

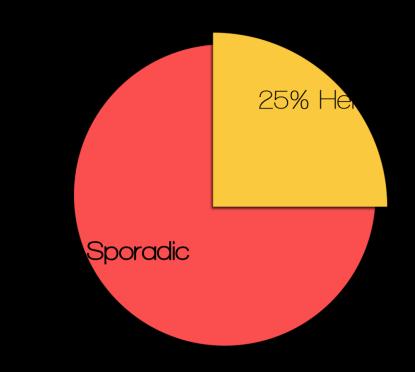
Integrating genomic, clinical, and patient questionnaire information for breast cancer diagnosis and treatment

> John Zhang Mentor Dr. Gil Alterovitz PRIMES Conference 2014 May 18, 2014



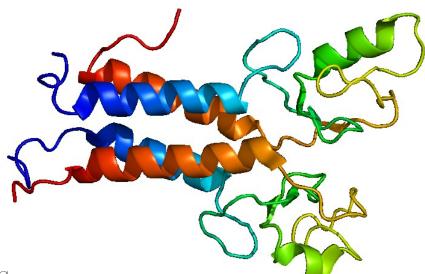
Background

- 1 in 10 women will have breast cancer
- 10% related to auto dominant gene
- Factors such as su toxic chemicals, and aging process



High Penetrance Genes

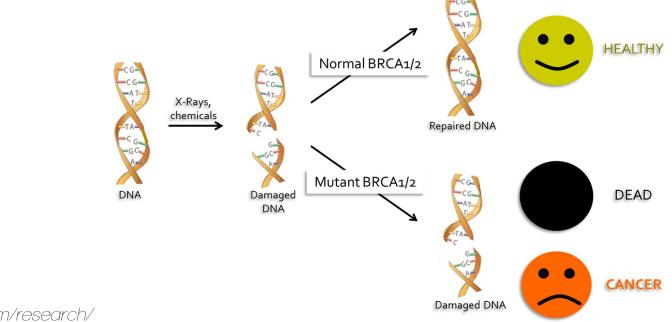
- BRCA1 and BRCA2
- Tumor suppressor gene
- Proof reading mechanism for DNIA
- Some do not display signs of cancer but
 80% will



http://en.wikipedia.org/wiki/File:Protein_BRCA1_PDB_1jm7.png

Mutation of BRCA1

- Elevates risk from 10% to 80%
- Cancer cells can proliferate leading to tumors



http://www.kirandhillon.com/research/



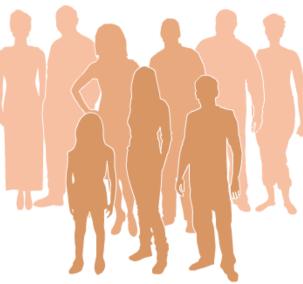
Prevention

- Mutations can be tested for
- Mutation carriers tend to develop cancer at an earlier age
- Affects ability to tolerate degradation of other mechanisms such as old age



Prevention

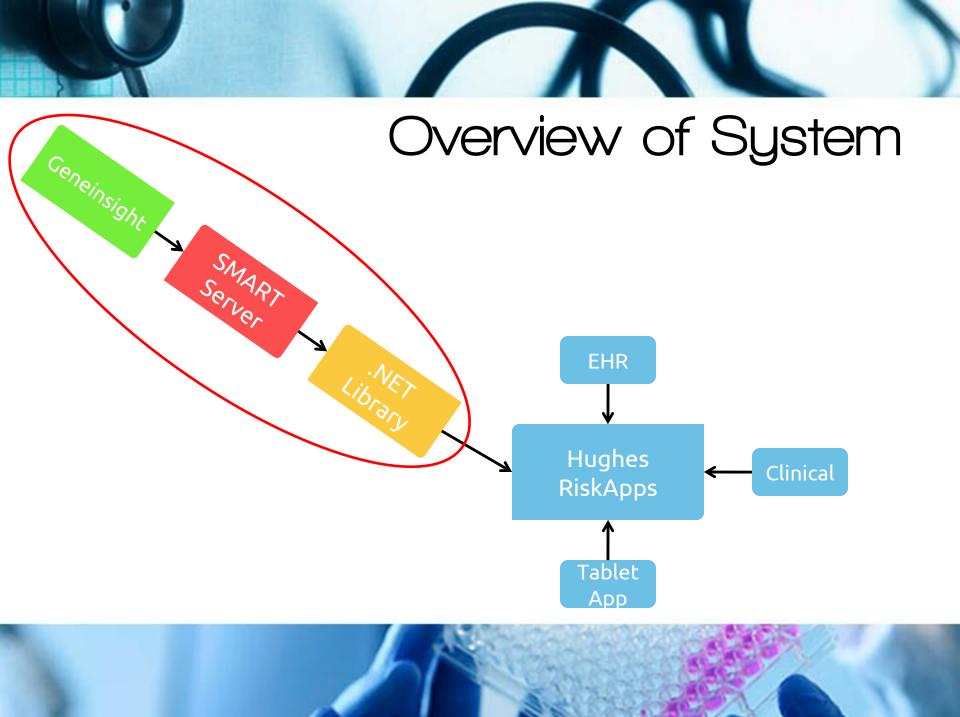
- Family history algorithms
- Once we find one case it facilitates the process
- Preemptive surgery to remove cancer





