ebr
European Biopharmaceutical Review
JULY 2021

IT'S A KNOCK-OUT

Gene knock-outs are powering a renaissance of CRISPR genome editing technology

FLESH IS BEST

Can preclinical testing using human living tissue win the fight against clinical attrition?

TAKE MY BREATH AWAY

Inexhaustible by nature, breath testing is set to play a vital role in precision medicine

Spearheading the Fight Against Infectious Diseases and Emerging Outbreaks

High throughput molecular biomarker technologies combined with bioinformatics have been successfully used to identify infectious agents, track transmissions, determine the origins of outbreak, and provide effective genomic surveillance

Anup Madan, Tina Green, Palanikumar Manoharan, and Abigail Cross at PPD

Infectious diseases are one of the leading causes of death worldwide. Prior to the recent outbreak of COVID-19, hospitals in the US alone reported well over 3 million cases of recognised infectious disease-related illnesses annually (1). Significantly greater numbers remained unrecognised, both in the inpatient and community settings, resulting in substantial morbidity and mortality. Despite progress made in our understanding of biological processes and availability of vaccines, the incidence of infectious diseases has increased globally during the last three decades. During the past decade alone, we have witnessed the emergence of many new pathogens not previously detected in humans, such as the avian influenza virus, Ebola, severe acute respiratory syndrome (SARS), and Middle East respiratory syndrome coronavirus (MERS-CoV). Critical and timely intervention for infectious diseases relies on rapid and accurate detection of the pathogen in the acute care setting and beyond (2).

Different laboratory methods, such as bacterial culture, molecular testing, and serology, are used to confirm a clinical diagnosis of infectious diseases. While these methods largely rely on the identification of the causative agents of the diseases themselves, serological methods assess the response of host innate immune systems for monitoring. The host's innate immune system is activated on infection by pathogens for non-specific suppression of pathogen replication and clearance. The presentation of pathogen-derived antigens to the cells of the adaptive immune response results in generating effective long-term immunity. The IgM-, IgA-, and IgG-type pathogen-specific antibody levels provide important measurements to predict population immunity against the disease. The pathogens also evolve to circumvent the host immune response and use host cells' transcriptional machinery to re-establish replication and infection. The functional state of the innate and adaptive immune response in a blood sample of a patient provides a prognostic biomarker of the disease and assists in vaccine development. In critical care settings, biomarkers are being increasingly utilised to improve clinical management for early diagnosis, risk stratification, and optimising therapeutic decisions.

Epidemics of infectious diseases vary geographically and through time due to movement of hosts who are susceptible to the disease or infected with the disease. In addition to the pathogen's genetic background, a host's genetic risk allows development of an advantageous ecosystem for pathogens. The interaction between hosts and pathogens is a coevolutionary process in which different sets of genetic events in any given population impact infection, disease development, rate of progression, convalescence, and asymptomatic carrier state. With the advances in genomic technologies and the development of computation tools, a high-resolution genomic architecture of the host and its cognate pathogens are used to decipher the evolutionary

