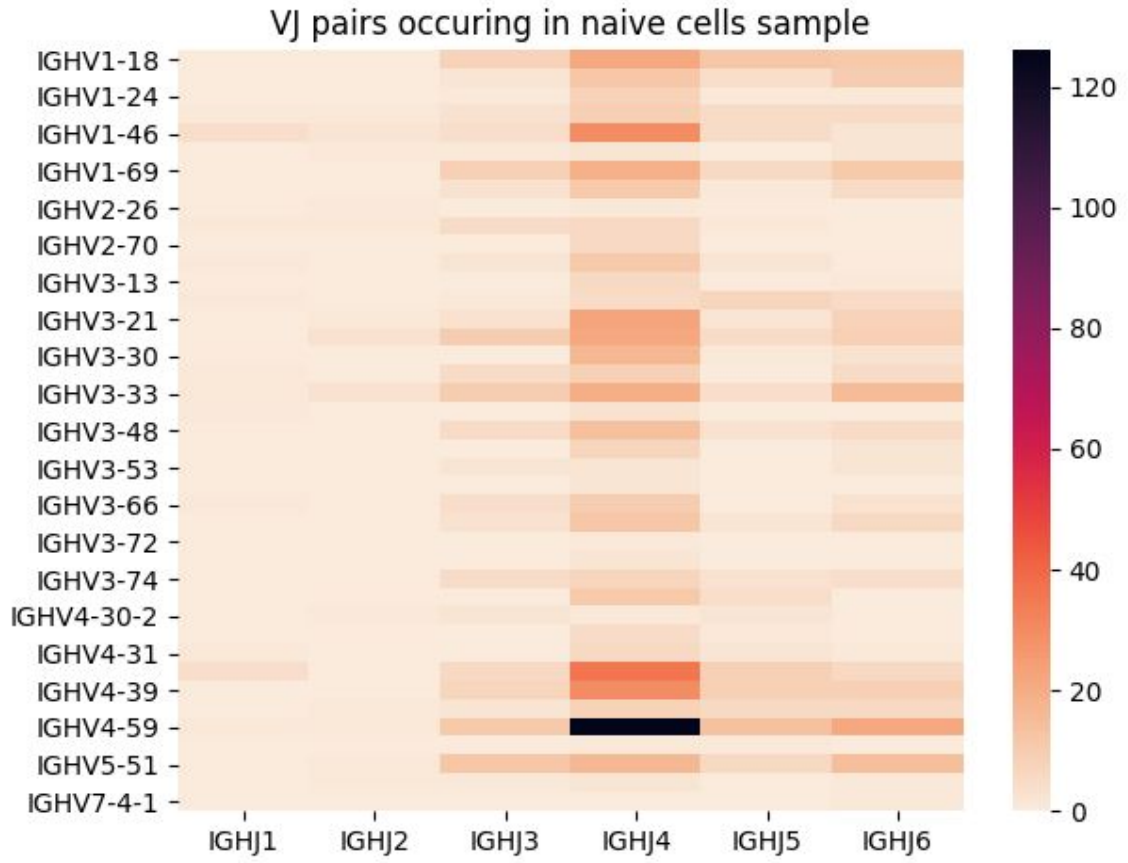
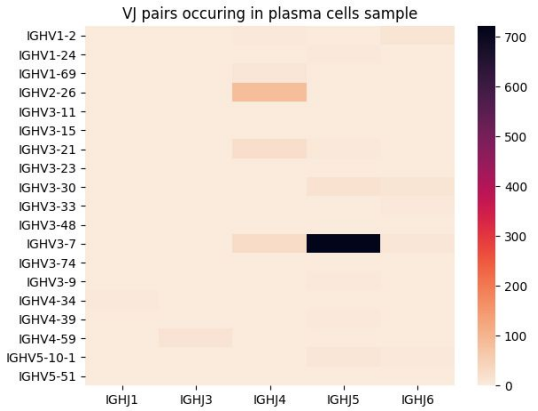
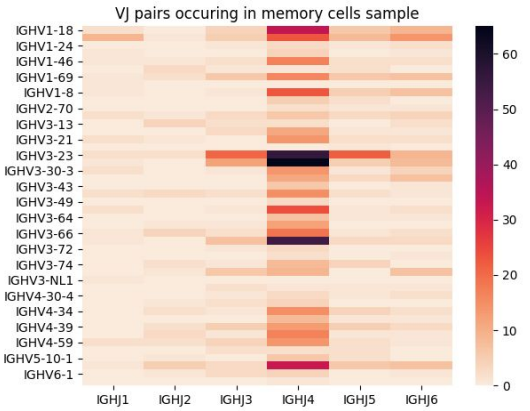
Homework1. Yakovleva Yulia
Code, results and data are available: https://github.com/Yulia-Yakovleva/immunogenomics
The report is available for comments at
https://docs.google.com/document/d/1ioVcOgVSQ94LWn33MX2Z0hTlqm7hcPH_YCkevO749Cc/edit?usp=sharing

The analysis was done with DiversityAnalyzer: https://immunotools.github.io/immunotools/

Task1. Heatmaps showing the number of sequences for each VJ pair.



