SMART Genomics API

-- Standardizing genomics API to facilitate utilization of genetic data in clinics and laboratories

Yishen Chen
Mentor: Dr. Gil Alterovitz
Third PRIMES Annual Conference

May 19th, 2013
Problems of Current Genomics Communication

- No standard API (Application Programming Interface)
- Lack means of communication
- Not integrated with clinical network
"Plug" doesn't work
Representation of Genetic Data

- Contents are essentially the same
  - Sequences (ATGC..) and genetic files

- Data are structured differently

- Different approaches of authentication
  - Most follow OAuth but implement differently
SMART Provides a Solution

- **Substitutability Medical Applications Reuseable Technology**

- Generic design of API
  - Developers focus on the app

- Integration with clinical data
SMART Genomics API

- Approaches substitutability differently because of differences between clinical and genomic data

- Integration of genetic data among heterogeneous sources

- Bridging clinics by providing CCDA Genomics
  - CCDA (Consolidated Clinical Document Architecture) -- an adopted format for transferring medical documents
  - Summary of a patient's genetic report
SMART Genomics API

- **Authentication**
  - OAuth2 - open standard for authentication
  - Two versions
    - Container - Container
    - App - Container

- **Data representation**
  - Structured in FHIR-style (Fast Health Interoperability Resource)
  - Interlinked data models
  - Grouped by records
Alignment → Variant → Sequences → Result(Files) → Lab → ?
Resource Example

Sequence

- patient: Resource(Patient) 1..1
- type: Code 1..1
- read: string 1..1
- quality: integer 0..1
- quantity: float 0..1
- lab: Resource(Lab) 0..1

Coordinate

- chromosome: string 1..1
- start: integer 1..1
- stop: integer 1..1
SMART Genomics API

Resources are organized by 'record':

- A **record** groups all resources owned by a patient

- An account contains one or more **records**

- A **record** can either be internal or external
  - Both are perceived as the same by external source
Current Implementation

- Imports data from cloud services and translates to SMART Genomics' format

- Demo app
  - Genomics Advisor (collaboration with Peijin Zhang)
Button to import data

Available Records

<table>
<thead>
<tr>
<th>Record Id</th>
<th>Patient Name</th>
<th>23andme Profile Id</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mike Polcari</td>
<td>e82ef7a3c2doc585</td>
</tr>
<tr>
<td>2</td>
<td>Lilly Demo</td>
<td>SP1_MOTHER_V3</td>
</tr>
<tr>
<td>3</td>
<td>LillyDemo</td>
<td>SP1_FATHER_V3</td>
</tr>
</tbody>
</table>

Imported data
**Demo, Lilly**

DOB: 1999-12-08  
AGE: 13  
SEX: FEMALE

### Type 1 Diabetes

<table>
<thead>
<tr>
<th>SNP</th>
<th>Locus</th>
<th>CHRM</th>
<th>Code</th>
<th>Risk</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs5753037</td>
<td>22Q12</td>
<td>22</td>
<td>CC</td>
<td>0.93</td>
<td>37.2%</td>
</tr>
<tr>
<td>rs3129934</td>
<td>HLA-DRB1</td>
<td>6</td>
<td>CC</td>
<td>1.34</td>
<td>70.6%</td>
</tr>
<tr>
<td>rs2642161</td>
<td>PTEN2</td>
<td>18</td>
<td>TT</td>
<td>0.91</td>
<td>70.1%</td>
</tr>
<tr>
<td>rs3758013</td>
<td>UBASH3A</td>
<td>21</td>
<td>CC</td>
<td>0.88</td>
<td>30.2%</td>
</tr>
</tbody>
</table>

