# Exploring the Depths of Bioinformatics: Unraveling the Process

**Elif Hangül**

## Why Bioinformatics?

Bioinformatics, at its core, is a multidisciplinary **field that combines biology, computer science, and statistics to extract meaningful insights from biological data.** It involves the application of computational techniques and algorithms to analyze and interpret vast datasets derived from biological research.
In the realm of molecular biology and genetics, where the complexity of data has surged in recent years, bioinformatics plays a pivotal role. It facilitates the organization, analysis, and interpretation of biological information, ranging from DNA sequences to protein structures. This integration of computational tools into biological research has revolutionized our ability to uncover hidden patterns, make predictions, and derive valuable knowledge from the vast ocean of biological data.

The importance of bioinformatics lies in its capacity to accelerate research and enhance our understanding of complex biological processes. It aids in **identifying genes associated with diseases, predicting protein structures and functions, and even designing novel drugs.** Bioinformatics tools enable researchers to navigate through the intricacies of genomics, transcriptomics, and proteomics, making it an indispensable ally in advancing scientific discoveries and medical breakthroughs.

### How Did I Start

Embarking on a journey into bioinformatics, one of the initial and essential steps is gaining insights into how leading companies leverage bioinformatics in their operations. This preliminary understanding serves as a foundation for several key reasons:

**Real-world Application:** Companies at the forefront of bioinformatics apply these tools and methodologies to solve practical challenges in the field of genetics, genomics, and molecular biology. By examining their practices, we can gain exposure to the real-world applications of bioinformatics.

**Technological Landscape:** Bioinformatics tools and technologies are continually evolving. By observing how companies incorporate the latest advancements, we can stay abreast of the dynamic technological landscape. This awareness is crucial for any one of us aspiring to contribute meaningfully to the field, as it helps in anticipating trends and staying adaptable to emerging technologies.

**Networking and Collaboration Opportunities:** Understanding how companies integrate bioinformatics opens doors to potential networking and collaboration opportunities. Familiarity with industry practices allows us to engage in more informed discussions with professionals in the field, enhancing our ability to collaborate effectively and contribute meaningfully to bioinformatics projects.

Selecting a suitable company may initially appear challenging due to the multitude of options available. In my case, I opted for **Sangamo Therapeutics**, drawn to their innovative applications like CAR-T cell therapy and the utilization of zinc proteins. **It's essential to recognize that the ideal choice varies for each individual.** By researching companies aligned with your interests and offering potential job opportunities, you can identify the one that best resonates with your career aspirations.

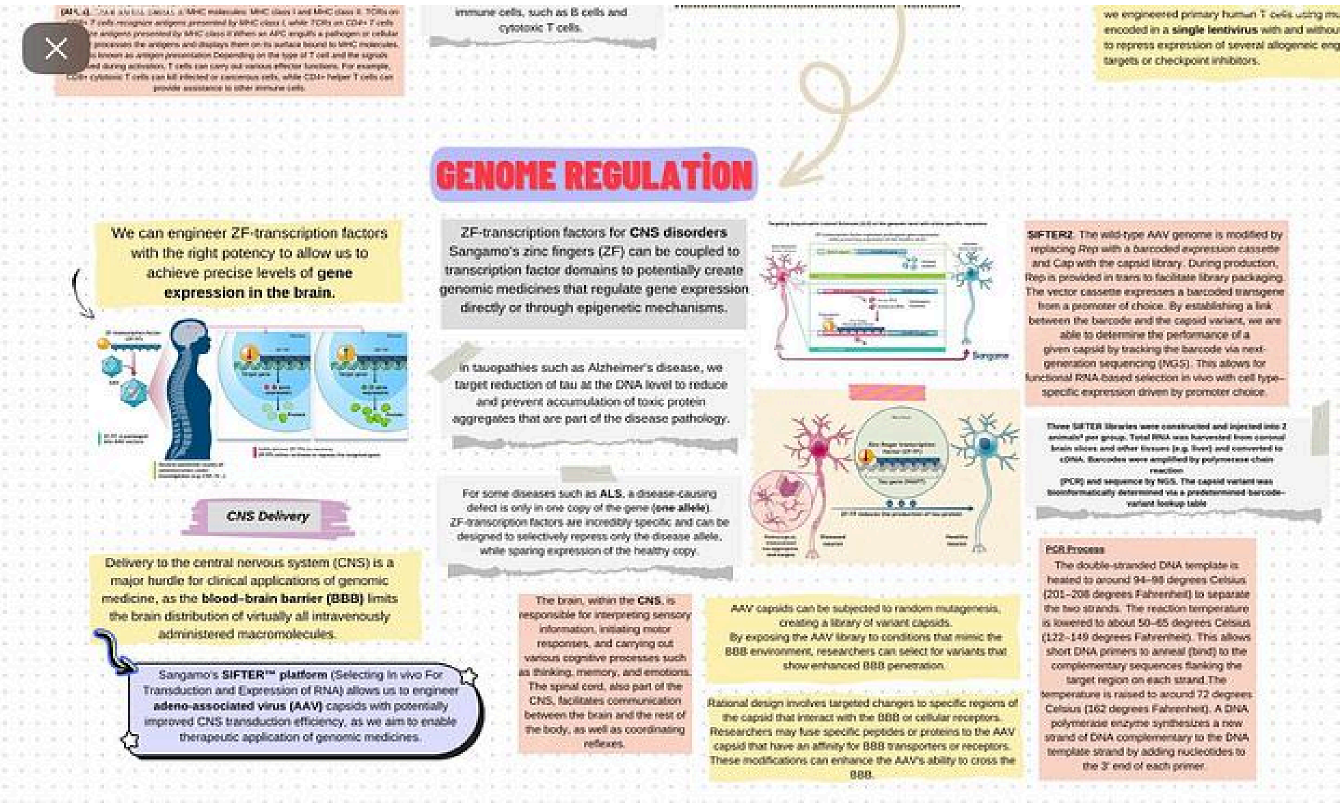
## Sangamo Therapeutics

Sangamo THERAPEUTICS

Sangamo Therapeutics, a prominent player in the biotechnology sector, is at the forefront of advancing genomic medicine and gene editing technologies. The company has garnered attention for its innovative approaches, particularly in the application of zinc finger proteins for precise and targeted gene editing. This novel technology involves *engineering zinc finger nucleases* (ZFNs) to specifically target desired DNA sequences, enabling meticulous modifications at the genetic level.

One notable area of Sangamo's work lies in the realm of **CAR-T cell therapy**, an emerging and promising field in the treatment of various diseases, notably certain forms of cancer. In the context of CAR-T cell therapy, Sangamo likely employs bioinformatics to analyze extensive genomic data. These analyses aid in optimizing the design of chimeric antigen receptor (CAR) T cells by identifying target genes, understanding gene expression patterns, and enhancing the therapeutic efficacy of these innovative therapies.

Beyond the laboratory, bioinformatics serves as an essential tool for Sangamo in the realm of **data analysis and interpretation**. Large-scale genomic data generated through various research initiatives require sophisticated computational approaches to extract meaningful insights. *Sangamo likely leverages bioinformatics tools to navigate through this vast genomic landscape, identifying potential therapeutic targets, understanding the functional consequences of genetic modifications, and predicting the safety and efficacy of their interventions.*



(APCs). T cell antigen ... MHC molecules: MHC class I and MHC class II. TCRs on CD8+ T cells recognize antigens presented by MHC class I, while TCRs on CD4+ T cells recognize antigens presented by MHC class II.When an APC engulfs a pathogen or cellular ... processes the antigens and displays them on its surface bound to MHC molecules. ...s known as antigen presentation Depending on the type of T cell and the signals ...ved during activation, T cells can carry out various effector functions. For example, CD8+ cytotoxic T cells can kill infected or cancerous cells, while CD4+ helper T cells can provide assistance to other immune cells.

immune cells, such as B cells and cytotoxic T cells.

we engineered primary human T cells using mu... encoded in a **single lentivirus** with and withou... to repress expression of several allogeneic eng... targets or checkpoint inhibitors.

