Understanding interactions between EBV and human genomic variation

Rajendra Mandage

TESI DOCTORAL UPF / 2017

DIRECTOR DE LA TESI

Dr. Gabriel Santpere

Prof. Arcadi Navarro

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

upf. Universitat Pompeu Fabra Barcelona

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor **Prof. Arcadi Navarro** for giving me the opportunity to work with him. Thank you Arcadi for your support, advice, and encouragement, and for providing me various opportunities to present and discuss my PhD work with the international scientific community.

I also would like to say a special thanks to my co-supervisor **Dr. Gabriel Santpere** for his continuous guidance, support and patience. His supervision helped me a lot in all the time of PhD and in completion of this thesis. I could not have imagined having a better guru (teacher) for my PhD project. I really appreciate Gabriel’s dissective intelligence about what is right or wrong in the project.

I would also like to acknowledge all lab members especially **Marco Telford** for being a part of GWAS work, **Xavi** for statistical analysis, **Juan** for SNPs exploration and **Juanma** for constant support to run a job on cluster.

I would like to particularly express my thanks to **Judit Sainz** for her administrative support that I received every year for visa and NIE related work. And finally, last but by no means the least, thanks also to everyone in IBE-PRBB … it was great sharing laboratory with **Txema, Diego, Marina** and **Gerard** during last four years.

I was also a part of many other activities outside the PRBB, with countless people particularly Mayukh, Vivek, Ajinkya, Saoni, Madhulika, Gaurav, Avinash, Aparna didi, Mahi sir, Hateem and Emanuela. Thank you all for all the parties and fun outside my work zone.

Finally, I would like to thank my family members and friends for their never-ending love and support, during ups and downs of my PhD.

ABSTRACT

The EBV has been linked to multiple human diseases such as different types of cancers. Recently single gene analysis and genome-wide analysis studies have been exploited to uncover the human genetic variants that are linked with EBV diseases. Those studies suggested a substantial role of host's genetics in these EBV-associated diseases and also provided a clue in understanding the interaction between virus and human. Furthermore, the outcome of the EBV infection is a complex phenomenon influenced by the variation in the genetic architecture of the viral and human genomes and/or the interacting environmental factors. Therefore, this PhD work is mainly a large-scale effort towards the understanding of the human and EBV genetic architecture to walk towards unveiling the role of genetic variation in EBV associated infections, disease susceptibility, immune recognition and invasion.

In the first chapter of this PhD, we confirmed that EBV copy number, measured in derived lymphoblastoid cell lines from more than 1700 individuals, differs within and among the 1000 Genome Project populations. The characterization of genetic basis of EBV copy number in LCLs by a GWAS analysis yielded us multiple genetic variants pointing to putatively relevant genes for the biology of EBV infection such as CAND1, FGD2, and KHDRBS2. In the second chapter of this PhD work, we studied the substantial variation observed in multiple antigenic regions of diverse EBV genome isolates, which affects the recognition by HLA molecules. This antigenic variation also clearly shows adaptive mechanism of virus to escape immune surveillance. The suggested EBV-human perturbations from this work must be follow-up in the context of the susceptibility of individual populations to a specific EBV associated pathology.

PROLOGUE

The coexistence of the ubiquitous EBV in human B-cell is the best example of balanced interaction between human and EBV, under normal circumstances. Although the virus persists in asymptomatic conditions, with the exceptional cases, an unusual interaction of human and EBV can result in the development of multiple cancers and other disorders with marked geographical variation in prevalence. An impact of host genetic variation on EBV infection was recognized previously in fragments providing a clue to the idea that differences in genetic architecture would confer differential susceptibility to EBV-associated diseases. However, there is still insufficient knowledge available at genomic scale on the genetic factors that can influence EBV biology, infection or the prevalence of EBV-associated disease due their large focus on analysis of a particular gene, typically centered on particular populations. This enforces us to shift our paradigm from single gene and/or population analysis to the large-scale genome-wide analysis considering worldwide populations at a time, to link the genetic polymorphism pinpointed throughout the genomes in different healthy populations, that could provide hints to ultimately link genomic variants with EBV associated diseases.

This PhD work is essentially a one step forward in understanding the interaction of human host and EBV to uncover the role of genetic variation in EBV infection, disease susceptibility and immune recognition. The combination of NGS methods (human and EBV whole genome sequences) and genome-wide association studies (GWAS) impelled us to perform a whole-genome level study to detect the genetic variants influencing EBV copy number (EBV load) of diverse genome samples from African, American, European, and Asian populations, ultimately providing clues to disease susceptibility. To detect the genetic variants from these 1000 Genome Project populations, first we conformed that EBV copy number is a stable phenotype over a time period within the lymphoblastoid cell lines (LCLs) and subsequently we developed an *in silico* algorithm to estimate the EBV copy number from LCLs. With this measurement, we have demonstrated that there is substantial variation in EBV copy number across the 1000 Genome populations suggesting a role of human genetic variation that can interact with EBV.

In another aspect of the PhD work, to extend our understanding on the immune response to EBV copy number, we have assessed the interaction of the human leukocyte antigen (HLA) alleles with EBV copy number as a phenotypic consequence controlling EBV infection in different population samples. Lastly, an extensive analysis was carried out to mine the EBV genome sequence polymorphisms in order to report on how natural variations (antigenic variations) might help the virus to escape the immune system and support lifelong persistent infection.

TABLE OF CONTENTS

ABBREVIATIONS

EBV	Epstein Barr virus
HHV-4	Human herpes virus 4
LCL	Lymphoblastoid cell lines
NGS	Next generation sequencing
NCP	Nasopharyngeal carcinoma
BL	Burkitt’s lymphoma
GC	Gastric carcinoma
EBVaGC	EBV-associated gastric carcinoma
HD	Hodgkin's lymphoma
IM	Infectious mononucleosis
MS	Multiple sclerosis
PTLD	Post-transplant lymphoproliferative disorder
EBNA	Epstein Barr nuclear antigen protein
LMP	Latent membrane protein
GWAS	Genome-wide association study
HLA	Human leukocyte antigen
MHC	Major histocompatibility complex molecules

1. INTRODUCTION

1.1 Why study Epstein-Barr virus (EBV)?

EBV, also known as human herpesvirus 4 (HHV-4), is a member of the gamma-herpesvirus family and is associated with malignant and non-malignant diseases in humans. It is considered as one of the most common and persistent viral infection. It is primarily transmitted by saliva and more than 90% of the world population is infected. Although EBV infection is asymptomatic in most individuals due to effective human host T-cell response, some individuals can develop infectious mononucleosis (IM) and other EBV-associated disorders such as nasopharyngeal carcinoma (NPC), Hodgkin lymphoma (HL) and Burkitt lymphoma (BL). Upon entering the human host, EBV establishes a latent infection in B cells and persists in B cells for a lifetime. However, the cellular mechanism through which EBV causes diseases is not fully understood.

A large-scale effort is needed to understand the interactions of EBV and human host genetic variation at population and continent level, as well as its impact on disease susceptibility. Little is known about how the genetic polymorphisms in human and EBV sequences influence the biology of the virus, and ultimately disease association. EBV is characterized by the heterogeneous distribution of worldwide isolates, hence systematic analysis and full understanding of the sequence variation patterns in viral strains could be a useful step for understanding immune escape and successful persistence lifelong infection to control EBV associated diseases.

1.2 Discovery of EBV

In 1961, Anthony Epstein, a pathologist, had the opportunity to attend a lecture by Denis Burkitt, an Irish surgeon who was working in Uganda. Burkitt delivered a talk about BL at Middlesex Hospital in the UK. During his speech, Epstein postulated that climate-dependent vectors might be responsible for the spread of the cancer-causing virus; although later it became clear that this was not the case. A few days later, Burkitt agreed to send some fresh tumor biopsy samples to Epstein for microscopic analysis, as

Epstein was already working on Rous sarcoma virus (RSV), the agent responsible for tumors in chickens, using electron microscopy. This formed the foundation for the discovery of EBV. In 1964, Anthony Epstein and Yvonne Barr successfully established cell lines from a BL samples. Subsequent microscopic observation of these cell lines confirmed that it was a herpesvirus family member (Epstein et al., 1979).

1.3 Taxonomy and Structure of EBV

EBV is the first tumor-isolated virus detected by electron microscopy in 1964 by Epstein and his group as a causative agent of African BL. It is a prototype of a herpesvirus gamma subfamily. EBV taxonomic name is human herpersvirus 4.

The EBV genome is a linear double-stranded DNA comprising approximately 170 kilobase pairs (kb), encoding around 85 genes. EBV open reading frames (ORFs) have been described based on *Bam*HI restriction fragments **(Fig 1)**. The EBV genes are mainly divided into lytic and latent genes; their protein products are listed in **Table 1.** The first complete sequence of EBV (type 1) was obtained from B95-8 marmoset cell lines from IM (Baer et al., 1984) (NCBI accession number V01555). This strain was not a complete representation of the majority of EBV strains as it was missing an 11.8-kb segment of the genome. Later, a hybrid sequence was assembled that would represent most of the isolates (NCBI accession number NC_007605). The missing 11.8-kb portion from V01555 was sequenced from the Raji strain of EBV (Parker et al., 1990) and inserted into a B95–8 prototype to correct the deletion, thus enabling the sequencing of all parts of the wild-type EBV genome (Roizman, 2007). Publications on the first complete genome sequence of EBV type 2 demonstrate that there is substantial allelic diversity in the genes EBNA-2, EBNA-3A, EBNA-3B, and EBNA-3C, which aids the classification of EBV into type 1 and type 2 strains (Dolan et al., 2006; Farrell, 2015; Sample et al., 1990).

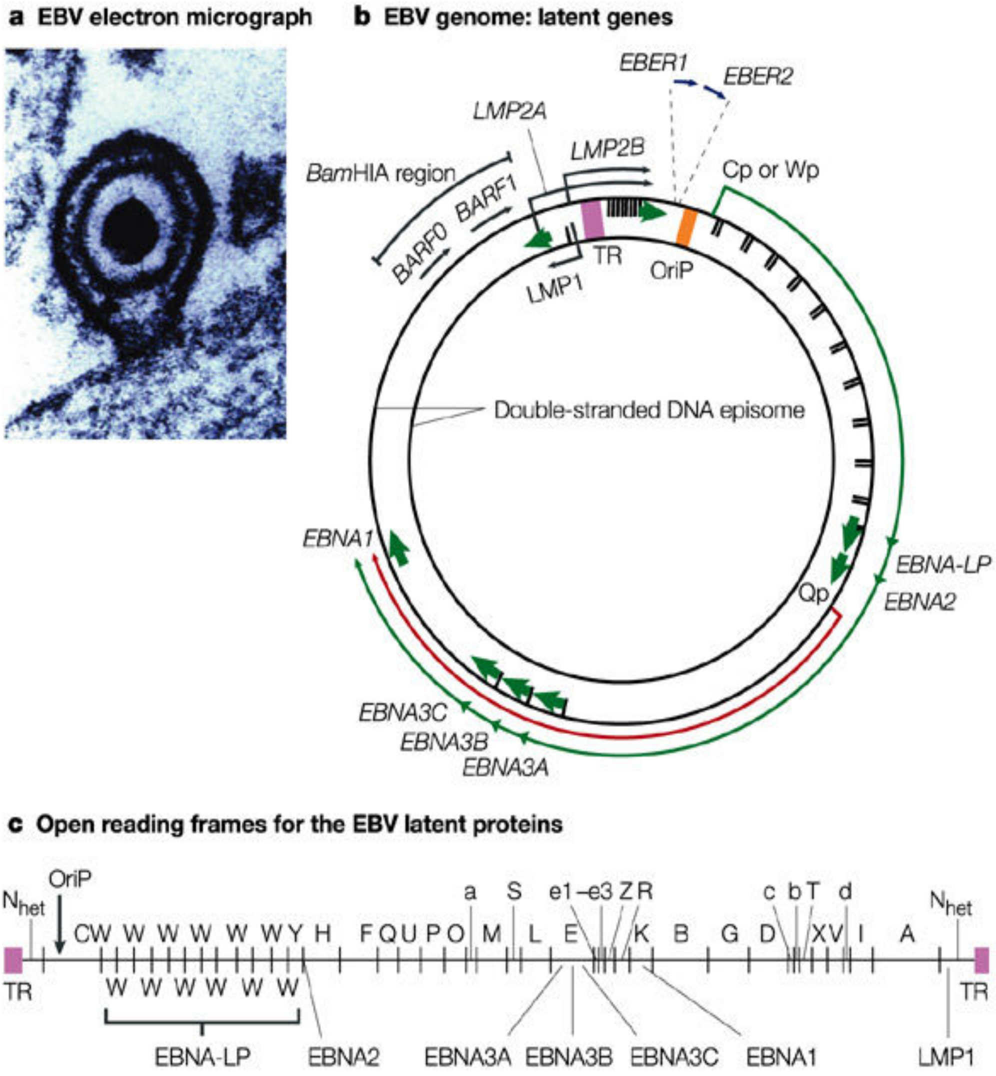

a EBV electron micrograph
b EBV genome: latent genes
EBER1
EBER2
LMP2A
LMP2B
BamHIA region
BARF0
BARF1
Cp or Wp
TR
OriP
LMP1
Double-stranded DNA episome
EBNA1
EBNA-LP
EBNA2
Qp
EBNA3C
EBNA3B
EBNA3A
c Open reading frames for the EBV latent proteins
OriP
Nhet
CW W W W W WY H F Q U P O M L E e1–e3 Z R K B G D c b T d X V I A
Nhet
TR
W W W W W W
TR
EBNA-LP
EBNA2
EBNA3A
EBNA3B
EBNA3C
EBNA1
LMP1

Figure: 1 | an Electron micrograph of EBV | **b** Diagrammatic representation of the location and transcription of EBV latent genes on the episome | **c** Location of ORFs for the EBV latent proteins on the *Bam*HI restriction endonuclease map of the B95-8 strain. (Figure has been adapted from Young and Rickinson, 2004)

1.4 Viral life cycle

The EBV life cycle can be broadly divided into two stages: lytic and latent. Once in latency, EBV can establish persistent lifelong infection in the human host. The EBV genome usually presents in the episomal form (circular DNA) in host cells but can also integrate (CIT). The virus enters the host lymphoid tissue of the oropharynx through the salivary transmission. B lymphocytes and epithelial cells are the primary targets of EBV, along with natural killer cells and T cells (Hutt-Fletcher, 2007). Although B lymphocytes are the main site of EBV infection, epithelial cells are also used as sites for the lytic cycle, in which viral progeny are produced and ultimately released for cell-to-cell spread and transmission (Hatton et al., 2014; Hutt-Fletcher, 2014; Imai et al., 1998). The prime focus of the virus is to establish lytic replication in the oropharynx, and then spread to lymph tissues in other compartments as a latent growth-transforming infection of B-lymphocytes. Effective T cell response clears most of the proliferating lymphocytes. However, some of the cells manage to escape the immune response and succeed in sustaining an established reservoir of resting viral genome–positive memory B cells, where the viral transcription program remains in silent mode (Babcock et al., 2000; Miyashita et al., 1995). Sometimes reactivation from latency is essential for viral production. This viral reactivation could result from terminal differentiation of memory B cells into plasma cells, although the precise mechanism of reactivation remains unclear (Laichalk and Thorley-Lawson, 2005; Thorley-Lawson et al., 2013; Young and Rickinson, 2004) (**Fig 2**).

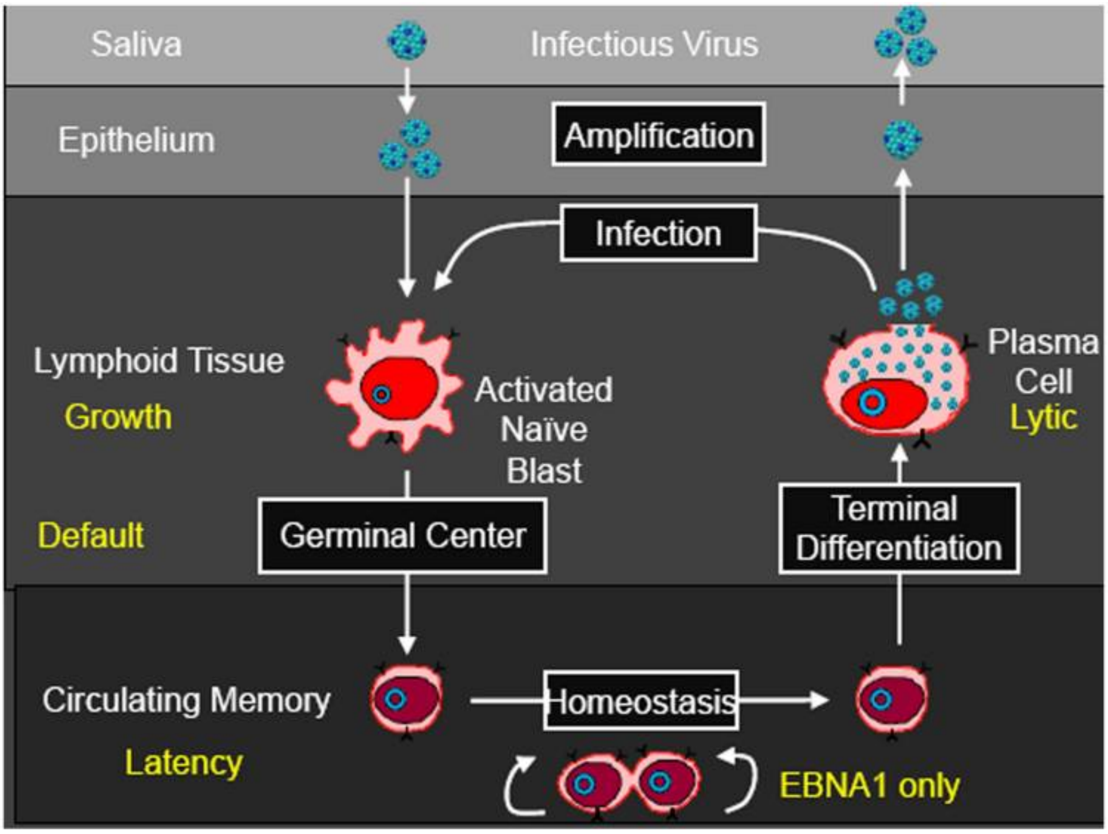


Figure: 2 Diagrammatic representations of EBV and human host interaction in primary and persistent infection (Figure has been adapted from Thorley-Lawson et al., 2013)

1.5 B cell transformation and LCLs

The B cell is the primary site of EBV infection. In vitro, EBV has demonstrated its ability to transform B cells into a continuously proliferating lymphoblastoid cell lines (LCLs) efficiently (Hussain and Mulherkar, 2012; Klein et al., 2010; Neitzel, 1986; Sugden, 1982). Within B cell, EBV establishes a latent infection characterized by a limited set of viral gene expression. The virus mainly expresses 6 nuclear antigen proteins (EBNAs), 3 membrane proteins (LMP-1, LMP-2A, LMP-2B), and 2 small nuclear RNA molecules (EBER-1, EBER-2) (Farrell, 1995; Klein et al., 2010; Nam et al., 2011; Ring, 1994; Wensing and Farrell, 2000; Young and Rickinson, 2004). This EBV-transformed LCL serves as an unlimited source of cell lines for human genomics studies (HapMap and 1000 Genome Project), proteomics studies as a surrogate model system for genotype-to-phenotype analysis, a varieties of molecular and functional assay studies, drug effects, and lymphocyte-associated disease analyses (Caron et al., 2002; Hussain and Mulherkar, 2012; Sie et al., 2009; Wheeler and Dolan, 2012).

1.6 EBV-associated disease spectrum

EBV infects 90% of human populations worldwide. Although EBV infection is asymptomatic in most of the population, it can result in several disorders in certain individuals. It has been postulated and shown to some degree that the expression of viral proteins during the lytic and latent life cycles disrupts normal cellular functional growth and development, which contribute to these EBV-associated diseases. Periodically, the virus is reactivated, and this reactivation with unknown mechanism is strongly associated with malignant diseases such as BL, NPC, HL, and other lymphoproliferative disorders.

Infection during early childhood is majorly asymptomatic, while infection in late childhood onwards can be more problematic and increases risk for both IM and multiple sclerosis (MS). It has been hypothesized that EBV infection occurs in early childhood in developing countries, while it occurs in late childhood in developed countries because of differences in hygiene conditions, and that this contributes to the different incidence of IM and MS around the world (Chabay and Preciado, 2013; Hjalgrim et al., 2007).

1.6.1 Infectious mononucleosis (IM)

EBV infection is self-limited, considered as asymptomatic, and is usually controlled by effective T cell immune response. If the infection occurs in adolescence, it may cause IM after primary infection. IM is characterized by a sore throat, fever, adenopathy, and splenomegaly (Chen, 2011; Hislop et al., 2007; Kutok and Wang, 2006).

1.6.2 Multiple sclerosis (MS)

MS is a disease of the central nervous system (CNS) that is also known as a chronic inflammatory demyelinating disease (Compston et al., 2006). Although substantial evidence supports the association between EBV and MS, the potential causal relationship remains unclear (Ascherio and Munger, 2010; Fierz, 2017; Pender and Burrows, 2014). The first report on EBV and MS was in 1979 when lymphocytes from patients with MS showed an increased tendency for spontaneous in vitro transformation (Fraser et al., 1979). Few studies suggest that EBV infection is essential for MS pathogenesis, although it is not sufficient for its development MS (Christensen, 2006; Pender, 2011).

1.6.3 Nasopharyngeal carcinoma (NPC)

NPC develops in the epithelium of the nasopharynx. The most common symptoms are neck swelling, nasal obstruction, and epistaxis. The association of EBV and NPC was established in 1973 (Wolf, H., zur Hausen, H., Becker, 1973), but the role of EBV in NPC pathogenesis remains unclear. It is thought that aberrant establishment of the virus latency cycle in epithelial cells might produce the genetic changes that result in carcinogenesis (Chang and Adami, 2006; Furukawa and MFurukaw, n.d.; Kutok and Wang, 2006; Young and Dawson, 2014).

1.6.4 Burkitt's lymphoma (BL)

EBV was discovered from tumor cells derived from BL biopsies in 1964. This discovery accelerated research work on the association of EBV with different cancers. The first evidence by Henle et al. demonstrated that all African BL samples were infected with EBV (Henle and Henle, 1966). The oncogenic development arises from a cascade of molecular and genetic events and not merely because EBV has the ability to transform B cells. Some observations have demonstrated that MYC-activating translocations might alter cellular functions, which drives lymphoma development (Bornkamm, 2009; Brady et al., 2007; Küppers, 2003).

1.6.5 Hodgkin lymphoma (HL)

HL is a unique oncogenic disorder due to the presence of scattered neoplastic Reed-Sternberg (RS) cells (usually found in HL biopsies) in a rich background of lymphocytes, eosinophils, plasma cells, and neutrophils (Ok et al., 2015). The EBV genome and its products have been detected in RS cells, confirming the association of EBV and HL (Jarrett et al., 1996).

1.7 EBV and host range

Herpesviruses belonging to the lymphocryptovirus family, such as EBV, are naturally found in different range of host such as humans, great apes, gibbons, and non-human primates species (Muhe and Wang, 2015; Wang, 2013). They typically infect humans and non-primates, but due to the lack of sufficient evidence, the host range is not yet well characterized.

1.7.1 Non-human animal model system for EBV disease study

EBV has developed an ability to immortalize the B-cells *in vitro* derived from non-human primates (Falk et al., 1974; Miller et al., 1972; Moghaddam et al., 1997; Rivailler et al., 2002). In a interesting study, B cells from chimpanzees had been immortalize by EBV, but the same effect could not be replicated in B cells derived from baboons and macaques (Gerber et al., 1977, 1976; Moghaddam et al., 1998; Rabin et al., 1978; Muhe and Wang, 2015). Therefore the comparative analysis of cellualar and molecular pathways that are ammenable to EBV immortlisation in human (non resistance to immortlisation) and non-human primates, as mentioned earlier, would serve as a mechanistic model to detect the underlying one more cellular pathways differences present in human and non-human primates may be essential for EBV disease (Muhe and Wang, 2015).

1.8 EBV genome sequencing and identification

The first EBV genome, EBV type 1 (B95-8), was sequenced and identified from IM samples from an American patient in 1984 (Baer et al., 1984), whereas the first type 2 (AG876) strain was identified and published from BL samples obtained in Central Africa (Dolan et al., 2006). First sequencing attempts have been conducted using conventional methods such as shotgun sequencing (B95-8, AG876, GD1), which were followed by next-generation sequencing (NGS) related strategies such as virus-free enrichment (GD2, C666-1, K4413-Mi, K4123-Mi, NA12878), EBV enrichment by lytic induction (Akata and Mutu), F-factor cloning (M81), amplicon sequencing (HKNPC1, LCL1, JQ009376), and EBV target enrichment capture by hybridization (HKNPC2, KF992564) (Baer et al., 1984a; Dolan et al., 2006; Kwok et al., 2014; Lei et al., 2013; Lin et al., 2013; Liu et al., 2016; Palser et al., 2015; Santpere et al., 2014b; Simbiri et al., 2015; Song et al., 2015; Tsai et al., 2013; Tso et al., 2013; Zeng et al., 2005). These methods have been successfully used for sequencing and identifying more than 120 EBV genomes, representing diverse geographical origins and disease types.

1.9 EBV type 1 and type 2 strain variation and distribution

Based on the genetic variation of EBNAs (EBNA-2, -3A, -3B, -3C), the EBV genome is mainly classified into type 1 and type 2 (sometimes also referred to as type A and type B) (Zimber et al., 1986; Terrace and Hospital, 1987; Sample et al., 1990; Baumforth et al., 1999; Palser et al., 2015). The EBNA-2 gene sequence from EBV type 2 is slightly shorter than that of EBV type 1 (Dambaugh et al., 1984; Farrell, 2015; Feederle et al., 2007; McGeoch, 2001; Zimber et al., 1986). Sample et al. proposed that variation in EBNA-3C could be a good choice for classifying EBV type 1 and 2, as EBNA-C from EBV type 1 (B95-8) is 77 amino acid shorter than that of EBV type 2 (Sample et al., 1990). The recent method for detecting EBV type 1 and 2 is based on variation in the EBNA-C gene sequence (Palser et al., 2015; Sample et al., 1990). There is more than 70% sequence homology between EBNA-2 from EBV type 1 and 2 at the gene level and 54% at the protein level (Farrell, 2015). Furthermore, many studies have found that EBV type 1 is a more potent transforming agent of B cells i*n vitro* as compared to EBV type 2 (Baumforth et al., 1999; Chang et al., 2009; Rickinson et al., 1987; Tiwawech et al., 2008; Tzellos and Farrell, 2012). Examination of EBV isolate distribution among the worldwide population has shown that EBV type 1 is predominant as compared to EBV type 2 (Zimber et al., 1986). However, some evidence suggests that EBV type 1 and type 2 co-infection is also possible in some immunocompromised patients, indicating that EBV type 2 infection may be acquired in immunocompromised states (Apolloni and Sculley, 1994; Yao et al., 1996). EBV type 1 infection has been dominantly detected in Caucasian and Asian populations. Around 74% of individuals are infected with EBV type 1, and 19% with EBV type 2, whereas only 7% are carrying both strains (Apolloni and Sculley, 1994; Correa et al., 2004; Klemenc et al., 2006). Nevertheless, EBV type 2 is mostly observed in sub-Saharan African populations (Apolloni and Sculley, 1994; Bobek et al., 2010; Kim et al., 2006; Sandvej et al., 2000; Tiwawech et al., 2008).

1.10 EBV type 1 and type 2 disease associations

Many attempts have been made to correlate the presence of EBV strains (type 1 or type 2) with EBV-associated diseases, but due to the scarcity of sufficient genome samples from healthy and disease conditions and strain type 1 and 2 (most present-day genomes are of EBV type 1 only), information on EBV types and disease association is limited (Chang et al., 2009; Tzellos and Farrell, 2012). Nevertheless, some studies have found that immunocompromised patients are co-infected with EBV type 2, suggesting that persistence of EBV infection might be influenced by the human immune system response against it (Apolloni and Sculley, 1994; Baumforth et al., 1999; Bellas et al., 2008; Bobek et al., 2010; Correa et al., 2007; Khanim et al., 1996).

1.11 EBV genome variation, polymorphism, and disease association

Although there are more than 120 EBV genomes sequenced till now, this number partially representing the diversity of EBV type 1 and 2 worldwide distribution detected by conventional and NGS technologies, this figure does not represent the complete picture of EBV natural variation (Chang et al., 2009; Kwok et al., 2012; Palser et al., 2015; Santpere et al., 2014; Szpara et al., 2014; Tso et al., 2013; Tzellos and Farrell, 2012). EBV was first discovered in BL cells from African populations. However, soon after this discovery, researchers realized that EBV infection is not just restricted to African populations, but is endemic worldwide (Tzellos and Farrell, 2012). Although much work has been done on specific genes such as LMP-1 and EBNA-1, from particular diseases such as NPC, BL, and MS, very little information is available on EBV complete genome sequence variation due to the lack of sufficient cases and control samples from different geographical areas to infer a global picture.

There has been much research conducted to link EBV strain and its genetic polymorphism with EBV infection (so-called EBV-associated disease). Although there is no clear evidence of EBV genome variation influencing the development of both malignant and non-malignant disease, the sequencing of some EBV genomes provided overwhelming evidence of diverse viral subtypes. This lack of evidences is due to the

scarcity of EBV genome sequences from healthy and diseased individuals. Following points explain this scenario:
(i) Most research to date has been directed more towards the isolation and functional analysis of specific genes such as LMP and EBNA. However, this does not provide a global view of genome functionality and diversity; (ii) The functional aspect of genetic polymorphism is not fully understood; (iii) EBV type 1 and 2 stains are remarkably different in terms of B cell transformation, but no evidence shows that EBV type 1 and 2 cause various diseases. Hence, it is difficult to link viral genome polymorphism with EBV biological and clinical behavior.

1.12 Biology and clinical significance of EBV copy number estimation in blood, serum, and plasma

The measurement of viral DNA now plays an essential role in the diagnosis and disease monitoring and management of many viral infections such as HIV or hepatitis B and C. As far as EBV is concerned, its DNA quantification from plasma and blood has already shown a prognostic significance in some of the lymphoproliferative disorders. For EBV infection quantification, there are few methods that are routinely applied. The quantification of EBV DNA in the blood can be performed in whole blood peripheral blood mononuclear cells (PBMC) and plasma or serum.

Many pieces of evidence support the association of EBV copy number with EBV-associated diseases, where the presence of viral copy number in saliva, peripheral blood, and serum from NPC and lymphoproliferative disorders has been demonstrated. For example, EBV copy number in the plasma may be useful for HL prognosis (Hohaus et al., 2011; Jarrett, 2003; Kanakry et al., 2013), and assessment of EBV copy number in plasma or whole blood may be a useful tool for monitoring therapy in other EBV-associated tumors, and it is currently used in NPC (Gärtner and Preiksaitis, 2010). Several methods are available for detecting and quantifying cellular and extracellular EBV copy number (EBV copies) in peripheral blood. Although old approaches were limited by the lack of samples, modern amplification techniques such as real time PCR have made it possible to quantify large sample sizes. These methods are aimed at detecting EBV copy number in healthy (control samples) and disease conditions.

In a healthy individual, the number of EBV-infected B cells varies significantly between individuals, ranging from 1-50 copies per million cell in peripheral blood (Khan et al., 1996). In contrast, some evidence shows that EBV copy number is apparently elevated in primary infection (Fan and Gulley, 2001) and in post-transplant lymphoproliferative disease (PTLD) (Stevens et al., 2002a, 2002b), whereas in some immune disorders, it could result from increased numbers of EBV-infected B cells (presence of high count of EBV-positive B cells in the blood) (Babcock et al., 1999; Yang et al., 2000). The presence of high EBV copy number in saliva, blood and plasma can also be associated with episodes of EBV reactivation (Lechowicz et al., 2002).

1.12.1 EBV copy number during IM

In healthy seropositive individuals, EBV copies are hardly found. However, in the acute phase of IM, high titers of EBV copy numbers have been detected in the serum of most patients (**Fig 2**) (Berger et al., 2001). An increase in EBV copy number is observed in the saliva for at least six months after the clinical symptoms of the disease disappear, accompanied by persistent infectivity of saliva due to chronic infection of pharyngeal epithelial cells (Fafi-Kremer et al., 2005).

