GENOME-WIDE ASSOCIATION STUDY OF POLIOMYELITIS

GENETIC PREDISPOSTION OF PARALYTIC POLIOMYELITIS USING GENOME-WIDE ASSOCIATION STUDIES

BY TINUKE OLUWASEFUNMI OLAGUNJU, MD, MPH

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirement

For The Degree of Masters of Science In Health Research Methodology

McMaster University

McMaster University MASTER OF SCIENCE (2019) Hamilton, Ontario (Health Research Methodology)

TITLE: The Genetic Predisposition of Paralytic Poliomyelitis Using Genome-Wide Association Studies

AUTHOR: Tinuke O. Olagunju MD, MPH

SUPERVISOR: Professor Mark Loeb, MD, FRCPC, MSc

ACKNOWLEDGEMENTS

First, I would like to thank God and Father of my Lord Jesus Christ for his faithfulness in my work and life. Next, I will like to appreciate my supervisor Prof. Mark Loeb for his exemplary support, help and guidance throughout this process of completing my studies in Health Research Methodology. He challenged and encouraged me to explore unfamiliar terrains and pursue novelty in research. I would also like to thank Prof. Guillaume Pare who was part of my thesis committee members, but indeed acted as a co-supervisor giving me exemplary counsel and guidance throughout the course of my studies. I am also grateful for Prof. Marek Smieja, a member of my thesis committee for his continued guidance and direction and Prof. Andrew McArthur for his feedback on my work. In addition, I will like to acknowledge Dr Amel Lamri and Tafadzwa Machipisa for their input in my analysis and Lorraine Carroll for her assistance. Lastly, I am most thankful for the love and support from my husband, Andrew, and family throughout all of my scholarly and career pursuits.

ABSTRACT

Poliomyelitis is a foremost cause of paralysis among preventable diseases among children and adolescents globally. It is caused by persistent infection with poliovirus (PV). The PV infection does not always cause paralysis. A lack of immunization always increases the risk of paralytic polio. Genetic factors also been shown to affect the risk of developing the disease.

The aim of this thesis is to investigate whether there are any genetic associations to paralytic poliomyelitis. This is based on a model for understanding its nature as a complex disease, where many genes are involved in contributing to the disease state. This is a population-based case-control study to identify genetic loci that influence disease risk.

The study examined the association of genetic variation in single nucleotide polymorphisms (SNPs) across the genome with paralytic poliomyelitis susceptibility in the United States and Canadian survivors of poliomyelitis population, using a genome-wide association study (GWAS) approach. No association was observed. Loci that have been previously implicated were not found to affect the susceptibility to poliomyelitis in this study.

The thesis consists of four chapters. Chapter 1 describes the epidemiology, pathogenesis and management of poliomyelitis. Chapter 2 gives an overview of the genomics of infectious diseases in general. Chapter 3 introduces the study population and presents the genome-wide analysis and associations with logistic regression to identify loci explore genes that might be associated

with paralytic poliomyelitis and presents results. Chapter 4 discusses the implications of the results and explains future directions.

Keywords: poliomyelitis, genome-wide association studies (GWAS), case-control, logistic regression

ABBREVIATIONS

ABCB1 - ATP-binding cassette sub-family B member 1

CD155 - Cluster of differentiation 155

CSF - Cerebrospinal fluid

CDC - Center for Disease Control and Prevention

cDNA - Complementary DNA

CXADR – Coxsackie-Adenovirus Receptor Gene

DNA - Deoxyribonucleic acid

EV - Enterovirus

FcγR - Fragment crystallizable gamma receptor

GATA6 - GATA-binding factor 6

GOC - Government of Canada

GPEI - Global Polio Eradication Initiative

GWAS – Genome-wide Association Studies

HCl – Hydrochloric acid

HLA – Human Leukocyte Antigen

HWE – Hardy-Weinberg Equilibrium

IBD - Identity-by-descent

IBS - Identity-by-state

IgA – Immunoglobulin A

IgG – Immunoglobulin G

IgM – Immunoglobulin M

IPV - Intravenous polio vaccine

IRES - Internal ribosome entry site

KCNJ4 - Potassium inwardly-rectifying channel, subfamily J, member 4

MAF – Minor Allele Frequency

MHC - Major Histocompatibility Complex

NAT - Nucleic acid test

NAAT - Nucleic acid amplification test

NCBI - National Center for Biotechnology Information

OPV - Oral polio vaccine

PCA - Principal component analysis

PCR - Polymerase Chain Reaction

PPS - Post-polio syndrome

PV - Poliovirus

PVR - Poliovirus receptor

QQ-plot – Quantile- Quantile plot

RNA – Ribonucleic acid

RT-PCR - Reverse Transcriptase Polymerase Chain Reaction

SNP - Single Nucleotide Polymorphism

TLR3 - Toll-like receptor 3

VAPP - Vaccine-associated paralytic polio

VDPP - Vaccine-derived paralytic polio

vRNA - Viral RNA

XRCC1 - X-ray repair cross-complementing protein 1

Table of Contents

Table of Contents

CHAPTER 1

1 Epidemiology, Pathogenesis and Management of Poliomyelitis

1.1 Poliomyelitis: History

Poliomyelitis is a disease of the central nervous system caused by a virus called poliovirus. Poliomyelitis is considered an ancient disease with records of episodes dating as far back as the eighteenth century (Scott, 1773; Underwood, 1889). Up to the 19th century, populations experienced only relatively small outbreaks but this changed around the beginning of the 20th century likely as a result of factors such as changing demographics, overcrowding, socio-cultural and economic determinants of health (Ochmann & Roser, 2019). As polio is transmitted via the fecal-oral route, the lack of safe drinking water and sanitation meant that children became exposed to the poliovirus soon after birth. During the first six months of life is time, children however benefit from a passive immunity that is passed on from their mothers in the form of antibodies. Serum antibodies, derived maternally, prevented viremia, thus preventing the ability of the virus to seed target tissue that is, the brain. The serum antibodies would however not prevent enteric infection in infants. As hygiene standards improved, the age at which children were first exposed to the poliovirus increased and this meant that the maternal antibodies were no longer present to protect children from polio. This led to increased outbreaks, and caused the age of contraction of polio to increase. In earlier years, the average age was between one and five year-olds but in later years a substantial number of cases was occurred in teenage and young adult years (Hall, Nathanson & Langmuir, 1957).

Between the time it was first discovered in the early eighteenth century, and modern times of the twentieth century, poliomyelitis had evolved from being a sporadic disease to an endemic disease that paralyzed hundreds of thousands of children globally. With the advent of vaccine, however, in the 1950s

and 1960s, it was drastically and almost entirely brought under control, but only in developed countries (GPEI, 2019). Large outbreaks still continued around the world with an estimated 350,000 cases a year and prevalence in 125 countries (GPEI, 2019). It was not until 1970s that routine vaccination was introduced worldwide as part of national immunization programs. By the time Global Polio Eradication Initiative (GPEI) was launched in 1988, poliomyelitis was paralyzing 1000 children per day.

1.2 Poliomyelitis: Significance

As a disease of public health significance, poliomyelitis provides an intriguing and instructive case study from the standpoint of medical history (Nathanson & Kew, 2010). Although causing sporadic disease prior to the 19th century, epidemics became established as a sporadic disease before the 19th century, Following the discovery and wide use of intravenous polio vaccine (IPV) and oral polio vaccine (OPV), the incidence of poliomyelitis was dramatically reduced, constituting one of the most successful public health programs ever conducted on a global scale. However, the goal of global polio eradication is yet to be achieved (GPEI, 2019; WHO, 2019).

The challenge is that although, 99% of the world no longer considers it a threat (Grassly, 2013), the world still faces a risk of contracting polio (Moszynski, 2013; WHO, 2019). Poliovirus is transmitted by infectious humans or their waste. It has a limited survival in the environment and humans are the only reservoir (Dowdle, 1997), though it can replicate in other primates. Virus reproduction is restricted to primates because other cell lines lack a functional receptor molecule (Troy *et al.,* 2015). Immunization with polio vaccine interrupts virus transmission as evidenced by the eradication of poliomyelitis in most countries (WHO, 2018).

There is a risk of resurgence and re-establishment of the disease globally as long as there are countries in which transmission has not been totally interrupted. Unless it is totally eradicated from the world, polio

could result in as many as 200,000 new cases every year, within 10 years, all over the world (Dowdle, 2002; WHO, 2018).

In addition, there is still also the risk of vaccine-associated paralytic polio (VAPP) and vaccine-derived poliovirus (VDPV). Furthermore, polio-like illnesses seem to be emerging and knowledge about polio can help provide understanding as to how to combat them (CDC, 2019; Preston, 2015). Globally, there are 107 and 157 cases of poliomyelitis due to wild polio virus (WPV) and circulating vaccine-derived poliovirus (cVDPV) respectively in 2019 (GPEI, 2019). In view of these, relentless effort that employs diverse approaches is required, until the final target of eradication is achieved. Notably, paralysis from polio is correlated with a wide range of important life outcomes including health, education occupation and income.

1.3 Poliomyelitis: Epidemiology

Poliomyelitis, also known as infantile paralysis, is predominantly a disease of childhood and adolescent years, affecting mainly children under five years of age though there have been cases of young-adult-onset. Presently, the prevalence of poliomyelitis has decreased by over 99% since 1988 when it was occurring at an estimated rate of 350,000 cases per year globally in more than 125 endemic countries to 33 cases per year in 2018. Four regions of the world including the Americas, Europe, South East Asia and the Western Pacific have been certified polio-free (See figure 7.1). This reduction is a result of a concerted effort to eradicate the disease through active surveillance and immunization. However, three countries have not been able to interrupt its transmission. These countries are Nigeria, Pakistan and Afghanistan. The untoward consequence of this is that polio can be re-imported back and spread very quickly back into the polio-free, unimmunized countries. In 2018, there were 33 cases of wild polio in Afghanistan and Pakistan, and 103 cases of vaccine-derived polio in Philippines, Indonesia, Democratic Republic of Congo, Nigeria, Niger and Somalia (WHO, 2018). Wild poliovirus type 2 was eradicated in 1999 and no case of

wild poliovirus type 3 has been found since the last reported case in Nigeria in November 2012. Wild Type I is still endemic in Nigeria, Pakistan and Afghanistan.

Globally, there are about 20 million polio survivors and about 40,000 in Canada. These individuals are susceptible to developing post-polio syndrome (PPS). Post-polio syndrome (PPS) is a condition that affects polio survivors, years after recovery from poliomyelitis. It is characterized by gradual new weakening in muscles that were previously affected by the polio infection. The most common symptoms are slowly progressive muscle weakness, fatigue and a gradual muscle atrophy. Other symptoms include joint pain as a result of bone degeneration and increasing skeletal deformities such as scoliosis. These symptoms run in a spectrum. Though rarely life-threatening, PPS can significantly interfere with an individual's ability to function independently. As examples, respiratory muscle weakness, can cause breathing difficulties and sleep apnea and weakness in swallowing muscles can result in food and liquid aspiration.

1.4 Poliomyelitis: Causative agent

Poliomyelitis is caused by poliovirus (PV). The virus belongs to the family Picornaviridae and genus enterovirus. PV is a type of human enteroviruses (EVs), a member virus of Enterovirus C. EVs are the most common human viruses and include more than 100 different genotypes; polioviruses being the best-characterized examples.

Polioviruses are small positively stranded RNA enteroviruses, just 30 nm across and with a complete genome of only approximately 7500 nucleotides. The single-stranded RNA core is surrounded by a protein capsid without a lipid envelope. This thin 20-sided shell is composed of four virion proteins (VP1, VP2, VP3, VP4) (Hogle *et al.,* 1985). It makes poliovirus resistant to lipid solvents and stable at low (acidic) pH. PV can survive for weeks at room temperature and for many months at 0°C to 8°C. They are resistant to ether, 70% alcohol and other laboratory disinfectants. Treatment with 0.3% formaldehyde, 0.1 N HCl, or free residual chlorine at a level of 0.3 to 0.45 parts per million rapidly inactivates polioviruses, as does

exposure to a temperature of 50°C or higher or to ultraviolet light (Minor & Bell, 1990). Two distinct types of antigens have been identified in harvests from virus infected cells They include D antigen which is largely but not exclusively associated with infectious virus and C antigen with empty capsids (Minor, 1990).

Poliovirus infects human cells by binding to the poliovirus receptor also known as CD155 on the cell surface. The 5' end of poliovirus RNA is extremely long, over 700 nucleotides and highly structured. This region of the viral genome is called internal ribosome entry site (IRES), and directs translation of the viral RNA. Following attachment to host cell membrane, the virus is taken up by receptor-mediated endocytosis. This interaction facilitates an irreversible conformational change of the virus leading to the viral RNA being released inside the host cell. Upon release, the virus hijacks the cell's translation machinery and causes inhibition of cellular protein synthesis in favor of virus-specific protein production. Poliovirus mRNA is translated as one long polypeptide and synthesized by internal proteases into about 10 individual viral proteins. After translation, transcription and genome replication occurs giving rise to fully assembled poliovirus which leaves the confines of its host cell by lysis. Each lysed cell can release up to 10,000 polio virions. It is shed in enormous quantities in the throat and intestines of infected individuals such that a gram of stool can contain several million virus particles (Dowdle, 2002)

1.5 Poliomyelitis: Pathogenesis

Poliomyelitis is an infection of the central nervous system caused by poliovirus (PV). Humans are the only natural hosts of PV. Transmission occurs by close personal contact through the fecal–oral route. Poliomyelitis is initiated by oral ingestion of poliovirus (Blondel *et al.,* 2005). Following ingestion, PV enters binds on the PVR (CD155) receptor on the gastro-intestinal system (oropharyngeal and intestinal mucosa) and penetrates the mucosa through specialized micro-fold cells (See figure 7.2). It resides there for about one week to three weeks (6-20 days) of incubating period (GPEI, 2019; Blondel *et al.,* 2005), replicating underlying submucosal lymphoid tissues. During this period, the virus is secreted in saliva and feces

causing most host-to-host transmission. A person is infectious 7-10 days before and after appearance of symptoms and continues to be as long as virus remains in saliva or feces. This is the alimentary stage. It then spreads to the follicular dendritic cells of the tonsils, the M cells of the Peyer's patches of the intestine and drains into the cervical and mesenteric lymph nodes constituting the lymphatic phase (Blondel *et al*., 2005). Following this, it enters into the viremic phase in which it is absorbed into bloodstream from where it spreads to other organs. Viremia is considered essential for leading to paralytic poliomyelitis in humans. It can reside in the lymphatic system and blood for up to 10 and 17 weeks causing flu-like symptoms. There are two dissemination routes: blood-brain barrier and peripheral nerves. In about 5% of infected people, it goes on to invade the meninges leading to non-paralytic polio forming the neural phase. In 1 in 200 people, it passes along certain nerve fiber pathways, replicating in and destroying motor neurons as it invades the anterior horn cells and ventral grey matter of spinal cord, bulbar region of brain stem and upper part of cervical vertebrae resulting in acute flaccid paralysis involving a single extremity to complete quadriplegia. The most serious complication of paralytic poliomyelitis is a respiratory failure from paralysis of the diaphragm and intercostal muscles due to the involvement of the motor centers of the brainstem and the spinal cord, which can lead to death (Blondel *et al.,* 2005).

