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

John Zhang

Mentor Dr. Gil Alterovitz
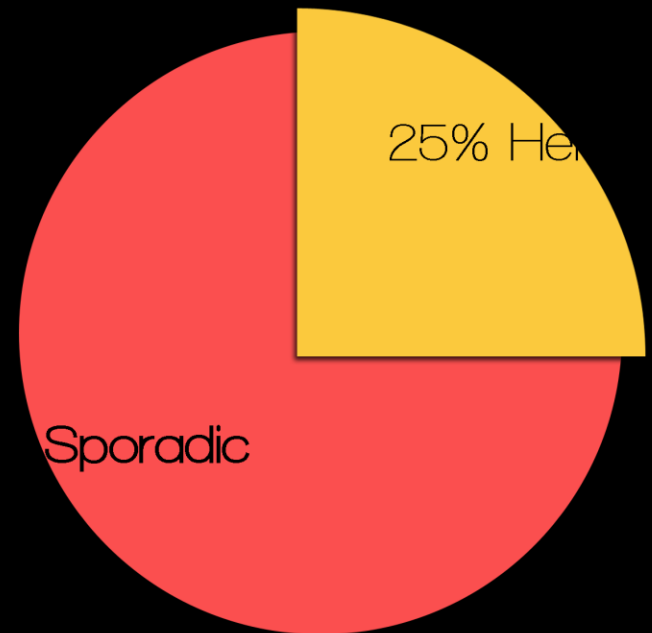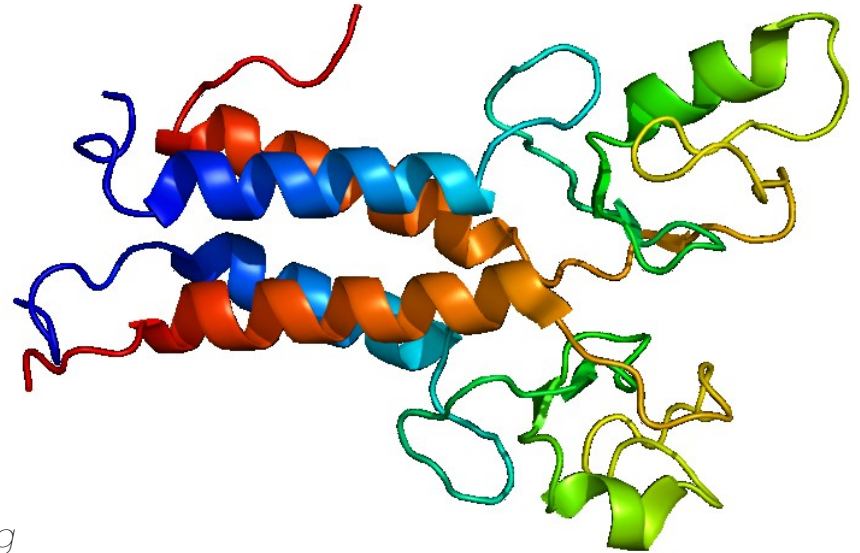
PRIMES Conference 2014

