



American Journal of
**Biochemistry and
Molecular Biology**

ISSN 2150-4210



Academic
Journals Inc.

www.academicjournals.com



Review Article

GWAS as Instrument for Elucidating the Role of the CCR5 Gene in HIV-1 Infection

¹Chidinma G. Daokoru-Olukole, ¹Nelson Pureaziba and ²Michael Ockiya

¹Department of Microbiology, Faculty of Sciences, Niger Delta University, PMB 071 Bayelsa NG, 560103, Amassoma, Nigeria

²Department of Animal Science, Faculty of Agriculture, Niger Delta University, PMB 071 Bayelsa NG, 560103, Amassoma, Nigeria

Abstract

Genome-wide association studies have become an indispensable tool in the determination of candidate genes that are found to be insignificant associations with certain phenotypes. This review is aimed at examining GWAS as a critical tool in elucidating the role of the CCR5 gene in HIV-1 infections. The presence of a CCR5 coreceptor is a genetic factor that modifies the outcome of HIV-1 infections. The CCR5 receptors are predominantly found on the surface of leucocytes, cells such as macrophages, CD4 and monocytes where they serve as entry receptors for HIV-1. The absence of abnormalities in CCR5 receptors increases the host's resistance to HIV-1 infection and its progression. In cases where mutation of the CCR5 gene occurs, resulting in a 32 base pair deletion and the absence of CCR5 on immune cell surfaces makes humans that are homozygous for 32 to be resistant to HIV-1 infections. Likewise, people heterozygous for 32 record delays in the progression of HIV-1 infections, a less rapid decline in the CD4 cell count and even lower viral load in circulation. The discovery of the role of CCR5 receptors and the impact of its mutation through GWAS have had a substantial impact on medical care, particularly the improvement of HIV-1 therapies. It resulted in the development of antiretroviral therapeutic agents that blocks the expression of the CCR5 receptors. These CCR5 blockers have immense potential in HIV-1 therapy as they have a high potential for prevention and treatment. This review thus provides evidence for the need to embrace GWAS and its applications in therapeutic strategies against viral infectious diseases.

Key words: GWAS, CCR5 receptors, CD4 cells, HIV-1 infections, host resistance, low-income countries, setbacks

Citation: Daokoru-Olukole, C.G., N. Pureaziba and M. Ockiya, 2023. GWAS as instrument for elucidating the role of the CCR5 gene in HIV-1 infection. *Am. J. Biochem. Mol. Biol.*, 13: 20-27.

Corresponding Author: Chidinma G. Daokoru-Olukole, Department of Microbiology, Faculty of Sciences, Niger Delta University, PMB 071 Bayelsa NG, 560103, Amassoma, Nigeria

Copyright: © 2023 G. Chidinma Daokoru-Olukole *et al.* This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Genome-Wide Association Studies (GWAS) have emerged as a critical tool applied in the detection of genes that are found in association with certain phenotypes. Thus, it has made it possible for reliable inferences to be made into the genetic basis for the differences in the phenotypic characteristics of organisms¹ to conduct genome-wide association studies, a large set of genomic data is required to reliably identify the presence of Single Nucleotide Polymorphism (SNP) distributed across various populations of an organism. This implied that GWAS in its elementary form is a mathematical assessment of the association between single nucleotide polymorphisms in the genome and a defined phenotype of interest¹. Holmes *et al.*² posted that the association between genes and phenotypes becomes significant where the differences in the genotype at a particular locus relate to the variations in the phenotype to a larger extent than might be likely by chance. The availability of human genomic data was made possible with the completion of the Human Genome Project which provided an increased understanding of the human genome and genetic variations³. However, more than just the knowledge of the genome is needed to fully conduct genome-wide analysis studies. A detailed description of human haplotypes and discovery of linkage equilibrium maps are also required in the identification of small numbers of SNPs capable of representing the most variations in the human genome^{1,4}. The increase in the amount of information about the human genome and all other genetic information needed to conduct GWAS studies has made it possible for several researchers to undertake these studies. This is evident by the increase in the number of GWAS publications now available⁵⁻⁸. Currently, there are over 3,000 human genome-wide association studies resulting in the detection of over 60,000 SNPs associated with a vast array of phenotypes⁷.

