Review Article

Human Genomic Loci Important in Common Infectious Diseases: Role of High-Throughput Sequencing and Genome-Wide Association Studies

Gerald Mboowa [iD],[1,2] **Ivan Sserwadda,**[1] **Marion Amujal,**[2] **and Norah Namatovu**[2]

[1]*Department of Immunology and Molecular Biology, College of Health Sciences, Makerere University, P.O. Box 7072, Kampala, Uganda*
[2]*Department of Medical Microbiology, College of Health Sciences, Makerere University, P.O. Box 7072, Kampala, Uganda*

Correspondence should be addressed to Gerald Mboowa; gmboowa@gmail.com

Academic Editor: José Ramón Blanco

HIV/AIDS, tuberculosis (TB), and malaria are 3 major global public health threats that undermine development in many resource-poor settings. Recently, the notion that positive selection during epidemics or longer periods of exposure to common infectious diseases may have had a major effect in modifying the constitution of the human genome is being interrogated at a large scale in many populations around the world. This positive selection from infectious diseases increases power to detect associations in genome-wide association studies (GWASs). High-throughput sequencing (HTS) has transformed both the management of infectious diseases and continues to enable large-scale functional characterization of host resistance/susceptibility alleles and loci; a paradigm shift from single candidate gene studies. Application of genome sequencing technologies and genomics has enabled us to interrogate the host-pathogen interface for improving human health. Human populations are constantly locked in evolutionary arms races with pathogens; therefore, identification of common infectious disease-associated genomic variants/markers is important in therapeutic, vaccine development, and screening susceptible individuals in a population. This review describes a range of host-pathogen genomic loci that have been associated with disease susceptibility and resistant patterns in the era of HTS. We further highlight potential opportunities for these genetic markers.

1. Introduction

HIV/AIDS, tuberculosis, and malaria are 3 major global public health threats causing substantial morbidity, mortality, negative socioeconomic impact, and human suffering [1]. Whole-genome sequencing (WGS) of hosts and their cognate pathogens has transformed our understanding of the contribution of genomics in infectious disease processes. As the antibiotic era changed our understanding of infections so has the genomic revolution. The host-pathogen coevolutionary "arms race" is a phenomenon that has been described to result from interaction within host innate, adaptive immune responses, exposure to antibiotics, and competition with commensal microbiota [2, 3]. When positive selection in one geographical region causes large allele frequency differences between populations than those expected for neutrally evolving alleles, high frequency of the derived allele (i.e., when a new allele increases to a frequency higher than that expected under genetic drift) [4] and HTS technologies are harnessed and used to decipher the nature and dynamics of these signals. The dynamics of host-pathogen interactions (e.g., length of exposure, geographical spread, morbidity, mortality, and cooccurring environmental events) influence the genetic architecture of resistance variants in modern populations [4]. The application of genomic sequence-based approaches to understand the host-pathogen interface continues to provide us with important clues for our survival. Human genetic variation is a major determinant of genetic susceptibility to many common infectious diseases [5]. Malaria, HIV/AIDS, and tuberculosis are some of the common

TABLE 1: Host loci strongly associated with the three common infectious diseases.

Disease	Population	Variant	Annotation	Genome-wide significance (P value)	Association	References
HIV-1/AIDS	African Americans	rs2523608	HLA-B*5703	5.6×10^{-10}	Viral load set point	[6]
	European ancestry	rs2395029	HCP5/B*5701	9.7×10^{-26}	Long-term nonprogression and viral load set point	[7–9]
	European ancestry	rs9264942	HLA-C	2.8×10^{-35}	Long-term nonprogression and viral load set point	[7]
	European ancestry	rs4418214	MICA	1.4×10^{-34}	Long-term nonprogression	
	European ancestry	rs3131018	PSORS1C3	4.2×10^{-16}	Long-term nonprogression	[7]
	African American	rs2255221	Intergenic	3.5×10^{-14}	Long-term nonprogression	
		rs2523590	HLA-B	1.7×10^{-13}	Long-term nonprogression	
		rs9262632	Intergenic	1.0×10^{-8}	Long-term nonprogression	
	European ancestry	rs9261174	ZNRD1, RNF39	1.8×10^{-8}	Disease progression	[8, 10]
		rs11884476	PARD3B	3.4×10^{-9}	Disease progression	[11]
		rs2234358	CXCR6	9.7×10^{-10}	Long-term nonprogression	[12]
Malaria	Ghanaian	rs10900585	ATP2B4	6.1×10^{-9}	Protective	[13]
	Ghanaian	rs2334880	MARVELD3	3.9×10^{-8}	Protective	[13]
	Ghanaian	rs8176719	ABO	2.9×10^{-13}	Protective	[13]
	Gambian	rs11036238	HBB	3.7×10^{-11}	Susceptibility	[14]
	Gambian	rs334	HBB	4×10^{-14}	Protective	[13, 14]
Tuberculosis	Moroccan	rs17590261	AGMO	2×10^{-6}	Age-at-onset of TB	[15]
	Ghanaian and Gambian	rs4331426	18q11.2	6.8×10^{-9}	Susceptibility	[16]

infectious diseases in which a range of genetic susceptibilities and resistant conferring loci have been identified using both traditional molecular-based approaches and HTS technologies. HTS has enabled us to identify genomic signatures from these interactions. These markers identified through infectious disease genome-wide association studies (GWASs) have considerable significances that can be exploited to understand host protective mechanisms against pathogens and identify new molecular targets for diagnostic, prophylactic, and therapeutic interventions (Table 1).

In the past, candidate gene studies have been used to identify disease susceptibility genetic loci for many major human infections. But with the advent of HTS approaches, many new loci have been and continue to be identified in diverse populations. A paradigm shift from candidate gene studies to GWAS and HTS has ushered in a "big data" genomic revolution era enabling us to redefine the genomic architecture of host-pathogen disease susceptibility. Both exome and whole-genome sequencing approaches are proving more successful and affordable. Host genetics influence clinical course of infectious diseases as well as genetic variation of the pathogens determining their survival in presence of selective pressure from the host and environment (antimicrobials) and identify genetic markers of drug-resistant pathogens and parasites. Furthermore, HTS has given us unprecedented resolution of understanding the role of host genetics to infectious diseases susceptibility. This genomic revolution has generated data informative for understanding the frequency of many genetic traits, including those that cause disease susceptibility in African populations and populations of recent African descent [17].

