Exploring CD8+ T-cell Specificities using Single Cell Immune Profiling with an Outlook to Reproducible Bio Data Science

William Hagedorn-Rasmussen

Abstract

This is supposed to be an abstract.

Introduction

With all the advancements in recent years in regards to computational power, laboratory techniques and -equipment, the amount of produced data is ever increasing. The focus in this article will be on biological data, even though it applies to a range of fields. For a reference on how quick the situation develops, the cost of sequencing the entire human genome decreased from (INSERT VALUE) to (INSERT VALUE) in a matter of (INSERT VALUE). The development obviously didn't stop there. As a more recent example, (INSERT EXAMPLE, e.g. RNA-sequencing/transcriptomics (see slides for system biology lecture 6)).

When a bioinformatician works with data, it is important it is done in a reproducible manner. Exactly as in the laboratory, it is paramount that other scientists is also able to achieve the same results and to come to the same conclusions. That imposes another problem, as the code used for working with data is not necessarily made available. If that's the case, how can you ever be sure, the no mistakes were made, or that any other problem occurred. Thereby, it is not only important to be working reproducible, but also transparent.

A part of working with data from experiments is to find meaningful correlations or even make new discoveries. These findings should be maked presentable to share recent discoveries. Especially when presenting to non-data stakeholders this becomes important.

- What is the situation now / why is this important?
  - Difficult to reproduce data analyses
  - Amount of available data grows
  - The presentation of data to non-data stakeholders
    - E.g. in industry
    - Make it interactive to reduce work load for bioinformatician
- What can this project do to help the issue?
-
- What kind of data am I working with?
  - Needed theory to understand
    - How are T-cells formed
    - What determines the TCR-sequence and what is the TCR
    - Alpha- beta sequence description
    - MHC, pMHC
    - Single Cell Immune Profiling

Methods

- Description of the structure of the package?
- Description of data set?
- How I cleaned
  - Describe dimensions of the data set(s)
- How I augmented
  - Describe dimensions of the data set(s)
- The integration into Shiny?

Results

- Mention pointers from method
  - What can the package offer
- Presentation of models
- How does the Shiny app work / what can it do

Discussion

Conclusion