

HUMAN GENETIC VARIATION VIEWER

Saket Choudhary¹, Leyla Garcia² and Andrew Nightingale²

October 30, 2014

C	G	C	A	T	C	G	A	G	C	T
C	G	C	G	T	C	G	A	G	C	T

¹University of Southern California and ²EMBL-EBI

- Motivation
- Solution
- Demo and Use-Cases
- Implementation
- Improvements

MOTIVATION

The power of the unaided mind is highly overrated. The real powers come from devising external aids that enhance cognitive abilities. – Donald Norman

- NGS has given rise to catalog of genetic variants: dbSNP, COSMIC...

- NGS has given rise to catalog of genetic variants: dbSNP, COSMIC...
- Loads of data, but limited relevant information: Benign, Damaging, Intermediate

- NGS has given rise to catalog of genetic variants: dbSNP, COSMIC...
- Loads of data, but limited relevant information: Benign, Damaging, Intermediate
- Lots of mutations \implies Loads of *differing* predictions

- NGS has given rise to catalog of genetic variants: dbSNP, COSMIC...
- Loads of data, but limited relevant information: Benign, Damaging, Intermediate
- Lots of mutations \implies Loads of *differing* predictions
- Non consensus scoring mechanisms: SIFT and Polyphen predictions for example can be entirely opposite

- NGS has given rise to catalog of genetic variants: dbSNP, COSMIC...
- Loads of data, but limited relevant information: Benign, Damaging, Intermediate
- Lots of mutations \implies Loads of *differing* predictions
- Non consensus scoring mechanisms: SIFT and Polyphen predictions for example can be entirely opposite
- Exploratory visualization is the first step towards discovering patterns

- NGS has given rise to catalog of genetic variants: dbSNP, COSMIC...
- Loads of data, but limited relevant information: Benign, Damaging, Intermediate
- Lots of mutations \implies Loads of *differing* predictions
- Non consensus scoring mechanisms: SIFT and Polyphen predictions for example can be entirely opposite
- Exploratory visualization is the first step towards discovering patterns
- Variation viewers are *absent*, if not, provide limited flexibility

- A graphical hub to present annotated variants from different sources

- A graphical hub to present annotated variants from different sources
- Present information at different levels in a coherent manner

- A graphical hub to present annotated variants from different sources
- Present information at different levels in a coherent manner
- Scalable, and Interactive exploration on the browser

DEMO

DETAILS

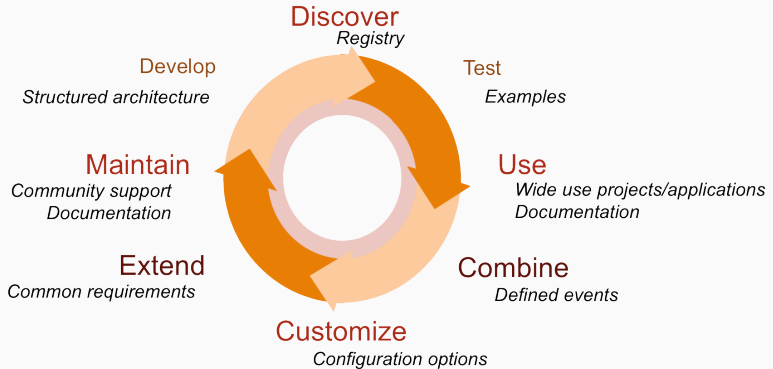
- Entirely written in javascript using the *d3js* library

- Entirely written in javascript using the *d3js* library
- Deployed as a BioJS component

- Entirely written in javascript using the *d3js* library
- Deployed as a BioJS component
- Events system that triggers events on user actions, allows cross-component communication

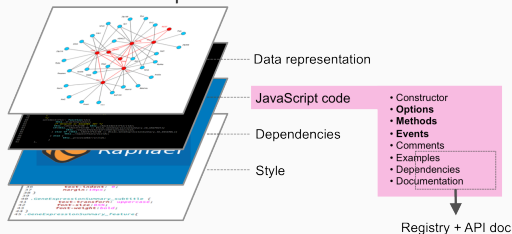
- BioJS is a javascript library for developing visualization of the biological data

- BioJS is a javascript library for developing visualization of the biological data



WHY BIOJS

Reusable components that can talk to each other



```
{  
  "id": "P00533_variant226",  
  "sourceIds": ["COSM1090877", "COSM1090879"],  
  "position": 541,  
  "wild_type": "L",  
  "mutation": "I",  
  "frequency": 0.0,  
  "polyphenPrediction": "benign",  
  "polyphenScore": 0.0,  
  "siftPrediction": "tolerated",  
  "siftScore": 0.86,  
  "somaticStatus": 1,  
  "consequenceTypes": "missense variant",  
  "cytogeneticBand": "7p11.2",  
  "genomicLocation": "7:g.55229314C>A"  
}
```

- Pre-generated JSON files
- Current version uses files generated by an unpublished webservice at EBI
- Protein variants, though not specific to it

- Supports JSON formatted files, alpha VCF support
- User defined scoring criteria
- Different levels of information
 - Overview: Condensed information
 - Zoomed View: All annotations
- Loading proteins through URL parameters
- SIFT, Polyphen,

- Identifying most or least mutated sites across proteins

- Identifying most or least mutated sites across proteins
- Discover differences between different scoring criteria

- Identifying most or least mutated sites across proteins
- Discover differences between different scoring criteria
- Benchmarking predictions

- VCF support, almost there

- VCF support, almost there
- Integration with Galaxy, web based bioinformatics workflows

- VCF support, almost there
- Integration with Galaxy, web based bioinformatics workflows
- Performance improvements

CONCLUSION

- A tool for visualizing genetic variants
- Supports visualization of different levels of information
- Cross component talks
- User defined and user controlled

Google, for running the Google Summer of Code 2014.

QUESTIONS?