Motivation

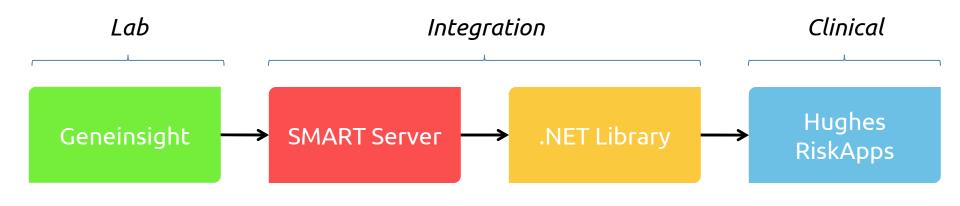
- Identification of mutation carrier
 - Family History
 - In Screening
- Most mutation carriers are not found
- Get cancer at early age with no access to risk reduction interventions
- Faster screening means help for people with mutations







 Integrate clinical system and labs that test for mutation carriers to identify and provide treatment for at risk individuals





Advantages

- Paperless and completely automated
- Patient data is easily accessible
- Less mistakes made





Geneinsight Variant Service

- SOAP Web Service
- Methods for accessing variant data
- Provides access to entire Partners Laboratory for Molecular Medicine knowledge base
- Accessed using modified suds library for Python



SMART Server & FHIR Format

- Substitutable Medical Apps & Reusable Technology
- Fast Health Interoperability Resource
- First SMART-enabled container in production
- Emerging standard for transferring clinical data
- Easily deployable

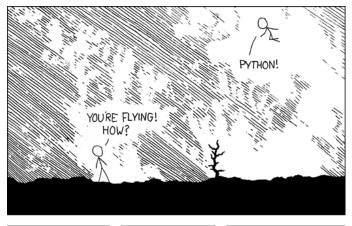




Conversion to EUID

Conversion to FHIR

- 1. All genes available on service are acquired
- 2. Converted to coordinates by Python script that searches Ensembl Genome Browser
- 3. Additional information is acquired from service and added to new format





Conversion to FHIR

Ensembligenome browser	nbl.org/Homo_sapiens/Gene/S	ummany?db=coreg=ENSG00000012048y=17:41196312-41277500	
Apps 📄 Read Later 🐨 Rando			
Ensembl 🗮 🗛	ST/BLAT BioMart Tools Do	vnloads Help & Documentation Elisg Mimors	Login 🔁 • Search all species
	17:41,196,312-41,277,500 Gene	BRCAT	
based displays immary lice variants (31)	Gene: BRCA1 ENSGO	000012048	
nscript comparison	Description	breast cancer 1, early enset [Source:HGNC Symbol;Acc:1100]	
porting evidence uence	Location	Chromosome 17: 41.195 212-41.277 500 reverse strand.	
econdary Structure	INSDC coordinates	chremosome: GRCh37: CM000679.1:41196312:41277500:1	
ernal references ulation ression	Transcripts	This gere has 31 transcripts (splice variants) Show transcript table	
parative Genomics mornic alignments ane tree (mage)	Summary Ø		
Gene tree (text)	Name	BRCA1 (HGINC Symbol)	
Gene tree (alignment) Gene gain/toss tree	Synonyms	BRCC1, PPP1R53, RNF53 [To view all Ensemblig ensem initiad to the name <u>discloses</u>]	
rthologues (59)	CCDS	This gene is a member of the Human CCDS set. CCDS11453. CCDS11454. CCDS11455. CCDS11456. CCDS11459.	
ralogues otein families (3)	RefSeq	Overlapping RefSeq Gene ID 672 matches and has similar biotype of protein_coding	
otype	LRG	LRG_292 provides a stable genomic reference framework for describing sequence variations for this gene	
tic Variation riation table	Ensembl version	ENS/0000012048-15	
ination table ination image	Gene type	Known protein coding	
ructural variation mail data	Prediction Method	Annotation for this gene includes both automatic annotation from Ensembl and Hayana manual curation, see article.	
rnaidata irsonal annotation	Alternative genes	This gene corresponds to the following database identifiers:	
istory ene history		Havana gene: 0TTHUM/G80000157426 (version 5)	
atgure this page	Go to Region i	Detail for more tracks and navigation options (e.g. zooming)	
id your data			
port data	🌣 🦓 < 🕀 🖬		
okmark this page		101.191b 11.394b 41.294b 41.294b 41.294b 41.294b 41.294b 41.294b 41.294b	Forward strand p 41.27Mb 41.28Mb
are this page	Genes (GENCODE		NBR2-001 > processed transcript NBR2-003 > processed transcript
			NBR2-004 >
		D 99,2394-001 >	processed transcript
		RPL2194.001 > processed as subopene	NBR2+005 > transcribed unprocessed pseudogen-
	Contigs	 Activities 	
	Genes (GENCODE	< (B)(Q) 400	
		< BRCA1-205 protein coding	
		preten coding	

