**HLA\_emerson.xlsx**

|  |  |
| --- | --- |
| Column | Description |
| Inferred\_format | Name of study/dataset |
| n\_sequences | Number of sequences in input |
| n\_nan | Number of sequences that are nan (not a number) |
| n\_non\_productive\_vj | Non-productive sequence |
| n\_invalid\_junction\_aa | Invalid CD3-region |
| n\_duplicates\_merged |  |
| n\_clonotypes\_after\_filtering | Clonotype is combination of V-CDR-J |
| n\_TRB | Number of clonotypes that are beta chain |
| patient\_id | Patient’s ID |
| repertoire\_id | One sequence run |
| metadata\_tags | Metadata of patient |
| HLA-A | HLA-type of HLA-class I-genes |
| HLA-B | HLA-type of HLA-class I-genes |
| HLA-C | HLA-type of HLA-class I-genes |
| HLA-DQA1 | HLA-type of HLA-class II-genes (A = alpha chain of gene,  1 = subcategory or variant of the DQA gene) |
| HLA-DQB1 | HLA-type of HLA-class II-genes (B = beta chain of gene,  1 = subcategory or variant of the DQB gene) |
| HLA-DRB1 | HLA-type of HLA-class II-genes (B = beta chain of gene,  1 = subcategory or variant of the DQB gene) |
|  |  |

**Questions:**

1. Moet ik alle kolommen gebruiken? Zo niet, welke kolommen zijn belangrijk?
   * Wel gebruiken: HLA-kolommen, repertoire\_id
2. ...

**P00001\_clonotypes.xlsx**

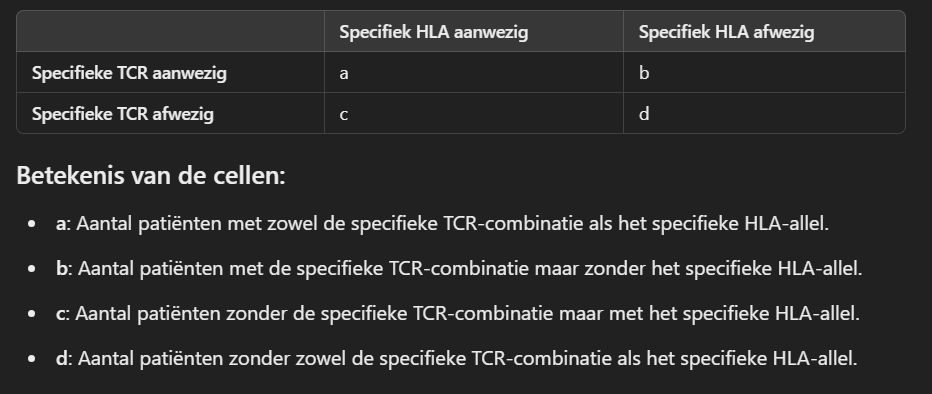
|  |  |
| --- | --- |
| Column | Description |
| junction | Nucleotide sequence |
| v\_call | Type of V-gen |
| junction\_aa | Aminozuur code voor junction |
| j\_call | Type of J-gen |
| duplicate\_count | Duplicates of TCR |
| repertoire\_id |  |
|  |  |
|  |  |

**Questions:**

1. Moet ik alle kolommen gebruiken? Zo niet, welke kolommen zijn belangrijk?
   * Gebruiken: v\_call, junction\_aa, j\_call
2. Waarom wordt repertoire\_id gebruikt (ipv patient\_id)?
   * …
3. ….

## Fisher exact method:

Voer de fisher-exact methode uit voor elke mogelijk combinatie van v\_call, junction\_aa, j\_call voor het allel **HLA A02:01**



* **P-waarde**: De Fisher's Exact Test berekent een p-waarde, die aangeeft hoe waarschijnlijk het is om de geobserveerde verdeling (of een meer extreme) te zien als de nulhypothese waar is. Een lage p-waarde (bijvoorbeeld < 0.05) geeft aan dat het verband waarschijnlijk niet toevallig is en dat er een associatie bestaat tussen de variabelen.
* De **odds ratio** (OR) in de context van de Fisher's Exact Test meet de sterkte van de associatie tussen twee categorische variabelen. Het vertelt je hoeveel groter (of kleiner) de kans is op een bepaalde uitkomst in de ene groep ten opzichte van de andere groep.
  + OR = 1: Er is geen verschil in kans tussen de groepen; de uitkomst is even waarschijnlijk in beide groepen.
  + OR > 1: De gebeurtenis is waarschijnlijker in de eerste groep dan in de tweede groep. Bijvoorbeeld, een OR van 2 betekent dat de kans op de gebeurtenis in de eerste groep twee keer zo groot is als in de tweede groep.
  + OR < 1: De gebeurtenis is minder waarschijnlijk in de eerste groep dan in de tweede groep. Bijvoorbeeld, een OR van 0.5 betekent dat de kans op de gebeurtenis in de eerste groep de helft is van die in de tweede groep.