Puzzle: VCF/GEMINI interface for genetic disease analysis

Authors: Robin Andeer, Måns Magnusson, Henrik Stranneheim
Affiliations: Department of Molecular Medicine and Surgery, Science for Life Laboratory, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden
Email: robin.andeer@scilifelab.se
Project website: https://robinandeer.gitbooks.io/puzzle/content/
Source code: https://github.com/robinandeer/puzzle
License: MIT

VCF and BCF files are the ubiquitous storage formats for variant calls. However, at their core it's no more than tab-delimited text files. They lack a powerful and standardized data model to enable sophisticated genetic analysis. Converting VCFs into a GEMINI database goes a long way to enable this type of complex processing. However, they are not human friendly when it comes to inspecting their contents. You need knowledge of SQL and the internal schema. This isn't something you can expect from your average clinical or research users.

Puzzle is a web interface for variant calling resources (VCF files and GEMINI databases). It lets you organize, inspect, filter, annotate, explore and analyze variant calls in a web browser. It provides a unified and intuitive interface for all loaded resources and can be used by anyone without computer programming literacy. It's already used extensively by clinicians and researchers at Karolinska Hospital working with a wide range of genetic disorders.

These are the main highlights of Puzzle:

- Simple installation (PyPI or Bioconda) and minimal setup
- Automatic discovery of resources, just point to a folder and Puzzle will do the rest
- Support for single nucleotide variant (SNV) and structural variant (SV) analysis
- Gene panels and human phenotype ontology (HPO) support
- Automatic annotations on gene and transcript level
- Support for grouping samples into logical groups like families
- Support for comments and highlighting of interesting variants

We set out to make Puzzle simple to install and intuitive to use. It's implemented in Python as a stand alone web server with a light weight SQLite backend. You can visualize your variant calls in minutes with minimal prerequisites. Simply point it to a directory with some VCF files to get started:

```
$ puzzle view tests/fixtures/
```