

# Leveraging the use of publicly available data for Cancer Research: an example in Chronic Lymphocytic Leukemia

Maria Tsagiopoulou

Institute of Applied Biosciences, Centre for Research and Technology Hellas (CERTH), Greece



**Big Data in Biology (FAIR is key)**



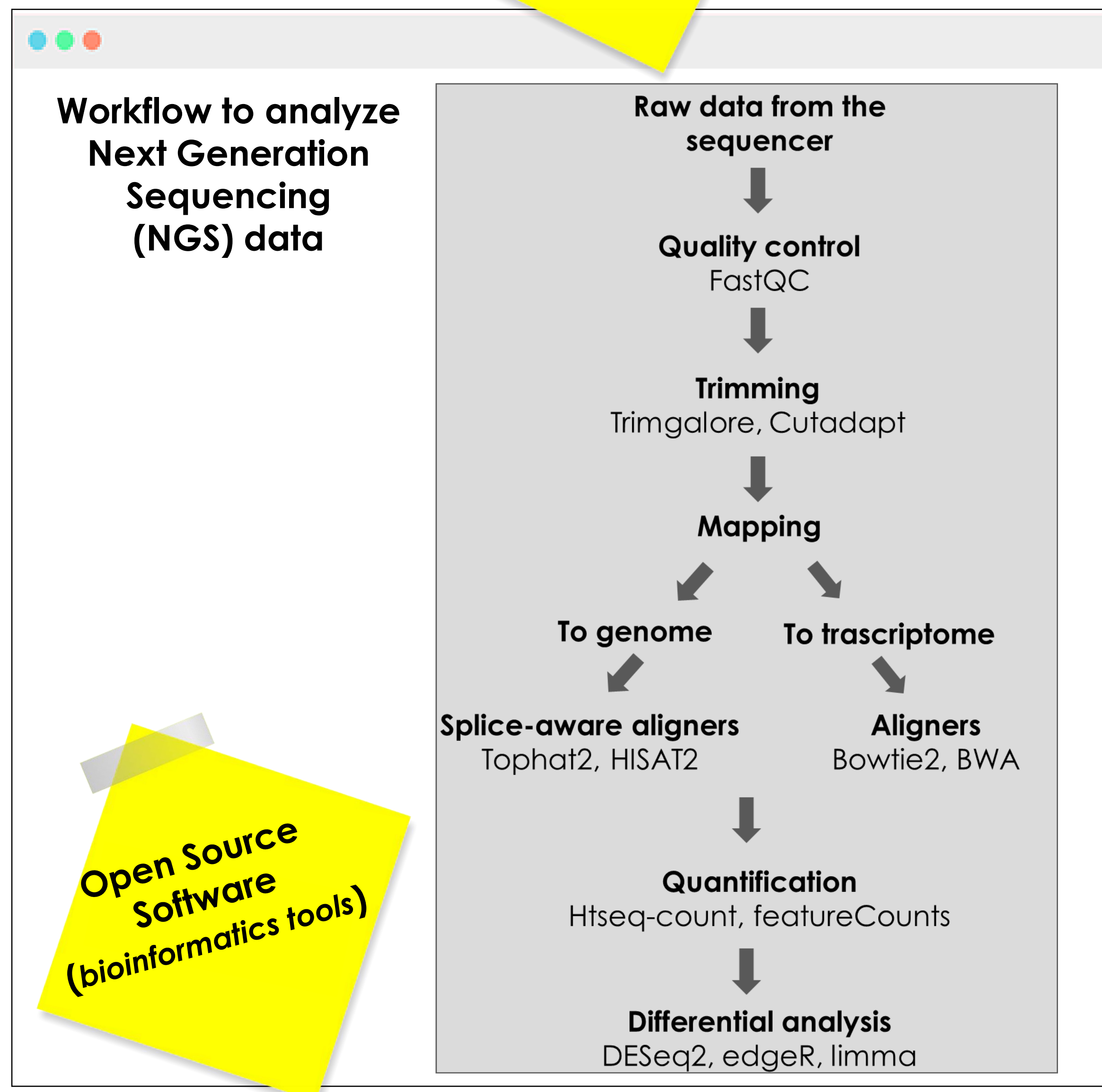
**Typing..**

Explore the public cancer genomics resources

Discover new trends and validate your findings with 1500+ datasets and 50+ cancer types on:

The Cancer Genome Atlas (TCGA), International Cancer Genome Consortium (ICGC), BLUEPRINT consortium, UCSC RNA-seq compendium.

