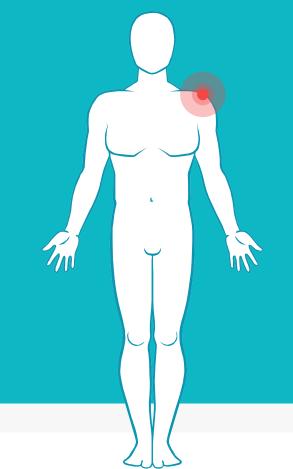
Precision Cancer Medicine: A SMART Web App



Makiah Bennett & Jack Flahive



"I GAVE IT A HEALTHY DOSE OF DENIAL, BUT IT DIDN'T HELP."

Motivations

Why bother?



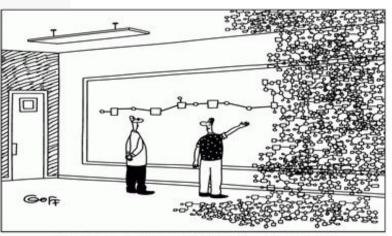
First Motivation: Vast Data

```
Start of VCF Header
##dbSNP BUILD ID=135
##reference=GRCh37.p5
##variationPropertyDocumentationUrl=ftp://ftp.ncbi.nlm.nih.gov/snp/specs/dbSNP BitField latest.pdf
##dbSNP TYPE=polymorphic in at least one population
##dbSNP POP ID=4446
                            Population ID for the data in this file
##dbSNP LOC POP ID=AAN GENO PANE
##dbSNP POP HANDLE=PGA-UW-FHCRC
##INFO=<ID=RSPOS, Number=1, Type=Integer, Description="Chr position reported in dbSNP">
##INFO=<ID=RV,Number=0,Type=Flag,Description="RS orientation is reversed">
##INFO=<ID=NS, Number=1, Type=Integer, Description="Number of Samples With Data">
##INFO=<ID=AF, Number=., Type=Float, Description="Allele Frequency, for each ALT allele, in the same order as listed">
##INFC=<ID=VP, Number=1, Type=String, Description="Variation Property">
##INFO=<ID=GENEINFO, Number=1, Type=String, Description="Pairs each of gene symbol:gene id. The gene symbol and id are delimited by a
colon (:) and each pair is delimited by a vertical bar (|) ">
##INFO=<ID=dbSNPBuildID.Number=1.Type=Integer.Description="First dbSNP Build for RS">
##INFO=<ID=SAO.Number=1.Type=Integer.Description="Variant Allele Origin: 0 - unspecified, 1 - Germline, 2 - Somatic, 3 - Both">
##INFO=<ID=SSR, Number=1, Type=Integer, Description="Variant Suspect Reason Code, 0 - unspecified, 1 - Paralog, 2 - byEST, 3 - Para EST
. 4 - oldAlign, 5 - other">
                                                                                    Start of Data table for this file
                          Format Statement for Population Frequency
##INFO=<ID=NOV, Number / Type=Flag, Description="Rs cluster has non-overlappin silele sets. True when rs set has more than 2 alleles
from different subsidions and these sets share no alleles in common.">
##INFO=<ID=GCF,Num_==0,Type=Flag,Description="Has Genotype Conflict Saw (rs, ind), different genotype. N/N is not included.">
##INFO=<ID=POFFREQ,Number=1,Type=String,Description="Frequencies and out of the ALT alleles by population ID. The form is na:ns:f
(c1/c2)[|f(c1/c2)] na is the number of alleles; ns is the number samples; f is the frequency; c1 is the allele count; and c2 is
sample count for that allele (c1 - homozygous count).">
                                                                                                                REQ for rs12121511
##FILTER=<ID=N Reference allele sistent Genotype Submission
3548 (22/22)
0; SAO=0; VP=05010000000050512000104; GENEINFO=AGTRAP: 57085; WGT=0; VC=SNV; SLO; VLD; G5; HD; GNO; KGP110t123; PH2; NOV; POPFREO=248: 124: 0.048387
0967741935(12/12)[0(0/)[0(0/)
83870968 (184/118)
       11799228
SAC=0; VP=05010000000050512000101; GENEINFO=AGTRAP: 57085; WGT=0; VC=SNV; SLO; VLD; G5; HD; GNO; KGPilot123; PH2: GCF: POPFREO=248: 124: 0.3387096
77419355 (84/66)
                                                                           RSPOS=11800048;GMAF=0.25;dbSNPBuildID=120;SSR=0;SAO=0;VP=050
100000000050510000104;GENEINFO=AGTRAP:57085;WGT=0;VC=
                                                          SLO:VLD:G5:HD:GNO:KGP11ot123:NOV:POPFREO=248:124:0(0/)|0.491935483870968(12
2/92) [0(0/)
                                                                                                       POPFREQ for rs11121815
                                                                Alternate alleles
```

