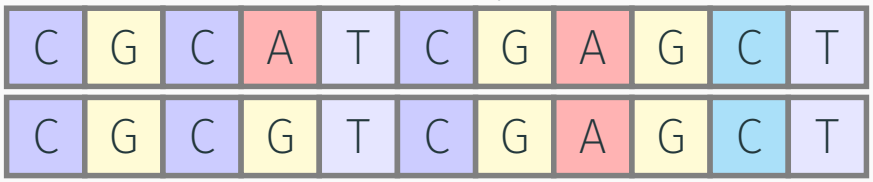
HUMAN GENETIC VARIATION VIEWER

Saket Choudhary [1], Leyla Garcia[2] and Andrew Nightingale[2]

October 30, 2014



[1]University of Southern California and [2]EMBI-EBI

OUTLINE

- Motivation
- Solution
- Demo and Use-Cases
- Implementation
- Future Work

MOTIVATION

VISUALIZATIONS ARE POWERFUL!

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

MOTIVATION

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

MOTIVATION

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

MOTIVATION

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

MOTIVATION

- 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

MOTIVATION

- 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

MOTIVATION

- 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

SOLUTION

- A graphical hub to present annotated variants from different sources

SOLUTION

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

SOLUTION

- 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

http://saketkc.github.io/biojs

DETAILS

IMPLEMENTATION

- Entirely written in javascript using the *d3js* library

IMPLEMENTATION

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

IMPLEMENTATION

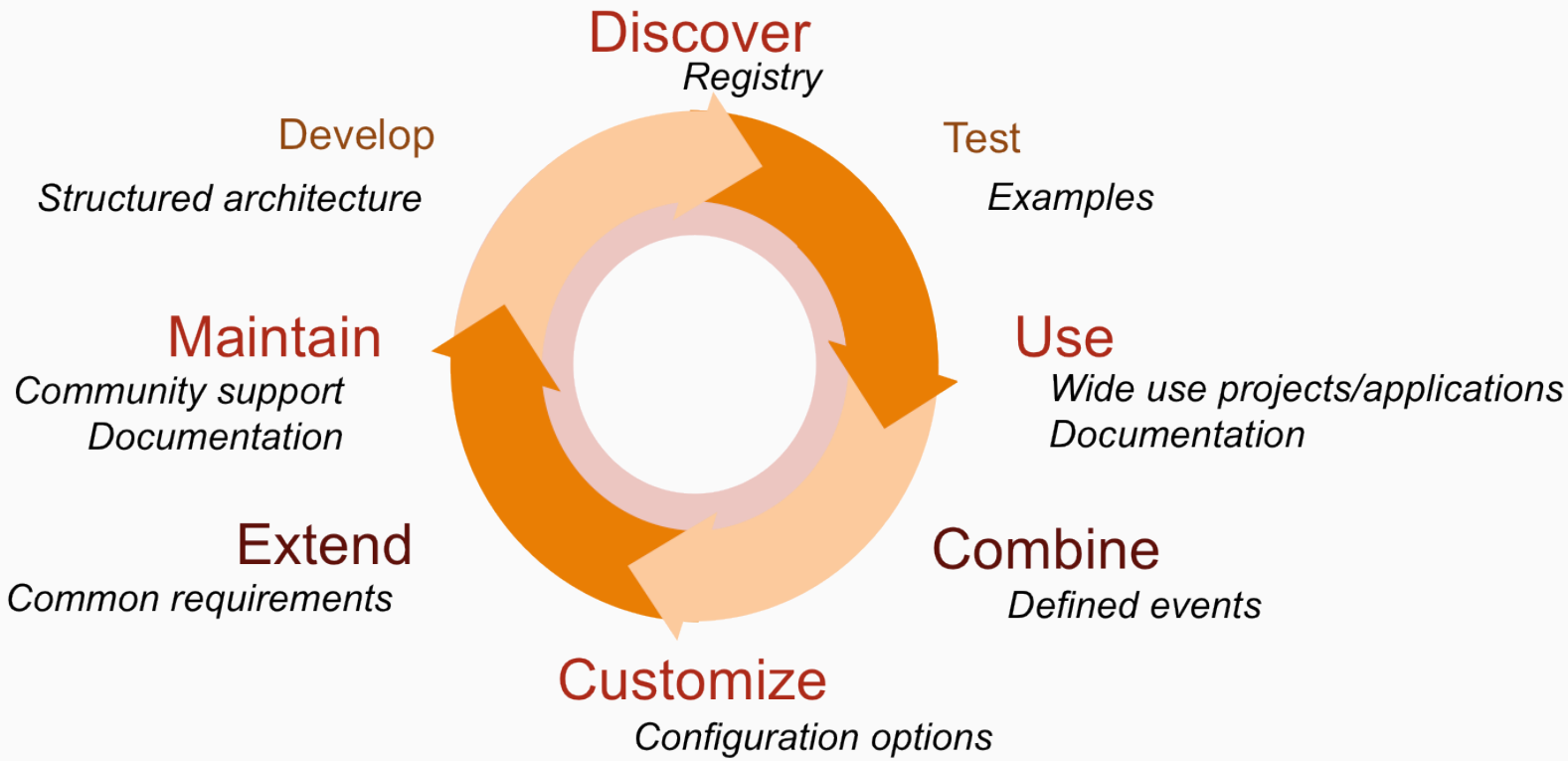
- Entirely written in javascript using the *d3js* library
- Deployed as a BioJS component
- Events system that triggers events on user actions, allows cross-component communication

