:** INFO
- **INFO:** AF=0.55555555555556;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 69.696969696970
- **FILTER:** INFO
- **INFO:** AF=0.69696969696970;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 79.797979797980
- **FILTER:** INFO
- **INFO:** AF=0.79797979797980;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 90.909090909091
- **FILTER:** INFO
- **INFO:** AF=0.90909090909091;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 72.727272727273
- **FILTER:** INFO
- **INFO:** AF=0.72727272727273;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 82.828282828283
- **FILTER:** INFO
- **INFO:** AF=0.82828282828283;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 89.898989898990
- **FILTER:** INFO
- **INFO:** AF=0.89898989898990;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 92.929292929293
- **FILTER:** INFO
- **INFO:** AF=0.92929292929293;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1

**Alternate alleles**

- POPFREQ for rs12121571

- **CHROM:** 11794553
- **POS:** 12121577
- **QUAL:** FILTER INFO
- **REF:** A
- **ALT:** T
- **QUAL:** 98.989898989899
- **FILTER:** INFO
- **INFO:** AF=0.98989898989899;A1=1;A2=1
- **FORMAT:** GT
- **GT:** 0/1
Second Motivation: EHRs Overly Complex

“This is where the idea for the new EHR starts getting a little complicated.”
Third Motivation: AI
Background

Prerequisites to understanding the app.
An emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person.

-Precision Medicine
-NIH (National Institutes of Health)
“Doctors have always recognized that every patient is unique, and doctors have always tried to tailor their treatments as best they can to individuals. You can match a blood transfusion to a blood type — that was an important discovery. What if matching a cancer cure to our genetic code was just as easy, just as standard? What if figuring out the right dose of medicine was as simple as taking our temperature?”

- President Obama, 2015
FHIR aims to simplify implementation without sacrificing information integrity. It leverages existing logical and theoretical models to provide a consistent, easy to implement, and rigorous mechanism for exchanging data between healthcare applications.

- hl7.org
[..] a set of **open specifications** to integrate apps with Electronic Health Records, portals, Health Information Exchanges, and other Health IT systems.
[...] an open-source implementation of the FHIR specification in Java.

- hapifhir.io
Development

In progress!
The S4G2 Server Backend

<Ha\l/> HAPI-FHIR
fhir made simple.

RESTful API
GET PUT POST DELETE

SMART

SMART on FHIR

OAuth

2

express-session deprecated undefined reserve option; provide reserve option app.js:3819
express-session deprecated undefined saveUninitialized option; provide saveUninitialized option app.js:3819

Docker

Starting S4G2 server...

Successfully loaded a scope!!

Attempting to load a scope...

Successfully loaded a scope!!

Starting S4G2 server...

Successfully loaded a scope!!

Attempting to load a scope...

Successfully loaded a scope!!
The Front-End

```javascript
var server = app.listen(process.env.PORT || 8080, "0.0.0.0", function () {
    console.log('Server listening at http://' + server.address().address + ':' + server.address().port);
});
```
Currently Viewing:

<table>
<thead>
<tr>
<th>Gene</th>
<th>Driver</th>
<th>PCM Score</th>
<th>Mutation</th>
<th>Drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUL5</td>
<td>Yes</td>
<td>0.73</td>
<td>c.T1630G point coding</td>
<td>AT-406</td>
</tr>
</tbody>
</table>

Go To:
Full Data | Top 50 Genes | Drugs
John Smith
Gender: M
DOB: 24 June 1990
Height: 182 cm

Mutations

<table>
<thead>
<tr>
<th>Name</th>
<th>Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>AURKA</td>
<td>E4499K</td>
</tr>
<tr>
<td>CCND1</td>
<td>V2037F</td>
</tr>
<tr>
<td>CHEK2</td>
<td>R156P</td>
</tr>
<tr>
<td>EWSR1</td>
<td>loss</td>
</tr>
<tr>
<td>FOXP1</td>
<td>R1031Q</td>
</tr>
<tr>
<td>GID4</td>
<td>F171fs*3</td>
</tr>
<tr>
<td>TP53</td>
<td>Amplification</td>
</tr>
<tr>
<td>JAK2</td>
<td>D423E</td>
</tr>
<tr>
<td>CDKN1B</td>
<td>Amplification</td>
</tr>
</tbody>
</table>

Mutated Genes Observed in Breast
Convolution

3 Main Layers

Layer 1: Prioritizes cancer “driver” mutations.

Layer 2: Links mutations to genes using logistic regression.

Layer 3: Generates a prioritized list of drugs targeting driver genes.

Similar to the iCAGES pipeline.
Appendix

- **Electronic Health Record (EHR)**
  - Electronic version of a patient’s medical history
- **Substitutable Medical Applications and Reusable Technology (SMART) Health IT Platform**
  - Open access API that classifies and packages genomic information for use in clinical applications
  - Allows for the integration of different data types in one application
- **Fast Healthcare Interoperability Resources (FHIR)**
  - Standard for exchanging healthcare information electronically
- **Single Nucleotide Polymorphisms (SNPs)**
  - The variation of a single base pair in a DNA sequence
- **VCF Files**
  - “Variant Call Format”: the raw file data format which our database provides for parsing. Used in the Texas patient database to which we refer for data.
Acknowledgements

- Thank you to Dr. Alterovitz for the much-needed help and support and the project idea!
- Thank you to Maksym Korablyov for answering and continuing to answer our many questions on machine learning!
- Thank you to MIT PRIMES for the enjoyable and interesting research opportunity!
THANKS!

Any questions?
You can find us at

- makiahtbennett@gmail.com
- jack.flahive@gmail.com