Team Name: DiCAIDS *{DIgital and Computational Approaches to Infectious Diseases Study}*

Challenge: Develop a platform for the discovery and analysis of correlates of protection

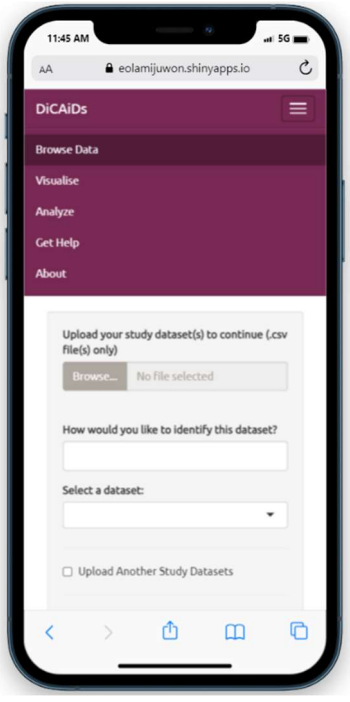
Team Members:

- Emmanuel Olamijuwon (EO), Lecturer, University of St Andrews
- Adeniyi Francis Fagbamigbe (AFF), Senior Lecturer, University of Aberdeen
- Kehinde Aruleba (KA), Lecturer, University of Leicester

Background work

Understanding the correlates of protection and their use as surrogate endpoints can accelerate vaccine development pipelines by informing the decision-making process about vaccine efficacy[1]. Although a few platforms, such as ImmuneSpace and ImmPort Galaxy, exist for exploring the correlates of protection, a lot more could be done. Novel computational methods such as machine learning and reinforcement learning, among others, could generate new insights into the correlates of protection. In many high-income countries (HICs), these methods have been leveraged effectively to identify the correlates of protection from various infectious diseases, such as influenza and hepatitis, among others[2,3].

Despite the huge burden of infectious diseases and recent advances in vaccine research and development in low-and-middle-income countries (LMICs), very limited research has examined the correlates of protection in these countries. This is partly due to the lack of data-science technical skills in most of these countries and the limited computational resource for fitting these models. In addition, we have identified from previous workshops led by our team that the lack of confidence in data science skills poses one of the greatest barriers to data-sharing practices among researchers in LMICS[4]. A no-code user-friendly computational platform thus offers many exciting opportunities to empower researchers in LMICs to identify the correlates of protection and take the computational burden off researchers in HICs. Achieving these objectives itself can increase the diversity/range of evidence available about vaccine efficacy and ultimately contribute to reducing the escalating impact of infectious diseases.



Prototype

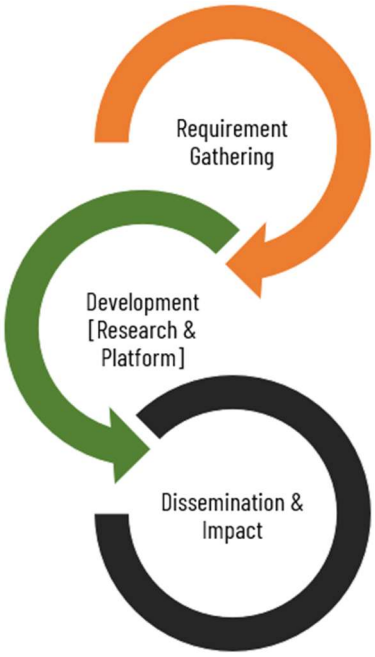
Our prototype - The DiCAiDS Discovery platform (see attached screenshot), built on the R-shiny framework, provides a user-friendly, click-and-drop interface for aggregating connected datasets, visualising patterns in the dataset and identifying potential correlates of protection. The app has five distinctive panels:

- *The Browse Panel* lets users upload, aggregate, process, and explore single/multiple connected *(unconnected in the future*) study datasets.
- *The Visualise Panel* lets users explore simple descriptive patterns in the datasets, including the prevalence of infection and antibody seroconversion and the geometric mean of antibody titers, among others. Users can also select which assay result and virus strain to visualise.

