Dataset descriptions:

This dataset comes from a proof-of-concept study published in 1999 by Golub et al. It showed how new cases of cancer could be classified by gene expression monitoring (via DNA microarray) and thereby provided a general approach for identifying new cancer classes and assigning tumors to known classes. These data were used to classify patients with acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL).

In total, there are 72 patient samples. The cancer type information is provided in the CW2_PatientID_labels.csv file.

In the CW2_dataset, it contains 74 columns. The first two columns are Gene Description and Gene Accession Number respectively, and the remainder columns are individual patients (ID: 1 to 72) gene expression values.