

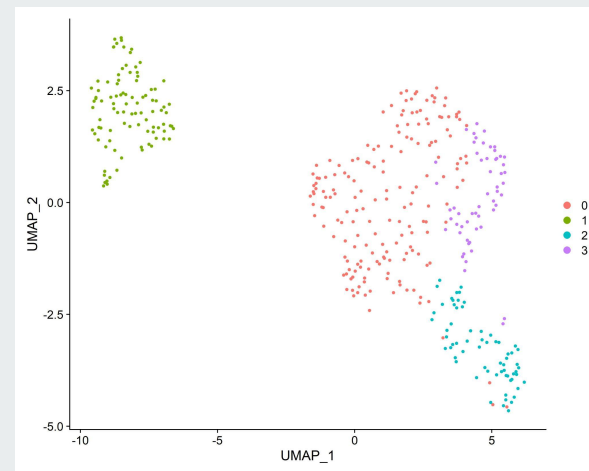
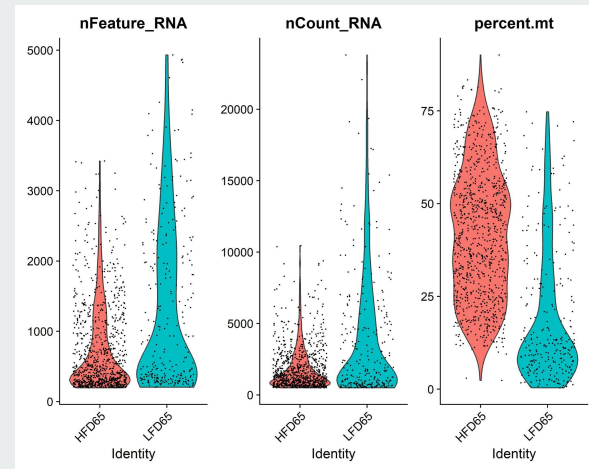


# ISCORE Fall 2023- Bioinformatics

Scholar: Kiran Paul, MS  
Research Service  
Professional-Informatics



Student: Menachem  
Allouch, BA, BS  
Undergraduate



---

# Introduction - What is Bioinformatics?

Bioinformatics is a unique field within STEM research in that it combines both aspects of biology and informative statistics of biological data.

It requires an in-depth knowledge of biological systems and mechanisms as well as a strong background in computer science/languages in order to properly analyze and interpret data.

Bioinformatics takes high-dimensional information and graphically optimizes it into low-dimensional graphs/plots.

Fun fact: The Human Genome Project, completed in 2003, was a major milestone in bioinformatics.

# Research Opportunity Experience

---

Some of the topics discussed during our weekly meetings at the Anschutz campus Death Star (Health Sciences Building):

- Cell sequencing
  - Single cell RNA sequencing
  - Bulk cell sequencing
- R/R studio
  - Seurat
  - DESeq2
- The workflow process
  - Sequencing
  - Preparing the data
  - Displaying the data in useful, informative graphs
- Practical applications
  - Effect of estradiol on breast cancer cells
  - In-vivo implant of breast cancer cells in mice, high fat and low fat diets

Fun fact: Bioinformatics tools are integral in predicting and analyzing the impact of mutations on human health.

---

# Cell Sequencing

Two of the most common cell sequencing methods used nowadays are bulk RNA sequencing and single cell RNA sequencing. Bulk sequencing involves the genetic library of a population of cells whereas single cell sequencing can provide the genetic makeup of a single cell.

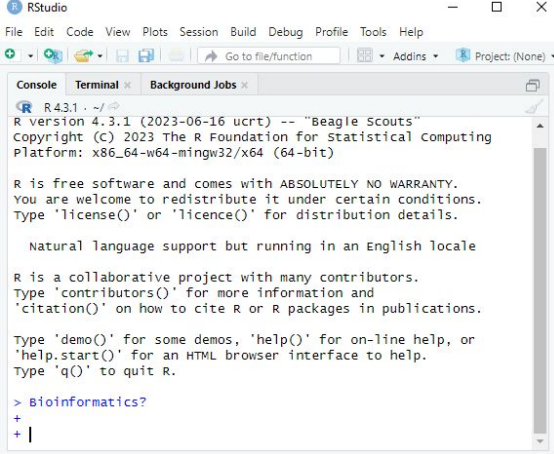
Fun fact: About 5% of the weight of a human cell is RNA.

## Why is this important?

Single cell RNA sequencing allows for the study of gene expression for individual cells, whereas bulk sequencing can only give the average gene expression across the entire population of the cells being studied. This has many real world applications, from developmental biology, to cellular heterogeneity, and tumor environments.