- *The Analyse Panel* lets users apply classical statistical or machine learning methods to identify immune correlates of protection. We also created three separate sub-panels, building on the framework for assessing the immunological correlates of protection in vaccine trials by Qin *et al.*[5].
- The *Help and About* panels under development are intended to provide additional resources (transparent metadata) about statistical models and assumptions underpinning all analyses covered on the platform. We will also use this panel to continually engage with the health data science community and other relevant stakeholders to refine the platform, our models, assumptions, and strategies.

Plan for future work

Over the next 12 months, we are committed to delivering a robust, user-friendly discovery platform, focusing on three distinct yet interrelated activities.



- ***Requirement gathering (EO)***

  In the first three months of our project, we will leverage our existing network and Wellcome's to coordinate a preliminary meeting with key domain experts and stakeholders, including biostatisticians, data scientists, immunologists, and virologists, among others. Officials from the Department of Health and other end users across LMICS will also be invited to these early meetings. We will meet with these teams bi-weekly to gather important information about what makes a great discovery platform for them and understand their unique needs while considering context-specific regulations. We will also schedule meetings with the Wellcome team and our selected team of experts to understand the structure of existing data sources—how and if they need to be standardised before use on the platform. We believe that these co-creation strategies with the partner have the potential to create value in the platform very early and motivate its sustained use.

- ***Development [Research (KA) & Platform (EO)]***

  Having gathered sufficient data, we will iteratively co-develop our discovery platform with our identified team of experts. The platform will offer a no-code, click-and-drop interface that can be run on the least computationally intensive digital device. As a result, we will take the computational and disease modelling burdens off researchers and regulatory authorities so that they can focus more on generating novel and relevant actionable insights for evaluating vaccine efficacy from their data. We will gather feedback from the team and refine the interface iteratively throughout the development process. Furthermore, our discovery platform will provide both classical and machine learning approaches to identify correlates of protection. A unique advantage of ML models is that they have the potential to uncover nonlinear patterns, interactions between multiple variables, and other complex dependencies, leading to a deeper understanding of the immune response and vaccine efficacy.

  During our development phase, we will co-create research on the application of ML methods to identify the correlates of protection from major infectious diseases decided by our [expert] team. This strategy

will achieve two key objectives. On the one hand, working with our interdisciplinary expert teams of biostatisticians and those from other relevant fields will help translate platform findings into actionable insights that align with regulatory expectations for the different countries. On the other hand, the co-creation of research using the proposed platform will build incremental capacity and appropriate skill sets for the country teams to use the platform and ease its use. Throughout, we will encourage a healthy, inclusive and engaged open-science community characterised by mutual respect, recognise diversity, and reward every unique contribution of our team (including country experts and regulatory authorities). We believe that our planned equitable research practice will enhance our models' transparency, acceptability, and eventual use to strengthen a research and development ecosystem that can deliver solutions that make a meaningful impact on individuals, communities, and countries.

- ***Dissemination and Impact (AFF)***

  As part of our dissemination and impact strategy, we will leverage our existing global network of researchers across 15+ countries, mostly LMIC and Wellcome's existing networks, to coordinate a series of virtual workshops in the last three months of our project. The workshops will focus on building capacities for using the discovery platform and training the next generation of health data scientists in LMICs and worldwide to use and interrogate this platform. We will co-produce at least two research articles with researchers (and communities involved in the project) from LMICs in world-leading journals. One of the articles will focus on the application of ML methods in identifying the correlates of protection. At the same time, the other will be software development focused, describing the discovery platform and its functionalities.

Overall, we will draw on our team's unique expertise across computer science (KA), computational social sciences/demography (EO), and global public health, health data science and biostatistics (AFF). We have applied computational methods on various issues at the intersection of data science, society, and population health. We are passionate and motivated to support LMICs to adequately prepare and respond to current and future social and health challenges. This is evident in our various capacity development initiatives (SICSS-Accra, SICSS-Nigeria, MD4SG, NeurIPS Affinity Chair) and links to research institutions/universities across Africa and beyond. As a result, this project contributes to our overarching goal of strengthening operational capacity for pandemic preparedness and response (decision-making) across LMICs, which aligns perfectly with Wellcome's strategy.