Pathogens have always been a major cause of human mortality, so they impose strong selective pressure on the human genome [18, 19]. HTS applied to screening populations of host immune-specific cells and their respective pathogens can highlight the host-pathogen unique genetic signatures important in host-pathogen coevolution, profiling immunological history, pathogen-induced immunodominance genetic patterns, predicting clinical outcomes of common infections (such as HIV/AIDS disease progression phenotypes like long-term nonprogressors and rapid progressors, as well as highly exposed persistently seronegative group), rapid diagnosis plus screening outbreaks involving Risk Group 4 highly infectious pathogens, and genetic characterization of live-attenuated vaccine vectors (Figures 1(a) and 1(b)).

GWASs demand recruitment of large well-phenotyped clinical cohorts within appropriate study designs and settings. However, they are very expensive, and therefore few funding agencies are able to finance such studies yet they may offer unique opportunities to unveil genetically important signals. Currently, there are a growing number of communicable disease-specific research initiatives that are specifically interested in looking at the stages of the infections and GWAS of disease progression. A classic example is the Collaborative African Genomics Network (CAfGEN), a H3Africa-funded consortium probing host genetic factors that are important to the progression of HIV and HIV-TB infection in sub-Saharan African children (https://www.h3africa.org/consortium/projects/16-projects/89-cafgen). CAfGEN is specifically looking at both rapid and long-term HIV/AIDS progression status while
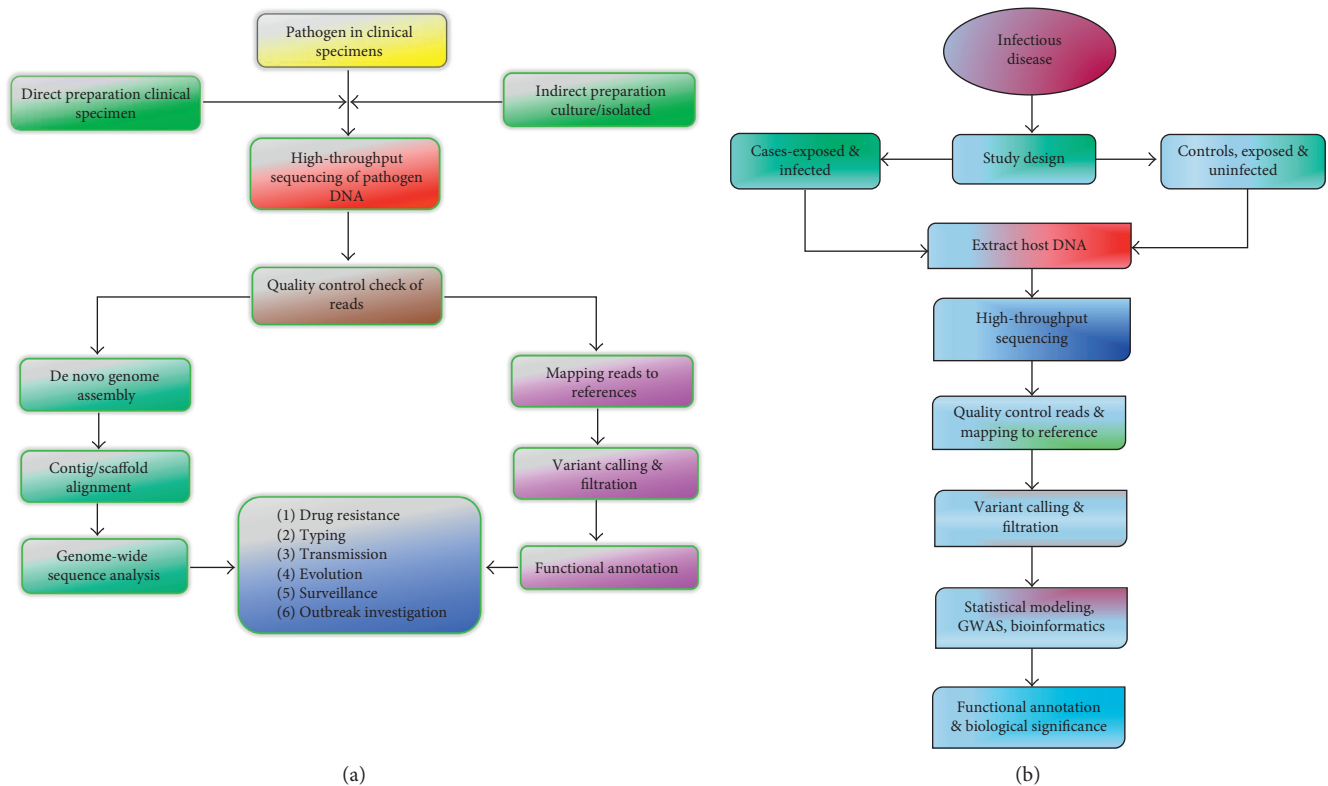
(a)

(b)

FIGURE 1: (a) Pipeline for interrogation of pathogen genomes using high-throughput sequencing and computational approaches. DNA extraction for HTS can be done from either direct clinical specimen of individuals who are suspected to be infected with the disease or from enriched/isolated cultures. Quality control and read preprocessing are critical steps in the analysis of datasets generated from high-throughput sequencing technologies. FASTQC is an example of a tool for general quality assessment of HTS data from all technologies. Genomes can be recreated with no prior knowledge using de novo sequence assembly as well as recreating the genome using prior knowledge based on a reference genome—alignment/mapping. The former is necessary for novel genomes and where the sequenced genome differs from reference. Sequence data analysis is important in infectious disease outbreak investigations, molecular typing, antimicrobial drug resistance, transmission, surveillance, and microbial evolution. (b) Pipeline for interrogation of host genomes using high-throughput sequencing and computational approaches. For a given infectious disease in a population, an appropriate study design is determined and host DNA is collected from cases (exposed to pathogen and infected) and controls (exposed to pathogen and uninfected). HTS of DNA from both cases and control is performed. Quality control (QC) procedures vary in different pipelines. These include QC on individuals for missingness, gender checks, duplicates and cryptic relatedness, population outliers, heterozygosity and inbreeding, QC on SNPs for missingness, minor allele frequency, and Hardy–Weinberg equilibrium. Many of these are computationally intensive, operationally challenging, and constantly evolving. Genome-wide association studies (GWASs) involving case-control studies compare the frequencies of common genetic variants, assume an appropriate statistical model, and account for multiple testing correction threshold to identify susceptibility and protective polymorphisms in the population.

utilizing a unique protocol that applies exome sequencing of AIDS extreme clinical phenotypes. The study designs and the protocols being utilized are valuable resources for future genomic research involving common infectious diseases. In this review, we describe the impact of HTS and genomics in understanding the human host-pathogen interaction.