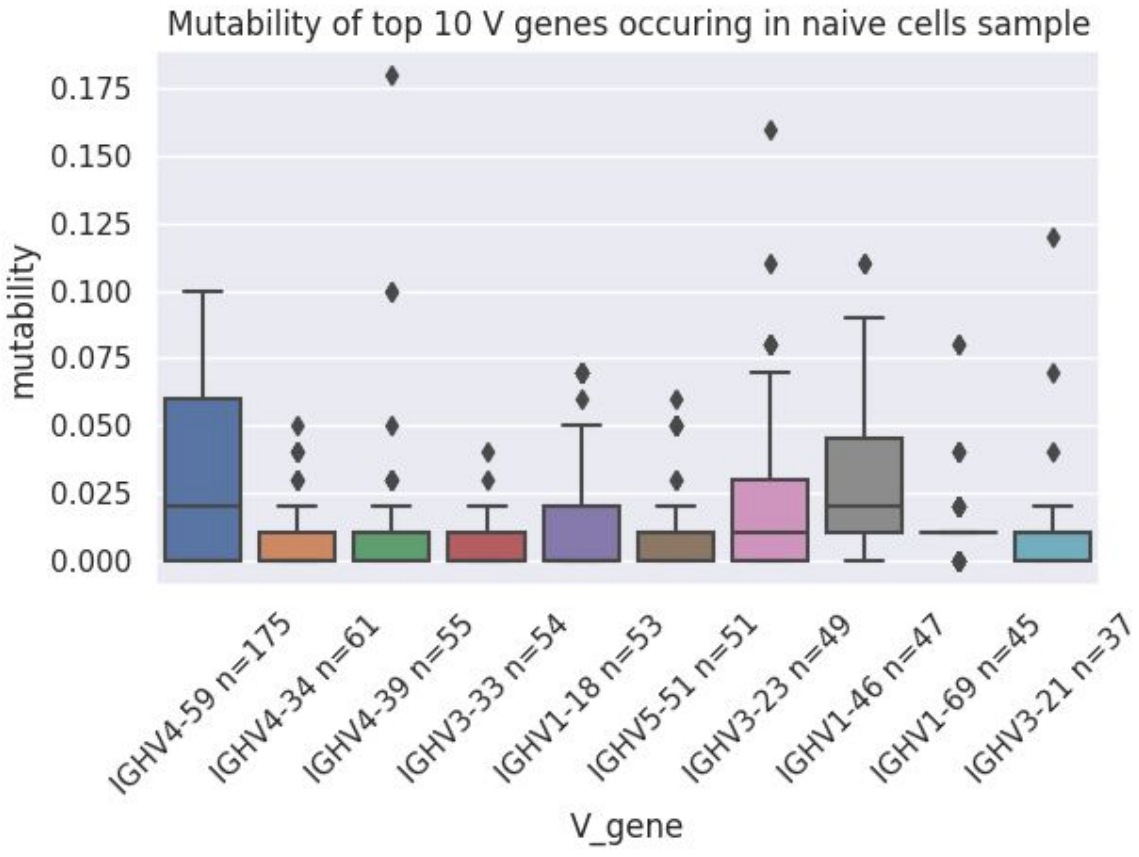


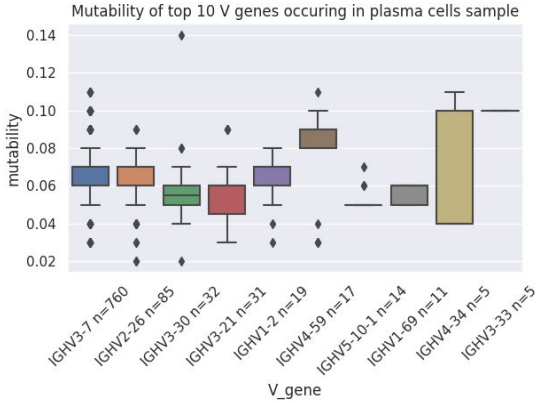


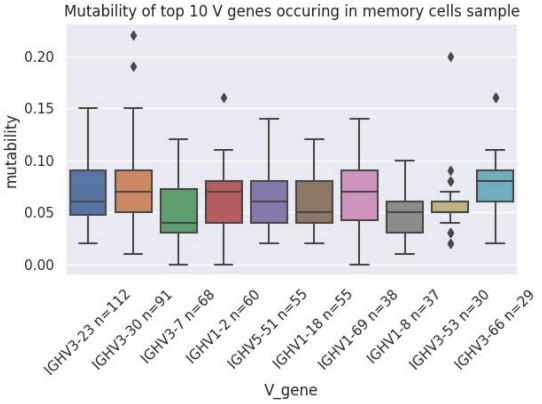
We have 148 unique pairs for naive sample. The common pair is IGHV7-4-1/IGHJ6 (occurred 126 times). We have 32 unique pairs for the plasma sample. The common pair is IGHV7-4-1/IGHJ6 (occurred 126 times), and we have 180 unique pairs for memory sample. The common pair is IGHV7-4-1/IGHJ4 (occurred 65 times). The sample with the smallest

