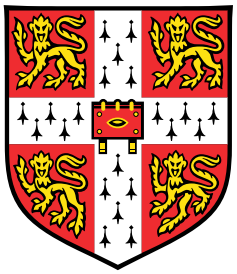
Immune-driven positive and balancing selection in human populations



Sarah Wallin Kaewert

Department of Archaeology and Anthropology
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Jesus College

September 2018

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 80,000 words including appendices, bibliography, footnotes, tables and equations.

Sarah Wallin Kaewert
September 2018

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Toomas Kivisild for his support and guidance throughout this project. Fond thanks to the members of the Kivisild lab for all their help and friendship, and to the Biological Anthropology graduate students for many happy hours spent in LCHES over the years. Thank you also to Dr. Charlotte Inchley, Dr. Georgi Hudjashov and Dr. Tiago Antão for running the Tajima's D, nSL, d_i, and iHS statistics on the Colla population data.

I would also like to express my profound gratitude to my family for their support and encouragement during my PhD. Thank you to friends near and far, old and new, for making this PhD so much fun. Special thanks to my family away from home–Stephanie Diepeveen, Ettie Unwin, Tina Andersson Tunivanua, and Jasmynne Bushrod.

Finally, I am very grateful to the Cambridge Trust for their funding of my PhD studies, to the Cambridge Philosophical Society for their research studentship, and to Jesus College and the Department of Archaeology and Anthropology for their support of my conference attendance.

Abstract

This thesis examines the evidence for positive and balancing selection on immune genes in 388 individuals grouped into thirteen geographically diverse populations. The data are high coverage whole genome sequences, many from populations that have been sparsely represented in global genetic diversity studies in the past.

Two main analyses were performed for both positive and balancing selection: enrichment tests for each population and class of immune genes, and filtering for top variants or genes driving selection signals. Four different measures of positive selection and three measures of balancing selection were used to scan the whole genome data for evidence of selection. Further filters, including functional importance predictors, were used to filter results for potential driver variants.

Positive selection results show significant enrichment for genes associated with bacteria or virus interaction, the innate immune system, and antigen processing and presentation. Results also include variants potentially driving signals of selection. One of these is a missense variant in the Northeast Siberian population in the gene *IL27*, which is involved in modulation of immune response to infection.

Balancing selection enrichment tests show that genes associated with T cell function and antigen processing and presentation are significantly enriched in every population. The HLA region features heavily in these enrichments. One top gene result is *GNLY*, a gene that produces the antimicrobial protein granulysin, in the West Siberian and Island Southeast Asian populations. Another is *PGLYRP4*, which is a top gene in seven populations and is involved in recognition and defense against Gram-negative and -positive bacteria.

In conclusion, as a general trend there is more enrichment in genes that interact with bacteria and viruses in the positive selection results, and more enrichment in genes involved with antigen processing and presentation and T cell function in the balancing selection results. The combined results show the different immune-driven selection histories of each population, as well as highlight a number of variants and genes that are potential drivers of selection and promising candidates for further study.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Introduction to the human evolutionary landscape

Fossil evidence suggests that modern humans emerged as a species not from a single place on the African continent, but rather from multiple groups of early modern humans over 300,000 years ago (Gunz et al., 2009; Hublin et al., 2017). The oldest fossil specimens of modern humans have been found in Africa, and predate other modern human fossil remains in other parts of the world by at least 50,000 years (review by Nielsen et al. (2017)). Further supporting Africa as the original location in which modern humans arose as a species, modern African populations have the highest overall levels of genetic heterozygosity of any human population (Tishkoff and Kidd, 2004) (review by Nielsen et al. (2017)). This supports the idea of genetic bottlenecks as people migrated out of Africa, reflected by lower levels of genetic heterozygosity in the rest of the world compared to Africa. African populations also show, compared to populations from other parts of the world, the lowest levels of linkage disequilibrium (LD) and the highest effective population sizes (Tishkoff and Kidd, 2004). These factors all point to a recent African origin for all currently living human populations.

At some point, some humans began to move out of Africa and into the rest of the world. The main out of Africa migration took place around 100,000 years ago, though the exact timing continues to be under debate (Fan et al., 2016)(review by Nielsen et al. (2017). Based on the dataset used in this very thesis, it has been suggested that there was a smaller, initial out of Africa wave that left almost no genetic trace except for in parts of Island Southeast Asia (Pagani et al., 2016). The migration routes followed by expanding populations are also under continuing debate. Genetically, this initial migration out of Africa and the following expansion left its trace in the human genome in the form of a bottleneck and drop in effective population size and genetic diversity in all non-African populations, as well as genetic drift between African and non-African populations (Tishkoff and Kidd, 2004).

Further genetic differences accumulated between groups as populations adapted to their new environments. As humans spread out into new geographic regions, they encountered vastly different environments to those in which the species had evolved and had to adapt to them in order to survive. These new environmental factors include sun exposure, altitude, temperature, different types of food available, and new pathogens. Accordingly, adaptations in genes associated with temperature regulation, skin pigmentation, hemoglobin levels, innate immune response, fatty acid metabolism, height, and lactose tolerance are just some examples of ways in which humans have adapted to their new environments (Fan et al., 2016)(review by Nielsen et al. (2017)). Some of these adaptations are represented in Figure 1.1.

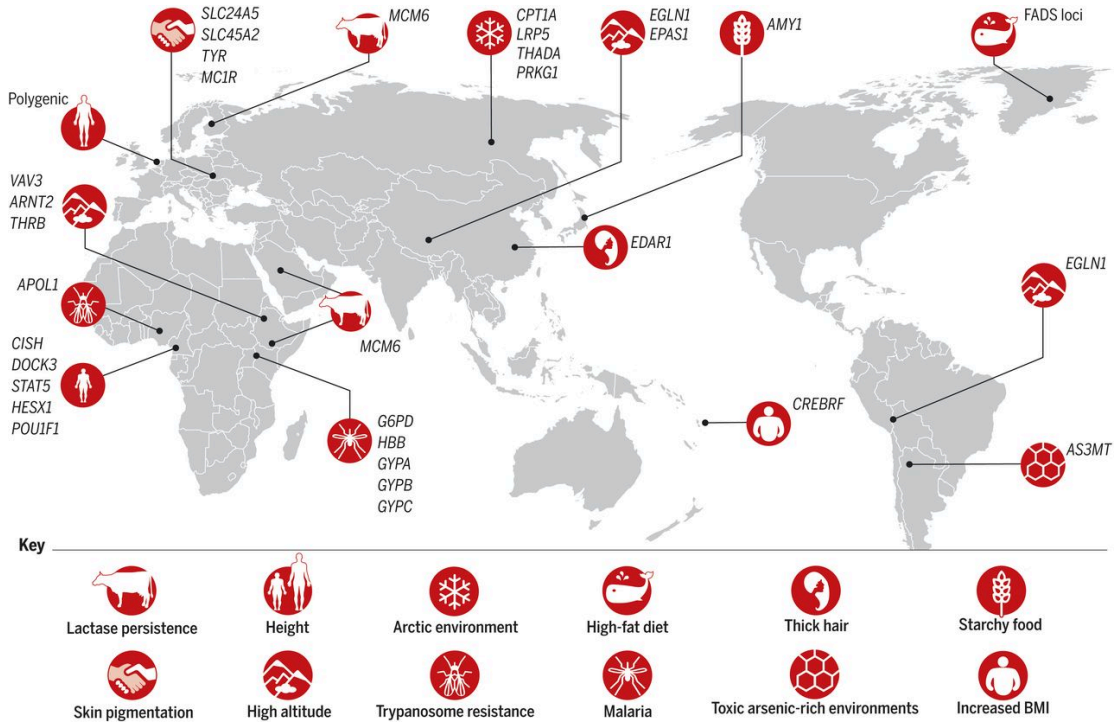


Fig. 1.1 A small subset of local human adaptations to the environment, from Fan et al. (2016).

In some cases, modern humans adapted to their new environments with some help from other hominins through interbreeding. Neanderthals and Denisovans are archaic hominins who split from the lineage that would become modern humans. Based on a recently-analyzed high coverage Neanderthal genome from Vindija Cave in Croatia, Neanderthals likely split with modern humans between 520 and 630 thousand years ago, and with Denisovans between 390 and 440 thousand years ago (Prüfer et al., 2017). Based on fossil evidence, Neanderthals were living in western Eurasia up to around 40,000 years ago (Higham et al., 2014), and Denisovans were living in the Altai mountains of Siberia between 30,000 and 50,000 years

ago, and were also likely present in other parts of East Asia (Reich et al., 2010). Since these archaic hominins were already living in these regions when modern humans encountered them, when interbreeding occurred, modern humans were able to benefit from some of the beneficial adaptations to the environment Neanderthals and Denisovans had accumulated.

Introgression analysis has found that Denisovans likely contributed between between 3 and 6% of modern-day Melanesian, Papuan, and Australian genomes, as well as smaller amounts to other East Asian genomes (see review by Racimo et al. (2015)). Similarly, all non-African populations contain a Neanderthal contribution of between around 1 and 3% (Prüfer et al., 2017). When we look at the patterns of which genes appear to be introgressed from archaic hominins, it appears that retention of archaic hominin genes has not been uniform across the genome. This suggests that beneficial regions have been positively selected to stay in the genome and deleterious regions have been selected out of the genome. Alleles contributed by Denisovans are found in genes associated with the HLA region, innate immune response, and altitude adaptation. Alleles contributed by Neanderthals are also associated with HLA regions, innate immune response, autoiummune disease, pigmentation, skin and hair morphology, and metabolism (see review by Racimo et al. (2015)). Regions of the modern human genome related to reproduction (specifically, genes on the X chromosome and genes only highly expressed in the testes) are significantly lacking in Neanderthal alleles, suggesting that Neanderthal alleles had a deleterious effect on human fertility and were selected against (Sankararaman et al., 2014). This evidence shows that not only did humans adapt to their new environments with contributions from archaic hominins, but that the past population history of interbreeding and introgression was a complex one.

Moving into new regions and encountering new environmental challenges is one scenario that introduces genetic adaptations. Another is cultural transition. With the adoption of agricultural and pastoralist lifestyles, some populations experienced changes in diet and new immune challenges due to larger population sizes. One major group of adaptations hypothesized to come from this cultural transition is dietary adaptations. For example, variants that cause lactase persistence in European and African populations correspond in age to times at which pastoralism and animal domestication spread into those populations (Tishkoff et al., 2006). There are several hypotheses as to why lactase persistence is advantageous. Milk is a source of calories, especially in times of famine (Gerbault et al., 2011). It has also been proposed that lactase persistence offers the benefit of better calcium absorption in places with little sunlight and/or insufficient vitamin D (Flatz and Rotthauwe, 1973). In addition, it has been suggested that the copy number of the amylase gene *AMY1* has been under positive selection in populations with high starch diets (Perry et al., 2007). Since high starch diets are associated with agricultural societies, it has been suggested that this selection is associated

with the transition to agriculture during the Neolithic transition (McClure, 2015; Perry et al., 2007). It has also been proposed that positive selection on an allele in the *HFE* gene that causes potentially pathogenic increased absorption of iron occurred due to increased levels of anemia caused by agriculturalist reliance on starch and dairy associated with the Neolithic transition (McCullough et al., 2015).

Another major group of adaptations to the Neolithic transition are those related to immune function. Those will be addressed in greater detail in Section 1.2.

1.2 Infectious disease as a driver of human adaptation

Infectious disease is widely considered to be one of the most important drivers of human evolution (Fumagalli et al., 2011)(review by Nielsen et al. (2017)). Genetic adaptations that allow individuals or populations to survive and pass on their genes to the next generation are preferentially retained in the genome, and the stronger the selective pressure, the more severe the selection event. Few selective pressures have exerted such powerful pressure as infectious disease. According to one estimate, infectious disease has been responsible for more deaths than all natural disasters, wars, and noninfectious diseases combined throughout human history (Inhorn and Brown, 1990). Before the discovery of vaccines and antibiotics, around half of all children were killed by infectious disease before the age of 15 (Siddle and Quintana-Murci, 2014), and from the Paleolithic to the Industrial Revolution, the average life expectancy was only 25 years (Casanova and Abel, 2005). Looking specifically at the two populations from the U.K., Figure 1.2 shows the stark difference between life expectancy before and after modern antibiotics, vaccines, and sanitation practices (Casanova and Abel, 2005). Because of its high death toll over thousands of years, and the continual evolution of pathogenic microorganisms, infectious disease has the potential to be one of the most important drivers of natural selection in human history. Indeed, many studies have found immune-related genes to be among top targets of selection in human populations (see Sections 3.1 and 4.1). These are examples of human populations adapting to pathogen pressure over time, and evolving to be better suited to their environment.

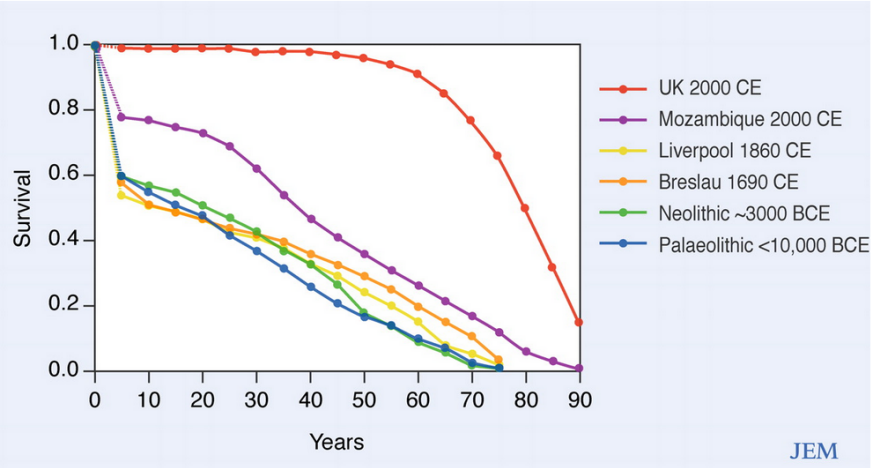


Fig. 1.2 Human life expectancy of different populations at different times during history, from Casanova and Abel (2005).

Rook's breakdown of human history into three major epidemiological phases provides a useful way to look at different pathogen environments humans would have been exposed to at different times in history, and shows the backdrop against which adaptation would have taken place. These three major groups of lifestyles, with corresponding sets of pathogens, are Paleolithic, Neolithic to pre-industrial (19th century), and modern day. These are presented in Figure 1.3.

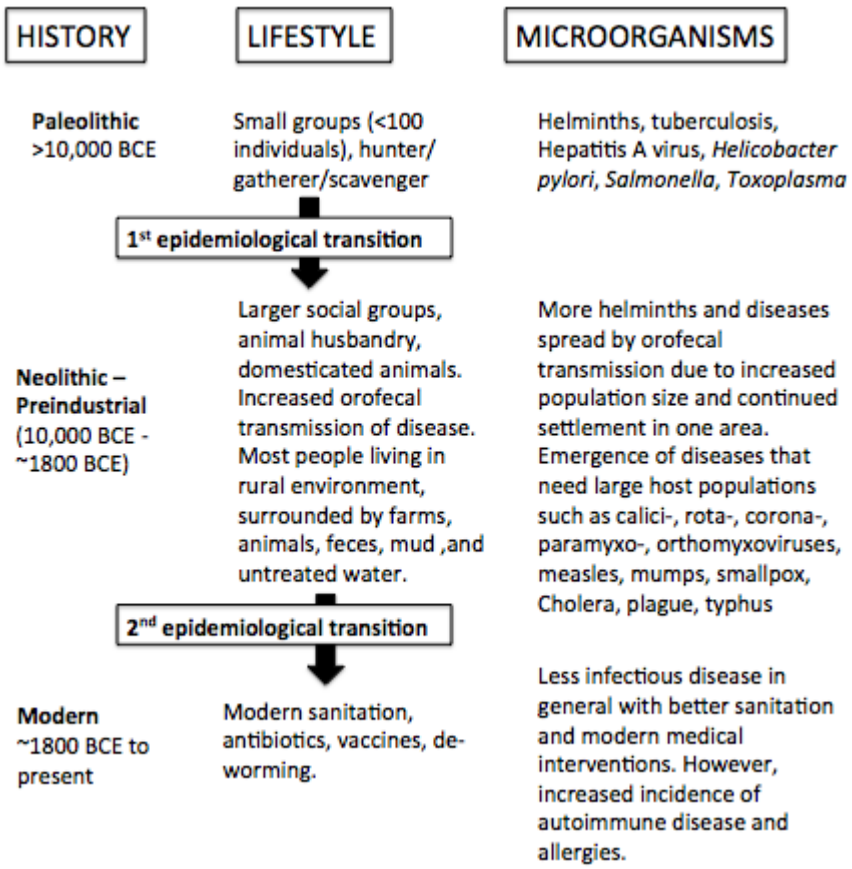


Fig. 1.3 Pathogen landscapes from different phases of human history, modified from Rook (2011).

Before the Neolithic transition, humans lived at low population density and carried diseases that were adapted to those demographic conditions. Specifically, they were adapted to live in the host for long periods of time and not be especially pathogenic, so that they could be transmitted over many years, often involving periods of latency and activation. Pathogens causing high mortality rates were selected against, since they were not able to be passed into large populations and eventually died out with the host populations (Blaser and Kirschner, 2007; Comas et al., 2013). It has been considered likely that various types of parasites probably infected paleolithic humans (McNeill, 1976), and that species of bacteria carried by humans during that time are likely to be those that were also encountered by early hominins in hunter-gatherer groups, as well as those that are encountered in modern hunter-gatherer groups. These include *salmonella*, *typhi*, and *staphylococci* bacteria (see review by Armelagos et al. (2005)), as well as tuberculosis (Comas et al., 2013; Rook, 2011). Similarly, it has been considered likely that paleolithic humans carried similar viruses to those carried by our hominin ancestors and primate relatives. In general, these viruses were able to thrive

in low density populations, because they often cause non-acute, chronic infections. Sporadic bursts of zoonotic disease are thought to have occurred from time to time throughout all of human evolution, as hominins interacted with wildlife around them, such as birds, insects, bats, and other primates (Van Blerkom, 2003). Arboviruses also likely infected humans from time to time due to contact with disease-carrying mosquitoes (McNeill, 1976). Some authors emphasize the importance of interactions between African great apes and ancestral hominin groups in Africa, as well as interactions between humans and Neanderthals and Denisovans outside of Africa, for the transmission of diseases such as various Papilloma- and Herpesviruses (see review by Pimenoff et al. (2018)).

During the Neolithic transition, population density grew as farming technology allowed for larger population sizes. Humans came into prolonged contact with domesticated animals and their associated pathogens, which led to zoonotic disease. According to the views of some authors, these conditions led to selection for pathogens that thrived in high population densities and spread much more quickly and caused more host damage (Blaser and Kirschner, 2007; Comas et al., 2013). Habitat area of the mosquito *Anopheles gambiae* is thought to have increased with clearing of forests for agricultural reasons, leading it to grow to rely on human blood as its principal source of food and a subsequent increase in the incidence of malaria and other mosquito-borne diseases (McNeill, 1976). More prevalent bacterial infections included cholera from contaminated water, typhus from lice, and plague from fleas (Armelagos et al., 2005). Fast-evolving RNA viruses became endemic in human populations from contact with livestock and domesticated animals, and the rodents that are associated with storage of farming surplus. Such viruses are also more likely to jump between species. These viruses include calici- and rotaviruses, which cause diarrhea, coronaviruses, which cause colds, and para- and orthomyxoviruses that cause the flu and other respiratory diseases (Van Blerkom, 2003). Sumeria has been suggested as the oldest city population large enough to host measles and other childhood diseases characteristic of the Neolithic transition. Thus, some earlier authors have considered 3000 BCE as the earliest date for "diseases of civilization" (McNeill, 1976).

A table illustrating the different types of infection and the periods of human history associated with them can be seen in Figure 1.4.

Time-line of human history(yr BP)	Effectivepopulation size	Major source of microbial transmission	Nature of immunity	Example	Persistence
Most ancient (>50,000)	Isolated hunter-gatherers (<100)	Maternal/intrafamilial	Ineffective	*Bacteroides* species, *H. pylori*	Active
Intermediate(10,000–50,000)	Communicating hunter-gatherer groups(<100 to 10,000)	Long-term carriers	Containment but not elimination	*M. tuberculosis*, *S. typhi*, varicella-zoster virus	Latency
Recent (<10,000)	Large societies (>500,000)	Acutely ill persons	Life long	Measles	No
Very recent (<200)	>10 million	Acute infection	Serotype-specific	Pandemic influenza	No

Fig. 1.4 Bacterial and viral persistence dynamics, from Blaser and Kirschner (2007).

The distribution of different pathogens through different ecological landscapes plays a key role in determining which populations contend with and adapt to specific pathogens. Pathogen richness has been shown to be associated with latitude, with higher levels of richness at lower levels of latitude (Cashdan, 2014; Guernier et al., 2004). A plot of number of pathogens vs. latitude can be seen in Figure 1.5. Cashdan (2014) suggest that outside of the tropics, pathogen species richness is especially affected by temperature and the presence or absence of frost. Conversely, she suggests that in the tropics, it is mostly affected by the amount of precipitation in a year. Mosquito-borne infections, for example, are highly associated with annual rainfall. High levels of dryness on a seasonal basis is also associated with increased levels of typhus, schistosomiasis, and leishmaniasis, potentially due to host and vector behavior (Cashdan, 2014). Because selection signals that are likely to be detected in this thesis are most probably thousands or tens of thousands of years old, it would require some climate reconstruction to hypothesize which regions experienced frost or extreme seasonal dryness at a given point in history. However, latitude remains constant, and the association of pathogen richness with latitude has likely remained strong throughout human history. Therefore, it will be interesting to see whether populations at lower latitudes exhibit more signals of selection in immune-related genes than populations at higher latitudes.

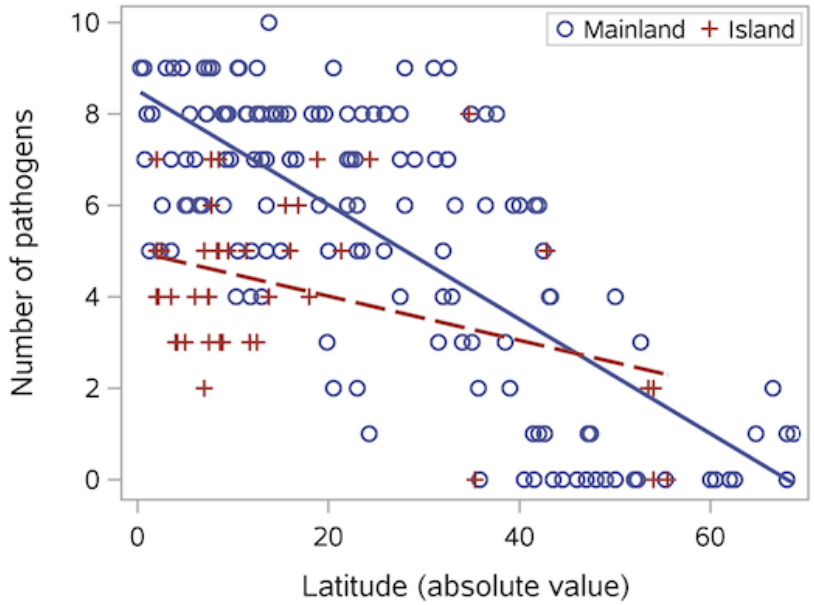


Fig. 1.5 Correlation between latitude and pathogen density, from Cashdan (2014).

Since pathogen diversity varies geographically, it follows that different populations adapt over time in response to local infectious diseases. Such adaptations, and the corresponding lack of adaptations to unfamiliar diseases in unexposed populations, have shaped not only the genetic landscape of the local population but human history as well with devastating consequences. This disease pressure came most catastrophically through outbreaks of "diseases of civilization", such as the bubonic plague, measles, smallpox, influenza, typhoid, and dysentery (McNeill, 1976) (see also review by Archer (2016)).

Human diseases can be most severe in previously unexposed human populations soon after exposure (review by Riley (2010)). For example, Europe and the Middle East were affected especially devastatingly by the plague in the 14th century, as was the Roman empire by its exposure to smallpox in the second century (review by Archer (2016)). However, over time these populations acquired some population immunity to such diseases. By around the year 1200, most of the Old World had been exposed to the various diseases listed above through overland and sea trade between Europe, the Middle East, China, and India and had adapted to them enough to be experiencing steady population increase (McNeill, 1976). When Old World populations came into contact with populations in which these diseases were not endemic, this mismatch in immunity often led to devastating population loss in the unexposed population.

One of the most illustrative examples of the drastic consequences of this differential immunity is the contribution of disease, carried by European colonists, to the decimation of

Native American populations. Using smallpox as an example of one of these diseases, the most important factor thought especially to have led to such catastrophically high mortality rates in Native Americans as compared to those of European colonists was the complete lack of immunity from prior exposure to any Poxvirus that could grant immunity to smallpox virus. In Europe, where smallpox was endemic and regular outbreaks occurred, a portion of the population at any time would have lived through a prior infection and be protected by acquired immunity. However, in the Americas, where it was not endemic, sporadic infections had much higher mortality rates because likely none of the population had been previously exposed (review by Riley (2010)). While there is debate as to exactly which diseases were introduced, and the precise scale of population decline, it is widely accepted that Native American populations were severely affected by post-contact infectious disease introduced by European colonists (Ramenofsky, 2003). One paper by Lindo et al. estimates a 57% reduction in effective population size between a pre-contact Pacific Northwest population and a related modern population of Tsimshian individuals. This same study finds evidence of positive selection on the immune gene *HLA-DQA1* pre-contact and negative selection on the same gene post-contact, again suggesting adaptation to a changed disease environment (Lindo et al., 2016).

Widespread mortality of indigenous peoples (and in some cases, replacement with immunologically advantaged European populations) when met with epidemic disease took place not only in the Americas, but also in parts of Siberia, Africa, the Pacific Islands, and Australia (McNeill, 1976), and thus has likely left a significant genetic signature in the genomes of surviving descendants. Interestingly, several of the above populations belong to the group of populations included in this thesis that have not often been included in previous genetic analyses. Since infectious disease has been such an important driver of adaptation in different human populations, it is likely that such adaptations will be visible in the genome.

Though tempting, it is often difficult, if not impossible, to predict which specific disease was responsible for a historical epidemic or signature of selection based on historical record. It is difficult to match descriptions of a past epidemic to a certain disease for several reasons. Historical medical terminology is different from modern medical terminology, and symptoms from a disease well established in its host population are different from symptoms of that same disease in the process of adapting to a previously unexposed population (McNeill, 1976). Because specific resistance to particular pathogens is acquired during the lifetime of an individual and not passed to the next generation through the germline, protective adaptations against infectious disease are most likely to be found in the first two layers of defense of the immune system: physical barriers and the innate immune system. Both of these two categories of defense rely on recognizing general attributes of infectious agents.

Because of this, it may be possible to ascertain the overall category of pathogen responsible for a signal of selection–for example, flagellated bacteria are responsible for selection at the *TLR5* locus (Grossman et al., 2013). However, in other cases it is difficult to know which specific organism is responsible. For example, there is continuing disagreement regarding the specific disease (bubonic plague or smallpox) responsible for selection on the *CCR5* gene (see Section 1.5) (Galvani and Slatkin, 2003).

1.3 A brief introduction to the human immune response to infection

The immune system exists to protect the human body from pathogenic organisms that exist in the environment. Because of the need to differentiate between host cells and invading organisms, as well as beneficial microorganisms and pathogenic ones, the immune response is very complex. This complexity is also necessary to respond to the many strategies employed by pathogens to cause disease. Most generally, the body detects and responds to pathogenic intruders by recognizing structural differences between pathogens and host cells. This is done in two complementary ways by the innate and adaptive branches of the immune system (Chaplin, 2010).

1.3.1 Innate immune response

The innate branch of the immune system is the first line of response to an infection. The overall purpose of this branch is to recognize general microbial patterns common to multiple pathogens to distinguish invading organisms from host cells. Elements of the innate immune system are all of those that are fully coded for in an individual's genome at birth (Chaplin, 2010).

The two main branches of the innate immune system are physical barriers and the cell-based innate immune response. Physical barriers include skin and skin secretions which keep microbes out of the body and keep the skin at a pH that is unfavorable to bacterial growth; mucus membranes and mucosal secretions (tears, mucus, and saliva) which again prevent microbial entry and contain defensins (antimicrobial peptides that act against bacteria, viruses, and fungi); stomach acid that kills microbes that have entered the digestive tract; and ciliated epithelia, which move mucus and microbes out of the respiratory tract (Murray et al., 2009; Reece et al., 2014). These barriers all keep infectious agents out of the body and create environments unfavorable for pathogen growth or entry.

However, if an infectious agent does make its way past those barriers and into the body, it next encounters the cell-based innate immune system, which acts quickly and relatively locally to try to eliminate infection (Murray et al., 2009). There are two main types of cells that act in the innate immune response: phagocytes and natural killer cells (Reece et al., 2014). Phagocytes are cells that engulf pathogens, and include neutrophils, macrophages, and dendritic cells. Generally these cells engulf and destroy pathogens and help to stimulate the adaptive immune system (Reece et al., 2014). The other main type of cell involved in the innate immune response is the natural killer cell, which is especially important in viral infections. Natural killer cells are lymphocytes that release chemicals that kill host cells presenting viral or cancerous antigens (Reece et al., 2014).

Several vital elements of the innate immune system that appear in the results of this thesis are pathogen-associated molecular pattern recognition receptors (PRRs), interferons, and the complement system (Murray et al., 2009; Reece et al., 2014). PRRs, such as Toll-like receptors (TLRs), recognize foreign substances such as flagellin, double-stranded RNA, and lipopolysaccharide (LPS), all of which are not found in the healthy human body and are indicative of microbial infection (Reece et al., 2014). Interferons are vital in the immune response to viral infections, and act to limit the replication of viruses in cells neighboring infected cells, and to activate macrophages (Reece et al., 2014). The complement system is a set of proteins that are activated by PAMPs and are involved in causing lysis of pathogenic cells, among other functions (Reece et al., 2014).

1.3.2 Adaptive immune response

The adaptive immune system is the next line of response to infection, which provides a specifically targeted response against any surviving infectious agents (Murray et al., 2009). The adaptive immune response relies on two main types of lymphocytes, or white blood cells. These are B cells and T cells. B cells mature in bone marrow, and T cells mature in the thymus (Reece et al., 2014).

Both B cells and T cells have antigen receptors on their cell surfaces. These receptors exhibit enormous diversity, generated by different combinations of variable segments. Upon exposure to a new pathogen, a B or T cell is activated when it binds to the pathogen antigen. Once activated, the cell will divide, creating two new types of cells: effector cells and memory cells. Effector cells for B cells are called plasma cells, which are responsible for producing antibodies specific to that pathogen. Effector cells for T cells are helper and cytotoxic T cells, which are described in more depth below. Memory cells enable a faster immune response should the the same antigen be presented again in a future infection (Reece et al., 2014).

There are two main facets of the adaptive immune response: the humoral response and the cell-mediated response. The humoral response takes place in the blood and lymph, and is centered around antibody response to infection. Antibodies in the blood or lymph will bind to pathogen antigens, which can block their normal function (for example, prevent a virus from entering a cell), or mark them for destruction by other immune cells. The cell-mediated response is based on the destruction of infected cells by cytotoxic T cells (Reece et al., 2014).

The helper T cell is at the base of each of these responses. If the helper T cell's T cell receptor binds to the MHC molecule and antigenic peptide presented by the antigen-presenting cell (a cell with an MHC class II molecule, which can be dendritic cells, macrophages, or B cells), the helper T cell becomes activated. This can lead to more activated helper T cells, activated B cells, and activation of cytotoxic T cells, which are the basis of the cell-mediated immune system. Cytotoxic T cells become activated by signals from helper T cells, and from binding to antigen-presenting cells. Cytotoxic T cells are responsible for destroying infected host cells (Reece et al., 2014).

Based on whether an individual has been exposed to a particular antigen before, they will experience either a primary or secondary immune response. The primary immune response takes place in an immunologically naive individual, who is encountering an antigen for the first time. The peak immune response in this case occurs between 10 and 17 days after exposure. The secondary response, occurring in an individual who has already encountered an antigen, peaks between 2 and 7 days after exposure. The immune response is also of greater magnitude in a secondary immune response (Reece et al., 2014). Since memory T and B cells can live for decades, exposed individuals have a clear advantage over non-exposed individuals. This stark difference in response time and magnitude underlines the advantage an immunologically experienced population would have over an immunologically naive one with regards to a particular disease and the potential influence of disease in shaping world history, as stated in Section 1.2.

Self-tolerance and the ability to avoid runaway immune and inflammatory responses are also vital requirements of the immune system. B cells and T cells that react with host tissues are removed from the lymphocyte pool (Reece et al., 2014). Regulatory T cells play an important role in negative/homeostatic regulation of immune response and suppressing autoimmunity (see review by Beissert et al. (2006).)

Adaptive mutations in any one of the branches of innate or adaptive immunity might allow an individual to better survive an infection. Consideration of the different types of cells important in the immune system went into choosing the classes of genes examined in this thesis (see Section 2.1). By studying whether certain genes or gene classes have been under

past selection, we may learn which genes are phenotypically important in an infection and have made contributions to a population's immunological adaptations.

1.4 Using statistics to look for signatures of selection in the genomic era

In order to examine more precisely the ways in which populations have adapted to pressure from infectious disease, researchers have turned to searching for genetic signals of selection. Until recently, these searches were confined to candidate gene studies (Akey, 2009; Sabeti et al., 2006). Though these studies were successful in finding several classic selection targets, they suffered from a major disadvantage in study design. For a candidate gene study, a selection hypothesis must be generated before analyses are performed. This hypothesis relies on adequate knowledge of how a given genotype corresponds to a phenotype of interest (Akey, 2009).

With the development of genotyping and sequencing technologies, and the explosion of genetic data from projects like the International HapMap Project and the 1000 Genomes Project in the last 15 years or so, it has become possible to look through the whole genome for increasingly high-resolution evidence of selection (Akey, 2009). In principle, this approach avoids the major pitfall of candidate gene studies mentioned above, since a whole genome scan makes no hypotheses about candidate genes. Additionally, demographic effects are less of a confounding factor because of the assumption that demographic effects affect the whole genome equally, and selection effects affect only certain loci (Akey, 2009). However, it is important to note the drawbacks of these types of selection scans. First, neither candidate gene studies nor classic genome-wide scans are designed to detect evidence of polygenic selection, which is an important mode of selection compared to the relatively rare "hard sweep" mode in which a single allele drives a selection event. Second, given that any selection scan gives a list of top results, it is not always obvious without further functional information which genes have actually been under selection and which are false positive results (see review by Pritchard et al. (2010)).

The type of genetic data used in selection studies has also evolved over time. With the falling cost of full genome sequencing, it has become possible to use whole genome sequences in selection studies instead of the historically more common genotype data. Using whole genome data has many advantages over using genotype data. These include the avoidance of ascertainment bias in SNP sampling and the potential to find the causative variant itself (Karlsson et al., 2014).Various statistics have been developed to enable genome-

wide scans for such evidence. These statistics have different strengths and can be used in combination to find not only a signature of selection but the likely causative SNP as well.

There are three main types of genetic selection: negative/purifying, positive, and balancing. Negative selection acts to remove deleterious mutations, and is expected to affect a large number of protein-altering mutations, since between around 38-75% of mutations causing amino acid changes are expected to be deleterious on some level (Eyre-Walker and Keightley, 1999; Kryukov et al., 2007)(also see review by Barreiro and Quintana-Murci (2010)). Negative selection is the most common type of selection occurring in the genome, since it acts to preserve the integrity of the genome. Positive selection is the process in which a beneficial allele rises in frequency within a population. It has been estimated that between 10 and 20% of amino acid changes between humans and chimpanzees have been caused by positive selection (Boyko et al., 2008). In general, positive selection can leave signals of regions of decreased diversity, longer haplotypes, or extreme differentiation. Targets of positive selection in the human genome have been found to include genes relating to skin pigmentation, dietary adaptations, olfactory receptors, brain size, and immune function (see review by Nielsen et al. (2007)). Balancing selection leads to the maintenance of diversity at a genetic location in the presence of a selective pressure that favors genetic diversity or a heterozygote advantage. In a region under balancing selection, an excess of diversity compared to neutral expectations would be observed.

Since this thesis is focused on regionally differential adaptation of multiple populations, I am most interested in searching for signals of positive and balancing selection. Adaptations in these two categories are not likely to be found in genes with a function so important to survival that any change will have severe detrimental effects. Because of this, evolutionary constraints are somewhat relaxed compared to genes under strong negative selection, and local variation can emerge and be acted upon by positive and balancing selection. Looking for negative selection in immune genes in this dataset would be an interesting opportunity for further study.

Most generally, statistics used to look for signals of selection seek regions of the genome that have different characteristics from surrounding regions, and that look different from what would be expected under neutral conditions. Because each statistic has limitations and the potential to return false negatives or false positives along with true positives or negatives, it is helpful to use several different statistics when analyzing a dataset. The different statistics used in this thesis to detect positive and balancing selection, as well as findings from previous studies on positive and balancing selection, are discussed in Chapters 3 and 4.

Each of these three types of selection can act on single genes, but another important mode of selection is polygenic selection. This is selection that acts on certain pathway and whose

effects are spread out over a number of genes, instead of the classic hard sweep example of strong selection acting on a single gene. It is thought that this mode of selection is an important complement to the hard sweep model, and that it may have acted on a significant number of traits (see reviews by Pritchard et al. (2010) and Nielsen et al. (2017)). Perhaps the most well-known example of polygenic selection is on height in Europeans (review by Nielsen et al. (2017)). However, methods to detect polygenic selection still have less power to detect sweeps than the most commonly used positive selection statistics, which in general look for evidence of hard sweeps. This is because looking for polygenic selection requires knowledge about not only the genetic components of an entire pathway, but the functional sites at each of those sites as well (review by Pritchard et al. (2010)). For this reason, the focus of this thesis is selection tests that detect hard sweeps.

1.5 Medical relevancy of selection studies

While the results of selection studies are interesting in their own right as a record of human adaptation to environmental pressures over time, they can also be medically useful. Selection studies can help distinguish regions of the genome which are phenotypically important, including regions associated with disease (Heyer and Quintana-Murci, 2009). Past adaptations to pathogen pressures can help us understand modern health conditions and why our immune systems function the way they do. They can also serve as potential inspirations for new approaches to disease treatment.

In general, it is expected that mutations which lower the fitness of individuals in a population will be removed from that population over time through negative selection. However, some deleterious immune-related mutations, such as those implicated in autoimmune disease or allergies, exist at relatively high frequencies that do not follow this expectation. One explanation for this is that these mutations were protective against infectious disease in the past. For example, a variant in the gene *SH2B3* that confers stronger protection against bacterial infection has been shown to have been under selection between 1200 and 1700 years ago. However, this variant is also associated with susceptibility to celiac disease (Zhernakova et al., 2010). The gene *HAVCR1*, which has also been under selection, is associated with both susceptibility to hepatitis A as well as allergies and asthma (Nakajima et al., 2005). The classic example of sickle cell anemia and malaria (Allison, 1954) also falls into this category. Sickle cell anemia is a disease most often found in populations with ancestry from regions where malaria is endemic. Individuals with one wild type and one sickle cell copy of the hemoglobin gene are relatively protected against malaria compared to individuals with two wild type copies of the hemoglobin gene (Ackerman et al., 2005). Despite the morbidity and

mortality caused by having two copies of the sickle cell hemoglobin gene, because of the advantage to heterozygotes the sickle cell variant is maintained at relatively high frequencies in certain populations.

Studying past adaptations to infectious disease can also yield useful insights for drug development and medical treatment of disease. Because variants under past selection necessarily are closely linked to human health, they are likely to be of medical relevance (see review by Nielsen et al. (2017)). One well-known target of positive selection in European populations is a deletion in the *CCR5* gene, which codes for one of the two receptors used by HIV to enter host cells. However, the HIV strains present in humans today only date back to about the 1930s (see review by Rambaut et al. (2004)), and signatures of genetic selection often take much longer to appear (on the order of thousands of years). Therefore, the disease causing this selection must have been something else, possibly bubonic plague or smallpox (experts still disagree) (Galvani and Slatkin, 2003). However, it was noted that individuals with two copies of this crucial deletion were in a way immune to HIV since the virus couldn't gain entry to their cells. At least two methods of HIV treatment are based on the blocking of the CCR5 receptor. One is the CCR5-inhibiting drug maraviroc, and the other is stem cell transplant from a donor with a double deletion in the *CCR5* gene (Dorr et al., 2005; Hütter et al., 2009).

TLR5 is a good example of a variant found through a selection scan that could serve as inspiration for medical treatment. Grossman et al. found that a nonsynonymous variant in *TLR5* in the Yoruba population shows signals of having been under selection (Grossman et al., 2013). In Europe and South Asia, this gene also can have relatively high incidence of the stop gain mutation TLR5-392STOP (Barreiro et al., 2009; Wlasiuk et al., 2009). Both this stop gain mutation and the Yoruban nonsynonymous variant act to decrease the the magnitude of NF-kappaB signaling in response to infection by flagellated bacteria (Barreiro et al., 2009; Grossman et al., 2013). It has been shown that decreased NF-kappaB signaling can lead to better outcomes in bacterial infections (Grossman et al., 2013; Koedel et al., 2000). In mice with knocked out *TLR5* genes, transport of *Salmonella typhimurium* from the intestines to other organs was impaired, leading to increased survival (Uematsu et al., 2006). Therefore, decreased NF-kappaB function mediated through *TLR5* variants can have a large effect on survival from bacterial infections and has likely played an important role in shaping human adaptation to those infections.

Following these examples, it is possible that future treatments may be inspired by findings of selection scans in which it is clear that an important immune response has been shaped by past selection.

1.6 Diversity in genomic studies

European populations are heavily overrepresented in genome-wide association study (GWAS) literature, and there is a lack of diversity in genomic research as whole. As of 2016, only twenty percent of participants in GWAS were non-European. Of the non-European twenty percent, Asian populations are disproportionately represented due to an increase in the number of GWAS completed in Asia in recent years. Other populations continue to be underrepresented, if represented at all (Popejoy and Fullerton, 2016).

This is a problem for many reasons. The first is that it gives the impression that research on populations of European descent is more important than research on other populations. It also means we may be missing valuable connections between genotype and disease. And because many associations may not be as meaningful or even valid at all in populations not included in the initial study, it means that underrepresented populations may receive suboptimal or even harmful treatment for a disease (Popejoy and Fullerton, 2016). A recent study on the effect of demography on genetic risk prediction found that application of GWAS results to populations other than the population on which the study was completed gives limited, biased, and unpredictable risk predictions (Martin et al., 2017).

There are some resources that do aim to represent human genetic diversity more completely. The 1000 Genomes Project is one of these resources. It includes a combination of whole-genome sequences, exome sequences, and genotype data for more than 2,000 individuals in 26 different populations (Auton et al., 2015). Recent large genomic analyses have focused more on including more genetic diversity in their studies. These include three papers from 2016 that introduced the Simons Genome Diversity Project dataset (Mallick et al., 2016), a dataset of genomes from Aboriginal Australians and Papuans (Malaspinas et al., 2016), and the Estonian Biocentre Human Genome Diversity Panel (Pagani et al., 2016).

Selection studies suffer from a similar lack of diversity to more medically-based genomics research. The majority of past selection studies have focused on three major continental populations: Europe, mainland East Asia, and Africa (and most often, specifically the Yoruba population) (Grossman et al., 2010; Lao et al., 2007; Racimo, 2015; Sabeti et al., 2007; Voight et al., 2006a). Using these three populations can give a broad picture of selection and distribution of allele frequencies between populations and potentially minimize the complications of interpreting the results from a large number of populations. However, it is not clear how well a relatively simple model of these three populations explains global genetic diversity and adaptations.

1.7 Dataset studied in this project

This thesis is based on population genetic analysis of data from the Estonian Biocentre Human Genome Diversity Panel (EGDP), which was published in 2016 by Pagani et al. (Pagani et al., 2016), and a set of 19 Colla genomes from a 2017 paper by Eichstaedt et al. (Eichstaedt et al., 2017). The Pagani et al. paper published 483 whole genome sequences, 369 of which were used for selection tests in that paper. That subset of the full genomic data was clustered using ChromoPainter and fineSTRUCTURE into populations of at least 15 individuals, which generated 12 population groups (Pagani et al., 2016). The populations represented from this set are the twelve following populations: West and Central Africa, West Asia and Armenia, Southwest Europe, Northeast Europe, Volga Uralic, South Asia, Central Siberia, West Siberia, South Siberia and Mongolia, Northeast Siberia, Mainland Southeast Asia, and Island Southeast Asia. The set of populations studied in this thesis includes those same twelve populations as well as the Colla population (from the Northwest Argentinian Highlands) mentioned above. It should be noted that some populations may contain multiple subpopulations–for example, the West and Central African population. Because this group likely contains lots of diversity and some population structure, this may have an effect on the selection results for this population.

The full set of genome sequences included in this thesis consists of 388 whole genome sequences with coverage between 30X and 80X coverage (Eichstaedt et al., 2017; Pagani et al., 2016). The dataset was filtered to contain only SNPs of high quality calls, excluding any structural variants. All the analyses in this thesis were performed on the cleaned, filtered dataset that was produced from the Pagani et al. 2016 publication. It is important to note that regions containing copy number variants (CNV) and structural variation are difficult to accurately assemble using short read sequencing technology. Therefore it is likely that the sequence data does not capture the haplotypic diversity of such regions. Because of this, despite the immunological importance of some of these regions (such as the KIR region on chromosome 19), their presence or absence in top selection results might not reflect their true biological significance. Different sequencing methods, such as long-read sequencing, are more appropriate for such regions (Roe et al., 2017).For more information on the sequencing and quality control steps of this dataset, please see the Supplementary Information of Pagani et al. 2016. The number of individuals per population can be seen in Table 1.1, along with three letter abbreviations for each population. A geographic map of the general location of each of the populations included in this thesis can be seen in Figure 1.6.

Population	Population abbreviation	Number of individuals
West and Central Africa	AFR	26
West Asia and Armenia	WAA	26
Southwest Europe	SWE	32
Northeast Europe	ENE	53
Volga Uralic	VOL	23
South Asia	SOA	28
Central Siberia	CSI	31
West Siberia	WSI	17
South Siberia and Mongolia	SSI	34
Northeast Siberia	NSI	25
Colla	COL	19
Mainland Southeast Asia	SEM	29
Island Southeast Asia	SEA	45

Table 1.1 Number of individuals per population grouping.

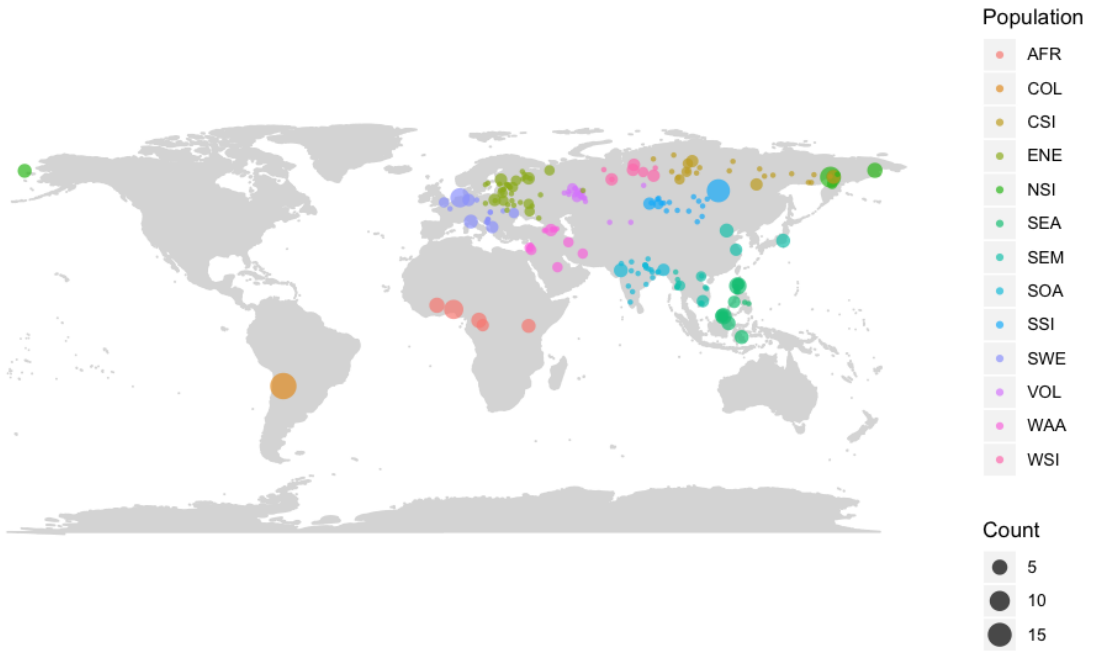


Fig. 1.6 A map of approximate geographical locations of individuals included in this thesis. Note that the numbers of points does not match the number of individuals in each population, since multiple individuals may be from the same place. Latitude and longitude coordinates taken from Pagani et al. (2016). To see exactly how many individuals are in each population, please see Table 1.1.

Several populations represented in the EGDP include regions that have rarely been included in selection studies, such as Island Southeast Asia, Siberia, and South America. Because of this, this thesis may be one of few studies on local immune-driven adaptation history in these populations. The results of this project should add greater resolution to the literature on immune-driven human adaptation and be a useful foundation for further study.

1.8 Questions to be answered in this thesis

The questions to be answered in this thesis can be separated into the following broad categories: those regarding methodology, those regarding population selection histories, and those regarding targets of selection.

Methodology

How much do the results of each of the selection tests overlap with each other?

How can a relatively large, window-based signal of selection be narrowed down to an individual variant potentially driving the signal?

How are the results influenced by the choice of which immune genes to examine for evidence of selection?

Population selection histories

What is the distribution of global and local signals in the results of each selection statistic, and how do those compare to the results of previous studies?

Is there evidence of more selective pressure on immune genes in populations living in lower latitudes, where pathogen richness is greater?

Targets of selection

Do different classes of immune genes show different selection histories (positive vs. balancing vs. none)?

Do different classes of immune genes show different ages of selective pressure (ancient vs. recent)?

What are some of the most striking targets of positive and balancing selection?

Chapter 2

Methods for inferring positive and balancing selection from genomic data

2.1 Selection of immune genes

In order to search for signatures of selection in immune-specific genes, several lists of immune-related genes were assembled based on key sets of immune cell and peptide classes (see Section 1.3). In the interest of avoiding bias and assembling as complete a list as possible for different immune phenotypes, these lists were assembled from existing databases. It should be noted that no list of pathogen-interacting genes is fully complete, but using a database is potentially the most comprehensive way to assemble a list of relevant genes. Multiple databases were used to give a more complete view of different phenotypes. The datasets used are the Gene Ontology (GO) Consortium database (Ashburner et al., 2000) and the Host-Pathogen Interaction Database (HPIDB) (Ammari et al., 2016).

Genes chosen from the Gene Ontology Consortium database were selected based on their vital involvement in the human immune system. Since the innate immune system is the first line of defense and is passed down in the germline, unlike defenses gained during a lifetime through the adaptive immune system, it is the likely site of many possible beneficial adaptations. Because of this, genes with the following ontologies relating to innate immunity were selected from this database: innate immunity, interaction with bacteria, interaction with viruses, and antigen processing and presentation. Bacteria and viruses are included here since they are potentially the two most important classes of pathogens to affect humans. However, it is also possible that mutations that affect the way T cells or B cells function could result in faster, better, or longer lasting adaptive immunity, which could also result in increased survival of individuals with that mutation. Therefore, a subset of genes with ontologies

relating to adaptive immunity (adaptive immunity, T cells, and B cells) were also included. The full list of gene ontology terms for all genes (assembly GRCh37) was downloaded from the Gene Ontology Consortium website in February 2014, and afterward immune gene lists based on ontology were assembled by filtering for genes with the following immune-related search terms in their gene ontology description: "adaptive immun," "bact," "vir," "T cell," "B cell," "innate immun," and "antigen processing and presentation." These search terms corresponded respectively to the gene lists for genes associated with adaptive immunity, defense against bacterial infection, defense against viral infection, T cell function, B cell function, and innate immunity (see Table 2.1). These terms were chosen to give broad coverage over innate and adaptive immune processes.

The Host-Pathogen Interaction Database (Ammari et al., 2016) is organized into interactions between host genes and pathogen genes. From the human host category, genes that interacted with bacteria, viruses, protozoa, and amoebozoa were chosen to be included in this thesis. Bacteria and viruses are important classes of pathogens to include for the reasons listed above. The HPIDB also includes human genes that interact with fungal pathogens, but since relatively few fungal diseases cause disease in immunocompetent individuals (Murray et al., 2009), it is unlikely that they have been responsible for significant selection events in human history. The full list of host-pathogen protein interactions was downloaded from HPIDB 2.0 Ammari et al. (2016) in July 2016, and all interactions between human proteins and pathogen proteins of the types listed above (bacteria, virus, protozoa, and amoebozoa) were separated into groups based on type of pathogen. The list of genes associated with each of these proteins made up the gene list used for each pathogen group (see Table 2.1).

Both of these lists of genes were further filtered to contain protein coding genes only, based on classification by Ensembl BioMart in January 2018.

For both of these categories, the immune function lists and number of genes per list are shown in Table 2.1. The full lists of genes in each of the immune categories can be seen in Appendix A.

Database	Function	Abbrev.	No. of unique genes
GO	Adaptive immune response	GO.Adapt	31
GO	Bacteria	GO.Bact	226
GO	Virus	GO.Virus	833
GO	T cell	GO.Tcell	423
GO	B cell	GO.Bcell	186
GO	Innate immune response	GO.Innate	605
GO	Antigen proc. and pres.	GO.APP	208
HPIDB	Bacteria	HP.Bact	3545
HPIDB	Virus	HP. Virus	5522
HPIDB	Protozoa	HP.Prot	19
HPIDB	Amoebozoa	HP.Amoe	7
Both databases	All above phenotypes		8912

Table 2.1 Number of unique genes per immune function category. This also includes abbreviations for each gene category that may be use later in the text. Abbreviation: "proc. and pres." = "processing and presentation."

2.2 Overview of selection statistics used in this thesis

2.2.1 Positive selection statistics

Positive selection is the process by which the frequency of an allele which confers survival or other adaptive benefits increases within a population over time. There are three main types of statistics used to look for this evidence of positive selection in a single species. These are methods based on the site frequency spectrum, linkage disequilibrium, and population differentiation (review by Akey (2009)). Since each statistic has its pros and cons, it can be useful to analyze the same dataset with multiple statistics. This can help correct against false positives and false negatives. Out of the 360 immune-related genes found to be under positive selection mentioned in the above section, only around half of those were highlighted by more than one study (review by Barreiro and Quintana-Murci (2010)). For example, of the fourteen percent of the genome found to have been under positive selection by at least one study, only fourteen percent of those regions have been found by more than one study (review by Akey (2009)). Clearly, different studies and selection statistics pick up different results. In order to avoid the pitfalls of only using one statistic, this thesis uses multiple tests

of positive and balancing selection. The specific statistics used to detect positive selection are outlined below.

Tajima's D was developed in 1989 by Fumio Tajima, and is a widely used statistic for selection studies. The two main components of the Tajima's D statistic are the number of segregating sites and the average number of pairwise differences in a given sequence (Tajima, 1989). It is one of several selection tests based on the site frequency spectrum (SFS), or the allele frequency distribution of SNPs (Nielsen et al., 2007). A negative value of Tajima's D implies possible positive selection, and is found in regions where there is less heterozygosity than would be expected for a given number of segregating sites under neutral conditions (Carlson et al., 2005). This result can also be seen due to demographic effects, such as a sudden population expansion after a previous bottleneck, or if a neutral site is linked to a site under selection (Tajima, 1989). This statistic is best used for picking up relatively ancient selection of signals (between 50,000 and 250,000 years old) (Grossman et al., 2013; Sabeti et al., 2006). Compared to other statistics, Tajima's D is more relevant to looking for complete selective sweeps (Pybus et al., 2015).

A second type of statistic used for whole genome scans for positive selection are those based on haplotype homozygosity. These statistics are based on the structure of linkage disequilibrium in various regions of the genome. A region containing an unexpectedly long haplotype at a high frequency is a candidate region for an incomplete positive selective sweep (Nielsen et al., 2007). The particular strength of these tests lies in detecting signatures of incomplete sweeps (review by Karlsson et al. (2014)). These types of tests are used to find evidence of relatively recent selective sweeps that took place within the last 25,000 years (review by Sabeti et al. (2006)). Both iHS and nSL were used in this thesis to compare the results of these related tests. As opposed to Tajima's D, which is relatively more applicable to finding evidence of a complete selective sweep, statistics based on haplotype homozygosity are better used to find evidence of incomplete sweeps (Pybus et al., 2015).

Two different statistics based on haplotype homozygosity will be used in this analysis. The iHS statistic, or integrated haplotype score, is based on the EHH test (Sabeti et al., 2002), which quantifies haplotype decay as a function of distance from a specific allele at one end of the haplotype (Voight et al., 2006b). The nSL statistic was published in 2014 by Ferrer-Admetlla et al., and is similar to the iHS test. The main difference between these two statistics is that the iHS statistic measures the distance between mutations in terms of genetic distance, whereas the nSL statistic measures it in the number of mutations. This means that the nSL statistic does not require a recombination map (Ferrer-Admetlla et al., 2014).

The DIND (derived intra-allelic nucleotide diversity) statistic, developed by Barreiro et al. in 2009, is a statistic that is used in this project to narrow down candidate regions

for relatively recent selection to candidate alleles or haplotypes. This statistic is based on the idea that a region having undergone a selective sweep will be characterized by a high frequency allele surrounded by a region of low heterozygosity, since the beneficial mutation will have risen in frequency faster than the normal time to accumulate genetic diversity in linked regions (Barreiro et al., 2009).

The d_i statistic was developed by Akey et al. in 2010, and is based on the F_{ST} between one population compared to all other populations. F_{ST} is a classic statistic of genetic differentiation in the field of population genetics, developed by Sewall Wright. It is a measure of the proportion of variance within a subpopulation to the variance within the total population (Holsinger and Weir, 2009). Because of this, the strength of the d_i lies in finding regions where one population is particularly differentiated compared to all other populations (Akey et al., 2010). F_{ST} is used to detect signals younger than 50-75,000 years old (review by Sabeti et al. (2006)). While extreme differentiation doesn't necessarily imply selection, it can be one of several indicators (Xue et al., 2009), and if the differentiated alleles are functionally significant, can be phenotypically noteworthy regardless of the underlying reason for differentiation.

2.2.2 Balancing selection statistics

Three different statistics were used to look for evidence of balancing selection in this project. These are the HKA statistic, the β statistic, and Tajima's D. Details about the methods used to calculate these statistics can be seen in the following sections of this chapter. These three statistics were chosen because they are best powered in different time ranges, allowing an examination of balancing selection at different times in history. Additionally, Tajima's D was calculated for the positive selection analysis already and can be applied to look for balancing selection as well–however, as will be noted in future, the window size should be much smaller for balancing selection than for positive selection. It should also be noted that because variants need time in order to accumulate, a somewhat older time frame is required when looking for balancing selection.

The HKA statistic, developed by Hudson et al. in 1987, is used to look for balancing selection by comparing the ratio of polymorphisms to divergences in a region of interest to the expected ratio of that region under neutral conditions (Hudson et al., 1987). Balancing selection can be inferred if there are more polymorphisms than would be expected under neutral conditions. The HKA statistic is used to detect signals younger than 250,000 years old (review by Sabeti et al. (2006)).

The β statistic is a recently developed statistic by Siewert and Voight. According to their work, alleles surrounding a SNP under balancing selection are expected to show

the same distribution of allele frequencies that the SNP itself should show–namely an overrepresentation of mid-range frequencies. β is therefore a summary statistic that searches for groups of alleles showing these target frequencies. According to the authors, this statistic has maximum power to detect balancing selection older than 100,000 generations, thus selection that happened before the split of major continental population groups and indeed before the emergence of anatomically modern humans (Siewert and Voight, 2017a). Despite the fact that the β statistic is most highly powered to detect extremely ancient selection, it has been run on each of the 13 populations used in this project because it may be still somewhat powered to detect more recent selection.

Tajima's D was used in this project to look for balancing selection as well as positive selection. Whereas extreme negative values of Tajima's D suggest positive selection, extreme positive values of Tajima's D suggest balancing selection. Positive values of Tajima's D suggest that the number of haplotypes is greater than the number of segregating sites. This implies more diversity than would be expected under neutral conditions (Carlson et al., 2005). However, it should be kept in mind that just as negative values of Tajima's D can be caused by a recent population bottleneck and expansion, positive values of Tajima's D can indicate population contraction (Stajich, 2004; Tajima, 1989).

Summary of statistics listed above

Statistic	Brief description	Time frame
Tajima's D	Regions of increased/decreased diversity	50-250 kya
iHS	Extended haplotypes	< 25 kya
nSL	Extended haplotypes	< 25 kya
d_i	Extreme differentiation of one population vs others	< 75 kya
HKA	High ratio of polymorphisms to divergences	< 250 kya
beta	Groups of alleles at target frequency	> 2.5 mya
DIND	High frequency derived alleles surrounded by regions of low heterozygosity	Relatively recent

Table 2.2 Timeframes under which statistic is best powered.

Table 2.2 shows a summary of each of the statistics used in this thesis, as well as a brief description of what each is designed to find and the time frame under which each is best powered/used. Citations and more information on each can be found in the text above.

2.3 Calculation of selection statistics

2.3.1 nSL, iHS, and Tajima's D

The analyses in this section are based on two datasets. The first is the Tajima's D, nSL, and iHS results published by Pagani et al. in 2016, which used window sizes of 200 KB, and so this thesis has done the same.

nSL and iHS are both haplotype homozygosity tests, which are designed to pick up signals of positive selection through detection of long haplotypes. Tajima's D is based on the site frequency spectrum, and is designed to find regions where there is less (or more, if looking for balancing selection) heterozygosity than expected for a given number of segregating sites under neutral expectation (see Section 3.1.2 for more details on positive selection statistics). These results consist of a score per each 200 KB window in the genome that passed a number of quality filters. Though nSL and iHS are SNP-based statistics, these tests were used to assign scores to 200 B windows in the Pagani et al. 2016 analysis, and so this is the approach used for this thesis as well. These quality filters are: windows with average sequencing coverage of 0 were removed, windows outside of two standard deviations from the genome wide depth of sequencing coverage were removed, and all windows with fewer than 20 SNPs were removed (Pagani et al., 2016). The second dataset is based on a population of 19 Colla individuals published by Eichstaedt et al. in 2017 (Eichstaedt et al., 2017) and the results for the Tajima's D, nSL, and iHS tests run on that population by Dr. Charlotte Inchley, Dr. Georgi Hudjashov and Dr. Tiago Antão respectively. Those windows were also filtered to include only the windows from the first dataset that passed quality filters. The coverage per window for all populations was between 30X and 72X (Eichstaedt et al., 2017; Pagani et al., 2016).

Choosing balancing selection candidate genes in top Tajima's D windows

Tajima's D can be used to detect signals of balancing selection as well as positive selection. In the case of balancing selection, extreme positive values of the statistic suggest potential balancing selection. Therefore, the top one percent of 200 KB windows with the highest test statistics was assembled for each population as candidate windows for balancing selection. It should be noted that 200 KB is probably much too large of a window to use for this purpose, and it is not the same window size used with other balancing selection statistics–however, the window scores already existed from the Pagani et al. 2016 analysis and it made sense to see whether that data could be repurposed to look for balancing selection. Since each window usually contains multiple genes, further filtering was needed to narrow down the list of genes

potentially driving a signal of balancing selection. This was done by calculating the mean minor allele frequency per gene in each population, and assuming the gene with the highest mean minor allele frequency to be the gene most likely to be under balancing selection in a given 200 KB window. The reasoning behind this step is that balancing selection maintains diversity in a gene, and a higher minor allele frequency means a more even balance between two alleles as opposed to dominance by one allele and a small representation by another allele. The top 50 protein coding genes per population were assembled, consisting of the gene with the maximum minor allele frequency per top window. When there were no genes in a top window, that window was not included in the results.

2.3.2 d_i

Population differentiation-based analyses were based on d_i scores estimated by Pagani et al. for the EGDP dataset (Pagani et al., 2016) and by Tiago Antão and Georgi Hudjashov for the Colla population. The d_i statistic is based on the F_{ST} statistic, and is designed to find SNPs at which one population is especially differentiated in comparison to all other populations (see Section 3.1.2 for more information on this statistic). In order to filter out SNPs in regions with poor coverage, and to preserve consistency between different analyses, genomic regions outside the 200 KB windows that passed the quality filters used in the window-based positive selection analysis were removed from the set of scores. In order to compare relative d_i values of genes, each gene was assigned the highest d_i value of all SNPs between its start and end coordinates.

The set of genes and gene start/end coordinates was downloaded form Ensembl's BioMart website on January 8th, 2018 and is based on the GRCh37.p13 dataset. It is filtered to contain gene start and stop coordinates of only protein coding genes. This gene list was further filtered to only include the longest gene length per gene where duplicate gene entries were included. All genes on non-autosomal chromosomes were removed, as well as all genes containing a decimal point in their gene name, since literature searches for these genes returned results suggesting they were not protein-coding genes.

2.3.3 HKA

The HKA statistic detects signals of balancing selection by comparing the ratio of polymorphisms to substitutions in a region of interest to the expected ratio of that region under neutral conditions (see Section 4.1.2 for a more detailed description of this statistic). Here, polymorphisms are polymorphic loci and substitutions are loci where the derived allele is fixed. The HKA statistic was calculated in windows of 2000 base pairs with a shift of 500

base pairs, as described in the Pagani et al. paper (Pagani et al., 2016). The script used to calculate the statistic per window used the chi squared formula given for the HKA statistic in the supplementary materials of that paper, which is the following:

$$\chi^2 = ((n_{poly} - e_{poly})^2/e_{poly}) + ((n_{subs} - e_{subs})^2/e_{subs})$$

This chi squared value was calculated per 2000 bp window. n_{poly} and n_{subs} are the observed number of polymorphisms and substitutions in a given window, $e_{poly} = p(n_{poly} + n_{subs})$, and $e_{subs} = (1-p)(n_{poly} + n_{subs})$. p is $n_{poly}/(n_{poly} + n_{subs})$ for the whole genome. In short, this statistic calculates whether there are more or fewer polymorphisms per window than would be expected based on the genome-wide ratio of polymorphisms to polymorphisms and substitutions. Where there are more polymorphisms than expected, the statistic receives a positive value and when there are fewer it is assigned a negative value.

All HKA windows were removed which were not within the bounds of the 200 KB windows that passed the quality filters, as with the d_i analysis. Each protein coding gene was assigned the maximum HKA value of all the 2000 bp windows whose center was between the start and stop coordinates of the gene, as in the Pagani et al. paper (Pagani et al., 2016).

2.3.4 β

The β statistic detects regions where there is an overrepresentation of SNPs at mid-range frequencies (please see Section 4.1.2 for a more detailed description of the β statistic). The unfolded (meaning that whether the allele is ancestral or derived at a given site is taken into account) β statistic was applied to each chromosome of each population using the Python script from the paper in which the β statistic was published (https://github.com/ksiewert/BetaScan) (Siewert and Voight, 2017a). The authors assume a lineage split between humans and chimpanzees of 250,000 generations, so based on that assumption and the author-stated lower bound of 100,000 generations, any balancing selection detected would be in between 100,000 and 250,000 generations ago. Based on these numbers and the recombination rate (2.5e-8, used both in the β paper and Pagani et al. 2016 HKA analysis), the authors chose a window size of 1000 base pairs, which I have used as well. Other program parameters used the default parameters based on author recommendations and the high coverage of the EGDP data.

I have removed SNPs that do not have a folded (folded meaning that the ancestral or derived identity of a given allele is not taken into account) frequency of 15 percent in at least one population, as the authors did, since the authors were not able to simulate balancing selection in that scenario and wanted to avoid false positives.

β windows that were not within the bounds of the 200 KB windows that passed quality filters were discarded, and the remaining windows were used to assign each protein-coding gene a β score that was the maximum score for all windows intersecting with that gene. Windows were intersecting if the window center point was between the gene start and end.

2.4 Enrichment tests

2.4.1 nSL, iHS, and Tajima's D

In order to look for enrichment in immune phenotypes in top selection results, a list of the top one percent of windows based on their selection statistic score was assembled for each of the three window-based selection statistics. Since extreme negative values of Tajima's D are suggestive of positive selection and extreme positive values of Tajima's D are suggestive of balancing selection, an enrichment test was performed on both the top one percent of windows (positive selection) and the bottom one percent of windows (balancing selection). For each combination of selection statistic and class of immune genes, a Fisher's exact test was performed to see if windows containing immune genes were overrepresented in the top one percent of selection results. In order to correct for false positive results due to multiple testing, a Benjamini-Hochberg correction was applied using a false discovery rate of 0.1 in order to get a corrected p-value. P-value cutoffs of 0.05 and 0.01 were considered. Please note that the Benjamini-Hochberg correction was applied across each population and not within each category of immune genes, so in enrichment tables presented in Chapters 3 and 4, two populations may have the same number of genes in the top 1% of results in a category, but one of those populations may be significantly enriched and the other not. This is due to the different significance levels within the Benjamini-Hochberg tests for each of those populations.

2.4.2 β , HKA, and d_i

The HKA, and d_i, and β tests are all different from the large window-based tests (nSL, iHS, and Tajima's D) in that the the calculation of test statistics took place on a much finer scale. In the case of β and HKA, each window was 100 to 200 times smaller than those used in the large window-based tests, and in the case of the d_i test, the test statistic was calculated for each SNP. Because of this, a positive correlation between gene size/number of SNPs per gene and test statistic was expected to be a confounding factor in ranking genes by maximum test statistic. In order to account for this and be able to find the true top hits for each statistic unskewed by confounding gene size, several different binning methods were

considered to create bins containing approximately the same number of genes. An advantage of using the bin approach is that it controls for gene size, however, one disadvantage is that the targets of selection are not necessarily equally distributed among the size bins, which is an assumption of the binning method. Other methods, such as re-sampling, could also work here.The β statistic as run on the South Asian population was used to run the binning method comparisons.

The first binning method was to divide the datasets by gene length into 6 bins with the following bin dividers: 10 KB, 50 KB, 100 KB, 500 KB, and 1 MB. The second method was to divide the results into quartiles by gene size. The third and fourth methods were to divide the results into quintiles and deciles by gene size, respectively. Each of these four methods was used on the unfiltered dataset, as well as a version of the dataset which had all genes with fewer than 20 SNPs filtered out of it. 20 SNPs was chosen as a cutoff based on its use as a cutoff for small windows in another selection paper by Pickrell et al. (Pickrell et al., 2009). Tables including average number of SNPs per gene, average maximum β score per bin, number of genes in the bin, bin start and bin end can be seen in Tables 2.3 and 2.4.

Based on this data, the method of binning by quintiles on data filtered for a minimum of 20 SNPs per gene was chosen as the best method. The quintiles method is better than the quartiles or deciles method because it captures the relationship between gene size/SNPs per gene and average maximum test score per gene without over- or under-dividing the data. It is better than method 1 because it does just as good a job capturing the correlation between the two variables of interest but is simpler to calculate. Using this method, we have roughly the same number of genes per each of the five bins.

The different binning methods for the HKA and d_i statistic (minus the quartiles method, since it proved to be not much different from the quintiles method) run on the South Asian population as an example are shown in tables 2.5 through 2.8.

For each statistic, once the data was appropriately divided into bins, the top 10 genes per bin, resulting in 50 genes total, was assembled for each population and each test.

The top one percent of genes was used to perform an enrichment test for each statistic identical to those performed on the top one percent of windows in the iHS, nSL and Tajima's D tests, except using genes instead of windows.

	Binning Method 1					Binning Method 2					Binning Method 3					Binning Method 4				
	avgSNPsPerGene	avgMaxBeta	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxBeta	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxBeta	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxBeta	Genes	BinStart	BinEnd
Bin 1	79.4	7.7	2953	<10KB	10KB	87.6	8.1	3267	198	11271.25	70.4	7.3	2588	198	8713.4	38.0	5.7	1255	198	4294.7
Bin 2	411.1	18.6	6041	10KB	50KB	315.1	16.9	3635	11271.25	30604.5	229.0	14.2	2833	8713.4	21211.6	102.7	8.9	1333	4294.7	8713.4
Bin 3	1101.1	27.8	2557	50KB	100KB	782.8	24.8	3829	30604.5	78551.75	495.2	20.9	2983	21211.6	43825.4	179.9	12.2	1383	8713.4	14045.3
Bin 4	3089.5	38.7	2883	100KB	500KB	3466.1	38.4	3985	78551.75	2304637	1027.9	27.2	3119	43825.4	98873.8	278.1	16.1	1450	14045.3	21211.6
Bin 5	10901.0	60.2	228	500KB	1MB						3992.3	40.7	3193	98873.8	2304637	408.1	19.5	1481	21211.6	30604.5
Bin 6	25048.8	72.6	55	1MB	>1MB											582.3	22.3	1502	30604.5	43825.4
Bin 7																822.7	25.4	1542	43825.4	63425.8
Bin 8																1233.0	29.0	1577	63425.8	98873.8
Bin 9																2007.5	33.3	1587	98873.8	173828.2
Bin 10																5978.4	48.1	1606	173828.2	2304637

Table 2.3 Comparison of binning methods for the unfiltered β statistic. Method 1 is divided by gene length into bins using bin dividers of 10 KB, 50 KB, 100 KB, 500 KB, and 1 MB. Methods 2, 3 and 4 are divided into quartiles, quintiles, and deciles by gene length, respectively.

	Binning Method 1					Binning Method 2					Binning Method 3					Binning Method 4				
	avgSNPsPerGene	avgMaxBeta	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxBeta	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxBeta	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxBeta	Genes	BinStart	BinEnd
Bin 1	85.3	8.0	2771	<10KB	10KB	98.8	8.7	3285	476	12016	81.3	7.9	2607	476	9414.6	48.4	6.2	1290	476	4986.4
Bin 2	411.1	18.6	6040	10KB	50KB	329.4	17.3	3573	12016	31628	241.7	14.5	2796	9414.6	22113.8	114.2	9.7	1317	4986.4	9414.6
Bin 3	1101.1	27.8	2557	50KB	100KB	802.0	24.9	3765	31628	80188	512.6	21.3	2931	22113.8	44951.4	191.8	12.7	1373	9414.6	14861.2
Bin 4	3089.5	38.7	2883	100KB	500KB	3509.6	38.6	3910	80188	2304637	1049.0	27.3	3066	44951.4	100851.4	291.6	16.3	1423	14861.2	22113.8
Bin 5	10901.0	60.2	228	500KB	1MB						4040.8	40.9	3133	100851.4	2304637	424.5	20.1	1455	22113.8	31628
Bin 6	25048.8	72.6	55	1MB	>1MB											600.6	22.6	1476	31628	44951.4
Bin 7																840.7	25.3	1514	44951.4	64675.4
Bin 8																1257.1	29.2	1552	64675.4	100851.4
Bin 9																2042.9	33.6	1557	100851.4	176273.2
Bin 10																6038.6	48.2	1576	176273.2	2304637

Table 2.4 Comparison of binning methods for the β statistic, where all genes with fewer than 20 SNPs have been filtered out. Binning methods are the same as in Figure 2.3.

	Binning Method 1					Binning Method 2					Binning Method 3				
	avgSNPsPerGene	avgMaxHKA	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxHKA	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxHKA	Genes	BinStart	BinEnd
Bin 1	75.0	45.0	3128	<10KB	10KB	61.2	42.9	2563	176	8043	31.3	36.7	1212	176	3743.8
Bin 2	408.9	74.1	6073	10KB	50KB	215.3	63.1	2868	8043	20190.6	91.2	49.0	1351	3743.8	8043
Bin 3	1100.7	97.2	2558	50KB	100KB	475.8	78.7	3052	20190.6	42543	167.2	59.6	1404	8043	13238.8
Bin 4	3086.9	121.6	2886	100KB	500KB	1002.3	94.4	3181	42543	96738.8	263.4	66.7	1464	13238.8	20190.6
Bin 5	10901.0	169.0	228	500KB	1MB	3936.7	125.7	3263	96738.8	2304637	390.5	75.0	1515	20190.6	29509.5
Bin 6	25048.8	213.9	55	1MB	>1MB						561.1	82.5	1537	29509.5	42543
Bin 7											801.9	89.3	1573	42543	61767.7
Bin 8											1202.8	99.5	1608	61767.7	96738.8
Bin 9											1966.9	108.8	1621	96738.8	170403.1
Bin 10											5907.8	142.5	1642	170403.1	2304637

Table 2.5 Comparison of binning methods for the unfiltered HKA statistic. Method 1 is divided by gene length into bins using bin dividers of 10 KB, 50 KB, 100 KB, 500 KB, and 1 MB. Methods 2 and are divided into quintiles and deciles by gene length, respectively.

	Binning Method 1					Binning Method 2					Binning Method 3				
	avgSNPsPerGene	avgMaxHKA	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxHKA	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxHKA	Genes	BinStart	BinEnd
Bin 1	83.6	48.2	2854	<10KB	10KB	77.3	47.6	2602	476	9102	45.5	42.4	1282	476	4707
Bin 2	410.3	74.4	6057	10KB	50KB	235.6	64.7	2815	9102	21692	109.2	52.9	1320	4707	9102
Bin 3	1100.7	97.2	2558	50KB	100KB	503.6	80.1	2959	21692	44344	186.5	60.7	1378	9102	14491
Bin 4	3086.9	121.6	2886	100KB	500KB	1037.5	95.2	3095	44344	100138	284.7	68.7	1437	14491	21692
Bin 5	10901.0	169.0	228	500KB	1MB	4012.7	126.5	3166	100138	2304637	416.6	76.6	1470	21692	31109
Bin 6	25048.8	213.9	55	1MB	>1MB						590.7	83.7	1489	31109	44344
Bin 7											830.5	90.4	1529	44344	63985
Bin 8											1244.7	100.1	1566	63985	100138
Bin 9											2020.7	110.1	1573	100138	175122
Bin 10											6004.6	143.0	1593	175122	2304637

Table 2.6 Comparison of binning methods for the HKA statistic, where all genes with fewer than 20 SNPs have been filtered out. Binning methods are the same as in Figure 2.5.

	Binning Method 1					Binning Method 2					Binning Method 3				
	avgSNPsPerGene	avgMaxDi	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxDi	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxDi	Genes	BinStart	BinEnd
Bin 1	77.1	8.2	3091	<10KB	10KB	64.7	7.7	2584	198	8261.4	34.2	5.4	1244	198	3950.2
Bin 2	409.7	14.3	6096	10KB	50KB	219.7	12.4	2868	8261.4	20443.2	95.3	10.0	1340	3950.2	8261.4
Bin 3	1100.4	18.1	2561	50KB	100KB	480.5	15.1	3039	20443.2	42789.2	171.3	11.7	1406	8261.4	13512.2
Bin 4	3086.9	22.4	2886	100KB	500KB	1007.1	17.7	3173	42789.2	97297	268.1	13.1	1462	13512.2	20443.2
Bin 5	10901.0	30.6	228	500KB	1MB	3944.8	23.0	3252	97297	2304637	394.9	14.6	1508	20443.2	29720
Bin 6	25048.8	34.7	55	1MB	>1MB						566.0	15.6	1531	29720	42789.2
Bin 7											806.1	16.8	1569	42789.2	62037.4
Bin 8											1208.0	18.5	1604	62037.4	97297
Bin 9											1972.0	20.4	1615	97297	171316.4
Bin 10											5917.5	25.7	1637	171316.4	2304637

Table 2.7 Comparison of binning methods for the unfiltered d_i statistic. Method 1 is divided by gene length into bins using bin dividers of 10 KB, 50 KB, 100 KB, 500 KB, and 1 MB. Methods 2 and are divided into quintiles and deciles by gene length, respectively.

	Binning Method 1					Binning Method 2					Binning Method 3				
	avgSNPsPerGene	avgMaxDi	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxDi	Genes	BinStart	BinEnd	avgSNPsPerGene	avgMaxDi	Genes	BinStart	BinEnd
Bin 1	83.8	8.8	2870	<10KB	10KB	77.6	8.6	2614	476	9088.4	45.8	6.6	1291	476	4737.1
Bin 2	410.1	14.3	6092	10KB	50KB	235.2	12.8	2826	9088.4	21619.2	109.5	10.6	1323	4737.1	9088.4
Bin 3	1100.4	18.1	2561	50KB	100KB	501.6	15.2	2967	21619.2	44173.6	186.4	12.0	1381	9088.4	14440.6
Bin 4	3086.9	22.4	2886	100KB	500KB	1033.2	17.8	3106	44173.6	99619.4	284.0	13.5	1445	14440.6	21619.2
Bin 5	10901.0	30.6	228	500KB	1MB	4003.3	23.2	3178	99619.4	2304637	415.1	14.8	1473	21619.2	30978
Bin 6	25048.8	34.7	55	1MB	>1MB						588.1	15.5	1494	30978	44173.6
Bin 7											827.4	16.9	1535	44173.6	63785.9
Bin 8											1239.0	18.7	1571	63785.9	99619.4
Bin 9											2013.4	20.6	1579	99619.4	174391.2
Bin 10											5993.2	25.8	1599	174391.2	2304637

Table 2.8 Comparison of binning methods for the d_i statistic, where all genes with fewer than 20 SNPs have been filtered out. Binning methods are the same as in Figure 2.5.