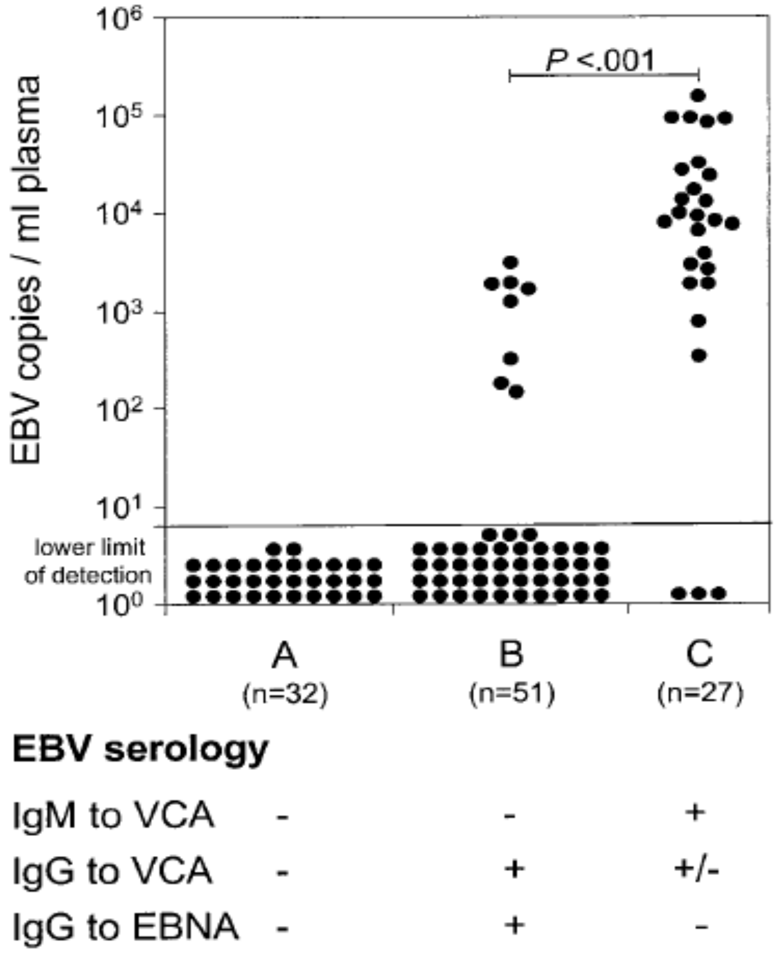


Figure 2 EBV DNA levels in serum from individuals with different EBV antibody patterns. A: Sera devoid of antibodies to EBV. B: Sera from EBV carriers. C: Sera from patients with acute EBV infection; IgG/IgM to VCA and IgG to EBNA, antibodies to viral capsid antigen and EBNA (Figure and text has been adapted from Berger et al., 2001)

1.12.2 EBV copy number in NPC

Serum quantification of EBV copy number using RT-PCR of patients with NPC showed that 96% of disease samples presented 21,058 copies/ml (median concentration, 55 of 57 samples) and 7% of control samples showed 0 copies/ml (median level, 3 of 7 samples) (Lo et al., 1999) **(Fig 3).**

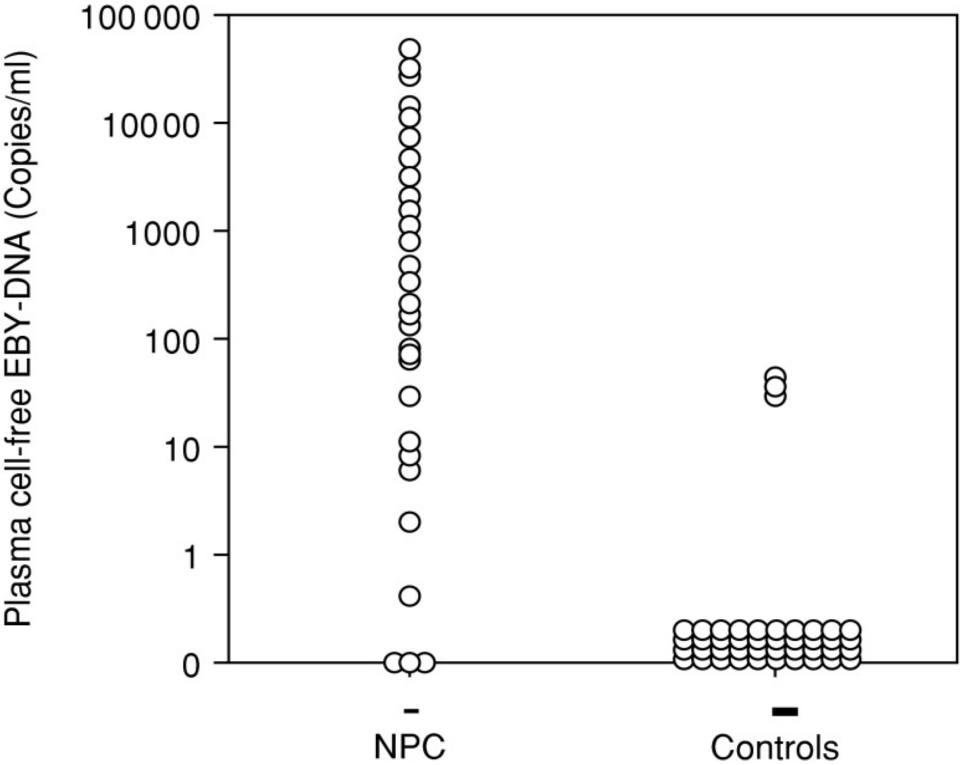


Figure: 3 Comparison of plasma EBV DNA in patients with NPC and control subjects. The categories (patients and controls) are plotted on the X-axis. The Y-axis denotes the concentration of cell-free EBV DNA (copies of EBV DNA/ml plasma) (Figure and text has been adapted from Lo et al., 1999)

2. CHAPTER 1

2.1 EBV copy number and GWAS analysis

2.1.1 GWAS and complex human diseases

The central role of human genetics is detecting the genetic factors affecting the risk of diseases. Although extensive, ever-accelerating work has been done in biomedical research for more than a decade to understand the genetic basis increasing risk of most human viral, bacterial, and other parasitic diseases, this still is a quite uncharacterized territory. Susceptibility to common disease is mostly affected by genetic variants present in the host and in the infectious agent, which sometimes influence disease-related phenotypic traits.

Detecting a disease-attributing gene is difficult in part because the causal gene contributes only a small fraction of heritability to a disease and the effect is small. To overcome this issue, genome-wide association study (GWAS) has been emerging as a powerful technique for mapping causal variants to understand the functional role of genes in disease conditions and treatment development through the detection of genetic variants, genes, and metabolic and signaling pathways that might be disrupted by an infection agent (Cantor et al., 2010; Cao and Moult, 2014; Ko and Urban, 2013; Nuzhdin et al., 2012; Trampush et al., 2017; Visscher et al., 2012; Wellcome et al., 2007). In 2005 and 2006, the first successful GWAS was published using a few markers and small sample size (<200 samples), but managed to detect a few common variants linked to age-related macular degeneration (Klein et al., 2005). The completion of the Human Genome Project, high-throughput sequencing technologies, the deposition of millions of single-nucleotide polymorphisms (SNPs) in public databases such as the 1000 Genome Project and HapMap Project, as well as advances in genotyping methodologies are opening new frontiers for the detection of the genetic risk of complex diseases. GWAS is widely used technique for identifying the genetic factors affecting diseases susceptibility and influencing disease-related phenotypic traits (Frazer et al., 2009; Hirschhorn and Daly, 2005; Ko and Urban, 2013; Raychaudhuri, 2011; Sebastiani et al., 2009).

Most disease-associated phenotypic traits are complex in nature, a complexity incremented by their variation during the different phases of the infection. These phenotypes cluster with disease consequences, suggesting genetic relatedness (Stranger et al., 2011). Enormous progress has been made in the last few years to map these phenotypes to disease susceptibility using the GWAS approach.

2.1.2 Introduction to GWAS method

In biomedical research, GWAS is primarily a study of the genome-wide data of genetic variants in different individuals to discover genetic variants whose genotype is associated with the phenotype of interest. GWAS emphasizes typically on the association between SNPs and complex phenotypic traits that can be a disease. GWAS usually detect SNP(s) present in linkage disequilibrium with genes or other regulatory elements that may influence the outcome of variable phenotypes. The phenotype could be a qualitative trait, representing for example healthy (control samples) or clinical manifestations (case samples) or a qualitative one, such as viral load. If a particular SNP is more frequently observed in disease cases than in controls, or is correlated with the phenotype, then it is considered associated with the manifestations of that trait. The availability and development of commercial high-throughput genotyping chips and the increasingly cheaper whole-genome sequencing techniques, aid the study of polymorphism on a large scale (up to several kb), facilitating the research on common genetic variation associated with complex phenotypic traits by GWAS (Bush and Moore, 2012; Cordell and Clayton, 2005; de Resende et al., 2014; Welter et al., 2014). The Wellcome Trust Case Control Consortium (WTCCC), a group of 50 research groups in the UK, published a landmark study using GWAS in which approximately 2000 samples per disease (8 diseases were selected for the analysis) were exploited to understand the pattern of human genome variability at population level. This data was used to link genetic polymorphisms with common illnesses such as bipolar disorder (BD), breast cancer (BC), coronary artery disease (CAD), Cohn’s disease (CD), hypertension (HT), rheumatoid arthritis (RA), type 1 diabetes (T1D), and type 2 diabetes (T2D), and they had successfully identified many genetic variants that could affect disease susceptibility (Wellcome et al., 2007).

Within ten years of its first application, GWAS had revolutionized the field of human genetics due to advances in high-end computational tools and of the availability of next-generation commercial chips. GWAS-related publications are in fact growing exponentially (Clarke et al., 2011; Welter et al., 2014). To date, the GWAS Catalog contains 3043 publications and 38,708 unique SNP–trait associations, as of July 17, 2017 (**Fig 4**).

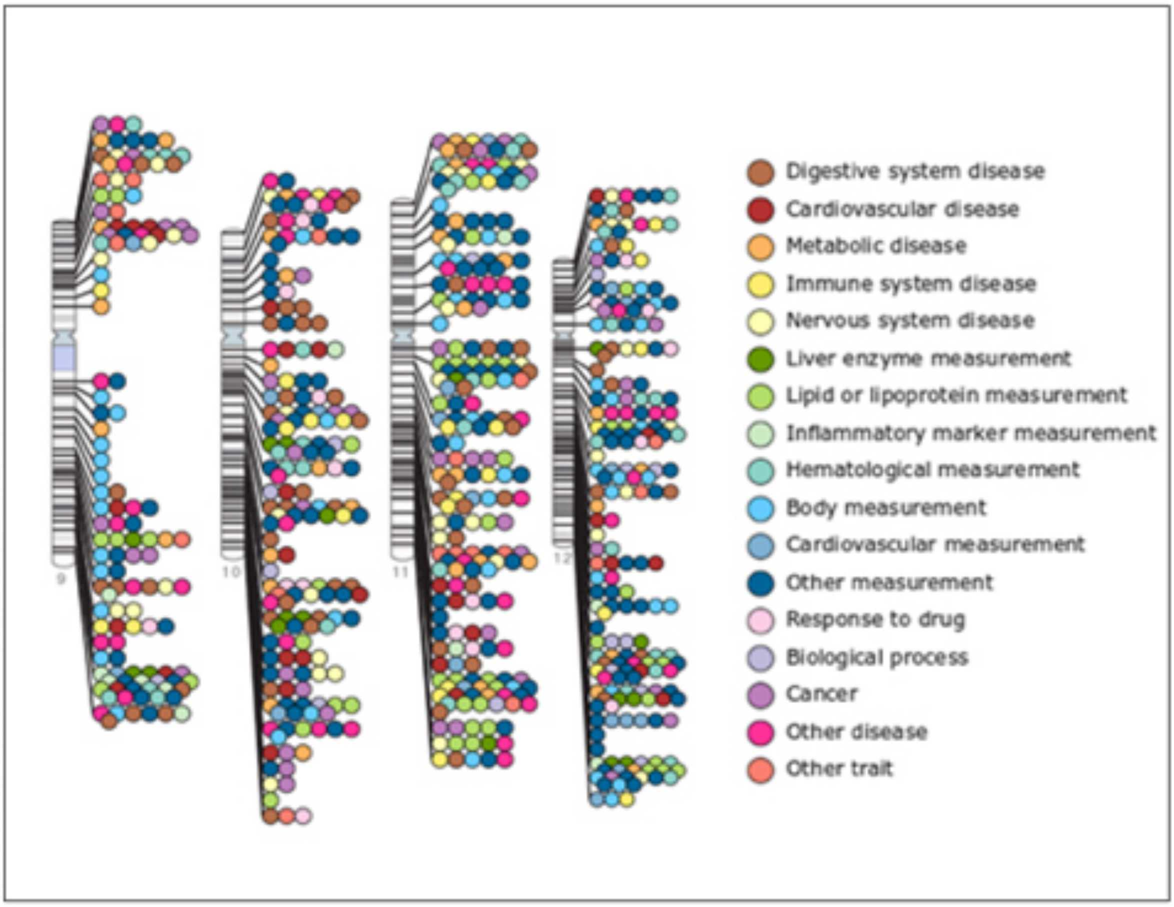


Figure: 4 Chromosome-wise distributions of GWAS identified variants associated with the most common human diseases (Figure has been adapted from National Human Genome Research Institute (NHGRI) GWAS Catalog website)

2.1.3 EBV, GWAS studies, and genomic variation detection

GWAS is considered a powerful tool for understanding the relations between human variation and EBV disease susceptibility. Although the advances in whole-sequencing technologies have enabled the study of genomic variation attributed to susceptibility to EBV-associated infections, only a few GWAS have been published so far due to the scarcity of data from healthy and disease individuals comprising different populations. Moreover, the lack of sufficient whole-genome sequences precludes a global level virome-wide association analysis to detect putative associations between EBV variants and risk for human diseases.

The few publications of EBV-related GWAS analysis from different populations using healthy and a diseased individuals provides partial clues on the presence of genetic variants affecting susceptibility to EBV infections. For example, a GWAS detected the presence of genetic factors that influence antibody levels against the EBNA-1 protein (Rubicz et al., 2013). Another GWAS involving 1200 patients with classic HL (cHL) and 6417 controls identified genetic variants from HLA region alleles that are strongly associated with cHL. Thus it accumulates the evidence of EBV as a risk factor for cHL (Urayama et al., 2012).

3. OBJECTIVE

The increasing incidence of EBV presence in the multiple cancers and other diseases has galvanized research efforts to answer the critical questions on the relationship between variation in human and EBV's genetic polymorphisms and disease involvement. These instances are increasingly characterized by the heterogeneous distribution of viral isolates. Several attempts have been made to dissect this association using isolation, sequencing, and analyzing of a particular polymorphism. However, this approach is ineffective for portraying the association at genome level due to the lack of genome-wide analysis data. Therefore, with **the main objective of providing a better characterization of genetic polymorphism at genome-wide level** using a combination of NGS data and other genome analysis techniques, this PhD work was started.

Within the context of the abovementioned main objective, this PhD work reflects 3 specific goals in terms of the genome-wide analysis of human populations and EBV isolates:

To confirm whether EBV copy number is a stable phenotype over the time (Objective 1)**. To detect the human genome variation that influences the EBV copy number in LCLs** (Objective 2), as well as **the EBV genome polymorphisms (amino acid change in epitope sequences) that modulate the immune response and alter the ability of the human leucocyte antigen (HLA) molecule to recognize EBV** (Objective 3), as explained in detail in Section 3.1, 3.2, and 3.3 respectively.

3.1 Objective 1

To confirm whether EBV copy number is stable phenotype over the time

The main objective of GWAS work was based on the hypothesis that there are certain genetic variants present in individual populations that governs the EBV copy number (viral load). Therefore it was essential to specifically interrogate the EBV copy number stability over the time within LCL samples, since the variation in viral load within a single LCL would have invalidated its estimations. Hence first we decided to confirm the viral load stability over the time to use it as phenotype in association testing.

3.2 Objective 2

To study human genome variation and EBV disease association

The prevalence of EBV-associated diseases shows considerable geographical differences regarding disease vulnerability, which suggests the presence of genetic factors rendering individual populations more prone to EBV infection. Hence, we wondered what contributes to the difference in EBV copy number in EBV-transformed LCL samples from the 1000 Genome Project. With the objective of studying human genome variation to test this hypothesis, we conducted a GWAS analysis to identify candidate genetic variants (SNPs and ultimately genes) putatively increasing risk to EBV infection, using LCLs as a surrogate model.

3.3 Objective 3

To study EBV genome variation and the immune system response

It is well established that EBV genomes are highly polymorphic in nature particularly in latency genes (Santpere et al., 2014; Palser et al., 2015) but some questions remain unanswered. What brings this polymorphism to arise and be maintained? Is it the immune pressure? How does EBV manage to survive in B cells despite robust immune surveillance? We aim to answer part of these questions by analyzing EBV genome variation using two approaches: (i) By exploring the role of HLA alleles on EBV copy numbers, estimated for European Asian, African and American populations from 1000 Genome Project. We tried here to build an association model statistical analysis tool to detect the HLA polymorphism linked to EBV copy number.

(i) We attempted to detect natural variation in the antigenic repertoire in the EBV genome by directly mapping epitope regions with the EBV proteome from multiple diseases, accounting for population-specific differences. We plan to analyze a catalog of EBV polymorphisms (epitope variation and ultimately antigenic diversity) to detect allelic variants (population or disease-specific) correlating with the worldwide distribution of HLA alleles.

4. MATERIAL AND METHODS

4.1 EBV copy number (viral load) stability over time

In addition to the GWAS analysis, our motivation was also to confirm whether EBV copy number is a stable phenotypic trait over time within lymphoblastoid cell lines. This is because the variation in viral load estimations within LCL could have been unacceptable for GWAS analysis. In order to confirm the stability, we had randomly selected 7 EBV transformed LCLs samples that are also a part of the 1000 Genome Projects with the accession IDs: HG01277, HG00245, HG00362, HG00657, NA18999, NA18502, NA19382. The culture of each 7 LCLs were then distributed into 3 replicates and each replicate was cultured at the same condition for 6 passages (approximately 3-4 days between passages), The spares cells from each passages then used to isolate DNA subsequently to estimate the viral load by RT-PCR. In order to count the cells, Neubauer chamber method was used. Finally ANOVA test was performed to estimate the relative viral load using the replicates of every LCL as repetition of DNA measures for every single passage. These EBV copy number estimations per LCL were afterwards compared with *in silico* copy number estimates as described in our research article to check the correlation of both real time PCR and *in silico* methods.

4.2 Tools and databases used in EBV copy number estimation and GWAS study

(A) 1000 Genome Project database

http://www.internationalgenome.org/

The 1000 Genome Project is an excellent repository for human genetic variation detected by whole-genome sequencing (WGS) from diverse worldwide populations. At the time of the performance of this project, it had collected the genomes of 2504 individuals from 26 populations **(Fig 5)** using a combination of low-coverage WGS, deep exon sequencing, and dense microarray genotyping. It had a total of over 88 million variants (84.7 million SNPs) (The 1000 Genomes Project Consortium, 2015).

The 1000 Genome Project serves as a worldwide reference for human genetic variation for GWAS, and is used for mapping expression quantitative trait loci (eQTL), filtering non-pathogenic variants from exomes, whole-genome and cancer genome sequencing projects, population structure analysis, and molecular evolution studies (Zheng-Bradley and Flicek, 2016). Most of the 1000 Genome Project genome samples are derived from LCLs maintained at the Coriell Institute for Medical Research; therefore, since all LCLs have been produced by EBV transformation, it is a good raw data source for estimating EBV copy number variation and how it varies in hosts with different genetic architectures.

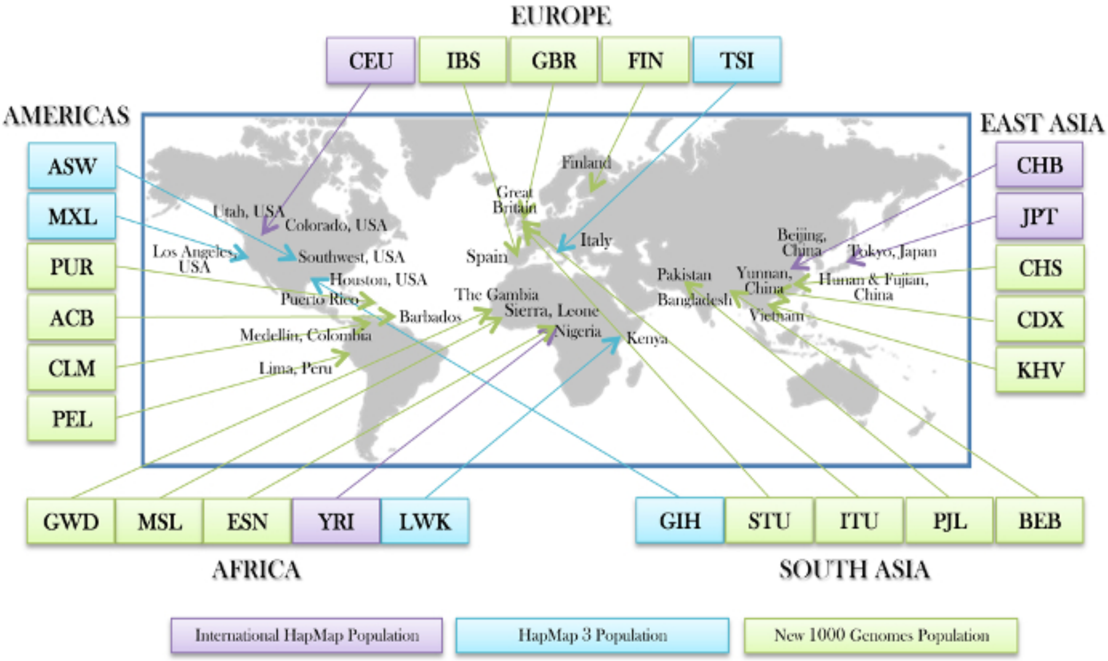


Figure: 5 1000 Genome Project describing sample population source (Figure has been adapted from 1000 Genome Project website)

Role of 1000 Genome Project in EBV copy number estimation

As described previously, most of the 1000 Genome Project data is derived from LCLs. It is a good source for the estimation of EBV copy number in individual LCL samples. First, we retrieved the file containing the alignments index from the 1000 Genome Project FTP site. (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/alignment_indices/), mapped and unmapped reads were re-mapped against human and EBV reference genomes separately. In total, 1753 LCL genome samples encompassing about 19 worldwide populations were considered as final dataset, and we also retrieved genotypic data (around 39 million variants) of the LCL samples under investigation in VCF format (Phase 3) from the 1000 Genome Project website for estimating EBV copy number in European, African, American, Asian population, and all population combined together to detect genetic variants associated with EBV copy number.

(B) PLINK

http://zzz.bwh.harvard.edu/plink/

PLINK is a free, open-source genome analysis tool mostly used in genetics, genomics, and the biomedical field as a reliable toolkit useful for genome and exome data management, quality control, basic statistical analysis, linkage disequilibrium patterns, population stratification detection, and analysis of genomic variation. It also allows the study of large datasets of genotypes and phenotypes through GWAS for both basic case/control studies and quantitative traits (Purcell et al., 2007).

Role of PLINK tool in EBV copy estimation

Initially, PLINK was used to test 39 million variants for Hardy-Weinberg equilibrium (HWE) failures and to filter out SNPs with p-values ≤ 0.01 from five population sets. It was also used for processing a 39M-genotype file from the 1000 Genome Project for conversion into binary files suitable for association testing. Lastly, clustering and distance matrix was also calculated using PLINK.

(C) SAMtools

http://samtools.sourceforge.net/

The Sequence Alignment/Map (SAM) is a generic alignment format for storing and reading alignments against reference genome sequences. It consists of various utilities for post-processing alignments in SAM format, such as indexing, variant caller, and alignment viewer, and therefore serves as an essential toolkit for processing read alignments analysis (Li et al., 2009).

Role of SAMtools tool in EBV copy number estimation

Mainly SAMtools view and SAMtools index commands were used to process 1000 Genome Project BAM files to render them suitable for GATK input files.

(D) Genome Analysis Toolkit (GATK)

https://software.broadinstitute.org/gatk/

Genome Analysis Toolkit (GATK) is a structured programming framework developed for efficient and robust analysis of next-generation DNA sequencing data using functional programming. It provides a wide range of data access and analysis modules such as coverage calculators and SNP calling. It also enables development of efficient and robust NGS tools, some of which are now an integral part of large-scale sequencing projects such as the 1000 Genome Project and The Cancer Genome Atlas (TCGA). It processes input BAM files to estimate coverage at different levels of partitioning and aggregation. Coverage can be calculated per locus, per interval, per gene, or the overall mapped region (McKenna et al., 2010).

Role of GATK in EBV copy number estimation

EBV copy number was calculated based on the relative coverage in each sample of human and EBV provided by GATK using 1000 Genome Project–ordered BAM files as input.

(E) Genome-wide efficient mixed-model association (GEMMA) tool

http://www.xzlab.org/software.html

GWAS was carried out using the genome-wide efficient mixed-model association (GEMMA) tool. It is based on a univariate linear mixed model (LMM) for marker association tests with a single phenotype in order to account for population stratification issues with sample structure and to estimate the proportion of phenotypic variance (Zhou and Stephens, 2012).

Role of GEMMA tool in GWAS analysis

The association test between EBV copy number as the GEMMA tool analyzed the phenotype against SNPs from the 1000 Genome Project for European, Asian, African, American, and all Populations datasets individually.

(F) R language and R Visual Studio

https://www.r-project.org

R is a free statistical computing and graphics tool supported by R Foundation. The extensive functionality of R has been strengthened by user-created packages (R packages) that allow users to perform a wide range of tasks related to statistical analysis (e.g., Bioconductor package) and data visualization in graphic format (e.g., ggplot2 package). R visual studio is R graphical user interface that makes R easy to use. It provides a code editor, debugs support, and visualization interface.

Role of R and R Visual Studio in GWAS data visualization and plot generation

Initially, R Visual Studio was used to do part of the data quality control analysis. For example R boxplotstats function was used to detect outliers in EBV copy numbers in European, Asian, African, American, and all population datasets separately. Then, it was used to generate covariate files containing the first 10 dimensions (MDS matrix) to correct the GWAS for population stratification using the cmdscale function.

R was extensively used for GWAS output processing, visualization and analysis, as it has an available package to generate manhattan plots (qqman). These are used to visualize the distribution of GWAS variants identified by association testing. It was also extensively used to plot images related to EBV copy number distribution among 1000 Genome Project populations, EBV copy number variation, and correlation plots to validate the *in silico* method with the RT-PCR method. It was also used to generate qq plots to show GWAS genome-wide p-value distributions in All Populations, Asian, African, American, and European population subsets. R was also used for EBV copy number raw values transformation such as log and rank normal transformation.

4.3 Replicability of GWAS results across continents

To determine whether the signals of association are replicated across continents, we selected all SNPs with p-values < 10^{-5} and grouped them in 400-kb clusters, putting together all SNPs that fell within 400 kb of each other, with each cluster having at least 2 SNPs. Cross-continent replication of these individual clusters was checked by generating qq-plots. This approach allows the evaluation of replicability between subsets even in the absence of genome-wide significant SNPs **(Fig 6)**.

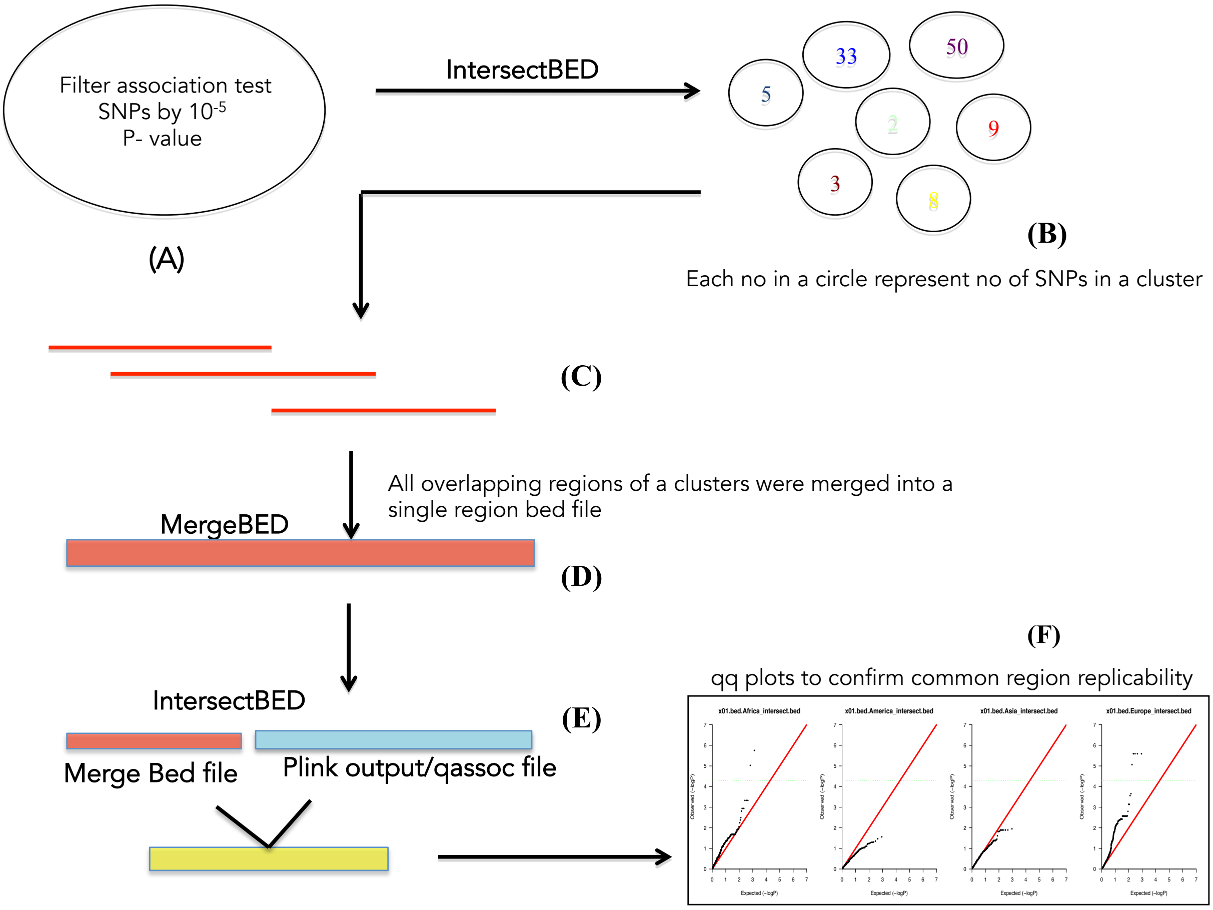

Filter association test
SNPs by 10-5
P- value
(A)
IntersectBED
33
50
5
2
9
3
8
(B)
Each no in a circle represent no of SNPs in a cluster
(C)
All overlapping regions of a clusters were merged into a single region bed file
MergeBED
(D)
(F)
qq plots to confirm common region replicability
x01.bed.Africa_intersect.bed
x01.bed.America_intersect.bed
x01.bed.Asia_intersect.bed
x01.bed.Europe_intersect.bed
Observed (-logP)
Expected (-logP)
IntersectBED
(E)
Merge Bed file
Plink output/qassoc file

Figure: 6 Schematic representation of GWAS cross-continent replicability by detection of overlapping regions

5. RESULTS

5.1 Genetic factors affecting EBV copy number in lymphoblastoid cell lines derived from the 1000 Genome Project samples

Rajendra Mandage*, Marco Telford*, Juan Antonio RodrõÂguez, Xavier Farre, Hafid Layouni, Urko M. Marigorta, Caitlin Cundiff, Jose Maria Heredia-Genestar, Arcadi Navarro, Gabriel Santpere

PLOS ONE | https://doi.org/10.1371/journal.pone.0179446

(*First Author Shared)

Mandage R, Telford M, Rodríguez JA, Farré X, Layouni H, Marigorta UM, et al. Genetic factors affecting EBV copy number in lymphoblastoid cell lines derived from the 1000 Genome Project samples. PLoS One. 2017 Jun 27;12(6):e0179446. DOI: 10.1371/journal.pone.0179446

6. CHAPTER 2

6.1 INTRODUCTION

This chapter addresses the aspect of studying the interaction of EBV with the host at the immune system level. It is presented in 3 main sections. Section 1 is an overview of the human immune system, and the role of the immune system in detecting pathogens and protecting the body from infection. Section 2 describes the explicit interplay of EBV and immune response. Lastly, Section 3 describes the genetic polymorphisms in EBV that help the virus evade the immune response.

Section 1 Overview of the immune system

6.1.1 What is the immune system

The human body gets constantly in contact with potentially harmful entities such as bacteria, viruses, and other parasites. In many cases, these entities are eradicated to prevent further infection. This is achieved by the body's own sophisticated mechanism, known as the immune system. The immune system is a defense mechanism that consists of a network of complex molecules interacting with each other to defend against invading pathogens. It protects the human host by detecting and eliminating invaders. The human defense mechanism comprises two types of immune response: innate and adaptive. The innate immune response is always considered the first line of defense. Although it is not pathogen-specific, the innate immune response activates immediately after pathogen entry. It always appears before the adaptive response during the first encounter with a new pathogenic entity, providing the first active line of defense in the form of macrophages and neutrophils (Beutler, 2004; Janeway et al., 2001; Mogensen, 2009). When cells begin to die or become damaged due to the presence of infection, signaling molecules known as pathogen-associated molecular patterns (PAMPs) become accessible to the surface proteins. This signaling orchestrates the initial innate response to infection (Mogensen, 2009). The failure of the initial innate response is followed by the adaptive immune response.

6.1.2 The adaptive immune response

The triggering of the adaptive response provides long-term protection in humans. The main component of adaptive immunity is B and T cells. T cells play a crucial role in the recognition of pathogenic antigens. The basic function of this response is to discriminate between a human's self (own peptides) and non-self (pathogenic peptides). Occasionally, the system fails to detect self and non-self peptides, which causes destructive reactivation, such as autoimmune diseases. The substance that elicits the adaptive immune response is known as an antigen. The adaptive response is mainly categorized into antibody-mediated and cell-mediated adaptive responses (**Fig 7**).

