Context of Data

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 tumours to known classes. These data were used to classify patients with acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL).

There are two datasets containing the initial (training, 38 samples) and independent (test, 34 samples) datasets used in the paper. These datasets contain measurements corresponding to ALL and AML samples from Bone Marrow and Peripheral Blood. Intensity values have been re-scaled such that overall intensities for each chip are equivalent.

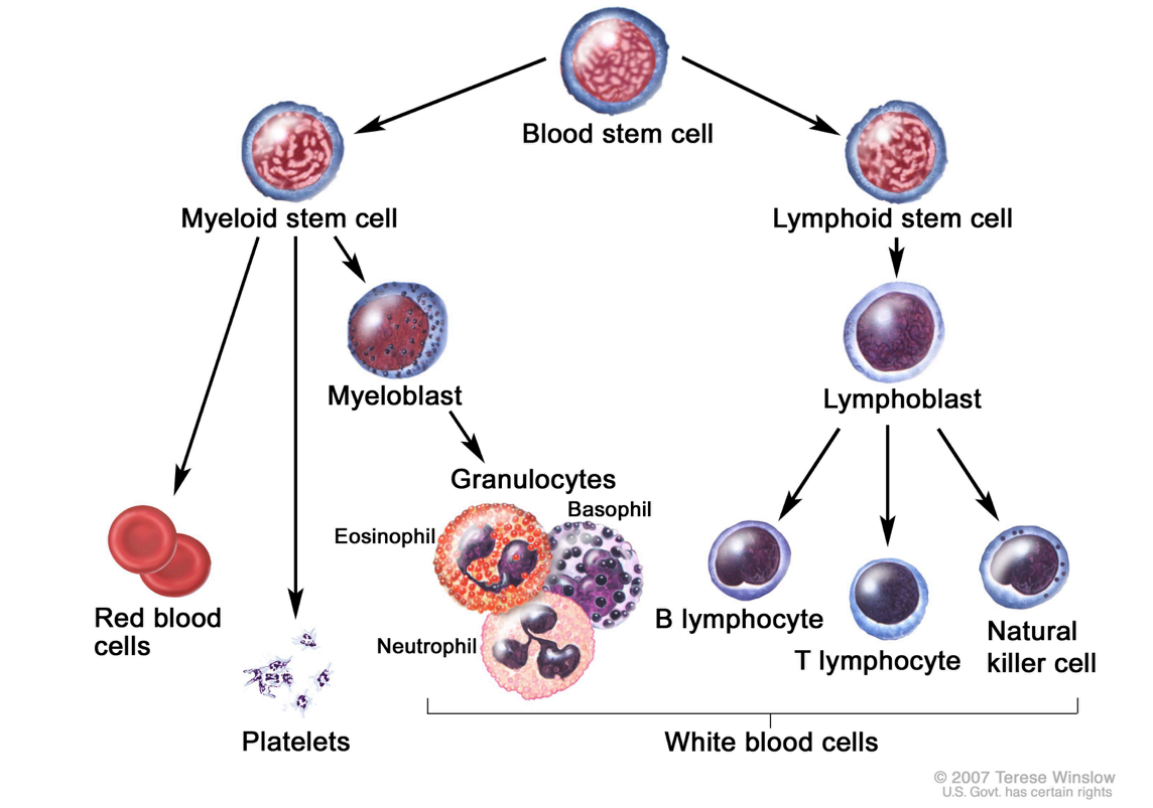
Leukemia

Leukemia is a type of cancer that affects the blood and bone marrow. It disrupts the normal function of healthy blood cells. In bone marrow, various types of blood cells are produced by blood stem cells called hematopoietic stem cells. Leukemia develops when hematopoietic stem cells differentiate into abnormal white blood cells, crossing out healthy blood cells and interfering with their functions. Leukemia patients may become susceptible to infection and have anemia or easy bleeding.

According to the statistics provided by the Global Cancer Observatory (GCO) in 2022, leukemia globally ranked 13th in terms of incidence rate among various 32 cancer types, while it occupied 10th place for mortality rate. About 500 thousand new incidence cases and 300 thousand new deaths cases were reported in 2022.

Hematopoietic Stem Cells

Hematopoietic stem cells differentiate into various types of blood cells. Initially, it differentiates into either lymphoid stem cell or myeloid stem cell. A lymphoid stem cell further specializes into lymphoblast. Lymphoblast then becomes one of three types of lymphocytes including B-cells, T-cells, and natural killer (NK) cells. A myeloid stem cell differentiates into different types of blood cells containing red blood cells, platelets and granulocytes. White blood cells have two lineages, either from lymphoid stem cell or myeloid stem cell. B cells, T cells, and natural killer (NK) cells are lymphoid white blood cells, while granulocytes are myeloid white blood cell.



