UNIVERSITY OF CAMBRIDGE

Host genomics: lessons for infectious diseases

PHG FOUNDATION

making science work for health

Authors

Heather Turner, Hayley Carr, Chantal Babb de Villiers, and Laura Blackburn

Acknowledgements

Pete Mills, Sobia Raza, Susan Mitchell, Rebecca Bazeley

URLs in this report were correct as of December 2024

Correspondence to

intelligence@phgfoundation.org

How to reference this report

Turner H. et al
Host genomics: lessons from infectious disease.
PHG Foundation. 2024

Executive summary

Infectious diseases are a major health problem. The COVID-19 pandemic drew attention to the global impact of pathogens, which remain a major cause of morbidity and mortality.

Infectious diseases elicit different responses among individuals exposed to the same pathogen – some people are more susceptible, and the severity of symptoms varies from person to person. Host genomics research aims to identify genetic variants that explain these differences in response, helping to improve infectious disease prevention and clinical management.

In this report, we explore the role of host genomics research to study four pathogens as examples of the opportunities and challenges faced in the field: Hepatitis C Virus (HCV), Human Immunodeficiency Virus (HIV), *Mycobacterium tuberculosis* and SARS-CoV-2 (the causative agent of COVID-19).

We find:

- Host genomics research has provided value by highlighting components of the immune system central to the host response to a pathogen. The clues provided by host genomics research are helping to identify critical elements that underpin an effective or ineffective host response.
- Despite pockets of success, discovery of the specific genetic factors involved in host response is proving challenging. The research often seems caught up in the logistics and challenges of the specific research question. Only a small proportion of host genomics research has supported innovation, such as in treatments.
- Remarkably few findings of host genomics research are shared between infectious diseases, demonstrating the specificity of host response and the variability of the factors involved.
- Overall, the result is a field of research that is fragmented, lacks coordination, and can feel unconnected to the broader context of essential infectious disease research.
- Taking a step back to consider the next phase of host genomics research could support more strategic thinking to align better with the clinical challenges.

Given the limited resources, host genomics research needs to be undertaken in the context of the public health and clinical priorities for each infectious disease.

Efforts to combat infectious disease are frequently under-resourced, disproportionately affecting low- and middle-income countries where infectious diseases are a leading cause of morbidity and mortality.

We advocate a more integrated and holistic approach to infectious disease research.

By taking a more considered approach, embracing innovation and novel methodologies, and prioritising translational research from the start, we believe host genomics can better address the real-world challenges of infectious disease.

To achieve this, we have the following three high-level recommendations:

- **Develop a host genomics research strategy** that supports greater coordination and strategic decisions on funding, through engaged consultation of active infectious disease researchers.
- **Support host genomics research** through collaboration, standardisation of research methods, development of robust infrastructure in affected countries, and ensure adequate funding for functional studies and translational research.
- **Embed translation** from the start by supporting a longer and more strategic model of research delivery; one that recognises infectious diseases as an ongoing and global health priority.

Contents

Abbreviations

AIDS – acquired immunodeficiency syndrome

BCG – Bacillus Calmette-Guérin vaccine

COVID-19 – coronavirus disease 2019

CRISPR – clustered regularly interspaced short palindromic repeats

GWAS – genome-wide association study

HCV – hepatitis C virus

HIV – human immunodeficiency virus

ICU – intensive care units

ITHGC – International Tuberculosis Host Genetics Consortium

LMICs – low- and middle-income countries

MSMD – mendelian susceptibility to mycobacterial disease

MTB – *Mycobacterium tuberculosis*

PGS – polygenic scores

SARS-CoV-2 – Severe acute respiratory syndrome coronavirus 2

TB – tuberculosis

TST – tuberculin sensitivity testing

WES – whole exome sequencing

WGS – whole genome sequencing

WHO – World Health Organization

1. Host genomics in response to infection

Infectious diseases are a major global health challenge, affecting millions of people and causing significant morbidity and mortality. Diseases, such as tuberculosis and COVID-19, represent ongoing threats with widespread health, social and economic impacts.

People differ greatly in their individual responses to pathogens that cause infectious disease. Some may be more susceptible than others to developing a specific disease, and severity of symptoms can vary greatly from person to person including more or less severe symptoms or no symptoms when infected by the same pathogen.

These differences in outcomes cannot be fully explained by their environmental exposures or clinical backgrounds. Host genomics research seeks to identify genetic variants that may explain some of these differences.

Host genomics research explores specific disease stages or phenotypes (the observable characteristics or traits), depending on the disease course for that infection (e.g. hospitalisation from SARS-CoV-2 infection or spontaneous clearance of Hepatitis C (HCV) infection). It aims to identify genetic variants for these phenotypes to improve understanding of infectious disease and lead to discoveries that improve disease management, treatment and control [1].

Host genomic research results have the potential to improve infection control and management, on an individual and population scale, for example:

- **Vaccines**

Development of vaccines based on host genomic findings for optimal response, particularly informed by the study of individuals resistant to infection.

- **Therapeutics**

Development of novel treatments against infection, or informed drug repurposing for an indication beyond that originally intended.

- **Risk stratification**

Determine likelihood of becoming infected or at risk of serious illness following infection to inform monitoring, preventative strategies or targeted interventions.

Our interest in this topic stems from the potential of genomics research to uncover novel insights that can transform disease management and clinical practice. We have explored the host genomics research landscape of four well-studied pathogens: SARS-CoV-2 (causative agent of COVID-19), human immunodeficiency virus (HIV), hepatitis C virus (HCV), and *Mycobacterium tuberculosis (MTB)*.

In this report, we:

- Describe themes from host genomics research, informed by these selected infectious diseases, and how this science is used to understand risk of infectious disease and impact on human health.
- Explore common principles identified through this research and how this science can inform better interventions for infectious diseases.
- Assess what can be learned from the study of the host response to infectious disease, as well as where there are gaps between the opportunities identified and the reality for patients.
- Identify recommendations that can support greater alignment between host genomics research and clinical priorities to lessen the impact of infections globally.

2. Response to infection and the role of genetics

Individuals and populations often respond differently when exposed to the same infectious disease. With HIV, for example, there are a small number of people who do not become infected, even after repeated exposure [2]. In HCV, a certain percentage of individuals without treatment will spontaneously recover, whereas others progress to acute disease [1,3,4].

The challenge is trying to identify why patients respond differently and how we can identify who is likely to have adverse outcomes.

Some insights come from tragic accidents. In Lübeck, Germany, in the four months from December 1929, 251 infants were vaccinated with BCG vaccine, which had been accidentally, and unwittingly contaminated, with varying amounts of virulent *M. tuberculosis* [5]. The affected infants were monitored for clinical or radiological signs of TB. While 173 infants survived, 72 infants died from TB. The overall findings were:

- 156 (68%) with clinical disease spontaneously resolved their symptoms suggesting newborns may be remarkably resistant to TB.
- The size of *M. tuberculosis* dose could be inferred (imperfectly) and this identified a clear relationship between the dose and outcome. At high doses more infants were susceptible, suggesting that high exposure overcomes host innate (genetic) resistance.
- Two infants received the lowest levels of contamination, but quickly progressed to disease and death which may suggest they were highly susceptible to TB.

A consistent finding in TB research is that a high proportion of people, when exposed, will not become infected or never progress to active disease.

2.1 Co-evolution of humans and pathogens

To understand the role of host genomics, it is necessary to understand the long history of co-evolution between humans and pathogens. Infectious diseases are one of the key drivers of evolution in organisms. Over a long period of exposure, populations will acquire genetic variants that confer resistance to a particular pathogen. However, these variants may not be favourable in all circumstances.

A classic example of this is the high incidence of sickle cell disease (and other genetic diseases of red blood cells) in areas where malaria is or has been prevalent [6]. For an individual to develop sickle cell anaemia, they must inherit two disease causing variants, one from each parent, affecting both copies of the beta-globin gene. In individuals who carry a single copy of the pathogenic variant, they are less likely to become seriously unwell when infected with malaria. This explains why sickle cell anaemia is more prevalent in populations where malaria is endemic.