1.6 Poliomyelitis: Immune response

IgM and IgG appear in the serum within 7 to 10 days of exposure to the virus. These antibodies generally prevent the poliovirus from entering into the central nervous system. Initially, the IgM response is greater than the IgG response. Two weeks after exposure, the level of IgM increases to its peak, following which it disappears from the serum within about 60 days. IgG levels on the other hand, increase steadily and persist for life in the serum (Paul *et al.,* 1951). IgA antibody appears in the serum 2 to 6 weeks after exposure (Ogra *et al.,* 1968) and remains at low levels with some individuals experiencing no rise in serum level. They are produced by plasma cells originating in gut-associated lymphoid tissues, mainly Peyer's patches. These cells are contained in mucosal sites, including the intestine, the pharynx, and the

mammary glands (Walker & Isselbacher, 1977). IgA antibody has been found to have persisted in the nasopharyngeal secretions of individuals 10 to 15 years after natural infection with wild type-1 poliovirus (Ogra & Karzon, 1971).

Passive immunity is transferred from mother to fetus via the placenta. The newborn will contain approximately the same concentration of IgG type1 and IgG type 2 but a lower concentration as the mother (Ananthakrishanan *et al.,* 1988). This maternal-derived immunity disappears from the serum after a half-life of about thirty days on average, ranging from 21 days to 50 days (Robertson, 1993)

Three antigenically distinct strains are known, (type 1, type 2, and type 3) based on their reaction with reference panels of neutralizing antisera (Bodian *et al.,* 1949). Infection with one type does not protect from the other types; however, immunity to each of the three strains is lifelong. The three serotypes of polio virus, although they differ in their virulence potential, affect human cell specifically through PVR CD-155 receptors. Presently, of the three strains, Type I is the only encountered form accounting for all cases of paralytic illnesses.

1.7 Poliomyelitis: Symptoms

Ninety-five percent (95%) of people infected with polio virus are asymptomatic. Of those who get symptoms, these include pharyngitis or gastroenteritis, headache; fever of 38-40 ºC; sore throat; anorexia; nausea; vomiting; and muscle aches. These symptoms may or may not subside in 1-2 weeks. Other symptoms include features of non-paralytic aseptic meningitis (e.g. irritability, restlessness, apprehensiveness, emotional instability, stiffness of the neck and back and Kernig and Brudzinski signs). However, five percent (5%) of infected people will have paralytic poliomyelitis – (loss of reflexes, severe muscle aches or weakness, loose and floppy limbs which may be one-sided (flaccid paralysis) or deformed limbs. There are three types of paralytic poliomyelitis, affecting spinal cord (spinal), brainstem (bulbar) or both (bulbo-spinal). Some people go on to develop post-polio syndrome. This is a cluster of disabling signs

and symptoms that affect some people years after having polio. Its signs and symptoms include progressive muscle or joint weakness and pain, fatigue, muscle wasting (atrophy), breathing or swallowing problems, sleep-related breathing disorders, such as sleep apnea decreased tolerance of cold temperatures. (WHO, 2018; CDC, 2017)

1.8 Poliomyelitis: Diagnosis

Poliovirus can be detected from respiratory specimens (endotracheal tube aspirate, nasopharyngeal aspirate or swab), gastrointestinal material (stool) or cerebrospinal fluid (CSF). The testing formats include viral isolation by cell culture, nucleic acid testing and serological tests (WHO, 2019; CDC, 2017).

Cell culture and virus isolation

Two stool specimens and 2 throat swabs collected 24 hours apart is the gold standard. Specimens should be obtained as early as possible after the onset of disease, preferably within 14 days. Stool should be collected in a sterile clinical cup. Throat swabs should be collected and transported in viral transport medium. Virus persists in the oropharynx for 1 week, and in the stool up to 6 to 8 weeks after onset. Blood samples might also be collected for analysis. A lumbar puncture (spinal tap) may also be performed to collect cerebrospinal fluid to look for evidence of the polio virus infection**.**

Nucleic acid test (NAT) and Polymerase Chain Reaction (PCR)

A nucleic acid test (NAT) or nucleic acid amplification test (NAAT) is a technique that can be utilized to detect poliovirus in specimen. It detects genetic materials rather than antigens or antibodies. This includes an amplification step of the genetic material such as polymerase chain reaction (PCR). Reverse Transcriptase (RT-PCR) is performed on cell cultures containing poliovirus. The viral RNA (vRNA) is converted to complementary DNA (cDNA). The cDNA is amplified in a PCR reaction using polymerase. The PCR products are detected and identified by hybridization with specific probes.

Oligonucleotide mapping

Following isolation, the virus is tested using oligonucleotide mapping (fingerprinting) or genomic sequencing to examine the genetic sequence of the virus to determine if it is wild-type or vaccine-like.

Serology

The micro neutralization test is used for detecting antibodies to poliovirus. Three specimens are collected serially. An acute-phase serum specimen should be obtained as early as possible in the course of illness. A convalescent-phase specimen should be obtained three to four weeks after the acute specimen and, if possible, a third specimen should be obtained three to four weeks after the second specimen is obtained. Acute and convalescent specimens are tested for evidence of a rise in neutralizing antibodies to the poliovirus serotype. A fourfold rise in neutralizing antibody (IgG) between the acute and convalescent specimens is suggestive of acute poliovirus infection.

1.9 Poliomyelitis: Prevention

Though irreversible and incurable, poliomyelitis is preventable. The use of two vaccines have been implemented in the fight against poliomyelitis. These are oral polio vaccine (OPV) and inactivated polio vaccine (IPV). The IPV also known as the Salk vaccine is an inactivated injectable (parenteral) vaccine while the OPV which is also known as Sabin vaccine is an oral, live attenuated vaccine.

Vaccination with either IPV or OPV is associated with the production high titers of serum or circulatory IgG antibody to poliovirus. The peak level is achieved after three to four doses. IPV prevent CNS invasion by poliovirus also enhances a marked reduction in viral nasopharyngeal excretion and spread. OPV decreases fecal shedding of poliovirus.

More than thirteen million cases of paralytic poliomyelitis have been prevented through the administrating of over ten billion doses of oral polio vaccine to over 2.5 billion children world-wide during

the period between 1988 and 2013 (Global Health, 2014). However, in 1 in 2.7million cases, oral polio vaccine (OPV) results in Vaccine-associated Polio Virus (VAPP), Vaccine-derived Polio Virus (VDPV) and circulating Vaccine-derived Polio Virus (cVDPV).

1.10 Risk and Prognostic factors

Risk factors include living in or travelling to endemic countries (GOC, 2019), contact with sub-optimally vaccinated populations (Alexander *et al.,* 2009), viral load, increasing age, pregnancy and strenuous exercise (Prevots, 2000). Prognostic factors include administration of intramuscular injections, severity and extent of paralysis.

CHAPTER 2

2 Literature Review-Genomic of Infectious diseases

2.1 An overview of infectious diseases.

Research suggests that host genetics play a major role in the pathogenesis of infectious diseases (Loeb, 2013). Genes that are involved in immune response are exquisitely diverse (Burgner *et al.,* 2006) and host genetic variation in human populations contributes to susceptibility to infectious disease. This variability is a function of the genetic composition of the individual and it is believed to be due at least in part to polymorphisms in genes responsible for immune response (Wyatt, 1978). These genetic variations underlie observed predisposition to changes in the response to invading exogenous pathogens, infection and contribute significantly to infectious disease burden (Allison, 1954).

Genetic studies provide the tools that are needed to understand the underlying genetic factors responsible for the variation in susceptibility to pathogen infection, as well as interactions between host and pathogen that define the host response (Castiblanco *et al.,* 2015). An early evidence of genetic susceptibility to infectious diseases dates back to identification of the sickle hemoglobin heterozygous variant as a major resistance and protective factor for malaria (Allison, 1954). A variant in the Duffy antigen gene promoter, which results in lack of erythrocyte surface expression, prevents binding by Plasmodium vivax, and is therefore protective. (Michon, 2001; Miller, 1976). There exists also, well-established links between specific genetic alterations and predisposition to malaria (P. falciparum and P. vivax), Creutzfeldt-Jacob disease (CJD), human immunodeficiency virus (HIV), and Norwalk virus.

Identification of biomarkers may inform novel targets to advance detection, prevention and treatment of disease (Prasad, 2016).

A genome-wide association study is a method that involves examining markers across the complete sets of DNA, or genomes, of many people to find genetic variations associated with a particular disease. A genome-wide association study (GWAS) is conducted as a case-control study with two groups of participants: those with and without the disease or phenotype being studied. If certain genetic variations, which are called single nucleotide polymorphisms, or SNPs are found to be significantly more frequent in people with the disease compared to people without disease, the variations are said to be "associated" with the disease. This can serve as a powerful pointer to, though do not necessarily mean causation. More in-depth investigation of the region DNA base pair sequencing is required for confirmation.

Until recently, the approach to infectious disease management has been primarily rooted in a pathogen-centered model (Klebanov, 2018). However, several genetic epidemiological studies focusing on infectious disease conducted among monozygotic (MZ) and dizygotic (DZ) twin pairs have suggested a genetic contribution to disease burden (Klebanov, 2018). In the same vein, susceptibility to some infections show markedly increased concordance in monozygotic compared with dizygotic twins (Jepson, 1998). Examples include tuberculosis (Uren *et al.,* 2017; Kallmann & Reisner, 1943; Comstock, 1978; Simonds, 1963), sinusitis (Marshall, 1962), leprosy (Chakravartti, 1973), Helicobacter pylori-specific antibody titers (Malaty *et al.,* 1994), and acute and chronic otitis media (Casselbrant, 1999; Rovers, 2002). In a study by Lin *et al.* (1989) among Chinese same-sex twins to investigate the distribution, as well and dizygotic (DZ) twins and singleton controls, highly significant differences were observed in the concordance of carrier status between monozygotic and dizygotic twins. In their study of children with HSV1 encephalitis, Cassonova *et al.* (2010) identified five mutations affecting genes encoding for either Toll-like receptor 3 (TLR3) or a signal transduction protein linking TLR3 to the interferon "a gene promoter", thereby preventing the synthesis of the key molecule in the anti-HSV1 immune response (Cassonova *et al.,* 2010).

2.2 Genome-Wide Association of infectious diseases

There has been some progress in GWAS studies of infectious diseases. GWAS studies of viruses similar to poliovirus (in the same genus - enteroviruses) include that of Coxsackie virus and Echovirus, Hepatovirus. In their study, Gu *et al.* (2018) found ABCB1 and XRCC1 genes variants to be associated with the protection against HAV infection and rs7412 C to T variation in APOE gene to be less prone to be infected by HEV (Gu *et al.,* 2018). Another study reported a positive association between one of the interferon lambda genes (IL-28B), and viral clearance /response to treatment for hepatitis C (Thomas *et al.,* 2009). A study by Marsman *et al.* (2016) identified a locus at chromosome 21q21, 98 kb proximal of the CXADR gene, encoding the Coxsackie and adenovirus receptor to be associated with arrhythmia susceptibility. A study by Cooper *et al.* (2009) suggested that a SNP located 30 kb centromeric of the genes encoding TLR-7 and -8 which may have a function in the host response to enterovirus infection (Triantafilou, 2005; Wang, 2007) is associated with increased risk for the diabetes (Cooper *et al.,* 2009) and may potentially alter the magnitude of the host response to enterovirus infection (Lind *et al.,* 2012).

GWAS studies of other infections include a study conducted among Africans that identified a variant (transcription factor GATA6) on chromosome 18 as being associated with tuberculosis cases (Thye et al., 2010). Another study in Gambia suggested that a hemoglobin AS variant could be associated with malaria (Jallow *et al.,* 2009). In addition, Troyer *et al.,* (2011) identified a variant on chromosome 2 as being associated with HIV disease progression.

It is of note that human genetic variation has also led to the development of new antimicrobial drugs and vaccines. It forms the basis for personalized medicine, in which the current one size-fits-all approach to medical care will give way to more customized strategies by tailoring management to each person's unique genetic makeup. Examples include Plasmodium vivax malaria vaccine; the Duffy binding protein

for malaria, mannose-binding lectin antimicrobial replacement therapy for bacteremia and CCR5 entry-inhibitor antiretroviral for HIV.

2.3 Genetic Risk factors

Genetic variations may underlie observed predisposition to poliomyelitis (Klebanov, 2018). This is arguably an important reason why some people develop paralysis following poliomyelitis, while others clear the infection, as a result of difference their genetic constitution. Importantly, genetic susceptibility could explain individual differences to response to both wild-polio as well as vaccine-polio virus infection.

It is also plausible to suggest that genetic susceptibility explains why less than one percent (1%) of individuals infected with poliovirus develop paralytic polio (WHO, 2018). In addition, in families with a clinical case of poliomyelitis, the ratio of in-apparent to apparent infection is between 3:1 and 7:1 compared to 100:1 in the general population and observational twin-twin concordance studies which suggest the important role of host genetics.

The question of genetically controlled constitutional susceptibility to acute poliomyelitis has been discussed by several authors (Aycock, 1942; Herndon & Jennings, 1951). In their study conducted on forty-seven pairs of twins, parents and siblings suggested the existence of a genetic factor controlling, at least in part, susceptibility to the paralytic form of poliomyelitis (Herndon & Jennings, 1951).

Several genetic epidemiological studies focusing on infectious disease among monozygotic (MZ) and dizygotic (DZ) twin pairs have suggested a genetic contribution to disease burden (Comstock, 1978; Lin *et al.,* 1989; Sorensen *et al.,* 1988). In the study conducted by Sorensen *et al.* (1988), the adopted child had 5.81 greater risk of also dying from infection if the biological parent died from infection whereas if it is the adoptive parent that died from infection, the relative risk (RR) for the child dying from infection was about only one. In particular, a twin-family study of poliomyelitis revealed that concordance of paralytic

poliomyelitis was reported to be thirty-six percent (36%) among monozygotic twins compared to six to seven percent (6-7%) among dizygotic twins (Herndon *et al.,* 1951).

Several other studies have also pointed out the importance of genetic risk factors in paralytic poliomyelitis, where first and second-degree relatives were studied (Addair, 1942; Aycock, 1942; Wyatt, 2014). As mentioned there is large variation in the susceptibility to developing paralytic poliomyelitis among individuals exposed to poliovirus infection. This could reflect a difference in host susceptibility to either exposure of poliovirus, infection following exposure, persistence of the infection or sensitivity to viral proteins.

While previous studies have suggested a genetic predisposition to poliomyelitis, this present study, to our knowledge is the first to investigate possible genetic underpinning of paralytic poliomyelitis using genome-wide association studies.

2.4 Genetic variation in immunological pathway

HLA antigens of the immune response system have been reported to be associated with paralytic polio, post-polio syndrome as well as effect on the susceptibility toward poliomyelitis infection. Many studies have focused on the major histocompatibility complex (MHC) region. The genes found in this region are very polymorphic giving rise to large variety of expressed MHC molecules. The function of the MHC is to display antigen peptides on the cell surface to initiate the cell-mediated immunity (CMI).

