Background

Cross-reactivity is the ability of a particular T cell to recognize and respond to several different antigens. T cell receptor (TCR) mediates recognition of antigens. You are working with a group of wet-lab scientists who are interested in a clonal population of T cells (all encoding the same TCR) with known reactivity to a peptide. They want to know whether this T cell is cross reactive to other peptides and have generated a set of peptides to test it against. As a proxy for T cell activity, they measure CD69+ expression with mean fluorescence intensity (MFI) using a flow cytometer. CD69 is a cell surface maker that is upregulated when a T cell is activated by an antigen (peptide). Some of the peptides will induce T cell activation and some will not. **Your task is to develop a method to identify reactive peptides and report your results.**

It is worth noting that even when there is no peptide present, there may some baseline level of CD69+ expression and that baseline is characteristic of a cell-line on a given day. Furthermore, flow cytometry is a delicate procedure and many factors irrelevant to the actual biology can affect the resulting MFI, including sample preparation and equipment setup.

Data

We have provided you with data from three experiments testing the same T cell of interest. Each experiment consists of 216 samples, representing 191 unique peptides. As a positive control, multiple replicates of the known, on-target peptide (peptide_00000) are distributed across each plate. As a negative control, there are multiple replicates without any peptide (i.e., NOPEP) distributed across each plate.

Samples are analyzed one at a time on the flow cytometer. Only one 96-well plate can be loaded onto the flow cytometer at a time. For each sample, the plate, well, time of analysis, and resulting MFI value are reported. Data is output in tab-separated format, one file per run of the cytometer.

Table 1. Example file formatting

name	Plate.ID	Well	MFI	Runtime
NOPEP	1	A1	15281	12/4/21 5:00
peptide_48207	1	A2	15899	12/4/21 5:02
peptide_00000	1	A3	37260	12/4/21 5:04

Instructions

The goal of this exercise is to understand how you'd solve a problem typical to our group. We would like to see how you frame and distill problems, turn data insight into a statistical solution, and present results. There is no single "correct" solution to this problem, so be

creative, document your work, and let us know what you find!

1. Develop a method to determine which peptides are cross reactive in an experiment. Include all your code and please make sure to include instructions on how to reproduce your results. Assume that this is the first batch of many experiments, so please ensure that your code is robust, flexible, and documented. You are welcome to choose R and/or Python for this task.
2. Prepare a document containing all your findings and visualizations. Pretend as if you are preparing a report to share with a broad audience of company stakeholders, including members from the wet-lab and computational biology.

Below are some suggested steps, but these are by no means cook-book style directions.

- Write a function to read in, parse, and visualize the raw data from each experiment. Please describe your visualization, as well as any insight you have gained into the structure of the data.
- Determine which peptides are reactive. How certain are you in these calls? Create a visualization to show this.
- Compare your results across experiments.

Assessment

We will evaluate the solution based on the following criteria:

- Use of data visualizations.
- Code cleanliness, useability, and documentation
- Clear description of your results and conclusions
- Clear justification of your method, including discussions of the pros and cons of your approach.

Final note

We are happy to answer clarifying questions, but the method approach is up to you. You have exactly 3 days to complete the task. Good luck – we look forward to seeing what you come up with!