**Project Checkpoint 1**

**Migraine**

**Team Members:** Suraj Khamkar **|** Rahul More **|** Pragathi Pendem **|** Sravani Pati

* **Unit of analysis**

Blood samples

We wanted to examine how the transcriptome of peripheral blood mononuclear cells (PBMC) and variations in the levels of plasma metabolites in migraineurs changed both during and between attacks.

A migraine is more than just a painful headache. In contrast to other headache illnesses, it is a disabling neurological condition with unique symptoms and therapeutic methods.

**RNA** - It carries the messenger instructions from DNA, which itself contains the genetic instructions.

PBMC - Human Peripheral blood mononuclear cells (PBMCs) are immune cells which involves cell-mediated immunity.

* **Dataset:**
  + Data set consists of **24 patients**.

Patients with migraines were given blood samples both before and after an incident. If the patient had not experienced a headache for at least 24 hours, the attack-free (interictal) sample was taken. Affected patients were urged to wait to begin their normal attack treatment until the blood was obtained for ictal samples.

* + Samples

The **samples are collected from 13 age and sex matched healthy volunteers**. Here RNA was isolated from PBMCs. The cubital veins of migraine sufferers and healthy volunteers were used to collect human blood (13 mL/person), which was then placed into ice-cold glass tubes containing either citrate or ethylenediaminetetraacetic acid (EDTA).

* + Transcriptome profile of PBMC Samples

163 genes were discovered to be differentially expressed in interictal PBMC samples when compared to healthy ones, with a fold change threshold of 1.5 and a p-value threshold of 0.05; 135 genes were upregulated and 28 were downregulated. 64 genes were upregulated and 80 were downregulated in ictal PBMC samples as compared to interictal ones (fold change: 1.3, p-value: 0.05).

* **Data Cleaning:**

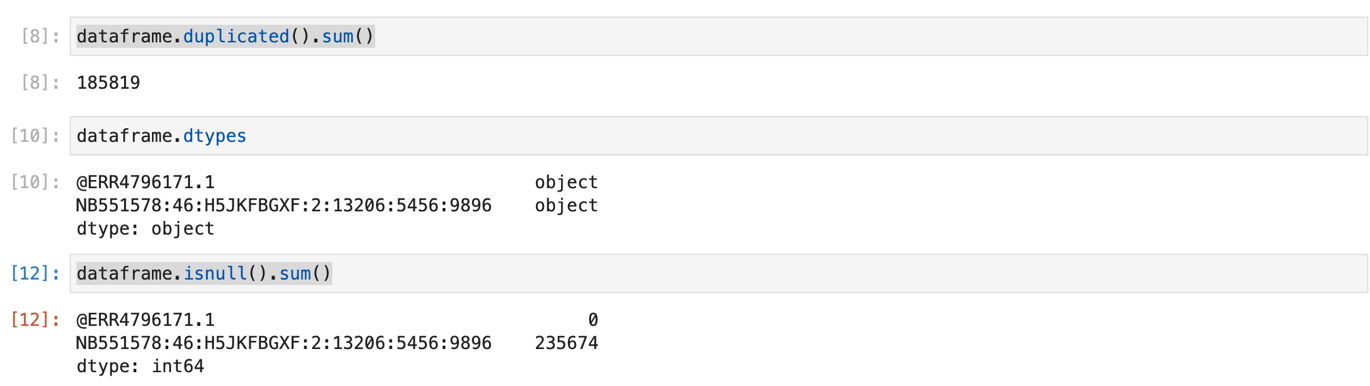
First, we downloaded dataset from [this link](https://www.ebi.ac.uk/ena/browser/view/PRJEB40032?show=reads). The datasets we have chosen are ERR4796171.fastq and ERR4796172.fastq.

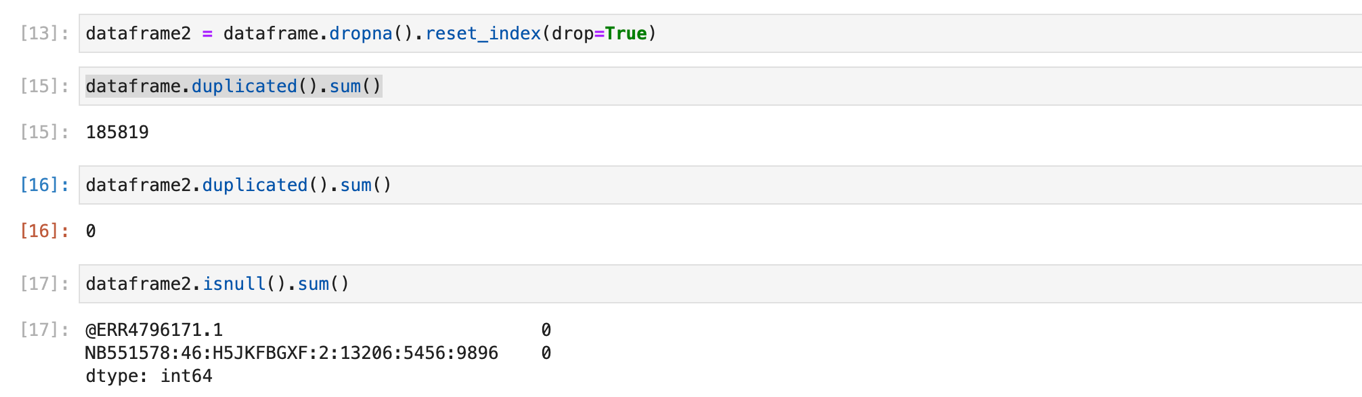
To get a greater number of records we have merged dataset in file named as merged\_ERR479617a.fastq.



This merged file had 314232 records in it before data cleaning consists of **duplicate** and **null** values.

Secondly, we checked the duplicate values and created another dataset after deleting null values from the same.





* **Data Visualization:**

For visualizing of our data, consisting of Sequence column we have used heatmap and histogram to show frequency of which will help to understand presence of migraine based on particular unique identifier followed by @ERR

