MAKERERE  **UNIVERSITY**

GENETIC DETERMINANTS OF DIFFERENTIAL SUSCEPTIBILITY TO HEPATITIS B INFECTION AMONG A GENERAL POPULATION COHORT FROM SOUTHWESTERN UGANDA: A GENOME-WIDE ASSOCIATION STUDY

By

Irene Namara (BBLT, MAK)

2020/HD07/20618U

Masters of Bioinformatics School of Biomedical Sciences

SUPERVISORS

Prof Segun Fatumo (Ph.D.)

Dr. Daudi Jjingo (Ph.D.)

Dr. Allan Kalungi (Ph.D.)

A RESEARCH DISSERTATION SUBMITTED TO THE SCHOOL OF BIOMEDICAL SCIENCES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR AWARD OF A MASTER'S OF SCIENCE DEGREE IN BIOINFORMATICS OF MAKERERE UNIVERSITY

NOVEMBER 2023

DECLARATION

I, IRENE NAMARA, declare that I have developed this research dissertation under my supervisors' guidance. It has never been submitted to any institution for any academic award henceforth free of any malpractice what-so ever.

Signature:  Date: 9th November 2023

I have submitted my research dissertation with approval from my supervisors;

Prof Segun Fatumo (Ph.D.)

Signature:  Date: 10th November 2023

Principal Investigator

The African Computational Genomics Group, MRC/UVRI and LSHTM Uganda Research Unit & The Department of Non-Communicable Diseases Epidemiology, London School of Hygiene and Tropical Medicine

Dr Daudi Jjingo (Ph.D.)

Signature:  Date: 10th November 2023

Director

African Centers of Excellence in Bioinformatics, Makerere University, Kampala

Dr. Allan Kalungi (Ph.D.)

Signature:  Date: 10th November 2023

Research Fellow

The African Computational Genomics Group, MRC/UVRI and LSHTM Uganda Research Unit & The Department of Non-Communicable Diseases Epidemiology, London School of Hygiene and Tropical Medicine

DEDICATION

I dedicate this research project to my family that has been a great support system throughout my study.

ACKNOWLEDGEMENT

I would like to thank God for the gift of life and all the blessings He has bestowed upon me. I thank my family for paying my tuition and for their unending support. A special appreciation goes towards my supervisors; Prof Segun Fatumo and Dr. Allan Kalungi of the African Computational Genomics Research Group at MRC/UVRI and LSHTM Uganda Research Unit, Dr. Daudi Jjingo of the African Centers of Excellence in bioinformatics at Makerere University, my mentor Dr.Ronald Galiwango, and all members of the African Computational Genomics Group for the guidance and support given to me.

Special gratitude goes out to Makerere University/UVRI Centre of Excellence in infection and immunity Research and training (MUII Plus), MRC/UVRI and LSHTM Uganda Research Unit under the leadership of Prof Alison Elliot for the generous student support rendered to me which has enabled me to complete this research project.

TABLE OF CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS & ACRONYMS

Abbreviation	Full form
AASLD	American association for the study of liver disease
AIDS	acquired immunodeficiency syndrome
ALT	alanine aminotransferase/alanine transaminase
Anti-HCV	hepatitis C antibody
AST	aspartate aminotransferase/aspartate transaminase
bmi	body mass index
CDC	centers for disease control and prevention
CHB	chronic hepatitis B
COVID-19	Coronavirus disease 2019
DNA	deoxyribonucleic acid
GEMMA	genome-wide efficient mixed model analysis
GPC	general population cohort
GWAS	genome-wide association study
HBI	hepatitis B infection
HBsAg	hepatitis B surface antigen
HBV	hepatitis B virus
HCV	hepatitis C virus
HDV	hepatitis D virus
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
IVW	inverse variance weighted
LSHTM	London school of hygiene and tropical medicine
MR	Mendelian randomization
MRC	medical research council
NICE	national institute for health and care excellence
RNA	ribonucleic acid
SNP	single nucleotide polymorphism
UGR	Uganda genome resource
UNCST	Uganda national council of science and technology

UVRI	Uganda virus research institute
UVRISEC	Uganda virus research institute science and ethics committee
WHO	world health organization

ABSTRACT

Introduction: Hepatitis B is one of the major causes of mortality and morbidity worldwide. Hepatitis B infection causes liver inflammation and damage, that may progress to liver cirrhosis, liver cancer and ultimately lead to death. The outcomes of hepatitis B virus exposure are highly variable and range from resistance to infection, spontaneous viral clearance, and chronic hepatitis B infection. This variability is moderated by a number of factors such as host genetics, viral factors and environmental factors. Genetic polymorphisms such as single nucleotide polymorphisms present in individuals of particular populations can be used to determine the risk of susceptibility to infectious and complex diseases. A plethora of studies have determined the host genetic impact such as inter-individual genome variability on differential susceptibility to hepatitis B infection. However, most of these studies have been carried out in European populations. There is little data from genetic studies among African populations, which are known to be highly genetically diverse, as evidenced by high levels of phenotypic variation. This limits the generalizability of findings to African populations. This study identified the host genetic determinants influencing differential susceptibility to hepatitis B infection among a Ugandan population. It estimated the genetic variants present among Ugandans that account for differential susceptibility to hepatitis B infection.

Objective: To investigate the host genetic impact on differential susceptibility to hepatitis B infection among a general population cohort in southwestern Uganda.

Methods: Genome-wide association analysis using GEMMA software to find genomic loci that are associated with differential susceptibility to hepatitis B infection was performed, followed by fine-mapping using the Bayesian approach to determine the specific causal variants identified to be in the association. Replication of genome-wide association study findings was done to validate findings and check for reproducibility in summary statistics available on the genome-wide association study (GWAS) catalog. A two-sample Mendelian randomization (MR) was performed to determine the causal effect of hepatitis B infection on liver function biomarkers of alanine aminotransferase and aspartate transferase (ALT) and Aspartate Transferase (AST).

Results: We identified 78 significant single nucleotide polymorphisms (SNPs) in 22 genomic loci associated with susceptibility to hepatitis B infection of which all 22 were novel. The most significant

SNPs mapped to nearest genes *POGK, MIXL1, RP11-24121.1, RP1-97D16.1, ADRB1,* and *RBFOX1* which have previously been associated with the liver tissue and its enzymes. Replication of association findings was not successful in two independent cohorts of European ancestry individuals. The two sample MR revealed no causal effect of hepatitis B infection susceptibility on ALT and AST serum levels.

Conclusion: This study identified existing and novel SNPs associated with differential susceptibility to hepatitis B infection present among Ugandans in a well-characterized General Population Cohort.

CHAPTER ONE: INTRODUCTION

1.1 Background

1.1.1 Epidemiology of hepatitis B

Hepatitis is defined as liver inflammation and damage caused by either viruses, heavy alcohol use, toxins, some medications, or particular medical conditions (WHO, 2022a). Hepatitis of viral origin is termed as viral hepatitis and the commonest viral strains are hepatitis A, hepatitis B, and hepatitis C (WHO, 2022a). Viral hepatitis infection is a great health issue globally and it has been reported to be associated with remarkable morbidity and mortality (Castaneda, Gonzalez, Alomari, Tandon, & Zervos, 2021). The global prevalence of hepatitis B infection was reported as 3.9% in 2016 with approximately 292 million people infected. The prevalence of hepatitis B infection differs across regions in the world with the highest endemicity noted in developing regions such as sub-Saharan Africa (Polaris Observatory, 2018). The variation in the prevalence of hepatitis B across regions in the world is driven by the differences in the predominant modes of transmission, the impact of vaccination programs, and the age at the time of infection in each of the world populations (Thomas et al., 2012; Zoulim & Durantel, 2015). The global prevalence of hepatitis B infection was reported as 3.5% in 2019 with approximately 296 million people infected whereby the highest prevalence was observed in the Western Pacific region, followed by the African region, and the lowest prevalence was observed in the Americas (WHO, 2021). Additionally, the World Health Organization reported an estimate of approximately 1.5 million new infections per year and an annual mortality of 820,000 individuals mainly from complications of hepatocellular carcinoma and liver cirrhosis in 2019 (WHO, 2021). The prevalence of hepatitis B reported was lower in children below 5 years showing some evidence of successful and effectiveness of global vaccination programs against hepatitis B (Eng-Kiong & Lok, 2023).

Hepatitis B transmission is primarily through contact with infected blood and other bodily fluids through either vertical transmission from mother to child during childbirth (Cheung & Lao, 2020), sharing contaminated needles, and unprotected sexual contact or horizontally among contact sport athletes, in households, and institutionalized individuals (Sabeena & Ravishankar, 2022).

A number of risk factors such as risky sexual behaviors, injection drug use, unsafe medical procedures, occupational exposure (health workers), blood transfusion, organ transplant, close household contact, and migration/travel contribute to the acquisition and transmission of hepatitis B (Gerlich, 2013; Hahne et al., 2013; Platt et al., 2016; Terrault et al., 2018). It is therefore important to understand them in a bid to come up with preventive strategies that are effective and efficient.

The hepatitis B virus is very stable at 37ºC on environmental surfaces for longer than 22 days and therefore vaccination, safe behaviors, and strict compliance with available hygienic guidelines and safety precautions are required in preventing and controlling hepatitis B infection (Than et al., 2019).

1.1.2 Hepatitis B disease progression, treatment and prevention

Hepatitis B can be classified as either acute - a recent positive result of hepatitis B surface antigen (HBsAg) with a discrete onset of symptoms such as fever, headache, malaise, vomiting, nausea, abdominal pain, and diarrhea probably accompanied with jaundice and elevated ALT levels (CDC, 2021a) or chronic - characterized by the presence of HBsAg in blood for over six months with or without evidence of liver disease ranging from just chronic hepatitis to liver fibrosis, cirrhosis and/or liver cancer (CDC, 2021b).

Liver disease progression is related to the levels of hepatitis B deoxyribonucleic acid (DNA) in the blood and therefore a lack of anti-retroviral treatment whose mechanism is to suppress viral replication and reduce liver inflammation and fibrosis would increase the incidence of cirrhosis, hepatocellular carcinoma and liver failure (NICE, 2017). The American Association for the Study of Liver Disease (AASLD) released the updated practice guidelines for the prevention, diagnosis, and treatment of chronic hepatitis B in 2018 (Terrault et al., 2018). These highlighted an addition of tenofovir alafenamide (TAF) to the list of preferred hepatitis B therapies which previously included entecavir, peg interferon (peg-IFN), and tenofovir disoproxil fumarate (TDF) (Terrault et al., 2018). A positive response to treatment is usually evidenced by normalization of serum levels of ALT, reduction in serum hepatitis B DNA to undetectable levels, loss of hepatitis B surface and envelope antigens followed by seroconversion to their respective antibodies, and reduced levels of liver inflammation seen from biopsies (NICE, 2017).

Prevention of hepatitis B infection can be achieved through vaccination using the safe and effective hepatitis B vaccine whose active substance is the viral surface protein HBsAg (WHO, 2017). The most widely used vaccines for hepatitis B are recombinant vaccines and immunity achieved through vaccination is assessed by checking the levels of hepatitis B surface antibody (WHO, 2017). The world health organization recommends 3 doses of hepatitis B vaccines administered intramuscularly either into the anterolateral aspect of the thigh among infants or into the deltoid muscle for older children and adults (WHO, 2017).

1.1.3 Genetic risk for hepatitis B

Hepatitis infection has a wide spectrum of clinical manifestations ranging from symptomatic cases, subclinical to asymptomatic in the first few days or weeks and differential severity in patients with active chronic hepatitis infection (Lanini, Ustianowski, Pisapia, Zumla, & Ippolito, 2019). A combination of multiple factors such as host, viral and environmental factors have been suggested as causes of great differences in Hepatitis B infection outcome among individuals exposed to Hepatitis B virus (J. Xu, Zhan, Fan, Yu, & Zeng, 2021). Human genetic variation is one of the crucial determining factors of susceptibility to numerous day-to-day infectious diseases (Hill, 2006). Many genes are noted to have an effect on disease susceptibility to infectious microorganisms and can therefore be used to gain an understanding of mechanisms underlying protection and susceptibility to disease (Hill, 1998). Inter-individual genome variability significantly contributes to the differences identified among individuals in susceptibility and resistance to particular pathogens and their observed response to disease (Romero-Gomez, Eslam, Ruiz, & Maraver, 2011). This is revealed by differences in clinical outcomes and responses of infected individuals to the available treatment options (Romero-Gomez et al., 2011).

Genome-wide association studies (GWAS) focus on estimating and analyzing variations in DNA sequences across the genome in order to find genetic risk factors for common diseases within a given population (Bush & Moore, 2012). Genome-wide association studies have been widely used to identify genetic associations traits using genetic markers such as single nucleotide polymorphisms (SNPs) throughout the genome (Mozzi, Pontremoli, & Sironi, 2018). Fortunately, GWAS are now looking into the evaluation of susceptibility to infectious diseases, a method that provides important insights into new unanticipated risk factors for susceptibility to infections (Mozzi et al., 2018). A given number of these GWAS have been carried out on hepatitis B infection, viral clearance and disease progression in majorly European and Asian populations and little attention given to African populations (Akcay, Katrinli, Ozdil, Doganay, & Doganay, 2018). African populations are known for having highly diverse genomes (Tishkoff et al., 2009) and their inclusion in such studies could lead to the discovery of new risk factors and also provide a deeper understanding of diseases (Gurdasani, Barroso, Zeggini, & Sandhu, 2019). Additionally, the inclusion of different populations in genetic studies addresses the issue of equality in genetic findings available to the public.

It should be noted that even though GWAS is robust and a very highly powered method of determining variants associated with disease, it does not identify specific causal variants and precisely

explain the underlying biological mechanisms for these genetic associations (Lin et al., 2018). This limitation therefore raises the need for post-GWAS analysis such as fine mapping to pinpoint the specific causal variants that are driving the association (Schaid, Chen, & Larson, 2018), replication to validate association findings (Marigorta, Rodriguez, Gibson, & Navarro, 2018) and Mendelian randomization to perform causal inference between traits (Davey Smith & Hemani, 2014).

This study therefore sought to bridge the gap of underrepresentation of African populations in hepatitis B genetics research by conducting a genome-wide association study to determine the individual genetic impact on susceptibility to Hepatitis B infection. The study was carried out among Ugandan participants recruited from the general population cohort (GPC); a population-based open cohort consisting of individuals located in southwestern Uganda in Kyamulibwa sub-county of Kalungu district. This study aimed at facilitating the estimation of genetic effects of common and loss of function variants that may be potentially contributing to differential susceptibility of the study participants to hepatitis B, evaluate whether already reported findings were consistent in an African population from Uganda and assess the causal effect of hepatitis B infection on liver function biomarkers.

1.2 Statement of the Problem

Genome-wide association studies have previously been used to provide remarkable insights into the genetic impact on individuals by identifying genetic loci and allelic polymorphisms that determine and influence genetic susceptibility and response to viral hepatitis (Romero-Gomez et al., 2011; Tong, Bock, & Velavan, 2014). However, most of these studies have been carried out in European populations. There is little data from genetic studies among African populations, which are known to be highly genetically diverse, as evidenced by high levels of phenotypic variation (Henn, Cavalli-Sforza, & Feldman, 2012). Additionally, a considerable amount of the hepatitis B global burden is carried by Africa, particularly, sub-Saharan Africa (Sonderup et al., 2017).

A review study revealed that previous studies centered on cytokine genes, the human leukocyte antigen (HLA), and toll-like receptors to investigate the host genetic determinants of hepatitis B infection had yielded contradicting findings (Zhang et al., 2019). For instance, GWAS carried out among Asian, Thai, and Japanese populations revealed that polymorphisms in the HLA region were significantly associated with chronic hepatitis B infection while those among Chinese, Turkish, and Caucasian populations did not reveal these associations (Zhang et al., 2019). This seems to reinforce

the importance of carrying out population specific genetic studies to harness the opportunity to make new discoveries. This study seeks to bridge the gap of underrepresentation of African populations in hepatitis B genetics research by investigating the genetic determinants of hepatitis B infection susceptibility among Ugandans.

1.3 General Objectives

To investigate the host genetic impact on differential susceptibility to hepatitis B infection among the general population cohort in southwestern Uganda.

1.4 Specific Objectives

i) To identify existing and novel single nucleotide polymorphisms (SNPs) associated with susceptibility to hepatitis B infection among the Ugandan GPC participants.
ii) To fine-map association signals at new and existing disease trait loci to zero down on the causal variants within the loci responsible for each association.
iii) To check for replication of association findings in publicly available datasets
iv) To find and assess the causal effect of Hepatitis B infection on the serum levels of liver function biomarkers of Alanine aminotransferase (ALT) and Aspartate Transferase (AST).

1.5 Significance

The study findings will contribute to the representation of Ugandans (Africans) in global hepatitis B genetics research. The inclusion of Africans in genetics research has the potential to discover SNPs and variants associated with differential susceptibility to hepatitis B and reveal the underlying pathways leading to the phenotypic variations noted. Additionally, since genomes of African ancestry individuals have lower linkage disequilibrium and high genetic heterogeneity, their inclusion in genetic studies could provide remarkable improvements in the localization of specific causal variants during fine mapping. These genes and variants identified can direct targeted efforts towards control of viral hepatitis B and may identify biological pathways underlying low or high risk of contracting hepatitis B infection. These identified pathways have the potential to lead researchers to new drugs or drug targets for the management of hepatitis B. Additionally, results could also enable scientists to come up with better clinical management strategies through patient stratification and personalized medicine in Uganda and Africa at large.

1.6 Conceptual Framework

Figure 1: Conceptual Framework of the study

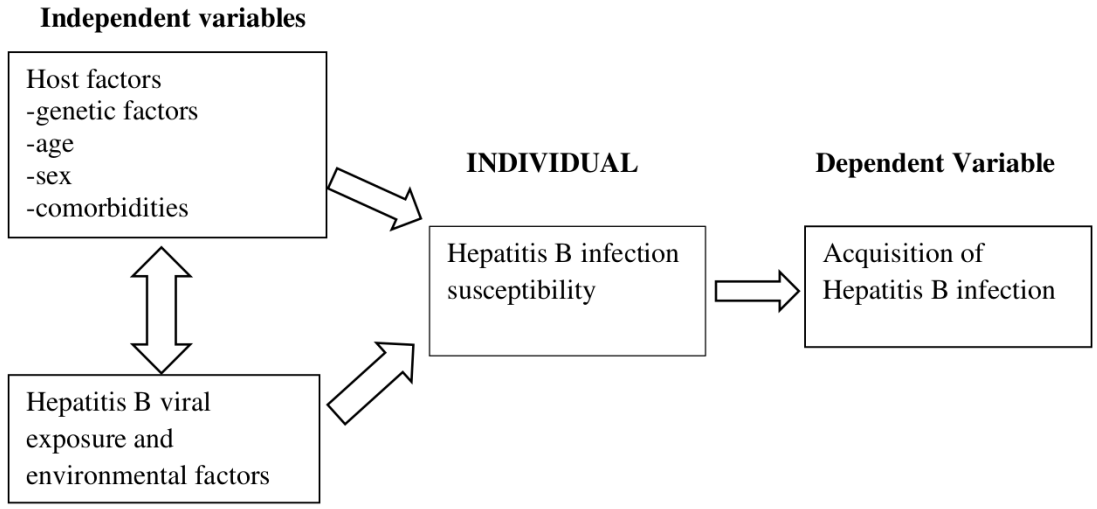


*Figure 1**:** Differential susceptibility to infection with viral hepatitis B is moderated by a combination of factors such as host factors, viral and environmental factors. An individual once exposed to hepatitis B virus may or may not get infected, and this variability may be caused by thepresence of certain variants within their genes that may protect them or increase their risk of infection.*

CHAPTER TWO: LITERATURE REVIEW

2.1 Burden of Hepatitis B

Viral hepatitis as a cause of liver inflammation and damage has had a tremendous impact on human health and has contributed to the morbidity and mortality of patients with acute and chronic hepatitis infection (Castaneda et al., 2021). Viral hepatitis may be caused by hepatitis virus strains A, B, C, D, and E, however, hepatitis strains B and C are the commonest causes of liver cancer, liver cirrhosis, and deaths due to viral hepatitis (WHO, 2022a). The World Health Organization reported approximately 354 million people to be living with hepatitis B or C in the world (WHO, 2022a). Hepatitis B virus causes acute and chronic hepatitis infections which may or may not generate mild or severe symptoms among the patients or resolve spontaneously (Pardee, 2019). Acute hepatitis B is short-term illness that occurs within 6 months after exposure to hepatitis B virus evidenced by the presence of jaundice, confirmed positive hepatitis B surface and core antigentests, and elevated alanine transaminase levels (Wilkins, Sams, & Carpenter, 2019). Chronic hepatitis B is the prolonged existence of hepatitis B surface antigen for longer than 6 months (Wilkins et al., 2019). According to the World Health Organization's global progress report on HIV, viral hepatitis, and sexually transmitted diseases released in 2021, 1.5 million new infections with acute and chronic hepatitis B were noted globally, with 820,000 deaths from hepatitis B (WHO, 2021). TheAfrican region alone contributed 990,000 (66%) new infections and 80,000 deaths from hepatitis B in 2019 alone consequently being the most burdened by hepatitis B infection in the world (WHO, 2021). The Ugandan Ministry of Health estimated the prevalence of chronic hepatitis B infection to be 4.1% however it differs from region to region with the highest number of cases being detectedin the north and the lowest in the Southwestern region (Ministry of Health, 2019).

