HUMAN GENETIC VARIATION VIEWER

Saket Choudhary ¹, Leyla Garcia² and Andrew Nightingale²

¹University of Southern California and ²EMBL-EBI

October 30, 2014
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
• NGS has given rise to catalog of genetic variants: dbSNP, COSMIC...
· NGS has given rise to catalog of genetic variants: dbSNP, COSMIC...
· Different categories of mutations: Benign, Damaging, Intermediate
· NGS has given rise to catalog of genetic variants: dbSNP, COSMIC...
· Different categories of mutations: Benign, Damaging, Intermediate
· Lack of consensus amongst scoring mechanisms: SIFT $\neq$ Polyphen
• NGS has given rise to catalog of genetic variants: dbSNP, COSMIC...
• Different categories of mutations: Benign, Damaging, Intermediate
• Lack of consensus amongst scoring mechanisms: SIFT \neq Polyphen
• Lots of mutations \implies Loads of differing predictions
· NGS has given rise to catalog of genetic variants: dbSNP, COSMIC...
· Different categories of mutations: Benign, Damaging, Intermediate
· Lack of consensus amongst scoring mechanisms: SIFT ≠ Polyphen
· Lots of mutations → Loads of differing predictions
· Exploratory visualization is the first step towards discovering patterns, comparing consensus, aggregating predictions
NGS has given rise to catalog of genetic variants: dbSNP, COSMIC...

- Different categories of mutations: Benign, Damaging, Intermediate
- Lack of consensus amongst scoring mechanisms: SIFT ≠ Polyphen
- Lots of mutations ➞ Loads of *differing* predictions

- Exploratory visualization is the first step towards discovering patterns, comparing consensus, aggregating predictions

- Variation viewers are practically *absent*, those present provide limited flexibility
· A graphical hub to present annotated variants from different sources
· A graphical hub to present annotated variants from different sources

· Incremental levels of abstractions
• A graphical hub to present annotated variants from different sources
• Incremental levels of abstractions
• Scalable and Interactive exploration on the web browser
Figure: Overview
Figure: Zoomed View
Demo

http://saketkc.github.io/biojs
DETAILS
· Written in javascript using the *d3js* library
IMPLEMENTATION

- Written in javascript using the d3js library
- Deployed as a BioJS component
· Written in javascript using the *d3js* library
· Deployed as a BioJS component
· Flexible system with ability to capture and react to user-actions
· 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

Data representation

JavaScript code
- Constructor
- Options
- Methods
- Events
- Comments
- Examples
- Dependencies
- Documentation

Dependencies

Style

Registry + API doc

onSelectionChanged

setSelection
· Pre-generated JSON files
· Current version uses files generated by an unpublished webservice at EBI
· Protein variants only
{ 
  "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"
}
· User defined scoring criteria
- User defined scoring criteria
- Different levels of abstractions, tooltips
FEATURES

- User defined scoring criteria
- Different levels of abstractions, tooltips
  - Overview: Condensed information
· User defined scoring criteria
· Different levels of abstractions, tooltips
  · Overview: Condensed information
  · Zoomed View: All annotations
· User defined scoring criteria
· Different levels of abstractions, tooltips
  · Overview: Condensed information
  · Zoomed View: All annotations

· SIFT, Polyphen, …
· User defined scoring criteria
· Different levels of abstractions, tooltips
  · Overview: Condensed information
  · Zoomed View: All annotations
· SIFT, Polyphen, ....
· Scalable, adaptable to new scores, mutation categories
· Identifying most or least mutated sites on a protein
• Identifying most or least mutated sites on a protein
• Discover differences between different scoring criteria
USE CASES

- Identifying most or least mutated sites on a protein
- Discover differences between different scoring criteria
- Benchmarking predictions
· VCF support(almost there!)
· 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
- 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
- Limited applications as a standalone tool, more usable with Protein Features Viewer
- Supports visualization of different levels of information
- Cross component talks
- User defined and user controlled
ACKNOWLEDGEMENTS

Google | Google Summer of Code 2014
BioJS Community
Egor Dolzhenko USC MCB
Questions?