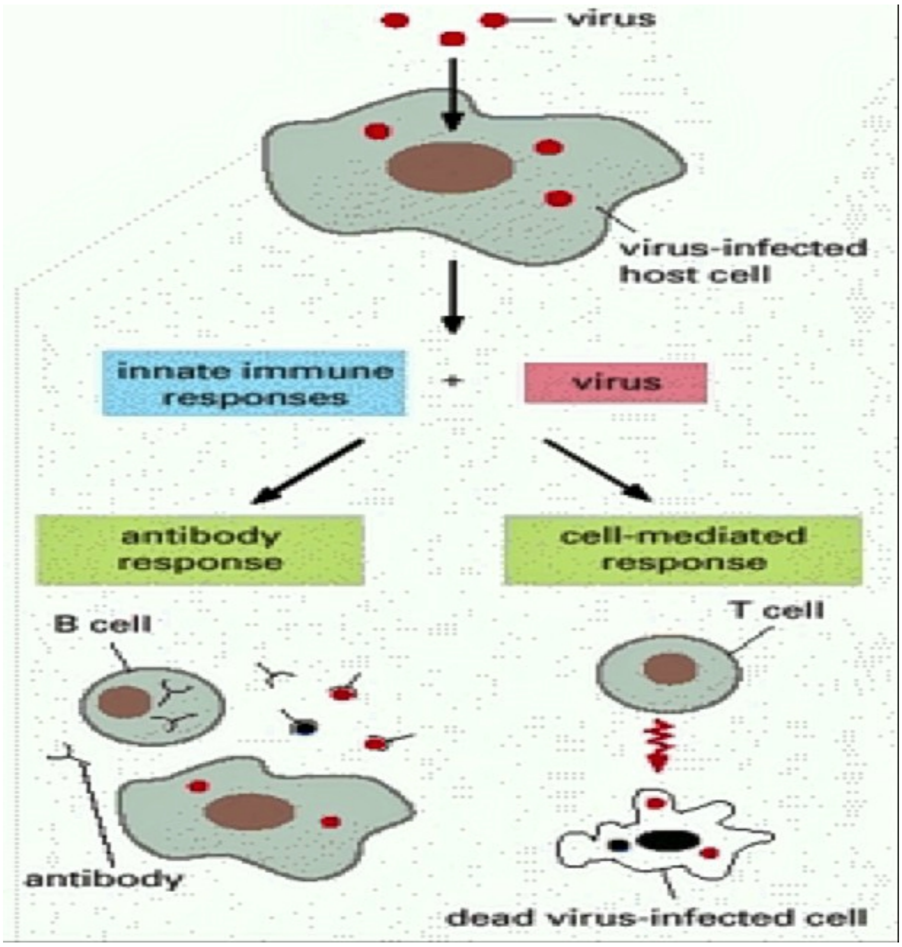


Figure: 7 Schematic diagram of cell-mediated and antibody-mediated immune response to virus (figure has been adapted from Mogensen, 2009).

B cells are active components of the antibody-mediated response, and they bind to viral antigens and act upon infected cells presenting those antigens. In the cell-mediated immune response, T cells bind directly to infectious cells containing the viral antigen, and might kill such cells to eliminate the pathogen (den Haan et al., 2014; Dudley, 1992; Fagarasan and Honjo, 2000; Janeway et al., 2001; Kaiko et al., 2008).

6.1.3 The immune response against virus

When a virus infects human cells, its primary aim is to escape the immune response and invade the cells to survive. The immune system comprises the action of human leucocyte antigens (HLA) molecules to detect the presence of virus inside the cells. MHC molecules process the viral protein fragments from infected cells for CD8+ and CD4+ T cell recognition to kill the virus and limit the infection. This process is also termed antigen processing and presenting. The ability of MHC molecules to detect viruses makes it difficult for viruses to survive and hide inside host cells (Harding, 1991; Jensen, 2007; Mueller and Rouse, 2008; Penn, 2002).

MHC molecules are polygenic and mainly consist of MHC type I and type II alleles, which are located on the short arm of chromosome 6 and include approximately 3,600 kb DNA (Beck and Trowsdale, 2000; Choo, 2007). It is a widely studied human genome region due to its polymorphic nature (Choo, 2007; Horton et al., 2004; Marsh et al., 2000). There are mainly two classes of HLA molecules:

(i) **Class I alleles:** HLA-A, HLA-B, and HLA-C

(ii) **Class II alleles:** HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, and HLA-DRB1

The classes have unique functionality regarding recognition and protection against pathogens. HLA class I molecules are mainly expressed on the surface of nucleated cells, whereas the expression of HLA class II molecules is restricted to B lymphocytes, antigen-presenting cells (monocytes, macrophages, dendritic cells), and T lymphocytes (Choo, 2007). The two classes of HLA molecules have multiple polymorphisms, providing a broader immune response against a wide range of pathogenic invaders.

6.1.4 What is the antigenic variation?

As mentioned earlier, MHC molecule helps detection of the virus inside cells and keeps the cell free from infection. However, in some cases, HLA molecules fail to recognize the presence of virus due to the adaptive characteristic of viruses as described in section 5.2.2. Viruses can develop means of escaping immune surveillance. One way of avoiding recognition is through antigenic variation. **A short peptide sequence from a**

virus recognized by an HLA class is known as an epitope, and the molecule from which this epitope is detected is referred to as an antigen. After the first exposure of virus to the immune system, MHC molecules present some of the viral epitopes for recognition and clearance. MHC binding specificity for such epitopes is preserved in memory cells for later exposure and quick response. More interestingly, T cells recognize only epitopes to which HLA class I and II molecules bind. HLA class I molecules detect epitopes with 8–10 amino acids; HLA class II molecules detect epitopes with 13–17 amino acids (Janeway et al., 2001) (Frank, 2002). The virus can evolve mutating or recombining in these epitope sequences and generates new variants so that T cells cannot detect the epitope, which aids viral invasion (Borst, 2003; Shimizu, 1997). By changing its antigenic sequences, specifically, surface proteins, or expressing new antigenic variants, a pathogen can evade the antibody response to escape attack by the host immune system. Antigenic variation is one of the most common mechanisms by which EBV adapts to evade the immune response (Bell et al., 2012.; Iwakiri et al., 1997; Sun et al., 2015).

6.2 Section 2 Interplay of EBV and immune response

6.2.1 Immune response to EBV infection

EBV has two distinct life cycle phases. In the latent phase, the virus genome is maintained at a constant copy number. In the lytic phase, the viral genome is amplified, and multiple antigenic proteins are produced (lytic antigens). This life cycle is characterized by the continuous interplay between human immune responses and viral escape strategies (Chen, 2011; Merlo et al., 2010). EBV infection is primarily controlled by the human adaptive immune response through continuous monitoring by the immune system. Antigen-specific cytotoxic T cells control primary EBV infection efficiently. EBV infection elicits both antibody-mediated and cell-mediated immune response. The antibody-mediated response also helps for the diagnosis of EBV infection. The T cell-mediated response is vital for controlling infection (Hochberg et al., 2004; Landais et al., 2005; Odumade et al., 2011; Ressing et al., 2008; Rickinson et al., 1997).

6.2.2 Immune invasion strategies of EBV

EBV evolves to establish successful latency and survival within the host, and it has the most sophisticated mechanisms for avoiding elimination by T cell response. These mechanisms include down-regulating highly antigenic proteins, expressing lytic proteins to disrupt MHC molecule antigen and processing machinery, such as through the expression of the lytic protein BNLF2, which blocks the antigen-processing activity of MHC class I molecules (CIT). BGLF5 inhibits the production of new MHC class I molecules via a poorly understood mechanism (CIT). BILF1 also down-regulates the MHC molecules (CIT). Therefore, the interference of the EBV lytic genes can almost disable the functional players of the immune system, resulting in inadequate recognition by T cells. Lastly, EBV can produce proteins that mimic human's self ones to avoid being detected as foreign (**Table 2**) (Merlo et al., 2010; Ressing et al., 2008; Thompson and Kurzrock, 2004). For example, BCRF1 is an EBV homolog to human IL-10, which was found to impair the activity of the IL-10 receptor (Liu et al., 1997). Another EBV encoded protein known as BHRF1, a human homolog of Bcl-2, have been shown to block human B cells from undergoing programmed cell death (Henderson et al., 1993).

Another adaptive mechanism is strain-to-strain variation in epitope sequences (antigenic diversity) to persist in human hosts. A large study has found an extensive sequence variation in the BNLF2 protein from EBV strains isolated from Australian Caucasians, Africans, and inhabitants of the Papua New Guinea highlands and lowlands (Horst et al., 2012). In another study on EBNA1 has demonstrated heterogeneous response of reactivation of monoclonal antibodies against various laboratory EBV strains and EBV isolates directed to the antigenic region (amino acid residues 442-530). EBNA1 from specific isolates was recognizable but not from others due to sequence variation in this area (Iwakiri et al., 1997). There is also a report on sequence antigenic heterogeneity in EBNA1 representing viral isolates circulating within UK and US populations (Wrightham et al., 1995).

6.2.3 EBV epitope variation, geography, and T cell-mediated immune response

T cells are critical players in adaptive immune response control of EBV infection. A T cell detects a short peptide sequence derived from EBV proteins via MHC alleles as a foreign protein and destroys the cells containing EBV to control the infection. To escape the immune system, the virus had to rely on altered form of epitope sequences to avoid recognition by the T cell response. The changes in the amino acid sequence of the epitope region block the recognition by MHC molecules and the consequent viral clearance (Couillin, 1994; D et al., 1998; Khanna et al., 1997). Therefore, extensive studies are underway to detect the EBV genetic variation that could influence the T cell response. A very early study in 1993 on EBV strains from HLA-A11 positive coastal Caucasian population demonstrated that a change in a single amino acid in a specific epitope sequence (IVTDFVIK) resulted in the abolishment of MHC recognition. Surprisingly, no amino acid change was observed in the same study with other populations carrying different HLA alleles (de Campos-Lima et al., 1993), hinting to regional differences in EBV epitopes due to selective pressure given by different HLA alleles. Contrary to those results, the epitope analysis of EBV isolates from Papua New Guinea revealed the presence of the identical substitution of epitope sequences for HLA alleles such as HLA-A11, B-35, and B-8 from EBV isolates from different regions, thus demonstrating the presence of a single dominant EBV strain in this population (Burrows et al., 1996). A study on a Chinese A11-positive population also showed the HLA-A11 restriction polymorphism for two epitopes, IVTDFSVIK and AVFDRKSDAK, supporting the possibility of virus-host coevolution (Midgley et al., 2003).

Additionally, recent works have shown a peculiar pattern of T cell response during IM, where almost 50% of cellular response is detected against lytic antigenic proteins such as BZLF1 and BRLF1, where the late or delayed response is shown in the case of latent antigens. Latent proteins account for only 5% of the cellular response, which is more specific to the EBNA-3A, EBNA-3B, and EBNA-3C proteins (Abbott et al., n.d.; Hislop et al., 2007, 2005).

To date, extensive work has been carried out to characterize genetic variation in EBV epitope sequences from lytic and latent proteins as a T cell response to EBV. The sequencing of 71 EBV genomes from healthy samples and multiple tumors demonstrated that two EBNA-3A epitopes (RRLHRLLLMR, SVRDRLARL) were conserved across all sequenced genomes and that other six epitopes (RRFPLDLR, QAKWRLQTL, KRPPIFIRR, RPPIFIRRL, VPAPAGPIV, RLRAEAQVK) showed variation specific for EBV type 1 and type 2 sequences. The other epitopes showed multiple variations at different frequencies in 71 isolates. A similar pattern was also observed with EBNA-3A and -3C proteins (**Fig 8)** (Palser et al., 2015). A recent work on the comparative analysis of 22 EBV genomes from healthy and different EBV associated diseased samples detected considerable variations in the EBNA-1, -2, -3A, -3B, -3C, LMP-1, -2, BLLF-1, and BZLF1 antigenic proteins, and observed that EBV strains from the same populations share some of these epitope variations (Zhou et al., 2017). Another work on the polymorphic region of BZFL1 suggested four overlapping epitopes influenced by a single amino acid change and 11 novel epitopes have been identified (Rist et al., 2015). Moreover, a functional assay of LMP-1 variants has identified a HLA-A2-restricted epitope-loss variant of LMP-1 in some EBV strains from southern China and Taiwan in patients with NPC (Lin et al., 2005).

Hence focusing on T cell responses toward antigenic variation among worldwide EBV strains would help to control EBV-associated infections. A systematic analysis of antigenic variation response would improve our understanding of various unexplored aspects of the interaction between EBV and host, immune response, immune system modulation by the virus, and virus escape strategies and control, as humoral and cellular responses against the virus are a crucial component for controlling viral primary infection and replication.

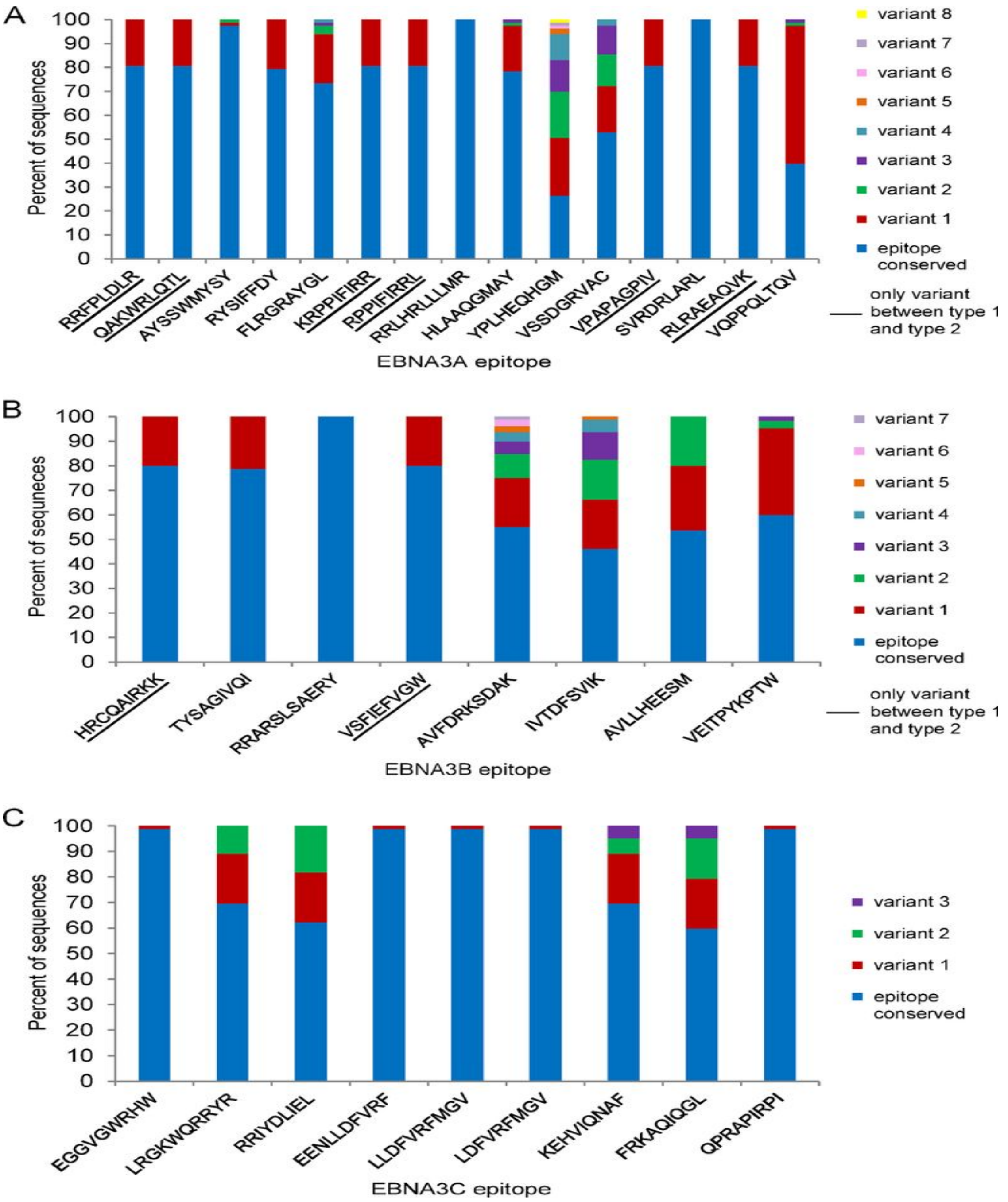


Figure: 8 EBNA-3 genes epitope variation across 83 EBV genomes. Graphs indicate the percentage of EBV genome sequences with fully conserved epitopes (blue) and the percentage of sequences that have each of the variant sequences as a stacked histogram. Some epitopes are fully conserved across all strains (fully blue bars), some have differences only between type 1 and type 2 (underlined), and some have multiple variants (Figure and text have been adapted from Palser et al., 2015).

Section 3 HLA and EBV associated disease

6.3 HLA allele variation, geography, and disease association

The HLA genes have remarkable variability. This variation increases their epitope recognition repertoire in a human host. HLA polymorphism also serves as a molecular marker for population genetics and wide-ranging areas of adaptive immunity, such as autoimmunity, HLA-associated studies, transplantation, infectious diseases, autoimmune disease, and vaccine development (Carrington and O'Brien, 2003; Fernando et al., 2008; Holmans, 2001; Mizuki et al., 2010; Rioux et al., 2009; Ryder et al., 1981; Shugart et al., 2011; Warren et al., 2012; Sanchez-Mazas et al., 2014; Sanchez-Mazas and Meyer, 2014; Seitzer et al., 2002; Shiina et al., 2009; Trowsdale and Knight, 2013). Previously published data on HLA have shown that the HLA allele distribution varies among ethnic groups living in different geographical areas (Buhler and Sanchez-Mazas, 2011; Chen et al., 2007; Edinur et al., 2009; Jinam et al., 2010; Maiers et al., 2007; Marsh et al., 2000).

These HLA regions have been extensively studied and provide more association signals with diseases such as HIV, cancer, hepatitis and autoimmune disease (Cotsapas et al., 2011; Cozen et al., 2012; McCormack et al., 2011; Pereyra et al., 2010; Raychaudhuri et al., 2012; Rioux et al., 2009; Sawcer et al., 2012). For example, a large-scale GWAS of 14,000 cases demonstrated that HLA region variation is strongly associated with diseases such as rheumatoid arthritis (RA) and type 1 diabetes (T1D (Wellcome et al., 2007). Another whole-genome association study also emphasized the importance of HLA polymorphisms in controlling HIV-1 infection (Fellay et al., 2007). HLA-B27 alleles are associated with increased risk of ankylosing spondylitis, an inflammatory joint disorder (Khan et al., 2007). The variation in the DRB1*1501 allele is linked with MS (Compston et al., 2006; Luckey et al., 2011; Olerup and Hillert, 1991). In allogeneic tissue and cell transplantation, it is essential to match HLA alleles between the donor and recipient (Opelz et al., 1999; Petersdorf, 2008). There is much evidence correlating HLA allele variation and disease association, but the exact roles of these variations and their contribution to disease progression, inflammatory response, and immune response have in many cases yet to be fully elucidated.

6.4 Homozygous and heterozygous HLA alleles and disease

The essential biological function of HLA class I and class II alleles is to process exogenous and endogenous peptides to present these peptides to MHC molecules to trigger effective T cell responses against viral infection (Hislop et al., n.d.; Kaur and Mehra, 2009; Parkin and Cohen, 2001). The specificity and diversity of HLA alleles indicate the molecular interplay in infectious diseases. There is much evidence regarding the specificity and heterozygosity of HLA molecules, where a specific HLA class can cumulatively alter viral disease progression, such as HLA class I alleles (HLA-A, HLA-B, HLA-C); heterozygosity confers a selective advantage against AIDS and hepatitis C viral infections due to an increased spectrum of antigens, consequently helping to trigger a broader T cell response (Blackwell et al., 2009; Carrington, 1999; Carrington and O’Brien, 2003; Hraber et al., 2007; Machulla et al., 2001; Rousseau et al., 2009; Shah et al., 2011; Tang et al., 2002).

6.5 1000 Genome Project and HLA polymorphism

The 1000 Genome Project is an excellent collection of human genome variation at a global level, representing a source for genotype-to-phenotype in LCLs. This data set includes human populations from Asian, European, American, and African continents and serves as a primary reference source for genetic studies (Altshuler et al., 2012).

NGS technology can be used as a tool for elucidating genetic heterogeneity in HLA types (Gourraud et al., 2014; Hosomichi et al., 2015; Warren et al., 2012; Yin et al., 2016). There are several NGS methods for high-throughput HLA typing, such as PCR-based HLA sequencing using NGS or HLA enrichment methods coupled to sequencing, **(Fig 9**). Researchers have developed HLA-typing softwares from whole genomes and exomes data **(Fig 10**) (Bentley et al., 2009; Danzer et al., 2013; Gabriel et al., 2009; Lank et al., 2010; Lind et al., 2010; Moonsamy et al., 2013; Yin et al., 2016).

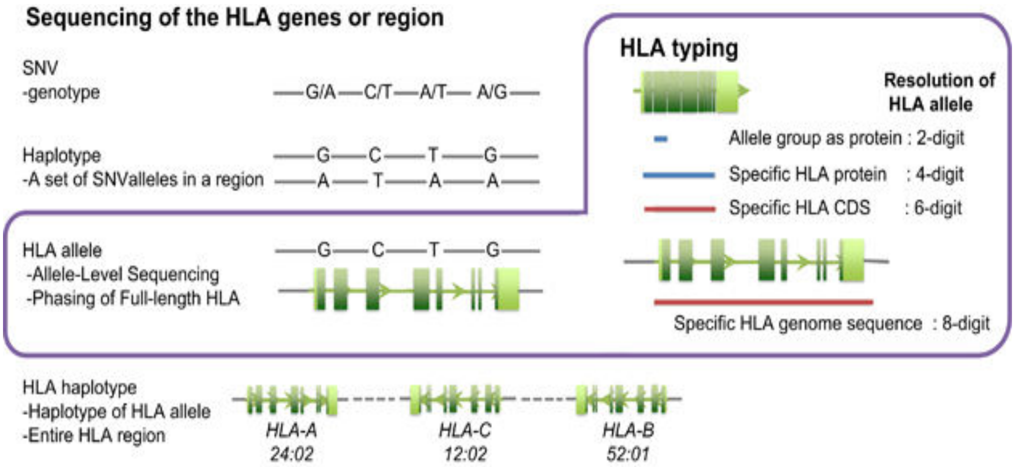

Sequencing of the HLA genes or region
SNV
-genotype
—G/A—C/T—A/T— A/G—
Haplotype
-A set of SNValleles in a region
—G—C—T—G—
—A—T—A—A—
HLA allele
-Allele-Level Sequencing
-Phasing of Full-length HLA
—G—C—T—G—
HLA typing
Resolution of HLA allele
Allele group as protein : 2-digit
Specific HLA protein : 4-digit
Specific HLA CDS : 6-digit
Specific HLA genome sequence : 8-digit
HLA haplotype
-Haplotype of HLA allele
-Entire HLA region
HLA-A 24:02
HLA-C 12:02
HLA-B 52:01

Figure 9 Overview of sequencing of HLA genes and regions for detecting HLA alleles **(**Figure has been adapted from Hosomichi et al., 2015)

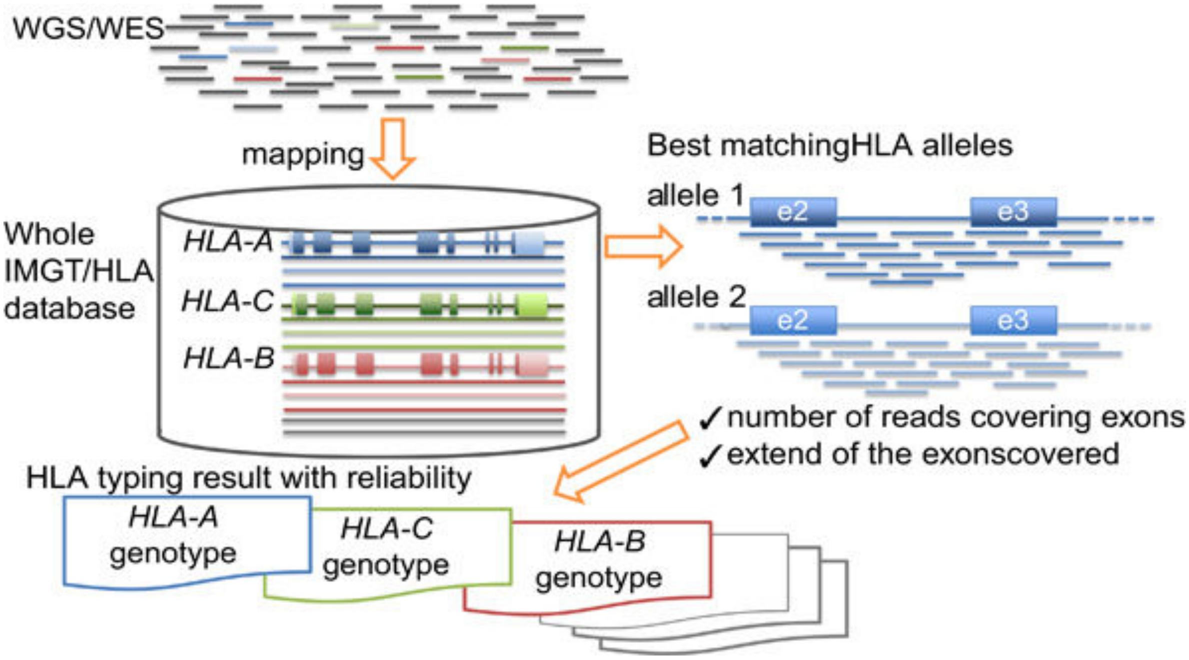


Figure: 10 Overview of HLA allele detection using whole-genome and exome sequencing data **(**Figure has been adapted from Hosomichi et al., 2015)

6.6 Background of HLA allele EBV copy number analysis

The incidences of EBV-associated diseases, as well as their prevalence, vary significantly across human population. For example, BL is most common in Africa, whereas NPC is more prevalent in Asia. Moreover, the HLA region on chromosome 6 contains many associations signals with EBV-related disease phenotypes (Houldcroft and Kellam, 2014; Rubicz et al., 2013). The extreme diversity of HLA allele peptides might influence the recognition of EBV antigens; therefore, there is always variation in the immune response against EBV infections. A genotyping study performed on the Dutch population confirmed the genetic association of HLA-A1 and HLA-A2 in EBV-positive classic HL (Huang et al., 2012). HLA allele variation is also associated with risk of IM and EBV-positive HL (Diepstra et al., 2005.; Niens et al., 2007). A GWAS identified multiple strong associations of EBV anti–EBNA-1 antibody count with genetic factors located in the HLA region (Rubicz et al., 2013). A combined study of linkage and association analysis demonstrated that HLA class II variation controls the level of antibody response against EBV antigens (Pedergnana et al., 2014). Although antibodies against EBV antigens may not directly reflect the EBV copy number (EBV load), it is essential to understand which HLA alleles (if any) control the EBV copy number in EBV-infected individuals.

Accordingly, we attempted to detect the HLA variants associated with EBV copy number from 1000 Genome Project samples to study whether variation in EBV copy number across European, African, American, and Asian populations affects recognition by HLA alleles. This type of studies would help to understand the ethnic variation in EBV copy number and HLA polymorphism, and their linked to EBV-associated pathologies.

6.7 MATERIAL AND METHODS

This section has ben divided into two parts. Part 1 describes the material and methods used to perform HLA-typing on the 1000 Genome Project samples, and their association with the EBV copy number variation. The second part describes the antigenic variation analysis of EBV in relation to the strain geographical origin.

Part 1: Computational derivation of HLA class I and II alleles

We retrieved 1000 Genome Project BAM files for 856 samples consisting of African, American, Asian, and European populations. The HLAminer software was used to predict HLA class I and class II alleles. HLAmoner is an automated algorithm explicitly developed for HLA predictions from NGS data (Warren et al., 2012) (Fig 10). All HLAminer-predicted HLA alleles were defined to 2-digit resolution like HLA-A*02 for further analysis.

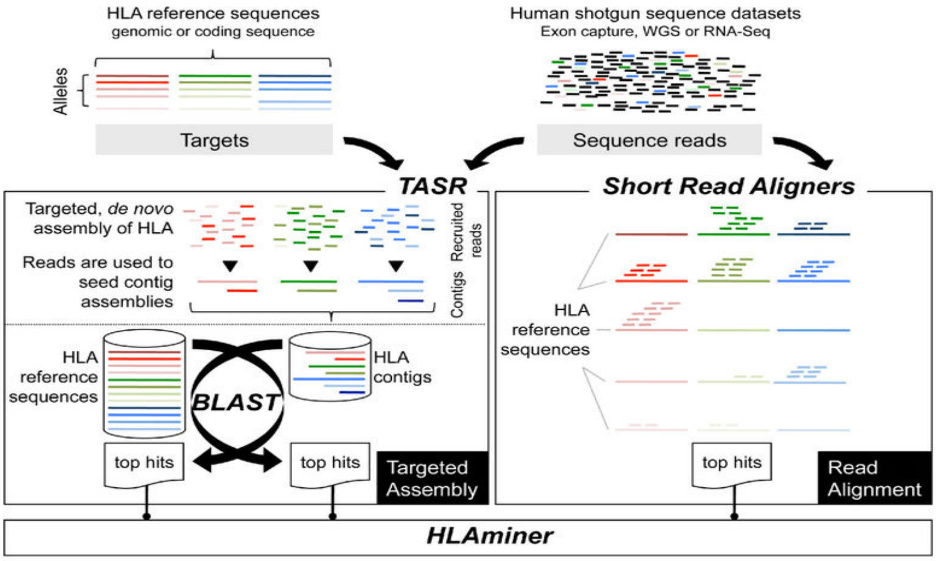


Figure: 11 Diagrammatic representation of HLAminer tool algorithm workflow (Figure has been adapted from Warren et al., 2012)

EBV copy number and HLA association test

Using R, a simple linear model was built to assess the association of the individual sample's HLA alleles with EBV copy number from African, American, European, and Asian populations.

Homozygous and heterozygous HLA association with EBV copy number

We compared viral copy number association with homozygous versus heterozygous alleles from individual genome samples using the linear regression model, including the first 10 MDS dimensions as variables.

Part 2: Detection of natural variation in EBV epitope sequences

Collection of EBV epitopes from IEDB

The Immune Epitope Database (IEDB; http://www.iedb.org/) was explored to retrieve T-cell epitopes encoding sequences with HHV-4 (EBV) as source organism (NCBI taxonomic ID 10376). The other parameters were set to host human, assay T-cell positive assays only, MHC restriction any, any disease, and any reference to retrieve specific T-cell epitopes only. As the IEDB database also provides epitope start and end position with reference to antigen sequences together with its corresponding HLA allele, a catalog has been generated using these data for each IEDB epitope sequence. In total, a catalog of 430 epitope sequences representing 30 EBV antigenic proteins were collected and used for antigenic diversity analysis.

EBV whole genome collection

EBV whole genomes are rapidly increasing due to advances in NGS, representing a wide array of geographic origins and diseases. The present study revolves around a wide array of EBV proteins representing a panel composed of the 123 EBV isolates described to date (Kwok et al., 2014; Lei et al., 2013; Liu et al., 2016; Palser et al., 2015; Santpere et al., 2014; Tso et al., 2013; Wang et al., 2016), representing African, American, Asian, Australian, and European populations. The data set includes the nucleotide sequences of samples derived from patients affected by BL, EBV-associated gastric carcinoma (EBVaGC), HL, NPC, IM, and PTLD, as well as part of their protein sequences.

EBV proteome retrieval

Proteome sequences (conceptual translation provided by GenBank for 123 EBV genomes and partial protein sequences) were retrieved from the Genbank database filtering for “Human gammaherpesvirus 4” or “EBV”. We obtained 14,657 protein records in identical protein report formats. The identical protein report from NCBI contains information about all proteins identical to the query protein and provides data on the source, nucleotide accession numbers, start and stop positions of the protein, protein name, and source organism. We filtered out sequences with redundancy, without source information, patented sequences, and Protein Data Bank (PDB) entries. A final dataset consisting of 6786 EBV protein sequences was downloaded in FASTA format using Entrez batch download utility.

Multiple sequence alignment of antigen proteins

Multiple sequence alignment of the 30 antigenic proteins was carried out using T-Coffee tool (Notredame et al., 2000). In order to detect epitope conservation and/or variation at an amino acid level, we used python script that scans for mutations within the multiple sequence alignment (MSA) of a particular protein from different isolates that appear at different position with respect to epitope region.

6.8 RESULTS

Part 1 HLA allele-EBV copy number association analysis

EBV copy number and HLA association test

We did not observe any significant alleles associated with EBV copy number (Supplementary Data Table 1). This could have been due to the relatively minimal samples size (n = 856).

Homozygous and heterozygous HLA association with EBV copy number

Many EBV-related diseases have been shown to be associated loci in the HLA region on chromosome 6. Therefore, we performed association analysis of different HLA-A, -B, and -C alleles among the 1000 Genome Project individuals with our estimated viral copy number. No particular allele resulted associated with viral copy number. We also tested the HLA zygosity, under the rationale that LCLs with heterozygous HLA alleles might show overall different infectiveness than homozygous alleles. However, we did not observe any significant difference in copy number between HLA homozygous or heterozygous LCLs (Fig 12).

