VaxG:

An Automated Pipeline for Multi-Epitope Vaccine Development

The development of multi-epitope vaccines in silico has revolutionized how we generate new immunogen possibilities. Here we introduce VaxG, a software that follows a pipeline based on leading works published in the literature.

1. USAGE

• Prediction

First, the program predicts epitopes (the smallest immunogenic part of a protein) that have affinity with human MHC I and MHC II receptors. This prediction is carried out using two different methods for increased result reliability.

The program utilizes tools available on the IEDB platform, such as NETMHCII, nn_align, NetMHCpan 4.1, and others, to predict the epitopes. The program always uses the IEDB's recommended method and a user-chosen alternative method. We recommend using NETMHCcons and nn_align for MHC I and MHC II, respectively, as they more easily adapt to the pipeline. Lastly, for predicting 16-amino acid B cell epitopes, the tool ABCpred is used with a default cutoff of 0.51.

• Filtering

After the 5 predictions, the MHC I and MHC II epitopes (from the 4 predictors) are filtered based on IC50 values and percentile ranks. Typically, values below 50 and 1, respectively, are ideal for identifying epitopes with high potential to trigger immune responses.

• Overlapping

With the filtered epitopes in hand, the data generated from MHC I tools and MHC II, are overlapped to identify common epitopes. In summary, we now have a set of MHC I epitopes and a set of MHC II epitopes. These data are then overlapped with B cell epitopes. MHC I epitopes typically have less than 10 amino acids, MHC II epitopes have more than 11, and B cell epitopes have 16 amino acids. Only MHC I and MHC II epitopes that are contained within the 16-amino acid epitopes are retained.

• Final Filtration

MHC I epitopes are evaluated using the MHCI immunogenicity tool, and those with a score higher than 0.1 are retained. Finally, both properly filtered MHCIxBcell and MHCIIxBcell epitopes are evaluated for their similarities.

Epitopes that differ by only one or two amino acids are excluded using the IEDB cluster analysis tool. For example:

ABCDEFGHI - epitope 1

BCDEFGHIJ - epitope 2

CDEFGHIJK - epitope 3

Only epitope 1, for instance, will be retained, and the others will be removed because they are considered virtually the same, leading to a potentially redundant immune response. Remember, the strength of multi-epitope vaccines lies in presenting different facets of the pathogen to the immune system! With all final epitopes in hand, we proceed to protein assembly.

• Protein Recombination:

If you have chosen an adjuvant to accompany your vaccine, remember, it must be a flat sequence of amino acids and should end with the sequence EAAAK (a linker).

When entering it, make sure to include the EAAAK sequence at the end!

The protein will be assembled as follows:

Adjuvant + EAAAK + MHCII Epitope 1 + GPGPG + MHCII Epitope 2 + ... + AAY + MHCI Epitope 1 + AAY + MHCI Epitope 2

These linkers are the standard choice based on the literature.

An ongoing process is the possibility of altering the order of epitopes in the protein to enhance immune response generation. Therefore, in the "chimeric proteins" option, you can choose how many times you want to permute the MHC I and MHC II epitopes while maintaining the structure order (Adj + MHCII + MHCI). The best sequence will be chosen by the Vaxijen software, which predicts protein antigenicity. The scores are saved in a .csv file.

• Qualitative evaluations

After the protein is assembled, the tools Allertop, Protparam, and PSIPRED will be used to predict allergenicity, physicochemical parameters, and generate a 2D model of the vaccine structure.

Allertop will generate a file with the information (allergenic or non-allergenic), Protparam provides various information such as isoelectric point and estimated protein half-life. Finally, the PSIPRED file contains the link leading to the PSIPRED result.

• Graphical Outputs

Upon completion, the program generates two images and a CSV table.

CSV Table (output.csv): The table follows the order of proteins in the FASTA file. That is, the first protein in the input file is target number 1, and so on. In this output.csv file, you will find the quantity of epitopes per vaccine target at each stage of the pipeline.

First Image (sequence.png): This graphical model shows the order of epitopes within the protein, visually representing their distribution along the protein sequence.

Second Image (flowchart.png): This is a flowchart diagram providing the following information:

- **MHCIrec and MHCIsec:** Total epitope values predicted by the recommended IEDB tool and the alternative tool for MHCI.
- **MHCIIrec and MHCIIsec:** Same as above for MHCII.

•**MHCIrec/sec and MHCIIrec/sec filtered**: Data filtered by previously chosen IC50 and percentile rank values.

•**MHCI over and MHCII over**: Quantity of epitopes predicted in common between MHCI tools and between MHCII tools.

•**MHCIxBc and MHCIIxBc:** Quantity of MHCI and MHCII epitopes contained within B cell epitopes (abcpred).

•**MHCI and MHCII final:** Final epitopes already filtered by cluster analysis and MHCI immunogenicity assessments.

2. INSTALATION

First you need to install the docker engine on your computer, whether it's Linux, Windows or Mac. Here's the installation link: https://docs.docker.com/engine/install/ .

After installing correctly, open the command prompt or terminal. To open the command prompt in Windows, you can use the keyboard shortcut Win + R to open the "Run" window, then type "cmd" and press Enter. At the command prompt, use the following command:

- ***docker pull openheimerfinger/vaxg***

And then:

- ***docker run -d -p 8000:8000 openheimerfinger/vaxg***

After that, just go to the following link in your browser:

- ***http://localhost:8000/***

3. Getting outputs

In addition to the results automatically going to the email entered in the input process, you can manually remove the results from the container using the following commands in the termial/prompt:

(First check the name of your created container, look in the NAMES header☺

- ***Docker ps***

```
C:\Users\Usuario\Downloads>docker ps
CONTAINER ID   IMAGE                      COMMAND                  CREATED          STATUS          PORTS                    NAMES
de7ffe699b0d   openheimerfinger/vaxg:v13  "python manage.py ru…"   25 minutes ago   Up 25 minutes   0.0.0.0:8000->8000/tcp   blissful_leavitt
```

In this case, the name of the container has been declared as blissful_leavitt, so to manually retrieve its results, you can run it in the terminal/prompt:

- ***Docker cp blissful_leavitt:/app/Final.zip Final.zip***

The file will be copied to the root folder where your terminal/prompt was run.