Pathogens have emerged at different time points and with differences in the geography of affected populations, with some pathogens being global and others regional. This has consequences for how different populations have been affected by a pathogen. When comparing pathogens that more recently infected humans (i.e. HIV) to pathogens that have been circulating in human populations for thousands of years (TB), older pathogens are more likely to have impacted on human evolution.

3. What can host genomics tell us about infectious disease?

Host genomics researchers take a pathogen specific approach, meaning that host genomics findings will be specific to a population and a pathogen. Typical research questions include:

- Who is resistant to or susceptible to disease?
- Who is able to clear an infection and who has persistent disease?
- Why do some individuals have mild symptoms when others experience severe disease, or die?
- Why do some individuals experience further health consequences even after an infection has cleared?

The relevance or utility of each of these questions depends on the natural disease course of each infection. For example, HIV is a chronic infection and persists even with effective treatment, whereas HCV can spontaneously clear, but may leave individuals with heightened risk of hepatocellular carcinoma [7, 8].

Host genomics research builds from and is enabled by research into specific infectious diseases and the consequences for patients. This is best demonstrated by the COVID-19 pandemic, which challenged health systems and the ability of medical science to respond to this emerging pathogen.

During the COVID-19 pandemic, large-scale population research was enabled by the existence of very large and well curated biobanks. This led to some of the largest genome wide association studies (GWAS) performed to date, with more than 200 thousand cases and 3 million controls [9]. This is not reflective of the wider research in host genomics. Most studies are smaller, exploring specific populations or cases depending on the question being addressed.

We found significant host genomics research activity across the four infectious diseases we examined. Our analysis shows that research activity is often highest where there is a clearly defined population (for example, ICU patients hospitalised with SARS-CoV-2 infection). Where populations are less readily identified, there tends to be less associated research activity. For example, susceptibility to HCV or HIV infection is difficult to study, because certain risk factors make specific populations high risk for exposure [3, 10]. Many individuals, despite genetic susceptibility, will never be exposed.

Where host genomics research has identified genetic associations, this is the starting point for further research to understand the role of the identified variant or gene(s) in host response.

3.1 Host genomics research methods

Genetics research, including host genomics, uses a variety of methods to identify genetic variation associated with response to disease. The choice of methodology is dependent on several factors, including the population and disease being studied, available infrastructure, recruitment needs, and the level of detail in data collected. Funding is often a limiting factor. In host genomics research, the most commonly used methods are:

- **Investigating rare variants** is undertaken in individuals with extreme responses to infection, such as individuals who are highly resistant or highly susceptible to a particular outcome. Whole genome sequencing (WGS) or whole exome sequencing (WES) are the preferred assays to enable this research. In these cases, a Mendelian (or rare) model of inheritance is assumed, and variants identified have a large impact on risk of the disease or trait. The challenge is identifying syndromes that are specific to the infection or related pathogens. An example of this is mendelian susceptibility to mycobacterial disease (MSMD) a syndrome characterised by susceptibility to mycobacterial infection. These patients are typically identified when individuals become extremely unwell following BCG (Bacillus Calmette-Guérin (BCG)) vaccination, which contains less virulent *M. bovis*. This susceptibility is less specific to MTB infection and some syndromes lead to more generalised immunodeficiency.
- **Candidate gene studies** are population-based studies that test whether a specific gene is implicated in a particular trait or disease. The study design uses prior knowledge of genes to identify those that are known or suspected to have an influence in the disease progression and that may impact on host response.
- **Genome-wide association studies (GWAS)** test thousands or millions of genetic variants to find variants associated with a specific trait or disease. This method studies common variants within a population (typically carried by at least 1% of the population). GWAS are generally based on genotyping arrays, which cover key variants spread across the genome. Imputation, a statistical method used to infer additional variants based on the known genetic architecture of a population, can be used to expand the number of variants analysed. Arrays should be selected for specific populations and, if not, may not be reflective of local population diversity [1].

Both candidate gene and genome-wide association studies will use a case-control design. This approach compares individuals with a specific condition (cases) to those without it (controls) to identify genetic variants associated with the condition.

Results from all host genomic studies, regardless of method, can increase understanding of infection and disease mechanisms. Once a gene or variant has been associated with a disease, functional studies are needed to get a mechanistic understanding of how the gene or variant affects biological processes. However, this is not a straightforward process, and a range of methodologies and scientific approaches are used ranging from computational analysis through to investigation of cell lines and disease models.

3.2 How does host genomics research improve understanding of response to infection?

Host genomics research reveals how our genetic makeup influences the immune system's response to infections, particularly how well individuals can combat infection. The immune system is nuanced and complex, having evolved over time to respond to a wide range of infectious agents. To achieve this, the immune response needs to be specific, to eliminate infection, while also being adaptable, so the infection cannot evade detection. Clues provided by host genomics research can help to navigate this complex system, as well as corroborating information from the wider field of infectious disease research.

Differentiating TB resistance and latent TB infection

Tuberculosis has a broad spectrum of clinical presentations. Approximately half of exposed individuals will clear the bacteria before they develop persistent infection [11]. When infection is successful, 5% will develop clinical disease within one to two years of infection. For the remaining 95%, TB is contained (in a granuloma) and an immunological equilibrium is achieved. This is known as latent TB infection. In some cases, latent TB infection can be triggered leading to active TB disease.

Most TB diagnostics have been developed to identify TB infection or identify drug resistance. These diagnostics do not allow differentiation of these sub-populations. However, there are very different implications for clinical management, according to whether a given individual has cleared infection and whether they are high risk for active TB. This is why the study of TB resistance and susceptibility has been a focus of host genomics research. Having a better understanding of the host factors that allow some individuals to clear infection spontaneously, control infection or when the immune system fails to control TB infection would make it possible to identify these different populations more effectively. This could inform better surveillance and targeted prevention.

Realising this vision has been challenging. Historic patterns of exposure of populations to TB have resulted in very different risk from infection. As a result, very few host genetic variants seem to be shared, although some findings have been replicated in different populations.

Two chromosomal regions were identified (*TST1 and TST2)* in a study in large families from South Africa who tested negative for TB despite known exposure suggesting that they may be resistant [12]. This exposure is usually indicated by tuberculin sensitivity testing (TST), a test that measures the size of the swelling or induration following injection of tuberculin protein into the skin. *TST1* was found to regulate the expression of the proinflammatory cytokine TNF (tumour necrosis factor). TNF is known to be important in TB infection, demonstrated by the fact that patients receiving anti-TNF treatment have an increased risk of active TB disease, where there is latent TB infection. In comparison, *TST2* is associated with the intensity of TST reactivity but the mechanism by which this variant protects against TB is currently unclear. This illustrates the challenge researchers face when the level of scientific understanding varies for different host genetic variants.

Control of HIV infection

When infected by HIV, individuals will not progress to disease (acquired immune deficiency syndrome - AIDS) at the same rate, with some individuals maintaining a low viral load without treatment [10, 13]. Even following treatment, some people can maintain a very low or undetectable viral load without treatment (elite controllers). Individuals with poor control are (conversely) more susceptible, more rapidly progressing to AIDS. This suggests that there are differences in the way our bodies respond to and control HIV infection.

HIV has evolved to evade detection by the immune system, to have a high viral load and increase the chance of onward transmission. Research has identified that human leukocyte antigens (HLA) are essential for control of HIV infection [14]. HLA is a complex of genes involved in immune recognition and the adaptive immune response. Through host genomic research, specific HLA variants have been associated with lower HIV viral loads [15, 16]. Certain HLA alleles may have an enhanced immune response to HIV, or these alleles may force the virus to adopt less effective immune evasion strategies, which reduce viral fitness.

The identification of these variants in different populations suggests a shared mechanism that could be crucial for understanding HIV control. This information is valuable in trying to determine the biological path that results in infection, or the control of an infection that would hopefully identify potential treatment strategies.

