Visualising SNP-SV Interactions

Introduction

When genes get copied, mistakes happen. We call these mistakes mutations. Single Nucleotide Polymorphisms (SNPs) are the most common types of mutations and they consist of a change in one single nucleotide; One letter in a string of DNA changes to another. Another type of mutation is a Structural Variant (SV), there are several types of SVs, but they have one thing in common, namely that a string of nucleotides is different from the parent genome. There are different types of SVs

In Genome Wide Association Studies (GWAS) the relation between SNPs and genetic diseases is being studied. However a significat relation is not always found. Thanks to recent advancements in technology, we are now able to also study the relation between SNPs and SVs to more clearly determine the cause of genetic diseases. There already exist many tools to visualise results of GWAS, however because the technology to collect data on SNP-SV interactions is so new, none of these tools is capable of visualising these yet. This poster focuses on the visualisations to make these relations between SNPs and SVs clear.

Goal

The goal of this study is to provide a visualisation for SNP-SV interactions that matches visualisation tools for other genomic data.

Results

The final product consists of a web-application where the user can choose what data to visualise.The following user-flow diagram describes the steps necessary to select which data to plot:
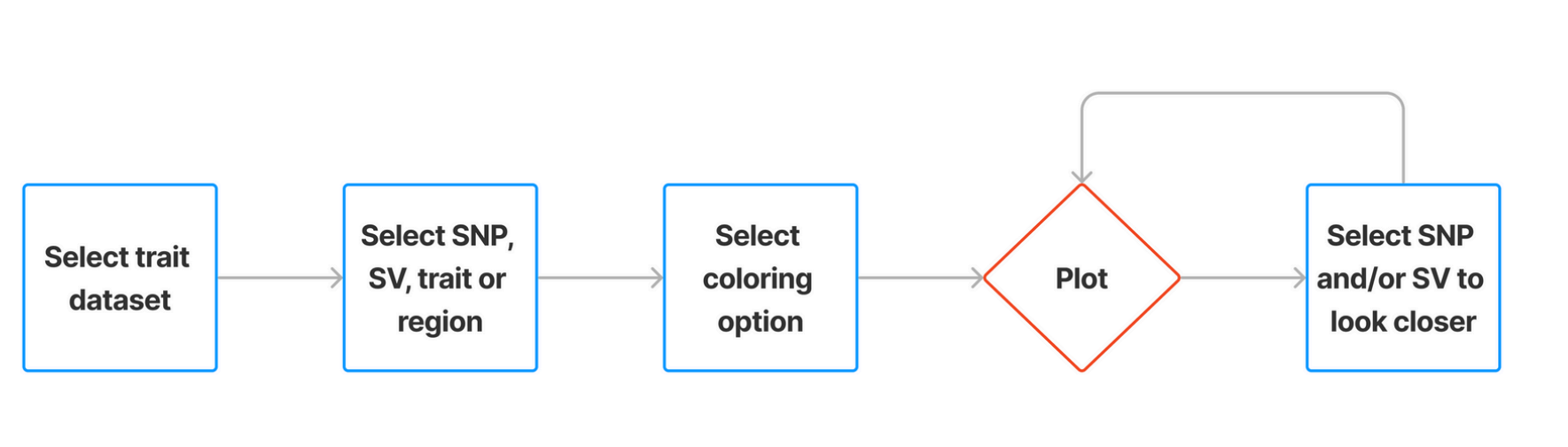
The plot, which can be seen in the white area on the right, consists of three layers. The top layer shows the p-value of the SNP-SV interaction, the middle layer shows the location of the SVs along the genome and the bottom layer shows the p-value of the SNP with a chosen trait.

Methods

There are many visualisation tools already out there. This project will build on snpxplorer.net [2] to add the required visualisations. For integration purposes with snpxplorer Python was chosen as the programming language of choice and Flask as the framework. Plotly was chosen as the package of choice for the easy integration into Flask.