2. Common Infectious Diseases

Infectious diseases account for 15 million deaths per year worldwide, and disproportionately affect the young, elderly people, and the poorest sections of society making them a high priority [20]. The World Health Organization in 2016 estimated the global mortality for tuberculosis, HIV/AIDS, and malaria to be at 3.23 million with most deaths occurring in sub-Saharan Africa. This region has continued to lead in

both prevalence and incidence of these major infectious killer diseases [21]. Research investment in infectious diseases was poorly matched. Data show that funding does not correspond closely with burden [22]. Review of findings to date suggests that the genetic architecture of infectious disease susceptibility may be importantly different from that of noninfectious diseases [20]. Other authors have extensively reviewed the genetic susceptibility to diseases [4, 5, 13, 16, 18–20, 23–49]. The ancient biological "arms race" between microbial pathogens and humans has shaped genetic variation in modern populations, and this has important implications for the growing field of medical genomics [4]. As humans migrated throughout the world, populations encountered distinct pathogens, and natural selection increased the prevalence of alleles that are advantageous in the new ecosystems in both host and pathogens [4]. Temporal patterns supporting

evidence of host-pathogen coevolution have been reported [50]. Common infectious diseases have shown geographical disparities, for example, *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages [51]. Molecular epidemiological studies show that, with the exception of sub-Saharan Africa, almost all HIV-1 subtypes, circulating recombinant forms, and several unique recombinant forms have been detected, but there is a specific geographic distribution pattern for HIV-1 subtypes [52–54]. HIV diversity plays a central role in the HIV pandemic [55] and has significant implications for diagnosis, vaccine development, and clinical management of patients [56]. High levels of *Plasmodium falciparum* malaria endemicity are common in Africa [57]. Formal characterization of disease-causing agents has always been fundamental to understanding the evolution of pathogens and the epidemiology of infectious disease. Linking this information with temporal, spatial, and clinical data can bring understanding of evolution, geographical spread, and disease associations for the pathogen, providing vital information for identifying sources of infection as well as designing interventions to prevent and treat disease [58].

3. High-Throughput Sequencing (HTS) and Computational Tools

New sequencing technology has enabled the identification of thousands of single nucleotide polymorphisms in the exome, and many computational and statistical approaches to identify disease association signals have emerged [59]. Foremost is a better understanding of disease pathogenesis and resistance in the expectation that this will lead in time to improved interventions such as better drugs or vaccines to prevent or attenuate the great global burden of infectious disease morbidity and mortality. With over 10 million deaths annually from infectious diseases and the threat of new epidemics and pandemics, this is a very high priority [20]. During the past decade, we have also witnessed the emergence of many new pathogens not previously detected in humans, such as the avian influenza virus, severe acute respiratory syndrome (SARS), and Ebola [60]. Now that the scientific community has access to complete genomes of infectious diseases though application of HTS, priority should be the dissection of host-pathogen interactions through development of powerful computational tools. HTS has profoundly altered our understanding of human diversity and disease. The interaction between hosts and pathogens is a coevolution process which may simply be described as "shooting a moving target." HTS and computational modelling tools will offer potential to understand this interaction leading to better vaccine designs and therapeutic targets. Sanger DNA sequencing is limited in throughput and high cost as compared to HTS platforms, which differ in their details but typically follow a similar general paradigm: template preparation, clonal amplification, followed by cyclical rounds of massively parallel sequencing [61]. Nanopore-based sequencing approaches such as single molecule, real-time (SMRT) sequencing technologies have been developed and consistently produce some of the longest average read lengths compared to HTS.

SMRT sequencing is particularly useful for projects involving de novo assembly of small bacterial and viral genomes as well as large genome finishing [62]. Many important sequencing platforms are reviewed in great depth [61, 63].

Computational tools are an important integral part of genomics. New computational methods are constantly being developed to collect, process, and extract useful biological information from a variety of samples and complex datasets [64]. The scale and complexity of genomic data is ever-expanding, requiring biologists to apply increasingly more sophisticated computational tools in the generation, analyses, interpretation, and storage of this data. The data are generated in different sizes, formats, and structures requiring a wide range of tools to manipulate. Despite the level of specialization needed in bioinformatics, it is important that life-scientists have a good understanding of it for a correct experimental design which allows them to reveal the information in a metagenome [64]. HTS technologies are generating an astonishing amount of unprecedented information in the history of Biology which has spurred a Biomedical Big Data to Knowledge (B2D2K) revolution. Thus, a new exhilarating rapidly evolving scientific field, Bioinformatics (Biology meets computer programming), has recently emerged and uses novel computational approaches to help solve important biological problems. Bioinformatics is a set of activities: data acquisition, database development, data analysis, data integration, and analysis of integrated data. The majority of available bioinformatics software requires some knowledge of the text-based command line of the UNIX or Linux operating systems, allowing custom programming scripts and pipelines to automate data manipulation and analysis in a single step [65]. Although bioinformatics tools/software are both "open sources" and commercially available, clinicians have limited bioinformatics knowledge [66–69]. It is clear that user-friendly bioinformatics pipelines are key to facilitating more widespread use of WGS, with more widespread bioinformatics expertise [65].