2.5 Finding driver SNPs

2.5.1 nSL, iHS, and Tajima's D

The DIND (derived intra-allelic diversity) statistic (Barreiro et al., 2009) was used to move from units of windows to actual SNPs potentially responsible for observed signals of selection. While Tajima's D is already a window-based test, both nSL and iHS are SNP-based tests. However, the scores were available in 200 KB windows from the Pagani et al. 2016 analysis, so DIND was needed to move from those windows back into individual SNPs. In future work, recalculating those statistics and not converting them to scores per 200 KB windows would be ideal. The DIND statistic looks for SNPs under positive selection by finding derived alleles at high frequency that are linked with lower diversity sites than would be expected under neutral conditions (Barreiro et al., 2009). The DIND statistic was run on each SNP in each of the top one percent windows in each of the three selection statistics, using a script based on one written by Dr. Florian Clemente. The statistic was run in the same way as in the Pagani et al. paper: a DIND score was calculated in a 100 KB window around each SNP. A neutral cutoff was established for each population from performing a DIND analysis on a set of randomly selected 100 non-genic SNPs from each derived allele frequency class, and defining the cutoff as 3 standard deviations above the neutral DIND score. (It should be noted that the use of SNPs from non-genic regions as a neutral comparison is somewhat biased by the effect of purifying selection resulting in a smaller effective population size in genic regions than non-genic regions.) The DIND statistic was run such that significant SNPs have frequencies between 0.4 and 0.9, meaning that no SNPs will show up that are near fixation in a given population (Pagani et al., 2016).

The resulting list of SNPs from each population was then filtered for predicted functional importance using the CADD (Combined Annotation Dependent Depletion) variant scoring tool (Kircher et al., 2014). CADD calculates a single scaled C-score per SNP using 63 different variant scoring systems based on diverse metrics in order to predict the deleteriousness or functional importance of a SNP. This score takes into account factors such as predicted pathogenicity, evolutionary conservation scores, regulatory information, and protein-based scores of deleteriousness such as SIFT and PolyPhen. All SNPs not passing the recommended cutoff of 15 for the scaled C-score were filtered out (Kircher et al., 2014). This step is to filter out SNPs with lesser functional importance and leave only SNPs that potentially have a function that may be important enough to drive an observed selection signal.

This final list of SNPs was run through the Ensembl Variant Effect Predictor (VEP) tool (McLaren et al., 2016) in order to classify SNPs as nonsense, missense, synonymous, and assign them to gene names if in genic regions.

2.5.2 A note on looking for driver SNPs in balancing selection results

Statistics for finding signals of balancing selection are generally based on looking for regions where there is more observed polymorphism than would be expected under neutral expectations. Therefore, depending on the statistic, it can be more difficult to look for single SNPs as the drivers of balancing selection signals than of positive selection signals, where one particular SNP can be responsible for driving an observed signal. Because of this, this thesis stops at the resolution of genes and doesn't look for specific SNPs which may be responsible for driving observed balancing selection signals. However, window sizes used to calculate balancing selection statistics are usually smaller than those used for calculating positive selection statistics, meaning that window-based balancing selection results are easier to localize to smaller regions. Additionally, there are SNP-based statistics such as beta that can be used to this end. This could be an interesting avenue for future work, though for this thesis windows were used in order to keep the downstream analysis of the beta results comparable with the other two balancing selection statistics.

It would also be difficult to screen significant variants for functional importance using the methods used in this thesis (CADD), since CADD is based on evolutionary conservation among other things. Loci which are under balancing selection are necessarily under somewhat relaxed evolutionary constraint, which allows for variation. It would be an interesting direction for further research to attempt to look for individual variants, using a test like the maximum likelihood HKA (Wright and Charlesworth, 2004), and to look into other functional screening methods that would be more appropriate for screening targets of balancing selection. Future work might also make use of the balancing selection statistics NCD (Bitarello et al., 2018) and BALLET (DeGiorgio et al., 2014).

Chapter 3

Results: Positive selection

3.1 Introduction

3.1.1 Previous evidence of positive selection in immune genes

As of 2010, around fourteen percent of the genome had been reported to be under positive selection by at least one study (review by Akey (2009)). However, replication of results in selection studies is notoriously low (review by Akey (2009)), so this should not be taken to mean that fourteen percent of the genome has actually been under selection. Of the variety of functional classes of genes found to be under positive selection (review by Akey (2009)), immune-related genes have been shown in multiple studies to exhibit some of the strongest signals of selection (see reviews by Barreiro and Quintana-Murci (2010); Karlsson et al. (2014); Quintana-Murci and Clark (2013); Siddle and Quintana-Murci (2014)). Specifically, according to Barreiro et al., 360 genes with immune function have been found to exhibit significant signs of positive selection in at least one study (review by Barreiro and Quintana-Murci (2010)). Not only do some immune-related genes display strong signals of selection, but the immune gene class as a whole is significantly overrepresented in regions found to have experienced recent positive selection (Voight et al., 2006b; Wang et al., 2006)(see also review by Barreiro and Quintana-Murci (2010)).

Further evidence for the impact of infectious disease on the human genome is given by studies showing that diversity at immune-related genes is driven by pathogen richness (Fumagalli et al., 2009a,b; Prugnolle et al., 2005a)(see also review by Barreiro and Quintana-Murci (2010)). While it can be difficult to assign a particular cause to a selection event, migrations across Eurasia, the Neolithic expansion and rise of agriculture, and trade links and colonization between continents all presented evolutionarily recent immunological challenges

which may have given rise to the selection signatures we see today (review by Karlsson et al. (2014)).

In general, the types of immune genes that might be expected to have undergone positive selection are those where a specific, directional change is beneficial, such as selection for or against cell-surface proteins that are used by viruses during a viral infection (Van Blerkom, 2003). One famous example of this is the positive selection on a deletion in the gene *CCR5*, which produces a chemokine receptor on white blood cells used by HIV during an infection (see Section 1.5.) In a list of genes suggested to be targets of positive selection by more than one study, genes included components of the complement system, chemokines, interleukins, interferon-related genes, various protein kinases, different viral receptors, and other genes with less obvious connections to immune function (review by Barreiro and Quintana-Murci (2010)). These gene categories largely fall into those chosen to be examined in this thesis.

3.2 Results of window-based tests: nSL, iHS, and Tajima's D

The nSL, iHS, and Tajima's D tests were all run in 200 KB windows across the whole genome (leaving out regions that didn't pass a coverage filter, please see Chapter 2 for more information) and each window was assigned a test score. This information was published in the Pagani et al. 2016 paper (Pagani et al., 2016). Windows with the highest test scores (or lowest, in the case of Tajima's D) are the regions that are the best candidates for having been under past positive selection. In order to focus on the most likely candidates and eliminate at least some of the false positives, only the top one percent of windows were considered for selection candidates. These windows are informative in several ways. They can be used to look for enrichment of immune genes in the top results. They can also be used as a starting point for looking for individual SNPs that may be driving signals of selection. As a reminder, the nSL and iHS tests are best powered to find signals of selection younger than 25,000 years old, whereas Tajima's D best finds signals between 50,000 and 250,000 years old. Therefore the results in the following sections represent selection on different timescales.

3.2.1 Enrichment for immune genes in top results

In order to see whether top selection results are enriched for genes with immune functions, an enrichment test was performed for each immune category and population in the top one percent of each of the three window-based tests (see Chapter 2 for more information on enrichment tests). The results of these enrichment tests can be seen in Tables 3.1 through 3.5.

These tables show positive enrichment, where more immune genes are represented in the top one percent of results than would be expected given a uniform distribution of immune genes in the selection results.

nSL

The results of the enrichment tests for the nSL statistic are shown in Table 3.1. The pale orange cells show significant positive enrichment for immune genes in a given category, and darker orange cells show significance at a smaller p-value. In this test, there is a wide distribution of enrichment results across most immune gene classes and populations. The bacteria and virus genes from both the GO database and the HPI database both show significant enrichments, though generally not in the same populations. The only overlap is in the Northeast European population in genes involved in defense against bacterial infections. Some of the most striking enrichments are in T cell genes in the Northeast Siberian population, antigen processing and presentation genes in the Island Southeast Asian population, and innate immune genes in the Colla population.

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
exp	2.15	8.92	5.11	2.56	7.25	2.1	0.38	19.4	25.7	0.07	0.06
AFR	1	5	4	2	7	2	0	20	21	0	0
WAA	1	8	2	4	5	5	0	20	19	0	0
SWE	2	7	6	3	7	3	0	20	33	0	0
ENE	3	10	3	2	2	2	0	20	25	0	0
VOL	0	5	2	4	4	3	0	12	17	0	0
SOA	0	6	3	3	2	2	1	21	27	0	0
WSI	3	8	6	4	10	0	0	20	30	0	0
SSI	2	7	4	1	10	5	1	20	33	0	0
CSI	3	7	8	6	10	2	1	14	31	0	0
NSI	2	8	10	0	4	0	0	17	25	0	0
COL	6	16	6	5	15	6	0	22	25	0	0
SEM	0	11	4	2	9	1	0	17	26	0	1
SEA	7	8	6	4	6	8	0	14	22	0	0

Table 3.1 Enrichment of top one percent windows from the nSL test containing immune genes based on the GO DB and HPIDB. The "exp" row gives the expected number of windows containing immune genes in any given one percent of the data, and the counts in each cell represent the number of windows containing immune genes of a given category in the top one percent of windows. Light orange indicates significant enrichment at p = 0.05, dark orange at p=0.01 in a Fisher's exact test with correction for multiple testing, as described in Chapter 2.

It is interesting to know which genes are driving significant signals of enrichment. A subset of these (based on clear immune function based on the literature) is shown in Table 3.2. Please note that the necessarily concise descriptions of genes in the results presented in Chapters 3 and 4 are not comprehensive and are largely focused on how each of the genes relates to immune function.The full list of immune genes from each significantly enriched category are listed in Appendix B. Genes from the HPIDB categories are not shown here or included in descriptions of enrichment tables for other statistics, since there are many more genes in those categories, and many are shared with the categories from the Gene Ontology DB. The descriptions of the genes in the enrichment table are divided into paragraphs according to the immune gene category in which they appear. Additionally, since some genes appear in multiple immune gene categories, they will be described in the first category in which they appear but not repeated in every category.

Based on Table 3.2, there is not much sharing of genes driving enrichment signals between populations. There are some genes that are shared between populations (such as

TNFSF4 in the South Siberia and Mongolian and Central Siberian populations), but these are often shared between populations that are relatively close geographically. This pattern fits with nSL as a measure of relatively recent, local selection.

	Pop.	Genes
GO.Bact	ENE	LCN2
	COL	CHIT1, DEFA1, DEFA1B, DEFA3, DEFA5, IL22RA1, IL27
	SEA	ERAP1, TLR4
GO.Virus	ENE	CXCR4, DUOX2, LCN2
	COL	DEFA1, DEFA1B, DEFA3, HLA-A, IFNLR1, IL27, TRIM56, UNC93B1, XPO1
	SEM	CD86, EPHA2, TRIM23
GO.Tcell	CSI	IL1B, IL7, TNFSF4
	NSI	IL15, NDFIP1, SMAD3
	SEA	HLA-A, HLA-F, HLA-G
GO.Bcell	WAA	IL4
	SWE	IL4
	CSI	IL7, TNFSF4
	SEA	TLR4
GO.Innate	WSI	DDX58, LGR4
	SSI	CAMK4, DEFA1, DEFA1B, DEFA3, DEFA5, NCF1, NRG1, RNF125, TICAM2
	CSI	DDX41, DMBT1, IRAK2
	COL	AGO1, AGO3, DEFA1, DEFA1B, DEFA3, DEFA5, FCN3, IL27, TRIM56, UNC93B1
	SEM	C4B
GO.APP	SSI	AP2M1, NCF1
	COL	AP2M1, HLA-A
	SEA	ERAP1, HLA-A, HLA-F, HLA-G
GO.Adapt	SOA	ZAP70
	SSI	TNFSF4
	CSI	TNFSF4

Table 3.2 Subset of immune-related genes driving significant nSL enrichment signals

In the enrichment results for the GO.Bacteria gene category and in the Northeast European population, LCN2 (lipocalin 2) expression is TLR-induced, playing an important role in the innate immune response to bacterial infection via sequestration of iron (Flo et al., 2004). In the Colla population, CHIT1, or chitinase 1, is a marker of macrophage activation and has been associated with immune response to tuberculosis and leprosy (Cakır et al., 2012; Iyer et al., 2009), as well as malaria, allergies, and fungal infections (Di Rosa et al., 2012). Various defensin genes such as *DEFA1*, *DEFA1B*, *DEFA3*, and *DEFA5* appear in multiple immune categories (GO.Bact, GO.Virus, GO.Innate) in different populations, namely the Colla and the South Siberian and Mongolian populations. Defensins are antimicrobial peptides found

in mucosal surfaces and secretions (Murray et al., 2009). In the Island Southeast Asian population, ERAP1 in the GO.Bact category plays role in peptide processing by the MHC and has been implicated in psoriasis susceptibility (Strange et al., 2010). Also in the results for the Island Southeast Asian population, TLR4 is one of the keystones of the innate immune response to infection. It is most well known as a receptor for LPS, or lipopolysaccharide, which is part of the cell membrane of Gram-negative bacteria. However, it also responds to respiratory syncytial virus and hepatitis C virus (Murray et al., 2009).

In the enrichment results for the GO.Virus category and the Northeast European population, CXCR4 is one of the receptors used by HIV to enter human T cells (Murray et al., 2009). DUOX2 is a NADPH oxidase that is induced by interferon and plays a role in antiviral immune response (Fink et al., 2013). *HLA-A*, *HLA-F* and *HLA-G* appear in this list in the GO.Virus, GO.Tcell and GO.APP categories in the Colla and Island Southeast Asian populations. HLA genes are highly polymorphic genes involved in antigen processing and presentation. *IFNLR1*, or *IL28RA*, is in a top window in the GO.Virus category in the Colla population. Again in the Colla population, IFNLR is a receptor for type-III interferons and IFNLR1 is one of the subunits of that receptor and is expressed in epithelial cells. Knockout mice for this gene experienced increased viral titers or increased shedding (in general, stronger infections) when infected with a variety of viruses including West Nile virus, norovirus, rotavirus, and influenza virus (review by Lazear et al. (2015)). When inhibited in humans cells infected with influenza, increased titers of the virus were observed (review by Lazear et al. (2015)). Blockage of the protein product of this gene has been suggested as a way to boost vaccine response (Egli et al., 2014). TRIM56 is a ubiquitin ligase whose expression is induced by interferon and is involved in the innate immune response to double-stranded DNA virus infection (Tsuchida et al., 2010), and plays a role in immune response against influenza virus as well (Liu et al., 2016a). UNC93B1 is a transmembrane protein used by multiple TLRs for passage within the cell and required for their function (Lee et al., 2013). XPO1, or exportin 1, has been shown to be associated with influenza replication. Inhibition of XPO1 led to a reduction in replication of influenza virus and this gene has been suggested as a potential drug target (Perwitasari et al., 2014). In the Mainland Southeast Asian population, CD86 is involved in the induction of T cell immune response (Freeman et al., 1993). EPHA2 is an ephrin receptor that is a host factor for hepatitis C virus and has been suggested to be a potential target for antiviral treatment (Lupberger et al., 2011). *TRIM23* is important in the TLR3 antiviral and innate immune response (Arimoto et al., 2010).

In the results for the T cell category of genes and in the Central Siberian population (though it also appears in other immune gene categories), *TNFSF4* is a member of the tumor

necrosis factor family that is expressed on antigen presenting cells and is a stimulator for T cells. It has been associated with risk for autoimmune disease (Graham et al., 2007; Gramaglia et al., 1998; Ohshima et al., 1997). In the Northeast Siberian population, *NDFIP1* is a regulator of gastrointestinal inflammation caused by T cells (Ramon et al., 2010). SMAD3 is a transcription factor involved in the regulation of T cell activation via TGF-β and mucosal immunity, as well as activation of interferon-β (Qing et al., 2004; Yang, 1999).

There are several cytokine genes that appear in different categories and populations. *IL1A* and *IL1B* play a role in the innate immune response and appear in the Central Siberian population in the GO.Tcell category. *IL4* and *IL7* are involved in the Th2 response and are in the Western Asian and Armenian and Central Siberian populations for the GO.Bcell category (Murray et al., 2009). In the Colla population, *IL27* (GO.Innate, GO.Virus) and *IL22RA1* (GO.Bact) also appear in top windows. *IL15* is in a top window in the GO.Tcell category of the Northeast Siberian population population.

In the results for the innate immune category and in the Western Siberian population, the gene *DDX58* produces the protein RIG-I, which is an interferon-induced recognizer of viral DNA/RNA and plays a role in the antiviral immune response (Malathi et al., 2007; Ramos and Gale, 2011; Suárez-Calvet et al., 2014). LGR4 is a modulator of TLR2 and TLR4 function (Du et al., 2013). In the South Siberian and Mongolian population, *RNF125* is involved in the regulation of RIG-I signaling (Arimoto et al., 2007). CAMK4 has been shown to play a role in the regulation of antibody response in a mouse influenza vaccine study (Nakaya et al., 2011). It is also involved in the inflammatory response and T cell development (Illario et al., 2008; Krebs et al., 1997). *NCF1* has a role in the oxidase complex and mice with mutations in this gene experience deficient neutrophil killing (Miao et al., 2010a; Sumimoto et al., 2005). Mutations in *NRG1* have been associated with immune dysregulation, including increased levels of autoimmune markers and altered cytokine expression (Marballi et al., 2010). TICAM2 is a TLR adaptor that is involved in the TLR4 immune response (Seya et al., 2005). Variants in this gene have been associated with TB susceptibility (Hall et al., 2014). In the Central Siberian population, DDX41 is a helicase that activates an interferon innate immune response to bacterial messengers cyclic di-GMP and cyclic di-AMP (Parvatiyar et al., 2012). DMBT1 is involved in pathogen recognition in mucosal tissues (Bikker et al., 2002; Rosenstiel et al., 2007). IRAK2 plays a role in the early and late phases of the innate immune TLR response (Kawagoe et al., 2008). In the results for the Colla population, AGO1 and AGO3 have been implicated in a mouse study of influenza A infection to be involved in tolerance of inflammation in the lung during infection (Van Stry et al., 2012). FCN3 is a pattern recognition molecule that is involved in the lectin complement pathway, with highest expression in the lungs (Garred et al., 2009; Hummelshoj et al., 2008). In the Mainland

Southeast Asian population, C4B is part of the complement pathway, which is part of the innate immune response against bacterial infection (Murray et al., 2009).

AP2M1 in the antigen processing and presentation category is a membrane trafficking protein that is essential for HCV replication and assembly (Neveu et al., 2012). This gene was in the enrichment results for both the South Siberian and Mongolian and Colla populations.

In the enrichment results for the Go.Adapt gene category in the South Asian population, ZAP70 is required for the formation of the immune synapse (Blanchard et al., 2002).

iHS

Table 3.3 shows the results for the enrichment test results based on the iHS statistic. Compared to the nSL test, there are more enrichments in the GO.Virus (though not the GO.Bact), HP.Bact, and HP.Virus categories. There are also fewer GO.APP enrichments and GO.Innate enrichments. Some of the most striking enrichments are in the GO.Virus category, in the Volga Uralic, Colla, and Mainland Southeast Asia populations. The Southwest European population in the GO.APP category is also very enriched for top iHS scores.

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
exp	2.15	8.92	5.11	2.56	7.25	2.1	0.38	19.4	25.7	0.07	0.06
AFR	1	9	4	1	6	2	0	24	28	0	0
WAA	1	11	6	4	3	3	1	22	32	0	0
SWE	3	14	6	5	6	7	0	26	36	0	0
ENE	5	13	3	2	5	3	0	31	31	0	0
VOL	1	20	6	8	11	3	1	35	40	0	0
SOA	1	12	5	3	7	2	0	25	34	0	0
WSI	0	11	5	3	4	1	0	23	33	0	0
SSI	2	12	9	6	9	4	1	27	37	0	0
CSI	1	9	7	3	6	1	1	31	41	0	0
NSI	2	10	8	0	4	3	0	36	46	0	0
COL	1	19	5	5	12	2	0	32	39	0	0
SEM	1	19	7	3	4	2	0	21	24	0	0
SEA	3	11	4	4	3	2	1	20	30	0	0

Table 3.3 Enrichment of top one percent windows from the iHS test containing immune genes based on the GO DB and HPIDB. The "exp" row gives the expected number of windows containing immune genes in any given one percent of the data, and the counts in each cell represent the number of windows containing immune genes of a given category in the top one percent of windows. Light orange indicates significant enrichment at p = 0.05, dark orange at p=0.01 in a Fisher's exact test with correction for multiple testing, as described in Chapter 2.

The full list of immune genes from each significantly enriched category are listed in Appendix B. A representative subset of these (based on clear immune function based on the literature) is shown in Table 3.4. Genes from the HPIDB categories are not shown here, since there are many more genes in those categories, and many are shared with the categories from the Gene Ontology DB. Several of the genes in this figure have been discussed in the nSL enrichment section above and will not be described again. This section is organized in the same way as the one covering the nSL enrichment results, where one immune gene category is described per paragraph.

	Pop.	Genes
GO.Virus	SWE	CPSF4, DDX5, DUOX2, RPLP1, TNFRSF14
	ENE	DUOX2, DYNLL1, TNFRSF14, TRIM23
	VOL	HNRNPK, IL27, PSMC2, SRPK2
	SOA	CD247, TNFRSF14
	WSI	SRPK2, SYNCRIP
	COL	ANKRD17, DYNLT1, IL27, RHOA, SYNCRIP, XPO1
	SEM	EPHA2, RHOA, TRIM23
GO.Tcell	SSI	IL2, PTPN22, TNFRSF14, TNFSF4
	NSI	IL27, SMAD3, ZEB1
	SEM	HLA-DQA1, HLA-DRA, HLA-DRB1, HLA-DRB5, ZEB1
GO.Bcell	WAA	GON4L, IL4, MZB1
	SWE	FNIP1, IL2, IL4
	VOL	IL2, LAT2
	SSI	CD19, IL2, PTPN22, TNFSF4
	COL	CD19, MZB1
	SEA	BANK1, IL2
GO.Innate	VOL	ADRBK1, IL27, LAT2, TRAFD1
	COL	ADRBK1, CD19, IL27, LAT, TYRO3
GO.APP	SSI	NCF1
GO.Adapt	CSI	TGFB1

Table 3.4 Subset of immune-related genes driving significant iHS enrichment signals

In the enrichment results for the GO.Virus category and in the Southwest European population, TNFRSF14 is a receptor used by herpes simplex virus 1 to enter human cells and has been implicated in autoimmune disease (see review by Steinberg et al. (2011)). DDX5 is involved in T_H17 cell differentiation and function, and this association has led DDX5 to be suggested as a target for minimizing pathogenesis caused by the pro-inflammatory effects of T_H17 (Huang et al., 2015). T_H17 cells are T cells that are involved in protection of mucosal surfaces from bacteria and fungi, but also are involved in autoimmune and inflammatory disease (Huh et al., 2011; Weaver et al., 2007). RPLP1 is a phosphoprotein that has been identified as an essential host factor for flavivirus infection (Campos et al., 2016). In the Volga Uralic population, HNRNPK has been implicated in the viral replication of influenza A virus (Tsai et al., 2013). In a study of macaque proteome response to influenza infection, PSMC2 had a strong correlation with disease (Brown et al., 2010). SRPK2 is involved in hepatitis B infection and has been suggested as a potential target for therapeutic intervention (Daub et al., 2002). In the South Asian population, *CD247* is involved in T cell signaling and

variants in this gene have been associated with multiple autoimmune diseases (Holmberg et al., 2016). In the Western Siberian population, SYNCRIP is involved in hepatitis C virus replication (Liu et al., 2009). In the Colla population, ANKRD17 is involved in regulating RLR (RIG-I-like)-associated immune response in response to viral RNA (Wang et al., 2012b). In the Mainland Southeast Asian population, RHOA is a GTPase that has been suggested as an antiviral target for viruses that cause syncytium (Pastey et al., 2000).

In the enrichment results for the T cell gene category and in the South Siberian and Mongolian populations, variants in the tyrosine phosphatase gene *PTPN22* have been linked to multiple autoimmune diseases, potentially through altered T cell receptor signaling (see review by Bottini et al. (2006)). In the Northeast Siberian population, *ZEB1* is involved in T cell development and memory B cell response (Arnold et al., 2012).

In the B cell category, both *FNIP1* in the Southwest European population and *GON4L* in the West Asian and Armenian population play a role in B cell lymphopoiesis (Lu et al., 2010; Park et al., 2012). In the West Asian and Armenian population, *MZB1* influences antibody secretion and calcium homeostasis in innate-like B cells (Flach et al., 2010). In the Volga Uralic population, LAT and LAT2 are transmembrane adaptor proteins that play a role in the activation of of T cells and B cells (and other immune cells), respectively (Brdička et al., 2002). In the South Siberian and Mongolian population, *CD19* has been associated with humoral immune response. Lower *CD19* expression results in suppressed immune response (Ogura et al., 2016). In the Island Southeast Asian population, BANK1 is an adaptor protein for B cells and has been associated with autoimmune disease risk (Jordan et al., 2003; Kozyrev et al., 2008).

In the enrichment results for the innate immune category and in the Volga Uralic and Colla populations, ADRBK1, or GRK2, is a G protein-coupled receptor kinase expressed at relatively high levels in immune cells and its expression levels are altered in experimental models of autoimmune disease. It plays a role in the innate immune response to infection (see review by Vroon et al. (2006)). In the Volga Uralic population, *TRAFD1*, or *FLN29*, has a dampening effect on the innate immune response to LPS. Knockout mice for this gene were more susceptible to sepsis than wild type mice (Sanada et al., 2008). In the Colla population, TYRO3 is an immunoregulatory receptor tyrosine kinase. Knockout mice for this gene develop autoimmunity (Lu, 2001).

In the adaptive immunity category and the Central Siberian population, TGFB1 is a cytokine that promotes T_H17 cell function (Murray et al., 2009) and is seen in the Central Siberian population in both the GO.Adapt and the HP.Bact categories.

Tajima's D

Based on Table 3.5, there is more sharing of significant enrichment between populations than there was for either the nSL or the iHS tests. Given that Tajima's D is supposed to pick up older signals of selection, these could be older shared signals of selection that are shown here compared to the younger selection signals shown in the results of the nSL or iHS tests. Every population is enriched in the HP.Bact category, and many are significantly enriched in the GO.Bact category as well. The West and Central African population is significantly enriched in the GO.APP category.

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
exp	2.15	8.92	5.11	2.56	7.25	2.1	0.38	19.4	25.7	0.07	0.06
AFR	0	16	6	7	9	8	0	25	28	0	0
WAA	5	15	7	4	13	3	0	27	30	0	0
SWE	4	12	6	3	8	3	0	32	31	0	0
ENE	2	13	9	7	13	4	0	28	32	0	0
VOL	3	11	5	3	9	1	0	23	29	0	0
SOA	6	7	3	3	12	0	0	25	30	0	0
WSI	2	9	3	0	7	2	0	36	27	0	0
SSI	2	8	4	2	10	1	0	33	26	0	0
CSI	4	10	4	0	10	3	0	36	28	0	0
NSI	3	9	3	0	8	1	0	33	32	0	0
COL	4	13	4	1	8	3	0	37	38	0	0
SEM	2	8	1	1	7	2	0	32	30	0	0
SEA	4	9	5	1	10	3	0	31	33	0	0

Table 3.5 Enrichment of top one percent windows from the Tajima's D test containing immune genes based on the GO DB and HPIDB. The "exp" row gives the expected number of windows containing immune genes in any given one percent of the data, and the counts in each cell represent the number of windows containing immune genes of a given category in the top one percent of windows. Light orange indicates significant enrichment at p = 0.05, dark orange at p=0.01 in a Fisher's exact test with correction for multiple testing, as described in Chapter 2.

A subset of genes driving the above enrichment signals is shown in Table 3.6. The full list of immune genes from each significantly enriched category are listed in Appendix B. Sharing at the specific gene level is higher in the Tajima's D results as well when compared to the nSL and iHS results, as shown in Table 3.6. Many of the populations share a core set

of genes driving the selection signal in a given class of genes. Descriptions for genes below may appear in the descriptions for the enrichment tables for the nSL and iHS tests.

In general, when compared to the enrichment results of the nSL and iHS populations, the Tajima's D enrichment results show more sharing of genes between populations. Unlike in the nSL and IHS enrichment results, there are some significant enrichments in the West and Central African population. In the case of enrichments in the GO.Virus category, two genes (*TMEM173* and *RTN3*) are found in both African and non-African populations, suggesting a potentially relatively ancient selection pressure shared that predates the split between modern African and non-African populations.

	Pop.	Genes
GO.Bact	WAA	IL23A
	SWE	IL23A
	SOA	DEFB121, DEFB123, DEFB124, IL23A
	CSI	NCF1
	COL	CD36, GPX1
	SEA	GPX1, NCF1
GO.Virus	AFR	ANKRD17, HNRNPA1, MAGI3, PSMC2, RTN3, SNRNP200, TMEM173
	WAA	CPSF4, DUOX2, HDAC1, IL23A, LCK, STAT2
	SWE	CPSF4, DUOX2, IL23A, STAT2
	VOL	CPSF4, TMEM173
	COL	NXF1, RHOA, RTN3, STAT3
GO.Tcell	ENE	SLA2, THEMIS2
GO.Bcell	AFR	PTEN, PTPN22, SLA2
	WAA	LCK
	ENE	SLA2
GO.Innate	WAA	ADRBK1, AGO1, AGO3, IL23A, LCK
	ENE	ADRBK1
	VOL	ADRBK1, TMEM173
	SOA	ADRBK1, AGO1, AGO3, IL23A
	SSI	CD247, GRB2
	CSI	GRB2, NCF1
	SEA	NCF1, NCK1, NFATC3
GO.APP	AFR	PSMC2
	COL	CD36, PSMD5, SEC31A
	SEA	DYNLL2, NCF1

Table 3.6 Subset of immune-related genes driving significant Tajima's D positive selection enrichment signals

In the enrichment results GO.Bacteria category and in the Colla population, CD36 is a receptor for bacteria and contributes to the immune response to especially Gram-negative bacteria and LPS (Baranova et al., 2008). *GPX1*, or glutathione peroxidase-1, plays a role in the inflammatory response in the intestine in a mouse model (Esworthy et al., 2001).

In the GO.Virus category in the West and Central African population, HNRNPA1 is an RNA-binding protein that plays a role in hepatitis C replication (Kim et al., 2007). MAGI3 interacts with multiple viruses, including HPV and influenza A (Golebiewski et al., 2011; Thomas et al., 2002). *RTN3* is involved in enterovirus replication (Tang et al., 2006). SNRNP200 is an RNA helicase that is an RNA sensor and activator of the innate antiviral immune response (Tremblay et al., 2016). *TMEM173*, or *STING* (stimulator of interferon genes), plays a role in innate immune response to viral infection and interferon production (Ishikawa and Barber, 2008). In the West Asian and Armenian population, HDAC1 is a histone deacetylase that plays a role in the regulation of TLR and interferon response (see review by Shakespear et al. (2011)). *LCK* plays a role in T cell receptor signaling (Philipsen et al., 2017). STAT2 in the West Asian and Armenian and Southwest European populations associates with HDAC1 and is also important in the innate immune viral response (Nusinzon and Horvath, 2003). In the Colla population, NXF1 is a nuclear export factor that has a role in influenza virus replication (Karlas et al., 2010; Satterly et al., 2007). STAT3 is involved in antigen-specific tolerance in T cells (Cheng et al., 2003) and plays a role in hepatitis C infection (Yoshida et al., 2002).

In the enrichment results for the T cell category and in the Northeast European population, SLA2, or SLAP, is an adaptor protein that plays a role in signal regulation of multiple types of immune cells (see review by Kazi et al. (2015)). *THEMIS2* is involved in macrophage inflammatory responses (Peirce et al., 2010).

In the B cell category and in the West and Central African population, *PTEN* plays a role in modulation of the innate immune response (Günzl and Schabbauer, 2008). *PTPN22* is involved in the antiviral immune response and regulation of interferon secretion (Wang et al., 2013b).

In the innate immune category and South Siberian and Mongolian population, GRB2 performs an inhibitory role in T cells along with LAT (Yamasaki et al., 2003). In the Island Southeast Asian population, *NCK1* is involved in formation of immune synapses between antigen-presenting cells and T cells (Yiemwattana et al., 2011). NFATC3 is a regulator of T cell receptor responsiveness (Rengarajan et al., 2001).

3.2.2 Finding driver SNPs

Enrichment results from the previous sections list many promising and exciting potential immune targets of selection. The long list of genes supports previous findings that immune genes are often among the top results for positive selection. However, while genes in the top one percent of results discussed in the previous section may indeed be targets of positive selection, this list of genes can also serve as a starting point for further analysis. The genes

highlighted in this section often share their 200 KB window with other genes that may not be associated with immunity. Therefore, the next step in this analysis is to narrow down the selection signal in each top window to find potential causative SNPs (or SNPs that are in linkage with causative SNPs) that are in immune genes.

Overall characteristics of significant SNPs

The method for finding individual SNPs based on the results from the genome-wide window-based nSL, iHS, and Tajima's D tests is described in Chapter 2 and is summarized by Figure 3.1. In short, the top one percent of results based on each of the window-based tests was run through a DIND filter, which resulted in SNPs that are derived alleles at high frequency and linked with lower diversity sites than would be expected under neutral conditions, suggesting positive selection. These SNPs were then passed through the CADD filter (see Section 2.5 on page 37) of functional importance, which scored each SNP based on a large number of scores including evolutionary conservation and predicted functional deleteriousness. Those SNPs that passed the CADD cutoff statistic were then run through the Kaviar tool (Glusman et al., 2011) to assign them to rsIDs, which were run through the Ensembl Variant Effect Predictor (Version GRCh37) (Zerbino et al., 2017) to find SNP consequences, gene names, and other characteristics.

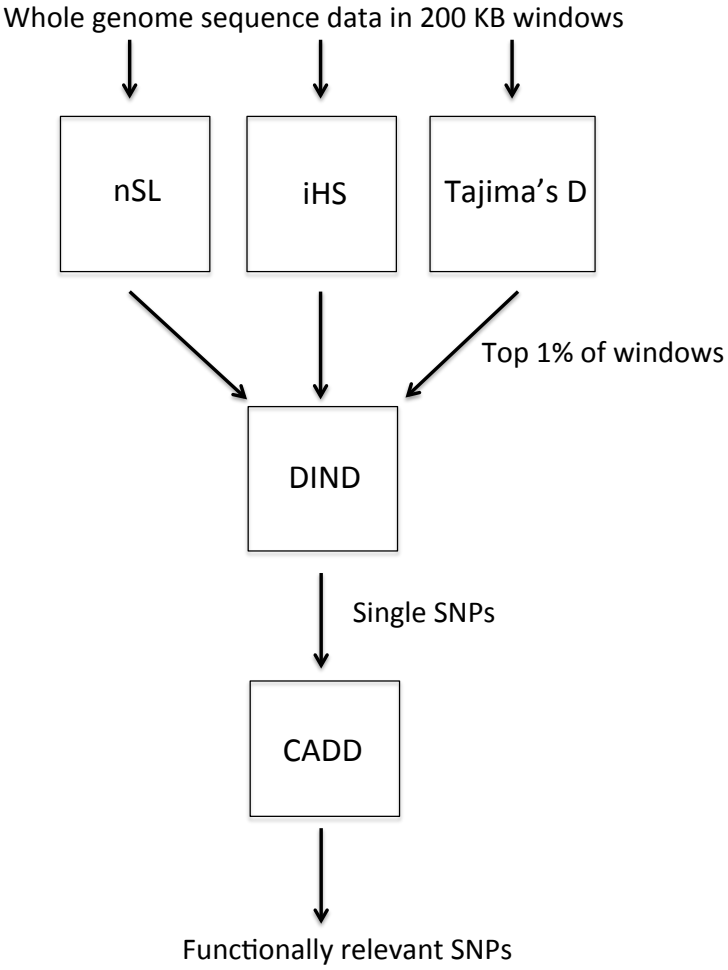


Fig. 3.1 Flow chart of window-based positive selection analysis.

After passing through the statistical and functional filters shown above, a total of 755 unique SNPs were counted as significant in all 13 populations. The number of significant SNPs per population can be seen in Table 3.7. There are the most significant SNPs in the West and Central African population, which has over twice as many as any other population. The Northeast Siberian population has the fewest significant SNPs.

	No. top 1% 200 KB wins (all tests)	→	SNPs above DIND cutoff	→	SNPs above CADD cutoff
AFR	375		11281		147
WAA	375		3464		37
SWE	375		3428		48
ENE	375		4317		63
VOL	375		4598		61
SOA	375		4090		56
WSI	375		2695		43
SSI	375		4049		52
CSI	375		2844		38
NSI	375		1893		28
COL	373		3016		53
SEM	375		3700		52
SEA	375		3477		77

Table 3.7 Number of windows and SNPs before and after the DIND and CADD filters

Figure 3.2 shows the relative number of SNPs with different Variant Effect Predictor (VEP)-defined consequences between the thirteen populations. The West and Central African population has the most variants at this stage overall, and that is evident from the figure. This figure is also a quick way to compare relative levels of different types of variants in the filtered dataset as a whole. The largest classes of variants are intron variants and non coding transcript variants. There are a fair number of missense variants, and very few stop gain or splice acceptor variants.

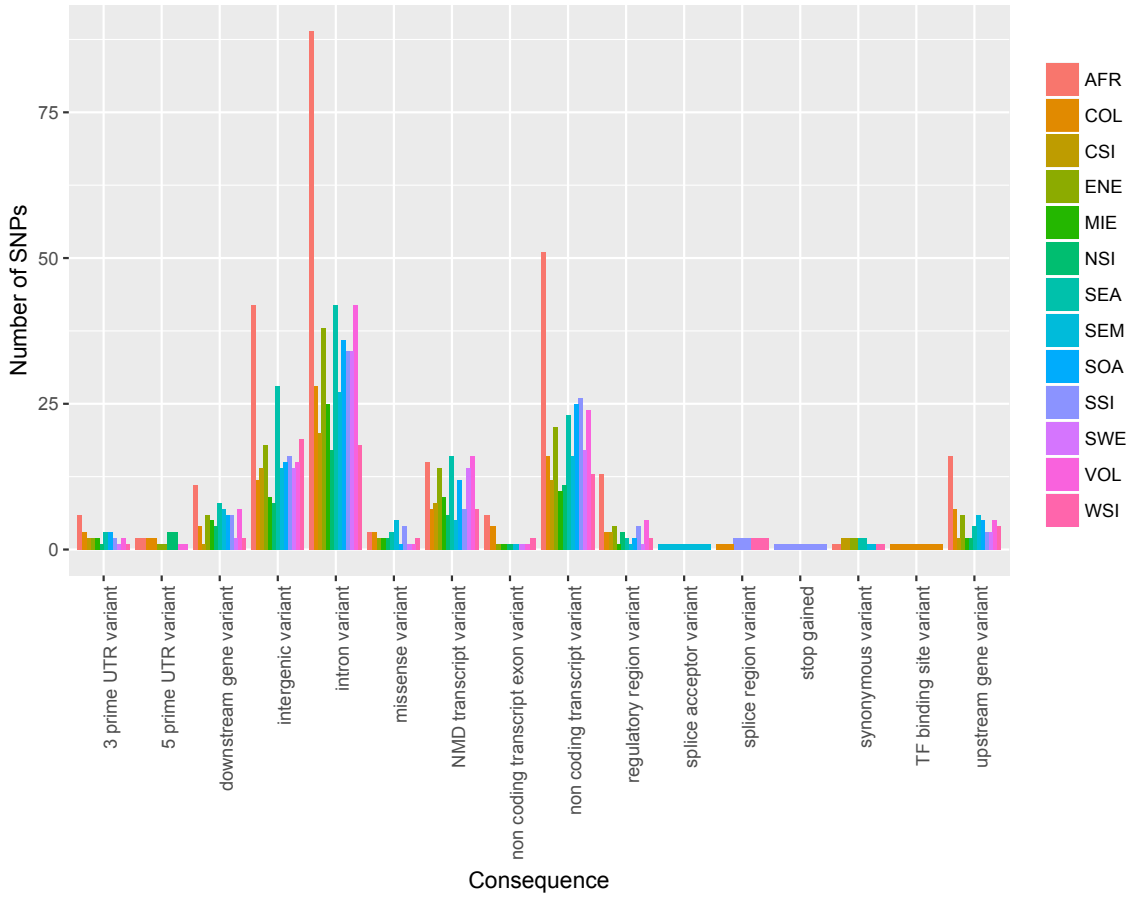


Fig. 3.2 Relative numbers of variants in each VEP-defined functional class. Each population is shown in a different color. "NMD" is an abbreviation for nonsense-mediated decay, "UTR" for untranslated region, and "TF" for transcription factor.

Figure 3.3 shows a distribution of CADD score for different Variant Effect Predictor variant types, separated into facets by VEP-defined impact. The four levels of impact are high, moderate, low, and modifier (labeled as A, B, C, and D, respectively). This shows how CADD ranks different types of variants in this dataset, as well as shows that the VEP-defined impact levels and the CADD scores generally tend to be in overall agreement.

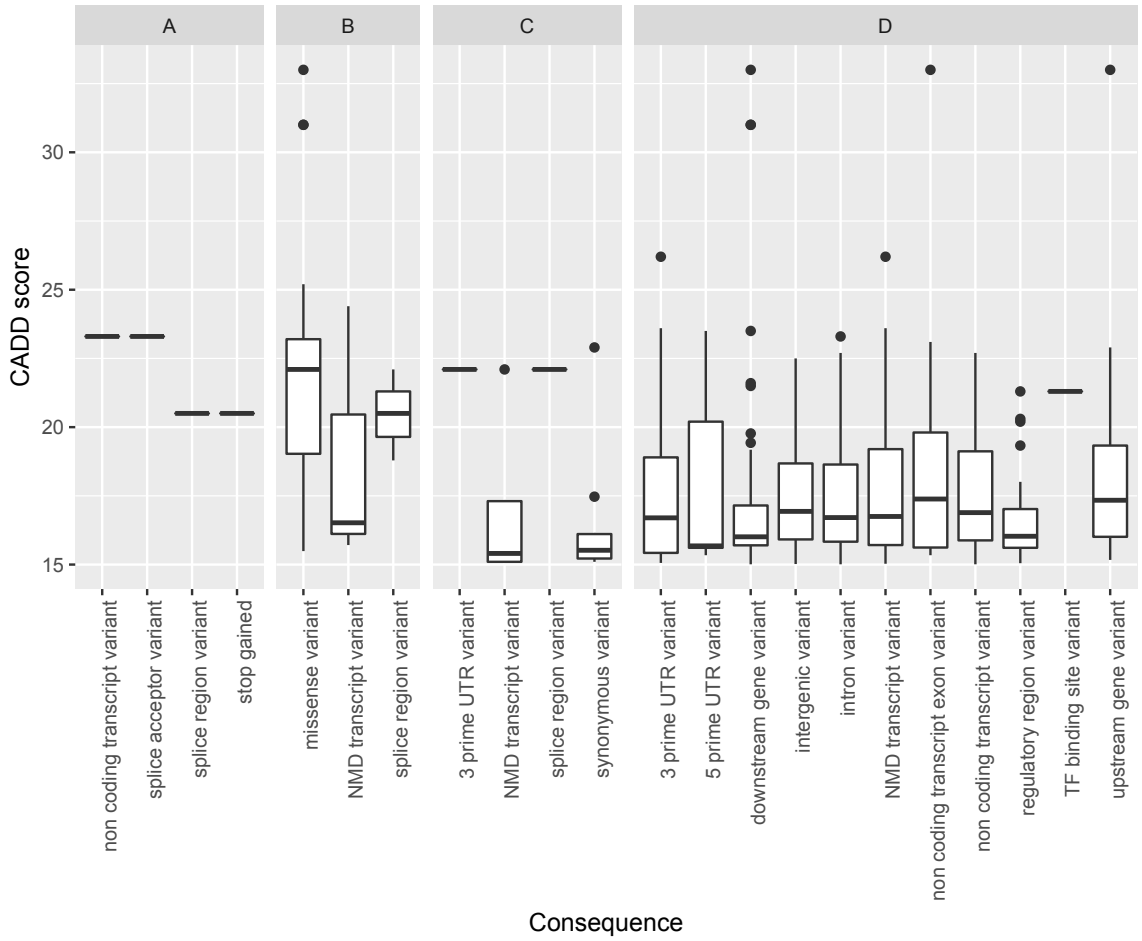


Fig. 3.3 Comparison of CADD scores and VEP impact categorizations (defined by VEP as "subjective impact classification of consequence type"). VEP-defined consequences are on the x-axis, and the data is separated into facets by impact. The four impact levels are "high", "moderate", "low", and "modifier", denoted by the labels A, B, C, and D, respectively. "NMD" is an abbreviation for nonsense-mediated decay, "UTR" for untranslated region, and "TF" for transcription factor.

Population-specific significant SNPs

In the following sections, a subset of the significant SNPs from the DIND and CADD filter will be presented along with the derived allele frequency in each population, the VEP-predicted consequence ("Cons."), the CADD score, the gene or genes associated with the SNP, and the window-based test in which the given SNP was in the top one percent of results. The full list of significant SNPs for each population can be found in Appendix C. SNPs were chosen for presentation in the following sections (in Tables 3.9 through 3.21) because of immune function or other function of interest based on a literature search. The colors in the

tables are a heat map-style visual representation of derived allele frequency. SNPs that are located near each other on a chromosome are grouped together by horizontal lines, as are SNPs that were assigned to multiple genes by the VEP. This is because likely only one SNP in a given region is causing a signal of positive selection, and others around it appear in the top results due to being in linkage disequilibrium with the driver SNP. Sometimes a given SNP appears multiple times, because VEP has assigned it multiple genes or consequences. Genes will be described in the first population in which they are significant, though they may be in the top results for more than one population. Table 3.8 shows the different abbreviations that are used in the following tables.

Column	Abbreviation	Full term
Tests	T	Tajima's D
Tests	N	nSL
Tests	I	iHS
Cons.	I	intron variant
Cons.	N	non coding transcript variant
Cons.	M	missense variant
Cons.	5	5 prime UTR variant
Cons.	3	3 prime UTR variant
Cons.	SR	splice region variant
Cons.	S	synonymous variant
Cons.	D	downstream gene variant
Cons.	IG	intergenic variant
Cons.	R	regulatory region variant
Cons.	U	upstream gene variant
Cons.	NC	non coding transcript exon variant
Cons.	SA	splice acceptor variant
Cons.	TF	TF binging site variant
Cons.	NMD	NMD transcript variant
Cons.	SG	stop gained

Table 3.8 Abbreviations used in significant SNP tables. "NMD" is an abbreviation for nonsense-mediated decay, "UTR" for untranslated region, and "TF" for transcription factor.

Significant SNPs in the West and Central African population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
T	1:113841736	17.21	IG	-	0.88	0.63	0.63	0.42	0.37	0.32	0.41	0.53	0.47	0.48	0.21	0.47	0.33
T	1:113993588	21.1	I,N	MAGI3	0.83	0.60	0.48	0.48	0.46	0.68	0.24	0.18	0.18	0.22	0.34	0.29	0.31
T	11:63452100	16.76	I,N,NMD	RTN3	0.40	0.67	0.81	0.86	0.80	0.66	0.97	0.88	0.95	1.00	0.97	0.97	0.96
N	15:73981701	15.81	I,NMD	CD276	0.50	0.62	0.55	0.43	0.52	0.50	0.44	0.41	0.45	0.52	0.42	0.50	0.54
N,I	17:3627619	15.87	R	-	0.46	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N,I	17:3627619	15.87	S	GSG2	0.46	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N,I	17:3657159	20.6	NC,M	ITGAE	0.79	0.58	0.67	0.70	0.52	0.38	0.32	0.21	0.03	0.10	0.37	0.28	0.23
N,I	17:3627619	15.87	I,N,U,D,NC	ITGAE	0.46	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
T	2:95720630	15.79	D	MAL	0.73	0.75	0.77	0.73	0.74	0.63	0.56	0.66	0.56	0.70	0.97	0.72	0.63
N	4:106910958	16.81	R	-	0.69	0.71	0.78	0.81	0.72	0.79	0.88	0.57	0.52	0.70	0.16	0.53	0.80
N	4:106891531	21.5	I,D,3	NPNT	0.67	0.48	0.42	0.49	0.39	0.52	0.53	0.35	0.37	0.46	0.61	0.21	0.18
N	4:106910958	16.81	I	NPNT	0.69	0.71	0.78	0.81	0.72	0.79	0.88	0.57	0.52	0.70	0.16	0.53	0.80
N	4:106959289	21.6	D	TBCK	0.71	0.71	0.64	0.61	0.76	0.68	0.62	0.78	0.79	0.64	0.89	0.83	0.74
N,I	5:15499642	16.53	R	-	0.69	0.96	0.98	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
N,I	5:15499642	16.53	U	FBXL7	0.69	0.96	0.98	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
T	5:74637711	16.16	I,N	HMGCR	0.73	0.60	0.61	0.56	0.46	0.43	0.38	0.47	0.60	0.82	0.42	0.48	0.52
T	6:108975464	16.09	R	-	0.77	0.85	0.88	0.83	0.78	0.91	0.71	0.84	0.89	0.94	0.97	0.86	0.77
T	6:108975464	16.09	I,U	FOXO3	0.77	0.85	0.88	0.83	0.78	0.91	0.71	0.84	0.89	0.94	0.97	0.86	0.77
N	6:158932817	15.13	D	CACYBPP3	0.67	0.52	0.48	0.51	0.61	0.57	0.56	0.47	0.40	0.38	0.37	0.45	0.50
N	6:158886564	15.52	I	TULP4	0.83	0.44	0.39	0.36	0.54	0.43	0.50	0.50	0.40	0.56	0.29	0.45	0.43
N	6:158932817	15.13	D,3	TULP4	0.67	0.52	0.48	0.51	0.61	0.57	0.56	0.47	0.40	0.38	0.37	0.45	0.50
N,I	7:29130712	21.6	I	CPVL	0.79	0.50	0.48	0.56	0.61	0.52	0.41	0.41	0.39	0.32	0.63	0.28	0.31
T	7:66212597	19.2	I,NMD	KCTD7	0.60	0.37	0.36	0.42	0.28	0.29	0.32	0.29	0.06	0.14	0.71	0.29	0.48
T	7:66212597	19.2	I,NMD	RABGEF1	0.60	0.37	0.36	0.42	0.28	0.29	0.32	0.29	0.06	0.14	0.71	0.29	0.48
N	9:116293017	18.25	I,N	RGS3	0.83	0.88	0.92	0.80	0.89	0.89	0.94	0.99	0.98	1.00	1.00	1.00	1.00

Table 3.9 Derived allele frequencies of immune-related DIND-significant SNPs in the West and Central African population compared to other populations

The SNPs presented in Table 3.9 can be separated into several different broad functional categories. One of these categories is T cell function. ITGAE, or CD103, is a receptor expressed on T cells and dendritic cells and is involved in immune homeostasis in both the intestine and the airways. In the intestine, its expression or lack thereof by dendritic cells determines the ratio of effector to regulatory T cells (Annacker et al., 2005). In the airways, its expression may maintain the balance between regulatory T cells and Th2 effector cells. This balance influences asthma susceptibility, tolerance to inhaled allergens, and clearance of respiratory viral infections (Nawijn et al., 2011), such as pulmonary poxvirus in mice (Beauchamp et al., 2010). *ITGAE* is in a region of the genome that has been shown to be under selection in a Yoruban population (The International HapMap Consortium, 2005). *CD276* also has an important role in T cell stimulation, and is released from monocytes, activated T cells, and monocyte-derived dendritic cells (Zhang et al., 2008). In a study of meningitis in children, higher levels of CD276 were found to be circulating in patients with bacterial meningitis compared to aseptic meningitis or controls, and it was suggested that the

levels of circulating CD276 can be a useful tool in discerning the intensity of inflammation of the infection (Chen et al., 2009b). FOXO3 is a transcription factor that is involved in the immune response via modulation of dendritic cells and magnitude of T cell response (Dejean et al., 2009). *RGS3* is a regulator of G protein signaling protein, and is involved in T cell migration (Williams et al., 2013). *NPNT* (nephronectin) plays a role in kidney development, but is also involved in hepatitis. In mouse models, it is upregulated and attracts T cells to the liver, which causes more acute disease. It is also involved in human hepatitis (Inagaki et al., 2013).

Another broad category of function is that of activation of other immune cells and the innate immune response. In a comparative genomics study, *TULP4* was identified as a host defense gene in *C. elegans* and a regulator of cytokine production in mouse macrophages (Alper et al., 2008). FBXL7 is a ubiquitin protein ligase that interacts with members of the TLR7/9 pathway and is involved in the innate immune response (Chiang et al., 2012). *MAL* codes for a protein that is necessary for TLR4 signal transduction. This implies its importance in bacterial infections and to LPS specifically (Fitzgerald et al., 2001). RABGEF1 is a negative activator of skin inflammation and mast cell activation (Tam et al., 2004). *RTN3* has been suggested to be involved in monocyte differentiation and recruitment (Chen et al., 2011).

Viral interaction is another category of immune function represented in Table 3.9. CPVL is a carboxypeptidase that plays a role in the immune system as well as other body systems. In macrophages, it is thought that CPVL may participate in pruning antigen peptides for presentation or digestion in the lysosome (Mahoney et al., 2001). In a comparison of hepatic transcriptomes between patients with chronic hepatitis B who responded to peginterferon treatment and those who did not, *CPVL* expression was associated with patient response in hepatitis B e antigen, and higher expression was associated with loss of hepatitis B surface antigen after two years (Jansen et al., 2015). This suggests that *CPVL* plays a role in immune response to viral infection. *HMGCR* is associated with cholesterol biosynthesis that plays an important role in the pathogenesis of West Nile Virus and potentially other flaviviruses. It is upregulated in infected cells, which allows for a larger intracellular membrane platform on which the virus can replicate (Mackenzie et al., 2007). However, this gene also has other biologically important functions, such as muscle differentiation and regeneration (Mammen et al., 2011; Martini et al., 2009), and it also is the pharmalogic target of statin medications (Mammen et al., 2011).

Finally, *MAGI3* is a gene that is associated with tight junctions in the intestine. Decreased expression of *MAGI3* was associated with inflammatory bowel disease and ulcerative colitis in a Swedish population, potentially via promotion of intestinal inflammation. This suggests

a role for this gene in the relationship between the gut microbiome and immune system (Norén et al., 2017). Several SNPs in *MAGI3* have also been associated with HPV clearance in HIV-positive women (Sudenga et al., 2014).

Significant SNPs in the West Asian and Armenian population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
I	1:46493460	19.03	M,D,U	MAST2	0.31	0.50	0.45	0.41	0.50	0.32	0.62	0.60	0.71	0.50	0.42	0.69	0.52
N	14:63568154	16.38	I	KCNH5	0.25	0.63	0.64	0.58	0.61	0.63	0.62	0.47	0.60	0.78	0.76	0.33	0.27
N,I	15:48591267	15.31	I,N	SLC12A1	0.50	0.71	0.81	0.80	0.80	0.66	0.82	0.59	0.35	0.52	0.53	0.50	0.70
I	2:152536498	23.2	M	NEB	0.42	0.81	0.83	0.91	0.76	0.66	0.53	0.47	0.34	0.42	0.97	0.52	0.29
I	2:152587814	19.8	I	NEB	0.33	0.85	0.83	0.91	0.78	0.64	0.50	0.53	0.37	0.42	0.92	0.59	0.34

Table 3.10 Derived allele frequencies of immune-related DIND-significant SNPs in the West Asian and Armenian population compared to other populations

MAST2 is a signaling cofactor of TLR7/9, and has been associated via GWAS with Crohn's disease and rheumatoid arthritis, which are both autoimmune diseases. TLR7 and TLR9 are microbial pattern recognition receptors that instigate an innate immune response to infection (Chiang et al., 2012).

KCNH5 is a potassium channel gene that has been found to be associated with severe cholera infection and to have been under positive selection in a Bangladeshi population (Karlsson et al., 2013).

NEB, or nebulin, plays an important role in skeletal muscle (Horowits et al., 1986). While not related to immune function, it is included here because of its striking geographic pattern of derived allele frequency. It's interesting to note that this missense SNP has a high frequency in the Colla population, who also have a number of missense SNPs in the titin gene, which also plays an important role in skeletal muscle.

Significant SNPs in the Southwest European population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
N	15:48591267	15.31	I,N	SLC12A1	0.50	0.71	0.81	0.80	0.80	0.66	0.82	0.59	0.35	0.52	0.53	0.50	0.70
N,I	2:136407479	21.1	I,N,M,D	R3HDM1	0.02	0.06	0.48	0.52	0.41	0.14	0.12	0.09	0.05	0.06	0.05	0.03	0.00
N,I	2:136576577	16.01	I	LCT	0.10	0.25	0.56	0.58	0.59	0.34	0.59	0.50	0.50	0.60	0.26	0.36	0.49
N,I	2:136608646	17.65	I,N	MCM6	0.02	0.04	0.48	0.45	0.30	0.11	0.06	0.01	0.03	0.06	0.00	0.02	0.00
N,I	2:136740900	15.48	I,NMD,N,U	DARS	0.04	0.25	0.59	0.56	0.57	0.29	0.62	0.38	0.48	0.62	0.24	0.36	0.50
I	2:152587814	19.8	I	NEB	0.33	0.85	0.83	0.91	0.78	0.64	0.50	0.53	0.37	0.42	0.92	0.59	0.34
N,I	3:130284284	25.2	M	COL6A6	0.06	0.81	0.89	0.88	0.83	0.59	0.68	0.59	0.65	0.60	0.92	0.50	0.44
N,I	3:130271485	17.06	IG	-	0.15	0.81	0.89	0.88	0.80	0.61	0.53	0.49	0.56	0.58	0.92	0.48	0.44
I	4:140359958	17.09	U	ACA64	0.21	0.60	0.52	0.54	0.50	0.43	0.65	0.47	0.66	0.52	0.95	0.57	0.37
N	8:16369596	19.39	I,N,U	MSR1	0.50	0.87	0.88	0.92	1.00	0.91	0.97	0.99	0.95	1.00	0.95	0.97	0.97

Table 3.11 Derived allele frequencies of immune-related DIND-significant SNPs in the Southwest European population compared to other populations

R3HDM1 is underexpressed in patients with celiac disease, which suggests a role in inflammation and autoimmunity and therefore perhaps also immune response (Garcia-Etxebarria et al., 2016). However, the SNP represented here (rs1446585 or 2:136407479) in *R3HDM1* is in linkage disequilibrium with rs4988235, which is correlated with lactase persistence (Goodrich et al., 2016). rs1446585 is also associated with levels of Bifidobacterium in the gut microbiome, however, this correlation may also be due to confounding by the lactase persistence SNP (Goodrich et al., 2016). It is possible that the signature of selection seen in this SNP is driven by the adaptive advantage of lactase persistence.

MSR1, or macrophage scavenger receptor1, binds to extracellular double-stranded RNA and transports it into the cell so it can be bound to by TLR3, which then sets off the innate immune response to viral infection via activation of interferons (Dansako et al., 2013; DeWitte-Orr et al., 2010).

Significant SNPs in the Northeast European population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
N,I	1:215308203	15.05	I,NMD	KCNK2	0.33	0.48	0.64	0.78	0.67	0.29	0.68	0.71	0.71	0.72	0.74	0.57	0.49
N,I	10:64024417	15.98	I	RTKN2	0.21	0.52	0.59	0.59	0.67	0.54	0.85	0.93	0.94	0.94	0.97	0.84	0.86
N,I	10:64024417	15.98	R	-	0.21	0.52	0.59	0.59	0.67	0.54	0.85	0.93	0.94	0.94	0.97	0.84	0.86
I	13:26906088	15.92	I	CDK8	0.12	0.63	0.58	0.42	0.52	0.57	0.29	0.28	0.50	0.50	0.47	0.33	0.22
I	18:34664093	18.03	3,NMD,I,M,NC	KIAA1328	0.21	0.85	0.61	0.64	0.63	0.71	0.85	0.88	0.81	0.58	0.74	0.79	0.83
N,I	2:136407479	21.1	I,N,M,D	R3HDM1	0.02	0.06	0.48	0.52	0.41	0.14	0.12	0.09	0.05	0.06	0.05	0.03	0.00
N,I	2:136576577	16.01	I	LCT	0.10	0.25	0.56	0.58	0.59	0.34	0.59	0.50	0.50	0.60	0.26	0.36	0.49
N,I	2:136608646	17.65	I,N	MCM6	0.02	0.04	0.48	0.45	0.30	0.11	0.06	0.01	0.03	0.06	0.00	0.02	0.00
N,I	2:136740900	15.48	I,NMD,N,U	DARS	0.04	0.25	0.59	0.56	0.57	0.29	0.62	0.38	0.48	0.62	0.24	0.36	0.50
I	3:114594139	15.28	I,N	ZBTB20	0.33	0.90	0.81	0.90	0.91	0.91	0.97	0.93	0.94	1.00	0.89	0.98	0.84
I	3:114594139	15.28	I,N	ZBTB20-AS3	0.33	0.90	0.81	0.90	0.91	0.91	0.97	0.93	0.94	1.00	0.89	0.98	0.84
I	4:151781297	16.74	I	LRBA	0.08	0.90	0.95	0.89	0.80	0.71	0.62	0.72	0.73	0.94	0.66	0.64	0.56
N	7:33671662	16.47	IG	-	0.21	0.56	0.70	0.75	0.61	0.43	0.56	0.22	0.29	0.44	0.34	0.28	0.30
N	7:33627588	16.28	I,NMD,N	BBS9	0.06	0.60	0.75	0.75	0.72	0.61	0.85	0.50	0.60	0.80	0.97	0.57	0.46

Table 3.12 Derived allele frequencies of immune-related DIND-significant SNPs in the Northeast European population compared to other populations

One broad category of genes shown in Table 3.12 are those that have been associated with immune disease. *KCNK2*, or *TREK1*, codes for a potassium channel that regulates immune cell entry into the central nervous system. Knockout mice for this gene experienced more severe experimental autoimmune encephalomyelitis (a mouse model for MS). Therefore, this gene has been suggested as a potential drug target for diseases involving malfunction of the blood-brain barrier (Bittner et al., 2013). This channel is opened by polyunsaturated fatty acids and lysophospholipids (Maingret et al., 2000; Patel, 1998). *KIAA1328* is in a chromosomal region that has been associated with type 1 diabetes and thus potentially plays a role in the immune response (Cooper et al., 2009). *RTKN2* is involved in T cell activation and was upregulated in response to a vaccine for enterovirus 71, which causes hand-foot-and-mouth disease (Liu et al., 2013b). It is also associated with rheumatoid arthritis in a Japanese population (Myouzen et al., 2012).

Another category of genes are those associated with innate immune response to infection. *CDK8* is responsible for regulation of over forty percent of interferon-γ -responsive genes, via phosphorylation of STAT1. This suggests that it plays an important role in the antiviral interferon immune response (Bancerek et al., 2013). ZBTB20 is a transcriptional repressor that promotes activation of TLR signaling. Knockout mice for this gene were less susceptible to endotoxin shock and sepsis caused by *E. coli* (Liu et al., 2013c). Mutations in *LRBA*, the gene coding for an LPS-responsive anchor protein, have been associated with an immune deficiency characterized by B cell activation defects and impaired autophagy (Lopez-Herrera et al., 2012).

The selection signal seen in the windows containing the genes *LCT* and *MCM6* (two adjacent 200 KB windows) are both likely due to the well-known selection for lactase tolerance (see above section on the Southwest European population). Neither of these are immune-related genes, but are included here because they are well-known signals of positive selection.

BBS9 has been identified as a leprosy susceptibility locus in a GWAS on a population with Chinese ancestry, though the mechanism for its involvement in the disease pathogenesis is unknown (Wang et al., 2016).

Significant SNPs in the Volga Uralic population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
N,I	2:136407479	21.1	I,N,M,D	R3HDM1	0.02	0.06	0.48	0.52	0.41	0.14	0.12	0.09	0.05	0.06	0.05	0.03	0.00
I	2:136740900	15.48	I,NMD,N,U	DARS	0.04	0.25	0.59	0.56	0.57	0.29	0.62	0.38	0.48	0.62	0.24	0.36	0.50
N,I	20:34023962	15.07	D	GDF5OS	0.04	0.52	0.59	0.58	0.70	0.38	0.76	0.56	0.53	0.96	0.89	0.62	0.62
N,I	20:34023962	15.07	R	-	0.04	0.52	0.59	0.58	0.70	0.38	0.76	0.56	0.53	0.96	0.89	0.62	0.62
N,I	20:34023962	15.07	I	GDF5	0.04	0.52	0.59	0.58	0.70	0.38	0.76	0.56	0.53	0.96	0.89	0.62	0.62

Table 3.13 Derived allele frequencies of immune-related DIND-significant SNPs in the Volga Uralic population compared to other populations

GDF5 is a growth differentiation factor and is involved in LPS signal transduction, thus serving as part of the innate immune response to Gram-negative bacterial infection (Triantafilou et al., 2001).

The windows containing the SNPs from the genes *R3HDM1* and *DARS* are the same two that have been represented in the SWE and ENE populations and are likely coming from the strong signal of selection on the *LCT/MCM6* loci.

Significant SNPs in the South Asian population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
I	1:11205058	18.23	S,NC	MTOR	0.02	0.71	0.69	0.75	0.72	0.59	0.62	0.82	0.82	0.76	0.74	0.84	0.72
I	1:11205058	18.23	I,N	MTOR-AS1	0.02	0.71	0.69	0.75	0.72	0.59	0.62	0.82	0.82	0.76	0.74	0.84	0.72
I	1:11322565	17.3	U	MTOR	0.08	0.71	0.64	0.71	0.67	0.61	0.59	0.79	0.81	0.76	0.71	0.78	0.79
I	1:11322565	17.3	R	-	0.08	0.71	0.64	0.71	0.67	0.61	0.59	0.79	0.81	0.76	0.71	0.78	0.79
N	1:35648316	15.1	I,NMD,D,U	SFPQ	0.65	0.96	1.00	0.96	0.80	0.80	0.53	0.40	0.39	0.48	0.24	0.47	0.46
I	1:39714238	17.81	I,N	MACF1	0.42	0.67	0.69	0.70	0.80	0.77	0.68	0.62	0.60	0.66	0.71	0.59	0.68
I	1:39643464	16.54	R	-	0.15	0.63	0.70	0.64	0.80	0.70	0.68	0.62	0.60	0.66	0.71	0.59	0.66
I	1:39643464	16.54	I,N	MACF1	0.15	0.63	0.70	0.64	0.80	0.70	0.68	0.62	0.60	0.66	0.71	0.59	0.66
T	1:92970209	15.42	D	EVI5	0.29	0.62	0.70	0.67	0.76	0.82	0.94	0.90	0.95	0.94	0.82	0.93	0.77
T	1:92875071	15.07	IG	-	0.38	0.48	0.47	0.49	0.59	0.77	0.79	0.84	0.90	0.74	0.50	0.95	0.86
T	1:93115127	16.46	I,N	EVI5	0.48	0.62	0.72	0.70	0.76	0.82	0.94	0.90	0.95	0.94	0.87	0.95	0.77
T	1:93115127	16.46	U	HMGB3P9	0.48	0.62	0.72	0.70	0.76	0.82	0.94	0.90	0.95	0.94	0.87	0.95	0.77
T	1:93165207	15.84	D	RNU4-59P	0.29	0.62	0.72	0.69	0.76	0.79	0.94	0.90	0.95	0.94	0.87	0.93	0.76
T	1:93165207	15.84	I,D	EVI5	0.29	0.62	0.72	0.69	0.76	0.79	0.94	0.90	0.95	0.94	0.87	0.93	0.76
T	1:93192134	15.11	I,N	EVI5	0.29	0.62	0.72	0.69	0.76	0.77	0.94	0.90	0.95	0.94	0.87	0.93	0.76
N	11:60893235	31	M	CD5	0.56	0.42	0.53	0.58	0.59	0.73	0.71	0.88	0.85	1.00	0.89	0.98	0.96
N	11:60893235	31	D	VPS37C	0.56	0.42	0.53	0.58	0.59	0.73	0.71	0.88	0.85	1.00	0.89	0.98	0.96
I	16:31099000	26.2	M,NC	PRSS53	0.00	0.00	0.00	0.00	0.00	0.57	0.00	0.00	0.00	0.00	0.00	0.05	0.00
I	16:31099000	26.2	D	VKORC1	0.00	0.00	0.00	0.00	0.00	0.57	0.00	0.00	0.00	0.00	0.00	0.05	0.00
I	16:31099000	26.2	D	ZNF646	0.00	0.00	0.00	0.00	0.00	0.57	0.00	0.00	0.00	0.00	0.00	0.05	0.00
I	2:206227030	20.5	I,N	PARD3B	0.15	0.71	0.73	0.75	0.76	0.82	0.79	0.78	0.76	0.80	1.00	0.59	0.56
I	2:206227030	20.5	I,N	PARD3B	0.15	0.71	0.73	0.75	0.76	0.82	0.79	0.78	0.76	0.80	1.00	0.59	0.56
I	2:206208049	16.17	I,U	PARD3B	0.10	0.69	0.66	0.72	0.70	0.82	0.79	0.76	0.77	0.80	1.00	0.59	0.57
I	2:206217855	15.04	I,N	PARD3B	0.15	0.71	0.73	0.75	0.76	0.82	0.79	0.78	0.77	0.80	1.00	0.59	0.57
N,I,T	4:71619607	20.7	I	RUFY3	0.40	0.90	0.92	0.96	0.96	0.88	0.94	0.90	0.85	0.44	0.84	0.83	0.82

Table 3.14 Derived allele frequencies of immune-related DIND-significant SNPs in the South Asian population compared to other populations

There are several genes shown in Table 3.14 that play a role in the immune response to viruses. *SFPQ*, or splicing factor proline/glutamine-rich, is a repressor of *IL8* transcription and plays a role in regulation of antiviral genes (Imamura et al., 2014). *PARD3B* has been associated with the rate of AIDS progression, via interaction with other genes (Troyer et al., 2011). It is also a tight junction gene and has been associated with celiac disease and ulcerative colitis (Wapenaar et al., 2007). *RUFY3* has been shown to be differentially regulated during an infection with Dengue virus, though the function of this gene in an infection is not clear (Folly et al., 2011). Generally, it is expressed in neurons and is involved in axon growth control (Wei et al., 2014). This SNP was in the top one percent of windows in each of the three selection tests.

MTOR is a kinase that plays an important role in the maturation, activation, and function of B cells, T cells, and antigen-presenting cells (Powell et al., 2012). It also plays an important role in metabolism (see review by Waickman and Powell (2012)).

EVI5 is a common retroviral integration site (Liao et al., 1995) was shown to be a risk locus for MS in a Dutch population (Hoppenbrouwers et al., 2008). The two neighboring windows shown in Figure 3.14 are likely showing the same signal from *EVI5*.

PRSS53 is a serine protease that is overexpressed in psoriatic skin (Quaranta et al., 2014; Tsoi et al., 2012). It also plays a role in hair growth and keratinization. rs201075024 (16:31099000) is associated with hair shape at the scalp and hair curliness and has been detected by the *FineMAV* method for detecting variants under positive selection for functional follow-up (Szpak et al., 2018).

rs2229177 (11:60893235) is a missense mutation in the cytoplasmic region of the *CD5* gene (Carnero-Montoro et al., 2011; Moreno-Estrada et al., 2009). This cytoplasmic region is highly conserved between species, potentially implying an important functional role for this region (Moreno-Estrada et al., 2009). *CD5* is a negative regulator of T- and B-cell receptor signaling and a pathogen-associated molecular pattern (PAMP) receptor for β-glucans, which are fungal cell wall components (Carnero-Montoro et al., 2011; Vera et al., 2009). In functional analyses, individuals who are homozygous for the derived allele at this locus show significantly stronger early and late signaling responses in response to β-glucan exposure than did ancestral homozygous individuals (Carnero-Montoro et al., 2011). The late response consisted of the release of the cytokine IL8, attracting neutrophils, which are part of the immune response to microbial (especially fungal) infections (Carnero-Montoro et al., 2011). According to the authors of this study, the geographic distribution and the functional importance of this allele suggest selection pressure from a fungal pathogen in East Asia. This selection mechanism could be either a direct regulation of interactions between CD5 and pathogens, or an indirect regulation, by regulating the T cell/B cell receptor (TCR/BCR)) response to infection (Carnero-Montoro et al., 2011). While the varied global diversity of this SNP may have been produced long ago due to selection pressure from an unknown fungal pathogen (Carnero-Montoro et al., 2011), its repercussions are relevant today in other diseases. Due to its decreased inhibition of TCR signaling compared to the derived allele, the ancestral allele is associated with stronger proliferation of T cells and higher prevalence of nephritis in autoimmune disease systemic lupus erythematosus (SLE) patients (Cenit et al., 2014). It has also been suggested that global diversity at this locus may be associated with the noted variance in severity and incidence of SLE in different ethnic groups (Cenit et al., 2014). This SNP has been found to be a candidate for selection in East Asian populations in several previous studies (Carnero-Montoro et al., 2011; Moreno-Estrada et al., 2009). It shows the highest derived allele frequencies in South and Central Asian populations, as well as in Siberian and Colla populations.

Significant SNPs in the Western Siberian population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
I	1:31504162	15.21	I,NMD,N	PUM1	0.31	0.15	0.17	0.27	0.22	0.32	0.47	0.35	0.66	0.48	0.00	0.59	0.39
I	16:71727571	15.97	D	SNORA70D	0.29	0.27	0.34	0.35	0.39	0.29	0.62	0.69	0.66	0.74	0.63	0.74	0.71
I	16:71727571	15.97	I,N	PHLPP2	0.29	0.27	0.34	0.35	0.39	0.29	0.62	0.69	0.66	0.74	0.63	0.74	0.71
I	16:71660310	15.75	R	-	0.13	0.27	0.33	0.35	0.39	0.27	0.65	0.68	0.66	0.74	0.63	0.71	0.68
I	16:71660310	15.75	M,U	MARVELD3	0.13	0.27	0.33	0.35	0.39	0.27	0.65	0.68	0.66	0.74	0.63	0.71	0.68
I	16:71760988	15.7	D	AP1G1	0.13	0.27	0.34	0.34	0.39	0.27	0.65	0.68	0.66	0.72	0.63	0.74	0.71
I	16:71760988	15.7	U	PHLPP2	0.13	0.27	0.34	0.34	0.39	0.27	0.65	0.68	0.66	0.72	0.63	0.74	0.71
I	16:71760988	15.7	R	-	0.13	0.27	0.34	0.34	0.39	0.27	0.65	0.68	0.66	0.72	0.63	0.74	0.71
I	16:71775080	15.37	I,NMD,D	AP1G1	0.13	0.27	0.34	0.35	0.39	0.27	0.65	0.68	0.66	0.74	0.63	0.74	0.71

Table 3.15 Derived allele frequencies of immune-related DIND-significant SNPs in the West Siberian population compared to other populations

The SNPs in Table 3.15 are separated into two overall loci. The first contains *PUM1*, which regulates *LGP2*. *LGP2* is a major regulator of innate immunity genes, including those involved in the interferon response (Liu et al., 2017).

The second group of SNPs includes one in the gene *MARVELD3*, which is involved in susceptibility to malaria, according to a GWAS in a Ghanaian population (Timmann et al., 2012). MARVELD3 is a tight junction protein that is expressed in endothelial cells and is thought to be involved in the pathology-causing adherence of malaria-infected erythrocytes to the endothelium of various internal organs (Gowda and Ockenhouse, 1999).

Significant SNPs in the South Siberian and Mongolian population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
T	1:93115127	16.46	I,N	EVI5	0.48	0.62	0.72	0.70	0.76	0.82	0.94	0.90	0.95	0.94	0.87	0.95	0.77
T	1:93115127	16.46	U	HMGB3P9	0.48	0.62	0.72	0.70	0.76	0.82	0.94	0.90	0.95	0.94	0.87	0.95	0.77
T	1:93106871	15.88	I,N	EVI5	0.83	0.65	0.72	0.69	0.76	0.93	0.94	0.90	0.95	0.94	0.87	0.95	0.86
T	1:93165207	15.84	D	RNU4-59P	0.29	0.62	0.72	0.69	0.76	0.79	0.94	0.90	0.95	0.94	0.87	0.93	0.76
T	1:93165207	15.84	I,D	EVI5	0.29	0.62	0.72	0.69	0.76	0.79	0.94	0.90	0.95	0.94	0.87	0.93	0.76
T	1:93192134	15.11	I,N	EVI5	0.29	0.62	0.72	0.69	0.76	0.77	0.94	0.90	0.95	0.94	0.87	0.93	0.76
N	11:60893235	31	M	CD5	0.56	0.42	0.53	0.58	0.59	0.73	0.71	0.88	0.85	1.00	0.89	0.98	0.96
N	11:60893235	31	D	VPS37C	0.56	0.42	0.53	0.58	0.59	0.73	0.71	0.88	0.85	1.00	0.89	0.98	0.96
I	15:63953153	16.03	I	HERC1	0.13	0.12	0.17	0.19	0.39	0.32	0.59	0.85	0.92	0.80	0.79	0.83	0.86
I	15:64012859	15.69	I,N	HERC1	0.13	0.23	0.27	0.22	0.43	0.43	0.59	0.85	0.92	0.80	0.79	0.83	0.86
N,I	2:109513601	19.61	M	EDAR	0.00	0.02	0.00	0.02	0.22	0.00	0.59	0.71	0.97	0.92	0.95	0.76	0.72
I	3:134278270	20.5	SG,SR,M,I,N,D	CEP63	0.02	0.29	0.34	0.38	0.50	0.34	0.50	0.69	0.73	0.72	0.92	0.53	0.54

Table 3.16 Derived allele frequencies of immune-related DIND-significant SNPs in the South Siberian and Mongolian population compared to other populations

EDAR, while not an immune-related gene, is a well known target of selection in several Asian populations (Sabeti et al., 2007). This gene influences hair thickness, sweat glands, and tooth development (Botchkarev and Fessing, 2005; Bryk et al., 2008). According to Fumagalli et al., rs3827760 (2:109513601, the SNP putatively under selection here) shares a haplotype with multiple other SNPs that were found to be significantly correlated with helminth diversity (Fumagalli et al., 2010).

CEP63, containing rs1127826 (3:134278270), is associated with bacteria, virus, and fungi interactions according to the HPIDB. CEP63 is a centrosomal protein involved in intracellular trafficking and is part of the cellular cytoskeleton. According to Le Baron et al., many viruses target the host cytoskeleton. The targeting of CEP63 by flaviruses (which include Dengue, yellow fever, and West Nile virus) is an example of this attack strategy (Le Breton et al., 2011). According to the HPIDB, this gene also interacts with Hepatitis C virus, *Yersinia pestis*, *Bacillus anthracis*, *Francisella tularensis*, and yeast, among other pathogens (Ammari et al., 2016). However, this gene likely has other cellular functions that are not involved with immune processes.

Significant SNPs in the Central Siberian population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
I	13:26906088	15.92	I	CDK8	0.12	0.63	0.58	0.42	0.52	0.57	0.29	0.28	0.50	0.50	0.47	0.33	0.22
I	18:9255982	22.9	3,NMD,M,D	ANKRD12	0.38	0.71	0.66	0.63	0.52	0.80	0.50	0.53	0.61	0.56	0.79	0.67	0.76
T	2:17775840	15.19	I	VSNL1	0.06	0.21	0.19	0.29	0.46	0.41	0.76	0.81	0.89	0.92	0.79	0.74	0.78
I	2:179124117	18.85	I,N	OSBPL6	0.08	0.33	0.45	0.39	0.30	0.30	0.56	0.74	0.82	0.88	0.82	0.60	0.43
I	2:179049910	15.98	IG	-	0.25	0.27	0.41	0.38	0.43	0.34	0.53	0.71	0.77	0.86	0.74	0.59	0.51

Table 3.17 Derived allele frequencies of immune-related DIND-significant SNPs in the Central Siberian population compared to other populations

CDK8 is responsible for regulation of over forty percent of interferon-γ -responsive genes, via phosphorylation of STAT1. This suggests that it plays an important role in the antiviral interferon immune response (Bancerek et al., 2013).