# GENOME REGULATİON

We can engineer ZF-transcription factors with the right potency to allow us to achieve precise levels of **gene expression in the brain.**

ZF-transcription factors for **CNS disorders** Sangamo's zinc fingers (ZF) can be coupled to transcription factor domains to potentially create genomic medicines that regulate gene expression directly or through epigenetic mechanisms.

In tauopathies such as Alzheimer's disease, we target reduction of tau at the DNA level to reduce and prevent accumulation of toxic protein aggregates that are part of the disease pathology.

For some diseases such as **ALS**, a disease-causing defect is only in one copy of the gene (**one allele**). ZF-transcription factors are incredibly specific and can be designed to selectively repress only the disease allele, while sparing expression of the healthy copy.

**SIFTER2**. The wild-type AAV genome is modified by replacing *Rep* with a barcoded expression cassette and Cap with the capsid library. During production, Rep is provided in trans to facilitate library packaging. The vector cassette expresses a barcoded transgene from a promoter of choice. By establishing a link between the barcode and the capsid variant, we are able to determine the performance of a given capsid by tracking the barcode via next-generation sequencing (NGS). This allows for functional RNA-based selection in vivo with cell type–specific expression driven by promoter choice.

**Three SIFTER libraries were constructed and injected into 2 animals* per group. Total RNA was harvested from coronal brain slices and other tissues (e.g. liver) and converted to cDNA. Barcodes were amplified by polymerase chain reaction (PCR) and sequence by NGS. The capsid variant was bioinformatically determined via a predetermined barcode-variant lookup table**

*CNS Delivery*

Delivery to the central nervous system (CNS) is a major hurdle for clinical applications of genomic medicine, as the **blood–brain barrier (BBB)** limits the brain distribution of virtually all intravenously administered macromolecules.

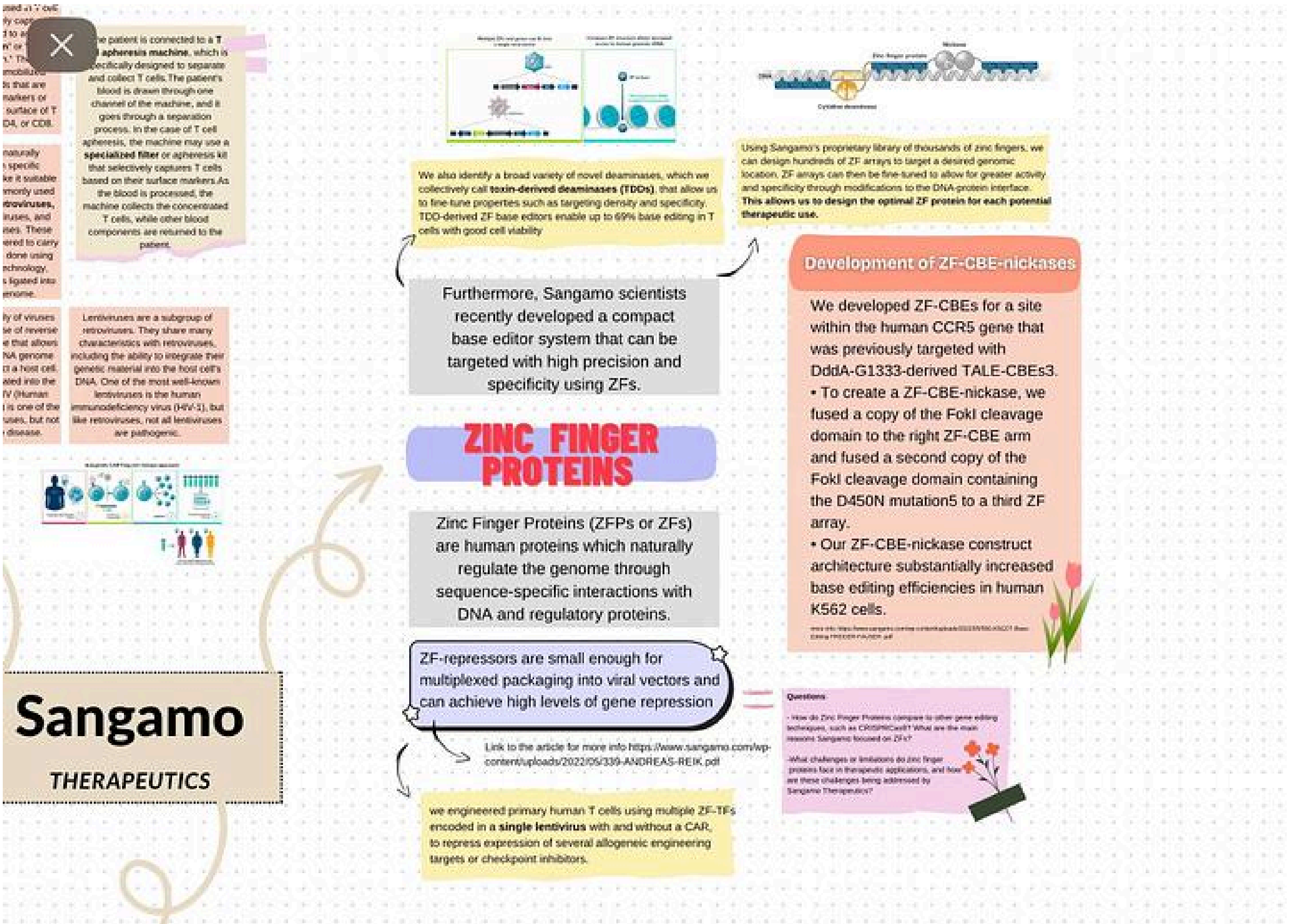
Sangamo's **SIFTER™ platform** (Selecting In vivo For Transduction and Expression of RNA) allows us to engineer **adeno-associated virus (AAV)** capsids with potentially improved CNS transduction efficiency, as we aim to enable therapeutic application of genomic medicines.

The brain, within the **CNS**, is responsible for interpreting sensory information, initiating motor responses, and carrying out various cognitive processes such as thinking, memory, and emotions. The spinal cord, also part of the CNS, facilitates communication between the brain and the rest of the body, as well as coordinating reflexes.

AAV capsids can be subjected to random mutagenesis, creating a library of variant capsids.
By exposing the AAV library to conditions that mimic the BBB environment, researchers can select for variants that show enhanced BBB penetration.

Rational design involves targeted changes to specific regions of the capsid that interact with the BBB or cellular receptors. Researchers may fuse specific peptides or proteins to the AAV capsid that have an affinity for BBB transporters or receptors. These modifications can enhance the AAV's ability to cross the BBB.

PCR Process
The double-stranded DNA template is heated to around 94–98 degrees Celsius (201–208 degrees Fahrenheit) to separate the two strands. The reaction temperature is lowered to about 50–65 degrees Celsius (122–149 degrees Fahrenheit). This allows short DNA primers to anneal (bind) to the complementary sequences flanking the target region on each strand.The temperature is raised to around 72 degrees Celsius (162 degrees Fahrenheit). A DNA polymerase enzyme synthesizes a new strand of DNA complementary to the DNA template strand by adding nucleotides to the 3' end of each primer.



...sed in T cell ...ly cap... ...d to a... ...w" or ...." Th... ...mobiliz... ...s that are ...markers or ...surface of T ...D4, or CD8.

...naturally ...specific ...ke it suitable ...monly used ...troviruses, ...iruses, and ...ses. These ...ered to carry ...done using ...chnology, ...s ligated into ...enome.

...ly of viruses ...se of reverse ...e that allows ...NA genome ...o a host cell. ...ated into the ...V (Human ...is one of the ...uses, but not ...disease.

...e patient is connected to a **T ...apheresis machine**, which is ...ecifically designed to separate and collect T cells.The patient's blood is drawn through one channel of the machine, and it goes through a separation process. In the case of T cell apheresis, the machine may use a **specialized filter** or apheresis kit that selectively captures T cells based on their surface markers.As the blood is processed, the machine collects the concentrated T cells, while other blood components are returned to the patient.