Although, GWAS began as a tool for investigating the correlation between the human genome and phenotypic characters, it can now be applied in various fields of study such as virology. Specifically, genome-wide association studies have been applied to study the genetic determinants of viral diseases in humans. One of the most extensively studied viral diseases is the Human Immunodeficiency Syndrome-1 (HIV-1). The first genome-wide association studies of infectious diseases were done on HIV-1. This made it possible for us to go beyond the existing knowledge to new exciting facts about the interactions between HIV and host genetics⁹. It has been established that HIV-1 interacts with various host factors during its infection and replication. Amongst the host factors studied, the Human CC-Type Chemokine

Receptor 5 (CCR5-Δ32) gene is the only confirmed genetic factor that modifies the outcomes of HIV-1 diseases. The CCR5 is a part of the G-protein Coupled Receptor (GPCR) family that is predominantly found on the surface of leucocyte cells such as the macrophages, T- cells and monocytes¹⁰. Earlier research on HIV-1 treatment focused on the development of inhibitors that blocks the binding of HIV-1 to CD4 cells because it was thought that the presence of CD4 alone was sufficient for the progression of HIV-1 infection. However, it is now understood that the binding of HIV-1 to CD4 receptors though necessary for HIV infections, is not sufficient to enable HIV to invade the host cells. A second step and coreceptor are needed, the CCR5 and CXR4 receptors³.

The absence or abnormalities in the CCR5 receptor has been shown in several studies to increase the host's resistance to the infection and progression of HIV-1⁸. The HIV-1 commonly uses the CCR5 receptors to invade immunological cells. The CCR5 receptors are found on the surface of immune cells, thus providing a means of entry for HIV-1 viral particles to infect the cell. In cases where mutation of the CCR5 gene occurs resulting in a 32-base pair deletion and the absence of CCR5 on cell surfaces makes humans that are homozygous for 32 to be resistant to HIV-1 infections. Also, people heterozygous for 32 have recorded delays in the progression of HIV infections, a less rapid decline in the CD4 cell count and even a lower viral load in circulation^{11,12}.

Considering the role of the CCR5, treatment of HIV-1 infections using a genetic approach has been proposed to block the CCR5 from being expressed. In one study, T-cells modified not to express CCR5 were mixed with wild T-cells that express the CCR5 to perform a challenge test against HIV-1. Modified T-cells became predominant over the wild-type because the HIV-1 viral particles destroyed the wild-type CCR5 cells while the modified CCR5 cells were resistant to the HIV-1 cells⁹.

The discovery of the role of CCR5 receptors and the impact of its modification through genome-wide analysis studies could have a substantial impact on medical care. For example, it was only possible to produce new antiviral drugs because of the increase in our understanding of the role of CCR5del32 homozygotes as the molecular basis of HIV-1 resistance¹³⁻¹⁷.

This review paper will highlight the relationship between the CCR5 receptor and HIV-1 infection and its significance in identifying significant phenotypic associations.

Units of genetic variation used in GWAS: At the early ages of GWAS, the use of a single nucleotide polymorphism was the unit for investigating genetic variations¹⁸. However, with the

advent of more genomic technologies, different techniques and measures of genetic variations about phenotypes have become popular. These include linkage disequilibrium, direct and indirect associations and population stratification^{19,20}.

CCR5 receptors and HIV-1 resistance: The CCR5 receptor is a member of the chemokine receptor family and is usually expressed on different cell types, including macrophages, activated T lymphocytes and dendritic cells. The chemokine receptor family is made up of an extracellular N- terminus, an intracellular C-terminus, 7 transmembrane helices and 3 extracellular loops. The elements that are located in the second extracellular loops and the N-terminus are specifically critical for interacting with HIV-1 during the phase of entry. This has directed attention to the application of CCR5 for designing antiretroviral therapies more productive¹⁰.

Additionally, there is an extra advantage of CCR5 as a cellular target because it is not necessary for the normal functioning of the immune system unlike the CD4 receptors and the CXCR4 viral co-receptor (the CD4 and CXCR4 have essential roles in immune function, thus, limiting their use as targets of antiretroviral agents)¹⁰.

Homozygous individuals for the $\Delta 32$ mutation of CCR5 demonstrate the comparative dispensability of the CCR5 co-receptor and have been shown to have greater resistance to HIV-1 infections¹⁹. Also, individuals who are heterozygous for the $\Delta 32$ mutation of the CCR5 receptor progress relatively slower to AIDS than those who are homozygous for the wild-type gene^{21,22}.

Furthermore, there exists a correlation, between the density of CCR5 on CD4 T-cells, RNA viral loads and the development of AIDS in infected patients²¹. This relationship has been established *in vitro*. In a study undertaken by Heredia *et al.*²³ it was reported that the antiretroviral activity of CCR5 antagonists was directly impacted by CCR5 surface antigens. The experiment established the correlation between the CCR5 levels with the inhibition rates of HIV-1 entry^{23,24}.