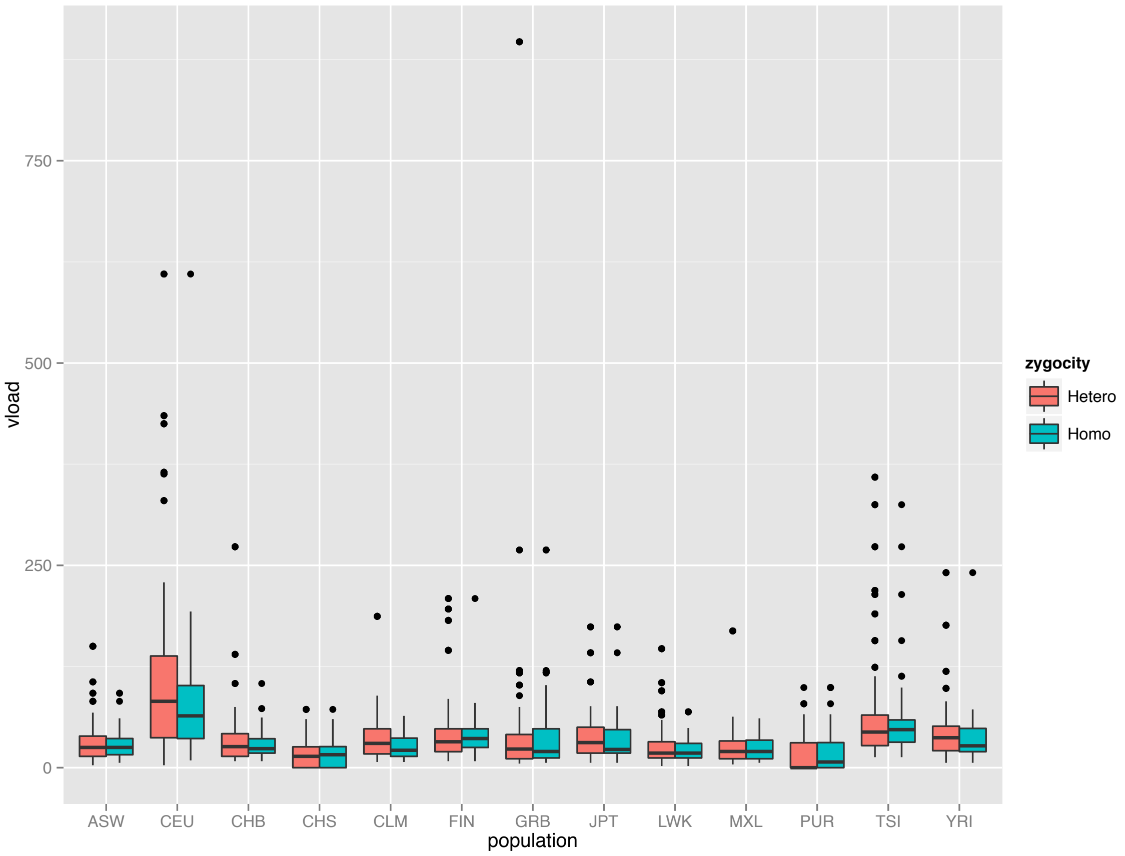


Figure: 12 EBV copy number distributions among 1000 genome populations and predicted homo and heterozygous HLA alleles for HLA-A allele

Part 2 Detection of natural variation in EBV epitope sequences

Genetic variation in 30 antigenic sequences

Genetic variation within IEDB specified epitope sequences comprising different regions was analyzed to detect amino acid changes present in a panel of EBV 123 genomes. After performing multiple sequence alignment (MSA) of each antigenic protein, we identified mutations at various positions within the MSA. We divided these variations into 4 categories on the basis of amino acid changes they convey: conserved positions (conservation in antigenic region of MSA), variant_1 (single mutation in antigenic region of MSA e.g. L<R), variants_2 (two mutations antigenic region of MSA e.g. I>V and R>Q), and variants_3 (three mutations in antigenic region of MSA e.g. P>T, V>H and K>R). Indeed overall comparative analysis of the amino acid substitution within antigenic regions revealed that there are 639 epitope regions with single mutations, 175 regions with two mutations, and 61 positions with three mutations detected within the alignment, while the rest of the 6033 positions were found to be conserved among 123 EBV isolates. We also categorized these mutations according to EBV type 1 and type 2, as well as life cycle phases. EBV type 1 strains resulted more conserved than type 2, whereas type 2 has demonstrated the presence of more variant_3 instances than type 1 similarly, lytic proteins were found to be more conservative than latent (Table 4). It could be due to the fact that antigenic variation might help the virus to evade the immune system during latency.

EBV type	Conserved	Variant_1	Variant_2	Variant_3
EBV type_1	5883	537	137	23
EBV type_2	631	106	40	38
Latent	2837	446	124	60
Lytic	3677	197	53	1

Table: 4 Antigenic variations according to EBV type 1 and 2, and lifecycle phases

EBV strain-specific and shared mutations detection

To detect EBV strain-specific antigenic pattern, we grouped mutations within epitope regions by 123 EBV strains and their associated diseases. This grouping yielded 47 amino acids mutations pattern that are shared by EBV strains and diseases. For example, the amino acid substitution from I>L in EBNA-3A protein is shared by 72 different EBV isolates whereas 63 strains shared substitution from A>V from BNA-3 protein. We also observed disease-specific antigenic variation pattern such as LCL2 and LCL3 isolates from lung carcinoma that have a common F>I mutation. The amino acid changes from L>R is also common to NCP disease. Interestingly 24 amino acid changes that are unique to EBV strains were also found (table 5).

Mutation	Protein name	Cycle	Strain name	Disease	Pop
11FY	EBNA-1	Latent	AG876	BL	Africa
5GE6EK	BARF1	Lytic	BL37	BL	Africa
4LI	BZLF1	Lytic	BL37	BL	Africa
1DG2TS	EBNA-2	Latent	BL37	BL	Africa
6LM	BSLF2/BMLF1	Lytic	Daudi	BL	Africa
20QH	EBNA-3C	Latent	EBVaGC4	EBVaGC	China
9LS	LMP-2A	Latent	EBVaGC4	EBVaGC	China
7NT	LMP-2A	Latent	EBVaGC6	EBVaGC	China
2GV	LMP-2A	Latent	EBVaGC7	EBVaGC	China
5RQ7SL9LP	BRRF2	Lytic	H002213	BL	Africa
1AV	BLRF2	Lytic	HL11	HL	UK
4LF11IL	EBNA-3A	Latent	L591	HL	Germany
12AS13RQ	BSLF2/BMLF1	Lytic	sLCL-1.12	sLCL	Africa
7HY	glyco_350	Lytic	sLCL-1.12	sLCL	Africa
3VG14TP15KR	EBNA-1	Latent	sLCL-1.17	sLCL	Africa
3ST	BZLF1	Lytic	sLCL-1.24	sLCL	Africa
8MV	EBNA-1	Latent	sLCL-2.15	sLCL	Africa
9MV	EBNA-1	Latent	sLCL-2.15	sLCL	Africa
3ED	EBNA-1	Latent	sLCL-IS1.01	PTLD	Australia
6MV17DE	EBNA-1	Latent	sLCL-IS1.04	PTLD	Australia
8RH	BARF0	Latent	sLCL-IS1.10	PTLD	Australia
8NS	EBNA-3C	Latent	sLCL-IS1.11	PTLD	Australia
7GS	EBNA-3C	Latent	VGO	BL	Brazil
5RH	EBNA-3C	Latent	Wewak1	BL	PNG

Table: 5 List of unique antigenic variation observed among few EBV strains

7. DISCUSSION

This section is divided in three main parts. In the first part, I will be describing the results of the EBV copy number and GWAS studies that aim to uncover the genetic variation present in the 1000 Genome Project populations, as well as their association with EBV disease susceptibility. In the second part, I will be discussing antigenic variation in epitope sequence and evasion of the immune response due to the alteration in amino acid sequence and HLA recognition.

The final section, I will discuss the global EBV-host interaction to uncover molecular mechanisms to link genetic variation with EBV associated diseases and therapeutics development.

Part I EBV copy number and GWAS studies (Human genome variation analysis)

Why is EBV-host interaction analysis at population level important?

From the beginning of medicine as a science, physicians had recognized that the clinical outcome of a disease often varies significantly from person to person. The cause of this variation may lie partly in both the host and pathogen genetic architecture, at least in diseases caused by a biological agent. Various approaches like candidate gene screening, genome-wide association analysis, and large-scale *in vitro* genome study have provided evidences that host genetic factors can influence the control of and/or the resistance to specific disease (Antonelli and Roilides, 2014; Hill, 2012; Loeb, 2013; Qureshi et al., 1999). The complete understanding of the interaction of host genomic variation with EBV pathophysiology demanded the determination of whole genome sequences of both human host and EBV at a large scale. This would allow to characterize genome-wide variation, and would be especially focused on samples from healthy and diseased individuals when aiming to understand the relation between EBV and its putatively associated diseases (accounting for type 1 and type 2 differences). To date most of the analysis focused on the detection of the variation of a limited set of genes (EBNA and

LMP are widely studied till now), only considering specific diseases, and taking into account only a restricted number of populations. Therefore there is strong urge to go beyond the single variant, single gene, or single population to study the effect of a single SNP on a gene function in order to understand host-EBV interaction fully. The advancement of next-generation technology (NGS) has made it possible to go beyond the single gene to elucidate the role of genetic variation in disease prevalence. Hence genome-wide scale analysis of host-EBV interaction using more significant populations and continents dataset, and considering almost a complete proteome of EBV, allows us to do a comprehensive characterization of EBV genetic polymorphism associated with different malignancies and with different geographical regions.

Hypothesis for GWAS analysis

The primary aim of this PhD work was to detect underlying genetic factors giving susceptibility to Epstein Barr virus-associated diseases using lymphoblastoid cell lines (LCL) as a surrogate model. EBV is etiologically linked with infectious mononucleosis, multiple sclerosis and the development of several cancer types. Although it infects more than 90% of world populations, most of the infected individuals coexist with EBV in asymptomatic stage. However, in some cases, the infection may influence the development certain pathologies. EBV associated diseases present distinct patterns. There have been numerous pieces of evidence showing remarkable geographical differences in the prevalence of EBV related diseases in human host populations. For example, In African populations, BL is commonly observed while in Asian populations such as South China's and some part of South Asia's ones, NCP incidence is more common. However, NCP is very rare in the Arabic regions of North Africa and the Arctic, than in other areas of the world (Bray et al., 2008; Chang and Adami, 2006; Hsu and Glaser, 2000; Lung and Chang, 1992). Burkitt lymphoma (BL) has higher incidence in equatorial Africa than rest of the world population (Bray et al., 2008; Mbulaiteye et al., 2009). Because of this, the following questions were raised, and tried to answer during my PhD work *(i) what make a particular population susceptible to these diseases? Is the presence of specific genetic variants in the populations making it more susceptible to EBV infection?*

There is considerable data available now that showed the presence of genetic variation as the critical factor in the susceptibility to these EBV associated diseases, but it had been mostly studied in individual genes instead of at a genome-wide scale. The substantial genetic variation in human genome sequence, as well as the variation in EBV isolates along with environmental factors may explain the geographical variation in disease outcomes. We tried to explain it in the following points.
(i) The genetic architecture of the human host might be giving rise to differences in disease risk (presence or absence of particular genetic variants). (ii) The genetic variation in EBV genomes, such as the one between type 1 and type 2, might be offering a different pathogenicity. (iii) The unusual pathogenic interactions between human host and different EBV strains might be the reason for the prevalence of EBV infection in a particular population.

We thought that one of the genetic factors that points underlying susceptibility to EBV diseases could be explained by EBV copy number (EBV load) in the human host. Therefore we hypothesize in our paper that differences in EBV copy number among individuals coming from 1000 genome project LCLs may reflect differential susceptibility to EBV infection (Mandage et al., 2017). To test this hypothesis, we took advantages of 1000 Genome Project samples since most of the samples derives from LCLs. The LCL is acting as a surrogate system to study the latent infection.. This approach grounds on the hypothesis that human genomic variation associated with EBV copy number from transformed LCLs might point to new candidate genes related to EBV-associated pathologies. An earlier comparative study on EBV copy number had already proven that genetic basis of regulatory variation in LCLs could be used as the proxy model to study B-lymphocytes gene expression patterns (Çalişkan et al., 2011). A previous study also tried to translate this idea to a GWAS analysis to explore the impact of host genetic factors on EBV copy number on 798 LCLs samples from HapMap. Nevertheless, this study could not find any significant association (Houldcroft et al., 2014). Here we aggregated 1753 LCL genomes from worldwide populations, interrogated explicitly whether EBV copy number is a stable phenotype in an LCL, estimating *in silico* EBV copy number, and performing GWAS analysis to examine the

EBV-host interaction at the individual population level to scrutinize the host response in the context of genetic architecture to EBV pathologies.

LCL genome sample retrieval and genome source reliability assessment

To begin with LCLS samples collection, we downloaded the 1000 Genome Project Phase3 aligned 2,535 BAM files representing 26 different worldwide human populations. The primary task was here to check the source of the samples since we were only interested in unrelated genome having LCLs as exclusive DNA source (we excluded blood-derived samples). This confirmation was achieved in two ways. (i) First, annotations information given by the 1000 Genome Project was searched to verify DNA source of every genome samples. (ii) Direct communication with the technical dept of 1000 Genome Project provided us information on DNA source on few samples. This lead us to exclude 367 genome samples due to the certain or possible derivation from blood. There are also 179 samples from ACB, KHV, STU, PUR, and PEL populations for which we couldn’t confirm the source (as no annotation wasn't available and no confirm data source was provided by 1000 Genome Project technician). Hence, these samples were also not considered in the GWAS study. We created five datasets of 1753 samples for European, American, Asian, African and all population combined to estimate EBV copy number and GWAS analysis for each of this data set. Afterwards using experimental (RT-PCR) and computational (GWAS) approaches, we detected several genetic variants in the genomes from African, American, Asian and European population dataset.

Confirmation of EBV copy number as stable phenotype over time

1000 Genome Project samples are mostly derived from shotgun sequencing of EBV-transformed LCLs and hence it serves as an ideal data set for us to estimate the EBV copy number of SNP-genotyped individuals. It was one of the major task of the GWAS project to check the stability of EBV copy number within these LCLs. While it is known that EBV copy numbers varies greatly between individual EBV-transformed LCLs, the stability in time of this trait has not been proven. Therefore, we decided to check the phenotypic stability of viral load. To test the stability, we used 7 LCLs derived from the 1000 Genome Project samples and estimated the viral load using real time PCR. EBV-infected LCLs were cultured, and after every 3-4 days the DNA from part of each cell line culture was extracted. The samples extracted from a same cell line at different time where used to test the stability of the viral load by relative quantification of EBV DNA in the samples. The quantification was performed on all cultured cell lines, and the resulting variance was compared using ANOVA. This result showed a significantly larger proportion of the overall variance explained by the different cell lines than the time points. This result confirmed the phenotypic stability of EBV copy number, supporting its use in a GWAS (Figure: 13).

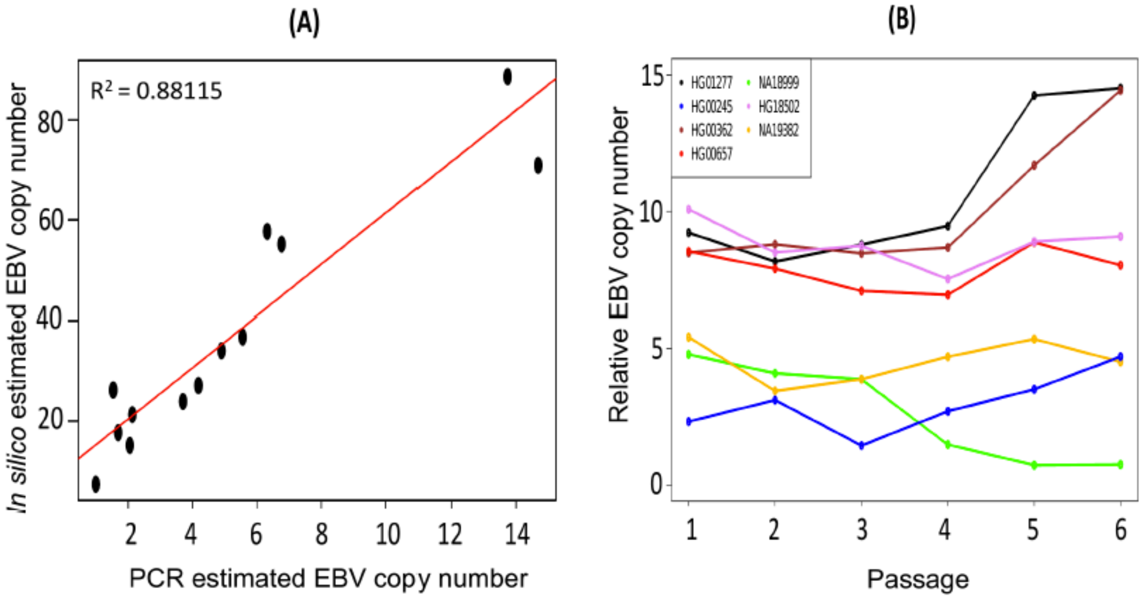


Figure: 13 The plot shows the correlation of EBV copy number estimations by RT-PCR and *in silico* methods **(A)**. The relative RT-PCR estimation of EBV copy number in 7 LCLs (indicated in different colors) for 6 passages showing the stability within the LCL compared to inter-strain variation **(B).**

Biology and clinical significance of EBV copy number estimation in LCL genomes

It is essential to understand that EBV copy numbers in saliva, blood, plasma, and serum might not be related with the ones in transformed LCLs, and the variation in these two measures may indicate different biological process. The significance of EBV DNA copy number quantification from LCLs might be the consequence of specific biological events of EBV–host interactions such as viral entry into B cells, EBV infection, lytic reactivation process, an episomal establishment with host genome, and finally B cell to LCL transformation as shown by our recent publication (Mandage et al., 2017)

Inference from EBV copy load estimation

We estimated EBV copy number for 1753 samples from 1000 Genome Project using our algorithm. EBV copy number ranged from 2 to 500 copies/LCL, and interestingly a substantial variation has been observed across the populations and continents. For example, the European population had shown a significantly higher EBV copy number than the non-European populations. EBV copy number was higher particularly in Iberian (IBS) and Utah (CEU) than in Finish (FIN) and British (GBR) subset of European populations. However, no difference was found between male and female samples.

Inference from genome-wide association analysis

We conducted here the most extensive GWAS analysis as a part of my PhD thesis to discover genetic heterogeneity among human population and the association with the prevalence of EBV infection. To suggest the possible connection, we took the advantages of the helpfulness of Genome-wide association studies (GWAS). We created a set of 5 populations by dividing 1753 samples into America, Asia, Africa, Europe and all population combined together to correlate genetic variation of the human host with EBV disease. We used EBV copy number as the proxy phenotype to study the influence of underlying genetic variation associated with it. Our work has detected few population-specific association signals in a genome-wide association study. However, only a single SNP encoding gene *MACROD2* surpass GWAS threshold. This gene has shown an active link with EBV infection such as gastric cancer. By using a region-based analysis, we have detected several other possible genes in our research article published recently, particularly DKK2, CSMD1 and CAND1, KHDRBS2, FGD2, NRG3, PIK3CB. These genes are mostly involved in cell cycle control and regulation transcription process related to cell signaling pathways like WNT, GPCRs, RHO GTPases, and interleukin receptor SHC signaling pathways. Some studies on NCP and other cancers have already shown that the deregulations of these pathways are implicated in EBV associated lymphomas.

Population-specific GWAS signals discovery

The fascinating result here we present, the most of GWAS hits are found to be population specific. We described EBV copy number associated genetic variants nearby to the genes that may give the most plausible explanation about EBV *in vivo* biological behavior and eventually EBV infection biology. It is remarkable that occurrence of population-specific genes was confirmed by using the VEGAS2 tool and it gave the similar GO terms categories for cell adhesion for Europeans and Americans populations. The EBV protein, LMP1 is thought to be regulating the cell adhesion functional modality (Morris et al., 2016) which is related to the entry of EBV to B-cell. However, this work was statistically powered to detect shared genetic variants across the America, Asian, European and African population. To identify the genetic variants that may share across the different population, if there is any, a large number of samples from distinct populations are needed.

GCTA, Data normalization, and phenotypic variance

To quant**ify** the proportion of phenotypic variance explained by all GWAS SNPs associated with EBV copy number, we used GCTA tool. We added first ten dimensions of a multidimensional scaling (MDS) of the identity-by-state matrix to account for population structure. Using a raw (untransformed) EBV copy number, we got a proportion of variance of 0.78 (n = 1730, SE ± 0.16, P = 9.076e-07) for all populations combined dataset. This estimate was consistent with the 0.65 of variance detected by previous GWAS analysis with 677 genomes samples (Houldcroft et al., 2014). However, this 0.78 variance value was seriously affected by the rank transformation method in another attempt (**table 5**). In which we normalized data by $\log_{10}$ and inverse rank transformation methods because EBV copy number is wide in range and variable across the populations, we decided to do normalization, and we obtained very low estimates of the proportion of genetic heritability.

Table 5 The phenotypic variance comparison with various transformations methods

Population/normalization method	N	Heritability	SE	P-value
All_populations_raw_value	1753	0.779485	0.157138	9.076e-07
All_populations_log$_{10}$_value	1753	0.377112	0.158701	0.006619
All_populations_rank_inverse	1753	0.048678	0.161480	0.3783
All_populations_squared	1753	0.809845	0.160508	1.251e-06

The reason might be that estimation obtained using untransformed data (raw and log$_{10}$) may not account for the structure which might be resolved by a population-wise inverse rank transformation. Therefore it is conclusive that the data transformation method significantly affects the proportion of the variance in liability to the phenotypic trait.

What are the main outcomes of this work?

(1) We attempted to estimate EBV copy number at the individual population level as well as continent level, and we have shown there is a clear evidence of variation in EBV copy number across 19 populations from 1000 genome populations well as continents. Thus, this could highlight the underlying interaction of population-specific variants with EBV copy number and associated EBV pathologies.

(2) We made the large-scale effort towards the determining role of human genetic variation playing in EBV related infections and detecting novel variants and genes that

might be the playing a pivotal role in disturbing normal cellular function that results in EBV infection.

(3) Although we observed a low level of inter-population replicability, it demonstrates the presence of population-specific genetic variants that can influence viral copy number.

(4) Our work also raised the possibility of linking EBV copy number variation in transformed B cell lines to specific genetic variants, as many of GWAS variants we detected are located near genes playing a crucial part in cell cycle control and cell signaling pathways and EBV-linked infections.

(5) The data transformation is another essential factor need to consider while doing GWAS with diverse population samples. As the genetic variance explained by it is highly dependent on transformation methods (combining all pop together of individual population analysis).

(6) Our results established the path for future experiments to understand the interaction of the human genetic variation and its extent to uncover the molecular mechanism of actions to gain a better insight into the connection between human genes and EBV biological and clinical behavior as seen in LCL to reduce the population-specific disease susceptibility and ultimately disease mortality.

Drawback of this work

(1) The major drawback of this work is relatively small samples size from 1000 Genome Project to do GWAS analysis. As much larger LCL samples size has a potential to map genetic variants with EBV copy number an ultimately clinical behavior of EBV. Nevertheless, we have detected here some of the high potential candidates at particular loci influencing EBV copy number that may hallmark the association of host genetic variation and EBV copy number.

(2) Although Genome-wide association studies (GWAS) is an advantageous methodology to discover genetic variants that influence disease outcomes, there are some limitations to replicate these results over different geographical populations/ethnic groups also the reported variants tend to explain small fractions of variability. Nevertheless by identifying the unique and common genetic variants between ethnicity and populations could help to resolve replicability issue and the role of genetic variants in the genetic architecture of complex disease (Marigorta and Navarro, 2013).

Part 2 antigenic variations in EBV isolates (EBV genome variation analysis)

Why is EBV antigenic diversity study important?

The entry of EBV through saliva, initial replication, and occasional reactivation from latent to lytic switch are some of the leading events of EBV infection. How EBV enters into B-lymphocytes and stays there for a lifelong despite a strong innate and adaptive immune response is a still mystery. It is likely that EBV has coevolved with the human-adapted immune system to influence and to temper the recognition by T-cells. It has been now accepted that EBV genome is polymorphic in nature in worldwide populations. In some cases, the distinct mutations in some epitopes from specific geographical isolates abrogate T-cell control. In particular, EBV changes the amino acid sequence in antigenic protein to escape from the immune system. Hence this work is essential in the aspect of identifying the potential amino acid changes that are distinctly present in the specific population and human diseases that might alter the immune recognition and confer strain to strain variation. Given the implication of wide range of EBV associated diseases, in diverse geographical populations, it is essential to understand the full spectrum of this antigenic variation pattern which, troubling immune response (healthy Vs. diseased) would help to exploit and design anti-EBV therapy.

In this PhD work, the epitope sequences from the Immune epitope database (IEDB) were analyzed to detect the natural variation in the experimentally validated epitope sequence regions. A comprehensive analysis of 30 antigenic proteins was carried out to evaluate the adaptive response of EBV to the human immune system. Overall out of 30 antigenic proteins, latent proteins have shown more variation in antigenic regions as compared to lytic proteins. Additionally, latent membrane proteins shared the largest pool of antigenic variation from different EBV isolates. Based on distinct mutation pattern, the groping of EBV strains demonstrated that more than 55 % of EBV strains had shown variation in latent proteins such as EBNA-1, EBNA-3, EBNA-3A as compared to approximately 45% antigenic variation of lytic proteins like BARF2 and BRRF2. For examples 58% of EBV isolates (63 out of 123) share mutation from A>V in EBNA-3, 56% of strains are having a common mutation M>I in EBNA-1 while only 47% (58 out of 123) showed mutation pattern of S>L in BRRF2 and only 34% (42 of 123) isolated are having V<A amino acid change in BARF2 protein. To notably mention, only proteins from latent cycle has shown the presence of strain-specific antigenic variation in AG876, BL37, Daudi, EBVaGC4, EBVaGC6, EBVaGC7, H002213, HL11, L591, sLCL-1.12, sLCL-1.17, sLCL-1.24, sLCL-2.15, sLCL-IS1.01, sLCL-IS1.04, sLCL-IS1.10, sLCL-IS1.11, VGO, and Wewak1 isolates (Table5). This data support the hypothesis, that variation in these proteins helps the virus to persist a lifelong infection.

So it would be challenging task now to identify the factors that influence this antigenic variation and the role of this polymorphism in host-virus interaction and immune invasion.

EBV antigenic mimicry in autoimmune disease

Under normal circumstances, the human immune system distinguishes between self and non self-proteins using MHC molecules to restrict reactivation of its own proteins. Occasionally the viral peptides having sufficient sequence and structural similarity gives the advantage to manipulate signaling pathways. This structural similarity can also trigger the autoreactive T-cell against self-protein. It is called as molecular mimicry in autoimmunity (Fujinami and Oldstone, 1985; Toussirot and Roudier, 2008; Wucherpfennig and Strominger, 1995).

EBV can occasionally produce proteins that are homologs of human proteins as explained in section 5.3.2 as a defense mechanism to avoid T-ell response. EBV has been linked to multiple autoimmune diseases such as multiple sclerosis, systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA) and recently in Parkinson's disease. EBV has demonstrated molecular mimicry by showing a cross-reaction as autoantigens, which result in cross-reactive antibody response and suppression of T-cell response. In such an experiment, an immunization to a rabbit with EBNA-1 peptide 'PPPGRRP' resulted in the development of lupus-like symptoms (James et al., 1997; Larsen et al., 2011; Lossius et al., 2012; Poole et al., 2006; Posnett, 2008; Woulfe et al., 2014).

Understanding the EBV-human interaction at global scale

EBV has co-evolved with the human host to grow as co-regulated at genomic scale for the lifelong latent infection. The implication of EBV genes and their role in regulatory pathways has been perceived by very few studies at individual gene level and virus-host interaction at genome-wide level analysis (Tempera and Lieberman, 2014). Our recent publication has demonstrated the presence of population-specific genomic variants affecting EBV copy numbers in 1000 Genome Project samples (Mandage et al., 2017). From these large-scale GWAS analysis data, it is clear that there is the substantial genomic variation that might affect response against EBV, and that the presence of population-specific variants might be contributing to disease development.

A thorough understanding EBV-human interaction would help to reveal how virus hijacks the human cellular system for the survival, secure replication process and at the end persistence for lifelong. By interacting with host proteins, EBV virus perturbs and disrupts human internal signaling pathways to influence dynamics of cellular functions (Mei and Zhang, 2016; Rowles et al., 2013). Therefore the ultimate goal to detect such human host specific proteins that might interact with EBV proteins to modulate and to influence cellular functions would help in understanding tissue tropism and viral pathogenesis and enhance our understanding of EBV-host interactions.

Integration of GWAS results with protein-protein interaction data

The host proteins those are used by the virus for replication, and other survival activities can vary across the host populations although the exact role of this variation is still unclear. Unlike GWAS studies, other research methods have also provided valuable details and often mechanistically oriented information on specific virus-host interactions by high-throughput screening and provided a large number of EBV-host interaction data (Arvey et al., 2012; Bailey et al., 2009; Calderwood et al., 2007; Choy et al., 2008). One of such method that runs in parallel with GWAS to detect host-EBV interacting proteins at genomic scale is “high-throughput yeast two-hybrid system” by using this method Calderwood et al. assessed the interaction of EBV proteins with human proteins and discovered 173 interactions in the form of “interactome map”. This high throughput screening examined host-virus interactions at individual regulatory loci contribute EBV gene expression and regulate EBV oncogene functions (Calderwood et al., 2007). In another study using a proteomics approach, the role of EBNA-1 in EBV-host interactions was dissected in nasopharyngeal and gastric carcinoma cells (Malik-Soni and Frappier, 2012). To uncover the regulatory interaction of host with EBV, a functional genomics study has suggested the co-expression of lytic genes with cancer-associated cellular pathways by integrating sequencing data of EBV positive LCLs (Arvey et al., 2012). A computational proteome-wide discovery of EBV interacting proteins, obtained 51,485 interaction reveals that EBV interferes with normal cellular pathways and blocks notch signaling and Hedgehog signaling pathways (Mei and Zhang, 2016). Nevertheless, the characterization of such interacting proteins, which, functions in the host-EBV interface has been challenging, mainly due to the technical challenges associated with discovery process.

Taking into consideration on the usefulness of this interaction data, we made an effort to query such interaction databases (CCSB Interactome Database) to discover the EBV genes/proteins interacting with our GWAS variants. Our GWAS analysis results in detection of only a single interaction of CAND1 with EBV gene BPLF1 (Gastaldello et al., 2012). Further querying the network of protein-protein interaction (PPI) using human-EBV interaction map surprisingly it resulted in detection of 2 more EBV genes interacting with CAND1.

Table: 6 EBV-host interacting proteins detected by GWAS and interactome analysis

EBV gene	Human protein	EBV gene function	EBV life cycle
BDLF4 **(PPI)**	CAND1	Lytic replication	Late Lytic
BPLF1 **(GWAS)**	CAND1	Lytic replication	Late lytic
BNLF2A **(PPI)**	CAND1	Immune evasion	Early lytic
BSLF2/BMLF1 **(GWAS)**	KHDRBS2	mRNA export factor	Early lytic

EBV-host interaction and immune response

The host genetic variation is nowadays well-recognized factor when it comes to identifying disease predisposition. Identification of such susceptibility in the host point of view would undoubtedly result in translation of genetic data to therapeutic development. One of the major disadvantages of this approach is the insufficient samples size to detect genes that interact with the pathogen (Burgner et al., 2006; Horby et al., 2013; Kambhampati et al., 2015; Weatherall et al., 1997). One of the major limitations of studies that attempt to identify the genes and mechanisms that underlie this susceptibility has been lack of power caused by small sample size.

The genetic variation analysis of both human and EBV to discover the consequence of the interaction is still an emerging research area of the future study. Most of the work carried out till date is concentrated to single variant mutational analysis, a minimal number of case and control studies restricted to uncover the role of these genes in EBV biology as well as in EBV associated diseases. In order to overcome these limitations, the integration of GWAS variants with RNA sequencing data, large scale of protein-protein interaction (PPI) and data generated by experimental and computational approaches would definitely provide a compressive network of EBV-host interaction specifying a key aspect to create the hypothesis and to validate the complex nature of the relationship between host and EBV to provide a comprehensive model of EBV persistence infection. The suggested perturbations of EBV-human interactions derived from our GWAS study or from the antigenic variation survey must be follow-up in the context of the susceptibility of individual populations to a specific EBV associated pathology.

8. CONCLUDING REMARKS

In brief, I would like to discuss some concluding remarks as a result of my PhD work

1. Leveraging 1000 Genome project data would be expedient to understand human-EBV interaction and it can serve as a cohort to generate and to test hypothesis on EBV biology related to LCL transformation.

2. EBV copy number is a stable phenotype. However, it differs within and between the populations and continents.

3. The human genetics influence EBV copy number in LCL across 1000 Genome population samples. It implicates the role of host genetic variation in EBV biology.

4. EBV copy number can be used as proxy phenotype to reflect the EBV biology in LCL as suggested by unveiled genes with a known role in EBV infection such as CAND1.

5. A large sample size from health and diseased individuals would be needed to highly improve GWAS power of detection of associations and to ensure proper testing of replication across populations.