2.2 Hepatitis B symptomatology and mode of transmission

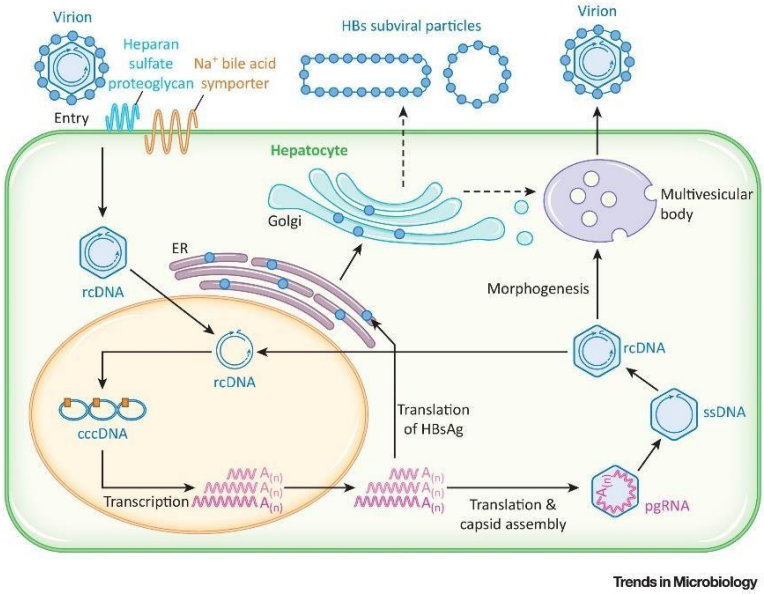
The symptoms of hepatitis B comprise loss of appetite, malaise, nausea, diarrhea, fever, jaundice, dark-colored urine, abdominal discomfort, and chronic liver inflammation which may result into liver scarring or cirrhosis and liver cancer and death (WHO, 2022a). Hepatitis B is transmitted through parenteral contact with infected body fluids such as saliva, cerebrospinal fluid,contaminated blood or blood products (serum and plasma), through the use of contaminated instruments for invasive products, and also by sexual contact (WHO, 2022a). Hepatitis B has also been reported to spread vertically at birth from mother to baby (Cheung & Lao, 2020) and horizontally through exposure to

bleeding wounds and sweat (Takata et al., 2020).

2.3 Hepatitis B genome, replication, pathogenesis, and diagnosis

Hepatitis B has a covalently closed circular DNA (ccc DNA) in its nucleus, it codes for pre-genomic RNA (pgRNA) that gets enclosed in a capsid with the aid of hepatitis B virus polymerase (Shih, Yang, Choijilsuren, Chang, & Liou, 2018). Replication of hepatitis B virus by reverse transcription takes place in the capsids where pgRNA is used as a template for stranded linear and then partially double-stranded relaxed circular (rc) DNA synthesis. Hepatitis B virus infection persists mainly due to ccc DNA and the liver's immune tolerance to hepatitis B virus antigens (Shih et al., 2018).

Figure 2: Replication cycle and natural history of infection and pathogenesis of hepatitis B



Adapted from (Shih et al., 2018).

Figure 2: Illustration of hepatitis B virus replication by reverse transcription. It takes place in the capsids to generate double-stranded relaxed circular DNA (rcDNA) and single-stranded DNA (ssDNA) from pgRNA. Capsids containing a mature rc DNA genome target the nucleus for cccDNA synthesis.

The genome of the virus contains 4 open reading frames that code for the viral polymerase, Hepatitis B core and envelope antigens, the regulatory protein HBx and the preS/S gene which codes for three surface antigens (LHBs, MHBs, and SHBs) (Schadler & Hildt, 2009). Hepatitis B surfaceantigen is an

envelope protein and is one of the markers for Hepatitis B virus infection (WHO, 2022b). It can be used to test for hepatitis B since it is usually produced excessively and can be detected in blood of individuals with acute and chronic hepatitis B infection (WHO, 2022b).

A given number of host factors have been reported to affect hepatitis B disease progression such as age, sex, alcohol intake, coinfection with human immunodeficiency virus (HIV) and hepatitis C virus (HCV), and other comorbidities such as diabetes and hypertension (Pisano et al., 2021). Previous studies have revealed that the course of infection is highly linked to the age at which the patient acquires the infection with mostly younger individuals more likely to get chronic infection (Z. Li, Hou, & Cao, 2015). Furthermore, studies on the impact of increased alcohol intake, the presence of coinfection with HIV and HCV, hepatitis D virus (HDV), and other comorbidities such as diabetes and hypertension and the male gender on hepatitis B diseaseprogression revealed a worsened disease progression to end-stage liver disease (Ganesan, Eikenberry, Poluektova, Kharbanda, & Osna, 2020; Iida-Ueno, Enomoto, Tamori, & Kawada, 2017; Loomba et al., 2013; Ruggieri, Gagliardi, & Anticoli, 2018; Sagnelli et al., 2012).

2.4 Applications of genome-wide association studies in genetic studies.

Genome-wide association studies are used to test for association between several thousands of SNPs and disease among a population of individuals and have been used to gain a deeper understanding of the genetics of complex diseases (Hardy & Singleton, 2009). Completion of the human genome project led to advances in technology which made it possible to use GWAS to detect genetic markers and study genetic variations across the genome that are associated to disease traits and complex diseases (Qiu et al., 2017). Genome-wide association studies have mostly been used to detect genetic associations for complex diseases, defined as those diseases that are influenced by both an individual's genes and the environment such as diabetes (Scott et al., 2007). Genome-wide association studies were originally more inclined to investigate genetic variants in complex diseases but are now expanding into infectious diseases (Mozzi et al., 2018).

Genome-wide association studies in infectious diseases such as human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS), malaria and tuberculosis have previously identified genetic markers that can be used to explain host protective mechanisms against pathogens and find novel diagnostic, prophylactic, and therapeutic moleculartargets for interventions (Mboowa, Sserwadda, Amujal, & Namatovu, 2018).

Genome-wide association tests are performed using software that incorporates statistical models that may be linear, logistic or mixed models. A good number of software are available foruse such as the Genome-wide Efficient Mixed model analysis software (GEMMA) (Zhou & Stephens, 2012) and Regenie (Mbatchou et al., 2021). These generate summary statistics that include a list of loci/SNPs, their chromosomal position and a p-value representing the statistical significance of the association test used. Manhattan plots and quantile-quantile plots are then used to interpret the significance of the association. Manhattan plots can be described as plots of the $\log_{10}$(p-value) of the statistical frequentist test used which can either be Wald's test, Likelihood ratio and score on the y-axis against the chromosomal position of the SNP on the x-axis (Turner, 2018). Quantile-quantile plots are best described as scatter plots of observable association p-value for all SNPs against the expected uniform distribution of p-values under the null hypothesis of no association and are used to detect the existence of population structure (Turner, 2018).

2.4.1 Genome-wide association studies of Hepatitis B

Genetic determinants such as allelic differences in the human leukocyte antigen (HLA), toll-like receptors, and cytokine genes have been studied in multiple ethnic populations to explain differential susceptibility to hepatitis B infection (J. Xu et al., 2021). In a study focusing on clarifying the effect of the *HLA-DP/DQ* gene on hepatitis B infection susceptibility in different nationalities, the *HLA-DQ* rs9272346 and 3 *HLA-DP* SNPs rs9277471, rs9277542 and rs9277535 were studied. This study's findings pointed to an association between *HLA-DP* rs9277535 and chronic Hepatitis B risk in the Han Chinese, another of *HLA-DP* rs9277471 and *HLA-DP* rs9277542 and decreased risk of hepatitis B and *HLA-DQ* rs9272346 was found to be associated with decreased risk of hepatitis B infection in the Han Chinese and Uygur populations (Xiang et al., 2016). A study investigating the effect of mitochondrial genetic variations among different nationalities of Chinese revealed differential susceptibility to hepatitis B infection (HBI), it noted that the mitochondrial DNA haplogroup F conferred susceptibility to HBI among the Yi nationality from Lijiang, China (Z. N. Chen et al., 2022). Cytokine genes produced by immune cells in the body are involved in the mediation of non-cytolytic virus clearance by targeting several steps in the hepatitis B virus lifecycle (Xia & Protzer, 2017). A given number of studies identified cytokines such as Interleukins (*IL-1B, IL-4, IL-6, IL-10, IL-10RB, IL-12A, IL-12B, IL-16, IL-18, IL-21, IL-21R*) and interferons (*IFN-ɣ, IFN- ɣR, IFN-AR2, IFN-LR1*) to be associated with susceptibility of humans to hepatitis B infection (J. Xu et al., 2021).

However, most of these studies were conducted in non-African populations and focused more on either the HLA region, cytokine genes, and toll like receptors and left out other regions of the genome that might be associated with hepatitis B infection. A study carried out among European-American (n=320) and African-American (n=85) individuals identified SNP rs9277534 *HLA-DPB1* to be associated with hepatitis B infection in African-Americans (Thomas et al., 2012) even with a small proportion of African-Americans included. This limited inclusion of Africans into these studies with no such study carried out in continental Africans leaves a gap in knowledge regarding the susceptibility loci associated with hepatitis B infection among them.

A study focusing on finding loci associated with hepatitis B surface antigen sero-clearance in chronic hepatitis B infection identified three novel susceptibility loci; rs7944135 on chromosome 11 (11q12.1), rs171941 on chromosome 5 (5q14.1) and rs6462008 on chromosome 7 (7p15.2) (Kim et al., 2018). These SNPs mapped to genes *MPEG1, MTX3, HOXA13, CSCL23,* and *DTX4* that were found to be of functional significance after functional analysis (Kim et al., 2018). Another study identified *HLA-C* and *UBE2L3* to be associated with chronic hepatitis B infection among Han Chinese (Hu et al., 2013). Findings from a GWAS investigating genetic determinants of persistent hepatitis B infection among Chinese revealed an association at a novel locus 8p21.3, rs7000921 and differential expression of *INTS10* gene which was reported to suppress viral replication via IRF3 in liver cells and highlighted to aid clearance of hepatitis B infection (Y. Li et al., 2016). However, most of these mentioned studies on hepatitis B infection have focused on finding genetic determinants of chronic hepatitis B infection and viral clearance of the infection and left out hepatitis B infection susceptibility as a trait.

2.4.2 Contribution of African genome diversity to genetic studies

Genetic diversity among populations evidenced by nucleotide sequence differences among individuals enables populations to adapt to their environment whereby a higher genetic diversity allows for higher acquisition of adaptive characteristics suitable for the environment (Rafael & Magnolia, 2021). Africa has been described as the origin of modern humans and Africans have been reported to have high genomic diversity as compared to other populations or individuals with other ethnicities (Tishkoff et al., 2009). This is mainly driven by the diet of Africans, their demographic history and exposure to diseases which resulted into alteration of their genotypes and phenotypes (Tishkoff et al., 2009). In addition to high genetic diversity, African populations are distinguished by considerable and large population structure and reduced linkage disequilibrium (LD) which may be

justified by their exposure to infectious diseases and response to diverse climate (Tishkoff et al., 2009). African populations have been deemed relevant and highly important for inclusion in genetic studies due to their differences in distribution of allelic frequencies as compared to non-African populations (Peprah, Xu, Tekola-Ayele, & Royal, 2015). However, only a few African populations have been included in GWAS with a representation of approximately 1.1% of the records available in the GWAS catalog (Fatumo, Chikowore, et al., 2022). Additionally, the poor transferability of genetic findings derived from the highly studied European populations into diverse populations such as Africa has been noted and explained by the population specific differences (Kamiza, Toure, et al., 2022). Uganda is very rich in ethnic, linguistic, and cultural diversity (Fatumo, Mugisha, et al., 2022) and therefore investigating such a population is a unique and unmatched chance tocomprehend the genetic structure of Ugandans. This reinforces the need to gain a deeper understanding of the impact of host genetics in phenotypic adaptation and disease susceptibility in African populations like Uganda. The inclusion of African populations will reduce the reliance on transfer of findings in precision medicine from non-African populations to African populations (Pereira, Mutesa, Tindana, & Ramsay, 2021).

2.5 Fine mapping in genetic studies.

Fine mapping is a statistical approach to identifying and prioritizing genetic variants that are most likely to causally influence a trait by analyzing genomic loci associated with that trait (Schaid et al., 2018). Fine mapping aims at finding the specific causal variant that is responsible for a given association with a trait since associations identified by genome-wide associations can be indirect because the tagged SNP could merely be associated with the actual causal SNP (Schaid et al., 2018). There are a number of fine mapping strategies such as heuristic (impractical and out of use), penalized regression models, and Bayesian methods (designed particularly for fine mapping and preferred over all other methods (Schaid et al., 2018). Bayesian method uses a Bayes factor to measure the evidence for association at a SNP where SNPs with bigger values of the Bayes factor were considered to be having stronger evidence forassociation (Frangou, 2013; Stephens & Balding, 2009). The Bayes factor was used to calculate the posterior probability for each SNP which was then used to make a comparison between SNPs within and across regions (Wellcome Trust Case Control et al., 2012). The posterior probability is the probability of the genotype configuration at a SNP in cases and controls under the alternative hypothesis divided by the probability of the genotype configuration at that SNP in cases and controls under the null hypothesis. The null hypothesis states that disease status is independent of the genotype at a particular SNP (Wellcome Trust Case Control et al., 2012).

2.6 Replication of genetic association findings

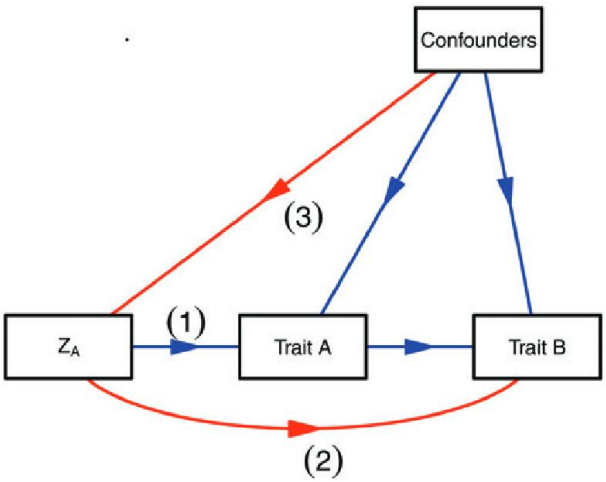
Replication of association findings assures researchers that the associations identified between the genotype and a given phenotype in a GWAS are acceptable and reliable and not due to chance or artifact brought about by uncontrolled bias (Kraft, Zeggini, & Ioannidis, 2009). Replication and reproducibility of association findings aid progressive science that allows for continued knowledge accumulation (Plomin, DeFries, Knopik, & Neiderhiser, 2016). Replication of initially reported results is a very credible method of verification of scientific discoveries and its success provides evidence of true positive status of the findings (Marigorta et al., 2018). A successful replication of a genetic marker would mean that it was found to be associated with that same trait/phenotype in an independent cohort (Kraft et al., 2009). Additionally, successful replication can also be demonstrated when the genetic marker is not significant in the independent cohort but has the same direction of effect as noted in the initial findings observed (Marigorta et al., 2018). Several other methods such as indirect replication where presence of positive association at variants in high LD with the genetic marker identified initially are also used as confirmation of replication (Marigorta et al., 2018). However, depending on the statistical significance threshold set for association in the independent cohort, a $p < 0.05$ may show replication and the genome-wide significance threshold ($p<5x10^{-8}$) is evidence of rediscovery of the genetic marker in the independent cohort (Marigorta et al., 2018). The simplest way of performing replication is by searching for the identified genetic markers in initial results in the NHGRI-EBI GWAS catalog (https://www.ebi.ac.uk/gwas/) (Marigorta et al., 2018). The NCI-NHGRI working group on Replication in association studies suggested some criteria for establishing positive replication which emphasized that replication studies had to be done in independent datasets with sufficient sample size and analyzing same or very similar traits of interest in a similar population (Chanock et al., 2007). Additionally, similar magnitude and direction of effect had to be achieved for the same SNP or with SNPs in very high LD with the SNP identified in the discovery dataset (Chanock et al., 2007). However, failure of replication in highly powered independent follow-up studies may not necessarily invalidate associations obtained since this could also be driven by LD and effect size differences across the different studies (Kraft et al., 2009).

2.7 Mendelian Randomization for causal inference

Mendelian randomization (MR) is a valuable method used to determine the causality of an association noted between a modifiable risk factor (exposure) and a clinically suitable outcome (Sekula, Del Greco, Pattaro, & Kottgen, 2016). Possible modifiable risk factors or exposures studied could be

effects of the environment, drug treatments, behavioral and other factors on human biology (Birney, 2022). Mendelian randomization uses genetic instruments (SNPs significantly associated with the tested exposure) as instrumental variables to deduce whether that exposure causes a given health or disease outcome (Bowden & Holmes, 2019). It is advantageous over randomized control trials because it is not prone to reverse causation bias, potential confounding and can be an alternative if randomized control trials are not feasible (Sekula et al., 2016). Additionally, the broad accessibility of published genetic associations provides instrument variables as the required input for MR studies hence making the approach more cost-effective and time efficient for use in determining the causal effects of exposures on outcomes (Sekula et al., 2016). Mendelian randomization leverages the fact that genetic variation is random at conception due to the random inheritance of random alleles from each parent thereby allowing the use of these variants to explain the possible causes of outcomes noted (Birney, 2022). Mendelian randomization enables the unbiased identification and estimation of the magnitude of causal effects by addressing the confounding problem (unobservable or unknown factors influencing the association between exposure and outcome (Davey Smith & Hemani, 2014). The accuracy of MR studies is based on three assumptions that must be carefully assessed and clearly interpreted before making a plausible causal inference (Lee, 2022). In a standard MR instance where trait A is hypothesized to cause trait B, the assumptions would be 1) the identified genetic instrumental variables should be robustly associated with trait A, 2) the genetic instrumental variable is not associated with the outcome except through the exposure and 3): the instrumental variable is not related to unmeasured or measured confounding factors (Davey Smith & Hemani, 2014). Several approaches are used to perform MR such as one sample MR which uses individual-level data from the same population to infer causality, two sample MR that requires GWAS results from two distinct sample populations of preferably the same ancestry, and bidirectional MR that assesses both directions of causation while assessing the causal effect (Burgess et al., 2019). Implementation of Mendelian randomization analysis in R can be done using methods such as inverse variance weighted, the most efficient method for meta-analysis, and others such as weighted median, simple median, weighted mode, and MR-EGGER for sensitivity analysis (Burgess et al., 2019).

Figure 3: Schematic representation of Mendelian randomization



Adapted from (Davey Smith & Hemani, 2014)

Figure 3: Mendelian randomization can be used to determine whether trait A has a causal effect on trait B given that the 3 assumptions; (1), (2) and (3) are fully met given that Z_A is a valid instrument. Z_A (1) is associated with trait A, (2) is not associated with trait B except through trait A and is not related to any confounders.

2.7.1 Application of Mendelian randomization in Hepatitis B

Two sample MR has previously been used to investigate the causal effect of chronic hepatitis B (CHB) infection on extrahepatic cancers and it revealed that CHB was causally associated with cervical and gastric cancers and also confirmed previous reports on its association with hepatocellular carcinomas among East Asians (Kamiza, Fatumo, Singini, Yeh, & Chikowore, 2022). Another study investigated the causal association between CHB and Coronavirus disease 2019 (COVID-19) in East Asians and revealed that CHB increased susceptibility and severity of COVID-19 among these individuals (Liu et al., 2023). However, these studies have been carried out in East Asian populations which leaves a gap in the application of the Mendelian randomization approach for causal inference of hepatitis B and possible outcomes among individuals of African ancestry. Additionally, Mendelian randomization seems less applied in hepatitis B genetic studies as noted in a PubMed database search with key terms Mendelian Randomization and Hepatitis B in August 2023 having revealed only 5 results.

2.7.2 Hepatitis B infection and liver function biomarkers

Liver function biomarkers are enzymes and proteins produced by the liver whereby the deviation of their levels from the expected normal range indicates different diseases and some include ALT, AST,

bilirubin, albumin, alkaline phosphatase (ALP) and gamma-glutamyl transferase (GGT) (ClevelandClinic, 2023). Some of the risk factors of the elevation of liver enzymes include hepatitis infection and exposure, alcohol use, certain medications and diabetes (ClevelandClinic, 2023). Acute and chronic liver infection with progression into liver cirrhosis are known effects of hepatitis B infection and elevated serum alanine transferase levels have previously been used to define acute hepatitis B infection (Wilkins et al., 2019). Elevation of AST and ALT levels is regarded as clinically significant during the natural history and treatment of CHB (Vaillant, 2021).