Lentiviruses are a subgroup of retroviruses. They share many characteristics with retroviruses, including the ability to integrate their genetic material into the host cell's DNA. One of the most well-known lentiviruses is the human immunodeficiency virus (HIV-1), but like retroviruses, not all lentiviruses are pathogenic.

We also identify a broad variety of novel deaminases, which we collectively call **toxin-derived deaminases (TDDs)**, that allow us to fine-tune properties such as targeting density and specificity. TDD-derived ZF base editors enable up to 69% base editing in T cells with good cell viability

Using Sangamo's proprietary library of thousands of zinc fingers, we can design hundreds of ZF arrays to target a desired genomic location. ZF arrays can then be fine-tuned to allow for greater activity and specificity through modifications to the DNA-protein interface. **This allows us to design the optimal ZF protein for each potential therapeutic use.**

Furthermore, Sangamo scientists recently developed a compact base editor system that can be targeted with high precision and specificity using ZFs.

## Development of ZF-CBE-nickases

We developed ZF-CBEs for a site within the human CCR5 gene that was previously targeted with DddA-G1333-derived TALE-CBEs3.
- To create a ZF-CBE-nickase, we fused a copy of the FokI cleavage domain to the right ZF-CBE arm and fused a second copy of the FokI cleavage domain containing the D450N mutation5 to a third ZF array.
- Our ZF-CBE-nickase construct architecture substantially increased base editing efficiencies in human K562 cells.

more info: https://www.sangamo.com/wp-content/uploads/2022/05/590-KNOTT-Base-Editing-FREDER-FAVOR.pdf

# ZINC FINGER PROTEINS

Zinc Finger Proteins (ZFPs or ZFs) are human proteins which naturally regulate the genome through sequence-specific interactions with DNA and regulatory proteins.

ZF-repressors are small enough for multiplexed packaging into viral vectors and can achieve high levels of gene repression

Link to the article for more info https://www.sangamo.com/wp-content/uploads/2022/05/339-ANDREAS-REIK.pdf

we engineered primary human T cells using multiple ZF-TFs encoded in a **single lentivirus** with and without a CAR, to repress expression of several allogeneic engineering targets or checkpoint inhibitors.

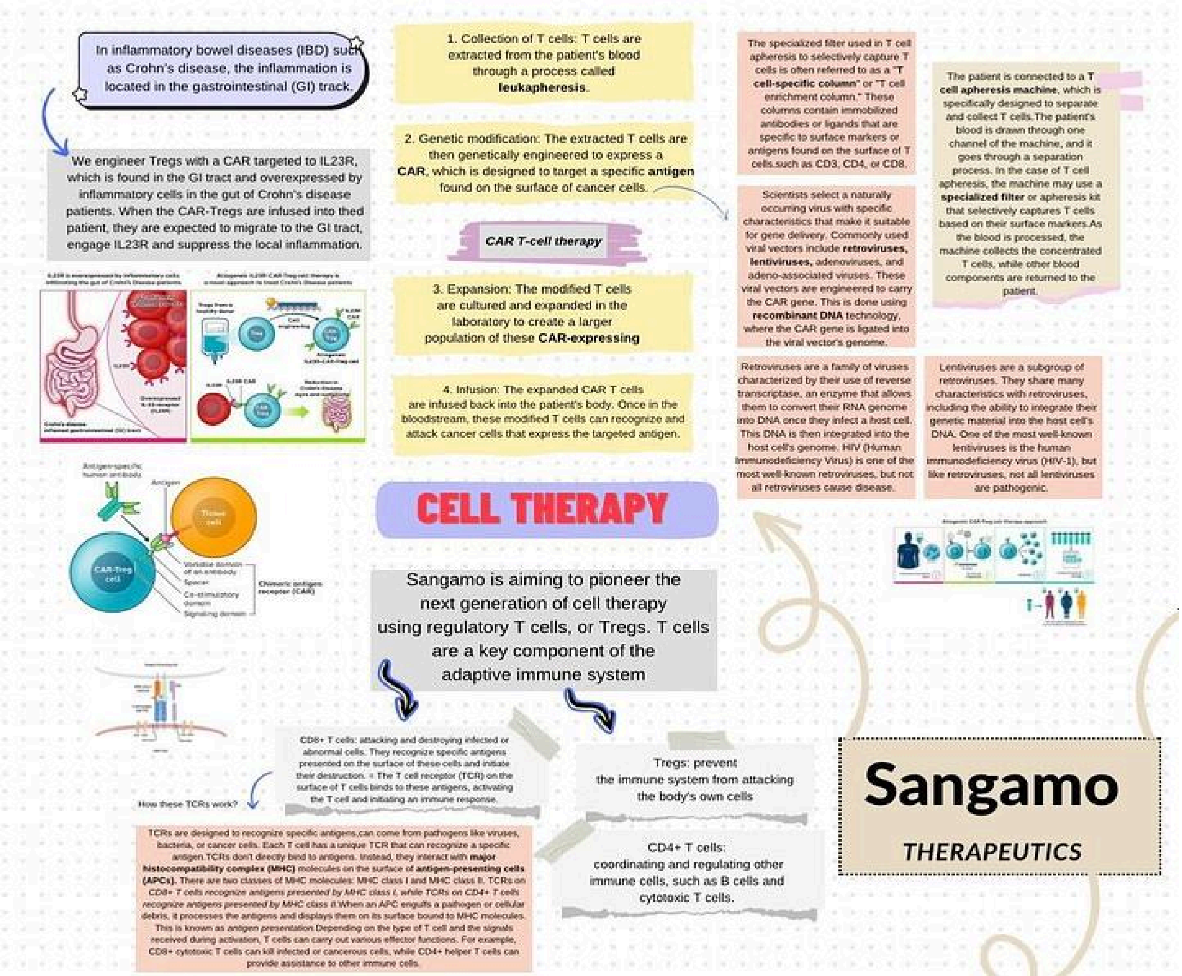
Questions

- How do Zinc Finger Proteins compare to other gene editing techniques, such as CRISPR/Cas9? What are the main reasons Sangamo focused on ZFs?

-What challenges or limitations do zinc finger proteins face in therapeutic applications, and how are these challenges being addressed by Sangamo Therapeutics?

# Sangamo

*THERAPEUTICS*



In inflammatory bowel diseases (IBD) such as Crohn's disease, the inflammation is located in the gastrointestinal (GI) track.

We engineer Tregs with a CAR targeted to IL23R, which is found in the GI tract and overexpressed by inflammatory cells in the gut of Crohn's disease patients. When the CAR-Tregs are infused into thed patient, they are expected to migrate to the GI tract, engage IL23R and suppress the local inflammation.

Antigen-specific human antibody
Antigen
Tissue cell
CAR-Treg cell
Variable domain of an antibody
Spacer
Co-stimulatory domain
Signaling domain
Chimeric antigen receptor (CAR)

How these TCRs work?

1. Collection of T cells: T cells are extracted from the patient's blood through a process called **leukapheresis**.

2. Genetic modification: The extracted T cells are then genetically engineered to express a **CAR**, which is designed to target a specific **antigen** found on the surface of cancer cells.

*CAR T-cell therapy*

3. Expansion: The modified T cells are cultured and expanded in the laboratory to create a larger population of these **CAR-expressing**

4. Infusion: The expanded CAR T cells are infused back into the patient's body. Once in the bloodstream, these modified T cells can recognize and attack cancer cells that express the targeted antigen.

## CELL THERAPY

Sangamo is aiming to pioneer the next generation of cell therapy using regulatory T cells, or Tregs. T cells are a key component of the adaptive immune system

CD8+ T cells: attacking and destroying infected or abnormal cells. They recognize specific antigens presented on the surface of these cells and initiate their destruction. = The T cell receptor (TCR) on the surface of T cells binds to these antigens, activating the T cell and initiating an immune response.

Tregs: prevent the immune system from attacking the body's own cells

CD4+ T cells: coordinating and regulating other immune cells, such as B cells and cytotoxic T cells.