4. Signatures of Selection on the Genomes

Positive selection (also known as Darwinian selection) in genes and genomes can point to the evolutionary basis for differences among species and among races within a species [70]. Many other aspects of human biology not necessarily related to the "branding" of our species, for instance, host-pathogen interactions, reproduction, dietary adaptation, and physical appearance, have also been the substrate of varying levels of positive selection. Comparative genetics/genomics studies in recent years have uncovered a growing list of genes that might have experienced positive selection during the evolution of humans and/or primates [71]. Microbial genome evolution is shaped by a variety of selective pressures. Understanding how these processes occur can help to address important problems in microbiology by explaining observed differences in phenotypes, including virulence and resistance to antibiotics. Greater access to whole-genome sequencing provides microbiologists with the opportunity to perform large-scale analyses of selection in novel settings such as within individual hosts [72].
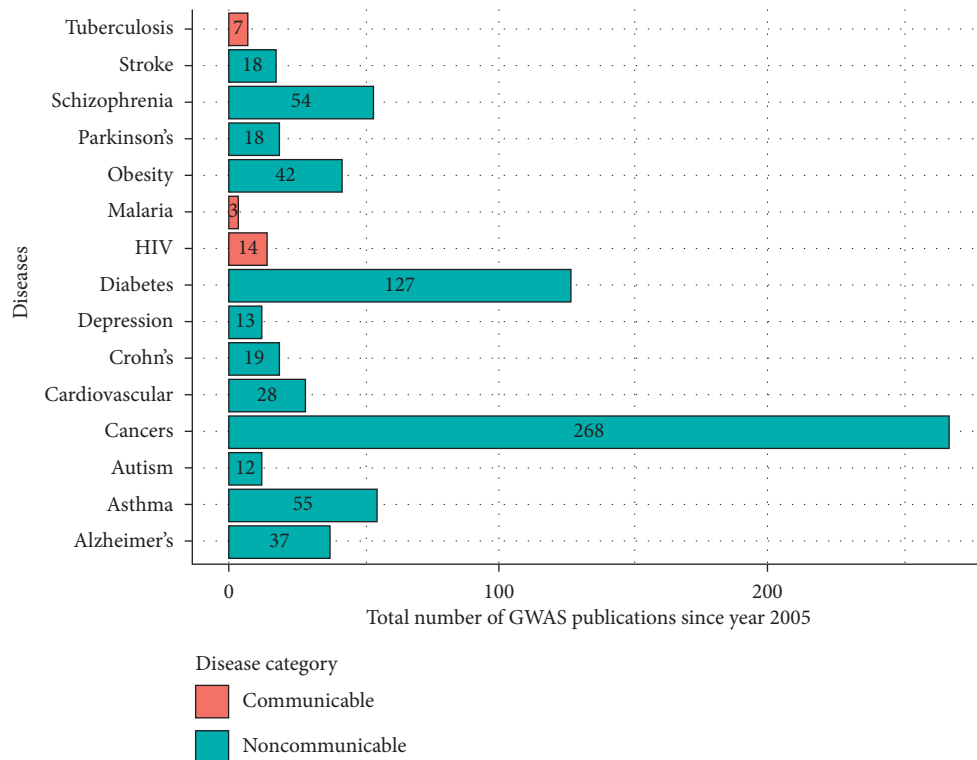
FIGURE 2: Comparison of selected communicable and noncommunicable disease GWASs since 2005.

Identification of positive selection genetic signatures in the genomes can help us to understand the kinetics and directions of continuing host-pathogen coadaptation and impact on their diagnosis, transmission, fitness, immunogenicity, and pathogenicity. Given the potential for strong selective pressure, that genetic programs controlling host-pathogen interactions in humans and other species are littered with signatures of positive selection [73]. The high mortality and widespread impact of malaria has resulted in this disease being the strongest evolutionary selective force in recent human history, and the genes that confer resistance to malaria provide some of the best-known case studies of strong positive selection in modern humans [27]. A number of specific genomic variants including β-globin locus, G6PD deficiency, Duffy, ovalocytosis, ABO, and human leukocyte antigen confer resistance to malaria in the human host. Elevated frequencies of hemoglobinopathies such as thalassemia and sickle cell disease, which are caused by mutations at the β-globin locus, are maintained via balancing selection ("the malarial hypothesis") [74–76]. This hypothesis suggests that some human diseases such as thalassemia are polymorphisms which provide heterozygote advantage because of the trade-offs between the advantages of resistance to malaria and negative effects due to the disease [77], where the hemoglobin S (HbS) homozygote disadvantage is recompensed through the malaria resistance of the heterozygote (HbAS) in regions of malaria endemicity [78, 79]. The Δ32 mutation at the CCR5 locus is a well-studied example of natural selection acting in humans. The mutation is found principally in Europe and western Asia, with higher frequencies generally in the north. Homozygous

carriers of the Δ32 mutation are resistant to HIV-1 infection because the mutation prevents functional expression of the CCR5 chemokine receptor normally used by HIV-1 to enter CD4+ T cells [80–87]. Host genetic factors play important roles in susceptibility to tuberculosis infection, and different gene polymorphisms in different ethnicity and genetic backgrounds may lead to different effects on tuberculosis risk [88]. Polymorphisms in natural resistance-associated macrophage protein 1 (NRAMP1), toll-like receptor 2 (TLR2), interleukin-6 (IL-6), tumor necrosis factor-alpha (TNF-α), interleukin-1 receptor antagonist (IL-1RA), IL-10, vitamin D receptor (VDR), dendritic cell-specific ICAM-3-grabbing nonintegrin (DC-SIGN), monocyte chemo-attractant protein-1 (MCP-1), nucleotide oligomerization binding domain 2 (NOD2), interferon-gamma (IFN-γ), inducible nitric oxide synthase (iNOS), mannose-binding lectin (MBL), and surfactant proteins A (SP-As) genes have been variably associated with tuberculosis infection, and there is strong evidence indicating that host genetic factors play critical roles in tuberculosis susceptibility, severity, and development among different populations [89–91]. Several NRAMP1 polymorphisms were significantly associated with PTB in African and Asian populations, but not in populations of European descent [92, 93].

