STA130H1S – Fall 2022

Week 5 Tutorial Handout

Today's agenda (5 min):

- Project background
- Team up & introduction
- Group discussion
- Class discussion, Project clarification, Q&A
- Scientific poster introduction (on a separate handout)
- Important take-homes

Project background (30 min) :

Prior to this tutorial students (and TAs!) should be familiar with the project introduction: https://github.com/pointOfive/STA130/blob/main/README.md#course-project

At the beginning of the tutorial, watch these two videos with the students (play at 1.25X on projector if available, or group students and let them watch on laptops):

- **A 12 min intro video on cancer:** https://www.youtube.com/watch?v=RZhL7LDPk8w
- **An 11 min video on thinking behind developing cancer treatment:** https://www.youtube.com/watch?v=7zk0ubr_lzo&list=PLSQl0a2vh4HBC-y3HzLVtvoCWk36MrWyn&index=4

Topics to discuss after watching the video (can do this later during class discussion):

- Cancers are abnormally-behaving cells (phenotype) caused by gene mutations (genotype). The cells used to generate this dataset are melanoma cell lines – a type of aggressive skin cancer cell modified for culturing in the lab.
- Genotype vs phenotype:
  - "Genotype" is genotype is an organism's unique sequence of DNA
  - "Phenotype" is an organism's actual observed properties, such as morphology, development, or behavior
- Proteomics:
  - The proteome is the entire set of proteins produced or modified by an organism.
  - According to the central dogma, the flow of genetic information follows the direction: DNA $\rightarrow$ RNA $\rightarrow$ protein. Therefore proteins can be thought of as an intermediate between genotype and phenotype.
  - In the context of this dataset:
    - Levels of 22 transcription factors (which are proteins), are associated with/surrogates of cell genotypes
    - Combinatorial levels of 4 specific proteins (MiTFg, NGFR, SOX10, and AXL) are used to define/characterize cell state/phenotype

- Cancer treatments
  - main goals are to slow/stop the proliferation and expansion of malignant cells
  - possible through multiple ways: killing mutated cells, modifying mutations in the cell, changing tumor environment
  - in this particular dataset, we hope to change deleterious phenotypes to healthy phenotypes by controlling the transcription factor network
  - what you can do, for example, is to look for changes in transcription factor network dependency structure under different experimental conditions

Team up & introduction (10 min) : Let the students team up into groups of four. If there're not enough people, groups of three are also fine. Assign single students to groups that are not yet full so there won't be anyone left without groups. Keep track of groups on your laptop. Give students time to introduce themselves if they don't know each other.

Group discussion (20 min) : Pose the following questions to the students:

- What do you take from the following paragraph now that you now more about cancer?

  The data we will work with is based on advances in the fields of Flow Cytometry for single cell analysis and Mass Spectrometry for measurement of cellular proteomic processes (the phenotypical process endpoint of cellular function and behavior). Based on these technologies, the multivariate landscape of proteomic activity can be measured for a single cell in any experiemental condition for any cell type (e.g., cancerous and benign cellular lines) at scale. By understanding typical cellular homeostasis of healthy and deleterious cells, and observing the phenotypical transformation of cellular proteomic homeostasis over time in response to different experimental conditions, we may eventually be able to understand how to direct deleterious cellular states to transition into non-deleterious states. And with this "data-driven identification and control of high-dimensional dynamical systems" we will be able to fight cancer!

- Can you pick a couple variables and come up with simple visualization that helps with addressing one of the questions:
  - Do protein levels in experimental condition x change over time t ?
  - Are protein levels at time t different between experimental conditions x_1 and x_2 ?
  - At time t in experimental condition x, what is the relationship between different proteins?
  - Can we predict phenotypical outcomes (Y) valule/state from transcription factors (TF)?
  - At what times t in experimental conditions x can we do this?
  - At time t in experimental condition x, what TF(s) is/are predictive of Y value/state?
- What would be one (of the three) ideal outcome for this project? Based on your example what could indicate a 'good' or 'bad' cellular homeostasis?

Class discussion, Project clarification, Q&A (20 min) : Ask students reconvene to clarify their understanding/questions of the project. You may ask each group to select a representative to talk about their conclusion and/or present their data visualization.

Scientific poster (20 min) : Walk through the handout on scientific handout with the students so they understand:

- components/sections of a scientific poster
- what makes a good scientific poster

Important take-homes (20 min) :

- Encourge students to start the analysis with what they've leanred in class as soon as possible
- Remind the students of the Poster presentations on Thursday December 8th (class time, venue TBD)
- Direct the students to the project proposal rubrics page and the project presentation rubrics page (Go over them with students if time). **Emphasize** the two major criteria for evaluating their project:
  - 1) Appropriate applications of statistical analyses, as well as supporting visualizations and explanation.
  - Note that the analyses should cover 3 of the 5 categories: Hypothesis Testing, Confidence Intervals, Correlation, Regression, Classification
  - Therefore it's recommended that they do not go beyond 2 time points for analysis (using Two-sample hypothesis test)
  - 2) Addressing the bigger picture of the project: what is "good" cellular homeostasis, and how can "bad" cellular homeostasis be changed to be "good"?