TCRs are designed to recognize specific antigens,can come from pathogens like viruses, bacteria, or cancer cells. Each T cell has a unique TCR that can recognize a specific antigen.TCRs don't directly bind to antigens. Instead, they interact with **major histocompatibility complex (MHC)** molecules on the surface of **antigen-presenting cells (APCs).** There are two classes of MHC molecules: MHC class I and MHC class II. TCRs on *CD8+ T cells recognize antigens presented by MHC class I, while TCRs on CD4+ T cells recognize antigens presented by MHC class II.*When an APC engulfs a pathogen or cellular debris, it processes the antigens and displays them on its surface bound to MHC molecules. This is known as *antigen presentation* Depending on the type of T cell and the signals received during activation, T cells can carry out various effector functions. For example, CD8+ cytotoxic T cells can kill infected or cancerous cells, while CD4+ helper T cells can provide assistance to other immune cells.

The specialized filter used in T cell apheresis to selectively capture T cells is often referred to as a "**T cell-specific column**" or "T cell enrichment column." These columns contain immobilized antibodies or ligands that are specific to surface markers or antigens found on the surface of T cells.such as CD3, CD4, or CD8.

The patient is connected to a **T cell apheresis machine**, which is specifically designed to separate and collect T cells.The patient's blood is drawn through one channel of the machine, and it goes through a separation process. In the case of T cell apheresis, the machine may use a **specialized filter** or apheresis kit that selectively captures T cells based on their surface markers.As the blood is processed, the machine collects the concentrated T cells, while other blood components are returned to the patient.

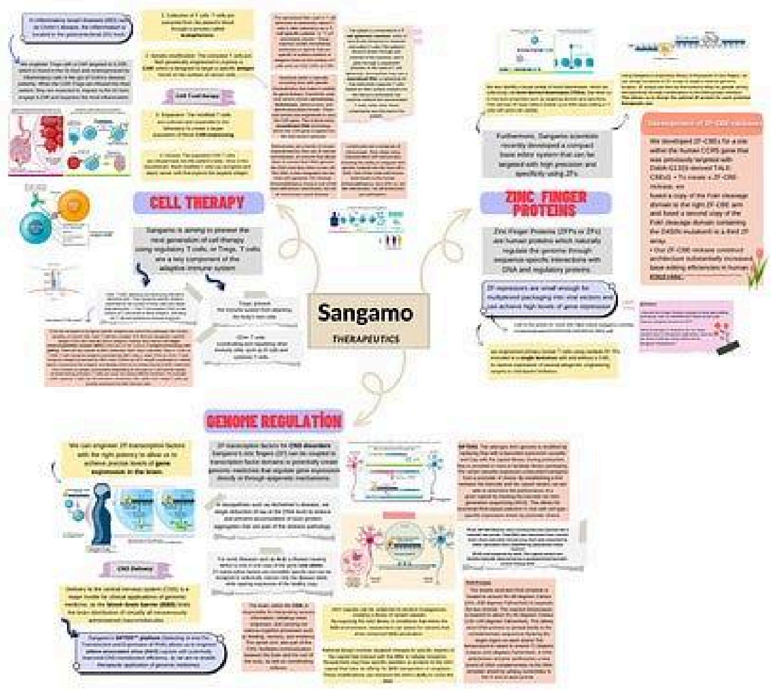
Scientists select a naturally occuring virus with specific characteristics that make it suitable for gene delivery. Commonly used viral vectors include **retroviruses, lentiviruses,** adenoviruses, and adeno-associated viruses. These viral vectors are engineered to carry the CAR gene. This is done using **recombinant DNA** technology, where the CAR gene is ligated into the viral vector's genome.

Retroviruses are a family of viruses characterized by their use of reverse transcriptase, an enzyme that allows them to convert their RNA genome into DNA once they infect a host cell. This DNA is then integrated into the host cell's genome. HIV (Human Immunodeficiency Virus) is one of the most well-known retroviruses, but not all retroviruses cause disease.

Lentiviruses are a subgroup of retroviruses. They share many characteristics with retroviruses, including the ability to integrate their genetic material into the host cell's DNA. One of the most well-known lentiviruses is the human immunodeficiency virus (HIV-1), but like retroviruses, not all lentiviruses are pathogenic.

**Sangamo**

*THERAPEUTICS*



MindMap about Sangamo Therapeutics

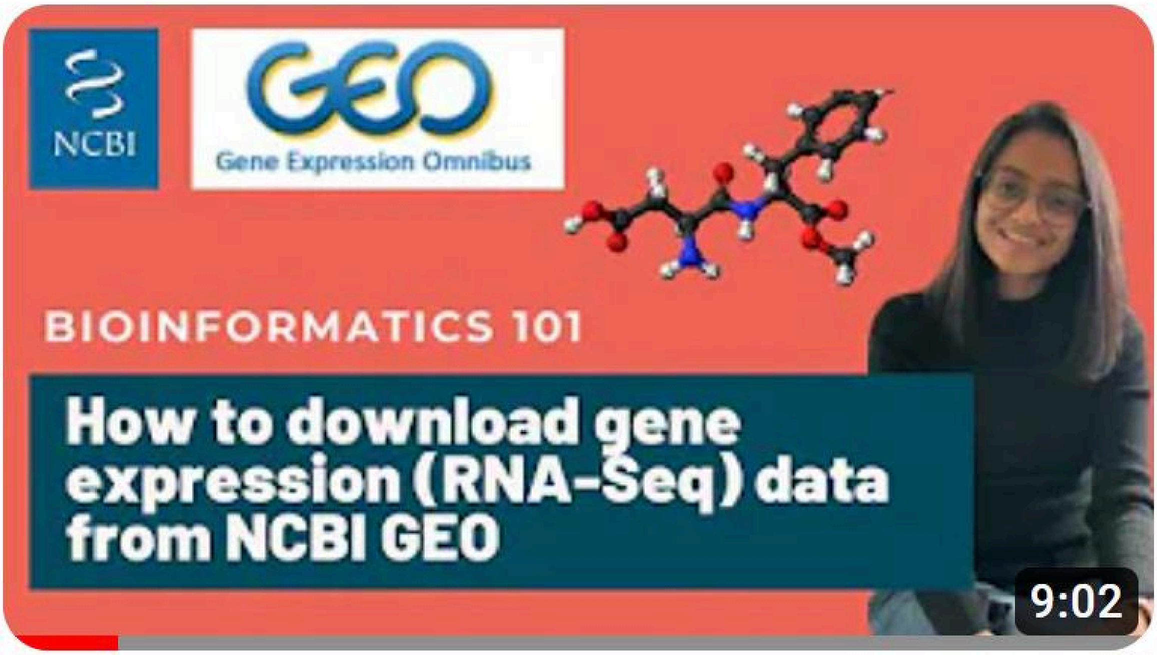
After dedicating a **two-week** period to thorough research on Sangamo and its applications (depends on your availability), I have synthesized my findings into a comprehensive mind map. **To explore the intricacies further, you can click [here].** Utilizing a mind map proved to be an optimal approach, ensuring a clear and organized understanding of the subject matter. By categorizing each application separately, this visual representation facilitates a streamlined process, allowing you to select a specific area for more in-depth exploration in the further steps.

## R Programming Language

In bioinformatics, we turn to R for its data prowess and statistical might. It's our go-to language for unravelling the intricacies of biological data, offering **a rich toolbox of libraries** designed just for us. With R, we seamlessly crunch large-scale genomic data, visualize complex datasets, and unearth insights from differential expression analyses. Its open-source spirit fuels collaboration, letting us share and refine tools worldwide. R isn't just a language; it's our key to unlocking the secrets within genomics and proteomics, a dynamic force in the ever-expanding realm of bioinformatics.