Spontaneous clearance and HCV

A quarter of individuals, when infected with HCV, will spontaneously clear infection without treatment. This spontaneous clearance seems to confer some lasting protection against reinfection with HCV [4].

Individuals cured through treatment do not have the same reduced viral load or reduced time to clearance as found in individuals with spontaneous clearance [17]. Understanding the differences in host response leading to spontaneous clearance could inform better treatments or preventative strategies that also confer lasting protection.

Genes of interest in spontaneous clearance are involved in the immune response, such as HLA class II, *GPR158 and KIR2DL3* genes. Females are more likely than males to clear HCV infection spontaneously. A GWAS looking at differences in spontaneous clearance between males and females identified a sex-specific variant located in proximity to *ARL5B*, a gene involved in the antiviral response [18]. Males with this variant were 30% more likely to have chronic HCV infection but this effect was not seen in females. This difference suggests that the regulatory region, controlling gene expression, may be silenced in females, explaining this difference.

Ideally, host genomics would provide clarity. However, results can be contradictory and this uncertainty limits translation as researchers dig for answers that may not be forthcoming. The gene most strongly correlated with HCV host response is *IFNL4* [19-21]. Variants in this gene have been implicated in response to infection, including spontaneous clearance, jaundice during acute infection and liver damage, as well as in response to some IFN-based treatments. However, the role of this gene is not straightforward, because *IFNL4 gene*

variants that lead to loss or reduced activity of *IFNL4* have a higher chance of spontaneous clearance [21] or cure [20], despite *IFNL4* having antiviral activity. Whereas reduced *IFNL4* activity also leads to increased viral load, and this increases the risk of HCV acquiring mutations that can lead to treatment resistance.

Outcomes of SARS-CoV-2 infection

During the COVID-19 pandemic, the primary focus of host genomics research was to quickly understand why patients infected by SARS-CoV-2 experienced very different outcomes. The urgency of the pandemic initiated large-scale efforts, at a level previously unimaginable. The COVID-19 Host Genetics Initiative investigated host genetic variants in COVID-19 severity and susceptibility, including more than 200,000 cases and over three million controls. This was made possible by many studies contributing their data to this study [22, 23]. While some patients had no or very mild symptoms, others experienced severe, life-threatening disease. Factors were identified that significantly changed a person's risk, i.e. older age or pre-existing medical conditions, but these did not entirely explain the differences seen.

Some understanding of what was occurring was found through host genomic research:

- **Susceptibility to SARS-CoV-2**

Understanding susceptibility to a pathogen, similar to understanding resistance, can tell us about key elements of the biology of infection. A variant in the *SLC6A20* gene is associated with susceptibility to SARS-CoV-2 infection [24]. Specifically, increased expression of *SLC6A20* seems to drive susceptibility to SARS-CoV-2 infection. A variant may be close to more than one gene and transcriptomics (expression) data can tell us which gene(s) are expressed differently, altering the host response when infected, and this can inform further study. While this study found that *SLC6A20* plays a role in susceptibility, further study is needed to understand the exact mechanism.

- **Host genomics can explain risk factors identified in epidemiology studies**

Different risk factors were quickly identified for SARS-CoV-2 susceptibility and host genomics research has helped to understand the biology behind these findings. Host genomics researchers found that some *ABO* gene variants provided protection against SARS-CoV-2 [24, 25]. The *ABO* gene defines individual blood types, specifically if an individual has the A, B, AB or O blood group. They found that O blood type was protective, while A blood type was associated with increased susceptibility.

- **The rare explanation behind extreme susceptibility**

Host genomics research is important for understanding why some individuals are highly susceptible to severe COVID-19 disease, without other apparent risk factors. In these cases, a monogenetic cause is suspected and whole genome or exome sequencing can help to identify these rare causes of disease.

Early in the COVID-19 pandemic, four patients (two brother pairs aged between 21-33 years) with no known existing health conditions were admitted to intensive care with severe COVID-19 disease. This raised suspicions that these patients may be highly susceptible and

that a rare cause may be responsible. Whole exome sequencing identified rare variants in the gene toll-like receptor 7 (*TLR7)* [26]. *TLR7* has a role in type I interferon expression, which is key for an effective viral immune response and was thought to explain their poor response to being infected with SARS-CoV-2.

3.3 Shared findings from host genomics research

While most findings in our analysis were specific to the pathogen, we have identified some themes that are common across the four pathogens investigated:

- **Gatekeepers to the cell control infection**

It is no surprise that ease of cell entry is a key factor in host resistance or susceptibility to infection. A 32-base pair deletion in the gene *CCR5* is relatively common in European populations and most prevalent in Norwegian populations (16.4%) [27]. People with two copies of this variant (homozygous) are resistant to HIV infection. This variant prevents CCR5 expression on the cell surface of CD4+ T cells and results in resistance to infection [15]. *CCR5* has also been implicated in HIV control with slower disease progression in individuals with one copy of this variant [16]. Conversely, individuals, who express more CCR5 at the cell surface, are susceptible to infection and progress more quickly to acquired immunodeficiency syndrome (AIDS). This is also the case for SARS-CoV-2, where reduced expression of *ACE2* was associated with resistance to infection. Recent host genomics research in Nigeria of Lassa fever identified LARGE1, a key protein in Lassa virus cell entry, to be associated with decreased risk of Lassa fever infection [28].

- **Immune genes regulate host response to infection**

Individual and population variation will be central to any response to infection, be it successful, ineffective or actively harmful. There is a high degree of specificity in what elements of the immune system (and therefore which immune genes) are involved in host response to a pathogen. There is also some limited overlap. For example, interferons (IFN) regulate and activate an immune response to many types of infection. They have also been found to be essential regulators of immune response in COVID-19 (type I and III IFN), TB (IFN-gamma) and HCV (IFNL4) infections [20, 24, 29-34]. Additionally, *TYK2* has been implicated in severe disease for COVID-19 and susceptibility to TB [24, 35].

- **Key components of the immune response are involved at different stages**

Host genomics research can indicate, through the groups of genes identified and their function, which elements of the immune system are most important. Some genes are involved in the innate immune response, which is the body's first line of defence against infection (e.g. toll-like receptors [TLR]). Other genes are implicated in the adaptive immune response, involving a group of specialised immune cells and organs that are able to select for specific proteins when exposed to a new pathogen, as well as create 'memory' to support a faster immune response in the future (e.g. immunoglobulin [Ig] which code for antibodies and T-cell receptors). Some proteins bridge these two broad elements of immune response (e.g. human leukocyte antigens [HLA] and killer-cell immunoglobulin-like receptors [KIR]).

- **Tissue or organ specific effects drive symptoms associated with infection**

Host genomics research can provide insight into why certain infectious diseases are associated with certain symptoms. For example, HCV is defined by liver damage and host genomics research has identified key genes in lipid metabolism and detoxification that could explain increased risk of HCV-related liver disease and liver cancer [36-39]. Additionally, changes to blood clotting implicated in severe COVID-19 disease may be explained, in part, by coagulation factors and platelet activation proteins identified in host genomics research [40, 41].

However, remarkably few specific findings of host genomics research are shared between infectious diseases, demonstrating the specificity of the host response and complexity of the factors involved. Host genomics research has shed light on the variability in disease presentation and the particular role of the immune system in host response.

Elucidating the complex responses to infection offers clues as to where to focus efforts to develop therapies and vaccines. Our analysis shows this has led to some success. One antiretroviral therapy, Maraviroc, has been developed to target *CCR5*, blocking entry of HIV into cells [42, 43].

A second example is Baricitinib, which first received global approval in 2017 as a treatment for rheumatoid arthritis.

Baracitinib works by inhibiting Janus kinase enzymes. Host genomics provided evidence that this action would be through *TYK2*, a gene that host genomics researchers then implicated in severe COVID-19 [24]. Baricitinib was first identified in drug screening trials as early as February 2020 using BenevolentAI [44]. So, while host genomics research is often cited for the repurposing of Baricitinib, several approaches enabled drug repurposing to take place.