5. Genome-Wide Association Studies in Infectious Diseases

GWASs are based on the "common disease, common variant" hypothesis, and they have been performed largely using single nucleotide polymorphism (SNP) arrays that focus

only on common genetic polymorphisms (for which the minor allele frequency is >5%) [35, 94–96]. GWAS approach has potential to provide candidates for the development of control measures against infectious diseases in humans [13]. For over 50 years, candidate gene studies have been used to identify loci for many major causes of human infectious mortality, including malaria, tuberculosis, and HIV-1 [20]. The first successful GWAS was published in 2005 ushering genome-wide approaches that have identified loci in diverse populations. Common genetic variants have also been demonstrated to regulate susceptibility/resistance to infectious diseases, for example, the CCR5Δ32 polymorphism that modulates HIV/AIDS disease progression [97]. Genome-wide association study approaches are being increasingly utilized to define genetic variants underlying susceptibility to major infectious diseases [34]. Infectious diseases follow a series of stages right from acquisition, disease development, rate of progression, convalescence, and asymptomatic carrier state. Therefore, every stage will be influenced by one or a set of mutations in a population or individual. Different populations will have mutations that affect diseases at different stages. This is where GWAS has and will continue to play an important role in identifying these mutations. A recent study suggested that host genetic risk in TB is depended upon the pathogen's genetic background and demonstrated the importance of analyzing the interaction between host and pathogen genomes in TB [98]. Studies are exploiting unique designs like extreme phenotype designs to identify complex trait genomic loci, while others have identified genetic associations of infectious diseases by integrating estimation of population admixture events to detect disease susceptibility loci after teasing out different ancestries and allelic, genotypic, or haplotype risk ratios.

A genomic database for all mutations identified to be conferring resistance to infectious diseases in different populations is a vital product of more than 10 years of GWAS. More studies should be appropriately designed to identify new potential infectious disease resistance-conferring mutations in human hosts. We searched the PubMed database for studies published since January 2005 using the terms "malaria," "tuberculosis," "HIV/AIDS," "genome-wide association." Search terms included combinations of ((disease-query[Title] AND Genome-wide association[Title])) where a disease-query was either a communicable or noncommunicable disease. Two search restrictions were set: publication date set (from 2005/01/01) and species (humans). Figure 2 indicates more GWASs carried out in noncommunicable diseases than infectious diseases. There is an urgent need for a major increase in funding for communicable disease control in the developing world and for more balanced allocation of the resources already provided [22].

6. Future Directions

Genomics and whole-genome sequencing have the capacity to greatly enhance knowledge and understanding of infectious diseases and clinical microbiology [65]. Human genetics is an indispensable tool for enhancing the understanding of the molecular basis of many common

diseases [99]. Over 4,500 SNPs have been associated with a variety of human traits and complex diseases [59]. We envisage that with the reducing costs, HTS and genomics will become an indispensable component of every healthcare system. HTS and computational tools offer a potential to stratify populations for risk of infectious disease based on genomic profiling thereby prioritizing interventions such as vaccines and therapeutics to the "most-at-risk" populations since there is no "one-size-fits-all" approach to treating infectious diseases. With the growing number of sequencing facilities on every continent, the future will offer a less costly approach that will integrate a genomic profile in routine patient management, improving management of diseases, and therapeutic development.

It is less likely to find a population or individual who carries mutations conferring resistance to an infectious disease at every stage of the infection. Different individuals have different mutations that offer resistance to different stages of the infections. A genomic catalogue of these mutations identified through HTS, computational tools, and GWAS now combined with the rapidly growing genome-editing technology known as CRISPR/Cas9 will enable introduction of an array of disease-stage-specific resistance-conferring mutations. This will offer interventions at all levels of the disease process unlike traditional vaccines.

Disclosure

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All authors participated in writing the manuscript. They further reviewed and approved the final manuscript.

Acknowledgments

References

[1] M. Vitoria, R. Granich, C. F. Gilks et al., "The global fight against HIV/AIDS, tuberculosis, and malaria: current status and future perspectives," American Journal of Clinical Pathology, vol. 131, no. 6, pp. 844–848, 2009.

[2] K. A. Bliven and A. T. Maurelli, "Evolution of bacterial pathogens within the human host," *Microbiology Spectrum*, vol. 4, no. 1, 2016.

[3] R. Dawkins and J. R. Krebs, "Arms races between and within species," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 205, no. 1161, pp. 489–511, 1979.

[4] E. K. Karlsson, D. P. Kwiatkowski, and P. C. Sabeti, "Natural selection and infectious disease in human populations," *Nature Reviews Genetics*, vol. 15, no. 6, pp. 379–393, 2014.

[5] A. V. Hill, "Genetics and genomics of infectious disease susceptibility," *British Medical Bulletin*, vol. 55, no. 2, pp. 401–413, 1999.

[6] K. Pelak, D. B. Goldstein, N. M. Walley et al., "Host determinants of HIV-1 control in African Americans," *Journal of Infectious Diseases*, vol. 201, no. 8, pp. 1141–1149, 2010.

[7] The International HIV Controllers Study, "The major genetic determinants of HIV-1 control affect HLA class I peptide presentation," *Science*, vol. 330, no. 6010, pp. 1551–1557, 2010.

[8] J. Fellay, K. V. Shianna, D. Ge et al., "A whole-genome association study of major determinants for host control of HIV-1," *Science*, vol. 317, no. 5840, pp. 944–947, 2007.

[9] S. Limou, S. L. Clerc, C. Coulonges et al., "Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02)," *Journal of Infectious Diseases*, vol. 199, no. 3, pp. 419–426, 2009.

[10] J. Fellay, D. Ge, K. V. Shianna et al., "Common genetic variation and the control of HIV-1 in humans," *PLoS Genetics*, vol. 5, no. 12, p. e1000791, 2009.

[11] J. L. Troyer, G. W. Nelson, J. A. Lautenberger et al., "Genome-wide association study implicates PARD3B-based AIDS restriction," *Journal of Infectious Diseases*, vol. 203, no. 10, pp. 1491–1502, 2011.

[12] S. Limou, C. Coulonges, J. T. Herbeck et al., "Multiple-cohort genetic association study reveals CXCR6 as a new chemokine receptor involved in long-term nonprogression to AIDS," *Journal of Infectious Diseases*, vol. 202, no. 6, pp. 908–915, 2010.

[13] C. Timmann, T. Thye, M. Vens et al., "Genome-wide association study indicates two novel resistance loci for severe malaria," *Nature*, vol. 489, no. 7416, pp. 443–446, 2012.

[14] M. Jallow, Y. Y. Teo, K. S. Small et al., "Genome-wide and fine-resolution association analysis of malaria in West Africa," *Nature Genetics*, vol. 41, no. 6, pp. 657–665, 2009.