Genome-wide association studies (GWAS) suggest that immunologic variation in the major histocompatibility complex (MHC) contributes to host genetic infectious disease susceptibility (Klebanov, 2018; Chapman *et al.,* 2012). In particular, human leukocyte antigen (HLA) class II has been implicated and several associations have been identified (Walsh *et al.,* 2003) making HLA the prototypical candidate genetic region for infectious disease susceptibility. A study by Zander *et al.* (1979) to investigate Caucasian patients who had paralytic poliomyelitis in the 1950s and the early 1960s, showed no significant deviation

in frequency of the HLA-A, -B or -D determinants from control populations (Zander *et al.,* 1979). Another study by Lasch *et al.* (1979) among a vaccine-protected infant population in the Gaza Strip in the 1970s suggested that though HLA-AW19 and -B7 antigens were found more frequently in the affected children compared with control subjects, these results were not statistically significant (Lasch *et al.,* 1979).

Genes that have been implicated in the etiology of poliomyelitis include the poliovirus receptor gene. It is located on chromosome F 19 and regulates its development (Couilin *et al.,* 1976). The viral receptor plays an important role in the persistence of entero-viral infections (Gun, 2001). Polioviruses share a common PV receptor (PVR) that belongs to the immunoglobulin superfamily and is known as CD155 (NCBI, 2019). It is encoded on 19q12----q13.2 (Siddique *et al.,* 1988). The PVR contains three extracellular immunoglobulin-like domains, a transmembrane domain and a cytoplasmic tail (OMIM,). PV has been shown to persist in neuronal cell lines (Rhoades *et al.,* 2011). It is expressed within embryonic structures that give rise to spinal cord anterior horn motor neurons (Gromeier, 2000). This is supported and made evident by poliovirus's affinity for the central nervous system. DNA polymorphisms in the poliovirus receptor gene (PVR) are associated with persistent poliovirus infection (Saunderson *et al.,* 2004). PVR gene domain 1 (located in exon 2) confers susceptibility to PV infection (Koike *et al.,* 1990). The protein encoded by this gene is a transmembrane glycoprotein belonging to the immunoglobulin superfamily. The external domain mediates cell attachment to the extracellular matrix molecule vitronectin, while its intracellular domain interacts with the dynein light chain Tctex-1/DYNLT1. The gene is specific to the primate lineage, and serves as a cellular receptor for poliovirus in the first step of poliovirus replication. It has been suggested to have a role in intestinal humoral immune responses and may also be used to positively select MHC-independent T cells in the thymus. Multiple transcript variants encoding different isoforms have been found for this gene. Location is at 19q13.31

A significant correlation has also been observed between FcγR polymorphism and poliomyelitis (Rekand *et al.,* 2002). FcγR is a type of an Fc receptor, a protein found on the surface of some cells that contribute

to the protective functions of the immune system. It got its name from its binding specificity for a part of an antibody known as the Fc (Fragment, crystallizable) region. Fc receptors bind to antibodies that are attached to infected cells or invading pathogens in order to stimulate phagocytic or cytotoxic cells to destroy them. FcγRs belong to the immunoglobulin superfamily and are the most important Fc receptors for inducing phagocytosis of opsonized microbes. There are many types of FcγR. Each one is unique in its antibody affinity as a result of differing molecular structure. FcγR is important for adequate antibody mediated immune response. The binding of IgG to receptors for the Fc region of IgG (FcγR) is a critical step for the initiation and the control of effector immune functions.

Some studies have demonstrated that Ala67Thr mutation in the poliovirus receptor is considered a possible risk factor for the development of both vaccine-associated and wild-type paralytic poliomyelitis (Kindberg, 2009). Ala67Thr (A67T) is found at the protein level. It is an amino acid substitution that results in the change of an Alanine to a Threonine (GCC>ACC). Ala67Thr is a single nucleotide variant located on chromosome 19 located in PVR exon 2. It seems to facilitate increased resistance against poliovirus-induced cell lysis and apoptosis. Ala67Thr mutation has also been demonstrated to be significantly higher in patients with post-polio syndrome (Bhattacharya, 2014). According to National Center for Biotechnology Information (NCBI) however, a Centre that advances science and health by providing access to biomedical and genomic information, it is unclear if it is a pathogenic or benign variant (Gene, 2004).

KCNJ4 gene has been proposed as a cellular genetic marker for infectious PV detection. KCNJ4 gene is located at 22q13.1 in human intestinal epithelial cells (INT-407 cells) and it codes a potassium channel protein (Sano *et al.,* 2017). Potassium channels are known to be involved with electrical signaling in the nervous system. One class is activated by depolarization whereas a second class is not. The class that is not activated by depolarization are termed inwardly rectifying K+ channels. Members of this class have a greater tendency to allow potassium to flow into the cell rather than out of it. This asymmetry in

potassium ion conductance plays a key role in the excitability of muscle cells and neurons. Two transcript variants encoding the same protein have been found for this gene.

Another study demonstrated that Poliovirus (PV) modifies membrane-trafficking machinery in host cells for its viral RNA replication (Arita *et al.,* 2012). These include ARF1, ACBD3, BIG1/BIG2, GBF1, RTN3, PI4KB and VCP.

Though cytogenetic and immune-genetic analysis pointing to genetic mutations and anomalies in immunological regions have suggested a genetic predisposition, the genetic basis for the majority of paralytic poliomyelitis remains largely unexplained.

CHAPTER 3

3. Study aim and Population Genome-wide association studies (GWAS)

3.1 Aim of study

The aim of this study was to identify SNPs associated with risk for poliomyelitis in the general population. The study focuses on the host model of poliomyelitis, investigating the genetic variations that may underlie predisposition to paralytic poliomyelitis.

3.2 Hypothesis and Objectives

The objective of the study is to investigate if genetic variants predisposed children and adolescents to paralytic poliomyelitis in North America following infection with the polio virus. The hypothesis is that there are genetic variants associated with susceptibility to paralytic poliomyelitis in polio-virus infected children and adolescents in North America. To address the hypothesis, a genome-wide association study (GWAS) was performed using saliva of 1064 individuals.

3.3 Study population

Socio-demographic characteristics of study population

The study population was recruited from individuals in Canada and the United States. It is estimated that there are approximately 40,000 and 400,000 individuals in Canada and United States (US) respectively who developed spinal paralytic polio as children or adolescents. Paralytic polio survivors networks in both Canada and the U.S, including those set up by the March of Dimes were approached by mail and telephone.

Our study population includes individuals identified as having had paralytic poliomyelitis in their childhood or adolescent years i.e. who contracted paralytic poliomyelitis in the 1940s and 1950s in Canada and the

United States. Control subjects were recruited from among their spouses. They are individuals who had no history of paralytic poliomyelitis. The rationale for using spouses is that there is a good likelihood for a similar age and ethnicity (Pare, 2017). Moreover, the participation rate of spouses likely will be better improved and higher than community controls.

After getting informed consent, participants completed a brief questionnaire on lifestyle factors and medical history. Age, sex, and past medical history was collected using a standardized data collection form in order to characterize the participants. For controls, information on age, sex, ethnicity as well as confirmation that they did not have a paralytic polio was obtained.

Confirmation of diagnosis paralytic poliomyelitis was made based on clinical presentation as well as medical records according to standards in order to rule out other causes of paralysis.

Following this, they submitted a sample of saliva by mail to McMaster using a self-addressed envelope with prepaid postage.

The study group consists of 1132 individuals of various ancestry that composed of 99.5% European (1075); Mixed (21) African (1); American (8); South Asian (13); East Asian (16) Unknown ancestry (30) individuals.

Recruitment of cases of poliomyelitis was based on medical records. The study participants were predominantly females (70%). Participants were between fifty-one and ninety-three years of age (51-93 years). The mean age at the time of sampling was 67.2 years (SD 14.5 years) and the mean age at onset of paralytic poliomyelitis was 5.6 years (SD 5.6 years).

For the present study, 1132 individuals (752 cases, 380 controls; 447 males, 685 females) were genotyped and 552667 SNPs were detected.

Clinical characteristics of study participants

Clinical features of study participants varied and includes paralysis in one limb (right arm or right leg), paralysis in both limbs on one side (right side and left side), paralysis in upper and lower limbs on different sides (right arm and left leg and vice versa) and paralysis in both lower limbs (both legs with one weaker than the other).

CHAPTER 4

4 Methods and Material

4.1 Specimen handling

Saliva samples (using Flinders Technology Associates (FTA) cards) were collected as specimen from each enrolled participant. The samples were sent in a self-contained envelop allowing participants to obtain the specimen themselves and send it back to us in a pre-paid envelop. One thousand one hundred and sixty-four samples from saliva of the March of Dimes cohort and their spouses were used to evaluate DNA yields.

4.2 DNA extraction

DNA from saliva samples was extracted using the Qiagen Pure Gene kit according to manufacturer's instructions. DNA concentration was quantified using a Bio Spec-Nano Spectrophotometer. The DNA was barcoded and stored at -75 to -85ºC prior to genotyping.

4.3 Genotyping

One thousand one hundred and sixty-four DNA samples were genotyped in two separate batches using Illumina Human Omni 2.5-8 chip. The call rate (rate at which SNPs were correctly identified across samples) for these SNPs using both platforms was greater than 99.5%.

The Illumina Human Omni 2.5-8 chip is a SNP genotyping system. It is an eight-sample Bead Chip array and features approximately 2.5 million markers per single DNA sample. It has a vast genomic coverage rate that cuts across diverse populations. It uses universal primers to amplify SNP-reactive DNA fragments and relies on direct hybridization of genomic targets to array-bound sequences. Single base extension is followed by fluorescent staining, signal amplification, scanning, and analysis using the Genome Studio Software (Ilumina, 2019). A schematic figure is provided in figure 7.3

4.4 Data storage and transfer

For data storage and transfer, the high capacity informatics system at McMaster university (Genologics Geneus v6.4.2 LIMS) was used. The program includes sample tracking, genotyping result management, genotyping quality control and project management. Illumina arrays are scanned and raw data generated using a Dell Precision T5400 with Intel Xeon CPU (16 GB RAM, 64-bit operating system).

4.5 Software for genetic analysis:

PLINK - PLINK is a free, open-source whole genome association analysis toolset, designed to perform a range of basic, large-scale analyses in a computationally efficient manner. PLINK has numerous useful features for managing and analyzing genetic data (Purcell *et al.*, 2007).

R is a programming language and software environment for statistical computing and graphics for developing statistical software and data analysis (R Development Core Team, 2011)

4.6 Factors influencing genetic associations

Biases in study design and errors in genotype calling have the potential to introduce systematic biases into genome-wide association studies, leading to an increase in the number of false-positive and false-negative associations Several factors can confound genetic associations. These include population sub-structure, genotyping batch/ genotyping center and DNA quality.

4.7 Data Cleaning and Quality Control

In order to reduce biases, data cleaning and quality control is instituted. This entails a systematic step-wise quality filtering of raw genotyping data, as shown in Figure 5, and this was conducted using PLINK v1.09 (Purcell *et al.,* 2007). The steps consist of the following:

4.8 Update Genome build

To maintain consistency with the sequencing data, output data was standardized to NCBI Genome Reference Consortium Human Build 37 (GRCh37), also known as hg19 and the primer sequences of each SNP were aligned to the reference panel -InfiniumCoreExome-24v1-1_A and HumanCoreExome-24v1-0_A. This aligns the strands (sequence of SNPs/alleles) in the dataset to the appropriate genotyping chip for correct alignment and variant positioning. It entails updating the chromosome, position and strands of SNPs in the binary .ped file using the strand and position from the files in the InfiniumCoreExome-24v1-1_A and HumanCoreExome-24v1-0_A which was used for the genotyping. It provides for correct identification of SNPs.

4.9 Check for discordant sex check information

A check for discrepancies between the self-reported sex of the individuals recorded in the dataset (ped. file) and their sex based on X chromosome heterozygosity and homozygosity (genomic data) was conducted. Typically one expects male samples to have a homozygosity rate of one (1) and females to have a homozygosity rate less than 0.2. Male DNA samples that are marked as female will have a higher than expected homozygosity rate and female samples marked as male will have a higher than expected heterozygosity rate. Individuals with inconsistences between the two values, designated as "detected problems", were removed (Anderson *et al.,* 2011; Purcell *et al.,* 2007).

4.10 Check for genotyping rate for individuals and markers

Check for the rate at which each SNP (marker) are genotyped (or missing) across individuals and the rate at which each individual (sample) is genotyped (or missing) for SNPs. A call rate of 99% is usually considered appropriate. This was done in six steps: First, we removed SNPs whose rate of missing-ness are 10% and above; i.e. genotyping rate of 90% and below. Next, we removed samples whose rate of missing-ness are 10% and above, i. e genotyping rate of 90% and below. Then, we removed SNPs whose

rate of missing-ness are 5% and above; i.e. genotyping rate of 95% and below. Following this, samples whose rate of missing-ness are 5% and above, i. e. genotyping rate of 95% and below were removed. After which SNPs whose rate of missing-ness are 1% and above, i.e. genotyping rate of 99% and below were also removed. Finally, samples whose rate of missing-ness are 1% and above i.e. genotyping rate of 99% and below were removed. The step-wise filtering allows for a better retaining of samples while still maintaining a high level of filtering.

4.11 Check for heterozygosity

Next samples with excess heterozygosity were Identified and removed. On the average, only about 30% of an individual's marker is expected to be heterozygous. Heterozygosity in more than 30% are considered excessive and to be due to sample contamination. It is of note that a lack of heterozygosity indicates that there is a problem as well. Observed heterozygosity rate per individual is plotted on the x-axis against the proportion of missing SNPs per individuals is plotted on the y-axis. Mean heterozygosity (which is given by (N–O)/N, where N is the number of non-missing genotypes and O is the observed number of homozygous genotypes for a given individual) is calculated. All individuals with heterozygosity rate ± 3 standard deviations from the mean should be excluded (Anderson *et al.,* 2011).

4.12 Check for duplicate and related sample

A fundamental characteristic of population-based studies is that samples are unrelated. In view of this, related and duplicate samples were identified and removed. This was done by creating a genome file in plink. This examines the distance in genetic terms between samples called the identity-by-state (IBS) distance. The degree of recent shared ancestry for a pair of individuals (identity by descent, IBD) can be estimated using genome-wide IBS data. Theoretically, IBD is expected to be equal to one (1) for duplicates or monozygotic twins, equal to 0.5 for first-degree relatives, equal to 0.25 for second-degree relatives,

equal to 0.125 for third-degree relatives and equal to zero if samples are totally unrelated. In practice, samples are considered duplicates if IBD is equal to 0.98 (Anderson *et al.,* 2011).

4.13 Check for population sub-structure

Genetic substructure of human populations can lead to observation of a spurious association, resulting in false-positive results or reduced power in genetic studies when genetically heterogeneous populations are analyzed together without further adjustment (Price *et al.,* 2006; Chao Tian *et al.,* 2008; Pare, 2010). This is as a result of allele frequency differences between cases and control that is unrelated to the phenotypic trait but due to differences among ethnic groups. There are statistical methods that can be applied to discern and correct for these differences and it has become imperative to carry out these corrections in genetic studies (Chao Tian *et al.,* 2008; Pare, 2010). Several different methods have been developed to address issues of population substructure in genome-wide association studies. One of the approaches that have achieved some measure of application to large data sets is the principal component analysis.

Following that population stratification can result in inflated type I error, a principal component analysis using 1000GENOMES ancestry information was conducted. Principal component analysis (PCA) was performed using PLINK in order to confirm self-reported ancestry. The effect of population structure was assessed through principal component analysis (PCA) implemented in Eigenstrat (Price *et al.,* 2006). PCA was used to filter the data for European population.