### *How To Learn It*

*Given the diversity in learning styles and abilities, my suggestion is to discover what works best for you. Explore various approaches to mastering the R language and tailor them to suit your preferences and objectives in bioinformatics. In this article, I'll share the method I employed in learning, offering insights that might resonate with your own learning journey.*


NCBI
GEO
Gene Expression Omnibus
BIOINFORMATICS 101
How to download gene expression (RNA-Seq) data from NCBI GEO
9:02

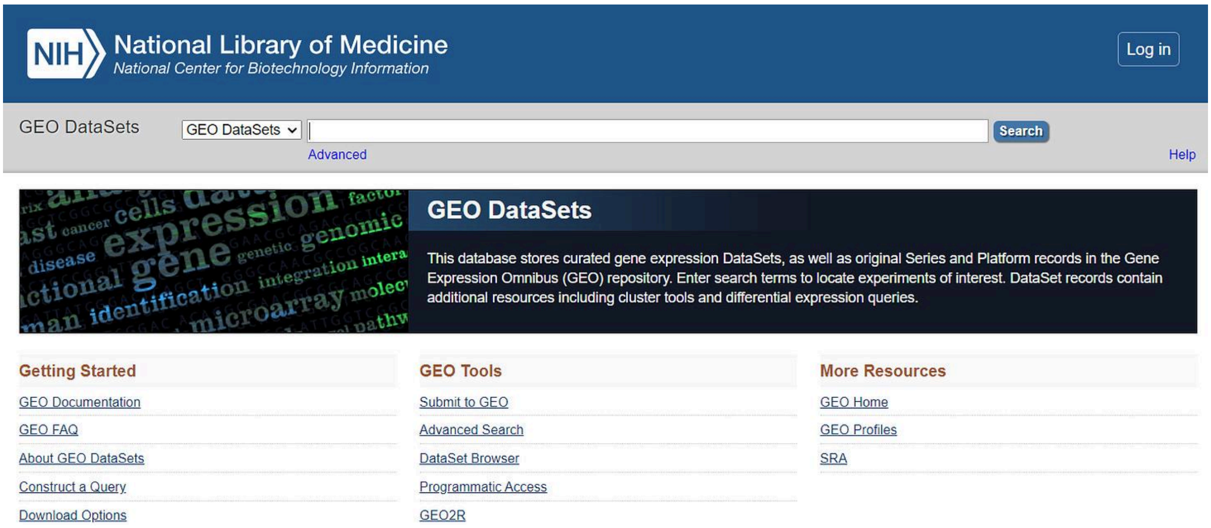
Bioinformatics 101 | How to download RNA-Seq data from NCBI GEO | Bioinformatics f...

**Youtube Tutorials:** I discovered that my most effective learning approach involves watching tutorials on YouTube first and then replicating the process independently. One standout channel, **Bioinformagician**, provides exceptional R language tutorials for bioinformatics. The creator, a young and passionate individual, makes the learning experience both engaging and straightforward. The tutorials meticulously cover every aspect, leaving no detail untouched. This method has proven remarkably easy for me to grasp complex concepts and execute them confidently. For those seeking a dynamic and thorough learning experience in R for bioinformatics, I highly recommend checking out Bioinformagician's tutorials [here].

### 5 down, 19 to go

The initial phase of this bioinformatics learning guide encompasses **the first five videos of the series**. In subsequent articles, we will delve into the further steps, exploring advanced topics and expanding our knowledge in the fascinating realm of bioinformatics.

Let's kickstart our learning journey! Before diving into specific research cases, we'll begin by focusing on breast cancer through a series of tutorials. In these sessions, you'll gain hands-on experience in **downloading patient data directly from the NCBI website.** This acquired skill will serve as a foundation for our later steps, enabling you to confidently select and explore your chosen disease in more detail.

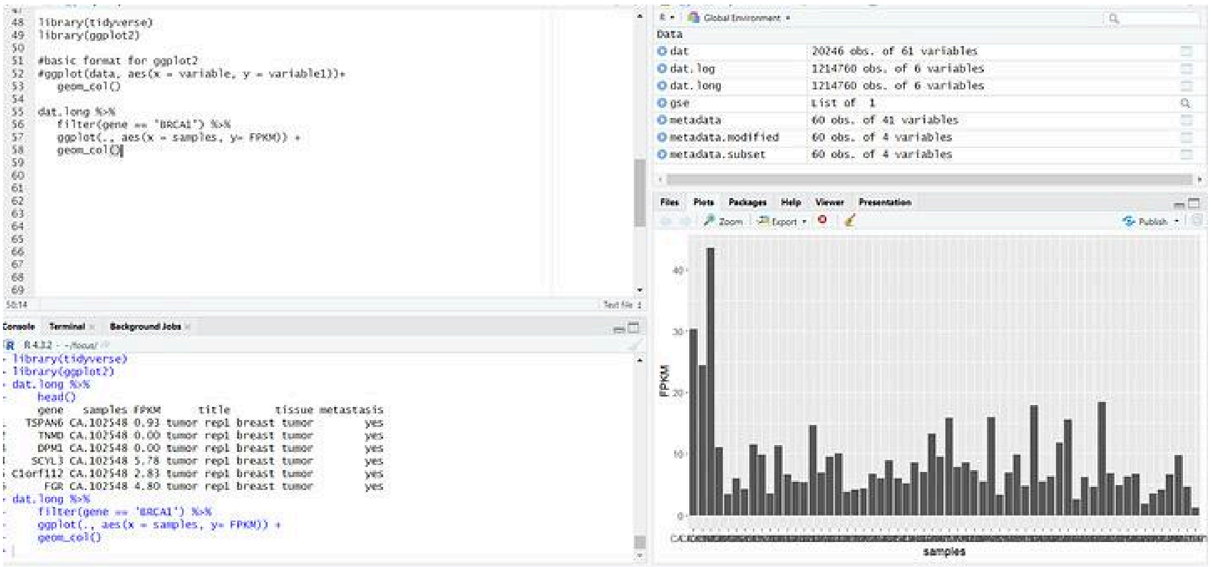

NIH
National Library of Medicine
National Center for Biotechnology Information
Log in
GEO DataSets
GEO DataSets
Search
Advanced
Help

GEO DataSets

This database stores curated gene expression DataSets, as well as original Series and Platform records in the Gene Expression Omnibus (GEO) repository. Enter search terms to locate experiments of interest. DataSet records contain additional resources including cluster tools and differential expression queries.

In the next phase of our learning journey, we'll delve into the art of **data manipulation**. Building on the foundation of breast cancer data obtained from the NCBI website, you'll acquire the skills to craft various datasets and tables. This process is crucial for extracting and *segregating biological information, setting the stage for conducting extensive and impactful experiments.*

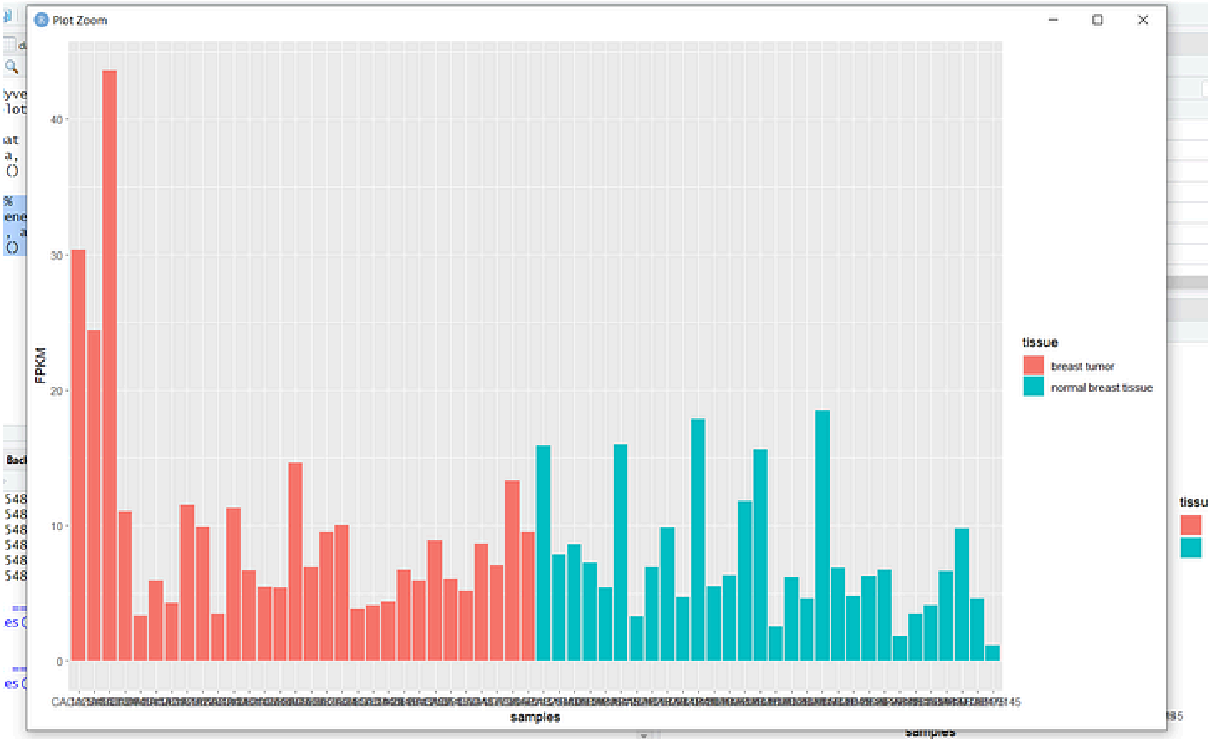
Now, let's dive into my personal favorite part – **data visualization**! In the upcoming tutorials, you'll learn the art of professionally visualizing biological data. These skills will empower you to create compelling graphics that can enhance the visual impact of your **research papers and experimental reports**. From intricate charts to insightful graphs, we'll explore a range of visualization techniques. Here's a sneak peek at some of the graphics we'll be crafting in the tutorials.

