Assignment 6

Topics: Chromatin conformation capture, RNA and protein structures, online tools and resources, and microscopy data analysis

Due date: 17 November 2025 at 11:59pm

Rules:

- You can work in group, but write your own answers
- You can use AI to help, but don't abuse it. Credit AI when used
- The objective of the assignment is to provide you with experience. Explain your work and observations. Don't just paste a screenshot of the result.
- You can contact me to ask for clarification

Credit: GPT-5 was used to aid the design of the assignment

Part A. Exploration of chromatin structure

Let's view the chromatin structure of human genome.

1. Go to **3D Genome Browser** at https://3dgenome.fsm.northwestern.edu/
2. Select Hi-C data from human embryo (H1-ESC, **GSE52457**)
3. Explore the chromatin region surrounding beta-globin gene (**HBB**)
    - Zoom in and zoom out to explore
    - Is HBB located in a topologically associating domain (TAD)? What are genes are located in the same TAD?
    - Is the TAD that HBB belong to part of the active or inactive chromosome?
    - Is there any long-range chromatin loop involving HBB locus? What about nearby genes?
    - Do you think the Hi-C profile show a clear pattern supporting the TADs and A/B compartments in this region?

Part B. Differential expression analysis using GEO2R tool

We want to analyze the transcriptomics data from this study of colorectal cancer from 2012, https://pubmed.ncbi.nlm.nih.gov/22496204/, by Kemper, K. et al. to identify genes that are differentially expressed in colorectal cancers.

1. Identify the **Gene Expression Omnibus** record for this transcriptomics dataset

2. Use the **Analyze with GEO2R** function on GEO website
3. Perform a differential expression analysis between 6 normal mucosa samples and the other 90 colorectal cancer samples
    - What are the top up-regulated genes in cancer? Report the gene symbols, adjusted p-values, and fold changes. Are they consistently up-regulated in all cancer samples or only some samples?
4. Perform another differential expression analysis, this time between 4 normal mucosa samples and the 6 paired colorectal cancer samples (find the matching tumor IDs)
    - Compare the result with the previous analysis. Are the same of different genes up-regulated?

Part C. Literature Analysis (Chromatin Architecture)

Study this landmark paper that first developed the Hi-C technique from 2009, https://www.science.org/doi/10.1126/science.1181369, "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome" by Lieberman-Aiden, E. et al., and answer the following questions:

- The original Hi-C protocol used restriction enzymes to digest the DNA. However, there are many fragmentation methods for DNA. Discuss the pros and cons of these methods in the context of Hi-C protocol.
- The analysis resolution of Hi-C data was set at 1 Mbp. How did the authors ensure that the sequence data are enough to support this resolution?
- Did all human chromosomes randomly interact or are there patterns that suggest/explain why some chromosomes tend to aggregate together?
- How did the authors explain the plaid patterns in Figure 3?
- How did the authors experimentally validate some of the findings about chromatin 3D structure?

Part D. Microscopy Data Analysis

Let's explore the capability of image processing to identify cells in microscopy images.

1. Go to CellPose website, https://www.cellpose.org/
2. Upload your favorite microscopic images (you can find some from https://idr.openmicroscopy.org/) or use some of the example images
3. Discuss the quality of the predicted cell outlines and cell masks

    - What kinds of errors and inaccuracies are there?
4. Study the CellPose paper
    - What does the algorithm identify as “cell pose”?
    - Why do you think the algorithm is so flexible (work on many types of images)?

Let’s get more serious with microscopy image analysis!

1. Install CellProfiler https://cellprofiler.org/
2. Download the example for human cells from https://cellprofiler.org/examples
    - The zip file contains both the image data and the parameter for CellProfiler
3. Process the downloaded images using the specified parameters
4. Write a report on the analysis process and result
    - Explain what CellProfiler conceptually does and what biological information was extracted at each step in the pipeline
    - Explain the outputs