# R! The Programming Language

- R is an integrated environment of software facilities used for data manipulation, calculation and graphical display.
- Majority of meetings and learning experience revolve around learning and working within the R-Studio environment.
- Seurat - heterogeneity (variability) of single cell transcriptions
- DESeq2 - package used to analyze and interpret bulk RNA sequencing



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)
Console Terminal Background Jobs
R 4.3.1 ~/\
R version 4.3.1 (2023-06-16 ucrt) -- "Beagle Scouts"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> Bioinformatics?
+
+ |
```

# The Workflow Process

From the lab to the cure

## Sequencing

The cells are sequenced (either single cell or bulk) at a lab and the raw data is sent to the researcher.

## Quality Control

The data is processed and any genes that do not meet a certain criteria/threshold are removed from the data pool. Cell data can be dumped at this point if it does not pass QC.

## Normalization/ Feature Selection

Normalizing the data ensures that the data entries appear similar across all fields. Feature selection allows the researcher to choose those genes with high cell-to-cell variation. This means they are highly expressed in some cells and low in others.

## Dimensional Reduction/Bio markers

Dimensional reduction allows the researcher to reduce the number of variables into a more easily explorable dataset without sacrificing too much on accuracy. 52D>2D. Biomarkers labels clusters of cells within similar environments, showing relationships.

## Graphical Presentation

The final step is to present the data in an informative graph or plot, showing cell/gene relationships or clustering. This aspect helps biology researchers to find relationships between cells/genes and helps cure many diseases.

**Fun Fact:** Dimensionality reduction is common in fields that deal with large numbers of observations of variables, such as speech recognition, neuroinformatics, and bioinformatics.



# Practical Applications

Bioinformatics has many real world applications. One that we discussed were the effects of estradiol on breast cancer cells gene expression.

Another was the effect of low-fat or high-fat diets on breast cancer cells placed in-vivo in mice.

Testing for the effect of estradiol on breast cancer cells can help treat or prevent breast cancer. Estradiol is a regulator of growth for some breast cancer cells that express estrogen receptors. Blocking these receptors can possibly help prevent or treat breast cancer.

Different types of diets can have an increased or decreased risk for breast cancer. High-fat diets are typically closer linked to increased risk of breast cancer. (1)

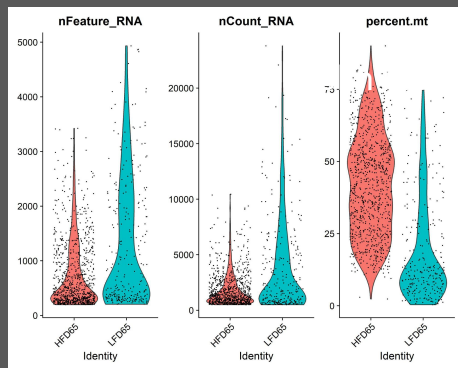
1. Uhomobhi TO, Okobi TJ, Okobi OE, et al. High-Fat Diet as a Risk Factor for Breast Cancer: A Meta-Analysis. *Cureus*. 2022;14(12):e32309. Published 2022 Dec 8. doi:10.7759/cureus.32309

**Fun Fact: Estrogen (another name for estradiol) is traditionally considered a 'female' hormone that helps a woman have female characteristics, supports menstruation, and conceive and carry a baby.**

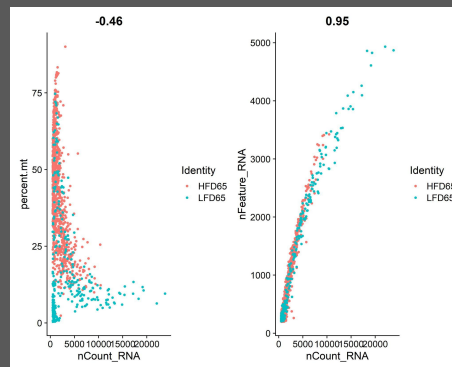
Fun fact: MCF-7 is a breast cancer cell line isolated in 1970 from a 69-year-old White woman.

# Graphs & Images

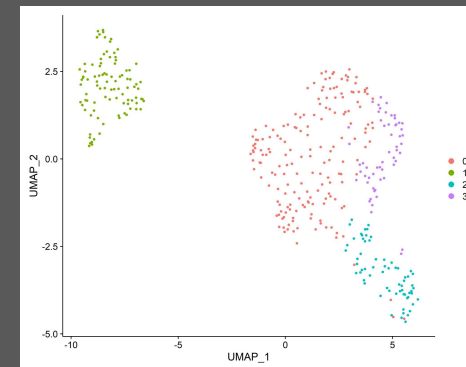
## Violin Plot



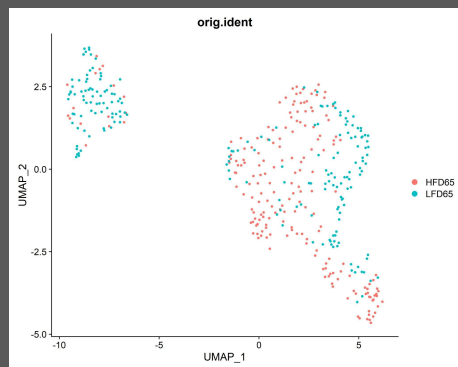
