Leids Universitair Medisch Centrum

Bioinformatics Internship master student

Description

For our research group 'Immune Monitoring' at the department Immunohematology and Blood Transfusion we are looking for a Master student in medical science or Bioinformatics for an internship project (6-8 months) with basic programming skills in R or Python.

This concerns an internship in the English language.

Information

Scoring clones of B cells in RNA-Seq data

Indu Khatri12, Erik van den Akker[2], Magda Berkowska[1], Marcel Reinders[2] and Jacques van Dongen[1]

Affiliations:

[1]Department of immunohematology and Blood Transfusion
[2]Leiden Computational Biology Center

The B-cells of the immune system recognize antigens through their B-cell receptors (BCR), also called immunoglobulins (Ig) or antibodies (Ab), composed of two identical heavy chains and two identical light chains (variable domain encoded by rearranged V, D, J genes). The activated B-cells undergo somatic hypermutation of their Ig genes and IGH class isotype switching (e.g. isotype IgM to IgG1) to generate sequence variants that raise an accurate immune response against the pathogen. B-cells memorize the characteristics of the pathogens previously encountered and can differentiate into antibody-secreting plasma cells on re-encounter.

With the advancement of the next-generation technologies we can capture the sequence of the antigen binding arm (Fab) of these antibodies. To understand the response of the immune system to an antigen we record the antibody sequences raised against a specific antigen (or groups of antigens) at several time-points after antigen encounter. Each sequence (accurate antibody Fab) is called a B cell-clone that

is adapting to the antigen in time to eliminate it. We often cluster these clones based on some rules to generate a graph (clonal lineage in time) to understand the evolution of the antibody in response to antigen. Defining clusters of clonal groups is challenging in terms of defining a threshold for grouping. The current grouping parameters are based on the V- and J- gene assignment, CDR3 length, and CDR3 distances. There are several limitations associated with this type of grouping parameters being:

1. D gene-segment is not considered. Note: D gene-segment is very small (10-15 nt) and it is very difficult to identify in the full sequence.
2. Sequencing errors are considered as SHMs. Hence, they are considered as a separate clone in the dynamics study. Note: Generally, Illumina based sequencing is popular in these studies that introduces several errors (known and unknown) which are difficult to identify.
3. Wrong assignment of the germline alleles to the clones. Note: Several V gene-segments are highly identical and is often mis-identified.

We want to generate a workflow where these issues could be resolved while clustering the clones and accurate lineage dynamic trees (graphs) could be generated. Further, the scores can be assigned to the clones and the clonal groups based on rules (obtained from sequence information). We will use the publicly available data to generate this workflow which could be implemented directly in the Change-O-Repertoire clonal assignment toolkit. This scoring method for identification of the accurate clones will be used to identify the accurate responses in vaccination settings or infection models.

Contact

If you want to apply, please send your English CV to dr.Indu Khatri (i.khatri@lumc.nl).
You can also contact her for more information.