

# Intro to Bioinformatics – 236523

## Final Project Workflow

by Tomer Goldstein - 205665409 & Eden Doron - 206336869

- 1) **Chosen disease:** Breast Cancer
- 2) We chose this specific disease for three main reasons:
  - We have close personal connections to women who have been ill with the disease.
  - it is extremely widespread (about 13% of women in the U.S. develop invasive breast cancer)
  - The disease is highly researched and therefore has a lot of previous studies and data that we can rely on.
- 3) We will be working with the dataset found in  
<https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-52194/?keywords=%22breast%20cancer%22%20&organism=Homo%20sapiens&exptype%5B0%5D=&exptype%5B1%5D=&array=&sortby=atlas&sortorder=descending&page=1&pagesize=500>

Which has RNA-seq data from 20 homo sapiens breast tumor samples.

- 4) From a brief search we found many manuscripts which look interesting and relevant to our project.  
After narrowing down the papers we ended up with a few that stand out:

- [Novel Insights into Breast Cancer Genetic Variance through RNA Sequencing](#)
- [RNA sequencing of cancer reveals novel splicing alterations](#)
- [Transcriptomic landscape of breast cancers through mRNA sequencing](#)
- [Repeated observation of breast tumor subtypes in independent gene expression data sets](#)
- [Emerging Role of Genomics and Cell-Free DNA in Breast Cancer](#)

5) We plan on analyzing the gene expressions in the different subtypes of breast cancer.

As we did similar work in some of the homework during the semester, we thought about analyzing the data with RNA-seq to produce differential expression results (maybe also using Xcell) in order to identify the genes regulation differences between the cancer subtypes. Next, we would like to examine the connection between the top regulated genes (and down regulated as well) and different proteins that affect tumor growth using STRING\CLIPdb as we saw in the last tutorial.

We wanted to know if this is the right way to approach the problem?  
Maybe other insights and ideas we haven't thought about?