[15] A. Grant, A. Sabri, A. Abid et al., "A genome-wide association study of pulmonary tuberculosis in Morocco," *Human genetics*, vol. 135, no. 3, pp. 299–307, 2016.

[16] T. Thye, F. O. Vannberg, S. H. Wong et al., "Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11. 2," *Nature Genetics*, vol. 42, no. 9, pp. 739–741, 2010.

[17] F. Gomez, J. Hirbo, and S. A. Tishkoff, "Genetic variation and adaptation in Africa: implications for human evolution and disease," *Cold Spring Harbor Perspectives in Biology*, vol. 6, no. 7, p. a008524, 2014.

[18] L. B. Barreiro and L. Quintana-Murci, "From evolutionary genetics to human immunology: how selection shapes host defence genes," *Nature Reviews Genetics*, vol. 11, no. 1, pp. 17–30, 2010.

[19] R. Cagliani and M. Sironi, "Pathogen-driven selection in the human genome," *International Journal of Evolutionary Biology*, vol. 2013, Article ID 204240, 6 pages, 2013.

[20] A. V. Hill, "Evolution, revolution and heresy in the genetics of infectious disease susceptibility," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1590, pp. 840–849, 2012.

[21] G. Mboowa, "Genetics of sub-Saharan African human population: implications for HIV/AIDS, tuberculosis, and malari," *International Journal of Evolutionary Biology*, vol. 2014, Article ID 108291, 8 pages, 2014.

[22] J. Shiffman, "Donor funding priorities for communicable disease control in the developing world," *Health Policy and Planning*, vol. 21, no. 6, pp. 411–420, 2006.

[23] D. Weatherall, J. Clegg, and D. Kwiatkowski, "The role of genomics in studying genetic susceptibility to infectious disease," *Genome Research*, vol. 7, no. 10, pp. 967–973, 1997.

[24] A. V. Hill, "The genomics and genetics of human infectious disease susceptibility," *Annual Review of Genomics and Human Genetics*, vol. 2, no. 1, pp. 373–400, 2001.

[25] J.-L. Casanova and L. Abel, "The genetic theory of infectious diseases: a brief history and selected illustrations," *Annual Review of Genomics and Human Genetics*, vol. 14, no. 1, pp. 215–243, 2013.

[26] S. Eridani, "Sickle cell protection from malaria," *Hematology Reports*, vol. 3, no. 3, 2011.

[27] P. W. Hedrick, "Population genetics of malaria resistance in humans," *Heredity*, vol. 107, no. 4, pp. 283–304, 2011.

[28] K. K. Singh and S. A. Spector, "Host genetic determinants of HIV infection and disease progression in children," *Pediatric Research*, vol. 65, no. 5, pp. 55–63, 2009.

[29] B. D. Walker and G. Y. Xu, "Unravelling the mechanisms of durable control of HIV-1," *Nature Reviews Immunology*, vol. 13, no. 7, pp. 487–498, 2013.

[30] P. Kumar, "Long term non-progressor (LTNP) HIV infection," *Indian Journal of Medical Research*, vol. 138, no. 3, pp. 291–293, 2013.

[31] M. P. Martin and M. Carrington, "Immunogenetics of HIV disease," *Immunological Reviews*, vol. 254, no. 1, pp. 245–264, 2013.

[32] P. Singh, G. Kaur, G. Sharma, and N. K. Mehra, "Immunogenetic basis of HIV-1 infection, transmission and disease progression," *Vaccine*, vol. 26, no. 24, pp. 2966–2980, 2008.

[33] S. R. Grossman, K. G. Andersen, I. Shlyakhter et al., "Identifying recent adaptations in large-scale genomic data," *Cell*, vol. 152, no. 4, pp. 703–713, 2013.

[34] F. O. Vannberg, S. J. Chapman, and A. V. Hill, "Human genetic susceptibility to intracellular pathogens," *Immunological Reviews*, vol. 240, no. 1, pp. 105–116, 2011.

[35] S. J. Chapman and A. V. Hill, "Human genetic susceptibility to infectious disease," *Nature Reviews Genetics*, vol. 13, no. 3, pp. 175–188, 2012.

[36] S. J. O'Brien and S. L. Hendrickson, "Host genomic influences on HIV/AIDS," *Genome Biology*, vol. 14, no. 1, p. 201, 2013.

[37] P. R. Shea, K. V. Shianna, M. Carrington, D. B. Goldstein et al., "Host genetics of HIV acquisition and viral control," *Annual Review of Medicine*, vol. 64, no. 1, pp. 203–217, 2013.

[38] D. Burgner, S. E. Jamieson, and J. M. Blackwell, "Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better?," *Lancet Infectious Diseases*, vol. 6, no. 10, pp. 653–663, 2006.

[39] M. Clementi and E. Di Gianantonio, "Genetic susceptibility to infectious diseases," *Reproductive Toxicology*, vol. 21, no. 4, pp. 345–349, 2006.

[40] S. Segal and A. V. Hill, "Genetic susceptibility to infectious disease," *Trends in Microbiology*, vol. 11, no. 9, pp. 445–448, 2003.

[41] A. Driss, J. M. Hibbert, N. O. Wilson, S. A. Iqbal, T. V. Adamkiewicz, and J. K. Stiles, "Genetic polymorphisms

linked to susceptibility to malaria," *Malaria Journal*, vol. 10, no. 1, p. 271, 2011.

[42] L. Abel and A. J. Dessein, "Genetic epidemiology of infectious diseases in humans: design of population-based studies," *Emerging Infectious Diseases*, vol. 4, no. 4, pp. 593–603, 1998.

[43] R. S. Sobota, C. M. Stein, N. Kodaman et al., "A locus at 5q33.3 confers resistance to tuberculosis in highly susceptible individuals," *American Journal of Human Genetics*, vol. 98, no. 3, pp. 514–524, 2016.

[44] K. Ding, M. de Andrade, T. A. Manolio et al., "Genetic variants that confer resistance to malaria are associated with red blood cell traits in African-Americans: an electronic medical record-based genome-wide association study," *G3: Genes, Genomes, Genetics*, vol. 3, no. 7, pp. 1061–1068, 2013.