Significant SNPs in the Northeast Siberian population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
N,I,T	1:119579276	16.11	I,N	WARS2	0.00	0.06	0.14	0.23	0.26	0.04	0.44	0.43	0.60	0.78	0.84	0.31	0.08
I	11:67220015	20.2	M	GPR152	0.00	0.00	0.02	0.00	0.09	0.00	0.24	0.13	0.13	0.40	0.03	0.17	0.10
I	11:67220015	20.2	R	-	0.00	0.00	0.02	0.00	0.09	0.00	0.24	0.13	0.13	0.40	0.03	0.17	0.10
I	11:67220015	20.2	U,NC,5	CABP4	0.00	0.00	0.02	0.00	0.09	0.00	0.24	0.13	0.13	0.40	0.03	0.17	0.10
N,I	11:67798336	15.68	I,N,NMD,5,U	NDUFS8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.56	0.00	0.00	0.00
N,I	11:67798336	15.68	U	MIR4691	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.56	0.00	0.00	0.00
N,I	11:67798336	15.68	D	ALDH3B1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.56	0.00	0.00	0.00
N,I	11:67798336	15.68	R	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.56	0.00	0.00	0.00
I	16:28513403	23.5	D	APOBR	0.06	0.19	0.25	0.35	0.41	0.21	0.56	0.37	0.35	0.56	0.74	0.10	0.04
I	16:28513403	23.5	5,M	IL27	0.06	0.19	0.25	0.35	0.41	0.21	0.56	0.37	0.35	0.56	0.74	0.10	0.04
T	5:98107341	18.03	R	-	0.00	0.40	0.47	0.54	0.57	0.34	0.65	0.68	0.74	0.88	0.87	0.66	0.73
T	5:98107341	18.03	I,N,D,U	RGMB	0.00	0.40	0.47	0.54	0.57	0.34	0.65	0.68	0.74	0.88	0.87	0.66	0.73
T	5:98107341	18.03	I,N	RGMB-AS1	0.00	0.40	0.47	0.54	0.57	0.34	0.65	0.68	0.74	0.88	0.87	0.66	0.73

Table 3.18 Derived allele frequencies of immune-related DIND-significant SNPs in the Northeast Siberian population compared to other populations

WARS2 has been shown to have been under positive selection in a Greenlandic Inuit population. The time of selection is thought to be during the peopling of the Americas, because the selected haplotype is present in some East Asians and high proportions of Native Americans. The selected haplotype is thought to be an example of adaptive introgression from Denisovans. *WARS2* is associated with body fat distribution, as well as many other traits, thus the exact phenotype under selection is not known (Racimo et al., 2016). The SNPs shown in Figure 3.18 were in the top one percent of windows in each of the three selection tests.

RGMB, or repulsive guidance molecule b, is involved in promotion of respiratory immunity and respiratory tolerance in the lung, via expansion of T cells to the antigen (Xiao et al., 2014).

GPR152 is a G-protein coupled receptor with unknown function. However, it is about thirty percent identical with two proteins (CRTH2 and FPRL1, (Elagoz et al., 2004; Nagata and Hirai, 2003), respectively) that are both involved in inflammation.Therefore, it is possible that this gene is also involved in inflammatory processes (Gloriam et al., 2005).

IL27 is included in the list of immune phenotypes for virus, bacteria, innate immunity, and T cells based on the GO term ontology searches. IL27 is a cytokine that is formed of two subunits, EBI3 and IL27p28. Mandatory coexpression of these two subunits for the formation of IL27 prevents out of control systemic inflammation. The receptor of IL27 is WSX1, which is expressed on naive, effector, regulatory, and memory T cells, as well as NK cells, macrophages, and mast cells. This variety shows the wide influence of IL27 (Hunter,

2005). IL27 is produced in response to certain pathogen stimulants as well as host immune signals such as interferon-γ (Hunter, 2005; van de Vosse et al., 2009).

IL27 has both pro- and anti-inflammatory effects. This is a common trait of cytokines and allows them to influence the immune system in nuanced ways (Hunter, 2005). The pro-inflammatory effects of IL27 have been shown in cancer and autoimmune studies (Hunter, 2005). In several murine cancer studies, the increased IFN-γ production, MHC expression, and cytotoxic T cell response caused by *IL27* overexpression contributed to increased tumor clearance (Chiyo et al., 2005; Hisada et al., 2004; Salcedo et al., 2004). Evidence from rodent studies has also shown that by targeting IL27 with antibodies, the severity of rodent arthritis and multiple sclerosis models can be lessened (Goldberg et al., 2004a,b). In humans, IL27 has a pro-inflammatory effect on monocytes. It does this by activating expression of *STAT1* and its target genes, causing a bigger release of inflammatory cytokines by monocytes upon TLR exposure, and causing a reduction in the production and function of IL10 (Kalliolias and Ivashkiv, 2008; Murray et al., 2009). Since IL10 is anti-inflammatory and is involved in monocyte deactivation, inhibiting IL10 has pro-inflammatory effects. However, this pro-inflammatory role for IL27 in monocytes is short-lived and self-limiting, lasting for only around 30 minutes after TLR-ligand binding. This is because the production of gp130, one of the subunits of the IL27 receptor, is inhibited by the presence of TLRs (Kalliolias and Ivashkiv, 2008).

Despite its involvement in pro-inflammatory processes, IL27's most important role seems to be limiting the innate and adaptive immune response. Multiple animal studies have demonstrated that animals deficient in the IL27/WSX-1 pathway, though able to mount normal or more effective than wild type early immune response to pathogen stimulus, suffer severe or lethal pathogenic effects from an inability to downregulate the initial immune response (Hamano et al., 2003; Hölscher et al., 2005; Villarino et al., 2003). The absence of WSX1 on other immune cell types such as NK cells and macrophages leads to increased levels of inflammatory cytokines compared to the wild type (Artis et al., 2004a; Villarino et al., 2005; Yamanaka et al., 2004). As in mice, IL27 has an anti-inflammatory effect on human T cells through promotion of IL10 and inhibition of IL17. Upon activation by IL27, CD4+ T cells stimulate the differentiation of Tr1 (T regulatory type 1) cells, which play an important role in self-tolerance (Murugaiyan et al., 2009; Pot et al., 2011). Tr1 cells produce IL10, which suppresses T cell response and the release of inflammatory cytokines (Murray et al., 2009; Murugaiyan et al., 2009). In addition to its promoting effect on IL10, IL27 also has an inhibitory effect on the pro-inflammatory IL17 (Murugaiyan et al., 2009). IL17 has been found to be involved in various human autoimmune diseases (Murugaiyan et al., 2009).

In mouse studies, IL17-deficient mice are partially protected from experimental models of autoimmune disease (Komiyama et al., 2006).

IL27 also has antiviral functions. It inhibits replication of influenza A and hepatitis C viruses by phosphorylation of *STAT* genes and antiviral factor protein kinase R (Frank et al., 2010; Liu et al., 2012). It inhibits the replication of HIV by inducing interferon-inducible genes in macrophages (Imamichi et al., 2008).

rs181206 (16:28513403) in IL27 has been associated with susceptibility to asthma and inflammatory bowel disease (Chae et al., 2007; Li et al., 2009; Zicca et al., 2014).

Significant SNPs in the Colla population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
I	1:161876448	19.92	I	ATF6	0.29	0.08	0.06	0.15	0.15	0.09	0.35	0.34	0.60	0.62	0.76	0.29	0.33
I	1:161825850	18.35	I,N	ATF6	0.29	0.08	0.06	0.15	0.15	0.07	0.35	0.34	0.60	0.62	0.76	0.29	0.33
I	1:161806356	15.35	I,U	ATF6	0.12	0.04	0.06	0.14	0.13	0.07	0.35	0.32	0.60	0.62	0.76	0.29	0.33
N	1:36638206	33	M,U,NC,D	MAP7D1	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.61	0.00	0.00
I	3:121500699	22	3,NMD,M	IQCB1	0.19	0.21	0.20	0.25	0.33	0.45	0.24	0.18	0.24	0.22	0.89	0.17	0.22
N	3:121908434	16.53	I	CASR	0.00	0.19	0.27	0.16	0.17	0.18	0.29	0.29	0.21	0.36	0.82	0.34	0.20
T	3:138191232	23.6	3,NMD,M,U,D,NC	ESYT3	0.02	0.42	0.48	0.37	0.43	0.54	0.44	0.56	0.60	0.58	0.87	0.43	0.49
T	6:119215902	18.07	R	-	0.04	0.23	0.22	0.22	0.26	0.36	0.38	0.37	0.61	0.58	0.82	0.33	0.34
T	6:119215402	15.34	R	-	0.15	0.44	0.38	0.35	0.39	0.52	0.53	0.62	0.73	0.64	0.82	0.62	0.53
T	6:119215902	18.07	I,N	ASF1A	0.04	0.23	0.22	0.22	0.26	0.36	0.38	0.37	0.61	0.58	0.82	0.33	0.34
T	6:119215402	15.34	5,NC	ASF1A	0.15	0.44	0.38	0.35	0.39	0.52	0.53	0.62	0.73	0.64	0.82	0.62	0.53
T	6:119215902	18.07	I,U	MCM9	0.04	0.23	0.22	0.22	0.26	0.36	0.38	0.37	0.61	0.58	0.82	0.33	0.34
T	6:119215402	15.34	I,NC	MCM9	0.15	0.44	0.38	0.35	0.39	0.52	0.53	0.62	0.73	0.64	0.82	0.62	0.53
N,I	6:159178345	24.7	M	SYTL3	0.21	0.54	0.36	0.36	0.41	0.36	0.62	0.76	0.73	0.98	0.89	0.72	0.68
I	6:51746757	19.15	I	PKHD1	0.21	0.27	0.28	0.49	0.41	0.52	0.41	0.68	0.66	0.60	0.89	0.72	0.76
I	6:51628094	15.81	I	PKHD1	0.02	0.33	0.41	0.43	0.39	0.52	0.53	0.60	0.66	0.70	0.71	0.81	0.67
I	6:51691632	15.66	I	PKHD1	0.27	0.38	0.47	0.49	0.43	0.63	0.47	0.68	0.74	0.74	0.87	0.79	0.68
I	6:51758405	15.38	I	PKHD1	0.15	0.23	0.28	0.49	0.41	0.52	0.41	0.68	0.66	0.60	0.89	0.72	0.76
N	6:51914956	18.79	R	-	0.15	0.44	0.53	0.58	0.57	0.34	0.41	0.37	0.34	0.38	0.89	0.31	0.48
N	6:51914956	18.79	M,SR	PKHD1	0.15	0.44	0.53	0.58	0.57	0.34	0.41	0.37	0.34	0.38	0.89	0.31	0.48

Table 3.19 Derived allele frequencies of immune-related DIND-significant SNPs in the Colla population compared to other populations

Several genes shown in Table 3.19 are involved with immune response to viral infection. *ATF6* is involved in West Nile Virus replication via the inhibition of innate immune response and the promotion of cell survival (Ambrose and Mackenzie, 2012). ASF1A is a histone chaperone and participates in DNA replication and repair (Mousson et al., 2006). It has been shown to play a role in the human antiviral immune response through regulation of interferon-β (Liu et al., 2016b). IQCB1 is included in the HPIDB because it has been shown to experimentally co-immunoprecipitate, or physically bind to, with the PA protein of the Influenza A virus. Over 1,000 human genes are involved in an influenza infection (Watanabe et al., 2014).

CASR is a calcium-sensing receptor that has been implicated in a mouse model of colitis. Knockout mice for this gene had compromised intestinal barrier functions, different makeup of the intestinal microbiome, and were prone to inflammation in the intestine (Cheng et al., 2014).

SYTL3 is also known as SLP3, or synaptotagmin-like protein 3. According to the HPIDB, it is known to interact with bacterial pathogens. Along with Rab27a and kinesin-1, it is part of a complex that plays a role in transporting lytic granules of cytotoxic T cells (CTLs) to the immune synapse (where the CTL meets the target cell). CTLs defend against cells infected by viruses, as well as cancerous cells. They act by attaching to the target cell and sending out lytic granules, which contain perforin and other cytotoxic molecules such as granzymes (Kurowska et al., 2012). This perforin helps granzymes to enter the target cell, which triggers an apoptotic cascade and causes the target cell to die (Dieckmann et al., 2016).

rs148608573 (1:36638206) shows a striking difference in derived allele frequency in the COL population compared to all others. *MAP7D1* is a gene involved in cytoskeletal function that is not referenced much in literature pertaining to immunity response, but it has been shown to demonstrate alternative splicing after T cell activation (Martinez and Lynch, 2013). This SNP also was found to be a high frequency allele having undergone pre-admixture selection in the admixed American population by the FineMAV paper (Szpak et al., 2018).

Significant SNPs in the Mainland Southeast Asian population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
N,I	17:64052744	18.75	I	CEP112	0.27	0.54	0.41	0.31	0.52	0.59	0.65	0.75	0.79	0.86	0.95	0.88	0.89
N,I	2:109513601	19.61	M	EDAR	0.00	0.02	0.00	0.02	0.22	0.00	0.59	0.71	0.97	0.92	0.95	0.76	0.72
N	3:121635230	17.75	I,N,D	SLC15A2	0.52	0.35	0.48	0.34	0.46	0.32	0.44	0.49	0.58	0.34	0.05	0.71	0.52
N	3:121664112	15.82	D	SLC15A2	0.15	0.35	0.48	0.33	0.46	0.32	0.44	0.49	0.56	0.34	0.05	0.72	0.61
N	3:121643804	15.49	M,U,D	SLC15A2	0.52	0.35	0.48	0.33	0.46	0.32	0.44	0.49	0.56	0.34	0.05	0.72	0.52
N	3:65521324	19.42	I,N,U	MAGI1	0.04	0.13	0.42	0.50	0.39	0.41	0.32	0.54	0.45	0.28	0.32	0.66	0.57
N	3:65536944	16.4	I,N,U	MAGI1	0.06	0.15	0.47	0.49	0.52	0.38	0.41	0.59	0.52	0.34	0.34	0.69	0.51
N	3:65464712	15.33	I,N,U	MAGI1	0.06	0.17	0.47	0.50	0.41	0.38	0.44	0.57	0.50	0.68	0.42	0.72	0.61

Table 3.20 Derived allele frequencies of immune-related DIND-significant SNPs in the Mainland Southeast Asian population compared to other populations

SLC15A2, or PEPT2, is a solute transporter that recognizes and moves peptides, including bacterial peptides such as those present in the cell wall of Gram-negative bacterias, across the lung and other epithelial tissues, activating the innate immune response. This transport is proton-dependent, and the inflammatory acidification during a local infection may enhance the efficacy of this process (Swaan et al., 2008).

MAGI1 has been shown to induce promotion of interferon-β , via activation of IRF3 during an influenza infection. This suggests it plays a role in the immune response to viral infection (Kumar et al., 2012).

Significant SNPs in the Island Southeast Asian population

Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
N,I	2:109513601	19.61	M	EDAR	0.00	0.02	0.00	0.02	0.22	0.00	0.59	0.71	0.97	0.92	0.95	0.76	0.72
N	2:44004010	15.71	S,NMD,M,I,N,5	DYNC2LI1	0.10	0.48	0.59	0.70	0.61	0.52	0.74	0.71	0.74	0.88	0.32	0.76	0.80
N	2:44004010	15.71	U	RN7SKP66	0.10	0.48	0.59	0.70	0.61	0.52	0.74	0.71	0.74	0.88	0.32	0.76	0.80
T	9:108037717	15.06	I,NMD	SLC44A1	0.27	0.85	0.86	0.85	0.78	0.79	0.68	0.72	0.66	0.50	0.58	0.88	0.87

Table 3.21 Derived allele frequencies of immune-related DIND-significant SNPs in the Island Southeast Asian population compared to other populations

DYNC2LI1 plays an important role in retrograde intraflagellar transport (IFT) and functions in cilia biogenesis (Kessler et al., 2015; The UniProt Consortium, 2015). Among its GO biological processes are antigen processing and presentation of exogenous peptide antigen via MHC class II, cilium assembly, and ER to Golgi vesicle-mediated transport (The UniProt Consortium, 2015). Retrograde IFT is the process by which proteins are carried from the tip of the cilia back to the base. DYNC2LI1 interacts with DYNC2H1 to form the dynein-2 complex, which is necessary for retrograde IFT complexes. Knocking down this gene results in reduced cilia length and altered cilia morphology, which are both suspected to affect cilia function. This causes a number of severe skeletal and other diseases, since cilia play important roles in development and organ function in growing and adult humans and are present in most cells (Kessler et al., 2015). MHC molecules are assembled in the endoplasmic reticulum, after which they are loaded into vesicles for transport to the Golgi apparatus and the plasma membrane (Hewitt, 2003), so this is how it might relate to the immune process. *DYNC2LI1* has also been reported to be upregulated in patients with psoriasis, lichen planus, and atopic dermatitis, all of which are autoimmune diseases of the skin (Li et al., 2013). This reinforces an association with immune processes suggested by its antigen processing gene ontology term.

3.3 Results of population differentiation-based analysis: d_i

As described in Section 2.3, d_i is a population differentiation statistic that is a function of F_{ST} between a given population and all other populations. This means that it is best used to detect local signals of selection that are not shared with other populations (Akey et al., 2010).

Per population, each gene was assigned a d_i score based on the highest-scoring SNP in that gene, allowing for a ranked list for enrichment analysis and listing the top scoring genes per population.

3.3.1 Enrichment for immune function in genes with top d_i scores

In each population, genes were divided into the top one percent of d_i scores and the bottom ninety-nine percent of d_i scores (correcting for gene size, see Section 2.4.2) in order to see whether the top one percent of genes were enriched for immune genes compared to neutral expectations. The results of these enrichment tests are shown in Table 3.22.

This table shows significant enrichments spread fairly evenly over the different categories of immune genes, with more enrichments in the innate immunity and antigen processing and presentation categories. The enrichments are spread between populations as well, with the Central Siberian population being the only population without any enrichments.

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
exp	2.26	8.33	4.21	1.86	5.96	2.05	0.31	15.72	24.1	0.05	0.04
AFR	1	9	3	2	7	2	0	17	30	0	0
WAA	1	7	5	2	1	6	1	8	21	0	0
SWE	3	8	9	4	6	2	0	12	28	0	0
ENE	3	8	5	3	8	4	0	13	23	0	0
VOL	6	4	2	0	8	2	1	14	23	0	0
SOA	2	6	5	2	11	1	1	15	25	0	0
WSI	0	9	4	0	7	1	0	11	21	0	0
SSI	3	5	5	0	10	4	1	16	26	0	0
CSI	1	9	2	0	2	2	0	15	24	0	0
NSI	2	7	4	3	6	5	1	12	14	0	0
COL	1	5	4	1	9	4	1	13	21	0	0
SEM	3	12	8	2	6	2	0	19	19	0	0
SEA	2	6	3	2	5	3	0	12	20	0	0

Table 3.22 Enrichment of top one percent d_i results in each population for immune genes based on GO DB and HPIDB compared with expected counts. The "exp" row gives the expected number of immune genes in any given one percent of the data, and the counts in each cell represent the number of immune genes of a given category in the top one percent of the results. Light orange indicates significant enrichment at p = 0.05, dark orange at 0.01 in a Fisher's exact test with correction for multiple testing, as described in Chapter 2.

A brief overview of the HLA

Because HLA genes are present in top selection (both positive and balancing) hits in many populations, it is worth pausing for a condensed description of these genes.

HLA (human leukocyte antigen) genes are found in the MHC (major histocompatibility complex), a region on chromosome 6 that contains a high concentration of polymorphic immune genes. Genes in this region play a crucial role in the human adaptive immune response and are also implicated in a large number of human diseases such as autoimmune disease, as well as other processes such as maternal/fetal survival during pregnancy. In general, the job of HLA genes is to present antigenic peptides to T cells in order for the immune system to distinguish whether that peptide is self or non-self and mount an immune response as appropriate. There are three classes of HLA genes, and the first two classes come up most frequently in the results of this project. Class I HLA genes are *HLA-A*, *-B*, *-C*, *-E*, *-F*, and *-G*, *MICA*, and *MICB*. Class I *HLA* genes are expressed on most nucleated cells, and are recognized by CD8+ T cells or NK cells. Class II HLA genes are largely expressed on antigen processing and presentation cells, and are recognized by CD4+ T cells. These consist of *HLA-D* genes, such as *HLA-DQA1* and *HLA-DRB2*, as well as other antigen processing and presentation genes such as *TAP1* and *TAP2*. As mentioned above, the majority of–though not all–HLA genes are highly polymorphic, with over 15,000 class I and II alleles found so far. Because of this, most individuals are heterozygous at MHC loci. High levels of polymorphism are maintained by balancing selection, and in principle give a population potential protection against a wide range of pathogens (see Janeway et al. (2001), review by Dendrou et al. (2018), and review by Jeffery and Bangham (2000), and (Rajagopalan and Long, 1999)).

Table 3.23 shows a subset of the genes responsible for driving significant signals of enrichment, with a description of relevant genes below. Several genes have been described in previous sections. The full version of Table 3.23 is in Appendix B.

	Pop.	Genes
GO.Bact	SWE	IL6, STAT1
	VOL	EPPIN, EPPIN-WFDC6, FCER1G, P2RX7, TLR1
	SEM	FZD5, IL6, TNFSF8
GO.Virus	AFR	APOBEC3H, CPSF4, DDX5, DUOX2, FKBP8, GRB2, IFNL1
	WSI	ACY3, CXCL9, UNC93B1
	SEM	IL6, LTBR
GO.Tcell	WAA	IFNG, LILRB2
	SWE	IL4, IL6, LAG3, SPNS2
	SEM	FZD5, HLA-DPA1, HLA-DPB1, IL4, IL6, RAB27A, TNFSF8
GO.Bcell	SWE	IL4, IL6, SPNS2
	ENE	IL6, SPNS2
	NSI	CD180, RAG1
GO.Innate	ENE	AGO1, IRAK2, TLR1, TLR6, TLR10
	SOA	ADCY9, C1QB, CD247, FCN3, NLRC4, TRIM32
	WSI	CORO1A, UNC93B1
	SSI	AGO1, AGO3, C1QBP, HCK, LILRA5, NRG1
	COL	ADRBK1, C1S, NLRX1, TYRO3
	SEM	IL4, TRAFD1
GO.APP	WAA	AP1S3, IFNG, LILRB2
	SSI	IFNG, LILRB2
	NSI	HLA-DQB1, HLA-DRA, PSMA7
	SEA	HLA-DPA1, HLA-DPB1, HLA-F
GO.Adapt	WAA	IFNG
	VOL	DUSP10
	SSI	IFNG
	NSI	RAG1

Table 3.23 Subset of immune-related genes driving significant d_i enrichment signals

In the GO.Bact category, *STAT1* is important for the expression of interferon gamma-induced genes in bacterial infection (Varinou et al., 2003). EPPIN and WFDC6 are part of a group of proteinase inhibitor genes that is found at various levels of conservation between mammals and is though to play a role in the innate immune response and have antimicrobial effects (Clauss et al., 2005). EPPIN also plays a role in sperm motility (Yenugu et al., 2004). FCER1G is a component of the immunoglobulin-E receptor. Immunoglobulin-E is a mediator of allergic response (Garman et al., 2000). *P2RX7* is involved in antimicrobial innate immune response (Gavala et al., 2013). *FZD5* contributes to the growth cycle of cells that manufacture antimicrobial peptides in the small intestine (SCHMAUssER et al., 2004; van Es et al., 2005). *TNFSF8* is thought to play a role in the regulation of immune activation (Collette et al., 2003).

In the GO.Virus category, *APOBEC3H* is a member of the APOBEC3 cytidine deaminase gene family that has innate immune activity, through causing hypermutation, against retroviruses. It it thought to have been under positive selection at certain times throughout primate evolution. While *APOBEC3H* in other primates has strong antiviral effects, the human version is expressed at low enough levels that it shows no antiviral activity. Two potential reasons are given for this lack of function: the retrovirus *APOBEC3H* evolved to combat went extinct and decreased the evolutionary pressure on the gene, or another member of the *APOBEC3* gene family now performs antiretroviral duties and human *APOBEC3H* is no longer required to be functional. Under relaxed evolutionary pressure, it may be beneficial to have fewer hypermutation-causing cytidine deaminase genes for the sake of the host cell (OhAinle et al., 2006). CPSF interacts with the NS1 protein of the influenza A virus (Nemeroff et al., 1998). *FKBP8* is involved in hepatitis C virus replication (Okamoto et al., 2006). IFNL1, or interferon-lambda 1 or IL29, is induced by viral infection and plays a role in the antiviral innate immune response (Kotenko et al., 2002). ACY3 is a hepatitis C virus core binding protein (Chen et al., 2009c). CXCL9 is a chemokine that is part of the innate antiviral immune response (Thapa et al., 2008). LTBR, or lymphotoxin beta receptor, is involved in epithelial cell innate immune response (Wang et al., 2010b).

In the T cell category, *IFNG* codes for interferon gamma, which is produced upon activation of the innate immune response and stimulates inflammation and macrophage action (Murray et al., 2009). LILRB is a leukocyte immunoglobulin-like receptor and is involved in control of proliferation of T cells (Brown et al., 2004). *LAG3* is associated with lymphocyte activation and has been implicated in susceptibility to autoimmune disease (Zhang et al., 2005). *SPNS2* has been linked with humoral response to immunization (Nijnik et al., 2012). RAB27A is part of a complex that plays a role in transporting lytic granules of cytotoxic T cells (CTLs) to the immune synapse (referenced with SYTL3 in the Colla section of DIND-significant SNPs above) (Kurowska et al., 2012).

In the B cell category, *CD180* is important in B cell response to LPS (Ogata et al., 2000). *RAG1* is involved in VDJ recombination, in which lymphocytes generate a large range of binding specificities (Oettinger et al., 1990).

In the innate immune category, *ADCY9*, or adenylyl cyclase type 9, is a gene that codes for a protein that produces cyclic AMP (cAMP). Among its other functions, cAMP is produced by T regulatory cells and plays a role in the downregulation of immune function. Through this relationship, polymorphisms in this gene have been significantly associated with markers of allergic reaction (Teixeira et al., 2017). C1QB is part of the C1q protein, which enables the binding of the C1 complex to the antigen-antibody complex on the cell surface (Murray et al., 2009). *NLRC4* is involved in the detection of bacterial components (Miao et al., 2010b).

TRIM32 is involved in a range of physiological processes, and one of them is in cellular defense against influenza virus (Fu et al., 2015). COROA1 is a coat protein that is involved in *Mycobacterium leprae* survival during infection (Tanigawa et al., 2009). C1QBP is the binding protein of C1QB. HCK is a tyrosine kinase that is involved in the innate immune response to bacteria (Ernst et al., 2002). LILRA5 is a receptor that plays a role in leukocyte activation and has been associated with autoimmune disease (Borges, 2002; Mitchell et al., 2008). C1s is another member of the C1 complex in the complement system (Murray et al., 2009). NLRX1 is located on the outer membrane of the mitochondria and is involved in the antiviral immune response (Moore et al., 2008). TRAFD1, or FLN29, is a negative regulator of TLR immune response (Mashima et al., 2005).

In the antigen processing and presentation category, *APIS3* has been associated with the viral innate immune response via interaction with TLR3 signaling (Setta-Kaffetzi et al., 2014). TLR3's ligand is double stranded RNA from viruses (Murray et al., 2009). *LILRB2*, like *LILRA5*, plays a role in regulation of T cell activity (Brown et al., 2009). PSMA7 is a negative regulator of antiviral innate immunity (Jia et al., 2009).

In the adaptive immunity category, *DUSP10*, or *MKP5*, plays various roles in adaptive and innate immune responses that can lead in some cases to better resistance to autoimmune disease or harmful immune overreaction (Zhang et al., 2004).

3.3.2 Top-scoring genes per population based on the d_i statistic

In addition to looking at enrichment for various immune functions in the top d_i results, it is also interesting to look at the highest scoring genes in each population. The 50 highest scoring genes in each population are shown in Tables 3.24 and 3.25, with genes in the immune gene lists presented in Chapter 2 highlighted in orange. As referenced in Chapter 2, these results are divided into five equally-sized bins by gene length, so as not to skew results in favor of larger genes.

Bin	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1	DARC	SIGLEC10	MAPK7	C1orf216	FCER1G	PRSS53	VAMP5	PRSS53	PCBD1	NDUFS8	PTPRCAP	FAM109B	LGALS3BP
1	KRT13	ZBTB6	C1orf216	CCNI2	PRSS53	SEPHS2	FEN1	C1orf216	PRSS53	HLA-DRA	RPS6KB2	CYP2D6	FGF17
1	APOF	C5orf27	PRSS53	MAPK7	NR1I3	ZNF768	TMEM150A	NANOG	NFKBIA	B3GNT1	CORO1B	FAM109A	B3GALT4
1	PKDREJ	IZUMO1	KCNAB3	PRSS53	SPINT3	BCKDK	C2orf68	IFNG	PRSS8	MRGPRF	RABGGTA	DKFZP779L1853	NR0B2
1	RPL13A	KLK4	GHRL	CYB561D2	EFNA1	PPIB	RNF181	LBX2	MOGAT3	TRIM10	RAD9A	RPLP0	TAAR2
1	AQP5	MAMSTR	PRDM13	SMCP	HOXD4	SNX22	C19orf48	C19orf71	AP1S1	HLA-DQB1	SSH3	LSMEM2	FEN1
1	ID1	ASCL3	TMEM133	ASPRV1	C1orf216	PRSS8	KLK1	TYSND1	ENO3	OR7E24	TBX10	NPRL2	RNASE9
1	LRG1	AKIP1	CCNI2	ATP5H	WFDC6	VKORC1	SRM	FBXL22	FAM187B	C17orf49	CCDC153	TUSC2	OR1S1
1	MRPS7	BTBD18	CHST5	CYP1A1	MAPK7	FASLG	SHISA3	PRSS8	C4orf6	RNASEK-C17orf49	RCE1	ZMYND10	OAZ3
1	TPPP3	FGF11	IL4	YBX2	TTC24	ID3	MEN1	LILRB2	ZNF768	C3orf27	DYNLT1	MAMSTR	PTGER1
2	CYP26B1	ZSWIM8	KDM6B	NCDN	NDUFS2	PRSS36	TNFSF15	SPIB	MPHOSPH10	MRPL21	TGM1	SIRPB2	HLA-DPA1
2	AP5M1	NDST2	GSG1	OAF	ADAMTS4	STX1B	NOL6	PLA2G2F	MCEE	ALDH3B1	C1RL	PDYN	HLA-DPB1
2	MAP1LC3B	ITGB1BP1	NAGK	LAX1	CRYBB2	KAT8	UNC119B	STX1B	NAGK	TCIRG1	RHOD	HLA-DPA1	RP1
2	GCNT6	RETNLB	NMS	TLR10	PFDN2	ZNF668	VAMP8	OGN	ALDOA	CRISP3	PITPNM1	ADH1B	PDYN
2	E2F8	ZBTB26	OAF	ZNF441	SYNPO2L	STX4	FOXO6	RRP7A	NAT16	CRISP2	ANKRD13D	RRP7A	HLA-F
2	PLIN4	AMOTL2	C17orf61-PLSCR3	LETMD1	GABARAPL1	ZNF646	GGCX	TNFSF15	DDA1	HSD17B13	ADRBK1	MYBBP1A	LAIR2
2	ARPC1B	TNK1	TMEM256-PLSCR3	TNFSF18	DUT	ZNF785	KLK15	LETMD1	TNK1	LTK	AZGP1	KARS	WDR46
2	PDAP1	RASIP1	CDC42BPG	RASSF1	NAGK	DUPD1	MLANA	NCDN	TMEM11	SELL	CARNS1	TSPO	CTU1
2	SDHAF2	PIK3IP1	SFPQ	MFSD2A	NAA20	TRMT13	COL9A2	ZNF35	TNFSF15	GATA2	NLRX1	PRKAG3	CCNJ
2	SLC25A19	SEC11C	CHST6	CHST6	FAM151A	JMJD7-PLA2G4B	MORN4	ANKS4B	CABP7	MAP1LC3B2	DEXI	RPL6	BOD1
3	BEND4	SEC24C	SLC45A2	SLC45A2	SLC45A2	MTFMT	CABP1	SLC24A5	C9orf91	IGHMBP2	PEX5	ILDR1	ADH7
3	NAA30	SULT6B1	SLC24A5	SLC24A5	SLC24A5	SPG21	C9orf91	NOL8	BPIFB1	C5orf47	HCLS1	NSFL1C	SDR42E2
3	ATP1A1	SLC24A5	LSMD1	TESPA1	ALDH4A1	KIAA0101	SLC24A5	POLD1	UGT2A3	BBS1	UBXN2B	ZNF793	TEF
3	SLC6A9	CELA3B	TATDN2	TATDN2	SLC22A9	SMAD7	SRD5A1	ECM2	TMEM201	ZDHHC24	EMG1	ADAT1	IL10RB
3	GPRC5C	TECPR1	TYRP1	TLR6	UGT2A3	ETV7	FADS1	ASPN	TNFSF8	CFB	CAPN1	SLC19A3	SLC24A5

Table 3.24 Top 50 genes by d_i score part 1, correcting for gene size by assigning genes according to length to 5 bins of equal size and taking the top 10 genes from each. Orange highlights indicate the gene belongs to one or more of the immune gene categories.

Bin	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
3	PTP4A3	DEPDC1	ROPN1	LMTK3	NPL	GPR111	MYRF	PSMB2	CHMP4B	ADSL	PILRA	TRAFD1	POLR3E
3	NECAB2	SGOL1	CMBL	PSMB2	TESPA1	GPR115	TMEM258	TFAP2E	HAVCR1	SGSM3	ZCWPW1	BRAP	TIE1
3	SS18L1	PDCL	CLEC1A	TFAP2E	ZNF600	CEP55	C1orf50	CSRNP2	OR7A5	RHAG	STRA6	SLC24A5	ATP1B1
3	GTSE1	EPCAM	PSMB2	MYEF2	ROPN1	SLC24A5	NSUN2	C4orf50	CBX1	SERPINA12	SDCBP	ROPN1	MYRF
3	ANXA5	GPR139	TFAP2E	ROPN1	PON3	FADS1	B3GNT2	TM4SF4	PNPLA3	DPP3	SLC26A3	SEMA3F	TMEM258
4	SYT7	NARG2	POU2F3	TLR1	CCDC138	EDAR	FADS2	EDAR	ITGAE	CPT1A	GOLGB1	NXNL2	EEF2K
4	CHRM5	PLCH2	ITGA6	STK35	EDAR	PIK3CD	KLRD1	IARS	INO80C	SPNS3	MED15	HADHB	C16orf52
4	LRRC4B	ANXA2	KIAA1467	TGM3	GCC2	ITGAL	GFRAL	CCDC138	EDAR	CHKA	SLC8B1	NAA25	VWA3A
4	GRB2	GABRR1	DRD2	EDAR	RNF182	ZCCHC14	SPSB1	CLSPN	SGPL1	SUV420H1	C1S	HADHA	PAMR1
4	EXOC5	NDRG4	KCNMB4	C7orf63	SHCBP1L	TNIP3	STX7	AGO4	SIAH1	GALNTL5	NSMAF	WBP2NL	MTTP
4	UNC5A	CAMK2G	DFNA5	CCDC14	SLC5A10	S100Z	GPR124	KIF3C	LRIG2	RTF1	DNAJC5B	TMEM116	IL17RD
4	DNAJC5B	AP1S3	ZNF28	NXNL2	LMCD1	MAPKBP1	EDAR	SIAH1	NF2	CPEB4	CCDC38	TTLL1	BLNK
4	ERN1	A2ML1	CCDC14	POU2F3	KPNA4	ANKDD1A	ATP2C2	GINS1	CCDC138	RAG1	EPT1	GCN1L1	FAM129B
4	ATG13	ADAM17	EDAR	CCDC138	KIF3C	EFHD1	CLN6	PAIP2B	LRP8	OR5V1	ARHGAP5	ACAD10	C6orf10
4	RRBP1	SIAH1	SLC5A10	FAM114A1	TLR1	SNX1	MED1	SYT10	STAP1	C11orf74	ELFN2	IQCB1	CDC5L
5	FAM189A1	LPHN2	DNAH2	HERC2	BNC2	KCNK2	SLC9A9	PPP2R2C	UNC13C	PDE4B	ADCY3	OCA2	CTNNA3
5	ABCA12	GLB1	OCA2	KCNS3	CCDC60	PHKB	NEDD4L	ASXL2	OCA2	PPP6R3	ZNF280D	LRIG1	RBFOX3
5	VAV2	MMP2	C20orf112	CRIM1	HERC2	PBX1	CREB5	RBFOX1	NKAIN2	LRP5	OPRM1	SLC25A26	GPR75-ASB3
5	CD226	GPR158	ARHGAP24	PRDM10	ITPR1	NCKAP5	CMTM8	TENM4	AKAP6	MYO3B	CCDC141	TLE1	MYLK
5	TBC1D30	CNTN5	PRDM16	BNC2	SUMF1	OCA2	AVEN	XXYLT1	SH2D4B	ARHGEF26	TTN	FHIT	FMN1
5	RTTN	MDGA2	PCDH15	RAP1GAP2	UACA	CSNK1G1	MFHAS1	TBL1XR1	TECRL	TNR	TCF7L2	FHOD3	FGF14
5	AVEN	RORA	ASIC2	NFASC	EYS	WARS2	SEMA4D	SLC24A3	ASXL2	RBFOX3	ABCC4	RYR3	FMN2
5	CAMTA1	RABGAP1	SMC6	OPCML	WWOX	MAP7	PC	ZNF638	FGGY	ATXN1	OSBPL10	MACROD2	ARL15
5	DLG2	MICAL2	PPM1L	ZMYM4	DIRC3	FAM120A	DPYS	RAPGEF1	CTNNA2	ZNF385B	TIAM2	RBM19	PDE1C
5	PDE7B	SEZ6L	KCNMA1	CNTN3	ASXL2	LSAMP	STPG2	CENPP	EBF1	CAMTA1	TCF12	RAB10	CHN1

Table 3.25 Top 50 genes by d_i score part 2, correcting for gene size by assigning genes according to length to 5 bins of equal size and taking the top 10 genes from each. Orange highlights indicate the gene belongs to one or more of the immune gene categories.

A subset of these top genes are presented below, based on whether a gene has a potentially important immunological role based on a literature search. The genes highlighted in orange are classified as immune-related genes based on the GO terms listed in Chapter 2. Table 3.26 shows the counts of how many times a given immune-related class of genes appears in the top 50 genes for that population. This can be used as a rough guide for comparing representation of immune gene classes between the classes themselves and between populations. Table 3.26 shows that no class of immune gene, save for HP.Bact and HP.Virus (which could be to do with the relative size of those gene categories), are obviously represented at a higher level in Tables 3.24 and 3.25. However, many of the highlighted genes do not have clear immunological function, and a good number of non-highlighted genes do appear to play an important role. Therefore, the orange highlights help to show the concentration of immune-related genes in the top results, but should not be used as a simple classifier. The descriptions of these genes are organized by the population in which they appear and grouped by general immune function. Some populations have more genes with clear immunological function than others.

	Gene Ontology DB						HPI DB			
	Bact.	Virus	Tcell	Bcell	APP	Adapt.	Bact.	Virus	Prot.	Amoe.
AFR	0	2	1	0	0	0	5	7	0	0
WAA	0	1	1	1	2	0	2	3	0	0
SWE	1	2	1	1	1	0	4	3	0	0
ENE	2	3	3	1	1	0	3	5	0	0
VOL	2	0	1	0	2	0	8	8	0	0
SOA	0	0	3	1	0	1	4	7	0	0
WSI	0	1	0	0	0	0	1	8	0	0
SSI	1	2	2	0	4	1	3	6	0	0
CSI	1	4	2	0	1	0	5	7	0	0
NSI	0	0	3	1	2	1	5	3	0	0
COL	0	3	1	0	1	0	4	6	0	0
SEM	0	3	1	0	1	0	8	6	0	0
SEA	1	1	3	1	3	0	3	3	0	0

Table 3.26 Counts of number of genes in each immune gene category and population in the top 50 genes based on d_i score per population

West and Central Africa

The highest scoring gene in Bin 1 of the West and Central African d_i results is *DARC*. DARC, or the Duffy blood group antigen, is the red blood cell receptor for *Plasmodium vivax*. Individuals without this receptor are resistant to malaria (Horuk et al., 1993; Miller

et al., 1976). *Plasmodium vivax* infects around 75 million people per year, largely in tropical countries. Its prevalence means that it could potentially exert a measure of selective pressure on genes with which it interacts (review by Allison (2009)). The *DARC* gene contains SNPs known to be some of the most allele frequency differentiated SNPs in the HapMap dataset, as well as those predictive for low white blood cell and neutrophil count in individuals with African ancestry (Reich et al., 2009).

Two high scoring genes are involved in maturation and differentiation of immune cells. LRG1 is expressed in neutrophils and involved in granulocyte differentiation (O'Donnell et al., 2002). In mice, *VAV2* knockout mice experience deficiencies in the humoral immune response and B cell maturation. VAV2 is a guanine nucleotide exchange factor (Doody et al., 2001).

CYP26B1 and ARPC1B are both involved in the development of inflammatory disease. CYP26B1 is an enzyme that degrades retinoic acid, and when expressed prevents the expression of CCR9 on T cells. Knockout mice for this gene are protected from intestinal inflammation caused by T cells (see review by Raverdeau and Mills (2014)). ARPC1B is part of an actin-related complex and is expressed in the blood. Deficiencies in this gene cause predisposition to inflammatory disease, such as eczema, eosinophilia, and elevated levels of IgA and IgE, through platelet abnormalities (Kahr et al., 2017).

Two genes are involved with the immune response to LPS. PTP4A3 is a protein tyrosine phosphatase that plays a role in LPS reaction. Mice treated with an analogous version of this protein were protected from severe host reactions to LPS treatment (Tang et al., 2010). MAP1LCDB, or LC3B, is an autophagic protein that is also involved in the immune response to LPS (He and Klionsky, 2009; Nakahira et al., 2010).

CD226 is a protein involved in intercellular adhesion and white blood cell signaling. Because of its expression in a range of immune cells (NK cells, monocytes, T cells, and B cells), it is likely to be involved in a range of immune responses (Shibuya et al., 1996). One study showed the importance of CD226 in the regulation of T-cell activation by demonstrating that mice treated with anti-CD226 resulted in delayed onset and reduce severity of the mouse model of MS (Dardalhon et al., 2005). This gene has been associated with risk for type 1 diabetes and MS, as well as possibly autoimmune thyroid disease and rheumatoid arthritis (Hafler et al., 2008).

West Asia and Armenia

The top scoring genes in the West Asian and Armenian populations with clear immune function based on the literature can be divided into several categories.

One of these categories are genes involved in immune response to pathogens. *RETNLB*, or resistin like beta, has been shown to be a part of the immune defense against helminths (Artis et al., 2004b; Herbert et al., 2009). TNK1 is an antiviral host protein that works with interferon-stimulated genes to suppress the life cycle of hepatitis C virus (Ooi et al., 2014). *TECPR1* plays a role in the targeting of pathogenic bacteria for selective autophagy (Ogawa et al., 2011). *ANXA2* is involved in various cellular activities and has also been associated with HCV virus production (Backes et al., 2010; Gerke et al., 2005).

Another category of genes in top results is that of T cell function. SIGLEC10 is an inhibitory receptor that is involved in T cell homeostasis and has anti-inflammatory effects when acting in concert with other molecules (Bandala-Sanchez et al., 2013; Chen et al., 2009a). PIK3IP1 is a transmembrane protein that is expressed on T cells and is thought to be an inhibitor of T cell activation (DeFrances et al., 2012).

Several genes are also involved in inflammation and the balance between immune activation and inactivation. *AP1S3* knockouts result in increased IL1 signaling and IL36a expression, which results in skin inflammation (Mahil et al., 2016). This gene plays a role in autophagosome formation (Guo et al., 2012). ADAM17 is a metalloproteinase that cleaves cell surface proteins and is involved in inflammation, immune response, and tissue regeneration (see review by Scheller et al. (2011)). MMP2 is a matrix metalloproteinase that is thought to be involved in leukocyte recruitment and repair of inflammation and has various pro- and anti-inflammatory roles in the immune system (see review by Parks et al. (2004)). RORA is one of two retinoic acid receptors that play a role in Th17 cell differentiation and therefore the balance between autoimmune disease and immune response (Yang et al., 2008).

SLC24A5 is a well known target of selection in European populations and is associated with skin pigmentation (Voight et al., 2006a). Selection for skin pigmentation has been driven by different levels of ultraviolet radiation in different environments, and has generally led to darker sin at low latitudes to protect skin from radiation and lighter skin at higher latitudes, potentially for vitamin D synthesis (review by Fan et al. (2016)).

Southwest Europe

Several top scoring genes in the Southwest European population were involved in immune function. *IL4* is involved in the Th2 immune response (Murray et al., 2009). It has also been shown to have high levels of differentiation in the promoter region, suggesting this gene has been under positive selection in different populations (Rockman et al., 2003). CLEC1A is a C-type lectin receptor and is expressed on the cell surface of dendritic cells and endothelial cells (see review by Sancho and Reis e Sousa (2012)). *POU2F3* plays an important role in the function of tuft cells, which are located in the intestine and involved in immune defense

against parasites. Knockout mice have defective immune response to helminths (Gerbe et al., 2016).

Several genes appearing in the top results are well known for being involved in hair phenotype and skin pigmentation. PRSS53 is a protease that plays a role in hair growth and keratinization and has been associated with hair shape at the scalp and hair curliness. It has also been detected as being under selection in an East Asian population by the CMS (Composite of Multiple Signals) statistic (Adhikari et al., 2016). *OCA2* is associated with pigmentation, especially in East Asian populations. The derived allele of interest here is thought to have come into existence about six thousand years ago (Murray et al., 2015). The *OCA2* gene is associated with pH levels within the melanosome (Liu et al., 2013a) and has been under positive selection in East Asian and European populations (Hider et al., 2013; Lao et al., 2007).

Northeast Europe

Several of the top scoring genes in the Northeast European population fall into the category of T cell function. *TESPA1* plays a role in T cell signaling, selection, and maturation (Wang et al., 2012a). TNFSF18, or GITRL, is a cytokine involved in the innate immune reaction and T cell activation (see review by Shevach and Stephens (2006)). LAX1, or LAX, is a negative regulator of T and B cell signialing (Zhu et al., 2005).

Other genes play a role in the function of other immune cells and barriers. ASPRV1, an aspartic retroviral-like protease, is expressed by neutrophils and plays a role in the progression of acute inflammation to chronic, causing diseases like multiple sclerosis and experimental autoimmune encephalomyelitis (Hawkins et al., 2017). MFSD2A is a regulator of blood brain barrier function (Ben-Zvi et al., 2014).

TLR1, *TLR6*, and *TLR10* form a cluster of genes that has been previously shown to be a target of selection in European populations (Barreiro et al., 2009). All are members of the Toll-like receptor family of pathogen recognition receptors that play an important role in the innate immune response to pathogen infection.

HERC2 is in the same genomic region as *OCA2* (from the top results in the Southwest European population) and is also involved in hair and eye pigmentation (Donnelly et al., 2011; Sturm et al., 2008).

Volga Uralic

FCERIG is one of the top scoring genes in the Volga Uralic population. It encodes the gamma subunit of FCERI, which is the receptor for immunoglobulin E. This receptor is

upregulated in antigen-presenting cells of individuals with atopic dermatitis, underlining its role in allergic response (Liang et al., 2011).

South Asia

Several of the top scoring genes in the South Asian population are involved in immune response and tolerance. SMAD7 is an inhibitor of TGF-β 1 signaling. TGF-β 1 is thought to be part of the mechanism that keeps T cells from attacking gut bacteria and prevents development of inflammatory gut disease (see review by Monteleone et al. (2004b)). It is also induced by interferon-γ and upregulated in gastritis caused by *Helicobacter pylori*, preventing TGF-β 1 from mitigating the immune response that causes tissue damage (Monteleone et al., 2004a). *PIK3CD* is involved in allergic and immune response. Knockout mice were protected from anaphylactic response, and from various types of cancer as well (Ali et al., 2004, 2014). STX4 is a member of the SNARE family of proteins, which are involved in vesicle and membrane function in various body systems, including the immune system (Stow et al., 2006). Expression of STX4 is upregulated upon exposure to LPS (Pagan et al., 2003) and can be found at the immunological synapse of some T cells (Das et al., 2004). PRSS8 modulates insulin sensitivity in the liver via interaction with TLR4 (Uchimura et al., 2014). Its expression is also upregulated in psoriatic skin (Tsoi et al., 2012).

ITGAL is a subunit of the LFA-1 (lymphocyte-associated antigen-1) integrin (Zhang et al., 2002). This molecule is expressed on white blood cells and plays a role in leukocyte differentiation and maintenance (Meli et al., 2016).

KCNK2, or *TREK1*, is a top scoring gene in the South Asian population, and it also contains a DIND-significant SNP in the Northeast European population. It is a potassium channel that regulates immune cell entry into the central nervous system and is involved in the function of the blood brain barrier (Bittner et al., 2013).

West Siberia

Several of the top scoring genes in the West Siberian population play roles in the immune response to infection by bacteria and other pathogenic organisms. SPSB1 is regulated by TLR actions and plays a role in the induction and regulation of NO (reactive nitrogen species) production. NO kills invading pathogens and is an important part of the host immune response (Lewis et al., 2011). KLRD1, or CD94, is a receptor expressed on NK cells and some T cells which can act as an activator or inhibitor, and likely plays a role in the survival of those cells and therefore immune response to pathogenic organisms (Gunturi et al., 2004). MFHAS1 is a regulator of TLR-dependent signaling and has been found at elevated levels in the blood of

individuals with sepsis (Ng et al., 2010; Zhong et al., 2015). SEMA4D is a semaphorin that is expressed in T cells. It expressed at lower levels in B cells, though expression is upregulated by LPS and CD40 antibody. Knockout mice have deficient antibody responses (see review by Suzuki et al. (2007)).

TNFSF15 and *TMEM258* are both involved in immune function in the intestine and bowel. TNFSF15 is a member of the tumor necrosis factor superfamily, and has been associated with inflammatory bowel disease in European and Japanese populations (Yamazaki et al., 2005). TMEM258, part of an oligosaccharyltransferase complex, is associated with intestinal homeostasis. Mice deficient in this gene suffered from intestinal pathology (Graham et al., 2016). *TMEM258* also appears in the top results for the Island Southeast Asian population.

South Siberia and Mongolia

Two of the top scoring genes in the South Siberian and Mongolian population are involved in the innate immune response to infection. *IFNG* is the gene for interferon gamma, which is produced upon activation of the innate immune response and stimulates inflammation and macrophage action (Murray et al., 2009). CSRNP2 is a transcription factor that is induced by lipopolysaccharide and is involved in cellular stress response or apoptosis (Eichelbaum and Krijgsveld, 2014). This suggests it plays a role in bacterial infection, though it is unclear exactly what that role is.

GINS and *LILRB2* play roles in development of immune cells. *LILRB2* plays a role in regulation of T cell activity (Brown et al., 2009). Deficiencies in GINS lead to developmental delays as well as deficiencies in neutrophils and NK cells (Cottineau et al., 2017).

Central Siberia

Two of the highest scoring genes in the Central Siberian population are involved with immune function in the lungs and airways. *NFKBIA* is a negative regulator of NF-kappa B and variants in this gene have been associated with different outcomes of childhood lung diseases such as asthma and respiratory syncytial virus infections (Ali et al., 2013; Hayden and Ghosh, 2008). BPIFB1, or bactericidal/permeability-increasing fold-containing B1, is secreted from epithelial cells in the upper respiratory tract (Bingle et al., 2010; Shum et al., 2013). It is thought to play role in innate immune response to Gram-negative bacterial infections based on its structural similarity to other well known innate immune response genes (Bingle, 2002). It has also been implicated in the immune response to *Vibrio cholerae*, and its expression is upregulated in the airways of patients with cystic fibrosis (Bingle et al., 2012; Shin et al., 2011).

HAVC1R (human hepatitis A virus cellular receptor 1) plays a role in both the immune response to the hepatitis A virus and allergic asthma (Feigelstock et al., 1998; McIntire et al., 2003). It has been shown to have been under positive selection during primate evolution and balancing selection in human populations (Kosiol et al., 2008; Nakajima et al., 2005). This balancing selection is thought to be a good example of the tradeoff between adaptation to pathogenic organisms and chronic disease (in this case, hepatitis A virus and asthma, respectively) (Sironi and Clerici, 2010).

Several other genes are associated with immune surveillance, activation, and function. SH2D4B and its homolog SH2D4A are T-cell-specific adapter proteins with relatively unknown functions. This lack of knowledge is potentially due to significant overlap in functions between these two proteins and others in the the T cell adapter protein family (Lapinski et al., 2009). *TNFSF8* is thought to play a role in the regulation of immune activation (Collette et al., 2003). SGPL1, or sphingosine 1-phosphate-lyase, is an enzyme responsible for the degradation of S1P (sphingosine 1-phosphate) (Saba et al., 1997). The gradients of S1P mediated by the effects of SGPL1 play a role in the exit of lymphocytes from lymphoid organs for the purpose of immune surveillance and function (Schwab, 2005). *EBF1* plays an important role in early B cell differentiation into different lineages (Vilagos et al., 2012).

Northeast Siberia

Several top scoring genes in the Northeast Siberian population play a role in the immune response to bacteria. CFB, or complement factor B, is part of the alternative complement pathway, which plays a role in the human immune response against polysaccharide-encapsulated bacteria, such as those involved in pneumococcal and meningococcal infections. Deficiency in this protein can cause recurrent infections of these types (Slade et al., 2013). PDE4B is a phosphodiesterase that plays a role in immune activation by LPS. Mice deficient in this gene experienced a strong reduction in induction of TNF-alpha in response to LPS stimulus, indicating a decreased sensitivity to LPS in the knockout mice (Jin and Conti, 2002). TCIRG1 modulates cell pH and plays a role in *Mycobacterium tuberculosis* infection (Lafourcade et al., 2008; Singh et al., 2006a).

Several other genes are involved in antigen processing and presentation. *HLA-DRA* and *HLA-DQB1* are both members of the class II MHC. *HLA-DQB1* has been associated with susceptibility to leprosy/tuberculosis in a Cambodian population, as well as multiple autoimmune diseases (see review by de Bakker et al. (2006)). *RAG1*, or recombination activating gene 1, takes part in initiating the process of VDJ gene recombination, which is responsible for the manufacture of a diverse set of antigen receptors in lymphocytes (Bassing

et al., 2002). Mutations in this gene can cause defective assembly of antigen receptors and lead to immunodeficiency disorders (Brauer et al., 2016).

MRGPRF is a G protein-coupled receptor that is part of a family of GPCRs that are linked to nociception and is expressed in sensory neurons (Dong et al., 2001). In a mouse model of intestinal inflammation, it has been shown to be downregulated in enteric neurons (Avula et al., 2011). In a study on human macrophage differentiation, it has been shown to be significantly upregulated in macrophages compared to monocytes. It is upregulated during the monocyte to macrophage differentiation process, and downregulated by LPS in human macrophages derived from monocytes. However, not much is known about this protein (Hohenhaus et al., 2013).

GATA2 is another top scoring gene that plays a role in the immune system. It is important in the regulation of haematopoiesis and mutations in this gene can cause immune disorders such as autoimmunity as well as increased incidence of infection, among other disorders (Collin et al., 2015).

CPT1A has previously been found to be under positive selection in this same Northeast Siberian population included in the EGDP dataset (Clemente et al., 2014). This gene has been under selection because of the role it plays in adapting to a cold environment or a diet high in fat (Clemente et al., 2014).

Colla

Two of the top scoring genes in the Colla population are involved with immune response to viruses. RABGGTA is a transferase activity marker that was significant in a study of genes associated with adaptive immunity following a H1N1 influenza A vaccine. It was significantly associated with both hemagglutination inhibition and virus neutralization antibody (Ovsyannikova et al., 2016). NLRX1 acts as a negative regulator of antiviral response in the mitochondria (Moore et al., 2008).

Two genes are involved in immune function in epidermal cells. *TGM1* is upregulated in psoriatic skin and is associated with aberrant organization and differentiation of epidermal cells (Kulski et al., 2005). ABCC4 is an ATP-binding cassette transporter that is expressed in dermal and epidermal dendritic cells, and plays a role in their migration as part of the immune response (van de Ven et al., 2008).

Several other top scoring genes play a role in immune response and inflammation. Variants in the TCF7L2 promoter have been associated with Crohn's disease (Koslowski et al., 2009). OPRM1 is an opioid receptor gene that has an influence on secretion of proinflammatory cytokines from peripheral immune cells (Matsunaga et al., 2009). C1S is one of three components of the first complement component, the binding of which to

immune complexes initiates the classical pathway innate immune response (see review by Fujita (2002)). ADRBK1, also known as GRK2, is a GPR kinase 2 that is associated with downregulation of chemokine receptors and is induced by activation of TLR2 (Alves-Filho et al., 2009). When this gene is inhibited by IL33, the chemokine CXCR2 sends a strong signal to attract neutrophils to the site of infection. This was shown with a mouse model of sepsis, in which mice treated with IL33 showed more efficient clearance of bacteria (Alves-Filho et al., 2010; Mantovani et al., 2011).

Mainland Southeast Asia

Several of the top scoring genes in the Mainland Southeast Asian population were involved susceptibility to and immune response to pathogenic organisms. *RPLP0* is upregulated upon bacterial challenge (Arbibe et al., 2006), and RAB10 is a GTPase that plays a role in the activation of TLR4 in response to LPS (Wang et al., 2010a). *TRAFD1*, or *FLN29*, has an inhibitory role in the innate immune response. Knockout mice for this gene were hyperresponsive to LPS, and were more likely to develop septic shock (Sanada et al., 2008). This gene also appeared in the iHS enrichment table in the Volga Uralic population. *HLA-DPA1* has been associated with clearance of hepatitis B in the Han Chinese (An et al., 2011). Variants in *MYBBP1A*, a NF-kappaB transcription repressor, have been associated with risk of developing pulmonary TB in a Han Chinese population (Cai et al., 2012; Owen et al., 2007). KARS secretion induces immune response via activation of macrophages (Park et al., 2005).

Island Southeast Asia

Two HLA genes appear in the top scoring genes in the Island Southeast Asian population. Like other HLA genes, *HLA-F* is involved in antigen processing and presentation. It is also involved in the surveillance of stressed cells by the innate immune system (Garcia-Beltran et al., 2016), and is expressed in the tonsil, spleen, and thymus (Lepin et al., 2000). *HLA-DP* has been associated with hepatitis B in Asian populations (Kamatani et al., 2009).

Several other genes are involved in immune activation. LGALS3BP, or 90K, has been shown to stimulate the host immune response, including by cells such as natural killer cells, and induce the production of IL2 (Ullrich et al., 1994). IL10RB is a co-receptor involved in the activation of cytokines IL10, IL22, IL28, and IL29 (Kotenko et al., 2000, 1997; Sheppard et al., 2002; Xie et al., 2000). Mutations in this gene have been associated with inflammatory bowel disease, enterocolitis, and hepatitis B infection (Begue et al., 2011; Frodsham et al., 2006; Glocker et al., 2009). POLR3E is part of the cytosolic RNA polymerase III complex

and functions in an immune capacity by sensing non-self DNA in the cytosol (Lee et al., 2012).

On the other hand, *LAIR1* and *IL17RD* are negative regulators of immune function. LAIR1, or leukocyte-associated Ig-like receptor-1, is a collagen receptor with inhibitory action on immune activation (Lebbink et al., 2006). IL17RD, or IL17 receptor D, is a negative regulator of TLR response and is also involved in neutrophil recruitment (Mellett et al., 2015, 2012).

Another gene, *BLNK*, or B cell linker protein, is involved in B cell development and function, via linkage of B cell receptor-associated kinases with different signaling pathways (Fu et al., 1998).

Chapter 4

Results: Balancing selection

4.1 Introduction

Balancing selection is another important mode of selection, and is the process by which diversity is maintained at a certain locus because the diversity is more adaptive than one allele being at high frequency. Whereas positive selection acts directionally (for example, specific surface antigens on host cells (Van Blerkom, 2003)), balancing selection acts on regions where diversity is adaptive. While resistance to some diseases is conferred by one particular allele, in some cases disease resistance is best maintained by genetic diversity. This diversity can result in heterozygote advantage or resilience against a greater variety of diseases than would be possible with less diversity (review by Karlsson et al. (2014)). For this reason, immune-related genes are a functional class of genes that may be expected to undergo balancing selection.

4.1.1 Previous evidence of balancing selection in immune genes

The most famous example of selection for diversity is in the MHC, which is a region associated with susceptibility to many different infectious and inflammatory diseases, as mentioned in Chapter 3 (Karlsson et al., 2014; Key et al., 2014). Though various studies have found different genes under balancing selection, immunity is consistently among the biological functions most strongly represented and overrepresented in results (Andrés et al., 2009; Key et al., 2014). Additionally, though many signatures of balancing selection are shared across populations, many of those that are unique are found in immune-related genes (Key et al., 2014).

In several studies, pathogen richness in a specific geographic region has been significantly correlated with diversity at immune loci in populations living in that same region (review

by Barreiro and Quintana-Murci (2010)). Blood group antigens (Fumagalli et al., 2009a), interleukins and interleukin receptors (Fumagalli et al., 2009b), and HLA class I genes (Prugnolle et al., 2005a) have all been correlated with pathogen diversity. These results suggest that the balancing selection in these genes has been pathogen-driven.

4.2 Results of the HKA test

The results from the HKA analysis are organized into two main parts. Section 4.2.1 is an enrichment analysis to show whether the top one percent of genes based on HKA score contain more immune genes than would be expected in the absence of selection. It shows which genes are present in significantly enriched windows, and how those selection signals are shared between populations. Section 4.2.2 is a further description of top HKA results in immune genes that may not have appeared in an enriched population but still warrant attention. As a reminder, the HKA test is best suited to finding signals of balancing selection younger than 250 kya (review by Sabeti et al. (2006)).

4.2.1 Enrichment of immune genes in top results

Table 4.1 shows the results for enrichment tests in each population and category of immune gene based on the results of the HKA statistic. A Fisher's exact test was used with correction for multiple hypothesis testing, as described in Chapter 2. Orange coloring indicates that significantly more immune genes appear in the top one percent of results for a given population than would be expected under neutral conditions. The expected values of how many genes per immune gene class in any one percent slice of the data is given at the top of the table. The full table of genes responsible for the enrichment signals seen in Table 4.1 can be seen in Table B.10.

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
exp	2.26	8.33	4.21	1.86	5.96	2.05	0.31	15.72	24.1	0.05	0.04
AFR	1	10	8	1	4	8	0	5	16	0	0
WAA	2	9	14	0	5	12	0	4	14	0	0
SWE	3	9	14	1	8	10	0	5	11	0	0
ENE	2	10	15	1	7	11	0	4	9	0	0
VOL	3	12	12	0	9	10	0	3	9	0	0
SOA	4	8	14	1	6	11	0	7	16	0	0
WSI	3	11	12	1	6	9	0	3	12	0	0
SSI	4	8	13	1	4	11	0	4	10	0	0
CSI	3	11	15	1	6	12	0	2	10	0	0
NSI	3	12	14	0	4	11	0	6	16	0	0
COL	3	9	12	0	6	9	1	5	10	0	0
SEM	4	7	12	0	8	9	0	4	13	0	0
SEA	2	10	10	1	7	8	0	5	16	0	0

Table 4.1 Enrichment of top one percent of HKA results in each population for immune genes based on the GO DB and HPIDB compared with expected counts. The "exp" row gives the expected number of immune genes in any given immunity category after randomly choosing one percent of all protein-coding genes, and the counts in each cell represent the observed number of immune genes of a given category in the top one percent of the results. Light orange indicates significant enrichment at p = 0.05, dark orange at p=0.01 in a Fisher's exact test with correction for multiple testing, as described in Chapter 2.

There are several categories of immune genes that show notable patterns of enrichment and lack of enrichment. Several categories show high levels of uniformity across many or all populations, potentially suggesting old signals of shared selection on the same genes. For example the T cell category and the antigen processing and presentation categories are significantly enriched in every population, largely because of genes in the MHC. For this reason, the enrichment analysis was done again after removing all genes in the MHC from the complete list of protein-coding genes considered. The MHC was defined here as anything overlapping with the region 6:28,866,528-33,775,446 (Gourraud et al., 2014). With MHC genes removed, it is possible to see which categories and populations retain a significant enrichment signal.

Table 4.2 shows the results of this enrichment analysis without MHC genes. All significant enrichments in the antigen processing and presentation category, as well as the GO.Virus

category, have disappeared. Interestingly, there are more GO.Bact and GO.Innate enrichments than in the enrichment table including MHC genes.

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
exp	2.23	8.17	3.99	1.85	5.9	1.79	0.3	15.57	23.89	0.05	0.04
AFR	2	6	2	1	5	0	0	4	14	0	0
WAA	2	4	4	0	6	0	0	3	14	0	0
SWE	3	7	5	1	9	0	0	4	11	0	0
ENE	2	7	5	1	7	0	0	3	10	0	0
VOL	3	8	3	0	9	0	0	2	10	0	0
SOA	4	5	4	1	7	0	0	6	15	0	0
WSI	3	8	4	1	6	0	0	2	11	0	0
SSI	4	7	5	2	5	0	0	4	14	0	0
CSI	4	6	4	1	8	0	0	2	12	0	0
NSI	4	7	4	0	5	0	0	7	19	0	0
COL	3	6	3	0	6	0	1	6	11	0	0
SEM	4	5	4	0	8	0	0	4	13	0	0
SEA	3	6	2	1	8	0	0	5	16	0	0

Table 4.2 Enrichment of top one percent of HKA results in each population for immune genes based on the GO DB and HPIDB compared with expected counts, with genes in MHC region removed. The "exp" row gives the expected number of immune genes in any given immunity category after randomly choosing one percent of all protein-coding genes, and the counts in each cell represent the observed number of immune genes of a given category in the top one percent of the results. Light orange indicates significant enrichment at p = 0.05, dark orange at p=0.01 in a Fisher's exact test with correction for multiple testing, as described in Chapter 2.

Because of the potentially confounding results of the MHC genes in driving enrichments, making other genes look as though they may be driving a signal of enrichment when really it is MHC genes driving the signal, the description of the enrichment table for the HKA results will be based on Table 4.2.

A subset, based on clear immunological function based on a literature search, of the genes responsible for the enrichment signals shown in Table 4.2 are shown in Table 4.3, and described in more detail below. The full table can be seen in Table B.11. This table shows that genes driving enrichment signals are largely shared between populations, even those that are widely geographically separated. However, there are no significant enrichments in the West and Central African population after the removal of MHC genes from the

analysis. Since other genes included in the enrichment table below are shared between such geographically distant populations, the pattern of sharing potentially suggests sharing between all non-African populations compared to the West and Central African population.

	Pop.	Genes
GO.Bact	SWE	DEFB1, P2RX7, PGLYRP4
	VOL	DEFB1, DMBT1, PGLYRP4
	SOA	DEFB1, DMBT1, PGLYRP4, TLR6
	WSI	DEFB1, DMBT1, PGLYRP4
	SSI	DEFB1, DMBT1, PGLYRP4, TLR6
	CSI	DMBT1, P2RX7, PGLYRP4, TLR6
	NSI	DEFB1, DMBT1, PGLYRP4, TLR6
	COL	DEFB1, DMBT1, PGLYRP4
	SEM	DEFB1, DMBT1, P2RX7, PGLYRP4
	SEA	DEFB1, DMBT1, PGLYRP4
GO.Tcell	WAA	CLC, LGALS8, TESPA1, UBASH3A
	SWE	CLC, INPP5D, LGALS8, P2RX7, TESPA1
	ENE	CLC, INPP5D, LGALS8, TESPA1
	SOA	CLC, INPP5D, TESPA1
	WSI	CLC, LGALS8, TESPA1
	SSI	CLC, INPP5D, LGALS8, TESPA1
	CSI	CLC, LGALS8, P2RX7, TESPA1
	NSI	CLC, LGALS8, SIGLEC1, TESPA1
	SEM	CLC, LGALS8, P2RX7, TESPA1
GO.Bcell	SSI	INPP5D
GO.Innate	WAA	APOBEC3H, DEFB1, PGLYRP4
	SWE	APOBEC3H, DEFB1, P2RX7, PGLYRP4, PLSCR1
	ENE	APOBEC3H, DEFB1, PGLYRP4, PLSCR1
	VOL	APOBEC3H, DEFB1, DMBT1, IFI16, PGLYRP4, PLSCR1
	SOA	APOBEC3H, DEFB1, DMBT1, PGLYRP4, TLR6
	WSI	APOBEC3H, DEFB1, DMBT1, PGLYRP4, PLSCR1
	CSI	C1QC, DMBT1, P2RX7, PGLYRP4, TLR6
	COL	DEFB1, DMBT1, PGLYRP4, PLSCR1
	SEM	C1QC, DEFB1, DMBT1, P2RX7, PGLYRP4, PLSCR1
	SEA	APOBEC3H, C1QC, DEFB1, DMBT1, IRGM, PGLYRP4, PLSCR1

Table 4.3 Subset of genes driving significant HKA enrichment signals

In the GO.Bacteria category, the populations with significant enrichments are characterized by sharing the same core of genes driving enrichment signals. These genes are *DEFB1*, *DMBT1*, and *PGLYRP4*, though the South Asian and Mainland Southeast Asian populations also share *TLR6*. *DEFB1* and *PGLYRP4* have been found to be targets of balancing selection in previous a previous study, in both African and European populations and African populations, respectively (Bitarello et al., 2018). DEFB1 is a defensin, or antimicrobial peptide, that has previously been shown to be under positive selection in primates (Hollox and Armour, 2008). Its promoter region has been shown to be under balancing selection by Cagliani et al. (Cagliani et al., 2008). DEFB1 provides protection from a range of Gram-negative bacteria (Goldman et al., 1997) and Candida species (Jurevic et al., 2003). It is expressed in a range of epithelia and tissues (Singh et al., 1998; Valore et al., 1998; Zhao et al., 1996). It is also expressed in milk and breast tissue of lactating mothers (Jia et al., 2001). This broad variety of expression and antimicrobial activity against a wide range of microbes suggests DEFB1 plays an important role in host defense. Certain variants of this gene have been associated with allergic disease, susceptibility to sepsis, HIV, and Candida infection (Braida et al., 2004; Cagliani et al., 2008; Chen et al., 2007; Jurevic et al., 2003; Leung et al., 2006). *DEFB1* is in the enrichments results of every population that is significantly enriched in the GO.Bact category except for the Central Siberian population. It also appears in a number of enrichments in the GO.Innate category. P2RX7 or P2X7 is a purinergic receptor that binds to extracellular ATP. In mouse studies, it plays a role in the regulation of intestinal T cells (Heiss et al., 2008). P2X7 plays an important role in tuberculosis infection. When infected macrophages were exposed to ATP, the viability of the infecting mycobacterium was reduced by ninety percent. When cells with a loss of function mutation in *P2X7* were exposed to ATP, there was no apoptosis or killing of mycobacteria (Saunders et al., 2003). *P2RX7* is in the enrichment results in the Southwest European, Central Siberian, and Mainland Southeast Asian populations, and also appears in several populations in the GO.Tcell and GO.Innate categories. PGLYRP4 (see review by Dziarski and Gupta (2010)) is a peptidoglycan recognition protein that is expressed in the skin, mucous membranes, the mouth, and the gastrointestinal tract (Lu et al., 2005). Peptidoglycan is an element of bacterial cell walls (Royet and Dziarski, 2007). PGLYRP4 has a bactericidal response against both Gram-negative and Gram-positive bacteria (Lu et al., 2005). This gene has also been associated with psoriasis (Kainu et al., 2009). TLR6 is a Toll-like receptor that is activated by bacterial and fungal pathogens (Murray et al., 2009), and appears in the top results in this category in the South Asian, South Siberia and Mongolian, Central Siberian, and Northeast Siberian populations. *DMBT1* plays an important role in the mucosal innate immune response via sensitivity to bacterial cell wall antigens. In the intestinal epithelia, DMBT1 expression is upregulated

upon exposure to LPS and may be implicated in Crohn's disease (Rosenstiel et al., 2007). DMBT1 is also expressed in saliva and expressed as salivary agglutinin, which binds to oral bacteria (Bikker et al., 2002). *DMBT1* copy number is correlated with population history of agriculture, via increase in dietary carbohydrates and the associated dental carie-causing bacteria *Streptococcus mutans* (Polley et al., 2015). *DMBT1* is in the enrichment signals of all populations that are significantly enriched in the GO.Bact category except for the Southwest European population, and it also appears in several populations in the GO.Innate category.

As in the GO.Bact category, the significant enrichments are largely driven by a core set of genes shared between populations. *CLC*, or galectin-10, is related to mucosal damage and number of phagocytic eosinophils (phagocytic cells involved in allergic response and defense against parasites) at an infection site. It has also been associated with celiac disease and allergic asthma (Chua et al., 2012; De Re et al., 2009; Murray et al., 2009). It appears in the enrichment signals in all populations that are significantly enriched in the GO.Tcell category. LGALS8, or βgalactoside-binding lectin, detects damage in host cells and marks them for degradation via autophagy. It has been shown to play a role in both bacterial and viral infections (Staring et al., 2017; Thurston et al., 2012). *LGALS8* is in the enrichment results in the West Asia and Armenian, Southwest European, Northeast European, West Siberian, South Siberia and Mongolian, Central Siberian, Northeast Siberian, and Mainland Southeast Asian populations, and has been found to be under balancing selection in populations in both Africa and Europe (Andres et al., 2009; Bitarello et al., 2018). *TESPA1* is a gene which plays an important role in T cell selection and growth and is involved in T cell receptor signaling and development (Wang et al., 2012a), and has been found to be under balancing selection in African populations (Bitarello et al., 2018). *TESPA1* is in the the enrichment signals in every population with a significant enrichment signal in this gene category. *UBASH3A* only appears in the enrichment signal of the West Asia and Armenian population and is a negative regulator of T cell receptor signaling and is associated with multiple autoimmune diseases (Carpino et al., 2009; Li et al., 2017). *INPP5D*, or *SHIP*, is involved in bacterial immune response. Knockout *SHIP* mice were hyper-responsive to LPS (lipopolysaccharide, or endotoxin, which is a component of Gram-negative bacterial membranes) and did not display endotoxin tolerance upon a second exposure to LPS (Sly et al., 2004). This suggests that *SHIP* plays an important role in regulating the initial and subsequent exposures to endotoxin and modulating the immune response to Gram-negative bacteria. *INPP5D* is in the enrichment signals in the Southwest European, Northeast European, South Asian, and South Siberia and Mongolian populations. SIGLEC1, or CD169, is lectin protein expressed by immune cells that binds to sialic acid, which is thought to be an important marker of self for

immune tolerance (Chen et al., 2014). A lack of sialylation can mark cells for degradation (Meesmann et al., 2010). Due to its immunosuppressive role, SIGLEC1 is thought to play a role in sepsis, by inappropriately inducing endotoxin tolerance (Wu et al., 2016). This induction of tolerance is thought to be an adaptation to persistent bacterial infections by preventing the immune system from mounting an ultimately damaging reaction (Sly et al., 2004). Macrophages expressing this protein play an important role in halting the spread of pathogens via the lymphatic system, stimulating the innate immune system, and assist B cells in recognizing antigens (Junt et al., 2007). *SIGLEC1* is only in the enrichment signal in the the Northeast Siberian population.

In the innate immune category, as in the others, a set of genes driving enrichment signals is shared between almost all populations. *APOBEC3H*, as previously described in Chapter 3, is a member of the APOBEC3 cytidine deaminase gene family that has innate immune activity, through causing hypermutation, against retroviruses. As mentioned in Chapter 3, it is thought to have been under positive selection at certain times throughout primate evolution. While *APOBEC3H* in other primates has strong antiviral effects, the human version is expressed at such low levels that it shows no antiviral activity. Two potential reasons are given for this lack of function: the retrovirus *APOBEC3H* evolved to combat went extinct and decreased the evolutionary pressure on the gene, or another member of the *APOBEC3* gene family now performs antiretroviral duties and *APOBEC3H* is no longer required to be functional. Under relaxed evolutionary pressure, it may be beneficial to have fewer hypermutation-causing cytidine deaminase genes for the sake of the host cell (OhAinle et al., 2006). *APOBEC3H* appears in the enrichment signals in the West Asia and Armenian, Southwest European, Northeast European, Volga Uralic, South Asian, West Siberian, and Island Southeast Asian populations. PLSCR1, or phospholipid scramblase 1, is a calcium-binding protein that amplifies the antiviral effects of the interferon response, since it is an interferon-stimulated gene and its expression is associated with increased expression of other interferon-stimulated genes (Dong et al., 2004). *PLSCR1* is in the enrichment signal in the Southwest European, Northeast European, Volga Uralic, West Siberian, Colla, Mainland Southeast Asia, and Island Southeast Asian populations. IFI16 is an intracellular sensor of exogenous double stranded DNA (present in DNA viruses and bacteria) and plays a role in the induction of interferon-β (Unterholzner et al., 2010). *IFI16* has been shown to be under past positive selection during primate evolution as well as long term balancing selection in Yoruban, European, and EastAsian populations from HapMap populations using the MLHKA test (Cagliani et al., 2014). This gene also contains SNPs that have been associated with both risk for and protection against various autoimmune diseases, suggesting antagonistic pleiotropy as the cause of balancing selection (Cagliani et al., 2014; Zhernakova et al., 2011).

C1qC is subunit of C1q, which is a member of the complement family and is a pattern recognition molecule. C1q helps to clear pathogens and cellular debris, including apoptotic cells. C1q deficiencies can cause autoimmune disease, potentially through the failure to remove apoptotic cells and subsequent autoantigens (Bohlson et al., 2007; Taylor et al., 2000). *C1QC* appears in the enrichment signals in the Central Siberian, Mainland Southeast Asian, and Island Southeast Asian populations. IRGM is an immunity-related GTPase with an interesting evolutionary history. Around 50 million years ago, most of the *IRG* gene cluster was deleted. The *IRGM* gene's open reading frame was disrupted by retrotransposition in the anthropoid common ancestor but was reestablished in the the common ancestor to humans and great apes due to disruption from an endogenous retrovirus insertion (Bekpen et al., 2009). IRGM controls pathogens via autophagy and has been linked to resistance to a range of pathogens (Bekpen et al., 2010; Singh et al., 2006b). In humans, it has been shown to control levels of intracellular mycobacteria (Singh et al., 2006b). *IRGM* has also been linked to Crohn's disease susceptibility (Parkes et al., 2007). *IRG* appears in the enrichment signal in the Island Southeast Asian population.

Despite that MHC genes were removed from the above enrichment analysis because they were potentially skewing enrichment results because of their strong representation in several gene categories, genes driving enrichment in the MHC are worth noting. The gene category of antigen processing and presentation is especially driven by MHC genes, therefore a subset of the genes driving the significant enrichments in Table 4.1 are shown in Table 4.4. The full list of genes can be seen in Table B.10.

GO.APP	AFR	HLA-A/B/C/DQA2/G, TAP1, TAP2
	WAA	HLA-A/B/C/DOA/DQA2/DRA/F/G, MICA, TAP1, TAP2
	SWE	HLA-A/B/C/DQA2/DRA/F/G, MICA, TAP1
	ENE	HLA-A/B/C/DOA/DQA2/DRA/F/G, MICA, TAP1
	VOL	HLA-A/B/C/DQA2/DRA/F/G, MICA, TAP1
	SOA	HLA-A/B/C/DOA/DQA2/DRA/F/G, MICA, TAP1
	WSI	HLA-A/B/C/DQA2/DRA/G, MICA, TAP1
	SSI	HLA-A/B/C/DPA1/DQA2/DRA/F/G, MICA, TAP1
	CSI	HLA-A/B/C/DPA1/DQA2/DQB2/DRA/F/G, MICA, TAP1
	NSI	HLA-A/B/C/DQA2/DQB2/DRA/F/G, MICA, TAP1
	COL	HLA-A/B/C/DPA1/DQA2/DQB2/DRA/F/G
	SEM	HLA-A/B/C/DOA/DPA1/DQA2/DRA/G, MICA
	SEA	HLA-A/B/C/DOA/DQA2/DRA/G, TAP2

Table 4.4 Subset of genes driving significant HKA enrichment signals in the antigen processing and presentation categorie

Table 4.4 is largely made up of HLA genes. *HLA-A*, *HLA-B*, *HLA-C*, *HLA-G*, *HLA-DF*, *HLA-DQ*, and *HLA-DP* have all been shown to be previously under balancing selection (Andres et al., 2009; Bitarello et al., 2018; DeGiorgio et al., 2014; Sabbagh et al., 2013). Supporting the idea that infectious disease is an important factor driving balancing selection at these loci, HLA diversity has been correlated with pathogen richness (Prugnolle et al., 2005b). It has also been shown through mouse studies that higher levels of heterozygosity in the MHC have better survival outcomes when infected with multiple pathogens, suggesting again that the high levels of polymorphism in the MHC may have evolved to provide protection against pathogens (McClelland et al., 2003; Penn et al., 2002). MICA is a cell surface protein in the MHC complex, which is an immune effector and upregulated upon infection (such as with *M. tuberculosis* or cytomegalovirus) and other stress events (Das et al., 2001; Groh et al., 2001). TAP1 and TAP2 are transporter subunits in the MHC (see review by Trowsdale (2011)). For this analysis, *TAP1* and *TAP2* are considered to be part of the MHC, but in future work it would be interesting to see whether the signal from these genes in coming from *TAP1* and *TAP2* themselves or other genes in the MHC.