These discoveries, coupled with the ostensible therapeutic effect observed when patients with leukaemia and AIDS received a transplant of $\Delta 32$ homozygous hematopoietic stem cells have given impetus for the application of blockers of CCR5 receptors for inhibiting entry and infection of HIV-1²⁴. Currently, these blockers include antagonists of CCR5, drugs that reduce the CCR5 surface density, CCR5 antibodies, fusion proteins targeting the CCR5 N-terminus and other essential sites of the CCR5 receptor²⁴⁻²⁸.

Remarkable suppression of HIV-1 entry has been recorded by some CCR5 blockers in clinical trials^{28,29}. Within the extracellular environment, entry inhibitors have a further

appeal because of their ability to cause immobilization of HIV-1 where it is reachable to the immune system³⁰.

Determining the relationship between the CCR5 gene and the outcome of HIV-1 infections: To ascertain the role of the CCR5 gene on the outcome of HIV-1 infections, two groups of human participants must be presented, a control group (participants without HIV-1 infection and a treatment group (HIV-infected participants). A good population size should be chosen, usually above 1000 participants for each group²⁵. A suitable genotyping platform is thereafter chosen to obtain the genotypic data of the CCR5 gene. The genotypic data of the CCR5 genes can be produced using microarray technology which allows the determination of the polymorphisms present in the sampled population²⁶.

Some of the available chips can be used to analyze as many single nucleotide polymorphisms as possible. The data generated from genotyping is statistically analyzed to ascertain the relationship between the presence of the SNPs in the experimental population and the outcome of HIV-1 infections in the sampled population. The statistical analysis begins with a holistic quality control analysis aimed at determining the accuracy of the genotypic data obtained²⁶. The statistical analysis involves testing the statistical hypothesis for each single nucleotide polymorphism obtained. The results of the statistical analysis were thereafter analyzed and inferences are made based on the analysis. Primarily, the differences in the SNPs between the healthy patients and infected patients are examined for a significant relationship between the deviations in the genes (SNPs) and the phenotypes in the population²⁷. In this case, the phenotypes are the measurable traits such as the rate of HIV progression, level of CD4 cells in the participants etc. If these traits are proven to be in significant association with the alterations in the CCR5 genes, it is considered to be a determinant of the outcome of HIV-1 infection. The GWAS is usually concluded by providing a biological or clinical validation to ensure the reliability of the findings²⁸.

Studies on GWAS and susceptibility to HIV infections:

Globally, there is a huge disease burden and deaths resulting from infectious agents. The World Health Organization²⁹ reported a statistic that the deaths recorded from infectious diseases were about 12 million people. This has created the need for a more holistic approach to understanding disease processes and the determinants of susceptibility of humans to infectious agents. There has been remarkable progress in the applications of GWAS in studying genes and phenotypes associated with diseases. This advancement results from the

development in the implementation of genomics, bioinformatics and statistical tools. One of the most extensively studied infectious agents and its pattern of infection and resistance by host cells is the Human Immunodeficiency Virus (HIV)³⁰. Samson *et al.*³¹ reported that the 32 base pair deletion that occurs in the Chemokine Receptor 5 Gene (CCR5) represses the expression of the receptor, making the host cells resistant to HIV infections.

Apart from the mutation of the CCR5del3 gene that confers cells' resistance to cells being infected by HIV, there have been several candidate genes have been identified. This includes, Killer-Immunoglobulin-Like-Receptors molecules (KIRs), Human Leukocyte Antigen (HLA), Cytokines, etc³².

Stewart *et al.*³³ reported that the progression of HIV infections may be affected by genetic factors. For instance, humans that are heterozygotes for CCR5del32 have been reported to exhibit susceptibility to HIV infections but the progression of the infection occurs very slowly when compared to other phenotypes.

In a similar study, Mallal *et al.*³⁴ suggested that genetic factors may likewise play a role in the response to HIV treatment as seen in the role of the Human Leukocyte Antigen (HLA) antigen B*5701 allele in cases of severe hypersensitivity reactions to abacavir (an anti-retroviral drug).

A similar study was undertaken by Fellay *et al.*³⁰ aimed at investigating the association between single nucleotide polymorphism in the Illumina 550k array and the viral load and the rate of CD4 T lymphocyte count decline in 486 research subjects. It was observed that 15% of the differences in the viral load resulted from two single nucleotide polymorphisms inside the I HLA regions. Interestingly, it was reported that near the HLA-C gene was one of the HLA single nucleotide polymorphisms. The other Human Leukocyte Antigen (HLA) SNP was found in the HCP5.