4.14 Check for Hardy-Weinberg Equilibrium (HWE)

Variants that deviated from Hardy-Weinberg equilibrium with a p-value of less than 0.000001 were removed. HWE states that except where there is an external force, allele and genotype frequencies should remain constant from one generation to another. Based on this, markers that were identified to deviate from HWE, were assumed to have been impacted upon by an eternal force. The external force

suspected here is genotyping errors. Hence they were removed. Controls should not deviate from HWE, hence generally, HWE is perform on only controls in order to avoid removing deviations that are due to true associations.

4.15 Check for minor allele frequency (MAF)

The minor allele is the less common allele for a SNP. The MAF is therefore the frequency of the minor allele. This measure can be used to get a rough idea of the variation of genotypes for a given SNP in a given population. It gives an indication of how common this SNP is. Variants that have a very low minor allele frequency are also removed in data cleaning process. A MAF of less than one percent (<0.01) is considered low. This value distinguishes between a common and a rare variant.

4.16 Association Analysis

GWAS was performed using PLINK version 1.09 and R v.3.5.1

Logistic regression was used for all analyses, modeled each SNP additively. Logistic regression is used to determine the probability of an event with a binary outcome. In the instance, being a case-control study, the events are presence of SNPs and absence of SNPs. Each SNP was tested individually using the log-additive model. Additive genetic models, assumes that each minor allele has an equal effect on the trait of interest or risk of disease.

For the statistical analysis of GWAS, only SNPs that are significantly correlated with the cases of disease with a widely-used standard of p-value below $5x10^{-8}$, termed “genome-wide significance” are accepted (Fadista, 2016; IHC, 2005; Panagiotou, 2012). This is based on performing the Bonferroni correction which is obtained by dividing the observed p-value by the number of tests performed (1,000,000) (Dudbridge, 2008; Masahori, 2016). This stringent significance cutoff is to allow for the exemption of false-positive hits that may arise as a result of the large number of SNPs being examined at once.

Quantile-quantile (QQ) were generated using R (Fig 7.8). The QQ plot is a graphical representation of the deviation of the observed p values from the null hypothesis. P values for each SNP are sorted from largest to smallest and plotted against expected values from a theoretical χ2-distribution. If the observed values correspond to the expected values, all points are on or near the middle line between the x-axis and the y-axis. If some observed P values are clearly more significant than expected under the null hypothesis, points will move towards the y-axis. If there is an early separation of the expected from the observed this means that many moderately significant p values are more significant than expected under the null hypothesis. However, such results are more likely to be due to population stratification rather than a true positive association significant than expected under the null hypothesis.

Manhattan plots were also generated using R. Manhattan plots represent the p-values of the entire GWAS on a genomic scale (Fig 7.9). The p values are represented in genomic order by chromosome and position on the chromosome (x-axis). The value on the y-axis represents the −log10 of the p-value (equivalent to the number of zeros after the decimal point plus one). Because of local correlation of the genetic variants, arising from infrequent genetic recombination, groups of significant p values tend to rise up high on the Manhattan plot.

4.17 Results

Update Genome build

All genotypes were oriented to the forward strand and there is no risk of strand ambiguities.

Genotyping rate for samples

Starting with 1164 samples, eighty-one (81) samples were excluded due to their call rates being less than one percent.

Check for duplicate samples

Twenty-three (23) duplicated samples were excluded. Of the duplicated samples, the sample with a less genotyping quality was removed.

Check for sex discrepancies

Two samples were removed due to sex discrepancies. There were no uncertain phenotypes.

Check for heterozygosity

No sample failed the heterozygosity check. One thousand and fifty-five (1055) samples were left from all ethnicities.

Genotyping rate for variants

Starting with 548504 SNPs in both case and control datasets. Forty-four thousand, two hundred and twenty-five (44225) variants were excluded due to call rate being in less than one percent of samples.

Hardy-Weinberg Equilibrium check

Forty-seven variants were removed for failing Hardy-Weinberg equilibrium.

Minor allele frequency check

Variants were removed because their minor allele frequency was less than one percent. Two hundred and sixty-two thousand, three hundred and seventy-four SNPs were left.

Principal Component analysis

PCA was used to filter the data for European population. Different ethnic groups were clearly distinguished with the two first components. PCA identified five ethnic groups that matched well with the reference data. Over 95%, one thousand and seventy-five (1075) individuals were of European ancestry,

eight were of American ancestry, one individual had African ancestry, thirteen people had South Asian ancestry, sixteen had East Asian ancestry and twenty-one and thirty individuals had mixed and unknown ancestry respectively. All individuals of self-reported European ancestry corresponded to the reference data. Every individual from European ancestry was kept in the analysis (See Figures 7.6 and 7.7).

After adequate data cleaning and quality control, the data set included 262374 SNPs that were in 662 cases and 328 controls. The top signals were in rs17626937 (A/G SNP variation located near genes SNX18, LOC105378967 on chromosome 5, p=1.232x10-6) and rs4758408 (C/T SNP variation located in the LOC101927825 gene on chromosome 11, p=4.21×10−6) (See Figure 7. 10).

QQ plot

A QQ plot of the resulting p values of SNPs' association with paralytic poliomyelitis is presented in Figure 7.8. The p values of association observed in this study, corresponded to, and were no greater than expected. QQ plot showed no evidence of inflation and no evidence of a deviation from the null hypothesis.

Manhattan plot

The Manhattan plot of the p-values of the GWAS is presented in Figure7.9. It shows the results of the logistic regression analysis performed to identify SNPs associated with paralytic poliomyelitis in the study dataset. The upper line indicates the genome-wide significance threshold ($P=5x10^{-8}$) of association with paralytic poliomyelitis. In this study, no SNP achieved genome-wide significance threshold. Thus, no SNP was identified to be associated with paralytic poliomyelitis.

CHAPTER 5

5 Discussion and future direction

5.1 Discussion

This study involved major innovations in efforts to identify the genetic basis of poliomyelitis. First, to our knowledge, this was the first genome-wide analysis in humans of paralytic poliomyelitis. Prior studies have focused on a relatively limited set of candidate gene polymorphisms, many of which have showed mixed results, in that neither experiment-wide nor gene-wide statistical significance was observed in the primary single-SNP analyses or in secondary analyses of haplotypes or of imputed genotypes.

The work presented in this thesis is based first on the genome-wide association study (GWAS), where the entire genome is searched for variants that are association to the disease. In this study over 552,667 variants were investigated in 773 patients and 301 controls to search for variants (single nucleotide polymorphisms (SNPs) associated with paralytic polio. This method is suitable because it can detect genetic effects for common and complex diseases where the risk associated with any given candidate gene is relatively small. One drawback however, is that it requires large sample size to detect effects. This is due to the very high number of statistical tests of associations (minimum one per SNP) that are performed, requiring a very stringent threshold of statistical significance. It is important to note however, that association does not imply causation.

There are some studies that have reported HLA-variant associations with poliomyelitis traits from genome-wide association studies in the GWAS catalogue, some of these are vaccine-related poliomyelitis and post-polio syndrome traits. In their study, Lasch *et al.,* (1979) tested the HLA region and found that there was no statistically significant difference in the HLA antigens (HLA-AW19 and -B7) between children affected with paralytic poliomyelitis compared to their controls. In addition, a study by Kinnunen *et al***.**

(1986) found no association between central nervous system affectation following poliomyelitis and HLA-associated genetic factors (Kinnunen *et al.,* 1986).

Considering the involvement of variants, we hypothesized that some variants might play a functional role in paralytic polio. In this study, genotyped SNPs were tested across the whole genome region for variants of association with paralytic poliomyelitis using a case-control study design, where the subjects with paralytic polio corresponds to cases and is compared to spouses used as controls.

No individual SNPs that were associated with poliomyelitis with the widely accepted genome-wide significance level ($P \leq 5 \times 10^{-8}$) were found. This may suggest that our study may have been underpowered to detect effect sizes given the sample size. Thus, it appears that larger samples sizes may be needed to detect SNPs associated with poliomyelitis.

This may also suggest that the common variants in our data do not play a role in the genetic disposition to paralytic polio.

Another type of genetic study, which investigates associations between genetic variation within pre-specified genes of interest and phenotypes or disease states is the candidate gene association studies. The chosen genes are usually based on literature review of previous studies or educated guess based on known pathways and prior knowledge of the genes biological functional impact on the trait or disease question. Following GWAS, focus was placed on SNPs based on previous findings. Previous findings showed that mutations in PVR genes and polymorphisms in Ala67Thr are associated with increased risk of paralytic poliomyelitis- both wild-type and vaccine-related, post-polio syndrome as well as susceptibility to poliovirus infection. One study had also shown that KCNJ4, a potassium inwardly rectifying channel subfamily J member 4 on chromosome 22 was associated with poliomyelitis. Those results were however not replicated in this study as SNPs associated with those regions did not show a genome-wide statistical significance.

Understanding the epigenetic underpinnings of poliomyelitis is pertinent to interpreting the study results relative to plausible genetic explanation of poliomyelitis in extant epidemiological literature (e.g**.** twins studies). In particular, the epigenome can accommodate environmental influences in the form of chemical and protein modifications of chromatin that do not involve alterations in the DNA sequence. Such modifications involving DNA methylation (mDNA) at cytosine sites can alter DNA binding to regulatory proteins (Sweat *et al.,* 2013), and are transferable from parent to offspring. As a result, observable differences between cases and control based on twin studies may be consequent to epigenetic modifications due to exposure to the same environmental factors rather than differences in genetic composition that is focused in the present GWAS of poliomyelitis.

The overrepresentation of females in the present study is consistent with findings in some previous studies (Rindge, 1955) and mainly attributed to several social factors such as increasing age, exposure to vaccination (WHO, 2018) rather than genetics thereby limiting inferences on the role of sex on our study findings.

The strength of this study includes the strict case definitions and procedures used to confirm case and control status in that only those with paralysis were enrolled as cases. All clinical characteristics were supported with documentation, including medical records. Selection of case patients and control patients was as unbiased as could be achieved. The genotyping, data cleaning and quality control process was stringent and thorough.

A limitation of this study may be related to the sample size. Thus methodical technique that allows increased sample size using higher ratio of healthy controls to cases of poliomyelitis may be employed to achieve better study power and reduce type 2 error (Lewallen & Courtright, 1998). This can be achieved by using controls from published GWAS studies.

5.2 Future directions

An understanding of the most probable genetic mechanism for poliomyelitis and infectious diseases in general is very crucial. In this regard, common variants are of great significance, hence an extension of the GWAS approach with microarrays to improve coverage and analysis of larger sample sizes might be the best way forward (Hill, 2012).

Future studies should also explore the possible role of epigenetics in the predisposition of individuals to poliomyelitis.

Closely related is that rare variants may be an explanatory mechanism for the missing component for most susceptibility with larger effect sizes (Johnson *et al.,* 2007). Given that GWAS microarrays results in poor assessments of rare variants (Hill, 2012), sequencing approaches is a viable option to advance the frontier in genetics of poliomyelitis (Hill, 2012)

6 Appendices

6.1 Appendix I: Flow chart of conceptual summary of GWAS process

#Update genome built

Align the strands (sequence of SNPs/alleles) in the dataset to the appropriate genotyping chip for correct alignment and variant positioning;

This entails updating the chromosome, position and strands of SNPS in the binary .ped file using the strand and position from the files in the InfiniumCoreExome-24v1-1_A and HumanCoreExome-24v1-0_A which was used for the genotyping.

It provides for correct identification of SNPs.

 # Convert MAP/PED to binary PLINK files

Change the file that describes the SNPS (MAP file) and the file that describes the individual with its genetic material (PED file) to a more compressed and efficient form. Three files are generated (.bim,.bed,.fam).

 # Check for discordant sex information

Check for discrepancies between the sex of the individuals recorded in the dataset (ped file) and their sex based on X chromosome heterozygosity and homozygosity (genomic data)

 # Check for genotyping rate for individuals and markers

Check for the rate at which each SNP (marker) are genotyped (or missing) across individuals and the rate at which each individual (sample) is genotyped (or missing) for SNPs.

Genotyping rate refers to the rate at which SNPs are called i.e correctly identified. It is the rate at which the DNA sequence was correctly determined.

SNP genotyping call rate - the proportion of samples with a genotype call for each marker. it is an indication of marker quality. A call rate of 99% is usually considered appropriate.

Individual genotyping - the proportion of SNPs with a genotype call for each sample. It is an indication of sample quality. A call rate of 99% is usually considered appropriate.

Missing-ness refers to the proportion of samples or SNPS that were NOT genotyped

 # Filter1

Include only SNPs whose minor allele have a frequency that are above 1%;
Include only SNPs whose rate of missingness are 10% and below; i.e.genotyping rate of 90% and above.
Include only samples whose rate of missingness are 10% and below. i.e genotyping rate of 90% and above.

 # Filter2

Repeat the filter 1 process but reduce the rate of missingness for samples and markers to 5%.

#Filter3

Conduct the filter processing again but reduce the rate of missingness for samples and markers to 1%.
The step-wise filtering allows for a better retainment of samples while still maintaining a high level of filtering.

Repeat check for differential genotyping rate (case vs control)

 # Find and identify and remove bad samples and snps

 # Check for related and duplicate samples

As part of the QC procedure, related and duplicate samples need to be removed. This can be done by creating a genome file in plink. This examines the distance in genetic terms between samples called the identity-by-state (IBS) distance. It is expected to be zero if samples are totally unrelated.

#Population Stratification

This accounts for the degree of systemic variability in the dataset (population). This step is to ensure that differences identified between cases and controls are as a result of an inherent genetic difference between them and not as a result of a systematic difference in their ancestry. Principal component analysis will also be used to determine this.

 # Test for association

This is the co-occurrence of a genotype with a phenotypic trait more often than would be expected by chance.

Test for a difference in the frequency of an allele between cases and controls.

This is carried out using various statistical tests including logistic regression.