6. There is significant variation in antigenic sequences from different EBV isolates, particularly in latency genes, which can influence the recognition by HLA molecules and illustrate a common adaptive mechanism used by viruses to scape immune surveillance.

9. REFERENCES

Abbott, Rachel, J.M.Q., Laura, L.L., Alison, M.S., Harry, M.P., Annette;, R., Alan, Ba., Rachel, J.M.Q., Laura, L.L., Alison, M.S., Harry, M.P., Annette;, R., n.d. CD8+ T cell responses to lytic EBV infection: late antigen specificities as subdominant components of the total response. Journal of immunology (Baltimore, Md : 1950)

Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., Gabriel, S.B., Gibbs, R.A., Green, E.D., Hurles, M.E., Knoppers, B.M., Korbel, J.O., Lander, E.S., Lee, C., Lehrach, H., Mardis, E.R., Marth, G.T., Mcvean, G.A., Nickerson, D.A., Schmidt, J.P., Sherry, S.T., Wang, J., Wilson, R.K., Dinh, H., Kovar, C., Lee, S., Lewis, L., Muzny, D., Reid, J., Wang, M., Fang, X.D., Guo, X.S., Jian, M., Jiang, H., Jin, X., Li, G.Q., Li, J.X., Li, Y.R., Li, Z., Liu, X., Lu, Y., Ma, X.D., Su, Z., Tai, S.S., Tang, M.F., Wang, B., Wang, G.B., Wu, H.L., Wu, R.H., Yin, Y., Zhang, W.W., Zhao, J., Zhao, M.R., Zheng, X.L., Zhou, Y., Gupta, N., Clarke, L., Leinonen, R., Smith, R.E., Zheng-Bradley, X., Grocock, R., Humphray, S., James, T., Kingsbury, Z., Sudbrak, R., Albrecht, M.W., Amstislavskiy, V.S., Borodina, T.A., Lienhard, M., Mertes, F., Sultan, M., Timmermann, B., Yaspo, M.L., Fulton, L., Fulton, R., Weinstock, G.M., Balasubramaniam, S., Burton, J., Danecek, P., Keane, T.M., Kolb-Kokocinski, A., McCarthy, S., Stalker, J., Quail, M., Davies, C.J., Gollub, J., Webster, T., Wong, B., Zhan, Y.P., Auton, A., Yu, F., Bainbridge, M., Challis, D., Evani, U.S., Lu, J., Nagaswamy, U., Sabo, A., Wang, Y., Yu, J., Coin, L.J.M., Fang, L., Li, Q.B., Li, Z.Y., Lin, H.X., Liu, B.H., Luo, R.B., Qin, N., Shao, H.J., Wang, B.Q., Xie, Y.L., Ye, C., Yu, C., Zhang, F., Zheng, H.C., Zhu, H.M., Garrison, E.P., Kural, D., Lee, W.P., Leong, W.F., Ward, A.N., Wu, J.T., Zhang, M.Y., Griffin, L., Hsieh, C.H., Mills, R.E., Shi, X.H., von Grotthuss, M., Zhang, C.S., Daly, M.J., DePristo, M.A., Banks, E., Bhatia, G., Carneiro, M.O., del Angel, G., Genovese, G., Handsaker, R.E., Hartl, C., McCarroll, S.A., Nemesh, J.C., Poplin, R.E., Schaffner, S.F., Shakir, K., Yoon, S.C., Lihm, J., Makarov, V., Jin, H.J., Kim, W., Kim, K.C., Rausch, T., Beal, K., Cunningham, F., Herrero, J., McLaren, W.M., Ritchie, G.R.S., Gottipati, S., Keinan, A., Rodriguez-Flores, J.L., Sabeti, P.C., Grossman, S.R., Tabrizi, S., Tariyal, R., Cooper, D.N., Ball, E. V, Stenson, P.D., Barnes, B., Bauer, M., Cheetham, R.K., Cox, T., Eberle, M., Kahn, S., Murray, L., Peden, J., Shaw, R., Ye, K., Batzer, M.A., Konkel, M.K., Walker, J.A., MacArthur, D.G., Lek, M., Herwig, R., Shriver, M.D., Bustamante, C.D., Byrnes, J.K., De la Vega, F.M., Gravel, S., Kenny, E.E., Kidd, J.M., Lacroute, P., Maples, B.K., Moreno-Estrada, A., Zakharia, F., Halperin, E., Baran, Y., Craig, D.W., Christoforides, A., Homer, N., Izatt, T., Kurdoglu, A.A., Sinari, S.A., Squire, K., Xiao, C.L., Sebat, J., Bafna, V., Burchard, E.G., Hernandez, R.D., Gignoux, C.R., Haussler, D., Katzman, S.J., Kent, W.J., Howie, B., Ruiz-Linares, A., Dermitzakis, E.T., 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65. doi:10.1038/nature11632

Antonelli, G., Roilides, E., 2014. Host genetics: Deciphering the variability in susceptibility to infections. Clinical Microbiology and Infection. doi:10.1111/1469-0691.12789

Apolloni, A., Sculley, T.B., 1994. Detection of A-type and B-type Epstein-Barr virus in throat washings and lymphocytes. Virology 202, 978–981. doi:10.1006/viro.1994.1422

Arvey, A., Tempera, I., Tsai, K., Chen, H.S., Tikhmyanova, N., Klichinsky, M., Leslie, C., Lieberman, P.M., 2012. An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. Cell Host and Microbe 12, 233–245. doi:10.1016/j.chom.2012.06.008

Ascherio, A., Munger, K.L., 2010. Epstein-barr virus infection and multiple sclerosis: a review. Journal of neuroimmune pharmacology : the official journal of the Society on NeuroImmune Pharmacology 5, 271–7. doi:10.1007/s11481-010-9201-3

Babcock, G.J., Decker, L.L., Freeman, R.B., Thorley-Lawson, D.A., 1999. Epstein-barr virus-infected resting memory B cells, not proliferating lymphoblasts, accumulate in the peripheral blood of immunosuppressed patients. The Journal of experimental medicine 190, 567–76. doi:10.1084/jem.190.4.567

Babcock, G.J., Hochberg, D., Thorley-Lawson, D.A., 2000. The Expression Pattern of Epstein-Barr Virus Latent Genes In Vivo Is Dependent upon the Differentiation Stage of the Infected B Cell. Immunity 13, 497–506. doi:10.1016/S1074-7613(00)00049-2

Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Séguin, C., 1984a. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature 310, 207–11. doi:10.1038/310207a0

Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Seguin, C., et al., 1984b. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature 310, 207–211. doi:10.1038/310207a0

Bailey, S.G., Verrall, E., Schelcher, C., Rhie, A., Doherty, A.J., Sinclair, A.J., 2009. Functional interaction between Epstein-Barr virus replication protein Zta and host DNA damage response protein 53BP1. J Virol 83, 11116–11122. doi:10.1128/JVI.00512-09

Baumforth, K.R., Young, L.S., Flavell, K.J., Constandinou, C., Murray, P.G., 1999. The Epstein-Barr virus and its association with human cancers. Molecular pathology : MP 52, 307–22. doi:10.1136/mp.52.6.307

Beck, S., Trowsdale, J., 2000. THE HUMAN MAJOR HISTOCOMPATIBILITY COMPLEX: Lessons from the DNA Sequence. Annual Review of Genomics and Human Genetics 1, 117–137. doi:10.1146/annurev.genom.1.1.117

Bell, Melissa, J.B., Rebekah;, M., John, J.M., Denis, J.B., Jacqueline, M.B., Scott, Rb., Melissa, J.B., Rebekah;, M., John, J.M., Denis, J.B., Jacqueline, M.B., n.d.

Widespread sequence variation in Epstein-Barr virus nuclear antigen 1 influences the antiviral T cell response. The Journal of infectious diseases.

Bellas, C., García-Cosío, M., Santón, A., Martín, P., Reguero, M.E., Cristóbal, E., 2008. Analysis of epstein-barr virus strains and variants in classical Hodgkin's lymphoma by laser microdissection. Histology and Histopathology 23, 209–217

Bentley, G., Higuchi, R., Hoglund, B., Goodridge, D., Sayer, D., Trachtenberg, E.A., Erlich, H.A., 2009. High-resolution, high-throughput HLA genotyping by next-generation sequencing. Tissue Antigens 74, 393–403. doi:10.1111/j.1399-0039.2009.01345

Berger, C., Day, P., Meier, G., Zingg, W., Bossart, W., Nadal, D., 2001. Dynamics of Epstein-Barr virus DNA levels in serum during EBV-associated disease. Journal of medical virology 64, 505–512.

Beutler, B., 2004. Innate immunity: An overview. Molecular Immunology. doi:10.1016/j.molimm.2003.10.005

Blackwell, J.M., Jamieson, S.E., Burgner, D., 2009. HLA and infectious diseases. Clinical Microbiology Reviews. doi:10.1128/CMR.00048-08

Bobek, V., Kolostova, K., Pinterova, D., Kacprzak, G., Adamiak, J., Kolodziej, J., Boubelik, M., Kubecova, M., Hoffman, R.M., 2010. A clinically relevant, syngeneic model of spontaneous, highly metastatic B16 mouse melanoma. Anticancer Research 30, 4799–4804. doi:10.1002/jmv

Bornkamm, G.W., 2009. Epstein-Barr virus and the pathogenesis of Burkitt's lymphoma: More questions than answers. International Journal of Cancer. doi:10.1002/ijc.24223

Borst, P., 2003. Mechanisms of Antigenic Variation. An Overview., in: Antigenic Variation. pp. 1–15. doi:10.1016/B978-012194851-1/50026-3

Brady, G., MacArthur, G.J., Farrell, P.J., 2007. Epstein-Barr virus and Burkitt lymphoma. Journal of clinical pathology 60, 1397–402. doi:10.1136/jcp.2007.047977

Bray, F., Haugen, M., Moger, T.A., Tretli, S., Aalen, O.O., Grotmol, T., 2008. Age-incidence curves of nasopharyngeal carcinoma worldwide: Bimodality in low-risk populations and aetiologic implications. Cancer Epidemiology Biomarkers and Prevention 17, 2356–2365. doi:10.1158/1055-9965.EPI-08-0461

Buhler, S., Sanchez-Mazas, A., 2011. HLA DNA sequence variation among human populations: Molecular signatures of demographic and selective events. PLoS ONE 6. doi:10.1371/journal.pone.0014643

Burgner, D., Jamieson, S.E., Blackwell, J.M., 2006. Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? Lancet Infectious Diseases. doi:10.1016/S1473-3099(06)70601-6

Burrows, J, M.B., S, R.P., L, M.S., T, B.M., D, J.K., RBurrows, J, M.B., S, R.P., L, M.S., T, B.M., D, J.K., n.d. Unusually high frequency of Epstein-Barr virus genetic variants in Papua New Guinea that can escape cytotoxic T-cell recognition: implications for virus evolution. Journal of virology

Bush, W.S., Moore, J.H., 2012. Chapter 11: Genome-Wide Association Studies. PLoS Computational Biology 8. doi:10.1371/journal.pcbi.1002822

Calderwood, M.A., Venkatesan, K., Xing, L., Chase, M.R., Vazquez, A., Holthaus, A.M., Ewence, A.E., Li, N., Hirozane-Kishikawa, T., Hill, D.E., Vidal, M., Kieff, E., Johannsen, E., 2007. Epstein-Barr virus and virus human protein interaction maps. Proceedings of the National Academy of Sciences of the United States of America 104, 7606–11. doi:10.1073/pnas.0702332104

Çalişkan, M., Cusanovich, D.A., Ober, C., Gilad, Y., 2011. The effects of EBV transformation on gene expression levels and methylation profiles. Human Molecular Genetics 20, 1643–1652. doi:10.1093/hmg/ddr041

Cantor, R.M., Lange, K., Sinsheimer, J.S., 2010. Prioritizing GWAS Results: A Review of Statistical Methods and Recommendations for Their Application. American Journal of Human Genetics. doi:10.1016/j.ajhg.2009.11.017

Cao, C., Moult, J., 2014. GWAS and drug targets. BMC genomics 15 Suppl 4, S5. doi:10.1186/1471-2164-15-S4-S5

Caron, M., Imam-Sghiouar, N., Poirier, F., Le Caër, J.P., Labas, V., Joubert-Caron, R., 2002. Proteomic map and database of lymphoblastoid proteins. Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences. doi:10.1016/S1570-0232(02)00040-5

Carrington, M., 1999. HLA and HIV-1: Heterozygote Advantage and B*35-Cw*04 Disadvantage. Science 283, 1748–1752. doi:10.1126/science.283.5408.1748

Carrington, M., O'Brien, S.J., 2003. The influence of HLA genotype on AIDS. Annual review of medicine 54, 535–551. doi:10.1146/annurev.med.54.101601.152346

Chabay, P.A., Preciado, M. V., 2013. EBV primary infection in childhood and its relation to B-cell lymphoma development: A mini-review from a developing region. International Journal of Cancer. doi:10.1002/ijc.27858

Chang, C.M., Yu, K.J., Mbulaiteye, S.M., Hildesheim, A., Bhatia, K., 2009. The extent of genetic diversity of Epstein-Barr virus and its geographic and disease patterns: A need for reappraisal. Virus Research. doi:10.1016/j.virusres.2009.07.005

Chang, E.T., Adami, H.O., 2006. The enigmatic epidemiology of nasopharyngeal carcinoma. Cancer Epidemiology Biomarkers and Prevention. doi:10.1158/1055-9965.EPI-06-0353

Chen, M.R., 2011. Epstein-barr virus, the immune system, and associated diseases. Frontiers in Microbiology 2, 1–5. doi:10.3389/fmicb.2011.00005

Chen, S., Ren, X., Liu, Y., Hu, Q., Hong, W., Xu, A., 2007. Human leukocyte antigen class I polymorphism in Miao, Bouyei, and Shui ethnic minorities of Guizhou, China. Human Immunology 68, 928–933. doi:10.1016/j.humimm.2007.09.006

Choo, S.Y., 2007. The HLA system: Genetics, immunology, clinical testing, and clinical implications. Yonsei Medical Journal 48, 11–23. doi:10.3349/ymj.2007.48.1.11

Choy, E., Yelensky, R., Bonakdar, S., Plenge, R.M., Saxena, R., De Jager, P.L., Shaw, S.Y., Wolfish, C.S., Slavik, J.M., Cotsapas, C., Rivas, M., Dermitzakis, E.T., Cahir-McFarland, E., Kieff, E., Hafler, D., Daly, M.J., Altshuler, D., 2008. Genetic analysis of human traits in vitro: Drug response and gene expression in lymphoblastoid cell lines. PLoS Genetics 4. doi:10.1371/journal.pgen.1000287

Christensen, T., 2006. The role of EBV in MS pathogenesis. International MS Journal

Clarke, G.M., Anderson, C. a, Pettersson, F.H., Cardon, L.R., Andrew, P., 2011. Basic statistical analysis in genetic case-control studies. Nature protocols 6, 121–133. doi:10.1038/nprot.2010.182.

Compston, A., McDonald, I., Noseworthy, J., Lassmann, H., Miller, D., Smith, K., Wekerle, H., Confavreux, C., 2006. McAlpine's Multiple Sclerosis, McAlpine's Multiple Sclerosis. doi:10.1016/B978-0-443-07271-0.X5001-0

Cordell, H.J., Clayton, D.G., 2005. Genetic association studies. The Lancet 366, 1121–1131. doi:10.1016/S0140-6736(05)67424-7

Correa, R.M., Fellner, M.D., Alonio, L.V., Durand, K., Teyssié, A.R., Picconi, M.A., 2004. Epstein-Barr virus (EBV) in healthy carriers: Distribution of genotypes and

30 bp deletion in latent membrane protein-1 (LMP-1) oncogene. Journal of Medical Virology 73, 583–588. doi:10.1002/jmv.20129

Correa, R.M., Fellner, M.D., Durand, K., Redini, L., Alonio, V., Yampolsky, C., Colobraro, A., Sevlever, G., Teyssie, A., Benetucci, J., Picconi, M.A., 2007. Epstein Barr virus genotypes and LMP-1 variants in HIV-infected patients. Journal of Medical Virology 79, 401–407. doi:10.1002/jmv.20782

Cotsapas, C., Voight, B.F., Rossin, E., Lage, K., Neale, B.M., Wallace, C., AbecasisGonç, G.R., Barrett, J.C., Behrens, T., Cho, J., De Jager, P.L., Elder, J.T., Graham, R.R., Gregersen, P., Klareskog, L., Siminovitch, K.A., van Heel, D.A., Wijmenga, C., Worthington, J., Todd, J.A., Hafler, D.A., Rich, S.S., Daly, M.J., 2011. Pervasive sharing of genetic effects in autoimmune disease. PLoS Genetics 7. doi:10.1371/journal.pgen.1002254

Couillin, I., 1994. Impaired cytotoxic T lymphocyte recognition due to genetic variations in the main immunogenic region of the human immunodeficiency virus 1 NEF protein. Journal of Experimental Medicine 180, 1129–1134. doi:10.1084/jem.180.3.1129

Cozen, W., Li, D., Best, T., Van Den Berg, D.J., Gourraud, P.A., Cortessis, V.K., Skol, A.D., Mack, T.M., Glaser, S.L., Weiss, L.M., Nathwani, B.N., Bhatia, S., Schumacher, F.R., Edlund, C.K., Hwang, A.E., Slager, S.L., Fredericksen, Z.S., Strong, L.C., Habermann, T.M., Link, B.K., Cerhan, J.R., Robison, L.L., Conti, D. V., Onel, K., 2012. A genome-wide meta-analysis of nodular sclerosing Hodgkin lymphoma identifies risk loci at 6p21.32. Blood 119, 469–475. doi:10.1182/blood-2011-03-343921

D, S.A., Derjaguin, B.., Churaev, N.., Radoev, B.P., Scheludko, A.D., Manev, E.D., Boomkamp, P., Lozano, A., Garcıa-Olivares, A., Dopazo, C., Maldarelli, C., Jain, R.K., Ivanov, I.B., Ruckenstein, E., Reynolds, O., Fm, S., Ibrahim, E. a., Akpan, E.T., Barlow, N.S., Yu and blackburn, Kargupta, K., Sharma, A., 1998. © 19 90

Nature Publishing Group. Journal of Colloid and Interface Science 245, 118–143. doi:10.1016/0021-9797(80)90501-9

Dambaugh, T., Hennessy, K., Chamnankit, L., Kieff, E., 1984. U2 region of Epstein-Barr virus DNA may encode Epstein-Barr nuclear antigen 2. Proceedings of the National Academy of Sciences of the United States of America 81, 7632–6. doi:10.1073/pnas.81.23.7632

Danzer, M., Niklas, N., Stabentheiner, S., Hofer, K., Pröll, J., Stückler, C., Raml, E., Polin, H., Gabriel, C., 2013. Rapid, scalable and highly automated HLA genotyping using next-generation sequencing: a transition from research to diagnostics. BMC Genomics 14, 221. doi:10.1186/1471-2164-14-221

de Campos-Lima, P.O., Gavioli, R., Zhang, Q.J., Wallace, L.E., Dolcetti, R., Rowe, M., Rickinson, A.B., Masucci, M.G., 1993. HLA-A11 epitope loss isolates of Epstein-Barr virus from a highly A11+ population. Science (New York, NY) 260, 98–100. doi:10.1126/science.7682013

de Resende, M.D.V., e Silva, F.F., Resende, M.F.R., Azevedo, C.F., 2014. Genome-Wide Association Studies (GWAS), Biotechnology and Plant Breeding. doi:10.1016/B978-0-12-418672-9.00004-0

den Haan, J.M.M., Arens, R., van Zelm, M.C., 2014. The activation of the adaptive immune system: Cross-talk between antigen-presenting cells, T cells and B cells. Immunology Letters. doi:10.1016/j.imlet.2014.10.011

Diepstra, A;, N., M;, V., E;, van I., G, W.N., I, M.S., M;, van der S., G;, van den B., A;, K., R, E. te M., G, J.P., SDiepstra, A;, N., M;, V., E;, van I., G, W.N., I, M.S., M;, van der S., G;, van den B., A;, K., R, E. te M., G, J.P., n.d. Association with HLA class I in Epstein-Barr-virus-positive and with HLA class III in Epstein-Barr-virus-negative Hodgkin’s lymphoma. Lancet (London, England).

Dolan, A., Addison, C., Gatherer, D., Davison, A.J., McGeoch, D.J., 2006. The genome of Epstein-Barr virus type 2 strain AG876. Virology 350, 164–170. doi:10.1016/j.virol.2006.01.015

Dudley, D.J., 1992. The immune system in health and disease. Bailliere’s clinical obstetrics and gynaecology 6, 393–416. doi:10.1016/S0950-3552(05)80003-3

Edinur, H.A., Zafarina, Z., Spínola, H., Nurhaslindawaty, A.R., Panneerchelvam, S., Norazmi, M.N., 2009. HLA polymorphism in six Malay subethnic groups in Malaysia. Human Immunology 70, 518–526. doi:10.1016/j.humimm.2009.04.003

Fafi-Kremer, S., Morand, P., Brion, J.-P., Pavese, P., Baccard, M., Germi, R., Genoulaz, O., Nicod, S., Jolivet, M., Ruigrok, R.W.H., Stahl, J.-P., Seigneurin, J.-M., 2005. Long-term shedding of infectious epstein-barr virus after infectious mononucleosis. The Journal of infectious diseases 191, 985–989. doi:10.1086/428097

Fagarasan, S., Honjo, T., 2000. T-independent immune response: New aspects of B cell biology. Science 290, 89–92. doi:10.1126/science.290.5489.89

Falk, L., Wolfe, L., Deinhardt, F., Paciga, J., Dombos, L., Klein, G., Henle, W., Henle, G., 1974. Epstein□barr virus: Transformation of non□human primate lymphocytes in vitro. International Journal of Cancer 13, 363–376. doi:10.1002/ijc.2910130312

Fan, H., Gulley, M.L., 2001. Epstein-Barr viral load measurement as a marker of EBV-related disease. Mol Diagn 6, 279–289. doi:10.1054/modi.2001.29161

Farrell, P.J., 2015. Epstein???barr virus strain variation, in: Current Topics in Microbiology and Immunology. pp. 45–69. doi:10.1007/978-3-319-22822-8_4

Farrell, P.J., 1995. Epstein-Barr virus immortalizing genes. Trends in Microbiology. doi:10.1016/S0966-842X(00)88891-5

Feederle, R., Neuhierl, B., Bannert, H., Geletneky, K., Shannon-Lowe, C., Delecluse, H.J., 2007. Epstein-Barr virus B95.8 produced in 293 cells shows marked tropism for differentiated primary epithelial cells and reveals interindividual variation in susceptibility to viral infection. International Journal of Cancer 121, 588–594. doi:10.1002/ijc.22727

Fellay, J., Shianna, K. V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., Cozzi-Lepri, A., De Luca, A., Easterbrook, P., Francioli, P., Mallal, S., Martinez-Picado, J., Miro, J.M., Obel, N., Smith, J.P., Wyniger, J., Descombes, P., Antonarakis, S.E., Letvin, N.L., McMichael, A.J., Haynes, B.F., Telenti, A., Goldstein, D.B., 2007. A Whole-Genome Association Study of Major Determinants for Host Control of HIV-1. Science 317, 944–947. doi:10.1126/science.1143767

Fierz, W., 2017. Multiple sclerosis: an example of pathogenic viral interaction? Virology Journal 14, 42. doi:10.1186/s12985-017-0719-3

Frank, S. a, 2002. Immunology and Evolution of Infectious Disease. Health San Francisco 300, 358. doi:10.1038/420741b

Fraser, K.B., Millar, J.H.D., Haire, M., Mccrea, S., 1979. INCREASED TENDENCY TO SPONTANEOUS IN-VITRO LYMPHOCYTE TRANSFORMATION IN CLINICALLY ACTIVE MULTIPLE SCLEROSIS. The Lancet 314, 715–717. doi:10.1016/S0140-6736(79)90643-3

Frazer, K.A., Murray, S.S., Schork, N.J., Topol, E.J., 2009. Human genetic variation and its contribution to complex traits. Nature Reviews Genetics 10, 241–251. doi:10.1038/nrg2554

Fujinami, R.S., Oldstone, M.B., 1985. Amino acid homology between the encephalitogenic site of myelin basic protein and virus: mechanism for autoimmunity. Science (New York, NY) 230, 1043–5.

doi:10.1126/science.2414848

Furukawa, MFurukaw, n.d. [Epstein-Barr virus (EBV) infection associated with nasopharyngeal carcinoma]. Gan to kagaku ryoho Cancer & chemotherapy.

Gabriel, C., Danzer, M., Hackl, C., Kopal, G., Hufnagl, P., Hofer, K., Polin, H., Stabentheiner, S., Pröll, J., 2009. Rapid high-throughput human leukocyte antigen typing by massively parallel pyrosequencing for high-resolution allele identification. Human Immunology 70, 960–964. doi:10.1016/j.humimm.2009.08.009

Gärtner, B., Preiksaitis, J.K., 2010. EBV viral load detection in clinical virology. Journal of Clinical Virology 48, 82–90. doi:10.1016/j.jcv.2010.03.016

Gastaldello, S., Callegari, S., Coppotelli, G., Hildebrand, S., Song, M., Masucci, M.G., 2012. Herpes virus deneddylases interrupt the cullin-RING ligase neddylation cycle by inhibiting the binding of CAND1. Journal of Molecular Cell Biology 4, 242–251. doi:10.1093/jmcb/mjs012

Gerber, P., Kalter, S.S., Schidlovsky, G., Peterson, W.D., Daniel, M.D., 1977. Biologic and antigenic characteristics of epstein???barr virus???related herpesviruses of chimpanzees and baboons. International Journal of Cancer 20, 448–459 doi:10.1002/ijc.2910200318

Gerber, P., Pritchett, R.F., Kieff, E.D., 1976. Antigens and DNA of a chimpanzee agent related to Epstein-Barr virus. J Virol 19, 1090–1099

Gourraud, P.A., Khankhanian, P., Cereb, N., Yang, S.Y., Feolo, M., Maiers, M., Rioux, J.D., Hauser, S., Oksenberg, J., 2014. HLA diversity in the 1000 genomes dataset. PLoS ONE 9. doi:10.1371/journal.pone.0097282

Gutierrez, M.I., Bhatia, K., Risueño, C. et al, 1992. Molecular epidemiology of

Burkitt's
lymphoma from South America: differences in breakpoint location and Epstein-Barr virus association from tumors in other world regions. Blood 79, 3261–3266.

Harding, C. V., 1991. Pathways of antigen processing. Current Opinion in Immunology 3, 3–9. doi:10.1016/0952-7915(91)90068-C

Hatton, O.L., Harris-Arnold, A., Schaffert, S., Krams, S.M., Martinez, O.M., 2014. The interplay between Epstein-Barr virus and B lymphocytes: Implications for infection, immunity, and disease. Immunologic Research 58, 268–276. doi:10.1007/s12026-014-8496-1

Henderson, S., Huen, D., Rowe, M., Dawson, C., Johnson, G., Rickinson, A., 1993. Epstein-Barr virus-coded BHRF1 protein, a viral homologue of Bcl-2, protects human B cells from programmed cell death. Proceedings of the National Academy of Sciences of the United States of America 90, 8479–83. doi:10.1073/pnas.90.18.8479

Henle, G., Henle, W., 1966. Immunofluorescence in cells derived from Burkitt's lymphoma. Journal of Bacteriology 91, 1248–1256.

Hill, A.V.S., 2012. Evolution, revolution and heresy in the genetics of infectious disease susceptibility. Philosophical Transactions of the Royal Society B: Biological Sciences 367, 840–849. doi:10.1098/rstb.2011.0275

Hirschhorn, J.N., Daly, M.J., 2005. GENOME-WIDE ASSOCIATION 6. doi:10.1038/nrg1521

Hislop, A.D., Kuo, M., Drake-Lee, A.B., Akbar, A.N., Bergler, W., Hammerschmitt, N., Khan, N., Palendira, U., Leese, A.M., Timms, J.M., Bell, A.I., Buckley, C.D., Rickinson, A.B., 2005. Tonsillar homing of Epstein-Barr virus-specific CD8+ T cells and the virus-host balance. Journal of Clinical Investigation 115, 2546–2555.

doi:10.1172/JCI24810

Hislop, A.D., Taylor, G.S., Sauce, D., Rickinson, A.B., 2007. Cellular responses to viral infection in humans: lessons from Epstein-Barr virus. Annual review of immunology 25, 587–617. doi:10.1146/annurev.immunol.25.022106.141553

Hislop, Andrew, D.R., Maaike, E. van L., Daphne;, P., Victoria, A.H., Daniëlle;, K.-L., Danijela;, C., Nathan, P.N., Jacques, J.R., Alan, B.W., Emmanuel, J.H.Jh., Andrew, D.R., Maaike, E. van L., Daphne;, P., Victoria, A.H., Daniëlle;, K.-L., Danijela;, C., Nathan, P.N., Jacques, J.R., Alan, B.W., n.d. A CD8+ T cell immune evasion protein specific to Epstein-Barr virus and its close relatives in Old World primates. The Journal of experimental medicine

Hjalgrim, H., Friborg, J., Melbye, M., 2007. The epidemiology of EBV and its association with malignant disease, in: Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. pp. 929–959. doi:10.1017/CBO9780511545313.054

Hochberg, D., Middeldorp, J.M., Catalina, M., Sullivan, J.L., Luzuriaga, K., Thorley-Lawson, D.A., 2004. Demonstration of the Burkitt's lymphoma Epstein-Barr virus phenotype in dividing latently infected memory cells in vivo. Proceedings of the National Academy of Sciences of the United States of America 101, 239–44. doi:10.1073/pnas.2237267100

Hohaus, S., Santangelo, R., Giachelia, M., Vannata, B., Massini, G., Cuccaro, A., Martini, M., Cesarini, V., Cenci, T., D'Alo, F., Voso, M.T., Fadda, G., Leone, G., Larocca, L.M., 2011. The viral load of Epstein-Barr virus (EBV) DNA in peripheral blood predicts for biological and clinical characteristics in Hodgkin lymphoma. Clinical Cancer Research 17, 2885–2892. doi:10.1158/1078-0432.CCR-10-3327

Horby, P., Nguyen, N.Y., Dunstan, S.J., Kenneth Baillie, J., 2013. An updated systematic review of the role of host genetics in susceptibility to influenza.

Influenza and other Respiratory Viruses. doi:10.1111/irv.12079

Horst, Daniëlle;, B., Scott, R.G., Derek;, van W., Bonnie;, B., Melissa, J.B., Ingrid, G.J.R., Maaike, E.W., Emmanuel, J.H.Jh., Daniëlle;, B., Scott, R.G., Derek;, van W., Bonnie;, B., Melissa, J.B., Ingrid, G.J.R., Maaike, E.W., n.d. Epstein-Barr virus isolates retain their capacity to evade T cell immunity through BNLF2a despite extensive sequence variation. Journal of virology.

Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K., Lush, M.J., Povey, S., Talbot, C.C., Wright, M.W., Wain, H.M., Trowsdale, J., Ziegler, A., Beck, S., 2004. Gene map of the extended human MHC. Nature Reviews Genetics 5, 889–899. doi:10.1038/nrg1489

Hosomichi, K., Shiina, T., Tajima, A., Inoue, I., 2015. The impact of next-generation sequencing technologies on HLA research. Journal of Human Genetics 60, 665–673. doi:10.1038/jhg.2015.102

Houldcroft, C.J., Kellam, P., 2014. Host genetics of Epstein-Barr virus infection, latency and disease. Reviews in medical virology. doi:10.1002/rmv.1816

Houldcroft, C.J., Petrova, V., Liu, J.Z., Frampton, D., Anderson, C. a, Gall, A., Kellam, P., 2014. Host genetic variants and gene expression patterns associated with Epstein-Barr virus copy number in lymphoblastoid cell lines. PloS one 9, e108384. doi:10.1371/journal.pone.0108384

Hraber, P., Kuiken, C., Yusim, K., 2007. Evidence for human leukocyte antigen heterozygote advantage against hepatitis C virus infection. Hepatology (Baltimore, Md) 46, 1713–21. doi:10.1002/hep.21889

Hsu, J.L., Glaser, S.L., 2000. Epstein-Barr virus-associated malignancies: Epidemiologic patterns and etiologic implications. Critical Reviews in

Oncology/Hematology. doi:10.1016/S1040-8428(00)00046-9

Huang, Xin;, K., Kushi;, N., Ilja;, K., Wierd;, V., Lydia;, B., Ilby;, K., Niels;, V., Rianne;, van I., Gustaaf;, O., Bianca;, H., Richard, S.P., Sibrand;, D., Arjan;, H., Bouke;, van den B., AnkeHuang, Xin;, K., Kushi;, N., Ilja;, K., Wierd;, V., Lydia;, B., Ilby;, K., Niels;, V., Rianne;, van I., Gustaaf;, O., Bianca;, H., Richard, S.P., Sibrand;, D., Arjan;, H., Bouke;, van den B., 2012. HLA associations in classical Hodgkin lymphoma: EBV status matters. PloS one.

Hussain, T., Mulherkar, R., 2012. Lymphoblastoid Cell lines: a Continuous in Vitro Source of Cells to Study Carcinogen Sensitivity and DNA Repair. International journal of molecular and cellular medicine 1, 75–87.