## Feature Scatter Plot



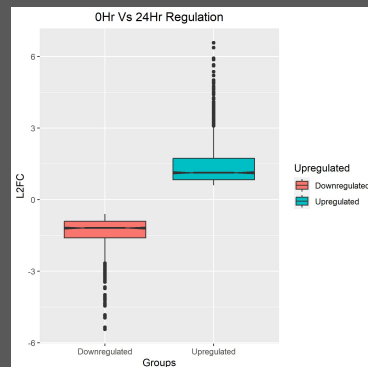
## UMAP



## UMAP



## Box and whisker Plot



## Excel Data of HFD/LFD

	baseMean	log2FoldChang	lfcSE	t-stat	pvalue	padj	t
17	U6	93.2095179	1.667009574	0.573691948	2.90575732	0.00366365	0.138301
24	NOC2L	1518.717676	0.713045	0.359771901	1.98193632	0.04748637	0.58001
33	RNF223	458.3879804	1.172532041	0.536634902	2.18497164	0.02889093	0.463743
149	Y_RNA	37598.48127	1.369135953	0.398330131	3.43718902	0.00058779	0.039256
150	ERRF1	936.9903409	0.781800192	0.307020956	2.54640661	0.01088383	0.265578
200	MAD2L2	376.0326093	0.756897472	0.324114299	2.33527948	0.01952883	0.373404
291	RNU1-1	69096.50927	1.977871087	0.364168717	5.43119438	5.60E-08	2.02E-05
327	MRTO4	659.503984	0.92014784	0.377845346	2.43524989	0.01488151	0.320461
356	ECE1	908.8956881	0.969749849	0.35675802	2.71822859	0.00656325	0.197956
381	KDM1A	205.402532	1.131647456	0.537103898	2.10694329	0.03511225	0.510327
526	SNHG12	251.0239262	0.931402436	0.39981778	2.32956732	0.01982903	0.376832
528	LINC01715	239.3028736	1.001652915	0.442335082	2.26446637	0.02354545	0.415725
529	RNU11	26046.64391	1.100141713	0.537607961	2.04636425	0.04072055	0.54297
548	RN7SKP91	5319.107687	2.824042939	1.288397165	2.19190403	0.02838644	0.461187
560	RNU6-40P	247.7942767	2.03363881	0.505414936	4.02370145	5.73E-05	0.006781
549	YARS	1042.987207	1.10409561	0.291881232	3.78268792	0.00015514	0.014802
698	GRIK3	128.4652365	3.841615882	0.924951417	4.15331639	3.28E-05	0.004315
664	MANEAL	272.1856447	0.930181923	0.33726152	2.75804344	0.00581485	0.182752



# Cultural Exchange

---

I enjoyed learning about Kiran and his background. I learned that Kiran came to the US about 12 years ago or so and has spent almost all his time here in Denver! I learned that it is very difficult, almost impossible, to become a doctor in India due to the selectively few spots open and huge numbers of applicants. I also learned that cricket is a fun sport to watch and if you mention Pakistan, that’s fighting words.

**Fun fact: There are 38 pro cricket teams in India. India has one of the best international cricket teams in the world.**

## Learning about new cultures

One of the most important facets of the ISCORE program is its ability to put people with completely different backgrounds into the same room and get them to get to know each other, even if it’s awkward at first. Even though the program was centered around the research, I particularly enjoyed the human aspect of interacting with new people that I would not usually have had the chance to interact with. Two of my takeaways that I learned from observing Kiran is that patience is key and almost anything can be debugged if you try hard enough.



# Thank You!

A big thank you to everyone involved in making my first research opportunity a possibility! Drumroll, please:

- Of course the researcher himself, Kiran! Kiran brought me into the world of bioinformatics with an unmatched wealth of knowledge.
- Elizabeth Evans, the cultural facilitator. Elizabeth’s ability to tell you that you might be culturally offending someone with a huge smile on her face is probably the best way to learn cultural sensitivity.
- And finally, Dr. Cristina Cenciarelli. If not for Dr. Cenciarelli’s effort and perseverance in promoting the ISCORE program, I would not have had such an amazing opportunity in the 1st place.
- Data displayed credited to the Kabos Lab in Anschutz.