6.2 Appendix II: Genome-wide association study Script

#Update Genome build

```
$script_p/External_scripts/update_build.sh $rgeno_p/ICE/Polio_ICE24v1_1_20170711
$arrays_p/InfiniumCoreExome-24v1-1_A/Oxford_Strand/InfiniumCoreExome-24v1-1_A-b37.strand
$preqc_p/ICE/PolioICE_update_03082018

$script_p/External_scripts/update_build.sh $rgeno_p/HCE/Polio_HumanCoreExome24v1_20150129
$arrays_p/HumanCoreExome-24v1-0-A/Oxford_Strand/HumanCoreExome-24v1-0_A-b37.strand
$preqc_p/HCE/PolioHCE_update_08082018
```

 # Creation of binary files

```
plink \ --file $rgeno_p/ICE/Polio_ICE24v1_1_20170711 \ --make-bed \ --out
$rgeno_p/ICE/Polio_ICE24v1_1_20170711

plink \ --file $rgeno_p/HCE/Polio_HumanCoreExome24v1_20150129 \ --make-bed \
--out $rgeno_p/HCE/Polio_HumanCoreExome24v1_20150129
```

 # Check for duplicated samples in datasets and retain samples with better genotyping rate

```
plink \  --bfile $preqc_p/ICE/PolioICE_update_03082018 \   --missing \ --out
$preqc_p/ICE/PolioICE_update_03082018

plink \  --bfile $preqc_p/HCE/PolioHCE_update_08082018 \   --missing \   --out
$preqc_p/HCE/PolioHCE_update_08082018

plink \  --bfile $preqc_p/HCE/PolioHCE_update_08082018 \  --extract
$preqc_p/HCE/hce_list_unique_SNPs_to_keep.txt \  --make-bed \
 --out $preqc_p/HCE/PolioHCE_No_DupVars
```

```
plink \  --bfile $preqc_p/ICE/PolioICE_update_03082018 \  --extract

$preqc_p/ICE/ice_list_unique_SNPs_to_keep.txt \  --make-bed \

 --out $preqc_p/ICE/PolioICE_No_DupVars

awk 'FNR==NR{a[NR]=$6;next}{$6=a[FNR]}1'

/home/tinuke/srv/Gene/Polio_GWAS_QC/pre/HCE/PolioHCE_update_0308201.fam

/home/tinuke/srv/Gene/Polio_GWAS_QC/pre/HCE/PolioHCE_No_DupVars.fam >

/home/tinuke/srv/Gene/Polio_GWAS_QC/pre/HCE/PolioHCE_No_DupVars_with_phenotypes.fam

awk 'FNR==NR{a[NR]=$6;next}{$6=a[FNR]}1'

/home/tinuke/srv/Gene/Polio_GWAS_QC/pre/ICE/PolioICEupdateid.fam

/home/tinuke/srv/Gene/Polio_GWAS_QC/pre/ICE/PolioICE_No_DupVars.fam >

/home/tinuke/srv/Gene/Polio_GWAS_QC/pre/ICE/PolioICE_No_DupVars_with_phenotypes.fam
```

 # Merge datasets

```
plink\  --bfile

/home/tinuke/srv/Gene/Polio_GWAS_QC/pre/HCE/PolioHCE_No_DupVars_with_phenotypes \

 --bmerge /home/tinuke/srv/Gene/Polio_GWAS_QC/pre/ICE/PolioICE_No_DupVars_with_phenotypes \

 --make-bed \ --out /home/tinuke/srv/Gene/Polio_GWAS_QC/QC/Polio_HCE_ICE
```

 # Check for SNP genotyping rate to retain only good SNPs

```
plink --bfile $qc_p/Polio_HCE_ICE --geno 0.1 --make-bed --out $qc_p/Polio_HCE_ICE_QC1
```

 # Check for samples genotyping rate to retain only good samples

```
plink --bfile $qc_p/Polio_HCE_ICE_QC1 --mind 0.1 --make-bed --out $qc_p/Polio_HCE_ICE_QC2
```

 # Sex check

plink --bfile $qc_p/Polio_HCE_ICE_QC2 --check-sex --out $qc_p/Polio_HCE_ICE_QC2

grep PROBLEM $qc_p/Polio_HCE_ICE_QC2.sexcheck

 # Heterozygozity check

plink --bfile $qc_p/Polio_HCE_ICE_QC3 --het --out $qc_p/Polio_HCE_ICE_QC3

het <- read.table(paste(qc_p, "/Polio_HCE_ICE_QC3.het", sep=""), header=T)

het$HET <- (het$N.NM. - het$O.HOM.)/ het$N.NM.

jpeg(paste(qc_p, "/Polio_HCE_ICE_QC3_Heterozygosity_rate_plot.jpeg", sep=""))

hist(het$HET , breaks=100)

dev.off()

het$to_keep <- (het$HET < 3*(mean(het$HET) + sd(het$HET))) | (het$HET > 3*(mean(het$HET) -

sd(het$HET)))

table(het$to_keep)

 # Identification and removal of duplicated samples

plink --bfile $qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2 --genome --out

$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2

$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2.log2

plink --bfile $qc_p/Polio_HCE_ICE_QC4 --missing --out $qc_p/Polio_HCE_ICE_QC4

```
$script_p/Polio_GWAS_QC_S7_find_duplicates.r plink --bfile $qc_p/Polio_HCE_ICE_QC4 --remove
$qc_p/Polio_HCE_ICE_QC4_list_duplicated_samples.txt --make-bed --out $qc_p/Polio_HCE_ICE_QC5
```

 # Genotyping to retain only good samples

```
plink --bfile $qc_p/Polio_HCE_ICE_QC5 --mind 0.05 --make-bed --out  $qc_p/Polio_HCE_ICE_QC6

mv $qc_p/Polio_HCE_ICE_QC6.log $qc_p/Polio_HCE_ICE_QC6.log1
```

 # Genotyping to retain only good SNPs

```
plink --bfile $qc_p/Polio_HCE_ICE_QC6 --geno 0.05 --make-bed --out  $qc_p/Polio_HCE_ICE_QC7

mv $qc_p/Polio_HCE_ICE_QC7.log $qc_p/Polio_HCE_ICE_QC7.log1
```

 # Get descriptive data

```
plink --bfile $qc_p/Polio_HCE_ICE_QC8 --freq --hardy --missing --out $qc_p/Polio_HCE_ICE_QC8

plink --bfile $qc_p/Polio_HCE_ICE_QC8 --freqx --out $qc_p/Polio_HCE_ICE_QC8
```

 # Ethnicity check (Principal Component Analysis (PCA))

```
plink --bfile  $qc_p/Polio_HCE_ICE_QC3 --indep-pairwise 50 5 0.2 --out $qc_p/Polio_HCE_ICE_QC3

plink --bfile $qc_p/Polio_HCE_ICE_QC3 --extract $qc_p/Polio_HCE_ICE_QC3.prune.in --make-bed --out
$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2

cut -f 1,4,4,2 $qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2.bim | head

awk '{print $1,$4,$4, $2}' $qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2.bim  >
$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2_range.txt
```

```
plink --bfile $kg_p/1000G_chrall.ex_maf_05.pruned --extract range

$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2_range.txt --make-bed --out

$kg_p/SNP_Subsets/1000G_SNP_in_Polio_Prune50_5_0.2

plink --bfile $kg_p/SNP_Subsets/1000G_SNP_in_Polio_Prune50_5_0.2 --bmerge

$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2 --make-bed --out

$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_ 1KG

plink --bfile $kg_p/SNP_Subsets/1000G_SNP_in_Polio_Prune50_5_0.2 --exclude

$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_1KG-merge.missnp  --make-bed --out

$kg_p/SNP_Subsets/1000G_SNP_in_Polio_Prune50_5_0.2_clean

plink --bfile $qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2 --exclude

$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_1KG-merge.missnp  --make-bed --out

$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2_clean

plink --bfile $kg_p/SNP_Subsets/1000G_SNP_in_Polio_Prune50_5_0.2_clean --bmerge

$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2_clean --make-bed --out

$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_1KG_2

plink --bfile $qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_1KG_2  --geno 0.05 --make-bed --out

$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_1KG_3

plink --bfile

/home/tinuke/srv/Gene/Polio_GWAS_QC/QC/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_1KG_3 –

cluster  --mds-plot 3  --out

/home/tinuke/srv/Gene/Polio_GWAS_QC/QC/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_1KG_3
```

```
plink --bfile $qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_1KG_3 --pca --out
$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_1KG_3

$script_p/Polio_GWAS_QC_S6_find_ethnic_outlisers.r

plink --bfile $qc_p/Polio_HCE_ICE_QC3 --remove
$qc_p/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_1KG_Ethnic_outliers.txt --make-bed --out
$qc_p/Polio_HCE_ICE_QC4

pca <- read.table(paste(qc_p,
"/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_1KG_3.mds",sep=""),header=T)

kg_ph<- read.xls(paste(kg_p, "/Phenotypes/20130606_sample_info.xlsx", sep=""), sheet=1 )

kg_p2 <- kg_ph[,c("Sample","Population")]

kg_p2$Ethnicity <- ……

colnames(kg_p2)[1]<-"IID"

kg_p2$Population <- NULL

Polioeth2019 <- read.table(paste(main_p, "/Polio_Phenotypes/POLIOHCEpheno.txt", sep=""), sep=" ",
header=TRUE)

colnames(Polioeth2019)<-c("IID", "Ethnicity")

eth<-rbind(kg_p2, Polioeth2019)

n<-merge(pca, eth, all.x=T, all.y=F)

m[m$Ethnicity=="AFR","col"]<-"yellow"

m[m$Ethnicity=="EUR","col"]<-"blue"
```

```
m[m$Ethnicity=="AMR","col"]<-"red"

m[m$Ethnicity=="SAS","col"]<-"purple"

m[m$Ethnicity=="EAS","col"]<-"green"

m[is.na(m$col),"col"]     <-"black"

jpeg(paste(qc_p, "/Polio_HCE_ICE_QC3_pruned_50_5_0.2merged_1KG_2_mds_EthAll_C1_C2.jpeg",

sep=""))

plot(m$C1, m$C2, col=m$col)

dev.off()

#### 1-EUROPEAN 2-MIXED ANCESTORY 3-AFRICAN 4-AMERICAN INDIAN/ALASKA NATIVE 5-EAST

ASIAN/SOUTH ASIAN 6-HAWAIAN/OTHER PACIFIC ISLANDER 7-OTHER 0-UNKNOWN

m<-m[m$Ethnicity==1 | m$Ethnicity=="AFR" | m$Ethnicity=="EUR" | m$Ethnicity=="AMR"  |

m$Ethnicity=="SAS"  | m$Ethnicity=="EAS" ,]
```

#Hardy-Weinburg Equilibrum (HWE) Check

```
plink --bfile $qc_p/Polio_HCE_ICE_QC7 --hwe 0.000005 --make-bed --out  $qc_p/Polio_HCE_ICE_QC8
```

#Subset participants by ethnicity

```
plink \  --bfile $qc_p/Polio_HCE_ICE_QC8 \ --keep

$pheno_p/Polio_HCE_ICE_FID_IID_Ethnicity_1_EUR.txt \   --make-bed \   --out

$qc_p/Polio_HCE_ICE_QC8_EUR
```

 # Association analysis (Logistic Regression)

```
plink --bfile /home/tinuke/srv/Gene/Polio_GWAS_QC/QC/Polio_HCE_ICE_QC8_EUR  --logistic --out
/home/tinuke/srv/Gene/Polio_GWAS_Results/From_QCed_data/Polio_HCE_ICE_QC8_maf0.01_EUR
```

 # Quantile-Quantile (QQ) plot

```
dataset1=read.table(file("Polio_HCE_ICE_QC8_maf0.01_EUR.assoc.logistic"), header=T)

colnames(dataset1) <- c("CHR","SNP","BP","A1","TEST","NMISS","OR","STAT","P")

library(qqman)

png("POLIO_qq_plot_imputed.png", height=3600, width=6000, res=600)

qq(dataset1$P, main = "Q-Q plot of pre-imputed GWAS p-values", xlim = c(0, 6))

garbage <- dev.off()
```

 # Manhattan plot

```
dataset1=read.table(file("Polio_HCE_ICE_QC8_maf0.01_EUR.assoc.logistic"), header=T)

colnames(dataset1) <- c("CHR","SNP","BP","A1","TEST","NMISS","OR","STAT","P")

library(qqman)

png("POLIO_man_plot_imputed.png", height=3600, width=6000, res=600)

manhattan(dataset1, main = "Manhattan Plot", ylim = c(0, 20), cex = 0.6, cex.axis = 0.9, col = c("blue4",
"orange3"))

garbage <- dev.off()
```