[45] S. N. Redmond, K. Eiglmeier, C. Mitri et al., "Association mapping by pooled sequencing identifies TOLL 11 as a protective factor against *Plasmodium falciparum* in Anopheles gambiae," *BMC Genomics*, vol. 16, no. 1, p. 779, 2015.

[46] Malaria Genomic Epidemiology Network, G. Band, K. A. Rockett, C. C. Spencer, and D. P. Kwiatkowski, "A novel locus of resistance to severe malaria in a region of ancient balancing selection," *Nature*, vol. 526, no. 7572, pp. 253–257, 2015.

[47] M. J. Newport and C. Finan, "Genome-wide association studies and susceptibility to infectious diseases," *Briefings in Functional Genomics*, vol. 10, no. 2, pp. 98–107, 2011.

[48] J. N. Hirschhorn and M. J. Daly, "Genome-wide association studies for common diseases and complex traits," *Nature Reviews Genetics*, vol. 6, no. 2, pp. 95–108, 2005.

[49] B. Penman, C. Buckee, S. Gupta, and S. Nee, "Genome-wide association studies in Plasmodium species," *BMC Biology*, vol. 8, no. 1, p. 90, 2010.

[50] M. E. Woolhouse, J. P. Webster, E. Domingo, B. Charlesworth, and B. R. Levin, "Biological and biomedical implications of the co-evolution of pathogens and their hosts," *Nature Genetics*, vol. 32, no. 4, pp. 569–577, 2002.

[51] D. Stucki, D. Brites, L. Jeljeli et al., "*Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages," *Nature Genetics*, vol. 48, no. 12, pp. 1535–1543, 2016.

[52] L. Buonaguro, M. Tornesello, and F. Buonaguro, "Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications," *Journal of Virology*, vol. 81, no. 19, pp. 10209–10219, 2007.

[53] S. Osmanov, C. Pattou, N. Walker, B. Schwardländer, and J. Esparza, "Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000," *Journal of Acquired Immune Deficiency Syndromes*, vol. 29, no. 2, pp. 184–190, 2002.

[54] J. Hemelaar, E. Gouws, P. D. Ghys, and S. Osmanov, "Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004," *AIDS*, vol. 20, no. 16, pp. W13–W23, 2006.

[55] M. M. Santoro and C. F. Perno, "HIV-1 genetic variability and clinical implications," *ISRN Microbiology*, vol. 2013, Article ID 481314, 20 pages, 2013.

[56] M. T. Pyne, J. Hackett, V. Holzmayer, and D. R. Hillyard, "Large-scale analysis of the prevalence and geographic distribution of HIV-1 non-B variants in the United States," *Journal of Clinical Microbiology*, vol. 51, no. 8, pp. 2662–2669, 2013.

[57] S. I. Hay, C. A. Guerra, P. W. Gething et al., "A world malaria map: *Plasmodium falciparum* endemicity in 2007," *PLoS Medicine*, vol. 6, no. 3, p. e1000048, 2009.

[58] S. D. Bentley and J. Parkhill, "Genomic perspectives on the evolution and spread of bacterial pathogens," *Proceedings of the Royal Society B: Biological Sciences*, vol. 282, no. 1821, p. 20150488, 2015.

[59] N. O. Stitziel, A. Kiezun, and S. Sunyaev, "Computational and statistical approaches to analyzing variants identified by exome sequencing," *Genome Biology*, vol. 12, no. 9, p. 227, 2011.

[60] E. C. Berglund, B. Nystedt, and S. G. Andersson, "Computational resources in infectious disease: limitations and challenges," *PLoS Computational Biology*, vol. 5, no. 10, p. e1000481, 2009.

[61] J. A. Reuter, D. V. Spacek, and M. P. Snyder, "High-throughput sequencing technologies," *Molecular Cell*, vol. 58, no. 4, pp. 586–597, 2015.

[62] A. C. English, S. Richards, Y. Han, M. Wang et al., "Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology," *PLoS One*, vol. 7, no. 11, article e47768, 2012.

[63] S. Goodwin, J. D. McPherson, and W. R. McCombie, "Coming of age: ten years of next-generation sequencing technologies," *Nature Reviews Genetics*, vol. 17, no. 6, pp. 333–351, 2016.

[64] A. Escobar-Zepeda, A. V.-P. de León, and A Sanchez-Flores, "The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics," *Frontiers in Genetics*, vol. 6, 2015.

[65] J. C. Kwong, N. Mccallum, V. Sintchenko, and B. P. Howden, "Whole genome sequencing in clinical and public health microbiology," *Pathology*, vol. 47, no. 3, pp. 199–210, 2015.

[66] D. Blankenberg, N. Coraor, G. Von Kuster, J. Taylor, and A. Nekrutenko, "Integrating diverse databases into an unified analysis framework: a Galaxy approach," *Database*, vol. 2011, p. bar011, 2011.

[67] R. Lazarus, A. Kaspi, and M. Ziemann, "Creating reusable tools from scripts: the Galaxy Tool Factory," *Bioinformatics*, vol. 28, no. 23, pp. 3139-3140, 2012.

[68] S. K. Gupta, B. R. Padmanabhan, S. M. Diene et al., "ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes," *Antimicrobial Agents and Chemotherapy*, vol. 58, no. 1, pp. 212–220, 2014.

[69] E. Zankari, H. Hasman, S. Cosentino et al., "Identification of acquired antimicrobial resistance genes," *Journal of Antimicrobial Chemotherapy*, vol. 67, no. 11, pp. 2640–2644, 2012.

[70] A. Wagner, "Rapid detection of positive selection in genes and genomes through variation clusters," *Genetics*, vol. 176, no. 4, pp. 2451–2463, 2007.

[71] E. J. Vallender and B. T. Lahn, "Positive selection on the human genome," *Human Molecular Genetics*, vol. 13, no. 2, pp. R245–R254, 2004.

[72] S. Moyo, E. Wilkinson, A. Vandormael et al., "Pairwise diversity and tMRCA as potential markers for HIV infection recency," *Medicine*, vol. 96, no. 6, p. e6041, 2017.

[73] J. Trowsdale and P. Parham, "Mini-review: defense strategies and immunity-related genes," *European Journal of Immunology*, vol. 34, no. 1, pp. 7–17, 2004.

[74] J. S. Haldane, "The rate of mutation of human genes," *Hereditas*, vol. 35, no. 1, pp. 267–273, 1949.