Hutt-fletcher, L.M., 2007. Epstein-Barr Virus Entry. Journal of virology 81, 7825–7832. doi:10.1128/JVI.00445-07

Hutt-Fletcher, L.M., 2014. Epstein–Barr virus replicating in epithelial cells. Proceedings of the National Academy of Sciences 111, 16242–16243. doi:10.1073/pnas.1418974111

Imai, S., Nishikawa, J., Takada, K., 1998. Cell-to-Cell Contact as an Efficient Mode of Epstein-Barr Virus Infection of Diverse Human Epithelial Cells. Journal of Virology 72, 4371–4378.

Iwakiri, D., Nakamura, H., Ono, Y., Fujiwara, S., 1997. Antigenic variation of the Epstein-Barr virus nuclear antigen EBNA1 as revealed by monoclonal antibodies. Virus research 50, 139–149.

James, J.A., Scofield, R.H., Harley, J.B., 1997. Lupus humoral autoimmunity after short peptide immunization, in: Annals of the New York Academy of Sciences. pp. 124–127. doi:10.1111/j.1749-6632.1997.tb52054.x

Janeway, C.J., Travers, P., Walport, M., 2001. Immunobiology: The Immune System in

Health and Disease. 5th edition. Garland Science.

Jarrett, a F., Armstrong, a a, Alexander, E., 1996. Epidemiology of EBV and Hodgkin’s lymphoma. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 7 Suppl 4, 5–10.

Jarrett, R.F., 2003. Risk factors for Hodgkin’s lymphoma by EBV status and significance of detection of EBV genomes in serum of patients with EBV-associated Hodgkin’s lymphoma. Leukemia & lymphoma 44 Suppl 3, S27–S32. doi:10.1080/10428190310001623801

Jensen, P.E., 2007. Recent advances in antigen processing and presentation. Nature Immunology 8, 1041–1048. doi:10.1038/ni1516

Jinam, T.A., Saitou, N., Edo, J., Mahmood, A., Phipps, M.E., 2010. Molecular analysis of HLA Class i and Class II genes in four indigenous Malaysian populations. Tissue Antigens 75, 151–158. doi:10.1111/j.1399-0039.2009.01417.x

Kaiko, G.E., Horvat, J.C., Beagley, K.W., Hansbro, P.M., 2008. Immunological decision-making: How does the immune system decide to mount a helper T-cell response? Immunology. doi:10.1111/j.1365-2567.2007.02719.x

Kambhampati, A., Payne, D.C., Costantini, V., Lopman, B.A., 2015. Host Genetic Susceptibility to Enteric Viruses: A Systematic Review and Metaanalysis. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 62, civ873-. doi:10.1093/cid/civ873

Kanakry, J.A., Li, H., Gellert, L.L., Lemas, M.V., Hsieh, W.S., Hong, F., Tan, K.L., Gascoyne, R.D., Gordon, L.I., Fisher, R.I., Bartlett, N.L., Stiff, P., Cheson, B.D., Advani, R., Miller, T.P., Kahl, B.S., Horning, S.J., Ambinder, R.F., 2013. Plasma Epstein-Barr virus DNA predicts outcome in advanced Hodgkin lymphoma: Correlative analysis from a large North American cooperative group trial. Blood

121, 3547–3553. doi:10.1182/blood-2012-09-454694

Kaur, G., Mehra, N., 2009. Genetic determinants of HIV-1 infection and progression to AIDS: Immune response genes. Tissue Antigens. doi:10.1111/j.1399-0039.2009.01337.x

Khan, G., Miyashita, E.M., Yang, B., Babcock, G.J., Thorley-Lawson, D.A., Ambinder, R.F., Lambe, B.C., Mann, R.B., Hayward, S.D., Zehnbauer, B.A., Burns, W.S., Charache, P., Azuma, M., Cayabyab, M., Buck, D., Phillips, J.H., Lanier, L.L., Chen, F., Zou, J.Z., diRenzo, L., Winberg, G., Hu, L.F., Klein, E., Klein, G., Ernberg, I., Coen, D.M., Decker, L.L., Klaman, L.D., Thorley-Lawson, D.A., Gray, D., Skarvall, H., Hurley, E.A., Thorley-Lawson, D.A., Lam, K.M., Syed, N., Whittle, H., Crawford, D.H., Levitskaya, J., Coram, M., Levitsky, V., Imreh, S., Steigerwald, M.P., Klein, G., Kurilla, M.G., Masucci, M.G., Masucci, M.G., Ernberg, I., Miller, C.L., Burkhardt, A.L., Lee, J.H., Stealey, B., Longnecker, R., Bolen, J.B., Kieff, E., Miyashita, E.M., Yang, B., Lam, K.M., Crawford, D.H., Thorley-Lawson, D.A., Qu, L., Rowe, D.T., Rickinson, A.B., Yao, Q.Y., Wallace, L.E., Rowe, M., Rowe, D.T., Gregory, C.D., Young, L.S., Farrell, P.J., Rupani, H., Rickinson, A.B., Saito, I., Servenius, B., Compton, T., Fox, R.I., Sugden, B., Phelps, M., Domoradzki, J., Summers, W.C., Klein, G., Telenti, A., Marshall, W.F., Smith, T.F., Thorley-Lawson, D.A., Miyashita, E.M., Khan, G., Tierney, R.J., Steven, N., Young, L.S., Rickinson, A.B., Wagner, H.J., Bein, G., Bitsch, A., Kirchner, H., Winer, B.J., Wyatt, R.T., Rudders, R.A., Zelenetz, A., Delellis, R.A., Krontiris, T.G., Yao, Q.Y., Rickinson, A.B., Gaston, J.S., Epstein, M.A., Yao, Q.Y., Ogan, P., Rowe, M., Wood, M., Rickinson, A.B., Yao, Q.Y., Czarnecka, H., Rickinson, A.B., Yates, J.L., Warren, N., Sugden, B., 1996. Is EBV persistence in vivo a model for B cell homeostasis? Immunity 5, 173–9. doi:10.1016/S1074-7613(00)80493-8

Khan, M.A., Mathieu, A., Sorrentino, R., Akkoc, N., 2007. The pathogenetic role of HLA-B27 and its subtypes. Autoimmunity Reviews. doi:10.1016/j.autrev.2006.11.003

Khanim, F., Yao, Q.Y., Niedobitek, G., Sihota, S., Rickinson, A.B., Young, L.S., 1996. Analysis of Epstein-Barr virus gene polymorphisms in normal donors and in virus-associated tumors from different geographic locations. Blood 88, 3491–501.

Khanna, R;, S., R, W.P., L;, M., D, J.B., S, R.N., J;, B., J, Mk., R;, S., R, W.P., L;, M., D, J.B., S, R.N., J;, B., n.d. Evolutionary dynamics of genetic variation in Epstein-Barr virus isolates of diverse geographical origins: evidence for immune pressure-independent genetic drift. Journal of virology.

Kim, S.M., Kang, S.H., Lee, W.K., 2006. Identification of two types of naturally-occurring intertypic recombinants of Epstein-Barr virus. Molecules and Cells 21, 302–307. doi:10.1109/TCOMM.2005.863731

Klein, G., Klein, E., Kashuba, E., 2010. Interaction of Epstein-Barr virus (EBV) with human B-lymphocytes. Biochemical and Biophysical Research Communications 396, 67–73. doi:10.1016/j.bbrc.2010.02.146

Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.-Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T., Bracken, M.B., Ferris, F.L., Ott, J., Barnstable, C., Hoh, J., 2005. Complement factor H polymorphism in age-related macular degeneration. Science (New York, NY) 308, 3 doi:10.1126/science.1109557

Klemenc, P., Marin, J., Šoba, E., Gale, N., Koren, S., Strojan, P., 2006. Distribution of Epstein-Barr virus genotypes in throat washings, sera, peripheral blood lymphocytes and in EBV positive tumor biopsies from Slovenian patients with nasopharyngeal carcinoma. Journal of Medical Virology 78, 1083–1090. doi:10.1002/jmv.20666

Ko, D.C., Urban, T.J., 2013. Understanding human variation in infectious disease susceptibility through clinical and cellular GWAS. PLoS pathogens 9, e1003424.

doi:10.1371/journal.ppat.1003424

Küppers, R., 2003. B cells under influence: transformation of B cells by Epstein-Barr virus. Nature reviews Immunology 3, 801–812. doi:10.1038/nri1201

Kutok, J.L., Wang, F., 2006. SPECTRUM OF EPSTEIN-BARR VIRUS – ASSOCIATED DISEASES. Annual Review of Pathology: Mechanisms of Disease 1, 375–404. doi:10.1146/annurev.pathol.1.110304.100209

Kwok, H., Tong, A.H.Y., Lin, C.H., Lok, S., Farrell, P.J., Kwong, D.L.W., Chiang, A.K.S., 2012. Genomic sequencing and comparative analysis of Epstein-Barr virus genome isolated from primary nasopharyngeal carcinoma biopsy. PLoS ONE 7. doi:10.1371/journal.pone.0036939

Kwok, H., Wu, C.W., Palser, A.L., Kellam, P., Sham, P.C., Kwong, D.L.W., Chiang, A.K.S., 2014. Genomic Diversity of Epstein-Barr Virus Genomes Isolated from Primary Nasopharyngeal Carcinoma Biopsy Samples. Journal of Virology 88, 10662–10672. doi:10.1128/JVI.01665-14

Laichalk, L.L., Thorley-Lawson, D.A., 2005. Terminal differentiation into plasma cells initiates the replicative cycle of Epstein-Barr virus in vivo. J Virol 79, 1296–1307. doi:10.1128/JVI.79.2.1296-1307.2005

Landais, E., Saulquin, X., Houssaint, E., 2005. The human T cell immune response to Epstein-Barr virus. The International journal of developmental biology 49, 285–292. doi:10.1387/ijdb.041947el

Lank, S.M., Wiseman, R.W., Dudley, D.M., O'Connor, D.H., 2010. A novel single cDNA amplicon pyrosequencing method for high-throughput, cost-effective sequence-based HLA class I genotyping. Human Immunology 71, 1011–1017. doi:10.1016/j.humimm.2010.07.012

Larsen, M., Sauce, D., Deback, C., Arnaud, L., Mathian, A., Miyara, M., Boutolleau, D., Parizot, C., Dorgham, K., Papagno, L., Appay, V., Amoura, Z., Gorochov, G., 2011. Exhausted cytotoxic control of epstein-barr virus in human lupus. PLoS Pathogens 7. doi:10.1371/journal.ppat.1002328

Lechowicz, M.J., Lin, L., Ambinder, R.F., 2002. Epstein-Barr virus DNA in body fluids. Current opinion in oncology 14, 533–7. doi:10.1097/01.CCO.0000028165.76026.0F

Lei, H., Li, T., Hung, G.-C., Li, B., Tsai, S., Lo, S.-C., 2013. Identification and characterization of EBV genomes in spontaneously immortalized human peripheral blood B lymphocytes by NGS technology. BMC genomics 14, 804. doi:10.1186/1471-2164-14-804

Levitskaya, J., Coram, M., Levitsky, V., Imreh, S., Steigerwald-Mullen, P.M., Klein, G., Kurilla, M.G., Masucci, M.G., 1995. Inhibition of antigen processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. Nature. doi:10.1038/375685a0

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Lin, Z., Wang, X., Strong, M.J., Concha, M., Baddoo, M., Xu, G., Baribault, C., Fewell, C., Hulme, W., Hedges, D., Taylor, C.M., Flemington, E.K., 2013. Whole-Genome Sequencing of the Akata and Mutu Epstein-Barr Virus Strains. Journal of Virology 87, 1172–1182. doi:10.1128/JVI.02517-12

Lin, Hsiang-Ju;, C., Jaw-Ming;, H., Man-Shan;, S., Yiyang;, L., Jung-ChungLin, Hsiang-Ju;, C., Jaw-Ming;, H., Man-Shan;, S., Yiyang;, L., n.d. Functional assays of HLA A2-restricted epitope variant of latent membrane protein 1 (LMP-1) of Epstein-Barr virus in nasopharyngeal carcinoma of Southern China and Taiwan.

Journal of biomedical science.

Lind, C., Ferriola, D., Mackiewicz, K., Heron, S., Rogers, M., Slavich, L., Walker, R., Hsiao, T., McLaughlin, L., D'Arcy, M., Gai, X., Goodridge, D., Sayer, D., Monos, D., 2010. Next-generation sequencing: The solution for high-resolution, unambiguous human leukocyte antigen typing. Human Immunology 71, 1033–1042. doi:10.1016/j.humimm.2010.06.016

Liu, Y., de Waal Malefyt, R., Briere, F., Parham, C., Bridon, J.M., Banchereau, J., Moore, K.W., Xu, J., 1997. The EBV IL-10 homologue is a selective agonist with impaired binding to the IL-10 receptor. Journal of immunology (Baltimore, Md : 1950) 158, 604–13.

Liu, Y., Yang, W., Pan, Y., Ji, J., Lu, Z., Ke, Y., 2016. Genome-wide analysis of Epstein-Barr virus (EBV) isolated from EBV-associated gastric carcinoma (EBVaGC). Oncotarget 7, 4903–14. doi:10.18632/oncotarget.6751

Lo, Y.M.D., Chan, L.Y., Lo, K.W., Leung, S.F., Zhang, J., Chan, A.T., Lee, J.C., Hjelm, N.M., Johnson, P.J., Huang, D.P., 1999. Quantitative analysis of cell-free Epstein-Barr virus DNA in plasma of patients with nasopharyngeal carcinoma. Cancer research 59, 1188–1191.

Loeb, M., 2013. Host genomics in infectious diseases. Infection and Chemotherapy. doi:10.3947/ic.2013.45.3.253

Lossius, A., Johansen, J.N., Torkildsen, Ø., Vartdal, F., Holmoy, T., 2012. Epstein-barr virus in systemic lupus erythematosus, rheumatoid arthritis and multiple sclerosis-association and causation. Viruses. doi:10.3390/v4123701

Luckey, D., Bastakoty, D., Mangalam, A.K., 2011. Role of HLA class II genes in susceptibility and resistance to multiple sclerosis: Studies using HLA transgenic mice. Journal of Autoimmunity 37, 122–128. doi:10.1016/j.jaut.2011.05.001

Lung, M.L., Chang, G.C., 1992. Detection of distinct Epstein???Barr virus genotypes in NPC biopsies from Southern Chinese and Caucasians. International Journal of Cancer 52, 34–37. doi:10.1002/ijc.2910520108

Machulla, H.K.G., Mller, L.P., Schaaf, A., Kujat, G., Schnermarck, U., Langner, J., 2001. Association of chronic lymphocytic leukemia with specific alleles of the HLA-DR4:DR53:DQ8 haplotype in German patients. International Journal of Cancer 92, 203–207. doi:10.1002/1097-0215(200102)9999:9999<::AID-IJC1167>3.0.CO;2-A

Maiers, M., Gragert, L., Klitz, W., 2007. High-resolution HLA alleles and haplotypes in the United States population. Human Immunology 68, 779–788. doi:10.1016/j.humimm.2007.04.005

Malik-Soni, N., Frappier, L., 2012. Proteomic profiling of EBNA1-host protein interactions in latent and lytic Epstein-Barr virus infections. Journal of virology 86, 6999–7002. doi:10.1128/JVI.00194-12

Mandage, R., Telford, M., Rodríguez, J.A., Farré, X., Layouni, H., Marigorta, U.M., Cundiff, C., Heredia-Genestar, J.M., Navarro, A., Santpere, G., 2017. Genetic factors affecting EBV copy number in lymphoblastoid cell lines derived from the 1000 Genome Project samples. PLoS ONE 12. doi:10.1371/journal.pone.0179446

Marigorta, U.M., Navarro, A., 2013. High trans-ethnic replicability of GWAS results implies common causal variants. PLoS genetics 9, e1003566. doi:10.1371/journal.pgen.1003566

Marsh, S.G.E., Parham, P., Barber, L.D., 2000. The HLA FactsBook, The HLA FactsBook. doi:10.1016/B978-012545025-6/50137-4

Mbulaiteye, S.M., Biggar, R.J., Bhatia, K., Linet, M.S., Devesa, S.S., 2009. Sporadic childhood Burkitt lymphoma incidence in the United States during 1992-2005. Pediatric Blood and Cancer 53, 366–370. doi:10.1002/pbc.22047

McCormack, M., Alfirevic, A., Bourgeois, S., Farrell, J.J., Kasperavičiūtė, D., Carrington, M., Sills, G.J., Marson, T., Jia, X., de Bakker, P.I.W., Chinthapalli, K., Molokhia, M., Johnson, M.R., O'Connor, G.D., Chaila, E., Alhusaini, S., Shianna, K. V, Radtke, R.A., Heinzen, E.L., Walley, N., Pandolfo, M., Pichler, W., Park, B.K., Depondt, C., Sisodiya, S.M., Goldstein, D.B., Deloukas, P., Delanty, N., Cavalleri, G.L., Pirmohamed, M., 2011. HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. The New England journal of medicine 364, 1134–43. doi:10.1056/NEJMoa1013297

McGeoch, D.J., 2001. Molecular evolution of the gamma-Herpesvirinae. Philosophical transactions of the Royal Society of London Series B, Biological sciences 356, 421–35. doi:10.1098/rstb.2000.0775

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20, 1297–1303. doi:10.1101/gr.107524.110

Mei, S., Zhang, K., 2016. Computational discovery of Epstein-Barr virus targeted human genes and signalling pathways. Scientific Reports 6, 30612. doi:10.1038/srep30612

Merlo, A., Turrini, R., Dolcetti, R., Martorelli, D., Muraro, E., Comoli, P., Rosato, A., 2010. The interplay between Epstein-Barr virus and the immune system: A rationale for adoptive cell therapy of EBV-related disorders. Haematologica 95, 1769–1777. doi:10.3324/haematol.2010.023689

Midgley, R.S., Bell, A.I., Yao, Q.Y., Croom-Carter, D., Hislov, A.D., Whitney, B.M., Chan, A.T.C., Johnson, P.J., Rickinson, A.B., 2003. HLA-A11-restricted epitope polymorphism among Epstein-Barr virus strains in the highly HLA-A11-positive Chinese population: Incidence and immunogenicity of variant epitope sequences. Journal of Virology 77, 11507–11516. doi:10.1128/Jvi.77.21.11507-11516.2003

Miller, G., Shope, T., Lisco, H., Stitt, D., Lipman, M., 1972. Epstein-Barr virus: transformation, cytopathic changes, and viral antigens in squirrel monkey and marmoset leukocytes. Proceedings of the National Academy of Sciences of the United States of America 69, 383–7.

Miyashita, E.M., Yang, B., Lam, K.M.C., Crawford, D.H., Thorley-Lawson, D.A., 1995. A novel form of Epstein-Barr virus latency in normal B cells in vivo. Cell 80, 593–601. doi:10.1016/0092-8674(95)90513-8

Mogensen, T.H., 2009. Pathogen recognition and inflammatory signaling in innate immune defenses. Clinical Microbiology Reviews. doi:10.1128/CMR.00046-08

Moghaddam, A., Koch, J., Annis, B., Wang, F., 1998. Infection of human B lymphocytes with lymphocryptoviruses related to Epstein-Barr virus. Journal of virology 72, 3205–12.

Moghaddam, A., Rosenzweig, M., Lee-Parritz, D., Annis, B., Johnson, R.P., Wang, F., 1997. An animal model for acute and persistent Epstein-Barr virus infection. Science 276, 2030–2033.

Moonsamy, P. V., Williams, T., Bonella, P., Holcomb, C.L., Höglund, B.N., Hillman, G., Goodridge, D., Turenchalk, G.S., Blake, L.A., Daigle, D.A., Simen, B.B., Hamilton, A., May, A.P., Erlich, H.A., 2013. High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array??? system for simplified amplicon library preparation. Tissue Antigens 81, 141–149. doi:10.1111/tan.12071

Morris, M.A., Dawson, C.W., Laverick, L., Davis, A.M., Dudman, J.P.R., Raveenthiraraj, S., Ahmad, Z., Yap, L.-F., Young, L.S., Young, L.S., Rickinson, A.B., Chan, A.S., Tao, Q., Chan, A.T., Dawson, C.W., Port, R.J., Young, L.S., Pathmanathan, R., Prasad, U., Sadler, R., Flynn, K., Raab-Traub, N., Morris, M.A., Dawson, C.W., Young, L.S., Dawson, C.W., Rickinson, A.B., Young, L.S., Young, L.S., Khabir, A., Morris, M.A., Wang, C., Horikawa, T., Kondo, S., Wilson, J.B., Weinberg, W., Johnson, R., Yuspa, S., Levine, A.J., Hu, L.F., Kondo, S., Kung, C.P., Raab-Traub, N., Wakisaka, N., Massague, J., Wotton, D., Derynck, R., Zhang, L., Yue, J., Mulder, K.M., Munz, B., Massague, J., Thomas, T.Z., Mylonas, I., Zhang, Z., Ying, S.Y., Munz, B., Hubner, G., Tretter, Y., Alzheimer, C., Werner, S., Watt, F.M., Prokova, V., Mosialos, G., Kardassis, D., Mori, N., Morishita, M., Tsukazaki, T., Yamamoto, N., Hu, C., Hocevar, B.A., Brown, T.L., Howe, P.H., Branton, Werner, S., Alzheimer, C., Laping, N.J., Inman, G.J., Yakymovych, I., Dijke, P. Ten, Heldin, C.H., Souchelnytskyi, S., Dennler, S., Levy, L., Hill, C.S., Derynck, R., Zhang, Y.E., Tadokoro, S., Brunton, Dawson, C.W., Laverick, L., Morris, M.A., Tramoutanis, G., Young, L.S., Schaffner, F., Ray, A.M., Dontenwill, M., Larjava, H., Haapasalmi, K., Salo, T., Wiebe, C., Uitto, V.J., Zeng, Z.Y., Ma, L.J., Rheinwald, J.G., Beckett, M.A., Dawson, C.W., Eliopoulos, A.G., Blake, S.M., Barker, R., Young, L.S., Dawson, C.W., Tramountanis, G., Eliopoulos, A.G., Young, L.S., 2016. The Epstein-Barr virus encoded LMP1 oncoprotein modulates cell adhesion via regulation of activin A/TGFβ and β1 integrin signalling. Scientific Reports 6, 19533. doi:10.1038/srep19533

Mueller, S.N., Rouse, B.T., 2008. Immune responses to viruses, in: Clinical Immunology. pp. 421–431. doi:10.1016/B978-0-323-04404-2.10027-2

Muhe, J., Wang, F., 2015. Host Range Restriction of Epstein-Barr Virus and Related Lymphocryptoviruses. Journal of virology 89, 9133–9136. doi:10.1128/JVI.01235-15

Nam, H.-Y., Shim, S.-M., Han, B.-G., Jeon, J.-P., 2011. Human lymphoblastoid cell lines: a goldmine for the biobankomics era. Pharmacogenomics 12, 907–17. doi:10.2217/pgs.11.24

Neitzel, H., 1986. A routine method for the establishment of permanent growing lymphoblastoid cell lines. Human Genetics 73, 320–326. doi:10.1007/BF00279094

Niens, Marijke;, J., Ruth, F.H., Bouke;, N., Ilja, M.D., Arjan;, P., Mathieu;, K., Niels;, D., Craig, P.G., Alice;, V., Lydia;, P., Sibrand;, te M., Gerard, J. van den B., AnkeNiens, Marijke;, J., Ruth, F.H., Bouke;, N., Ilja, M.D., Arjan;, P., Mathieu;, K., Niels;, D., Craig, P.G., Alice;, V., Lydia;, P., Sibrand;, te M., Gerard, J. van den B., n.d. HLA-A*02 is associated with a reduced risk and HLA-A*01 with an increased risk of developing EBV+ Hodgkin lymphoma. Blood.

Nuzhdin, S. V., Friesen, M.L., McIntyre, L.M., 2012. Genotype-phenotype mapping in a post-GWAS world. Trends in Genetics. doi:10.1016/j.tig.2012.06.003

Odumade, O.A., Hogquist, K.A., Balfour, H.H., 2011. Progress and problems in understanding and managing primary epstein-barr virus infections. Clinical Microbiology Reviews 24, 193–209. doi:10.1128/CMR.00044-10

Ok, C.Y., Li, L., Young, K.H., 2015. EBV-driven B-cell lymphoproliferative disorders: from biology, classification and differential diagnosis to clinical management. Experimental & Molecular Medicine 47, e132. doi:10.1038/emm.2014.82

Olerup, O., Hillert, J., 1991. HLA class II□associated genetic susceptibility in multiple sclerosis: A critical evaluation. Tissue Antigens. doi:10.1111/j.1399-0039.1991.tb02029.x

Opelz, G., Wujciak, T., Döhler, B., Scherer, S., Mytilineos, J., 1999. HLA compatibility and organ transplant survival. Collaborative Transplant Study. Reviews in immunogenetics 1, 334–42.

Palser, A.L., Grayson, N.E., White, R.E., Corton, C., Correia, S., Ba Abdullah, M.M., Watson, S.J., Cotten, M., Arrand, J.R., Murray, P.G., Allday, M.J., Rickinson, A.B., Young, L.S., Farrell, P.J., Kellam, P., 2015. Genome diversity of Epstein-Barr virus from multiple tumor types and normal infection. Journal of virology 89, 5222–37. doi:10.1128/JVI.03614-14

Parker, B.D., Bankier, A., Satchwell, S., Barrell, B.T., Farrell, P.J., 1990. Sequence and transcription of Raji Epstein-Barr virus DNA spanning the B95-8 deletion region. Virology 179, 339–346. doi:10.1016/0042-6822(90)90302-8

Parkin, J., Cohen, B., 2001. An overview of the immune system. Lancet. doi:10.1016/S0140-6736(00)04904-7

Pedergnana, Vincent;, S., Laurène;, C., Aurélie;, G., Julien;, B., Pauline;, F., Christophe;, C., Patrice;, H., Olivier;, L.-P., Catherine;, A., Corinne;, T., Yassine;, A., Alexandre;, T., Ioannis;, B., Caroline;, A., LaurentPedergnana, Vincent;, S., Laurène;, C., Aurélie;, G., Julien;, B., Pauline;, F., Christophe;, C., Patrice;, H., Olivier;, L.-P., Catherine;, A., Corinne;, T., Yassine;, A., Alexandre;, T., Ioannis;, B., Caroline;, A., 2014. Combined linkage and association studies show that HLA class II variants control levels of antibodies against Epstein-Barr virus antigens. PloS one.

Pender, M.P., 2011. The Essential Role of Epstein-Barr Virus in the Pathogenesis of Multiple Sclerosis. The Neuroscientist 17, 351–367. doi:10.1177/1073858410381531

Pender, M.P., Burrows, S.R., 2014. Epstein–Barr virus and multiple sclerosis: potential opportunities for immunotherapy. Clinical & Translational Immunology 3, e27. doi:10.1038/cti.2014.25

Penn, D.J., 2002. Major Histocompatibility. Encyclopedia of Life Sciences 1–7. doi:10.1038/npg.els.0003986

Pereyra, F., Jia, X., McLaren, P.J., Telenti, A., de Bakker, P.I.W., Walker, B.D., Ripke, S., Brumme, C.J., Pulit, S.L., Carrington, M., Kadie, C.M., Carlson, J.M., Heckerman, D., Graham, R.R., Plenge, R.M., Deeks, S.G., Gianniny, L., Crawford, G., Sullivan, J., Gonzalez, E., Davies, L., Camargo, A., Moore, J.M., Beattie, N., Gupta, S., Crenshaw, A., Burtt, N.P., Guiducci, C., Gupta, N., Gao, X., Qi, Y., Yuki, Y., Piechocka-Trocha, A., Cutrell, E., Rosenberg, R., Moss, K.L., Lemay, P., O'Leary, J., Schaefer, T., Verma, P., Toth, I., Block, B., Baker, B., Rothchild, A., Lian, J., Proudfoot, J., Alvino, D.M.L., Vine, S., Addo, M.M., Allen, T.M., Altfeld, M., Henn, M.R., Le Gall, S., Streeck, H., Haas, D.W., Kuritzkes, D.R., Robbins, G.K., Shafer, R.W., Gulick, R.M., Shikuma, C.M., Haubrich, R., Riddler, S., Sax, P.E., Daar, E.S., Ribaudo, H.J., Agan, B., Agarwal, S., Ahern, R.L., Allen, B.L., Altidor, S., Altschuler, E.L., Ambardar, S., Anastos, K., Anderson, B., Anderson, V., Andrady, U., Antoniskis, D., Bangsberg, D., Barbaro, D., Barrie, W., Bartczak, J., Barton, S., Basden, P., Basgoz, N., Bazner, S., Bellos, N.C., Benson, A.M., Berger, J., Bernard, N.F., Bernard, A.M., Birch, C., Bodner, S.J., Bolan, R.K., Boudreaux, E.T., Bradley, M., Braun, J.F., Brndjar, J.E., Brown, S.J., Brown, K., Brown, S.T., Burack, J., Bush, L.M., Cafaro, V., Campbell, O., Campbell, J., Carlson, R.H., Carmichael, J.K., Casey, K.K., Cavacuiti, C., Celestin, G., Chambers, S.T., Chez, N., Chirch, L.M., Cimoch, P.J., Cohen, D., Cohn, L.E., Conway, B., Cooper, D.A., Cornelson, B., Cox, D.T., Cristofano, M. V, Cuchural, G., Czartoski, J.L., Dahman, J.M., Daly, J.S., Davis, B.T., Davis, K., Davod, S.M., DeJesus, E., Dietz, C.A., Dunham, E., Dunn, M.E., Ellerin, T.B., Eron, J.J., Fangman, J.J.W., Farel, C.E., Ferlazzo, H., Fidler, S., Fleenor-Ford, A., Frankel, R., Freedberg, K.A., French, N.K., Fuchs, J.D., Fuller, J.D., Gaberman, J., Gallant, J.E., Gandhi, R.T., Garcia, E., Garmon, D., Gathe, J.C., Gaultier, C.R., Gebre, W., Gilman, F.D., Gilson, I., Goepfert, P.A., Gottlieb, M.S., Goulston, C., Groger, R.K., Gurley, T.D., Haber, S., Hardwicke, R., Hardy, W.D., Harrigan, P.R., Hawkins, T.N., Heath, S., Hecht, F.M., Henry, W.K., Hladek, M., Hoffman, R.P., Horton, J.M., Hsu, R.K., Huhn, G.D., Hunt, P., Hupert, M.J., Illeman, M.L., Jaeger, H., Jellinger, R.M., John, M., Johnson, J.A., Johnson, K.L., Johnson, H., Johnson, K., Joly, J., Jordan, W.C., Kauffman, C.A., Khanlou, H., Killian, R.K., Kim, A.Y., Kim, D.D., Kinder, C.A., Kirchner, J.T., Kogelman, L., Kojic, E.M., Korthuis, P.T., Kurisu, W., Kwon, D.S., LaMar, M., Lampiris, H., Lanzafame, M., Lederman, M.M., Lee, D.M., Lee, J.M.L., Lee, M.J., Lee, E.T.Y., Lemoine, J., Levy, J.A., Llibre, J.M., Liguori, M.A., Little, S.J., Liu, A.Y., Lopez, A.J., Loutfy, M.R., Loy, D., Mohammed, D.Y., Man, A., Mansour, M.K., Marconi, V.C., Markowitz, M., Marques, R., Martin, J.N., Martin, H.L., Mayer, K.H., McElrath, M.J., McGhee, T.A., McGovern, B.H., McGowan, K., McIntyre, D., Mcleod, G.X., Menezes, P., Mesa, G., Metroka, C.E., Meyer-Olson, D., Miller, A.O., Montgomery, K., Mounzer, K.C., Nagami, E.H., Nagin, I., Nahass, R.G., Nelson, M.O., Nielsen, C., Norene, D.L., O'Connor, D.H., Ojikutu, B.O., Okulicz, J., Oladehin, O.O., Oldfield, E.C., Olender, S.A., Ostrowski, M., Owen, W.F., Pae, E., Parsonnet, J., Pavlatos, A.M., Perlmutter, A.M., Pierce, M.N., Pincus, J.M., Pisani, L., Price, L.J., Proia, L., Prokesch, R.C., Pujet, H.C., Ramgopal, M., Rathod, A., Rausch, M., Ravishankar, J., Rhame, F.S., Richards, C.S., Richman, D.D., Rodes, B., Rodriguez, M., Rose, R.C., Rosenberg, E.S., Rosenthal, D., Ross, P.E., Rubin, D.S., Rumbaugh, E., Saenz, L., Salvaggio, M.R., Sanchez, W.C., Sanjana, V.M., Santiago, S., Schmidt, W., Schuitemaker, H., Sestak, P.M., Shalit, P., Shay, W., Shirvani, V.N., Silebi, V.I., Sizemore, J.M., Skolnik, P.R., Sokol-Anderson, M., Sosman, J.M., Stabile, P., Stapleton, J.T., Starrett, S., Stein, F., Stellbrink, H.-J., Sterman, F.L., Stone, V.E., Stone, D.R., Tambussi, G., Taplitz, R.A., Tedaldi, E.M., Telenti, A., Theisen, W., Torres, R., Tosiello, L., Tremblay, C., Tribble, M.A., Trinh, P.D., Tsao, A., Ueda, P., Vaccaro, A., Valadas, E., Vanig, T.J., Vecino, I., Vega, V.M., Veikley, W., Wade, B.H., Walworth, C., Wanidworanun, C., Ward, D.J., Warner, D.A., Weber, R.D., Webster, D., Weis, S., Wheeler, D.A., White, D.J., Wilkins, E., Winston, A., Wlodaver, C.G., van't Wout, A., Wright, D.P., Yang, O.O., Yurdin, D.L., Zabukovic, B.W., Zachary, K.C., Zeeman, B., Zhao, M., 2010. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science (New York, NY) 330, 1551–1557. doi:10.1126/science.1195271

Petersdorf, E.W., 2008. Optimal HLA matching in hematopoietic cell transplantation. Current Opinion in Immunology. doi:10.1016/j.coi.2008.06.014

Poole, B.D., Scofield, R.H., Harley, J.B., James, J.A., 2006. Epstein-Barr virus and molecular mimicry in systemic lupus erythematosus. Autoimmunity. doi:10.1080/08916930500484849

Posnett, D.N., 2008. Herpesviruses and autoimmunity. Current opinion in investigational drugs (London, England : 2000) 9, 505–14.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics 81, 559–575. doi:10.1086/519795

Qureshi, S.T., Skamene, E., Malo, D., 1999. Comparative genomics and host resistance against infectious diseases. Emerging infectious diseases 5, 36–47. doi:10.3201/eid0501.990105

Rabin, H., Neubauer, R.H., Hopkins, R.F., Nonoyama, M., 1978. Further characterization of a herpesvirus□positive orang□utan cell line and comparative aspects of in vitro transformation with lymphotropic old world primate herpesviruses. International Journal of Cancer 21, 762–767. doi:10.1002/ijc.2910210614

Raychaudhuri, S., 2011. Mapping rare and common causal alleles for complex human diseases. Cell. doi:10.1016/j.cell.2011.09.011

Raychaudhuri, S., Sandor, C., Stahl, E.A., Freudenberg, J., Lee, H.-S., Jia, X., Alfredsson, L., Padyukov, L., Klareskog, L., Worthington, J., Siminovitch, K.A., Bae, S.-C., Plenge, R.M., Gregersen, P.K., de Bakker, P.I.W., 2012. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nature Genetics 44, 291–296. doi:10.1038/ng.1076

Ressing, M.E., Horst, D., Griffin, B.D., Tellam, J., Zuo, J., Khanna, R., Rowe, M., Wiertz, E.J.H.J., 2008. Epstein-Barr virus evasion of CD8+ and CD4+ T cell immunity via concerted actions of multiple gene products. Seminars in Cancer

Biology. doi:10.1016/j.semcancer.2008.10.008

Rickinson, A.B., Young, L.S., Rowe, M., 1987. Influence of the Epstein-Barr Virus Nuclear Antigen EBNA 2 on the Growth Phenotype of Virus-Transformed B Cells. Journal of Virology 61, 1310–1317.