**Total Relative Risk: 1.0**

### Type 2 Diabetes

<table>
<thead>
<tr>
<th>SNP</th>
<th>Locus</th>
<th>CHRM</th>
<th>Code</th>
<th>Risk</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2877716</td>
<td>APCS</td>
<td>3</td>
<td>CC</td>
<td>1.05</td>
<td>59.3%</td>
</tr>
<tr>
<td>rs4402960</td>
<td>IGFBP2</td>
<td>3</td>
<td>GG</td>
<td>0.92</td>
<td>46.2%</td>
</tr>
<tr>
<td>rs52153</td>
<td>KCNJ11</td>
<td>11</td>
<td>TT</td>
<td>0.90</td>
<td>45.2%</td>
</tr>
<tr>
<td>rs2337932</td>
<td>KCNJ11</td>
<td>11</td>
<td>CC</td>
<td>1.03</td>
<td>86.5%</td>
</tr>
<tr>
<td>rs7903146</td>
<td>TCF7L2</td>
<td>10</td>
<td>CC</td>
<td>0.82</td>
<td>51.8%</td>
</tr>
<tr>
<td>rs7578597</td>
<td>THADA</td>
<td>2</td>
<td>TT</td>
<td>1.03</td>
<td>81.4%</td>
</tr>
<tr>
<td>rs7961581</td>
<td>TSPAN8</td>
<td>12</td>
<td>CC</td>
<td>0.95</td>
<td>53.4%</td>
</tr>
</tbody>
</table>

**Total Relative Risk: 0.7**

### Hypertension

<table>
<thead>
<tr>
<th>SNP</th>
<th>Locus</th>
<th>CHRM</th>
<th>Code</th>
<th>Risk</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12413409</td>
<td>CYPI7A1</td>
<td>10</td>
<td>GG</td>
<td>1.03</td>
<td>82.0%</td>
</tr>
<tr>
<td>rs17367594</td>
<td>MTHFR</td>
<td>1</td>
<td>AA</td>
<td>1.03</td>
<td>74.0%</td>
</tr>
</tbody>
</table>

**Total Relative Risk: 1.06**

### Coronary Heart Disease

<table>
<thead>
<tr>
<th>SNP</th>
<th>Locus</th>
<th>CHRM</th>
<th>Code</th>
<th>Risk</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1746048</td>
<td>CXCL12</td>
<td>10</td>
<td>CC</td>
<td>1.05</td>
<td>70.6%</td>
</tr>
<tr>
<td>rs2991384</td>
<td>MRAA</td>
<td>1</td>
<td>CC</td>
<td>1.07</td>
<td>51.8%</td>
</tr>
<tr>
<td>rs9818870</td>
<td>MRAA</td>
<td>3</td>
<td>CC</td>
<td>0.96</td>
<td>71.8%</td>
</tr>
<tr>
<td>rs7739181</td>
<td>PHACTR1</td>
<td>6</td>
<td>GG</td>
<td>1.00</td>
<td>42.2%</td>
</tr>
<tr>
<td>rs6735587</td>
<td>WDR12</td>
<td>2</td>
<td>TT</td>
<td>0.95</td>
<td>74.0%</td>
</tr>
</tbody>
</table>

**Total Relative Risk: 1.11**

### Disease Information

Patient is not at increased genomic risk for any Diabetes related comorbidities

### Drug Advice

No information available
Future plans

● **Design more data models**
  ○ Planning to add microarray (lab for gene expression)

● **Write more apps to demonstrate the API's functionalities**
  ○ A user friendly API playground for developer to understand basic aspects of the API

● **Distribute the code for the community to adopt**

● **Deploy as a reference container**
  ○ Developers can register their app and try it out
Register new app

App name: Genomics Advisor
App description: Risk Calculation of Diseases
Launch uri: localhost:8000/
Redirect uri: localhost:8000/receive_code/

Register
Acknowledgements

- Dr. Gil Alterovitz for initiating this project and giving me advices
- Jason Evans for advices on designing the API calls
- Dr. Khovanova and other organizers of PRIMES for providing me this valuable opportunity
- Peijin Zhang for writing the demo app, Genomics Advisor
- My parents for supporting me
SMART Genomics API

Two aspect of the API

- Authentication
  - OAuth2 - open standard for authentication

- Data representation
  - use gene coordinate as identifier -- e.g. chr13_123_124
  - Implemented FHIR
  - Resources grouped by records
Authentication

- Follows OAuth2 protocol
  - access token
  - refresh token

- Two versions
  - app-container -- throwaway access without refresh token
  - container-container -- permanent access

- Enables data sharing
Owner authorizes access

Access Token

App accesses data with token

Smart Container

Refresh Token