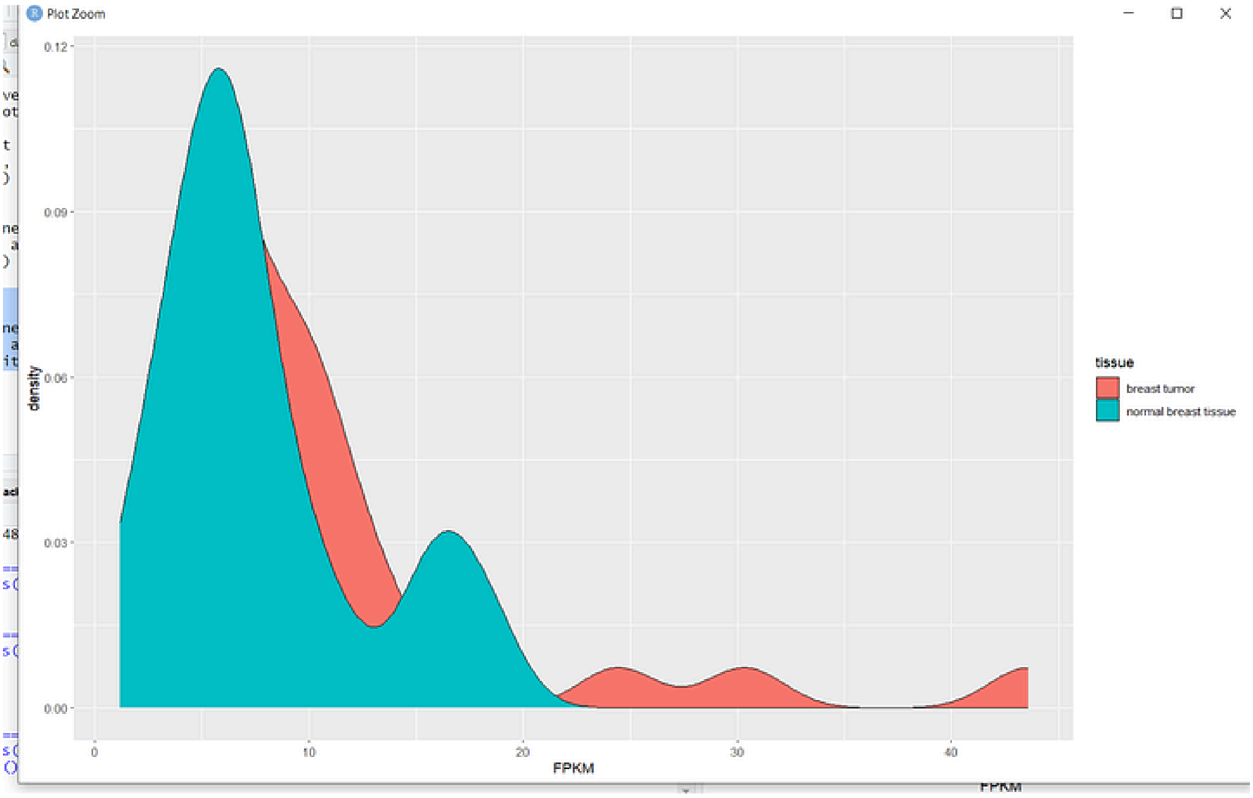


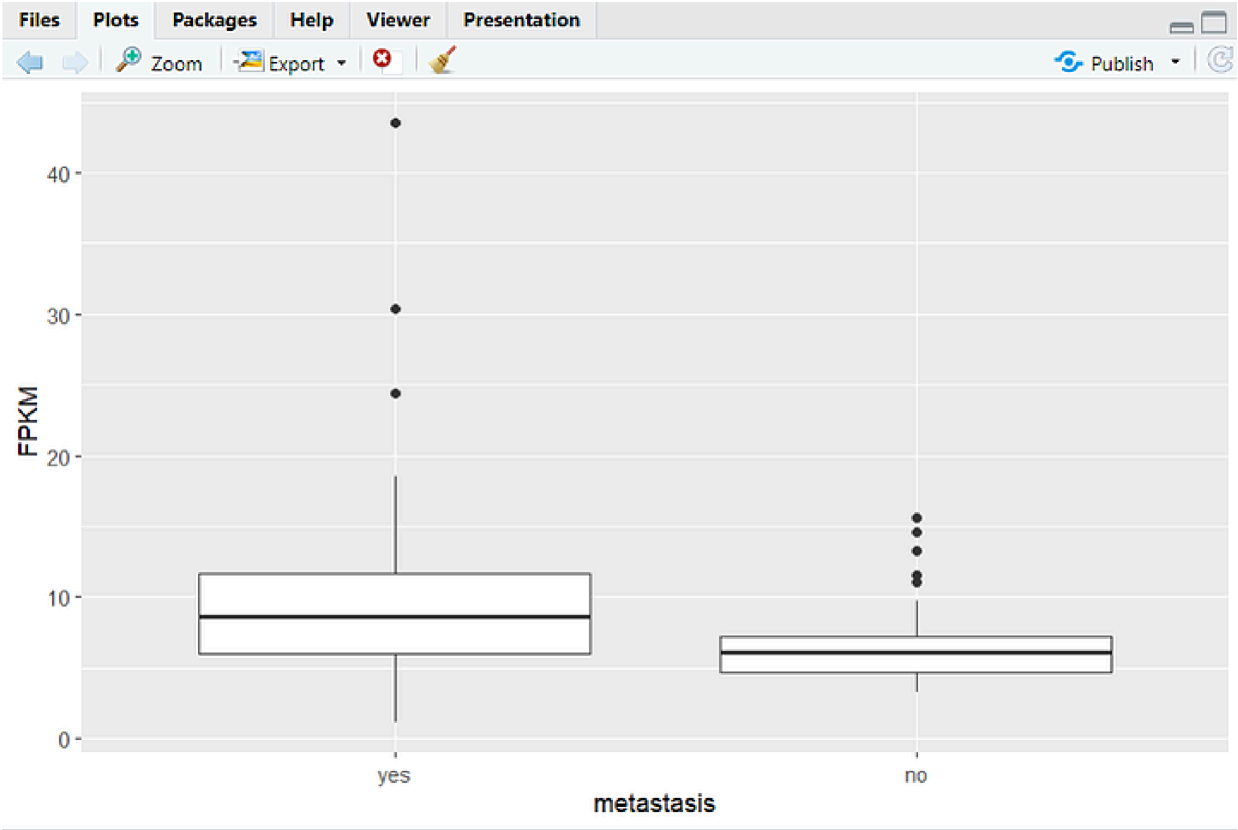
```
library(tidyverse)
library(ggplot2)

#basic format for ggplot2
#ggplot(data, aes(x = variable, y = variable1))+
   geom_col()

dat.long %>%
   filter(gene == 'BRCA1') %>%
   ggplot(., aes(x = samples, y= FPKM)) +
   geom_col()
```

```
R 4.3.2 · ~/focus/
> library(tidyverse)
> library(ggplot2)
> dat.long %>%
+     head()
      gene   samples FPKM      title        tissue metastasis
1   TSPAN6 CA.102548 0.93 tumor rep1 breast tumor        yes
2     TNMD CA.102548 0.00 tumor rep1 breast tumor        yes
3     DPM1 CA.102548 0.00 tumor rep1 breast tumor        yes
4    SCYL3 CA.102548 5.78 tumor rep1 breast tumor        yes
5 C1orf112 CA.102548 2.83 tumor rep1 breast tumor        yes
6      FGR CA.102548 4.80 tumor rep1 breast tumor        yes
> dat.long %>%
+     filter(gene == 'BRCA1') %>%
+     ggplot(., aes(x = samples, y= FPKM)) +
+     geom_col()
> 
```

```
Global Environment
Data
dat                  20246 obs. of 61 variables
dat.log              1214760 obs. of 6 variables
dat.long             1214760 obs. of 6 variables
gse                  List of  1
metadata             60 obs. of 41 variables
metadata.modified    60 obs. of 4 variables
metadata.subset      60 obs. of 4 variables
```