Rickinson, A, B.M., D, Jr., A, B.M., 1997. Human cytotoxic T lymphocyte responses to Epstein-Barr virus infection. Annual review of immunology.

Ring, C.J.A., 1994. The B cell-immortalizing functions of Epstein-Barr virus. Journal of General Virology.

Rioux, J.D., Goyette, P., Vyse, T.J., Hammarström, L., Fernando, M.M.A., Green, T., De Jager, P.L., Foisy, S., Wang, J., de Bakker, P.I.W., Leslie, S., McVean, G., Padyukov, L., Alfredsson, L., Annese, V., Hafler, D.A., Pan-Hammarström, Q., Matell, R., Sawcer, S.J., Compston, A.D., Cree, B.A.C., Mirel, D.B., Daly, M.J., Behrens, T.W., Klareskog, L., Gregersen, P.K., Oksenberg, J.R., Hauser, S.L., 2009. Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. Proceedings of the National Academy of Sciences of the United States of America 106, 18680–5. doi:10.1073/pnas.0909307106

Rist, Melissa, J.N., Michelle, A.B., Jacqueline, M.B., Scott, Rr., Melissa, J.N., Michelle, A.B., Jacqueline, M.B., n.d. T cell epitope clustering in the highly immunogenic BZLF1 antigen of Epstein-Barr virus. Journal of virology.

Rivailler, P., Jiang, H., Cho, Y.-G., Quink, C., Wang, F., 2002. Complete Nucleotide Sequence of the Rhesus Lymphocryptovirus: Genetic Validation for an Epstein-Barr Virus Animal Model. JOURNAL OF VIROLOGY 76, 421–426. doi:10.1128/JVI.76.1.421–426.2002

Roizman, B., 2007. Human Herpesviruses: Biology, Therapy and Immunoprophylaxis, Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.

doi:10.2277/0521827140

Rousseau, C.M., Lockhart, D.W., Listgarten, J., Maley, S.N., Kadie, C., Learn, G.H., Nickle, D.C., Heckerman, D.E., Deng, W., Brander, C., Ndung'u, T., Coovadia, H., Goulder, P.J.R., Korber, B.T., Walker, B.D., Mullins, J.I., 2009. Rare HLA drive additional HIV evolution compared to more frequent alleles. AIDS research and human retroviruses 25, 297–303. doi:10.1089/aid.2008.0208

Rowles, D.L., Terhune, S.S., Cristea, I.M., 2013. Discovery of host-viral protein complexes during infection. Methods in Molecular Biology 1064, 43–70. doi:10.1007/978-1-62703-601-6_4

Rubicz, R., Yolken, R., Drigalenko, E., Carless, M.A., Dyer, T.D., Bauman, L., Melton, P.E., Kent, J.W., Harley, J.B., Curran, J.E., Johnson, M.P., Cole, S.A., Almasy, L., Moses, E.K., Dhurandhar, N. V., Kraig, E., Blangero, J., Leach, C.T., Göring, H.H.H., 2013. A Genome-Wide Integrative Genomic Study Localizes Genetic Factors Influencing Antibodies against Epstein-Barr Virus Nuclear Antigen 1 (EBNA-1). PLoS Genetics 9. doi:10.1371/journal.pgen.1003147

Rubicz, Rohina;, Y., Robert;, D., Eugene;, C., Melanie, A.D., Thomas, D.B., Lara;, M., Phillip, E.K., Jack, W.H., John, B.C., Joanne, E.J., Matthew, P.C., Shelley, A.A., Laura;, M., Eric, K.D., Nikhil, V.K., Ellen;, B., John;, L., Charles, T.G., Harald, H.Hr., Rohina;, Y., Robert;, D., Eugene;, C., Melanie, A.D., Thomas, D.B., Lara;, M., Phillip, E.K., Jack, W.H., John, B.C., Joanne, E.J., Matthew, P.C., Shelley, A.A., Laura;, M., Eric, K.D., Nikhil, V.K., Ellen;, B., John;, L., Charles, T.G., 2013. A genome-wide integrative genomic study localizes genetic factors influencing antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1). PLoS genetics

Sample, J., Young, L., Martin, B., Chatman, T., Kieff, E., Rickinson, a, Kieff, E., 1990. Epstein-Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B, and EBNA-3C genes. Journal of virology 64, 4084–4092.

Sanchez-Mazas, A., Buhler, S., Nunes, J.M., 2014. A new HLA map of Europe: Regional genetic variation and its implication for peopling history, disease-association studies and tissue transplantation, in: Human Heredity. pp. 162–177. doi:10.1159/000360855

Sanchez-Mazas, A., Meyer, D., 2014. The relevance of HLA sequencing in population genetics studies. Journal of Immunology Research. doi:10.1155/2014/971818

Sandvej, K., Zhou, X.G., Hamilton-Dutoit, S., 2000. EBNA-1 sequence variation in Danish and Chinese EBV-associated tumours: Evidence for geographical polymorphism but not for tumour-specific subtype restriction. Journal of Pathology 191, 127–131. doi:10.1002/(SICI)1096-9896(200006)191:2<127::AID-PATH614>3.0.CO;2-E

Santpere, G., Darre, F., Blanco, S., Alcami, A., Villoslada, P., Mar Albà, M., Navarro, A., 2014. Genome-wide analysis of wild-type Epstein-Barr virus genomes derived from healthy individuals of the 1,000 Genomes Project. Genome biology and evolution 6, 846–60. doi:10.1093/gbe/evu054

Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C.A., Patsopoulos, N.A., Moutsianas, L., Su, Z., Freeman, C., Hunt, S.E., Edkins, S., Gray, E., David, R., Potter, S.C., Goris, A., Band, G., Oturai, A.B., Strange, A., Comabella, M., Hammond, N., Kockum, I., Mccann, O.T., Ban, M., Dronov, S., Robertson, N., Bumpstead, S.J., Lisa, F., International, T., Sclerosis, M., Consortium, G., Case, W.T., Wtccc, C.C., 2012. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 476, 214–219. doi:10.1038/nature10251

Sebastiani, P., Timofeev, N., Dworkis, D.A., Perls, T.T., Steinberg, M.H., 2009. Genome-wide association studies and the genetic dissection of complex traits. American Journal of Hematology. doi:10.1002/ajh.21440

Seitzer, U., Gerdes, J., Müller-Quernheim, J., 2002. Genotyping in the MHC locus: potential for defining predictive markers in sarcoidosis. Respiratory research 3, 6. doi:10.1186/rr178

Shah, N., Decker, W.K., Lapushin, R., Xing, D., Robinson, S.N., Yang, H., Parmar, S., Tung, S.S., O'Brien, S., Fernandez-Viña, M., Shpall, E.J., Wierda, W.G., 2011. HLA homozygosity and haplotype bias among patients with chronic lymphocytic leukemia: implications for disease control by physiological immune surveillance. Leukemia 25, 1036–1039. doi:10.1038/leu.2011.30

Sheng, W., Bouguermouh, A., Bouzid, M., Djennaoui, D., Ooka, T., 2004. BAMHI DNA fragment H-polymorphism of Epstein-Barr virus is associated with the mutations present in an 89 BP sequence localized in EBNA2 gene. Virus Genes 29, 99–108. doi:10.1023/B:VIRU.0000032793.30419.6c

Shiina, T., Hosomichi, K., Inoko, H., Kulski, J.K., 2009. The HLA genomic loci map: expression, interaction, diversity and disease. Journal of Human Genetics 54, 15–39. doi:10.1038/jhg.2008.5

Shimizu, K., 1997. [Mechanisms of antigenic variation in influenza virus]. Nihon rinsho Japanese journal of clinical medicine 55, 2610–6.

Sie, L., Loong, S., Tan, E.K., 2009. Utility of lymphoblastoid cell lines. Journal of neuroscience research 87, 1953–1959. doi:10.1002/jnr.22000

Simbiri, K.O., Smith, N.A., Otieno, R., Wohlford, E.E.M., Daud, I.I., Odada, S.P., Middleton, F., Rochford, R., 2015. Epstein-barr virus genetic variation in lymphoblastoid cell lines derived from Kenyan pediatric population. PLoS ONE 10, 1–18. doi:10.1371/journal.pone.0125420

Song, K.-A., Yang, S.-D., Hwang, J., Kim, J.-I., Kang, M.-S., 2015. The full-length DNA sequence of Epstein Barr virus from a human gastric carcinoma cell line, SNU-719. Virus Genes 51, 329–337. doi:10.1007/s11262-015-1248

Stevens, S.J.C., Blank, B.S.N., Smits, P.H.M., Meenhorst, P.L., Middeldorp, J.M., 2002a. High Epstein-Barr virus (EBV) DNA loads in HIV-infected patients: correlation with antiretroviral therapy and quantitative EBV serology. AIDS (London, England) 16, 993–1001. doi:10.1097/00002030-200205030-00005

Stevens, S.J.C., Verschuuren, E.A.M., Brule, A.J.C.V.A.N.D.E.N., Middeldorp, J.M., 2002b. Role of Epstein – Barr Virus DNA Load Monitoring in Prevention and Early Detection of Post-transplant Lymphoproliferative Disease 43, 831–840. doi:10.1080/10428190290016971

Stranger, B.E., Stahl, E.A., Raj, T., 2011. Progress and promise of genome-wide association studies for human complex trait genetics. Genetics. doi:10.1534/genetics.110.120907

Sugden, B., 1982. Epstein-Barr virus: a human pathogen inducing lymphoproliferation in vivo and in vitro. Rev Infect Dis.

Sun, L., Zhao, Z., Liu, S., Liu, X., Sun, Z., Luo, B., 2015. Sequence variation analysis of Epstein-Barr virus nuclear antigen 1 gene in the virus associated lymphomas of Northern China. PLoS ONE 10, 1–12. doi:10.1371/journal.pone.0140529

Szpara, M.L., Gatherer, D., Ochoa, A., Greenbaum, B., Dolan, A., Bowden, R.J., Enquist, L.W., Legendre, M., Davison, A.J., 2014. Evolution and Diversity in Human Herpes Simplex Virus Genomes. Journal of Virology 88, 1209–1227. doi:10.1128/JVI.01987-13

Tang, J., Tang, S., Lobashevsky, E., Myracle, A.D., Fideli, U., Aldrovandi, G., Allen, S., Musonda, R., Kaslow, R.A., Hiv, Z., 2002. Favorable and Unfavorable HLA Class I Alleles and Haplotypes in Zambians Predominantly Infected with Clade C Human Immunodeficiency Virus Type 1 76, 8276–8284.

doi:10.1128/JVI.76.16.8276

The 1000 Genomes Project Consortium, 2015. A global reference for human genetic variation. Nature. doi:10.1038/nature15393

Thompson, M.P., Kurzrock, R., 2004. Epstein-Barr Virus and Cancer. Clinical Cancer Research. doi:10.1158/1078-0432.CCR-0670-3

Thorley-Lawson, D.A., Hawkins, J.B., Tracy, S.I., Shapiro, M., 2013. The pathogenesis of Epstein-Barr virus persistent infection. Current Opinion in Virology 3, 227–232. doi:10.1016/j.coviro.2013.04.005

Tiwawech, D., Srivatanakul, P., Karalak, A., Ishida, T., 2008. Association between EBNA2 and LMP1 subtypes of Epstein-Barr virus and nasopharyngeal carcinoma in Thais. Journal of Clinical Virology 42, 1–6. doi:10.1016/j.jcv.2007.11.011

Toussirot, E., Roudier, J., 2008. Epstein-Barr virus in autoimmune diseases. Best practice & research Clinical rheumatology 22, 883–96. doi:10.1016/j.berh.2008.09.007

Trampush, J., Yang, M., Yu, J., Knowles, E., Davies, G., Liewald, D., Starr, J., 2017. GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. Nature Publishing Group 1–10. doi:10.1038/mp.2016.244

Trowsdale, J., Knight, J.C., 2013. Major histocompatibility complex genomics and human disease. Annual review of genomics and human genetics 14, 301–23. doi:10.1146/annurev-genom-091212-153455

Tsai, M.H., Raykova, A., Klinke, O., Bernhardt, K., Gärtner, K., Leung, C., Geletneky, K., Sertel, S., Münz, C., Feederle, R., Delecluse, H.J., 2013. Spontaneous Lytic Replication and Epitheliotropism Define an Epstein-Barr Virus Strain Found in Carcinomas. Cell Reports 5, 458–470. doi:10.1016/j.celrep.2013.09.012

Tso, K.K.-Y., Yip, K.Y.-L., Mak, C.K.-Y., Chung, G.T.-Y., Lee, S.-D., Cheung, S.-T., To, K.-F., Lo, K.-W., 2013. Complete genomic sequence of Epstein-Barr virus in nasopharyngeal carcinoma cell line C666-1. Infectious agents and cancer 8, 29. doi:10.1186/1750-9378-8-29

Tzellos, S., Farrell, P.J., 2012. Epstein-Barr Virus Sequence Variation—Biology and Disease. Pathogens 1, 156–175. doi:10.3390/pathogens1020156

Urayama, Kevin, Y.J., Ruth, F.H., Henrik;, D., Arjan;, K., Yoichiro;, C., Amelie;, G., Valerie;, B., Anne;, N., Alexandra;, B., Nikolaus;, F., Lenka;, B., Yolanda;, M., Marc;, S., Anthony;, S., Lesley;, L., Annette;, M., Dorothy;, T., Malcolm;, S., Karin, E.A., Rose-Marie;, A., Hans-Olov;, G., Bengt;, F., Bjarke;, N., Ilja, M.V., Lydia;, van I., Gustaaf, W.L., Tracy;, C., Pierluigi;, K., Lambertus;, V., Sita, H.H., Ivana;, V., Lars;, M., Gary, J.T., Peter;, C., David, I.B., Simone;, A., Antonio;, H., Claire, M.O., Kim;, T., Anne;, M., Beatrice;, C., Federico;, K., Kay-Tee;, T., Ruth, C.P., Petra, H.M.G., Carlos, A.Q., José, R.S., María-José;, H., José, M.A., Eva;, D., Miren;, C.-C., Françoise;, B.-M., H, B.R., Elio;, R., Eve;, B., Paolo;, de S., Silvia;, Z., Diana;, M., Mads;, van den B., Anke;, L., Mark;, B., Paul;, M., James, Du., Kevin, Y.J., Ruth, F.H., Henrik;, D., Arjan;, K., Yoichiro;, C., Amelie;, G., Valerie;, B., Anne;, N., Alexandra;, B., Nikolaus;, F., Lenka;, B., Yolanda;, M., Marc;, S., Anthony;, S., Lesley;, L., Annette;, M., Dorothy;, T., Malcolm;, S., Karin, E.A., Rose-Marie;, A., Hans-Olov;, G., Bengt;, F., Bjarke;, N., Ilja, M.V., Lydia;, van I., Gustaaf, W.L., Tracy;, C., Pierluigi;, K., Lambertus;, V., Sita, H.H., Ivana;, V., Lars;, M., Gary, J.T., Peter;, C., David, I.B., Simone;, A., Antonio;, H., Claire, M.O., Kim;, T., Anne;, M., Beatrice;, C., Federico;, K., Kay-Tee;, T., Ruth, C.P., Petra, H.M.G., Carlos, A.Q., José, R.S., María-José;, H., José, M.A., Eva;, D., Miren;, C.-C., Françoise;, B.-M., H, B.R., Elio;, R., Eve;, B., Paolo;, de S., Silvia;, Z., Diana;, M., Mads;, van den B., Anke;, L., Mark;, B., Paul;, M., n.d. Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. Journal of the National Cancer Institute.

Visscher, P.M., Brown, M.A., McCarthy, M.I., Yang, J., 2012. Five years of GWAS discovery. American Journal of Human Genetics. doi:10.1016/j.ajhg.2011.11.029

Wang, F., 2013. Nonhuman primate models for Epstein-Barr virus infection. Current Opinion in Virology 3, 233–237. doi:10.1016/j.coviro.2013.03.003

Warren, R.L., Choe, G., Freeman, D.J., Castellarin, M., Munro, S., Moore, R., Holt, R.A., 2012. Derivation of HLA types from shotgun sequence datasets. Genome medicine 4, 95. doi:10.1186/gm396

Weatherall, D., Clegg, J., Kwiatkowski, D., 1997. The role of genomics in studying genetic susceptibility to infectious disease. Genome Research. doi:10.1101/gr.7.10.967

Wellcome, T., Case, T., Consortium, C., 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661–78. doi:10.1038/nature05911

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., Parkinson, H., 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Research 42. doi:10.1093/nar/gkt1229

Wensing, B., Farrell, P.J., 2000. Regulation of cell growth and death by Epstein-Barr virus. Microbes and infection / Institut Pasteur 2, 77–84.

Wheeler, H.E., Dolan, M.E., 2012. Lymphoblastoid cell lines in pharmacogenomic discovery and clinical translation. Pharmacogenomics 13, 55–70. doi:10.2217/pgs.11.121

Wolf, H., zur Hausen, H., Becker, V., 1973. EB viral genomes in epithelial nasopharyngeal carcinoma cells. Nature 244, 245–247. doi:10.1038/10.1038/newbio244245a0

Woulfe, J.M., Gray, M.T., Gray, D.A., Munoz, D.G., Middeldorp, J.M., 2014. Hypothesis: A role for EBV-induced molecular mimicry in Parkinson’s disease. Parkinsonism and Related Disorders. doi:10.1016/j.parkreldis.2014.02.031

Wrightham, M.N., Stewart, J.P., Janjua, N.J., Pepper, S.D., Sample, C., Rooney, C.M., Arrand, J.R., 1995. Antigenic and sequence variation in the C-terminal unique domain of the Epstein-Barr virus nuclear antigen EBNA-1. Virology 208, 521–530. doi:10.1006/viro.1995.1183

Wucherpfennig, K.W., Strominger, J.L., 1995. Molecular mimicry in T cell-mediated autoimmunity: Viral peptides activate human T cell clones specific for myelin basic protein. Cell 80, 695–705. doi:10.1016/0092-8674(95)90348-8

Yang, J., Tao, Q., Flinn, I.W., Murray, P.G., Post, L.E., Ma, H., Piantadosi, S., Caligiuri, M. a, Ambinder, R.F., 2000. Characterization of Epstein-Barr virus-infected B cells in patients with posttransplantation lymphoproliferative disease: disappearance after rituximab therapy does not predict clinical response. Blood 96, 4055–4063.

Yao, Q.Y., Tierney, R.J., Croom-Carter, D., Dukers, D., Cooper, G.M., Ellis, C.J., Rowe, M., Rickinson, A.B., 1996. Frequency of multiple Epstein-Barr virus infections in T-cell-immunocompromised individuals. J Virol 70, 4884–4894.

Yin, Y., Lan, J.H., Nguyen, D., Valenzuela, N., Takemura, P., Bolon, Y.T., Springer, B., Saito, K., Zheng, Y., Hague, T., Pasztor, A., Horvath, G., Rigo, K., Reed, E.F., Zhang, Q., 2016. Application of high-throughput next-generation sequencing for HLA typing on buccal extracted DNA: Results from over 10,000 donor recruitment samples. PLoS ONE 11. doi:10.1371/journal.pone.0165810

Young, L.S., Dawson, C.W., 2014. Epstein-Barr virus and nasopharyngeal carcinoma. Chinese Journal of Cancer. doi:10.5732/cjc.014.10197

Young, L.S., Rickinson, A.B., 2004. Epstein-Barr virus: 40 years on. Nature reviews Cancer 4, 757–768. doi:10.1038/nrc1452

Young, L.S., Yap, L.F., Murray, P.G., 2016. Epstein–Barr virus: more than 50 years old and still providing surprises. Nature Reviews Cancer 16, 789–802. doi:10.1038/nrc.2016.92

Zeng, M.-S., Li, D.-J., Liu, Q.-L., Song, L.-B., Li, M.-Z., Zhang, R.-H., Yu, X.-J., Wang, H.-M., Ernberg, I., Zeng, Y.-X., 2005. Genomic sequence analysis of Epstein-Barr virus strain GD1 from a nasopharyngeal carcinoma patient. Journal of virology 79, 15323–30. doi:10.1128/JVI.79.24.15323-15330.2005

Zheng-Bradley, X., Flicek, P., 2016. Applications of the 1000 Genomes Project resources. Briefings in Functional Genomics 16, elw027. doi:10.1093/bfgp/elw027

Zhou, L., Chen, J., Qiu, X., Pan, Y., Zhang, Z., Shao, C., 2017. Comparative analysis of 22 Epstein–Barr virus genomes from diseased and healthy individuals. Journal of General Virology 98, 96–107. doi:10.1099/jgv.0.000699

Zhou, X., Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. Nature genetics 44, 821–4. doi:10.1038/ng.2310

Zimber, U., Adldinger, H.K., Lenoir, G.M., Vuillaume, M., Knebel-Doeberitz, M. V., Laux, , U., Bornkamm, G.W., 1986. Geographical prevalence of two types of Epstein-Barr virus. Virology 154, 56–66. doi:10.1016/0042-6822(86)90429-0

10. APPENDIX

10.1 Supplementary data of thesis chapters

Open reading frame	Protein Common name	Alternative name	Main proposed function
Latent genes			
BKRF1	EBNA-1		Replication, transcriptional regulation
BYRF1	EBNA-2		Trans-activation
BERF1	EBNA-3A	EBNA-3	Transcriptional regulation
BERF2	EBNA-3B	EBNA-4	Transcriptional regulation
BERF3/4	EBNA-3C	EBNA-6	Transcriptional regulation
BWRF1	EBNA-LP	EBNA-5	Trans-activation
BNLF1	LMP-1		B-cell survival, anti-apoptosis
BNRF1	LMP-2A/2B	TP1/2	Maintenance of latency
BARF0			Not shown to be translated
EBER1/2			Regulation of innate immunity
Lytic genes **Immediate early genes**			
BZLF1	ZEBRA		Initiation of lytic cycle
BRLF1			Initiation of lytic cycle
BILF4			Initiation of lytic cycle
Early Genes			
BMRF1			Trans-activation
BALF2			DNA binding
BALF5			DNA polymerase
BORF2			Ribonucleotide reductase subunit
BARF1			Ribonucleotide reductase subunit
BXLF1			Thymidine kinase
BGLF5			Alkaline exonuclease
BSLF1			Primase

BBLF4		Helicase
BKRF3		Uracil DNA glycosylase
Late genes		
BLLF1	gp350/220	Major envelope glycoprotein
BXLF2	gp85 (gH)	Virus–host envelope fusion
BKRF2	gp25 (gL)	Virus–host envelope fusion
BZLF2	gp42	Binds MHC class II
BALF4	gp110	B infection
BDLF3	gp100	Immune Invasion
BILF2	gp55	Immune Invasion
BCRF1		Immune Invasion
BHRF1		Viral bcl-2 analogue

Table: 1 List of EBV genes and their respective protein products (Figure has been adapted from ARC MONOGRAPHS – 100B)

EBV protein	Human protein
BCRF1	Interleukin 10
BDLF2	Cyclin B1
BHRF1	BCL-2
BALF1	BCL-2
BARF1	C-FMS receptor
NA	ICAM-1 (CD54)

Table: 2 List of EBV proteins showing homology with human proteins

CEU		CHB		CHS		YRI		TSI	
Allele	P-val	Allele	P-val	Allele	P-val	Allele	P-val	Allele	P-val
A1	0.9569	**A1**	**0.0755**	A1	0.8126	A1	0.435	A1	0.814
A2	0.1662	A2	0.924	A2	0.135	A2	0.455	A2	0.642
A3	0.1727	A3	0.1667	A3	0.6551	A3	0.484	A3	0.643
A11	0.3704	A11	0.7048	A11	0.1303	A11	0.465	A11	0.439
A23	0.4056	A24	0.9636	A24	0.1609	A23	0.478	A23	0.833
A24	0.1235	A26	0.2097	A26	0.1113	A26	0.421	A24	0.821
A26	0.7416	A30	0.8655	A30	0.9537	A29	0.556	A26	0.619
A29	0.4006	A31	0.1534	A31	0.903	A30	0.437	A29	0.437
A30	0.5981	**A33**	**0.0611**	**A32**	**0.048**	A31	0.817	A30	0.498
A31	0.1134	A66	0.1667	A33	NA	A32	0.517	A31	0.574
A32	0.8802	A68	NA	B7	0.1569	A33	0.512	A32	0.643
A66	0.7234	B7	0.2698	B8	0.5425	A36	0.561	A33	0.492
A68	NA	B8	0.3629	B13	0.7799	A66	0.548	A66	0.87
B7	0.5727	B13	0.1837	B15	0.3066	A68	0.507	A68	0.829
B8	0.8668	**B14**	**0.0574**	B27	0.8146	A74	NA	A69	NA
B13	0.3898	B15	0.1517	B35	0.9374	B7	0.686	B7	0.872
B40	0.3554	B46	0.1715	B51	0.1511	B42	0.393	B38	0.921
B41	0.8119	B48	0.4814	B52	0.2115	B44	0.784	B39	0.788
B42	0.3729	B50	0.4135	B53	0.1802	B45	0.547	B40	0.933
B44	0.737	B51	0.2866	B54	0.4265	B46	0.426	B41	0.818
B46	0.6429	**B52**	**0.0727**	B55	0.5137	B49	0.588	B42	0.957
B48	0.7368	B53	0.2408	**B56**	**0.0523**	B50	0.605	B44	0.835
B49	0.7094	B54	0.1378	B57	0.4094	B51	0.899	B46	0.979
B50	0.7763	B56	0.3684	B58	0.2756	B52	0.586	B49	0.678
B51	0.6447	B58	0.2119	B59	0.2445	B53	0.962	B50	0.86
B55	0.3666	B67	NA	B67	0.8016	B57	0.497	B51	0.825
B57	0.8245	C1	0.262	B78	NA	B58	0.887	B52	0.949
B58	0.3946	C2	NA	C1	0.4015	B67	0.58	B53	0.583
B67	NA	C3	0.3753	C3	0.8291	B78	NA	B54	0.709
C1	0.2913	C4	0.4205	C4	0.1415	C1	0.611	B55	0.511
C2	0.7716	C5	0.2821	C5	0.2925	C2	0.667	B57	0.98
C3	**0.0548**	C6	0.723	C6	0.4046	C3	0.65	B58	0.835
C4	**0.0758**	C7	0.9207	C7	0.4825	C4	0.653	B67	0.557
C5	**0.0978**	C8	0.7304	C8	0.6129	C5	NA	B78	0.645
C6	**0.031**	C12	0.3347	C12	0.3966	C7	0.961	B82	NA
C7	**0.0549**	C14	0.1195	C14	0.5632	C8	0.437	C1	0.527
C8	**0.0395**	C15	0.9887	C15	0.4621	C12	0.368	C2	0.487

Table: 3 linear model output with HLA alleles and its corresponding P-value
(Only partial out is shown here due to the space constraint)

Strain_Name	Disease	Genome accession	Population	EBV type	Reference
BL36	BL	LN827557	Africa	1	PMID: 25787276
AG876	BL	NC_009334	Africa	2	PMID: 25787276
BL37	BL	LN827526	Africa_unkonwn	1	PMID: 25787276
M-ABA	LCL_ NCP	LN827527	Africa_unkonwn	1	PMID: 25787276
AG876	BL	DQ279927	African	2	PMID: 16490228
CV-ARG	BL	KR063343	Argentina	1	PMID: 26593963
sLCL-IS2.01	sLCL_ PTLD	LN827589	Australia	2	PMID: 25787276
sLCL-IM1.09	sLCL_IM	LN827567	Australia	1	PMID: 25787276
sLCL-IM1.17	sLCL_IM	LN827583	Australia	1	PMID: 25787276
sLCL-IM1.05	sLCL_IM	LN827590	Australia	1	PMID: 25787276
sLCL-IM1.02	sLCL_IM	LN827596	Australia	1	PMID: 25787276
sLCL-IM1.16	sLCL_IM	LN827799	Australia	1	PMID: 25787276
sLCL-IS1.08	sLCL_PTLD	LN827553	Australia	1	PMID: 25787276
sLCL-IS1.11	sLCL_PTLD	LN827569	Australia	1	PMID: 25787276
sLCL-IS1.01	sLCL_PTLD	LN827570	Australia	1	PMID: 25787276
sLCL-IS1.18	sLCL_PTLD	LN827572	Australia	1	PMID: 25787276
sLCL-IS1.14	sLCL_PTLD	LN827575	Australia	1	PMID: 25787276
sLCL-IS1.20	sLCL_PTLD	LN827576	Australia	1	PMID: 25787276
sLCL-IS1.13	sLCL_PTLD	LN827578	Australia	1	PMID: 25787276
sLCL-IS1.06	sLCL_PTLD	LN827584	Australia	1	PMID: 25787276
sLCL-IS1.15	sLCL_PTLD	LN827586	Australia	1	PMID: 25787276
sLCL-IS1.19	sLCL_PTLD	LN827588	Australia	1	PMID: 25787276
sLCL-IS1.10	sLCL_PTLD	LN827592	Australia	1	PMID: 25787276
sLCL-IS1.12	sLCL_PTLD	LN827593	Australia	1	PMID: 25787276
sLCL-IS1.07	sLCL_PTLD	LN827594	Australia	1	PMID: 25787276
sLCL-IS1.03	sLCL_PTLD	LN827595	Australia	1	PMID: 25787276
sLCL-IS1.04	sLCL_PTLD	LN827597	Australia	1	PMID: 25787276
CCH	BL	KP968257	Brazil	1	PMID: 26593963
MP	BL	KP968258	Brazil	1	PMID: 26593963
SCL	BL	KP968259	Brazil	1	PMID: 26593963
VGO	BL	KP968260	Brazil	1	PMID: 26593963
RPF	BL	KR063344	Brazil	1	PMID: 26593963

FNR	BL	KR063345	Brazil	1	PMID: 26593963
EBVaGC3	EBVaGC	KT254013	China	1	PMID: 26716899
EBVaGC1	EBVaGC	KT273942	China	1	PMID: 26716899
EBVaGC2	EBVaGC	KT273943	China	1	PMID: 26716899
EBVaGC4	EBVaGC	KT273944	China	1	PMID: 26716899
EBVaGC5	EBVaGC	KT273945	China	1	PMID: 26716899
EBVaGC6	EBVaGC	KT273946	China	1	PMID: 26716899
EBVaGC7	EBVaGC	KT273947	China	1	PMID: 26716899
EBVaGC8	EBVaGC	KT273948	China	1	PMID: 26716899
EBVaGC9	EBVaGC	KT273949	China	1	PMID: 26716899
NA	LC	KT823506	China	1	PMID: 27189712
NA	LC	KT823507	China	1	PMID: 27189712
NA	LC	KT823508	China	1	PMID: 27189712
NA	LC	KT823509	China	1	PMID: 27189712
HKNPC2	NCP	KF992564	China	1	PMID: 24991008
HKNPC3	NCP	KF992565	China	1	PMID: 24991008
HKNPC4	NCP	KF992566	China	1	PMID: 24991008
HKNPC5	NCP	KF992567	China	1	PMID: 24991008
HKNPC6	NCP	KF992568	China	1	PMID: 24991008
HKNPC7	NCP	KF992569	China	1	PMID: 24991008
HKNPC8	NCP	KF992570	China	1	PMID: 24991008
HKNPC9	NCP	KF992571	China	1	PMID: 24991008
C666-1 resequence	NCP	LN827525	China	1	PMID: 25787276
C666-1	NCP	KC617875	China	1	PMID: 25787276
D3201.2	NCP	LN827549	China	1	PMID: 25787276
GD1	NCP	AY961628	China	1	PMID: 25787276
GD2	NCP	HQ020558	China	1	PMID: 25787276
L591	HL	LN827523	Germany	1	PMID: 25787276
HU11393	BL	KP968261	Ghana	1	PMID: 26593963
H018436D	BL	KP968262	Ghana	1	PMID: 26593963
H058015C	BL	KP968263	Ghana	1	PMID: 26593963
H002213	BL	KP968264	Ghana	1	PMID: 26593963
H03753A	BL	KR063342	Ghana	1	PMID: 26593963