These examples illustrate a critical point. Host genomics is part of the infectious disease research landscape, and these insights provide evidence for translational research. But expectations of what is possible using this science can be high. Rather, host genomics should be considered a tool that helps researchers to identify critical elements of the immune system in order to improve understanding of host responses to infection. Nevertheless, host genomics research is challenging to undertake, and results of this research can be challenging to interpret.

4. What makes host genomics research difficult?

There are still large gaps in our understanding of how immune signalling pathways unfold in response to a given pathogen. This signalling is both for a given cell type (e.g. alveolar macrophages infected by *M. tuberculosis*) and for the interactions between immune cells in response to this infection. We know that the exact immune responses vary in response to different pathogens; however, the specific events of infection can be very difficult to study. An understanding of this context is important because the results of host genomics studies point to the specific components of the immune system that define an effective or ineffective host response.

It could be helpful to frame thinking about the host response in a similar way to complex disease. Host genomics is one element of the multifaceted immune response. Other factors include pathogen load or exposure, underlying health and the occurrence of repeat exposure or coinfection.

HIV is particularly challenging, for example, because it can lead to active TB in individuals with latent TB infection. The interaction of HIV and TB is exacerbating these epidemics, causing a syndemic where both pandemics are more severe because they are occurring simultaneously. Some populations are more at risk of exposure to these pathogens, and it can be difficult to determine what risk is related to host genetics and what risk is from the added burden of infectious disease. Another example is HCV, which often occurs with hepatitis B virus (HBV) or HIV.

As such, there are different factors that affect the complexity of host genomics research, which in turn informs decision making around methodologies. The specificity of host genomics findings to a particular pathogen (and often populations) is challenging, because interpretation requires further research into each result, and very few findings will have broad applicability.

4.1 Limitations in research methods and data collection

The course of infectious diseases is variable, and the quality of host genomics research depends on methods that are able to reliably identify individuals with the disease or phenotype under investigation. When understanding why host genomics research has had limited translational outcomes, there are several core elements to consider:

- **Quality of phenotype data**

High quality phenotyping is key to successful health-related research, because phenotype measures are used to categorise those with disease and others whom they are compared to, and host genomics research is no exception. The reliability of a phenotyping method can vary according to the disease and phenotype under investigation.

Often research is performed based on large biobank studies which may include very crude phenotype data – sometimes as basic as 'positive' or 'negative' – for a disease. This can limit what a study, such as a GWAS or candidate gene study, can identify about a disease or trait [45].

This challenge is demonstrated in the case of TB resistance and control. Tuberculin sensitivity testing (TST) for TB cannot differentiate between individuals who have cleared an infection and individuals with latent TB infection [12]. However, TST testing is cheap with minimal logistic considerations and so remains the default test for presence of a TB infection. Nevertheless, TB researchers cannot use datasets that only contain these test results to differentiate between individuals who can eliminate an infection quickly and those with latent infection. Diagnostics are needed that can identify those that are most likely to develop active TB disease to be treated, as this is a pressing concern for management of TB in affected communities. Better differentiation of these subpopulations is likely to improve host genomics research, because identified variants should be specific to these disease states.

- **Ensuring representation of affected populations**

Global majority populations are underrepresented in genomic studies, despite these populations frequently being the most affected by infectious diseases. For example, the highest burden of HIV infection is in women in sub-Saharan Africa, and yet women are significantly underrepresented in research [46].

Often, host genomics research will be investigating admixed (mixed ancestry) populations, particularly when studying infectious diseases in regions of the world where there has been significant migration (e.g. South Africa, Brazil and Peru). Significant differences in risk of infection are observed within these geographic populations when they are exposed to the same pathogen. This has been shown to be related to population ancestry. For example, a TB susceptibility gene was identified when studying admixed populations in South Africa, originating from Khoe-San ancestry [47]. Whilst identifying susceptible sub-groups, this does add to the complexity, largely because most statistical methods used in genomics research were not developed to be used in admixed populations.

Methodological advances, such as those developed and used to study admixed populations in South Africa, can leverage signatures of admixture in populations leading to novel insights [48, 49]. Understanding the genetics of the affected population is essential to ensure confidence that findings are related to the pathogen and are not artifacts of the population's genetic background.

Many smaller studies, with more detailed, nuanced clinical data, have been performed. However, the majority of these have been in well-resourced settings. This is despite host genomics research likely having most value in lower income countries, which are often more affected by infectious diseases. This reflects both the challenge of accurate phenotyping and available resources. In particular, some phenotypes may be more difficult to research for a specific infectious disease or disease context. For example, susceptibility to HIV or HCV is under-studied because there are known risk factors that significantly alter the likelihood of exposure. Many people who may be susceptible will never be exposed to the pathogen.

- **Allocation of resources**

Host genomics research forms part of the wider research landscape for any infectious diseases, where inequalities are also found.

Certain populations may be overrepresented in research because they are more readily identifiable for a specific trait pertinent to the research. For example, while women are generally underrepresented in HIV research, they are more likely to be spontaneous controllers (i.e. maintain a very low viral load following antiretroviral treatment). Women from high income settings are therefore overrepresented in studies of post-treatment control in high-income settings [13].

Conversely, as we mention above, some populations may be underrepresented, despite their high burden of disease. If host genomics research is to meet the needs of the most affected populations different decisions need to be made as to where to allocate resources in infectious disease research.

A review of key populations vulnerable to TB reveals that the risk for some populations has not been estimated [50]. This is a potentially serious omission because understanding the risk profiles of different populations for a given infectious disease informs decision making around the allocation of public health and research funding.

The review is informative of the wider landscape where similar gaps are likely to exist across infectious diseases and affected populations. Not only will the most in need populations fail to benefit from existing public health and research initiatives, but they will also be subject to intersectional inequalities. The impact of infectious disease across all global populations is critical. Further research to address these gaps can inform host genomics research by identifying populations, currently absent in research, who may have unique host genetic variants informing their risk of disease.

- **Replication of findings**

Replication is essential to validate results and can inform how widely applicable a finding is. The significance of genomics research is measured by the power of the study, strength of association and, most significantly, replication of findings [51]. However, replication of host genomics findings is proving difficult. Host diversity, pathogen diversity, and co-evolution between pathogens and specific populations may result in findings that are population specific and difficult to replicate.

In the quest to increase the statistical power of host genomics research, initiatives have been set up to facilitate studies combining data from multiple cohorts. In TB, large-scale collaboration was spurred on by the International Tuberculosis Host Genetics Consortium (ITHGC) [52], that brought together smaller host genomics groups and enabled their data to be combined. The hope was to increase the power of the research (and therefore confidence in) any results. This effort was not as successful as might have been hoped, with only one variant reaching statistical significance [53]. However, it offered a significant learning opportunity, for example, when comparing phenotype definitions between studies.

The evolutionary arms race adds to the difficulty in replicating results. Host genomics studies have been more successful in COVID-19 and HIV, relatively recent pathogens, compared to TB, an ancient pathogen that has been circulating in human populations for thousands of

years. Replication of findings has been more successful for SARS-CoV-2 across populations, although differences do exist. This can also be seen in which strains have spread most successfully, with new COVID strains repeatedly becoming globally dominant. In comparison, very few findings have been replicated when studying the host genomics of TB [53].

Replication of results may not be possible, because of population and pathogen specific evolution, leading to unique findings specific to that population. TB presents a particular challenge, although we identified nuances in each pathogen studied.

Estimates in 2016 suggest that approximately 1.7 billion people have latent TB infection [54]. Unlike other pathogens, TB evolves very slowly, so we also have a good understanding of local TB strains affecting different populations. TB has influenced human evolution over thousands of years. One reason TB host genetic factors have been difficult to replicate may be that these are population and pathogen strain specific. However, the methods to study this, which compare the genome of the host (human) and strain (pathogen) - a genome-to-genome study, are still being established.

Resolving these challenges

The solutions to these challenges are not obvious. Collaborations that resulted in large studies have been achieved in COVID-19 initiatives and for TB in the ITHGC with more limited results than postulated. One explanation is that combining retrospective studies, which may have used different definitions for cases and controls, limits the interoperability of these datasets. Learning from these efforts, researchers are starting to define the key elements of successful host genomics research. Consensus around these methods would support greater collaboration and enable the establishment of consortiums for different pathogens, as pioneered by ITHGC.