4.2.2 Top-scoring genes per population based on the HKA statistic

In addition to looking at genes that are driving significant enrichment signals, it is also interesting to look at the top scoring genes in each population. As in the d_i results, the genes have been divided into 5 bins based on gene size, so larger genes don't have an unfair skew toward higher scores compared to shorter genes (see Section 2.4.2 for more details on dividing genes into bin sizes). In order to see generally the representation of immune genes in these top scoring 50 genes in each population, all genes in any of the immune gene categories of interest described in Chapter 2 (Table 2.1) are highlighted in orange.

Bin	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1	IGLL5	HLA-C	HLA-B	HLA-C	CDSN	HLA-DQA2	HLA-B	TRIML1	HLA-DQA2	TRIML1	CDSN	HLA-DQA2	HLA-DQA2
1	HLA-C	C6orf15	C6orf15	HLA-B	HLA-B	C6orf15	CDSN	CDSN	CDSN	CDSN	TRIML1	CDSN	TRIML1
1	HLA-DQA2	TRIML1	TRIML1	TRIML1	C6orf15	HLA-B	TRIML1	HLA-B	HLA-B	OR51B6	C6orf15	C6orf15	C6orf15
1	C11orf40	HLA-G	HLA-A	CDSN	HLA-C	HLA-C	C6orf15	C6orf15	C6orf15	HLA-DQA2	NANOG	HLA-B	CDSN
1	HLA-G	HLA-A	CDSN	C6orf15	TRIML1	CDSN	OR51B6	HLA-A	HLA-G	NANOG	HLA-DQA2	OR51B6	HLA-B
1	DEFB1	HCG27	ZFP57	HLA-G	OR51B6	TRIML1	HLA-DQA2	HLA-DQA2	TRIML1	HLA-B	HLA-A	TRIML1	RNF39
1	HLA-A	HLA-DQA2	OR51B6	HLA-DRA	HLA-DRA	IGLL5	OR51F1	OR51B6	OR51B6	TAP1	HLA-G	HLA-C	HLA-A
1	CDSN	CDSN	OR51F1	OR51F1	OR51F1	OR51B6	HLA-DRA	HLA-G	HLA-DRA	RNF39	OR51B6	HLA-A	OR51B6
1	TRIML1	HLA-B	HLA-G	TAP1	HLA-A	HLA-G	TAP1	HLA-C	NANOG	HLA-DRA	HLA-B	HCG27	ZFP57
1	HLA-B	OR51B6	TAP1	OR51B6	HLA-DQA2	OR51F1	HLA-C	HLA-DRA	HLA-C	HLA-A	HLA-C	HLA-G	PPP1R15A
2	TRIM40	BTNL2	BTNL2	PSMB9	BTNL2	BTNL2	KRT40	BTNL2	BTNL2	TMEM128	TMEM128	BTNL2	BTNL2
2	C20orf166	PSMB9	PSMB9	KRT40	TMEM128	TMEM128	LRRC15	FCER2	HLA-F	PGLYRP4	KRT40	ARPC5	LRRC15
2	DHRS2	KRT40	HLA-F	FCER2	FCER2	KRT40	ARPC5	POU5F1	ARPC5	BTNL2	BTNL2	KRT40	ARPC5
2	AMTN	ARPC5	POU5F1	HLA-F	LRRC15	POU5F1	AMTN	TMEM128	TMEM128	KRT40	PGLYRP4	TMEM128	TMEM128
2	HTR5A	MICA	ARPC5	TRIM40	KRT40	HLA-F	BTNL2	PGLYRP4	TRIM40	PSMB9	HLA-F	IFIT3	MYOZ3
2	POLR1E	POU5F1	OSGIN1	BTNL2	ARPC5	TAS1R2	PSMB9	LRRC15	LRRC15	POLR1E	AMTN	CYB5R2	PGLYRP4
2	KRT40	TRIM40	KRT40	LRRC15	TRIM40	LILRA2	POU5F1	OSGIN1	FCER2	FCER2	LYZL6	FCER2	TXN2
2	LRRC15	LRRC15	LYZL6	LYZL6	PGLYRP4	OSGIN1	TMEM128	C12orf36	MYOZ3	MYOZ3	FCER2	MTIF3	KRT40
2	RWDD4	FCER2	POLR1E	AMTN	SPHK1	LRRC15	C12orf36	HLA-F	KRT40	OSGIN1	L1TD1	PGLYRP4	C12orf36
2	FCER2	HLA-F	PGLYRP4	POU5F1	EIF3D	POLR1E	POLR1E	CCHCR1	PSMB9	L1TD1	BIRC5	OSGIN1	TAS1R2
3	NECAB2	MUC22	PSORS1C1	MUC22	PSORS1C1	MUC22	PSORS1C1	PSORS1C1	PSORS1C1	SNX19	PSORS1C1	OR11A1	PSORS1C1
3	PSORS1C1	PSORS1C1	MUC22	PSORS1C1	TESPA1	PSORS1C1	MUC22	TESPA1	CYBRD1	TESPA1	TESPA1	PSORS1C1	MUC22
3	MAR2	TAP2	TESPA1	TESPA1	MUC22	CHRNB3	TESPA1	CHRNB3	MUC22	PSORS1C1	MUC22	ZNF766	NECAB2
3	TMPRSS13	TMPRSS13	CHRNB3	CHRNB3	ZNF766	SELE	ZNF766	MUC22	TESPA1	MUC22	SNX19	MUC22	TNFRSF10D
3	ZNF766	ZNF766	RDH13	RDH13	CHRNB3	TESPA1	INCENP	SELE	ZNF766	TMPRSS9	C2orf83	CYBRD1	TESPA1

Table 4.5 Top 50 genes by HKA score part 1, correcting for gene size by dividing into 5 bins and taking the top 10 genes from each. Orange highlights indicate the gene belongs to one or more of the immune gene categories.

Bin	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
3	OR11A1	NECAB2	SULT6B1	NQO2	SELE	PGPEP1L	SNX19	SLC19A3	SNX19	INCENP	NUP54	TESPA1	CYBRD1
3	MUC22	SELE	SELE	C2orf83	C2orf83	MAR2	ALDH4A1	SNX19	SELE	LGALS8	SLC19A3	SLC19A3	SNX19
3	NQO2	TESPA1	ZNF766	SELE	SNX19	ALDH4A1	LGALS8	INCENP	WDR75	TMPRSS13	ADAMTSL2	MAR2	WFDC1
3	PGPEP1L	MAR2	PGPEP1L	PGPEP1L	INCENP	C2orf83	TMPRSS13	TMPRSS13	FOPNL	C2orf83	WDR75	INCENP	RDH13
3	TAP2	CHRNB3	C2orf83	SULT6B1	TMPRSS13	TMPRSS13	MPHOSPH6	LGALS8	TMPRSS9	WDR75	PLSCR1	SNX19	KNG1
4	CLDN16	CLDN16	CLDN16	CLDN16	CLDN16	CLDN16	IGFBP7	CLDN16	CLDN16	CLDN16	ATP2C2	CLDN16	CLDN16
4	SUZ12	NCMAP	TLDC1	PPIL2	IGFBP7	IGFBP7	SLC5A12	PSMG4	IGFBP7	PFKFB3	SLC5A12	OR5V1	AKR1E2
4	TEKT5	LY86	IGFBP7	IGFBP7	AKR1E2	ANGPT2	CLEC3A	IGFBP7	AKR1E2	PSMG4	AIM1	AKR1E2	SMOX
4	GPR55	OR5V1	SUZ12	AIM1	ATP2C2	PPIL2	AIM1	AIM1	LY86	IGFBP7	AKR1E2	PFKFB3	CLEC3A
4	OR5V1	IGFBP7	AKR1E2	CLEC3A	C6orf10	TLDC1	ATP2C2	KLHL23	DMBT1	CLEC3A	SMARCB1	ATP2C2	AIM1
4	SLC39A12	COL9A1	AIM1	AKR1E2	TLDC1	DMBT1	ACOT11	DMBT1	AIM1	ATP2C2	IGFBP7	CLEC3A	SYT6
4	LY86	AIM1	RFX8	PRIMPOL	AIM1	COL9A1	CAPN13	ATP2C2	CLEC3A	DMBT1	MAEA	PALD1	SLC5A12
4	IGFBP7	TRPV3	CLEC3A	SSR1	SLC37A1	AKR1E2	PRIMPOL	IGSF5	PSMG4	COL9A1	CLEC3A	DMBT1	IGFBP7
4	GPR78	PSMG4	SLC37A1	COL9A1	PSMG4	SLC5A12	TSPAN15	PFKFB3	ATP2C2	AIM1	ACOT11	TMEM156	LDB3
4	ALDH1L1	AKR1E2	IGSF5	PALD1	PALD1	IGSF5	RFX8	SLC5A12	COL9A1	AKR1E2	LY86	IGFBP7	PFKFB3
5	SNX29	CSMD1	CSMD1	CSMD1	CSMD1	CSMD1	CSMD1	CSMD1	CSMD1	CSMD1	CSMD1	KCNE4	CSMD1
5	CDYL2	CDYL2	STK32A	WWOX	EGLN3	WWOX	FAM19A5	KCNE4	CDYL2	CDYL2	KCNE4	CSMD1	CDYL2
5	CSMD1	WWOX	LUZP2	STK32A	CDYL2	LUZP2	RBFOX1	RBFOX1	EGLN3	NTN4	WSCD1	NTN4	LUZP2
5	RBFOX1	KCNE4	RBFOX1	LUZP2	WWOX	NTN4	ZNF98	EGLN3	RBFOX1	RBFOX1	LUZP2	RBFOX1	RBFOX1
5	LUZP2	PCDH15	EGLN3	RBFOX1	WSCD1	KCNE4	PCDH15	PCDH15	NTN4	HUS1	PCDH15	CHODL	ARHGAP8
5	WWOX	ZNF98	SH3RF3	EGLN3	LTBP1	ARHGAP8	CPE	NTN4	WSCD1	PKD1L1	ZNF98	CDYL2	PRR5-ARHGAP8
5	WSCD1	LUZP2	LTBP1	INPP5D	RBFOX1	PRR5-ARHGAP8	EGLN3	LTBP1	STK32A	SH3RF3	MYO3A	ARHGAP8	NTN4
5	KCNE4	WSCD1	DOCK1	COL4A3	STK32A	MYO3A	MYO3A	STK32A	FAM19A5	WSCD1	EGLN3	PRR5-ARHGAP8	CHODL
5	ARHGAP8	RBFOX1	ZNF98	NTN4	PCDH15	RBFOX1	WSCD1	SLC22A23	SH3RF3	STK32A	CPE	WSCD1	WSCD1
5	PRR5-ARHGAP8	LTBP1	MYO3A	WSCD1	MYO3A	WSCD1	STK32A	LUZP2	PCDH15	ZNF98	FAM19A5	LTBP1	FAM19A5

Table 4.6 Top 50 genes by HKA score part 2, correcting for gene size by dividing into 5 bins and taking the top 10 genes from each. Orange highlights indicate the gene belongs to one or more of the immune gene categories.

A subset of genes from Tables 4.5 and 4.6 are described below, based on clear immune function from a literature search. Because many of the genes are shared between several populations, the descriptions below will be divided by general immune function instead of by population groups as in the positive selection chapter. While many genes are shared between the top results of multiple populations, suggesting a potentially older or shared signal, some of the genes appear in the top results of only one population, potentially suggesting a more recent and specific signal. Table 4.7 shows the counts of how many times a given immune-related class of genes appears in the top 50 genes for that population. This can be used as a rough guide for comparing representation of immune gene classes between the classes themselves and between populations. Table 4.7 shows that the T cell and antigen processing and presentation gene categories are represented more strongly than other categories of immune genes, which is not unexpected given the importance of MHC genes in these results. Genes in the GO.Virus category are also well represented. Please note that not all of the genes highlighted in orange had a clear connection to immune function based on the literature, and so they were not included in the descriptions below. Similarly, some genes that were not highlighted did have clear connections to immune function, and so they were included in the gene descriptions below.

	Gene Ontology DB						HPI DB			
	Bact.	Virus	Tcell	Bcell	APP	Adapt.	Bact.	Virus	Prot.	Amoe.
AFR	1	4	6	0	6	0	2	4	0	0
WAA	0	5	9	0	9	0	0	2	0	0
SWE	1	4	6	0	6	0	1	3	0	0
ENE	0	4	8	1	7	0	2	4	0	0
VOL	1	3	6	0	5	0	1	3	0	0
SOA	1	3	6	0	5	0	0	2	0	0
WSI	0	4	7	0	6	0	1	2	0	0
SSI	2	4	9	0	7	0	0	2	0	0
CSI	1	4	7	0	7	0	1	2	0	0
NSI	2	5	7	0	6	0	1	3	0	0
COL	1	6	7	0	6	0	0	3	0	0
SEM	2	5	6	0	5	0	1	2	0	0
SEA	1	2	4	0	3	0	0	4	0	0

Table 4.7 Counts of number of genes in each immune gene category and population in the top 50 genes based on HKA score per population

Many of the genes with top HKA scores are involved in response to bacterial infection, though this is not reflected well by Table 4.7. *DEFB1* appears in the top results here, as well as in the HKA gene enrichment table under the GO.Bacteria table, where it is described as an

antimicrobial peptide that has been implicated in defense against bacterial infection. Unlike many other genes in the top HKA results, this gene is a top hit only in the West and Central African population, suggesting relatively localized balancing selection. However, *DEFB1* was described as driving significant enrichment signals in the South Asian, South Siberia and Mongolian, and Mainland Southeast Asian populations in the GO.Bact categories. It has a high score in both the Mainland Southeast Asian and Southeast Asian populations, though it doesn't fall into the top 50 genes by score. Therefore, it seems likely that any balancing selection signal in *DEFB1* is not specific to any one population and may be fairly widespread geographically. *FCER2*, or *CD23*, induces killing of bacteria and fungi via induction of nitric oxide synthase immune cells (see review by Flemming (2017)). *CD23* is also associated with allergic response, including T cell activation in response to allergen exposure and IgE levels associated with specific allergen (Selb et al., 2017). It is in the top results in the West and Central African, West Asia and Armenian, Northeast European, Volga Uralic, South Siberia and Mongolian, Central Siberian, Northeast Siberian, Colla, and Mainland Southeast Asian populations. As referenced in the HKA enrichment results, PGLYRP4 is a peptidoglycan recognition protein that is expressed in various tissues and has a bactericidal response against both Gram-negative and Gram-positive bacteria. It is in the top 50 genes in the Southwest European, Volga Uralic, South Siberia and Mongolian, Northeast Siberian, Colla, Mainland Southeast Asian and Island Southeast Asian populations. LILRA2 is a leukocyte immunoglobulin-like receptor and is a negative effector of antimicrobial TLR activity and a positive effector of eosinophils. It is also implicated in leprosy infection (see review by Brown et al. (2004)). It is in the top results of the South Asian population. *LGALS8* appeared in the HKA enrichment genes and is involved in host response to bacterial and viral infections, marking damaged host cells for autophagy. It is a top scoring gene in the West Siberian, South Siberia and Mongolian, and Northeast Siberian populations. *DMBT1* appears in the HKA enrichment tables, and plays a role in the mucosal innate immune response to bacteria. It is a top scoring gene in the South Asian, South Siberia and Mongolian, Central Siberian, Northeast Siberian, and Mainland Southeast Asian populations.

Defense against viral infection is another category of immune genes that is well represented in top HKA results, as shown in Table 4.7. *TAP1* is part of the MHC and is involved in antigen processing and presentation. It is also a virus-inducible negative regulator of the innate immune response, and is induced by influenza A virus and human enterovirus 71 (Xia et al., 2017). It is in the top results in the Northeast European, Southwest European, West Siberian, and Northeast Siberian populations. *TRIM40* is also in the MHC, and suppresses the viral-RNA recognition by RLR receptors. A deficiency in this gene results in a stronger antiviral immune response and decreased viral replication (Zhao et al., 2017). It is in the top

results in the West and Central African, West Asia and Armenian, Volga Uralic, Northeast European, Central Siberian populations. PSMB9, or LMP2, is a protein that is part of the proteasome. The proteasome degrades intracellular proteins into peptides that are presented by MHC class I molecules (Pierini and Lenz, 2018). Various proteasome genes appear in the top balancing selection results for HKA and other tests, though not all play a clear role in immune response based on the literature. The expression of *PSMB9* is increased upon exposure to interferon-gamma. This may make a proteasome that is better suited to processing endogenous antigenic peptides (Akiyama et al., 1994). *PSMB9* was in the top results in the West Asia and Armenian, Southwest European, Northeast Siberian, West Siberian, Central Siberian, Northeast Siberian populations. *NUP54* appeared in the HKA enrichment genes and plays a role in influenza replication (Tafforeau et al., 2011), and is a top result here in the Colla population. *PLSCR1* appeared in the HKA enrichment genes and is an amplifier of interferon's antiviral effects, and also appears here in the top results of only the Colla population.

As would be expected, genes involved in antigen processing and presentation are also among the genes with the highest HKA scores. *HLA-C*, *HLA-DQA2*, *HLA-G*, *HLA-A*, *HLA-B* are all among the highest scoring genes. They have associations with susceptibility to infectious and inflammatory disease, and appear in the top results of all populations. As seen in the enrichment results, *MICA* is an immune effector and upregulated upon infection (Das et al., 2001; Groh et al., 2001). *TAP2* is part of the MHC and plays a role in the immune response against various pathogens (see review by Trowsdale (2011)). It is in the top scoring genes in the West and Central African and the West Asia and Armenian populations.

A number of genes in Tables 4.5 and 4.6 are involved in immune activation. *SPHK1* (sphingosine kinase 1), along with *SPHK2*, produces S1P, which is an immune regulator that was previously discussed in the d_i results section of the Central Siberian population (Spiegel and Milstien, 2011). In summary, S1P gradients play a role in migration of immune cells. The gene *SPHK1* is in the top results of the Volga Uralic population. *SELE*, or E-selectin, plays a role in the recruitment of leukocytes to the endothelium and has been studied in the case of an individual with recurrent infections and impaired pus formation (DeLisser et al., 1999). It is in the top results of the West Asia and Armenian, Southwest European, South Asian, South Siberia and Mongolian, Northeast European, Volga Uralic and Central Siberian populations. *TNFRSF10D* is a member of the tumor necrosis factor receptor superfamily, whose members have important roles in immune response (see review by Collette et al. (2003)). Transcription of *TNFRSF10D* is a result of stimulation with the intestinal pathogen *S. Typhimurium* (Bruno et al., 2009), and this gene is found in the top results of the Island Southeast Asian population. *BTNL2* has been associated with inflammatory disease, and is

involved in T cell activation (Nguyen et al., 2006). It is in the top results of all but the West and Central African populations. As seen in the enrichment results, *TESPA1* is involved in T cell receptor signaling and T cell development. It is also in the top results of all but the West and Central African populations. *NQO2* plays a role in activation of chemokines and NF-kappaB signaling. Knockout mice for this gene and *NQO1* had impaired B cell function and increased tendency toward autoimmune disease (Iskander et al., 2006). *NQO2* is in the top results of the West and Central African and Northeast European populations.

4.3 Results of the β test

Like the HKA results, the results of enrichment analyses in the β results are separated into two parts: enrichment for immune genes in the top one percent of results, and a list of the top 50 scoring genes in each population. As with the HKA results, the top results are corrected for gene size as described in the Chapter 2. The β statistic is best powered to detect balancing signals older than 100,000 generations, which is about 2.5 million years ago if a generation time of 25 years is used (Siewert and Voight, 2017b).

4.3.1 Enrichment of immune genes in top results

The results of the enrichment test are shown below in Table 4.8. As with other tests, the expected number of genes per immune category is shown in the row labeled "exp", and light and dark orange colors represent statistically significant (with corrections for multiple testing) enrichment at two different p values. The gene count for the number of genes in the top one percent of results in a given population and gene category is given in each cell.

Most populations show a significant enrichment in genes associated with T cell function. Interestingly, while the top results of many populations are enriched for genes involved with response to bacteria, there are no significant enrichments for response to viral infection. There is less of a strong enrichment in the antigen processing and presentation category than there is in the HKA results. There are no significant enrichments in the HPI DB results.

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
exp	2.26	8.33	4.21	1.86	5.96	2.05	0.31	15.72	24.1	0.05	0.04
AFR	6	5	8	1	6	5	1	6	20	0	0
WAA	3	7	10	3	10	5	1	6	16	0	0
SWE	3	4	8	1	7	2	1	6	16	0	0
ENE	3	6	9	2	6	3	1	8	20	0	0
VOL	3	3	6	1	9	2	1	6	16	0	0
SOA	5	6	7	1	7	3	1	7	20	0	0
WSI	6	4	8	1	7	4	1	5	17	0	0
SSI	5	4	9	2	11	3	1	9	13	0	0
CSI	5	5	6	2	8	3	0	7	17	0	0
NSI	5	7	7	1	8	4	1	7	16	0	0
COL	3	4	7	1	3	5	1	3	19	0	0
SEM	4	3	5	1	11	2	1	8	12	0	0
SEA	6	4	9	2	5	4	0	4	15	0	0

Table 4.8 Enrichment of top one percent of β results in each population for immune genes based on the GO DB and HPIDB compared with expected counts. The "exp" row gives the expected number of immune genes in any given immunity category after randomly choosing one percent of all protein-coding genes, and the counts in each cell represent the observed number of immune genes of a given category in the top one percent of the results. Light orange indicates significant enrichment at p=0.05 and dark orange at p=0.01 in a Fisher's exact test with correction for multiple testing, as described in Chapter 2.

As with the HKA results, it is expected that genes in the MHC will be common among top balancing selection targets and may skew enrichments in other categories. Therefore, the enrichment analysis was performed again, excluding genes in the MHC. The results of this are shown in Table 4.9. While the enrichment patterns have changed somewhat, these results are much more similar to those including MHC genes than were the HKA results.

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
exp	2.23	8.17	3.99	1.85	5.9	1.79	0.3	15.57	23.89	0.05	0.04
AFR	6	5	6	1	7	2	1	6	21	0	0
WAA	3	7	7	3	11	2	1	7	18	0	0
SWE	3	3	6	1	7	0	1	6	17	0	0
ENE	3	5	6	2	7	0	1	8	20	0	0
VOL	3	4	4	1	9	0	1	6	15	0	0
SOA	5	6	4	1	7	0	1	7	20	0	0
WSI	6	4	6	1	10	1	1	5	17	0	0
SSI	5	6	6	2	12	0	1	10	13	0	0
CSI	5	6	5	2	10	1	1	7	18	0	0
NSI	5	8	4	1	9	1	1	7	16	0	0
COL	4	5	3	1	5	1	1	3	18	0	0
SEM	4	3	3	1	11	0	1	8	12	0	0
SEA	6	5	6	2	7	0	1	4	16	0	0

Table 4.9 Enrichment of top one percent of β results in each population for immune genes based on the GO DB and HPIDB compared with expected counts, with genes in MHC region removed. The "exp" row gives the expected number of immune genes in any given immunity category after randomly choosing one percent of all protein-coding genes, and the counts in each cell represent the observed number of immune genes of a given category in the top one percent of the results. Light orange indicates significant enrichment at p = 0.05, dark orange at p=0.01 in a Fisher's exact test with correction for multiple testing, as described in Chapter 2.

A subset of genes, chosen for their clear relation to immune function based on a literature search, driving the enrichment signals seen above is shown below in Table 4.10. This table shows that the shared signals of enrichment shown in Table 4.9 are often caused by the same genes, though not always with perfect overlap. Most of the genes in Table 4.10 appear more than once and often are shared between many populations. Unlike in the HKA results, there are significant enrichments in the West and Central African population, and these genes are shared more or less with all other populations in the analysis. This suggests shared signals predating the split between modern African and non-African populations, which fits with the timeframe for which the β statistic is optimally powered.

	Pop.	Genes
GO.Bact	AFR	CHIT1, DEFB1, DMBT1
	SWE	DEFB1, DMBT1, PGLYRP4
	ENE	DEFB1, DMBT1, PGLYRP4
	VOL	DEFB1, DMBT1, PGLYRP4
	SOA	DEFB1, DMBT1, PGLYRP4
	WSI	DEFB1, DMBT1, GNLY, PGLYRP4
	SSI	DEFB1, DMBT1, PGLYRP4
	CSI	DEFB1, DMBT1, PGLYRP4
	NSI	DEFB1, DMBT1, PGLYRP4
	COL	DEFB1, DMBT1
	SEM	DEFB1, DMBT1
	SEA	DEFB1, DMBT1, GNLY
GO.Tcell	AFR	CD209, CLC, IRF1, SFTPD, TESPA1
	WAA	CD209, IRF1, LGALS8, TESPA1, TRPM4
	SWE	CLC, IRF1, LGALS8, SFTPD, TESPA1, TRPM4
	ENE	CLC, INPP5D, IRF1, LGALS8, TESPA1
	WSI	CLC, IRF1, LGALS8, SFTPD, TESPA1, TRPM4
	SSI	IRF1, LGALS8, SFTPD, TESPA1, VAV1
	CSI	IRF1, SFTPD, TESPA1
	NSI	IRF1, SFTPD, TESPA1
	SEA	IRF1, LGALS8, TESPA1, TRPM4
GO.Innate	AFR	CD209, CLEC6A, DEFB1, DMBT1, IRF1, SFTPD
	WAA	APOL1, CD209, CLEC6A, DEFB1, DMBT1, IFI16, IRF1, PGLYRP4, ZC3HAV1
	SWE	APOL1, DEFB1, DMBT1, IRF1, PGLYRP4, SFTPD
	VOL	APOBEC3H, APOL1, DEFB1, DMBT1, IRF1, PGLYRP4
	SOA	DEFB1, DMBT1, IRF1, PGLYRP4, ZC3HAV1
	WSI	APOL1, DEFB1, DMBT1, IFI16, IRF1, KL, PGLYRP4, SFTPD
	SSI	C3, DEFB1, DMBT1, IFI16, IRF1, KL, PGLYRP4, SFTPD, VAV1
	CSI	APOBEC3H, C3, CLEC6A, DEFB1, DMBT1, IRF1, KL, PGLYRP4, SFTPD
	NSI	APOBEC3H, DEFB1, DMBT1, IFI16, IRF1, PGLYRP4, SFTPD
	SEM	APOBEC3H, APOL1, C3, CLEC6A, DEFB1, DMBT1, IRF1, KL, SFTPD
	SEA	APOBEC3H, APOL1, CLEC6A, DEFB1, DMBT1, IRF1

Table 4.10 Subset of genes driving significant β enrichment signals

Many of the genes driving enrichments in the GO.Bact category are shared between populations. One of the exceptions is CHIT1, or chitinase, an enzyme that digests chitin. *CHIT1* is enriching the GO.Bact signal in the West and Central African population. It also was part of the enrichment signal in the Colla population in the nSL test. Chitin is a polysaccharide that is a structural component of a variety of fungal, bacterial, and parasitic (including malaria) pathogens but is not made by human cells. This enzyme plays a role in the immune response to these organisms, and is also implicated in inflammatory disease (see review by Di Rosa et al. (2016)). Several genes driving significant enrichments in

the GO.Bact category are shared between almost all of the populations with significant enrichment results. These are *DEFB1*, *DMBT1*, and *PGLYRP4*, all of which are involved in defense against bacterial infection. These three genes also appear multiple times in the enrichment signals in the GO.Innate category, also with widespread sharing of these genes between populations. *GNLY* produces granulysin, which is an antimicrobial protein that has cytolytic effects against bacteria (including tuberculosis), parasites, fungi, and malaria (see review by Krensky and Clayberger (2009)). This gene drove significant enrichment in the West Siberian and Island Southeast Asian populations, and has been suggested to be a target of balancing selection in European populations by Bitarello et al. (2018).

In the GO.Tcell category, CD209 is an attachment factor for various viruses, bacteria, and parasites including *Mycobacterium tuberculosis*, dengue virus, *Leishmania*, hepatitis C virus, and others (Colmenares et al., 2002; Kooyk et al., 2003; Lozach et al., 2004; Sakuntabhai et al., 2005). *CD209* is part of the enrichment signal in the West and Central African and West Asia and Armenian populations, both in the GO.Tcell category and the GO.Innate category. *CLC*, involved in eosinophil function (De Re et al., 2009), is also shared between four of the nine populations. It is part of the significant enrichment signal in both the West and Central African and West Asia and Armenian populations. IRF1 is a transcription factor that is a regulator of interferon expression (Miyamoto et al., 1988). *IRF1* appears in the enrichment results in all the significantly enriched populations, suggesting a widely shared signal. It also appears in multiple populations in the GO.Innate category. SFTPD is a surfactant protein that is important in the defense of lung tissue from bacterial and viral pathogens (Wright, 2005). *SFTPD* is driving enrichment in the West and Central African, Southwest European, West Siberian, and South Siberia and Mongolian populations. *TESPA1* is involved in T cell receptor signaling and T cell development (Wang et al., 2012a), and is driving enrichment signals in all the populations with significant enrichments in this gene ontology category. *TRPM4* is associated with calcium signaling and has been found to be important for survival in a mouse model of sepsis (Serafini et al., 2012). *TRPM4* is also shared between four of the nine populations with significant enrichments in genes involved in T cell function. *VAV1* plays a role in the adhesion of T cells and antigen-presenting cells (Krawczyk et al., 2002), and unusually only appears in one one enriched population in Table 4.10, the South Siberia and Mongolian population. *LGALS8* is part of the enrichment signal in the West Asia and Armenian, Southwest European, Northeast European, West Siberian, South Siberia and Mongolian, and Island Southeast Asian populations. It is involved in detecting damage in host cells and plays a role in both bacterial and viral infections (Staring et al., 2017; Thurston et al., 2012). *INPP5D* plays a role in regulating response to endotoxin (Sly et al., 2004).

INPP5D appears only in the Northeast European enrichment signal in this gene ontology category.

In the innate immune category, CLEC6a is a C-type lectin receptor, which is involved in the immune response to fungi and mycobacteria (see review by Dambuza and Brown (2015)). This gene is in the enrichment signals for the West and Central African, West Asia and Armenian, Central Siberian, Mainland Southeast Asian, and Island Southeast Asian populations. ZC3HAV1 is an antiviral protein that shows evidence of long-standing balancing selection in several major human population groups, as well as trans-specific polymorphism between humans and chimpanzees (Cagliani et al., 2012). It only appears in the enrichment signal in the West Asia and Armenian population. *C3* is another component of the innate immune complement pathway (Murray et al., 2009), and appears in the enrichment signals in the South Siberia and Mongolian, Central Siberian, and Mainland Southeast Asian populations. *KL*, or klotho, is involved in aging, which is also reflected in its role in T cell aging and autoimmune disease (Witkowski et al., 2007). It has been suggested to be a target of balancing selection in European and African populations by Bitarello et al. (2018). *APOL1* provides innate resistance against the parasite *Trypanosoma brucei*, which causes trypanosomiasis in African primates, humans, and other animals. However, functional APOL1 has only been identified in humans and several other primates, which suggests that this gene may have a fitness cost in other animals that outweigh this benefit. This fitness cost may be related to the involvement of this gene in kidney disease in African American populations (review by Limou et al. (2015)). *APOL1* is in the enrichment signals in the West Asia and Armenian, Southwest European, Volga Uralic, West Siberian, Mainland Southeast Asian and Island Southeast Asian populations. *APOBEC3H* has appeared in the top HKA results, as well as in top positive selection results, and is an antiviral gene that plays an important role in other primates but seems to be less important in humans (OhAinle et al., 2006). This gene is part of the enrichment signal in the Volga Uralic, Central Siberian, Northeast Siberian, Mainland Southeast Asian, and Island Southeast Asian populations. IFI16 is an intracellular sensor of viral DNA and has been shown to be under past positive selection during primate evolution as well as long term balancing selection in Yoruban, European, and EastAsian populations from HapMap populations using the MLHKA test (Cagliani et al., 2014; Unterholzner et al., 2010). *IFI16* was in HKA enrichment results in the Volga Uralic population, and here is in the enrichment results in the West Asia and Armenian, West Siberian, South Siberia and Mongolian, and Northeast Siberian populations.

4.3.2 Top-scoring genes per population based on the β statistic

As in the presentation of the HKA results, the 50 genes with the highest β scores per population are presented in Tables 4.11 and 4.12. Again, genes highlighted in orange show that a given gene is in one of the immune gene classes described in Table 2.1. These are divided into 5 bins based on gene size to avoid a bias toward higher scores for larger genes (please see Section 2.4.2 for more information).

Bin	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1	CD209	CD209	IL36RN	OR51B6	MRGPRX2	DEFB1	PNMAL1	DEFB1	MRGPRX2	HLA-DQA2	HLA-DQA2	SYT5	HLA-DQA2
1	CLC	CDSN	ARL14EPL	PMAIP1	OR51B6	CDSN	SPRR1B	PNMAL1	OR51F1	OR52E6	OR52E6	PNMAL1	SYT5
1	CDSN	COX4I1	DEFB1	ARL14EPL	HLA-DQA2	OR51B6	SYT5	CDPF1	CDSN	CDSN	SPRR1B	APOBEC3H	OR51B6
1	GAS2L2	DEFB1	MRGPRX2	IZUMO1	DEFB1	ARL14EPL	MRGPRX2	NOV	CST9L	MRGPRX2	SYT5	HLA-DQA2	OR51Q1
1	OR5P2	MRGPRX2	ILK	DEFB1	ITLN1	MRGPRX2	ARL14EPL	IZUMO1	GP5	PNMAL1	IL36RN	S100A3	NOV
1	IL37	IL36RN	CLC	MRGPRX2	GAS2L2	HLA-DQA2	SPRR3	SYT5	S100A3	FCGR3B	ARL14EPL	S100A4	IL36RN
1	DEFB1	ILK	OR51B6	ILK	IL36RN	SYT5	IL36RN	C1orf68	S100A4	SYT5	NANOG	C1orf68	PNMAL1
1	GP5	NANOG	NAA38	LSM10	LSM10	IRF1	DEFB1	HLA-DQA2	IL36RN	ARL14EPL	PNMAL1	IRF1	SPRR1B
1	HLA-DQA2	NOV	OR51Q1	IRF1	ZNF280A	OR12D2	ANKRD53	OR51F1	IL37	LYZL6	OR51B6	KRT40	NAA38
1	IRF1	SYT5	CST9L	CLC	OR5P2	C19orf84	FCGR3B	CDSN	AVPR1B	KRT40	CDSN	DEFB1	APOBEC3H
2	MYOZ3	HLA-DPA1	HLA-DPA1	HLA-DPA1	HLA-DPA1	EPPIN	HLA-DPA1	HLA-DPA1	ARPC5	TIMM21	HLA-DPA1	HLA-DPA1	HLA-DPA1
2	SMPDL3A	MYOZ3	BSPRY	TNNT2	MYOZ3	EPPIN-WFDC6	ARPC5	PLIN5	HLA-DPA1	OSGIN1	EPPIN	OSGIN1	OSGIN1
2	DHRS2	PAK1IP1	OSGIN1	BSPRY	TIMM21	HLA-DPA1	EPPIN	TNNT2	ISX	MICA	EPPIN-WFDC6	ISX	FAAH
2	IL17RC	SMPDL3A	TIMM21	R3HDML	PAK1IP1	PAK1IP1	EPPIN-WFDC6	OSGIN1	MYOZ3	GNPDA1	CABP5	PDLIM2	PDLIM2
2	R3HDML	TMEM128	TMEM128	ECHDC3	TNNT2	GCSH	TMEM128	PAK1IP1	OSGIN1	TREML4	CPN2	BTNL2	L1TD1
2	RCN3	POLR1E	TNNT2	TIMM21	SMPDL3A	PDLIM2	OSGIN1	RCN3	TEX33	EPPIN	CCL26	MYOZ3	ISX
2	ARPC5	KCNK18	SMPDL3A	CPN2	OSGIN1	POLR1E	AP1M2	POLR1E	POLR1E	EPPIN-WFDC6	AP1M2	TIMM21	POLR1E
2	TNNT2	TAS1R2	KCNK18	PGLYRP4	TMEM128	TNNT2	GNLY	MYOZ3	ECHDC3	MYOZ3	MICA	SMPDL3A	ARPC5
2	EPPIN	ARPC5	PGLYRP4	DNAJC22	KCNK18	MYOZ3	KCNK18	MICA	TREML4	BTNL2	TMEM128	ARPC5	FAM92B
2	EPPIN-WFDC6	TNNT2	CCL26	IL32	CCL26	C17orf80	BTNL2	BTNL2	CBY1	TNNT2	POLR1E	GCSH	C15orf48
3	TESPA1	TESPA1	LGALS8	LGALS8	LGALS8	LGALS8	CLEC3A	TESPA1	CLEC3A	TESPA1	ADAMTSL2	TESPA1	TESPA1
3	CHRNB3	CLEC3A	CHRNB3	SLC19A3	CHRNB3	FOPNL	SNX19	CHRNB3	TESPA1	CLEC3A	CLEC3A	CLEC3A	CLEC3A
3	CD84	CHRNB3	ZNF534	TEKT1	ASAH1	TESPA1	TESPA1	CLEC3A	SLC19A3	TEKT1	CLEC1A	FOPNL	LGALS8
3	CLEC3A	PTGS1	RXRG	CHRNB3	CLEC3A	CLEC3A	NECAB2	LAD1	PODN	PODN	SLC19A3	ZNF766	FOPNL
3	IL1RN	SULT6B1	ZNF766	NECAB2	TESPA1	CHRNB3	CLEC1A	ASAH1	SULT6B1	SLC19A3	TESPA1	ADAMTSL2	CLEC1A

Table 4.11 Top 50 genes by β score part 1, correcting for gene size by dividing into 5 bins and taking the top 10 genes from each. Orange highlights indicate the gene belongs to one or more of the immune gene categories.

Bin	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
3	FASTKD1	PNOC	IL1RN	LAD1	LAD1	GPR111	SLC19A3	IL1RN	NECAB2	ASAH1	SULT6B1	RDH13	WDR4
3	ASAH1	SLC19A3	TESPA1	CLEC3A	IL1RN	CYLC2	SULT6B1	ZNF534	FOPNL	SULT6B1	C2orf83	PODN	CYLC2
3	SPINK4	LGALS8	CENPU	FOPNL	LGI2	RXRG	ZNF665	PNOC	MYO15B	RXRG	ASAH1	ZNF534	ASAH1
3	RDH13	PSORS1C1	SLC19A3	SMCO2	NECAB2	LDLRAP1	RASSF2	FOPNL	ZNF534	FOPNL	RDH13	MGST3	NECAB2
3	SLC19A3	CYLC2	LAD1	ASAH1	SLC19A3	IL1RN	IL1RN	NECAB2	TEKT1	ZNF266	TEKT1	ASAH1	DNAJC12
4	TEKT5	EGLN1	DMBT1	ORC5	DMBT1	IGSF5	RFX8	PTPN21	PSMG4	DMBT1	DMBT1	DMBT1	PSMG4
4	NCMAP	CYP4Z1	ORC5	CYP4Z1	SMOX	ORC5	PSMG4	DMBT1	PTPN21	PSMG4	PSMG4	PSMG4	DMBT1
4	PSMG4	IGSF5	GLYATL2	IGSF5	PSMG4	DMBT1	PTPN21	SULT2B1	DMBT1	ORC5	SMOX	UGT2B4	KCNE1
4	CASQ2	NCMAP	PSMG4	DMBT1	EGLN1	PSMG4	CNR2	PSMG4	MAEA	KIAA0513	IGSF5	CD38	SLC17A5
4	RFX8	PSMG4	TLDC1	PMEPA1	IGSF5	PTPN21	DMBT1	RPL39L	CD38	YIPF7	PLA2G4C	PMEPA1	CD38
4	SLC39A12	ORC5	CYP4Z1	TLDC1	GLYATL2	TLDC1	TRPM4	KIAA0513	RTN4IP1	LRRC20	RTN4IP1	ASB18	TRPM4
4	CHIT1	PTPN21	CCDC85C	CNR2	TRPM4	SMOX	NID2	TLDC1	YIPF7	SNX31	LDB3	SLC17A5	UGT2B4
4	LRRC2	PLA2G4C	CCNK	GLYATL2	RFX8	FBLN1	RTN4IP1	MDGA1	FLT3	FBLN1	SLC17A5	SEC14L5	COL9A1
4	TDGF1	TRPM4	B3GNT5	CD38	SULT2B1	B3GNT5	CD38	TEAD4	SNX31	PLEKHG5	PTPN21	KL	TTC37
4	SNX31	RFX8	SERPINE2	SULT2B1	KLHL5	HNF4A	PMEPA1	ITGAE	KL	SEC14L5	UGT2B4	RTN4IP1	ASB18
5	RGSL1	HUS1	CSMD1	HUS1	HUS1	CSMD1	MANBA	CSMD1	HUS1	CSMD1	EGLN3	HUS1	CSMD1
5	ABCA4	PKD1L1	HUS1	PKD1L1	PKD1L1	GRIN2A	HUS1	HUS1	PKD1L1	F13A1	HUS1	PKD1L1	SGCG
5	CSMD1	MANBA	PKD1L1	CSMD1	CSMD1	HUS1	PKD1L1	PKD1L1	CSMD1	HUS1	PKD1L1	F13A1	F13A1
5	LHFPL3	JUP	HS3ST4	JUP	JUP	PKD1L1	CSMD1	MSR1	GRIN2A	PKD1L1	CSMD1	OFCC1	OFCC1
5	WWOX	CSMD1	MYO3A	CCDC129	FAM189A1	HS3ST4	F13A1	KIRREL3	NFASC	GRIN2A	POLR2M	CSMD1	HUS1
5	AGAP1	ADAM29	SH3RF3	RBFOX1	EFCAB4B	AMPH	GRIN2A	VWF	ABCA4	ABCA4	CNTN5	JUP	PKD1L1
5	LRRC16A	CNTN5	GRIN2A	EYS	F13A1	SH3RF3	NFASC	MALRD1	WWOX	POLR2M	GRIN2A	POLR2M	RBFOX1
5	CES5A	NFASC	MALRD1	CES5A	PDE1C	TLK1	LTBP1	GRIN2A	MANBA	CCDC129	MANBA	SH3RF3	MYO3A
5	RBFOX1	F13A1	KIRREL3	GRIN2A	NFASC	CDH4	ARHGAP8	TMC2	AMPH	WSCD1	LTBP1	ADAM29	SH3RF3
5	ATP8A2	ABCA4	RAD18	SH3RF3	AMPH	CPE	PRR5-ARHGAP8	PKHD1	CDH13	AMPH	NOL10	WWOX	FSTL4

Table 4.12 Top 50 genes by β score part 2, correcting for gene size by dividing into 5 bins and taking the top 10 genes from each. Orange highlights indicate the gene belongs to one or more of the immune gene categories.

A subset of these top genes are described below, based on their clear connection to immune function based on a literature search. As in the description of the top scoring genes from the HKA test, the gene descriptions will be grouped by broad immune function, since many signals are shared between multiple populations. Table 4.13 shows the counts of how many times a given immune-related class of genes appears in the top 50 genes for that population. This can be used as a rough guide for comparing representation of immune gene classes between the classes themselves and between populations. Table 4.13 shows that the gene categories relating to bacterial and viral defense response, as well as T cell function, are represented more strongly than other categories of immune genes. However, as with the results of the d_i and HKA tests, genes assigned to these categories do not always have a clear relation to immune function based on a literature search, and some genes that are not highlighted to have a clear immune function based on a literature search. Therefore, the genes highlighted below are grouped by immune function, and not by strict membership to the gene ontology terms in Table 4.13.

	Gene Ontology DB						HPI DB			
	Bact.	Virus	Tcell	Bcell	APP	Adapt.	Bact.	Virus	Prot.	Amoe.
AFR	4	2	5	0	2	1	1	5	0	0
WAA	1	1	5	0	2	0	2	4	0	0
SWE	3	2	4	0	1	0	1	7	0	0
ENE	3	3	5	1	1	1	4	6	0	0
VOL	2	1	5	0	2	0	2	6	0	0
SOA	4	3	5	0	2	1	3	7	0	0
WSI	5	2	3	1	2	0	1	5	0	0
SSI	2	1	4	0	3	0	4	5	0	0
CSI	1	1	3	2	1	0	2	5	0	0
NSI	3	2	3	0	2	0	2	4	0	0
COL	3	2	4	0	4	0	0	4	0	0
SEM	2	3	4	1	2	1	1	6	0	0
SEA	1	2	5	1	2	0	1	2	0	0

Table 4.13 Counts of number of genes in each immune gene category and population in the top 50 genes based on β score per population

As Table 4.13 shows, a large number of the genes with top β scores are involved in defense against bacterial pathogens. CD209 is a C-type lectin receptor expressed in phagocytic cells that functions as a pathogen recognition receptor for a wide range of pathogens including HIV, *Mycobacterium tuberculosis*, *Helicobacter pylori*, hepatitis C virus, Dengue virus, certain strains of pneumonia bacteria, as well as parasites responsible for schistosomiasis and leishmaniasis (see Barreiro et al. (2005) for review). Barreiro et al. undertook an extensive

population genetic analysis of *CD209* in a global context and found that it has been under strong selective constraint. However, it does exhibit high levels of nucleotide diversity in African populations. This, the authors conclude, is not due to balancing selection but rather to ancient population structure within the larger African population. Time to most recent common ancestor for this gene is one of the oldest reported, at around 2.8 million years (Barreiro et al., 2005). This certainly fits with the older timeframe of the β test, which here seems to have picked up not a signal of true balancing selection but rather an artifact of population structure. IL37 is a suppressor of the immune response. Mice with transgenic expression of this gene were less susceptible to LPS-induced shock and had fewer cytokines and less activation of dendritic cells (Nold et al., 2010). *DEFB1*, an antimicrobial defense gene, was in the top HKA results as well and here is seen in the top results for the Central and West African, West Asia and Armenian, Southwest European, Northeast European, Volga Uralic, South Asian, West Siberian, South Siberia and Mongolian, and Mainland Southeast Asian populations. *MRGPRX2* is expressed in mast cells and is a receptor for LL-37, an antimicrobial peptide (Subramanian et al., 2011; Zanetti et al., 1995). It is in the top results in the West Asia and Armenian, Southwest European, Northeast European, Volga Uralic, South Asian, West Siberian, Central Siberian, and Northeast Siberian populations. *IL17RC* is in the top results in the West and Central African population. IL17RC, along with IL17RA, forms a heterodimer receptor IL17R, which is the ligand for IL17A and IL17F. *IL17RC* is expressed mostly in non-immune cells, and is bound with higher affinity by IL17F than IL17A. Both *IL17A* and *IL17F* play a role in immune response to bacteria and fungi. *IL17F* is mostly involved in mucosal defense, and *IL17A* is involved in autoimmune and allergic responses as well as immune responses (see review by Iwakura et al. (2011)). *PGLYRP4* appears here in the top results of the Southwest European and Northeast European population populations, as well as in the top HKA results, and is a peptidoglycan recognition protein that has a bactericidal response against both Gram-negative and Gram-positive bacteria. As seen in the β gene enrichment results, *GNLY* produces the antimicrobial protein granulysin, which acts against a wide range of pathogens. *GNLY* is a top scoring gene in the West Siberian population. *LGALS8* is in the top results for the West Asia and Armenian, Southwest European, Northeast European, Volga Uralic, South Asian, Island Southeast Asian populations, and it also appeared in the HKA top results. It is involved in detecting damage in host cells and plays a role in both bacterial and viral infections (Staring et al., 2017; Thurston et al., 2012). As mentioned in the gene enrichment results, *TRPM4* is involved in calcium signaling and has been implicated in sepsis survival. It is in the top results in the West Asia and Armenian, Volga Uralic, West Siberian, and Island Southeast Asian populations. *DMBT1* appears here again and also is in the top enrichment and individual HKA results. It plays an

important role in the mucosal innate immune response via sensitivity to bacterial cell wall antigens. In the β results, it is among the top scoring genes in the all but the West and Central African and West Asia and Armenian populations.

Multiple top scoring genes are involved in defense against viral infection. *IRF1* was in the β enrichment results, and is a regulator of interferon expression, and here is in the top results in the West and Central African, Northeast European, South Asian, and Mainland Southeast Asian populations. *APOBEC3H* was also found in the HKA enrichment results, and is an important antiviral gene in other primates but seems to be less important in humans (OhAinle et al., 2006). Here it is in the top results in the Mainland Southeast Asian and Island Southeast Asian populations. IL32, in the top results of the Northeast European population, is a proinflammatory cytokine and plays a role in defense against DNA viruses, RNA viruses, and *Mycobacterium tuberculosis* via induction of other inflammatory cytokines and differentiation of monocytes into macrophage-like cells (Bai et al., 2010; Netea et al., 2005, 2008; Zepp et al., 2011). *TREML4*, in the top results of the Central Siberian and Northeast Siberian populations, is a positive regulator of TLR7 signaling. Mice deficient in this gene showed decreased propensity to autoimmune disease and showed an inhibited immune response to influenza virus. The *TREM* family of genes are receptors that amplify or dampen TLR signaling to regulate immune response (Ornatowska et al., 2007; Ramirez-Ortiz et al., 2015; Turnbull et al., 2006). Also seen in the window-based DIND results for the Southwest European population, MSR1 binds to extracellular double-stranded RNA and instigates the innate immune response to viral infection via activation of interferons via interaction with TLR3 (Dansako et al., 2013; DeWitte-Orr et al., 2010). *MSR1* is in the top results of the South Siberia and Mongolian population.

IL36RN is also involved in the defense response against potentially pathogenic organisms, though the organism in this case is not bacteria nor viruses but rather fungus. IL36RN is the receptor antagonist of IL36 and acts against the inflammation-causing IL36 family of cytokines. IL36 cytokines are thought to play a role in anti-fungal immune response, and mutations in *IL36RN* have been associated with psoriasis (Blumberg et al., 2007; Gresnigt et al., 2012; Marrakchi et al., 2011).

Several top scoring genes are involved in antigen processing and presentation, though fewer in the β results than in the HKA results. Interestingly, whereas other HLA genes were very common in top results for the HKA test, only *HLA-DQA2* appears in the top β results, and only in the West and Central African, Volga Uralic, South Asian, South Siberia and Mongolian, Northeast Siberian, Colla, Mainland Southeast Asian and Island Southeast Asian populations. *MICA* is also involved in antigen processing and presentation and appears in the

top results in the South Siberia and Mongolian, Northeast Siberian, and Colla populations, as well as in the top HKA results. It is a cell surface protein in the MHC complex.

A number of genes in the top β results are involved in T cell and immune activation and inflammation. *CLC*, or galectin-10, is in the top West and Central African, Southwest European, and Northeast European results and was also described in the HKA enrichment results. It plays a role in eosinophil function and has been implicated in inflammatory disease (De Re et al., 2009; Murray et al., 2009). CCL26 is a chemokine that plays a role in allergic disease by attracting eosinophils, killer lymphocytes, and monocytes (Nakayama et al., 2010). It is in the top results in the Southwest European, Volga Uralic, and Colla populations. *PDLIM2* is in the top results in the South Asian, Mainland Southeast Asian, Island Southeast Asian populations and is involved in limiting the inflammation caused by the immune response. Knockout mice for this gene were more likely to have EAE (Qu et al., 2012). AP1M2 deficiency in mice causes immune impairment and chronic colitis, and the gene is implicated in maintaining gut immune homeostasis. Human patients with Crohn's disease exhibit reduced expression levels (Takahashi et al., 2011). *AP1M2* is in the top results in the West Siberian and Colla populations. *BTNL2* is in the top results in the West Siberian, South Siberia and Mongolian, Northeast Siberian, and Mainland Southeast Asian populations, as well as in the HKA top results, and is involved in T cell activation (Nguyen et al., 2006). ISX is an intestinal transcription factor in the top Central Siberian, Mainland Southeast Asian, and Island Southeast Asian results that has been identified as an important part of maintaining vitamin A homeostasis that helps to maintain immune function tolerance in the gut (Widjaja-Adhi et al., 2017). *CD84* is in the top West and Central African results, and is a stimulator for interferon-gamma production, and is involved in T cell activation (Martin et al., 2001; Tangye et al., 2003). IL1RN, like IL36RN, is an interleukin receptor antagonist. It binds to the IL1 receptor, as does IL1, but it does not initiate an inflammatory response. Therefore, this gene acts as a way to limit immune and inflammatory response (McIntyre, 1991). *IL1RN* is in the top results in the West and Central African, Southwest European, Volga Uralic, South Asian, West Siberian, and South Siberia and Mongolian populations. *PTPN21*, in the top results of the West Asia and Armenian, South Asian, West Siberian, South Siberia and Mongolian, Central Siberian, and Colla population, is a negative regulator of ICAM-1 in in skin cells, thus having a dampening effect on immune reactions in the skin (Cho et al., 2017). *ITGAE*, or *CD103*, was a significant gene in the window-based positive selection tests in the West and Central African population. It is involved in immune homeostasis in both the intestine and the airways, and here is in the top South Siberia and Mongolian results. *KL* was described in the β enrichment results, and is involved in aging,

which is also reflected in its role in T cell aging and autoimmune disease. It is in the top results in the Volga Uralic, Central Siberian, and Mainland Southeast Asian populations.

Another top result in the β results is *EGLN1*, which is not known for its immune involvement, but is included here because it is a noted target of past positive selection in both Tibetan and Andean populations as an adaptation to living at high altitude (Bigham et al., 2010; Yi et al., 2010). It is a top result in both the West Asian and Armenian population and the Volga Uralic population, suggesting that potentially the selection history of this gene may be more complex than previously thought.

4.4 Results of the Tajima's D test

As in the previous two sections, the results of enrichment analyses in the Tajima's D balancing selection results are separated into two parts: enrichment for immune genes in the top 1% of results, and a list of the top 50 scoring genes in each population. Since Tajima's D was computed per 200 KB window, each gene did not have its own Tajima's D score. In order to rank individual genes, the gene with the highest mean minor allele frequency from each of the top 50 genic windows was chosen. Those genes are listed in descending order with the gene from the top window first. In some cases, a gene may appear multiple times in one population. This is because that gene was in multiple windows and had the highest mean minor allele frequency in each window. The Tajima's D test is best powered to detect signals between 50,000 and 250,000 years ago (Grossman et al., 2013; Sabeti et al., 2006). As is discussed in the Discussion chapter of this thesis, it is likely that a window size of 200 KB is likely far too large to be used to search for signals of balancing selection. However, the data was available in 200 KB windows so it was repurposed in this analyses to see if it could be used in multiple ways.

4.4.1 Enrichment of immune genes in top results

Compared to the other two tests of balancing selection, there are very few significant enrichments for immune genes in top Tajima's D balancing selection results. Also unlike the other two tests, there is a (single) significant enrichment in a HPI DB category, here in the Colla population in the HP.Bacteria category. The most enrichments are in the antigen processing and presentation category, and there are two in the GO.Bact category.

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
exp	2.15	8.92	5.11	2.56	7.25	2.1	0.38	19.4	25.7	0.07	0.06
AFR	3	10	11	0	4	8	0	12	19	0	0
WAA	1	3	2	1	3	2	0	7	15	0	0
SWE	1	6	1	0	2	1	0	12	18	0	0
ENE	2	5	4	1	1	2	0	8	14	0	0
VOL	3	6	3	2	3	3	0	15	18	0	0
SOA	2	4	3	1	1	5	0	12	21	0	0
WSI	1	10	0	1	4	2	0	16	14	0	0
SSI	1	6	1	1	2	2	0	9	18	0	0
CSI	2	4	3	0	6	1	0	15	19	0	0
NSI	2	4	4	1	4	0	0	12	16	0	0
COL	1	6	1	0	3	1	1	23	22	0	0
SEM	1	6	4	0	5	1	0	8	18	0	0
SEA	2	5	2	0	5	2	0	6	17	0	0

Table 4.14 Enrichment of top one percent of Tajima's D and mean MAF results in each population for immune genes based on the GO DB and HPIDB compared with expected counts. The "exp" row gives the expected number of immune genes in any given immunity category after randomly choosing one percent of all protein-coding genes, and the counts in each cell represent the observed number of immune genes of a given category in the top one percent of the results. Orange indicates significant enrichment at p = 0.01 in a Fisher's exact test with correction for multiple testing, as described in Chapter 2.

As with the HKA and β statistics, it is possible that MHC genes will skew the enrichment results of the Tajima's D test and suggest significant enrichments where there are none. Because of this, the enrichment analysis was performed excluding the MHC genes. The results of this analysis are shown in Table 4.15. Also as with the HKA and β results, the pattern of significant enrichments is different when MHC genes are excluded. Notably, all enrichments in the antigen processing and presentation category disappear save for that in the South Asian population. All significant enrichments in the West and Central African population disappear as well, as does the sole enrichment in the GO.Tcell category.

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
exp	2.13	8.83	5.02	2.55	7.21	2	0.37	19.31	25.59	0.07	0.06
AFR	2	7	5	0	4	2	0	11	17	0	0
WAA	1	3	2	1	3	2	0	7	15	0	0
SWE	1	6	1	0	2	1	0	12	18	0	0
ENE	2	4	3	1	1	1	0	8	13	0	0
VOL	3	6	2	2	3	2	0	15	18	0	0
SOA	2	3	1	1	1	3	0	11	20	0	0
WSI	1	10	0	1	4	2	0	16	14	0	0
SSI	1	6	1	1	2	2	0	9	18	0	0
CSI	2	4	3	0	6	1	0	15	19	0	0
NSI	2	4	4	1	4	0	0	12	16	0	0
COL	1	6	1	0	3	1	1	23	22	0	0
SEM	1	5	3	0	5	0	0	8	18	0	0
SEA	2	5	2	0	5	2	0	6	17	0	0

Table 4.15 Enrichment of top one percent of Tajima's D and mean MAF results in each population for immune genes based on the GO DB and HPIDB compared with expected counts, excluding genes in the MHC. The "exp" row gives the expected number of immune genes in any given immunity category after randomly choosing one percent of all protein-coding genes, and the counts in each cell represent the observed number of immune genes of a given category in the top one percent of the results. Orange indicates significant enrichment at p = 0.01 in a Fisher's exact test with correction for multiple testing, as described in Chapter 2.

A subset of the genes responsible for the enrichment signals shown in Table 4.15 are shown in Table 4.16, and described in more detail below.

	Pop.	Genes
GO.Bact	VOL	NFKB1, TLR2
GO.Virus	WSI	AP1S3, DDX1, DUOX2, RNASEL, TRIM23
GO.Adapt	COL	NLRP10

Table 4.16 Subset of genes driving significant Tajima's D balancing selection enrichment signals, not including MHC genes

In the GO.Bact category and Volga Uralic population, *NFKB1* is involved in the regulation of a number of inflammatory responses (Borm et al., 2005). TLR2 is activated by both

bacterial and fungal ligands and is an important part of the innate immune response to those pathogens (Murray et al., 2009).

In the GO.Virus category in the West Siberian population, *AP1S3* is necessary for normal TLR3 endosomal translocation and downstream signaling, as well as normal *IL1* and *IL36alpha* expression. Mutations in this gene can cause skin autoimmune disease and inflammation (Mahil et al., 2016; Setta-Kaffetzi et al., 2014). This gene was also seen in the d_i top results in the West Asia and Armenian population. DDX1 is an RNA helicase and plays a role in coronavirus replication (Xu et al., 2010). *DUOX2* was described in the nSL enrichment results in the Northeast European population and plays a role in antiviral immune response (Fink et al., 2013). RNASEL is an RNAse that sets of interferon production pathway during a viral infection (Malathi et al., 2007). *TRIM23* was in the Mainland Southeast Asia nSL enrichment results, and plays a role in the TLR3 antiviral and innate immune response (Arimoto et al., 2010).

In the adaptive immunity category and the Colla population, *NLRP10* is important in the initiation of the adaptive immune response (Eisenbarth et al., 2012).

4.4.2 Top-scoring genes per population based on the Tajima's D statistic

As in the HKA, β , and d_i results, it is interesting to look at individual top scoring genes in the Tajima's D balancing selection analysis as well as looking at enrichment signals. However, Tajima's D was run on 200 KB windows across the genome and not gene-wise, as HKA, d_i, and β were. Therefore, in order to get gene-wise results to be comparable to the other balancing selection statistics, the top 50 windows are represented in descending order of Tajima's D score, with the gene within that window with the highest mean minor allele frequency representing that window. The rationale behind this system is that genes under balancing selection would likely tend to have variants at intermediate frequency than variants at high or low frequencies. Please see Chapter 2 (Section 2.3.1) for more information. As with the previous sections, genes highlighted in orange have been classified as immune-related genes according to Table 2.1 in Chapter 2. Not all highlighted genes have a clear connection to immune function based on the literature, so not all are described. Some genes not highlighted have a clear connection to immune function based on the literature, and so those are described.

AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
HLA-DQA2	KLHL5	MAPT	OR5H6	OR5H6	OR5H6	BARD1	RALYL	BICC1	BICC1	WWP2	BICC1	OR5H6
HLA-DRA	UGT2A2	OR5H6	LUZP2	BICC1	BICC1	TPD52	BICC1	SLC25A48	EYS	COL11A1	FAM175A	BICC1
HLA-G	ZNF568	STPG2	BICC1	NBR1	OR5B3	BICC1	OR5H6	RALYL	USP38	BICD2	KRTAP8-1	STPG2
HLA-DPB1	PIGN	BICC1	OR10X1	PCMT1	SLC35B3	NRG1	DPP10	TRHDE	APIP	TMEM38B	CCDC169	OR52E6
TAS2R42	TMPRSS11F	MAPT	SLC27A6	NMUR2	KLHL5	SLC25A48	LAMC1	TFCP2L1	EYS	PCDH9	OR10X1	CCNG2
XIRP2	DYNC2H1	FREM3	DND1	KRTAP22-1	RXFP1	SPRR3	UGT2B7	MYO3A	NTS	WWP2	GPHN	NOP10
SLC38A9	OR52E6	SLC35B3	POT1	PKHD1	FREM3	PCMT1	FNDC1	TMEM135	PDS5B	CCDC175	SLCO4C1	SPP1
PKHD1	SPAG16	SPAG16	GALC	SLC14A1	CNTLN	FNDC1	POT1	MKL2	NFKB1	GLRA1	POT1	POT1
ECD	PKHD1	KANSL1	OR10A2	FREM3	SLC14A1	LEPR	KRTAP22-1	C10orf11	ALDH1L1	MICU2	OR5H6	TMEM196
ATF5	TBX20	CRHR1	ATRNL1	POT1	POT1	ASB17	BTBD3	KSR1	IPMK	KCNJ15	GPHN	TMEM38B
HLA-A	OR5B3	KLHL5	KLHL5	GRM8	CCDC169	FABP2	C12orf56	LAMC1	MYO3A	OR52J3	ALDH1L1	KLHL5
SPINK5	GRM8	SLC27A6	CNTNAP2	SLC27A6	DND1	SNX2	DYNC2H1	ATF6	GPHN	NARS2	DNTTIP2	DYNC2H1
VPREB1	COG6	RBM26	PKHD1	NBAS	CWC27	GPC5	KLHL5	NFKB1	FER1L6	OR5P2	RALYL	DND1
BRK1	OR10X1	LARS	SLC14A1	LANCL2	HLA-A	SLC28A2	ALDH1L1	TEX9	MAN2A1	LMBRD1	SLC14A1	NRXN1
SLC28A2	TRPC6	FAM175A	EPHA4	PARD3	LEPR	N4BP1	CCNG2	FAM175A	KLRG1	CCDC175	NUBPL	ALG8
RGS13	CCDC25	FANK1	MTRNR2L6	NMUR2	FAM175A	KRTAP22-1	SLC14A1	PELI2	ZNF443	POLD3	TMC1	STPG2
STPG2	SLC14A1	OR10A2	FAM175A	LPPR1	LAMC1	PDS5B	SLC35B3	EYS	STPG2	LCORL	AMTN	SLITRK6
TMEM51	TPD52	CCDC171	CCDC169	DCC	PSMC1	GRM8	BICD1	EIF3E	SLC15A5	LDB2	CCDC69	SPAG16
NLGN1	MTRNR2L6	LUZP2	GIMAP7	CCDC169	PYDC2	CWC27	GRM8	BICD1	RERGL	CNTNAP2	CDC42BPA	PTGDR
STPG2	CEP112	PKHD1	PAPL	N4BP1	SLITRK6	OLFM4	UGT8	TMC1	FAM71C	RABGAP1L	MRPL15	DDIT4L
LARP6	OR5H6	CNTNAP2	OR8G5	NAALAD2	NOP10	PKP2	KRTAP8-1	PBLD	GPC5	SCAPER	DEPDC7	LGALS8
ZMAT4	PDZRN4	OR8G5	SPAG16	GLRA1	CRYGD	PKHD1	COL4A4	OLFM4	C12orf71	SPI1	NUP54	GPHN
KRTAP8-1	UGT2B4	KY	OR10H3	SPECC1	NRXN1	SLC38A9	SLITRK6	MTRNR2L6	CEACAM3	ALK	C5orf17	SENP7
APBB1IP	OR52L1	MTRNR2L6	CWC27	SLITRK6	SPAG16	EYS	NUP54	OR4C12	SERPIND1	CAND1	FSHR	CWC27
CGGBP1	DMXL2	CRYBG3	PLCB1	SLC46A3	XCR1	ZNF568	SLC25A48	MPZL2	DYM	CSNK1A1L	SPINK5	TANK

Table 4.17 Part 1 of table showing the gene with the maximum mean MAF from each of the top 50 200 KB windows with the highest positive Tajima's D values, excluding windows not containing protein-coding genes. Orange highlights indicate the gene belongs to one or more of the immune gene categories.

AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
LMBRD1	OR10H3	FBXO28	GRM8	DLG2	DCBLD2	KLRG1	TMEM128	OR5H6	TBC1D22A	RABGAP1L	TGFBR3	DEPDC7
FBXO28	PANK1	OR10X1	HLA-DQA2	INSL6	OR10A2	ANKRD62	FSHR	DPP10	RALYL	OR10V1	SENP7	CCDC39
ANO3	CCDC169	GRM8	PDIA6	ZNF479	PCDH15	SLC14A1	LHFPL3	ZNF568	SLITRK6	COL4A4	NPFFR2	AMTN
OR10H3	CNTNAP2	PDIA6	APIP	FABP2	STPG2	CBY1	ATP8A2	CBLN4	EIF3E	HIVEP3	LAMC1	PARD3
PTPN21	SNX13	OR5B3	FRMD4B	HLA-G	CADM2	OR5H6	CENPK	DYM	HINT1	LRRFIP2	HTATIP2	CPS1
CENPU	THUMPD1	TAS2R20	SLC9C1	OCA2	HLA-DPB1	FAM83B	OR10A2	CLOCK	FAM216B	BAI3	TMEM128	TMEM132C
PREX2	KCNIP4	ULK4	NOL10	FAM175A	N4BP1	AP1S3	DAPK1	APBA2	GPHN	SLITRK6	F3	POLD3
OR9K2	GLIPR1L2	NRXN1	GPR111	TAS2R20	VPREB1	SCNN1G	PYDC2	FLVCR1	DND1	INSL6	NARS2	FBXL13
MSMO1	FSIP2	GIMAP7	STPG2	TPD52	OR13C9	CCDC130	RFC3	DGKB	HPSE2	EIF3E	ZFAT	CNTNAP2
KIAA1217	FGD4	F3	CRYBG3	SERINC2	NAALADL2	NBAS	SPINK5	VWDE	PRTFDC1	SH3TC2	DND1	CAND1
EPM2A	DCBLD2	CCDC169	XCR1	SLC28A2	RCBTB1	FGF14	TECRL	TAF1B	MTURN	CCDC17	ZNF438	TECTA
LAMC1	PELI2	TMPRSS11F	C6orf10	ANKRD62	CEP112	NBAS	ERO1LB	TMEM128	SOX11	IARS	CLEC3A	FAT3
ATP8A1	CCDC17	DMXL2	SPINK5	PSMD3	NOL10	RAB11FIP5	CLEC3A	ZNF266	CNBD1	MUSK	RAB18	FAM160A1
MPZL2	F3	NOL10	USP38	ATRNL1	CNTNAP2	ST6GALNAC3	HTATIP2	OR10X1	NDST4	SUN5	RASGEF1B	CCDC34
GATM	ACSS3	RASSF9	CLOCK	ZNF568	FLG2	ATP5S	CWC27	IPMK	DIP2B	RBBP8	PYDC2	GPHN
HLA-C	CEP112	DND1	CCDC91	C6orf163	FRRS1	ADAMTS12	OR52E6	FAT3	BARD1	DGKI	CRYGD	SNRPC
TMEM135	DCC	DLG2	ANKRD62	MORC1	ANKRD62	CENPK	MYOM3	KRTAP8-1	ERO1LB	MRPL39	OR52E6	PLCB1
ADAMTS3	FABP2	POT1	ITLN1	RBM26	FNDC1	POT1	MTRNR2L9	SHPRH	HMGN4	ALK	TTC18	MEPE
LGALS8	ACTR1A	LRAT	LARS	LONRF2	PDIA6	BRIP1	SLC35B4	DISP1	CBLN4	POLR1E	PRKG1	NRP1
RPP21	KALRN	ZNF479	CDSN	SAMD12	USP38	ACTR1A	EYS	FSHR	UBE3C	SCFD1	PANK1	KDR
CDSN	AFTPH	MAST2	TG	GPR111	RUNX1	CCDC169	CASQ2	MAK16	FREM3	CCDC109B	ALG8	FAM98A
PRNT	FREM3	OR52E6	PDZRN4	TECTA	PADI4	OR10H3	MRPL15	DOCK9	ZNF568	TACR3	TEX9	SLC30A5
PPAP2A	OR51Q1	TULP4	CENPA	OR8G5	MAST2	HSPB7	EYS	RIMKLA	SCNN1G	NBAS	ZNF45	OR2A25
STYK1	TRHR	GRIP1	OR51Q1	NBAS	SLC35G1	EYS	NPHP3	EPM2A	OR5H6	MTBP	BARD1	PYDC2
ULK4	BICC1	SP140L	KRT83	CLEC3A	C1QTNF7	C12orf79	DGKB	CSRNP1	SLC14A1	ETFA	MPHOSPH6	ZNF92

Table 4.18 Part 2 of table showing the gene with the maximum mean MAF from each of the top 50 200 KB windows with the highest positive Tajima's D values, excluding windows not containing protein-coding genes. Orange highlights indicate the gene belongs to one or more of the immune gene categories.

A subset of the top 50 genes by Tajima's D score and minor allele frequency, included in the descriptions below based on their clear involvement in immune function based on a literature search, are listed below for each population and grouped by gene function. Two categories of genes which are not overtly involved in immune function but that are well represented in Tables 4.17 and 4.18 are olfactory receptors and solute carriers. It is interesting to note that selection signals seem to be shared less in this statistic than in others. In the gene descriptions below, many genes appear in the top results of multiple populations. However, these are usually shared between two or a small number of populations. Sometimes these populations are geographically close, or are likely to share a relatively recent common ancestor population, but sometimes they are significantly geographically separated. Based on these results, there doesn't seem to be a pattern for which types of immune genes are likely to be shared as top results between populations and which are not. Table 4.19 shows the counts of how many times a given immune-related class of genes appears in the top 50 genes for that population. This can be used as a rough guide for comparing representation of immune gene classes between the classes themselves and between populations. Table 4.19 shows relatively small numbers of immune-related genes in these results, and the highest concentration of immune genes in the HP.Bact and HP.Virus categories. However, as with the results of previous balancing selection statistics, genes assigned to these categories do not always have a clear relation to immune function based on a literature search, and some genes that are not highlighted to have a clear immune function based on a literature search.

	Gene Ontology DB						HPI DB			
	Bact.	Virus	Tcell	Bcell	APP	Adapt.	Bact.	Virus	Prot.	Amoe.
AFR	0	4	9	0	6	0	3	6	0	0
WAA	1	1	0	1	2	0	1	3	0	0
SWE	0	0	0	0	0	0	3	3	0	0
ENE	0	1	2	0	1	0	3	4	0	0
VOL	0	1	1	1	2	0	5	4	0	0
SOA	0	2	2	0	3	0	2	8	0	0
WSI	0	2	0	1	2	0	7	6	0	0
SSI	1	3	1	0	1	0	3	8	0	0
CSI	1	1	2	0	0	0	6	9	0	0
NSI	1	0	1	0	0	0	5	7	0	0
COL	0	3	0	0	0	0	6	8	0	0
SEM	0	2	1	0	0	0	1	4	0	0
SEA	1	1	1	0	1	0	3	4	0	0

Table 4.19 Counts of number of genes in each immune gene category and population in the top 50 genes based on Tajima's D score and MAF per population

Though this is not reflected by Table 4.19, a large number of top scoring genes are involved in the defense response to bacterial infection, meaning that selection in these genes could be an adaptation to bacterial pathogens. *LGALS8* is in the top results in the West and Central African and Island Southeast Asian populations and also appeared in top HKA and β results. Between the three statistics, is appears in the top results of a wide variety of Eurasian and Island Southeast Asian populations This suggests that balancing selection in this gene is global and potentially very old. It is involved in detecting damage in host cells and plays a role in both bacterial and viral infections (Staring et al., 2017; Thurston et al., 2012). TRPC6 is among the top results in the West Asia and Armenian population, and is a receptor-activated calcium channel (see review by Dietrich et al. (2005)) that has also been associated with vascular inflammation caused by endotoxin (LPS). LPS induces a TLR4 response, which induces TRPC6 in endothelial cells. This in turn sets off a cascade that leads to increased lung vascular inflammation and permeability via activation of NF-kappaB. This can cause acute lung injury in individuals with sepsis (Tauseef et al., 2012). Also in the top results in the West Asia and Armenian and Central Siberian populations, *PELI2*, or Pellino 2, is important for TLR/IL1R-mediated post-transcriptional control. *PELI2* knockout cells underwent decreased expression of inflammatory genes usually induced by LPS or IL1 (Kim et al., 2012). *APIP* is in the top results in the Northeast European population and the Northeast Siberian population. Reduced expression of APIP, or apoptotic protease activating factor 1-interacting protein, is associated with an increase of *Salmonella*-associated cell death and increased survival of individuals with systemic inflammatory response syndrome during the course of a *Salmonella* infection. The authors of this study suggest that the SNP associated with these outcomes has been under selection in Europe and Asia, though not in the Americas, Oceania, or Africa, due to earlier development of agricultural practices in the former regions (Ko et al., 2012). ITLN1, or intelectin, is a lectin that binds to galactofuranose, which is a component of bacterial cell walls, suggesting a role for this gene in the innate immune response (Tsuji et al., 2001). *ITNL1* is in the top results in the Northeast European population. NBR1, in the top results in the Volga Uralic population, is an LC3-interacting adaptor protein that acts to tag microbial and cytosolic material for autophagy, implying a role in the coordination of the innate immune response (see review by Levine et al. (2011)). The *SLC46* family of genes has been shown to be involved in innate immune recognition of peptidoglycan, however the role of *SLC46A3* in this context has not been elucidated (Paik et al., 2017). *SLC46A3* is in the top results of the Volga Uralic population. Antimicrobial fragments of the skin protein FLG2, or filaggrin-2, target the replication of the bacteria *Pseudomonas aeruginosa*. Healthy skin is usually able to fight off infection, but when the skin barrier is compromised infection can ensue (Hansmann et al., 2015; Yu et al., 2007). *FLG2* is in the top results of the South

Asian population. *PADI4*, or *PAD4*, plays an important role in the formation of neutrophil extracellular traps, which are part of the innate immune response against bacteria. Knockout mice for this gene are more susceptible to bacterial infection (Li et al., 2010b). *PADI4* is in the top results of the South Asian population. *OLFM4*, in the top results of the West Siberian and Central Siberian populations, plays a role in *H. pylori* infection. Knockout mice had a lower bacterial load and more inflammatory signaling that wild type mice (Liu et al., 2010). *PYDC2*, or *CPOP2*, acts as a regulator of inflammasome and controls the manufacture of IL-1beta. Its expression was also induced in response to LPS (Dorfleutner et al., 2006). It is in the top results in the South Asian, South Siberia and Mongolian, Mainland Southeast Asian, and Island Southeast Asian populations. The phagocyte receptor CEACAM3, in the top results in the Northeast Siberian population, is expressed on granulocytes and is involved in the recognition and response to several Gram-negative bacteria such as *Neisseria gonorrhoeae* and *Haemophilus influenzae*, which can cause sepsis, meningitis, and infertility (see review by Pils et al. (2008)). *RASGEF1B* expression is induced by microbial infection and TLR activation and plays a role in the mediation of the innate immune response (Andrade et al., 2010). *RASGEF1B* is in the top results of the Mainland Southeast Asian population, and has been suggested as a target of balancing selection by Bitarello et al. (2018).

Several of the top scoring genes are involved in defense response against parasites. *FREM3* is in a region that has been shown to be under ancient balancing selection, potentially because of its proximity to and LD with *GYPE*, *GYPB* and *GYPA*, which are implicated in malaria infection (Leffler et al., 2013; Manjurano et al., 2015). *FREM3* is in the top results in the West Asia and Armenian, Southwest European, South Asian, Volga Uralic, and Northeast Siberian populations. LEPR, or the leptin receptor, has been associated with susceptibility to amebiasis and amebic colitis, and leptin has been associated in mice with differential susceptibility to *M. tuberculosis* and *Streptococcus pneumoniae*. Thus leptin and its receptor are important in the immune response, as well as in their perhaps better-known role in appetite and metabolism (see review by Mackey-Lawrence and Petri (2012)). *LEPR* is in the top results of the South Asian and West Siberian populations.

A number of top scoring genes are involved in defense response against viral infections. XCR1 is in the top results in the Northeast European and South Asian populations, and is the chemokine receptor of XCL1, a chemokine that is produced during immune and inflammatory reactions. *XCL1* has been associated with host response to viral, bacterial, and parasitic pathogens. *XCR1* is expressed by dendritic cells and is involved in antigen presentation and the cytotoxic T cell response. The interaction of XCR1 and XCL1 also plays a role in thymic selection and self-tolerance (see review by Lei and Takahama (2012)). *USP38* is in the top results in the Northeast European, South Asian, and Northeast Siberian

populations, and is a negative regulator of type I interferon signaling (Lin et al., 2016). It is also in the same region as *FREM3*, mentioned in the paragraph above. N4BP1, in the top results of the Volga Uralic, South Asian, and West Siberian populations, has been shown to inhibit interferon-beta production upon viral infection, however, it is not clear from the literature how widespread this immunomodulatory effect is (Li et al., 2011). *AP1S3* was described in the enrichment results in the West Siberian population, and is involved in TLR3 function. It is in the top results in the West Siberian population. *NUP54*, in the top results of the South Siberia and Mongolian and Mainland Southeast Asian populations, also appeared in HKA results and is a nuclear transport protein that has been shown to be necessary for influenza replication (Tafforeau et al., 2011). *DAPK1*, in the top South Siberia and Mongolian population, is a regulator of the antiviral immune response via its interaction with interferon-beta, IRF7, IRF3, INFB1, and RIG-1 (Zhang et al., 2014). As seen in the Tajima's D balancing selection enrichment results, *NFKB1* is involved in the regulation of a number of inflammatory responses including response to viral infection and innate immune response, migraine, and the interaction between gut microflora and the immune system (Hiscott et al., 2001; Lawrence et al., 2005; Nenci et al., 2007; Reuter et al., 2002). *NFKB1* is in the top results for the Central Siberian and Northeast Siberian populations. WWP2, in the top Colla results, is a negative regulator of the TLR3-mediated innate immune response to double stranded viral RNA. Knockout mice produced higher levels of interferon-beta, TNFalpha, IL6, and CCL5 in response to the activation of this response (Yang et al., 2013).

As with the results for other balancing selection statistics, there are a number of top scoring genes involved in antigen processing and presentation. The top four results in the West and Central African population are genes in the MHC, which is not surprising based on the other balancing selection tests. However, it interesting to note that while results for other populations do include HLA genes, none are at the very top of the results as they are in the West and Central African population. *HLA-A* and *HLA-DPB1* are in the top results in the West and Central African and South Asian population. *HLA-G* is in the top results in the West and Central African and Volga Uralic population. *HLA-DRA* and *HLA-C* are only in the top results in the West and Central African population. *HLA-DQA2* is in the top results in the West and Central African and Northeast European populations. HLA genes are clearly less often shared among top results than in other balancing selection statistics. *RPP21* is among the top results in the West and Central African population. Loci in this gene were found to have significant effect on antibody response to HBV vaccine. These loci are though to have some regulatory effect on the expression of various HLA genes (Yucesoy et al., 2013).

A number of genes in the top Tajima's D results for balancing selection are those involved with T cell and immune activation and function. *PTPN21* in the West and Central African

population was also in the top individual genes for the β results. It is is a negative regulator of immune reactions in the skin. Among the top results in the Southwest European and Northeast European populations are *GIMAP7* and *GRIP1*. GIMAP7 is a GTPase of immunity-associated protein that plays a role in lymphocyte survival (Schwefel et al., 2013). *GRIP1* is a regulator of innate immune reaction via its role as a coactivator of IRF3 (interferon regulatory factor 3) (Reily et al., 2005), and has been suggested to be a target of balancing selection in European populations by Bitarello et al. (2018). *LANCL2*, significant in the Volga Uralic population, has been implicated in the regulatory response of T cells and macrophages during infection with *H. pylori* (Leber et al., 2016). *RUNX1*, in the top results of the South Asian population, has been associated with the control of regulatory T cell function, predisposition to autoimmune disease, and hematopoiesis and T cell development (Helms et al., 2003; Ono et al., 2007; Taniuchi et al., 2002; Tokuhiro et al., 2003). *KLRG1* is expressed on T cells and NK cells, and has a negative regulatory effect on numbers of NK cells (Blaser et al., 1998; Wang et al., 2013a). *KLRG1* is in the top results in the West Siberian and Northeast Siberian populations. *LRRFIP2*, in the top results in the Colla population, is an inhibitor of NLRP3 inflammasome activation, which has been associated with immune and inflammatory disease (Jin et al., 2013; Wen et al., 2012).