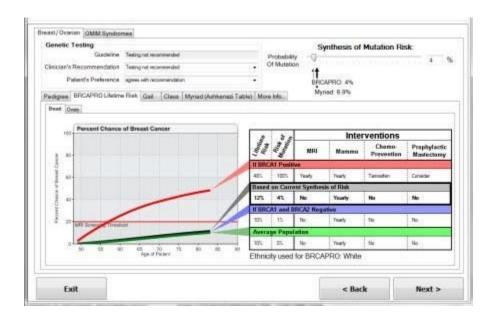
def get what i want():
result = open('gi_genes.txt', 'w')
genes = set()
for data in getChromosomeVariants():
new genes = get genes(data)
for g in (genes & new_genes):
result.write(g+'\n')
print g
genes = get_genes(data)
<pre>ief get_coordinate_for_gene(term):</pre>
term += ' ensembl'
<pre>url = r'https://www.google.com/search?rlz=1C1LENP_enUS539US539&es_sm=93&q='+</pre>
<pre>g_search = requests.get(url)</pre>
<pre>search_results = BS(g_search.text)</pre>
anchor = search_results.find('a', href=google_a_p)
ensembl_url = urlparse.parse_qs(anchor['href'])['/url?q'][0]
ensembl_page = requests.get(ensembl_url).text
try:
<pre>return coordinate_p.search(ensembl_page).group(1)</pre>
except:
pass
\$ 2637 Genes
around 3 sec conversion for each
ief convertGenes():
<pre>genes = open('gi_genes.txt', 'r')</pre>
<pre>coordinates = open('gi_coordinates.txt', 'w')</pre>
coordinates.write('coordinates={')
<pre>for gene in string.split(genes.read(), '\n'):</pre>
coordinate = get_coordinate_for_gene(gene)
if coordinate:
<pre>coordinates.write('"' + gene + '":"' + coordinate + '",')</pre>
else:
coordinates.write('"' + gene + '":"' + str(0) + '",')
fname == 'main':
<pre>#test here</pre>
convertGenes()

coordinates={"SFTPD":"10:81697496-81742370", "KCNE2":"21:35736323-35743688", "FKBP14":"7:30050203-30066300", "KCNE1":"21:35818988-35884573", "COL11A1":"1:103342023-103574052", "ESPN":"4:152120331-152152371", "SFTPC":"8:22014426-22021992", "RPGR":"X:38128416-38186817", "RSPH4A":"6:116937642-116954148", "PRKAG2":"7:151253197-151574210", "SPRED1":"15:38544527-38649450", "SLC39A13":"11:47428683-47438047", "HBB":"11:5246694-5250625", "JPH2":"20:42740335-42816218", "HPS4":"22:26839389-26879803", "HPS6":"10:103825147-103827792", "HPS1":"19:42755105-42779978", "TNNC1":"3:52485118-52488086", "HPS3":"3:148847371-148891519"

467 in total

Integration with Hughes riskApps

- Calls the SMART Server using .NET Library
- Authenticates user with OAuth 2.0 process



Cancer F	hughesriskApps		
Are yo	ou still having your pe	riods?	
	Yes		
	No		
	Not sure		
	Clear		
is means, are you still menstruating or are y	au still having bleeding every month or so	7	
		Back	Next



Results

- Variant information from Geneinsight can be processed in FHIR format
- Data can be stored on SMART server
- Library that calls SMART server and can be easily integrated with clinical applications

Future Work

- Set up meeting with Brian
- Improved deployability for library and server
- Integration of Geneinsight Report Service developed specifically for riskApps
- Expand to other models other than breast cancer



Acknowledgements

- Tom Chen
- Dr. Gil Alterovitz
- Brian Drohan
- Dr. Kevin Hughes
- Larry Babb
- PRIMES

