Single-Cell Analysis of the 3D Topologies of Genomic Loci Using Genome Architecture Mapping

We will study on the largest portion of the HLB – the HIST1 region.

HIST1 is located on mouse chromosome 13 at the following coordinates:

Start:21.7 Mb

Stop: 24.1 Mb

Genomic Region of Interest

1. extract the HIST1 region from the segmentation table
2. extract relevant NPs for HIST1 (contain at least one window in the region of interest)

**NP Similarities**

1. compute the Jaccard *index*[[1]](#footnote-1) (aka the Jaccard *similarity* coefficient) for each pair of relevant NPs
2. store the computed values in a matrix; the matrix should contain *similarities* for all pairs of relevant NPs

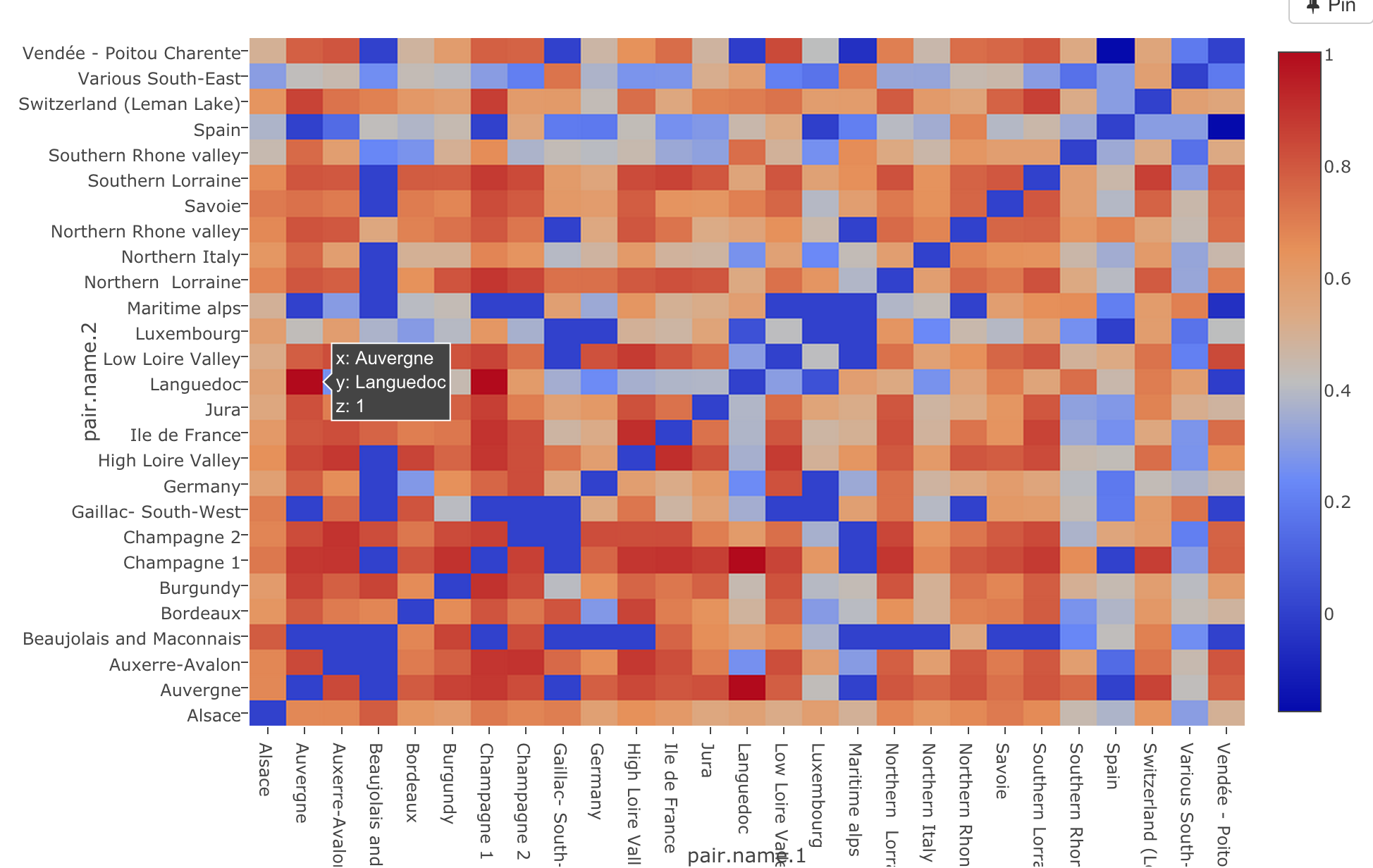
NOTE: You should implement the technique for computing similarity of asymmetric binary attributes1.

**NP Differences**

1. compute the Jaccard *distance*1 (use the method for asymmetric binary attributes**1**) for each pair of relevant NPs
2. store the computed values in a matrix; the matrix should contain *distances* for all pairs of relevant NPs

**Visualization**

1. Visualize the similarity matrix as a heatmap
2. Visualize the distance matrix as a heatmap



1. <https://en.wikipedia.org/wiki/Jaccard_index#Weighted_Jaccard_similarity_and_distance> [↑](#footnote-ref-1)