Though a pigmentation-related gene and not an immune-related gene, *OCA2* appears here in the top results in the Volga Uralic population. It is also in the top d_i results of several populations. It is interesting that this gene appears in both positive and balancing selection results.

Chapter 5

Discussion

The preceding chapters outlining the results of positive and balancing selection analysis results contain lots of information, including results for enrichment for different types of immune genes and individual top SNPs and genes. From the results of these analyses, we can learn about not only what types of genes have been under which type of selection, but also about global distribution of selection signals and gain insight into the methods used in this thesis. This discussion chapter will be organized around the themes set out at the end of Chapter 1: methodology, population selection histories, and selection targets.

5.1 Methodology

This thesis project used the nSL, iHS, Tajima's D, and d_i statistics to look for evidence of positive selection and the HKA, β , and Tajima's D statistics to look for evidence of balancing selection. Because each of these statistics is designed to look for specific qualities in a region of the genome (for example, long haplotypes or groups of variants at mid-range frequencies), each will return a unique set of results that doesn't fully overlap with the results of other statistics. Therefore, results reflect a variety of time frames and selection scenarios. The extent to which the different statistics gave overlapping results was varied.

The following questions posed at the end of Chapter 1 are discussed in this section:

- How much do the results of each of the selection tests overlap with each other?
- How can a relatively large, window-based signal of selection be narrowed down to an individual variant potentially driving the signal?
- How are the results influenced by the choice of which immune genes to examine for evidence of selection?

5.1.1 Sharing of top windows between populations and tests

Positive selection

The different proportions of overlap between the three window-based statistics can be seen in Table 5.1, and a subset of the gene overlaps between the d_i and the window-based positive selection results can be seen in Table 5.11. In the window-based tests of selection, the results of the nSL and iHS tests overlapped the most, and Tajima's D overlapped less with both. Table 5.1 shows that around twenty or thirty percent of top one percent windows are usually shared between the nSL and iHS tests, which makes sense because both are based on haplotype homozygosity tests. Despite the levels of sharing of windows between the nSL and iHS tests, it is interesting to note that the two tests are different enough that they rarely pick up the same enrichment signal. In future work it would be interesting to look into why this is the case, since both are based on measurements of haplotype homozygosity and would therefore be expected to return similar results. This is shown in Figures 3.1 and 3.3. For the most part, less than ten percent of top one percent windows are shared between either haplotype homozygosity test and the Tajima's D test, except in the South Asian population. It is not clear why this population is an outlier in this respect. The results of the d_i statistic did overlap to a small extent with the window-based test results. These overlapping results would be in regions that display extreme differentiation in a given population compared to all other populations as well as score high in a given window-based test.

Pop.	nSL:iHS	nSL:TD	TD:iHS
AFR	0.352	0.024	0.04
WAA	0.336	0.056	0.136
SWE	0.304	0.048	0.128
ENE	0.392	0.08	0.096
VOL	0.224	0.064	0.136
SOA	0.368	0.12	0.16
WSI	0.16	0.048	0.072
SSI	0.248	0.024	0.096
CSI	0.2	0.056	0.056
NSI	0.24	0.024	0.048
COL	0.312	0.008	0.048
SEM	0.368	0.072	0.152
SEA	0.304	0.08	0.056
Average	0.293	0.054	0.094

Table 5.1 Proportions of top 1% windows shared between different tests per population

Table 5.2 shows the extent to which top one percent window are shared between populations. The Tajima's D test shows much more sharing of windows, whereas the nSL and iHS tests have many more windows that only appear in one population or are shared between a small number. This fits with Tajima's D being used to measure more ancient selection, since sharing signals between multiple populations hints at that type of selection history. The nSL and iHS tests are used to measure recent selection, so it makes sense that those signals would appear in individual, local populations.

Num. pops.	Count (TD)	Count (nSL)	Count (iHS)
1	378	1001	832
2	112	192	189
3	65	43	65
4	52	8	23
5	35	8	8
6	24	2	6
7	15	1	1
8	11	1	2
9	5	0	2
10	3	0	0
11	3	0	1
12	0	1	0
13	0	0	0

Table 5.2 Counts of top 1% windows shared between populations. There were 12554 windows total for each population.

Balancing selection

As in the results of the positive selection tests, each statistic used to search for signals of balancing selection is designed to pick up a different type of signal. This means that each statistic produced a unique set of results. Of the three different measures of balancing selection used in this project, the HKA is the only commonly used measure of balancing selection. β is fairly newly developed, and Tajima's D is not often used to measure balancing selection. Many of the top results from the HKA statistic also appear in the top results of the β statistic, giving confidence in the results of the β statistic. These include HLA genes, *DEFB1*, and others. The results of the combination of Tajima's D and minor allele frequency, on the other hand, hardly overlap at all with either the HKA or the β results. This, along with the conspicuous absence of HLA genes in the top results of any population save the West and Central African population, marks the results of this test as perhaps less reliable than those of the other two. However, given the large difference in window size between the Tajima's D test and the others, it is likely that the test itself is not to blame for lack of reasonable results but rather the way in which it was applied. The extent to which the top 50 genes in each population overlapped in each of the three tests can be seen in Table 5.3. HKA and β results

overlap much more often than either does with the Tajima's D statistic (except in the case of the West and Central African population), again marking its results as somewhat different from those of the other two tests.

Population	HKA:β	HKA:Tajima's D	Tajima's D:β
AFR	0.18	0.10	0.04
WAA	0.16	0.00	0.00
SWE	0.20	0.02	0.00
ENE	0.10	0.04	0.00
VOL	0.14	0.00	0.02
SOA	0.20	0.00	0.00
WSI	0.16	0.00	0.02
SSI	0.18	0.02	0.02
CSI	0.18	0.02	0.00
NSI	0.28	0.00	0.00
COL	0.24	0.00	0.02
SEM	0.20	0.04	0.02
SEA	0.16	0.00	0.02
Average	0.18	0.02	0.01

Table 5.3 Proportion of overlapping results between the top 50 genes in each population in each of the three balancing selection tests

Table 5.4 shows the extent to which top 50 genes are shared between populations in each of the balancing selection tests. The HKA test shows the most sharing of genes between populations all the way up to all thirteen populations. The β test shows an intermediate amount of sharing of top genes between populations, with very few genes shared between more than half of the populations. This is surprising, since β is supposed to be optimized to detect selection signals that predate the split of all modern populations. In that case, it would be expected for there to be a very high number of top β signals shared between a high number of populations. Finally, the combination of Tajima's D and minor allele frequency produced the least amount of sharing of top hits between populations, with a high number of top hits found only in one population and no sharing at all above nine populations.

Num. pops.	Count (HKA)	Count (β)	Count (Tajima's D)
1	51	95	230
2	20	52	84
3	15	24	24
4	9	16	8
5	13	10	8
6	7	10	8
7	11	4	2
8	3	8	3
9	5	2	2
10	3	2	0
11	4	5	0
12	5	1	0
13	7	0	0

Table 5.4 Counts of top 50 genes shared between populations in the balancing selection tests.

Of these three statistics, only HKA is widely used as of yet to look for signatures of balancing selection. It is interesting to note the extent to which the other two, relatively untested, statistics pick up the same signals, and to compare them to HKA in terms of sharing of signals between populations.

The relative proportion of what are considered to be top results (here, one percent) to the rest of the genome is also an important consideration. For example, if a larger proportion of top results was used (for example, five percent), it is possible that there would be more sharing of top results between populations. Also, one limitation of using the top one percent of results for either balancing or positive selection is that is assumes the same amount and strength of each type of selection in each population, which is unlikely to be true. For future work, it could be useful to instead use an empirical approach based on p-values alone instead.

A note on window size of Tajima's D used in balancing selection

The use of a window size of 200 KB was far too large for this purpose. This window size was used because the data existed in this form for the positive selection analysis, so it made sense to see if it could be reused for this purpose. However, future analyses should use a much smaller window if using Tajima's D to look for balancing selection.

5.1.2 Narrowing down positive selection signals from windows to driver SNPs

This project used a number of filters to move from a 200 KB window which was highlighted as a top positive selection hit to a single SNP potentially driving that signal. This involved using the DIND statistic to look for derived alleles at high frequency linked with low-diversity sites, followed by using CADD to filter for variants predicted to be functionally important. The combination of these two filters was effective in producing a list of SNPs that could be driving a signal of selection in a top-scoring window, however, there are several limitations to this method that could also be addressed with future work. One of these is that CADD has somewhat limited functionality in regulatory regions (Nishizaki and Boyle, 2017), meaning that potential targets in regulatory regions may have been inappropriately filtered out from the results. Perhaps a further study could focus on selection in regulatory regions, and try to find tools optimized for scoring functional importance in those regions. Another limitation is that the DIND statistic discarded any SNP with a derived allele frequency above ninety percent, thus screening out any SNPs that are near fixation. It is possible that a beneficial allele under selection could reach such high frequencies, and this work would have missed those variants. Regarding the iHS and nSL analyses, DIND was applied to top windows in order to find driver SNPs, however, iHS and nSL are both SNP-based statistics. In future work, it would be more efficient to use the top SNPs from the iHS and nSL statistic without converting the scores into windows. Finally, there are many genes that were in the top one percent of window-based and d_i results that appeared in the descriptions of the enrichment tables but were eventually screened out via the DIND and CADD results before the final smaller list of results was compiled and would be interesting for further inquiry.

5.1.3 Gene lists

This project also highlights some caveats and lessons with regards to the selection of gene lists used to highlight selection in certain phenotypes:

- Using gene lists that are small is not ideal for enrichment tests. Examples of gene lists that are potentially too small for such tests in this project were the lists for adaptive immune genes, genes associated with amoebozoa infections, and genes associated with protozoa infections. While those classes of immune genes are certainly worth examining, any signals of enrichment in those classes should be interpreted with caution, since a statistically significant enrichment could consist of a single gene in the top results.

- The database from which gene lists are taken has a large effect on the composition of those gene lists. For example, there were overlaps between the bacteria and virus lists in the GO database and the HPI database. However, the enrichment tests results did not always match between the two gene lists. This shows that the way gene lists are curated can have a large effect on the results of a selection study like this. Based on literature searches for top hits, inclusion in the GO database gene lists often accompanied a clearer connection to immune function than did inclusion in the HPI database gene lists. For this reason, the results of this thesis more often focused on interpretation of those results, especially in the discussion of genes driving significant enrichment signals. If a study were looking to capture as many genes as possible that have been associated with a certain type of pathogen, using a gene list from the HPI DB might be better.

- Not all genes associated with important immunological function will be included in gene lists designed to cast a wide immune-based net, such as the *DARC* gene. On the other hand, not all genes included even in a gene ontology list will have a clear connection with immune function based on a literature search. Many genes involved in immune function also have other important physiological roles. This means that top scoring genes should be double checked to see if they do have immune function based on the literature.

5.1.4 Enrichment tests and use of outlier approach

The use of Fisher's Exact test to ascertain whether certain classes of genes are overrepresented in top scoring windows and genes rests on the assumption that genes and windows are evenly distributed throughout the genome. However, it is possible that top scoring genes and/or windows are clustered together, causing enrichments to appear where there are none. This method does not take this bias into account, and future work might address this issue by using a different method (for example, the Gowinda package (Kofler and Schlotterer, 2012)) as well as ascertain whether gene clustering unfairly biased enrichment test results.

The use of the outlier approach to find potential targets of selection is also problematic in some instances. It is especially useful to look for enrichment of certain gene classes in the top results. However, it is not necessarily fair to assume that all results in the top one percent of statistic scores are actually under selection, or that every population has the same proportion of their genome (one percent) under positive and balancing selection. Issues with this outlier approach are highlighted in the balancing selection results with the removal of MHC genes. When these genes are removed from the analysis, the enrichment results

in this section change significantly. This suggests that the outlier approach can limit how informative results are.

5.2 Population selection histories

Examining the geographical distributions of private and shared selection signals between different regions and statistics can shed light on the regional specificity or lack thereof of selection pressures. In order to place the results of this thesis in context with previous work with regards to the distribution of global and local selection signals, it is useful here to compare the results from this thesis with the published results of studies which have used similar methodology. An exploration of these distributions and their implications are included in the following sections.

The following questions posed at the end of Chapter 1 are discussed in this section:

- What is the distribution of global and local signals in the results of each selection statistic, and how do those compare to the results of previous studies?

- Is there evidence of more selective pressure on immune genes in populations living in lower latitudes, where pathogen richness is greater?

5.2.1 Positive selection

A summary of the sharing of top selection signals between three commonly included broad continental regions in positive selection studies (Africa, East Asia, and Europe) are shown in Table 5.5 and are based on several previously published analyses (Frazer et al., 2007; Voight et al., 2006b). This table shows that across studies and selection statistics, the majority of top positive selection signals are private to each of the three continental groups. The number of private signals per region are fairly equal between the three different regions. There are a small number of shared signals between continental groups, and rarely a global shared signal.

Study	Voight et al. (2006b)	Frazer et al. (2007)	Frazer et al. (2007)
Selection statistic	iHS	LRH	iHS
Continental region	No. of signals (% of total signals)	No. of signals (% of total signals)	No. of signals (% of total signals)
A	243 (0.34)	37 (0.33)	26 (0.26)
EA	232 (0.32)	38 (0.34)	35 (0.35)
E	223 (0.31)	30 (0.27)	27 (0.27)
EA, A	4 (0.01)	2 (0.2)	3 (0.03)
E, A	6 (0.01)	2 (0.2)	1 (0.01)
E, EA	9 (0.01)	2 (0.2)	5 (0.05)
A, EA, E	0 (0)	1 (0.1)	3 (0.03)
Total top signals	717	112	100

Table 5.5 Sharing of top signals between geographic groups in previous whole-genome, multi-regional positive selection studies. Abbreviations: "A" = Africa, "EA" = East Asia, "E" = Europe. Both analyses are based on Yoruba (YRI), north and western European origin (CEU), and Han Chinese and Japanese (CHB + JPT) populations.

Table 5.6 shows the results from analyses in this thesis, and is designed to be as directly comparable to the results from previous studies, as shown in Table 5.5, as possible. Therefore, the same three continental regions are represented. The African continental region is represented by the West and Central African population, the European region by the Southwest European population, and the East Asian region by the Mainland East Asian population. Similarly to the distribution shown in Table 5.5, most of the top signals are specific to one continental region in particular, with some shared between two continental regions and very few appearing in all three. However, the Tajima's D results show slightly different patterns from the other two tests. Both the nSL and iHS results, like the results summarized in Table 5.5, show very similar numbers of private top selection hits. However, the Tajima's D results show more private results in the African continental region compared to the other two regions. Also, results for non-African signals shared between Europe and East Asia contain several times more signals than those of the nSL or iHS tests. This could reflect that the Tajima's D statistic detects older signals of selection than the nLS or iHS tests.

Statistic	nSL	iHS	Tajima's D
Continental region	No. of top wins. (% of total top wins.)	No. of top wins. (% of total top wins.)	No. of top wins. (% of total top wins.)
A	119 (0.32)	107 (0.29)	106 (0.28)
E	116 (0.31)	110 (0.29)	86 (0.23)
EA	115 (0.31)	109 (0.29)	89 (0.24)
A, E	2 (0.01)	7 (0.02)	9 (0.02)
A, EA	3 (0.01)	8 (0.02)	6 (0.02)
E, EA	6 (0.02)	5 (0.01)	26 (0.07)
A, E, EA	1 (0.01)	3 (0.01)	4 (0.01)
Total top wins.	375	375	375

Table 5.6 The distribution of private and shared top positive selection results. Continental populations were used to make this table as comparable to those based on previously published results as possible. Abbreviations: "A" = Africa, "EA" = East Asia, "E" = Europe. The African continental region is here represented by the West and Central African population, the European region by the Southwest European population, and the East Asian region by the Mainland Southeast Asian population. Total top windows here include only top widows from the AFR, SWE, and SEM populations.

When comparing the overlap of top signals between the studies represented in Table 5.5 to the analogous distribution of signals in the results of the analyses in this thesis (5.6), there was a small amount of overlap between the two. One reason for this may be that the analyses are based on different populations. The populations (AFR, SWE, SEM) used for comparison, while the closest matches in the EGDP dataset to those used in above the previously published analyses, do not match perfectly with the YRI (Yoruba in Ibadan, Nigeria), CHB (Han Chinese in Beijing, China) + JPT (Japanese in Tokyo, Japan), and CEU (Utah Residents (CEPH) with Northern and Western European Ancestry) HapMap populations used in the Voight et al. (2006b) and Frazer et al. (2007) analyses. Therefore, the comparison between Tables 5.5 and 5.6 is potentially most useful in looking at the relative quantitative distribution of signals between continental regions, instead of looking at overlap of specific top selection hits.

There were no overlaps of significant, globally shared windows between the analyses represented by these two tables. In the Frazer et al. (2007) study shown in Table 5.5, all three global signals (one is significant in both the iHS and the LRH tests) are all in non-genic regions. Genes in the three global windows in the iHS analysis from this thesis are *NDUFB3*,

ORC2, *CFLAR*, and *FAM126B*; *VPS13B*; and *PIH1D2*, *SDHD*, *C11orf57*, *TIMM8B*, *DLAT*, and *DIXDC1*. Genes from the one significant global window in the nSL analysis are *CROCC*, *FAM231A*, and *FAM231C*. Genes from the four significant global windows in the Tajima's D results are *PHKB* and *ITFG1*; and *DACH1*. None of these genes were highlighted in the results or discussion of this thesis as being especially relevant in immune function.

When discussing global distribution of shared and local selection signals, it is also interesting to look at signals that are specific to the African continental regions compared to potential "Out-of-Africa" (OOA) signals–those found in non-African regions and not found in African regions. As is clear from Tables 5.5 and 5.6, there are many more top signals that are specific to the African continental region than those shared between European and East Asian populations. Of signals specific to the African continental region, six 200 KB windows from this thesis overlap with top regions in the Frazer et al. (2007) analysis, and the only gene represented in those windows is *TMEFF2*. One 200 KB window overlaps with a top window in the Voight et al. (2006b) analysis, containing the gene *TNKS*.

OOA signal overlaps between the results from this thesis and the Frazer et al. (2007) analysis are sparse. One window, in the top results of both the nSL and iSH tests, overlaps with a top hit in the Frazer et al. (2007) analysis, but this region is non-genic. There are no overlaps in OOA signals between the Voight et al. (2006b) analysis and the results of this thesis. The most recognizable OOA signal from the results of the analyses in this thesis is the window containing *EDAR*. As noted in Chapter 1, the grouping of West and Central Africa into one population could have diluted real signals that are present in either population, and could account for the lack of overlapping results with other studies. Additionally, differentiation between populations could look like a false signal of balancing selection.

Other continental-specific overlaps are well-known selection hits as well, especially in the European continental region. In the European continental region, *LCT* and the surrounding region appear both in the top results of this thesis and of the Voight et al. (2006b) analysis, as well as *TYRP1* (Wilde et al., 2014). Both of these genes also appear in the top results of the Frazer et al. (2007) analysis.

5.2.2 Balancing selection

Similarly to the positive selection above, Table 5.7 provides a set of selection signal distributions with which to compare the results from this thesis. The studies represented in the table below are both whole-genome scans for balancing selection, as are the analyses performed in this thesis (Andres et al., 2009; Bitarello et al., 2018). Both of these studies looked at two main broad continental regions of ancestry: Africa and Europe. Table 5.7 shows different relative distributions of top signals between the two studies. The Bitarello et al.

(2018) study shows relatively even numbers of signals between the African and European populations, with the largest number of signals being shared between the two populations. Conversely, the Andres et al. (2009) paper shows the most top signals in the European American population.

Study	Bitarello et al. (2018)		Andres et al. (2009)
Selection statistic	NCD2		HKA + MWU
Continental region	No. of outlier wins. (% of total outlier wins.)	Population	No. of extreme genes (% of total extreme genes)
A	79 (0.29)	AA	15 (0.25)
E	84 (0.31)	EA	32 (0.53)
A, E	102 (0.38)	AA, EA	13 (0.21)
Total signals	265		60

Table 5.7 Sharing of top signals between geographic groups in two previous whole-genome, multi-regional balancing selection studies. Abbreviations: "A" = Africa, "E" = Europe, "AA" = African American and "EA" = European American.. Populations making up the African population in the Bitarello et al. (2018) paper are LWK (Luhya in Webuye, Kenya) and YRI (Yoruba in Ibadan, Nigeria), and those making up the European population are TSI (Toscani in Italia) and GBR (British in England and Scotland) 1000 Genome populations.

Table 5.8 shows the distribution of private and shared signals between broad continental regions based on the top 50 gene results for each of the balancing selection statistics in this thesis. It includes the East Asian region as well for ease of comparison with positive selection results. Each of the statistics shows a somewhat different distribution of signals between the three continental regions, though in each statistic the majority of signals are local to one region. The HKA results have the most shared signals, with a relatively high number of global signals. The β results have relatively fewer shared signals, though still more than the Tajima's D results, which show no globally shared signals.

Statistic	HKA	β	Tajima's D
Continental region	No. of top wins. (% of total top wins.)	No. of top wins. (% of total top wins.)	No. of top wins. (% of total top wins.)
A	22 (0.14)	32 (0.21)	42 (0.28)
E	23 (0.15)	31 (0.20)	36 (0.24)
EA	16 (0.01)	28 (0.18)	37 (0.22)
A, E	4 (0.02)	5 (0.03)	4 (0.02)
A, EA	11 (0.07)	8 (0.05)	3 (0.02)
E, EA	10 (0.06)	9 (0.06)	9 (0.06)
A, E, EA	13 (0.08)	5 (0.03)	0 (0)
Total top wins.	150	150	150

Table 5.8 The distribution of private and shared top balancing selection results from this thesis. Abbreviations: "A" = Africa, "EA" = East Asia, "E" = Europe. The African continental region is here represented by the West and Central African population, the European region by the Southwest European population, and the East Asian region by the Mainland Southeast Asian population. Total top windows here include only top widows from these three populations.

There were a relatively high number of top signals from the Andres et al. (2009) and Bitarello et al. (2018) studies that also appeared in the top results of the HKA, β and Tajima's D balancing selection results. However, the distribution of sharing between populations and the populations in which a signal was significant did not often match.

Genes appearing in all three global regions in the HKA results were *HLA-G*, *HLA-A*, *CDSN*, *TRIML1*, *HLA-B*, *PSORS1C1*, *ZNF766*, *MUC22*, *IGFBP7*, *CSMD1*, *RBFOX1*, *KRT40*, and *CLDN16*. In the β results, globally shared gene signals were *DEFB1*, *SMPDL3A*, *TESPA1*, *PSMG4*, and *CSMD1*. Several of these signals did appear as top hits in both the European and African populations in the Bitarello et al. (2018) analysis: *HLA-B*, *HLA-G*, *CDSN*, *PSORS1C1*, *RBFOX1*, and *CSMD1*.

Examining the sharing of balancing selection signals between continental regions highlights the utility of using multiple statistics for whole-genome scans. Several signals that appear to be either African-only or non-African only in one statistic are proved not to be by the results of another statistic. For example, based on the West and Central African, Southwest European, and Mainland Southeast Asian results of the HKA test, *DEFB1* appears to be an African-only signal. However, it appears in all three of those populations in the β results. *TESPA1* appears to be a non-African only signal based on the HKA results, but it

is a global signal in the β results. This suggests that signal sharing is likely more common than the results for each individual statistic as shown in Table 5.8 would suggest. Because of this, it is somewhat complex to say which signals are specific to a single region or shared between multiple regions, because the geographical distributions of signals rarely are similar between the different statistics. It should be noted that these sorts of comparisons depend on the somewhat arbitrary nature of choosing the cutoff percentage for top hits and non-top hits. Therefore, a gene that doesn't appear in the top results in one test may be outside, though near, the cutoff–therefore just because a gene is not significant in one population does not mean it is unique as a selection hit to another.

5.2.3 Comparison of distribution of selection signal sharing between positive and balancing selection results

The comparison of the distributions of sharing signals in Tables 5.6 and 5.8 highlights two points. First, in both types of selection, the majority of signals are local to the three continental regions. This suggests very broadly that the majority of adaptations potentially represented by these signals are relatively local adaptations. Second, the balancing selection signals, at least in the HKA and β results, are more highly shared in general than any of the positive selection signals. This could suggest similar selection pressures acting on widely geographically separated populations, or, more parsimoniously, older shared signals of selection that predate population splits into the populations we see today.

Note:

A graphical summary for each population of all significant enrichment signals from each selection statistic can be found in Appendix D.

5.2.4 Pathogen-driven selection and latitude

In order to address the question of whether populations at lower latitudes experienced more pathogen-driven selection, we can look to see whether the selection patterns of those populations appear to be different from those of populations at higher latitudes. The populations at the lowest latitudes are the West and Central African, South Asian, Mainland Southeast Asian, and Island Southeast Asian populations (please refer to Figure 1.6 for a map). The Northeast European and Northeast Siberian populations are high latitude populations for contrast. When comparing the enrichment results for these populations, neither set appears to have a defining set of characteristics, or have more total enrichments than

the others (for an overview of the total significant selection enrichments in every test in these populations, please see the tables corresponding to these populations in Appendix D). Therefore, based on the results of this thesis, no evidence for more pathogen-driven selection at lower latitudes was found. However, it should be noted that this approach focuses on the number of significant enrichments in immune gene categories, and not the magnitude of the selection. For example, one population may have experienced strong selection in only one immune gene category, whereas another might show weak selection signals in multiple gene categories. It is possible that future work focusing on this question might go about it another way and find a clear differentiation. However, tying pathogen concentration to latitude is complicated, since the richness of pathogen species is more closely correlated with variation of precipitation around the mean, according to Guernier et al. (2004). The authors suggest that disease dynamics are affected by not only the pathogen, but by any vectors and other hosts, complicating the picture further, and that additionally, this type of environmental prediction applies largely to pathogens with an external phase of the lifecycle, which are likely to be parasites. Many evolutionarily important pathogens lack an external phase of their lifecycle, and so are less influenced by the environment (Guernier et al., 2004).

Several other studies by Fumagalli et al., have examined the relationship between pathogen richness, selection, and geography, but have not found any correlation between variants under selection and latitude (Fumagalli et al., 2009a, 2010, 2009b, 2011).

5.3 Targets of selection

Past selection on genes and variants associated with disease resistance and the immune response helped populations survive their epidemiological environments and are still relevant today. In instances where disease pressures remain constant (for example, malaria in regions where it is endemic) these adaptations can still be beneficial. In other instances where the prevalence of selection-causing epidemics can be controlled with vaccines and other interventions, a more immediate effect of past adaptations can be their influence on inflammatory and autoimmune disease. Either way, looking for past adaptations can help us to better understand modern disease and find new variants that can shed more light on our complex immunological relationship with our environment. This can include finding host pathways important in infections and screening findings of other studies (review by Heyer and Quintana-Murci (2009)). Including diverse populations in future selection studies will be important in order to capture variants that could otherwise be missed and to build a more complete picture of human adaptation.

The following questions posed at the end of Chapter 1 are discussed in this section:

- Do different classes of immune genes show different selection histories (positive vs. balancing vs. none)?
- Do different classes of immune genes show different ages of selective pressure (ancient vs. recent)?
- What are some of the most striking targets of positive and balancing selection?

5.3.1 Discussion of targets of positive selection

Positive selection has shaped much of the human genome, and immune-related genes are often among the top hits in selection studies. However, there is a need for such results from populations that have not often been included in these studies. This thesis gave positive selection results for some populations that have been included previously, and some that have not, shedding new light on how pathogens and disease have shaped positive selection in human populations.

The synthesis of the results of the nSL, iHS, Tajima's D, and d_i tests allows broad conclusions to be drawn about the signatures of positive selection left in the thirteen populations included in this study. First, it is useful to compare the extent of positive selection on immune-related genes to that of genes in other functional categories. In the Supplementary Materials of Pagani et al. (2016), which included the same populations as this thesis (save the Colla), six different functional categories were considered for enrichment in top selection signals: pigmentation, thermoregulation, fatty acid metabolism, vasoconstriction/vasodilation, bacterial genes, and innate immune genes. The proportions of significant enrichments to non-significant enrichments based on the results of those enrichment analyses are 0.027, 0.083, 0.083, 0, 0.027, and 0.055, respectively. Based on these results, the two immune gene categories are neither the categories with the most enrichments nor the fewest. This suggests that immune genes are an important part of the results of those analyses, but not necessarily the most significant category of genes.

Table 5.9 shows how the results of each of the three window-based positive selection enrichment test results overlap. It shows that in most gene categories there is significant enrichment in many populations in at least one test. However, there are many fewer categories with significant enrichment in more than one test. Only several categories show significant enrichment in all three tests, such as the GO.Virus category in the Colla population and the antigen processing and presentation category in the Southwest European population. The HP.Bact category shows the most agreement between selection tests. Significant enrichment in multiple tests may be less likely to be a false positive result and potentially indicates real enrichment for positive selection in a given category of immune genes and population.

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
AFR											
WAA											
SWE											
ENE											
VOL											
SOA											
WSI											
SSI											
CSI											
NSI											
COL											
SEM											
SEA											

Table 5.9 Overlap of window-based positive selection enrichment test results. Pale orange shows significant enrichment in one test, medium orange in two tests, and maroon in all three tests.

The enrichment results can also be used to examine time frames of potential past selection categories at low resolution, since the Tajima's D statistic finds selection signals that are relatively ancient (older than 50,000 years) compared to those found by the iHS and nSL statistics (younger than 30,000 years) (Grossman et al., 2013). Table 5.10 shows the division between statistics that find relatively ancient (Tajima's D) versus relatively positive selection signals (nSL and iHS), with grey columns representing ancient and white columns representing recent selection. Some gene categories show a difference in the number of significant enrichments between these two categories. For example, GO.Bact shows more ancient significant enrichments than recent ones. GO.Tcell, GO.Bcell, GO.Adapt., and HP.Virus show more recent significant enrichments than ancient ones. Other categories show similar numbers of enrichments between the two time frames.

	Gene Ontology DB														Human Pathogen Interaction DB							
	Bact.		Virus		T cell		B cell		Innate		APP		Adapt.		Bact.		Virus		Amoe.		Prot.	
AFR			x				x				x				x	x						
WAA	x		x				x	x	x			x		x	x			x				
SWE	x		x	x				x			x	x			x	x	x	x				
ENE		x	x	x	x		x		x						x	x						
VOL	x		x	x				x	x	x					x	x		x				
SOA	x			x				x	x					x	x	x		x				
WSI		x		x				x		x					x			x				
SSI						x		x	x	x		x		x	x	x	x	x				
CSI	x					x		x	x	x				x	x	x		x				
NSI						x									x	x	x	x				
COL	x	x	x	x				x		x	x	x			x	x	x	x				
SEM				x		x				x					x							x
SEA	x	x				x		x	x		x	x		x	x	x	x					

Table 5.10 Ancient versus recent positive selection enrichment results. Ancient signals are represented by the grey columns and shows Tajima's D results. Recent signals are represented by white columns, and shows nSL and iHS results. An "x" signifies significant enrichment in that category.

The genes driving significant enrichment signals, seen in Tables 3.2, 3.4, 3.6, and 3.23, show that while genes driving significant enrichment signals are sometimes shared between populations, more often enrichment signals in an immune gene category are driven by different genes in each population. This suggests population-specific adaptations to particular environments, rather than shared adaptations either through common ancestry or convergent evolution. It may also be suggestive of polygenic adaptation, in that a number of genes, varying between populations, contribute to overall adaptation in a certain category. Because of this, future work might look more explicitly for evidence of polygenic selection in this dataset as it likely plays an important role in adaptation in the immune system.

A number of the genes that emerged as top selection targets in this study have been previously found to be top targets in other studies, though not always in the same populations. These include *EDAR* (Sabeti et al., 2007), *ITGAE* (The International HapMap Consortium, 2005), *IL1A* (Tang et al., 2007), *IL4* (Rockman et al., 2003), *CXCR4* (Tang et al., 2007), *ARPC1B* (Kimura et al., 2007), *RAG1* (Kelley, 2006), *CD5* (Carnero-Montoro et al., 2011), *DUOX2* (Kelley, 2006), *ITGAL* (Kimura et al., 2007), and *MSR1* (Kimura et al., 2007).

It is important to note here that references to malaria, a very strong selection pressure throughout human history, have been largely absent from results of this thesis. Even the HPI database list of genes associated with protozoa infections in humans was very short. This is surprising. It is highly likely that malaria's selective pressure is represented in the results of this thesis, but not classified as such because of the lack of enrichment in the genes

classified as interacting with that particular pathogen. However, there are several references to individual genes associated with malaria in both the positive and balancing selection results.

Summary of top SNP results for further study from window-based selection tests

The population-specific SNP tables in the window-based selection results contain a large number of genes, many of which have interesting immune function. However, many of the genes that appeared in the top one percent enrichment tables did not pass through the DIND and CADD filters into the individual top SNP results. This may be for several reasons. DIND discards SNPs with frequencies below forty percent and above ninety percent. Therefore, any SNPs at very high frequency or relatively low frequency will have been filtered out by this statistic. Additionally, the CADD filtering step likely filtered out some regulatory SNPs, since it has been shown to underperform on scoring regulatory variants (Gulko et al., 2015). Finally, a top window signal could have been driven by a SNP in a different, non-immune gene, or by demography and for this reason did not pass through the DIND filter.

Of the SNPs that did make it through the filters, several appear particularly interesting for further study based on derived allele frequency, CADD score, VEP consequence, absence or relative absence from previous selection studies, and immune function based on the literature. These SNPs are in the following genes:

- ***IL27* in the Northeast Siberian population:** This gene is of further interest for several reasons. It appears as a significant target of positive selection in a population that has seldom been included in genetic studies, and based on the results of a literature search this gene has not previously been reported to have been under selection in any human population. Finally, it has an important role in immune response to infection by a range of pathogenic organisms, a number of which that have had the potential to significantly shape human populations (for example, hepatitis C and influenza A). It also plays a role in modern inflammatory diseases such as asthma and inflammatory bowel disease. These associations point to *IL27* being a prime candidate for a gene that was target of past selection driven by infectious disease that has modern health implications relating to inflammatory or autoimmune disease. This SNP, rs181206, is at middling frequency in all European and Siberian populations, as well as in the Colla and South Asian populations, and is at relatively low frequency in the Mainland Southeast Asian and Island Southeast Asian populations.

- ***MAGI3* in the West and Central African population:** Because of its involvement in intestinal immunity and inflammation, *MAGI3* would make an interesting target for

further investigation. Based on a literature search, rs7543189, the SNP highlighted in the West and Central African population table above, has not been associated with any particular disease.

- ***MAST2* in the West Asia and Armenian population:** Despite middling allele frequencies across most populations, *MAST2* would be an interesting gene for follow up study because of its role in TLR signaling and immune response to infection, as well as its role in inflammatory disease such as Crohn's disease and rheumatoid arthritis. Based on a literature search, it is unclear what role the missense SNP rs1707336 plays in an infection. Because of similar allele frequencies across populations, it is likely that this particular SNP doesn't confer a region-specific benefit.

- ***CEP63* in the South Siberian and Mongolian population:** Because rs1127826 is a stop gain mutation, and because the derived allele frequency is relatively low in most populations and relatively high in only the South Siberia and Mongolian, Central Siberian, Northeast Siberian, and Colla populations, as well as its role in viral infections, *CEP63* and rs1127826 would be good candidates for further study.

- ***SYLT3* in the Colla population:** *SYTL3* is an interesting candidate for follow-up study because of its role in the movement of lytic granules to the immune synapse and its relatively high frequencies in Siberian, Asian, and Colla populations compared to other populations. The missense SNP represented here, rs901363, does not have any associations with disease at this time based on a literature search.

- ***SLC15A2* in the Southeast Asia Mainland population:** Because of the role of *SLC15A2* in the immune response to bacteria, especially airborne infections, this gene would make an interesting candidate for further study. It shows particularly high derived allele frequency in the Mainland Southeast Asian population and relatively low to middling derived allele frequencies in all other populations. rs2257212, the missense SNP, has been studied as a drug ADME (absorption, distribution, metabolism, excretion) SNP (Li et al., 2010a). Polymorphisms in this SNP are predicted to have a potential effect on drug action (Pinsonneault, 2004).

Summary of top results for further study from d_i test

Immune genes are well-represented in the top 50 genes in each population in the d_i results, not to mention in the results from the enrichment analysis. There are many immune-related, potentially interesting genes to choose from for future work. Perhaps the most straightforward way to narrow down this impractically long list of genes is to look for genes which appear in

the top selection results of window-based tests as well, and which appear from the literature to have relevant immune functions. These are shown in Table 5.11:

Test	Gene	Pop.	Immune Category
nSL, iHS	*IL4*	SWE	GO.Bcell
nSL, iHS	*ADRBK1*	COL	GO.Innate
nSL, iHS	*RPS6KB2*	COL	GO.Innate
nSL	*KARS*	SEM	GO.Virus
nSL	*NRG1*	SSI	GO.Innate
nSL	*TAB1*	SSI	GO.Innate
nSL	*KIF3C*	SSI	GO.APP
nSL	*HLA-F*	SEA	GO.APP
iHS	*RPL6*	SEM	GO.Virus
iHS	*PTPN11*	SEM	GO.Tcell
iHS	*PDCD6IP*	WSI	GO.Virus
iHS	*PSMB2*	MIE	GO.APP
iHS	*TYRO3*	COL	GO.Innate

Table 5.11 Gene overlaps between top results of window-based selection tests and top d_i results with relevant immune function based on a literature search

Of the genes above, several are especially interesting for follow-up study, based on their clear involvement in immune functions. These are:

- ***ADRBK1* in the Colla population:** As referenced in the Colla d_i results, ADRBK1 is induced by TLR2 activation and is associated with downregulation of chemokine receptors, resulting in impaired neutrophil recruitment and more severe sepsis in a mouse model. Based on this downregulation, this gene appears to have an important role in the balance of an immune reaction between being strong enough and not too strong.

- ***HLA-F* in the Island Southeast Asian population:** As mentioned above, *HLA-F* is involved in antigen presentation, as well as surveillance of stressed cells by the innate immune system. Based on a literature search, there are few known associations with susceptibility to either infectious or inflammatory disease associated with this gene.

- ***TYRO3* in the Colla population:** As described in the iHS enrichment section, TYRO3 is an immunoregulatory receptor tyrosine kinase that is important for preventing

autoimmunity in mice. TYRO3 also plays an inhibitory role in type 2 immunity, helping to control the magnitude of the immune reaction. It has been suggested as a drug target–for example, blocking TYRO3 could lead to improved helminth expulsion, or activating TYRO3 could lessen allergic response (Chan et al., 2016).

The genes listed above are very few compared to those in the top one percent results per population, including those highlighted in the description of the enrichment table. Again, there are too many there to practically follow up on, but very likely at least some of those are worthy of further study as well. For example, a number of complement genes were represented in those results, and it would be interesting to further investigate extreme differentiation in those genes between populations.

Future work on positive selection targets

Future work could focus on the genes listed in the above sections on the top hits from the window-based and d_i results. The most straightforward results to follow up on would be specific SNPs, such as those stemming from the DIND analysis, since those can more easily be studied than an entire gene. However, since the d_i analysis was originally performed on SNPs, it would be fairly straightforward to select SNPs with high scores from that analysis for further research. There are two main types of future work stemming from these SNPs that would be interesting to pursue. The first is to ascertain whether any of the most promising SNPs from the analyses in this chapter are associated with immune phenotypes through genome-wide association studies. If there are significant associations between a SNP and an immune phenotype, it would give credence to the idea of that SNP having evolutionary importance in that phenotype. The other type of analysis to pursue would be functional analysis to find out what effect, if any, any of the SNPs chosen for further study have on a given protein and downstream phenotypes. This could take the shape of computer-based protein structure modeling or studies in model organisms.

5.3.2 Discussion of targets of balancing selection

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
AFR											
WAA											
SWE											
ENE											
VOL											
SOA											
WSI											
SSI											
CSI											
NSI											
COL											
SEM											
SEA											

Table 5.12 Overlap of significant balancing selection enrichment signals. Pale orange shows one significant enrichment, medium shows two, and dark orange shows significant enrichments in all three statistics.

Table 5.12 shows which classes of immune gene appear to be most targeted by balancing selection. Every single population is significantly enriched in at least one, and often two or even three, selection tests in the categories of T cell and antigen processing and presentation genes. These are likely driven by the very strong representation of HLA genes in the results. In particular, the West and Central African population is enriched in both of these categories in all three tests. These categories make sense as targets of balancing selection, since diversity in antigen processing and presentation cells allows for the immune system to mount an attack against a wider range of pathogens. There are several immune gene categories, such as GO.Bact, GO.Virus, and GO.Innate, that show varied levels of enrichment among populations instead of the unanimous enrichment in the GO.Tcell and APP categories. This may suggest that evolutionary pressures causing balancing selection act less often or strongly on these categories of genes. There is a notable absence of enrichments in the gene classes defined by the HPIDB. It is not clear why this is the case.

Enrichment tests were completed for each balancing selection statistic with and without the inclusion of MHC genes. The overlap of significant enrichment signals excluding MHC

genes can be seen in Table 5.13. Just as the enrichment signals for individual statistics were different when the MHC was excluded, the pattern of significant enrichments looks different in this table than it does in Table 5.12.

	Gene Ontology DB							HPI DB			
	Bact.	Virus	T cell	B cell	Innate	APP	Adapt.	Bact.	Virus	Amoe.	Prot.
AFR											
WAA											
SWE											
ENE											
VOL											
SOA											
WSI											
SSI											
CSI											
NSI											
COL											
SEM											
SEA											

Table 5.13 Overlap of significant balancing selection enrichment signals, not including genes in the MHC. Pale orange shows one significant enrichment, medium shows two, and dark orange shows significant enrichments in all three statistics.

Most glaringly, all but one of the enrichment signals in the antigen processing and presentation category have disappeared. This is a category that was enriched in every single population in at least one test with the inclusion of MHC genes, showing the importance of those genes in the balancing selection results. The enrichments in the genes associated with T cell function have largely remained, as have those in the GO.Bact category. There are in fact more significant enrichments in the GO.Innate category when MHC genes are excluded from the analysis.

In general, many of the top gene results in each of the three selection tests were shared between populations to a large degree. This is also reflected in some gene classes in the enrichment results summarized in Table 5.12. This sharing of signals could either reflect similar selection pressures in all populations, ancient signals in ancestral groups that then split into the modern groups seen here, or some combination of both phenomena. Certainly with the ideal timeframe given by the authors of the β statistic, the results of that test may well suggest selection on a population that predates splits into modern continental groups. It

is most parsimonious to assume that signals shared between a large number of populations, especially populations that are separated by large geographical distances, are those that predate the split of those populations. Signals shared by a small number of populations that are geographically near may reflect a more localized selection pressure.

Several of the top results from this study confirm previous findings from other studies. Among these are *HLA-B* (Andres et al., 2009), *DEFB1* (Cagliani et al., 2008), *TLR6* (Ferrer-Admetlla et al., 2008), *CDSN* (Andres et al., 2009), and *LGALS8* (Andres et al., 2009). A fair number of genes in the top results of this study appear not to have been previously found as targets of balancing selection, and may be interesting for future inspection. Several of these are listed below:

- ***PGLYRP4*:** *PGLYRP4* was in the top one percent of results for both the HKA test (in all populations but West and Central Africa) as well as in the β test (in all populations save West and Central Africa, Colla, Mainland Southeast Asia, and Island Southeast Asia). PGLYRP4 is a peptidoglycan recognition protein that is expressed in the skin, mucous membranes, the mouth, and the gastrointestinal tract (see review by Dziarski and Gupta (2010)). It has a bactericidal response against both Gram-negative and Gram-positive bacteria (Lu et al., 2005). Given the wide range of populations in which this gene was a top balancing selection target, it is likely that this gene was under selection in ancestral populations in response to pressure from bacterial infections.

- ***PGLYRP2*:** *PGLYRP2* was in the enrichment results for the West and Central African population in the Tajima's D test, though not in the top 50 genes for that population or any other. It was also not in the top one percent of results for any population in the HKA or the β test. It is an amidase that digests peptidoglycan (Gelius et al., 2003). It is constitutively expressed in the liver (Liu et al., 2001), but is expressed in keratinocytes and corneal epithelial cells upon their exposure to both Gram-positive and -negative bacteria, as well as cytokines. It is thought that the digestion of peptidoglycan could lead either to reduction of the proinflammatory nature of peptidoglycan, via PGLYRP2 acting as a scavenger for this molecule, or to enhanced antimicrobial defense (Wang et al., 2005). Despite its similar name to the gene above, *PGLYRP2* is located on chromosome 19, whereas *PGLYRP4* is located on chromosome 1.

- ***GNLY*:** *GNLY*, or granulysin, was in the enriched GO.Bact results for the β statistic in the Island Southeast Asian population and was in the top 50 results for the β statistic in the West Siberian population. It is not in the top one percent of results for any population in the HKA test. GNLY releases bactericidal granzymes into the cytoplasm of infected cells (Walch et al., 2014), and has been suggested as a measure

of overall cellular immunity (Ogawa et al., 2003). In one study, individuals with latent tuberculosis had lower levels of serum granulysin than healthy individuals (Thuong et al., 2016).

- ***DMBT1*:** *DMBT1* was in the top 50 results in the South Asian, South Siberia and Mongolian, Central Siberian, Northeast Siberian, and Mainland Southeast Asian populations in the HKA test, and was in the top one percent of results in the Volga Uralic, West Siberian, Colla, and Island Southeast Asian population in the HKA test. It was in the top 50 results of the Southwest European, Northeast European, Volga Uralic, South Asian, West Siberian, South Siberia and Mongolian, Central Siberian, Northeast Siberian, Colla, Mainland Southeast Asian, and Island Southeast Asian populations in the β results, and in the top one percent of results for all other populations. DMBT1 is involved in the mucosal innate immune response via sensitivity to bacterial cell wall antigens. In the intestinal epithelia, which is an important tissue in terms of being an immunological barrier to pathogenic organisms, *DMBT1* expression is upregulated upon exposure to LPS and may be implicated in Crohn's disease (Rosenstiel et al., 2007). There are two other variants of *DMBT1*. DMBT1^{gp340} is the variant of DMBT1 involved in respiratory defense and DMBT1^{SAG}, or salivary agglutinin, is the variant involved in oral defense (Ligtenberg et al., 2001). DMBT1^{SAG} binds a variety of bacteria, including *Staphylococcus aureus*, *E. coli*, and *H. pylori* (Bikker et al., 2002). *DMBT1* copy number is correlated with population history of agriculture, via increase in dietary carbohydrates and the associated dental carie-causing bacteria *Streptococcus mutans* (Polley et al., 2015). This gene also plays a role in epithelial differentiation (Mollenhauer et al., 2000).

- ***TLR2*:** *TLR2* is in the enrichment results for the Tajima's D test in the Volga Uralic population. It did not appear in the top one percent of results in any population in either the HKA or the β results. TLR2 is a Toll-like receptor that can recognizes a broad range of pathogenic organisms including Gram-positive and -negative bacteria, mycobacteria, spirochetes, trypanosomes, fungi, and more (see review by Takeda et al. (2003)). *TLR2* has been associated with susceptibility to tuberculosis (Ogus et al., 2004).

The individual genes listed above are potential new contributions to the field of knowledge of immune genes that have been under balancing selection in humans. All are involved in not only the immune response but also have been implicated in various inflammatory diseases. These dual associations highlight the delicate balance that must be struck by the immune system in terms of when and how much of a response to mount. This theme

also applies to appropriate calibration of an immune response during an infection–if the immune system gives too small a response, the infection won't be cleared, but if it gives an overreaction, it leads to damage to the host. One of the conditions that came up many times in the literature search in association with top balancing selection hits was sepsis. This is a condition in which the inflammation caused by the immune system to clear a bacterial infection becomes dysregulated and causes tissue damage and can lead to death (Cohen, 2002; Thimmulappa, 2006). Not all of the host factors involved in the development of this condition are understood (Cohen, 2002; Thimmulappa, 2006). It is possible that studies such as this one, based on finding immune genes that appear to have been evolutionarily important in response to infection, could be helpful in further elucidating the mechanisms of such conditions that rely on appropriate balance of inflammation and tolerance. Since sepsis and other inflammatory/autoimmune diseases were strongly associated with the top results from the three selection tests used in this project, in future it could be interesting to assemble a gene list for a set of conditions of interest (such as sepsis and inflammatory disease) and look for evidence of stronger balancing selection on the set of genes as whole compared to sets of genes associated with conditions/diseases of other etiology.

5.3.3 Comparison of targets of positive versus balancing selection

We can look at Tables 5.9, 5.12, and 5.13 to compare overall enrichments in different gene categories. Perhaps the most striking difference between enrichments in positive and balancing selection are those in the GO.Virus category. There are many significant enrichments in this category in the positive selection results, but far fewer in the balancing selection results. This is in contrast to the GO.Bact category, which is significantly enriched in the results of both types of selection. This suggests that both single variants and diversity at a locus have been adaptive in defense against bacterial pathogens, but that single adaptive variants have been more under selection due to pressure from viral infections.

Another significant difference between the enrichments results of the positive and balancing selection results are those in the antigen processing and presentation category. In the positive selection results, there are enrichments in the West and Central African, West Asia and Armenian, and Southwest European populations, but most of the populations in the rest of the Eurasian continent show no enrichment. On the other hand, in the balancing selection results that include MHC genes (Table 5.12), every single population is significantly enriched for genes in this category. As Table 5.13 shows, these enrichments are almost completely driven by genes in the MHC. These results are suggestive of the significance of this class of immune genes as targets of balancing selection.

Some temporal patterns of selection between the different gene categories can also been seen from the enrichment results. In the positive selection results (see Table 5.10), using Tajima's D as a measure of ancient selection and nSL and iHS as a measure of recent selection, the GO.Bact category shows more ancient significant enrichments than recent ones. The GO.Tcell, GO.Bcell, GO.Adapt., and HP.Virus categories show more recent significant enrichments than ancient ones. Using β as a measure of extremely ancient selection, again the GO.Bact category is enriched in 12 of 13 populations, and the T cell category is enriched in 11 of 13 (enrichments not including MHC genes). Together, these results suggest that disease caused by bacteria likely exerted more early evolutionary pressure on humans than did viruses. This conclusion supports the knowledge that the major viral epidemics that have shaped human immunity require relatively large population groups, which did not exist until recently, after the first epidemiological transition in the Neolithic.

In terms of targets of selection, in future work it could be of interest to include negative selection in a study like this in order to see the contrasts between which genes have been under each type of selection. It would also be interesting to focus more on selection in regulatory regions, since they are likely to play an important role in immune response.

References

Ackerman, H., Usen, S., Jallow, M., Sisay-Joof, F., Pinder, M., and Kwiatkowski, D. P. (2005). A comparison of case-control and family-based association methods: The example of sickle-cell and malaria. *Annals of Human Genetics*, 69(5):559–565.

Adhikari, K., Fontanil, T., Cal, S., Mendoza-Revilla, J., Fuentes-Guajardo, M., Chacón-Duque, J.-C., Al-Saadi, F., Johansson, J. A., Quinto-Sanchez, M., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R. B., Pérez, G. M., Gómez-Valdés, J., Villamil-Ramírez, H., Hunemeier, T., Ramallo, V., de Cerqueira, C. C. S., Hurtado, M., Villegas, V., Granja, V., Gallo, C., Poletti, G., Schuler-Faccini, L., Salzano, F. M., Bortolini, M.-C., Canizales-Quinteros, S., Rothhammer, F., Bedoya, G., Gonzalez-José, R., Headon, D., López-Otín, C., Tobin, D. J., Balding, D., and Ruiz-Linares, A. (2016). A genome-wide association scan in admixed latin americans identifies loci influencing facial and scalp hair features. *Nature Communications*, 7:10815.

Akey, J. M. (2009). Constructing genomic maps of positive selection in humans: Where do we go from here? *Genome Research*, 19(5):711–722.

Akey, J. M., Ruhe, A. L., Akey, D. T., Wong, A. K., Connelly, C. F., Madeoy, J., Nicholas, T. J., and Neff, M. W. (2010). Tracking footprints of artificial selection in the dog genome. *Proceedings of the National Academy of Sciences*, 107(3):1160–1165.

Akiyama, K., Yokota, K., Kagawa, S., Shimbara, N., Tamura, T., Akioka, H., Nothwang, H., Noda, C., Tanaka, K., and Ichihara, A. (1994). cdna cloning and interferon gamma down-regulation of proteasomal subunits x and y. *Science*, 265(5176):1231–1234.

Ali, K., Bilancio, A., Thomas, M., Pearce, W., Gilfillan, A. M., Tkaczyk, C., Kuehn, N., Gray, A., Giddings, J., Peskett, E., and et al. (2004). Essential role for the p110δ phosphoinositide 3-kinase in the allergic response. *Nature*, 431(7011):1007–1011.

Ali, K., Soond, D. R., Piñeiro, R., Hagemann, T., Pearce, W., Lim, E. L., Bouabe, H., Scudamore, C. L., Hancox, T., Maecker, H., and et al. (2014). Inactivation of pi(3)k p110δ breaks regulatory t-cell-mediated immune tolerance to cancer. *Nature*, 510(7505):407–411.

Ali, S., Hirschfeld, A. F., Mayer, M. L., Fortuno, E. S., Corbett, N., Kaplan, M., Wang, S., Schneiderman, J., Fjell, C. D., Yan, J., and et al. (2013). Functional genetic variation in nfkbia and susceptibility to childhood asthma, bronchiolitis, and bronchopulmonary dysplasia. *The Journal of Immunology*, 190(8):3949–3958.

Allison, A. (1954). Protection afforded by sickle-cell trait against subtertian malareal infection. *BMJ*, 1(4857):290–294.

Allison, A. C. (2009). Genetic control of resistance to human malaria. *Current Opinion in Immunology*, 21(5):499–505.

Alper, S., Laws, R., Lackford, B., Boyd, W. A., Dunlap, P., Freedman, J. H., and Schwartz, D. A. (2008). Identification of innate immunity genes and pathways using a comparative genomics approach. *Proceedings of the National Academy of Sciences*, 105(19):7016–7021.

Alves-Filho, J. C., Freitas, A., Souto, F. O., Spiller, F., Paula-Neto, H., Silva, J. S., Gazzinelli, R. T., Teixeira, M. M., Ferreira, S. H., and Cunha, F. Q. (2009). Regulation of chemokine receptor by toll-like receptor 2 is critical to neutrophil migration and resistance to polymicrobial sepsis. *Proceedings of the National Academy of Sciences*, 106(10):4018–4023.

Alves-Filho, J. C., Sônego, F., Souto, F. O., Freitas, A., Verri, W. A., Auxiliadora-Martins, M., Basile-Filho, A., McKenzie, A. N., Xu, D., Cunha, F. Q., and et al. (2010). Interleukin-33 attenuates sepsis by enhancing neutrophil influx to the site of infection. *Nature Medicine*, 16(6):708–712.

Ambrose, R. L. and Mackenzie, J. M. (2012). Atf6 signaling is required for efficient west nile virus replication by promoting cell survival and inhibition of innate immune responses. *Journal of Virology*, 87(4):2206–2214.

Ammari, M. G., Gresham, C. R., McCarthy, F. M., and Nanduri, B. (2016). Hpidb 2.0: a curated database for host–pathogen interactions. *Database*, 2016:baw103.

An, P., Winkler, C., Guan, L., O'Brien, S. J., and Zeng, Z. (2011). A common hla–dpa1 variant is a major determinant of hepatitis b virus clearance in han chinese. *The Journal of Infectious Diseases*, 203(7):943–947.

Andrade, W. A., Silva, A. M., Alves, V. S., Salgado, A. P. C., Melo, M. B., Andrade, H. M., Dall'Orto, F. V., Garcia, S. A., Silveira, T. N., and Gazzinelli, R. T. (2010). Early endosome localization and activity of rasgef1b, a toll-like receptor-inducible ras guanine-nucleotide exchange factor. *Genes Immunity*, 11(6):447–457.

Andrés, A. M., Hubisz, M. J., Indap, A., Torgerson, D. G., Degenhardt, J. D., Boyko, A. R., Gutenkunst, R. N., White, T. J., Green, E. D., Bustamante, C. D., Clark, A. G., and Nielsen, R. (2009). Targets of Balancing Selection in the Human Genome. *Molecular Biology and Evolution*, 26(12):2755–2764.

Andres, A. M., Hubisz, M. J., Indap, A., Torgerson, D. G., Degenhardt, J. D., Boyko, A. R., Gutenkunst, R. N., White, T. J., Green, E. D., Bustamante, C. D., and et al. (2009). Targets of balancing selection in the human genome. *Molecular Biology and Evolution*, 26(12):2755–2764.

Annacker, O., Coombes, J. L., Malmstrom, V., Uhlig, H. H., Bourne, T., Johansson-Lindbom, B., Agace, W. W., Parker, C. M., and Powrie, F. (2005). Essential role for cd103 in the t cell–mediated regulation of experimental colitis. *The Journal of Experimental Medicine*, 202(8):1051–1061.

Arbibe, L., Kim, D. W., Batsche, E., Pedron, T., Mateescu, B., Muchardt, C., Parsot, C., and Sansonetti, P. J. (2006). An injected bacterial effector targets chromatin access for transcription factor nf-κb to alter transcription of host genes involved in immune responses. *Nature Immunology*, 8(1):47–56.

Archer, S. (2016). Colonialism and other afflictions: Rethinking native american health history. *History Compass*, 14(10):511–521.

Arimoto, K.-i., Funami, K., Saeki, Y., Tanaka, K., Okawa, K., Takeuchi, O., Akira, S., Murakami, Y., and Shimotohno, K. (2010). Polyubiquitin conjugation to nemo by triparite motif protein 23 (trim23) is critical in antiviral defense. *Proceedings of the National Academy of Sciences*, 107(36):15856–15861.

Arimoto, K.-i., Takahashi, H., Hishiki, T., Konishi, H., Fujita, T., and Shimotohno, K. (2007). Negative regulation of the rig-i signaling by the ubiquitin ligase rnf125. *Proceedings of the National Academy of Sciences*, 104(18):7500–7505.

Armelagos, G. J., Brown, P. J., and Turner, B. (2005). Evolutionary, historical and political economic perspectives on health and disease. *Social Science Medicine*, 61(4):755–765.

Arnold, C. N., Pirie, E., Dosenovic, P., McInerney, G. M., Xia, Y., Wang, N., Li, X., Siggs, O. M., Karlsson Hedestam, G. B., and Beutler, B. (2012). A forward genetic screen reveals roles for nfkbid, zeb1, and ruvbl2 in humoral immunity. *Proceedings of the National Academy of Sciences*, 109(31):12286–12293.

Artis, D., Villarino, A., Silverman, M., He, W., Thornton, E. M., Mu, S., Summer, S., Covey, T. M., Huang, E., Yoshida, H., Koretzky, G., Goldschmidt, M., Wu, G. D., Sauvage, F. d., Miller, H. R. P., Saris, C. J. M., Scott, P., and Hunter, C. A. (2004a). The IL-27 Receptor (WSX-1) Is an Inhibitor of Innate and Adaptive Elements of Type 2 Immunity. *The Journal of Immunology*, 173(9):5626–5634.

Artis, D., Wang, M. L., Keilbaugh, S. A., He, W., Brenes, M., Swain, G. P., Knight, P. A., Donaldson, D. D., Lazar, M. A., Miller, H. R. P., and et al. (2004b). Relm /fizz2 is a goblet cell-specific immune-effector molecule in the gastrointestinal tract. *Proceedings of the National Academy of Sciences*, 101(37):13596–13600.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarkis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29.

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., and et al. (2015). A global reference for human genetic variation. *Nature*, 526(7571):68–74.

Avula, L. R., Buckinx, R., Alpaerts, K., Costagliola, A., Adriaensen, D., Van Nassauw, L., and Timmermans, J.-P. (2011). The effect of inflammation on the expression and distribution of the mas-related gene receptors mrge and mrgf in the murine ileum. *Histochemistry and Cell Biology*, 136(5):569–585.

Backes, P., Quinkert, D., Reiss, S., Binder, M., Zayas, M., Rescher, U., Gerke, V., Bartenschlager, R., and Lohmann, V. (2010). Role of annexin a2 in the production of infectious hepatitis c virus particles. *Journal of Virology*, 84(11):5775–5789.

Bai, X., Kim, S.-H., Azam, T., McGibney, M. T., Huang, H., Dinarello, C. A., and Chan, E. D. (2010). Il-32 is a host protective cytokine against mycobacterium tuberculosis in differentiated thp-1 human macrophages. *The Journal of Immunology*, 184(7):3830–3840.

Bancerek, J., Poss, Z. C., Steinparzer, I., Sedlyarov, V., Pfaffenwimmer, T., Mikulic, I., Dölken, L., Strobl, B., Müller, M., Taatjes, D. J., and et al. (2013). Cdk8 kinase phosphorylates transcription factor stat1 to selectively regulate the interferon response. *Immunity*, 38(2):250–262.

Bandala-Sanchez, E., Zhang, Y., Reinwald, S., Dromey, J. A., Lee, B.-H., Qian, J., Böhmer, R. M., and Harrison, L. C. (2013). T cell regulation mediated by interaction of soluble cd52 with the inhibitory receptor siglec-10. *Nature Immunology*, 14(7):741–748.

Baranova, I. N., Kurlander, R., Bocharov, A. V., Vishnyakova, T. G., Chen, Z., Remaley, A. T., Csako, G., Patterson, A. P., and Eggerman, T. L. (2008). Role of human cd36 in bacterial recognition, phagocytosis, and pathogen-induced jnk-mediated signaling. *The Journal of Immunology*, 181(10):7147–7156.

Barreiro, L. B., Ben-Ali, M., Quach, H., Laval, G., Patin, E., Pickrell, J. K., Bouchier, C., Tichit, M., Neyrolles, O., Gicquel, B., Kidd, J. R., Kidd, K. K., Alcaïs, A., Ragimbeau, J., Pellegrini, S., Abel, L., Casanova, J.-L., and Quintana-Murci, L. (2009). Evolutionary Dynamics of Human Toll-Like Receptors and Their Different Contributions to Host Defense. *PLoS Genet*, 5(7):e1000562.

Barreiro, L. B., Patin, E., Neyrolles, O., Cann, H. M., Gicquel, B., and Quintana-Murci, L. (2005). The heritage of pathogen pressures and ancient demography in the human innate-immunity cd209/cd209l region. *The American Journal of Human Genetics*, 77(5):869–886.

Barreiro, L. B. and Quintana-Murci, L. (2010). From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nature Reviews Genetics*, 11(1):17–30.

Bassing, C. H., Swat, W., and Alt, F. W. (2002). The mechanism and regulation of chromosomal v(d)j recombination. *Cell*, 109(2):S45–S55.

Beauchamp, N. M., Busick, R. Y., and Alexander-Miller, M. A. (2010). Functional divergence among cd103+ dendritic cell subpopulations following pulmonary poxvirus infection. *Journal of Virology*, 84(19):10191–10199.

Begue, B., Verdier, J., Rieux-Laucat, F., Goulet, O., Morali, A., Canioni, D., Hugot, J.-P., Daussy, C., Verkarre, V., Pigneur, B., and et al. (2011). Defective il10 signaling defining a subgroup of patients with inflammatory bowel disease. *The American Journal of Gastroenterology*, 106(8):1544–1555.

Beissert, S., Schwarz, A., and Schwarz, T. (2006). Regulatory t cells. *Journal of Investigative Dermatology*, 126(1):15–24.

Bekpen, C., Marques-Bonet, T., Alkan, C., Antonacci, F., Leogrande, M. B., Ventura, M., Kidd, J. M., Siswara, P., Howard, J. C., and Eichler, E. E. (2009). Death and resurrection of the human irgm gene. *PLoS Genetics*, 5(3):e1000403.

Bekpen, C., Xavier, R. J., and Eichler, E. E. (2010). Human irgm gene "to be or not to be". *Seminars in Immunopathology*, 32(4):437–444.

Ben-Zvi, A., Lacoste, B., Kur, E., Andreone, B. J., Mayshar, Y., Yan, H., and Gu, C. (2014). Mfsd2a is critical for the formation and function of the blood–brain barrier. *Nature*, 509(7501):507–511.

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J. M., Mei, R., Scherer, S. W., Julian, C. G., Wilson, M. J., López Herráez, D., and et al. (2010). Identifying signatures of natural selection in tibetan and andean populations using dense genome scan data. *PLoS Genetics*, 6(9):e1001116.

Bikker, F. J., Ligtenberg, A. J. M., Nazmi, K., Veerman, E. C. I., van't Hof, W., Bolscher, J. G. M., Poustka, A., Amerongen, A. V. N., and Mollenhauer, J. (2002). Identification of the bacteria-binding peptide domain on salivary agglutinin (gp-340/dmbt1), a member of the scavenger receptor cysteine-rich superfamily. *Journal of Biological Chemistry*, 277(35):32109–32115.

Bingle, C. D. (2002). Plunc: A novel family of candidate host defence proteins expressed in the upper airways and nasopharynx. *Human Molecular Genetics*, 11(8):937–943.

Bingle, C. D., Wilson, K., Lunn, H., Barnes, F. A., High, A. S., Wallace, W. A., Rassl, D., Campos, M. A., Ribeiro, M., and Bingle, L. (2010). Human lplunc1 is a secreted product of goblet cells and minor glands of the respiratory and upper aerodigestive tracts. *Histochemistry and Cell Biology*, 133(5):505–515.

Bingle, L., Wilson, K., Musa, M., Araujo, B., Rassl, D., Wallace, W. A., LeClair, E. E., Mauad, T., Zhou, Z., Mall, M. A., and et al. (2012). Bpifb1 (lplunc1) is upregulated in cystic fibrosis lung disease. *Histochemistry and Cell Biology*, 138(5):749–758.

Bitarello, B. D., de Filippo, C., Teixeira, J. C., Schmidt, J. M., Kleinert, P., Meyer, D., and Andrés, A. M. (2018). Signatures of long-term balancing selection in human genomes. *Genome Biology and Evolution*, 10(3):939–955.

Bittner, S., Ruck, T., Schuhmann, M. K., Herrmann, A. M., Maati, H. M. o., Bobak, N., Göbel, K., Langhauser, F., Stegner, D., Ehling, P., and et al. (2013). Endothelial twik-related potassium channel-1 (trek1) regulates immune-cell trafficking into the cns. *Nature Medicine*, 19(9):1161–1165.

Blanchard, N., Di Bartolo, V., and Hivroz, C. (2002). In the immune synapse, zap-70 controls t cell polarization and recruitment of signaling proteins but not formation of the synaptic pattern. *Immunity*, 17(4):389–399.

Blaser, C., Kaufmann, M., and Pircher, H. (1998). Cutting edge: Virus-activated cd8 t cells and lymphokine-activated nk cells express the mast cell function-associated antigen, an inhibitory c-type lectin. *The Journal of Immunology*, 161:6451–6454.

Blaser, M. J. and Kirschner, D. (2007). The equilibria that allow bacterial persistence in human hosts. *Nature*, 449(7164):843–849.

Blumberg, H., Dinh, H., Trueblood, E. S., Pretorius, J., Kugler, D., Weng, N., Kanaly, S. T., Towne, J. E., Willis, C. R., Kuechle, M. K., and et al. (2007). Opposing activities of two novel members of the il-1 ligand family regulate skin inflammation. *The Journal of Experimental Medicine*, 204(11):2603–2614.

Bohlson, S. S., Fraser, D. A., and Tenner, A. J. (2007). Complement proteins c1q and mbl are pattern recognition molecules that signal immediate and long-term protective immune functions. *Molecular Immunology*, 44(1-3):33–43.

Borges, L. (2002). Lir9, an immunoglobulin-superfamily-activating receptor, is expressed as a transmembrane and as a secreted molecule. *Blood*, 101(4):1484–1486.

Borm, M. E. A., Bodegraven, A. A., Mulder, C. J. J., Kraal, G., and Bouma, G. (2005). A nfkb1 promoter polymorphism is involved in susceptibility to ulcerative colitis. *International Journal of Immunogenetics*, 32(6):401–405.

Botchkarev, V. A. and Fessing, M. Y. (2005). Edar signaling in the control of hair follicle development. *Journal of Investigative Dermatology Symposium Proceedings*, 10(3):247–251.

Bottini, N., Vang, T., Cucca, F., and Mustelin, T. (2006). Role of ptpn22 in type 1 diabetes and other autoimmune diseases. *Seminars in Immunology*, 18(4):207–213.

Boyko, A. R., Williamson, S. H., Indap, A. R., Degenhardt, J. D., Hernandez, R. D., Lohmueller, K. E., Adams, M. D., Schmidt, S., Sninsky, J. J., Sunyaev, S. R., White, T. J., Nielsen, R., Clark, A. G., and Bustamante, C. D. (2008). Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genetics*, 4(5):e1000083.

Braida, L., Boniotto, M., Pontillo, A., Tovo, P. A., Amoroso, A., and Crovella, S. (2004). A single-nucleotide polymorphism in the human beta-defensin 1 gene is associated with hiv-1 infection in italian children. *AIDS*, 18(11):1598–1600.

Brauer, P. M., Pessach, I. M., Clarke, E., Rowe, J. H., Ott de Bruin, L., Lee, Y. N., Dominguez-Brauer, C., Comeau, A. M., Awong, G., Felgentreff, K., and et al. (2016). Modeling altered t-cell development with induced pluripotent stem cells from patients with rag1 -dependent immune deficiencies. *Blood*, 128(6):783–793.

Brdička, T., Imrich, M., Angelisová, P., Brdičková, N., Horváth, O., Špička, J., Hilgert, I., Lusková, P., Dráber, P., Novák, P., and et al. (2002). Non–t cell activation linker (ntal). *The Journal of Experimental Medicine*, 196(12):1617–1626.

Brown, D., Trowsdale, J., and Allen, R. (2004). The lilr family: modulators of innate and adaptive immune pathways in health and disease. *Tissue Antigens*, 64(3):215–225.

Brown, D. P., Jones, D. C., Anderson, K. J., Lapaque, N., Buerki, R. A., Trowsdale, J., and Allen, R. L. (2009). The inhibitory receptor lilrb4 (ilt3) modulates antigen presenting cell phenotype and, along with lilrb2 (ilt4), is upregulated in response to salmonella infection. *BMC Immunology*, 10(1):56.

Brown, J. N., Palermo, R. E., Baskin, C. R., Gritsenko, M., Sabourin, P. J., Long, J. P., Sabourin, C. L., Bielefeldt-Ohmann, H., Garcia-Sastre, A., Albrecht, R., and et al. (2010). Macaque proteome response to highly pathogenic avian influenza and 1918 reassortant influenza virus infections. *Journal of Virology*, 84(22):12058–12068.

Bruno, V. M., Hannemann, S., Lara-Tejero, M., Flavell, R. A., Kleinstein, S. H., and Galán, J. E. (2009). Salmonella typhimurium type iii secretion effectors stimulate innate immune responses in cultured epithelial cells. *PLoS Pathogens*, 5(8):e1000538.

Bryk, J., Hardouin, E., Pugach, I., Hughes, D., Strotmann, R., Stoneking, M., and Myles, S. (2008). Positive selection in east asians for an edar allele that enhances nf-χb activation. *PLoS ONE*, 3(5):e2209.

Cagliani, R., Forni, D., Biasin, M., Comabella, M., Guerini, F. R., Riva, S., Pozzoli, U., Agliardi, C., Caputo, D., Malhotra, S., and et al. (2014). Ancient and recent selective pressures shaped genetic diversity at aim2-like nucleic acid sensors. *Genome Biology and Evolution*, 6(4):830–845.

Cagliani, R., Fumagalli, M., Riva, S., Pozzoli, U., Comi, G. P., Menozzi, G., Bresolin, N., and Sironi, M. (2008). The signature of long-standing balancing selection at the human defensin β-1 promoter. *Genome Biology*, 9(9):R143.

Cagliani, R., Guerini, F. R., Fumagalli, M., Riva, S., Agliardi, C., Galimberti, D., Pozzoli, U., Goris, A., Dubois, B., Fenoglio, C., and et al. (2012). A trans-specific polymorphism in zc3hav1 is maintained by long-standing balancing selection and may confer susceptibility to multiple sclerosis. *Molecular Biology and Evolution*, 29(6):1599–1613.

Cai, L., Deng, S.-L., Liang, L., Pan, H., Zhou, J., Wang, M.-Y., Yue, J., Wan, C.-L., He, G., and He, L. (2012). Identification of genetic associations of sp110/mybbp1a/rela with pulmonary tuberculosis in the chinese han population. *Human Genetics*, 132(3):265–273.

Cakır, G., Gumus, S., Ucar, E., Kaya, H., Tozkoparan, E., Akgul, E. O., Karaman, B., Deniz, O., Kurt, I., Ozkan, M., and et al. (2012). Serum chitotriosidase activity in pulmonary tuberculosis: Response to treatment and correlations with clinical parameters. *Annals of Laboratory Medicine*, 32(3):184.

Campos, R. K., Wong, B., Xie, X., Lu, Y.-F., Shi, P.-Y., Pompon, J., Garcia-Blanco, M. A., and Bradrick, S. S. (2016). Rplp1 and rplp2 are essential flavivirus host factors that promote early viral protein accumulation. *Journal of Virology*, 91(4):e01706–16.

Carlson, C. S., Thomas, D. J., Eberle, M. A., Swanson, J. E., Livingston, R. J., Rieder, M. J., and Nickerson, D. A. (2005). Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Research*, 15(11):1553–1565.

Carnero-Montoro, E., Bonet, L., Engelken, J., Bielig, T., Martínez-Florensa, M., Lozano, F., and Bosch, E. (2011). Evolutionary and Functional Evidence for Positive Selection at the Human CD5 Immune Receptor Gene. *Molecular Biology and Evolution*, page msr251.

Carpino, N., Chen, Y., Nassar, N., and Oh, H.-W. (2009). The sts proteins target tyrosine phosphorylated, ubiquitinated proteins within tcr signaling pathways. *Molecular Immunology*, 46(16):3224–3231.

Casanova, J.-L. and Abel, L. (2005). Inborn errors of immunity to infection the rule rather than the exception. *The Journal of Experimental Medicine*, 202(2):197–201.

Cashdan, E. (2014). Biogeography of human infectious diseases: A global historical analysis. *PLoS ONE*, 9(10):e106752.

Cenit, M. C., Martínez-Florensa, M., Consuegra, M., Bonet, L., Carnero-Montoro, E., Armiger, N., Caballero-Baños, M., Arias, M. T., Benitez, D., Ortego-Centeno, N., de Ramón, E., Sabio, J. M., García-Hernández, F. J., Tolosa, C., Suárez, A., González-Gay, M. A., Bosch, E., Martín, J., and Lozano, F. (2014). Analysis of Ancestral and Functionally Relevant CD5 Variants in Systemic Lupus Erythematosus Patients. *PLoS ONE*, 9(11):e113090.

Chae, S.-C., Li, C.-S., Kim, K. M., Yang, J. Y., Zhang, Q., Lee, Y.-C., Yang, Y.-S., and Chung, H.-T. (2007). Identification of polymorphisms in human interleukin-27 and their association with asthma in a Korean population. *Journal of Human Genetics*, 52(4):355–361.

Chan, P. Y., Silva, E. A. C., De Kouchkovsky, D., Joannas, L. D., Hao, L., Hu, D., Huntsman, S., Eng, C., Licona-Limon, P., Weinstein, J. S., and et al. (2016). The tam family receptor tyrosine kinase tyro3 is a negative regulator of type 2 immunity. *Science*, 352(6281):99–103.

Chaplin, D. D. (2010). Overview of the immune response. *Journal of Allergy and Clinical Immunology*, 125(2, Supplement 2):S3–S23.

Chen, G.-Y., Brown, N. K., Zheng, P., and Liu, Y. (2014). Siglec-g/10 in self-nonself discrimination of innate and adaptive immunity. *Glycobiology*, 24(9):800–806.

Chen, G.-Y., Tang, J., Zheng, P., and Liu, Y. (2009a). Cd24 and siglec-10 selectively repress tissue damage-induced immune responses. *Science*, 323(5922):1722–1725.

Chen, Q.-X., Lv, C., Huang, L.-X., Cheng, B.-L., Xie, G.-H., Wu, S.-J., and Fang, X.-M. (2007). Genomic variations within defb1 are associated with the susceptibility to and the fatal outcome of severe sepsis in chinese han population. *Genes Immunity*, 8(5):439–443.

Chen, X., Zhang, G., Li, Y., Feng, X., Wan, F., Zhang, L., Wang, J., and Zhang, X. (2009b). Circulating b7-h3(cd276) elevations in cerebrospinal fluid and plasma of children with bacterial meningitis. *Journal of Molecular Neuroscience*, 39(3):402–402.

Chen, Y., Xiang, R., and Zhao, S. (2011). The potential role of rtn3 in monocyte recruitment and atherosclerosis. *Molecular and Cellular Biochemistry*, 361(1-2):67–70.

Chen, Y.-r., Chen, T.-y., Zhang, S.-l., Lin, S.-m., Zhao, Y.-r., Ye, F., Zhang, X., Shi, L., Dang, S.-s., and Liu, M. (2009c). Identification of a novel protein binding to hepatitis c virus core protein. *Journal of Gastroenterology and Hepatology*, 24(7):1300–1304.

Cheng, F., Wang, H.-W., Cuenca, A., Huang, M., Ghansah, T., Brayer, J., Kerr, W. G., Takeda, K., Akira, S., Schoenberger, S. P., and et al. (2003). A critical role for stat3 signaling in immune tolerance. *Immunity*, 19(3):425–436.

Cheng, S. X., Lightfoot, Y. L., Yang, T., Zadeh, M., Tang, L., Sahay, B., Wang, G. P., Owen, J. L., and Mohamadzadeh, M. (2014). Epithelial casr deficiency alters intestinal integrity and promotes proinflammatory immune responses. *FEBS Letters*, 588(22):4158–4166.

Chiang, C.-y., Engel, A., Opaluch, A. M., Ramos, I., Maestre, A. M., Secundino, I., De Jesus, P. D., Nguyen, Q. T., Welch, G., Bonamy, G. M., and et al. (2012). Cofactors required for tlr7- and tlr9-dependent innate immune responses. *Cell Host Microbe*, 11(3):306–318.

Chiyo, M., Shimozato, O., Yu, L., Kawamura, K., Iizasa, T., Fujisawa, T., and Tagawa, M. (2005). Expression of IL-27 in murine carcinoma cells produces antitumor effects and induces protective immunity in inoculated host animals. *International Journal of Cancer*, 115(3):437–442.

Cho, Y.-C., Kim, B. R., and Cho, S. (2017). Protein tyrosine phosphatase ptpn21 acts as a negative regulator of icam-1 by dephosphorylating ikkβ in tnf-α-stimulated human keratinocytes. *BMB Reports*, 50(11):584–589.

Chua, J. C., Douglass, J. A., Gillman, A., O'Hehir, R. E., and Meeusen, E. N. (2012). Galectin-10, a potential biomarker of eosinophilic airway inflammation. *PLoS ONE*, 7(8):e42549.

Clauss, A., Lilja, H., and Lundwall, Å. (2005). The evolution of a genetic locus encoding small serine proteinase inhibitors. *Biochemical and Biophysical Research Communications*, 333(2):383–389.

Clemente, F. J., Cardona, A., Inchley, C. E., Peter, B. M., Jacobs, G., Pagani, L., Lawson, D. J., Antão, T., Vicente, M., Mitt, M., DeGiorgio, M., Faltyskova, Z., Xue, Y., Auyb, Q., Szpak, M., Mägi, R., Eriksson, A., Manica, A., Raghavan, M., Rasmussen, M., Rasmussen, S., Willerslev, E., Vidal-Puig, A., Tyler-Smith, C., Villems, R., Nielsen, R., Metspalu, M., Malyarchuk, B., Derenko, M., and Kivisild, T. (2014). A selective sweep on a deleterious mutation in cpt1a in arctic populations. *The American Journal of Human Genetics*, 95(5):584–589.

Cohen, J. (2002). The immunopathogenesis of sepsis. *Nature*, 420(6917):885–891.

Collette, Y., Gilles, A., Pontarotti, P., and Olive, D. (2003). A co-evolution perspective of the tnfsf and tnfrsf families in the immune system. *Trends in Immunology*, 24(7):387–394.

Collin, M., Dickinson, R., and Bigley, V. (2015). Haematopoietic and immune defects associated withgata2mutation. *British Journal of Haematology*, 169(2):173–187.