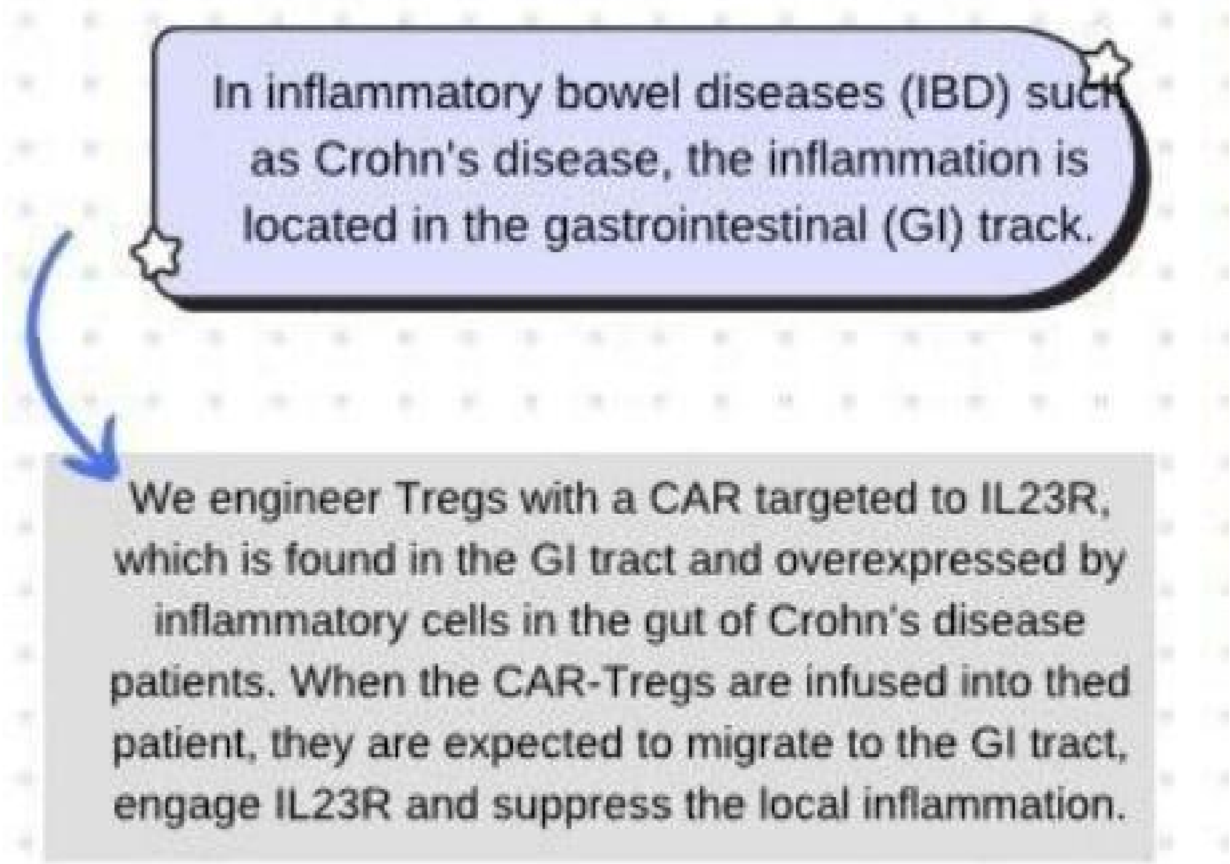
In the final phase of this learning series, we'll broaden our horizons by delving into a different type of data—**SRA**, accessible through the **NCBI website**. SRA stands for Sequence Read Archive, a repository that stores raw, high-throughput sequencing data. It includes vast amounts of genetic information from various organisms, providing a goldmine for researchers.

In the final installment of this tutorial series, we'll explore the **differences between RNA-Seq normalization methods**, with a specific focus on techniques like FPKM and RPKM. Understanding these normalization methods is paramount in the realm of bioinformatics, as they play a crucial role in ensuring accuracy and comparability across RNA-Seq datasets.

Before we venture into the intricacies of **DESeq2** in the upcoming videos, it's crucial to take a moment to understand our learning. In the following sessions, we'll revisit and reinforce the concepts we've covered, ensuring a solid understanding of the fundamentals. This reflective pause offers the perfect opportunity to adapt our newfound knowledge into practical applications **for your individual projects.**

### How to Personalize Our Learning Experience

Recall our initial research **on a company**? We immersed ourselves in exploring their applications, creating a structured mind map to comprehend their mechanisms. Now armed with insights into how they initiate experiments, let's use this knowledge to **replicate and apply** similar methodologies in our own projects.

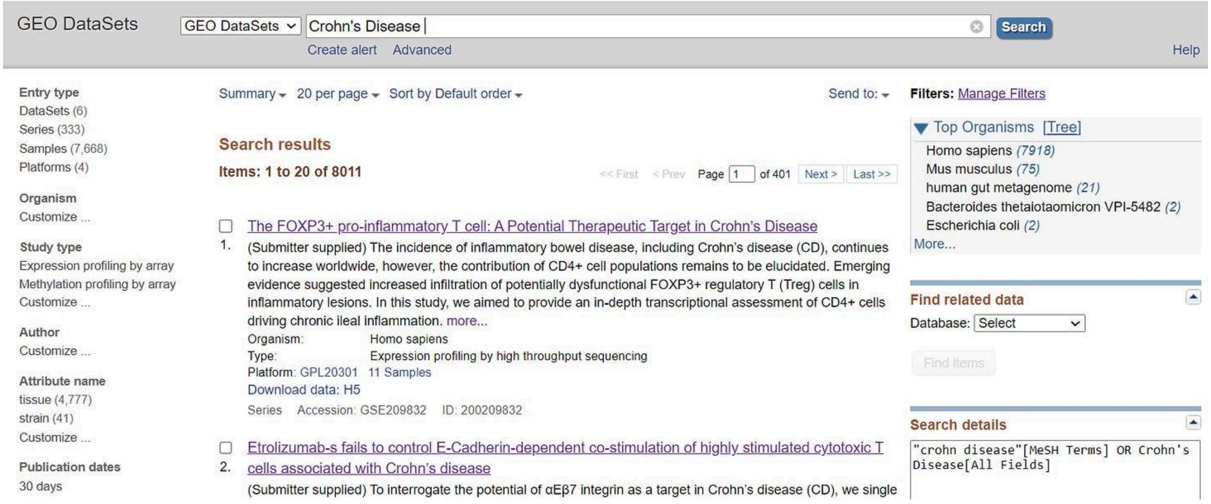


In inflammatory bowel diseases (IBD) such as Crohn's disease, the inflammation is located in the gastrointestinal (GI) track.