Efforts may be more successful when standard collection methods and clearly defined phenotypes are required when initiating a large study. By sharing the lessons from previous experience, it will be possible to develop a framework for more collaborative and successful research in the future.

4.2 Interpreting results

To understand findings from host genomics research and enable clinical translation, it is valuable to have a robust understanding of the mechanism by which a variant or gene alters the host response and how this leads to the phenotype. However, host genomics results, often, do not have additional evidence from functional studies. Success in addressing the gap in evidence needed for translation has been mixed. Host genomics research can inform or add to the complexity of understanding the role of the immune system in infectious disease.

However, it is important to remember that a successful immune system should not be too specific but should be able to respond effectively to the many different pathogens it will encounter. The genetics of the immune system has significant diversity at a population level and that the immune system has evolved to circumvent different ways that infections evade detection in order to spread. This diversity is what allows populations to be resilient to novel pathogens, but adds to the complexity researchers face when interpreting results.

Our analysis identified a few examples that demonstrate this problem:

- **Immune genes are polymorphic**

Some immune genes (in particular HLA and KIR genes) are genetically highly variable (polymorphic), and this can make them very challenging to research. These genes play an important role in an immune response to infection and are frequently identified in host genomics research across various infectious diseases. HLA and KIR genes are important in both innate and adaptive immune responses. They are clustered on certain chromosomes. This close proximity means that several genes may be linked with a variant, and researchers are developing methods to understand the specific role of variants that have been associated with susceptibility to infections.

- **Different immune genes are not equally understood**

Arrays are a key assay used to measure specific variants across the genome and are used in GWAS. HLA genes are better represented on arrays than KIR genes with the result that the role of KIR genes is under-studied in host genomics research [55, 56]. There are arrays developed specifically for HLA genes and, given the locus with KIR genes shares similar complexity, similar assays should be developed to enable better study of these genes. This would lead to new insights into the host response that host genomics researchers believe are currently being missed.

- **Individuals carry two different versions of a gene (alleles)**

For most genes, each person carries two copies. For the immune system, this is known to influence the immune response. Individuals often carry different HLA alleles, although T cells will only express one version of the gene. Following HIV infection, individuals who carry different HLA alleles have been observed to have a lower viral load than homozygous individuals – who have two copies of the same allele. The 'heterozygotes advantage' hypothesis suggests that the breadth of immune presentation results in better control of infection [57, 58].

- **Understanding gene interaction (epistasis)**

Variants may only influence host response when they occur in combination, a phenomenon known as epistasis. GWAS analyses are limited when trying to identify functionally related variants, which may only have a phenotypic effect when inherited together [59]. For example, individuals who carry specific activating variants in the KIR receptor *KIR3DS1* and *HLA-B* (*Bw4 I80*) had lower viral loads indicative of HIV infection control [60]. This combination of variants has also been found to be protective against hepatocellular carcinoma following HCV infection in three studies [7]. The same variants may be implicated in different ways, and these may be advantageous in some contexts, and harmful in others.

5. Research towards clinical implementation

Aware of the priority issues for an infectious disease, researchers are keen to realise the opportunities of host genomics for population health. Certain areas are regularly highlighted where host genomics research could lead to novel clinical interventions including:

- Identifying biomarkers and the development of risk prediction algorithms, which includes polygenic scores (combining multiple genomic variants, each of which make a small contribution to risk, into a single score) to assess risk of infection, severity of disease and predict outcomes.
- Improving diagnostics and timing of treatment for optimal management (e.g. latent TB infection).
- Identifying druggable targets for novel drug development or to inform drug repurposing in all diseases.
- Developing vaccines, treatments or cures for infectious disease which either harness or repress components of the immune system that are allowing the pathogen to persist.
- Optimising vaccination by developing vaccines designed to activate essential components of the immune system for an optimal response where there is no vaccine available (HCV and HIV), or the vaccine has limited efficacy (TB).

Genetics also has promise in other areas to support clinical management of infection:

- Identifying genetic biomarkers for predicting response to treatment (i.e. pharmacogenomics).
- Predicting the long-term risks following infection (i.e. hepatocellular carcinoma following HCV infection).

For host genomics research to benefit patients directly, each study needs to be tailored to address the complex challenges of a specific pathogen in a host system. Host genomics research is often one piece of evidence that contributes to the bigger picture which may lead to a new intervention. In particular, 'omics technologies are increasingly being used to gain new insights into the diseases we have discussed. There are examples of host genomics research addressing these areas.

5.1 Biomarkers for diagnosing TB infection

Current TB diagnostics are not able to distinguish between new infection, re-infection or latent infection. One promising diagnostic approach is the use of transcriptomic signatures using cell-free RNA. This assay could identify changes in the host response to TB infection that pre-warn of active (and therefore contagious) TB infection. Cell-free RNA from blood

samples is believed to be more informative than whole blood RNA because most RNA will originate in dead or damaged cells, which occur in the case of an infection. A six-gene panel of host genes was developed and validated in different populations globally, meeting the World Health Organisation (WHO) target product profile requirements to identify individuals with potential TB infection for treatment [61, 62]. Future research is needed to assess how this test performs, particularly when delivered in a real-world setting. This approach aligns with the development of biomarkers and risk prediction algorithms, improving diagnostics and timing of treatment for optimal TB management.

5.2 Searching for an end to HIV

With a few exceptions, HIV is incurable, as the infection is able to persist in reservoir cells. This means that patients continue to live with HIV and infection can become active again without appropriate monitoring and treatment. HIV has been reported to be cured in a few cases reported in the literature through a bone marrow transplant from donors with a common deletion variant in CCR5, known to confer resistance to HIV infection [63, 64]. An estimated 39 million people are living with HIV and therefore scalable solutions are needed to eradicate HIV globally.

- **Developing a cure**

  There is unlikely to be a one-size-fits-all cure for HIV. Gene therapies may offer one possibility. Clinical trials are exploring CRISPR-based therapies to make cuts to the integrated HIV genome, damaging the virus and preventing future replication [65]. CRISPR can also be used to target host genes involved in cell entry, known from host genomics research to confer resistance to infection (*CCR5 and CXCR4*), to prevent replication [66]. Other studies want to identify the host mechanisms that HIV uses to maintain the proviral reservoir [67]. By targeting these genes (*HDAC2 and BRD2*), HIV would start to replicate, and patients would become responsive to antiretroviral therapy [68]. Host genomics research may be able to provide information about what genes or pathways to target to eradicate the HIV reservoir. The goal would be to eradicate the HIV reservoir completely, curing HIV infection. A successful cure for HIV can only work if this treatment is accessible and available in all affected countries. This would be a priority to meet the WHO ambition of eradicating HIV by 2030 [69].

- **Prevention at the source**

  The World Health Organisation (WHO) has identified five priority pathogens for vaccine development, including HIV [70]. Developing a vaccine for HIV is a key step towards being able to prevent HIV transmission in high-risk populations. HIV is a single stranded RNA virus with a very high mutation rate, making vaccine development very challenging. Potential targets that are key mediators of the HIV immune response have been identified in host genomics research. These insights are being explored to refine vaccine development strategies. For example, HLA-E, which may play a role in natural killer cells recognising HIV infected cells, could be a target for a vaccine approach. A vaccine developed to protect rhesus macaques against simian immunodeficiency virus (from a similar lineage to HIV) has been found to be protective. Researchers believe this

effect was mediated by Mamu-E (a HLA-E homolog) [71]. More research is needed to understand how these immune changes could be used to prevent HIV replication and, therefore, transmission.

Despite pockets of success, discovery of the specific genetic factors involved in host response is proving challenging. The research often seems caught up in the logistics and challenges of the specific research question. Given the limited resources, host genomics research needs to be undertaken in the context of the public health and clinical priorities for each infectious disease. Efforts to combat infectious disease are already frequently under-resourced, disproportionately affecting low- and middle-income countries where infectious diseases are a leading cause of morbidity and mortality.

