Association test between Inversions and Extreme Recombination Rates

Location patterns 2

Ruth Gómez Graciani

January 20, 2022

1 Background

In this report we try to answer to the question: are inversions less (or more) likely to become polymorphic under certain recombinatory contexts? Specifically, we test whether there is a tendency to avoid extremely high recombination rates (putative hotspots) and favor low-recombination regions (putative coldspots).

```
We will be analyzing 133 inversions, 54 NAHR and 79 NH.
```

2 Extreme recombination rates calculation

One of our main concerns is how to define extreme recombination regions (ERRs). In pedigree-based maps analysis, it is a common practice to calculate the standarized recombination rate (SRR), which is obtained by relativizing each rate to the population mean recombination rate [1]. Usually, it is considered high recombination rate when SRR>10 [2][3], which corresponds to the ~1% most extreme values in the distribution. In our case, we are using the population-based recombination map from [4], and I will try to keep values as raw as possible to avoid unnecessary biases.

In addition, I will explore multiple thresholds when defining ERRs to see how are inversions located relative to milder and harder hotspots and coldspots.

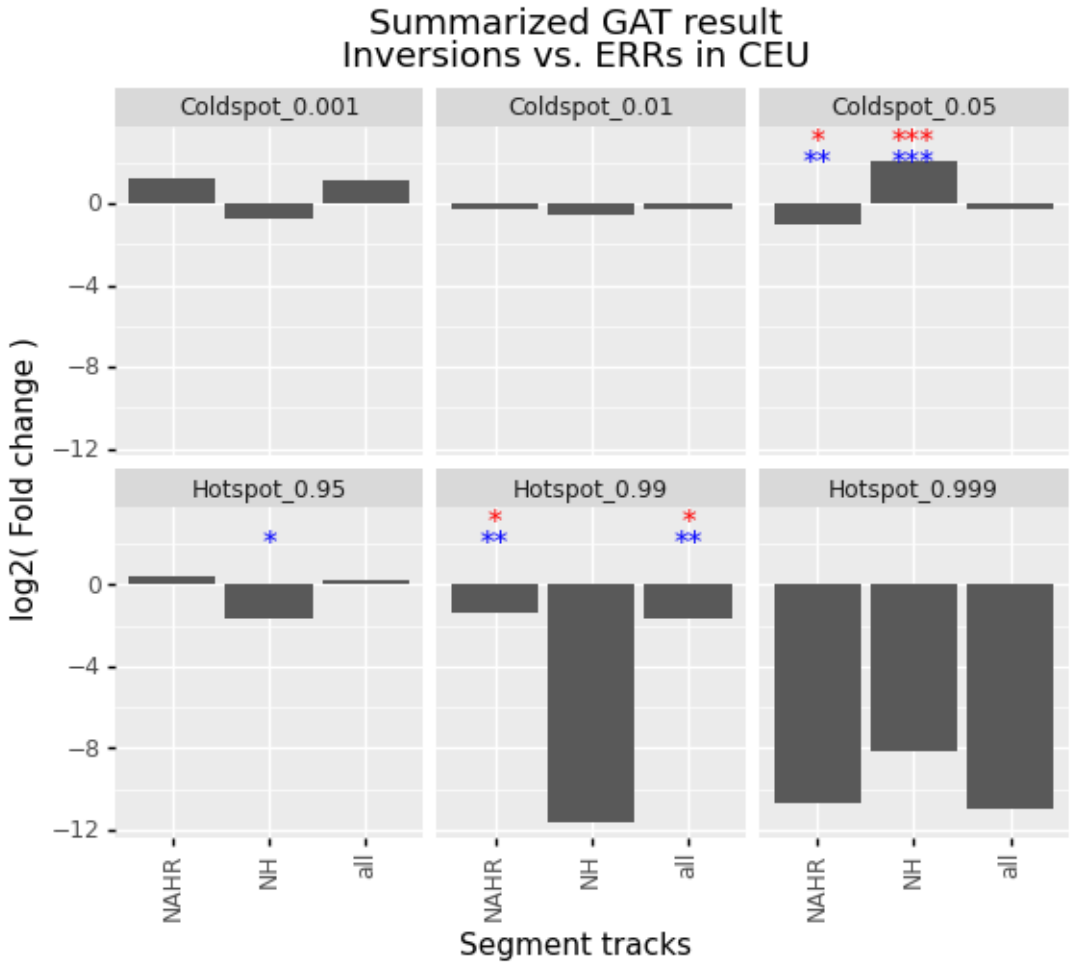
3 Association analysis with GAT

The genomic association tester (GAT) is a python package that makes permutation analyses to test for association between sets of genomic intervals. One or more tracks of segments and one or more tracks of annotations are required as input, and the program returns observed and expected number of overlapping nucleotides between the segments and the annotations, the corresponding fold-change, and the empirical p-value and q-value, among other useful measurements.