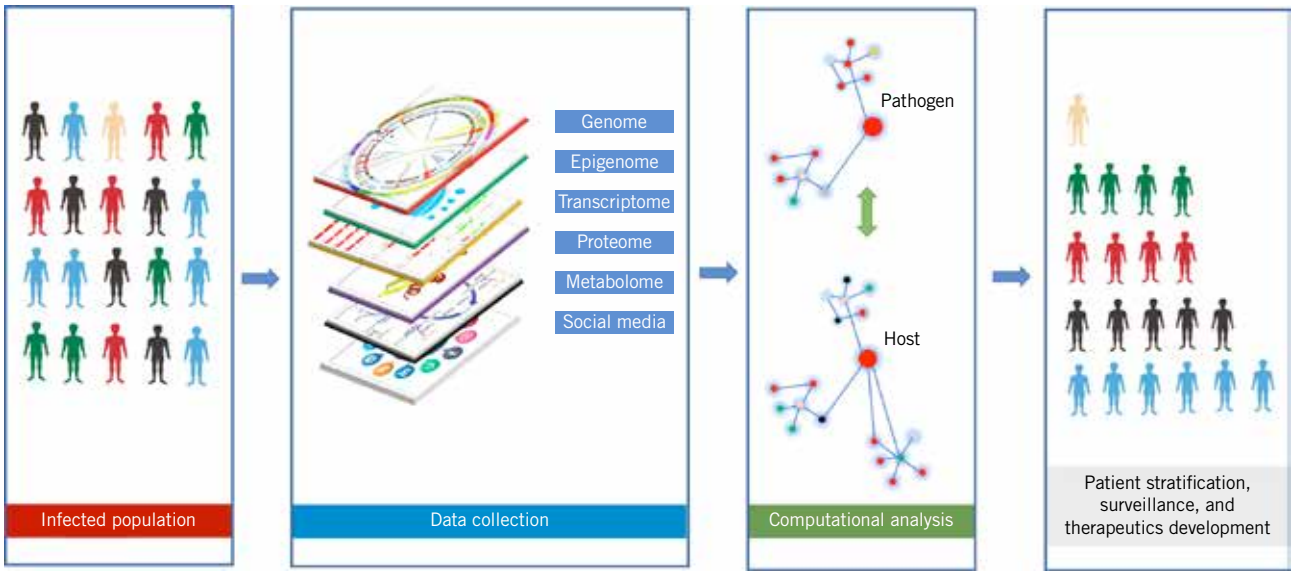

Figure 1: Schematic representation of genomic surveillance for infectious diseases

dynamics of host-pathogen interactions. These genome-wide association studies exploit study designs that filter out genotypic or haplotype risk ratios to identify genetic susceptibilities and resistant conferring loci in human genomes (3). The identified mutational signatures are used to understand host protective mechanisms against pathogens for therapeutic interventions and provide markers for diagnostics and surveillance. For example, the Δ32 mutation at the CCR5 locus is a well-studied example of natural selection acting in humans (4). Homozygous carriers of the Δ32 mutation (CCR5Δ32/Δ32) have been found to be resistant to HIV-1 infection in Caucasian subjects due to a non-functional CCR5 chemokine receptor (5). The homozygous mutation is present in only 1% of people descended from Northern Europe. Another 15% of people with European heritage carry one copy of the gene, which reduces the chances of infection and delays the progress of AIDS. The prevalence of this mutation in Europeans has existed in the population since before HIV infection occurred in humans, suggesting that past epidemics have played a role. The timing of the prevalence of mutation coincides with the Black Death pandemic, potentially driving natural selection in the human population (6). Those with the mutation were more likely to survive the plague and pass on their genes than those without, which caused an increase in the percentage of people with the mutation.

Similar to introducing chromosomal alterations, highly pathogenic viruses also have been shown to regulate the host epigenome (7). Some of these epigenetic changes, in particular DNA methylation of CpG islands, can be induced upon initial infection. The process is driven mainly by the increase of DNA methyltransferase (DNMT) activity, the enzymes that catalyse the transfer of methyl groups to cytosine residues of DNA. In a recent study, Corley *et al* showed that the methylation landscape of promoter-associated CpG islands is altered by the SARS-CoV-2 infection (8).

While the promoter regions of genes involved in the host type I interferon (IFN-1) response were hypermethylated, promoter regions of inflammatory genes were hypomethylated. These changes in the methylation patterns correlate with the changes in the expression of the related genes. Similar analysis in other infectious diseases has shown that pathogen infection leaves marked changes in the host DNA methylation patterns that could affect the expression of host factors involved in viral replication as well as in innate and adaptive immune defence (9). While the causal relationship between the epigenetic signals and infectious agents remains to be unravelled, these epigenomic signals present novel opportunities for therapeutics.

The changes in genomes and epigenomes in a population following infection results in changes in gene expression that can be investigated to develop disease classifiers (10). Peripheral blood is an ideal source for performing these studies due to the presence of the circulating white blood cells that are directly responding to the myriad immune signals cascading from remote primary sites of infection. Several studies have indicated that gene expression profiling of peripheral blood mononuclear cells (PBMCs) is a powerful novel approach for analysing host responses during infection, since bacteria and viruses trigger unique biomarkers during infection (11). Historically, microarray technology has been utilised for measuring gene expression changes. As the cost of next-generation sequencing (NGS) has rapidly decreased, the ability to obtain the entire snapshot of the transcriptome at a higher resolution and detect expressed sequence variants makes it a method of choice for performing such studies. These experiments yield massive amounts of data that are subjected to dimensionality reduction using mathematical models, such as SPARse factor analysis, Bayesian analysis, etc., to construct classifiers (12-13). As an example, a Bayesian network model has been recently used to quantify

immune response to a viral infection using gene-expression datasets from whole blood or PBMC samples (14). The authors used the gene-expression data generated during clinical studies following infections with influenza, respiratory syncytial virus, dengue, yellow fever, rotavirus, and hepatitis B virus to identify a transcriptional signature associated with increased activity of JAK-STAT1/2 and JAK-STAT3 signal transduction pathways. Similar analysis from sufficiently large studies holds the promise to develop a diagnostic test that can distinguish disease state in an individual with similar phenotypic features, such as viral vs bacterial respiratory infection and systemic inflammatory response syndrome vs sepsis, etc. Once disease-specific fingerprints can be identified, they can serve as a powerful diagnostic tool that can be assessed using quantitative polymerase chain reaction (qPCR) or digital PCR in a clinical setting.