May 18, 2014

Background

- 1 in 10 women will have breast cancer

- 10% related to aut[o] dominant gene

- Factors such as su[...] toxic chemicals, and aging process[...]
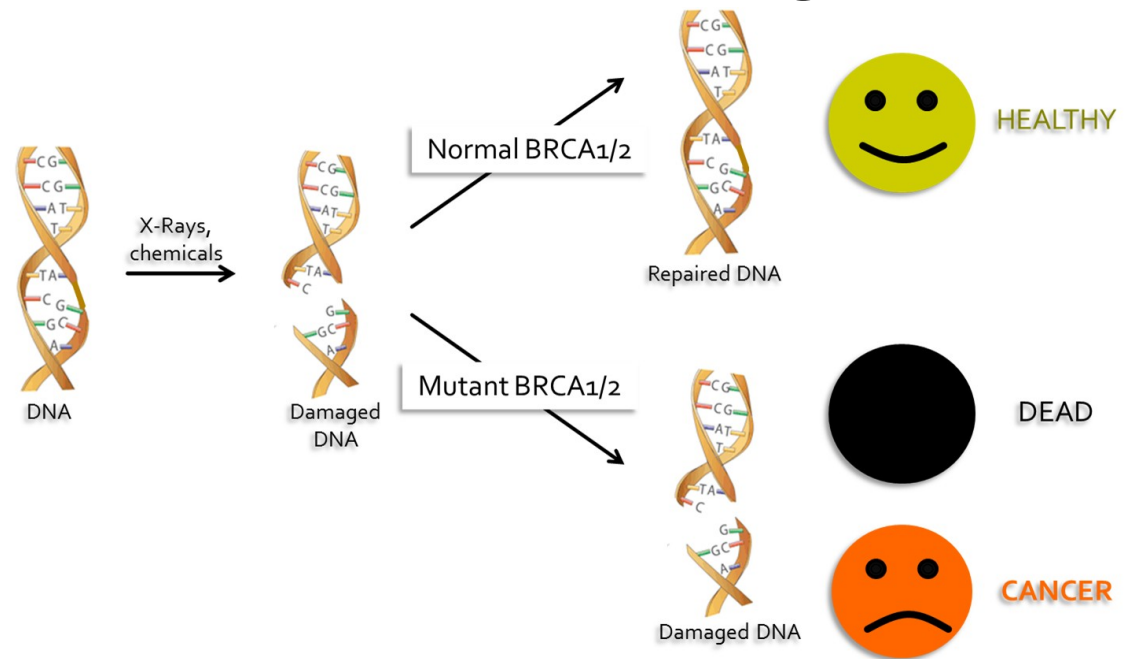
25% He[r]

