A novel way of visualising eQTLs relative to SNP-SV pairs using Gosling.js

1. Introduction

Quantitative trait locus (QTL) studies connect genotypes with tissue/cell type specific cellular functions. Currently, QTL studies are mostly done on single nucleotide changes, but as SVs are bigger and have greater impact on traits, SV-QTL connections are of great interest.

2. Objective

The goal of this research was to create a tool that streamlines the steps of searching for relevant SNP-SV and SNP-eQTL data, integrating them into a unified dataset, and providing visualizations wider relationships.

Legend

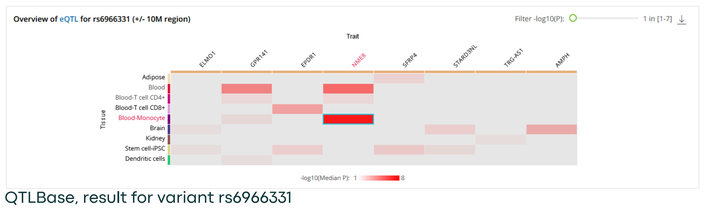
- **GWAS**: A **Genome-Wide Association Study** scans the genomes of many individuals to find genetic variants associated with specific traits.
- **SNP**: A **Single Nucleotide Polymorphism** is a single base-pair variation in the DNA sequence
- **SV**: A **Structural Variation** is a large-scale alterations of DNA.
- **eQTL**: An **expression Quantitative Trait Locus** is a genomic locus that correlates with variations in gene expression levels among individuals.

3. Related works

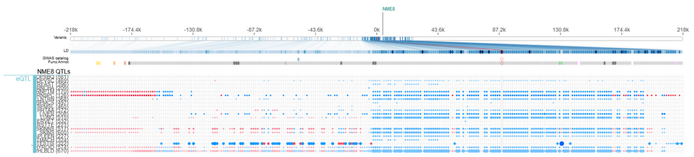
There are numerous tools to visualise the connections between:

- SNPs and SV
- Genes and their expression in tissues (eQTL)
- SNPs and diseases

There is however a lack of tools that visualise the connection between SVs and QTLs.



QTLBase, result for variant rs6966331



GTEx Portal, result for variant rs6966331

QTLBase and GTExPortal can show correlation of tissues with genes or individual SNPs within genes respectively, but not the correlation of SNPs to SVs.

4. Materials and Method

The created visualisation is built as an extention on snpXplorer, a publicly available web-server for displaying GWAS associations. For the visualisation library, Gosling.js was chosen, as it offers a nice selection of functions for visualising genomic data.

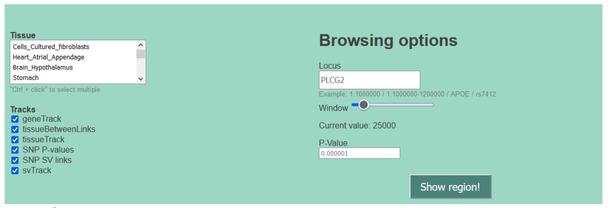
Two main datasets were used:

- SNP-SV data
- SNP-eQTL data

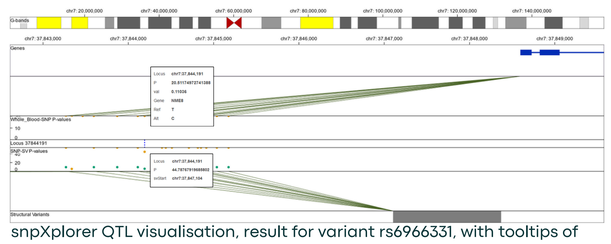
Upon querying, the datasets are first filtered. The SNP-eQTL data has to be transformed, such that it can be indexed on tissue, instead of SNP. Both datasets can then be joined together. Finally, the genes are translated from ensemble to gene names for convenience and their location is added.

5. Results

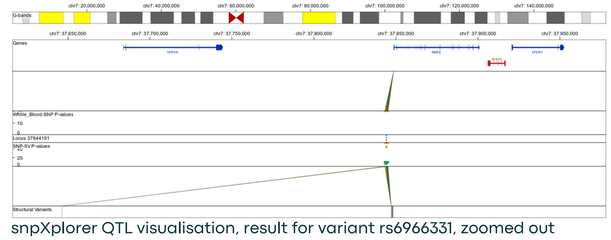
- The input form offers options for:
  - Tissues
  - Tracks to show
  - Target locus, range, gene, or variant
  - Window size
  - P-value for filtering
- The resulting plot displays at the top the chromosome of the target, and 7 main tracks containing:
  - genes
  - SNP-gene links
  - SNP-eQTL P-values
  - marker of the target
  - SNP-SV P-values
  - SNP-SV links
  - SVs
- Tracks can be deselected if unnecessary for a particular use case. This prevents data loading for these tracks, reducing query time.



Input form



snpXplorer QTL visualisation, result for variant rs6966331, with tooltips of the QTL and right SV related to the variant



snpXplorer QTL visualisation, result for variant rs6966331, zoomed out

6. Discussion & Recommendation

- Pandas is a easy to use dataframe library, but isn't optimized for querying genomic data.
- The CSV format is a common data format, but it limits how Gosling.js can interact with it. Switching to another format, such as BEDDB, would eliminate the reformatting of data.
- The SNP-gene links, SNP-eQTL P-values, SNP-SV P-values and SNP-SV links tracks currently each retrieve data seperately, due to limitations in Gosling.js.

7. Conclusion

The created tool succeeds in its original goal of overlapping and visualising the SNP-SV and SNP-eQTL datasets, which other tools failed to achieve.

Integration with snpXplorer codebase should be straightforward and no new dependencies were introduced.

Gosling.js offers integration of other visualisation libraries, so with this framework in place, new options for further development open up.

Authors Sonny Ruff s.ruff@student.tudelft.nl

Responsible Professor Marcel Reinders m.j.t.reinders@tudelft.nl

Supervisor Niccolo Tesi n.tesi@tudelft.nl

References

1. snpXplorer. (n.d.). https://snpxplorer.net/
2. Zheng, Z., Huang, D., Wang, J., Zhao, K., Zhou, Y., Guo, Z., . . . Li, M. J. (2020, January). QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. Nucleic Acids Research, 48 (D1), D983–D991. Retrieved 2024-04-23, from https://doi.org/10.1093/nar/gkz888 doi: 10.1093/nar/gkz888
3. QTLbase Home. (n.d.). http://www.mulinlab.org/qtlbase
4. GTEX Portal. (n.d.). https://gtexportal.org/

TU Delft