6. Outlook for host genomics research

Host genomics research is an active field, and researchers have laudable aspirations to improve population health, but this research may not directly inform interventions. Instead, the value comes from these biological insights, which provide evidence to support hypotheses for translational research. Yet, to date host genomics research has been conducted in a disease-centric and population-specific manner. The result is a field of research that is fragmented, lacks coordination and is seemingly unconnected to the broader context of infectious disease research.

All infectious disease research needs to consider the host. Researching disease without considering the host is like trying to understand a story by only reading the right-hand pages of a book. Both perspectives are needed to complete the story.

We are advocating a more integrated and holistic perspective.

This includes having a broader view that appreciates the interconnectedness of different infectious diseases, and different populations, and that promotes interdisciplinary collaboration and coordination, recognising that:

- **Infectious diseases are a major health problem**

The COVID-19 pandemic drew attention to the impact and reach of these pathogens and globally infectious diseases remain a major cause of morbidity and mortality. The impact of these infections is not felt equally, with the greatest burden of these diseases falling on low- and middle-income countries. Host genomics research has a responsibility to ensure that the needs of those most affected are represented in this work.

- **Efforts to combat infectious disease**

Efforts to combat infectious diseases are frequently under-resourced. Host genomics research needs to be sensitive to the wider context in which this research is being delivered. However, it is essential for host genomics research not to be an afterthought. Rather, host genomics research needs to be one pillar in infectious disease research programmes.

By embracing new approaches that use novel methodologies, being sensitive to the broader context, and prioritising the translation of research findings into clinical practice, science can more effectively address the real-world challenges of infectious diseases. To be able to achieve this we have the following three recommendations.

Recommendation 1

Develop a coordinated host genomics research strategy

A unified host genomics research strategy developed by clinical experts, geneticists, infectious disease specialists and immunologists is essential. Currently there is a lack of co-ordination across the field. Host genomics research is not only disease-centric, but often focused on a single country or population. Although there are examples of collaboration between research groups, most is performed in isolation.

A host genomics research strategy would support greater coordination and standardisation of research approaches. By sharing learnings and methodologies, these researchers will create opportunities to understand relationships between different populations and infectious diseases. Funders are well-placed to encourage and incentivise greater coordination, cooperation and collaboration of host genomics research. This should be both in terms of what host genomics research is funded and how this research fits into the wider infectious disease funding landscape to ensure these efforts are aligned.

Development of a host genomics strategy should include engaged consultation with active researchers who can inform strategic choices on funding. This should involve multidisciplinary expertise, capturing a range of perspectives on where the challenges are, where research is needed and how research needs to be integrated to address specific challenges of the pathogen(s) under investigation. Learning from efforts such as ITHGC and the COVID-19 Host Genetics Initiative, should inform other host genomics research and support greater collaboration in future studies.

Recommendation 2

Support research that leverages advances in technology and data science

Aligning research goals between infectious disease and host genomics research would support focus on public health and clinical priorities. Two crucial barriers to overcome are establishing robust research infrastructure that supports host genomics and infectious disease research, and a more holistic research landscape that follows through discovery research and functional studies to better understand the role of a gene or variant in host response.

Recommendation 2A: Enhance data standardization and phenotyping

High quality host genomics research is underpinned by infectious disease research which is, in turn, dependent on high quality data collection. Given that phenotyping, particularly to determine disease status of an infection, is the foundation of host genomics research, there is a pressing need for accurate, precise and standardised phenotyping to support high confidence results that can be taken forward in future research. Host genomics researchers

should agree standards and definitions (for the specific pathogen) that inform study design. This should, in turn, inform research efforts by applying these definitions in data collection and analysis. This will also improve the interoperability of datasets facilitating greater comparison and, where appropriate, collaboration.

Recommendation 2B: Support interdisciplinary collaborations that use innovative research methodologies

Host genomics researchers are increasingly exploring innovative methods, complementary to their existing work, which may lead to new insights or improve confidence in existing work:

- **The host-pathogen relationship**

Genome-to-genome studies use both host and pathogen data to identify population and strain specific associations. This may explain observed differences in population-specific host response. This is a novel approach, but early studies are promising [72, 73].

- **Develop methods for global populations**

Statistical methods in genomics have been developed for homogenous populations, particularly European populations, to ensure any associations are linked to the trait under investigation. However, this does not reflect global populations, particularly regions of the world where there has been significant migration resulting in admixed populations, who are removed from genomics studies. However, we know that different ancestral backgrounds can change individuals' risk from infection. Inclusion of these individuals in research is essential to redress inequities and will inform insights that may otherwise be missed. Statistical methods are now being developed that account for admixed populations and efforts in this space are essential for genomics research to become more equitable.

- **Leveraging multimodal data and 'omics**

Multimodal approaches are becoming embedded in genomic medicine and, as tools to harness these data improve, they are likely to become the norm. There are already examples of how they have been used, for example, transcriptomics used to explain how a variant alters gene expression, and this data can guide hypothesis generation and next steps in research. Some sequencing technologies can capture methylation and other epigenetic sequence data, and this integration will increase the power of these multimodal approaches.

- **Mapping complexity using long-read sequencing**

A key obstacle in genomics research is the quality of genomics data, which has predominantly been generated using arrays or short-read sequencing. Longer read sequencing captures larger genome segments, providing clearer, less ambiguous data and improving the accuracy of reconstructing DNA sequences. For example, long read sequencing is better at capturing variation in complex regions of the genome of known significance (i.e. HLA genes) and variation in populations poorly represented in genomics research.

Recommendation 2C: Establish infrastructure to benefit the global majority

The readiness of the host genomics research community to respond to the COVID-19 pandemic was only possible because of prior investment in large scale genomics infrastructure. There are several lessons from this:

1. Host genomics research is one of the core pillars of pandemic preparedness.
2. The response was not perfect and care should be taken to learn from this experience.
3. While pandemics have been infrequent, the impact of infection is ongoing and therefore continuous funding into this infrastructure and research is essential.

Low-and middle-income countries (LMICs) stand to gain the most from host genomics research, given that these countries remain most significantly burdened by infectious disease, which is a longstanding equity issue.

The research undertaken in and with low resource settings is increasing and this is leading to novel findings pertinent to their specific populations and informing the broader understanding of the host response [28, 72]. However, the greater part of research funding is still directed to or through high resource countries. Therefore, investment into infectious disease research should be directed towards LMICs, where these diseases are most prevalent.

Efforts are underway to establish equivalent initiatives in lower resource settings. An example is the African Biobank and Epidemiological Ecosystem (ABLE) which aims to facilitate scientific research collaborations involving partners across Africa [39]. Host genomics research is one type of research that is enabled by these initiatives and biobanks will support wider research to combat the burden of infectious disease.

Recommendation 2D: Support functional studies as follow up research

Follow up research using functional studies are essential to understand the role of a variant in disease. This has been done to varying extents; however, many variants identified in genomics research, and their role in disease, remain unexplained. There is a need to increase funding spent supporting research to clarify the role of identified variants. Addressing this gap is a key step in the translation pathway.

Recommendation 3

Facilitate translation of research into practice

For host genomics research findings to reach clinical practice, translation needs to be considered from the beginning. Researchers can become trapped in a cycle of searching for funding and publications, leaving limited space to attend to the translation of their work (or even just how their work fits into the wider context).

Different skills are needed through the discovery and translation pathway. Consequently, a sustainable funding model that supports these different types of expertise is required.

Infectious diseases will remain a priority and therefore research needs to be part of a longer-term model of funding. The current funding model requires scientists to constantly adjust to a shifting funding landscape. Building a more sustainable funding model would promote partnerships between different researchers and disciplines, and support follow through on research from discovery to adoption.