In step 3 the user is able to customize the plot by choosing different color options:
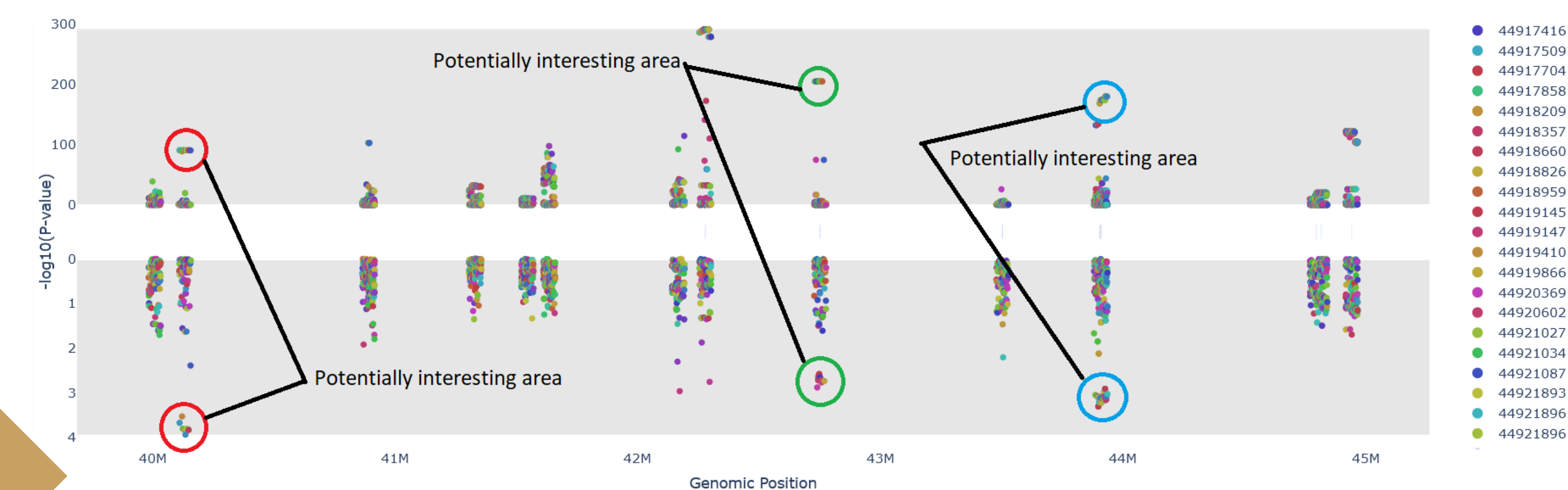1. SNP: each SNP-SV interaction will get it's own color
2. SV: each SV will get it's own color. SNPs get the same color as the SV they interact with.
3. Trait: each trait will get it's own color. SNPs get the same color as the trait.
4. SNP type: each type of SV will get it's own color. (4 colors, 1 for each unique letter in a DNA string).
5. SV type: each type of SV will get it's own color.
6. Correlation: The SNPs are colored based on whether their correlation with the SV is positive or negative.

Results - Case Study

The plot on the left was generated by selecting a range of chr1:40,000,000-45,000,000. Spanning 5 million base pairs. We only consider SNPs withing 25,000 base pairs either side of an SV. The color option is set to SNP, this way we can most easily find SNPs which have a high p-value on the top (with the SV) as well as on the bottom (with the chosen trait). However, because we are so far zoomed out some clusters are difficult to make out so we will need to zoom to them.