The study discovered that the genes that regulate the viral load were not overlapping with the genes identified to regulate the disease progression although there was a correlation between the parameters. The findings of this study were replicated severally to establish the effects of HCP5 and HLA-C on either HLA-A10 or HLA-B*57, thus pointing to the linkage disequilibrium within the Human Leukocyte Antigen (HLA) region³⁵.

Another study by Fellay *et al.*³⁰ was expanded to involve over 2500 patients infected with HIV. Their study showed that the genetic variants that are present close to the HLA-B and HLA-C genes are regarded to be the most significant genetic variants associated with the control of the HIV-1 virus. However, their investigations did not find previously identified loci that were detected through gene studies.

There have been other areas of the HIV infectious process that have been studied by GWAS such as the role of the X chromosome, HIV-associated Human Leukocyte Antigen (HLA) loci in CD4, the ratios of CD8 T lymphocytes and genetic factors that influence viral transmission from mother to child. Ferreira *et al.*³⁶ suggested that the variants of the Human Leukocyte Antigen (HLA) that are associated with viral load regulation in HIV-1 infections are linked independently with CD4 and CD8 ratio (CD4 and CD8 are dysregulated in HIV-1 infections). In a related study by Joubert *et al.*³⁷, 100 HIV-infected persons were compared with 126 infants that were uninfected using GWAS. In their study, about nine single nucleotide polymorphisms found within sex-different genes were found to be associated with a decreased transmission rate.

Ethical issues in GWAS in Africa: The advent of technological advancement has caused a spike in the interest in human genomic differences as a tool for analyzing common complex human diseases. It has been established that the genetic diversity of humans can be applied in the study of the mechanisms of intricate diseases. The GWAS has been applied expansively in these studies. In fact, for close to a decade, it has been a reliable tool in the identification of loci of the genome that plays a role in the determination of susceptibility or resistance to a wide array of common diseases³⁸⁻⁴⁰. Genomic studies focusing on diseases such as GWAS have largely been used to study a wide array of disease conditions but the application of this tool in studying diseases that affect people in low-income nations has been scarce^{40,41}. This makes it imminent for medical genomic studies to be conducted to study diseases that are common in poor countries. These countries are usually characterized by high illiteracy levels and poor healthcare systems. The application of GWAS in these regions will help mitigate the impact of such imbalance^{42,43}.

Despite the relevance of medical genomic research in low-income countries, there are a lot of setbacks encountered in conducting such studies. These setbacks could range from obtaining consent from research participants/volunteers to the privacy of the volunteers and collection and storage of the genomic data obtained⁴⁴⁻⁵².

Protecting the interests of research participants: One of the major challenges faced in the conduct of GWAS in low-income countries is the limited accessibility of testing sites, health care and relevant resources that would make it easier for the participants to get involved. Because of the nature of GWAS, sample collection cannot be done in locations without the required resources. Thus, participants may have to travel far to

get involved. This makes it quite difficult for GWAS to be conducted. Although these challenges are faced even in developed countries, the issues are more critical in less developed countries¹⁴.

Valid consent: According to the World Medical Association⁵³, before medical research will be undertaken, it is a must to obtain valid consent. The valid consent should generally be adequately informed and understood, it should be voluntary. The processes through which valid consent is obtained should be appropriate to the locally established laws⁵⁴. However, the processes undertaken to obtain consent may present challenges. One of the major challenges faced is related to providing adequate information to the participants in a comprehensible manner. Due to the illiteracy level in low-income countries, providing comprehensive information on the research to the participants is usually a daunting task⁵⁵. It may however be possible to describe certain aspects of the research to the participants but providing enough comprehensive information to ensure the participants give their valid consent for GWAS remains a significant challenge^{51,56}.

Regulating human genomics research: It is common to see collaborations between researchers from low-income and high-income countries especially when the research is centred on diseases prevalent in low-income countries. During such collaborations, it is common for advanced molecular analysis such as genotyping and whole-genome analysis to be done in developed countries. This creates issues about the use of the stored samples and data and the possession and ethical review of the research^{51,56}.

Sample export and ownership: To perform GWAS, access to sophisticated facilities and personnel with relevant expertise and experience is required. These are often available in a few nations around the world. As a result, the collected samples for genome-wide analysis are usually exported for analysis and quality control. This transport of samples to other countries for analysis results in serious ethical issues^{51,56}.