7. Tables and Figures

7.1 Polio Belt

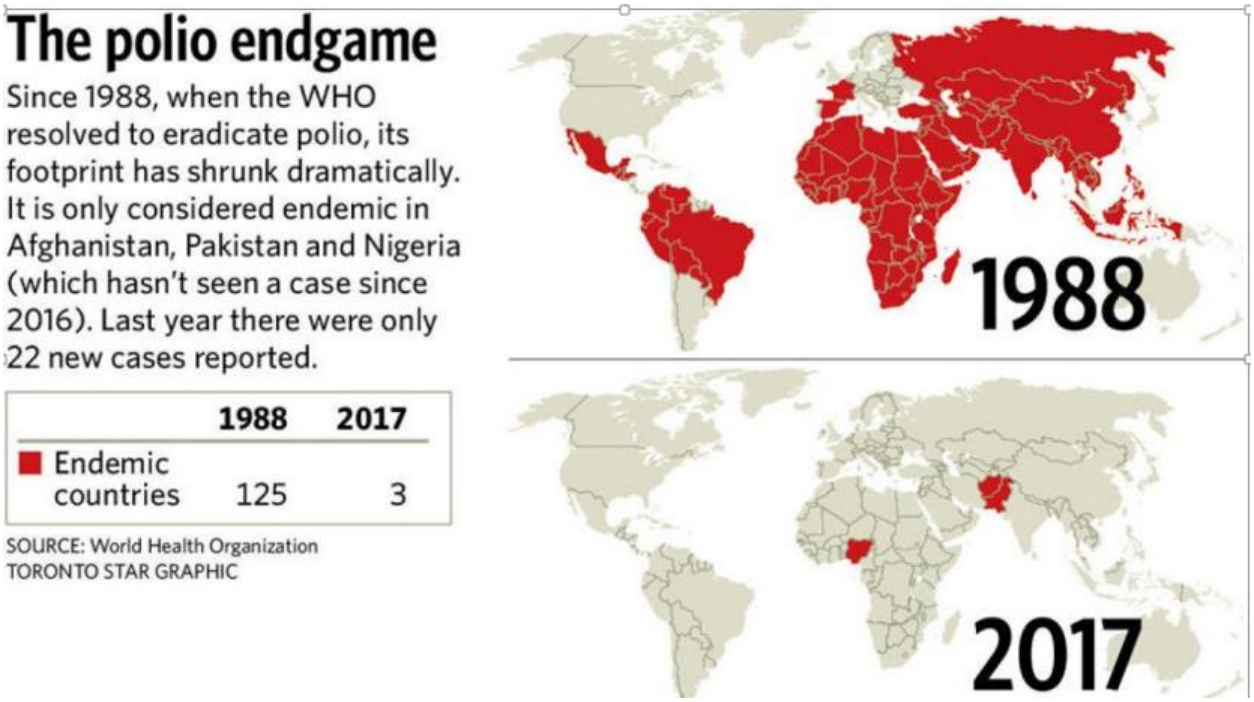


Figure adapted from WHO, 2019; GPEI, 2019

7.2 Pathogenesis

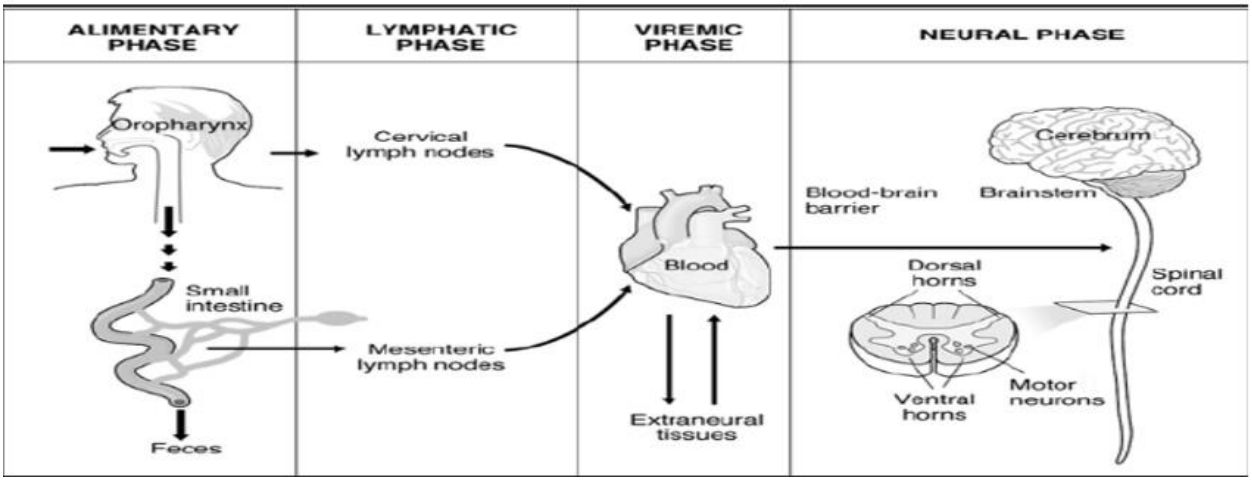


Figure adapted from Blondel et al., 2005

7.3 Schematic diagram of Infinium Assay protocol

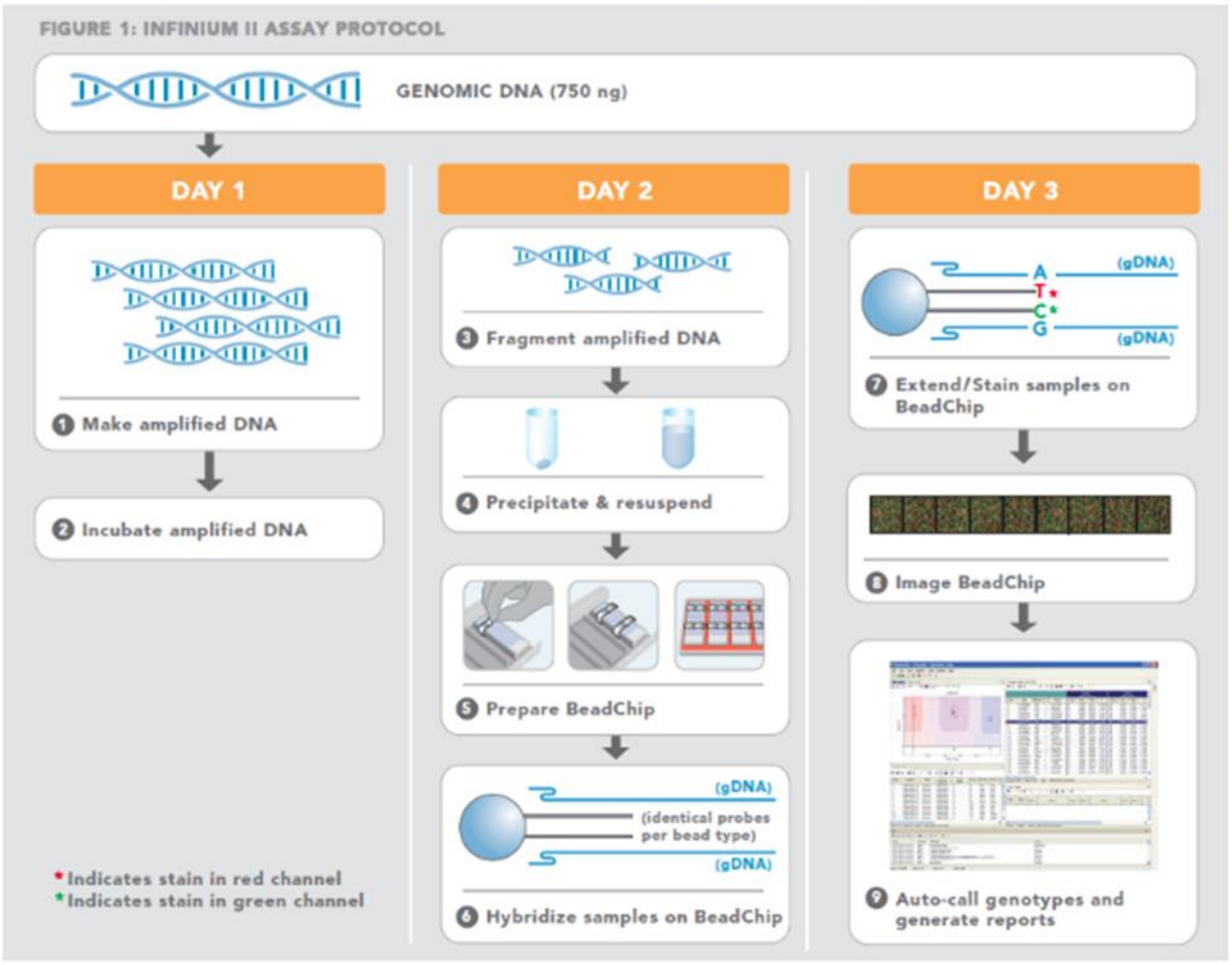
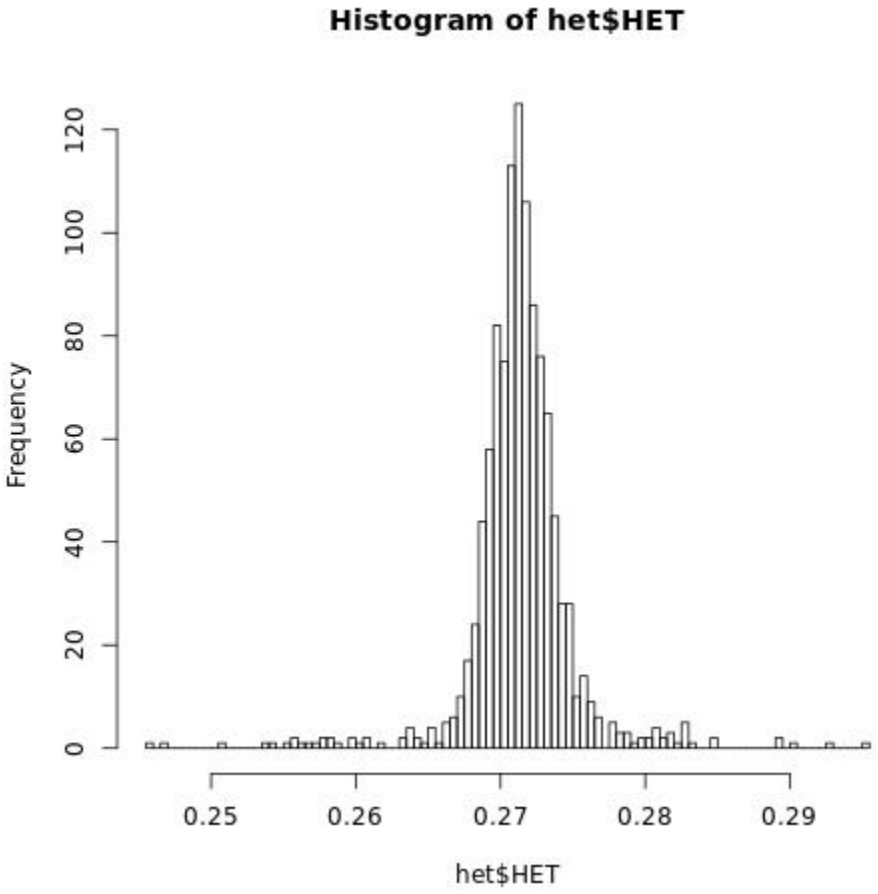


Figure adapted from UCDavis, 2019

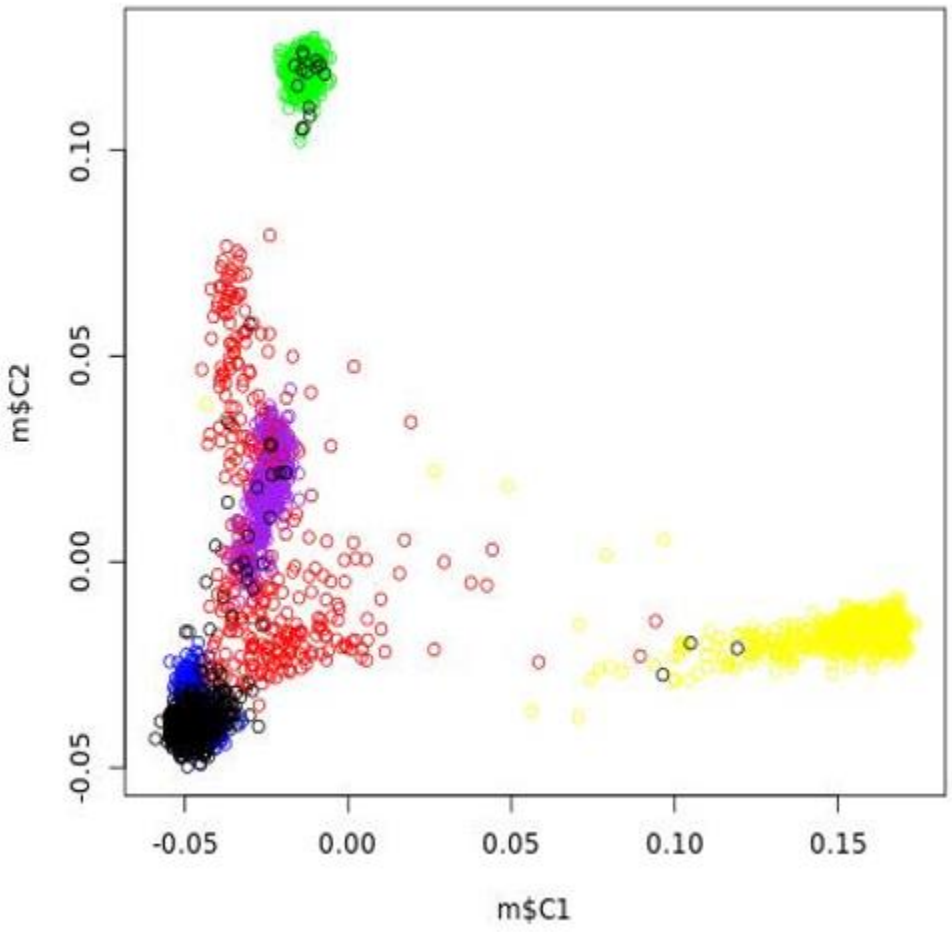
7.4 Heterozygosity histogram



het$HET - Observed heterozygosity rate per individual

Frequency - Proportion of missing SNPs per individuals

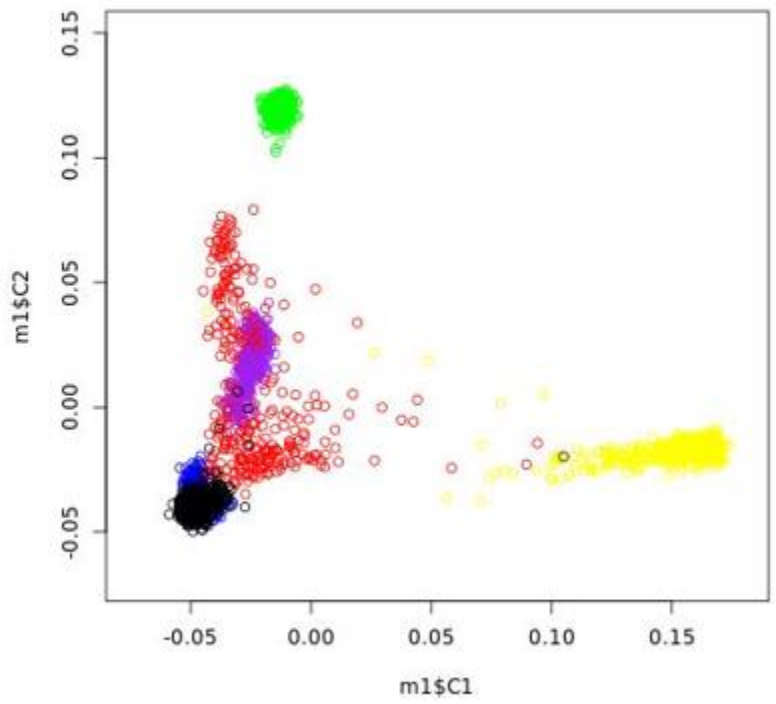
7.5 Principal component analysis for all ethnicities with 1000 Genomes population data



m$C1 – Principal Component 1; m$C2 – Principal Component 2

	Participants in study data
	African Ancestry from 1000 Genomes data
	Hispanic Ancestry from 1000 Genomes data
	East Asian Ancestry from 1000 Genomes data
	European Ancestry from 1000 Genomes data
	South Asian Ancestry from 1000 Genomes data

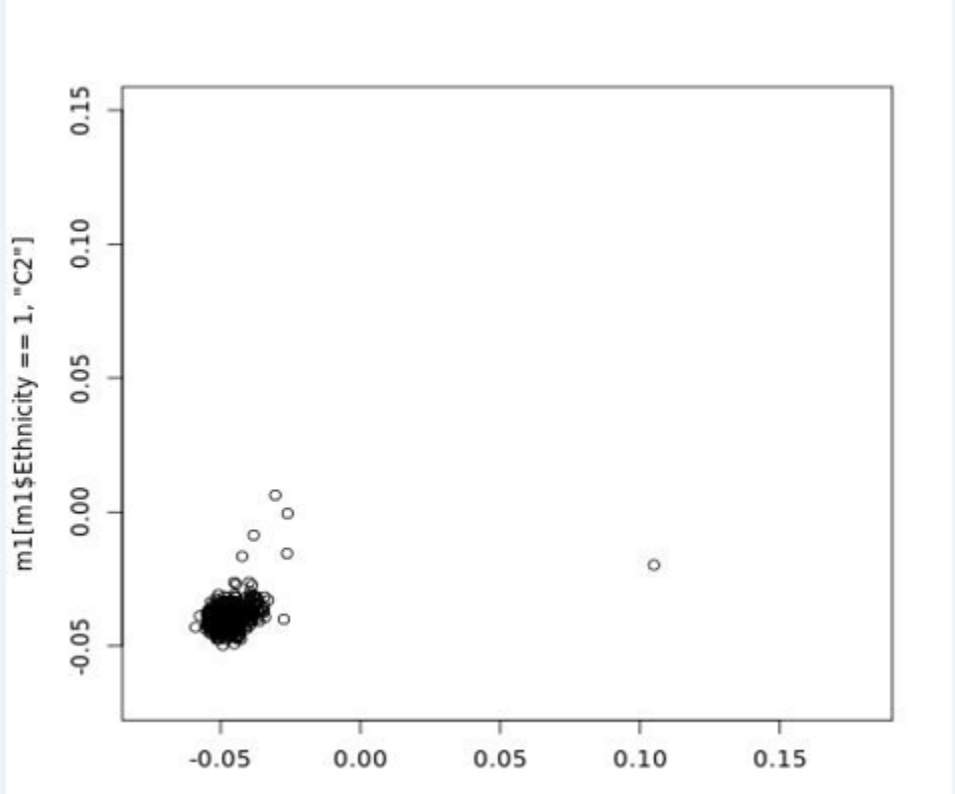
7.6 Principal component analysis for European population with 1000 Genomes population



m$C1 – Principal Component 1; m$C2 – Principal Component 2

	Participants of European ancestry in study data
	African ancestry from 1000 Genomes data
	Hispanic ancestry from 1000 Genomes data
	East Asian ancestry from 1000 Genomes data
	European ancestry from 1000 Genomes data
	South Asian ancestry from 1000 Genomes data