We engineer Tregs with a CAR targeted to IL23R, which is found in the GI tract and overexpressed by inflammatory cells in the gut of Crohn's disease patients. When the CAR-Tregs are infused into thed patient, they are expected to migrate to the GI tract, engage IL23R and suppress the local inflammation.

## Part of my MindMap

My enthusiasm lies in **CAR-T cell therapy,** and upon reviewing my mind map, you'll notice a specific focus on Sangamo's application of this therapy for patients with **Crohn's Disease**. Recognizing the potential, I've decided to make this the cornerstone of my *replicate project.*
Once you have selected the disease you want to focus on, utilize the **NCBI website** to identify the appropriate dataset. Specifically, navigate to **GEODatasets** and enter the name of the chosen disease for relevant information.



Search results

Items: 1 to 20 of 8011

□ 1. The FOXP3+ pro-inflammatory T cell: A Potential Therapeutic Target in Crohn's Disease

(Submitter supplied) The incidence of inflammatory bowel disease, including Crohn's disease (CD), continues to increase worldwide, however, the contribution of CD4+ cell populations remains to be elucidated. Emerging evidence suggested increased infiltration of potentially dysfunctional FOXP3+ regulatory T (Treg) cells in inflammatory lesions. In this study, we aimed to provide an in-depth transcriptional assessment of CD4+ cells driving chronic ileal inflammation. more...

Organism: Homo sapiens
Type: Expression profiling by high throughput sequencing
Platform: GPL20301 11 Samples
Download data: H5
Series Accession: GSE209832 ID: 200209832

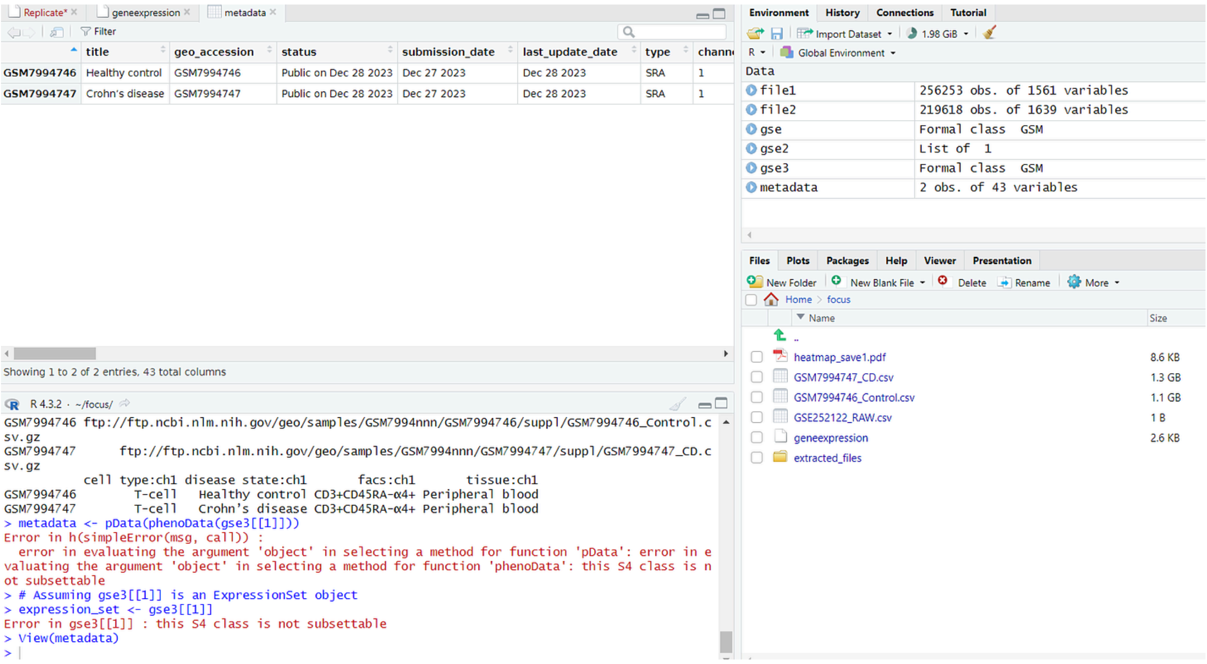
□ 2. Etrolizumab-s fails to control E-Cadherin-dependent co-stimulation of highly stimulated cytotoxic T cells associated with Crohn's disease

(Submitter supplied) To interrogate the potential of αEβ7 integrin as a target in Crohn's disease (CD), we single

"crohn disease"[MeSH Terms] OR Crohn's Disease[All Fields]

You will encounter numerous datasets. One significant mistake I made was neglecting to verify **the file type**. If it's in H5 or another format, it becomes challenging to convert and read in R. The most preferable format is **CSV**, and you will learn how to handle CSV files in tutorials. I've spent considerable time attempting to convert tar (H5) files, and I must admit, it was quite traumatizing! Make sure to find the dataset that best fits your needs, and don't forget to confirm if it's in CSV format.

Following that, all you need to do is follow the steps you learned from the tutorials applicable to your dataset! **While it may initially be challenging to adapt to the specifics of your dataset, with time, you'll gain a full understanding of what each step entails.** Feel free to ask questions, including reaching out to me through the contact information provided at the end.



|  | title | geo_accession | status | submission_date | last_update_date | type | chann |
|---|---|---|---|---|---|---|---|
| GSM7994746 | Healthy control | GSM7994746 | Public on Dec 28 2023 | Dec 27 2023 | Dec 28 2023 | SRA | 1 |
| GSM7994747 | Crohn's disease | GSM7994747 | Public on Dec 28 2023 | Dec 27 2023 | Dec 28 2023 | SRA | 1 |

Showing 1 to 2 of 2 entries, 43 total columns

```
GSM7994746 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM7994nnn/GSM7994746/suppl/GSM7994746_Control.c
sv.gz
GSM7994747      ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM7994nnn/GSM7994747/suppl/GSM7994747_CD.c
sv.gz
           cell type:ch1 disease state:ch1       facs:ch1       tissue:ch1
GSM7994746        T-cell   Healthy control CD3+CD45RA-α4+ Peripheral blood
GSM7994747        T-cell   Crohn's disease CD3+CD45RA-α4+ Peripheral blood
> metadata <- pData(phenoData(gse3[[1]]))
Error in h(simpleError(msg, call)) :
  error in evaluating the argument 'object' in selecting a method for function 'pData': error in e
valuating the argument 'object' in selecting a method for function 'phenoData': this S4 class is n
ot subsettable
> # Assuming gse3[[1]] is an ExpressionSet object
> expression_set <- gse3[[1]]
Error in gse3[[1]] : this S4 class is not subsettable
> View(metadata)
> |
```

| Data |  |
|---|---|
| file1 | 256253 obs. of 1561 variables |
| file2 | 219618 obs. of 1639 variables |
| gse | Formal class  GSM |
| gse2 | List of  1 |
| gse3 | Formal class  GSM |
| metadata | 2 obs. of 43 variables |

| Name | Size |
|---|---|
| .. |  |
| heatmap_save1.pdf | 8.6 KB |
| GSM7994747_CD.csv | 1.3 GB |
| GSM7994746_Control.csv | 1.1 GB |
| GSE252122_RAW.csv | 1 B |
| geneexpression | 2.6 KB |
| extracted_files |  |

The table I made from the data set, to use in further research
In my case, my dataset lacked numerous pieces of information necessary for creating visualizations. However, I addressed this by creating **metadata** that includes crucial details required for research papers. These tables have significantly streamlined the process of writing research papers and preparing reports. In your dataset, specifically check **if there's an FPKM section in your metadata**. If it exists, experiment with every visualization method you learn. This will undoubtedly enhance your experience and understanding of the data.