Most concerns arise because of the uncertainty over the control and use of the samples in the country to which they are exported to. This concern becomes aggravated knowing the participants will have limited or no control over their samples^{57,58}.

Archived samples: Genome-wide analysis studies require the assemblage of a very large sum of samples and well-characterized phenotypes. The large quantity of samples

implies that the GWAS will be labour-intensive, requiring high technology for genomic studies. Such facilities are not available in low-income countries. Consequently, it is often not feasible to conduct such genomic research and it may not also be possible to collect genomic samples for every research. As a result, genomic data are often archived. Issues on the decisions for the re-use of human genomic samples arise and this further complicates obtaining consent for GWAS. Therefore, there is a need to strike an equilibrium between the ethical consequences of assembling thousands of samples against the ethics of analyzing archived samples with no valid consent from the original participants⁵⁸.

Limitations of GWAS: The application of genome-wide association studies may fail under certain conditions to detect new loci that are susceptible to certain diseases even though a large sample size is used for the study⁵⁹. Wray *et al.*⁶⁰ reported that genome-wide analysis studies may perform poorly in determining associations of genotype and phenotype in cases of rare diseases because of the difficulty in obtaining the sample size required.

Another potential pitfall of GWAS is the issue of validation and discovery sample overlap⁵⁹. Sometimes, there may be an overestimation of the validation samples (an independent sample with determined/known phenotypic characters), especially in situations where they are more strongly associated with the discovery sample when compared to the target samples.

Wray *et al.*⁶⁰ proposed that the similarities of population stratification can inflate the accuracy of the association in cases where there is a very close match between both the validation and discovery samples with population stratification but does not suggest a strong match to the stratification of the targeted sample.

Moore *et al.*⁵⁹ pointed to another GWAS limitation observed in the use of standard logical regression. Their study thus suggested that more advanced algorithms should be used to illuminate the potential relationship with variations in DNA sequences, variations in susceptibility to disease and environmental exposure. The findings of GWAS have aided the improvement of antiretroviral therapy for HIV-1 treatment.

Scarcity of world-class genomic laboratories mitigates the application and advancement of GWAS in Africa. Therefore, more genomic research centres should be established to promote the use of GWAS in Africa.

The GWAS should be deployed to study more viral diseases to find ways of developing effective antiviral therapies.

CONCLUSION

Genome-wide association studies have been developed as an approach that involves the scanning of markers across selected sets of DNAs of a large population to find possible genetic variations that are associated with a particular disease. The information about the identified relationship between genes and diseases is used by researchers to build better strategies for disease detection, treatment and even prevention.

The GWAS has been applied in medical virology research to find the genes associated with HIV-1 disease and its progression. Various genome-wide analysis studies have established that the abnormalities of the CCR5 receptor such as the CCR5-Δ32 mutation and the absence of the receptor confer resistance of the host to HIV-1 infection and even slow down the progression of the infection. Furthermore, there exists a correlation, between the density of CCR5 on CD4 T-cells, RNA viral loads and the development of AIDS in infected patients. This has resulted in the development of antiretroviral therapeutic agents that blocks the expression of the CCR5 receptors. These CCR5 blockers have immense potential in HIV-1 therapy as they have a high potential for prevention and treatment. This seminar text thus emphasizes the benefits of embracing GWAS and its applications in therapeutic strategies against viral infectious diseases.

SIGNIFICANCE STATEMENT

Although, there has been a significant increase in the quality and the number of publications on the applications of Genome-Wide Association Studies (GWAS), there are very few journals or none that link the application of GWAS and the setbacks it faces in Africa despite being hotspot for HIV. Also, this review paper presented both empirical and theoretical research as evidence for the role of the CCR5 co-receptor in HIV-1 infections. Generally, this review deepens the scientific appreciation for GWAS as a critical tool not just in studying HIV-1 related infection and other viral diseases. Providing emphasis on the setbacks of GWAS in Africa provides key knowledge to the public, especially those in Africa about some of the critical aspects of conducting molecular genetic research in human populations. This is believed to increase the level of cooperation in future GWAS experiments.