In the final.zip file there are no other intermediate results that have been generated, such as the epitope files for proteins before the MHCi immunogenicity and clustering analysis.

Therefore, if your proteins have few epitopes and you want to relax the filtering process without performing clustering or MHCI immunogenicity, you can retrieve the predicted/filtered/overlapped epitopes for MHCI and MHCII per protein and use them to build your vaccine manually. To do this, check the name of the files generated in the folder by running:

- ***Docker exec –it blissful_leavitt ls***

 # replace blissful_leavitt with the name of your container

Then use the docker cp command again to retrieve the files you want.

The intermediate files have the following name pattern:

MHCI_Epitopes_1.txt and

MHCII_Epitopes_1.txt

The number at the end of the name represents the order of the protein according to the fasta file. Therefore, the file ending in 1 is associated with the first protein in the input fasta file.

If you want to see the contents inside the files before extract them, run:

Docker exec –it blissful_leavitt cat name_of_the_file.txt

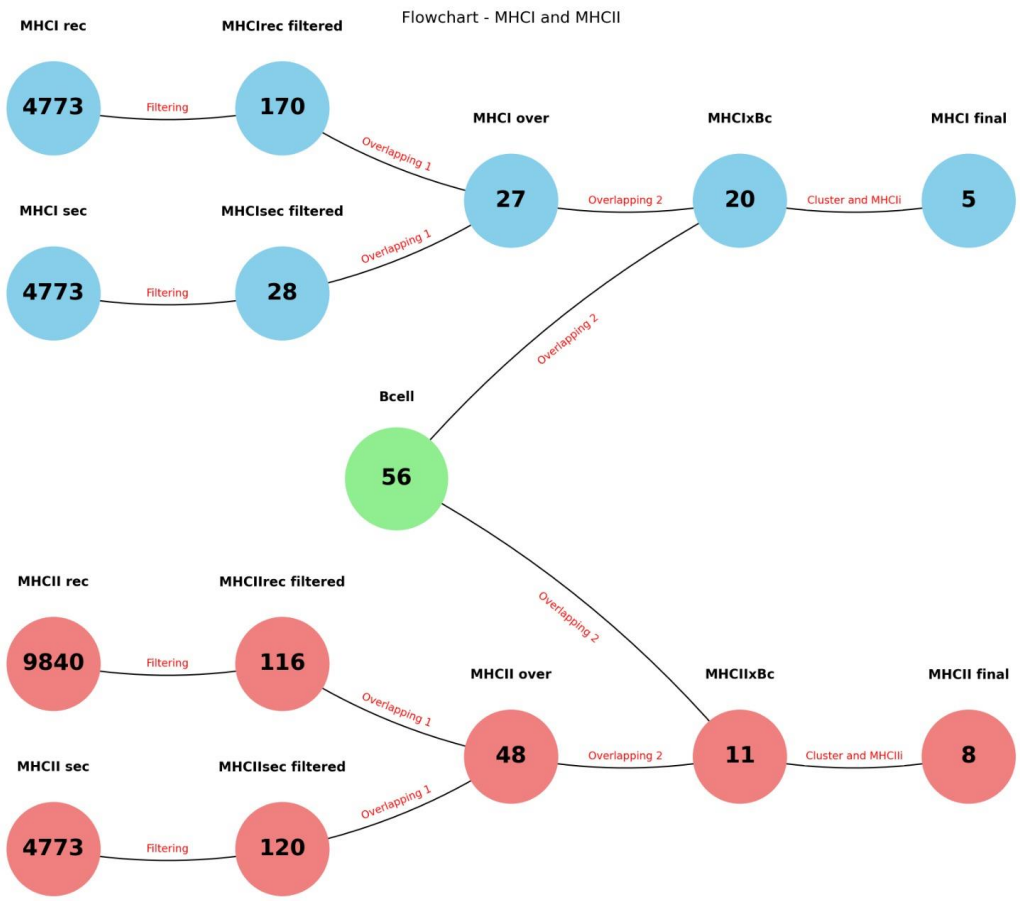
1. Outputs examples

1.

Targets	MHCI prec	MHCI prec	MHCII pre	MHCII pre	Bcell	MHCI fil1	MHCI fil2	MHCII fil1	MHCII fil2	MHCI ovei	MHCII ove	MHCI over2	MHCII over2
1	2566	2415	1884	2415	15	87	11	47	35	11	21	7	4
2	4691	4415	3509	4415	25	166	17	102	51	17	23	12	4
3	815	767	545	767	2	27	5	0	3	5	0	0	0

2.

ADJUVANT
GILSKYAKSLSFSVN
NFWQNFKLLANPTKT
SKIKRICVFKNKRIK
LHEIDFIILDGKKIY
GSIRYFVINGKKTRL
CMILVGLTSFAYAKN
GLTSFAYAKNSVTAT
KQKIINHSNPLKKLL
LLRNTTFFV
TTFFVNHLR
KIKRICVFK
SVTATFIGK
SLAESEFAI

MHCII epitope
MHCI epitope

3. Resume of the filtering process until the final epitopes