CHAPTER THREE: METHODOLOGY

3.1 Study design and setting

The research was a case control study investigating the host genetic determinants influencing differential susceptibility to viral hepatitis B among the GPC in southwestern Uganda. This study was nested in a primary study entitled "The epidemiology and genetics of communicable and non-communicable diseases in the general population cohort in Kyamulibwa, Uganda. The primary study recruited participants into the general population cohort and obtained samples and participant demographics which were used to generate the Uganda genome resource. Genotype and phenotype data used in this particular study was obtained from the Uganda genome resource.

Figure 4: Overall methodology workflow

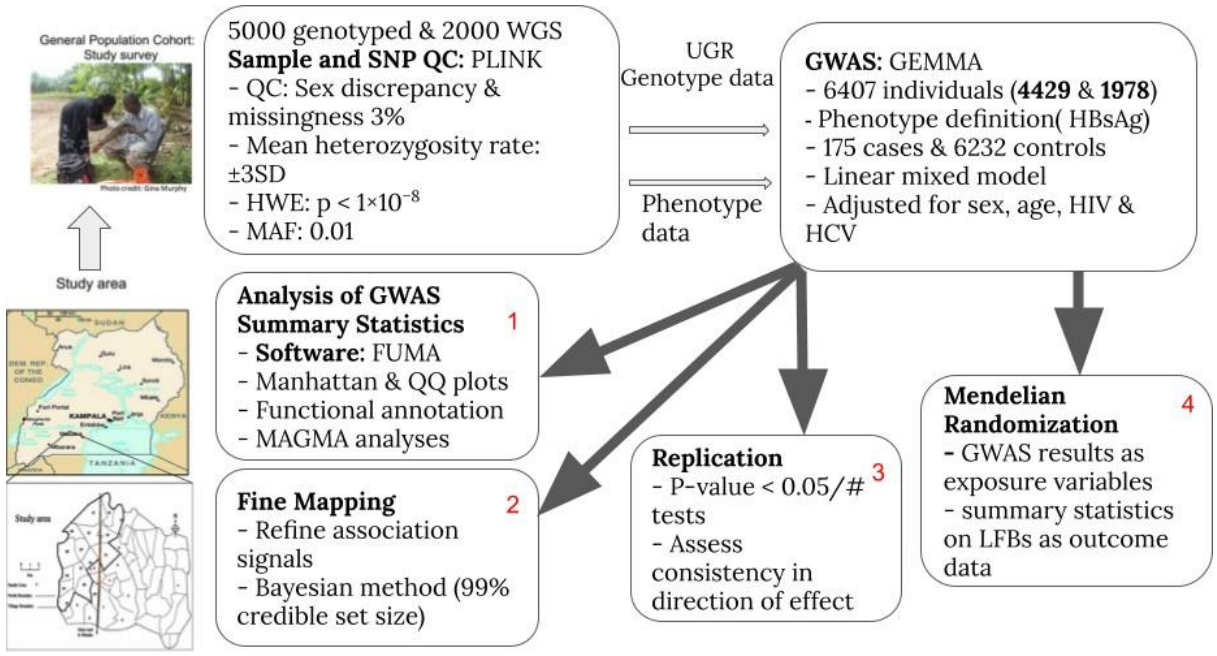


Figure 4: *Samples were obtained from individuals recruited in the GPC, 5000 genotyped, 2000 sequenced using Whole Genome Sequencing and all tested for HBV, HCV and HIV. The samples and SNPs obtained were taken through a quality control step in PLINK software using parameters such as sex discrepancy, sample and SNP missingness, heterozygosity rate, Hardy Weinberg Equilibrium (HWE) and minor allele frequency (MAF). The Uganda Genome Resource (UGR) comprises of samples and SNPs that passed the quality control criteria thresholds. UGR genotype and phenotype data was used as input to test for association and results analyzed to interpret summary statistics obtained and subsequent post GWAS analysis such as fine mapping, replication*

and Mendelian randomization.

3.1.1 GPC study participants

The General Population Cohort is a population based open cohort of about 22,000 individuals from 25 neighboring villages in Kyamulibwa sub county in Kalungu district in Southwestern Uganda (Asiki et al., 2013). The population is distributed across these villages which are marked by administrative boundaries with a few of the individuals concentrated in the small trading centers (Asiki et al., 2013). The cohort has been in existence since 1989 and was established by the Medical Research Council, United Kingdom in partnership with the Uganda Virus Research Institute, an umbrella organization under the country's Health Ministry (Asiki et al., 2013). The main aim of its establishment was to evaluate incidence and prevalence trends of HIV infection and their determinants in rural southwestern Uganda (Asiki et al., 2013). Genetic studies with theaim of providing insights into genetic variation in communicable and non-communicable diseaseswere introduced in 2010 and have since been used to reveal genetic determinants of diseases suchas Diabetes, Chronic Kidney disease among participants in this cohort (Asiki et al., 2013). The study population is recruited using house-to-house census rounds conducted annually during which study participants for the medical surveys are picked. The cohort is active and dynamic evidenced by new births, deaths and migrations in an out of the population noted and recorded on every round of follow up (Asiki et al., 2013). During survey round 22 conducted between 2010 and 2011, most attention was directed to genetics and epidemiology of communicable and non-communicable diseases (Fatumo, Mugisha, et al., 2022). Participant details of medical, sociodemographic and geographic factors were recorded, and participants assessed for major modifiable cardio-metabolic risk factors and infectious disease bio-markers such as Hepatitis B and Hepatitis C (Asiki et al., 2013).

3.1.2 Data collection

Anthropometric measurements such as body mass index and blood samples for Hepatitis B and Hepatitis C testing were collected by well trained nurses while using fully calibrated equipment and following stipulated standard operating procedures (Kalyesubula et al., 2018). Written informed consent on use of participant clinical record for research and sample storage for future use was obtained from the participants (Asiki et al., 2013).

3.1.3 Laboratory tests

Blood serum samples (8.5mls) for serological and biochemical analysis were collected in serum separation vacutainer tubes. The serum separation vacutainer tubes were centrifuged at 1000- 13,000 G for 10 minutes and kept at a fridge temperature of 4 - 8°C at the field station laboratory. Laboratory blood tests performed included HIV, HBV, and HCV testing. Human immunodeficiency virus (HIV) testing was done following the approved national algorithm using Determine, Statpak, and SD Bioline test kits. Hepatitis B and C virus testing was done usinghepatitis B surface antigen and HCV antibody tests respectively on the Cobas e 601 Auto Analyzer (Roche Diagnostics, North America) (Kalyesubula et al., 2018). All tests were carried out at the MRC/UVRI Entebbe laboratory (Kalyesubula et al., 2018).

3.1.4 The Uganda Genome Resource (UGR)

The Uganda Genome Resource (UGR) is an active distinguished and well-described genomic database that contains a wide spectrum of phenotypes ranging from communicable and non-communicable diseases and risk factors (Fatumo, Mugisha, et al., 2022). The UGR was obtained from 5,000genotyped, 2,000 sequenced, and 342 overlapping (sequenced and genotyped) individuals from nine ethnolinguistic groups in the GPC (Gurdasani, Carstensen, et al., 2019). Details of steps takenduring DNA extraction, genotyping, sequencing, quality control, imputation, and merging of UGR quality-controlled genotype and sequence data to generate a resource consisting of 6,407 individualsare well described in these publications (Fatumo, Mugisha, et al., 2022; Gurdasani, Carstensen, et al., 2019). The 6,407 individuals are a combination of 4,429 individuals with genotype, and 1,978 individualswith sequence data (Gurdasani, Carstensen, et al., 2019).

3.1.5 Phenotype definition

A phenotype is an observable characteristic exhibited in an individual or organism due to their genotype's interaction with the environment (Britannica, 2023). For this study, our phenotype (outcome) of interest was hepatitis B infection susceptibility which was defined by hepatitis Bsurface antigen status. In case control studies, an individual with the outcome of interest is termed as a case while a control does not have the outcome of interest (Tenny, Kerndt, & Hoffman, 2023).

3.1.5.1 Case definition

Participants with confirmed positive tests for hepatitis B surface antigen were classified as hepatitis B cases. These participants with positive hepatitis B surface antigen tests were considered to have been actively infected with Hepatitis B at the time of testing.

3.1.5.2 Control definition

Participants with a negative hepatitis B surface antigen test were classified as controls.

3.1.6 Inclusion and exclusion criteria

Inclusion criteria: Since this was using data from a primary study, all individuals and SNPs in the UGR were included in the study.

Exclusion criteria: SNPs that had a minor allele frequency of less than 0.01 and violated the Hardy-Weinberg equilibrium principle were excluded from the run in GEMMA for association testing.

3.1.7 Ethical Considerations

Written informed consent to undergo study procedures, use of participant clinical records for research purposes, and use of their samples for future genetic studies was obtained from participants in the GPC before all study procedures were conducted on every medical survey/round (provide the ref number here Appendix I). Study procedures included interviews (HIV risk questionnaire, sexual behavior, medical history), blood tests, and sample storage for future use. These all followed the guidelines set by the Uganda National Council of Science and Technology (UNCST) (Asiki et al., 2013).

The parent study titled Epidemiology and genetics of communicable and non-communicable diseases in the General Population Cohort in Kyamulibwa, Uganda obtained ethical approval from the Uganda Virus Research Institute Science and Ethics Committee, UVRISEC (AppendixII) and the Uganda National Council of Science and Technology, UNCST (Appendix III). Ethical approval and a waiver of informed consent for this study with reference number SBS-2023-298 were obtained from the School of Biomedical Sciences Research EthicsCommittee (SBSREC) for a period of 24/04/2023 to 24/04/2024 (Appendix IV).

3.3 Data management and analysis

3.3.1 Data preprocessing for GWAS

3.3.1.1 Phenotype data

An Excel file containing participant demographics including age, sex and body mass index was merged with another Excel file containing participant results for hepatitis B, HCV and HIV results. The merged Excel file with all complete participant information was imported into R and all participant results were coded as 0 for negative results and 1 for positive results whereas participant sex coded as 1 for males and 2 for females as required by the GEMMA software.

A phenotype file for hepatitis B infection susceptibility was prepared and it included one column where each line was a phenotype value of either 0s or 1s depending on each individual's hepatitis B result obtained. The phenotype file was prepared in the same order of individuals as the mean genotype file with the same number of rows as the number of individuals in the mean genotype file. A covariate file was prepared, and it included participant age, sex, HCV, and HIV results for all participants ordered in the same way as the phenotype file. The first column of the covariate file comprised a column of 1s as an intercept as expected by the GEMMA software. Missing phenotype or covariate information was coded as N/A.

3.3.1.2 Genotype data

Genotype data included a mean genotype file in BIMBAM format, a SNP annotation file, and a relatedness matrix. These files contained genotype information for chromosomes 1 to 22.

3.3.1.2.1 Mean genotype file

This file contained genotype information with the first column having the SNP identifier, the second column having the minor allele, the third column having the major allele and the remaining columns contained the posterior/imputed mean genotypes numbered between 0 and 2 for all the 6407 individuals. The data contained one mean genotype file for each of the 22 autosomal chromosomes.

3.3.1.2.2 SNP Annotation file

This file contained SNP information with the first column having the SNP identifier, the second column having its base pair position, and the third column having its chromosome number. This was required to ensure the SNP information columns in the output file for the association test generated by GEMMA did not have "-9" which denotes missing values.

3.3.1.2.3 Genotype Relatedness Matrix

The relatedness matrix contained information on the relatedness between given two individualsas required by the linear mixed model software provided by GEMMA. It was an *n x n* matrix where each row and each column corresponded to individuals in the same order as in the mean genotype file. In the matrix, the i^{th} row and j^{th} column was a number indicating the relatedness value between the i^{th} and j^{th} individuals. This was important such that the linear model can account for population stratification and sample structure.

3.3.2 Analysis of genome-wide association study data

Figure 5: Method workflow for the analysis of the GWAS data

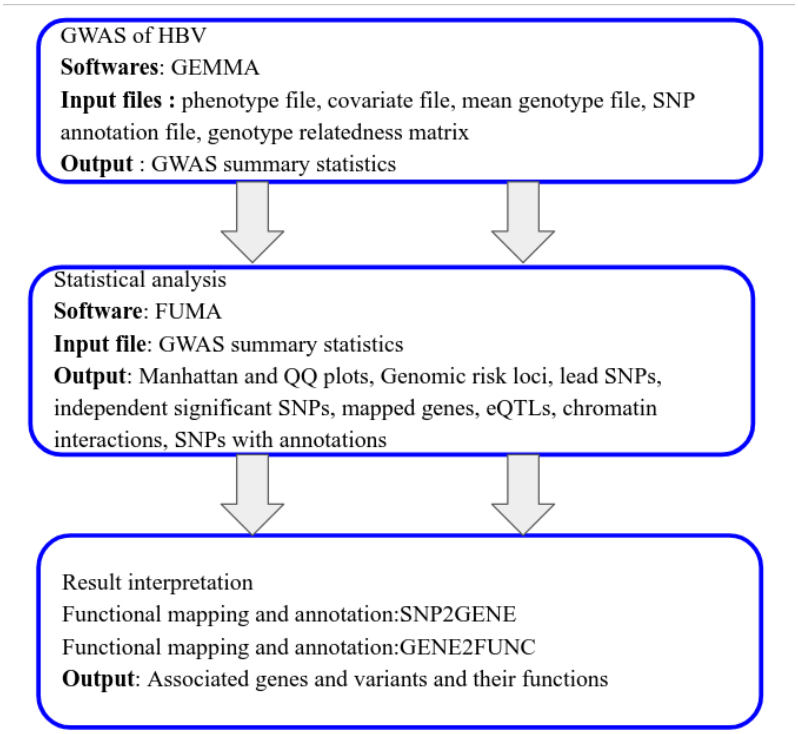


Figure 5:The GWAS of hepatitis B infection susceptibility was done in GEMMA software using phenotype, covariate, mean genotype, SNP annotation files and genotype related matrices as input files. GWAS summary statistics generated were uploaded to FUMA for generation, visualization and interpretation of results. Functional and tissue expression analysis was done on SNP and genes identified as associated with hepatitis B infection susceptibility.

3.3.2.1 Genome-wide association analysis of hepatitis B infection susceptibility

Genome-wide association study analysis was done to detect any genetic loci that could be associated with differential susceptibility to hepatitis B infection using Genome wide Efficient Mixed Model

Association Analysis (GEMMA) version 0.98.3 software (Zhou & Stephens, 2012). The GEMMA software uses applies the genome-wide Efficient Mixed Model Association algorithm for standard linear mixed models (Zhou & Stephens, 2012) GEMMA is fast, properly accounts for population stratification and relatedness, efficient and is a more computationally practical method for performing exact genome-wide association analysis for many individuals (Zhou & Stephens, 2012).

GWAS was performed for each chromosome using a linear mixed model while adjusting for age,sex, HCV, and HIV in GEMMA on the Uganda Medical Informatics Centre (UMIC) server. The model was set to exclude SNPs with a minor allele frequency of less than 1%. The likelihood ratio test was used as the frequentist test statistic.

Each run per chromosome generated two output files, one was a log text file that contained detailed information about the running parameters and the computational time used while the other was an association text file that contained the results. Columns in the association file included the chromosome number, SNP ID, base pair positions on the chromosome, number of missing valuesfor a given SNP, minor allele, major allele, allele frequency, beta estimates, standard errors for thebeta, restricted maximum likelihood estimates for lambda and p-values from the likelihood ratio test.

All the 22 association files containing summary statistics obtained for each chromosome were merged to make one complete file starting from chromosomes 1 to 22. The association file obtained was used for statistical analysis, post-analytical visualization, and interpretation of GWAS data.

3.3.2.2 Statistical Analysis

Visualization, interpretation, functional mapping and annotation of the generated summary statistics was done using Functional Mapping and Annotation of genome-wide Association Studies (FUMA GWAS). FUMA is an integrative web-based software that has two functions i) the SNP2GENE function that uses summary statistics to provide comprehensive functional annotation for all SNPs in genomic areas identified by lead SNPs and ii) the GENE2FUNC function that uses a list of gene IDs identified by the SNP2GENE function to provide their biological function (Watanabe, Taskesen, van Bochoven, & Posthuma, 2017; Watanabe, Umicevic Mirkov, de Leeuw, van den Heuvel, & Posthuma, 2019).

3.3.2.2.1 SNP2GENE

This is the main function of FUMA which takes GWAS summary statistics as input and mainly characterizes significant SNPs at a genome-wide significance threshold of 5×10^{-8} and performs

genome-wide analyses such as Multi-Marker Analysis of Genomic annotation (MAGMA) gene and gene set analysis. The SNPs in GWAS summary statistics uploaded were annotated with their biological function and gene mapped basing on positional, eQTL, and chromatin interaction information of SNPs. FUMA did this in two steps, first by characterizing independent significant SNPs and their surrounding genomic loci based on LD structure, identifying lead SNPs and genomic risk loci. Secondly, Independent significant SNPs and other SNPs in LD with the independent significant SNPs were annotated for functional consequences such as i) gene functions (based on Ensembl genes (build 85) using ANNOVAR), ii) deleteriousness score (based on combined annotation dependent depletion, CADD score), iii) potential regulatory functions (based on RegulomeDB score and 15-core chromatin state predicted by ChromHMM for 127 tissue/cell types), iv) effects on gene expression using eQTLs of various tissue types and 3D structure of chromatin interactions with Hi-C data. Lastly, functionally annotated SNPs were subsequently mapped to genes based on functional consequences on genes based on their physical position on the genome, eQTL associations, and 3D chromatin interactions. Independent significant SNPs and correlated SNPs were also linked to the GWAS catalog to findpreviously reported associations of the SNPs in these risk loci with a variety of other phenotypes.Manhattan, quantile-quantile (qq), regional, and MAGMA tissue expression plots and interactive tables were generated by FUMA for interpretation and visualization.

3.3.2.2.2 GENE2FUNC

This function was used to understand the putative biological mechanisms of prioritized genes by annotating these genes in biological context. FUMA does this by analyzing the biological information for each prioritized gene to find previous associations with any diseases as well as drug targets by mapping OMIM ID and Drug Bank ID. Patterns of tissue specific expression based on GTEx v6 RNA-seq data for each gene were visualized in an interactive heat map. FUMA was used to identify tissue specificity of prioritized genes for both single gene level analysis and overrepresentation in sets of differentially expressed genes (DEG; sets of genes which are more (or less) expressed in a specific tissue compared to other tissue types) for each of 53 tissue types based on GTEx v6 RNA-seq data.

FUMA also tested for the enrichment of prioritized genes in biological pathways and functional classes using the hypergeometric test against gene sets obtained from MsigDB and WikiPathways. Results from this analysis included an overview of the shared biological functions of prioritized genes

represented in a gene expression heat map, and text files showing enrichment of differentially expressed genes, overrepresentation of gene sets, and links to external biological information of input genes.

3.3.3 Fine mapping for localization of causal variants

Fine mapping was done in R/Rstudio version 4.2.1 using the Bayesian method. It was done to pinpoint individual variants that have an effect on the hepatitis B infection susceptibility among the SNPs whose p-value was below 5×10^{-8}. In this step, specific causal variants were characterizedand identified from the significant genomic loci. Bayesian approach of fine mapping was the method of choice because it had previously been reported to have fewer false positives and giving better fine mapping performance when compared to other fine mapping strategies such as the simple method based on R square (van de Bunt et al., 2015).

3.3.3.1 Procedure

i) All significant SNPs (SNPs with a likelihood ratio test p-value of less than 5×10^{-8}) were selected from the summary statistics using awk in the Linux command line.

ii) Significant SNPs were clumped to remove SNPs with high LD with the lead SNPs to generate a list of independent SNPs only. In this step, only themost significant SNPs (lowest p-value) in each LD block were identified for use in furtheranalysis. The clumping threshold excluded SNPs with an r^2 value greater than 0.1 while using an in-sample LD and within 500kb of each other.

iii) Z scores were calculated for all the SNPs in the association summary statistics by dividing their beta estimates by the corresponding standard error.

iv) The Bayes factor for each SNP in the summary statistics was calculated using thecalculated z score using the formula below.

$$BF_i = e^{\left[\frac{Z_i^2 - \log(K)}{2}\right]}$$, where k is the number of studies, in this case k was 1, BF_i is the Bayes factor and Z_i is the z score.

v) For each of the SNPs in the list of independent SNPs, SNPs that were 500000 base pair positions before and after that particular SNP were filtered and their posterior probabilities calculated. SNPs found to be with the highest posterior probability were considered to be

more likely to be the causal variant of the association at that locus.

vi) Posterior probabilities of each of these SNPs were computed using the Bayes factor as in the formula below:

$$\pi_i = \frac{BF_i}{\sum_j BF_j}$$

Where π_i, is the posterior probability, BF_i is the Bayes factor for SNP*i* (given SNP in the list of independent SNPs) and BF*j* represents the Bayes factor of SNPs 500000 base pair positions before or after the given SNP in the region/locus. Therefore, the summation inthe denominator defines the overall SNPs at the locus.

vii) The SNPs at that given locus were then sorted and arranged all the SNPs based on their posterior probability at the locus in descending order.

viii) The 99% credible set size was derived by counting the number of SNPs required to attain a cumulative posterior probability that is greater than or equal to 0.99.

ix) The SNPs with the smallest 99% credible set size were considered to be specifically causal anddriving the association in that locus.