The same randomization of a set of segments (e.g. inversions) can be compared against multiple annotations (e.g. hotspots and coldspots), to test for a significant enrichment of any annotation in the segments, and different segments can be tested for enrichment of the same annotation. In either case, even when significant enrichments are found, inferences about the difference between two fold-change measurements are not possible. For exampe, we can not say with statistical confidence that

a segment 2-fold significantly enriched in annotation A and 3-fold in annotation B is more enriched in B elements, because A elements could be more abundant in the genome. To contrast fold enrichments of different sets of segments against the same annotation, there is a tool implemented in GAT.

I have tested GAT using the CEU population historical recombination measurements and 100000 permutations. The annotations were high recombination regions (Hotspots) above the 5, 1, and 0.1% most extreme and low recombination regions (Coldspots) below the 5, 1 and 0.1% most extreme. The most remarkable results have been summarized in Figure 1.




```
<ggplot: (8769247801766)>
```

Figure 1: Summary of GAT results when testing the overlap between inversions and ERRs. Remember that the magnitudes of fold-changes are not comparable. Significant p-values are indicated in blue and significant q-values in red.

NAHR inversions tend to avoid extreme coldspots, probably because higher recombination rates favor their generation. NH inversions, on the other hand, are significantly enriched in extremely

low recombination regions, which may be favorable because that way the probability of generating chromosomal aberrations during meiosis is lower.

Regarding hotspots, there are some non-significant fold changes < -8. In all four cases, the observed overlap between segment and track is 0, but the corresponding confidence interval is very broad and includes 0 as an expected value. This is caused by the scarcity of hotspots above the 0.1% most extreme in the 'Hotspot_0.999' panel, and by the small size of NH inversions in the 'Hotspot_0.99' panel, both of which make likely an overlap of 0 nucleotides between tracks. Both NH and NAHR inversions seem to avoid regions that have the highest 1% recombination rates in the genome.

4 Limitations and further work

This analysis is only a proof of concept, and many adjustments are still required, some of which may need discussion:

- As stated in the GAT documentation, if the section of the genome to use as a base for the permutations (workspace) is too large, fold enrichment values will be too optimistic. The workspace should be limited to those regions where the analysis is actually feasible and regions where any of the tracks could not be calculated should be excluded. As a starting point, regions where inversion detection is not possible and regions where the recombination rate windows are very large, indicating low SNP density and less reliable estimates, should be discarded.
- Properties of the sequence that are potentially correlated with the segment of interest and the annotations, but that are not interesting in out experiment (e.g. GC content, repeat density, gene/exon density), can be controlled by adding them as isochore tracks. These will be used to divide the workspace into smaller segments with similar properties.
- HsInv0501 due to its large size and HsInv0573 as a recombination-increasing factor, can be confounding, so we should test if removing them alters the identified patterns.

Once the enrichment analysis is adjusted to our needs, I will repeat it for multiple populations, which should not require long computing times, because GAT is fast and parallelizable. This will add robustness to our result by confirming whether this pattern is consistent along populations or meta-populations.

> **Note:** before the analysis of each opulation, we should delete those inversions with frequency 0 from the segments track.

References

[1] Claude Bhérer, Christopher L Campbell, and Adam Auton. Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. *Nature Communications*, 8:14994, apr 2017.

[2] Marta Farré, Diego Micheletti, and Aurora Ruiz-Herrera. Recombination rates and genomic shuffling in human and chimpanzee - A new twist in the chromosomal speciation theory. *Molecular Biology and Evolution*, 30(4):853–864, apr 2013.

[3] Augustine Kong, Gudmar Thorleifsson, Daniel F. Gudbjartsson, Gisli Masson, Asgeir Sigurdsson, Aslaug Adalbjorg Jonasdottir, G. Bragi Walters, Aslaug Adalbjorg Jonasdottir, Arnaldur Gylfason, Kari Th. Kristinsson, Sigurjon A. Gudjonsson, Michael L. Frigge, Agnar Helgason, Unnur Thorsteinsdottir, and Kari Stefansson. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature*, 467(7319):1099–1103, oct 2010.

[4] Jeffrey P Spence and Yun S Song. Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Science Advances*, 5(10), 2019.