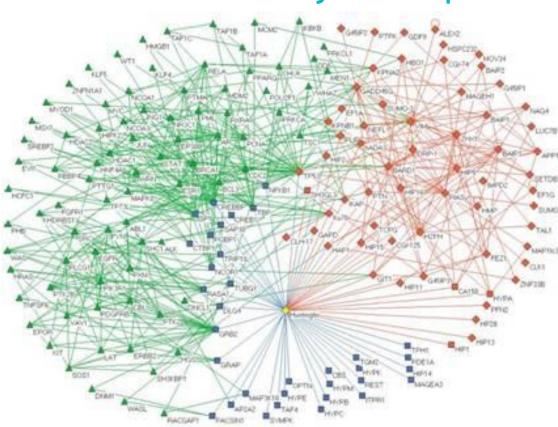




Second Motivation: EHRs Overly Complex



"This is where the idea for the new EHR starts getting a little complicated."





Third Motivation: Al



Background

Prerequisites to understanding the app.

Precision Medicine

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





The Precision Cancer Medicine Initiative

"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?"







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.





[..] a set of **open specifications** to integrate apps with Electronic Health Records, portals, Health Information Exchanges, and other Health IT systems.

"

- docs.smarthealthit.org



[...] an open-source implementation of the FHIR specification in Java.

"

hapifhir.io

Development

In progress!



The S4G2 Server Backend







```
* Debugger pin code: 818-054-806
Sat May 20 22:51:56.543 [initandlisten] preallocateIsFaster=true 12.3
Sat May 20 22:51:56.543 [initandlisten] preallocateIsFaster check took 6.923 secs
Sat May 20 22:51:56.544 [initandlisten] preallocating a journal file /data/db/journal/prealloc.0
                                                        File Preallocator Progress: 524288000/1073741824
Sat May 20 22:51:59.923 [initandlisten]
Sat May 20 22:52:02.009 [initandlisten]
                                                        File Preallocator Progress: 838860800/1073741824
                                                        File Preallocator Progress: 943718400/1073741824
Sat May 20 22:52:06.661 [initandlisten]
Sat May 20 22:52:10.295 [initandlisten]
                                                        File Preallocator Progress: 985661440/1073741824
Sat May 20 22:52:14.519 [initandlisten] preallocating a journal file /data/db/journal/prealloc.1
                                                        File Preallocator Progress: 503316480/1073741824
Sat May 20 22:52:17.787 [initandlisten]
Opening data file and reading scopes
Sat May 20 22:52:20.894 [initandlisten]
                                                        File Preallocator Progress: 765460480/1073741824
                                                        File Preallocator Progress: 922746880/1073741824
Sat May 20 22:52:25.900 [initandlisten]
Attempting to load a scope...
Sat May 20 22:52:35.101 [initandlisten] preallocating a journal file /data/db/journal/prealloc.2
                                                        File Preallocator Progress: 524288000/1073741824
Sat May 20 22:52:38.060 [initandlisten]
Sat May 20 22:52:41.571 [initandlisten]
                                                        File Preallocator Progress: 796917760/1073741824
                                                                                                                74%
                                                        File Preallocator Progress: 975175680/1073741824
Sat May 20 22:52:45.969 [initandlisten]
Sat May 20 22:52:53.133 [FileAllocator] allocating new datafile /data/db/local.ns, filling with zeroes...
Sat May 20 22:52:53.133 [FileAllocator] creating directory /data/db/_tmp
Sat May 20 22:52:53.163 [FileAllocator] done allocating datafile /data/db/local.ns, size: 16MB, took 0.014 secs
Sat May 20 22:52:53.181 [FileAllocator] allocating new datafile /data/db/local.0, filling with zeroes...
Sat May 20 22:52:53.208 [FileAllocator] done allocating datafile /data/db/local.0, size: 64MB, took 0.026 secs
Sat May 20 22:52:53.225 [websyr] admin web console waiting for connections on port 28017
Sat May 20 22:52:53.261 [initandlisten] waiting for connections on port 27017
Sat May 20 22:52:53.600 [FileAllocator] allocating new datafile /data/db/s4gserver.ns, filling with zeroes...
Sat May 20 22:52:53.676 [FileAllocator] done allocating datafile /data/db/s4gserver.ns, size: 16MB. took 0.075 secs
Sat May 20 22:52:53.696 [FileAllocator] allocating new datafile /data/db/s4gserver.0. filling with zeroes...
Sat May 20 22:52:53.715 [FileAllocator] done allocating datafile /data/db/s4gserver.0, size: 64MB, took 0.018 secs
Sat May 20 22:52:53.716 [FileAllocator] allocating new datafile /data/db/s4gserver.1, filling with zeroes...
Sat May 20 22:52:53.716 [conn2] build index staserver.scopes { id: 1 }
Sat May 20 22:52:53.722 [conn2] build index done. scanned 0 total records. 0.005 secs
Sat May 20 22:52:53.723 [conn2] insert s4gserver.scopes ninserted:1 keyUpdates:0 locks(micros) w:122589 122ms
Successfully loaded a scope!!
Attempting to load a scope...
```

Sat May 20 22:52:53.737 [FileAllocator] done allocating datafile /data/db/s4gserver.1. size: 128MB. took 0.02 secs

Starting S4G2 server...

npm info it worked if it ends with ok npm info using npm@4.2.0

info using node@v7.10.0

Successfully loaded a scope!!