3.3.4 Replication of genetic association findings.

Replication of findings was performed by downloading summary statistics of publicly available datasets from the GWAS catalog (https://www.ebi.ac.uk/gwas/). Due to the unavailability of publicly available summary statistics for download on the GWAS catalog for studies carried out among individuals of African ancestry looking at the same trait as our study. The summary statistics used were from two GWAS conducted among individuals of European ancestry in studies whose trait of interest was Hepatitis B virus surface antigen positivity. One (GCST90012669) was a clinical laboratory test-wide association scan aiming to find relationships between genetic risk and complex disease & physiological measurements (Dennis et al., 2021). This study's GWAS was performed using a linear model in fastGWA software among 2,255 European ancestry individuals (Dennis et al., 2021). The other study (GCST006356) aimed at evaluating the impact of genetic, environmental, and intrinsic factors such as age and sex on differential humoral responses to common disease pathogens and vaccines (Scepanovic et al., 2018). This GWAS study was conducted among 1000 French individuals recruited in the Milieu Intérieur cohort using logistic regression in PLINK

software (Scepanovic et al., 2018).

3.3.4.1 Procedure for replication

i) Significant SNPs obtained in the GWAS done in 3.3.2 above were searched for in the summary statistics of the downloaded data from the GWAS catalog for a GWAS of accession IDs GCST90012669 and GCST006356.

ii) The SNPs found in both were compared to check for consistency in the direction of effect by looking at beta values obtained and significance by looking at p-values obtained.

iii) A SNP with the same direction of effect as seen in the beta value and a p-value <= 0.05was considered significant and replication was a success.

3.3.5 Mendelian randomization

Determination of the causal effects of hepatitis B infection susceptibility on liver function biomarkers of ALT and AST was done using MR.

Figure 6: Schematic representation of the MR

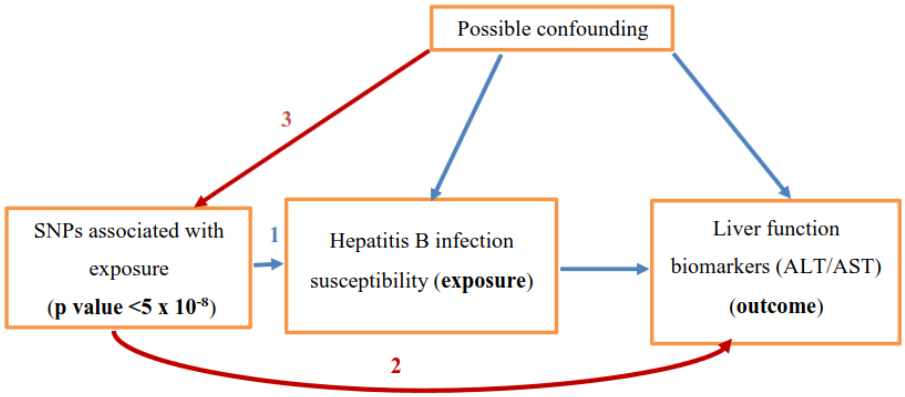


Figure 6: *A Mendelian randomization study to test the causal effect of Hepatitis B infection susceptibility on serum levels of liver function biomarkers AST and ALT provided that assumptions 1,2 and 3 are adequately met. 1): the SNPs used as instrumental variables are significantly associated with the exposure at genome-wide significance threshold of 5 x 10-8, 2): the instrumental variable is not associated with outcome except through the exposure and 3): the instrumental variable is not related to any confounding*

3.3.5.1 Mendelian Randomization flowchart

Figure 7: Method workflow for Mendelian Randomization

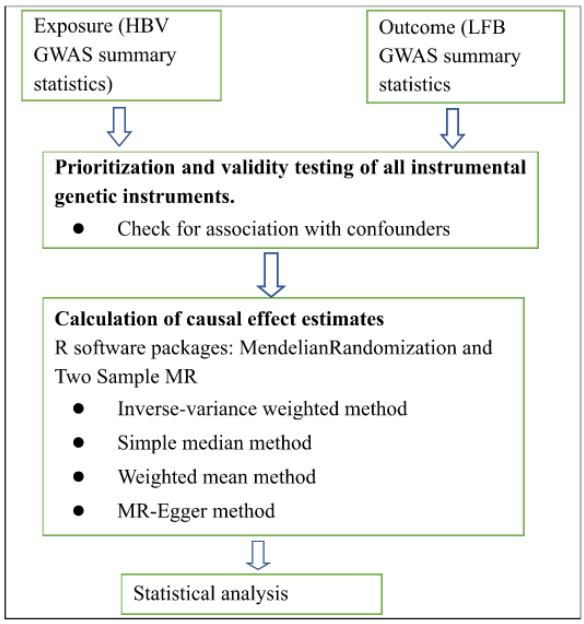


Figure 7*: Mendelian randomization to assess the causal effect of hepatitis B infection susceptibility (exposure) on liver function biomarkers ALT and AST (outcomes) was performed first by selecting and testing the validity of genetic instruments. Estimation of causal effect was done using Mendelian randomization and two sample MR packages in R/R studio using Inverse variance weighted, simple median, weighted median, and MR-egger methods.*

3.3.5.2 Procedure

i) Summary statistics obtained in the GWAS done in 3.3.2.1 above were used as the exposure (instrumental variable).

ii) Summary statistics of two independent GWAS among individuals of African ancestry for ALT (GCST90013663) and AST (GCST90013664) were downloaded from the GWAS catalog https://www.ebi.ac.uk/gwas/ and used as the outcome.

iii) Two independent MR analyses were performed with hepatitis B infection susceptibility as the exposure in both analyses, with ALT and AST serum levels as independent outcomes.

iv) Significant SNPs (SNPs with a likelihood ratio test p-value of less than 5×10^{-8}) selected from the full GWAS summary statistics from 3.3.2.1 above were searched for in the summary statistics for each of the outcome variables. SNPs found to be in common for both the exposure and outcome obtained from each of the outcome variables were selected for downstream analysis.

v) SNPs in the exposure and outcome were imported in R and Mendelian randomization

performed in R/Rstudio version 4.2.1 using the MendelianRandomization (Broadbent et al., 2020) and TwoSampleMR (Hemani et al., 2018) packages.

vi) Before the MR analysis, SNPs were clumped at a threshold of 500kb window at an r2 of 0.01 to remove SNPs that were in high LD with the lead SNPs. SNPs were harmonized to ensure palindromic SNPs were dealt with before the actual MR analysis was performed.

vii) MR analysis was done by use of the inverse-variance weighted (IVW) method for meta-analysis of individual causal effects of SNPs in the exposure on the outcome as the main method and simple median, MR-Egger, weighted mode, and weighted median as complementary methods.

viii) Sensitivity analysis such as the MR-Egger method was used to test and account for horizontal pleiotropy. Additionally, Cochran's Q test was performed to detect heterogeneity among the instrumental variables. Leave-one-out analysis was also used to test the robustness of the results to determine the undue influence of individual SNPs on the MR estimation.

ix) P-values and Confidence intervals were used to determine statistical significance. Causal estimateswith a p-value of less than 0.05 were considered significant.

x) Odds ratio (OR) of less than 1 indicated that the exposure decreased the risk of the outcome (protective exposure) while an OR of greater than 1 implied that the exposure increased the risk of the outcome.

xi) Causal estimates obtained were plotted on a forest plot for visualization and interpretation.

CHAPTER FOUR: RESULTS

4.1 Descriptive statistics of phenotype data

The data included 6407 participants with information on their age, sex, body mass index, hepatitis B, hepatitis C, and HIV status. Overall, 57.13% of the participants in this study were female, and the average age of participants included was 33.6 years. The estimated prevalence of Hepatitis B, Hepatitis C, and HIV in this study was 2.7%, 3.6%, and 7.2% respectively. The females had a higher prevalence of Hepatitis C and HIV while the males in this study had a higher prevalence of Hepatitis B.

Table 1: Summary of participant data.

	Female(N=3,660)	**Male(**(N=2,747)	**Overall**(N=6,407)
HBV results			
Negative	3,582 (97.9%)	2,649 (96.4%)	6,231 (97.3%)
Positive	77 (2.1%)	98 (3.6%)	175 (2.7%)
Missing	1 (0.0%)	0 (0%)	1 (0.0%)
HCV results			
Negative	3,528 (96.4%)	2,649 (96.4%)	6,177 (96.4%)
Positive	131 (3.6%)	97 (3.5%)	228 (3.6%)
Indeterminate/discrepant	0 (0%)	1 (0.0%)	1 (0.0%)
Missing	1 (0.0%)	0 (0%)	1 (0.0%)
HIV status			
Discrepant	1 (0.0%)	1 (0.0%)	2 (0.0%)
Negative	3,351 (91.6%)	2,586 (94.1%)	5,937 (92.7%)
Positive	305 (8.3%)	158 (5.8%)	463 (7.2%)
Missing	3 (0.1%)	2 (0.1%)	5 (0.1%)
Age			
Mean (SD)	34.6 (18.0)	32.2 (18.6)	33.6 (18.3)

Median [Min, Max]	31.0 [13.0, 94.0]	26.0 [13.0, 97.0]	29.0 [13.0, 97.0]
body mass index			
Mean (SD)	22.1 (4.08)	20.1 (2.98)	21.2 (3.77)
Median [min, max]	21.5 [12.8, 55.9]	19.9 [10.5, 48.6]	20.7 [10.5, 55.9]
Missing	188 (5.1%)	22 (0.8%)	210 (3.3%)

Table 1 is a summary showing an overview of participant data used for this study including hepatitis B results and baseline characteristics, age, sex, body mass index, HCV, and HIV status of the GPC participants.

As illustrated in table 2 below, baseline characteristics such as age, sex, body mass index, hepatitis C and HIV statuses of cases and controls were compared to check for significant differences to reduce bias. There was no significant difference in the mean age and body mass index of participants identified as cases as compared to those identified as controls. There was a significant difference between the number of males that identified as cases as compared to the females with the males having a higher prevalence of hepatitis B. There was no significant difference between Hepatitis C and HIV statuses of cases as compared to the controls.

Table 2: Comparison of baseline characteristics between cases and controls.

Baseline characteristic	Statistical test	Estimate	P-value
Age	Welch two sample t-test	-0.93289	0.3521
Body mass index	Welch two sample t-test	0.70373	0.4825
Sex	Pearson's Chi-squared test	12.132	0.0004955
Hepatitis C status	Pearson's Chi-squared test	0.79428	0.6722
HIV status	Pearson's Chi-squared test	7.0678	0.06977

Table 2 showing statistical tests run for comparison between the baseline characteristics between hepatitis B cases and controls.

Figure 8: Proportion of GPC participants with Hepatitis B infection grouped by sex.

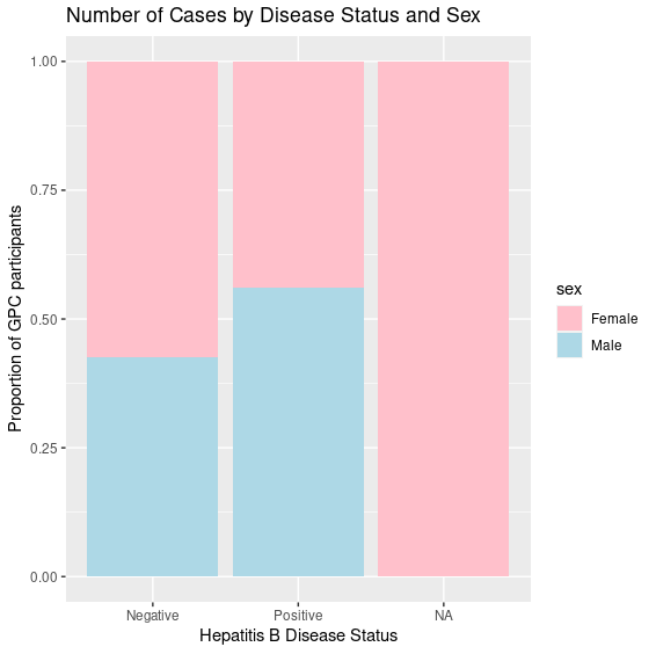


Figure 8 shows the proportion of GPC participants with hepatitis B infection grouped by sex.

4.2 Genome-wide association results

The GWAS summary statistics after running the association obtained had 16,558,784 SNPs and only 78 of these SNPs were considered to be of genome-wide significance while using the likelihood ratio test as a test statistic (p-value < $5x10^{-8}$).

4.2.1 Functional mapping and annotation: SNP2GENE

Functional mapping and annotation of GWAS summary statistics done using FUMAGWAS identified 22 genomic risk loci and 23 lead SNPs. Lead SNPs are independent SNPs maintained after clumping to remove SNPs that were in very high LD with them. Analysis of functional consequences of the SNPs on genes revealed that most of the SNPs were intergenic, followed by those that were intronic and only a few were ncRNA intronic and none were exonic as shown in Figure 9. None of the SNPs obtained had earlier been reported in the GWAS catalog for any other trait.

Figure 9: Functional Consequences of SNPs on genes

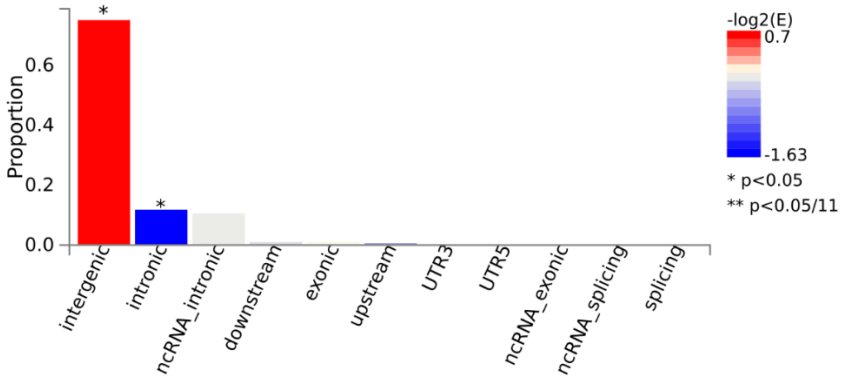


Figure 9 shows the proportion of SNPs with different functional consequences on genes and reveals that most of the SNPs obtained were intergenic.

Two genome-wide significance plots for visualization of GWAS summary statistics were made: a Manhattan plot and a qq plot. The Manhattan plot in Figure 10 depicts SNP association results in the genomic context. The SNP with the highest genome-wide significance was on chromosome 6. Figure 11 shows a Manhattan plot with SNPs on given chromosomes mapped to their nearest genes *POGK, MIXL1, RP11-24121.1, RP1-97D16.1, ADRB1,* and *RBFOX1*. The input SNPs were mapped to 18015 protein-coding genes and genome-wide significance of 2.774e-6 defined after Bonferroni multiple testing correction (0.05/18015). Only one gene (*TMEM236*) was shown to be of genome-wide significance as shown in Figure 12. The qq plot in figure 13 indicates compliance between the observed results and those expected by chance as the points were reasonably plotted along the slope except for a few outliers.

Figure 10: Manhattan plot of GWAS summary statistics

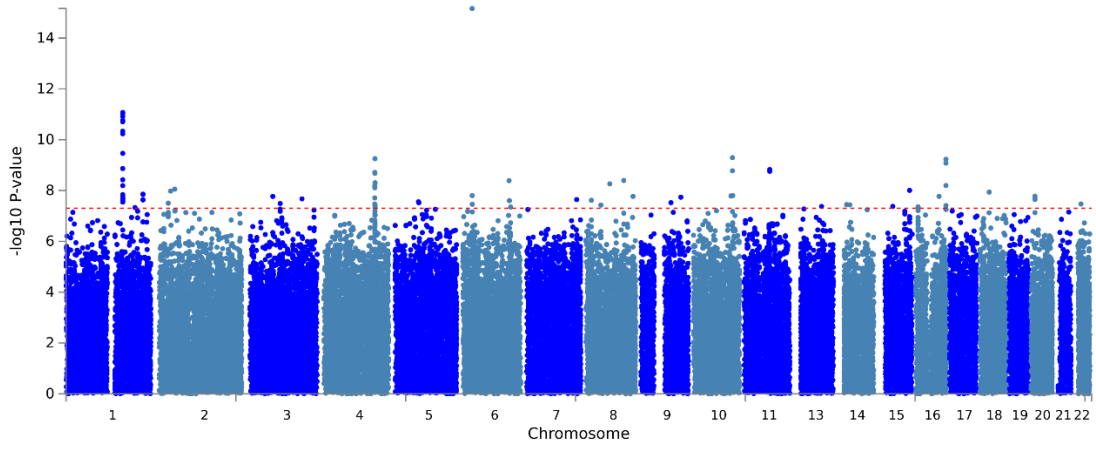


Figure 10 is a Manhattan plot illustrating the association between the SNPs in the UGR data and

hepatitis B infection susceptibility at genome-wide significance threshold of $5x10^{-8}$. The y-axis is the negative log_{10} of the p-value of the association between each SNP and hepatitis B infection susceptibility while the x-axis is the chromosome position. Each point represents a SNP and SNPs above the red dotted line represent genome-wide significant associations.

Figure 11: Manhattan plot of SNPs mapped to their nearest genes.

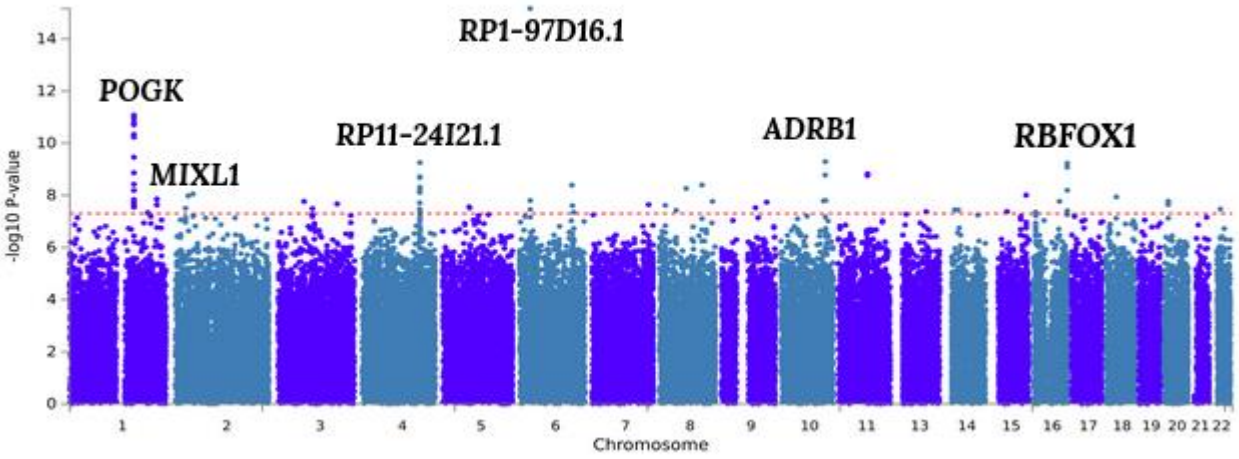


Figure 11 is a Manhattan plot showing genome-wide significant SNPs mapped to their nearest genes after SNP annotation.

Figure 12: Gene-based (protein-coding genes) Manhattan plot.

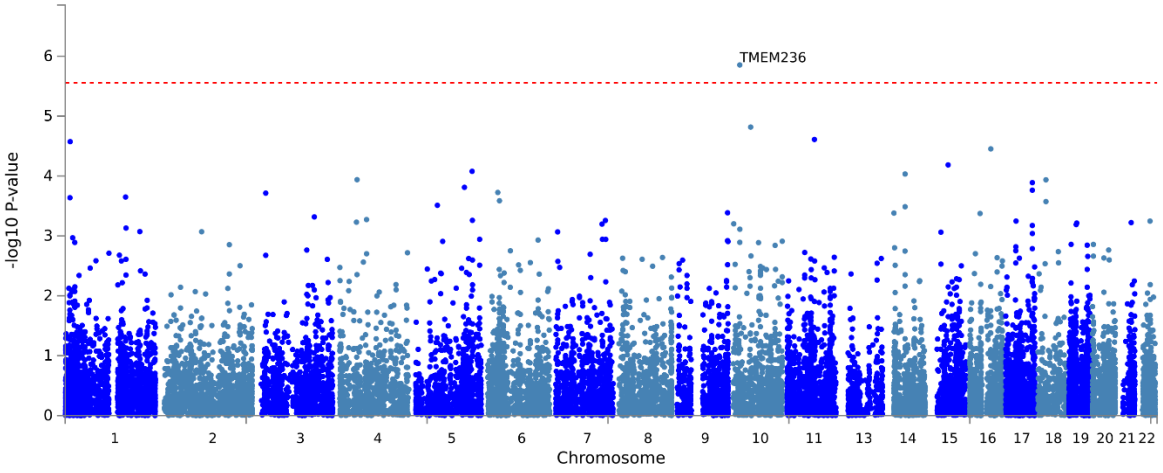


Figure 12 is a gene-based Manhattan plot after input SNPs were mapped to 18015 protein-coding genes. Genome-wide significance threshold (red dotted line) was defined at p-value=0.05/18015

=2.78e-6

Figure 13: The qq plot of GWAS summary statistics

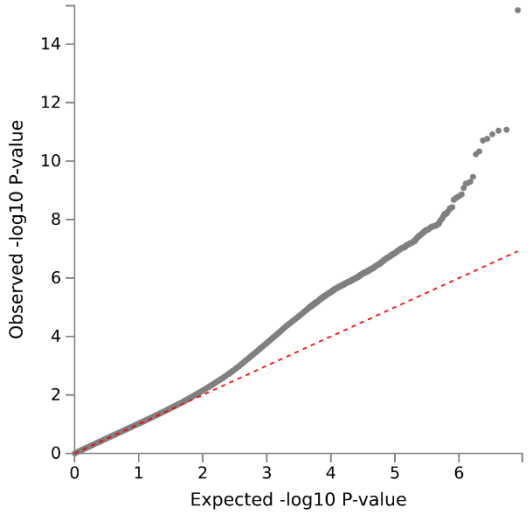


Figure 13 is a qq plot to assess population substructure in the GWAS of hepatitis B infection susceptibility. The x axis depicts the expected –log10 of the p-values while the y axis has the observed negative log10 p-values.