In the plot on the right we zoomed into the red area. The entire x-axis now spans roughly 60,000 base pairs and we can clearly make out more details. We can now also see that there are indeed SNPs which have a high p-value in both layers of the plot. We are now also zoomed in far enough to see the SV a little bit more clearly.
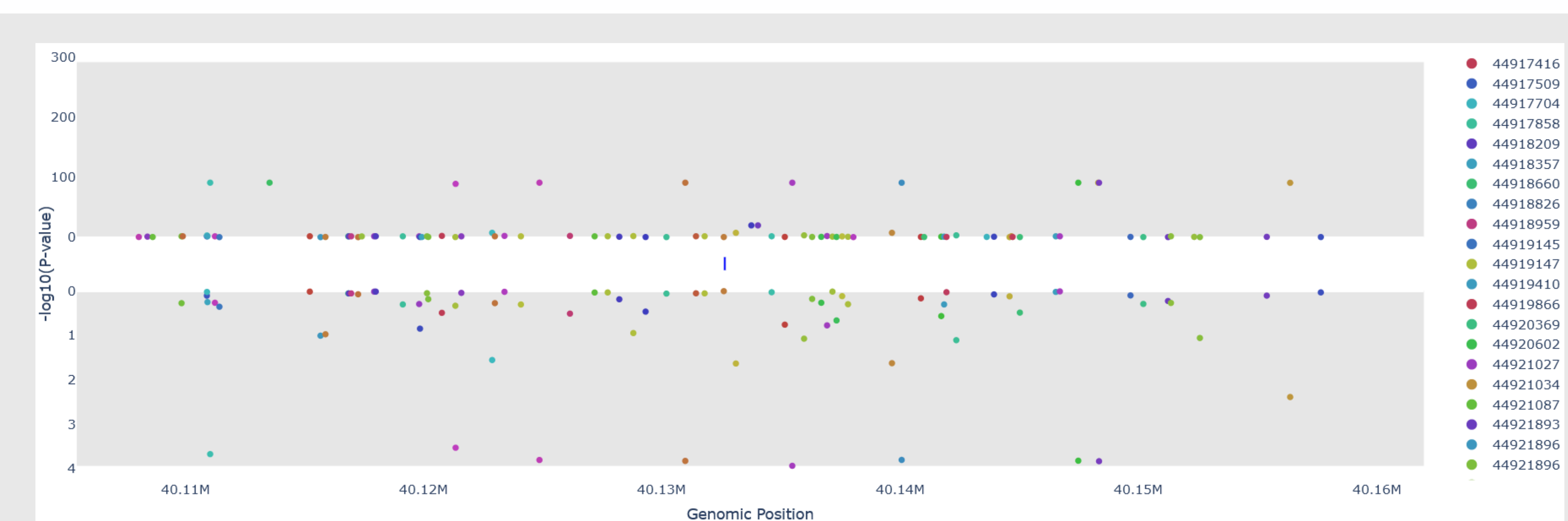
In order to get more information about a certain SNP or SV it is possible to hover over them.