Colmenares, M., Puig-Kröger, A., Pello, O. M., Corbı, A. L., and Rivas, L. (2002). Dendritic cell (dc)-specific intercellular adhesion molecule 3 (icam-3)-grabbing nonintegrin (dc-sign, cd209), a c-type surface lectin in human dcs, is a receptor forleishmaniaamastigotes. *Journal of Biological Chemistry*, 277(39):36766–36769.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S. R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., and Gagneux, S. (2013). Out-of-africa migration and neolithic coexpansion of mycobacterium tuberculosis with modern humans. *Nature Genetics*, 45(10):1176–1182.

Cooper, J. D., Walker, N. M., Smyth, D. J., Downes, K., Healy, B. C., and Todd, J. A. (2009). Follow-up of 1715 snps from the wellcome trust case control consortium genome-wide association study in type i diabetes families. *Genes Immunity*, 10(S1):S85–S94.

Cottineau, J., Kottemann, M. C., Lach, F. P., Kang, Y.-H., Vély, F., Deenick, E. K., Lazarov, T., Gineau, L., Wang, Y., Farina, A., and et al. (2017). Inherited gins1 deficiency underlies growth retardation along with neutropenia and nk cell deficiency. *Journal of Clinical Investigation*, 127(5):1991–2006.

Dambuza, I. M. and Brown, G. D. (2015). C-type lectins in immunity: recent developments. *Current Opinion in Immunology*, 32:21–27.

Dansako, H., Yamane, D., Welsch, C., McGivern, D. R., Hu, F., Kato, N., and Lemon, S. M. (2013). Class a scavenger receptor 1 (msr1) restricts hepatitis c virus replication by mediating toll-like receptor 3 recognition of viral rnas produced in neighboring cells. *PLoS Pathogens*, 9(5):e1003345.

Dardalhon, V., Schubart, A. S., Reddy, J., Meyers, J. H., Monney, L., Sabatos, C. A., Ahuja, R., Nguyen, K., Freeman, G. J., Greenfield, E. A., and et al. (2005). Cd226 is specifically expressed on the surface of th1 cells and regulates their expansion and effector functions. *The Journal of Immunology*, 175(3):1558–1565.

Das, H., Groh, V., Kuijl, C., Sugita, M., Morita, C. T., Spies, T., and Bukowski, J. F. (2001). Mica engagement by human vγ2vδ2 t cells enhances their antigen-dependent effector function. *Immunity*, 15(1):83–93.

Das, V., Nal, B., Dujeancourt, A., Thoulouze, M.-I., Galli, T., Roux, P., Dautry-Varsat, A., and Alcover, A. (2004). Activation-induced polarized recycling targets t cell antigen receptors to the immunological synapse. *Immunity*, 20(5):577–588.

Daub, H., Blencke, S., Habenberger, P., Kurtenbach, A., Dennenmoser, J., Wissing, J., Ullrich, A., and Cotten, M. (2002). Identification of srpk1 and srpk2 as the major cellular protein kinases phosphorylating hepatitis b virus core protein. *Journal of Virology*, 76(16):8124–8137.

de Bakker, P. I. W., McVean, G., Sabeti, P. C., Miretti, M. M., Green, T., Marchini, J., Ke, X., Monsuur, A. J., Whittaker, P., Delgado, M., and et al. (2006). A high-resolution hla and snp haplotype map for disease association studies in the extended human mhc. *Nature Genetics*, 38(10):1166–1172.

De Re, V., Simula, M. P., Cannizzaro, R., Pavan, A., De Zorzi, M. A., Toffoli, G., and Canzonieri, V. (2009). Galectin-10, eosinophils, and celiac disease. *Annals of the New York Academy of Sciences*, 1173(1):357–364.

DeFrances, M. C., Debelius, D. R., Cheng, J., and Kane, L. P. (2012). Inhibition of t-cell activation by pik3ip1. *European Journal of Immunology*, 42(10):2754–2759.

DeGiorgio, M., Lohmueller, K. E., and Nielsen, R. (2014). A model-based approach for identifying signatures of ancient balancing selection in genetic data. *PLoS Genetics*, 10(8):e1004561.

Dejean, A. S., Beisner, D. R., Ch'en, I. L., Kerdiles, Y. M., Babour, A., Arden, K. C., Castrillon, D. H., DePinho, R. A., and Hedrick, S. M. (2009). Transcription factor foxo3 controls the magnitude of t cell immune responses by modulating the function of dendritic cells. *Nature Immunology*, 10(5):504–513.

DeLisser, H. M., Christofidou-Solomidou, M., Sun, J., Nakada, M. T., and Sullivan, K. E. (1999). Loss of endothelial surface expression of e-selectin in a patient with recurrent infections. *Blood*.

Dendrou, C. A., Petersen, J., Rossjohn, J., and Fugger, L. (2018). Hla variation and disease. *Nature Reviews Immunology*, 18(5):325–339.

DeWitte-Orr, S. J., Collins, S. E., Bauer, C. M. T., Bowdish, D. M., and Mossman, K. L. (2010). An accessory to the "trinity": Sr-as are essential pathogen sensors of extracellular dsrna, mediating entry and leading to subsequent type i ifn responses. *PLoS Pathogens*, 6(3):e1000829.

Di Rosa, M., Distefano, G., Zorena, K., and Malaguarnera, L. (2016). Chitinases and immunity: Ancestral molecules with new functions. *Immunobiology*, 221(3):399–411.

Di Rosa, M., Malaguarnera, G., De Gregorio, C., Drago, F., and Malaguarnera, L. (2012). Evaluation of chi3l-1 and chit-1 expression in differentiated and polarized macrophages. *Inflammation*, 36(2):482–492.

Dieckmann, N. M. G., Frazer, G. L., Asano, Y., Stinchcombe, J. C., and Griffiths, G. M. (2016). The cytotoxic t lymphocyte immune synapse at a glance. *Journal of Cell Science*, 129(15):2881–2886.

Dietrich, A., Kalwa, H., Rost, B. R., and Gudermann, T. (2005). The diacylgylcerol-sensitive trpc3/6/7 subfamily of cation channels: functional characterization and physiological relevance. *Pflügers Archiv - European Journal of Physiology*, 451(1):72–80.

Dong, B., Zhou, Q., Zhao, J., Zhou, A., Harty, R. N., Bose, S., Banerjee, A., Slee, R., Guenther, J., Williams, B. R. G., and et al. (2004). Phospholipid scramblase 1 potentiates the antiviral activity of interferon. *Journal of Virology*, 78(17):8983–8993.

Dong, X., Han, S.-k., Zylka, M. J., Simon, M. I., and Anderson, D. J. (2001). A diverse family of gpcrs expressed in specific subsets of nociceptive sensory neurons. *Cell*, 106(5):619–632.

Donnelly, M. P., Paschou, P., Grigorenko, E., Gurwitz, D., Barta, C., Lu, R.-B., Zhukova, O. V., Kim, J.-J., Siniscalco, M., New, M., and et al. (2011). A global view of the oca2-herc2 region and pigmentation. *Human Genetics*, 131(5):683–696.

Doody, G. M., Bell, S. E., Vigorito, E., Clayton, E., McAdam, S., Tooze, R., Fernandez, C., Lee, I. J., and Turner, M. (2001). Signal transduction through vav-2 participates in humoral immune responses and b cell maturation. *Nature Immunology*, 2(6):542–547.

Dorfleutner, A., Bryan, N. B., Talbott, S. J., Funya, K. N., Rellick, S. L., Reed, J. C., Shi, X., Rojanasakul, Y., Flynn, D. C., and Stehlik, C. (2006). Cellular pyrin domain-only protein 2 is a candidate regulator of inflammasome activation. *Infection and Immunity*, 75(3):1484–1492.

Dorr, P., Westby, M., Dobbs, S., Griffin, P., Irvine, B., Macartney, M., Mori, J., Rickett, G., Smith-Burchnell, C., Napier, C., Webster, R., Armour, D., Price, D., Stammen, B., Wood, A., and Perros, M. (2005). Maraviroc (uk-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor ccr5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrobial Agents and Chemotherapy*, 49(11):4721–4732.

Du, B., Luo, W., Li, R., Tan, B., Han, H., Lu, X., Li, D., Qian, M., Zhang, D., Zhao, Y., and et al. (2013). Lgr4/gpr48 negatively regulates tlr2/4-associated pattern recognition and innate immunity by targeting cd14 expression. *Journal of Biological Chemistry*, 288(21):15131–15141.

Dziarski, R. and Gupta, D. (2010). Review: Mammalian peptidoglycan recognition proteins (pgrps) in innate immunity. *Innate Immunity*, 16(3):168–174.

Egli, A., Santer, D. M., O'Shea, D., Barakat, K., Syedbasha, M., Vollmer, M., Baluch, A., Bhat, R., Groenendyk, J., Joyce, M. A., and et al. (2014). Il-28b is a key regulator of b- and t-cell vaccine responses against influenza. *PLoS Pathogens*, 10(12):e1004556.

Eichelbaum, K. and Krijgsveld, J. (2014). Rapid temporal dynamics of transcription, protein synthesis, and secretion during macrophage activation. *Molecular Cellular Proteomics*, 13(3):792–810.

Eichstaedt, C. A., Pagani, L., Antao, T., Inchley, C. E., Cardona, A., Mörseburg, A., Clemente, F. J., Sluckin, T. J., Metspalu, E., Mitt, M., Mägi, R., Hudjashov, G., Metspalu, M., Mormina, M., Jacobs, G. S., and Kivisild, T. (2017). Evidence of early-stage selection on epas1 and gpr126 genes in andean high altitude populations. *Scientific Reports*, 7(1).

Eisenbarth, S. C., Williams, A., Colegio, O. R., Meng, H., Strowig, T., Rongvaux, A., Henao-Mejia, J., Thaiss, C. A., Joly, S., Gonzalez, D. G., and et al. (2012). Nlrp10 is a nod-like receptor essential to initiate adaptive immunity by dendritic cells. *Nature*, 484(7395):510–513.

Elagoz, A., Henderson, D., Babu, P. S., Salter, S., Grahames, C., Bowers, L., Roy, M.-O., Laplante, P., Grazzini, E., Ahmad, S., and et al. (2004). A truncated form of ckβ8-1 is a potent agonist for human formyl peptide-receptor-like 1 receptor. *British Journal of Pharmacology*, 141(1):37–46.

Ernst, M., Inglese, M., Scholz, G. M., Harder, K. W., Clay, F. J., Bozinovski, S., Waring, P., Darwiche, R., Kay, T., Sly, P., and et al. (2002). Constitutive activation of the src family kinase hck results in spontaneous pulmonary inflammation and an enhanced innate immune response. *The Journal of Experimental Medicine*, 196(5):589–604.

Esworthy, R. S., Aranda, R., Martín, M. G., Doroshow, J. H., Binder, S. W., and Chu, F.-F. (2001). Mice with combined disruption of gpx1 andgpx2 genes have colitis. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 281(3):G848–G855.

Eyre-Walker, A. and Keightley, P. D. (1999). High genomic deleterious mutation rates in hominids. *Nature*, 397(6717):344–347.

Fan, S., Hansen, M. E. B., Lo, Y., and Tishkoff, S. A. (2016). Going global by adapting local: A review of recent human adaptation. *Science*, 354(6308):54–59.

Feigelstock, D., Thompson, P., Mattoo, P., Zhang, Y., and Kaplan, G. G. (1998). The human homolog of havcr-1 codes for a hepatitis a virus cellular receptor. *Journal of Virology*, 72(8):6621–6628.

Ferrer-Admetlla, A., Bosch, E., Sikora, M., Marques-Bonet, T., Ramirez-Soriano, A., Muntasell, A., Navarro, A., Lazarus, R., Calafell, F., Bertranpetit, J., and et al. (2008). Balancing selection is the main force shaping the evolution of innate immunity genes. *The Journal of Immunology*, 181(2):1315–1322.

Ferrer-Admetlla, A., Liang, M., Korneliussen, T., and Nielsen, R. (2014). On Detecting Incomplete Soft or Hard Selective Sweeps Using Haplotype Structure. *Molecular Biology and Evolution*, 31(5):1275–1291.

Fink, K., Martin, L., Mukawera, E., Chartier, S., De Deken, X., Brochiero, E., Miot, F., and Grandvaux, N. (2013). Ifnβ/tnfα synergism induces a non-canonical stat2/irf9-dependent pathway triggering a novel duox2 nadph oxidase-mediated airway antiviral response. *Cell Research*, 23(5):673–690.

Fitzgerald, K. A., Palsson-McDermott, E. M., Bowie, A. G., Jefferies, C. A., Mansell, A. S., Brady, G., Brint, E., Dunne, A., Gray, P., Harte, M. T., and et al. (2001). Mal (myd88-adapter-like) is required for toll-like receptor-4 signal transduction. *Nature*, 413(6851):78–83.

Flach, H., Rosenbaum, M., Duchniewicz, M., Kim, S., Zhang, S. L., Cahalan, M. D., Mittler, G., and Grosschedl, R. (2010). Mzb1 protein regulates calcium homeostasis, antibody secretion, and integrin activation in innate-like b cells. *Immunity*, 33(5):723–735.

Flatz, G. and Rotthauwe, H. (1973). Lactose nutrition and natural selection. *The Lancet*, 302(7820):76–77.

Flemming, A. (2017). Jnk inhibitors boost antifungal immunity. *Nature Reviews Immunology*, 17(3):149–149.

Flo, T. H., Smith, K. D., Sato, S., Rodriguez, D. J., Holmes, M. A., Strong, R. K., Akira, S., and Aderem, A. (2004). Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron. *Nature*, 432(7019):917–921.

Folly, B. B., Weffort-Santos, A. M., Fathman, C., and Soares, L. R. (2011). Dengue-2 structural proteins associate with human proteins to produce a coagulation and innate immune response biased interactome. *BMC Infectious Diseases*, 11(1).

Frank, A. C., Zhang, X., Katsounas, A., Bharucha, J. P., Kottilil, S., and Imamichi, T. (2010). Interleukin-27, an Anti-HIV-1 Cytokine, Inhibits Replication of Hepatitis C Virus. *Journal of Interferon & Cytokine Research*, 30(6):427–431.

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., and et al. (2007). A second generation human haplotype map of over 3.1 million snps. *Nature*, 449(7164):851–861.

Freeman, G., Gribben, J., Boussiotis, V., Ng, J., Restivo, V., Lombard, L., Gray, G., and Nadler, L. (1993). Cloning of b7-2: a ctla-4 counter-receptor that costimulates human t cell proliferation. *Science*, 262(5135):909–911.

Frodsham, A. J., Zhang, L., Dumpis, U., Taib, N. A. M., Best, S., Durham, A., Hennig, B. J. W., Hellier, S., Knapp, S., Wright, M., and et al. (2006). Class ii cytokine receptor gene cluster is a major locus for hepatitis b persistence. *Proceedings of the National Academy of Sciences*, 103(24):9148–9153.

Fu, B., Wang, L., Ding, H., Schwamborn, J. C., Li, S., and Dorf, M. E. (2015). Trim32 senses and restricts influenza a virus by ubiquitination of pb1 polymerase. *PLOS Pathogens*, 11(6):e1004960.

Fu, C., Turck, C. W., Kurosaki, T., and Chan, A. C. (1998). Blnk. *Immunity*, 9(1):93–103.

Fujita, T. (2002). Evolution of the lectin–complement pathway and its role in innate immunity. *Nature Reviews Immunology*, 2(5):346–353.

Fumagalli, M., Cagliani, R., Pozzoli, U., Riva, S., Comi, G. P., Menozzi, G., Bresolin, N., and Sironi, M. (2009a). Widespread balancing selection and pathogen-driven selection at blood group antigen genes. *Genome Research*, 19(2):199–212.

Fumagalli, M., Pozzoli, U., Cagliani, R., Comi, G. P., Bresolin, N., Clerici, M., and Sironi, M. (2010). The landscape of human genes involved in the immune response to parasitic worms. *BMC Evolutionary Biology*, 10(1):264.

Fumagalli, M., Pozzoli, U., Cagliani, R., Comi, G. P., Riva, S., Clerici, M., Bresolin, N., and Sironi, M. (2009b). Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. *The Journal of Experimental Medicine*, 206(6):1395–1408.

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L., and Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genetics*, 7(11):e1002355.

Galvani, A. P. and Slatkin, M. (2003). Evaluating plague and smallpox as historical selective pressures for the ccr5- 32 hiv-resistance allele. *Proceedings of the National Academy of Sciences*, 100(25):15276–15279.

Garcia-Beltran, W. F., Hölzemer, A., Martrus, G., Chung, A. W., Pacheco, Y., Simoneau, C. R., Rucevic, M., Lamothe-Molina, P. A., Pertel, T., Kim, T.-E., and et al. (2016). Open conformers of hla-f are high-affinity ligands of the activating nk-cell receptor kir3ds1. *Nature Immunology*, 17(9):1067–1074.

Garcia-Etxebarria, K., Jauregi-Miguel, A., Romero-Garmendia, I., Plaza-Izurieta, L., Legarda, M., Irastorza, I., and Bilbao, J. R. (2016). Ancestry-based stratified analysis of immunochip data identifies novel associations with celiac disease. *European Journal of Human Genetics*, 24(12):1831–1834.

Garman, S. C., Wurzburg, B. A., Tarchevskaya, S. S., Kinet, J.-P., and Jardetzky, T. S. (2000). Structure of the fc fragment of human ige bound to its high-affinity receptor fcεriα. *Nature*, 406(6793):259–266.

Garred, P., Honore, C., Ma, Y. J., Munthe-Fog, L., and Hummelshoj, T. (2009). Mbl2, fcn1, fcn2 and fcn3—the genes behind the initiation of the lectin pathway of complement. *Molecular Immunology*, 46:2737–2744.

Gavala, M. L., Liu, Y.-P., Lenertz, L. Y., Zeng, L., Blanchette, J. B., Guadarrama, A. G., Denlinger, L. C., Bertics, P. J., and Smith, J. A. (2013). Nucleotide receptor p2rx7 stimulation enhances lps-induced interferon-β production in murine macrophages. *Journal of Leukocyte Biology*, 94(4):759–768.

Gelius, E., Persson, C., Karlsson, J., and Steiner, H. (2003). A mammalian peptidoglycan recognition protein with n-acetylmuramoyl-l-alanine amidase activity. *Biochemical and Biophysical Research Communications*, 306(4):988–994.

Gerbault, P., Liebert, A., Itan, Y., Powell, A., Currat, M., Burger, J., Swallow, D. M., and Thomas, M. G. (2011). Evolution of lactase persistence: an example of human niche construction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1566):863–877.

Gerbe, F., Sidot, E., Smyth, D. J., Ohmoto, M., Matsumoto, I., Dardalhon, V., Cesses, P., Garnier, L., Pouzolles, M., Brulin, B., and et al. (2016). Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites. *Nature*, 529(7585):226–230.

Gerke, V., Creutz, C. E., and Moss, S. E. (2005). Annexins: linking ca2+ signalling to membrane dynamics. *Nature Reviews Molecular Cell Biology*, 6(6):449–461.

Glocker, E.-O., Kotlarz, D., Boztug, K., Gertz, E. M., Schäffer, A. A., Noyan, F., Perro, M., Diestelhorst, J., Allroth, A., Murugan, D., and et al. (2009). Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *New England Journal of Medicine*, 361(21):2033–2045.

Gloriam, D. E., Schiöth, H. B., and Fredriksson, R. (2005). Nine new human rhodopsin family g-protein coupled receptors: identification, sequence characterisation and evolutionary relationship. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1722(3):235–246.

Glusman, G., Caballero, J., Mauldin, D. E., Hood, L., and Roach, J. C. (2011). Kaviar: an accessible system for testing SNV novelty. *Bioinformatics*, 27(22):3216–3217.

Goldberg, R., Wildbaum, G., Zohar, Y., Maor, G., and Karin, N. (2004a). Suppression of Ongoing Adjuvant-Induced Arthritis by Neutralizing the Function of the p28 Subunit of IL-27. *The Journal of Immunology*, 173(2):1171–1178.

Goldberg, R., Zohar, Y., Wildbaum, G., Geron, Y., Maor, G., and Karin, N. (2004b). Suppression of Ongoing Experimental Autoimmune Encephalomyelitis by Neutralizing the Function of the p28 Subunit of IL-27. *The Journal of Immunology*, 173(10):6465–6471.

Goldman, M. J., Anderson, G., Stolzenberg, E. D., Kari, U., Zasloff, M., and Wilson, J. M. (1997). Human β-defensin-1 is a salt-sensitive antibiotic in lung that is inactivated in cystic fibrosis. *Cell*, 88(4):553–560.

Golebiewski, L., Liu, H., Javier, R. T., and Rice, A. P. (2011). The avian influenza virus ns1 esev pdz binding motif associates with dlg1 and scribble to disrupt cellular tight junctions. *Journal of Virology*, 85(20):10639–10648.

Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., Spector, T. D., Bell, J. T., Clark, A. G., and Ley, R. E. (2016). Genetic determinants of the gut microbiome in uk twins. *Cell Host Microbe*, 19(5):731–743.

Gourraud, P.-A., Khankhanian, P., Cereb, N., Yang, S. Y., Feolo, M., Maiers, M., D. Rioux, J., Hauser, S., and Oksenberg, J. (2014). Hla diversity in the 1000 genomes dataset. *PLoS ONE*, 9(7):e97282.

Gowda, D. C. and Ockenhouse, C. F. (1999). Adherence adherence of plasmodium falciparum-infected erythrocytes to chondroitin 4-sulfate. *Bioscience Reports*, 19(4):261–271.

Graham, D. B., Lefkovith, A., Deelen, P., de Klein, N., Varma, M., Boroughs, A., Desch, A. N., Ng, A. C., Guzman, G., Schenone, M., and et al. (2016). Tmem258 is a component of the oligosaccharyltransferase complex controlling er stress and intestinal inflammation. *Cell Reports*, 17(11):2955–2965.

Graham, D. S. C., Graham, R. R., Manku, H., Wong, A. K., Whittaker, J. C., Gaffney, P. M., Moser, K. L., Rioux, J. D., Altshuler, D., Behrens, T. W., and et al. (2007). Polymorphism at the tnf superfamily gene tnfsf4 confers susceptibility to systemic lupus erythematosus. *Nature Genetics*, 40(1):83–89.

Gramaglia, I., Weinberg, A. D., Lemon, M., and Croft, M. (1998). Ox-40 ligand: A potent costimulatory molecule for sustaining primary cd4 t cell responses. *The Journal of Immunology*, 161:6510–6517.

Gresnigt, M. S., Rösler, B., Jacobs, C. W., Becker, K. L., Joosten, L. A., van der Meer, J. W., Netea, M. G., Dinarello, C. A., and van de Veerdonk, F. L. (2012). The il-36 receptor pathway regulatesaspergillus fumigatus-induced th1 and th17 responses. *European Journal of Immunology*, 43(2):416–426.

Groh, V., Rhinehart, R., Randolph-Habecker, J., Topp, M. S., Riddell, S. R., and Spies, T. (2001). Costimulation of cd8$\alpha\beta$ t cells by nkg2d via engagement by mic induced on virus-infected cells. *Nature Immunology*, 2(3):255–260.

Grossman, S. R., Andersen, K. G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., Park, D. J., Griesemer, D., Karlsson, E. K., Wong, S. H., Cabili, M., Adegbola, R. A., Bamezai, R. N. K., Hill, A. V. S., Vannberg, F. O., Rinn, J. L., Lander, E. S., Schaffner, S. F., and Sabeti, P. C. (2013). Identifying Recent Adaptations in Large-Scale Genomic Data. *Cell*, 152(4):703–713.

Grossman, S. R., Shylakhter, I., Karlsson, E. K., Byrne, E. H., Morales, S., Frieden, G., Hostetter, E., Angelino, E., Garber, M., Zuk, O., Lander, E. S., Schaffner, S. F., and Sabeti, P. C. (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*, 327(5967):883–886.

Guernier, V., Hochberg, M. E., and Guégan, J.-F. (2004). Ecology drives the worldwide distribution of human diseases. *PLoS Biology*, 2(6):e141.

Gulko, B., Hubisz, M. J., Gronau, I., and Siepel, A. (2015). A method for calculating probabilities of fitness consequences for point mutations across the human genome. *Nature Genetics*, 47(3):276–283.

Gunturi, A., Berg, R. E., and Forman, J. (2004). The role of cd94/nkg2 in innate and adaptive immunity. *Immunologic Research*, 30(1):029–034.

Gunz, P., Bookstein, F. L., Mitteroecker, P., Stadlmayr, A., Seidler, H., and Weber, G. W. (2009). Early modern human diversity suggests subdivided population structure and a complex out-of-africa scenario. *Proceedings of the National Academy of Sciences*, 106(15):6094–6098.

Günzl, P. and Schabbauer, G. (2008). Recent advances in the genetic analysis of pten and pi3k innate immune properties. *Immunobiology*, 213(9-10):759–765.

Guo, Y., Chang, C., Huang, R., Liu, B., Bao, L., and Liu, W. (2012). Ap1 is essential for generation of autophagosomes from the trans-golgi network. *Journal of Cell Science*, 125(7):1706–1715.

Hafler, J. P., Maier, L. M., Cooper, J. D., Plagnol, V., Hinks, A., Simmonds, M. J., Stevens, H. E., Walker, N. M., Healy, B., and et al. (2008). Cd226 gly307ser association with multiple autoimmune diseases. *Genes Immunity*, 10(1):5–10.

Hall, N. B., Igo, R. P., Malone, L. L., Truitt, B., Schnell, A., Tao, L., Okware, B., Nsereko, M., Chervenak, K., and et al. (2014). Polymorphisms in ticam2 and il1b are associated with tb. *Genes Immunity*, 16(2):127–133.

Hamano, S., Himeno, K., Miyazaki, Y., Ishii, K., Yamanaka, A., Takeda, A., Zhang, M., Hisaeda, H., Mak, T. W., Yoshimura, A., and Yoshida, H. (2003). WSX-1 Is Required for Resistance to Trypanosoma cruzi Infection by Regulation of Proinflammatory Cytokine Production. *Immunity*, 19(5):657–667.

Hansmann, B., Schröder, J.-M., and Gerstel, U. (2015). Skin-derived c-terminal filaggrin-2 fragments are pseudomonas aeruginosa-directed antimicrobials targeting bacterial replication. *PLOS Pathogens*, 11(9):e1005159.

Hawkins, R. F. W., Patenaude, A., Dumas, A., Jain, R., Tesfagiorgis, Y., Kerfoot, S., Matsui, T., Gunzer, M., Poubelle, P. E., Larochelle, C., and et al. (2017). Icam1+ neutrophils promote chronic inflammation via asprv1 in b cell–dependent autoimmune encephalomyelitis. *JCI Insight*, 2(23).

Hayden, M. S. and Ghosh, S. (2008). Shared principles in nf-χb signaling. *Cell*, 132(3):344–362.

He, C. and Klionsky, D. J. (2009). Regulation mechanisms and signaling pathways of autophagy. *Annual Review of Genetics*, 43(1):67–93.

Heiss, K., Janner, N., Mahnss, B., Schumacher, V., Koch-Nolte, F., Haag, F., and Mittrucker, H.-W. (2008). High sensitivity of intestinal cd8+ t cells to nucleotides indicates p2x7 as a regulator for intestinal t cell responses. *The Journal of Immunology*, 181(6):3861–3869.

Helms, C., Cao, L., Krueger, J. G., Wijsman, E. M., Chamian, F., Gordon, D., Heffernan, M., Daw, J. A. W., Robarge, J., Ott, J., and et al. (2003). A putative runx1 binding site variant between slc9a3r1 and nat9 is associated with susceptibility to psoriasis. *Nature Genetics*, 35(4):349–356.

Herbert, D. R., Yang, J.-Q., Hogan, S. P., Groschwitz, K., Khodoun, M., Munitz, A., Orekov, T., Perkins, C., Wang, Q., Brombacher, F., and et al. (2009). Intestinal epithelial cell secretion of relm-β protects against gastrointestinal worm infection. *The Journal of Experimental Medicine*, 206(13):2947–2957.

Hewitt, E. W. (2003). The MHC class I antigen presentation pathway: strategies for viral immune evasion. *Immunology*, 110(2):163–169.

Heyer, E. and Quintana-Murci, L. (2009). Perspective: Evolutionary genetics as a tool to target genes involved in phenotypes of medical relevance. *Evolutionary Applications*, 2(1):71–80.

Hider, J. L., Gittelman, R. M., Shah, T., Edwards, M., Rosenbloom, A., Akey, J. M., and Parra, E. J. (2013). Exploring signatures of positive selection in pigmentation candidate genes in populations of east asian ancestry. *BMC Evolutionary Biology*, 13(1):150.

Higham, T., Douka, K., Wood, R., Ramsey, C. B., Brock, F., Basell, L., Camps, M., Arrizabalaga, A., Baena, J., Barroso-Ruíz, C., Bergman, C., Boitard, C., Boscato, P., Caparrós, M., Conard, N. J., Draily, C., Froment, A., Galván, B., Gambassini, P., Garcia-Moreno, A., Grimaldi, S., Haesaerts, P., Holt, B., Iriarte-Chiapusso, M.-J., Jelinek, A., Pardo, J. F. J., Maíllo-Fernández, J.-M., Marom, A., Maroto, J., Menéndez, M., Metz, L., Morin, E., Moroni, A., Negrino, F., Panagopoulou, E., Peresani, M., Pirson, S., de la Rasilla, M., Riel-Salvatore, J., Ronchitelli, A., Santamaria, D., Semal, P., Slimak, L., Soler, J., Soler, N., Villaluenga, A., Pinhasi, R., and Jacobi, R. (2014). The timing and spatiotemporal patterning of neanderthal disappearance. *Nature*, 512(7514):306–309.

Hisada, M., Kamiya, S., Fujita, K., Belladonna, M. L., Aoki, T., Koyanagi, Y., Mizuguchi, J., and Yoshimoto, T. (2004). Potent Antitumor Activity of Interleukin-27. *Cancer Research*, 64(3):1152–1156.

Hiscott, J., Kwon, H., and Génin, P. (2001). Hostile takeovers: viral appropriation of the nf-kb pathway. *Journal of Clinical Investigation*, 107(2):143–151.

Hohenhaus, D. M., Schaale, K., Le Cao, K.-A., Seow, V., Iyer, A., Fairlie, D. P., and Sweet, M. J. (2013). An mrna atlas of g protein-coupled receptor expression during primary human monocyte/macrophage differentiation and lipopolysaccharide-mediated activation identifies targetable candidate regulators of inflammation. *Immunobiology*, 218(11):1345–1353.

Hollox, E. J. and Armour, J. A. (2008). Directional and balancing selection in human beta-defensins. *BMC Evolutionary Biology*, 8(1):113.

Holmberg, D., Ruikka, K., Lindgren, P., Eliasson, M., and Mayans, S. (2016). Association of cd247 (cd3ζ) gene polymorphisms with t1d and aitd in the population of northern sweden. *BMC Medical Genetics*, 17(1).

Hölscher, C., Hölscher, A., Rückerl, D., Yoshimoto, T., Yoshida, H., Mak, T., Saris, C., and Ehlers, S. (2005). The IL-27 receptor chain WSX-1 differentially regulates antibacterial immunity and survival during experimental tuberculosis. *Journal of Immunology (Baltimore, Md.: 1950)*, 174(6):3534–3544.

Holsinger, K. E. and Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Reviews Genetics*, 10(9):639–650.

Hoppenbrouwers, I. A., Aulchenko, Y. S., Ebers, G. C., Ramagopalan, S. V., Oostra, B. A., van Duijn, C. M., and Hintzen, R. Q. (2008). Evi5 is a risk gene for multiple sclerosis. *Genes Immunity*, 9(4):334–337.

Horowits, R., Kempner, E. S., Bisher, M. E., and Podolsky, R. J. (1986). A physiological role for titin and nebulin in skeletal muscle. *Nature*, 323(6084):160–164.

Horuk, R., Chitnis, C., Darbonne, W., Colby, T., Rybicki, A., Hadley, T., and Miller, L. (1993). A receptor for the malarial parasite plasmodium vivax: the erythrocyte chemokine receptor. *Science*, 261(5125):1182–1184.

Huang, W., Thomas, B., Flynn, R. A., Gavzy, S. J., Wu, L., Kim, S. V., Hall, J. A., Miraldi, E. R., Ng, C. P., Rigo, F., and et al. (2015). Ddx5 and its associated lncrna rmrp modulate th17 cell effector functions. *Nature*, 528(7583):517–522.

Hublin, J.-J., Ben-Ncer, A., Bailey, S. E., Freidline, S. E., Neubauer, S., Skinner, M. M., Bergmann, I., Le Cabec, A., Benazzi, S., Harvati, K., and et al. (2017). New fossils from jebel irhoud, morocco and the pan-african origin of homo sapiens. *Nature*, 546(7657):289–292.

Hudson, R. R., Kreitman, M., and Aguadé, M. (1987). A Test of Neutral Molecular Evolution Based on Nucleotide Data. *Genetics*, 116(1):153–159.

Huh, J. R., Leung, M. W. L., Huang, P., Ryan, D. A., Krout, M. R., Malapaka, R. R. V., Chow, J., Manel, N., Ciofani, M., Kim, S. V., and et al. (2011). Digoxin and its derivatives suppress th17 cell differentiation by antagonizing rorγt activity. *Nature*, 472(7344):486–490.

Hummelshoj, T., Fog, L. M., Madsen, H. O., Sim, R. B., and Garred, P. (2008). Comparative study of the human ficolins reveals unique features of ficolin-3 (hakata antigen). *Molecular Immunology*, 45(6):1623–1632.

Hunter, C. A. (2005). New IL-12-family members: IL-23 and IL-27, cytokines with divergent functions. *Nature Reviews Immunology*, 5(7):521–531.

Hütter, G., Nowak, D., Mossner, M., Ganepola, S., Müßig, A., Allers, K., Schneider, T., Hofmann, J., Kücherer, C., Blau, O., Blau, I., Hofmann, W. K., and Thiel, E. (2009). Long-term control of hiv by ccr5 delta32/delta32 stem-cell transplantation. *New England Journal of Medicine*, 360(7):692–698.

Illario, M., Giardino-Torchia, M. L., Sankar, U., Ribar, T. J., Galgani, M., Vitiello, L., Masci, A. M., Bertani, F. R., Ciaglia, E., Astone, D., and et al. (2008). Calmodulin-dependent kinase iv links toll-like receptor 4 signaling with survival pathway of activated dendritic cells. *Blood*, 111(2):723–731.

Imamichi, T., Yang, J., Huang, D.-W., Brann, T. W., Fullmer, B. A., Adelsberger, J. W., Lempicki, R. A., Baseler, M. W., and Lane, H. C. (2008). IL-27, a novel anti-HIV cytokine, activates multiple interferon-inducible genes in macrophages. *AIDS (London, England)*, 22(1):39–45.

Imamura, K., Imamachi, N., Akizuki, G., Kumakura, M., Kawaguchi, A., Nagata, K., Kato, A., Kawaguchi, Y., Sato, H., Yoneda, M., and et al. (2014). Long noncoding rna neat1-dependent sfpq relocation from promoter region to paraspeckle mediates il8 expression upon immune stimuli. *Molecular Cell*, 53(3):393–406.

Inagaki, F. F., Tanaka, M., Inagaki, N. F., Yagai, T., Sato, Y., Sekiguchi, K., Oyaizu, N., Kokudo, N., and Miyajima, A. (2013). Nephronectin is upregulated in acute and chronic hepatitis and aggravates liver injury by recruiting cd4 positive cells. *Biochemical and Biophysical Research Communications*, 430(2):751–756.

Inhorn, M. C. and Brown, P. J. (1990). The Anthropology of Infectious Disease. *Annual Review of Anthropology*, 19(1):89–117.

Ishikawa, H. and Barber, G. N. (2008). Sting is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature*, 455(7213):674–678.

Iskander, K., Li, J., Han, S., Zheng, B., and Jaiswal, A. K. (2006). Nqo1 and nqo2 regulation of humoral immunity and autoimmunity. *Journal of Biological Chemistry*, 281(41):30917–30924.

Iwakura, Y., Ishigame, H., Saijo, S., and Nakae, S. (2011). Functional specialization of interleukin-17 family members. *Immunity*, 34(2):149–162.

Iyer, A., van Eijk, M., Silva, E., Hatta, M., Faber, W., Aerts, J. M., and Das, P. K. (2009). Increased chitotriosidase activity in serum of leprosy patients: Association with bacillary leprosy. *Clinical Immunology*, 131(3):501–509.

Janeway, C. A. J., Travers, P., Walport, M., and Shlomchik, M. J. (2001). *Immunobiology: The Immune System In Health and Disease.* Garland Science, New York, 5th edition.

Jansen, L., de Niet, A., Makowska, Z., Dill, M. T., van Dort, K. A., Terpstra, V., Bart Takkenberg, R., Janssen, H. L., Heim, M. H., Kootstra, N. A., and et al. (2015). An intrahepatic transcriptional signature of enhanced immune activity predicts response to peginterferon in chronic hepatitis b. *Liver International*, 35(7):1824–1832.

Jeffery, K. J. and Bangham, C. R. (2000). Do infectious diseases drive mhc diversity? *Microbes and Infection*, 2(11):1335–1341.

Jia, H. P., Starner, T., Ackermann, M., Kirby, P., Tack, B. F., and McCray, P. B. (2001). Abundant human β-defensin-1 expression in milk and mammary gland epithelium. *The Journal of Pediatrics*, 138(1):109–112.

Jia, Y., Song, T., Wei, C., Ni, C., Zheng, Z., Xu, Q., Ma, H., Li, L., Zhang, Y., He, X., and et al. (2009). Negative regulation of mavs-mediated innate immune response by psma7. *The Journal of Immunology*, 183(7):4241–4248.

Jin, J., Yu, Q., Han, C., Hu, X., Xu, S., Wang, Q., Wang, J., Li, N., and Cao, X. (2013). Lrrfip2 negatively regulates nlrp3 inflammasome activation in macrophages by promoting flightless-i-mediated caspase-1 inhibition. *Nature Communications*, 4(1).

Jin, S.-L. C. and Conti, M. (2002). Induction of the cyclic nucleotide phosphodiesterase pde4b is essential for lps-activated tnf- responses. *Proceedings of the National Academy of Sciences*, 99(11):7628–7633.

Jordan, M. S., Singer, A. L., and Koretzky, G. A. (2003). Adaptors as central mediators of signal transduction in immune cells. *Nature Immunology*, 4(2):110–116.

Junt, T., Moseman, E. A., Iannacone, M., Massberg, S., Lang, P. A., Boes, M., Fink, K., Henrickson, S. E., Shayakhmetov, D. M., Di Paolo, N. C., and et al. (2007). Subcapsular sinus macrophages in lymph nodes clear lymph-borne viruses and present them to antiviral b cells. *Nature*, 450(7166):110–114.

Jurevic, R. J., Bai, M., Chadwick, R. B., White, T. C., and Dale, B. A. (2003). Single-nucleotide polymorphisms (snps) in human -defensin 1: High-throughput snp assays and association with candida carriage in type i diabetics and nondiabetic controls. *Journal of Clinical Microbiology*, 41(1):90–96.

Kahr, W. H. A., Pluthero, F. G., Elkadri, A., Warner, N., Drobac, M., Chen, C. H., Lo, R. W., Li, L., Li, R., Li, Q., and et al. (2017). Loss of the arp2/3 complex component arpc1b causes platelet abnormalities and predisposes to inflammatory disease. *Nature Communications*, 8:14816.

Kainu, K., Kivinen, K., Zucchelli, M., Suomela, S., Kere, J., Inerot, A., Baker, B. S., Powles, A. V., Fry, L., Samuelsson, L., and et al. (2009). Association of psoriasis to pglyrp and sprr genes at psors4 locus on 1q shows heterogeneity between finnish, swedish and irish families. *Experimental Dermatology*, 18(2):109–115.

Kalliolias, G. D. and Ivashkiv, L. B. (2008). IL-27 Activates Human Monocytes via STAT1 and Suppresses IL-10 Production but the Inflammatory Functions of IL-27 Are Abrogated by TLRs and p38. *The Journal of Immunology*, 180(9):6325–6333.

Kamatani, Y., Wattanapokayakit, S., Ochi, H., Kawaguchi, T., Takahashi, A., Hosono, N., Kubo, M., Tsunoda, T., Kamatani, N., Kumada, H., and et al. (2009). A genome-wide association study identifies variants in the hla-dp locus associated with chronic hepatitis b in asians. *Nature Genetics*, 41(5):591–595.

Karlas, A., Machuy, N., Shin, Y., Pleissner, K.-P., Artarini, A., Heuer, D., Becker, D., Khalil, H., Ogilvie, L. A., Hess, S., and et al. (2010). Genome-wide rnai screen identifies human host factors crucial for influenza virus replication. *Nature*, 463(7282):818–822.

Karlsson, E. K., Harris, J. B., Tabrizi, S., Rahman, A., Shlyakhter, I., Patterson, N., O'Dushlaine, C., Schaffner, S. F., Gupta, S., Chowdhury, F., and et al. (2013). Natural selection in a bangladeshi population from the cholera-endemic ganges river delta. *Science Translational Medicine*, 5(192):192ra86–192ra86.

Karlsson, E. K., Kwiatkowski, D. P., and Sabeti, P. C. (2014). Natural selection and infectious disease in human populations. *Nature Reviews Genetics*, 15(6):379–393.

Kawagoe, T., Sato, S., Matsushita, K., Kato, H., Matsui, K., Kumagai, Y., Saitoh, T., Kawai, T., Takeuchi, O., and Akira, S. (2008). Sequential control of toll-like receptor–dependent responses by irak1 and irak2. *Nature Immunology*, 9(6):684–691.

Kazi, J. U., Kabir, N. N., and Rönnstrand, L. (2015). Role of src-like adaptor protein (slap) in immune and malignant cell signaling. *Cellular and Molecular Life Sciences*, 72(13):2535–2544.

Kelley, J. L. (2006). Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Research*, 16(8):980–989.

Kessler, K., Wunderlich, I., Uebe, S., Falk, N. S., Gießl, A., Brandstätter, J. H., Popp, B., Klinger, P., Ekici, A. B., Sticht, H., Dörr, H.-G., Reis, A., Roepman, R., Seemanová, E., and Thiel, C. T. (2015). DYNC2li1 mutations broaden the clinical spectrum of dynein-2 defects. *Scientific Reports*, 5:11649.

Key, F. M., Teixeira, J. C., de Filippo, C., and Andrés, A. M. (2014). Advantageous diversity maintained by balancing selection in humans. *Current Opinion in Genetics & Development*, 29:45–51.

Kim, C. S., Seol, S. K., Song, O.-K., Park, J. H., and Jang, S. K. (2007). An rna-binding protein, hnrnp a1, and a scaffold protein, septin 6, facilitate hepatitis c virus replication. *Journal of Virology*, 81(8):3852–3865.

Kim, T. W., Yu, M., Zhou, H., Cui, W., Wang, J., DiCorleto, P., Fox, P., Xiao, H., and Li, X. (2012). Pellino 2 is critical for toll-like receptor/interleukin-1 receptor (tlr/il-1r)-mediated post-transcriptional control. *Journal of Biological Chemistry*, 287(30):25686–25695.

Kimura, R., Fujimoto, A., Tokunaga, K., and Ohashi, J. (2007). A practical genome scan for population-specific strong selective sweeps that have reached fixation. *PLoS ONE*, 2(3):e286.

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, 46(3):310–315.

Ko, D. C., Gamazon, E. R., Shukla, K. P., Pfuetzner, R. A., Whittington, D., Holden, T. D., Brittnacher, M. J., Fong, C., Radey, M., Ogohara, C., and et al. (2012). Functional genetic screen of human diversity reveals that a methionine salvage enzyme regulates inflammatory cell death. *Proceedings of the National Academy of Sciences*, 109(35):E2343–E2352.

Koedel, U., Bayerlein, I., Paul, R., Sporer, B., and Pfister, H. W. (2000). Pharmacologic interference with nf-$\varkappa$b activation attenuates central nervous system complications in experimental pneumococcal meningitis. *The Journal of Infectious Diseases*, 182(5):1437–1445.

Kofler, R. and Schlotterer, C. (2012). Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics*, 28(15):2084–2085.

Komiyama, Y., Nakae, S., Matsuki, T., Nambu, A., Ishigame, H., Kakuta, S., Sudo, K., and Iwakura, Y. (2006). IL-17 Plays an Important Role in the Development of Experimental Autoimmune Encephalomyelitis. *The Journal of Immunology*, 177(1):566–573.

Kooyk, Y. v., Appelmelk, B., and Geijtenbeek, T. B. (2003). A fatal attraction: Mycobacterium tuberculosis and hiv-1 target dc-sign to escape immune surveillance. *Trends in Molecular Medicine*, 9(4):153–159.

Kosiol, C., Vinař, T., da Fonseca, R. R., Hubisz, M. J., Bustamante, C. D., Nielsen, R., and Siepel, A. (2008). Patterns of positive selection in six mammalian genomes. *PLoS Genetics*, 4(8):e1000144.

Koslowski, M. J., Kübler, I., Chamaillard, M., Schaeffeler, E., Reinisch, W., Wang, G., Beisner, J., Teml, A., Peyrin-Biroulet, L., Winter, S., and et al. (2009). Genetic variants of wnt transcription factor tcf-4 (tcf7l2) putative promoter region are associated with small intestinal crohn's disease. *PLoS ONE*, 4(2):e4496.

Kotenko, S. V., Gallagher, G., Baurin, V. V., Lewis-Antes, A., Shen, M., Shah, N. K., Langer, J. A., Sheikh, F., Dickensheets, H., and Donnelly, R. P. (2002). Ifn-λs mediate antiviral protection through a distinct class ii cytokine receptor complex. *Nature Immunology*, 4(1):69–77.

Kotenko, S. V., Izotova, L. S., Mirochnitchenko, O. V., Esterova, E., Dickensheets, H., Donnelly, R. P., and Pestka, S. (2000). Identification of the functional interleukin-22 (il-22) receptor complex. *Journal of Biological Chemistry*, 276(4):2725–2732.

Kotenko, S. V., Krause, C. D., Izotova, L. S., Pollack, B. P., Wu, W., and Pestka, S. (1997). Identification and functional characterization of a second chain of the interleukin-10 receptor complex. *The EMBO Journal*, 16(19):5894–5903.

Kozyrev, S. V., Abelson, A.-K., Wojcik, J., Zaghlool, A., Linga Reddy, M. V. P., Sanchez, E., Gunnarsson, I., Svenungsson, E., Sturfelt, G., Jönsen, A., and et al. (2008). Functional variants in the b-cell gene bank1 are associated with systemic lupus erythematosus. *Nature Genetics*, 40(2):211–216.

Krawczyk, C., Oliveira-dos Santos, A., Sasaki, T., Griffiths, E., Ohashi, P. S., Snapper, S., Alt, F., and Penninger, J. M. (2002). Vav1 controls integrin clustering and mhc/peptide-specific cell adhesion to antigen-presenting cells. *Immunity*, 16(3):331–343.

Krebs, J., Wilson, A., and Kisielow, P. (1997). Calmodulin-dependent protein kinase iv during t-cell development. *Biochemical and Biophysical Research Communications*, 241(2):383–389.

Krensky, A. M. and Clayberger, C. (2009). Biology and clinical relevance of granulysin. *Tissue Antigens*, 73(3):193–198.

Kryukov, G. V., Pennacchio, L. A., and Sunyaev, S. R. (2007). Most rare missense alleles are deleterious in humans: Implications for complex disease and association studies. *The American Journal of Human Genetics*, 80(4):727–739.

Kulski, J. K., Kenworthy, W., Bellgard, M., Taplin, R., Okamoto, K., Oka, A., Mabuchi, T., Ozawa, A., Tamiya, G., and Inoko, H. (2005). Gene expression profiling of japanese psoriatic skin reveals an increased activity in molecular stress and immune response signals. *Journal of Molecular Medicine*, 83(12):964–975.

Kumar, M., Liu, H., and Rice, A. P. (2012). Regulation of interferon-β by magi-1 and its interaction with influenza a virus ns1 protein with esev pbm. *PLoS ONE*, 7(7):e41251.

Kurowska, M., Goudin, N., Nehme, N. T., Court, M., Garin, J., Fischer, A., de Saint Basile, G., and Menasche, G. (2012). Terminal transport of lytic granules to the immune synapse is mediated by the kinesin-1/slp3/rab27a complex. *Blood*, 119(17):3879–3889.

Lafourcade, C., Sobo, K., Kieffer-Jaquinod, S., Garin, J., and van der Goot, F. G. (2008). Regulation of the v-atpase along the endocytic pathway occurs through reversible subunit association and membrane localization. *PLoS ONE*, 3(7):e2758.

Lao, O., de Gruijter, J. M., van Duijn, K., Navarro, A., and Kayser, M. (2007). Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Annals of Human Genetics*, 71(3):354–369.

Lapinski, P. E., Oliver, J. A., Bodie, J. N., Marti, F., and King, P. D. (2009). The t-cell-specific adapter protein family: Tsad, alx, and sh2d4a/sh2d4b. *Immunological Reviews*, 232(1):240–254.

Lawrence, T., Bebien, M., Liu, G. Y., Nizet, V., and Karin, M. (2005). Ikkα limits macrophage nf-χb activation and contributes to the resolution of inflammation. *Nature*, 434(7037):1138–1143.

Lazear, H. M., Nice, T. J., and Diamond, M. S. (2015). Interferon-λ: Immune functions at barrier surfaces and beyond. *Immunity*, 43(1):15–28.

Le Breton, M., Meyniel-Schicklin, L., Deloire, A., Coutard, B., Canard, B., de Lamballerie, X., Andre, P., Rabourdin-Combe, C., Lotteau, V., and Davoust, N. (2011). Flavivirus ns3 and ns5 proteins interaction network: a high-throughput yeast two-hybrid screen. *BMC Microbiology*, 11(1):234.

Lebbink, R. J., de Ruiter, T., Adelmeijer, J., Brenkman, A. B., van Helvoort, J. M., Koch, M., Farndale, R. W., Lisman, T., Sonnenberg, A., Lenting, P. J., and et al. (2006). Collagens are functional, high affinity ligands for the inhibitory immune receptor lair-1. *The Journal of Experimental Medicine*, 203(6):1419–1425.

Leber, A., Bassaganya-Riera, J., Tubau-Juni, N., Zoccoli-Rodriguez, V., Viladomiu, M., Abedi, V., Lu, P., and Hontecillas, R. (2016). Modeling the role of lanthionine synthetase c-like 2 (lancl2) in the modulation of immune responses to helicobacter pylori infection. *PLOS ONE*, 11(12):e0167440.

Lee, B. L., Moon, J. E., Shu, J. H., Yuan, L., Newman, Z. R., Schekman, R., and Barton, G. M. (2013). Unc93b1 mediates differential trafficking of endosomal tlrs. *eLife*, 2.

Lee, M. N., Roy, M., Ong, S.-E., Mertins, P., Villani, A.-C., Li, W., Dotiwala, F., Sen, J., Doench, J. G., Orzalli, M. H., and et al. (2012). Identification of regulators of the innate immune response to cytosolic dna and retroviral infection by an integrative approach. *Nature Immunology*, 14(2):179–185.

Leffler, E. M., Gao, Z., Pfeifer, S., Segurel, L., Auton, A., Venn, O., Bowden, R., Bontrop, R., Wall, J. D., Sella, G., and et al. (2013). Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science*, 339(6127):1578–1582.

Lei, Y. and Takahama, Y. (2012). Xcl1 and xcr1 in the immune system. *Microbes and Infection*, 14(3):262–267.

Lepin, E., Bastin, J., Allan, D., Roncador, G., Braud, V., Mason, D., Merwe, P. d., McMichael, A., Bell, J., Powis, S., and et al. (2000). Functional characterization of hla-f and binding of hla-f tetramers to ilt2 and ilt4 receptors. *European Journal of Immunology*, 30(12):3552–3561.

Leung, T. F., Li, C. Y., Liu, E. K. H., Tang, N. L. S., Chan, I. H. S., Yung, E., Wong, G. W. K., and Lam, C. W. K. (2006). Asthma and atopy are associated with defb1 polymorphisms in chinese children. *Genes Immunity*, 7(1):59–64.

Levine, B., Mizushima, N., and Virgin, H. W. (2011). Autophagy in immunity and inflammation. *Nature*, 469(7330):323–335.

Lewis, R. S., Kolesnik, T. B., Kuang, Z., D'Cruz, A. A., Blewitt, M. E., Masters, S. L., Low, A., Willson, T., Norton, R. S., and Nicholson, S. E. (2011). Tlr regulation of spsb1 controls inducible nitric oxide synthase induction. *The Journal of Immunology*, 187(7):3798–3805.

Li, C.-S., Zhang, Q., Lee, K.-J., Cho, S.-W., Lee, K.-M., Hahm, K.-B., Choi, S.-C., Yun, K.-J., Chung, H.-T., and Chae, S.-C. (2009). Interleukin-27 polymorphisms are associated with inflammatory bowel diseases in a Korean population. *Journal of Gastroenterology and Hepatology*, 24(10):1692–1696.

Li, J., Zhang, L., Zhou, H., Stoneking, M., and Tang, K. (2010a). Global patterns of genetic diversity and signals of natural selection for human adme genes. *Human Molecular Genetics*, 20(3):528–540.

Li, P., Li, M., Lindberg, M. R., Kennett, M. J., Xiong, N., and Wang, Y. (2010b). Pad4 is essential for antibacterial innate immunity mediated by neutrophil extracellular traps. *The Journal of Experimental Medicine*, 207(9):1853–1862.

Li, S., Wang, L., Berman, M., Kong, Y.-Y., and Dorf, M. E. (2011). Mapping a dynamic innate immunity protein interaction network regulating type i interferon production. *Immunity*, 35(3):426–440.

Li, X., Li, J., Yang, Y., Hou, R., Liu, R., Zhao, X., Yan, X., Yin, G., An, P., Wang, Y., and Zhang, K. (2013). Differential gene expression in peripheral blood T cells from patients with psoriasis, lichen planus, and atopic dermatitis. *Journal of the American Academy of Dermatology*, 69(5):e235–e243.

Li, Y., Cheng, H., Xiao, F.-L., Liang, B., Zhou, F.-s., Li, P., Zheng, X.-d., Sun, L.-d., Yang, S., and Zhang, X.-j. (2017). Association of ubash3a gene polymorphism and atopic dermatitis in the chinese han population. *Genes and Immunity*, 18(3):158–162.

Liang, Y., Wang, P., Zhao, M., Liang, G., Yin, H., Zhang, G., Wen, H., and Lu, Q. (2011). Demethylation of the fcer1g promoter leads to fcεri overexpression on monocytes of patients with atopic dermatitis. *Allergy*, 67(3):424–430.

Liao, X., Buchberg, A. M., Jenkins, N. A., and Copeland, N. G. (1995). Evi-5, a common site of retroviral integration in akxd t-cell lymphomas, maps near gfi-1 on mouse chromosome 5. *Journal of Virology*, 69(11):7132–7137.

Ligtenberg, T. J. M., Bikker, F. J., Groenink, J., Tornoe, I., Leth-Larsen, R., Veerman, E. C. I., Nieuw Amerongen, A. V., and Holmskov, U. (2001). Human salivary agglutinin binds to lung surfactant protein-d and is identical with scavenger receptor protein gp-340. *Biochemical Journal*, 359(1):243–248.

Limou, S., Dummer, P. D., Nelson, G. W., Kopp, J. B., and Winkler, C. A. (2015). Apol1 toxin, innate immunity, and kidney injury. *Kidney International*, 88(1):28–34.

Lin, M., Zhao, Z., Yang, Z., Meng, Q., Tan, P., Xie, W., Qin, Y., Wang, R.-F., and Cui, J. (2016). Usp38 inhibits type i interferon signaling by editing tbk1 ubiquitination through nlrp4 signalosome. *Molecular Cell*, 64(2):267–281.

Lindo, J., Huerta-Sánchez, E., Nakagome, S., Rasmussen, M., Petzelt, B., Mitchell, J., Cybulski, J. S., Willerslev, E., DeGiorgio, M., and Malhi, R. S. (2016). A time transect of exomes from a native american population before and after european contact. *Nature Communications*, 7:13175.

Liu, B., Li, N. L., Shen, Y., Bao, X., Fabrizio, T., Elbahesh, H., Webby, R. J., and Li, K. (2016a). The c-terminal tail of trim56 dictates antiviral restriction of influenza a and b viruses by impeding viral rna synthesis. *Journal of Virology*, 90(9):4369–4382.

Liu, C., Xu, Z., Gupta, D., and Dziarski, R. (2001). Peptidoglycan recognition proteins. *Journal of Biological Chemistry*, 276(37):34686–34694.

Liu, F., Wen, B., and Kayser, M. (2013a). Colorful dna polymorphisms in humans. *Seminars in Cell Developmental Biology*, 24(6-7):562–575.

Liu, H. M., Aizaki, H., Choi, K. S., Machida, K., Ou, J. J.-H., and Lai, M. M. (2009). Syncrip (synaptotagmin-binding, cytoplasmic rna-interacting protein) is a host factor involved in hepatitis c virus rna replication. *Virology*, 386(2):249–256.

Liu, L., Cao, Z., Chen, J., Li, R., Cao, Y., Zhu, C., Wu, K., Wu, J., Liu, F., and Zhu, Y. (2012). Influenza A Virus Induces Interleukin-27 through Cyclooxygenase-2 and Protein Kinase A Signaling. *Journal of Biological Chemistry*, 287(15):11899–11910.

Liu, L., Zhang, Y., Wang, J., Zhao, H., Jiang, L., Che, Y., Shi, H., Li, R., Mo, Z., Huang, T., and et al. (2013b). Study of the integrated immune response induced by an inactivated ev71 vaccine. *PLoS ONE*, 8(1):e54451.

Liu, W., Yan, M., Liu, Y., Wang, R., Li, C., Deng, C., Singh, A., Coleman, W. G., and Rodgers, G. P. (2010). Olfactomedin 4 down-regulates innate immunity against helicobacter pylori infection. *Proceedings of the National Academy of Sciences*, 107(24):11056–11061.

Liu, X., Zhang, P., Bao, Y., Han, Y., Wang, Y., Zhang, Q., Zhan, Z., Meng, J., Li, Y., Li, N., and et al. (2013c). Zinc finger protein zbtb20 promotes toll-like receptor-triggered innate immune responses by repressing i b gene transcription. *Proceedings of the National Academy of Sciences*, 110(27):11097–11102.

Liu, Y., Qu, L., Liu, Y., Roizman, B., and Zhou, G. G. (2017). Pum1 is a biphasic negative regulator of innate immunity genes by suppressing lgp2. *Proceedings of the National Academy of Sciences*, 114(33):E6902–E6911.

Liu, Z., Yang, L., Sun, Y., Xie, X., and Huang, J. (2016b). Asf1a enhances antiviral immune response by associating with cbp to mediate acetylation of h3k56 at the ifnb promoter. *Molecular Immunology*, 78:57–64.

Lopez-Herrera, G., Tampella, G., Pan-Hammarström, Q., Herholz, P., Trujillo-Vargas, C. M., Phadwal, K., Simon, A. K., Moutschen, M., Etzioni, A., Mory, A., and et al. (2012). Deleterious mutations in lrba are associated with a syndrome of immune deficiency and autoimmunity. *The American Journal of Human Genetics*, 90(6):986–1001.

Lozach, P.-Y., Amara, A., Bartosch, B., Virelizier, J.-L., Arenzana-Seisdedos, F., Cosset, F.-L., and Altmeyer, R. (2004). C-type lectins l-sign and dc-sign capture and transmit infectious hepatitis c virus pseudotype particles. *Journal of Biological Chemistry*, 279(31):32035–32045.

Lu, P., Hankel, I. L., Knisz, J., Marquardt, A., Chiang, M.-Y., Grosse, J., Constien, R., Meyer, T., Schroeder, A., Zeitlmann, L., and et al. (2010). The justy mutation identifies gon4-like as a gene that is essential for b lymphopoiesis. *The Journal of Experimental Medicine*, 207(7):1359–1367.

Lu, Q. (2001). Homeostatic regulation of the immune system by receptor tyrosine kinases of the tyro 3 family. *Science*, 293(5528):306–311.

Lu, X., Wang, M., Qi, J., Wang, H., Li, X., Gupta, D., and Dziarski, R. (2005). Peptidoglycan recognition proteins are a new class of human bactericidal proteins. *Journal of Biological Chemistry*, 281(9):5895–5907.

Lupberger, J., Zeisel, M. B., Xiao, F., Thumann, C., Fofana, I., Zona, L., Davis, C., Mee, C. J., Turek, M., Gorke, S., and et al. (2011). Egfr and epha2 are host factors for hepatitis c virus entry and possible targets for antiviral therapy. *Nature Medicine*, 17(5):589–595.

Mackenzie, J. M., Khromykh, A. A., and Parton, R. G. (2007). Cholesterol manipulation by west nile virus perturbs the cellular immune response. *Cell Host Microbe*, 2(4):229–239.

Mackey-Lawrence, N. M. and Petri, W. A. (2012). Leptin and mucosal immunity. *Mucosal Immunology*, 5(5):472–479.

Mahil, S. K., Twelves, S., Farkas, K., Setta-Kaffetzi, N., Burden, A. D., Gach, J. E., Irvine, A. D., Képíró, L., Mockenhaupt, M., Oon, H. H., and et al. (2016). Ap1s3 mutations cause skin autoinflammation by disrupting keratinocyte autophagy and up-regulating il-36 production. *Journal of Investigative Dermatology*, 136(11):2251–2259.

Mahoney, J. A., Ntolosi, B., DaSilva, R. P., Gordon, S., and McKnight, A. J. (2001). Cloning and characterization of cpvl, a novel serine carboxypeptidase, from human macrophages. *Genomics*, 72(3):243–251.

Maingret, F., Patel, A. J., Lesage, F., Lazdunski, M., and Honoré, E. (2000). Lysophospholipids open the two-pore domain mechano-gated k+channels trek-1 and traak. *Journal of Biological Chemistry*, 275(14):10128–10133.

Malaspinas, A.-S., Westaway, M. C., Muller, C., Sousa, V. C., Lao, O., Alves, I., Bergström, A., Athanasiadis, G., Cheng, J. Y., Crawford, J. E., Heupink, T. H., Macholdt, E., Peischl, S., Rasmussen, S., Schiffels, S., Subramanian, S., Wright, J. L., Albrechtsen, A., Barbieri, C., Dupanloup, I., Eriksson, A., Margaryan, A., Moltke, I., Pugach, I., Korneliussen, T. S., Levkivskyi, I. P., Moreno-Mayar, J. V., Ni, S., Racimo, F., Sikora, M., Xue, Y., Aghakhanian, F. A., Brucato, N., Brunak, S., Campos, P. F., Clark, W., Ellingvåg, S., Fourmile, G., Gerbault, P., Injie, D., Koki, G., Leavesley, M., Logan, B., Lynch, A., Matisoo-Smith, E. A., McAllister, P. J., Mentzer, A. J., Metspalu, M., Migliano, A. B., Murgha, L., Phipps, M. E., Pomat, W., Reynolds, D., Ricaut, F.-X., Siba, P., Thomas, M. G., Wales, T., Wall, C. M., Oppenheimer, S. J., Tyler-Smith, C., Durbin, R., Dortch, J., Manica, A., Schierup, M. H., Foley, R. A., Lahr, M. M., Bowern, C., Wall, J. D., Mailund, T., Stoneking, M., Nielsen, R., Sandhu, M. S., Laurent Excoffier, L., Lambert, D. M., and Willerslev, E. (2016). A genomic history of aboriginal australia. *Nature*, 538(7624):207–214.

Malathi, K., Dong, B., Gale, M., and Silverman, R. H. (2007). Small self-rna generated by rnase l amplifies antiviral innate immunity. *Nature*, 448(7155):816–819.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., Skoglund, P., Lazaridis, I., Sankararaman, S., Fu, Q., Rohland, N., Renaud, G., Erlich, Y., Willems, T., Gallo, C., Spence, J. P., Song, Y. S., Poletti, G., Balloux, F., van Driem, G., de Knijff, P., Gallego Romero, I., Jha, A. R., Behar, D. M., Bravi, C. M., Capelli, C., Hervig, T., Moreno-Estrada, A., Posukh, O. L., Balanovska, E., Balanovsky, O., Karachanak-Yankova, S., Sahakyan, H., Toncheva, D., Yepiskoposyan, L., Tyler-Smith, C., Xue, Y., Abdullah, M. S., Ruiz-Linares, A., Beall, C. M., Di Rienzo, A., Jeong, C., Starikovskaya, E. B., Metspalu, E., Parik, J., Villems, R., Henn, B. M., Hodoglugil, U., Mahley, R., Sajantila, A., Stamatoyannopoulos, G., Wee, J. T. S., Khusainova, R., Khusnutdinova, E., Litvinov, S., Ayodo, G., Comas, D., Hammer, M. F., Kivisild, T., Klitz, W., Winkler, C. A., Labuda, D., Bamshad, M., Jorde, L. B., Tishkoff, S. A., Watkins, W. S., Metspalu, M., Dryomov, S., Sukernik, R., Singh, L., Thangaraj, K., Pääbo, S., Kelso, J., Patterson, N., and Reich, D. (2016). The simons genome diversity project: 300 genomes from 142 diverse populations. *Nature*, 538(7624):201–206.

Mammen, A. L., Chung, T., Christopher-Stine, L., Rosen, P., Rosen, A., Doering, K. R., and Casciola-Rosen, L. A. (2011). Autoantibodies against 3-hydroxy-3-methylglutaryl-coenzyme a reductase in patients with statin-associated autoimmune myopathy. *Arthritis Rheumatism*, 63(3):713–721.

Manjurano, A., Sepúlveda, N., Nadjm, B., Mtove, G., Wangai, H., Maxwell, C., Olomi, R., Reyburn, H., Drakeley, C. J., Riley, E. M., and et al. (2015). Usp38, frem3, sdc1, ddc,andloc727982gene polymorphisms and differential susceptibility to severe malaria in tanzania. *Journal of Infectious Diseases*, 212(7):1129–1139.

Mantovani, A., Cassatella, M. A., Costantini, C., and Jaillon, S. (2011). Neutrophils in the activation and regulation of innate and adaptive immunity. *Nature Reviews Immunology*, 11(8):519–531.

Marballi, K., Quinones, M. P., Jimenez, F., Escamilla, M. A., Raventós, H., Soto-Bernardini, M. C., Ahuja, S. S., and Walss-Bass, C. (2010). In vivo and in vitro genetic evidence

of involvement of neuregulin 1 in immune system dysregulation. *Journal of Molecular Medicine*, 88(11):1133–1141.

Marrakchi, S., Guigue, P., Renshaw, B. R., Puel, A., Pei, X.-Y., Fraitag, S., Zribi, J., Bal, E., Cluzeau, C., Chrabieh, M., and et al. (2011). Interleukin-36–receptor antagonist deficiency and generalized pustular psoriasis. *New England Journal of Medicine*, 365(7):620–628.

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., Daly, M. J., Bustamante, C. D., and Kenny, E. E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *The American Journal of Human Genetics*, 100(4):635–649.

Martin, M., Romero, X., de la Fuente, M. A., Tovar, V., Zapater, N., Esplugues, E., Pizcueta, P., Bosch, J., and Engel, P. (2001). Cd84 functions as a homophilic adhesion molecule and enhances ifn- secretion: Adhesion is mediated by ig-like domain 1. *The Journal of Immunology*, 167(7):3668–3676.

Martinez, N. M. and Lynch, K. W. (2013). Control of alternative splicing in immune responses: many regulators, many predictions, much still to learn. *Immunological Reviews*, 253(1):216–236.

Martini, C., Trapani, L., Narciso, L., Marino, M., Trentalance, A., and Pallottini, V. (2009). 3-hydroxy 3-methylglutaryl coenzyme a reductase increase is essential for rat muscle differentiation. *Journal of Cellular Physiology*, 220(2):524–530.

Mashima, R., Saeki, K., Aki, D., Minoda, Y., Takaki, H., Sanada, T., Kobayashi, T., Aburatani, H., Yamanashi, Y., and Yoshimura, A. (2005). Fln29, a novel interferon- and lps-inducible gene acting as a negative regulator of toll-like receptor signaling. *Journal of Biological Chemistry*, 280(50):41289–41297.

Matsunaga, M., Isowa, T., Murakami, H., Kasugai, K., Yoneda, M., Kaneko, H., and Ohira, H. (2009). Association of polymorphism in the human μ-opioid receptor oprm1 gene with proinflammatory cytokine levels and health perception. *Brain, Behavior, and Immunity*, 23(7):931–935.

McClelland, E. E., Penn, D. J., and Potts, W. K. (2003). Major histocompatibility complex heterozygote superiority during coinfection. *Infection and Immunity*, 71(4):2079–2086.

McClure, S. B. (2015). The pastoral effect. *Current Anthropology*, 56(6):901–910.

McCullough, Heath, and Smith (2015). Hemochromatosis: Niche construction and the genetic domino effect in the european neolithic. *Human Biology*, 87(1):39.

McIntire, J. J., Umetsu, S. E., Macaubas, C., Hoyte, E. G., Cinnioglu, C., Cavalli-Sforza, L. L., Barsh, G. S., Hallmayer, J. F., Underhill, P. A., Risch, N. J., and et al. (2003). Hepatitis a virus link to atopic disease. *Nature*, 425(6958):576–576.

McIntyre, K. W. (1991). Inhibition of interleukin 1 (il-1) binding and bioactivity in vitro and modulation of acute inflammation in vivo by il-1 receptor antagonist and anti-il-1 receptor monoclonal antibody. *Journal of Experimental Medicine*, 173(4):931–939.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. *Genome Biology*, 17(1).

McNeill, W. H. (1976). *Plagues and Peoples*. Anchor Books.

Meesmann, H. M., Fehr, E.-M., Kierschke, S., Herrmann, M., Bilyy, R., Heyder, P., Blank, N., Krienke, S., Lorenz, H.-M., and Schiller, M. (2010). Decrease of sialic acid residues as an eat-me signal on the surface of apoptotic lymphocytes. *Journal of Cell Science*, 123(19):3347–3356.

Meli, A. P., Fontés, G., Avery, D. T., Leddon, S. A., Tam, M., Elliot, M., Ballesteros-Tato, A., Miller, J., Stevenson, M. M., Fowell, D. J., and et al. (2016). The integrin lfa-1 controls t follicular helper cell generation and maintenance. *Immunity*, 45(4):831–846.

Mellett, M., Atzei, P., Bergin, R., Horgan, A., Floss, T., Wurst, W., Callanan, J. J., and Moynagh, P. N. (2015). Orphan receptor il-17rd regulates toll-like receptor signalling via sefir/tir interactions. *Nature Communications*, 6(1).

Mellett, M., Atzei, P., Horgan, A., Hams, E., Floss, T., Wurst, W., Fallon, P. G., and Moynagh, P. N. (2012). Orphan receptor il-17rd tunes il-17a signalling and is required for neutrophilia. *Nature Communications*, 3(1).

Miao, E. A., Leaf, I. A., Treuting, P. M., Mao, D. P., Dors, M., Sarkar, A., Warren, S. E., Wewers, M. D., and Aderem, A. (2010a). Caspase-1-induced pyroptosis is an innate immune effector mechanism against intracellular bacteria. *Nature Immunology*, 11(12):1136–1142.

Miao, E. A., Mao, D. P., Yudkovsky, N., Bonneau, R., Lorang, C. G., Warren, S. E., Leaf, I. A., and Aderem, A. (2010b). Innate immune detection of the type iii secretion apparatus through the nlrc4 inflammasome. *Proceedings of the National Academy of Sciences*, 107(7):3076–3080.

Miller, L. H., Mason, S. J., Clyde, D. F., and McGinniss, M. H. (1976). The resistance factor toplasmodium vivaxin blacks. *New England Journal of Medicine*, 295(6):302–304.

Mitchell, A., Rentero, C., Endoh, Y., Hsu, K., Gaus, K., Geczy, C., McNeil, H. P., Borges, L., and Tedla, N. (2008). Lilra5 is expressed by synovial tissue macrophages in rheumatoid arthritis, selectively induces pro-inflammatory cytokines and il-10 and is regulated by tnf-α, il-10 and ifn-γ. *European Journal of Immunology*, 38(12):3459–3473.

Miyamoto, M., Fujita, T., Kimura, Y., Maruyama, M., Harada, H., Sudo, Y., Miyata, T., and Taniguchi, T. (1988). Regulated expression of a gene encoding a nuclear factor, irf-1, that specifically binds to ifn-β gene regulatory elements. *Cell*, 54(6):903–913.

Mollenhauer, J., Herbertz, S., Holmskov, U., Tolnay, M., Krebs, I., Merlo, A., Schrøder, H. D., Maier, D., Breitling, F., Wiemann, S., Gro ne, H.-J., and Poustka, A. (2000). Dmbt1 encodes a protein involved in the immune defense and in epithelial differentiation and is highly unstable in cancer. *Cancer Research*, 60:1704–1710.

Monteleone, G., Del Vecchio Blanco, G., Palmieri, G., Vavassori, P., Monteleone, I., Colantoni, A., Battista, S., Spagnoli, L. G., Romano, M., Borrelli, M., and et al. (2004a). Induction and regulation of smad7 in the gastric mucosa of patients with helicobacter pylori infection. *Gastroenterology*, 126(3):674–682.

Monteleone, G., Pallone, F., and MacDonald, T. T. (2004b). Smad7 in tgf-β-mediated negative regulation of gut inflammation. *Trends in Immunology*, 25(10):513–517.

Moore, C. B., Bergstralh, D. T., Duncan, J. A., Lei, Y., Morrison, T. E., Zimmermann, A. G., Accavitti-Loper, M. A., Madden, V. J., Sun, L., Ye, Z., and et al. (2008). Nlrx1 is a regulator of mitochondrial antiviral immunity. *Nature*, 451(7178):573–577.

Moreno-Estrada, A., Tang, K., Sikora, M., Marquès-Bonet, T., Casals, F., Navarro, A., Calafell, F., Bertranpetit, J., Stoneking, M., and Bosch, E. (2009). Interrogating 11 Fast-Evolving Genes for Signatures of Recent Positive Selection in Worldwide Human Populations. *Molecular Biology and Evolution*, 26(10):2285–2297.

Mousson, F., Ochsenbein, F., and Mann, C. (2006). The histone chaperone asf1 at the crossroads of chromatin and dna checkpoint pathways. *Chromosoma*, 116(2):79–93.

Murray, N., Norton, H. L., and Parra, E. J. (2015). Distribution of two oca2 polymorphisms associated with pigmentation in east-asian populations. *Human Genome Variation*.

Murray, P. R., Rosenthal, K. S., and Pfaller, M. A. (2009). *Medical Microbiology*. Elsevier Health Sciences, 6 edition.

Murugaiyan, G., Mittal, A., Lopez-Diego, R., Maier, L. M., Anderson, D. E., and Weiner, H. L. (2009). IL-27 Is a Key Regulator of IL-10 and IL-17 Production by Human CD4+ T Cells. *The Journal of Immunology*, 183(4):2435–2443.

Myouzen, K., Kochi, Y., Okada, Y., Terao, C., Suzuki, A., Ikari, K., Tsunoda, T., Takahashi, A., Kubo, M., Taniguchi, A., and et al. (2012). Functional variants in nfkbie and rtkn2 involved in activation of the nf-χb pathway are associated with rheumatoid arthritis in japanese. *PLoS Genetics*, 8(9):e1002949.

Nagata, K. and Hirai, H. (2003). The second pgd2 receptor crth2: structure, properties, and functions in leukocytes. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 69(2-3):169–177.

Nakahira, K., Haspel, J. A., Rathinam, V. A. K., Lee, S.-J., Dolinay, T., Lam, H. C., Englert, J. A., Rabinovitch, M., Cernadas, M., Kim, H. P., and et al. (2010). Autophagy proteins regulate innate immune responses by inhibiting the release of mitochondrial dna mediated by the nalp3 inflammasome. *Nature Immunology*, 12(3):222–230.

Nakajima, T., Wooding, S., Satta, Y., Jinnai, N., Goto, S., Hayasaka, I., Saitou, N., Guan-jun, J., Tokunaga, K., Jorde, L. B., Emi, M., and Inoue, I. (2005). Evidence for natural selection in the havcr1 gene: high degree of amino-acid variability in the mucin domain of human havcr1 protein. *Genes Immunity*, 6(5):398–406.

Nakaya, H. I., Wrammert, J., Lee, E. K., Racioppi, L., Marie-Kunze, S., Haining, W. N., Means, A. R., Kasturi, S. P., Khan, N., Li, G.-M., and et al. (2011). Systems biology of vaccination for seasonal influenza in humans. *Nature Immunology*, 12(8):786–795.

Nakayama, T., Watanabe, Y., Oiso, N., Higuchi, T., Shigeta, A., Mizuguchi, N., Katou, F., Hashimoto, K., Kawada, A., and Yoshie, O. (2010). Eotaxin-3/cc chemokine ligand 26 is a functional ligand for cx3cr1. *The Journal of Immunology*, 185(11):6472–6479.

Nawijn, M. C., Hackett, T. L., Postma, D. S., van Oosterhout, A. J., and Heijink, I. H. (2011). E-cadherin: gatekeeper of airway mucosa and allergic sensitization. *Trends in Immunology*, 32(6):248–255.

Nemeroff, M. E., Barabino, S. M., Li, Y., Keller, W., and Krug, R. M. (1998). Influenza virus ns1 protein interacts with the cellular 30 kda subunit of cpsf and inhibits 3 end formation of cellular pre-mrnas. *Molecular Cell*, 1(7):991–1000.

Nenci, A., Becker, C., Wullaert, A., Gareus, R., van Loo, G., Danese, S., Huth, M., Nikolaev, A., Neufert, C., Madison, B., and et al. (2007). Epithelial nemo links innate immunity to chronic intestinal inflammation. *Nature*, 446(7135):557–561.

Netea, M. G., Azam, T., Ferwerda, G., Girardin, S. E., Walsh, M., Park, J.-S., Abraham, E., Kim, J.-M., Yoon, D.-Y., Dinarello, C. A., and et al. (2005). Il-32 synergizes with nucleotide oligomerization domain (nod) 1 and nod2 ligands for il-1 and il-6 production through a caspase 1-dependent mechanism. *Proceedings of the National Academy of Sciences*, 102(45):16309–16314.

Netea, M. G., Lewis, E. C., Azam, T., Joosten, L. A. B., Jaekal, J., Bae, S.-Y., Dinarello, C. A., and Kim, S.-H. (2008). Interleukin-32 induces the differentiation of monocytes into macrophage-like cells. *Proceedings of the National Academy of Sciences*, 105(9):3515–3520.

Neveu, G., Barouch-Bentov, R., Ziv-Av, A., Gerber, D., Jacob, Y., and Einav, S. (2012). Identification and targeting of an interaction between a tyrosine motif within hepatitis c virus core protein and ap2m1 essential for viral assembly. *PLoS Pathogens*, 8(8):e1002845.

Ng, A. C. Y., Eisenberg, J. M., Heath, R. J. W., Huett, A., Robinson, C. M., Nau, G. J., and Xavier, R. J. (2010). Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. *Proceedings of the National Academy of Sciences*, 108(Supplement$_1$) : 4631 − −4638.

Nguyen, T., Liu, X. K., Zhang, Y., and Dong, C. (2006). Btnl2, a butyrophilin-like molecule that functions to inhibit t cell activation. *The Journal of Immunology*, 176(12):7354–7360.

Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., and Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature*, 541(7637):302–310.

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A. G. (2007). Recent and ongoing selection in the human genome. *Nature Reviews Genetics*, 8(11):857–868.

Nijnik, A., Clare, S., Hale, C., Chen, J., Raisen, C., Mottram, L., Lucas, M., Estabel, J., Ryder, E., Adissu, H., and et al. (2012). The role of sphingosine-1-phosphate transporter spns2 in immune system function. *The Journal of Immunology*, 189(1):102–111.

Nishizaki, S. S. and Boyle, A. P. (2017). Mining the unknown: Assigning function to noncoding single nucleotide polymorphisms. *Trends in Genetics*, 33(1):34–45.

Nold, M. F., Nold-Petry, C. A., Zepp, J. A., Palmer, B. E., Bufler, P., and Dinarello, C. A. (2010). Il-37 is a fundamental inhibitor of innate immunity. *Nature Immunology*, 11(11):1014–1022.

Norén, E., Almer, S., and Söderman, J. (2017). Genetic variation and expression levels of tight junction genes identifies association between magi3 and inflammatory bowel disease. *BMC Gastroenterology*, 17(1).

Nusinzon, I. and Horvath, C. M. (2003). Interferon-stimulated transcription and innate antiviral immunity require deacetylase activity and histone deacetylase 1. *Proceedings of the National Academy of Sciences*, 100(25):14742–14747.

O'Donnell, L. C., Druhan, L. J., and Avalos, B. R. (2002). Molecular characterization and expression analysis of leucine-rich α2-glycoprotein, a novel marker of granulocytic differentiation. *Journal of Leukocyte Biology*, 72(3):478–485.

Oettinger, M., Schatz, D., Gorka, C., and Baltimore, D. (1990). Rag-1 and rag-2, adjacent genes that synergistically activate v(d)j recombination. *Science*, 248(4962):1517–1523.