4.2.2: Functional mapping and annotation: GENE2FUNC

24 genes were used as input for the GENE2FUNC and all of them had recognized Ensembl IDs. There was no significant differential expression of the genes nearest to the significant SNPs in the summary statistics. These genes *ADRB1, MIXL1, POGK,* and *RBFOX1* as seen in figure 14 do not show any differential gene expression in the liver and other tissues expected to be affected by Hepatitis B infection. There were no significantly enriched differentially expressed genes when investigated in 54 tissue types as shown in figure 15 below.

Figure 14: Gene expression heat map

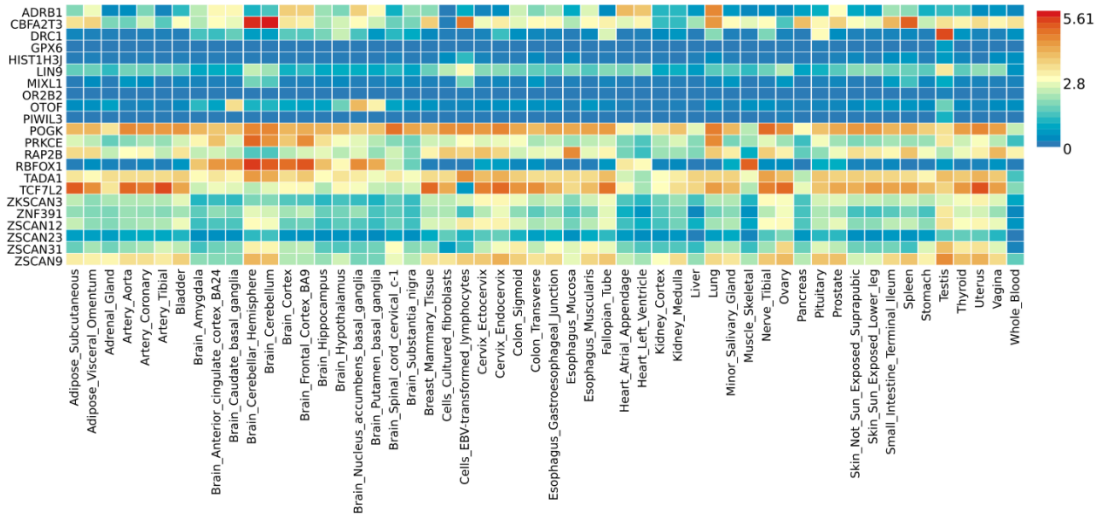


Figure 14 is a heat map showing gene expression of 24 genes mapped to identified genome wide associated SNPs in 54 tissue types.

Figure 15: Differentially expressed genes.

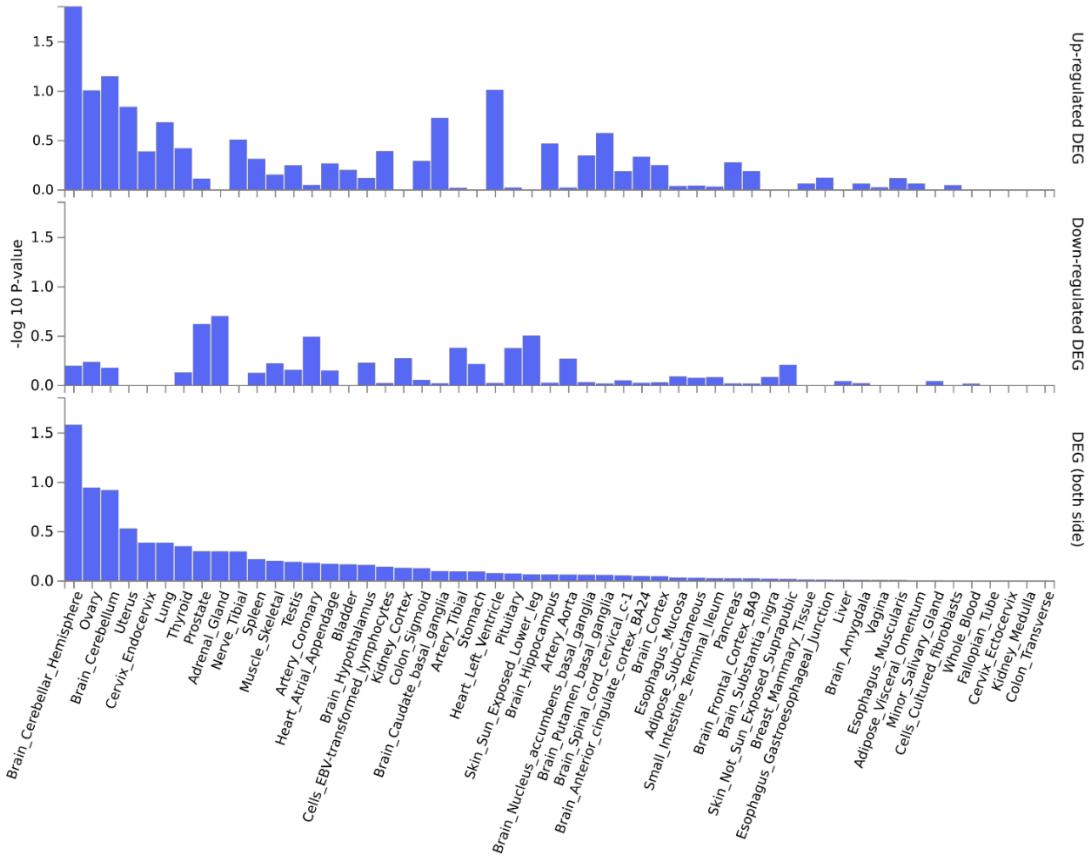


Figure 15 is a plot showing differentially expressed genes used for testing tissue specificity defined

for each label of each expression data set.

4.3 Fine mapping results

20 independent lead SNPs were obtained after clumping SNPs all 78 significant SNPs from the summary statistics in step ii of section 3.3.3.2 above. The number of credible sets (99% credible set size) derived for each of the 20 SNPs by counting the number of SNPs required to attain a cumulative posterior probability that is greater than or equal to 0.99 is represented in table 3 below. The SNP annotations that include the nearest gene to the lead SNP in that locus and their respective positions and distances from these particular genes are also included in table 2. 3.SNPs with the smallest 99% credible set size were considered as probable specific causal variants that were possibly driving the association in that locus. As noted in the results in table 3 below, one of the SNPs with the smallest credible set size was in an intronic position while the one in an intergenic position was quite distant from its nearest gene. The SNPs nearest to the genes identified in the functional mapping and annotation: SNP2GENE section in 4.2.1 above *MIXL1, POGK,* and *ADRB1* had considerably small credible set sizes of 5,8 and 9 respectively, and were considered to be the ones driving the associations in their respective loci.

Table 3: Bayesian fine mapping results summary

SNP rsid	Chr:bp	Nearest gene	Distance from gene	position	EA	NEA	maf	b	se	P-value	#
rs113453487	2:45902660	*PRKCE*	0	intronic	C	G	0.011	0.083	0.014	8.82e-09	1
rs190525861	6:27676205	*RP1-97D16.1*	28573	intergenic	A	C	0.01	0.085	0.015	1.58e-08	1
rs187715946	10:114957080	*TCF7L2*	29642	intergenic	G	T	0.014	0.083	0.014	1.67e-09	3
rs114975164	18:27822248	*RP11-675P14.1*	470992	intergenic	A	T	0.018	0.068	0.012	1.16e-08	4
rs115683525	16:89071260	*CTD-25555A7.1*	9482	intergenic	G	A	0.033	0.052	0.008	5.91e-10	5
rs151106850	**1:226401873**	***MIXL1***	**9445**	**intergenic**	**C**	**T**	**0.021**	**0.062**	**0.011**	**1.39e-08**	**5**
rs140791405	2:26653884	*DRC1*	0	intronic	T	C	0.015	0.068	0.012	3.15e-08	5
rs114905599	3:152978084	*RN7SL300P*	60277	intergenic	T	C	0.016	0.073	0.013	2.13e-08	5
rs76237083	20:11707956	*RP11-268G13.1*	39984	intergenic	C	T	0.011	0.082	0.014	1.68e-08	7
rs73310381	8:111108679	*RP11-403P13.1*	9030	intergenic	A	G	0.051	0.041	0.007	4.03e-09	7
rs75277414	**1:166807469**	***POGK***	**1211**	**intergenic**	**G**	**A**	**0.012**	**0.094**	**0.014**	**8.46e-12**	**8**
rs12159250	22:25144676	*PIWIL3*	0	intronic	T	C	0.022	0.056	0.01	3.39e-08	8
rs114388659	9:89143763	*RP11-395D3.1*	97304	intergenic	G	T	0.035	0.045	0.008	3e-08	8
rs17875425	**10:115798872**	***ADRB1***	**4933**	**intergenic**	**G**	**A**	**0.083**	**0.031**	**0.005**	**1.59e-08**	**9**
rs76175242	3:88931515	*NARG2P2*	65776	intergenic	A	G	0.012	0.081	0.015	3.25e-08	9

rs79126961	15:94755691	*MCTP2*	19075	intergenic	G	A	0.011	0.084	0.015	9.89e-09	12
rs115069782	4:150195675	*RP11-24I21.1*	0	intronic	C	G	0.013	0.086	0.014	5.57e-10	16
rs116723051	8:15801883	*RP11-77B22.1*	841	downstream	C	G	0.012	0.078	0.014	2.42e-08	23
rs4424342	9:118284179	*DEC1*	119632	intergenic	A	G	0.031	0.048	0.009	1.84e-08	47
rs148992333	6:140822037	*AL356137.1*	159722	intergenic	T	A	0.013	0.072	0.013	4.48e-08	71

Table 3 showing localization of specific causal variants and the credible set size for each of the 20 lead SNPs. The abbreviations are defined as, EA: effect allele, NEA: non effect allele, maf: minor allele frequency, b: beta effect estimate, se: standard error of the beta, #: number of credible sets (99% credible set size)

4.4 Replication of genetic association findings

Replication of association findings done in the two independent cohorts of European individuals failed to reproduce and validate the findings obtained in our discovery study. Two SNPs rs4424342 and rs12159250 at chromosome 9 and 22 at base pair positions 118284179 and 25144676 respectively from the discovery dataset were found to be common and shared in both replication cohorts. However, due to the insignificant p-values of these SNPs in the replication cohorts, the replication was considered a failure. As summarized in table 4, SNP rs4424342 had the same direction of effect and almost similar magnitude of effect in both studies, however, SNP rs12159250 had a remarkably different direction and magnitude of effect. In contradiction, table 5 shows SNP rs4424342 had a different direction and magnitude of effect while SNP rs12159250 had the same direction of effect with different magnitudes.

Table 4: Replication with GWAS study accession ID GCST90012669.

rsid	**Chr:pos (Build 37)**	**Discovery GWAS**							**Replication cohort:**					
		Nearest gene	**EA**	**NEA**	**BETA**	**SE**	**maf**	**p-value**	**EA**	**NEA**	**BETA**	**SE**	**maf**	**p-value**
rs4424342	9:118284179	*DEC1*	G	A	-0.048	0.009	0.023	1.838024e-08	G	A	-0.027	0.050	0.086	0.58366
rs12159250	22:25144676	*PIWIL3*	C	T	-0.056	0.010	0.036	3.385387e-08	C	T	0.0103	0.062	0.057	0.86831

Table 4 showing replication of GWAS on hepatitis B infection susceptibility using GCST90012669 summary statistics to validate association findings. Abbreviations are defined as: EA: effect allele,

NEA: non effect allele, maf: minor allele frequency, BETA: beta effect estimate, SE: standard error of the beta, chr:pos; Chromosome:base pair position.

Table 5: Replication with GWAS study accession ID GCST006356.

rsid	chr:pos (Build 37)	Discovery GWAS							Replication cohort:					
		Nearest gene	EA	NEA	BETA	SE	maf	p-value	EA	NEA	BETA	SE	maf	p-value
rs4424342	9:118284179	*DEC1*	G	A	-0.048	0.009	0.023	1.83e-08	G	A	0.00069	0.189	0.072	0.97087
rs12159250	22:25144676	*PIWIL3*	C	T	-0.056	0.010	0.036	3.39e-08	C	T	-0.2417	0.203	0.065	0.23417

Table 5 showing replication of GWAS on hepatitis B infection susceptibility using GCST006356 summary statistics to validate association findings. Abbreviations are defined as: EA: effect allele, NEA: non effect allele, maf: minor allele frequency, BETA: beta effect estimate, SE: standard error of the beta, chr:pos; Chromosome: base pair position.

4.5 Causal effect of hepatitis B infection on serum levels of liver function biomarkers AST and ALT using Mendelian randomization.

For each of the MR analyses, 19 variants were selected for downstream analysis. However, 3 of them were removed for being palindromic and 7 of them for being in high LD with the lead SNPs. A total of 9 SNPs were therefore used for the MR analysis.

The MR analysis was performed using 9 SNPs and the results of the IVW method showed no causal effect of hepatitis B infection susceptibility on AST (OR_{IVW}=1.01130, 95% CI 0.54575 - 1.87397, p= 0.97153) and MR Egger (OR=0.58022, 95% CI 0.10572-3.18442, p= 0.55078) as shown in table 6. For ALT, MR analysis was also performed using 9 SNPs and the results of the IVW method also showed no causal effect of hepatitis B infection susceptibility on ALT (OR_{IVW}=0.85428, 95% CI 0.45877-1.59075, p=0.61952) and MR Egger (OR=1.50017, 95% CI 0.25295-8.89690, p=0.66868) as shown in table 7. The sensitivity analysis results in table 8 show no evidence of horizontal pleiotropy and heterogeneity.

Plots were made for visualization of the obtained results as shown below. Forest plots showing the causal effect of hepatitis B infection susceptibility on AST and ALT serum levels in figures 16 and 20 respectively. Scatter plots showing the effect of SNPs in the exposure on SNPs in each of the outcomes are shown in figures 17 and 20 respectively. Forest plots in Figure 18 show the effect of causal SNPs before and after running the leave-one-out analysis when AST was used as an outcome

and in Figure 21 for ALT as an outcome. The leave-one-out sensitivity analysis by removing one SNP at a time showed stable results for both outcome variables.

Table 6: Causal effect of hepatitis B infection susceptibility on AST serum levels

id. exposure	id. outcome	method	N snp	Beta	Standard error	p-value	Odds ratio	or_lci95	or_uci95
HBV	AST	Inverse variance weighted	9	0.01123	0.31471	0.97153	1.01130	0.54575	1.87397
HBV	AST	MR Egger	9	-0.54435	0.86868	0.55078	0.58022	0.10572	3.18442
HBV	AST	Weighted median	9	-0.08002	0.41550	0.84728	0.92310	0.40885	2.08413
HBV	AST	Weighted mode	9	-0.28601	0.49098	0.57625	0.75125	0.28698	1.96659
HBV	AST	Simple median	9	0.01561	0.45610	0.97270	1.01573	0.41547	2.48326

Table 6 showing the causal effect of hepatitis B infection susceptibility on serum AST levels using 9 instrumental variables. Abbreviations: or_lci95: lower 95% confidence interval, or_uci95: upper 95% confidence interval

Table 7: Causal effect of hepatitis B infection susceptibility on ALT serum levels

id. exposure	id. outcome	method	N snp	Beta	Standard error	p-value	odds ratio	or_lci95	or_uci95
HBV	ALT	Inverse variance weighted	9	-0.15750	0.31720	0.61952	0.85428	0.45877	1.59075
HBV	ALT	MR Egger	9	0.40558	0.90823	0.66868	1.50017	0.25295	8.89690
HBV	ALT	Weighted median	9	-0.02695	0.41635	0.94839	0.97341	0.43042	2.20139
HBV	ALT	Weighted mode	9	0.03012	0.59787	0.96105	1.03058	0.31927	3.32662
HBV	ALT	Simple median	9	-0.03815	0.48568	0.93739	0.96257	0.37155	2.49373

Table 7 showing the causal effect of hepatitis B infection susceptibility on serum ALT levels using 9 instrumental variables. Abbreviations: or_lci95: lower 95% confidence interval, or_uci95: upper 95% confidence interval

Table 8: Heterogeneity and horizontal pleiotropy tests for the MR.

Id. exposure	Id. outcome	test	Method	Effect size	P-value
HBV	AST	Heterogeneity	Cochran's Q test (MR Egger)	Q=6.518734	0.4806444

			Cochran's Q test (IVW)	Q=6.989581	0.5377572
HBV	AST	Horizontal Pleiotropy	MR Egger regression	Egger intercept = 0.03781314	0.5146758
HBV	ALT	Heterogeneity	Cochran's Q test (MR Egger)	Q=8.260569	0.3101791
			Cochran's Q test (IVW)	Q=8.782601	0.360663
HBV	ALT	Horizontal Pleiotropy	MR Egger regression	Egger intercept = -0.03833156	0.5272789

Table 8 showing the heterogeneity and horizontal pleiotropy tests done for the MR.

Figure 16: A forest plot for causal effect of hepatitis B infection susceptibility on AST levels

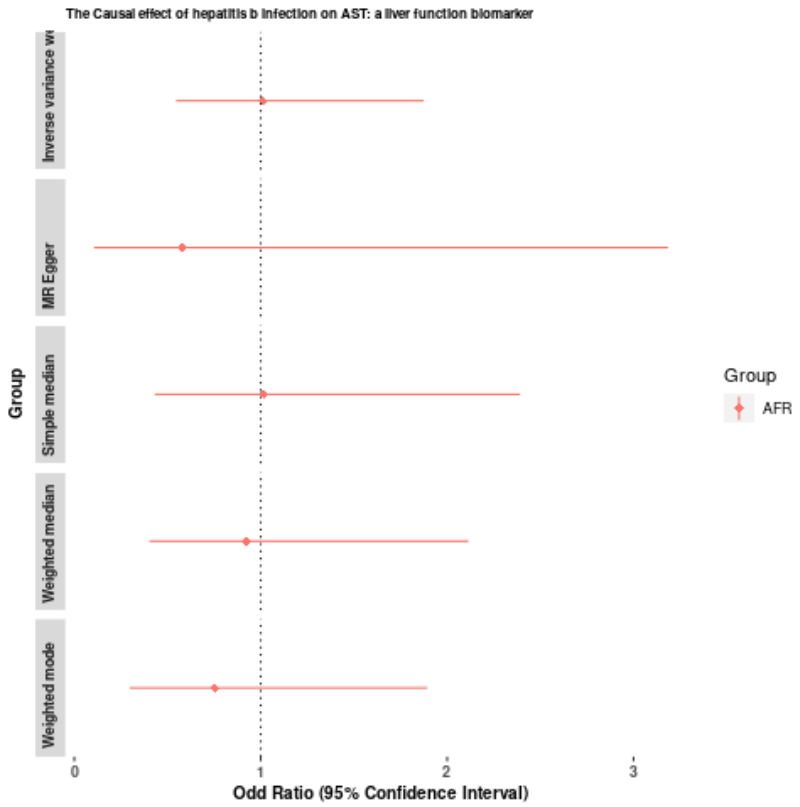


Figure 16 is a forest plot showing the association between hepatitis B infection susceptibility and AST levels with all the methods used depicted.

Figure 17: Scatter plot for causal effect of hepatitis B infection susceptibility on AST levels.