|                         | Samples | Data Types   |
|-------------------------|---------|--|
| TCGA                    | 12,470  | copy number, gene-, exon-, miRNA-, and protein-expression, somatic mutation, DNA methylation, survival, and clinical data                      |
| Pan-Cancer Atlas        | 12,591  | copy number, gene-, miRNA-, and protein-expression, somatic mutation, DNA methylation, molecular subtypes, multiple curated survival endpoints |
| ICGC                    | 17,697  | copy number, gene expression, somatic mutation   |
| UCSC RNA-seq compendium | 19,131  | TCGA, TARGET, and GTEx gene- and transcript-expression   |
| TCGA ATAC-seq           | 404     | ATAC-seq peak signal   |



**Open Source Software (bioinformatics tools)**

ICGC Data Portal

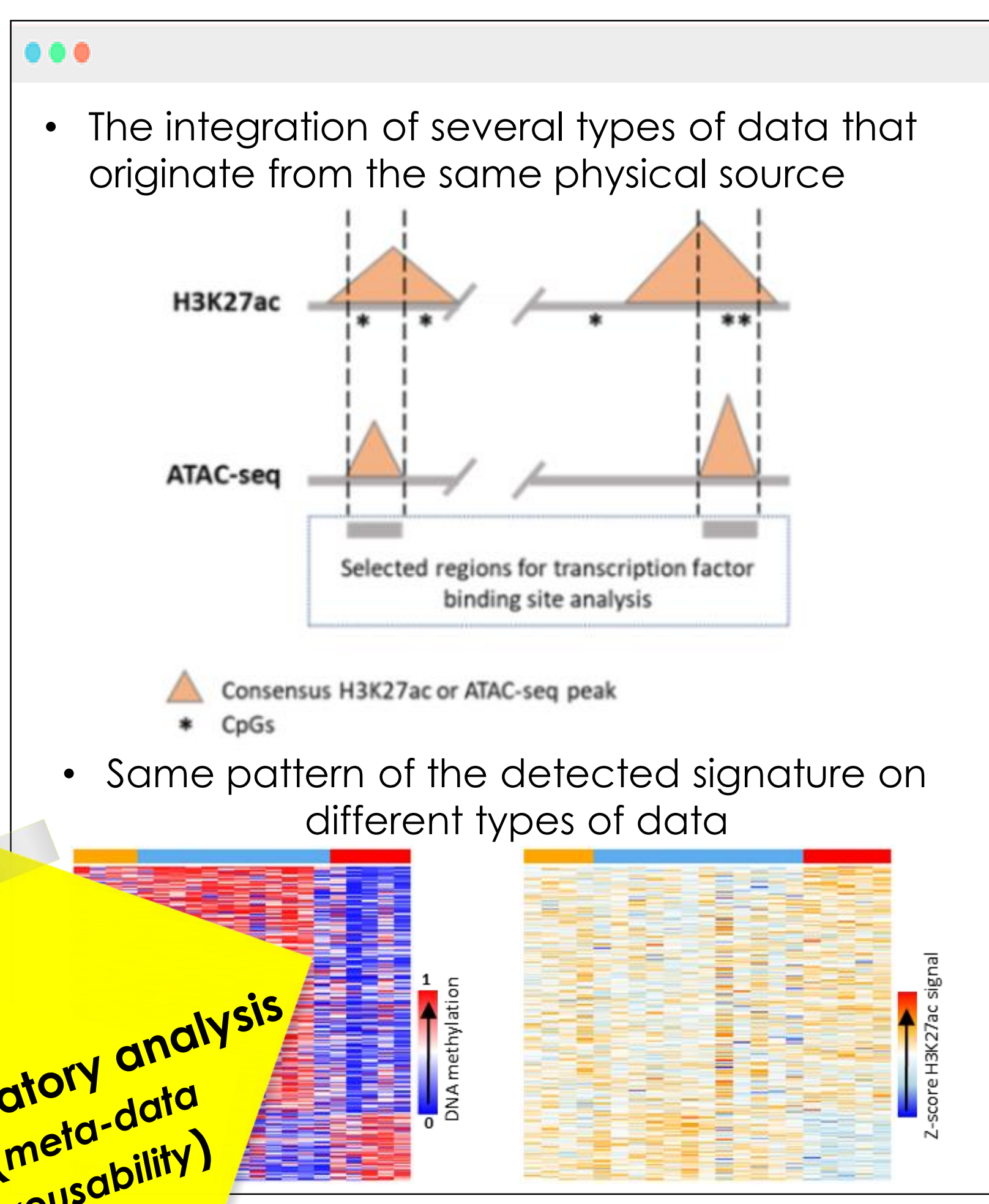
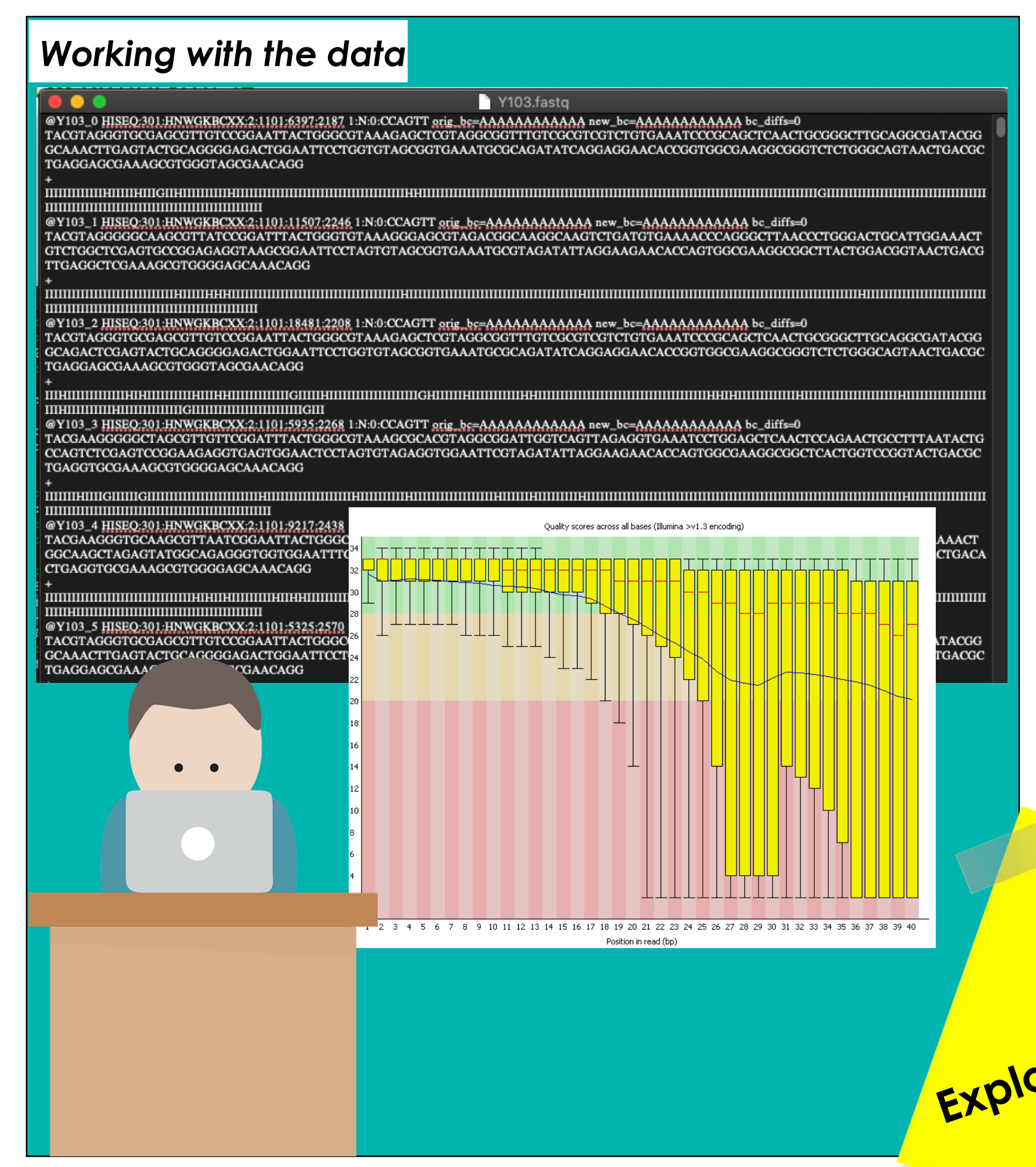
**CHRONIC LYMPHOCYTIC LEUKEMIA - ES**

Summary Donors Mutations

**Summary**

|   |                                     |
|---|-------------------------------------|
| Code  | CLLE-ES                             |
| Name  | Chronic Lymphocytic Leukemia - ES   |
| Primary Site                                | Blood                               |
| Tumour Type                                 | Chronic Lymphocytic Leukemia        |
| Tumour Subtype                              | CLL with mutated and unmutated IgVH |
| Countries                                   | Spain                               |
| Number of donors in PCAWG                   | 100                                 |
| Number of donors with molecular data in DCC | 551                                 |
| Total number of donors                      | 551                                 |

**Accessible research data through consortia**



**Exploratory analysis (meta-data reusability)**

Jaspar database which is an open-access database of curated, non-redundant transcription factor binding profiles

AGJASPARAC

TFCP2 (P-value=2.60e-4)

TFCP2 expression

- Hidden associations between the input matrices.

**Open Access DBs (connection to biological knowledge)**

Altogether, there are currently several open-access resources for research data but also several community-supported bioinformatics tools and databases that can manipulate the data. This field is promising since it can directly benefit by adhering to the open science principles, in order to produce high quality and reproducible scientific results.