Latest advancements in biology allow more precise RNA measurements (single cell) but bring new challenges for data analysis.

My project’s goal was to show a map of clusters of the genes using unsupervised learning methods. Including Explore known and novel clustering and matrix completion methods for SC and bulk sequencing data.

After no publications/tools found on the subject- I headed to build my own framework that will allow me to experiment on different preprocessing, imputations and clustering algorithms to find the right ones for this kind of task.

During the project I integrated different tools and techniques from publications in computational biology and in general unsupervised learning.

I found that clustering single cell gene expression is possible.

The first use of the clustering ability was on an MDS (Myelodysplastic syndrome) research in Soroka hospital- there the clustering can be used to identify changes between patients.