Introduction to data analysis for natural and social sciences

Marco Casari

June 26, 2023

1 Introduction

The present document constitutes the second part of the exam. A summary of article "Inferring the immune response from repertoire sequencing" is provided.[1] Results sections "Modeling repertoire variation" and "Inferring the noise profile from replicate experiments" are supplied with technical results.

2 Summary

The article introduces a probabilistic model able to estimate the noise and describe clonotypes expansion in longitudinal repertoire sequencing data. Next Generation Sequencing (NGS) provides large amounts of repertoire sequencing data, which traditional inference methods fails to describe accurately due to experimental noise and biological diversity. Many factors contributes to the noise, some are due to the experiment design, e.g. sampling procedure, library preparation, others have biological origin, e.g. gene expression. These factors introduce variability in sequence counts which are not related to the natural variability of immune cell receptors and should be removed to have an accurate estimate of the distribution of clonotype counts.

2.1 Modeling repertoire variation

2.2 Inferring the noise profile from replicate experiments

---

[1] Version 2 of the article is referenced. Full citation: Puelma Touzel M, Walczak AM, Mora T (2020) Inferring the immune response from repertoire sequencing. PLoS Comput Biol 16(4): e1007873. https://doi.org/10.1371/journal.pcbi.1007873.