number of VJ combinations is plasma. The sample with the highest number of VJ combinations is memory since this fraction of cells keeps immunological memory about previous infections.

Task2. Mutabilities for top-10 V genes

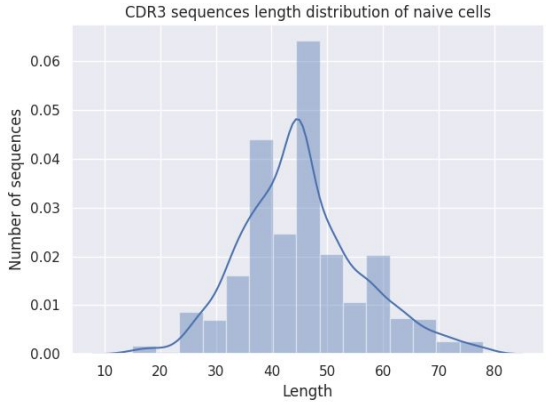


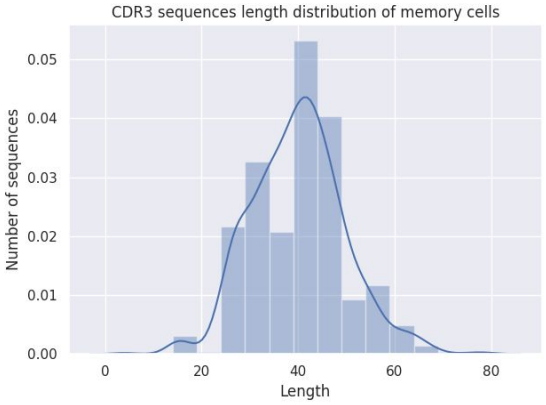


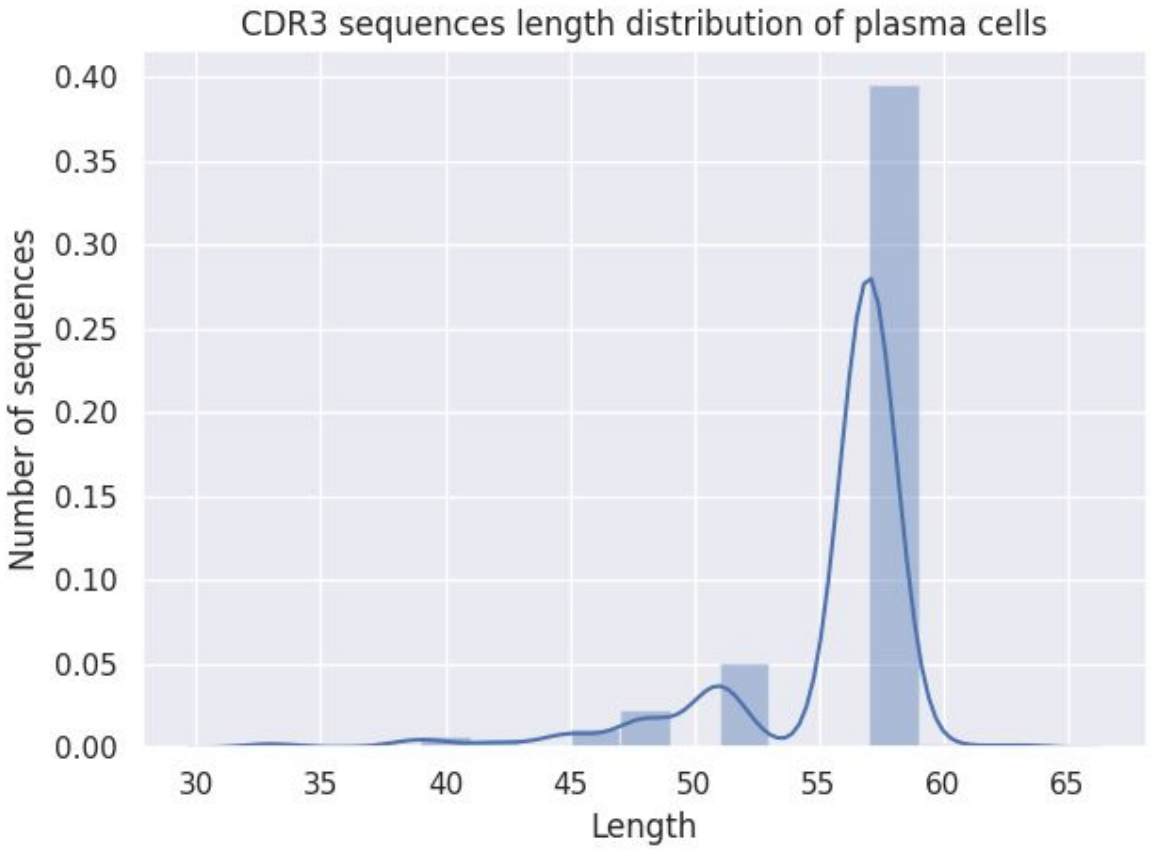


Since naive cells have not undergone the antigene dependent differentiation yet, this fraction has the lowest mutability. Since plasma cells and memory cells undergo the somatic hypermutagenesis of V genes during the differentiation, they have higher mutability, than naive cells.