While bulk transcriptional profiling provides very useful information, sequencing of transcriptome at the single-cell level, on a genome-wide scale, enables a greater appreciation of the cellular diversity in complex biological organisms and the myriad host transcriptional states during infection (15). For instance, single cell RNA-sequencing analysis classified the PBMCs from HIV-1 envelope-vaccinated neonatal and adult monkeys into four groups: B, T, natural killer, and monocyte, with each cell cluster showing different expression patterns between neonatal and adult monkeys. A significant increase in the ratio of activated B cells was found in neonatal monkeys, indicating that the neonatal immune system produces a stronger protective response than that of adult monkeys during HIV infection (16). Similar analysis has been performed to monitor the changes in the peripheral immune cell landscape in patients infected with SARS-CoV-2 (17). While in its infancy in infectious diseases, understanding infection as an integrated process between pathogen and host with resolution at the single-cell level ultimately will inform development of vaccines with greater productive and protective host immunity, enable the development of novel therapeutics that harness host mechanisms, and yield more accurate biomarkers to guide better diagnostics.

Since the functioning of a biological system is largely driven by proteins, efforts also have been made to construct host-viral protein-protein interaction networks using the literature-curated datasets (18). These networks include direct physical interactions between the proteins, as well as indirect associations, such as contributions to the same biological processes. There are several databases that provide information about direct physical interactions such as IntAct, the Database of Interacting Proteins, and the Biomolecular Interaction Network Database, while indirect association is obtained from functional genomic data sources, co-expression, functional similarity, text mining, and colocalisation databases. In these networks, the nodes represent proteins and the edges represent functional interactions between the proteins. The key interactions enriched in the infection pathways and associated nodes provide valuable targets for drug development. One can use computational tools to perform *in silico* knockouts and identify potential targets for drug development. While extremely powerful, the major drawback of such an analysis is the dependence on the known literature. There are still several genes that code for hypothetical proteins, and these are not represented in the analyses. In addition, the similarities between various infectious agents make it harder to tease out networks to identify biomarkers for individual diseases.

In addition to the biological network, social networks, with the diagnosed patients as 'nodes' and their epidemiological contact as 'edges', are being used to study how movement affects epidemics (19). Integration of social network analysis allows the capture of genetic differences in the host population on disease transmission, simulate and predict disease spread, and test disease control strategies. Many of these analysis tools draw on concepts and algorithms from graph theory that provide information about the importance of various nodes and how they communicate with each other. Characterisation of individuals based on different levels of infectiousness could guide the contact tracing interventions to prioritise contact screening, testing, and monitoring in a targeted manner at field level.

The changing epidemiology of infectious diseases requires new control strategies. However, establishing the burden of infectious diseases in low-resource settings is challenging due to the absence of effective surveillance systems. Efficient diagnostic tools are needed to provide accurate and timely guidance for identification, transmission disruption, and appropriate treatment administration of infectious diseases. Point-of-care (POC) tests provide actionable results at the site of care delivery. Significant progress has been made in developing accurate, simple, and cost-effective diagnostic tools for the detection of infectious disease-specific nucleic acids in the past decade. Multiple approaches have been exploited that simplify experimental procedures coupled with integrated microfluidic devices, and synthetic biology approaches (20). In addition to molecular approaches, serological approaches have been used to detect pathogen-specific proteins for diagnosis. Stringent clinical validations are still needed for these technologies to be translated from research to clinical practice. Since many infectious diseases may present with similar clinical symptoms, POC tests with multiplex functionality are highly desirable. Systematic characterisation of a set of biomarker signatures for a single infectious disease using high throughput technologies and computational approaches will prove to be a useful approach for future screening.

The breadth of applications for biomarkers in the practice of infectious diseases medicine – including technologies to facilitate rapid and accurate pathogen identification and determination of susceptibility to antimicrobials, patient risk stratification, and the study of communicable diseases epidemiology – will help in preventing, detecting, and responding to epidemics.