7.7 Principal component analysis for self-reported European population only

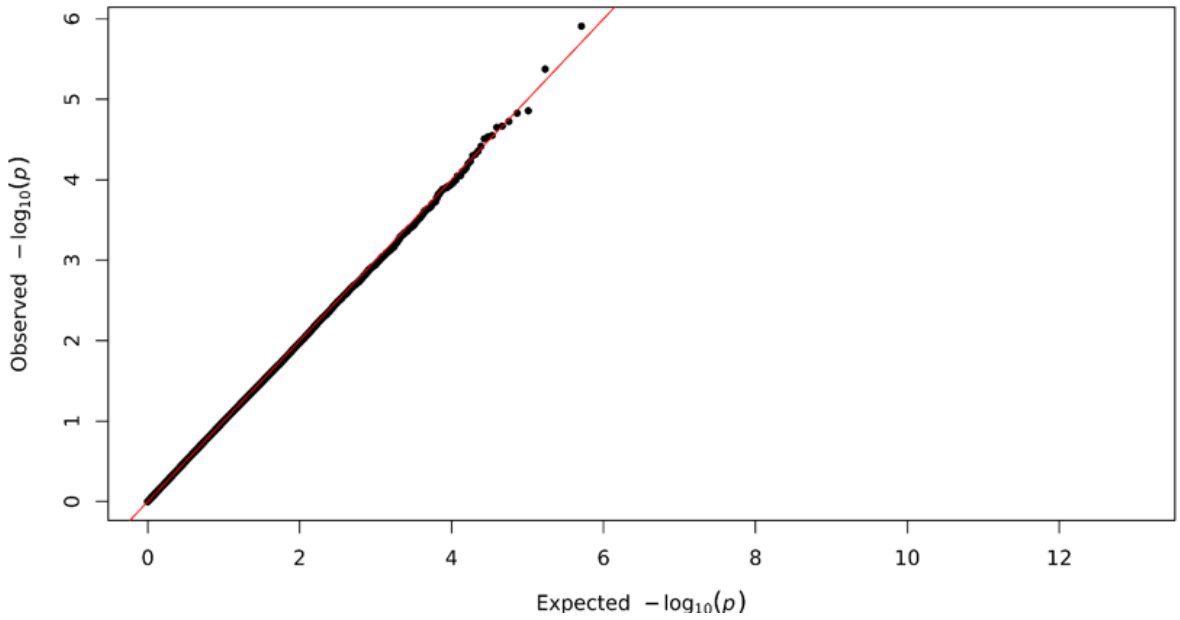


m1[m1$Ethnicity==1, “C1”]

m1[m1$Ethnicity==1, “C1”] – Principal Component 1

m1[m1$Ethnicity==1, “C2”] – Principal Component 2

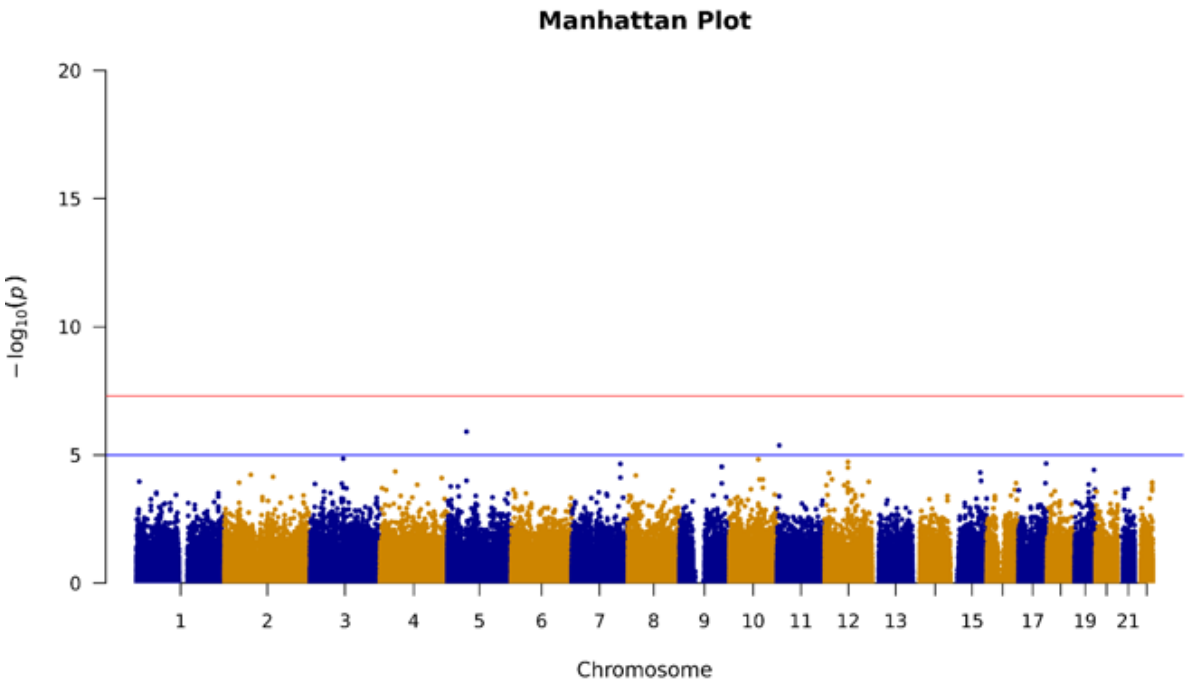
7.8 QQ plot of GWAS



Expected $-\log_{10}(p)$ value - log 10 of p value of association expected

Observed –log10 (p) value - log 10 of p value of association observed in study data

7.9 Manhattan plot of GWAS



Chromosome – Chromosome for each SNP

$-\log_{10}(p)$ - −log10 of the p-value of association for each SNP

Red line – line of genome-wide significance p-value ($p= 5x10^{-8}$)

7.10 Table of variants with the smallest p-value

CHR	SNP	BP	A1	TEST	NMISS	OR	STAT	P
5	rs17626937	53914475	A	ADD	990	0.5493	-4.85	1.232e-06
11	rs4758408	6348639	T	ADD	979	0.4745	-4.601	4.211e-06

CHR-chromosome; SNP- single nucleotide polymorphism identifier; BP-base-pair position; A1 -Tested allele (minor allele) ; TEST – code for test (Additive); NMISS – number of non-missing individuals included in analysis; OR-odds ratio; STAT- Coefficient t-statistic; p- p--value for t-statistic

7.11 List of genes reported as being associated with paralytic poliomyelitis in literature

Name/Gene ID	Description	Location	Aliases	MIM
KCNJ4 ID: 3761	potassium inwardly rectifying channel subfamily J member 4 [*Homo sapiens* (human)]	Chromosome 22, NC_000022.11 (38426327..38455199, complement)	HIR, HIRK2, HRK1, IRK-3, IRK3, Kir2.3	600504

Name/Gene ID	Description	Location	Aliases	MIM
PVR ID: 5817	PVR cell adhesion molecule [*Homo sapiens* (human)]	Chromosome 19, NC_000019.10 (44643798..44666162)	CD155, HVED, NECL5, Necl-5, PVS, TAGE4	173850

Table adapted from NCBI (2019)

7.12 Table of epidemiological studies exploring genetic risk of poliomyelitis

Table of epidemiological studies exploring genetic risk of poliomyelitis

Study	Design	Participant	Genotype	Statistics	Inference
Rekand et al., (2002)	Case-Control	Six (6) polio cases Vs Thirteen (13) controls	Fc-YrIIIA V/V	Odds Ratio=1	The FcγRIIIA V/V genotype may lower the risk for contracting acute poliomyelitis through better clearance of poliovirus
Kindberg et al, (2009)	Case-Control	Nine (9) vaccine-associated and Six (6) wild-type polio cases Vs Seventy-one (71) controls	Ala67Thr SNP	Odds Ratio = 1.6	Ala67Thr mutation in the poliovirus receptor is a possible risk factor for the development of vaccine-associated and wild-type paralytic poliomyelitis.
Bhattacharya et al, (2014)	Case-Control	One hundred and ten (110) cases Vs two hundred (200) controls	Ala67Thr SNP	Ala67Thr mutation was detected in 45.46% of progressive PPS and 10% of control subjects. Odds Ratio was not reported.	Changes in the PVR gene may lead to post-polio syndrome (PPS).

FcγR - Fragment crystallizable gamma receptor; Ala67Thr – Alanine67thLocusThreonine

7.13 Table of study population

	CASE	CONTROL
NUMBER OF MALES	226	233
NUMBER OF FEMALES	547	158
MEAN AGE AT RECRUITMENT (YRS)	69.4 (10.07)	67.2 (14.5)
MEAN AGE AT DIAGNOSIS (YRS)	5.6 (5.6)	

7.14 Glossary of words

GENOMICS: Genomics is the study of the structure, function and mapping of genomes, including interactions of those genes with each other and with the environment. In contrast to genetics, which refers to the study of individual genes and their roles in inheritance, genomics aims at the collective characterization and quantification of all of an organism's genes.

GENOME: A genome is an organism's complete set of DNA, including all of its genes. Each genome contains all of the information needed to build and maintain that organism. In humans, a copy of the entire genome consists of more than 3 billion DNA base pairs and is contained in all cells that have a nucleus.

SINGLE-NUCLEOTIDE POLYMORPHISM: A single-nucleotide polymorphism (SNP) is a substitution of a single nucleotide that occurs at a specific position in the genome. For example, if at a specific base position in the human genome, two different nucleotides may occupy the position, this means that there two possible nucleotide variations. They are said to be alleles for this position. This variation is considered a SNP when it is present at a level of more than one percent (1%) in the population. SNPs are the most common type of genetic variation among people. Most SNPs have no effect on health or development however, some SNPs underline differences in our susceptibility to or protection against diseases as well as resistance or response to vaccinations and therapy; thus acting as biological markers.

GENOME-WIDE ASSOCIATION STUDY: Genome-wide association study (GWAS) is an observational study of the whole genome-wide set of genetic variants in different individuals to see if any variant typically SNPs is associated with a trait. If one type of the variant (one allele) is more frequent in people with the disease than those without the disease, that variant is said to be associated with the disease. The associated SNPs are then considered to mark a region of the human genome that may influence the risk of disease.

Flinders Technology Associates (FTA) cards: These are cotton-based, cellulose paper containing chemicals that burst cells, denature proteins and protect DNA, leaving a sample suitable for molecular identification without the risk of disease contamination. They are a user-friendly way to send samples from the field to the laboratory for the identification of avian pathogens or biological analysis

References

- Addair J, Snyder LH (1942) Evidence for an autosomal recessive gene for susceptibility to paralytic poliomyelitis. Studies in human inheritance XXI*. J Hered 33: 306-309*.
- Albright FS, Orlando P, Pavia AT, Jackson GG, Cannon Albright LA (2008). Evidence for a heritable predisposition to death due to influenza. J Infect Dis.; 197:18–24.
- Alexander JP, Ehresmann K, Seward J et al. (2009) Transmission of Imported Vaccine-Derived Poliovirus in an Under-vaccinated Community in Minnesota. JID; 199:391-7.
- Allison AC (1954). Protection afforded by sickle-cell trait against sub-tertian malarial infection. Br. Med. J. 1, 290–29410.1136/bmj.1.4857.290doi:10.1136/bmj.1.4857.290
- Ananthakrishnan S, Puri RK, Badrinath R (1988). Poliovirus antibody titre in cord blood and its correlation with antenatal and natal factors*. Indian Pediatr. Nov;25 (11):1033-9.*
- Arévalo-Herrera M, Castellanos A, Yazdani SS, Shakri AR, Chitnis CE, Dominik R, Herrera S (2005). Immunogenicity and protective efficacy of recombinant vaccine based on the receptor-binding domain of the Plasmodium vivax Duffy binding protein in Aotus monkeys. *Am J Trop Med Hyg. Nov;73(5 Suppl):25-31*
- Aycock, WL (1942) Familial aggregation in poliomyelitis. *Am. J. M. Sc. 203: 452-465.*
- Bhattacharya, S.K., Sarkar, A., & Sengupta, S.D. (2014). Ala67Thr Mutation in the Human Polio Virus Receptor (PVR) Gene inPost-Polio Syndrome Patients. *J Microb Biochem Technol 2014, 6:4 http://dx.doi.org/10.4172/1948-5948.1000141*
- Blondel B, Colbère-Garapin F, Couderc T, Wirotius A, Guivel-Benhassine F (2005). Poliovirus, pathogenesis of poliomyelitis, and apoptosis. *Curr Top Microbiol Immunol; 289:25-56..*
- Burgner D, Jamieson SE, Blackwell JM (2006) Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? *Lancet Infect Dis. 6:653-663. 10.1016/S1473-3099(06)70601-6*

- Bush WS, Moore JH (2012). Chapter 11: Genome-wide association studies. *PLoS Comput Biol.;8(12):e1002822. doi: 10.1371/journal.pcbi.1002822.*
- Carnie JA, Lester R, Mann R et al., (2007) Public Health Response to Imported Case of Poliomyelitis, Australia, 2007. *Emerging Infectious Diseases; 15(11): 1733-37*.
- Carrington M, Nelson GW, Martin MP (1999). HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science; 283:1748–1752.*
- Casanova JL (2015). Human genetic basis of interindividual variability in the course of infection. *Proc Natl Acad Sci U S A. Dec 22;112(51):E7118-27. doi: 10.1073/pnas.1521644112.*
- Casanova JL, Tardieu M, Abel L (2010). Genetic predisposition to herpetic meningo-encephalitis in children. *Bull Acad. Natl Med. Jun; 194 (6):915-22.*
- Casselbrant ML, Mandel EM, Fall PA (1999). The heritability of otitis media: a twin and triplet study. JAMA. 1999; 282: 2125–2130.
- Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NT, Thuong NT et al., (*2008)*. The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. *PLoS Pathog. Mar 28;4(3):e1000034. doi: 10.1371/journal.ppat.1000034.*
- Chakravartti MR, Vogel F (1973). A twin study on leprosy by Georg Thieme; Stuttgart: 1973. *Acta Geneticae Medicae Et Gemellologiae, 24(1-2), 179-179. doi:10.1017/S1120962300022186*
- Chapman SJ, Khor CC, Vannberg FO, Maskell NA, Davies CW, Hedley EL, Segal S, et al., (2006). PTPN22 and invasive bacterial disease. *Nat Genet. May;38(5):499-500.*
- Comstock GW (1978). Tuberculosis in twins: a re-analysis of the Prophit survey. *Am Rev Respir Dis. Apr;117(4):621-4.*
- Cooper JD, Walker NM, Smyth DJ, Downes K, Healy BC, Todd JA (2009) Type I Diabetes Genetics Consortium. *Genes Immun. Dec; 10 Suppl 1():S85-94*