Ogata, H., Su, I.-h., Miyake, K., Nagai, Y., Akashi, S., Mecklenbräuker, I., Rajewsky, K., Kimoto, M., and Tarakhovsky, A. (2000). The toll-like receptor protein rp105 regulates lipopolysaccharide signaling in b cells. *The Journal of Experimental Medicine*, 192(1):23–30.

Ogawa, K., Takamori, Y., Suzuki, K., Nagasawa, M., Takano, S., Kasahara, Y., Nakamura, Y., Kondo, S., Sugamura, K., Nakamura, M., and et al. (2003). Granulysin in human serum as a marker of cell-mediated immunity. *European Journal of Immunology*, 33(7):1925–1933.

Ogawa, M., Yoshikawa, Y., Kobayashi, T., Mimuro, H., Fukumatsu, M., Kiga, K., Piao, Z., Ashida, H., Yoshida, M., Kakuta, S., and et al. (2011). A tecpr1-dependent selective autophagy pathway targets bacterial pathogens. *Cell Host Microbe*, 9(5):376–389.

Ogura, M., Inoue, T., Yamaki, J., Homma, M. K., Kurosaki, T., and Homma, Y. (2016). Mitochondrial reactive oxygen species suppress humoral immune response through reduction of cd19 expression in b cells in mice. *European Journal of Immunology*, 47(2):406–418.

Ogus, A., Yoldas, B., Ozdemir, T., Uguz, A., Olcen, S., Keser, I., Coskun, M., Cilli, A., and Yegin, O. (2004). The arg753gln polymorphism of the human toll-like receptor 2 gene in tuberculosis disease. *European Respiratory Journal*, 23(2):219–223.

OhAinle, M., Kerns, J. A., Malik, H. S., and Emerman, M. (2006). Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase apobec3h. *Journal of Virology*, 80(8):3853–3862.

Ohshima, Y., Tanaka, Y., Tozawa, H., Takahashi, Y., Maliszewski, C., and Delespesse, G. (1997). Expression and function of ox40 ligand on human dendritic cells. *The Journal of Immunology*, 159:3838–3848.

Okamoto, T., Nishimura, Y., Ichimura, T., Suzuki, K., Miyamura, T., Suzuki, T., Moriishi, K., and Matsuura, Y. (2006). Hepatitis c virus rna replication is regulated by fkbp8 and hsp90. *The EMBO Journal*, 25(20):5015–5025.

Ono, M., Yaguchi, H., Ohkura, N., Kitabayashi, I., Nagamura, Y., Nomura, T., Miyachi, Y., Tsukada, T., and Sakaguchi, S. (2007). Foxp3 controls regulatory t-cell function by interacting with aml1/runx1. *Nature*, 446(7136):685–689.

Ooi, E. L., Chan, S. T., Cho, N. E., Wilkins, C., Woodward, J., Li, M., Kikkawa, U., Tellinghuisen, T., Gale, M., and Saito, T. (2014). Novel antiviral host factor, tnk1, regulates ifn signaling through serine phosphorylation of stat1. *Proceedings of the National Academy of Sciences*, 111(5):1909–1914.

Ornatowska, M., Azim, A. C., Wang, X., Christman, J. W., Xiao, L., Joo, M., and Sadikot, R. T. (2007). Functional genomics of silencing trem-1 on tlr4 signaling in macrophages. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 293(6):L1377–L1384.

Ovsyannikova, I. G., Salk, H. M., Kennedy, R. B., Haralambieva, I. H., Zimmermann, M. T., Grill, D. E., Oberg, A. L., and Poland, G. A. (2016). Gene signatures associated with adaptive humoral immunity following seasonal influenza a/h1n1 vaccination. *Genes Immunity*, 17(7):371–379.

Owen, H. R., Elser, M., Cheung, E., Gersbach, M., Kraus, W. L., and Hottiger, M. O. (2007). Mybbp1a is a novel repressor of nf-kappab. *Journal of Molecular Biology*, 366:725–736.

Pagan, J. K., Wylie, F. G., Joseph, S., Widberg, C., Bryant, N. J., James, D. E., and Stow, J. L. (2003). The t-snare syntaxin 4 is regulated during macrophage activation to function in membrane traffic and cytokine secretion. *Current Biology*, 13(2):156–160.

Pagani, L., Lawson, D. J., Jagoda, E., Mörseburg, A., Eriksson, A., Mitt, M., Clemente, F., Hudjashov, G., DeGiorgio, M., Saag, L., Wall, J. D., Cardona, A., Mägi, R., Sayres, M. A. W., Kaewert, S., Inchley, C., Scheib, C. L., Järve, M., Karmin, M., Jacobs, G. S., Antao, T., Iliescu, F. M., Kushniarevich, A., Ayub, Q., Tyler-Smith, C., Xue, Y., Yunusbayev, B., Tambets, K., Mallick, C. B., Saag, L., Pocheshkhova, E., Andriadze, G., Muller, C., Westaway, M. C., Lambert, D. M., Zoraqi, G., Turdikulova, S., Dalimova, D., Sabitov, Z., Sultana, G. N. N., Lachance, J., Tishkoff, S., Momynaliev, K., Isakova, J., Damba, L. D., Gubina, M., Nymadawa, P., Evseeva, I., Atramentova, L., Utevska, O., Ricaut, F.-X., Brucato, N., Sudoyo, H., Letellier, T., Cox, M. P., Barashkov, N. A., Škaro, V., Mulahasanovic´, L., Primorac, D., Sahakyan, H., Mormina, M., Eichstaedt, C. A., Lichman, D. V., Abdullah, S., Chaubey, G., Wee, J. T. S., Mihailov, E., Karunas, A., Litvinov, S., Khusainova, R., Ekomasova, N., Akhmetova, V., Khidiyatova, I., Marjanović, D., Yepiskoposyan, L., Behar, D. M., Balanovska, E., Metspalu, A., Derenko, M., Malyarchuk, B., Voevoda, M., Fedorova, S. A., Osipova, L. P., Lahr, M. M., Gerbault, P., Leavesley, M., Migliano, A. B., Petraglia, M., Balanovsky, O., Khusnutdinova, E. K., Metspalu, E., Thomas, M. G., Manica, A., Nielsen, R., Villems, R., Willerslev, E., Kivisild, T., and Metspalu, M. (2016). Genomic analyses inform on migration events during the peopling of Eurasia. *Nature*, 538(7624):238–242.

Paik, D., Monahan, A., Caffrey, D. R., Elling, R., Goldman, W. E., and Silverman, N. (2017). Slc46 family transporters facilitate cytosolic innate immune recognition of monomeric peptidoglycans. *The Journal of Immunology*, 199(1):263–270.

Park, H., Staehling, K., Tsang, M., Appleby, M. W., Brunkow, M. E., Margineantu, D., Hockenbery, D. M., Habib, T., Liggitt, H. D., Carlson, G., and et al. (2012). Disruption of fnip1 reveals a metabolic checkpoint controlling b lymphocyte development. *Immunity*, 36(5):769–781.

Park, S. G., Kim, H. J., Min, Y. H., Choi, E.-C., Shin, Y. K., Park, B.-J., Lee, S. W., and Kim, S. (2005). From the cover: Human lysyl-trna synthetase is secreted to trigger proinflammatory response. *Proceedings of the National Academy of Sciences*, 102(18):6356–6361.

Parkes, M., Barrett, J. C., Prescott, N. J., Tremelling, M., Anderson, C. A., Fisher, S. A., Roberts, R. G., Nimmo, E. R., Cummings, F. R., and et al. (2007). Sequence variants in the autophagy gene irgm and multiple other replicating loci contribute to crohn's disease susceptibility. *Nature Genetics*, 39(7):830–832.

Parks, W. C., Wilson, C. L., and López-Boado, Y. S. (2004). Matrix metalloproteinases as modulators of inflammation and innate immunity. *Nature Reviews Immunology*, 4(8):617–629.

Parvatiyar, K., Zhang, Z., Teles, R. M., Ouyang, S., Jiang, Y., Iyer, S. S., Zaver, S. A., Schenk, M., Zeng, S., Zhong, W., and et al. (2012). The helicase ddx41 recognizes the bacterial secondary messengers cyclic di-gmp and cyclic di-amp to activate a type i interferon immune response. *Nature Immunology*, 13(12):1155–1161.

Pastey, M. K., Gower, T. L., Spearman, P. W., Crowe, J. E., and Graham, B. S. (2000). A rhoa-derived peptide inhibits syncytium formation induced by respiratorysyncytial virus and parainfluenza virus type 3. *Nature Medicine*, 6(1):35–40.

Patel, A. J. (1998). A mammalian two pore domain mechano-gated s-like k+ channel. *The EMBO Journal*, 17(15):4283–4290.

Peirce, M. J., Brook, M., Morrice, N., Snelgrove, R., Begum, S., Lanfrancotti, A., Notley, C., Hussell, T., Cope, A. P., and Wait, R. (2010). Themis2/icb1 is a signaling scaffold that selectively regulates macrophage toll-like receptor signaling and cytokine production. *PLoS ONE*, 5(7):e11465.

Penn, D. J., Damjanovich, K., and Potts, W. K. (2002). Mhc heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences*, 99(17):11260–11264.

Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., Werner, J., Villanea, F. A., Mountain, J. L., Misra, R., Carter, N. P., Lee, C., and Stone, A. C. (2007). Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*, 39(10):1256–1260.

Perwitasari, O., Johnson, S., Yan, X., Howerth, E., Shacham, S., Landesman, Y., Baloglu, E., McCauley, D., Tamir, S., Tompkins, S. M., and et al. (2014). Verdinexor, a novel selective inhibitor of nuclear export, reduces influenza a virus replication in vitro and in vivo. *Journal of Virology*, 88(17):10228–10243.

Philipsen, L., Reddycherla, A. V., Hartig, R., Gumz, J., Kästle, M., Kritikos, A., Poltorak, M. P., Prokazov, Y., Turbin, E., Weber, A., and et al. (2017). De novo phosphorylation and conformational opening of the tyrosine kinase lck act in concert to initiate t cell receptor signaling. *Science Signaling*, 10(462):eaaf4736.

Pickrell, J. K., Coop, G., Novembre, J., Kudaravalli, S., Li, J. Z., Absher, D., Srinivasan, B. S., Barsh, G. S., Myers, R. M., Feldman, M. W., and et al. (2009). Signals of recent positive selection in a worldwide sample of human populations. *Genome Research*, 19(5):826–837.

Pierini, F. and Lenz, T. L. (2018). Divergent allele advantage at human mhc genes: Signatures of past and ongoing selection. *Molecular Biology and Evolution*.

Pils, S., Gerrard, D. T., Meyer, A., and Hauck, C. R. (2008). Ceacam3: An innate immune receptor directed against human-restricted bacterial pathogens. *International Journal of Medical Microbiology*, 298(7-8):553–560.

Pimenoff, V., Houldcroft, C., Rifkin, R., and Underdown, S. (2018). The role of adna in understanding the coevolutionary patterns of human sexually transmitted infections. *Genes*, 9(7):317.

Pinsonneault, J. (2004). Genetic variants of the human h+/dipeptide transporter pept2: Analysis of haplotype functions. *Journal of Pharmacology and Experimental Therapeutics*, 311(3):1088–1096.

Polley, S., Louzada, S., Forni, D., Sironi, M., Balaskas, T., Hains, D. S., Yang, F., and Hollox, E. J. (2015). Evolution of the rapidly mutating human salivary agglutinin gene (dmbt1) and population subsistence strategy. *Proceedings of the National Academy of Sciences*, 112(16):5105–5110.

Popejoy, A. B. and Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature*, 538(7624):161–164.

Pot, C., Apetoh, L., and Kuchroo, V. K. (2011). Type 1 regulatory T cells (Tr1) in autoimmunity. *Seminars in Immunology*, 23(3):202–208.

Powell, J. D., Pollizzi, K. N., Heikamp, E. B., and Horton, M. R. (2012). Regulation of immune responses by mtor. *Annual Review of Immunology*, 30(1):39–68.

Pritchard, J. K., Pickrell, J. K., and Coop, G. (2010). The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, 20(4):R208–R215.

Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., and et al. (2017). A high-coverage neandertal genome from vindija cave in croatia. *Science*, 358(6363):655–658.

Prugnolle, F., Manica, A., Charpentier, M., Guégan, J. F., Guernier, V., and Balloux, F. (2005a). Pathogen-Driven Selection and Worldwide HLA Class I Diversity. *Current Biology*, 15(11):1022–1027.

Prugnolle, F., Manica, A., Charpentier, M., Guégan, J. F., Guernier, V., and Balloux, F. (2005b). Pathogen-driven selection and worldwide hla class i diversity. *Current Biology*, 15(11):1022–1027.

Pybus, M., Luisi, P., Dall'Olio, G. M., Uzkudun, M., Laayouni, H., Bertranpetit, J., and Engelken, J. (2015). Hierarchical boosting: a machine-learning framework to detect and classify hard selective sweeps in human populations. *Bioinformatics*, page btv493.

Qing, J., Liu, C., Choy, L., Wu, R.-Y., Pagano, J. S., and Derynck, R. (2004). Transforming growth factor /smad3 signaling regulates irf-7 function and transcriptional activation of the beta interferon promoter. *Molecular and Cellular Biology*, 24(3):1411–1425.

Qu, Z., Fu, J., Ma, H., Zhou, J., Jin, M., Mapara, M. Y., Grusby, M. J., and Xiao, G. (2012). Pdlim2 restricts th1 and th17 differentiation and prevents autoimmune disease. *Cell Bioscience*, 2(1):23.

Quaranta, M., Knapp, B., Garzorz, N., Mattii, M., Pullabhatla, V., Pennino, D., Andres, C., Traidl-Hoffmann, C., Cavani, A., Theis, F. J., and et al. (2014). Intraindividual genome expression analysis reveals a specific molecular signature of psoriasis and eczema. *Science Translational Medicine*, 6(244):244ra90–244ra90.

Quintana-Murci, L. and Clark, A. G. (2013). Population genetic tools for dissecting innate immunity in humans. *Nature Reviews Immunology*, 13(4):280–293.

Racimo, F. (2015). Testing for ancient selection using cross-population allele frequency differentiation. *Genetics*, 202(2):733–750.

Racimo, F., Gokhman, D., Fumagalli, M., Ko, A., Hansen, T., Moltke, I., Albrechtsen, A., Carmel, L., Huerta-Sánchez, E., and Nielsen, R. (2016). Archaic adaptive introgression in tbx15/wars2. *Molecular Biology and Evolution*, page msw283.

Racimo, F., Sankararaman, S., Nielsen, R., and Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics*, 16(6):359–371.

Rajagopalan, S. and Long, E. O. (1999). A human histocompatibility leukocyte antigen (hla)-g–specific receptor expressed on all natural killer cells. *The Journal of Experimental Medicine*, 189(7):1093–1100.

Rambaut, A., Posada, D., Crandall, K. A., and Holmes, E. C. (2004). The causes and consequences of hiv evolution. *Nature Reviews Genetics*, 5(1):52–61.

Ramenofsky, A. (2003). Native american disease history: past, present and future directions. *World Archaeology*, 35(2):241–257.

Ramirez-Ortiz, Z. G., Prasad, A., Griffith, J. W., Pendergraft, W. F., Cowley, G. S., Root, D. E., Tai, M., Luster, A. D., El Khoury, J., Hacohen, N., and et al. (2015). The receptor treml4 amplifies tlr7-mediated signaling during antiviral responses and autoimmunity. *Nature Immunology*, 16(5):495–504.

Ramon, H. E., Riling, C. R., Bradfield, J., Yang, B., Hakonarson, H., and Oliver, P. M. (2010). The ubiquitin ligase adaptor ndfip1 regulates t cell-mediated gastrointestinal inflammation and inflammatory bowel disease susceptibility. *Mucosal Immunology*, 4(3):314–324.

Ramos, H. J. and Gale, M. (2011). Rig-i like receptors and their signaling crosstalk in the regulation of antiviral immunity. *Current Opinion in Virology*, 1(3):167–176.

Raverdeau, M. and Mills, K. H. G. (2014). Modulation of t cell and innate immune responses by retinoic acid. *The Journal of Immunology*, 192(7):2953–2958.

Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., and Jackson, R. B. (2014). *Campbell Biology*. Pearson, 10th edition.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M. V., Derevianko, A. P., Hublin, J.-J., Kelso, J., Slatkin, M., and Pääbo, S. (2010). Genetic history of an archaic hominin group from denisova cave in siberia. *Nature*, 468(7327):1053–1060.

Reich, D., Nalls, M. A., Kao, W. H. L., Akylbekova, E. L., Tandon, A., Patterson, N., Mullikin, J., Hsueh, W.-C., Cheng, C.-Y., Coresh, J., and et al. (2009). Reduced neutrophil count in people of african descent is due to a regulatory variant in the duffy antigen receptor for chemokines gene. *PLoS Genetics*, 5(1):e1000360.

Reily, M. M., Pantoja, C., Hu, X., Chinenov, Y., and Rogatsky, I. (2005). The grip1:irf3 interaction as a target for glucocorticoid receptor-mediated immunosuppression. *The EMBO Journal*, 25(1):108–117.

Rengarajan, J., Tang, B., and Glimcher, L. H. (2001). Nfatc2 and nfatc3 regulate th2 differentiation and modulate tcr-responsiveness of naïve th cells. *Nature Immunology*, 3(1):48–54.

Reuter, U., Chiarugi, A., Bolay, H., and Moskowitz, M. A. (2002). Nuclear factor-?b as a molecular target for migraine therapy. *Annals of Neurology*, 51(4):507–516.

Riley, J. C. (2010). Smallpox and american indians revisited. *Journal of the History of Medicine and Allied Sciences*, 65(4):445–477.

Rockman, M. V., Hahn, M. W., Soranzo, N., Goldstein, D. B., and Wray, G. A. (2003). Positive selection on a human-specific transcription factor binding site regulating il4 expression. *Current Biology*, 13(23):2118–2123.

Roe, D., Vierra-Green, C., Pyo, C.-W., Eng, K., Hall, R., Kuang, R., Spellman, S., Ranade, S., Geraghty, D. E., and Maiers, M. (2017). Revealing complete complex kir haplotypes phased by long-read sequencing technology. *Genes and Immunity*, 18(3):127–134.

Rook, G. A. W. (2011). Hygiene hypothesis and autoimmune diseases. *Clinical Reviews in Allergy Immunology*, 42(1):5–15.

Rosenstiel, P., Sina, C., End, C., Renner, M., Lyer, S., Till, A., Hellmig, S., Nikolaus, S., Folsch, U. R., Helmke, B., and et al. (2007). Regulation of dmbt1 via nod2 and tlr4 in intestinal epithelial cells modulates bacterial recognition and invasion. *The Journal of Immunology*, 178(12):8203–8211.

Royet, J. and Dziarski, R. (2007). Peptidoglycan recognition proteins: pleiotropic sensors and effectors of antimicrobial defences. *Nature Reviews Microbiology*, 5(4):264–277.

Saba, J. D., Nara, F., Bielawska, A., Garrett, S., and Hannun, Y. A. (1997). Thebst1gene ofsaccharomyces cerevisiaeis the sphingosine-1-phosphate lyase. *Journal of Biological Chemistry*, 272(42):26087–26090.

Sabbagh, A., Luisi, P., Castelli, E. C., Gineau, L., Courtin, D., Milet, J., Massaro, J. D., Laayouni, H., Moreau, P., Donadi, E. A., and et al. (2013). Worldwide genetic variation at the 3 untranslated region of the hla-g gene: balancing selection influencing genetic diversity. *Genes Immunity*, 15(2):95–106.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., and Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909):832–837.

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T. S., Altshuler, D., and Lander, E. S. (2006). Positive Natural Selection in the Human Lineage. *Science*, 312(5780):1614–1620.

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., Lander, E. S., and TheInternationalHapMapConsortium (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164):913–918.

Sakuntabhai, A., Turbpaiboon, C., Casadémont, I., Chuansumrit, A., Lowhnoo, T., Kajaste-Rudnitski, A., Kalayanarooj, S. M., Tangnararatchakit, K., Tangthawornchaikul, N., Vasanawathana, S., and et al. (2005). A variant in the cd209 promoter is associated with severity of dengue disease. *Nature Genetics*, 37(5):507–513.

Salcedo, R., Stauffer, J. K., Lincoln, E., Back, T. C., Hixon, J. A., Hahn, C., Shafer-Weaver, K., Malyguine, A., Kastelein, R., and Wigginton, J. M. (2004). IL-27 Mediates Complete Regression of Orthotopic Primary and Metastatic Murine Neuroblastoma Tumors: Role for CD8+ T Cells. *The Journal of Immunology*, 173(12):7170–7182.

Sanada, T., Takaesu, G., Mashima, R., Yoshida, R., Kobayashi, T., and Yoshimura, A. (2008). Fln29 deficiency reveals its negative regulatory role in the toll-like receptor (tlr) and retinoic acid-inducible gene i (rig-i)-like helicase signaling pathway. *Journal of Biological Chemistry*, 283(49):33858–33864.

Sancho, D. and Reis e Sousa, C. (2012). Signaling by myeloid c-type lectin receptors in immunity and homeostasis. *Annual Review of Immunology*, 30(1):491–529.

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., and Reich, D. (2014). The genomic landscape of neanderthal ancestry in present-day humans. *Nature*, 507(7492):354–357.

Satterly, N., Tsai, P.-L., van Deursen, J., Nussenzveig, D. R., Wang, Y., Faria, P. A., Levay, A., Levy, D. E., and Fontoura, B. M. A. (2007). Influenza virus targets the mrna export machinery and the nuclear pore complex. *Proceedings of the National Academy of Sciences*, 104(6):1853–1858.

Saunders, B. M., Fernando, S. L., Sluyter, R., Britton, W. J., and Wiley, J. S. (2003). A loss-of-function polymorphism in the human p2x7 receptor abolishes atp-mediated killing of mycobacteria. *The Journal of Immunology*, 171(10):5442–5446.

Scheller, J., Chalaris, A., Garbers, C., and Rose-John, S. (2011). Adam17: a molecular switch to control inflammation and tissue regeneration. *Trends in Immunology*, 32(8):380–387.

SCHMAUssER, B., ANDRULIS, M., ENDRICH, S., LEE, S. K., JOSENHANS, C., MULLER-HERMELINK, H.-K., and ECK, M. (2004). Expression and subcellular distribution of toll-like receptors tlr4, tlr5 and tlr9 on the gastric epithelium in helicobacter pylori infection. *Clinical and Experimental Immunology*, 136(3):521–526.

Schwab, S. R. (2005). Lymphocyte sequestration through s1p lyase inhibition and disruption of s1p gradients. *Science*, 309(5741):1735–1739.

Schwefel, D., Arasu, B. S., Marino, S. F., Lamprecht, B., Köchert, K., Rosenbaum, E., Eichhorst, J., Wiesner, B., Behlke, J., Rocks, O., and et al. (2013). Structural insights into the mechanism of gtpase activation in the gimap family. *Structure*, 21(4):550–559.

Selb, R., Eckl-Dorna, J., Neunkirchner, A., Schmetterer, K., Marth, K., Gamper, J., Jahn-Schmid, B., Pickl, W. F., Valenta, R., and Niederberger, V. (2017). Cd23 surface density on b cells is associated with ige levels and determines ige-facilitated allergen uptake, as well as activation of allergen-specific t cells. *Journal of Allergy and Clinical Immunology*, 139(1):290–299.e4.

Serafini, N., Dahdah, A., Barbet, G., Demion, M., Attout, T., Gautier, G., Arcos-Fajardo, M., Souchet, H., Jouvin, M.-H., Vrtovsnik, F., and et al. (2012). The trpm4 channel controls monocyte and macrophage, but not neutrophil, function for survival in sepsis. *The Journal of Immunology*, 189(7):3689–3699.

Setta-Kaffetzi, N., Simpson, M. A., Navarini, A. A., Patel, V. M., Lu, H.-C., Allen, M. H., Duckworth, M., Bachelez, H., Burden, A. D., Choon, S.-E., and et al. (2014). Ap1s3 mutations are associated with pustular psoriasis and impaired toll-like receptor 3 trafficking. *The American Journal of Human Genetics*, 94(5):790–797.

Seya, T., Oshiumi, H., Sasai, M., Akazawa, T., and Matsumoto, M. (2005). Ticam-1 and ticam-2: toll-like receptor adapters that participate in induction of type 1 interferons. *The International Journal of Biochemistry Cell Biology*, 37(3):524–529.

Shakespear, M. R., Halili, M. A., Irvine, K. M., Fairlie, D. P., and Sweet, M. J. (2011). Histone deacetylases as regulators of inflammation and immunity. *Trends in Immunology*, 32(7):335–343.

Sheppard, P., Kindsvogel, W., Xu, W., Henderson, K., Schlutsmeyer, S., Whitmore, T. E., Kuestner, R., Garrigues, U., Birks, C., Roraback, J., and et al. (2002). Il-28, il-29 and their class ii cytokine receptor il-28r. *Nature Immunology*, 4(1):63–68.

Shevach, E. M. and Stephens, G. L. (2006). The gitr–gitrl interaction: co-stimulation or contrasuppression of regulatory activity? *Nature Reviews Immunology*, 6(8):613–618.

Shibuya, A., Campbell, D., Hannum, C., Yssel, H., Franz-Bacon, K., McClanahan, T., Kitamura, T., Nicholl, J., Sutherland, G. R., Lanier, L. L., and et al. (1996). Dnam-1, a novel adhesion molecule involved in the cytolytic function of t lymphocytes. *Immunity*, 4(6):573–581.

Shin, O. S., Uddin, T., Citorik, R., Wang, J. P., Della Pelle, P., Kradin, R. L., Bingle, C. D., Bingle, L., Camilli, A., Bhuiyan, T. R., and et al. (2011). Lplunc1 modulates innate immune responses to vibrio cholerae. *The Journal of Infectious Diseases*, 204(9):1349–1357.

Shum, A. K., Alimohammadi, M., Tan, C. L., Cheng, M. H., Metzger, T. C., Law, C. S., Lwin, W., Perheentupa, J., Bour-Jordan, H., Carel, J. C., and et al. (2013). Bpifb1 is a lung-specific autoantigen associated with interstitial lung disease. *Science Translational Medicine*, 5(206):206ra139–206ra139.

Siddle, K. J. and Quintana-Murci, L. (2014). The Red Queen's long race: human adaptation to pathogen pressure. *Current Opinion in Genetics & Development*, 29:31–38.

Siewert, K. M. and Voight, B. F. (2017a). Detecting long-term balancing selection using allele frequency correlation. *bioarXiv*.

Siewert, K. M. and Voight, B. F. (2017b). Detecting long-term balancing selection using allele frequency correlation. *Molecular Biology and Evolution*.

Singh, C. R., Moulton, R. A., Armitige, L. Y., Bidani, A., Snuggs, M., Dhandayuthapani, S., Huner, R. L., and Jagannath, C. (2006a). Processing and presentation of a mycobacterial antigen 85b epitope by murine macrophages is dependent on the phagosomal acquisition of vacuolar proton atpase and in situ activation of cathepsin d. *The Journal of Immunology*, 177(5):3250–3259.

Singh, P. K., Jia, H. P., Wiles, K., Hesselberth, J., Liu, L., Conway, B.-A. D., Greenberg, E. P., Valore, E. V., Welsh, M. J., Ganz, T., Tack, B. T., and McCray Jr., P. B. (1998). Production of β-defensins by human airway epithelia. *PNAS*, 95(25):14961–14966.

Singh, S. B., Davis, A. S., Taylor, G. A., and Deretic, V. (2006b). Human irgm induces autophagy to eliminate intracellular mycobacteria. *Science*, 313(5792):1438–1441.

Sironi, M. and Clerici, M. (2010). The hygiene hypothesis: an evolutionary perspective. *Microbes and Infection*, 12(6):421–427.

Slade, C., Bosco, J., Unglik, G., Bleasel, K., Nagel, M., and Winship, I. (2013). Deficiency in complement factor b. *New England Journal of Medicine*, 369(17):1667–1669.

Sly, L. M., Rauh, M. J., Kalesnikoff, J., Song, C. H., and Krystal, G. (2004). Lps-induced upregulation of ship is essential for endotoxin tolerance. *Immunity*, 21(2):227–239.

Spiegel, S. and Milstien, S. (2011). The outs and the ins of sphingosine-1-phosphate in immunity. *Nature Reviews Immunology*, 11(6):403–415.

Stajich, J. E. (2004). Disentangling the effects of demography and selection in human history. *Molecular Biology and Evolution*, 22(1):63–73.

Staring, J., von Castelmur, E., Blomen, V. A., van den Hengel, L. G., Brockmann, M., Baggen, J., Thibaut, H. J., Nieuwenhuis, J., Janssen, H., van Kuppeveld, F. J. M., and et al. (2017). Pla2g16 represents a switch between entry and clearance of picornaviridae. *Nature*, 541(7637):412–416.

Steinberg, M. W., Cheung, T. C., and Ware, C. F. (2011). The signaling networks of the herpesvirus entry mediator (tnfrsf14) in immune regulation. *Immunological Reviews*, 244:169–187.

Stow, J. L., Manderson, A. P., and Murray, R. Z. (2006). Snareing immunity: the role of snares in the immune system. *Nature Reviews Immunology*, 6(12):919–929.

Strange, A., Capon, F., Spencer, C. C. A., Knight, J., Weale, M. E., Allen, M. H., Barton, A., Band, G., Bellenguez, C., and et al. (2010). A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between hla-c and erap1. *Nature Genetics*, 42(11):985–990.

Sturm, R. A., Duffy, D. L., Zhao, Z. Z., Leite, F. P., Stark, M. S., Hayward, N. K., Martin, N. G., and Montgomery, G. W. (2008). A single snp in an evolutionary conserved region within intron 86 of the herc2 gene determines human blue-brown eye color. *The American Journal of Human Genetics*, 82(2):424–431.

Suárez-Calvet, X., Gallardo, E., Nogales-Gadea, G., Querol, L., Navas, M., Díaz-Manera, J., Rojas-Garcia, R., and Illa, I. (2014). Altered rig-i/ddx58-mediated innate immunity in dermatomyositis. *The Journal of Pathology*, 233(3):258–268.

Subramanian, H., Gupta, K., Guo, Q., Price, R., and Ali, H. (2011). Mas-related gene x2 (mrgx2) is a novel g protein-coupled receptor for the antimicrobial peptide ll-37 in human mast cells. *Journal of Biological Chemistry*, 286(52):44739–44749.

Sudenga, S. L., Wiener, H. W., King, C. C., Rompalo, A. M., Cu-Uvin, S., Klein, R. S., Shah, K. V., Sobel, J. D., Jamieson, D. J., and Shrestha, S. (2014). Dense genotyping of immune-related loci identifies variants associated with clearance of hpv among hiv-positive women in the hiv epidemiology research study (hers). *PLoS ONE*, 9(6):e99109.

Sumimoto, H., Miyano, K., and Takeya, R. (2005). Molecular composition and regulation of the nox family nad(p)h oxidases. *Biochemical and Biophysical Research Communications*, 338(1):677–686.

Suzuki, K., Kumanogoh, A., and Kikutani, H. (2007). Semaphorins and their receptors in immune cell interactions. *Nature Immunology*, 9(1):17–23.

Swaan, P. W., Bensman, T., Bahadduri, P. M., Hall, M. W., Sarkar, A., Bao, S., Khantwal, C. M., Ekins, S., and Knoell, D. L. (2008). Bacterial peptide recognition and immune activation facilitated by human peptide transporterpept2. *American Journal of Respiratory Cell and Molecular Biology*, 39(5):536–542.

Szpak, M., Mezzavilla, M., Ayub, Q., Chen, Y., Xue, Y., and Tyler-Smith, C. (2018). Finemav: prioritizing candidate genetic variants driving local adaptations in human populations. *Genome Biology*, 19(1).

Tafforeau, L., Chantier, T., Pradezynski, F., Pellet, J., Mangeot, P. E., Vidalain, P.-O., Andre, P., Rabourdin-Combe, C., and Lotteau, V. (2011). Generation and comprehensive analysis of an influenza virus polymerase cellular interaction network. *Journal of Virology*, 85(24):13010–13018.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3):585–595.

Takahashi, D., Hase, K., Kimura, S., Nakatsu, F., Ohmae, M., Mandai, Y., Sato, T., Date, Y., Ebisawa, M., Kato, T., and et al. (2011). The epithelia-specific membrane trafficking factor ap-1b controls gut immune homeostasis in mice. *Gastroenterology*, 141(2):621–632.

Takeda, K., Kaisho, T., and Akira, S. (2003). Toll-likereceptors. *Annual Review of Immunology*, 21(1):335–376.

Tam, S.-Y., Tsai, M., Snouwaert, J. N., Kalesnikoff, J., Scherrer, D., Nakae, S., Chatterjea, D., Bouley, D. M., and Galli, S. J. (2004). Rabgef1 is a negative regulator of mast cell activation and skin inflammation. *Nature Immunology*, 5(8):844–852.

Tang, K., Thornton, K. R., and Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biology*, 5(7):e171.

Tang, W.-F., Yang, S.-Y., Wu, B.-W., Jheng, J.-R., Chen, Y.-L., Shih, C.-H., Lin, K.-H., Lai, H.-C., Tang, P., and Horng, J.-T. (2006). Reticulon 3 binds the 2c protein of enterovirus 71 and is required for viral replication. *Journal of Biological Chemistry*, 282(8):5888–5898.

Tang, X., Woodward, T., and Amar, S. (2010). A ptp4a3 peptide pimap39 modulates tnf-alpha levels and endotoxic shock. *Journal of Innate Immunity*, 2(1):43–55.

Tangye, S. G., Nichols, K. E., Hare, N. J., and van de Weerdt, B. C. M. (2003). Functional requirements for interactions between cd84 and src homology 2 domain-containing proteins and their contribution to human t cell activation. *The Journal of Immunology*, 171(5):2485–2495.

Tanigawa, K., Suzuki, K., Kimura, H., Takeshita, F., Wu, H., Akama, T., Kawashima, A., and Ishii, N. (2009). Tryptophan aspartate-containing coat protein (coro1a) suppresses toll-like receptor signalling inmycobacterium lepraeinfection. *Clinical Experimental Immunology*, 156(3):495–501.

Taniuchi, I., Osato, M., Egawa, T., Sunshine, M. J., Bae, S.-C., Komori, T., Ito, Y., and Littman, D. R. (2002). Differential requirements for runx proteins in cd4 repression and epigenetic silencing during t lymphocyte development. *Cell*, 111(5):621–633.

Tauseef, M., Knezevic, N., Chava, K. R., Smith, M., Sukriti, S., Gianaris, N., Obukhov, A. G., Vogel, S. M., Schraufnagel, D. E., Dietrich, A., and et al. (2012). Tlr4 activation of trpc6-dependent calcium signaling mediates endotoxin-induced lung vascular permeability and inflammation. *The Journal of Experimental Medicine*, 209(11):1953–1968.

Taylor, P. R., Carugati, A., Fadok, V. A., Cook, H. T., Andrews, M., Carroll, M. C., Savill, J. S., Henson, P. M., Botto, M., and Walport, M. J. (2000). A hierarchical role for classical pathway complement proteins in the clearance of apoptotic cells in vivo. *The Journal of Experimental Medicine*, 192(3):359–366.

Teixeira, H. M., Alcantara-Neves, N. M., Barreto, M., Figueiredo, C. A., and Costa, R. S. (2017). Adenylyl cyclase type 9 gene polymorphisms are associated with asthma and allergy in brazilian children. *Molecular Immunology*, 82:137–145.

Thapa, M., Welner, R. S., Pelayo, R., and Carr, D. J. J. (2008). Cxcl9 and cxcl10 expression are critical for control of genital herpes simplex virus type 2 infection through mobilization of hsv-specific ctl and nk cells to the nervous system. *The Journal of Immunology*, 180(2):1098–1106.

The International HapMap Consortium (2005). A haplotype map of the human genome. *Nature*, 437(7063):1299–1320.

The UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids Research*, 43(D1):D204–D212.

Thimmulappa, R. K. (2006). Nrf2 is a critical regulator of the innate immune response and survival during experimental sepsis. *Journal of Clinical Investigation*, 116(4):984–995.

Thomas, M., Laura, R., Hepner, K., Guccione, E., Sawyers, C., Lasky, L., and Banks, L. (2002). Oncogenic human papillomavirus e6 proteins target the magi-2 and magi-3 proteins for degradation. *Oncogene*, 21(33):5088–5096.

Thuong, P. H., Tam, D. B., Sakurada, S., Hang, N. T. L., Hijikata, M., Hong, L. T., Ngoc, P. T. M., Anh, P. T., Cuong, V. C., Matsushita, I., and et al. (2016). Circulating granulysin levels in healthcare workers and latent tuberculosis infection estimated using interferon-gamma release assays. *BMC Infectious Diseases*, 16(1).

Thurston, T. L. M., Wandel, M. P., von Muhlinen, N., Foeglein, Á., and Randow, F. (2012). Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion. *Nature*, 482(7385):414–418.

Timmann, C., Thye, T., Vens, M., Evans, J., May, J., Ehmen, C., Sievertsen, J., Muntau, B., Ruge, G., Loag, W., and et al. (2012). Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature*, 489(7416):443–446.

Tishkoff, S. A. and Kidd, K. K. (2004). Implications of biogeography of human populations for "race" and medicine. *Nature Genetics*, 36(11s):S21–S27.

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B., Osman, M., Ibrahim, M., Omar, S. A., Lema, G., Nyambo, T. B., Ghori, J., Bumpstead, S., Pritchard, J. K., Wray, G. A., and Deloukas, P. (2006). Convergent adaptation of human lactase persistence in africa and europe. *Nature Genetics*, 39(1):31–40.

Tokuhiro, S., Yamada, R., Chang, X., Suzuki, A., Kochi, Y., Sawada, T., Suzuki, M., Nagasaki, M., Ohtsuki, M., Ono, M., and et al. (2003). An intronic snp in a runx1 binding site of slc22a4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nature Genetics*, 35(4):341–348.

Tremblay, N., Baril, M., Chatel-Chaix, L., Es-Saad, S., Park, A. Y., Koenekoop, R. K., and Lamarre, D. (2016). Spliceosome snrnp200 promotes viral rna sensing and irf3 activation of antiviral response. *PLOS Pathogens*, 12(7):e1005772.

Triantafilou, K., Triantafilou, M., and Dedrick, R. L. (2001). A cd14-independent lps receptor cluster. *Nature Immunology*, 2(4):338–345.

Trowsdale, J. (2011). The mhc, disease and selection. *Immunology Letters*, 137(1-2):1–8.

Troyer, J. L., Nelson, G. W., Lautenberger, J. A., Chinn, L., McIntosh, C., Johnson, R. C., Sezgin, E., Kessing, B., Malasky, M., Hendrickson, S. L., and et al. (2011). Genome-wide association study implicates pard3b-based aids restriction. *Journal of Infectious Diseases*, 203(10):1491–1502.

Tsai, P.-L., Chiou, N.-T., Kuss, S., García-Sastre, A., Lynch, K. W., and Fontoura, B. M. A. (2013). Cellular rna binding proteins ns1-bp and hnrnp k regulate influenza a virus rna splicing. *PLoS Pathogens*, 9(6):e1003460.

Tsoi, L. C., Spain, S. L., Knight, J., Ellinghaus, E., Stuart, P. E., Capon, F., Ding, J., Li, Y., Tejasvi, T., Gudjonsson, J. E., and et al. (2012). Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nature Genetics*, 44(12):1341–1348.

Tsuchida, T., Zou, J., Saitoh, T., Kumar, H., Abe, T., Matsuura, Y., Kawai, T., and Akira, S. (2010). The ubiquitin ligase trim56 regulates innate immune responses to intracellular double-stranded dna. *Immunity*, 33(5):765–776.

Tsuji, S., Uehori, J., Matsumoto, M., Suzuki, Y., Matsuhisa, A., Toyoshima, K., and Seya, T. (2001). Human intelectin is a novel soluble lectin that recognizes galactofuranose in carbohydrate chains of bacterial cell wall. *Journal of Biological Chemistry*, 276(26):23456–23463.

Turnbull, I. R., Gilfillan, S., Cella, M., Aoshi, T., Miller, M., Piccio, L., Hernandez, M., and Colonna, M. (2006). Cutting edge: Trem-2 attenuates macrophage activation. *The Journal of Immunology*, 177(6):3520–3524.

Uchimura, K., Hayata, M., Mizumoto, T., Miyasato, Y., Kakizoe, Y., Morinaga, J., Onoue, T., Yamazoe, R., Ueda, M., Adachi, M., and et al. (2014). The serine protease prostasin regulates hepatic insulin sensitivity by modulating tlr4 signalling. *Nature Communications*, 5.

Uematsu, S., Jang, M. H., Chevrier, N., Guo, Z., Kumagai, Y., Yamamoto, M., Kato, H., Sougawa, N., Matsui, H., Kuwata, H., Hemmi, H., Coban, C., Kawai, T., Ishii, K. J., Takeuchi, O., Miyasaka, M., Takeda, K., and Akira, S. (2006). Detection of pathogenic intestinal bacteria by toll-like receptor 5 on intestinal cd11c+ lamina propria cells. *Nature Immunology*, 7(8):868–874.

Ullrich, A., Sures, I., D'Egidio, M., Jallal, B., Powell, T., Herbst, R., Dreps, A., Azam, M., Rubinstein, M., Natoli, C., Shawver, L. K., Shlessinger, J., and Iacobelli, S. (1994). The secreted tumor-associated antigen 90k is a potent immune stimulator. *The Journal of Biological Chemistry*, 269(28):18401–18407.

Unterholzner, L., Keating, S. E., Baran, M., Horan, K. A., Jensen, S. B., Sharma, S., Sirois, C. M., Jin, T., Latz, E., Xiao, T. S., and et al. (2010). Ifi16 is an innate immune sensor for intracellular dna. *Nature Immunology*, 11(11):997–1004.

Valore, E. V., Park, C. H., Quayle, A. J., Wiles, K. R., McCray, P. B., and Ganz, T. (1998). Human beta-defensin-1: an antimicrobial peptide of urogenital tissues. *Journal of Clinical Investigation*, 101(8):1633–1642.

Van Blerkom, L. M. (2003). Role of viruses in human evolution. *American Journal of Physical Anthropology*.

van de Ven, R., Scheffer, G. L., Reurs, A. W., Lindenberg, J. J., Oerlemans, R., Jansen, G., Gillet, J.-P., Glasgow, J. N., Pereboev, A., Curiel, D. T., and et al. (2008). A role for multidrug resistance protein 4 (mrp4; abcc4) in human dendritic cell migration. *Blood*, 112(6):2353–2359.

van de Vosse, E., van Dissel, J. T., and Ottenhoff, T. H. (2009). Genetic deficiencies of innate immune signalling in human infectious disease. *The Lancet Infectious Diseases*, 9(11):688–698.

van Es, J. H., Jay, P., Gregorieff, A., van Gijn, M. E., Jonkheer, S., Hatzis, P., Thiele, A., van den Born, M., Begthel, H., Brabletz, T., and et al. (2005). Wnt signalling induces maturation of paneth cells in intestinal crypts. *Nature Cell Biology*, 7(4):381–386.

Van Stry, M., Oguin, T. H., Cheloufi, S., Vogel, P., Watanabe, M., Pillai, M. R., Dash, P., Thomas, P. G., Hannon, G. J., and Bix, M. (2012). Enhanced susceptibility of ago1/3 double-null mice to influenza a virus infection. *Journal of Virology*, 86(8):4151–4157.

Varinou, L., Ramsauer, K., Karaghiosoff, M., Kolbe, T., Pfeffer, K., Müller, M., and Decker, T. (2003). Phosphorylation of the stat1 transactivation domain is required for full-fledged ifn-γ-dependent innate immunity. *Immunity*, 19(6):793–802.

Vera, J., Fenutría, R., Cañadas, O., Figueras, M., Mota, R., Sarrias, M.-R., Williams, D. L., Casals, C., Yelamos, J., and Lozano, F. (2009). The CD5 ectodomain interacts with conserved fungal cell wall components and protects from zymosan-induced septic shock-like syndrome. *Proceedings of the National Academy of Sciences*.

Vilagos, B., Hoffmann, M., Souabni, A., Sun, Q., Werner, B., Medvedovic, J., Bilic, I., Minnich, M., Axelsson, E., Jaritz, M., and et al. (2012). Essential role of ebf1 in the generation and function of distinct mature b cell types. *The Journal of Experimental Medicine*, 209(4):775–792.

Villarino, A., Hibbert, L., Lieberman, L., Wilson, E., Mak, T., Yoshida, H., Kastelein, R. A., Saris, C., and Hunter, C. A. (2003). The IL-27r (WSX-1) Is Required to Suppress T Cell Hyperactivity during Infection. *Immunity*, 19(5):645–655.

Villarino, A. V., Larkin, J., Saris, C. J. M., Caton, A. J., Lucas, S., Wong, T., Sauvage, F. J. d., and Hunter, C. A. (2005). Positive and Negative Regulation of the IL-27 Receptor during Lymphoid Cell Activation. *The Journal of Immunology*, 174(12):7684–7691.

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006a). A map of recent positive selection in the human genome. *PLoS Biology*, 4(3):e72.

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006b). A Map of Recent Positive Selection in the Human Genome. *PLoS Biol*, 4(3):e72.

Vroon, A., Heijnen, C. J., and Kavelaars, A. (2006). Grks and arrestins: regulators of migration and inflammation. *Journal of Leukocyte Biology*, 80(6):1214–1221.

Waickman, A. T. and Powell, J. D. (2012). mtor, metabolism, and the regulation of t-cell differentiation and function. *Immunological Reviews*, 249(1):43–58.

Walch, M., Dotiwala, F., Mulik, S., Thiery, J., Kirchhausen, T., Clayberger, C., Krensky, A. M., Martinvalet, D., and Lieberman, J. (2014). Cytotoxic cells kill intracellular bacteria through granulysin-mediated delivery of granzymes. *Cell*, 157(6):1309–1323.

Wang, D., Lou, J., Ouyang, C., Chen, W., Liu, Y., Liu, X., Cao, X., Wang, J., and Lu, L. (2010a). Ras-related protein rab10 facilitates tlr4 signaling by promoting replenishment of tlr4 onto the plasma membrane. *Proceedings of the National Academy of Sciences*, 107(31):13806–13811.

Wang, D., Zheng, M., Lei, L., Ji, J., Yao, Y., Qiu, Y., Ma, L., Lou, J., Ouyang, C., Zhang, X., and et al. (2012a). Tespa1 is involved in late thymocyte development through the regulation of tcr-mediated signaling. *Nature Immunology*, 13(6):560–568.

Wang, E. T., Kodama, G., Baldi, P., and Moyzis, R. K. (2006). Global landscape of recent inferred Darwinian selection for Homo sapiens. *Proceedings of the National Academy of Sciences of the United States of America*, 103(1):135–140.

Wang, H., Gupta, D., Li, X., and Dziarski, R. (2005). Peptidoglycan recognition protein 2 (n-acetylmuramoyl-l-ala amidase) is induced in keratinocytes by bacteria through the p38 kinase pathway. *Infection and Immunity*, 73(11):7216–7225.

Wang, J. M., Cheng, Y. Q., Shi, L., Ying, R. S., Wu, X. Y., Li, G. Y., Moorman, J. P., and Yao, Z. Q. (2013a). Klrg1 negatively regulates natural killer cell functions through the akt pathway in individuals with chronic hepatitis c virus infection. *Journal of Virology*, 87(21):11626–11636.

Wang, Y., Koroleva, E. P., Kruglov, A. A., Kuprash, D. V., Nedospasov, S. A., Fu, Y.-X., and Tumanov, A. V. (2010b). Lymphotoxin beta receptor signaling in intestinal epithelial cells orchestrates innate immune responses against mucosal bacterial infection. *Immunity*, 32(3):403–413.

Wang, Y., Shaked, I., Stanford, S. M., Zhou, W., Curtsinger, J. M., Mikulski, Z., Shaheen, Z. R., Cheng, G., Sawatzke, K., Campbell, A. M., and et al. (2013b). The autoimmunity-associated gene ptpn22 potentiates toll-like receptor-driven, type 1 interferon-dependent immunity. *Immunity*, 39(1):111–122.

Wang, Y., Tong, X., Li, G., Li, J., Deng, M., and Ye, X. (2012b). Ankrd17 positively regulates rig-i-like receptor (rlr)-mediated immune signaling. *European Journal of Immunology*, 42(5):1304–1315.

Wang, Z., Sun, Y., Fu, X., Yu, G., Wang, C., Bao, F., Yue, Z., Li, J., Sun, L., Irwanto, A., and et al. (2016). A large-scale genome-wide association and meta-analysis identified four novel susceptibility loci for leprosy. *Nature Communications*, 7:13760.

Wapenaar, M. C., Monsuur, A. J., van Bodegraven, A. A., Weersma, R. K., Bevova, M. R., Linskens, R. K., Howdle, P., Holmes, G., Mulder, C. J., Dijkstra, G., and et al. (2007). Associations with tight junction genes pard3 and magi2 in dutch patients point to a common barrier defect for coeliac disease and ulcerative colitisan unusual case of ascites. *Gut*, 57(4):463–467.

Watanabe, T., Kawakami, E., Shoemaker, J. E., Lopes, T. J., Matsuoka, Y., Tomita, Y., Kozuka-Hata, H., Gorai, T., Kuwahara, T., Takeda, E., and et al. (2014). Influenza virus-host interactome screen as a platform for antiviral drug development. *Cell Host Microbe*, 16(6):795–805.

Weaver, C. T., Hatton, R. D., Mangan, P. R., and Harrington, L. E. (2007). Il-17 family cytokines and the expanding diversity of effector t cell lineages. *Annual Review of Immunology*, 25(1):821–852.

Wei, Z., Sun, M., Liu, X., Zhang, J., and Jin, Y. (2014). Rufy3, a protein specifically expressed in neurons, interacts with actin-bundling protein fascin to control the growth of axons. *Journal of Neurochemistry*, 130(5):678–692.

Wen, H., Ting, J. P.-Y., and O'Neill, L. A. J. (2012). A role for the nlrp3 inflammasome in metabolic diseases—did warburg miss inflammation? *Nature Immunology*, 13(4):352–357.

Widjaja-Adhi, M. A. K., Palczewski, G., Dale, K., Knauss, E. A., Kelly, M. E., Golczak, M., Levine, A. D., and von Lintig, J. (2017). Transcription factor isx mediates the cross talk between diet and immunity. *Proceedings of the National Academy of Sciences*, 114(43):11530–11535.

Wilde, S., Timpson, A., Kirsanow, K., Kaiser, E., Kayser, M., Unterländer, M., Hollfelder, N., Potekhina, I. D., Schier, W., Thomas, M. G., and et al. (2014). Direct evidence for positive selection of skin, hair, and eye pigmentation in europeans during the last 5,000 y. *Proceedings of the National Academy of Sciences*, 111(13):4832–4837.

Williams, J. W., Yau, D., Sethakorn, N., Kach, J., Reed, E. B., Moore, T. V., Cannon, J., Jin, X., Xing, H., Muslin, A. J., and et al. (2013). Rgs3 controls t lymphocyte migration in a model of th2-mediated airway inflammation. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 305(10):L693–L701.

Witkowski, J. M., Soroczynska-Cybula, M., Bryl, E., Smolenska, Z., and Jozwik, A. (2007). Klotho–a common link in physiological and rheumatoid arthritis-related aging of human cd4+ lymphocytes. *The Journal of Immunology*, 178(2):771–777.

Wlasiuk, G., Khan, S., Switzer, W. M., and Nachman, M. W. (2009). A history of recurrent positive selection at the toll-like receptor 5 in primates. *Molecular Biology and Evolution*, 26(4):937–949.

Wright, J. R. (2005). Immunoregulatory functions of surfactant proteins. *Nature Reviews Immunology*, 5(1):58–68.

Wright, S. I. and Charlesworth, B. (2004). The hka test revisited. *Genetics*, 168(2):1071–1076.

Wu, Y., Lan, C., Ren, D., and Chen, G.-Y. (2016). Induction of siglec-1 by endotoxin tolerance suppresses the innate immune response by promoting tgf-β1 production. *Journal of Biological Chemistry*, 291(23):12370–12382.

Xia, Z., Xu, G., Yang, X., Peng, N., Zuo, Q., Zhu, S., Hao, H., Liu, S., and Zhu, Y. (2017). Inducible tap1 negatively regulates the antiviral innate immune response by targeting the tak1 complex. *The Journal of Immunology*, 198(9):3690–3704.

Xiao, Y., Yu, S., Zhu, B., Bedoret, D., Bu, X., Francisco, L. M., Hua, P., Duke-Cohan, J. S., Umetsu, D. T., Sharpe, A. H., and et al. (2014). Rgmb is a novel binding partner for pd-l2 and its engagement with pd-l2 promotes respiratory tolerance. *The Journal of Experimental Medicine*, 211(5):943–959.

Xie, M.-H., Aggarwal, S., Ho, W.-H., Foster, J., Zhang, Z., Stinson, J., Wood, W. I., Goddard, A. D., and Gurney, A. L. (2000). Interleukin (il)-22, a novel human cytokine that signals through the interferon receptor-related proteins crf2–4 and il-22r. *Journal of Biological Chemistry*, 275(40):31335–31339.

Xu, L., Khadijah, S., Fang, S., Wang, L., Tay, F. P. L., and Liu, D. X. (2010). The cellular rna helicase ddx1 interacts with coronavirus nonstructural protein 14 and enhances viral replication. *Journal of Virology*, 84(17):8571–8583.

Xue, Y., Zhang, X., Huang, N., Daly, A., Gillson, C. J., MacArthur, D. G., Yngvadottir, B., Nica, A. C., Woodwark, C., Chen, Y., and et al. (2009). Population differentiation as an indicator of recent positive selection in humans: An empirical evaluation. *Genetics*, 183(3):1065–1077.

Yamanaka, A., Hamano, S., Miyazaki, Y., Ishii, K., Takeda, A., Mak, T. W., Himeno, K., Yoshimura, A., and Yoshida, H. (2004). Hyperproduction of Proinflammatory Cytokines by WSX-1-Deficient NKT Cells in Concanavalin A-Induced Hepatitis. *The Journal of Immunology*, 172(6):3590–3596.

Yamasaki, S., Nishida, K., Sakuma, M., Berry, D., McGlade, C. J., Hirano, T., and Saito, T. (2003). Gads/grb2-mediated association with lat is critical for the inhibitory function of gab2 in t cells. *Molecular and Cellular Biology*, 23(7):2515–2529.

Yamazaki, K., McGovern, D., Ragoussis, J., Paolucci, M., Butler, H., Jewell, D., Cardon, L., Takazoe, M., Tanaka, T., Ichimori, T., and et al. (2005). Single nucleotide polymorphisms in tnfsf15 confer susceptibility to crohn's disease. *Human Molecular Genetics*, 14(22):3499–3506.

Yang, X. (1999). Targeted disruption of smad3 results in impaired mucosal immunity and diminished t cell responsiveness to tgf-beta. *The EMBO Journal*, 18(5):1280–1291.

Yang, X. O., Pappu, B. P., Nurieva, R., Akimzhanov, A., Kang, H. S., Chung, Y., Ma, L., Shah, B., Panopoulos, A. D., Schluns, K. S., and et al. (2008). T helper 17 lineage differentiation is programmed by orphan nuclear receptors rorα and rorγ. *Immunity*, 28(1):29–39.

Yang, Y., Liao, B., Wang, S., Yan, B., Jin, Y., Shu, H.-B., and Wang, Y.-Y. (2013). E3 ligase wwp2 negatively regulates tlr3-mediated innate immune response by targeting trif for ubiquitination and degradation. *Proceedings of the National Academy of Sciences*, 110(13):5115–5120.

Yenugu, S., Richardson, R. T., Sivashanmugam, P., Wang, Z., O'Rand, M. G., French, F. S., and Hall, S. H. (2004). Antimicrobial activity of human eppin, an androgen-regulated, sperm-bound protein with a whey acidic protein motif1. *Biology of Reproduction*, 71(5):1484–1490.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T. S., and et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, 329(5987):75–78.

Yiemwattana, I., Ngoenkam, J., Paensuwan, P., Kriangkrai, R., Chuenjitkuntaworn, B., and Pongcharoen, S. (2011). Essential role of the adaptor protein nck1 in jurkat t cell activation and function. *Clinical Experimental Immunology*, 167(1):99–107.

Yoshida, T., Hanada, T., Tokuhisa, T., Kosai, K.-i., Sata, M., Kohara, M., and Yoshimura, A. (2002). Activation of stat3 by the hepatitis c virus core protein leads to cellular transformation. *The Journal of Experimental Medicine*, 196(5):641–653.

Yu, Y., Cheng, A. S., Wang, L., Dunne, W. M., and Bayliss, S. J. (2007). Hot tub folliculitis or hot hand–foot syndrome caused by pseudomonas aeruginosa. *Journal of the American Academy of Dermatology*, 57(4):596–600.

Yucesoy, B., Talzhanov, Y., Johnson, V. J., Wilson, N. W., Biagini, R. E., Wang, W., Frye, B., Weissman, D. N., Germolec, D. R., Luster, M. I., and et al. (2013). Genetic variants within the mhc region are associated with immune responsiveness to childhood vaccinations. *Vaccine*, 31(46):5381–5391.

Zanetti, M., Gennaro, R., and Romeo, D. (1995). Cathelicidins: a novel protein family with a common proregion and a variable c-terminal antimicrobial domain. *FEBS Letters*, 374(1):1–5.

Zepp, J. A., Nold-Petry, C. A., Dinarello, C. A., and Nold, M. F. (2011). Protection from rna and dna viruses by il-32. *The Journal of Immunology*, 186(7):4110–4118.

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C. G., and et al. (2017). Ensembl 2018. *Nucleic Acids Research*, 46(D1):D754–D761.

Zhang, G., Hou, J., Shi, J., Yu, G., Lu, B., and Zhang, X. (2008). Soluble cd276 (b7-h3) is released from monocytes, dendritic cells and activated t cells and is detectable in normal human serum. *Immunology*, 123(4):538–546.

Zhang, J., Hu, M.-M., Shu, H.-B., and Li, S. (2014). Death-associated protein kinase 1 is an irf3/7-interacting protein that is involved in the cellular antiviral immune response. *Cellular Molecular Immunology*, 11(3):245–252.

Zhang, Y., Blattman, J. N., Kennedy, N. J., Duong, J., Nguyen, T., Wang, Y., Davis, R. J., Greenberg, P. D., Flavell, R. A., and Dong, C. (2004). Regulation of innate and adaptive immune responses by map kinase phosphatase 5. *Nature*, 430(7001):793–797.

Zhang, Z., Deng, C., Lu, Q., and Richardson, B. (2002). Age-dependent dna methylation changes in the itgal (cd11a) promoter. *Mechanisms of Ageing and Development*, 123(9):1257–1268.

Zhang, Z., Duvefelt, K., Svensson, F., Masterman, T., Jonasdottir, G., Salter, H., Emahazion, T., Hellgren, D., Falk, G., Olsson, T., and et al. (2005). Two genes encoding immune-regulatory molecules (lag3 and il7r) confer susceptibility to multiple sclerosis. *Genes Immunity*, 6(2):145–152.

Zhao, C., Jia, M., Song, H., Yu, Z., Wang, W., Li, Q., Zhang, L., Zhao, W., and Cao, X. (2017). The e3 ubiquitin ligase trim40 attenuates antiviral immune responses by targeting mda5 and rig-i. *Cell Reports*, 21(6):1613–1623.

Zhao, C., Wang, I., and Lehrer, R. I. (1996). Widespread expression of beta-defensin hbd-1 in human secretory glands and epithelial cells. *FEBS Letters*, 396(2-3):319–322.

Zhernakova, A., Elbers, C. C., Ferwerda, B., Romanos, J., Trynka, G., Dubois, P. C., de Kovel, C. G., Franke, L., Oosting, M., Barisani, D., Bardella, M. T., FinnishCeliacDiseaseStudy-Group, Joosten, L. A. B., Saavalainen, P., van Heel, D. A., Catassi, C., Netea, M. G., and

Wijmenga, C. (2010). Evolutionary and functional analysis of celiac risk loci reveals sh2b3 as a protective factor against bacterial infection. *The American Journal of Human Genetics*, 86(6):970–977.

Zhernakova, A., Stahl, E. A., Trynka, G., Raychaudhuri, S., Festen, E. A., Franke, L., Westra, H.-J., Fehrmann, R. S. N., Kurreeman, F. A. S., Thomson, B., and et al. (2011). Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-hla shared loci. *PLoS Genetics*, 7(2):e1002004.

Zhong, J., Shi, Q.-Q., Zhu, M.-M., Shen, J., Wang, H.-H., Ma, D., and Miao, C.-H. (2015). Mfhas1 is associated with sepsis and stimulates tlr2/nf-χb signaling pathway following negative regulation. *PLOS ONE*, 10(11):e0143662.

Zhu, M., Granillo, O., Wen, R., Yang, K., Dai, X., Wang, D., and Zhang, W. (2005). Negative regulation of lymphocyte activation by the adaptor protein lax. *The Journal of Immunology*, 174(9):5612–5619.

Zicca, E., Quirino, A., Marascio, N., Nucara, S., Fabiani, F., Trapasso, F., Perrotti, N., Strazzulla, A., Torti, C., Liberto, M. C., and Focà, A. (2014). Interleukin 27 polymorphisms in hcv rna positive patients: is there an impact on response to interferon therapy? *BMC Infectious Diseases*, 14 (Suppl 5).

Appendix A

Full lists of genes in each class of immune genes

A.1 Genes from the Gene Ontology Database

A.1.1 GO.Bact genes

CATSPER4, TEKT2, ISG15, HIST2H2BE, S100A9, S100A12, S100A8, S100A14, SLAMF8, FCER1G, IL10, LYST, TNFRSF14, IL22RA1, IL23R, CD160, S100A7, IL6R, SELP, MR1, PLA2G2A, FGR, PGLYRP3, PGLYRP4, CRP, CD1D, CHIT1, BCL10, DRC1, FZD5, DNAH7, NLRC4, RAB1A, GNLY, SLC11A1, CCL20, REG3G, ITGAV, STAT1, VIL1, LTF, WNT5A, DNAH1, CAMP, TLR9, STAB1, PRKCD, TMF1, MYD88, IL12A, PLD1, CCR4, CD47, MECOM, GPX1, TLR2, TLR6, TLR1, DEFB131, RBPJ, HTN3, HTN1, PPBP, CXCL13, ANXA3, PLAC8, TLR3, ADH7, NFKB1, IL8, CXCL2, ROPN1L, DNAH5, CATSPER3, LEAP2, IL12B, F2RL1, ERAP1, CD14, DNAH8, HIST1H2BC, HIST1H2BE, HIST1H2BF, HIST1H2BG, HIST1H2BI, HIST1H2BJ, HIST1H2BK, TNF, DEFB133, DEFB114, DEFB113, DEFB110, DEFB112, LTA, HLA-DRB1, TNFAIP3, CD36, DNAH11, NOD1, NCF1, VGF, IL6, SERPINE1, ASIC3, BAIAP2L1, WASL, DEFB1, DEFA6, DEFA4, DEFA1, DEFA1B, DEFA3, DEFA5, DEFB4B, DEFB103B, SPAG11B, DEFB104B, DEFB106B, DEFB105B, DEFB107B, DEFB107A, DEFB105A, DEFB106A, DEFB104A, DEFB103A, DEFB4A, DEFB136, DEFB135, DEFB134, DEFB130, RIPK2, IFNB1, IFNE, SYK, TLR4, TNFSF8, CARD9, LCN2, MBL2, PARG, DMBT1, MMP7, TIRAP, DNHD1, CATSPER1, DYNC2H1, SPA17, PRG2, DEFB108B, NLRP10, NLRP6, RELA, LRRC32, TNFRSF1A, LALBA, DCD, IFNG, LYZ, STAB2, CD4, IL23A, KLRK1, TBK1, P2RX7, IRG1, DNAAF2, NEMF, RNASE3, RNASE7, CEBPE, CTSG, GPX2, CATSPER2, B2M, DNAH3, CCDC135, SPN, NOD2, HP, IRF8, PYCARD, IL27, DNAH2,

DNAH17, NLRP1, NOS2, SPACA3, CCL2, PPM1D, BAIAP2, CABYR, MALT1, CATSPERD, ELANE, HAMP, BCL3, AZU1, MYO1F, ACP5, IL27RA, PGLYRP2, PGLYRP1, C5AR1, IRF3, DEFB125, DEFB126, DEFB127, DEFB128, DEFB129, DEFB132, MAVS, CST11, DEFB115, DEFB116, DEFB118, DEFB119, DEFB121, DEFB123, DEFB124, ROMO1, BPI, WFDC12, EPPIN-WFDC6, EPPIN, LBP, HCK

A.1.2 GO.Virus genes

ISG15, IFNLR1, IFI44L, GBP3, GBP1, POLR3C, ADAR, IFI16, RNASEL, PTPRC, TRIM11, LYST, NLRP3, USF1, TNFRSF14, SPEN, EPHA2, UBR4, EIF4G3, LCK, HDAC1, PSMB2, KDM4A, GTF2B, VCAM1, PRMT6, RBM15, MAGI3, PSMB4, SNAPIN, CRTC2, FDPS, CD1D, SLAMF1, F11R, PVRL4, APOA2, FCGR2B, CD247, IVNS1ABP, CD46, TARDBP, JUN, PROX1, IL23R, AIM2, PPIE, PPIH, NR5A2, TCEB3, ARF1, ENO1, STMN1, FGR, IFI44, LAMTOR5, XCL1, TNFSF4, XPR1, BATF3, NFIA, RPL22, RPL11, RPS8, RPL5, RPS27, HNRNPR, RAVER2, PRPF3, RCC1, TAF12, GFI1, TAF13, PSMA5, PSMD4, TPR, NUP133, CD55, BCL2L11, IFIH1, RSAD2, EIF2AK2, CD207, CD8A, STAT1, XRCC5, XPO1, CENPA, VRK2, PDCL3, BUB1, BIN1, ERCC3, CCNT2, CXCR4, ITGB6, DYNC1I2, ITGAV, HSPD1, CFLAR, CASP8, CREB1, SP110, SP100, ACKR3, PPM1B, CD28, POLR2D, CD8B, AP1S3, ODC1, DDX1, APOB, PRKRA, RPS7, RPS27A, RPL31, RPL37A, HNRNPLL, SNRNP200, FSIP2, PSME4, CHMP3, RANBP2, PSMD14, NUP35, PSMD1, EIF2AK3, HYAL2, CD86, PLSCR1, ABCF3, KPNA1, SATB1, KAT2B, DYNC1LI1, FBXL2, PDCD6IP, WDR48, CX3CR1, CCR3, CCR2, CCR5, RHOA, DAG1, RBM15B, VPRBP, CD200R1, CD80, GTF2E1, EIF4G1, EIF4A2, CLDN1, TFRC, PAK2, DLG1, ZNF639, POLR2H, NCBP2, AP2M1, MYD88, MST1R, HYAL3, HYAL1, DHX36, IL12A, UBP1, CXCR6, RPL32, RPL15, RPSA, RPL14, RPL29, RPL24, RPL35A, THOC7, NUP210, PSMD6, CHMP2B, PSMD2, CXCL9, CXCL10, HERC5, ELMOD2, DDX60, TLR3, CTBP1, PDGFRA, KDR, ANKRD17, EIF4E, GYPA, ABCE1, FBXW7, CENPU, SLC25A4, REST, LEF1, PPID, NELFA, POLR2B, IL15, IL8, ERVMER34-1, RPL9, RPL34, RPS3A, HNRNPD, LARP1B, NUP54, F2RL1, POLR3G, IRF1, TMEM173, IL12B, DDX41, XRCC4, ITGA2, IL6ST, TRIM23, PIK3R1, TNPO1, VDAC1, CAMLG, CDC25C, HAVCR1, NPM1, GNB2L1, TNIP1, CDK7, GTF2H2, CCNH, DOCK2, CCT5, MEF2C, RPL37, RPS23, RPS14, RPL26L1, THOC3, NUP155, TAF9, SKP1, MB21D1, HMGA1, TAP2, TAP1, DYNLT1, NUP153, HLA-A, HLA-C, HLA-B, MICB, PSMB8, PSMB9, DAXX, SRPK1, CUL7, LMBRD1, SYNCRIP, RNGTT, FYN, MAP3K5, ULBP3, PSMB1, TBP, TRIM38, TNF, TAF11, GTF2H4, NELFE, TNFAIP3, HS3ST5, IFNGR1, ERVFRD-1, DEK, RPS18, RPS10, RPL10A, RPS12, DDX39B, VTA1, NUP43, HSPA1A, HSPA1B, IL6, CRCP, TRIM56, IRF5, ZC3HAV1, PPIA, CPSF4, RNF216, RALA, CDK13, EIF4H,

COPS6, PILRA, PSMC2, MDFIC, CAV1, SND1, HIPK2, CLEC5A, ZYX, CUL1, SRPK2, POLR2J, RAC1, ELMO1, AP1S1, PSMA2, HSPB1, CDK6, CHRM2, ERV3-1, ERVW-1, NUPL2, POM121, VPS37D, TAF6, NUP205, DEFA1, DEFA1B, DEFA3, POLR3D, BNIP3L, OPRK1, EXOSC4, DERL1, SLC20A2, ATP6V1H, LYN, TRAM1, TCEB1, WWP1, SCRIB, TCEA1, POLR2K, USP17L2, CLU, IKBKB, RPS20, RPL7, RPL30, RPL8, VPS37A, CHMP7, GTF2E2, CHMP4C, VPS28, SLC52A2, DDX58, IFNB1, IFNW1, IFNA21, IFNA4, IFNA7, IFNA10, IFNA16, IFNA17, IFNA14, IFNA5, IFNA6, IFNA13, IFNA2, IFNA8, IFNA1, IFNE, IFNK, CARD9, PSIP1, CREB3, MPDZ, RRAGA, NFX1, HNRNPK, SYK, C9orf156, AMBP, PSMB7, RXRA, C9orf69, NCBP1, CDK9, NELFB, CCL19, LCN2, LRSAM1, RPS6, PCSK5, RPL35, RPL12, RPL7A, CHMP5, PSMD5, NUP188, NUP214, PRF1, DDIT4, POLR3A, IFIT2, IFIT3, IFIT1, IFIT5, GPAM, DMBT1, BNIP3, ZMYND11, KIN, VIM, ITGB1, CUL2, SIRT1, WAPAL, IDE, BTRC, GBF1, TAF5, PDZD8, IL2RA, IFIT1B, GATA3, CXCL12, DDX21, ACTA2, CHUK, CTBP2, RPS24, HNRNPH3, NMT2, BAD, IFITM2, IFITM1, IFITM3, TRIM5, TRIM22, FAM111A, RELA, UNC93B1, FADD, BANF1, IRF7, CD81, NUP98, HPX, IPO7, COPB1, TSG101, HTATIP2, PSMC3, CCDC86, DDB1, NXF1, RTN3, KAT5, SF3B2, PACS1, ACY3, PAAF1, MMP1, CUL5, NLRX1, PVRL1, HSPA8, POU2F3, POLR2L, GTF2H1, SSRP1, POLR2G, RSF1, TSPAN32, AP2A2, EEF1G, CFL1, FOSL1, CREBZF, NCR3LG1, RPLP2, RPL27A, RPS13, FAU, RPS3, RPS25, PSMD13, TAF10, PSMA1, NUP160, VPS37C, CASP12, APOBEC1, KCNJ8, ABCC9, PCBP2, IL23A, STAT2, TBK1, IFNG, POLR3B, OAS1, OAS3, OAS2, OASL, BICD1, CD4, RAN, TNFRSF1A, LTBR, ATF7IP, DDX11, CCNT1, KRT7, KRT8, KRT18, SP1, ATF7, CBX5, HNRNPA1, ITGA5, NACA, MDM2, UNG, RPLP0, DYNLL1, SCARB1, HMGA2, TARBP2, GTF2H3, IRAK3, RPS26, RPL41, RPL6, AAAS, NUP107, NUP37, HCFC2, PSMD9, VPS37B, UBC, HSP90B1, ITGB7, DNAJC3, LIG4, GAS6, RB1, KPNA3, IPO5, EFNB2, CUL4A, GTF2F2, DCLK1, TPT1, RPL21, NUPL1, IRF9, DICER1, PABPN1, PSMB5, AP1G2, NFKBIA, C14orf166, PSMA3, SNW1, RCOR1, KLC1, SIVA1, PACS2, SUPT16H, MNAT1, TRAF3, NPC2, RPS29, HNRNPC, PSMB11, PSME1, PSME2, CHMP4A, PSMA6, PSMC6, GTF2A1, PSMC1, PML, ISG20, SMAD3, UBE3A, NEDD4, GTF2A2, PSMA4, IL16, MFGE8, ANPEP, CRTC3, NTRK3, SIN3A, B2M, DUOX2, CYP1A1, RPL4, RPLP1, RPS17, RPS17L, FURIN, SNRPN, SNRPA1, POLR3K, POLR3E, PYCARD, ITGAX, NLRC5, UBE2I, E4F1, ATP6V0C, KCTD5, CREBBP, UBN1, USP7, MAPK3, SRCAP, PSMB10, WWP2, VAC14, KARS, TFAP4, TCEB2, POLR2C, IL27, SPN, AP1G1, RPS15A, CCL22, IST1, RPL3L, RPS2, RPL13, THOC6, NUP93, PARD6A, VPS4A, PSMD7, SLFN11, DHX58, BECN1, TRIM25, UNC13D, KPNB1, YWHAE, PSMB6, C1QBP, EIF4A1, TP53, EFNB3, SUPT6H, RAB11FIP4, PSMB3, KRT19, KAT2A, STAT3, ITGB3, KPNA2, GRB2, ALYREF, INPP5K,

CCL5, CCL3, CCL4, BTBD17, TOP2A, PFN1, POLR2A, SUPT4H1, AP2B1, DDX5, SPACA3, CCL11, CCL8, IFI35, TBX21, CCL2, RPL26, RPL23A, RPL23, RPL19, RPL27, RPL38, NUP88, UBB, PSMD11, CCL1, PSMD3, PSME3, NMT1, CALCOCO2, PSMC5, PSMD12, NUP85, CHMP6, SLC52A1, PMAIP1, BCL2, TYMS, NEDD4L, CTDP1, SERPINB3, RPL17-C18orf32, RPL17, THOC1, SEH1L, RNMT, PSMA8, TAF4B, VPS4B, TGFB1, PVRL2, TICAM1, ILF3, BST2, LSM14A, IFNL3, IFNL2, IFNL1, EXOSC5, IRF3, ZNF175, LILRB1, CD209, CLEC4M, USF2, POLR2E, SGTA, ICAM1, CARM1, LDLR, RAD23A, GADD45GIP1, BRD4, AP1M1, FKBP8, CRTC1, CEBPA, PVR, ERCC2, BAX, SMARCA4, HPN, IL12RB1, GTF2F1, ELL, POLR2I, SUPT5H, CD37, AP1M2, AP2S1, AP2A1, CCDC130, URI1, HNRNPUL1, BCL3, INSR, ZNF571, RPS15, RPL36, RPS28, RPL18A, UBA52, RPS16, RPS19, RPL18, RPL13A, RPS11, RPS9, RPL28, RPS5, HNRNPL, PSMD8, PSMC4, NUP62, CHMP2A, SLC1A5, BCL2L1, MAVS, POLR3F, ITCH, SAMHD1, CD40, RBCK1, TBC1D20, CD93, HCK, RBL1, SRC, TOP1, PLCG1, ACOT8, VAPB, TAF4, PSMA7, UCKL1, SLPI, NELFCD, RPS21, PSMF1, CHMP4B, RAE1, CXADR, IFNAR1, MX2, MX1, ADARB1, IFNAR2, IFNGR2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H, POLR3H, XRCC6, MAPK1, SMARCB1, IL2RB, RBX1, EP300, ZC3H7B, FBLN1, POLR2F, AP1B1, RPL3, THOC5, SNRPD3, RANBP1, NUP50

A.1.3 GO.Tcell genes

TGFB2, RPL22, RORC, CD247, TNFSF4, WNT4, RC3H1, TNFRSF14, XCL1, PTPN22, VTCN1, TNFAIP8L2, LAX1, PTPRC, GLMN, IL10, IL23R, ITPKB, LCK, BCL10, S100A7, MR1, VCAM1, CD1D, SPTA1, ZP4, ZBTB7B, CD2, ECM1, TNFSF18, PIK3CD, DDOST, CD48, CD1C, PSEN2, FASLG, MTOR, CDC42, LGALS8, SEMA4A, LYST, S1PR1, TNFRSF4, NCSTN, THEMIS2, PDE4B, ENAH, RSAD2, CD8A, POU3F3, IHH, CTLA4, PELI1, ZC3H8, ZAP70, CD28, IGFBP2, PAX8, HSPD1, ADAM17, IL1RL2, IL36B, GLI2, IL1B, NCK2, CD8B, DPP4, CASP8, ICOS, PDCD1, SLC11A1, BCL11A, SP3, NHEJ1, ZFP36L2, FZD7, FZD5, BCL2L11, FKBP1B, CXCR4, RPS27A, CACNB4, INPP5D, SATB1, EOMES, CBLB, BTLA, TIGIT, CD86, APOD, IL20RB, DLG1, CCR2, CD80, IL12A, TGFBR2, CD47, WNT5A, SELK, NCK1, HES1, TRAT1, BCL6, CTNNB1, PIK3CA, PAK2, PIK3CB, LEF1, IL15, CASP3, PDE5A, TXK, IL2, CXCL13, IL21, PPP3CA, RHOH, KIT, ADD1, NFKB1, DOCK2, TCF7, IRF1, IL4, CD74, IL7R, NDFIP1, ITK, IL12B, AP3B1, PRELID1, IL6ST, SPINK5, CAMK4, PRLR, F2RL1, RICTOR, PIK3R1, EGR1, XRCC4, APC, FYB, PDE4D, LCP2, BTN3A1, FYN, STX11, MICB, BTN2A2, HLA-G, HLA-DRB1, THEMIS, CD83, MAP3K7, HLA-DMA, VNN1, HFE, HLA-F, HLA-A, HLA-E, HLA-C, HLA-B, TAP2, TAP1, AIF1, SRF, HLA-DOA, IRF4,

TREML2, RIPK1, HLA-DRA, HLA-DRB5, HLA-DQA1, HLA-DQA2, HLA-DQB2, HLA-DPA1, HLA-DPB1, SOX4, RUNX2, MICA, RAET1E, BTN3A2, BTN3A3, GNL1, CCND3, CRIP3, TAB2, SHH, GLI3, GIMAP5, IKZF1, CARD11, AZGP1, IL6, ZP3, ZNHIT1, PIK3CG, RAC1, CAV1, GIMAP1, CDK6, IDO1, PAG1, RIPK2, IL7, KIF13B, LYN, CHD7, PRKDC, IKBKB, ANXA1, STOML2, FUT7, TNFSF8, CCL19, CCL21, CDKN2A, SLC46A2, IFNB1, IFNA2, NRARP, CD274, PDCD1LG2, PHPT1, SYK, TRAF2, SIT1, MAPKAP1, IL2RA, FAS, NKX2-3, SRGN, GPAM, PRKCQ, PPP3CB, SFTPD, PAX2, GATA3, ADAM8, ADK, ZEB1, MAP3K8, FZD8, CHST3, CXCL12, PTEN, CHUK, INS, CD3E, CD3D, CD3G, AMICA1, NLRP10, LRRC32, FADD, RAG1, SCGB1A1, PTPRJ, THY1, IGF2, IL18, SART1, TRAF6, BAD, CD59, CD5, PAK1, RAG2, MPZL2, CTSC, CD151, RELA, PTPN6, IFNG, LAG3, CD27, DTX1, PAWR, IL23A, IGF1, NCKAP1L, CD4, TESPA1, P2RX7, ABCB9, CLEC7A, KLRK1, PTPN11, WNT1, CACNB3, UBE2N, UBC, ELF1, HSPH1, TNFSF11, TNFSF13B, FLT3, LCP1, LIG4, SIVA1, AKT1, BCL11B, JAG2, BMP4, PNP, ZBTB1, RIPK3, PSEN1, ZFP36L1, PPP2R3C, RNF31, NFKBIA, BLM, RAB27A, CD276, PDCD7, ADAM10, B2M, CHRNA7, NEDD4, SMAD3, CSK, CCNB2, CTSH, ITGAL, ITGAM, ITGAD, DNAJA3, SPN, NOD2, PYCARD, CORO1A, LAT, IL27, MLST8, PDPK1, PSMB10, BCAR1, PLCG2, NCOR1, CCR7, CCL2, ERBB2, DUSP3, CCL5, FOXJ1, MINK1, STAT5B, STAT5A, FOXN1, CD7, CCL3, GRB2, TBX21, TP53, JMJD6, SPNS2, UBB, SKAP1, BCL2, PTPN2, SOCS6, MALT1, YES1, PMAIP1, LILRB1, IFNL1, JAK3, TGFB1, NFKBID, LILRB2, TNFSF9, IL12RB1, EBI3, AP3D1, LGALS16, TNFSF14, KCNN4, PRKD2, STK11, CLC, TRPM4, CLPTM1, CD209, PVR, PVRL2, VAV1, ICAM1, TCF3, BAX, ZBTB32, PIK3R2, UBA52, VASP, PAX1, PRNP, ITCH, ADA, SIRPG, SIGLEC1, FKBP1A, SLA2, SRC, MAFB, RBCK1, PLCG1, UBE2V1, LIME1, CX-ADR, UBASH3A, ICOSLG, SOD1, IFNAR1, PKNOX1, MYH9, ADORA2A, TNFRSF13C, LGALS1, GRAP2, PATZ1, MAPK1