Sporadic

High Penetrance Genes

- BRCA1 and BRCA2

- Tumor suppressor gene

- Proof reading mechanism for DNA

- Some do not display signs of cancer but 80% will

Mutation of BRCA1

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

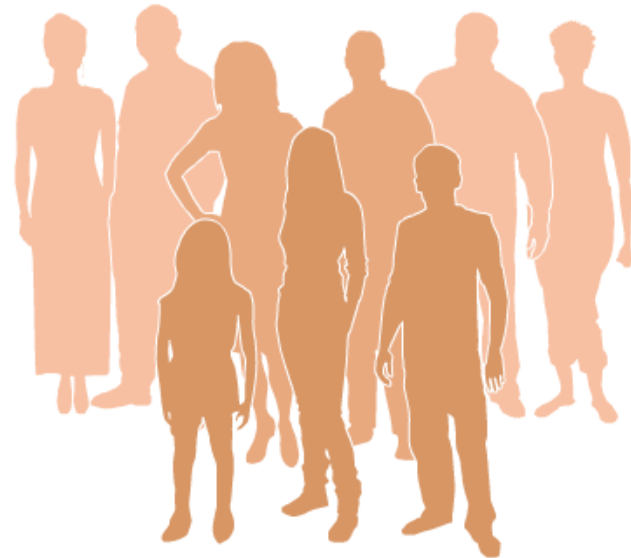*http://www.kirandhillon.com/research/*

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
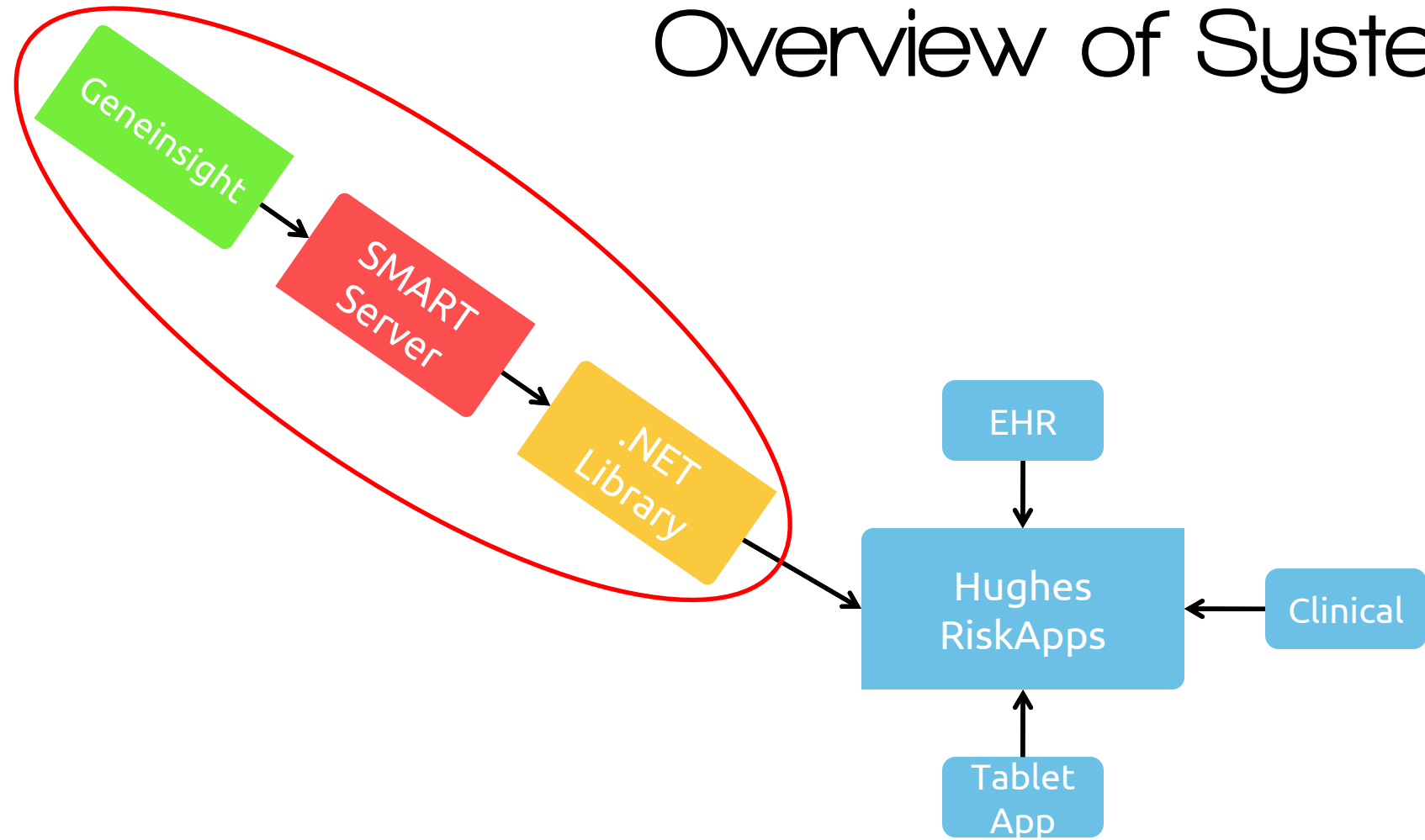- Preemptive surgery to remove cancer

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
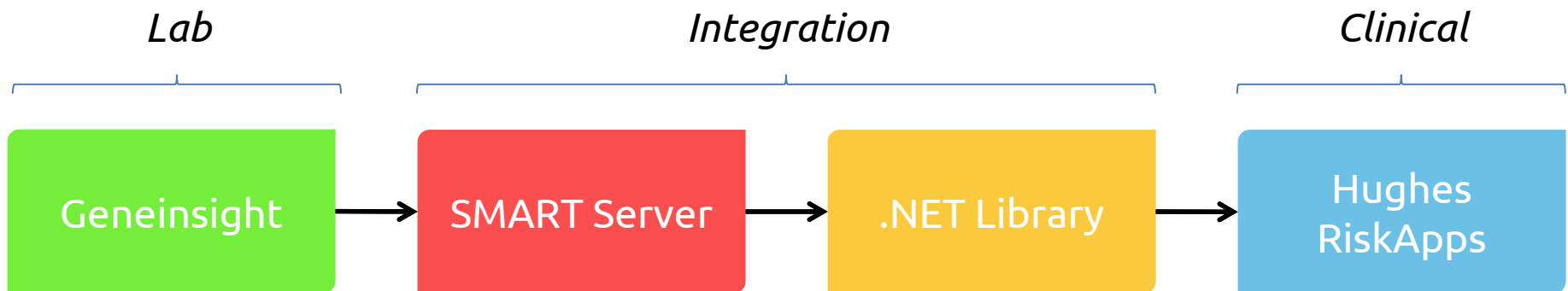- Faster screening means help for people with mutations

