The dataset concerns an important problem of chronic myeloid leukemia identification. We are interested in differentiating the chronic myeloid leukemia from the normal control samples based on single-cell data analysis. The dataset contains 2151 cells and the expression value of 27118 genes. Phenotype of interest is the disease type of each cell. The population contains 187 normal cells and 1964 cells associated with chronic myeloid leukemia. For further information about the dataset, please refer to the paper below. A short python script “loading\_leukemia.ipynb” is provided to help with data loading.

Reference:

Giustacchini A, Thongjuea S, Barkas N, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. Nature Medicine. 2017 Jun;23(6):692-702. DOI: 10.1038/nm.4336. PMID: 28504724.