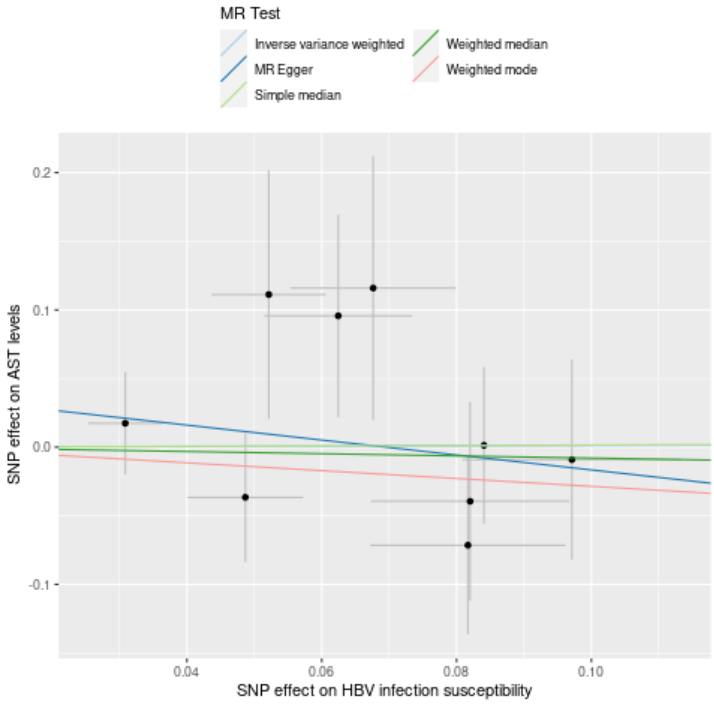


Figure 17 is a scatter plot showing the association between hepatitis B infection susceptibility and AST levels with all the methods used depicted.

Figure 18: Forest plots of causal effect before and after doing the leave one out analysis (left to right)

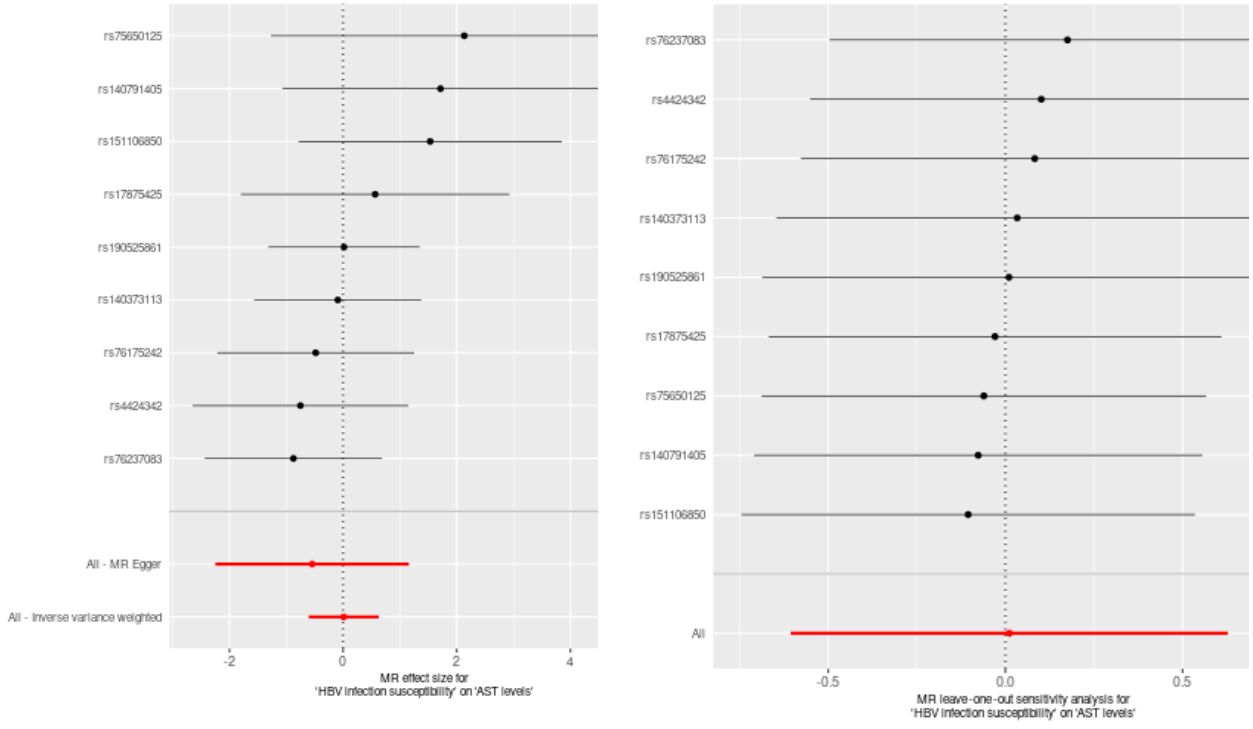


Figure 18: On the left; a forest plot showing the MR estimate and 95% confidence intervals (gray line segment) for each of the 9 SNPs and also shows the MR Egger and IVW results at the bottom. On the right, a forest plot showing leave one out analysis to evaluate whether any single instrumental variable was driving the causal effect.

Figure 19: A forest plot of causal effect of hepatitis B infection susceptibility on ALT levels

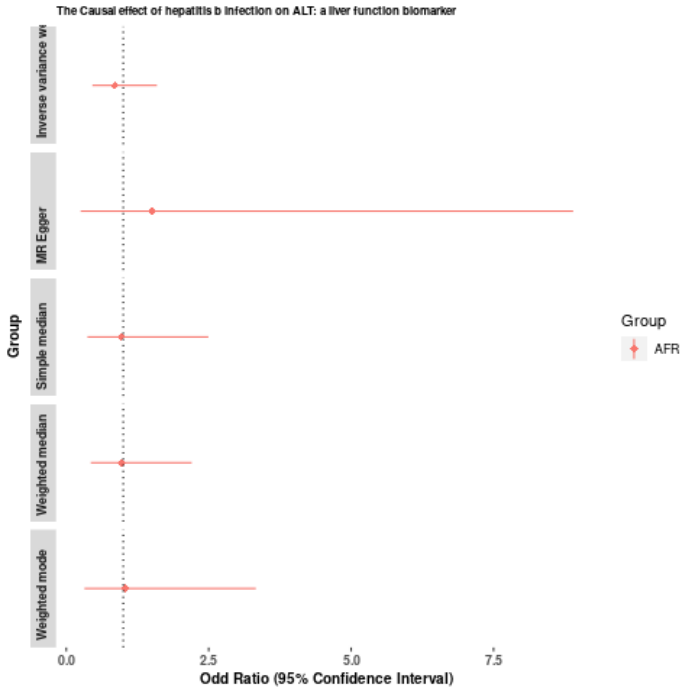


Figure 19 is a forest plot showing the association between hepatitis B infection susceptibility and ALT levels with all the methods used depicted.

Figure 20: Scatter plot for causal effect of hepatitis B infection susceptibility on ALT levels.

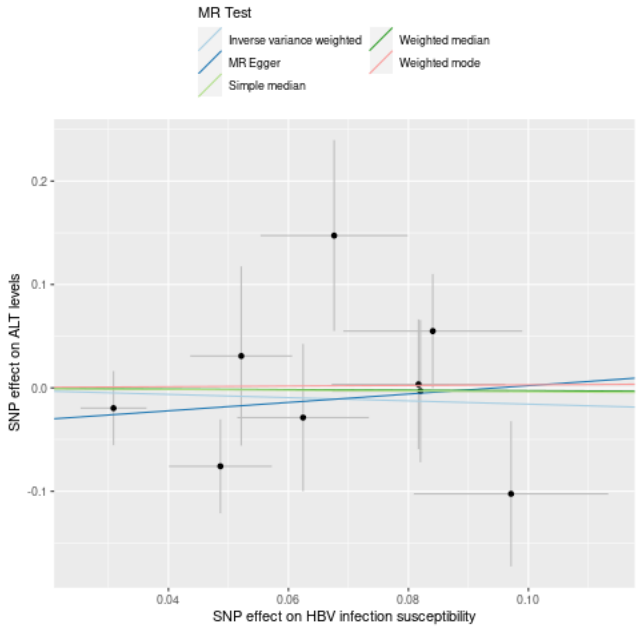


Figure 20 is a scatter plot showing the association between hepatitis B infection susceptibility and ALT levels with all the methods used depicted.

Figure 21: Forest plots showing causal effect before and after doing the leave one out analysis(left to right)

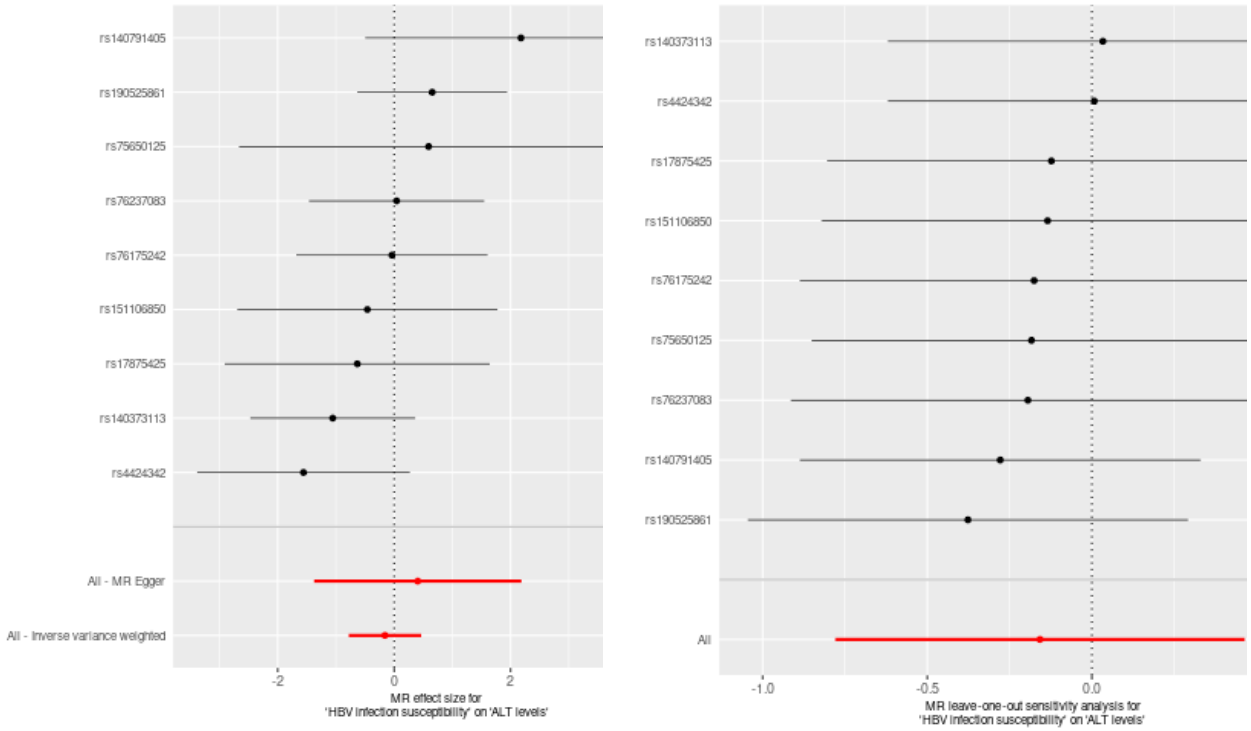


Figure 21: On the left; a forest plot showing the MR estimate and 95% confidence intervals (gray

line segment) for each of the 9 SNPs and also shows the MR Egger and IVW results at the bottom. On the right, a forest plot showing leave one out analysis to evaluate whether any single instrumental variable was driving the causal effect.

CHAPTER FIVE: DISCUSSION

5.1 Discussion

This study investigated genetic determinants of differential susceptibility of hepatitis B infection among participants in a general population cohort in Kyamulibwa subcounty in Kalungu district, southwestern Uganda. It aimed at identifying SNPs that were associated with differential susceptibility to Hepatitis B infection among Ugandans to discover new risk factors unique to Ugandans and additionally address the underrepresentation of Africans in genetic studies. The estimated prevalence of hepatitis B infection in this study was 2.7% and this indicated a small number of individuals infected thereby confirming the reports by the Ministry of Health. The report generated by the Ministry of Health in Uganda in 2019 reported variations in prevalence of hepatitis B across regions with the southwestern region having the lowest prevalence (Ministry of Health, 2019). Males had a higher estimated prevalence of Hepatitis B infection confirming findings from previous studies indicating that females were less susceptible to viral infections as compared to males because they produce more effective and prolonged innate, humoral and cell-mediated immune responses (Giefing-Kroll, Berger, Lepperdinger, & Grubeck-Loebenstein, 2015; Ruggieri, Anticoli, D'Ambrosio, Giordani, & Viora, 2016)

The GWAS analysis done in this study revealed 78 significant SNPs in 22 genomic risk loci which were associated with hepatitis B infection susceptibility. None of these SNPs had previously been reported to be associated with any trait on the GWAS catalog. This seemed to reinforce the fact that the inclusion of diverse African populations in genetic studies could unveil new risk factors and give an in-depth understanding of diseases (Gurdasani, Barroso, et al., 2019). None of the SNPs previously reported to be associated with Hepatitis B infection were identified in this study and this could be a result of a small case sample size (175 cases; 6231 controls) and probably population specific differences. It is therefore plausible that with a larger sample size with a balanced number of cases and controls, there is potential to identify more independent genomic risk loci and significant SNPs associated with hepatitis B infection susceptibility in an African ancestry population.

The most significant SNPs on given chromosomes were mapped to their nearest genes *POGK*, *MIXL1, RP11-24121.1, RP1-97D16.1, ADRB1,* and *RBFOX1*. The functions of proteins encoded by the Pogo transposable element derived with the KRAB (Kruppel-associated box) domain (*POGK*) gene are not well established (W. Xu et al., 2022). However, *POGK* has been previously associated

with liver cancer and its high expression confirmed to predict poor prognosis in patients with hepatocellular carcinoma (W. Xu et al., 2022). Some of the SNPs identified in this study could have mapped to *POGK* as the nearest gene because hepatitis B virus infection together with hepatitis C virus have been reported as the primary carcinogenic infectious agents of hepatocellular carcinoma (de Martel, Georges, Bray, Ferlay, & Clifford, 2020).

The Mix paired-like homeobox (*MIXL1)* is reported to take part in signal transduction pathways and innate immune responses and has previously been associated with hepatitis C viral clearance (Y. Chen et al., 2016). The RNA binding protein fox-1, *RBFOX1* has been previously included among the recurrent targeted genes for hepatitis B integration and reported to exhibit decreased expression in hepatitis B integration events in tumor-adjacent tissues among patients with liver cancer (Ding et al., 2012). Identification of SNPs in this study that mapped to *RBFOX1* confirmed that changes in given genes such as this one that are used by viral DNA for integration into host chromosomes could be one of the plausible mechanisms for differential susceptibility to infections among individuals.
TMEM236 is a transmembrane protein that may have a role in the viral binding of hepatitis B virus into the host cells and it has been previously associated with a unit decrease in Aspartate aminotransferase levels (Sinnott-Armstrong et al., 2021) which is a known liver function biomarker. The qq plot in figure 11 had a number of points reasonably plotted along the slope (null hypothesis) except for some outliers and it also did not show an early departure of the observed from the expected p-values. This seems to suggest that there is low evidence of population substructure and therefore suggestive of little or no systematic differences in allele frequencies due to population stratification (Turner, 2018)

The results identified in the fine mapping show potential for identifying causal variants of hepatitis B infection susceptibility. However, most of the SNPs with small 99% credible set size were intergenic and quite distant from their nearest genes while a few were intronic. None of the lead SNPs was nearest to an exonic position and this is probably due to a small case sample size. Fortunately, some of the SNPs were nearest to genes *POGK*, *MIXL1*, and *ADRB1)* which have a biological and plausible role in the hepatitis B infection lifecycle, and these if studied in a larger sample size could inform what could be the specific causal variants driving the associations noted in differential susceptibility to hepatitis B infection.

Replication of association findings in two independent cohorts of European ancestry focusing on hepatitis B surface antigen seropositivity as traits of interest was unsuccessful. This was probably

because two of the replication criteria had not been met while selecting the cohorts of replication; one suggested the use of replication data sets from similar populations as the discovery and another emphasized a larger sample size of the replication datasets as compared to discovery datasets in order to account for the winner's curse problem (Chanock et al., 2007). It was also probably because of population specific differences such as allele frequency and effect size differences and also the low LD that is exhibited by the highly genetically diverse Africans as compared to the Europeans (Tishkoff et al., 2009).

This study explored the causal effect of hepatitis B infection susceptibility on serum levels of liver function biomarkers AST and ALT using two sample MR analysis. The results showed no significant causal effect of hepatitis B infection susceptibility on AST and ALT serum levels. Sensitivity analysis demonstrated no evidence of horizontal pleiotropy and heterogeneity while the leave-one-out analysis suggested the robustness of results obtained.

The obtained results show potential for the determination of the causal effect of hepatitis B infection susceptibility on liver function biomarkers ALT and AST if more instrumental variables are obtained from GWAS studies with bigger case sample sizes of hepatitis B infection susceptibility. An observational study evaluating hepatitis B infection as a precursor of alteration of hepatic enzymes and compounds among antenatal patients revealed that it caused changes in their levels but this was highly dependent on the stage of infection and influenced by other factors such as alcohol consumption (Abulude, Ahmed, & Sadisu, 2017). In the past ALT levels were used to make treatment decisions for patients chronically infected with hepatitis B, however, this was ruled out since some patients with high hepatitis B DNA levels were found to have normal ALT levels thereby being denied proper treatment (Sarin & Kumar, 2008). As of July 2023, a few studies had been conducted using Mendelian randomization approaches on hepatitis B infection, and of which all these studies are conducted in non-African populations. Therefore, this study is probably the first to assess the causal effect of hepatitis B infection susceptibility on liver function biomarkers in an African ancestry population using Mendelian randomization.

CHAPTER SIX: CONCLUSION, LIMITATIONS & RECOMMENDATIONS

6.1 Conclusions

The study identified 78 genome-wide significant SNPs in 22 genomic loci and genes *POGK*, *MIXL1*, *ADRB1*, *RBFOX,1* and *TMEM236* that are associated with hepatitis B infection susceptibility. We could not replicate the genetic association findings obtained in our study due to the unavailability of publicly available data on African populations looking at our similar trait of interest. This failure to validate association findings reinforces the need for including African populations in genetic studies to support replication steps and other post-GWAS analysis that require publicly available data.

This was probably the first study to perform causal inference of hepatitis B infection susceptibility on ALT and AST levels and it suggested that hepatitis B infection susceptibility is not causally associated with alterations in levels of liver function biomarkers ALT and AST. These results show a high potential for the detection of novel genetic risk loci associated with hepatitis B infection susceptibility and a deeper understanding of underlying biological mechanisms involved among individuals of African Ancestry given that there were significant findings noted even with a small case sample size.

6.2 Limitations of the study

The study had the following limitations, first, hepatitis B infection susceptibility definition was not well done due to the inability to perform further tests to further characterize whether the infection was current or past among the infected individuals. This is because the samples available were not adequate for tests required to be performed. Second, the study focused on finding common SNPs and left out rare SNPs since genome wide association studies generally do not identify rare variants (Tam et al., 2019). Third, the study had a very small case sample size as compared to the controls and this probably inhibited the detection of causal SNPs associated with hepatitis B infection susceptibility. Fourth, we could not compare and validate our findings with other studies. This is because there are few genome-wide association studies focusing on hepatitis B that have been done probably because it is an infectious disease with none done in an African population.