POS: 40130980
-log10(P-value): 3.8242880407020277
RSID: rs55820078
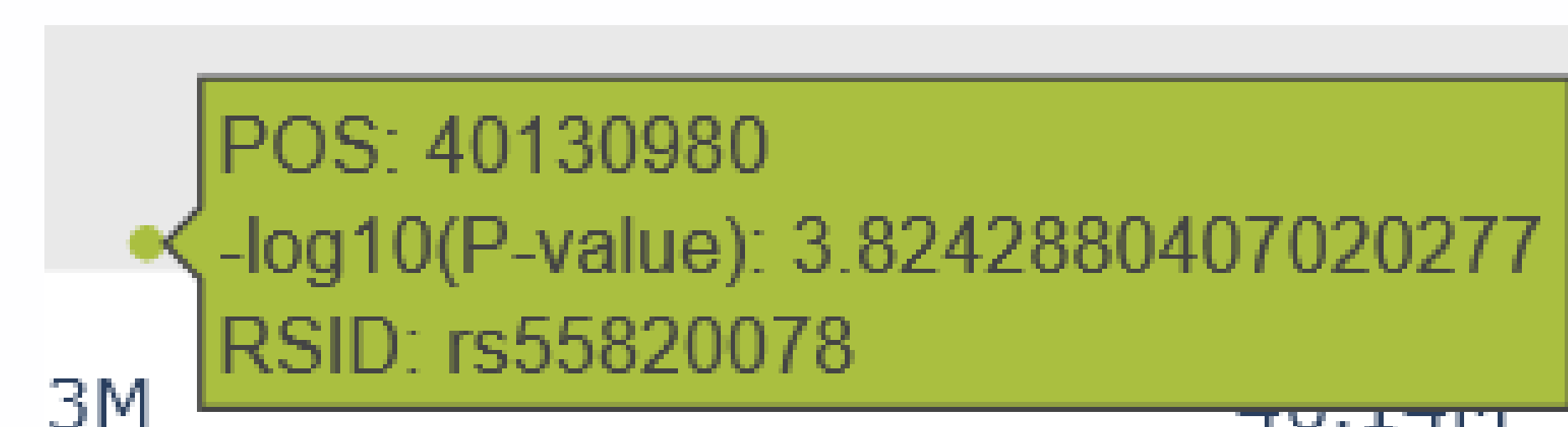
Or change the color setting.
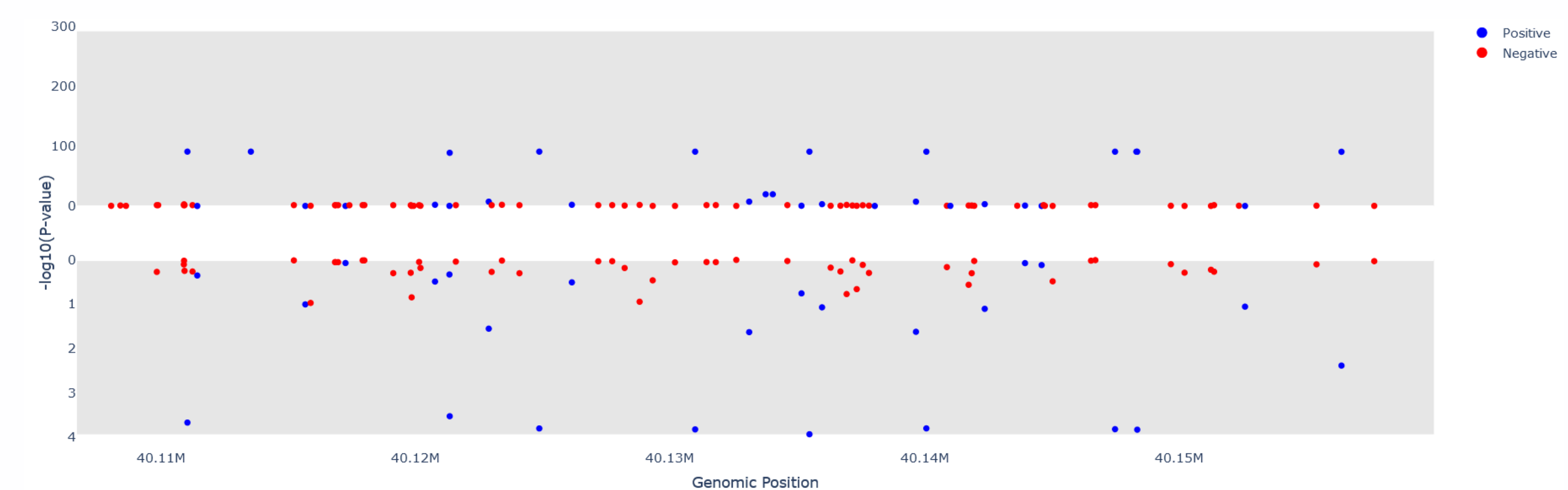
Conclusions

From this visualisation it is possible to clearly make out which SNP-SV interactions are important and which aren't. As seen in the example in the case study there are definetely some interesting SNP-SV interactions that are worth studying further. With this addition snpxplorer can now be used to aid in studying these interactions.

Future improvements

There are still some possible improvements for the visualisations:
- Adding an extra track at the bottom or in the middle to show the genomic position of genes can be very helpful to also compare these interactions to them
- More coloring options: I propose one for Linkage Disequilibrium (LD). LD is a way to express the correlation between SNPs. Showing this in the plot will be helpful to see whether SVs belong to these groups that SNPs often form or are maybe correlated with lots of different SNPs.

Author   Tom Jacobs
   t.g.jacobs@student.tudelft.nl
Supervisor   Niccolò Tesi
Responsible professor   Marcel Reinders

References   [1] N. Tesi, S. van der Lee, M. Hulsman, H. Holstege, and M. J. T. Reinders. snpXplorer: a web application to explore human SNP-associations and annotate SNP-sets. Nucleic Acids Research, 49(W1):W603–W612, 05 2021.

Code   The code is available on: https://github.com/TomJacobsGit/research_project