References

1. *Visit: www.cdc.gov/nchs/fastats/infectious-disease.htm*
2. *Morens DM and Fauci AS, Emerging pandemic diseases: how we got to COVID-19,* Cell *183(3): p837. Erratum for: Cell 182(5): pp1,077-92, 2020*
3. *Papkou A* et al, *The genomic basis of Red Queen dynamics during rapid reciprocal host-pathogen coevolution,* Proc Natl Acad Sci U S A *116(3): pp923-8, 2019*
4. *Novembre J* et al, *The geographic spread of the CCR5 Delta32 HIV-resistance allele,* PLoS Biol *3(11): e339, 2005*
5. *Samson M* et al, *Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene,* Nature *382(6593): pp722-5, 1996*
6. *Duncan SR* et al, *Reappraisal of the historical selective pressures for the CCR5-Delta32 mutation,* J Med Genet *42(3): pp205-8, 2005*
7. *Chlamydas S* et al, *Epigenetic mechanisms regulating COVID-19 infection,* Epigenetics *16(3): pp263-70, 2021*
8. *Corley MJ* et al, *Genome-wide DNA methylation profiling of peripheral blood reveals an epigenetic signature associated with severe COVID-19,* J Leukoc Biol*: 10.1002/JLB.5HI0720-466R, 2021*
9. *Oriol-Tordera B* et al, *Methylation regulation of antiviral host factors, interferon stimulated genes (ISGs) and T-cell responses associated with natural HIV control,* PLoS Pathog *16(8): e1008678, 2020*
10. *Ng DL* et al, *A diagnostic host response biosignature for COVID-19 from RNA profiling of nasal swabs and blood,* Sci Adv *7(6): eabe5984, 2021*
11. *Singhania A* et al, *The value of transcriptomics in advancing knowledge of the immune response and diagnosis in tuberculosis,* Nat Immunol *19(11): pp1,159-68, 2018*
12. *Carvalho CM* et al, *High-dimensional sparse factor modeling: applications in gene expression genomics,* J Am Stat Assoc *103(484): pp1,438-56, 2008*
13. *Chen M* et al, *Detection of viruses via statistical gene expression analysis,* IEEE Trans Biomed Eng *58(3): pp468-79, 2011*
14. *Sampson DL* et al, *A four-biomarker blood signature discriminates systemic inflammation due to viral infection versus other etiologies,* Sci Rep *7(1): p2,914, 2017*
15. *Kotliar D* et al, *Single-cell profiling of ebola virus disease* in vivo *reveals viral and host dynamics,* Cell *183(5): pp1,383-401.e19, 2020*
16. *Luo G* et al, *Probing infectious disease by single-cell RNA sequencing: progresses and perspectives,* Comput Struct Biotechnol J *18: pp2,962-71, 2020*
17. *Zhang JY* et al, *Single-cell landscape of immunological responses in patients with COVID-19,* Nat Immunol *21(9): pp1,107-18, 2020*
18. *Lasso G* et al, *A structure-informed atlas of human-virus interactions,* Cell *178(6): pp1,526-41.e16, 2019*
19. *Potter GE* et al, *Networks of face-to-face social contacts in Niakhar, Senegal,* PLoS One *14(8): e0220443, 2019*
20. *Chen H* et al, *Point of care testing for infectious diseases,* Clin Chim Acta *493: pp138-47, 2019*

Anup Madan PhD serves as Director of Biomarkers for **PPD® Laboratories**. He has more than 20 years of experience using high throughput genomics and system approaches to better understand disease etiology and identify biomarkers for P4 medicine. He has contributed to more than 60 peer-reviewed publications in prestigious journals including *Science* and *Nature*. Dr Madan received his PhD in Biochemical Genetics from Tata Institute of Fundamental Research, India, and completed his post-doctoral research with Dr Leroy Hood at the University of Washington, US.

Tina Green serves as Director of Laboratory Biostatistics for **PPD® Laboratories.** She has more than 20 years of experience supporting vaccine and biomarker assay development, optimisation, qualification, and validation.

Palanikumar Manoharan PhD serves as a Senior Research Scientist in the **PPD® Laboratories**' central lab. Dr Manoharan received his PhD in Medical Microbiology from the University of Madras, India, and completed post-doctoral research at West Virginia University, US, and the University of Cincinnati, US. His research has made significant contributions to the field of medical microbiology and chronic inflammatory diseases. He joined PPD in 2019 as a validation scientist in the molecular department. He currently serves as an Associate Editor for the journal *Nutrition & Metabolism*.

Abigail Cross serves as a Group Leader in the molecular genomics department for **PPD® Laboratories**. She works to manage clinical sample analysis and vaccine assay maintenance activities.