**Week 4 LCBQ**

This week, I improved my advanced size reduction algorithm. I fixed a few issues that were responsible for incoherent results. There are still some visualization issues that I’d like to fix with Lucile, especially concerning the colors. But the algorithm is now functional and is able to reduce the size of a tree to almost any wanted size. However, it cannot reduce it below a certain number of nodes, since it can’t delete the ten most abundant clonotypes and their ancestors. The default value for the number of clonotypes displayed by this version is 30.

To make sure that my work was actually relevant to the research project, I challenged it by comparing it to the existing solution of displaying only the 30 most abundant clonotypes. This solution infers a tree by taking into account only these 30 clonotypes. Thanks to the visualization and the metrics I developed, we understood that this solution was actually giving wrong information about the evolutional history of the clonotypes. It is therefore more accurate to infer a lineage tree based on a large number of clonotypes, and then reduce its size to make it easier to read. This work raised the question of the initial number of clonotypes to take into account for the inference of the lineage tree. After running a few tests and discussing this issue with the team, we decided that setting a number of reads threshold would set the initial number of clonotypes taken into account.

My challenge for this week is to implement my pruning solution in the analysis pipeline, with the help of the team.