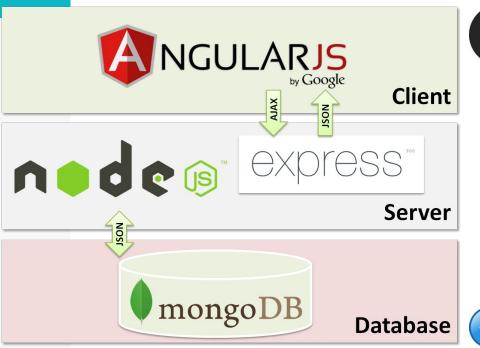
npm info lifecycle s4gserver@1.0.0~prestart: s4gserver@1.0.0

> s4gserver@1.0.0 start /root/s4gserver > node ./bin/www

express-session deprecated undefined resave option; provide resave option app.js:35:9
express-session deprecated undefined saveUninitialized option; provide saveUninitialized option app.js:35:9



The Front-End





Fitzgerald, Dean

Mutations

Name	Variant	
AURKA	E4499K	
CCND1	V2037F	
CHEK2	R156P	
EWSR1	loss	
FOXP1	R1031Q	
GID4	F171fs*3	
TP53	Amplification	
JAK2	D423E	
CDKN1B	Amplification	



Trump, Donald



DeVos, Betsy



Those here are grouped in order of their similarity to their genome's mutations

Pence, Michael



Schmoe, Joe



Biden, Joe



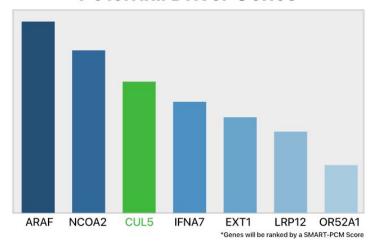
Obama, Barack

```
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);
});
```

PCM Web App Home About Profile Log Out



Potential Driver Genes



Currently Viewing:

Gene Driver PCM	DCM Score	Mutation			Drug	
	PCIVI SCORE	Mut. Syntax	Category	Prot. Syntax	Drug	
CUL5	Yes	0.73	c.T1630G	point coding	p.F544V	AT-406



Go To:

Full Data | Top 50 Genes | Drugs

PCM Web App Home About Profile Log Out



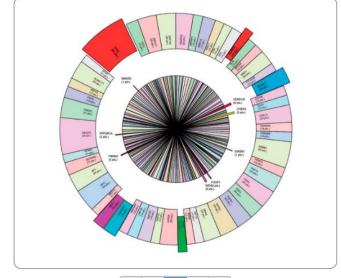
Gender: M DOB: 24 June 1990 Height: 182 cm

	Height. 102 Cili
Mutati	ons
Compa	risons
Wiki	
Setting	s

Mutations

Name	Variant
AURKA	E4499K
CCND1	V2037F
CHEK2	R156P
EWSR1	loss
FOXP1	R1031Q
GID4	F171fs*3
TP53	Amplification
JAK2	D423E
CDKN1B	Amplification

Mutated Genes Observed in Breast



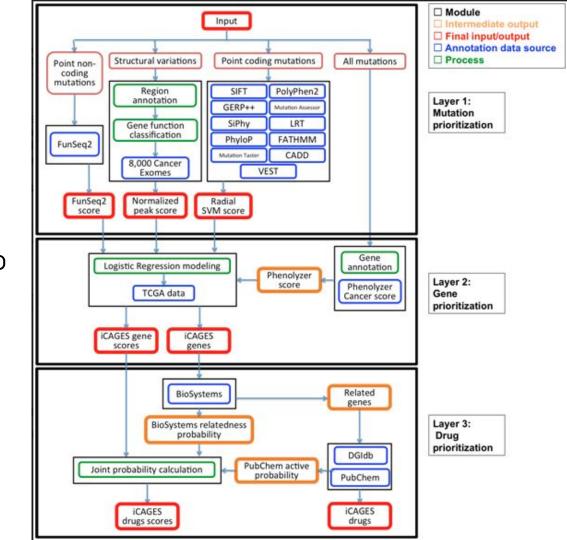
« 1 2 3 »





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