6.3 Recommendations

There is need to conduct this study with a better-defined phenotype and a larger sample size where the numbers of cases and controls are well balanced to achieve very high statistical power

REFERENCES

Abulude, O. A., Ahmed, I., & Sadisu, F. U. (2017). Assessment of Hepatitis B Viral Infection as a Predictor of Hepatic Enzymes and Compounds Alteration among Antenatal Patients. *Med Sci (Basel), 5*(4). doi:10.3390/medsci5040024

Akcay, I. M., Katrinli, S., Ozdil, K., Doganay, G. D., & Doganay, L. (2018). Host genetic factors affecting hepatitis B infection outcomes: Insights from genome-wide association studies. *World J Gastroenterol, 24*(30), 3347-3360. doi:10.3748/wjg.v24.i30.3347

Asiki, G., Murphy, G., Nakiyingi-Miiro, J., Seeley, J., Nsubuga, R. N., Karabarinde, A., . . . team, G. P. C. (2013). The general population cohort in rural south-western Uganda: a platform for communicable and non-communicable disease studies. *Int J Epidemiol, 42*(1), 129-141. doi:10.1093/ije/dys234

Birney, E. (2022). Mendelian Randomization. *Cold Spring Harb Perspect Med, 12*(4). doi:10.1101/cshperspect.a041302

Bowden, J., & Holmes, M. V. (2019). Meta-analysis and Mendelian randomization: A review. *Res Synth Methods, 10*(4), 486-496. doi:10.1002/jrsm.1346

Britannica, T. E. o. E. (2023). Encyclopedia Britannica: Phenotype. *Encyclopedia Britannica.* doi: https://www.britannica.com/science/phenotype

Broadbent, J. R., Foley, C. N., Grant, A. J., Mason, A. M., Staley, J. R., & Burgess, S. (2020). MendelianRandomization v0.5.0: updates to an R package for performing Mendelian randomization analyses using summarized data. *Wellcome Open Res, 5*, 252. doi:10.12688/wellcomeopenres.16374.2

Burgess, S., Davey Smith, G., Davies, N. M., Dudbridge, F., Gill, D., Glymour, M. M., . . . Theodoratou, E. (2019). Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res, 4*, 186. doi:10.12688/wellcomeopenres.15555.2

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS Comput Biol, 8*(12), e1002822. doi:10.1371/journal.pcbi.1002822

Castaneda, D., Gonzalez, A. J., Alomari, M., Tandon, K., & Zervos, X. B. (2021). From hepatitis A to E: A critical review of viral hepatitis. *World J Gastroenterol, 27*(16), 1691-1715. doi:10.3748/wjg.v27.i16.1691

CDC. (2021a). Hepatitis B, Acute 2012 Case Definition. *Centre for disease control and prevention.* Retrieved from https://ndc.services.cdc.gov/case-definitions/hepatitis-b-acute-2012/

CDC. (2021b). Hepatitis B, Chronic 2012 Case Definition. *Centers for Disease Control and Prevention.* Retrieved from https://ndc.services.cdc.gov/case-definitions/hepatitis-b-chronic-2012/

Chanock, S. J., Studies, N.-N. W. G. o. R. i. A., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D. J., . . . Collins, F. S. (2007). Replicating genotype-phenotype associations. *Nature, 447*(7145), 655-660. doi:10.1038/447655a

Chen, Y., Shen, C., Guha, D., Ding, M., Kulich, S., Ashimkhanova, A., . . . Gupta, P. (2016). Identification of the transcripts associated with spontaneous HCV clearance in individuals co-infected with HIV and HCV. *BMC Infect Dis, 16*(1), 693. doi:10.1186/s12879-016-2044-7

Chen, Z. N., Liu, F. W., Bai, L. P., Jin, C. X., Li, X. T., Yang, W. X., . . . Zhou, T. C. (2022). Mitochondrial DNA haplogroup F confers genetic susceptibility to chronic HBV infection for the Yi nationality in Lijiang, China. *Infect Genet Evol, 101*, 105293. doi:10.1016/j.meegid.2022.105293

Cheung, K. W., & Lao, T. T. (2020). Hepatitis B - Vertical transmission and the prevention of mother-to-child transmission. *Best Pract Res Clin Obstet Gynaecol, 68*, 78-88. doi:10.1016/j.bpobgyn.2020.02.014

ClevelandClinic. (2023). Elevated liver enzymes. Retrieved from https://my.clevelandclinic.org/health/symptoms/17679-elevated-liver-enzymes

Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet, 23*(R1), R89-98. doi:10.1093/hmg/ddu328

de Martel, C., Georges, D., Bray, F., Ferlay, J., & Clifford, G. M. (2020). Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob Health, 8*(2), e180-e190. doi:10.1016/S2214-109X(19)30488-7

Dennis, J. K., Sealock, J. M., Straub, P., Lee, Y. H., Hucks, D., Actkins, K., . . . Davis, L. K. (2021). Clinical laboratory test-wide association scan of polygenic scores identifies biomarkers of complex disease.

Genome Med, 13(1), 6. doi:10.1186/s13073-020-00820-8
Ding, D., Lou, X., Hua, D., Yu, W., Li, L., Wang, J., . . . Lin, B. (2012). Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. *PLoS Genet, 8*(12), e1003065. doi:10.1371/journal.pgen.1003065
Eng-Kiong, T., & Lok, A. S. (2023, Sep 21, 2022). Epidemiology, transmission, and prevention of hepatitis B virus infection. Retrieved from https://www.uptodate.com/contents/epidemiology-transmission-and-prevention-of-hepatitis-b-virus-infection
Fatumo, S., Chikowore, T., Choudhury, A., Ayub, M., Martin, A. R., & Kuchenbaecker, K. (2022). A roadmap to increase diversity in genomic studies. *Nat Med, 28*(2), 243-250. doi:10.1038/s41591-021-01672-4
Fatumo, S., Mugisha, J., Soremekun, O. S., Kalungi, A., Mayanja, R., Kintu, C., . . . Kaleebu, P. (2022). Uganda Genome Resource: A rich research database for genomic studies of communicable and non-communicable diseases in Africa. *Cell Genom, 2*(11), None. doi:10.1016/j.xgen.2022.100209
Frangou, E. (2013). Bayesian methods for fine mapping in genetic association studies.
Ganesan, M., Eikenberry, A., Poluektova, L. Y., Kharbanda, K. K., & Osna, N. A. (2020). Role of alcohol in pathogenesis of hepatitis B virus infection. *World J Gastroenterol, 26*(9), 883-903. doi:10.3748/wjg.v26.i9.883
Gerlich, W. H. (2013). Medical virology of hepatitis B: how it began and where we are now. *Virol J, 10*, 239. doi:10.1186/1743-422X-10-239
Giefing-Kroll, C., Berger, P., Lepperdinger, G., & Grubeck-Loebenstein, B. (2015). How sex and age affect immune responses, susceptibility to infections, and response to vaccination. *Aging Cell, 14*(3), 309-321. doi:10.1111/acel.12326
Gurdasani, D., Barroso, I., Zeggini, E., & Sandhu, M. S. (2019). Genomics of disease risk in globally diverse populations. *Nat Rev Genet, 20*(9), 520-535. doi:10.1038/s41576-019-0144-0
Gurdasani, D., Carstensen, T., Fatumo, S., Chen, G., Franklin, C. S., Prado-Martinez, J., . . . Sandhu, M. S. (2019). Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. *Cell, 179*(4), 984-1002 e1036. doi:10.1016/j.cell.2019.10.004
Hahne, S. J., Veldhuijzen, I. K., Wiessing, L., Lim, T. A., Salminen, M., & Laar, M. (2013). Infection with hepatitis B and C virus in Europe: a systematic review of prevalence and cost-effectiveness of screening. *BMC Infect Dis, 13*, 181. doi:10.1186/1471-2334-13-181
Hardy, J., & Singleton, A. (2009). Genomewide association studies and human disease. *N Engl J Med, 360*(17), 1759-1768. doi:10.1056/NEJMra0808700
Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., . . . Haycock, P. C. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *Elife, 7*. doi:10.7554/eLife.34408
Henn, B. M., Cavalli-Sforza, L. L., & Feldman, M. W. (2012). The great human expansion. *Proc Natl Acad Sci U S A, 109*(44), 17758-17764. doi:10.1073/pnas.1212380109
Hill, A. V. (1998). The immunogenetics of human infectious diseases. *Annu Rev Immunol, 16*, 593-617. doi:10.1146/annurev.immunol.16.1.593
Hill, A. V. (2006). Aspects of genetic susceptibility to human infectious diseases. *Annu Rev Genet, 40*, 469-486. doi:10.1146/annurev.genet.40.110405.090546
Hu, Z., Liu, Y., Zhai, X., Dai, J., Jin, G., Wang, L., . . . Shen, H. (2013). New loci associated with chronic hepatitis B virus infection in Han Chinese. *Nat Genet, 45*(12), 1499-1503. doi:10.1038/ng.2809
Iida-Ueno, A., Enomoto, M., Tamori, A., & Kawada, N. (2017). Hepatitis B virus infection and alcohol consumption. *World J Gastroenterol, 23*(15), 2651-2659. doi:10.3748/wjg.v23.i15.2651
Kalyesubula, R., Hau, J. P., Asiki, G., Ssebunya, B., Kusemererwa, S., Seeley, J., . . . Newton, R. (2018). Impaired renal function in a rural Ugandan population cohort. *Wellcome Open Res, 3*, 149. doi:10.12688/wellcomeopenres.14863.3
Kamiza, A. B., Fatumo, S., Singini, M. G., Yeh, C. C., & Chikowore, T. (2022). Hepatitis B infection is causally associated with extrahepatic cancers: A Mendelian randomization study. *EBioMedicine, 79*, 104003. doi:10.1016/j.ebiom.2022.104003
Kamiza, A. B., Toure, S. M., Vujkovic, M., Machipisa, T., Soremekun, O. S., Kintu, C., . . . Fatumo, S. (2022). Transferability of genetic risk scores in African populations. *Nat Med, 28*(6), 1163-1166. doi:10.1038/s41591-022-01835-x

Kim, T. H., Lee, E. J., Choi, J. H., Yim, S. Y., Lee, S., Kang, J., . . . Um, S. H. (2018). Identification of novel susceptibility loci associated with hepatitis B surface antigen seroclearance in chronic hepatitis B. *PLoS One, 13*(7), e0199094. doi:10.1371/journal.pone.0199094

Kraft, P., Zeggini, E., & Ioannidis, J. P. (2009). Replication in genome-wide association studies. *Stat Sci, 24*(4), 561-573. doi:10.1214/09-STS290

Lanini, S., Ustianowski, A., Pisapia, R., Zumla, A., & Ippolito, G. (2019). Viral Hepatitis: Etiology, Epidemiology, Transmission, Diagnostics, Treatment, and Prevention. *Infect Dis Clin North Am, 33*(4), 1045-1062. doi:10.1016/j.idc.2019.08.004

Lee, Y. H. (2022). Overview of Mendelian Randomization Analysis. *Journal of Rheumatic Diseases*. Retrieved from https://www.jrd.or.kr/journal/view.html?doi=10.4078/jrd.2020.27.4.241

Li, Y., Si, L., Zhai, Y., Hu, Y., Hu, Z., Bei, J. X., . . . Zhou, G. (2016). Genome-wide association study identifies 8p21.3 associated with persistent hepatitis B virus infection among Chinese. *Nat Commun, 7*, 11664. doi:10.1038/ncomms11664

Li, Z., Hou, X., & Cao, G. (2015). Is mother-to-infant transmission the most important factor for persistent HBV infection? *Emerg Microbes Infect, 4*(5), e30. doi:10.1038/emi.2015.30

Lin, J. R., Jaroslawicz, D., Cai, Y., Zhang, Q., Wang, Z., & Zhang, Z. D. (2018). PGA: post-GWAS analysis for disease gene identification. *Bioinformatics, 34*(10), 1786-1788. doi:10.1093/bioinformatics/btx845

Liu, Z., Song, L., Chen, J., Zhou, Y., Wang, Y., Tang, L., & Li, Y. (2023). Causal associations between chronic hepatitis B and COVID-19 in East Asian populations. *Virol J, 20*(1), 109. doi:10.1186/s12985-023-02081-4

Loomba, R., Yang, H. I., Su, J., Brenner, D., Barrett-Connor, E., Iloeje, U., & Chen, C. J. (2013). Synergism between obesity and alcohol in increasing the risk of hepatocellular carcinoma: a prospective cohort study. *Am J Epidemiol, 177*(4), 333-342. doi:10.1093/aje/kws252

Marigorta, U. M., Rodriguez, J. A., Gibson, G., & Navarro, A. (2018). Replicability and Prediction: Lessons and Challenges from GWAS. *Trends Genet, 34*(7), 504-517. doi:10.1016/j.tig.2018.03.005

Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J. A., Ziyatdinov, A., . . . Marchini, J. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet, 53*(7), 1097-1103. doi:10.1038/s41588-021-00870-7

Mboowa, G., Sserwadda, I., Amujal, M., & Namatovu, N. (2018). Human Genomic Loci Important in Common Infectious Diseases: Role of High-Throughput Sequencing and Genome-Wide Association Studies. *Can J Infect Dis Med Microbiol, 2018*, 1875217. doi:10.1155/2018/1875217

Ministry of Health, U. (2019). Uganda Guidelines for Prevention , Testing, Care and Treatment of Hepatitis B and C Virus Infection. Retrieved from http://library.health.go.ug/publications/hepatitis- b/uganda-guidelines-prevention-testing-care-and-treatment-hepatitis-b-and-c

Mozzi, A., Pontremoli, C., & Sironi, M. (2018). Genetic susceptibility to infectious diseases: Current status and future perspectives from genome-wide approaches. *Infect Genet Evol, 66*, 286-307. doi:10.1016/j.meegid.2017.09.028

NICE, C. G. (2017). *Hepatitis B (chronic): diagnosis and management*. London: National Institute for Health and Care.

Pardee, M. (2019). Diagnosis and Management of Hepatitis B and C. *Nurs Clin North Am, 54*(2), 277-284. doi:10.1016/j.cnur.2019.02.004

Peprah, E., Xu, H., Tekola-Ayele, F., & Royal, C. D. (2015). Genome-wide association studies in Africans and African Americans: expanding the framework of the genomics of human traits and disease. *Public Health Genomics, 18*(1), 40-51. doi:10.1159/000367962

Pereira, L., Mutesa, L., Tindana, P., & Ramsay, M. (2021). African genetic diversity and adaptation inform a precision medicine agenda. *Nat Rev Genet, 22*(5), 284-306. doi:10.1038/s41576-020-00306-8

Pisano, M. B., Giadans, C. G., Flichman, D. M., Re, V. E., Preciado, M. V., & Valva, P. (2021). Viral hepatitis update: Progress and perspectives. *World J Gastroenterol, 27*(26), 4018-4044. doi:10.3748/wjg.v27.i26.4018

Platt, L., Easterbrook, P., Gower, E., McDonald, B., Sabin, K., McGowan, C., . . . Vickerman, P. (2016). Prevalence and burden of HCV co-infection in people living with HIV: a global systematic review and meta-analysis. *Lancet Infect Dis, 16*(7), 797-808. doi:10.1016/S1473-3099(15)00485-5

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 Replicated Findings From

Behavioral Genetics. *Perspect Psychol Sci, 11*(1), 3-23. doi:10.1177/1745691615617439
Polaris Observatory, C. (2018). Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study. *Lancet Gastroenterol Hepatol, 3*(6), 383-403. doi:10.1016/S2468-1253(18)30056-6
Qiu, B., Jiang, W., Olyaee, M., Shimura, K., Miyakawa, A., Hu, H., . . . Tang, L. (2017). Advances in the genome-wide association study of chronic hepatitis B susceptibility in Asian population. *Eur J Med Res, 22*(1), 55. doi:10.1186/s40001-017-0288-3
Rafael, T. M., & Magnolia, A. C. (2021). Genetic variation- The Source of Biological Diversity. In M. d. A. C. Rafael Trindade Maia (Ed.), *Genetic Variation* (pp. 296). Federal University of Campina: Open Access.
Romero-Gomez, M., Eslam, M., Ruiz, A., & Maraver, M. (2011). Genes and hepatitis C: susceptibility, fibrosis progression and response to treatment. *Liver Int, 31*(4), 443-460. doi:10.1111/j.1478-3231.2011.02449.x
Ruggieri, A., Anticoli, S., D'Ambrosio, A., Giordani, L., & Viora, M. (2016). The influence of sex and gender on immunity, infection and vaccination. *Ann Ist Super Sanita, 52*(2), 198-204. doi:10.4415/ANN_16_02_11
Ruggieri, A., Gagliardi, M. C., & Anticoli, S. (2018). Sex-Dependent Outcome of Hepatitis B and C Viruses Infections: Synergy of Sex Hormones and Immune Responses? *Front Immunol, 9*, 2302. doi:10.3389/fimmu.2018.02302
Sabeena, S., & Ravishankar, N. (2022). Horizontal Modes of Transmission of Hepatitis B Virus (HBV): A Systematic Review and Meta-Analysis. *Iran J Public Health, 51*(10), 2181-2193. doi:10.18502/ijph.v51i10.10977
Sagnelli, E., Stroffolini, T., Mele, A., Imparato, M., Sagnelli, C., Coppola, N., & Almasio, P. L. (2012). Impact of comorbidities on the severity of chronic hepatitis B at presentation. *World J Gastroenterol, 18*(14), 1616-1621. doi:10.3748/wjg.v18.i14.1616
Sarin, S. K., & Kumar, M. (2008). Should chronic HBV infected patients with normal ALT treated: debate. *Hepatol Int, 2*(2), 179-184. doi:10.1007/s12072-008-9065-2
Scepanovic, P., Alanio, C., Hammer, C., Hodel, F., Bergstedt, J., Patin, E., . . . Milieu Interieur, C. (2018). Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. *Genome Med, 10*(1), 59. doi:10.1186/s13073-018-0568-8
Schadler, S., & Hildt, E. (2009). HBV life cycle: entry and morphogenesis. *Viruses, 1*(2), 185-209. doi:10.3390/v1020185
Schaid, D. J., Chen, W., & Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet, 19*(8), 491-504. doi:10.1038/s41576-018-0016-z
Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L., . . . Boehnke, M. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science, 316*(5829), 1341-1345. doi:10.1126/science.1142382
Sekula, P., Del Greco, M. F., Pattaro, C., & Kottgen, A. (2016). Mendelian Randomization as an Approach to Assess Causality Using Observational Data. *J Am Soc Nephrol, 27*(11), 3253-3265. doi:10.1681/ASN.2016010098
Shih, C., Yang, C. C., Choijilsuren, G., Chang, C. H., & Liou, A. T. (2018). Hepatitis B Virus. *Trends Microbiol, 26*(4), 386-387. doi:10.1016/j.tim.2018.01.009
Sinnott-Armstrong, N., Tanigawa, Y., Amar, D., Mars, N., Benner, C., Aguirre, M., . . . Rivas, M. A. (2021). Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat Genet, 53*(2), 185-194. doi:10.1038/s41588-020-00757-z
Sonderup, M. W., Afihene, M., Ally, R., Apica, B., Awuku, Y., Cunha, L., . . . Hepatology Association of sub-Saharan, A. (2017). Hepatitis C in sub-Saharan Africa: the current status and recommendations for achieving elimination by 2030. *Lancet Gastroenterol Hepatol, 2*(12), 910-919. doi:10.1016/S2468-1253(17)30249-2
Stephens, M., & Balding, D. J. (2009). Bayesian statistical methods for genetic association studies. *Nat Rev Genet, 10*(10), 681-690. doi:10.1038/nrg2615
Takata, K., Yamauchi, E., Shakado, S., Uehara, Y., Fukuda, H., Yamauchi, R., . . . Hirai, F. (2020). Horizontal Transmission of Hepatitis B Virus Genotype C Among Members of a Wrestling Club in Japan. *Am J Case Rep, 21*, e925044. doi:10.12659/AJCR.925044

Tam, V., Patel, N., Turcotte, M., Bosse, Y., Pare, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat Rev Genet, 20*(8), 467-484. doi:10.1038/s41576-019-0127-1
Tenny, S., Kerndt, C. C., & Hoffman, M. R. (2023). Case Control Studies. In *StatPearls*. Treasure Island (FL) ineligible companies. Disclosure: Connor Kerndt declares no relevant financial relationships with ineligible companies. Disclosure: Mary Hoffman declares no relevant financial relationships with ineligible companies.
Terrault, N. A., Lok, A. S. F., McMahon, B. J., Chang, K. M., Hwang, J. P., Jonas, M. M., . . . Wong, J. B. (2018). Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. *Hepatology, 67*(4), 1560-1599. doi:10.1002/hep.29800
Than, T. T., Jo, E., Todt, D., Nguyen, P. H., Steinmann, J., Steinmann, E., & Windisch, M. P. (2019). High Environmental Stability of Hepatitis B Virus and Inactivation Requirements for Chemical Biocides. *J Infect Dis, 219*(7), 1044-1048. doi:10.1093/infdis/jiy620
Thomas, R., Thio, C. L., Apps, R., Qi, Y., Gao, X., Marti, D., . . . Carrington, M. (2012). A novel variant marking HLA-DP expression levels predicts recovery from hepatitis B virus infection. *J Virol, 86*(12), 6979-6985. doi:10.1128/JVI.00406-12
Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., . . . Williams, S. M. (2009). The genetic structure and history of Africans and African Americans. *Science, 324*(5930), 1035-1044. doi:10.1126/science.1172257
Tong, H., Bock, C. T., & Velavan, T. P. (2014). Genetic insights on host and hepatitis B virus in liver diseases. *Mutat Res Rev Mutat Res, 762*, 65-75. doi:10.1016/j.mrrev.2014.06.001
Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *The Journal Of Open Source Software*. doi:https://doi.org/10.21105/joss.00731
Vaillant, A. (2021). Transaminase Elevations during Treatment of Chronic Hepatitis B Infection: Safety Considerations and Role in Achieving Functional Cure. *Viruses, 13*(5). doi:10.3390/v13050745
van de Bunt, M., Cortes, A., Consortium, I., Brown, M. A., Morris, A. P., & McCarthy, M. I. (2015). Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci. *PLoS Genet, 11*(9), e1005535. doi:10.1371/journal.pgen.1005535
Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nat Commun, 8*(1), 1826. doi:10.1038/s41467-017-01261-5
Watanabe, K., Umicevic Mirkov, M., de Leeuw, C. A., van den Heuvel, M. P., & Posthuma, D. (2019). Genetic mapping of cell type specificity for complex traits. *Nat Commun, 10*(1), 3222. doi:10.1038/s41467-019-11181-1
Wellcome Trust Case Control, C., Maller, J. B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., . . . Donnelly, P. (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet, 44*(12), 1294-1301. doi:10.1038/ng.2435
WHO. (2017). Hepatitis B vaccines: WHO position paper - July 2017. *Wkly Epidemiol Rec, 92*(27), 369-392. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28685564
WHO. (2021). Global Progress Report on HIV, Viral Hepatitis and Sexually Transmitted Infections, 2021. . Retrieved from https://www.who.int/publications/i/item/9789240027077
WHO. (2022a). Hepatitis. Retrieved from https://www.who.int/health-topics/hepatitis#tab=tab_1
WHO. (2022b). Hepatitis: Testing & Diagnostics. Retrieved from https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hepatitis/testing-and-diagnostics
Wilkins, T., Sams, R., & Carpenter, M. (2019). Hepatitis B: Screening, Prevention, Diagnosis, and Treatment. *Am Fam Physician, 99*(5), 314-323. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/30811163
Xia, Y., & Protzer, U. (2017). Control of Hepatitis B Virus by Cytokines. *Viruses, 9*(1). doi:10.3390/v9010018
Xiang, X., Guo, Y., Yang, L., Ge, Q., Mijit, S., & Xu, F. (2016). Association of human leukocyte antigen DP/DQ gene polymorphisms with chronic hepatitis B in Chinese Han and Uygur populations. *Infect Genet Evol, 43*, 407-411. doi:10.1016/j.meegid.2016.06.022
Xu, J., Zhan, Q., Fan, Y., Yu, Y., & Zeng, Z. (2021). Human genetic susceptibility to hepatitis B virus infection. *Infect Genet Evol, 87*, 104663. doi:10.1016/j.meegid.2020.104663
Xu, W., Huang, Y., Mei, Y., Zhang, Y., Luo, Q., Zhu, S., . . . Li, J. (2022). High Expression of POGK Predicts Poor Prognosis in Patients with Hepatocellular Carcinoma. *Curr Oncol, 29*(11), 8650-8667. doi:10.3390/curroncol29110682

Zhang, Z., Wang, C., Liu, Z., Zou, G., Li, J., & Lu, M. (2019). Host Genetic Determinants of Hepatitis B Virus Infection. *Front Genet, 10*, 696. doi:10.3389/fgene.2019.00696

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat Genet, 44*(7), 821-824. doi:10.1038/ng.2310

Zoulim, F., & Durantel, D. (2015). Antiviral therapies and prospects for a cure of chronic hepatitis B. *Cold Spring Harb Perspect Med, 5*(4). doi:10.1101/cshperspect.a021501

APPENDICES

Appendix I: Participant Informed Consent Document

VNO

The Kyamulibwa Cohort Study Round 22

HNO

Adult and emancipated minor consent: survey and blood samples

MRC Medical Research Council | UNIVERSITY OF CAMBRIDGE | wellcome trust sanger institute

*To be read out to potential participants **aged 18 years or older** and **emancipated minors**:*

1. Why is this study being done?

The Medical Research Council Programme on AIDS in Uganda (MRC Unit), with the Uganda Virus Research Institute (UVRI) and Ugandan Ministry of Health, has been conducting household annual surveys and collecting blood samples from residents of Kyamulibwa subcounty since 1989. The work has mainly been on the medical and social research of HIV/AIDS and sexually transmitted diseases. These surveys have helped the MRC Unit and the Ugandan Government to find ways to improve health services for people with HIV and other infectious diseases.

