

COVID-BIOCHIP: A Web Tool to Analyse COVID-19 Antigen Microarrays

Aurora SUCRE^{a,b,1}, Isabel AMAYA-RODRIGUEZ^a, Hee-Sool RHO^c, Sonia SERNA^c,
Niels REICHARDT^c, Alba GARIN-MUGA^{a,b} and Gorka EPELDE^{a,b}

^a*Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), 20009 Donostia-San Sebastián, Spain*

^b*Biodonostia Health Research Institute, Basque Research and Technology Alliance (BRTA), 20014 Donostia-San Sebastián, Spain*

^c*CIC BiomaGUNE, Basque Research and Technology Alliance (BRTA), 20009 Donostia-San Sebastián, Spain*

Abstract. During the COVID-19 pandemic, there was a growing need to characterise the disease. A very important aspect is the ability to measure the immunisation extent, which can be achieved using antigen microarrays that quantitatively measure the presence of COVID-related antibodies. A significant limitation for these tests was the complexity of manually analysing the results, and the limited availability of software for its analysis. In this paper, we describe the development of COVID-BIOCHIP, an ad-hoc web-based solution for the automatic analysis and visualisation of COVID-19 antigen microarray data results.

Keywords. COVID-19, microarrays, antigens, web development, data analysis

1. Introduction

An important aspect to characterise the COVID-19 disease spread and foresee the impact of restriction measures relaxation or the emergence of new variants, is the ability to measure the extent of immunisation within the population, usually attempted using serological tests. These tests only give a binary response regarding the presence of certain antibodies, which is not enough to fully define immunisation profiles and predict long-term effects [1]. Nevertheless, there are other approaches that better describe immunisation, but in general, they are more expensive and difficult to interpret.

In response to these limitations, during the first months of COVID-19 pandemic, a consortium of Basque companies decided to embark on a project to develop a system for the characterisation of immunisation profiles, that considers multiple COVID-related antigens and antibodies, while remaining affordable and enabling its implementation in real conditions. The project's goals were to develop a new type of antigen microarray, define the analytic protocol to employ these arrays on immunisation characterisation, and develop a web solution to automate bulk data analysis. This software overcomes one of the biggest limitations associated to microarray data: the analysis, when performed manually, is very time-consuming and cumbersome. Moreover, there was limited availability of public software for automatic analysis of microarray data, and those that

¹ Corresponding Author, Aurora Sucre, Vicomtech Foundation; E-mail: amsucre@vicomtech.org

were found, were not compatible with the arrays that were going to be developed. Therefore, there was a need to develop an ad-hoc solution. In this paper, the development of a web tool for the automatic analysis of COVID-19 antigen arrays is described.

2. Methods

Due to the nature of the COVID-19 pandemic and the need of developing a comprehensive solution in the shortest possible time, different stages of the project, that in other conditions would happen one after the other one, had to be executed in parallel. The antigen microarrays, the analytical webtool and other elements of the project were developed simultaneously. To overcome time limitations and to be able to develop a user-friendly application, that considers all the requirements that arose during development, a user-centred design methodology [2] was followed, where end-users were involved in all stages of the decision-making process and constantly provided their insight so the tool could be adapted to their needs and preferences.

Finally, the web is based on Python. Specifically, the Django-Plotly-Dash framework [3] was employed, allowing for the rapid development of visually rich web solutions that exploit Django [4] strengths. Dash [5] itself is oriented to single-page visual dashboards with limited processing capabilities, but combining its visualisation functionalities with Django utilities in the backend, allows the development of complex multi-page web applications with an efficient Django backend. Also, the PostgreSQL database management system was used to develop the solution's persistence, being handled from the Django server. Lastly, the different components were packed in independent Docker [6] images, and orchestrated with Docker-Compose for deployment.

3. Results

The implemented user-centred methodology and the chosen web framework allowed us to build an intuitive solution for the automatic analysis of COVID-19 antigen microarrays data in a short period of time, meeting the expectations of the project.

Regarding the workflow, the web application receives raw CSV files generated in the microarray scanner, processes them to extract and calculate relevant data, and generates different plots that allow researchers to assess the immunisation profile of the patients to know in which stage of the disease (or recovery) they are, and, for instance, study for how long immunisation lasts based on the type of generated antibodies.