Overview of System

Goal

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

Lab　　　　　　　*Integration*　　　　*Clinical*

Geneinsight → SMART Server → .NET Library → Hughes RiskApps

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

- **S**ubstitutable **M**edical **A**pps & **R**eusable **T**echnology

- **F**ast **H**ealth **I**nteroperability **R**esource

- First SMART-enabled container in production

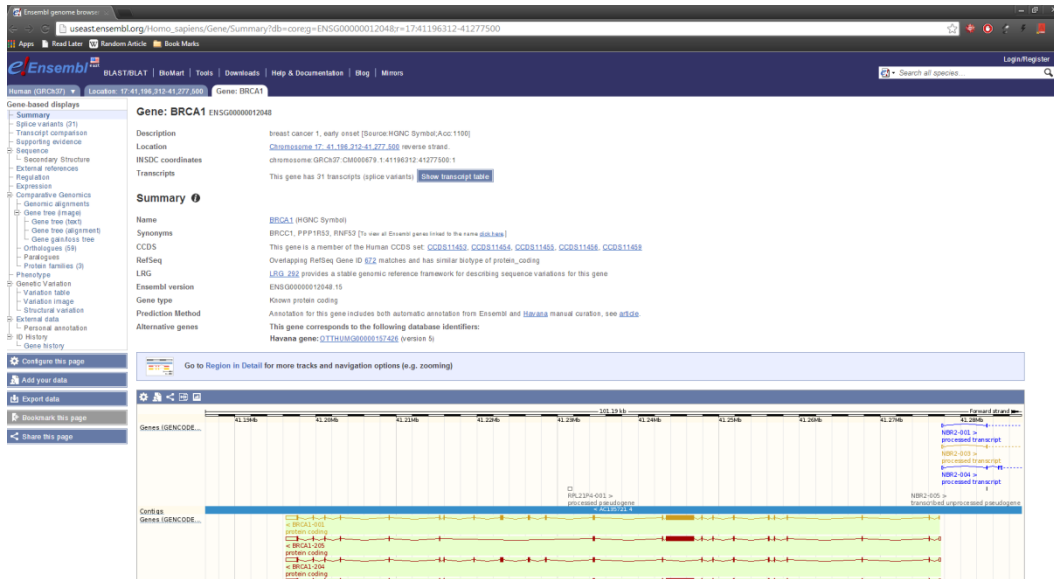- Emerging standard for transferring clinical data

- Easily deployable

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
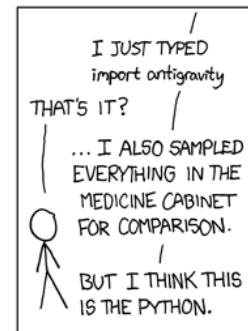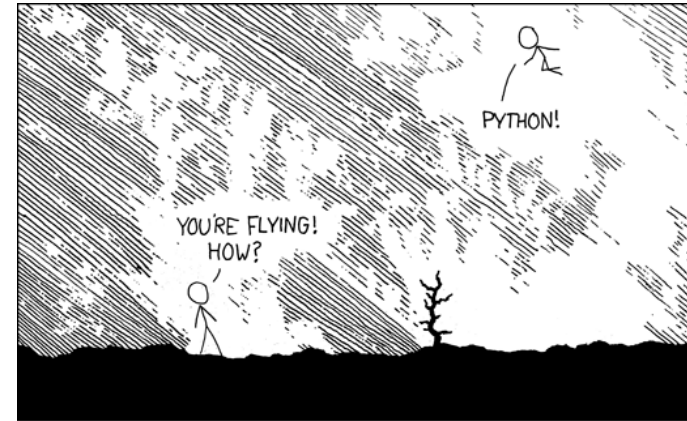
Conversion to FHIR


```python
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)

def get_coordinate_for_gene(term):
    term += ' ensembl'
    url = r'https://www.google.com/search?rlz=1C1LENP_enUS539US539&es_sm=93&q='+
    g_search = requests.get(url)
    search_results = BS(g_search.text)
    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:
        return coordinate_p.search(ensembl_page).group(1)
    except:
        pass

# 2637 Genes
# around 3 sec conversion for each
def convertGenes():
    genes = open('gi_genes.txt', 'r')
    coordinates = open('gi_coordinates.txt', 'w')
    coordinates.write('coordinates={')
    for gene in string.split(genes.read(), '\n'):
        coordinate = get_coordinate_for_gene(gene)
        if coordinate:
            coordinates.write('"' + gene + '":"' + coordinate + '",')
        else:
            coordinates.write('"' + gene + '":"' + str(0) + '",')

if __name__ == '__main__':
    #test here..
    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

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