- Dudbridge F, Gusnanto A(*2008)* Estimation of significance thresholds for genome-wide association scans. *Genet Epidemiol. Apr; 32(3):227-34*
- el-Sayed N, el-Gamal Y, Abbassy AA, Seoud I, Salama M, Kandeel A, et al., (2008) Monovalent type 1 oral poliovirus vaccine in newborns. *N Engl J Med. 2008 Oct 16;359(16):1655-65. doi: 10.1056/NEJMoa0800390*
- Fadista J, Manning AK, Florez JC, Groop L. The (in) famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet. 2016 Aug; 24 (8):1202-5. doi: 10.1038/ejhg.2015.269.*
- Gedda L, Rajani G, Brenci G, Lun MT, Talone C, Oddi G (1984). Heredity and infectious diseases: a twin study. *Acta Genet Med Gemellol (Roma);33(3):497-500.*
- Gene [Internet] (2004). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2019 12 17]. Available from: https://www.ncbi.nlm.nih.gov/gene/
- Government of Canada (2019). Polio: vaccine advice. Retrieved on 22/11/2019 rom https://travel.gc.ca/travelling/health-safety/travel-health-notices 171
- Global Polio Eradication *Initiative [GPEI], (2019) Polio Today.* Retrieved on 23/09/2019 from http://polioeradication.org/polio-today/
- Guarino K, Voorman A, Gasteen M, Stewart D, Wenger J (2017). Violence, insecurity, and the risk of polio: A systematic analysis. *PLoS One. Oct 11;12(10):e0185577.*
  doi: 10.1371/journal.pone.0185577.
- Gu M, Qiu J, Guo D, Xu Y, Liu X, Shen C, Dong C (2018). Evaluation of candidate genes associated with hepatitis A and E virus infection in Chinese Han population. *Virol J. Mar 20;15 (1):47. doi: 10.1186/s12985-018-0962-2.*

- Hall WJ, Nathanson N, Langmuir, A. (1957). The Age Distribution of Poliomyelitis in the United States in 1955. *Am J Hyg. Sep;66(2):214-34.*
- Herndon CN, Jennings RG (1951) A twin-family study of susceptibility to poliomyelitis. *Am J Hum Genet. Mar;3(1):17-46*
- Hill AV (2012). Evolution, revolution and heresy in the genetics of infectious disease susceptibility. *Philos Trans R Soc Lond B Biol Sci. Mar 19;367(1590):840-9.*
- Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T, Kang S et al., (1996). The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat Med. Nov;2(11):1240-3.*
- *Ilumina (2019) Convert array data into meaningful results-GenomeStudio. Software Modules. Retrieved from https://www.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html*
- International HapMap Consortium: A haplotype map of the human genome. *Nature 2005; 437: 1299–1320*
- Jallow M. Teo YY, Small KS, Rockett KA, Deloukas P, Clark TG, et al., (2009). Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat Genet. Jun;41(6):657-65.* doi:10.1038/ng.388
- Jepson A (1998). Twin studies for the analysis of heritability of infectious diseases. *Bull Inst Pasteur.;96:71–81.*
- Johnson CM, Lyle EA, Omueti KO, Stepensky VA, Yegin O, Alpsoy E, et al., (2007). Cutting edge: A common polymorphism impairs cell surface trafficking and functional responses of TLR1 but protects against leprosy. *J Immunol. Jun 15;178(12):7520-4.*
- Kabir M, Afzal MS (2016). Epidemiology of polio virus infection in Pakistan and possible risk factors for its transmission. *Asian Pac J Trop Med. Nov;9(11):1044-1047.*

- Kallmann FJ, Reisner D (1943). Twin studies on the significance of genetic factors in tuberculosis. *Am Rev Tuberc.; 47:549–574.*
- Kanai M, Tanaka T, Okada Y (2016). Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. J Hum Genet. Oct;61(10):861-866.
- Khor CC, Chapman SJ, Vannberg FO, Dunne A, Murphy C, Ling EY et al., (2007). A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis*. Nat Genet. 2007 Apr;39(4):523-8.*
- Khor CC, Vannberg FO, Chapman SJ, Guo H, Wong SH, Walley AJ et al., (2010) CISH and susceptibility to infectious diseases. *N Engl J Med. Jun 3;362(22):2092-101*
- Kindberg E, Ax C, Fiore L, Svensson L (2009). Ala67Thr mutation in the poliovirus receptor CD155 is a potential risk factor for vaccine and wild-type paralytic poliomyelitis*. J Med Virol. May;81(5):933-6.*
- Klebanov N (2018) Genetic Predisposition to Infectious Disease. *Cureus 10(8): e3210. DOI 10.7759/cureus.3210*
- Lasch EE, Joshua H, Gazit E, El-Massri M, Marcus O, Zamir R (1979). Study of the HLA antigen in Arab children with paralytic poliomyelitis*. Isr J Med Sci. Jan; 15 (1):12-3.*
- Lee YH, Rho YH, Choi SJ, Ji JD, Song GG, Nath SK, Harley JB (2006). The PTPN22 C1858T functional polymorphism and autoimmune diseases: a meta-analysis. *Rheumatology (Oxford). 2007 Jan; 46 (1):49-56.*
- Lewallen S, Courtright P (1998) Epidemiology in practice: case-control studies. *Community Eye Health.;11(28):57-8.*
- Lin TM1, Chen CJ, Wu MM, Yang CS, Chen JS, Lin CC, Kwang TY, Hsu ST, Lin SY, Hsu LC. Hepatitis B virus markers in Chinese twins *Anticancer Res. 1989 May-Jun; 9 (3):737-41.*

- Lind K, Hühn M H, Flodström-Tullberg M (2012) Immunology in the clinic review series; focus on type 1 diabetes and viruses: the innate immune response to enteroviruses and its possible role in regulating type 1 diabetes. *Clin Exp Immunol. Apr; 168(1): 30–38.*
- Loeb M (2013) Host genomics in infectious diseases. *Infect Chemother. Sep;45 (3):253-9. doi: 10.3947/ic.2013.45.3.253.*
- Loeb M, Eskandarian S, Rupp M, Fishman N, Gasink L, Patterson J, Bramson J, Hudson TJ, Lemire M (2011). Genetic variants and susceptibility to neurological complications following West Nile virus infection. *J Infect Dis. Oct 1;204(7):1031-7*
- Lockhart, J. (1837). Memoirs of the life of Sir Walter Scott. Edinburgh: Robert Cadell. Available online: https://archive.org/details/memoirslifesirw85lockgoog
- Ma X, Liu Y, Gowen BB, Graviss EA, Clark AG, Musser JM (2007). Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. *PLoS One. Dec 19;2(12):e1318.*
- Malaty HM, Engstrand L, Pedersen NL, Graham DY (1994). Helicobacter pylori infection: genetic and environmental influences. A study of twins. *Ann Intern Med. Jun 15;120(12):982-6*
- Marshall AG, Hutchinson EO, Honisett J (1962). Heredity in common diseases. A retrospective survey of twins in a hospital population. *Br Med J. Jan 6;1(5270):1-6*
- Marsman RF, Wilde AA, Bezzina CR (2011). Genetic predisposition for sudden cardiac death in myocardial ischaemia: the Arrhythmia Genetics in the NEtherlandS study. *Neth Heart J. Feb;19(2):96-100.*
- Martin AR, Daly MJ, Robinson EB, Hyman SE, Neale BM (2019). Predicting Polygenic Risk of Psychiatric Disorders. *Biol Psychiatry. 2019 Jul 15;86(2):97-109.*
- Metzker ML (2010). Sequencing technologies - the next generation. *Nat Rev Genet ;11:31-46.*

- Michon P, Woolley I, Wood EM, Kastens W, Zimmerman PA, Adams JH (2001). Duffy-null promoter heterozygosity reduces DARC expression and abrogates adhesion of the P. vivax ligand required for blood-stage infection. *FEBS Lett. Apr 20;495(1-2):111-4*.
- Miller LH, Mason SJ, Clyde DF, McGinniss MH (1976). The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med. Aug 5;295(6):302-4*
- Nandi SS, Sharma DK, Deshpande JM (2016). Assay for identification of heterozygous single-nucleotide polymorphism (Ala67Thr) in human poliovirus receptor gene*. Indian J Med Res. 2016 Jul;144(1):38-45.*
- Nathanson N, Kew OM. From emergence to eradication: the epidemiology of poliomyelitis deconstructed. *Am J Epidemiol. 2010 Dec 1;172(11):1213-29*. doi: 10.1093/aje/kwq320.
- National Center for Biotechnology Information- [NCBI], (2019). Gene. Retrieved on 12/11/2019 from https://www.ncbi.nlm.nih.gov/gene/3761
- Ochmann S, Roser M, Our World In Data- Polio (2019) - "Polio". Published online at OurWorldInData.org. Retrieved on 23/11/2019 from: https://ourworldindata.org/polio
- Opare JK, Akweongo P, Afari EA, Odoom JK (2019). Poliovirus neutralizing antibody levels among individuals in three regions of Ghana. *Ghana Med J. Jun;53(2):170-180.*
- Pare G (2010) Genome-wide association studies--data generation, storage, Interpretation, and bioinformatics. *J Cardiovasc Transl Res. Jun;3 (3):183-8.*
- Panagiotou OA, Ioannidis JP; Genome-Wide Significance Project. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int J Epidemiol. 2012 Feb; 41 (1):273-86. doi:10.1093/ije/dyr178.*
- Petersen KA, Matthiesen F, Agger T, Kongerslev L, Thiel S, Cornelissen K, Axelsen M (2006) Phase I safety, tolerability, and pharmacokinetic study of recombinant human mannan-binding lectin. *J Clin Immunol. Sep;26(5):465-75*

- Prasad S, Tyagi AK, Aggarwal BB (2016). Detection of inflammatory biomarkers in saliva and urine: Potential in diagnosis, prevention, and treatment for chronic diseases*. Exp Biol Med (Maywood). Apr;241(8):783-99*
- Prevots DR, Burr RK, Sutter RW, Murphy TV (2000) Advisory Committee on Immunization Practices. Poliomyelitis prevention in the United States. Updated recommendations of the Advisory Committee on Immunization Practices (ACIP*). MMWR Recomm Rep. May 9;49(RR-5):1-22; quiz CE1-7.*
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al., (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559-75
- Rani R, Singh A, Israni N, Singh A, Sharma P, Kar HK (2009). The role of polymorphic protein tyrosine phosphatase non-receptor type 22 in leprosy. *J Invest Dermatol. Nov;129(11):2726-8.*
- Rekand T, Langeland N, Aarli JA, Vedeler CA (2002). Fcgamma receptor IIIA polymorphism as a risk factor for acute poliomyelitis. *J Infect Dis. Dec 15;186(12):1840-3.*
- Rindge ME (1955) Note on the age and sex incidence of poliomyelitis. *J Infect Dis. 1955 Jan-Feb;96(1):101-3.*
- Rovers M, Haggard M, Gannon M, Koeppen-Schomerus G, Plomin R (2002). Heritability of symptom domains in otitis media: a longitudinal study of 1,373 twin pairs. *Am J Epidemiol. May 15;155(10):958-64.*
- Sano D, Tazawa M, Inaba M, Kadoya S, Watanabe R, Miura T, Kitajima M, Okabe S (2018). Selection of cellular genetic markers for the detection of infectious poliovirus*. J Appl Microbiol. Apr;124(4):1001-1007.*
- Saunderson R, Yu B, Trent RJ, Pamphlett R (2004). A polymorphism in the poliovirus receptor gene differs in motor neuron disease. *Neuroreport. 2004 Feb 9;15(2):383-6.*

- Simonds B (1963). Tuberculosis in twins. Pitman Medical Publishing Co. Ltd; London: 1963.
- Sørensen TI, Nielsen GG, Andersen PK, Teasdale TW (1988). Genetic and environmental influences on premature death in adult adoptees. *N Engl J Med. Mar 24;318(12):727-32*
- Sweatt JD, Nestler EJ, Meaney MJ, Akbarian S (2013). Chapter 1-an overview of the molecular basis of epigenetics. In: Epigenetic regulation in the nervous system San Diego: Academic Press; 2013. p. 3–33
- Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'Huigin C, Kidd J et al., (2009). Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature. Oct 8;461(7265):798-801.*
- Thursz MR, Thomas HC, Greenwood BM, Hill AV (1997). Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nat Genet. 1997 Sep;17(1):11-2.*
- Thye T, Owusu-Dabo E, Vannberg FO, van Crevel R, Curtis J, Sahiratmadja E et al., (2012) Common variants at 11p13 are associated with susceptibility to tuberculosis. *Nat Genet. 2012 Feb 5;44(3):257-9.*
- Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, et al., (2010) Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat Genet. 2010 Sep;42(9):739-741. doi: 10.1038/ng.639. Epub 2010 Aug 8. Erratum in: Nat Genet.* 2011 Oct;43(10):1040.
- Triantafilou K, Orthopoulos G, Vakakis E, Ahmed MA, Golenbock DT, Lepper PM, Triantafilou M (2005). Human cardiac inflammatory responses triggered by Coxsackie B viruses are mainly Toll-like receptor (TLR) 8-dependent. *Cell Microbiol. Aug;7(8):1117-26.*
- Troyer JL, Nelson GW, Lautenberger JA, Chinn L, McIntosh C, Johnson RC, et al., (2011) Genome-wide association study implicates PARD3B-based AIDS restriction. *J Infect Dis. May 15;203(10):1491-502.*

- UCDavis, (2019). DNA Genome Center- DNA Technologies and Expression Analysis Core Laboratory. Retrieved on 11/12/2019 from: https://dnatech.genomecenter.ucdavis.edu/infinium-assay/
- Underwood M. (1789). Treatise on the Diseases of Children. Two Volumes. London: J. Matthews. Retrieved on 12/11/2019 from: https://archive.org/details/2575056R.nlm.nih.gov
- Uren C, Henn BM, Franke A, Wittig M, van Helden PD, Hoal EG, Möller M (2017) A post-GWAS analysis of predicted regulatory variants and tuberculosis susceptibility. *PLoS One. Apr 6;12(4):e0174738. doi: 10.1371/journal.pone.0174738*
- Walsh EC, Mather KA, Schaffner SF (2003). An integrated haplotype map of the human major histocompatibility complex. *Am J Hum Genet. 2003; 73:580–590.*
- Wang JP, Asher DR, Chan M, Kurt-Jones EA, Finberg RW (2007). Cutting Edge: Antibody-mediated TLR7-dependent recognition of viral RNA. *J Immunol. Mar 15;178(6):3363-7.*
- Wong SH, Hill AV, Vannberg FO (2010) Genome-wide association study of leprosy. *N. Engl. J. Med. 362, 1446–1447*
- Wong SH, Gochhait S, Malhotra D, Pettersson FH, Teo YY, Khor CC (2010). Leprosy and the adaptation of human toll-like receptor 1. *PLoS Pathog. 2010 Jul 1;6:e1000979.*
- World Health Organisations [WHO] (2019) The Immunological Basis for Immunization Series "Polio" World Health Organization/ Global Program For Vaccines and Immunization/ Expanded Program on Immunization. Retrieved on 25/11/2019 rom https://www.who.int/immunization/documents/immunological_basis_series/en/
- World Health Organization (2019). Polio Global Eradication Initiative. Standard Operating Procedures: Responding to a Poliovirus Event or Outbreak, Version 3, January 2019. Retrieved on 22/11/2019 from

http://polioeradication.org/wp-content/uploads/2016/07/sop-polio-outbreak-response-version-20193101.pdf

- Wyatt HV (2014) Epidemics of poliomyelitis in the Maltese island of Gozo: genetic susceptibility. *Malta Medical Journal 26: 3-8.*
- Zander H, Grosse-Wilde H, Kuntz B, Scholz S, Albert ED (1979). HLA-A, -B, and –D antigens in paralytic poliomyelitis. *Tissue Antigens. Apr;13(4):310-3.*