WHY BIOJS

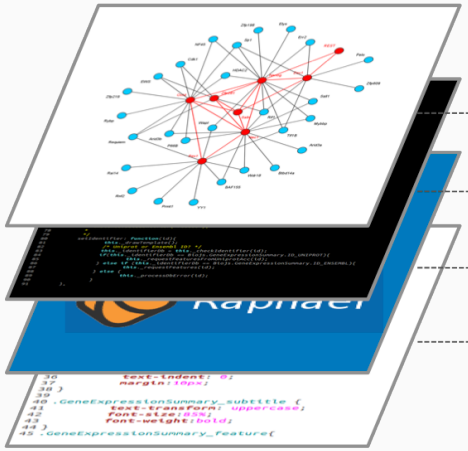
- BioJS is a javascript library for developing visualization of the biological data

WHY BIOJS

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



Reusable components that can talk to each other



Data representation

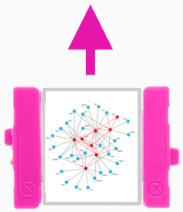
JavaScript code

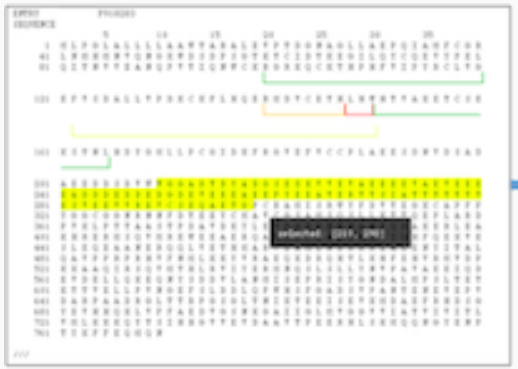
- Constructor
- **Options**
- **Methods**
- **Events**
- Comments
- Examples
- Dependencies
- Documentation

Dependencies

Style

Registry + API doc





onSelectionChanged

setSelection

```
{
"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"
}
```

DATA INPUT

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

FEATURES

- 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, ...

USE CASES

- Identifying most or least mutated sites across proteins

USE CASES

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

USE CASES

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

IMPROVEMENTS

- VCF support(almost there)

IMPROVEMENTS

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

IMPROVEMENTS

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

IMPROVEMENTS

- VCF support(almost there)
- Integration with Galaxy, web based bioinformatics workflows
- Performance improvements
- Interaction with 3D Protein viewer to highlight domains

CONCLUSION

SUMMARY

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

ACKNOWLEDGEMENTS

Google, for running the Google Summer of Code 2014.

Questions?