[75] A. Allison, "The sickle-cell and haemoglobin c genes in some African populations," *Annals of Human Genetics*, vol. 21, no. 1, pp. 67–89, 1956.

[76] E. T. Wood, D. A. Stover, M. Slatkin, M. W. Nachman, and M. F. Hammer, "The β-globin recombinational hotspot reduces the effects of strong selection around HbC, a recently arisen mutation providing resistance to malaria," *American Journal of Human Genetics*, vol. 77, no. 4, pp. 637–642, 2005.

[77] P. W. Hedrick, "Resistance to malaria in humans: the impact of strong, recent selection," *Malaria Journal*, vol. 11, no. 1, p. 349, 2012.

[78] V. R. De Mendonça, M. S. Goncalves, and M. Barral-Netto, "The host genetic diversity in malaria infection," *Journal of Tropical Medicine*, vol. 2012, Article ID 940616, 17 pages, 2012.

[79] J. B. S. Haldane, "Suggestions as to quantitative measurement of rates of evolution," *Evolution*, vol. 3, no. 1, pp. 51–56, 1949.

[80] J. Novembre, A. P. Galvani, and M. Slatkin, "The geographic spread of the CCR5 Δ32 HIV-resistance allele," *PLoS biology*, vol. 3, no. 11, p. e339, 2005.

[81] J. C. Stephens, D. E. Reich, D. B. Goldstein et al., "Dating the origin of the CCR5-Δ32 AIDS-resistance allele by the coalescence of haplotypes," *American Journal of Human Genetics*, vol. 62, no. 6, pp. 1507–1515, 1998.

[82] F. Libert, P. Cochaux, G. Beckman et al., "The Δ CCR5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe," *Human Molecular Genetics*, vol. 7, no. 3, pp. 399–406, 1998.

[83] G. Lucotte and G. Mercier, "Distribution of the CCR5 gene 32-bp deletion in Europe," *JAIDS Journal of Acquired Immune Deficiency Syndromes*, vol. 19, no. 2, pp. 174–177, 1998.

[84] J. J. Martinson, N. H. Chapman, D. C. Rees, Y.-T. Liu, and J. B. Clegg, "Global distribution of the CCR5 gene 32-basepair deletion," *Nature Genetics*, vol. 16, no. 1, pp. 100–103, 1997.

[85] G. Lucotte, "Distribution of the CCR5 gene 32-basepair deletion in West Europe. A hypothesis about the possible dispersion of the mutation by the Vikings in historical times," *Human Immunology*, vol. 62, no. 9, pp. 933–936, 2001.

[86] G. Lucotte, "Frequencies of 32 base pair deletion of the (Δ32) allele of the CCR5 HIV-1 co-receptor gene in Caucasians: a comparative analysis," *Infection, Genetics and Evolution*, vol. 1, no. 3, pp. 201–205, 2002.

[87] G. Lucotte and F. Dieterlen, "More about the Viking hypothesis of origin of the Δ32 mutation in the CCR5 gene conferring resistance to HIV-1 infection," *Infection, Genetics and Evolution*, vol. 3, no. 4, pp. 293–295, 2003.

[88] H. Rong, Q. Zhang, and Z. Zhang, "Host genetic effect on tuberculosis susceptibility in Chinese Uyghur," *Frontiers in Laboratory Medicine*, vol. 1, no. 1, pp. 5–10, 2017.

[89] F. Wu, W. Zhang, L. Zhang et al., "NRAMP1, VDR, HLA-DRB1, and HLA-DQB1 gene polymorphisms in susceptibility to tuberculosis among the Chinese Kazakh population: a case-control study," *BioMed Research International*, vol. 2013, Article ID 484535, 8 pages, 2013.

[90] A. K. Azad, W. Sadee, and L. S. Schlesinger, "Innate immune gene polymorphisms in tuberculosis," *Infection and Immunity*, vol. 80, no. 10, pp. 3343–3359, 2012.

[91] S. A. Khalilullah, H. Harapan, N. A. Hasan, W. Winardi, I. Ichsan, and M. Mulyadi, "Host genome polymorphisms and tuberculosis infection: what we have to say?," *Egyptian Journal of Chest Diseases and Tuberculosis*, vol. 63, no. 1, pp. 173–185, 2014.

[92] H. Li, T. T. Zhang, Y. Q. Zhou, Q. H. Huang, and J. Huang, "SLC11A1 (formerly NRAMP1) gene polymorphisms and tuberculosis susceptibility: a meta-analysis," *International Journal of Tuberculosis and Lung Disease*, vol. 10, no. 1, pp. 3–12, 2006.

[93] L. Abel, J. El-Baghdadi, A. A. Bousfiha, J.-L. Casanova, and E. Schurr, "Human genetics of tuberculosis: a long and winding road," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1645, p. 20130428, 2014.

[94] A. L. Gloyn and M. I. McCarthy, "Variation across the allele frequency spectrum," *Nature Genetics*, vol. 42, no. 8, pp. 648–650, 2010.

[95] J. McClellan and M.-C. King, "Genetic heterogeneity in human disease," *Cell*, vol. 141, no. 2, pp. 210–217, 2010.

[96] E. T. Cirulli and D. B. Goldstein, "Uncovering the roles of rare variants in common disease through whole-genome sequencing," *Nature Reviews Genetics*, vol. 11, no. 6, pp. 415–425, 2010.

[97] P. Garred, J. Eugen-Olsen, A. K. N. Iversen, T. L. Bennfield, A. Svejgaard, and B. Hofmann, "Dual effect of CCR5 Δ32 gene deletion in HIV-1-infected patients," *The Lancet*, vol. 349, no. 9069, p. 1884, 1997.

[98] Y. Omae, L. Toyo-oka, H. Yanai et al., "Pathogen lineage-based genome-wide association study identified CD53 as susceptible locus in tuberculosis," *Journal of Human Genetics*, vol. 62, no. 12, pp. 1015–1022, 2017.

[99] H.-Q. Qu, Q. Li, J. B. McCormick, and S. P. Fisher-Hoch, "What did we learn from the genome-wide association study for tuberculosis susceptibility?," *Journal of Medical Genetics*, vol. 48, no. 4, pp. 217-218, 2011.