REFERENCES

1. Klein, R.J., C. Zeiss, E.Y. Chew, J.Y. Tsai and R.S. Sackler *et al*, 2005. Complement factor H polymorphism in age-related macular degeneration. *Science*, 308: 385-389.
2. Holmes, M.V., M. Ala-Korpela and G.D. Smith, 2017. Mendelian randomization in cardiometabolic disease: Challenges in evaluating causality. *Nat. Rev. Cardiol.*, 14: 577-590.
3. International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. *Nature*, 431: 931-945.
4. Purcell, S., S.S. Cherny and P.C. Sham, 2003. Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, 19: 149-150.
5. Zhou, X., F.K. Tan, N. Wang, M. Xiong and S. Maghidman *et al*, 2003. Genome-wide association study for regions of systemic sclerosis susceptibility in a Choctaw Indian population with high disease prevalence. *Arthritis Rheumatism*, 48: 2585-2592.
6. Marchini, J., L.R. Cardon, M.S. Phillips and P. Donnelly, 2004. The effects of human population structure on large genetic association studies. *Nat. Genet.*, 36: 512-517.
7. Weiss, L.A., J.V.V. Weele, D.L. Newman, S.J. Kim and H. Dytch *et al*, 2004. Genome-wide association study identifies *ITGB3* as a QTL for whole blood serotonin. *Eur. J. Hum. Genet.*, 12: 949-954.
8. Haines, J.L., M.A. Hauser, S. Schmidt, W.K. Scott and L.M. Olson *et al*, 2005. Complement factor H variant increases the risk of age-related macular degeneration. *Science*, 308: 419-421.
9. Sheppard, S.K., L. Cheng, G. Meric, C.P.A. de Haan and A.K. Llerena *et al*, 2014. Cryptic ecology among host generalist *Campylobacter jejuni* in domestic animals. *Mol. Ecol.*, 23: 2442-2451.
10. Laabei, M., M. Recker, J.K. Rudkin, M. Aldeljawi and Z. Gulay *et al*, 2014. Predicting the virulence of MRSA from its genome sequence. *Genome Res.*, 24: 839-849.
11. Salipante, S.J., D.J. Roach, J.O. Kitzman, M.W. Snyder and B. Stackhouse *et al*, 2015. Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. *Genome Res.*, 25: 119-128.
12. Fatkenheuer, G., A.L. Pozniak, M.A. Johnson, A. Plettenberg and S. Staszewski *et al*, 2005. Efficacy of short-term monotherapy with maraviroc, a new CCR5 antagonist, in patients infected with HIV-1. *Nat. Med.*, 11: 1170-1172.
13. Skol, A.D., L.J. Scott, G.R. Abecasis and M. Boehnke, 2006. Corrigendum: Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.*, 38: 209-213.
14. Young, A.I., M.L. Frigge, D.F. Gudbjartsson, G. Thorleifsson and G. Bjornsdottir *et al*, 2018. Relatedness disequilibrium regression estimates heritability without environmental bias. *Nat. Genet.*, 50: 1304-1310.
15. Kerminen, S., A.R. Martin, J. Koskela, S.E. Ruotsalainen and A.S. Havulinna *et al*, 2019. Geographic variation and bias in the polygenic scores of complex diseases and traits in finland. *Am. J. Hum. Genet.*, 104: 1169-1181.