References

1. Tomaras, G. D. & Plotkin, S. A. Complex immune correlates of protection in HIV-1 vaccine efficacy trials. *Immunol. Rev.* **275**, 245–261 (2017).
2. Arevalillo, J. M., Sztein, M. B., Kotloff, K. L., Levine, M. M. & Simon, J. K. Identification of immune correlates of protection in Shigella infection by application of machine learning. *J. Biomed. Inform.* **74**, 1–9 (2017).
3. Choi, I. *et al.* Machine Learning Methods Enable Predictive Modeling of Antibody Feature:Function Relationships in RV144 Vaccinees. *PLOS Comput. Biol.* **11**, e1004185 (2015).
4. INASP. Using AI to reduce knowledge inequity – a transdisciplinary approach. *INASP Blog* https://blog.inasp.info/using-ai-to-reduce-knowledge-inequity-a-transdisciplinary-approach/ (2022).
5. Qin, L., Gilbert, P. B., Corey, L., McElrath, M. J. & Self, S. G. A Framework for Assessing Immunological Correlates of Protection in Vaccine Trials. *J. Infect. Dis.* **196**, 1304–1312 (2007

Salaries

Name	Justification	Period on Projects(months)	%time	Total
Dr E. Olamijuwon	PI - supervising the project, cordinating teams and contributing to platform developement.	12 months.	20%	£0.00
Dr K. Aruleba	Co-I - lead algorithm development and supervise research fellow	12 months.	15%	£0.00
Dr A. Fagbamigbe	Co-I - lead dissemination and impact strategy. Liaise with biostatistics and teams in LMICs.	12 months.	15%	£0.00
Research Fellow	Post-doc -bring expertise in immunology to the team and lead the development of the discovery platform. Integrating feedback from expert teams.	12 months.	80%	£61,424.00
subtotal				£61,424.00

Materials & Consumables

Description	Justification	Period	Total
Server - Standard_D16pds_v5 16 vCPUs 64GB Memory Linux OS	Robust and secure cloud server for running statistical models and data visualisations. Registeration will cover 24 months period so that there is sufficient time to transition between fundings for sustained availability of the platform.	24 months	£7,812.00
Domain registration	Domain registeration will cover 24 months period so that there is sufficient time to transition from this funding to a new funding for sustained availability of the platform	24 months	£98.00
subtotal			£7,910.00

Equipment

Description	Justification	Total
1 Unit HP ENVY x360 15.6" 2 in 1 Laptop - AMD Ryzen™ 7, 512 GB SSD	A high-performance laptop for research fellow use.	£950.00
1 unit of HP M27fwa Full HD 27" IPS LCD Monitor	An additional screen to support programming and research.	£250.76
subtotal		£1,200.76

Access Charges

Description	Justification	Total
subtotal		£0.00

Travel & Subsistence

Description	Justification			Total
Team meetings	Our team will meet quarterly to evaluate project progress and discuss strategies for moving forward.	There will be 4 physical meetings beside weekly virtual meetings		
Train/Flight		3 times, 4 people @ £200		£2,400.00
Hotel Accomodation		3 times, 6 nights, 4 people @ £100		£7,200.00
Taxis		3 time, 4 ways, 4 people @£20		£960.00
Virtual recuring meeting with LMIC stakeholders		Virtual		£0.00
Meeting with LMIC Stakeholders	These will include dissemination meetings as well as early-road testing of the platform. Travels for all members of the team are quoted. However, not all team members will attend the same meeting/workshop.			
Flight		2 times, 4 people @ £925		£7,400.00
Hotel Accomodation		2 times, 6 nights, 4 people @ £100		£4,800.00
Airport Taxis		2 time, 4 ways, 4 people @£30		£960.00
Virtual meeting with Wellcome Trust		Virtual		£0.00
subtotal				£23,720.00

Miscellaneous

Description	Justification			Total
Open Access Publiction		2 Publications @ £2400 each		£4,800.00
Webinar	Free			
subtotal				£4,800.00

Summary of Costs Requested

Description	Total
Salaries	£61,424.00
Materials & Consumables	£7,910.00
Equipment	£1,200.76
Access Charges	£0.00
Travel & Subsistence	£23,720.00
Miscellaneous	£4,800.00
Grand Total	£99,054.76