This ambition should be supported by a host genomics strategy, engaged and sustained funding models and engagement with stakeholders at each stage of research. A more holistic approach will keep host genomics researchers engaged with the clinical priorities of the infectious disease(s) under investigation and support engagement with host genomics discoveries. This model will enable synergies that support greater innovation. Without this shift, we risk continuation of historic models of host genomics research, one that struggles to provide the benefit and value that patients (and researchers) want to see.

References

1. Kwok, A. J., Mentzer, A., Knight, J. C. Host genetics and infectious disease: new tools, insights and translational opportunities. Nature Reviews Genetics. 2021. 22(3): pp. 137-153.
2. Klunk, J., Vilgalys, T. P., Demeure, C. E., et al. Evolution of immune genes is associated with the Black Death. Nature. 2022. 611(7935): pp. 312-319.
3. Hepatitis C. World Health Organisation (WHO). 2022; Available from: https://www.who.int/news-room/fact-sheets/detail/hepatitis-c.
4. Thomas, D. L., Thio, C. L., Martin, M. P., et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature. 2009. 461(7265): pp. 798-801.
5. Fox, G. J., Orlova, M., Schurr, E. Tuberculosis in Newborns: The Lessons of the "Lübeck Disaster" (1929–1933). PLOS Pathogens. 2016. 12(1): p. e1005271.
6. Cyrklaff, M., Sanchez, C. P., Kilian, N., et al. Hemoglobins S and C interfere with actin remodeling in Plasmodium falciparum-infected erythrocytes. Science. 2011. 334(6060): pp. 1283-6.
7. Walker, A. J., Peacock, C. J., Pedergnana, V., et al. Host genetic factors associated with hepatocellular carcinoma in patients with hepatitis C virus infection: A systematic review. J Viral Hepat. 2018. 25(5): pp. 442-456.
8. Sharkawy, R. El, Bayoumi, A., Metwally, M., et al. A variant in the MICA gene is associated with liver fibrosis progression in chronic hepatitis C through TGF-β1 dependent mechanisms. Scientific Reports. 2019. 9(1): p. 1439.
9. The COVID-19 Host Genetics Initiative. A second update on mapping the human genetic architecture of COVID-19. Nature. 2023. 621(7977): pp. E7-E26.
10. Deeks, S. G., Overbaugh, J., Phillips, A., et al. HIV infection. Nature Reviews Disease Primers. 2015. 1(1): p. 15035.
11. Bell, L. C. K., Noursadeghi, M. Pathogenesis of HIV-1 and Mycobacterium tuberculosis co-infection. Nature Reviews Microbiology. 2018. 16(2): pp. 80-90.
12. Möller, M., Kinnear, C. J., Orlova, M., et al. Genetic Resistance to Mycobacterium tuberculosis Infection and Disease. Frontiers in Immunology. 2018. 9.
13. Scully, E. P. Sex Differences in HIV Infection. Curr HIV/AIDS Rep. 2018. 15(2): pp. 136-146.
14. Luo, Y., Kanai, M., Choi, W., et al. A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. Nature Genetics. 2021. 53(10): pp. 1504-1516.
15. McLaren, P. J., Fellay, J. HIV-1 and human genetic variation. Nature Reviews Genetics. 2021. 22(10): pp. 645-657.
16. Gingras, S. N., Tang, D., Tuff, J. F., et al. Minding the gap in HIV host genetics: opportunities and challenges. Human Genetics. 2020. 139(6): pp. 865-875.

17. Sacks-Davis, R., Grebely, J., Dore, G. J., et al. Hepatitis C Virus Reinfection and Spontaneous Clearance of Reinfection--the InC3 Study. J Infect Dis. 2015. 212(9): pp. 1407-19.

18. Vergara, C., Valencia, A., Thio, C. L, et al. A Multiancestry Sex-Stratified Genome-Wide Association Study of Spontaneous Clearance of Hepatitis C Virus. The Journal of Infectious Diseases. 2020. 223(12): pp. 2090-2098.

19. Yan, Z., Wang, Y. Viral and host factors associated with outcomes of hepatitis C virus infection (Review). Mol Med Rep. 2017. 15(5): pp. 2909-2924.

20. Møhlenberg, M., O'Brien, T. R., Hartmann, R. The role of IFNL4 in liver inflammation and progression of fibrosis. Genes & Immunity. 2022. 23(3): pp. 111-117.

21. Page, K., Mirzazadeh, A., Rice, T. M., et al. Interferon Lambda 4 Genotype Is Associated With Jaundice and Elevated Aminotransferase Levels During Acute Hepatitis C Virus Infection: Findings From the InC3 Collaborative. Open Forum Infect Dis. 2016. 3(1): p. ofw024.

22. Initiative, Covid-19 Host Genetics. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur J Hum Genet. 2020. 28(6): pp. 715-718.

23. Initiative, Covid-19 Host Genetics. A second update on mapping the human genetic architecture of COVID-19. Nature. 2023. 621(7977): pp. E7-E26.

24. Niemi, M. E. K., Daly, M. J., Ganna, A. The human genetic epidemiology of COVID-19. Nature Reviews Genetics. 2022.

25. Liu, N., Zhang, T., Ma, L., et al. The impact of ABO blood group on COVID-19 infection risk and mortality: A systematic review and meta-analysis. Blood Rev. 2021. 48: p. 100785.

26. van der Made, C. I., Simons, A., Schuurs-Hoeijmakers, J., et al. Presence of Genetic Variants Among Young Men With Severe COVID-19. Jama. 2020. 324(7): pp. 663-673.

27. Solloch, U. V., Lang, K., Lange, V., et al. Frequencies of gene variant CCR5-Delta32 in 87 countries based on next-generation sequencing of 1.3 million individuals sampled from 3 national DKMS donor centers. Hum Immunol. 2017. 78(11-12): pp. 710-717.

28. Kotliar, D., Raju, S., Tabrizi, S., et al. Genome-wide association study identifies human genetic variants associated with fatal outcome from Lassa fever. Nat Microbiol. 2024. 9(3): pp. 751-762.

29. Pairo-Castineira, E., Rawlik, K., Bretherick, A. D., et al. GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19. Nature. 2023. 617(7962): pp. 764-768.

30. McNab, F., Mayer-Barber, K., Sher, A., et al. Type I interferons in infectious disease. Nature Reviews Immunology. 2015. 15(2): pp. 87-103.

31. Harris, J., Keane, J. How tumour necrosis factor blockers interfere with tuberculosis immunity. Clin Exp Immunol. 2010. 161(1): pp. 1-9.

32. Lalvani, A., Millington, K. A. T Cells and Tuberculosis: Beyond Interferon-gamma. J Infect Dis. 2008. 197(7): pp. 941-3.

33. Fabri, M., Stenger, S., Shin, D. M., et al. Vitamin D is required for IFN-gamma-mediated antimicrobial activity of human macrophages. Sci Transl Med. 2011. 3(104): p. 104ra102.

34. Boisson-Dupuis, S., Bustamante, J. Mycobacterial diseases in patients with inborn errors of immunity. Curr Opin Immunol. 2021. 72: pp. 262-271.

35. Boisson-Dupuis, S., Ramirez-Alejo, N., Li, Z., et al. Tuberculosis and impaired IL-23-dependent IFN gamma immunity in humans homozygous for a common TYK2 missense variant. Science Immunology. 2018. 3(30): p. eaau8714.

36. Ohta, A., Ogawa, E., Murata, M., et al. Impact of the PNPLA3 genotype on the risk of hepatocellular carcinoma after hepatitis C virus eradication. Journal of Medical Virology. 2022. 94(10): pp. 5007-5014.

37. Wang, Q., Chang, B., Li, X., et al. Role of ALDH2 in Hepatic Disorders: Gene Polymorphism and Disease Pathogenesis. J Clin Transl Hepatol. 2021. 9(1): pp. 90-98.

38. Shen, X. F., Zeng, X. T., Jian, Z. Y., et al. Quantitative assessment of the effect of epidermal growth factor 61A/G polymorphism on the risk of hepatocellular carcinoma. Oncol Lett. 2015. 10(5): pp. 3199-3205.