The solution's functionalities can be sorted in three different categories which will be described in detail next. The different web views developed are shown in **Figure 1**.

3.1. Study Management

During initial stages of development, we discovered that the characteristics of the input data may vary along time, and therefore it was crucial that the solution was flexible enough to read and process input data having different formats.

The management module allows users to intuitively define different array and analysis configurations, so input data can be processed correctly. Regarding the input data, users can define how many samples are studied, how many antigens and replicates

are included in the array and which antibodies are analysed. Regarding processing, users can define which parameter to extract from the files, which statistics to calculate and the affinity thresholds to classify the results. The configurations are stored and can be reused.



Figure 1. Web views: A) log-in, B) option selection, C) study configuration, D) data upload, E) study search, F) output - plots, G) output – tables

3.2. Data analysis

For each new analysis, users can define new configurations (if needed), edit, or use any of the existing configurations. Once the configuration is chosen, users can upload the required CSVs and define the names of the samples that were included. The required files and information vary based on the chosen configuration.

Once the data is uploaded, the system processes each file to extract the required value for each antigen replicate, antibody and sample. Then, statistics are calculated between replicates of the same antigen and results are classified based on predefined thresholds regarding the antigen-antibody affinity strength. Every result is stored in the database, sorted by sample. Users may access previous results using the “search” page.

3.3. Data visualisation

The results of individual samples can be visualised in summary tables or bar plots. The summary table shows antigens on rows and antibodies on columns. The plot view shows a different bar plot for each antibody, where each bar corresponds to a different antigen. Bar and table values are calculated for each set of replicates.

In both views users can choose which sample to display and which antigens to include. In the plot view, users may also (1) modify the colour of every chart at once, (2) modify the colour based on the affinity strength and (3) modify the colour of individual bars to highlight those that are more relevant to the user on each specific analysis.

Finally, users are allowed to download the tables and plots in different formats, so that the information can be embedded in their reports, publications or wherever they need.

4. Discussion

The user-centred-design approach that was followed and the chosen development framework allowed for an efficient and rapid development of an intuitive and high-performing solution that filled the gaps found on the context of microarray data analysis.

The COVID-19 BIOCHIP web solution allows users to automatically process large amounts of data in a short period of time, by only defining certain configuration parameters and loading the raw data. The generated results are shown in a very simple manner that help users on quickly detecting common patterns in the data and identifying the most relevant information. Also, the ability of storing results and accessing them when needed, help users on studying the progression of immunity in patients along time, as they are able to compare the analytical results from samples collected in different moments and observe how the antibodies quantification varies.

The software has been tested in specific conditions and has return promising results. Nevertheless, in-depth validation will be performed by a larger group of experts that will mainly evaluate two aspects: software usability and efficiency. During this validation stage, users will also test the solution using non-covid-related microarray data to verify that the tool is as flexible as expected and that can be exploited within other contexts.

Regarding additional future work, the idea is to improve the visualisation functionalities, including new type of charts that provide further insight on the data, and that ease the comparison between different samples enabling, for instance, the joint exploitation of different patients' data for deriving public health insights.

Acknowledgements

The COVID-19 BIOCHIP project was funded by the Basque Government in the framework of special grants oriented to combat COVID-19 in the year 2020.

References

- [1] Krajewski R, Gołębiowska J, Makuch S, Mazur G, Agrawal S. Update on serologic testing in COVID-19. *Clin Chim Acta*. 2020 Nov; 510: 746-750
- [2] Luna D, Quispe M, Gonzalez Z, Alemraes A, Risk M, Garcia M, Otero C. User-centered design to develop clinical applications. Literature review. *Stud Health Technol Inform*. 2015; 216: 967-967
- [3] Gibbs Consulting. Django-Plotly-Dash. 2020. Available from: django-plotly-dash.readthedocs.io
- [4] Django Software Foundation. Django. 2020. Available from: djangoproject.com
- [5] Plotly Technologies Inc. Dash. 2020. Available from: plotly.com/dash
- [6] Merkel D. Docker: lightweight Linux containers for consistent development and deployment. *Linux journal*. 2014; 2014(239): 2.