Now, the MRC Unit with researchers in the UK (the University of Cambridge and the Wellcome Trust Sanger Institute) and the Ugandan Ministry of Health would also like to know how common are obesity, diabetes, anaemias, high cholesterol and high blood pressure, and other related diseases, in Kyamulibwa subcounty. This information will help the MRC Unit and the Ugandan Government to plan better health services to treat and prevent these diseases.

2. Why have I been chosen to take part in the study?

Your household has been selected because you live in Kyamulibwa subcounty, Kalungu District. We will be asking 18,000 people, including children aged 13-17, from 25 villages in Kyamulibwa subcounty to participate in the study. You may be a regular participant in this study, or this may be your first time, so please use this opportunity to familiarise yourself with the study, and ask any questions you have.

3. Why are blood samples being collected in this study?

HIV is one of the leading causes of death in Uganda. HIV and other infectious diseases can be found in the blood. We can also look for diabetes, anaemias and high cholesterol in the blood. We want to know how many people have these and other related diseases. We also want to store some of the blood to test for other diseases in the future. We will also see whether genetic differences (inherited differences between families because of ancestry) have an effect on these diseases.

4. What will happen to me if I decide to take part in the study?

You can choose to take part in the study or not. It is your choice. If you choose not to take part in the study, there is no problem. If you choose to take part, we will ask you some questions about your health and lifestyle (for example, your diet, tobacco and alcohol consumption, physical activity, sexual behaviour and medical history).

UGANDA VIRUS RESEARCH INSTITUTE SCIENCE & ETHICS COMMITTEE 12 AUG 2011 P.O. BOX 49 ENTEBBE

Version 3: June 2011 Page 1 of 7

We will measure how much you weigh and how tall you are. We will measure your waist and hip circumference. We will record your blood pressure and pulse. The survey staff will explain what your results mean. They will advise you on whether further testing is needed at the MRC study clinic in Kyamulibwa. At the MRC study clinic, the clinic staff may recommend changes to your lifestyle or treatment to help keep you and your family healthy.

After we have taken these measurements, we will also ask you to give some blood samples. You can choose to give blood or not. If you agree, we would like to draw a little blood (16.5 ml) from a vein in your arm. We will fill the blood into three small tubes. We would prefer to have venous blood for our study. We use only new sterile needles to collect blood.

The blood samples are sent to our laboratories in Entebbe for testing. We will test for infectious diseases like HIV, hepatitis B and hepatitis C. All information you provide and the results from the laboratory testing are kept confidential. We will also look for possible diabetes, anaemias, high cholesterol and to see how well your liver works. Your blood results will be ready in four weeks.

Your results will be given to you if you wish to know them. Your HIV test result will be available from your local HIV counselling centre. The other blood results will be available through the MRC study clinic in Kyamulibwa. You can choose to make an appointment at the clinic to get these results. The clinic staff will explain what your results mean. They will advise you on whether further testing is needed. They may recommend changes to your lifestyle or treatment to help keep you and your family healthy. All results will remain available in the MRC study clinic for one year.

If you do not collect your results, we would like to contact you with your results. We will only do this if your results are abnormal. We will not contact you if your results are normal. You will need to indicate on your consent form whether you want us to contact you if your results are abnormal.

We will also examine your blood for genetic differences (differences due to ancestry) that may have an effect on your health. We will search for these differences in the chemicals (DNA) in your blood at specialised laboratories in the UK. It is necessary to send blood to the UK because the equipment for the genetic analysis of the blood is not available in Uganda. Because the long term effects of these genetic differences are not known, we will not provide these results to you.

If you agree, we will keep any leftover blood in a secure archive in Uganda and the UK. We may use it for other studies related to health or diseases. These studies will be done in Uganda or at specialised laboratories in the UK, the USA and other countries. Storage could be for many years. We will not contact you with results from future studies as the work is for research purposes and not for making diagnoses. However, if you allow your blood to be used, we may be able to find out things that will help improve health service provision in Uganda.

With your permission, we may use your information and blood samples collected in previous surveys, to see if your results have changed over time. If you agree, we may also look at your medical records at the MRC study clinic in Kyamulibwa, to help us understand your results better.

5. Can I decline to be in the study or decline to give blood?

It is up to you whether or not to take part in the study. You can stop being in the study or decide not to give blood at any time you want without giving a reason. This will not affect the standard medical care that you receive from the state.

6. Are there any benefits to me from being in the study?

If you go to the MRC study clinic to discuss your results, you will receive lifestyle advice or treatment that may improve your health. The information about you will help the MRC Unit and the Ugandan Government to plan better health services to treat and prevent these diseases.

7. What risks can I expect from being in the study?

If you take part in the study, the risks to you are very slight. Most of the study questions are general in nature, but there are some that are personal and may make you feel uncomfortable. You are free to refuse to answer any questions. However, in order to obtain good results from the study, it is important that you attempt to answer all questions if possible. If you give blood, the risks to you are small. You may get some slight bruising where the blood is taken from your arm. If you have any discomfort, bleeding or swelling at the site, please contact our study staff or your health worker. All research team members are trained to protect your privacy and all information you share will be kept secret.

8. How will the information and blood I give in the study be kept private?

Everything we talk about will be kept secret to the extent allowed by the law. Your results will be kept secret to the extent allowed by the law. To protect your privacy, we will use a code number to identify you and all information about you, including your blood samples. We will put a study number, not your name, on the blood tubes. We will keep these records and samples securely locked. For the current study, only research team members in Uganda and the UK will be able to look at the records and samples. Your name or any other facts that might point to you will not appear when we present this study or publish its results.

Your records and samples will then be securely archived in Uganda and the UK for future research studies. This future research may be done by us or by other research teams working in Uganda, the UK, the USA or other countries. Only your information, blood samples and results which do not have your name or identifiable information will be stored and shared with other research teams.

9. Whom can I ask if I have questions about the study?

We would like to answer all your questions. If you have any questions now, please ask us. If you have any questions in the future, there are other people that you can contact:

- Dr. Anatoli Kamali, Principal Investigator, MRC/UVRI Uganda Research Unit on AIDS, P.O. Box 49, Entebbe. Telephone: 0417 704000.
- Dr. Gershim Asiki, Project Leader, MRC/UVRI Uganda Research Unit on AIDS, P.O. BOX 49, Entebbe. Telephone: 0392 720042 122.

If you have any concerns about your rights in this study, please contact Mr. Tom Lutalo, Chairman of the UVRI Science and Ethics Committee, UVRI, P.O. Box 49, Entebbe. Telephone: 0414 321962.

UGANDA VIRUS RESEARCH INSTITUTE ENTEBBE SCIENCE ETHICS COMMITTEE 12 AUG 2011 P.O. BOX 49 ENTEBBE (U)

10. Your consent

If you are willing to take part in the study, we would like you to complete some consent forms to confirm your agreement. You will need to mark your initial or thumb print against each of the statements on the consent forms. You will also need to write your full name at the bottom of each form, then sign (or give a thumb print) and date them. We will also sign and date the forms. Then you will have a copy to keep and we will keep a copy for the study records.

VNO

The Kyamulibwa Cohort Study Round 22

HNO

Adult and emancipated minor consent: survey and blood samples

Medical Research Council | UNIVERSITY OF CAMBRIDGE | wellcome trust sanger institute

Enter participant ID number

Consent for Study Survey

Please initial or thumbprint boxes

1	I understand the information sheet I have read, or have had read to me, for this study and have been given a copy to keep. I have had the opportunity to ask questions and understand why the study is being done.		cons1
2	I agree to be in this study and to be asked questions about my health and lifestyle.		cons2
3	I agree to have my height, weight, hip and waist circumference, pulse and blood pressure measured for this study.		cons3
4	I give permission for responsible individuals from the study teams in Uganda and the UK to look at the information gathered about me in previous surveys, and to access my medical records at the MRC study clinic in Kyamulibwa, where it is relevant to this study.		cons8
5	I understand that the information gathered about me will be kept confidential, and stored in a secure archive in Uganda and the UK. I also understand that any research analysis or publication outcome will be based on anonymised data, that is, without my name, identifiable information or contact details attached.		cons4
6	I understand that my participation in this study is voluntary and that I am free to withdraw from the study at any time, without giving a reason and without my standard medical care provided by the state or legal rights being affected.		cons7
7	I agree that the information gathered about me can be stored anonymously for future studies in a secure archive. I understand that future studies using my information may be undertaken by other research teams in Uganda, the UK, the USA and other countries.		cons5
8	I understand that the results of these future studies are unlikely to have any implications for me personally. I also understand that I will not benefit financially if this research leads to the development of a new treatment or medical test.		cons6

Name of participant (BLOCK CAPITALS)	**Date**	**Signature**
Witness to consent	**Date**	**Signature**
Name of person taking consent	**Date**	**Signature**

1 copy for participant; 1 to be kept by research team

UGANDA VIRUS RESEARCH INSTITUTE ETHICS COM 12 AUG 2011 P.O.BOX 49 ENTEBBE (U)

VNO

The Kyamulibwa Cohort Study Round 22

HNO

Adult and emancipated minor consent: survey and blood samples

Enter participant ID number

Consent for Current Use of Blood Samples

		Please initial or thumbprint boxes	
1	I understand the information sheet I have read, or have had read to me, for this study and have been given a copy to keep. I have had the opportunity to ask questions and understand why the study is being done.		conb1
2	I agree to give a sample of blood for research in the study.		conb2
3	I give permission for responsible individuals from the study teams in Uganda and the UK to look at the information gathered about me in previous surveys, and to access my medical records at the MRC study clinic in Kyamulibwa, where it is relevant to this study.		conb7
4	I understand that the information gathered about me and the blood samples I have given will be kept confidential, and stored in a secure archive in Uganda and the UK. I also understand that any research analysis or publication outcome will be based on anonymised data, that is, without my name, identifiable information or contact details attached.		conb3
5	I agree to serological and biochemical analyses of my blood samples to provide information on infectious diseases (HIV, hepatitis B, hepatitis C), metabolic diseases (diabetes, high cholesterol, liver dysfunction) and anaemias.		conb4
6	I agree to genetic analyses, including DNA sequencing, of my blood samples to be undertaken in the UK, information from which will not be relayed to me.		conb5
7	I understand that my participation in this study is voluntary and that I am free to withdraw from the study at any time, without giving a reason and without my standard medical care provided by the state or legal rights being affected.		conb6

Disclosure of blood test results – please initial or thumbprint one box for each test to indicate whether or not you wish to be informed of your results if they are abnormal. You will NOT be contacted if your results are normal, but you can choose to collect these results yourself from the MRC clinic.

	YES	*NO*
I would like to be informed of my HIV result		
I would like to be informed of my hepatitis B and hepatitis C results		
I would like to be informed of my results for possible diabetes, anaemias, high cholesterol and liver function		

Name of participant (BLOCK CAPITALS)	**Date**	**Signature**
Witness to consent	**Date**	**Signature**

UGANDA RESEARCH ETHICS COMMITTEE 12 AUG 2011 P.O.BOX 49 ENTEBBE (U)

VNO

The Kyamulibwa Cohort Study Round 22

HNO

Adult and emancipated minor consent: survey and blood samples

Enter participant ID number

Consent for Future Use of Blood Samples

Please initial or thumbprint boxes

1	I understand the information sheet I have read, or have had read to me, for this study and have been given a copy to keep. I have had the opportunity to ask questions and understand why the study is being done.		conf1
2	I agree that the blood samples I have provided for this study, and information gathered from them, can be stored in a secure archive in Uganda and the UK for future research in these and other countries. I understand that future studies using my blood samples and information may be done by other research teams in Uganda, the UK, the USA and other countries.		conf2 conf3
3	I agree that the blood samples I have provided for previous surveys, and information gathered from them, may be used for future research. I understand that future studies using my previously collected blood samples and information may be done by other research teams in Uganda, the UK, the USA and other countries.		conf6
4	I understand that future research will only be undertaken on anonymised blood samples and will include biochemical and genetic studies aimed at investigating the causes of infectious and cardiometabolic diseases. I understand that the results of these future studies are unlikely to have any implications for me personally and that I will not benefit financially if this research leads to the development of a new treatment or medical test.		conf4
5	I understand that my participation in this study is voluntary and that I am free to withdraw from the study at any time, without giving a reason and without my standard medical care provided by the state or legal rights being affected.		conf5

Name of participant (BLOCK CAPITALS)	**Date**	**Signature**
Witness to consent	**Date**	**Signature**
Name of person taking consent	**Date**	**Signature**

1 copy for participant; 1 to be kept by research team

UGANDA VIRUS RESEARCH INSTITUTE ENTEBBE SCIENCE & ETHICS COMMITTEE 12 AUG 2011 P.O.BOX 49

Appendix II: UVRISEC Ethical Approval for parent study

Tel: (256) 41- 321962 (Direct),
(256) 41- 320385/6 (General)

Fax: (256) 41- 320483
Email:directoruvri@ug.cdc.gov

THE REPUBLIC OF UGANDA

UGANDA VIRUS RESEARCH INSTITUTE
P.O.Box 49 ENTEBBE (U)

Our Ref : GC/127/10/10/25
Your Ref:

25th October 2010

Dr. Dermot Maher,

RE: UVRI SEC review of the protocol titled **"Epidemiology and genetics of communicable and non-communicable diseases in the General Population Cohort in kyamulibwa, Uganda".**

Thank you for submitting the above study dated 11th October 2010 to UVRI Science and Ethics Committee (SEC).

This is to inform you that your study was reviewed during the SEC meeting of 21st October 2010. After review of your protocol, UVRI SEC approval has been given for you to carry out your study for a period of year that is up to 24th October 2011.

The reviewed and approved documents included;

1. UVRI application form.
2. Study protocol
3. Study questionnaires
4. Consent and assent forms
5. Applicants CVs
6. Copy of the MRC/UVRI unit standard MTA.

UVRI SEC would expect a progress report and a request for renewal a month earlier before the expiry of initial date of approval (24/10/2011), to allow timely review.

You can now commence with your study after registration with the Uganda National Council for Science and Technology (UNCST).

Yours sincerely,

Namale (NAMALE)

Dr. Alice Namale
Vice Chair, UVRI SEC

C.C Secretary, UVRI SEC

Appendix III: UNCST Ethical Approval for parent study

Uganda National Council For Science and Technology

(Established by Act of Parliament of the Republic of Uganda)

Your ref: ..... *Date :* .....

Our ref: .....HS 870..... **20/12/2010**

Dr. Maher Dermot
C/O MRC/UVRI Uganda Research Unit on AIDS
Uganda Virus Research Institute
P.O Box 49
Entebbe

Dear Dr. Maher,

RE: RESEARCH PROJECT, "THE EPIDEMIOLOGY AND GENETICS OF COMMUNICABLE AND NON-COMMUNICABLE DISEASES IN THE GENERAL POPULATION COHORT IN KYAMULIBWA, UGANDA"

This is to inform you that the Uganda National Council for Science and Technology (UNCST) approved the above research proposal on **November 22, 2010.** The approval will expire on **November 22, 2011.** If it is necessary to continue with the research beyond the expiry date, a request for continuation should be made in writing to the Executive Secretary, UNCST.

Any problems of a serious nature related to the execution of your research project should be brought to the attention of the UNCST, and any changes to the research protocol should not be implemented without UNCST's approval except when necessary to eliminate apparent immediate hazards to the research participant(s).

This letter also serves as proof of UNCST approval and as a reminder for you to submit to UNCST timely progress reports and a final report on completion of the research project.

Yours sincerely,

Leah Nawegulo
for: Executive Secretary
UGANDA NATIONAL COUNCIL FOR SCIENCE AND TECHNOLOGY

LOCATION / CORRESPONDENCE
Plot 3/5/7, Nasser Road
P.O. Box 6884
Kampala, Uganda

COMMUNICATION
TEL: (256) 414-250 499, (256) 414 705 500
FAX: (256)414-234 579
EMAIL: uncst@starcom.co.ug
WEBSITE: http://www.uncst.go.ug

Appendix IV: SBSREC Ethical Approval for this particular study

MAKERERE UNIVERSITY
P.O. Box 7072 Kampala, Uganda
E-mail:biomedicalresearch62@gmail.com
Tel: + 256 779340363/+256701940363
COLLEGE OF HEALTH SCIENCES
SCHOOL OF BIOMEDICAL SCIENCES
RESEARCH ETHICS COMMITTEE

25/04/2023

To: IRENE NAMARA

MAKERERE UNIVERSITY
+256701093772

Type: Initial Review

Re: SBS-2023-298: Genetic Determinants of Differential Susceptibility to Viral Hepatitis B And C Infection Among A General Population Cohort From South Western Uganda: A Genome-Wide Association Study

I am pleased to inform you that at the **127** convened meeting on **16/03/2023**, the MAK School of Biomedical Sciences REC (SBSREC), committee meeting voted to approve the above referenced application.
Approval of the research is for the period of **25/04/2023** to **25/04/2024.**

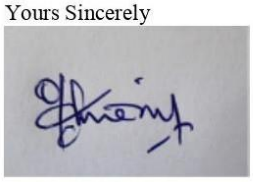
As Principal Investigator of the research, you are responsible for fulfilling the following requirements of approval:

1. All co-investigators must be kept informed of the status of the research.
2. Changes, amendments, and addenda to the protocol or the consent form must be submitted to the REC for re-review and approval **prior** to the activation of the changes.
3. Reports of unanticipated problems involving risks to participants or any new information which could change the risk benefit: ratio must be submitted to the REC.
4. Only approved consent forms are to be used in the enrollment of participants. All consent forms signed by participants and/or witnesses should be retained on file. The REC may conduct audits of all study records, and consent documentation may be part of such audits.
5. Continuing review application must be submitted to the REC **eight weeks** prior to the expiration date of **25/04/2024** in order to continue the study beyond the approved period. Failure to submit a continuing review application in a timely fashion may result in suspension or termination of the study.
6. The REC application number assigned to the research should be cited in any correspondence with the REC of record.
7. You are required to register the research protocol with the Uganda National Council for Science and Technology (UNCST) for final clearance to undertake the study in Uganda.

The following is the list of all documents approved in this application by MAK School of Biomedical Sciences REC (SBSREC):

No.	Document Title	Language	Version Number	Version Date
1	Protocol	English	3	2023-04-15

Yours Sincerely



Dr. Moses Ocan
For: MAK School of Biomedical Sciences REC (SBSREC)