16. Zaidi, A.A. and I. Mathieson, 2020. Demographic history mediates the effect of stratification on polygenic scores. *ELife*, Vol. 9. 10.7554/eLife.61548.
17. Loh, P.R., G. Kichaev, S. Gazal, A.P. Schoech and A.L. Price, 2018. Mixed-model association for biobank-scale datasets. *Nat. Genet.*, 50: 906-908.
18. Bush, W.S. and J.H. Moore, 2012. Chapter 11: Genome-wide association studies. *PLoS Comput. Biol.*, Vol. 8. 10.1371/journal.pcbi.1002822.
19. Slatkin, M., 2008. Linkage disequilibrium-understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.*, 9: 477-485.
20. Sved, J.A. and W.G. Hill, 2018. One hundred years of linkage disequilibrium. *Genetics*, 209: 629-636.
21. Reynes, J., P. Portales, M. Segondy, V. Baillat and P. Andre *et al.*, 2000. CD4⁺ T cell surface CCR5 density as a determining factor of virus load in persons infected with human immunodeficiency virus type 1. *J. Infect. Dis.*, 181: 927-932.
22. Westby, M. and E. van der Ryst, 2010. CCR5 antagonists: Host-targeted antiviral agents for the treatment of HIV infection, 4 years on. *Antiviral Chem. Chemother.*, 20: 179-192.
23. Heredia, A., O. Latinovic, R.C. Gallo, G. Melikyan and M. Reitz *et al.*, 2008. Reduction of CCR5 with low-dose rapamycin enhances the antiviral activity of vicriviroc against both sensitive and drug-resistant HIV-1. *Proc. Natl. Acad. Sci.*, 105: 20476-20481.
24. Anastassopoulou, C.G., A.J. Marozsan, A. Matet, A.D. Snyder and E.J. Arts *et al.*, 2007. Escape of HIV-1 from a small molecule CCR5 inhibitor is not associated with a fitness loss. *PLoS Pathog.*, Vol. 3. 10.1371/journal.ppat.0030079.
25. Maeda, K., D. Das, P.D. Yin, K. Tsuchiya and H. Ogata-Aoki *et al.*, 2008. Involvement of the second extracellular loop and transmembrane residues of CCR5 in inhibitor binding and HIV-1 fusion: Insights into the mechanism of allosteric inhibition. *J. Mol. Biol.*, 381: 956-974.
26. Zheng, Y., G.W. Han, R. Abagyan, B. Wu and R.C. Stevens *et al.*, 2017. Structure of CC chemokine receptor 5 with a potent chemokine antagonist reveals mechanisms of chemokine recognition and molecular mimicry by HIV. *Immunity*, 46: 1005-1017.
27. Huang, C.C., S.N. Lam, P. Acharya, M. Tang and S.H. Xiang *et al.*, 2007. Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science*, 317: 1930-1934.
28. Ji, C., J. Zhang, M. Dioszegi, S. Chiu and E. Rao *et al.*, 2007. CCR5 small-molecule antagonists and monoclonal antibodies exert potent synergistic antiviral effects by cobinding to the receptor. *Mol. Pharmacol.*, 72: 18-28.
29. World Health Organization 2004. The global burden of disease: 2004 update. <https://www.who.int/publications/i/item/9789241563710>
30. Fellay, J., K.V. Shianna, D. Ge, S. Colombo and B. Ledergerber *et al.*, 2007. A whole-genome association study of major determinants for host control of HIV-1. *Science*, 317: 944-947.
31. Samson, M., F. Libert, B.J. Doranz, J. Rucker and C. Liesnard *et al.*, 1996. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature*, 382: 722-725.
32. An, P. and C.A. Winkler, 2010. Host genes associated with HIV/AIDS: Advances in gene discovery. *Trends Genet.*, 26: 119-131.
33. Stewart, G.J., L.J. Ashton, R.A. Biti, R.A. Ffrench and B.H. Bennetts *et al.*, 1997. Increased frequency of CCR-5 Δ 32 heterozygotes among long-term non-progressors with HIV-1 infection. *AIDS*, 11: 1833-1838.
34. Mallal, S., D. Nolan, C. Witt, G. Masel and A.M. Martin *et al.*, 2002. Association between presence of *HLA-B*5701*, *HLA-DR7*, and *HLA-DQ3* and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet*, 359: 727-732.
35. Catano, G., H. Kulkarni, W. He, V.C. Marconi and B.K. Agan *et al.*, 2008. HIV-1 disease-influencing effects associated with *ZNRD1*, *HCP5* and *HLA-C* alleles are attributable mainly to either *HLA-A10* or *HLA-B*57* alleles. *PLoS ONE*, Vol. 3. 10.1371/journal.pone.0003636.
36. Ferreira, M.A.R., M. Mangino, C.J. Brumme, Z.Z. Zhao and S.E. Medland *et al.*, 2010. Quantitative trait loci for CD4:CD8 lymphocyte ratio are associated with risk of type 1 diabetes and HIV-1 immune control. *Am. J. Hum. Genet.*, 86: 88-92.
37. Joubert, B.R., E.M. Lange, N. Franceschini, V. Mwapasa and K.E. North *et al.*, 2010. A whole genome association study of mother-to-child transmission of HIV in Malawi. *Genome Med.*, Vol. 2. 10.1186/gm138.
38. Lalezari, J., G.K. Yadavalli, M. Para, G. Richmond and E. DeJesus *et al.*, 2008. Safety, pharmacokinetics, and antiviral activity of HGS004, a novel fully human igG4 monoclonal antibody against CCR5, in HIV-1-infected patients. *J. Infect. Dis.*, 197: 721-727.
39. Moore, J.P. and R.W. Doms, 2003. The entry of entry inhibitors: A fusion of science and medicine. *Proc. Nat. Acad. Sci.*, 100: 10598-10602.
40. Rosenberg, N.A., L. Huang, E.M. Jewett, Z.A. Szpiech and I. Jankovic *et al.*, 2010. Genome-wide association studies in diverse populations. *Nat. Rev. Genet.*, 11: 356-366.
41. Need, A.C. and D.B. Goldstein, 2009. Next generation disparities in human genomics: Concerns and remedies. *Trends Genet.*, 25: 489-494.