Task3. Length distribution of CDR3 sequences

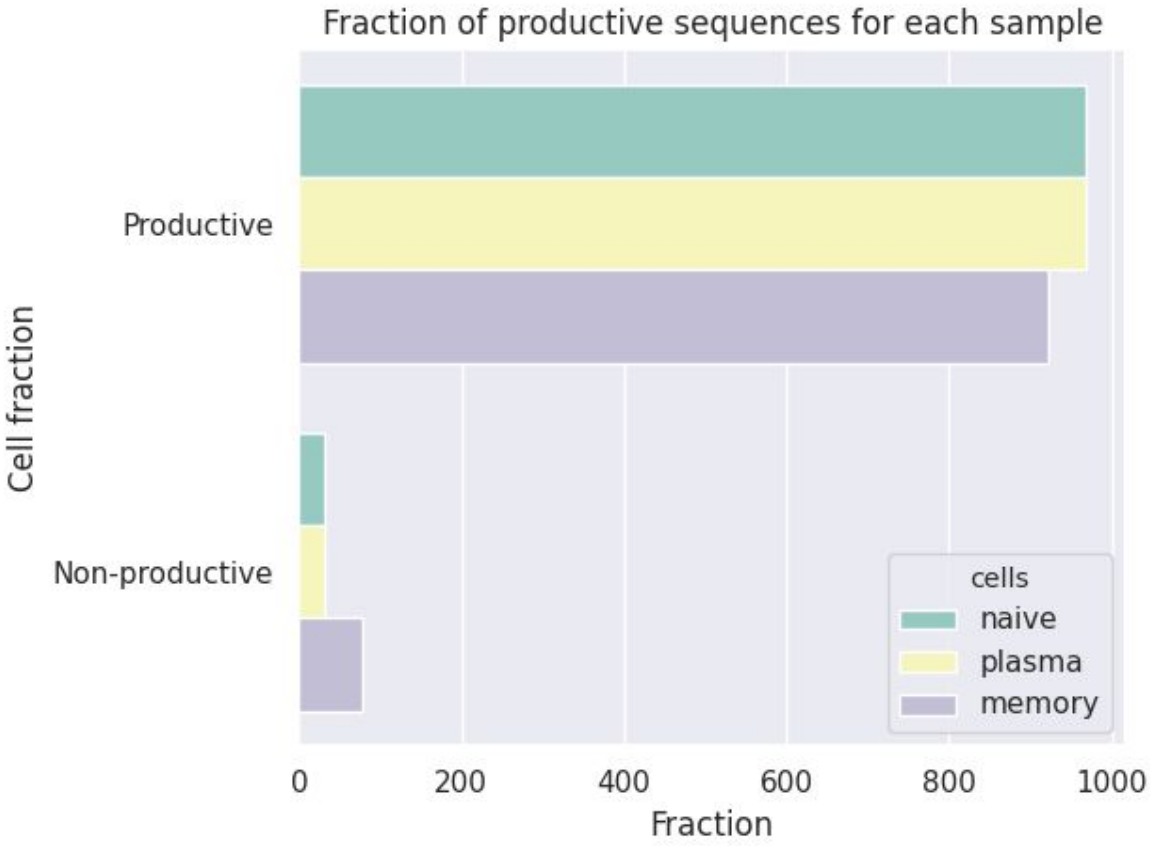






Length distributions of naive and memory cells are quite similar and look like normal distributions. The length distribution of plasma cells is biased, presumably, during the clonal expansion of plasma cells.

Task4. Non-productive fraction



The fraction of non-productive sequences: 3.1% for naive cells, 3.2% for plasma cells, 7.8% for memory cells. Memory cells have the highest non-productive fraction since these cells could undergo somatic hypermutagenesis in the lowest selective pressure. For naive and plasma cells productive fraction is approximately equal.