Initial Experimental Design

1. **What is an appropriate question?**
    a. Background: miRNA
        i. Small non-coding RNA
        ii. Complementary to mRNA
        iii. Regulate gene expression:
            1. Degrade mRNA
            2. Repress translation
        iv. Potential novel biomarkers for diagnosis, miRNA mimics can be used as therapeutics
    b. Using differentially expressed miRNA data:
        i. Which miRNAs are up-/down-regulated during certain stages of breast cancer?
        ii. Use miRNA expression data as a novel way of categorizing breast cancer tumours ➔ look for trends?
        iii. Just an idea from Wayne: use miRNA expression data to trace back the phylogenetic development of breast cancer tumour to identify tumor subtypes
2. **What is my hypothesis?**
    a. (i.) Each stage of breast cancer has distinct miRNA expression patterns
        i. Same patient
    b. (ii.) miRNA expression varies among the different breast cancer tumour subtypes
        i. There's many subtypes
    c.
3. **What are my datasets and tools to work with?**
    a. TCGA database
        i. https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga
        ii. E.g. TCGA-BRCA
        iii. Contains miRNA expression data for breast cancer tumour samples
    b. ArrayExpress
        i. https://www.ebi.ac.uk/arrayexpress/
    c. METABRIC
        i. https://www.cbioportal.org/study/summary?id=brca_metabric
    d. Can use these databases to identify the target genes of miRNA:
        i. miRWalk: http://mirwalk.umm.uni-heidelberg.de/

    ii. miRBase database: http://www.mirbase.org/
e. miRNA analysis: https://tools4mirs.org/
    i. Differential expression analysis tool
f. Matplotlib package for data analysis: https://matplotlib.org/
g. miRGE, analysis of miRNA sequencing data: https://github.com/mhalushka/miRge
    i. https://www.biorxiv.org/content/10.1101/250779v1.full

4. **What is a method to answer the question?**
    a. Obtain breast tumour miRNA data from database(s)
    b. Determine which miRNAs are differentially expressed in breast cancer tumours (this has been studied before)
        i. New profiling algorithm ➔ different results
        ii. Some studies use small sample sizes ➔ our research can use more samples from multiple databases

5. **How much do I need to learn from scratch?**
    a. Learn more about statistical analysis
    b. Python and Object-oriented programming
    c. Matplotlib
    d. Biological background of breast cancer
    e. Bioinformatic concepts

6. **What is a likely outcome?**
    a. Much of the literature supports differential miRNA expression ➔ hypotheses are likely true
        i. 4 miRNA signatures predict breast cancer metastasis: https://pubmed.ncbi.nlm.nih.gov/29555574/
        ii. miRNA suppresses ER+ breast cancer: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6134714/#Sec2title
        iii. 5 miRNA signatures are predictive of breast tumor aggressiveness: https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.28171

7. **What are the visualizations like when presenting results?**

a. Flow chart to give an overview of the methods
b. Visualize data in graphs
    i. Bar graphs for miRNA expression levels
    ii. Clustering plots for the association of different miRNA and breast tumour subtypes
c. Identify the differentially expressed miRNA and conduct statistical analysis
d. (or) comparison of different stages of breast cancer instead of analyzing one patient at a time

8. **How do I validate what I am doing?**
    a. Look for statistical significance (e.g. P-value)
    b. Compare findings from the experiment to literature (i.e. what did we get vs. what was expected?)
    c. ROC curve: used to evaluate machine learning algorithms. (yet to decide whether useful or not) https://reader.elsevier.com/reader/sd/pii/S0031320396001422?token=9CABF06B85626DE67AFB4DCD76B0A76E4020C96BAEF6DE873FC203972254209248F5D71F0923C0C5FAD85D4FC46FCC51

Scratch:

https://portal.gdc.cancer.gov/files/76fc8f6f-8332-4a64-809a-7f106393a269

De novo

https://onlinelibrary.wiley.com/doi/epdf/10.1002/psc.595