Project3: Pathogen antigen discovery and immune profiling
Lead supervisor: Nicholas Croucher (Sanger)
Co-supervisors: John Lees (EMBL-EBI), Ewan Harrison (Sanger)

Project description:
Vaccine and immunotherapy development efforts for infectious diseases need to find pathogen targets that elicit a strong and lasting immune response, cover virulent strains within diverse pathogen populations, and are not susceptible to immune escape. Most existing pharmaceutical interventions treat pathogens as a simple static target, while they are actually both diverse and dynamic, under a strong pressure to evade immunity. Deep learning algorithms and high-throughput functional and surveillance genomics can now be integrated to enable more sophisticated approaches to targeting microbial populations.

Measurement of host immune responses due to prior infection, commensal exposure and immunisation can be used to understand whether specific pathogen antigens are immunogenic (Croucher *et al* 2024), and with sufficient data may enable prediction of whether an individual will be protected from future infection (Plotkin 2023). The recent development of cheaper and more scalable immune profiling assays, in particular PhIP-seq (Huang *et al* 2024), provides an opportunity to measure human antibody response strength to multiple pathogen antigens in a time-resolved manner.

The development of new approaches for preventing and treating infectious diseases therefore requires an integrated understanding of both pathogen diversity and host immune responses. This project will develop population-genomic methods to prioritise pathogen antigens, and use new sequencing-based screens to measure host immune responses to these antigens over time.

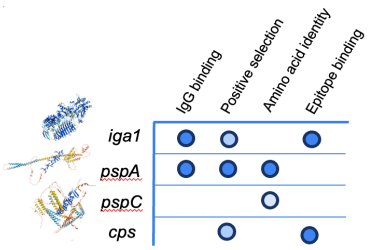
The project will develop from our combined expertise in pathogen evolution, immune array screening, host-pathogen interactions, and mathematical modelling of immune dynamics.

Goals:
There are currently extensive pan-species genomic datasets hosted at EMBL-EBI and Sanger for both bacteria (AllTheBacteria) and viruses (EBI Pathogens Portal). These will enable this project to take a pathogen-agnostic approach to analysing virome- and microbiome-wide PhIP-seq data to characterise the immunology of human-pathogen interactions across the full adaptive immune system response. This will subsequently drive more detailed analyses of well-characterised priority pathogens, in which the immunobiology is particularly complex or of urgent importance to ongoing vaccine development (e.g. *Staphylococcus aureus*, *Streptococcus pneumoniae*, Enterobacterales including *Escherichia coli and Klebsiella pneumoniae*, Respiratory syncytial virus [RSV] and Parainfluenza 3).

On each of these pathogens, the ESPOD fellow will develop a high-throughput computational screen for protein antigens:

1. Compile and curate large high quality genomic datasets, drawing from public datasets such as AllTheBacteria (Lees group), and targeted population sequencing cohorts with rich metadata (Croucher and Harrison groups).

2. Systematically identify the core and accessory proteins (panproteomes) from these datasets. The fellow will use and extend highly-scalable computational methods to cluster proteins across the population, using genomic context to resolve orthologues and paralogues, which is particularly important for highly-variable or multi-copy antigens.
3. Use structural prediction deep learning algorithms to model host-pathogen interactions between these antigens and adaptive immune factors, identifying which surface-exposed residues are likely to form epitopes (structures recognised by antibodies).
4. Use population genomic signatures of diversifying selection (e.g. dN/dS, diversity statistics, site frequency spectra) to identify loci that are likely to be important in evading adaptive immune responses .
5. Use existing virome and microbiome multi-pathogen PhIP-seq datasets (e.g. Mina *et al* 2019, Vogl *et al* 2021, Shrewsbury *et al* 2024) to identify which structural and evolutionary signals are the most predictive of antigenic proteins, and develop a method for identifying such epitopes using population genomic data.
6. Design of proteome panels for screening with PhIP-Seq and protein microarrays that focus on the most informative immune responses to common viral and bacterial pathogens. These can form the basis of future PhIP-seq experiments at Sanger, with the aim of characterising immune responses across large cohorts.



The fellow will create a lightweight resource to display and share the results of their analysis (figure, left). This will enable their evidence synthesis approach to be modified, interrogated and updated as new results are generated. It will also help develop connections with other researchers to plan which targets to take forward.

Training and development:
The fellow will be trained in the design and application of methods which can be scaled to datasets of millions of genomes, and extract specific biological information. By working with the expertise at EMBL-EBI, they will develop skills and resources to present biological data and maximise its reuse. Combining this with Sanger's deep expertise in experimental design, access to cohorts and samples, novel sequencing screens, and antigen array design and validation will put the fellow in a strong position to work on future vaccine and immunotherapeutics development problems at the intersection of host and pathogen, using sequencing to validate computational predictions. The fellow will also strongly benefit from the combined pathogen domain expertise of all the involved groups.

The project is designed to have strong links to pharmaceutical research and development, and we anticipate engaging with industry to further develop this work into an application to the Open Targets funding scheme, or a competitive funding application targeted at translational awards, or a further fellowship for the ESPOD fellow. This initial project is suitable for extension through including different cohorts of serum donors, and incorporation of new multi-omics approaches. By the end of the project, the fellow will have developed a target prioritisation method and large scale genomics and data resource, and the means to perform initial validation and immune profiling with these targets. This would put them in a strong place for future careers to further develop this work in either academia or industry.