Types of Leukemia

Leukemia is broadly classified into four main types depending on the type of white blood cells affected and the rate at which the disease progresses. Based on whether leukemia affects lymphoid or myeloid white blood cells, it is classified as lymphoblastic leukemia or myeloid leukemia. Also, depending on the proliferation speed, it is classified as ‘acute’ if it is rapid, or ‘chronic’ if it is slow growing. Those four types of leukemia are Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL), and Chronic Myeloid Leukemia (CLL). Our data is focused on the ALL and AML patients.

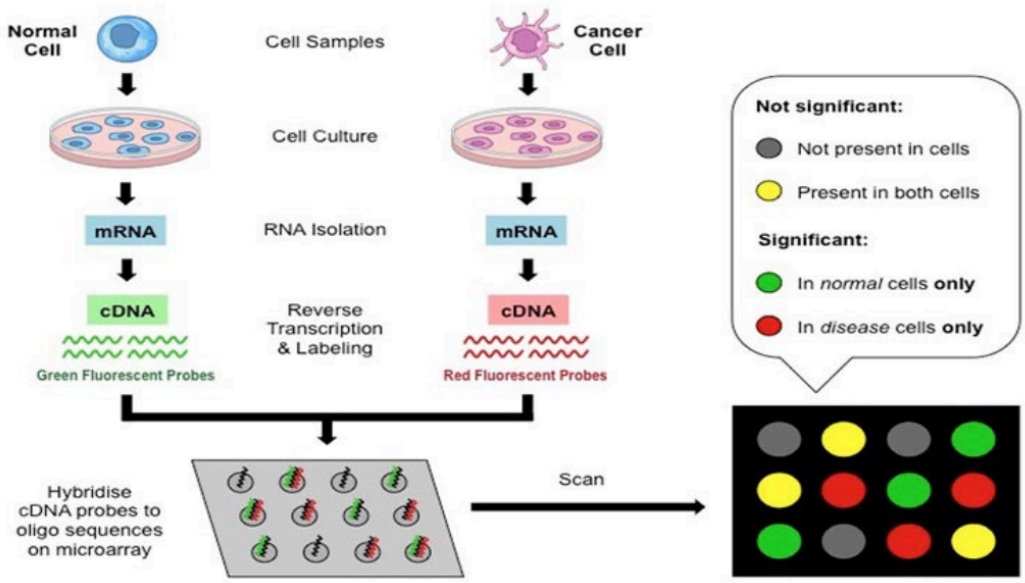
Importance of distinction between ALL and AML

Compared to chronic leukemia, acute leukemias need earlier and faster diagnosis as they progress more rapidly and aggressively. The appearance and symptoms of ALL and AML are highly similar. However, the cure rate is remarkably diminished when ALL therapy is used for AML. Thus, accurate distinction of acute leukemias is crucial for successful treatment. Since the mere inspection of the appearance of acute leukemias for classification has significant limitations, more systematic approach of gene expression monitoring is used.

Gene Expression Monitoring and DNA Microarray

DNA microarray is a powerful tool to analyze the expression levels of thousands of genes within a biological sample. This technology relies on the principle of complementary base pairing to detect and quantify nucleic acid sequences. A DNA chip consists of thousands of tiny spots on which DNA probes are fixed. A probe is a short single-stranded DNA sequence complementary to a certain gene. Each spot is coated with multiple identical DNA probes

DNA microarray mechanism and methods



1. *Sample preparation*

- RNA sample preparation - Initially, two RNA samples are prepared. One sample is a control RNA sample obtained from healthy tissue, and the other is a RNA sample obtained from the patient's tumor tissue.
- cDNA Transcription and fluorescent labeling - Next, the RNA is transcribed to cDNA using reverse transcription. During this step, the cDNA is fluorescently labeled. The cDNA from the healthy tissue is labeled in green, and the cDNA from the patient is labeled in red.

2. *Loading samples to microarrays*

Then, the cDNA samples are applied to each spot, where fluorescently labeled cDNA molecules then bind to their complementary DNA probes on the array. In each spot, one of four scenarios can occur. If neither control nor patient samples have complementary cDNA molecules to the DNA probes on the array, the spot will not display any color. Alternatively, if only the control sample contains cDNA molecules complementary to the probes, the spot will appear green. Similarly, the spot will show red if only the patient sample contains cDNA molecules

complementary to the probes. Finally, if both control and patient samples have complementary cDNA molecules to the probes, the spot will display yellow.

3. *Fluorescence scanning and data extraction*

After hybridization, each spot of the DNA chip is scanned using a fluorescence scanner to detect and transform the intensity of the colors into numeric values. The scanner fluorescence intensity at each spot corresponds to the expression level of the corresponding gene in the sample.

Reference for images_

Adult Acute Lymphoblastic Leukemia treatment. National Cancer Institute. (n.d.).

https://www.cancer.gov/types/leukemia/patient/adult-all-treatment-pdq

Karki, G. (2020, June 6). DNA microarray: Principle, types and steps involved in cdna microarrays.

Online Biology Notes.

https://www.onlinebiologynotes.com/dna-microarray-principle-types-and-steps-involved-in-cdna-microarrays/