39. De Re, V., Caggiari, L., De Zorzi, M., et al. Genetic diversity of the KIR/HLA system and susceptibility to hepatitis C virus-related diseases. PLoS One. 2015. 10(2): p. e0117420.

40. Kousathanas, A., Pairo-Castineira, E., Rawlik, K., et al. Whole-genome sequencing reveals host factors underlying critical COVID-19. Nature. 2022. 607(7917): pp. 97-103.

41. Bradbury, C. A., McQuilten, Z. Anticoagulation in COVID-19. The Lancet. 2022. 399(10319): pp. 5-7.

42. Fätkenheuer, G., Pozniak, A. L., Johnson, M. A., et al. Efficacy of short-term monotherapy with maraviroc, a new CCR5 antagonist, in patients infected with HIV-1. Nature Medicine. 2005. 11(11): pp. 1170-1172.

43. Dean, L. 'In:' Maraviroc Therapy and CCR5 Genotype. 2012.

44. Richardson, P., Griffin, I., Tucker, C., et al. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. Lancet. 2020. 395(10223): pp. e30-e31.

45. Möller, M., Kinnear, C. J. Human global and population-specific genetic susceptibility to Mycobacterium tuberculosis infection and disease. Current Opinion in Pulmonary Medicine. 2020. 26(3): pp. 302-310.

46. Scully, E. P. Sex, gender and infectious disease. Nat Microbiol. 2022. 7(3): pp. 359-360.

47. Croock, D., Swart, Y., Schurz, H., et al. Confirmation of HLA-II associations with TB susceptibility in admixed African samples. 2024.

48. Tan, T., Atkinson, E. G. Strategies for the Genomic Analysis of Admixed Populations. Annu Rev Biomed Data Sci. 2023. 6: pp. 105-127.

49. Gopalan, S., Smith, S. P., Korunes, K., et al. Human genetic admixture through the lens of population genomics. Philos Trans R Soc Lond B Biol Sci. 2022. 377(1852): p. 20200410.

50. Litvinjenko, S., Magwood, O., Wu, S., et al. Burden of tuberculosis among vulnerable populations worldwide: an overview of systematic reviews. Lancet Infect Dis. 2023. 23(12): pp. 1395-1407.

51. Nayak, B. K. Understanding the relevance of sample size calculation. Indian J Ophthalmol. 2010. 58(6): pp. 469-70.

52. N., Vivek. The Role of Host Genetics (and Genomics) in Tuberculosis. Microbiology Spectrum. 2016. 4(5): p. 4.5.26.

53. Schurz, H., Naranbhai, V., Yates, T. A., et al. Multi-ancestry meta-analysis of host genetic susceptibility to tuberculosis identifies shared genetic architecture. eLife. 2024. 13: p. e84394.

54. Houben, R. M., Dodd, P. J. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLoS Med. 2016. 13(10): p. e1002152.

55. Vukcevic, D., Traherne, J. A., Naess, S., et al. Imputation of KIR Types from SNP Variation Data. Am J Hum Genet. 2015. 97(4): pp. 593-607.

56. Santiago-Lamelas, L., Castro-Santos, P., Carracedo, A., et al. Unveiling the Significance of HLA and KIR Diversity in Underrepresented Populations. Biomedicines. 2024. 12(6).

57. Arora, J., Pierini, F., McLaren, P. J., et al. HLA Heterozygote Advantage against HIV-1 Is Driven by Quantitative and Qualitative Differences in HLA Allele-Specific Peptide Presentation. Molecular Biology and Evolution. 2020. 37(3): pp. 639-650.

58. Arora, J., McLaren, P. J., Chaturvedi, N., et al. HIV peptidome-wide association study reveals patient-specific epitope repertoires associated with HIV control. Proceedings of the National Academy of Sciences. 2019. 116(3): pp. 944-949.

59. McLaren, P. J., Carrington, M. The impact of host genetic variation on infection with HIV-1. Nature Immunology. 2015. 16(6): pp. 577-583.

60. Martin, M. P., Gao, X., Lee, J., et al. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. Nature Genetics. 2002. 31(4): pp. 429-434.

61. High priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting. World Health Organisation (WHO). 2014; Available from: https://www.who.int/publications/i/item/WHO-HTM-TB-2014.18.

62. Chang, A., Loy, C. J., Eweis-LaBolle, D., et al. Circulating cell-free RNA in blood as a host response biomarker for detection of tuberculosis. Nat Commun. 2024. 15(1): p. 4949.

63. Gupta, R. K., Peppa, D., Hill, A. L., et al. Evidence for HIV-1 cure after CCR5Δ32 allogeneic haemopoietic stem-cell transplantation 30 months post analytical treatment interruption: a case report. The Lancet HIV. 2020. 7(5): pp. e340-e347.

64. Hütter, G., Nowak, D., Mossner, M., et al. Long-Term Control of HIV by CCR5 Delta32/Delta32 Stem-Cell Transplantation. New England Journal of Medicine. 2009. 360(7): pp. 692-698.

65. Study of EBT-101 in aviremic HIV-1 infected adults on stable ART. ClinicalTrials.gov. 2023; Available from: https://classic.clinicaltrials.gov/ct2/show/NCT05144386.

66. Freen-van Heeren, J. J. Closing the Door with CRISPR: Genome Editing of CCR5 and CXCR4 as a Potential Curative Solution for HIV. BioTech (Basel). 2022. 11(3).

67. Dai, W., Wu, F., McMyn, N., et al. Genome-wide CRISPR screens identify combinations of candidate latency reversing agents for targeting the latent HIV-1 reservoir. Sci Transl Med. 2022. 14(667): p. eabh3351.

68. Dash, P. K., Chen, C., Kaminski, R., et al. CRISPR editing of CCR5 and HIV-1 facilitates viral elimination in antiretroviral drug-suppressed virus-infected humanized mice. Proc Natl Acad Sci U S A. 2023. 120(19): p. e2217887120.

69. HIV and AIDS. World Health Organisation (WHO). 2023; Available from: https://www.who.int/news-room/fact-sheets/detail/hiv-aids.

70. Hasso-Agopsowicz, M., Hwang, A., Hollm-Delgado, M. G., et al. Identifying WHO global priority endemic pathogens for vaccine research and development (R&D) using multi-criteria decision analysis (MCDA): an objective of the Immunization Agenda 2030. EBioMedicine. 2024. p. 105424.

71. Yang, H., Rei, M., Brackenridge, S., et al. HLA-E-restricted, Gag-specific CD8+ T cells can suppress HIV-1 infection, offering vaccine opportunities. Science Immunology. 2021. 6(57): p. eabg1703.

72. Phelan, J., Gomez-Gonzalez, P. J., Andreu, N., et al. Genome-wide host-pathogen analyses reveal genetic interaction points in tuberculosis disease. Nat Commun. 2023. 14(1): p. 549.

73. Ansari, M. A., Pedergnana, V., C, L. C. Ip, et al. Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. Nat Genet. 2017. 49(5): pp. 666-673.

Acknowledgements

The desk-based research underpinning our analysis was supplemented by interviews. We thank the following interviewees for their time, engagement, and expert insights:

- Professor Jacques Fellay, Global Health Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
- Professor Sarah Dunston, NHMRC Leadership Fellow and Dame Kate Campbell Fellow in the Department of Infectious Diseases, The University of Melbourne, Australia
- Professor Marlo Möller, Tuberculosis Host Genetics group, Division of Molecular Biology & Human Genetics, University of Stellenbosch, South Africa
- Professor Megan Murray, Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, USA
- Dr Sara Clohisey Hendry, Royal (Dick) School of Veterinary Studies, Roslin Institute, University of Edinburgh, UK
- Professor William Irving, Division of Microbiology and Infectious Diseases, School of Life Sciences, University of Nottingham, UK

PHG FOUNDATION

making science work for health

The PHG Foundation is a non-profit think tank with a special focus on how genomics and other emerging health technologies can provide more effective, personalised healthcare and deliver improvements in health for patients and citizens.

intelligence@phgfoundation.org

UNIVERSITY OF CAMBRIDGE