NA	NCP	KC617875	Hong Kong	1	PMID: 23915735
M81	NCP	KF373730	Hong Kong	1	PMID: 25787276
HKNPC1	NCP	JQ009376	Hong Kong	1	PMID: 22590638
HKN14	sLCL	LN824209	Hong Kong	1	PMID: 25787276
HKN19	sLCL	LN824224	Hong Kong	1	PMID: 25787276
HKN15	sLCL	LN827547	Hong Kong	1	PMID: 25787276
Akata resequence	BL	LN824208	Japan	1	PMID: 25787276
Akata	BL	KC207813	Japan	1	PMID: 25787276
Daudi	BL	LN827545	Kenya	1	PMID: 25787276
Makau	BL	LN827551	Kenya	1	PMID: 25787276
Mutu	BL	KC207814	Kenya	1	PMID: 25787276
Mak1	BL	LN824203	Kenya	1	PMID: 25787276
Cheptages	BL	LN827556	Kenya	2	PMID: 25787276
sLCL-1.12	sLCL	LN824205	Kenya	1	PMID: 25787276
sLCL-1.11	sLCL	LN827550	Kenya	1	PMID: 25787276
sLCL-1.08	sLCL	LN827552	Kenya	1	PMID: 25787276
sLCL-1.02	sLCL	LN827558	Kenya	1	PMID: 25787276
sLCL-1.19	sLCL	LN827562	Kenya	1	PMID: 25787276
sLCL-1.18	sLCL	LN827563	Kenya	1	PMID: 25787276
sLCL-1.07	sLCL	LN827565	Kenya	1	PMID: 25787276
sLCL-1.06	sLCL	LN827566	Kenya	1	PMID: 25787276
sLCL-1.24	sLCL	LN827568	Kenya	1	PMID: 25787276
sLCL-BL1.20	sLCL	LN827571	Kenya	1	PMID: 25787276
sLCL-1.10	sLCL	LN827573	Kenya	1	PMID: 25787276
sLCL-1.09	sLCL	LN827574	Kenya	1	PMID: 25787276
sLCL-1.17	sLCL	LN827577	Kenya	1	PMID: 25787276
sLCL-1.13	sLCL	LN827579	Kenya	1	PMID: 25787276
sLCL-1.05	sLCL	LN827581	Kenya	1	PMID: 25787276
sLCL-BL1.03	sLCL	LN827582	Kenya	1	PMID: 25787276
sLCL-1.04	sLCL	LN827585	Kenya	1	PMID: 25787276
sLCL-2.14	sLCL	LN827560	Kenya	2	PMID: 25787276
sLCL-2.16	sLCL	LN827580	Kenya	2	PMID: 25787276
sLCL-2.21	sLCL	LN827587	Kenya	2	PMID: 25787276

sLCL-2.15	sLCL	LN827591	Kenya	2	PMID: 25787276
sLCL-2.22	sLCL	LN831023	Kenya	2	PMID: 25787276
GC1	EBVaGC	KP735248	Korea	1	PMID: 26459384
YCCEL1	EBVaGC	LN827561	Korea	1	PMID: 25787276
B95.8 (+ Raji)	IM	NC_007605	N. America	1	PMID: 25787276
B95-8	IM	V01555	N. America	1	PMID: 6087149
P3HR1 c16	BL	LN827548	Nigeria	2	PMID: 25787276
Jijoye	BL	LN827800	Nigeria	2	PMID: 25787276
Wewak1	BL	LN827544	PNG	2	PMID: 25787276
Saliva1	Healthy saliva	LN824142	UK	1	PMID: 25787276
HL05	HL	LN824204	UK	1	PMID: 25787276
HL08	HL	LN824225	UK	1	PMID: 25787276
HL01	HL	LN824226	UK	1	PMID: 25787276
HL09	HL	LN827522	UK	1	PMID: 25787276
HL11	HL	LN827524	UK	1	PMID: 25787276
HL02	HL	LN827546	UK	1	PMID: 25787276
HL04	HL	LN827564	UK	1	PMID: 25787276
AFB1	LCL	LN827554	Unknown	2	PMID: 25787276
K4123-Mi	Healthy donor	KC440851	USA	1	PMID: 25787276
K4413-Mi	Healthy donor	KC440852	USA	1	PMID: 25787276
X50-7	LCL	LN827555	USA	1	PMID: 25787276
LCL B95- (del EBER2) reseq	LCL	LN827739	USA	1	PMID: 25787276
pLCL-TRL1-post	sLCL_PTLD	LN824206	USA	1	PMID: 25787276
pLCL-TRL1-pre	sLCL_PTLD	LN824207	USA	1	PMID: 25787276
pLCL-TRL595	sLCL_PTLD	LN827559	USA	1	PMID: 25787276

Table: 4 List of EBV genomes collected and used for antigenic variation analysis

10.2 Supplementary data of PLONE ONE research article

“Genetic factors affecting EBV copy number in lymphoblastoid cell lines derived from the 1000 Genome Project samples”

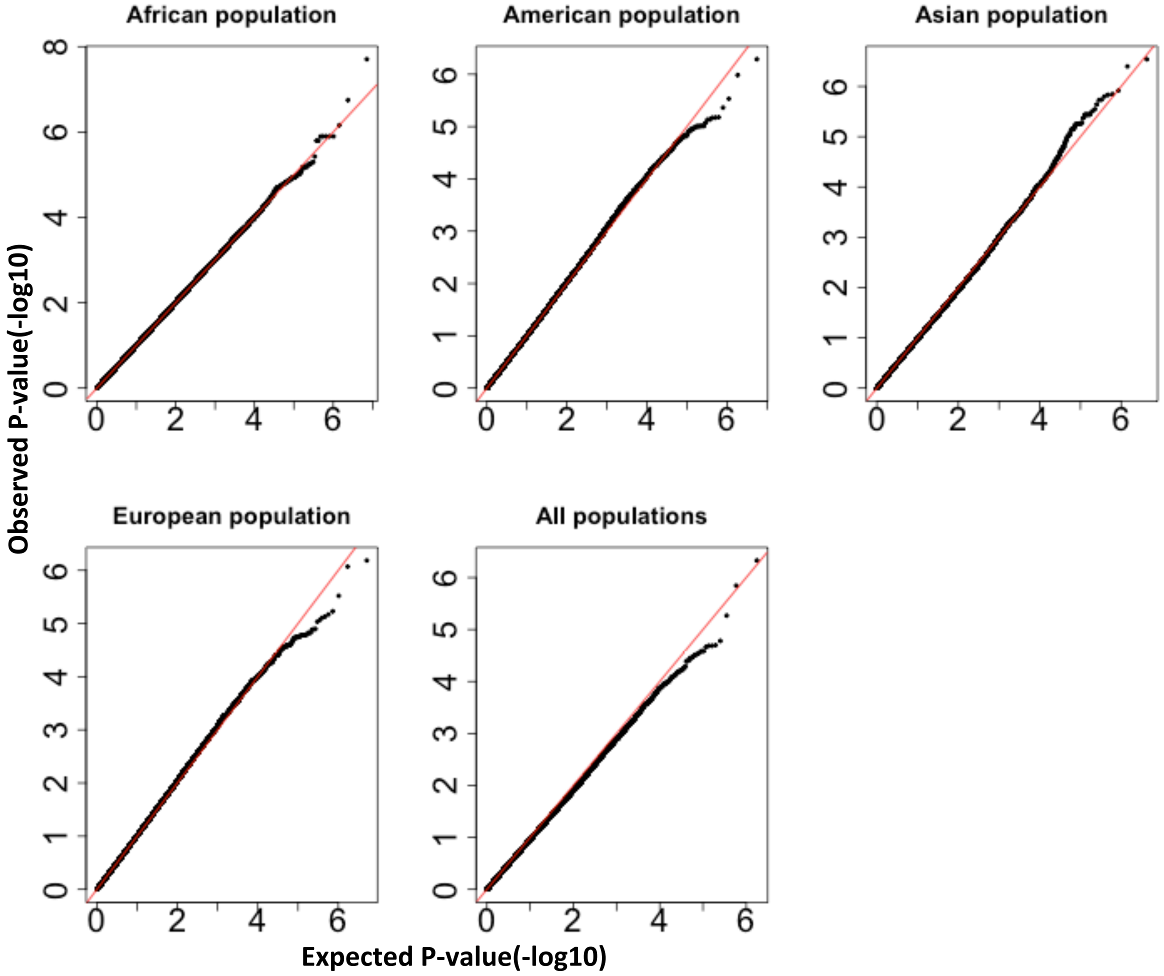


Figure S1 qq plot showing GWAS association test genome-wide p-values distribution in All Populations, Asian, African, American and European population subsets

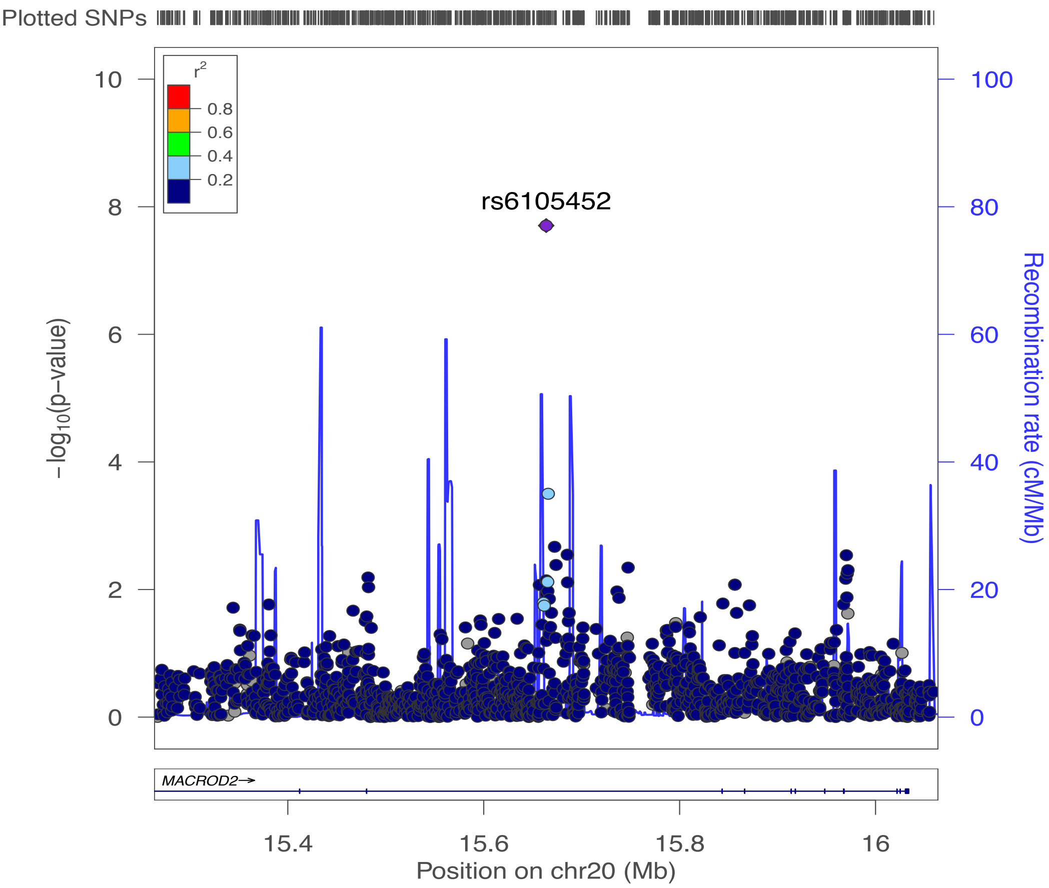
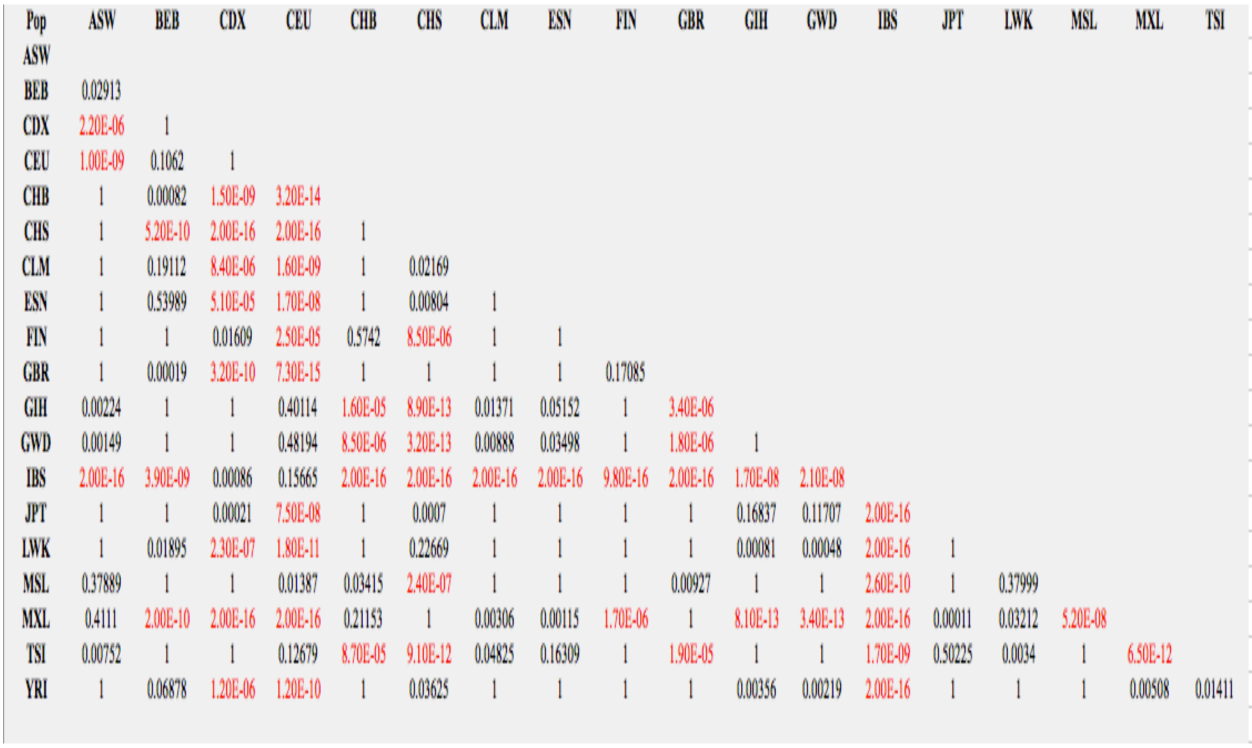


Figure: S2 Regional association plot produced by Locuszoom tool showing GWAS top SNP rs105452 from African population subset in purple color and SNPs in the surrounding region colored depending on their degree of correlation (r2) with rs105452. Lower panel contains gene within this area. Solid blue lines representing recombination rates



Pop	ASW	BEB	CDX	CEU	CHB	CHS	CLM	ESN	FIN	GBR	GIH	GWD	IBS	JPT	LWK	MSL	MXL	TSI
ASW																		
BEB	0.02913																	
CDX	2.20E-06	1																
CEU	1.00E-09	0.1062	1															
CHB	1	0.00082	1.50E-09	3.20E-14														
CHS	1	5.20E-10	2.00E-16	2.00E-16	1													
CLM	1	0.19112	8.40E-06	1.60E-09	1	0.02169												
ESN	1	0.53989	5.10E-05	1.70E-08	1	0.00804	1											
FIN	1	1	0.01609	2.50E-05	0.5742	8.50E-06	1	1										
GBR	1	0.00019	3.20E-10	7.30E-15	1	1	1	1	0.17085									
GIH	0.00224	1	1	0.40114	1.60E-05	8.90E-13	0.01371	0.05152	1	3.40E-06								
GWD	0.00149	1	1	0.48194	8.50E-06	3.20E-13	0.00888	0.03498	1	1.80E-06	1							
IBS	2.00E-16	3.90E-09	0.00086	0.15665	2.00E-16	2.00E-16	2.00E-16	2.00E-16	9.80E-16	2.00E-16	1.70E-08	2.10E-08						
JPT	1	1	0.00021	7.50E-08	1	0.0007	1	1	1	1	0.16837	0.11707	2.00E-16					
LWK	1	0.01895	2.30E-07	1.80E-11	1	0.22669	1	1	1	1	0.00081	0.00048	2.00E-16	1				
MSL	0.37889	1	1	0.01387	0.03415	2.40E-07	1	1	1	0.00927	1	1	2.60E-10	1	0.37999			
MXL	0.4111	2.00E-10	2.00E-16	2.00E-16	0.21153	1	0.00306	0.00115	1.70E-06	1	8.10E-13	3.40E-13	2.00E-16	0.00011	0.03212	5.20E-08		
TSI	0.00752	1	1	0.12679	8.70E-05	9.10E-12	0.04825	0.16309	1	1.90E-05	1	1	1.70E-09	0.50225	0.0034	1	6.50E-12	
YRI	1	0.06878	1.20E-06	1.20E-10	1	0.03625	1	1	1	1	0.00356	0.00219	2.00E-16	1	1	1	0.00508	0.01411

Figure: S3 ANOVA test output showing significant differences at the level of populations and continents

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Variance explained
Corrected Model	15,446[a]	41	.377	18.715	.000	.901	
Intercept	58.010	1	58.010	2881.691	.000	.972	
LCL	12.255	6	2.043	101.465	.000	.879	16%
Passage (LCL)	3.134	35	.090	4.448	.000	.650	4%
Error	1.691	84	.020				
Total	75.450	126					
Corrected Total	17.137	125					

a. R Squared = .901 (Adjusted R Squared = .853)

Table: S1 EBV viral load stability: ANOVA

		Asian Population			
Chr	**Gene**	**Start**	**Stop**	**Top SNP**	**Top SNP (p-value)**
6	KHDRBS2	62389864	62996100	rs855408	9.23E-06
6	PACRG	163148163	163736524	rs73784520	1.01E-05
20	DNMT3B	31350190	31397162	rs6057648	1.24E-05
20	LOC63930	61640734	61716423	rs6089823	1.97E-05
3	ZNF385D	21462489	21792816	rs2878599	2.09E-05
2	MATN3	20191812	20212455	rs6734005	2.42E-05
2	LOC101928222	20189964	20204567	rs6734005	2.42E-05
12	NUAK1	106457124	106533811	rs3782691	2.61E-05
5	HMP19	173472606	173536182	rs75695796	2.74E-05
12	CNTN1	41086243	41466213	rs12296676	2.81E-05
		European Population			
10	NRG3	83635069	84746935	rs594418	6.54E-07
5	ARSI	149675908	149682525	rs77943970	7.46E-06
4	KLF3	38665789	38703129	rs73232890	1.27E-05
17	SEC14L1	75084724	75213181	rs1254790	1.51E-05
7	GRM8	126078651	126892428	rs1361964	1.64E-05
3	ROBO2	77089293	77699114	rs17823689	2.01E-05
3	PLCL2	16926451	17132098	rs9844888	2.30E-05
13	LINC00540	22784423	22850659	rs76465489	2.55E-05
4	KLF3-AS1	38614321	38666249	rs7654470	2.66E-05
15	SPATA8-AS1	97315234	97326567	rs1393884	2.75E-05
		American Pop subset			
7	LOC101927914	157258924	157292410	rs6459788	6.64E-06
6	FAM184A	119280993	119470358	rs6569038	6.74E-06
8	CPQ	97657454	98155731	rs56064043	6.82E-06
16	WWOX	78133309	79246564	rs12596233	7.37E-06
7	TNS3	47314751	47621742	rs334524	1.08E-05
5	PDE4D	58264865	59783925	rs16889892	1.43E-05
2	TMEM194B	191371618	191399468	rs10171376	1.51E-05
2	MFSD6	191273080	191367041	rs7594994	1.51E-05
5	SPOCK1	136310986	136835018	rs11741632	1.76E-05
11	CAPN5	76777991	76837198	rs10899351	2.11E-05
		African Population			
20	MACROD2	13976145	16033841	rs6105452	1.97E-08
1	SLC35F3	234040678	234460262	rs12759054	1.28E-06
7	LRRC61	150020295	150035245	rs4395803	3.75E-06
11	CADM1	115044344	115375241	rs220839	5.25E-06
8	DLC1_2	13162115	13372429	rs116179838	6.60E-06

8	DLC1_1	12940871	13372429	rs116179838	6.60E-06
11	SLC6A5	20620945	20676610	rs77394600	7.99E-06
10	FGFR2	123237843	123357972	rs2981435	1.12E-05
2	ABCA12	215796265	216003151	rs66780625	1.16E-05
7	ZBED6CL	150026937	150029811	rs3800781	1.17E-05

Table S2 VEGAS2 output showing lists of genes ranked by p-value

SNP	Chr	Position	GWAS P-value	Gene feature	SNP annotation	Exonic variant annotation
rs200655768	chr2	179575949	2.89E-07	exonic	TTN	**synonymous**

rs12154141	chr6	63170073	4.01E-07	intergenic	KHDRBS2(dist=173973)
					LGSN(dist=815783)	

rs80274284	chr1	116579977	1.22E-06	exonic	SLC22A15	**nonsynonymous**

rs5861895	chr4	130265383	1.43E-06	intergenic	C4orf33(dist=231540)
					LOC101927282(dist=379943)	

rs2153486	chr14	88203392	1.47E-06	intergenic	LOC283585(dist=814293 ,GALC(dist=195966)	

rs5861894	chr4	130265109	1.59E-06	intergenic	C4orf33(dist=231266)
					LOC101927282(dist=380217)	

rs201255786	chr10	55566507	1.84E-06	exonic	PCDH15	**unknown**
rs11324540	chr6	62712094	1.88E-06	intronic	KHDRBS2	
rs201761909	chr19	42224910	2.30E-06	exonic	CEACAM5	**nonsynonymous**

rs201062520	chr4	130232478	2.86E-06	intergenic	C4orf33(dist=198635)
					LOC101927282(dist=412848)	

rs62025977	chr15	79650517	3.04E-06	intronic	TMED3	
rs184202621	chr9	115166387	3.25E-06	exonic	HSDL2	**nonsynonymous**

rs10498820	chr6	63115973	3.64E-06	intergenic	KHDRBS2(dist=119873)
					LGSN(dist=869883)	

rs12195364	chr6	63120399	3.64E-06	intergenic	KHDRBS2(dist=124299)
					LGSN(dist=865457)	

rs12525078	chr6	63117376	3.64E-06	intergenic	KHDRBS2(dist=121276)
					LGSN(dist=868480)	

rs16884511	chr6	63121450	3.64E-06	intergenic	KHDRBS2(dist=125350)
					LGSN(dist=864406)	

rs2842785	chr6	63092226	3.64E-06	intergenic	KHDRBS2(dist=96126)
					LGSN(dist=893630)	

rs72516830	chr4	130244072	4.13E-06	intergenic	C4orf33(dist=210229)
					LOC101927282(dist=401254)	

rs72876131	chr6	63182559	4.30E-06	intergenic	KHDRBS2(dist=186459)
					LGSN(dist=803297)	

rs6820523	chr4	130277165	5.34E-06	intergenic	C4orf33(dist=243322)
					LOC101927282(dist=368161)	

rs6820852	chr4	130277236	5.34E-06	intergenic	C4orf33(dist=243393)
					LOC101927282(dist=368090)	

rs2492797	chr6	63104576	5.55E-06	intergenic	KHDRBS2(dist=108476)
					LGSN(dist=881280)	

rs2639422	chr6	63099995	5.55E-06	intergenic	KHDRBS2(dist=103895)
					LGSN(dist=885861)	

rs2639423	chr6	63099760	5.55E-06	intergenic	KHDRBS2(dist=103660)
					LGSN(dist=886096)	

rs2639424	chr6	63099399	5.55E-06	intergenic	KHDRBS2(dist=103299)
					LGSN(dist=886457)	

rs2639430	chr6	63096304	5.55E-06	intergenic	KHDRBS2(dist=100204)
					LGSN(dist=889552)	

rs2842782	chr6	63095298	5.55E-06	intergenic	KHDRBS2(dist=99198)
					LGSN(dist=890558)	

rs4246722	chr4	130241135	5.77E-06	intergenic	C4orf33(dist=207292)
					LOC101927282(dist=404191)	

rs4975210	chr4	130241377	5.77E-06	intergenic	C4orf33(dist=207534), LOC101927282(dist=403949)	

rs4130024	chr4	130243112	6.23E-06	intergenic	C4orf33(dist=209269)
					LOC101927282(dist=402214)	

rs2842814	chr6	63102236	6.81E-06	intergenic	KHDRBS2(dist=106136)
					LGSN(dist=883620)	

rs2639383	chr6	63087238	6.95E-06	intergenic	KHDRBS2(dist=91138)
					LGSN(dist=898618)	

rs4328927	chr4	130244334	7.09E-06	intergenic	C4orf33(dist=210491)
					LOC101927282(dist=400992)	

rs4522874	chr4	130244543	7.09E-06	intergenic	C4orf33(dist=210700)
					LOC101927282(dist=400783)	

rs6534723	chr4	130244708	7.09E-06	intergenic	C4orf33(dist=210865)
					LOC101927282(dist=400618)	

rs6534724	chr4	130244776	7.09E-06	intergenic	C4orf33(dist=210933)
					LOC101927282(dist=400550)	

rs7175793	chr15	79641767	7.52E-06	intronic	TMED3	

rs4336241	chr4	130245490	8.37E-06	intergenic	C4orf33(dist=211647)
					LOC101927282(dist=399836)	

rs201130852	chr1	184764507	8.68E-06	exonic	FAM129A	**synonymous**
rs855408	chr6	62967717	9.23E-06	intronic	KHDRBS2	

rs5876797	chr6	63075225	9.43E-06	intergenic	KHDRBS2(dist=79125)
					LGSN(dist=910631)	

rs7656960	chr4	130253095	9.49E-06	intergenic	C4orf33(dist=219252)
					LOC101927282(dist=392231)	

rs2842803	chr6	63074376	9.55E-06	intergenic	KHDRBS2(dist=78276)
					LGSN(dist=911480)	

rs199893425	chr2	231404151	9.73E-06	intronic	SP100	
			European pop			
rs594418	chr10	84385052	6.54E-07	intronic	NRG3	
rs661469	chr10	84386093	8.51E-07	intronic	NRG3	

rs13204008	chr6	36970610	3.02E-06	intergenic	MTCH1(dist=16283)
					FGD2(dist=2813)	

rs13202913	chr6	151791737	5.89E-06	downstream	ARMT1	

rs7669967	chr4	156049070	6.75E-06	intergenic	RBM46(dist=299105)
					NPY2R(dist=80711)	

rs77943970	chr5	149677818	7.46E-06	exonic	ARSI	**synonymous**

rs367916962	chr6	151796244	7.82E-06	intergenic	ARMT1(dist=5010)
					CCDC170(dist=18931)	

rs831380	chr6	36963787	8.70E-06	intergenic	MTCH1(dist=9460)
					FGD2(dist=9636)	

rs671631	chr10	84387269	9.26E-06	intronic	NRG3	
			American pop			
rs1161098	chr12	67847460	5.16E-07	intergenic	CAND1(dist=138988), LOC100507175(dist=66402)	

rs62309385	chr4	137055898	1.03E-06	intergenic	LINC00613(dist=221063)
					PCDH18(dist=1384175)	

rs60350499	chr17	71111631	2.94E-06	intergenic	SLC39A11(dist=22778)
					SSTR2(dist=49529)	

rs2700565	chr12	67857853	4.35E-06	intergenic	CAND1(dist=149381)
					LOC100507175(dist=56009)	

rs6459788	chr7	157260190	6.64E-06	ncRNA_exonic	LOC101927914	
rs6569038	chr6	119416472	6.75E-06	intronic	FAM184A	
rs56064043	chr8	97798136	6.82E-06	intronic	CPQ	
rs71035108	chr10	67804180	7.34E-06	intronic	CTNNA3	
rs12596233	chr16	78657884	7.38E-06	intronic	WWOX	

rs9943465	chr10	132015271	8.42E-06	intergenic	GLRX3(dist=36625)
					MIR378C(dist=745580)	

rs1387611	chr5	160519563	9.76E-06	intergenic	LOC285629(dist=153930)
					GABRB2(dist=195873)	

rs3980578	chr5	160519442	9.76E-06	intergenic	LOC285629(dist=153806)	

					GABRB2(dist=195994)

rs4921374	chr5	160521734	9.76E-06	intergenic	LOC285629(dist=156101)
					GABRB2(dist=193702)

rs367058	chr3	138496810	9.89E-06	intergenic	PIK3CB(dist=18609)
					LINC01391(dist=157221)

rs388649	chr3	138496809	9.89E-06	intergenic	PIK3CB(dist=18608)
					LINC01391(dist=157222)

rs10251462	chr7	157258183	9.90E-06	upstream	LOC101927914

rs10898913	chr11	72971437	1.00E-05	intergenic	P2RY2(dist=17965)
					P2RY6(dist=4113)

African pop					
rs6105452	chr20	15663771	1.97E-08	intronic	MACROD2

rs138117677	chr16	3398904	1.79E-07	intergenic	ZNF75A(dist=30323)
					OR2C1(dist=6985)

rs6797827	chr3	109454370	6.93E-07	intergenic	MIR4445(dist=132626)
					PVRL3-AS1(dist=1309793)

rs114469326	chr4	108110889	1.27E-06	intergenic	DKK2(dist=153436)
					PAPSS1(dist=423933)

rs116671518	chr4	108110866	1.27E-06	intergenic	DKK2(dist=153413)
					PAPSS1(dist=423956)

rs12759054	chr1	234119810	1.27E-06	intronic	SLC35F3

rs74475807	chr4	108110200	1.27E-06	intergenic	DKK2(dist=152747)
					PAPSS1(dist=424622)

rs75528040	chr4	108111569	1.27E-06	intergenic	DKK2(dist=154116)
					PAPSS1(dist=423253)

rs189784920	chr4	108117116	1.60E-06	intergenic	DKK2(dist=159663), APSS1(dist=417706)

rs33941707	chr4	108117120	1.60E-06	intergenic	DKK2(dist=159665)
					PAPSS1(dist=417702)

rs4395803	chr7	150025367	3.75E-06	intronic	LRRC61

rs10099002	chr8	5162760	5.07E-06	intergenic	CSMD1(dist=310432)
					LOC100287015(dist=1098317)

rs220839	chr11	115319096	5.25E-06	intronic	CADM1
rs55992909	chr11	12476123	5.64E-06	intronic	PARVA

rs10741594	chr11	12563953	5.66E-06	intergenic	PARVA(dist=7050)
					TEAD1(dist=132016)

rs11764936	chr7	150025915	5.88E-06	intronic	LRRC61

rs10481377	chr8	5162212	6.49E-06	intergenic	CSMD1(dist=309884)
					LOC100287015(dist=1098865)

rs10780171	chr8	5161465	6.49E-06	intergenic	CSMD1(dist=309137)
					LOC100287015(dist=1099612)

rs11136856	chr8	5161506	6.49E-06	intergenic	CSMD1(dist=309178)
					LOC100287015(dist=1099571)

rs55759102	chr8	5161372	6.49E-06	intergenic	CSMD1(dist=309044)
					LOC100287015(dist=1099705)

rs116179838	chr8	13272163	6.60E-06	intronic	DLC1

rs10107440	chr8	5151979	6.62E-06	intergenic	CSMD1(dist=299651)
					LOC100287015(dist=1109098)

rs77394600	chr11	20675142	7.99E-06	intronic	SLC6A5

rs1420840	chr8	5152925	8.27E-06	intergenic	CSMD1(dist=300597)
					LOC100287015(dist=1108152)

rs73501482	chr9	103598352	9.40E-06	intergenic	MURC(dist=247683)
					PLPPR1(dist=192679)

rs75251726	chr1	25457281	9.46E-06	intergenic	MIR4425(dist=107204)
					SYF2(dist=91486)

rs4552921	chr8	5156316	9.79E-06	intergenic	CSMD1(dist=303988)
					LOC100287015(dist=1104761)

All pop						

rs314879	chr13	23309382	4.70E-07	intergenic	LINC00540(dist=458723)
					BASP1P1(dist=161787)	

rs200699422	chr4	174352039	5.40E-06	intergenic	SCRG1(dist=31422)
					HAND2(dist=95613)	

rs1062630	chr6	31138107	1.43E-06	exonic	POU5F1	**synonymous**

Table: S3 Continent wise list of GWAS top SNPs filtered by 10^{-6} and 10^{-7} p-value