42. Berlinguer, G., 2004. Bioethics, health, and inequality. *Lancet*, 364: 1086-1091.
43. Hesketh, T., 2005. The 10/90 report on health research 2003-2004 Global forum for health research. Helping correct the 10/90 gap, Geneva, 2004. xxvi + 282 pp., paperback, ISBN: 2-940286-16-7. *Trans. R. Soc. Trop. Med. Hyg.*, Vol. 99. 10.1016/J.TRSTMH.2005.04.003.
44. Kaye, J., P. Boddington, J. de Vries, N. Hawkins and K. Melham, 2010. Ethical implications of the use of whole genome methods in medical research. *Eur. J. Hum. Genet.*, 18: 398-403.
45. Caulfield, T., A.L. McGuire, M. Cho, J.A. Buchanan and M.M. Burgess *et al.*, 2008. Research ethics recommendations for whole-genome research: Consensus statement. *PLoS Biol.*, Vol. 6. 10.1371/journal.pbio.0060073.
46. McGuire, A.L., T. Caulfield and M.K. Cho, 2008. Research ethics and the challenge of whole-genome sequencing. *Nat. Rev. Genet.*, 9: 152-156.
47. Mascalzoni, D., A. Hicks, P. Pramstaller and M. Wjst, 2008. Informed consent in the genomics era. *PLoS Med.*, Vol. 5, No. 9. 10.1371/journal.pmed.0050192.
48. Lunshof, J.E., R. Chadwick, D.B. Vorhaus and G.M. Church, 2008. From genetic privacy to open consent. *Nat. Rev. Genet.*, 9: 406-411.
49. Heeney, C., N. Hawkins, J. de Vries, P. Boddington and J. Kaye, 2011. Assessing the privacy risks of data sharing in genomics. *Public Health Genomics*, 14: 17-25.
50. Malin, B. and L. Sweeney, 2004. How (not) to protect genomic data privacy in a distributed network: Using trail re-identification to evaluate and design anonymity protection systems. *J. Biomed. Inf.*, 37: 179-192.
51. Chokshi, D.A., M. Parker and D.P. Kwiatkowski, 2006. Data sharing and intellectual property in a genomic epidemiology network: Policies for large-scale research collaboration. *Bull. World Health Organ.*, 84: 382-387.
52. Parker, M., S.J. Bull, J. de Vries, T. Agbenyega, O.K. Doumbo and D.P. Kwiatkowski, 2009. Ethical data release in genome-wide association studies in developing countries. *PLoS Med.*, Vol. 6. 10.1371/journal.pmed.1000143.
53. Lange, C., D. DeMeo, E.K. Silverman, S.T. Weiss and N.M. Laird, 2004. PBAT: Tools for family-based association studies. *Am. J. Hum. Genet.*, 74: 367-369.
54. CIOMS/WHO (Council of International Organization of Medical Sciences/World Health Organization), 1993. *International Ethical Guidelines for Biomedical Research Involving Human Subjects*. Geneva.
55. Molyneux, C.S., N. Peshu and K. Marsh, 2004. Understanding of informed consent in a low-income setting: Three case studies from the Kenyan coast. *Social Sci. Med.*, 59: 2547-2559.
56. Tekola, F., S. Bull, B. Farsides, M.J. Newport and A. Adeyemo *et al.*, 2009. Impact of social stigma on the process of obtaining informed consent for genetic research on podocniosis: A qualitative study. *BMC Med. Ethics*, Vol. 10. 10.1186/1472-6939-10-13.
57. Upshur, R.E.G., J.V. Lavery and P.O. Tindana, 2007. Taking tissue seriously means taking communities seriously. *BMC Med. Ethics*, Vol. 8. 10.1186/1472-6939-8-11.
58. Muula, A.S. and J.M. Mfutso-Bengo, 2007. Responsibilities and obligations of using human research specimens transported across national boundaries. *J. Med. Ethics*, 33: 35-38.
59. Moore, J.H., F.W. Asselbergs and S.M. Williams, 2010. Bioinformatics challenges for genome-wide association studies. *Bioinformatics*, 26: 445-455.
60. Wray, N.R., J. Yang, B.J. Hayes, A.L. Price, M.E. Goddard and P.M. Visscher, 2013. Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.*, 14: 507-515.