A.1.4 GO.Bcell genes

PIK3CD, TXLNA, CHRNB2, LAX1, CR2, BCL10, VCAM1, GON4L, NTRK1, PTPRC, IL10, LCK, VAV3, MNDA, NOC2L, RC3H1, HES5, TNFSF4, XCL1, TNFRSF4, WNT3A, PTPN22, ZAP70, HSPD1, CASP8, HDAC4, BCL2L11, ADAM17, MSH2, BCL11A, SP3, ITGA4, NHEJ1, CTLA4, INPP5D, ID2, INHA, PELI1, CD86, POU1F1, BCL6, PRKCD, KLHL6, BTLA, HES1, TGFBR2, FOXP1, GCSAM, BANK1, CXCL13, RBPJ, CASP3, LEF1, CD38, TEC, KIT, IL2, TNIP2, IL21, GAPT, IL4, PFDN1, PIK3R1, MEF2C, IL7R, CD180, FNIP1, CD74, IL5, IL13, XRCC4, MZB1, BAK1, TRAF3IP2, MYB, TNFAIP3, PRDM1, CDKN1A, SOX4, HDAC9, SKAP2, LAT2, POLM, IKZF1, FZD9, GIMAP1, SH2B2, INHBA, IL6, CARD11, AHR, TPD52, IKBKB, LYN, PRKDC, BLK, PTK2B,

SFRP1, IL7, IFNB1, SHB, TLR4, SYK, CDKN2A, NOTCH1, LRRC8A, KLF6, DCLRE1C, ITGB1, HHEX, BLNK, NKX2-3, FAS, PLEKHA1, PTEN, MS4A1, CXCR5, RAG1, RAG2, CLCF1, ATM, LPXN, BAD, TIRAP, AICDA, NCKAP1L, PTPN6, PAWR, CD27, GAS6, TNFSF13B, FLT3, GPR183, IRS2, PCID2, LIG4, ZBTB1, TSHR, PPP2R3C, HIF1A, PRKCH, CHRNA7, ONECUT1, MEF2A, PRKCB, PLCG2, CD19, BCAR1, NOD2, IKZF3, HDAC5, SPNS2, TNFRSF13B, TP53, CD79B, FOXJ1, AURKB, STAT5B, STAT5A, FLCN, MALT1, PTPN2, BCL2, BST2, CD79A, BAX, TCF3, LYL1, JAK3, CEBPG, IL11, BCL3, POU2F2, TICAM1, CD40, CHRNA4, SLA2, LIME1, NFATC2, ADA, ICOSLG, SAMSN1, TNFRSF13C, NFAM1, MAPK1, MIF

A.1.5 GO.Innate genes

IFI16, AIM2, ISG15, PIK3CD, MASP2, MTOR, CASP9, DDOST, CDC42, C1QA, C1QC, C1QB, RPS6KA1, FCN3, WASF2, FGR, LCK, AGO4, AGO1, AGO3, C8A, C8B, JUN, PRKACB, BCL10, VAV3, CSF1, CAPZA1, MOV10, NRAS, SIKE1, TXNIP, POLR3GL, POLR3C, FCGR1A, CTSS, CTSK, TNFAIP8L2, THEM4, PGLYRP3, PGLYRP4, S100A9, S100A12, S100A8, S100A7, ADAR, SHC1, CD1D, FCER1A, FCER1G, FCGR2A, FCGR3A, CD247, DHX9, NCF2, ARPC5, CFH, IKBKE, MAPKAPK2, C4BPB, C4BPA, CD55, CR2, CR1, CD46, TLR5, NLRP3, SH2D1B, PTPN22, DUSP10, NLRC4, RSAD2, ADCY3, EIF2AK2, SOS1, PRKCE, CALM2, RPS27A, PELI1, ACTR2, PPP3R1, ACTR3, MARCO, TANK, IFIH1, WIPF1, ATF2, PDE1A, NCKAP1, CASP10, CASP8, CD28, CREB1, ERBB4, XRCC5, ARPC2, ATG9A, IRS1, LRRFIP1, ITPR1, ARPC4, BRK1, IRAK2, PPARG, RAF1, MYD88, CTNNB1, CCR2, TREX1, NCKIPSD, PRKAR2A, UBA7, MST1R, MAPKAPK3, DUSP7, TLR9, PRKCD, PROS1, TRAT1, CD80, GSK3B, CD86, ADCY5, PIK3R4, NCK1, PIK3CB, DHX36, PIK3CA, POLR2H, MASP1, NRROS, PAK2, LTF, PLSCR1, SPON2, FGFR3, TLR10, TLR1, TLR6, KLB, TEC, PDGFRA, KIT, EREG, BTC, FGF5, MAPK10, HERC5, PPP3CA, NFKB1, UBE2D3, CFI, EGF, FGF2, GAB1, TLR2, DDX60, TLR3, TMEM173, ADCY2, FAM105B, MYO10, RICTOR, C9, C7, C6, FGF10, MAP3K1, CD180, PIK3R1, NAIP, F2RL1, MEF2C, POLR3G, CAMK4, TICAM2, ATG12, UBE2D2, NRG2, HBEGF, CD14, FGF1, CSF1R, PDGFRB, IRGM, ITK, CYFIP2, LCP2, FGF18, FGFR4, F12, DDX41, MAPK9, IL4, ERAP1, IRF1, MB21D1, RIPK1, LY86, C2, CFB, C4A, C4B, AGER, ITPR3, SRPK1, MAPK14, CDKN1A, TREML1, TREM2, TREM1, NCR2, CNPY3, PPP2R5D, POLR1C, HSP90AB1, CRISP3, AKIRIN2, MAP3K7, ATG5, FOXO3, FYN, VNN1, MAP3K5, TNFAIP3, TAB2, RPS6KA2, CCR6, PDGFA, PRKAR1B, GPER1, CARD11, ACTB, RAC1, TAX1BP1, WIPF3, NOD1, PDE1C, ELMO1, ADCY1, EGFR, CRCP, LIMK1, LAT2, NCF1, CD36, ARPC1A, TRIM56, SRPK2, PIK3CG, PRKAR2B, CAPZA2, WASL, ZC3HAV1, CLEC5A, DEFB1, DEFA6, DEFA4,

DEFA1, DEFA1B, DEFA3, DEFA5, DEFB4B, DEFB103B, DEFB103A, DEFB4A, CTSB, FGF20, LZTS1, FGF17, POLR3D, CLU, DUSP4, NRG1, FGFR1, IKBKB, PRKDC, LYN, LY96, RIPK2, POLR2K, ADCY8, AGO2, PTK2, TLR4, IFNB1, IFNA21, IFNA4, IFNA7, IFNA10, IFNA16, IFNA17, IFNA14, IFNA5, IFNA6, IFNA2, IFNA8, IFNA1, DDX58, PRSS3, PRKACG, CTSL, SYK, TXN, TRIM32, C5, DAB2IP, MAPKAP1, LCN2, ABL1, VAV2, FCN2, FCN1, CARD9, TRAF2, C8G, IFNK, ADAM8, PRKCQ, ABI1, MAPK8, MBL2, UBE2D1, CDK1, PPP3CB, POLR3A, SFTPD, PTEN, IFIT5, CHUK, FGF8, NFKB2, PPAPDC1A, FGFR2, DMBT1, DOCK1, TRIM5, HRAS, IRF7, POLR2L, CHID1, TOLLIP, ART1, TRIM21, NLRP10, SAA1, LGR4, CD59, TRAF6, UBE2L6, SERPING1, DTX4, MS4A2, DAK, BAD, MAP4K2, RELA, CFL1, ADRBK1, RPS6KB2, UNC93B1, FGF19, FGF4, FGF3, FADD, LRRC32, PAK1, GAB2, PANX1, MRE11A, BIRC3, BIRC2, CASP1, CD3G, NLRX1, TIRAP, APOA4, TBK1, FGF23, FGF6, CD4, C1S, C1RL, CLEC4C, CLEC4A, CLEC6A, CLEC4D, KLRG1, CLEC7A, KLRD1, KLRK1, KLRC2, CDKN1B, KRAS, ITPR2, IRAK4, ADCY6, ATF1, NR4A1, PCBP2, PDE1B, ERBB3, IL23A, STAT6, MDM2, FRS2, KITLG, DUSP6, UBE2N, HSP90B1, POLR3B, ARPC3, PTPN11, P2RX7, UBC, IRAK3, TRAFD1, FGF9, POLR1D, HMGB1, KL, FOXO1, SUGT1, IRS2, IRG1, RNASE7, ADCY4, RIPK3, PRKD1, NFKBIA, LGALS3, PELI2, ZBTB1, FOS, CALM1, RPS6KA5, LGMN, HSP90AA1, TRAF3, AKT1, SIN3A, RASGRP1, B2M, FGF7, MAP2K1, NRG4, PSTPIP1, MEF2A, TYRO3, PYCARD, NOD2, POLR3K, MSRB1, TSC2, MLST8, PDPK1, MEFV, CREBBP, ADCY9, POLR3E, TNRC6A, IL27, CD19, LAT, MAPK3, CORO1A, PYDC1, ITGAM, ADCY7, CYLD, NLRC5, NFATC3, IL34, PHLPP2, PLCG2, CYBA, CRK, MYO1C, PLD2, C1QBP, NLRP1, CLEC10A, MYH2, MAP2K4, UBB, MAPK7, MAP2K3, VTN, RNF135, ERBB2, WIPF2, DHX58, DUSP3, TBKBP1, TRIM25, PRKCA, PRKAR1A, MAP2K6, CD300LB, CD300E, GRB2, TNRC6C, BAIAP2, ACTG1, NOS2, CCL5, COLEC12, YES1, RNF125, PIK3C3, SIGLEC15, MALT1, PHLPP1, BCL2, NFATC1, GZMM, FGF22, CFD, POLR2E, CACTIN, MAP2K2, TICAM1, C3, VAV1, CD209, CLEC4M, MAP2K7, PIN1, PRKCSH, ECSIT, PRKACA, PGLYRP2, BST2, JAK3, PIK3R2, UBA52, TYROBP, NFKBIB, AXL, GSK3A, PGLYRP1, CALM3, IRF3, AKT1S1, SIGLEC14, PRKCG, LILRA5, SSC5D, NLRP4, MYO1F, MAVS, DEFB127, TRIB3, SIRPB1, POLR3F, DEFB118, BCL2L1, HCK, BPIFB3, BPIFA1, ITCH, SAMHD1, SRC, LBP, PLCG1, YWHAB, ELMO2, UBE2V1, NFATC2, ZBP1, APP, MX1, ITGB2, ADARB1, S100B, MAPK1, IGLL5, MIF, APOL1, POLR2F, PLA2G6, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H, PDGFB, TAB1, GRAP2, TNRC6B, EP300, POLR3H, XRCC6, MAPK11

A.1.6 GO.Adapt genes

PIK3CD, BCL10, CTSS, TNFSF4, DUSP10, ZAP70, TXK, ITK, IL6ST, IRF1, TRIM27, PIK3CG, RIPK2, IFNB1, CTSL, SYK, SIRT1, RAG1, JAM3, NLRP10, FADD, IRF7, IFNG, RIPK3, NEDD4, CTSH, PYCARD, ALOX15, TNFRSF11A, TGFB1, SAMSN1

A.1.7 GO.APP genes

CTSS, CD1A, CD1E, MR1, RAB4A, PSMB2, PSMA5, FCGR1B, FCGR1A, PSMD4, PSMB4, FCER1G, NCF2, KIF2C, KIFAP3, CTSE, ARF1, CD1D, CD1C, CD1B, CD8A, PSME4, RPS27A, CD207, PSMD14, ITGAV, PSMD1, KIF3C, DYNC2LI1, DCTN1, ACTR1B, DYNC1I2, AP1S3, SLC11A1, PSMD6, ITGB5, SEC61A1, PSMD2, SEC13, KIF15, RAB7A, AP2M1, SEC31A, CENPE, SEC24B, SEC24D, AP3B1, CD74, ERAP1, ERAP2, LNPEP, KIF2A, KIF3A, SAR1B, SEC24A, DCTN4, KIF4B, CANX, HFE, HLA-F, HLA-G, HLA-A, HLA-E, HLA-C, HLA-B, MICA, MICB, HLA-DRA, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DQA2, HLA-DQB2, HLA-DOB, TAP2, PSMB8, PSMB9, HLA-DMB, HLA-DMA, HLA-DOA, HLA-DPA1, HLA-DPB1, RAET1E, RAET1G, ULBP2, ULBP1, RAET1L, ULBP3, TAP1, TAPBP, PSMB1, AZGP1, PSMA2, SEC61G, NCF1, CD36, PSMC2, DYNC1I1, AP1S1, NOD1, DCTN6, CTSL, SEC61B, PSMD5, PSMB7, SH3GL2, DCTN3, CLTA, CTSV, CCL19, CCL21, SEC61A2, SEC24C, KIF11, ACTR1A, CCDC67, PSMD13, PSMA1, PSMC3, AP2A2, CTSD, KIF18A, TRAF6, KLC2, CTSF, SPTBN2, DYNC2H1, IFNG, PSMD9, UBC, RACGAP1, DCTN2, KIF5A, DYNLL1, TPP2, PSME1, PSMB5, PSMB11, PSME2, PSMA6, PSMC6, PSMA3, PSMC1, SEC23A, LGMN, DYNC1H1, KLC1, KIF26A, CTSH, PDIA3, B2M, PSMA4, KIF23, THBS1, PSMB10, PSMD7, CYBA, DCTN5, KIF22, DYNC1LI2, AP1G1, PYCARD, NOD2, PSMB6, UBB, PSMD11, PSMB3, PSMD3, PSME3, PSMC5, PSMD12, RILP, AP2B1, KIF2B, DYNLL2, CLTC, BLMH, NPEPPS, CCR7, PSMA8, OSBPL1A, AP3D1, CD209, CLEC4M, RELB, FCGRT, CALR, IFI30, UBA52, PSMD8, PSMC4, AP1M2, DNM2, AP1M1, AP2S1, AP2A1, LILRB2, ICAM1, PROCR, PSMF1, PSMA7, KIF3B, NCF4, AP1B1

A.2 Genes from the Host-Pathogen Interaction Database

A.2.1 HP.Bact

SEPT2, SEPT7, A1CF, AAMP, AATF, ABCA2, ABCA7, ABHD3, ABHD5, ABI1, ABI3, ACADM, ACAP1, ACBD6, ACKR3, ACLY, ACSF2, ACSL1, ACSL5, ACSM1, ACTB,

ADAT2, ADCK3, ADCY4, ADH1B, ADK, ADNP, ADNP2, ADPGK, AEBP1, AES, AFF1, AGFG2, AHDC1, AHR, AIP, AKAP6, AKAP9, AKNA, ALDH2, ALDOA, ALDOB, ALG3, ALG9, AMBP, AMPD3, AMY2B, ANK2, ANO6, ANXA1, ANXA2, ANXA5, ANXA6, ANXA7, AP1B1, AP1G1, AP1G2, AP2A2, AP2B1, AP3B1, AP3D1, AP3M1, APBA1, APBA2, APLP2, APOA1, APOBR, APOE, APOL3, AQP3, ARAP2, AREG, AREL1, ARF1, ARF6, ARID2, ARL1, ARL4A, ARPC2, ARPC3, ARPC4, ARRB1, ARRB2, ARSA, AS3MT, ASB3, ASCC2, ASCC3, ASH1L, ASH2L, ASPG, ASTE1, ASXL1, ASXL2, ATF2, ATF6B, ATG10, ATG13, ATG2A, ATM, ATN1, ATR, ATRN, AUP1, AUTS2, AXIN1, AZI2, BACH2, BAG1, BAG2, BAG6, BANK1, BAP1, BCL2, BCL6, BDP1, BECN1, BFAR, BICD2, BIN2, BIRC2, BIRC6, BLK, BLVRB, BMS1, BNIP2, BOC, BRD2, BRD7, BRD8, BRI3, BRK1, BRPF1, BRPF3, BRWD1, BSCL2, BSDC1, BST2, BTAF1, BTBD1, BTBD2, BTBD3, BTBD9, BTF3, BUB3, BZW1, C1QA, C1QB, C1S, C4BPA, C5AR2, CAB39, CALD1, CALR, CAMLG, CAP1, CAPS2, CAPZB, CARD8, CASP1, CASP4, CASP8, CASP9, CAV1, CBLC, CBWD5, CBX3, CBX5, CBY1, CCAR1, CCAR2, CCDC9, CCL14, CCL21, CCL4, CCL5, CCM2L, CCNC, CCNF, CCNH, CCNI, CCNL1, CCNT2, CCPG1, CCR1, CCSAP, CCZ1B, CD1C, CD22, CD320, CD34, CD36, CD4, CD44, CD48, CD59, CD5L, CD6, CD63, CD68, CD69, CD7, CD79B, CD81, CD83, CD8A, CD97, CDC16, CDC23, CDC27, CDC37, CDC42, CDC5L, CDC73, CDIP1, CDK12, CDK14, CDK2, CDK7, CDK9, CDKL1, CDPF1, CEBPG, CELF2, CENPH, CENPJ, CEP44, CEP57, CEP63, CEP85, CEP89, CEPT1, CERS2, CFLAR, CHD1L, CHD4, CHD6, CHD8, CHD9, CHP1, CHRD, CHTOP, CIAO1, CIB1, CIC, CINP, CIR1, CIRBP, CISD1, CIZ1, CKAP5, CLCC1, CLCN3, CLIC1, CLIP1, CLIP2, CLK1, CLN5, CLOCK, CLPB, CLPX, CLU, CMC1, CNN1, CNN3, CNOT1, CNOT2, CNOT3, CNOT8, CNPY2, CNPY3, CNR2, CNST, CNTRL, COA6, COG2, COG3, COG4, COG5, COG8, COPA, COPB2, COPE, COPG1, COPZ1, COX20, CPEB4, CPED1, CPNE8, CPSF2, CPSF6, CPT1A, CPVL, CR1, CR2, CRBN, CREB1, CREG1, CRIM1, CRIPT, CRLF3, CRP, CRYAB, CSDE1, CSF1, CSF1R, CSTF3, CTCF, CTDP1, CTGF, CTIF, CUL1, CUL2, CUL5, CUL9, CUTA, CUTC, CUX1, CWC15, CWC22, CXCR2, CXCR4, CXCR5, CYLD, CYR61, CYTIP, CYYR1, DAAM1, DAB2, DAD1, DAG1, DAXX, DBR1, DCAF6, DCTN2, DCTN3, DCTN5, DDAH1, DDIT4, DDX17, DDX23, DDX24, DDX28, DDX41, DDX42, DDX46, DDX47, DDX5, DDX50, DDX56, DDX6, DDX60, DEK, DESI1, DFFA, DGKA, DGKD, DGKZ, DHCR7, DHRS3, DHRS7, DHX15, DHX32, DIDO1, DLG2, DLG5, DMPK, DMTF1, DNMT1, DOCK5, DOCK8, DOK3, DPEP2, DPM1, DPP4, DPY30, DPYD, DRAM2, DRG2, DSCR4, DTNA, DTNB, DUT, DVL2, DVL3, DXO, DYM, DYSF, E2F2, E2F4, E2F5, EAF2, EAPP, ECD, ECE1, ECH1, ECI2, ECM1, EDC4, EDEM1, EDF1, EDRF1, EEA1, EED, EGFL7, EGFR, EGLN1, EGR1, EHBP1, EHD3, EHD4, EHMT1, EID1, EIF1, EIF2A, EIF3A, EIF3B,

EIF3C, EIF3D, EIF3G, EIF3M, ELF1, ELF2, ELL, ELMO1, EMC2, EMC4, EMC6, EMID1, ENAH, ENKD1, EOMES, EP300, EPAS1, EPHB6, EPN1, EPS15, EPS8, ERAP1, ERAP2, ERCC3, ERI3, ERP29, ERP44, ESR1, ESYT1, ETFA, ETFB, ETHE1, ETS1, ETS2, ETV6, EVA1C, EVI2A, EVI5, EXOC1, EXOC2, EXOC3, EXOC4, EXOC7, EXT1, EXT2, FADD, FAF1, FAIM3, FAM3C, FANCF, FARP2, FBLN1, FBLN2, FBN1, FBRS, FBXL3, FBXL4, FBXL5, FBXL7, FCAR, FCGBP, FCHO1, FCHO2, FCRLA, FEM1B, FES, FEZ2, FFAR2, FGF1, FGR, FHL2, FHL5, FIZ1, FKBP2, FKBP4, FKBP5, FKBP8, FLI1, FLII, FLOT2, FMO2, FMOD, FNBP1, FNTA, FOPNL, FOS, FOSB, FOXJ2, FOXK1, FOXK2, FOXP1, FOXP4, FPR1, FRMD3, FRY, FRYL, FSD1L, FSTL1, FUBP1, FUS, FUT8, FXR1, FXYD6, FYB, FYCO1, FYN, G2E3, G3BP2, GALE, GAS6, GAS7, GATA2, GATA6, GATM, GBA2, GBF1, GBGT1, GBP1, GBP2, GBP3, GCC1, GCC2, GDE1, GDPD5, GET4, GGA1, GGA2, GGT1, GGT6, GHITM, GIPC1, GIPC2, GIT1, GIT2, GLI3, GLTP, GLYR1, GMFG, GMPR2, GNA11, GNA12, GNAI3, GNPTG, GNS, GOLM1, GOPC, GP1BB, GPER1, GPN1, GPS2, GPX3, GRAP, GRASP, GRB2, GRN, GSDMD, GSTA1, GSTM1, GSTP1, GTF2I, HAT1, HAUS3, HAUS6, HAX1, HBB, HBS1L, HCAR3, HCFC2, HCLS1, HCST, HDAC1, HDAC2, HDAC3, HDAC4, HDAC9, HEBP1, HELB, HELLS, HELQ, HELZ, HERC1, HES1, HFE, HGS, HHEX, HHLA2, HID1, HIF1A, HINT1, HIPK2, HIPK3, HLF, HMCN1, HMGB1, HMGB2, HMGCL, HMGN2, HMGN3, HMGN4, HMHA1, HMOX1, HMOX2, HNF1B, HOOK2, HPS1, HRG, HSDL1, HSF1, HSF2, HSPB1, HSPB6, HSPB8, HTRA1, HTRA2, HTRA3, HYAL2, IBTK, ICAM2, ICAM3, ID3, IDH3A, IER2, IFI27, IFI44, IFIT3, IFT20, IFT57, IFT80, IGF1R, IGJ, IGSF6, IKZF1, IKZF3, IL16, IL1R2, IL2RB, IL32, ILF3, ILK, IMP3, IMPA2, INADL, INF2, ING4, ING5, INSR, IPO13, IPO4, IPO5, IPO8, IQCB1, IREB2, IRF2, IRF4, IRF8, IRF9, ISCA1, ISOC1, IST1, ITCH, ITFG2, ITM2B, ITPA, ITPR1, ITPR2, ITPR3, ITSN1, ITSN2, IVD, IWS1, JADE1, JADE2, JAG1, JAG2, JAK1, JMJD4, JPH4, JTB, JUN, JUNB, KANK3, KAT6B, KCNRG, KDM2A, KDM6B, KDM7A, KEAP1, KIF17, KIF1B, KIF1C, KLC2, KLF10, KLF11, KLF15, KLF6, KLHL2, KLHL5, KLRB1, KMT2B, KMT2C, KMT2D, KMT2E, KRCC1, KRI1, KRIT1, KSR1, KYNU, LACE1, LAG3, LAMA5, LAMP1, LARP1, LARP7, LCK, LCMT1, LCP2, LDB1, LDB2, LDHA, LDHB, LENG8, LHX4, LIMA1, LIMD1, LIMD2, LIME1, LIMK2, LIPE, LIX1L, LMAN1, LMNA, LMNB1, LMNB2, LMO4, LPAR6, LPIN2, LPXN, LRBA, LRCH1, LRCH4, LRIF1, LRIG1, LRIG2, LRMP, LRP1, LRP10, LRP4, LRP5, LRRK2, LSM1, LTBP4, LUC7L, LUZP1, LY75, LY9, LYN, LYST, MACF1, MADD, MAF, MAGI1, MAGI2, MAML2, MAML3, MAP4, MARK2, MARK3, MATR3, MBD5, MBNL1, MBNL2, MCFD2, MCM9, MCRS1, MDFIC, MDM2, MDM4, MDN1, MED21, MED23, MED24, MED28, MED4, MEF2D, MEGF8, MEIS1, MEIS2, MEOX1, MERTK, MET, MEX3C, MFAP4, MFN1, MFSD6, MGA, MGLL, MGMT, MIA3, MIB2, MIIP, MKL1, MKNK2,

MKRN1, MLF2, MLH1, MLXIP, MMP19, MMP2, MMP9, MMRN1, MMS19, MNDA, MOAP1, MOB1B, MON2, MPDZ, MPEG1, MRC1, MS4A7, MSH6, MSL1, MSL2, MT1G, MTBP, MTMR2, MTMR3, MTO1, MTOR, MTSS1, MTUS1, MUC1, MUC4, MVP, MX1, MYCT1, MYD88, MYH10, MYH14, MYH9, MYL6, MYL9, MYLK, MYO1D, MYO1G, MYO5A, MYO9B, MYOZ2, MYPOP, N4BP2, NAA25, NAB1, NACA, NADK, NAGA, NAGK, NAMPT, NANP, NARF, NASP, NAT8, NBAS, NBEA, NBPF1, NBPF8, NBR1, NCALD, NCAM1, NCBP2, NCF2, NCF4, NCK1, NCOA1, NCOA2, NCOA3, NCOA4, NCOA5, NCOA6, NCOA7, NCOR1, NCOR2, NDE1, NDEL1, NDRG1, NDRG3, NEK7, NELFE, NELL2, NEO1, NETO2, NEXN, NFAT5, NFE2, NFKB1, NFYC, NGRN, NID1, NINJ1, NIPBL, NIT2, NKD2, NKTR, NLRP1, NMI, NNAT, NOA1, NOB1, NOC2L, NOD1, NOL7, NOL8, NOS3, NOXA1, NPAT, NPC2, NPDC1, NPHP3, NQO2, NR1D1, NR1H2, NR1H3, NR1I3, NR2C1, NRCAM, NRF1, NRG1, NRIP1, NRP1, NSA2, NSD1, NSUN2, NUDT3, NUDT5, NUMBL, NUP62, NUP98, NUPL2, NXF1, NXNL2, OBSCN, ODF2L, OLFM4, ORAI2, ORC3, ORC4, OS9, OSCAR, OSER1, OSMR, OSTF1, OTUB1, OTUD1, OTUD4, P2RX4, P2RX5, PA2G4, PAF1, PAIP1, PAK1, PALLD, PALMD, PAN2, PAPD5, PARP1, PARP4, PARVA, PATL1, PAX5, PBX2, PCBP1, PCBP2, PCBP3, PCCA, PCF11, PCGF5, PCID2, PCM1, PCSK6, PDCD4, PDCD5, PDCD6, PDE2A, PDE4D, PDE6D, PDE7A, PDE8A, PDGFB, PDIA3, PDIA5, PDIA6, PDK4, PDP1, PDS5B, PEAK1, PEAR1, PEG10, PENK, PEPD, PER3, PEX14, PEX2, PEX26, PEX5, PGAM1, PGM1, PGS1, PHC1, PHC3, PHF1, PHF11, PHF12, PHF14, PHIP, PHKG2, PHTF2, PI4KA, PIAS3, PIBF1, PIGH, PIGQ, PILRA, PILRB, PINK1, PJA2, PKN1, PLAG1, PLCG2, PLD1, PLD3, PLEC, PLIN2, PLIN3, PLK2, PLRG1, PLS1, PLVAP, PMP22, PNKP, PNMA1, PNRC1, POGZ, POLH, POLK, POMP, POMT1, PPCS, PPIB, PPIG, PPM1A, PPM1G, PPRC1, PPT1, PPWD1, PRDM1, PRDM2, PRDM4, PRDX1, PRDX3, PRDX6, PRELP, PREX1, PRKRA, PROSC, PRR12, PRR22, PRR3, PRRC1, PSD3, PSMD1, PSMD2, PSMD3, PSMD4, PSMD7, PSMD9, PSME1, PSME2, PTAR1, PTBP1, PTBP3, PTEN, PTGDS, PTGR1, PTH1R, PTK7, PTN, PTPRA, PTPRB, PTPRC, PTPRH, PTPRM, PUF60, PUM1, PUM2, PWP1, PXDC1, QKI, QPCT, RAB10, RAB18, RAB1A, RAB1B, RAB34, RAB37, RAB5A, RAB5C, RAB6C, RAB7A, RAB8A, RABL3, RAC1, RAC2, RAD21, RAD50, RAF1, RAI1, RAI14, RAP1B, RARB, RASA1, RBBP6, RBL2, RBM12, RBM22, RBM25, RBM27, RBM33, RBM38, RBM39, RBM4, RBM45, RBM47, RBM6, RBP2, RBPMS, RCN3, RDH11, RECK, RER1, RERG, REXO4, RFTN1, RFWD2, RFWD3, RFX5, RGL3, RGP1, RGPD1, RGS19, RGS3, RHOA, RHOC, RHPN1, RIC8A, RIF1, RILP, RIN1, RIN3, RING1, RINT1, RITA1, RND3, RNF11, RNF31, RNF4, RNF41, RNF44, RNF8, RNH1, ROBO2, ROGDI, RPAP3, RPE, RPN1, RRAGC, RRAS, RREB1, RRS1, RSAD1, RSF1, RTEL1, RTF1, RTKN, RTN1, RTN2, RTN3, RTN4, RTTN, RUFY2, RUFY3, RUNX3,

RYR2, RYR3, SAAL1, SAE1, SAMD8, SAMD9, SAP18, SARNP, SART3, SATB1, SAV1, SBDS, SBNO1, SBNO2, SCIMP, SCLT1, SCLY, SCOC, SCRIB, SCYL2, SDC1, SDC2, SDHA, SDK2, SDPR, SEC13, SEC63, SENP1, SENP6, SEPP1, SERF2, SETD2, SF3A1, SF3B1, SF3B2, SFI1, SFPQ, SFRP4, SGK3, SGTA, SIDT1, SIKE1, SIL1, SIX5, SKI, SKP1, SLFN5, SLIRP, SLIT2, SLK, SLMAP, SLTM, SLU7, SMAD4, SMAP2, SMC3, SMC4, SMC6, SMG1, SMG7, SMG9, SMOC1, SMTN, SNAI3, SND1, SNIP1, SNRK, SNX1, SNX2, SNX25, SNX29, SNX33, SNX5, SNX6, SNX7, SNX9, SOGA1, SON, SOS1, SOS2, SP1, SP110, SP140, SP3, SPAG5, SPAST, SPG20, SPG21, SPHAR, SPHK2, SPI1, SPIB, SPIN1, SPOPL, SRCAP, SREK1, SRF, SRGN, SRP14, SRP19, SRP54, SRP68, SRRM2, SRRT, SRSF1, SRSF2, SRSF3, SRSF5, SRSF6, SSBP2, SSFA2, SSUH2, ST5, STAB1, STAG1, STAG3, STAM2, STAT1, STAT2, STAT3, STAT4, STAT6, STAU1, STAU2, STIP1, STK11, STK16, STMN1, STON1, STRN, STT3A, STT3B, STX16, STX17, STX3, STX8, SUGP2, SULF2, SUMF1, SUMO1, SUMO2, SUN1, SUN2, SUSD1, SUSD2, SVEP1, SYCP2, SYNC, SYNE1, SYNJ1, SYNPO, SYPL1, SYTL1, SYTL3, SYVN1, SZRD1, TAB1, TAB2, TACC1, TADA3, TAF11, TAF4B, TAF6, TAF7, TAF9, TANK, TAOK2, TAOK3, TAP1, TBCB, TBK1, TBRG1, TBX2, TBX6, TCAIM, TCEA1, TCF21, TCHP, TCL1A, TEF, TES, TET2, TET3, TEX2, TFEB, TFG, TFPT, TGFB1, TGS1, THADA, THAP5, THAP8, THOC3, THRB, THSD1, THSD4, THTPA, TIA1, TIAM1, TINF2, TIPIN, TIPRL, TIRAP, TJAP1, TKT, TLE1, TLE3, TLE4, TLK1, TLR2, TLR4, TLX1, TMCC1, TMCC3, TMCO1, TMED2, TMEM2, TMF1, TMTC2, TMUB2, TMX1, TMX3, TNIP1, TNPO1, TOX4, TPP1, TRA2A, TRAF3, TRAF5, TRAF6, TRAF7, TRAK2, TRAM1, TRAP1, TRIM3, TRIM7, TRIM8, TRIP4, TRIP6, TRNT1, TRPC1, TRPS1, TSC1, TSC2, TSNAX, TTC1, TTC19, TTC24, TTC3, TTC7A, TTC9C, TTLL3, TTYH3, TWF2, TXLNA, TXNIP, TXNL1, TYW1, U2AF1, UBC, UBE2K, UBE2N, UBXN4, UCK1, UCKL1, UCP2, UFC1, UFL1, UIMC1, ULK1, UPP1, URM1, USB1, USF1, USF2, USMG5, USO1, USPL1, UTP20, UTP6, UXS1, VAC14, VAMP2, VAPA, VAPB, VASP, VDAC2, VEGFA, VEGFB, VHL, VMAC, VMP1, VPS11, VPS16, VPS18, VPS25, VPS29, VPS35, VPS39, VPS52, VPS53, VTA1, VWCE, VWF, WAC, WBP11, WDFY4, WDR1, WDR33, WDR41, WDR43, WDR59, WDR6, WDR72, WDR88, WIPF1, WIPF2, WIPI2, WISP2, WIZ, WLS, WNK1, WNK2, WRN, WSB1, WWC2, WWP1, WWTR1, XAF1, XPC, XPO1, XPO5, XRCC3, XRCC5, XRCC6, XRN1, XRN2, YAF2, YIF1A, YIPF2, ZAP70, ZBED4, ZBED5, ZBTB1, ZBTB5, ZC3H3, ZEB1, ZEB2, ZFHX3, ZFR, ZHX1, ZHX2, ZHX3, ZMIZ1, ZMYM2, ZMYM4, ZNF24, ZNF26, ZNF3, ZNF84, ZNF91, ZNFX1, ZNRF2, ZW10, ZYX, ZZZ3

A.2.2 HP.Virus

SEPT1, SEPT8, SEPT9, SEPT10, AAAS, AAK1, AATF, ABCA1, ABCB6, ABCD3, ABCE1, ABCF1, ABCF2, ABHD5, ABL1, ABR, ABT1, ACACA, ACAD9, ACADM, ACAP2, ACBD5, ACKR3, ACLY, ACOT8, ACSL1, ACSL3, ACTB, ACTN1, ACTN2, ACTN4, ACY1, ACY3, ADAM9, ADCK2, ADCK3, ADCK4, ADCY6, ADCY9, ADNP, ADRM1, AEBP1, AES, AFF1, AFF4, AGAP3, AGFG1, AGGF1, AGK, AGO1, AGO2, AGR2, AHR, AHSA1, AICDA, AIDA, AIMP1, AIMP2, AIP, AKAP1, AKAP8, AKAP9, AKIP1, AKT1, ALDH2, ALDOA, ALDOB, ALG1, ALG5, ALMS1, AMBP, AMFR, AMPD2, AMPH, ANK2, ANK3, ANO10, ANO6, ANO8, ANXA1, ANXA2, ANXA3, ANXA5, ANXA6, ANXA7, AP1B1, AP1G1, AP1M1, AP1M2, AP2A1, AP2A2, AP2B1, AP2M1, AP2S1, AP3B1, AP3D1, AP3M1, AP4M1, AP5S1, APEX1, APH1A, API5, APLP2, APMAP, APOA1, APOA2, APOA5, APOB, APOE, APOH, APOL3, ARAP1, AREG, ARF1, ARF3, ARF4, ARF5, ARID2, ARL1, ARL16, ARL2, ARL4D, ARL8A, ARL8B, ARMC4, ARMC5, ARMC6, ARMC8, ARMC9, ARNT, ARPC4, ARRB2, ARV1, ASAP1, ASCC2, ASF1B, ASPH, ASPM, ASXL1, ATAD1, ATE1, ATF1, ATF2, ATF3, ATF4, ATF5, ATF6B, ATF7, ATG12, ATG4C, ATG5, ATG9A, ATM, ATN1, ATP5E, ATP5H, ATP5I, ATP5J, ATP5L, ATP9A, ATR, ATRIP, ATRN, AUP1, AURKA, AURKB, AXIN1, AXIN2, AZI2, BACE2, BACH1, BAD, BAG1, BAG2, BAG3, BAG4, BAG5, BAG6, BANK1, BANP, BAP1, BASP1, BATF3, BAX, BAZ1A, BAZ1B, BBC3, BBX, BCAM, BCAR1, BCAT2, BCL10, BCL2, BCL6, BCR, BECN1, BEND4, BEND5, BEND7, BET1, BICC1, BICD1, BICD2, BID, BIK, BIN1, BIN3, BIRC6, BIRC7, BLCAP, BLM, BLNK, BMPER, BMS1, BNIP2, BNIP3, BOLA1, BOLA2, BPHL, BPTF, BRAT1, BRCA1, BRD2, BRD3, BRD4, BRD8, BRMS1, BROX, BSCL2, BSPRY, BST2, BTAF1, BTBD1, BTBD2, BTF3, BUB1, BUB3, BUD13, BZW1, BZW2, C1D, C1QA, C1QBP, C2CD5, CALD1, CALR, CALU, CAMLG, CAND1, CAND2, CANT1, CAP1, CAP2, CAPS2, CAPZB, CARD9, CARM1, CASC3, CASC5, CASP1, CASP3, CASP4, CASP6, CASP7, CASP8, CAV1, CAV2, CAV3, CBL, CBR1, CBS, CBWD1, CBWD6, CBX1, CBX3, CBX4, CBX5, CBY1, CCAR2, CCDC8, CCL1, CCL17, CCL2, CCL21, CCL25, CCL26, CCL28, CCNA1, CCNA2, CCNB1, CCNB2, CCNC, CCND1, CCND2, CCND3, CCNE1, CCNE2, CCNG1, CCNG2, CCNH, CCNL2, CCNO, CCNT1, CCNT2, CCNY, CCPG1, CCR5, CD14, CD209, CD248, CD320, CD4, CD44, CD47, CD59, CD5L, CD63, CD68, CD80, CD81, CD82, CD86, CD97, CDC16, CDC20, CDC23, CDC27, CDC37, CDC42, CDC45, CDC5L, CDC6, CDC73, CDCA2, CDCA7, CDCP1, CDIPT, CDK1, CDK13, CDK19, CDK2, CDK3, CDK4, CDK5, CDK6, CDK7, CDK8, CDK9, CDR2, CDR2L, CDS2, CEBPA, CEBPB, CEBPG, CELF1, CENPA, CENPB, CENPC, CENPH, CENPJ, CENPU, CEP44, CEP55, CEP57, CEP63, CEP68, CEP70, CEP76, CEP85, CEP89, CEPT1, CERS1, CERS2, CERS4,

CERS6, CGNL1, CHD1, CHD3, CHD4, CHD5, CHD6, CHD8, CHERP, CHMP6, CHP1, CHPF2, CHPT1, CHTOP, CIAO1, CIB1, CIC, CIDEB, CINP, CIR1, CIRBP, CIZ1, CKAP4, CKAP5, CLCC1, CLCN7, CLGN, CLIC1, CLIC4, CLIP1, CLIP4, CLK1, CLK2, CLN6, CLPB, CLPP, CLPX, CLU, CMC1, CMC2, CNBP, CNIH4, CNN1, CNN2, CNN3, CNOT1, CNOT2, CNOT3, CNOT6, CNOT7, CNOT8, CNPY2, CNTLN, CNTRL, COA3, COA7, COG1, COG2, COG3, COG4, COG5, COG6, COG7, COG8, COIL, COMT, COPA, COPB2, COPE, COPG1, COPG2, COPZ1, COQ5, COX15, COX17, COX20, COX5A, COX5B, COX6C, COX7C, COX8A, CPNE3, CPSF1, CPSF2, CPSF3, CPSF4, CPSF6, CPSF7, CPT1A, CPT2, CPVL, CRADD, CRBN, CREB1, CREB3, CREM, CRIM1, CRIP2, CRK, CRKL, CRTC1, CRTC2, CRTC3, CRX, CRYAB, CSDE1, CSF1R, CSK, CSN2, CSN3, CSPG4, CSPP1, CSRP1, CSRP2, CSTF1, CSTF3, CTBP1, CTBP2, CTDP1, CTGF, CTR9, CTRL, CUBN, CUL1, CUL2, CUL3, CUL4A, CUL5, CUL7, CUL9, CUTA, CWC15, CXCL3, CXCR4, CXXC1, CYB5B, CYLD, CYR61, DAAM2, DAB2, DAD1, DAG1, DAXX, DCAF6, DCAF7, DCAF8, DCD, DCP1A, DCTN1, DCTN2, DDA1, DDAH2, DDB1, DDB2, DDIT3, DDR2, DDX1, DDX11, DDX17, DDX18, DDX20, DDX21, DDX23, DDX24, DDX25, DDX28, DDX4, DDX42, DDX46, DDX47, DDX49, DDX5, DDX54, DDX55, DDX56, DDX58, DDX6, DEAF1, DEDD2, DEK, DERL1, DERL2, DEXI, DFFA, DGAT1, DHCR7, DHRS7, DHX15, DHX29, DHX30, DHX34, DHX36, DHX37, DHX58, DHX8, DHX9, DIP2A, DIP2B, DIRC2, DKK3, DLG1, DLG4, DLG5, DLK2, DLX2, DMAP1, DMBT1, DMRT3, DNAL1, DNM1L, DNMT1, DOCK2, DOCK3, DOCK5, DOCK7, DOCK8, DOCK9, DOHH, DOK1, DPF1, DPM1, DPM3, DPPA2, DPY30, DRG1, DSC3, DSCR3, DSG1, DTD1, DTX3, DTX3L, DUOX1, DUS1L, DUT, DVL1, DVL2, DVL3, DYM, DZIP3, E2F1, E2F2, E2F3, E2F4, E2F5, E2F6, E2F7, E4F1, EAF1, ECD, ECE1, ECI2, EDC3, EDC4, EDEM3, EDF1, EED, EFHD2, EFNB2, EFNB3, EGFL7, EGFR, EGLN1, EGR1, EGR2, EGR3, EHD2, EHMT1, EHMT2, EI24, EID1, EIF1, EIF2A, EIF3A, EIF3B, EIF3C, EIF3D, EIF3E, EIF3F, EIF3G, EIF3H, EIF3I, EIF3K, EIF3L, EIF3M, ELF1, ELF2, ELF3, ELMO1, ELMO2, ELMO3, ELOF1, ELP4, EMC1, EMC10, EMC2, EMC7, EMC9, EME1, EMP2, ENAH, ENKD1, ENPP1, ENY2, EP300, EP400, EPHA2, EPHB2, EPHB4, EPN1, EPN2, ERBB2, ERBB3, ERCC1, ERCC2, ERCC3, ERCC6, ERG, ERH, ERI1, ERP29, ERP44, ESRP1, ESYT1, ESYT2, ETAA1, ETFA, ETFB, ETS1, ETV4, EVA1A, EVA1B, EVL, EXD2, EXOC1, EXOC2, EXOC4, EXOC5, EXOC6, EXOC7, EXOC8, EXPH5, EXTL3, EZH2, FABP5, FADD, FADS1, FADS2, FAF2, FAHD1, FAM3C, FANCA, FANCG, FANCI, FARP1, FAS, FAT1, FBF1, FBLN1, FBLN2, FBLN5, FBN1, FBN2, FBN3, FBXL2, FBXW7, FCN1, FCN2, FES, FEZ1, FGF12, FGFR4, FGL1, FHL2, FHL3, FIBP, FIS1, FITM2, FKBP2, FKBP4, FKBP5, FKBP6, FKBP7, FKBP8, FKBP9, FLCN, FLII, FLNB, FLNC, FLOT1, FLOT2, FMNL1, FNBP4, FNDC1, FNIP1, FOS,

FOSL2, FOXC1, FOXF1, FOXK1, FOXK2, FOXM1, FOXP1, FRG1, FRMD1, FRS3, FRYL, FUBP1, FUBP3, FURIN, FUS, FUT8, FXR1, FXR2, FXYD5, FXYD6, FYB, FYCO1, FYN, FZD2, FZD6, FZD7, G2E3, G3BP1, G3BP2, GAB1, GAK, GALE, GALK1, GANAB, GAS6, GATA3, GATC, GATM, GBF1, GBP2, GBP6, GCC1, GCDH, GCSH, GDAP1, GDPD1, GEM, GET4, GFAP, GFPT1, GFPT2, GGA1, GGA2, GGA3, GGT7, GID4, GID8, GIN1, GIPC1, GKAP1, GLI2, GLI4, GLMN, GLRB, GLRX3, GLRX5, GLTP, GLYAT, GLYR1, GMCL1, GMDS, GMEB1, GNAI2, GNAI3, GNAT3, GNAZ, GNL3, GNPAT, GNPTG, GNS, GON4L, GOPC, GOSR1, GPAA1, GPC1, GPD1L, GPM6A, GPN1, GPN3, GPR1, GPR27, GPR62, GPR87, GPS2, GPX1, GPX2, GPX4, GPX8, GRB2, GREB1, GREM1, GRHPR, GRN, GRSF1, GRWD1, GSK3A, GSK3B, GSTA4, GSTK1, GSTM1, GSTM2, GSTO1, GSTO2, GSTP1, GTF2I, GUF1, GYS1, HAT1, HAUS4, HAUS5, HAX1, HBB, HBD, HBEGF, HCAR1, HCFC2, HCK, HDAC1, HDAC2, HDAC3, HDAC4, HDAC5, HDAC7, HDAC9, HDDC2, HECW1, HELLS, HELQ, HELZ, HELZ2, HERC2, HERC5, HGS, HIC2, HIF1A, HINT3, HIPK2, HIRA, HKDC1, HLTF, HM13, HMGA1, HMGA2, HMGB1, HMGB2, HMGN2, HMGN3, HMOX2, HNF4A, HOMEZ, HOOK1, HPS6, HRG, HSBP1, HSDL1, HSDL2, HSPB1, HSPB2, HSPB6, HSPB8, HTRA2, HYOU1, ICAM1, ICAM2, ICAM3, ICMT, ID1, IDE, IDH3A, IDH3B, IFI27, IFIH1, IFIT1, IFNA1, IFNG, IFRD2, IFT81, IFT88, IGJ, IGLL1, IKZF1, IKZF3, IL17B, IL18, IL1B, IL23R, IL32, IL8, ILF2, ILF3, ILVBL, IMPA1, IMPA2, INCA1, INF2, INSR, IP6K2, IPO11, IPO4, IPO5, IPO7, IPO8, IPO9, IPPK, IQCB1, IRAK2, IRF1, IRF3, IRF6, IRF7, IRS1, ISCA2, ISCU, ISG15, ISLR, ISLR2, ISOC2, ISY1, ITCH, ITFG2, ITIH3, ITM2B, ITM2C, ITPK1, ITSN1, ITSN2, IWS1, JAG1, JAG2, JAK1, JAK2, JUN, JUNB, JUND, KANK2, KAT2A, KAT2B, KAT5, KAT7, KCMF1, KCNH3, KCNH6, KCNK3, KCNRG, KCTD1, KCTD2, KCTD5, KCTD6, KCTD9, KDM1A, KDM3A, KDM4B, KDM4D, KDM5A, KDSR, KEAP1, KHNYN, KIF11, KIF12, KIF14, KIF17, KIF1B, KIF1C, KIF22, KIF2C, KIF3B, KIF3C, KIF5A, KIF6, KIF7, KIF9, KIFC3, KLC1, KLC2, KLC3, KLC4, KLF3, KLHL7, KLHL9, KLK5, KLK6, KMT2A, KNG1, KNOP1, KNTC1, KPRP, KRCC1, KRR1, KRT35, KRT36, KRT81, KRT83, KRT86, KTN1, LACTB, LAD1, LAMA1, LAMA2, LAMA3, LAMA5, LAMB1, LAMB2, LAMB3, LAMC1, LAMC2, LAMC3, LAMP1, LARP1, LARP4, LARP7, LASP1, LAYN, LBR, LCE2B, LCK, LCN1, LDB1, LDHA, LDHB, LDLR, LEMD2, LENG1, LENG8, LEO1, LETM1, LGMN, LIFR, LIMA1, LIMD1, LIMS2, LMAN1, LMAN2, LMBR1, LMF2, LMNA, LMNB1, LMNB2, LMO3, LMO4, LMO7, LMOD1, LNX1, LNX2, LOXL1, LPAR1, LPP, LPXN, LRIG1, LRIG2, LRP1, LRP10, LRP12, LRP5, LRRC7, LSM12, LSM3, LSM4, LSR, LTBP3, LTBP4, LTN1, LTV1, LUC7L, LUZP1, LY6D, LYAR, LYN, LYNX1, LYST, LZTR1, LZTS2, MACC1, MACF1, MAEA, MAF, MAFB, MAFG, MAG, MAGI1, MAML1, MANF, MAP1B, MAP1S,

MAP4, MAP7, MAPK3, MARCO, MARK2, MARK3, MAST1, MAST2, MAST4, MAT2B, MATN2, MATR3, MAVS, MAX, MBD2, MBD3, MBTD1, MCFD2, MCL1, MCM2, MCM3, MCM4, MCM5, MCM6, MCM7, MCM8, MCPH1, MCRS1, MDC1, MDFI, MDFIC, MDM2, MDM4, MDN1, MECR, MED1, MED11, MED13, MED15, MED16, MED17, MED20, MED23, MED24, MED25, MED26, MED27, MED28, MED29, MED4, MED6, MED7, MED8, MED9, MEGF6, MEGF8, MEIS2, MEN1, MEOX2, MEPCE, MFAP1, MFN1, MFN2, MFSD1, MFSD3, MFSD5, MGAT1, MGME1, MGP, MGST1, MGST2, MGST3, MIA3, MIB2, MIF, MINA, MISP, MITD1, MKLN1, MKNK1, MKRN1, MKRN2, MKRN3, MKS1, MLEC, MLF1, MLF2, MLH1, MLTK, MLX, MLXIP, MMAB, MMP1, MMP14, MMRN2, MMS19, MNDA, MOB1A, MOB1B, MOB2, MOGS, MON2, MORN3, MOV10, MPI, MPP7, MPRIP, MPZL1, MRC2, MROH1, MROH7, MRS2, MSH2, MSH3, MSH6, MSRB1, MSTO1, MSX2, MT1F, MTA1, MTA2, MTA3, MTAP, MTCH1, MTCH2, MTOR, MTPN, MTUS2, MTX1, MTX2, MUC5B, MUS81, MVP, MX1, MXRA7, MXRA8, MYB, MYC, MYCN, MYD88, MYEF2, MYH10, MYH11, MYH14, MYH2, MYH9, MYL6, MYL6B, MYLK, MYLK2, MYNN, MYO1B, MYO1C, MYO5A, MYO5C, MYO6, MYO9A, MYO9B, MYOF, MYOM1, MYOZ1, MYSM1, MZT2B, N4BP1, NAA16, NACA, NAGK, NASP, NAT10, NAT14, NAV1, NBN, NBPF8, NCAN, NCBP1, NCBP2, NCDN, NCEH1, NCK1, NCLN, NCOA1, NCOA2, NCOA3, NCOA5, NCOA7, NCOR1, NCOR2, NDC80, NDE1, NDEL1, NDRG1, NDRG2, NDST1, NEDD4, NEDD8, NEK1, NEK5, NEK7, NELFB, NELL1, NELL2, NEMF, NFAT5, NFE2, NFIA, NFIC, NFIL3, NFKB1, NFKB2, NFU1, NFX1, NGLY1, NHSL1, NID1, NID2, NIN, NINJ1, NINL, NIPBL, NISCH, NKAPL, NLRX1, NMB, NMI, NMT1, NMT2, NMU, NOA1, NOC2L, NOC4L, NOL12, NOL7, NOL9, NOLC1, NOM1, NOMO1, NOMO2, NOMO3, NOP14, NOP16, NOP2, NOP56, NOP58, NOP9, NOS3, NOXA1, NPC1, NPC2, NPHP1, NPHP3, NPL, NPM3, NPS, NPTN, NQO2, NR1H4, NR2C2, NR4A1, NRF1, NSA2, NSD1, NSF, NSUN2, NUBP2, NUCB1, NUCB2, NUDC, NUDT5, NUMA1, NUMBL, NUP50, NUP54, NUP62, NUP85, NUP88, NUP93, NUP98, NUPL2, NUTM1, NVL, NXF1, NXT1, OAS1, OAT, OAZ1, OAZ2, OBSCN, OLA1, OLIG3, OPA1, OPTN, ORC1, ORC2, ORC3, ORC4, ORC5, OS9, OSBP2, OSGEP, OSMR, OSTC, OSTF1, OSTM1, OTX1, OTX2, OXA1L, OXSM, P2RX1, P2RX5, P4HA1, P4HA2, P4HA3, P4HTM, PA2G4, PACS1, PAF1, PAIP1, PAIP2, PAK1, PAK2, PAK4, PALLD, PALM, PALMD, PANK4, PAPD5, PARK7, PARP1, PARP2, PARP4, PARP9, PARVG, PATZ1, PAWR, PCBP1, PCBP2, PCBP3, PCCA, PCCB, PCDH1, PCF11, PCGF6, PCID2, PCLO, PCM1, PCNA, PCNT, PCSK5, PCSK6, PCSK9, PDCD2, PDCD6, PDE3B, PDIA2, PDIA3, PDIA4, PDIA5, PDIA6, PDK2, PDPK1, PDPN, PDS5A, PDS5B, PEBP1, PEF1, PEG10, PEG3, PELO, PELP1, PEPD, PEX14, PEX19, PEX3, PEX5, PGAM1, PGAM2, PGAM5, PHAX, PHB, PHB2, PHC2, PHC3, PHF14,

PHF5A, PHTF1, PI4KA, PI4KB, PIAS1, PIAS2, PIAS3, PIAS4, PICK1, PIGG, PIGO, PIGQ, PIGS, PIGT, PIGU, PILRA, PIN1, PIP, PITX1, PJA2, PKD2, PKN1, PKN2, PKP4, PLAC8, PLCB3, PLCD3, PLD3, PLEC, PLEK, PLIN3, PLK1, PLK2, PLK3, PLOD1, PLOD2, PLOD3, PLRG1, PLS1, PLTP, PML, PMM2, PMS1, PMVK, PNKD, PNKP, PNMA1, PNO1, PNRC1, POGZ, POMP, PON2, PON3, POP1, POP5, POP7, POTEE, POTEF, POTEI, POTEJ, PPARA, PPARG, PPIA, PPIB, PPIC, PPIE, PPIG, PPIH, PPIL3, PPIL4, PPM1A, PPM1B, PPM1G, PPME1, PPRC1, PPT1, PQLC3, PRC1, PRDX1, PRDX2, PRDX3, PRDX5, PRDX6, PREB, PRKDC, PRKRA, PROM2, PRPF3, PRR15, PRR3, PRRC1, PRTN3, PSIP1, PSMD1, PSMD2, PSMD3, PSMD4, PSMD5, PSMD6, PSMD7, PSMD8, PSMD9, PSME1, PSME2, PSME3, PSME4, PSMF1, PSMG1, PSPC1, PTBP1, PTBP2, PTBP3, PTCD1, PTCD3, PTGDS, PTGES, PTH2, PTK7, PTMA, PTOV1, PTPRF, PTPRK, PTRF, PUF60, PUM1, PUM2, PURA, PURB, PUS7, PVR, PWP1, PYGB, PYGL, PYGM, PYGO2, QKI, QPCTL, QSOX1, QSOX2, RAB10, RAB14, RAB15, RAB18, RAB1A, RAB1B, RAB21, RAB24, RAB25, RAB2A, RAB32, RAB34, RAB3D, RAB4A, RAB5A, RAB5B, RAB5C, RAB6A, RAB7A, RAB8A, RABL3, RAC1, RAC2, RAD18, RAD50, RAD51, RAF1, RAG2, RAI14, RALA, RALY, RAN, RAP1A, RARA, RARB, RASA2, RASD1, RBAK, RBBP4, RBBP6, RBL1, RBL2, RBM14, RBM15, RBM19, RBM22, RBM23, RBM25, RBM27, RBM39, RBM4, RBM42, RBM4B, RBM5, RBM6, RBM8A, RBMS1, RBMS2, RBP2, RBPMS, RBX1, RCAN1, RCAN3, RCC1, RCC2, RCN1, RCN2, RCN3, RCOR1, RDH11, RDH13, RDH14, REEP2, REEP5, REL, RELL2, RER1, REST, REV1, REXO4, RFC1, RFC2, RFC3, RFC4, RFC5, RFT1, RFTN1, RFWD2, RFX4, RFX6, RGL2, RGP1, RGS14, RHOA, RHOC, RHOG, RHOJ, RIBC2, RIC8A, RIF1, RIMS1, RIMS2, RIMS3, RIN1, RING1, RINT1, RIOK1, RIPK1, RIPK3, RIPK4, RLF, RNF10, RNF11, RNF31, RNF4, RNF41, RNF5, RNF6, RNF8, RNFT1, RNPS1, ROBO1, ROCK1, ROCK2, ROGDI, ROMO1, RPA1, RPAP2, RPAP3, RPF2, RPN1, RPN2, RPP30, RPP40, RRAGA, RRAS, RRAS2, RRBP1, RRP1, RRP12, RRP7A, RRP8, RRS1, RSBN1, RSF1, RSPH3, RSRC2, RSU1, RTCB, RTF1, RTN3, RTN4, RUFY3, RUNX1, RUNX2, RUSC2, RXRA, RXRB, RYBP, S1PR3, SAAL1, SAE1, SAFB2, SAP18, SAR1A, SARNP, SART3, SATB1, SBSN, SCAI, SCFD1, SCLT1, SCMH1, SCN9A, SCNM1, SCO1, SCO2, SCOC, SCRIB, SCYL2, SDC2, SDC4, SDF2, SDHA, SDHB, SDK1, SDPR, SEC13, SEC62, SEC63, SELK, SENP1, SENP2, SENP3, SEPP1, SERF2, SERP1, SET, SETD2, SETD5, SETX, SF3A1, SF3A2, SF3A3, SF3B1, SF3B2, SF3B3, SF3B4, SF3B5, SFI1, SFPQ, SFRP1, SFRP4, SFXN1, SFXN3, SGCB, SGCD, SGPL1, SGTA, SGTB, SHC1, SHOC2, SIAH1, SIAH2, SIDT2, SIK1, SIK3, SIKE1, SIL1, SIN3A, SKP1, SKP2, SLBP, SLIRP, SLIT1, SLIT2, SLIT3, SLK, SLMAP, SLPI, SLU7, SLX4, SMAD1, SMAD2, SMAD3, SMAD4, SMAD7, SMC2, SMC3, SMC4, SMC5, SMC6, SMCO1, SMG7, SMOC1, SMU1,

SMUG1, SMYD3, SND1, SNF8, SNIP1, SNRPA, SNTB1, SNTB2, SNW1, SNX14, SNX17, SNX18, SNX19, SNX27, SNX3, SNX4, SNX5, SNX7, SNX9, SOAT1, SOCS1, SOGA1, SON, SOX4, SOX5, SP1, SP100, SP110, SPAG5, SPC24, SPCS1, SPCS2, SPCS3, SPEM1, SPERT, SPG21, SPG7, SPI1, SPIB, SPIN1, SPNS1, SPOP, SPRR4, SPTA1, SRBD1, SRC, SRCAP, SREK1, SRP14, SRP19, SRP54, SRP68, SRP72, SRPK1, SRPK2, SRPRB, SRR, SRRM1, SRRM2, SRRT, SRSF1, SRSF2, SRSF3, SRSF4, SRSF5, SRSF6, SRSF7, SRSF9, SSBP2, SSFA2, SSNA1, SSRP1, ST14, STAB1, STAM2, STAT1, STAT2, STAT3, STAU1, STAU2, STIM1, STIP1, STK11, STK24, STK38, STMN1, STOM, STRAP, STRBP, STRN, STRN3, STRN4, STT3A, STT3B, STX10, STX12, STX17, STX18, STX4, STX5, STX7, STX8, STYX, SUGP2, SULF2, SUMF2, SUMO1, SUMO2, SUMO3, SUN1, SUN2, SURF4, SURF6, SUZ12, SV2A, SVEP1, SVIL, SVIP, SYBU, SYK, SYNC, SYNE1, SYNE2, SYNM, SYPL1, SYT11, SYT16, SYT6, SYTL1, SYVN1, TAAR1, TAB1, TAB2, TACC1, TACC3, TADA3, TAF11, TAF12, TAF1A, TAF1B, TAF2, TAF4, TAF5L, TAF6, TAF6L, TAF7, TAF9, TANK, TAOK1, TAOK3, TAP1, TAP2, TBCB, TBCD, TBK1, TBL2, TBL3, TBP, TBRG4, TCF21, TCF25, TCFL5, TCHP, TCTA, TEAD1, TEAD2, TEAD4, TEC, TECR, TEFM, TEKT4, TELO2, TERF1, TERF2, TERT, TES, TEX10, TEX2, TFAM, TFAP4, TFB1M, TFB2M, TFCP2, TFDP1, TFDP2, TFG, TGFA, TGM1, TGM2, THADA, THAP1, THEM6, THOC1, THOC3, THOC5, THOC6, THRB, THRSP, THY1, TIA1, TIPIN, TIRAP, TLE2, TLN1, TLN2, TLR2, TLR4, TMA7, TMCO1, TMCO3, TMCO4, TMED1, TMED2, TMED3, TMED4, TMED7, TMED9, TMF1, TMOD3, TMPPE, TMX1, TMX2, TMX3, TNIP1, TNIP2, TNKS2, TNNI1, TNNI2, TNNI3, TNNT3, TNPO1, TNPO2, TNPO3, TOB1, TOB2, TOE1, TOP1, TOP2A, TOP2B, TOR1A, TOR3A, TOR4A, TOX4, TPD52, TPM1, TPM2, TPM3, TPM4, TPP1, TPP2, TPR, TPRA1, TPRKB, TRA2A, TRA2B, TRADD, TRAF1, TRAF2, TRAF3, TRAF4, TRAF5, TRAF6, TRAIP, TRAM1, TRAP1, TRHDE, TRIM5, TRIO, TRIP6, TRPA1, TRPM4, TRPM7, TRPS1, TRPT1, TRPV2, TRPV4, TRRAP, TRUB1, TRUB2, TSC1, TSC2, TSN, TSR1, TSSC1, TSTD2, TTC1, TTC12, TTC13, TTC16, TTC23, TTC27, TTC38, TTC4, TTC7B, TTF1, TTI1, TTI2, TTLL7, TULP3, TUSC1, TUSC2, TXNIP, TYK2, U2AF1, U2AF2, UACA, UAP1, UBA3, UBAC1, UBAC2, UBAP1, UBB, UBC, UBD, UBE2C, UBE2F, UBE2K, UBE2O, UBE2S, UBE2T, UBE2W, UBE3A, UBE3C, UBE4A, UBN1, UBN2, UBR2, UBR4, UBR5, UBXN1, UCHL5, UCK2, UCKL1, UCP3, UFL1, UGDH, UIMC1, UNG, UPP1, USF1, USF2, USH1C, USMG5, USO1, USPL1, UTP15, UTP6, UTS2, UVRAG, UVSSA, UXS1, VAC14, VAMP3, VAMP5, VAPA, VAPB, VASP, VAT1, VAV2, VDAC1, VDAC2, VDAC3, VDR, VEGFB, VGF, VHL, VMP1, VPRBP, VPS11, VPS18, VPS28, VPS29, VPS35, VPS39, VPS4A, VPS51, VPS52, VPS53, VPS54, VPS72, VRK1, VRK2, VRK3, VTI1A, VWA8, VWA9, VWF, WASF1, WASF3, WASL, WBP11, WDFY3, WDR11, WDR18, WDR26, WDR3, WDR33, WDR36,

WDR5, WDR6, WDR60, WDR61, WFS1, WIPF2, WLS, WNK1, WNK2, WNT5A, WRN, WWC2, WWOX, WWP1, WWP2, XCL1, XKR7, XPA, XPC, XPO1, XPO4, XPO5, XPO6, XPO7, XPOT, XRCC1, XRCC4, XRCC5, XRCC6, XRN2, YAF2, YAP1, YDJC, YIF1B, YIPF3, YIPF4, YKT6, YLPM1, YPEL5, ZAP70, ZBED4, ZBP1, ZBTB1, ZC3H8, ZDBF2, ZER1, ZFP14, ZFP42, ZFP91, ZFPL1, ZFR, ZG16, ZG16B, ZGPAT, ZHX2, ZHX3, ZMAT2, ZMAT3, ZMAT4, ZMIZ1, ZMYM2, ZMYM6, ZNF44, ZNF74, ZNF83, ZNF84, ZW10, ZWINT, ZXDC, ZYX, ZZEF1, ZZZ3

A.2.3 HP.Prot

ACTB, ATG12, MBP, UBL5

A.2.4 HP.Amoe

CDC42, MBP, PASK, RAC1, RAN

Appendix B

Full lists of genes in population and immune gene category combinations that are significantly enriched for immune genes

B.1 Positive selection

B.1.1 nSL enrichments

	Pop.	Genes
GO.Bact	ENE	BAIAP2L1, LCN2
	WSI	ASIC3, BCL3, FGR
	COL	CHIT1, DEFA1, DEFA1B, DEFA3, DEFA5, IL22RA1, IL27, SERPINE1, TEKT2
	SEA	ADH7, CEBPE, ERAP1, PLAC8, PPM1D, SELP, TLR4
GO.Virus	ENE	CCNT2, CEBPA, CXCR4, DUOX2, ELMO1, FDPS, HSPA1A, HSPA1B, IFI35, LCN2, RPL27, RXRA
	COL	ABCF3, AP1S1, AP2M1, BAD, CD86, CHMP3, DEFA1, DEFA1B, DEFA3, DYNLT1, HLA-A, HMGA1, IFNLR1, IL27, PPIE, PSMA3, RPS10, TRIM56, UNC93B1, XPO1
	SEM	CCT5, CD86, DDIT4, ENO1, EPHA2, KARS, RANBP2, SUPT6H, TAF4B, TRIM23
GO.Tcell	CSI	FKBP1B, IL1B, IL7, PRKDC, SRC, TNFSF4, TNFSF18, ZNHIT1
	NSI	IL15, NDFIP1, PDE4D, SMAD3, WNT5A
	SEA	ADAM17, HLA-A, HLA-F, HLA-G, PAG1, SIRPG, UBE2N
GO.Bcell	WAA	HDAC9, IL4, LAT2, PRKCH
	SWE	HIF1A, IL4, PRKCH
	WSI	BCL3, HDAC9, LYL1, PAWR
	CSI	IL7, PRKDC, TNFSF4, TPD52, TSHR
	COL	BAD, CD86, CLCF1, HIF1A, PRKCH
	SEA	ADAM17, TLR4, TPD52

Table B.1 Genes from each significantly enriched immune category and population combination from the nSL test, part 1

	Pop.	Genes
GO.Innate	WSI	ARPC1A, ARPC3, DDX58, FGFR4, FGR, LGR4, NCKIPSD, PDPK1, PRKAR2A, TREX1, WASF2
	SSI	ARPC1A, ARPC2, CAMK4, DEFA1, DEFA1B, DEFA3, DEFA5, ERBB4, NCF1, NRG1, RNF125, TAB1, TICAM2
	CSI	AKIRIN2, ATF1, BPIFA1, DDX41, DMBT1, F12, FGFR4, IRAK2, PRKDC, SRC
	COL	ADRBK1, AGO1, AGO3, AGO4, BAD, CD86, DEFA1, DEFA1B, DEFA3, DEFA5, FCN3, FGF4, FGF19, IL27, KITLG, MAP4K2, MAPK8, RPS6KB2, TRIM56, UNC93B1, WASF2
	SEM	ACTB, AGER, APOL1, ARPC2, C4B, CD86, ERBB4, NRG2, PRKD1, UBE2D2
GO.APP	WAA	CTSE, DCTN2, DCTN6, KIF3A, KIF5A, PSMA8
	SWE	CTSE, DCTN1, KIF3A
	SSI	AP2M1, DYNLL2, KIF3C, KIF22, NCF1
	COL	AP1S1, AP2M1, HLA-A, OSBPL1A, PSMA3, SEC24D
	SEA	BLMH, DYNC2LI1, ERAP1, HLA-A, HLA-F, HLA-G, PSMB5, PSMB11, PSMC1, SEC61G
GO.Adapt	SOA	ZAP70
	SSI	TNFSF4
	CSI	TNFSF4
HP.Bact	AFR	ASH2L, BSDC1, CCNF, CPVL, DDIT4, DPY30, DTNA, EMC6, FBXL7, GTF2I, LSM1, NCOA2, NDRG3, P2RX5, PHF14, RAB6C, RBM27, RGS3, RNF11, TACC1
	SWE	ADNP, APBA2, APLP2, ATF6B, BAG6, CCNT2, CLIC1, CUL9, CXCR4, DPM1, DUT, DXO, EMC4, FBN1, HIF1A, NELFE, PBX2, PPIB, PTK7, PTPRM, SCOC, SNX1, SPG21, SRF, TRIP4, UBXN4, URM1
	ENE	APBA2, ASH1L, AUP1, BAG6, BRI3, CCNT2, CIZ1, CLIC1, CXCR4, DUT, EDEM1, ELMO1, FBN1, HTRA2, KAT6B, NBR1, PAPD5, PTPRM, RTKN, SCOC, SFPQ, SLTM, UBXN4, WDR88, ZMYM4
HP.Virus	SWE	ADNP, APLP2, ASPH, ATF6B, BAG6, BEND4, CCNT2, CLGN, CLIC1, CUL7, CUL9, CXCR4, DCTN1, DDAH2, DNM1L, DPM1, DUT, EHMT2, EIF3H, FBN1, GPD1L, HIF1A, ISLR2, ISLR, KLC4, MCM6, MOB1A, MYEF2, NGLY1, OXSM, PLAC8, POTEE, PPIB, PTK7, RNF5, RXRA, SCOC, SPG21, STX12, THAP1, TRUB2, UBN2, WWOX
	WSI	AGAP3, ATRIP, BCAM, BCAT2, BEND4, BMS1, CCL26, CDK5, CNOT1, DDX58, DOCK2, FGFR4, GMEB1, GPN3, HDAC9, IP6K2, KCTD5, LAMB2, LMAN2, NOS3, NSD1, ORC1, P4HTM, PAWR, PDPK1, POGZ, PVR, RAB24, RAB34, RGS14, SDK1, SLU7, SNX27, SREK1, STX10, TMA7, TMX3, TRAF4, UIMC1, UPP1, VPS29, WDR6, WLS
	SSI	ADCK4, AHSA1, AP2M1, ATF4, BBX, BEND4, BEND7, CDIPT, COX8A, DVL3, EMC10, EXOC6, FABP5, FZD2, GTF2I, HINT3, ITM2C, KDM3A, KIF3C, KIF22, LAMA2, LTBP4, MARK2, MKS1, MVP, MYH14, MYOF, NCOA7, NINJ1, NUMBL, PLD3, PMS1, PNKD, RBX1, REEP5, RSPH3, SDK1, SPIB, SRP19, SRSF1, TAB1, TMED7, WNK2
HP.Prot	SEM	ACTB

Table B.2 Genes from each significantly enriched immune category and population combination from the nSL test, part 2

B.1.2 iHS enrichments

	Pop.	Genes
GO.Bact	ENE	BAIAP2L1, HIST2H2BE, PLAC8, TNFRSF14
GO.Virus	SWE	CCNT2, CFLAR, CPSF4, DDX5, DUOX2, DYNLL1, FDPS, HSPA1A, HSPA1B, PSMA8, PSMB2, RANBP2, RPLP1, RXRA, TNFRSF14
	ENE	CCNT2, CFLAR, DUOX2, DYNLL1, EIF4G3, ELMO1, HSPA1A, HSPA1B, IDE, NELFA, NUP37, RPL6, TNFRSF14, TRIM23
	VOL	CCNT2, CFLAR, CHMP3, EIF4H, FBXW7, HNRNPK, HSPA1A, HSPA1B, IL27, LARP1B, MNAT1, NELFA, POLR2F, PSMC2, RCC1, RPL23A, SND1, SRPK2, SUPT6H, TAF12, VPS4A, VTA1
	SOA	ACKR3, BTRC, CD247, CFLAR, DYNLL1, HSPA1A, HSPA1B, PSMB2, RPL4, TAF4B, TNFRSF14, TNPO1, XPR1
	WSI	AP1G1, CRTC2, DOCK2, LARP1B, MNAT1, PDCD6IP, RBM15B, RPL23A, RPS27, SRPK2, SUPT6H, SYNCRIP, VPRBP, XRCC4
	COL	ANKRD17, BAD, DAG1, DYNLT1, FAU, HMGA1, IL27, PACS1, PILRA, RBX1, RCC1, RHOA, RNGTT, RPL4, RPL7A, RPL23A, RPLP0, RPS10, RPS25, SUPT6H, SYNCRIP, TAF12, XPO1
	SEM	CFLAR, DAG1, DDIT4, EPHA2, GFI1, HMGA1, HSPA1A, HSPA1B, NUP37, RANBP2, RHOA, RPL5, RPL6, RPS10, SUPT6H, TAF4B, TNPO1, TRIM23, WWP2, XRCC4, ZNF571
GO.Tcell	SSI	ADAM17, IL2, LAT, PIK3CB, PTPN22, TNFRSF14, TNFSF4, TNFSF18, ZBTB7B, ZEB1
	NSI	CBLB, FKBP1B, IL27, PDE4D, SMAD3, WNT5A, ZEB1
	SEM	ADAM17, FOXN1, HLA-DQA1, HLA-DRA, HLA-DRB1, HLA-DRB5, PAG1, PTPN11, XRCC4, ZEB1
GO.Bcell	WAA	GON4L, IL4, MZB1, SHB
	SWE	FNIP1, HES5, HIF1A, IL2, IL4, PRKCH
	VOL	CLCF1, FZD9, HIF1A, IL2, KLHL6, LAT2, PRKCH, PRKDC
	SOA	CLCF1, HES5, HIF1A, PRKCH
	SSI	ADAM17, CD19, HES5, IL2, PTPN22, TNFSF4
	COL	BAD, CD19, CLCF1, MZB1, PRKDC
	SEA	BANK1, IL2, SHB, TPD52
GO.Innate	VOL	ADRBK1, CAPZA1, CTSK, CTSS, DHX9, IL27, LAT2, POLR2F, PRKDC, RPS6KB2, SRPK2, TRAFD1
	COL	ADRBK1, BAD, CD19, FGF3, FGF4, FGF19, IL27, LAT, MAP2K1, MAP4K2, MAPK8, PIK3CB, PRKDC, RPS6KB2, TYRO3
GO.APP	WAA	KIF3A, PSMA8, PSMB2
	SWE	DCTN1, DYNLL1, KIF3A, KIF23, PSMA8, PSMB2
	SSI	DYNC1I2, NCF1, PSMD14, SPTBN2
GO.Adapt	WAA	TRIM27
	CSI	TGFB1
	SEA	SIRT1

Table B.3 Genes from each significantly enriched immune category and population combination from the iHS test, part 1

	Pop.	Genes
HP.Bact	AFR	ASH2L, CCNF, CFLAR, COPB2, CPVL, CRLF3, EIF3M, FBXL7, GTF2I, IFT80, ITCH, JTB, LSM1, MACF1, MLXIP, MTBP, PHF14, PRR3, RAP1B, RBP2, RNF11, SMC4, TIPIN, TMCC1, TTC1, TTC3
	SWE	APBA2, ASH1L, AUP1, BAG6, BRI3, CCNT2, CFLAR, CLIC1, DDX5, DLG2, ERP29, G2E3, HIF1A, HTRA2, MYO5A, NAA25, PIBF1, PPIB, PTPRM, PUM2, RBM27, RTKN, SDC1, SF3A1, SNX1, THSD4, TRIP4, UBXN4, URM1, ZEB1, ZMYM4, ZZZ3
	ENE	ALDH2, AUP1, BAG6, BRI3, CCNT2, CFLAR, CLIC1, DUT, EEA1, ELMO1, ERP29, EXOC4, FBN1, HTRA2, KIF17, KRIT1, LRBA, MED21, NAA25, NBR1, PEAK1, PIBF1, PPWD1, PTH1R, PTPRM, RAB5A, RGS3, RTKN, SEC63, SF3A1, SFPQ, TMCC1, UBXN4, ZMYM4, ZZZ3
	VOL	APBA2, APOBR, BAG6, BFAR, CCNT2, CDPF1, CFLAR, CLIC1, CLIP2, COG8, DUT, ENAH, ERP29, FBN1, HIF1A, HPS1, KDM2A, KMT2E, MED21, NAA25, NANP, NBR1, NOL8, NSD1, OLFM4, PCGF5, RAB34, RBM4, RNF11, SND1, THSD4, TMCC1, TXNL1, UBXN4, VTA1, ZNF91
	SOA	ACKR3, ADNP, AIP, ANXA1, BAG6, CDPF1, CFLAR, CLCC1, CLIC1, DPM1, EPHB6, FCHO2, FUS, FYB, GSTP1, HIF1A, KDM2A, MACF1, MOB1B, MTMR3, MTOR, NELL2, NOL8, RGS3, RITA1, RUFY3, TAF4B, TIPIN, TNPO1, VPS39
	SSI	ASH2L, BAG6, BFAR, CEP63, CFLAR, CHD9, CLIC1, COPG1, DHRS7, GCC2, GTF2I, HAT1, HERC1, HRG, LSM1, MARK2, MCRS1, MYH14, NINJ1, NR1H2, NSA2, OTUB1, PPM1A, RBM4, RBM45, RTN1, SPIB, STAU1, STAU2, TANK, TMED2, WNK2, ZEB1
	CSI	ASTE1, BICD2, CFLAR, DDX56, ERI3, FBXL3, FUT8, GATM, GDE1, GGA2, HAT1, KMT2C, MYLK, NACA, NCOA7, NCOR1, NOL8, ORC3, PAPD5, RBM45, SENP6, STAU1, TGFB1, TIPIN, TMCC1, TNPO1, UFL1, UPP1, ZEB1, ZNF91, ZW10
	NSI	AIP, APOBR, CAB39, CAMLG, CDK12, CFLAR, CHD9, COG8, CPED1, CPT1A, DDX46, EGLN1, FAM3C, FCHO2, FEM1B, FUT8, GATA2, GSTP1, HERC1, HMOX1, ISCA1, LRP5, NIPBL, PATL1, PDE4D, PDS5B, PGAM1, PIBF1, PTPRA, RBM4, RPN1, SLMAP, STX3, TAOK3, TNPO1, TSNAX, VPS16, XPC, ZEB1, ZZZ3
	COL	APOBR, AZI2, CCNL1, CINP, CMC1, DAG1, EAF2, FKBP2, GSTA1, HCLS1, HMGB2, IQCB1, KDM2A, KLC2, KYNU, MATR3, MDN1, MED21, MGA, NUDT3, PAIP1, PILRA, PILRB, PTPRA, RAB1B, RAB34, REXO4, RHOA, RIN1, SART3, SIL1, STAG3, SUMO2, SYTL3, SYVN1, TIPIN, TMCC1, VEGFB, VPS11, VPS16, XPO1, YIF1A
	SEA	ADH1B, ASCC3, ATR, BANK1, BLVRB, CFLAR, COG5, CYYR1, FBXL7, FCHO2, HELLS, PATL1, PHIP, PIBF1, PLD3, RSF1, RTN3, SGK3, SRP19, SSBP2, STX3, TAOK3, TNPO1, XRN1

Table B.4 Genes from each significantly enriched immune category and population combination from the iHS test, part 2

	Pop.	Genes
HP.Virus	WAA	AGO1, AMFR, CCPG1, CLK2, DDX5, DDX54, DOCK3, DRG1, DUT, ECD, EGR1, FBN1, GON4L, MANF, MAST2, MATR3, MSH6, MSTO1, MYEF2, MYOZ1, NCDN, NGLY1, ORC1, ORC2, OSTM1, OXSM, P4HA1, PAIP2, PATZ1, PIGG, SEC63, SFPQ, SIL1, TFB1M, TNPO1, UACA, VPRBP, XRN2, ZZZ3
	SWE	ABT1, AUP1, BACH1, BAG6, BEND4, CCNT2, CLIC1, CLK2, COQ5, CPSF4, CSPP1, DCTN1, DDAH2, DDX5, DOCK3, DOK1, EIF3H, ERP29, FNIP1, G2E3, GATC, HIF1A, HTRA2, MCM6, MOB1A, MOGS, MYO5A, NCDN, OAZ2, ORC2, PANK4, PLAC8, PPIB, PTCD1, PUM2, RBM27, RNF10, RXRA, SCFD1, SF3A1, SRSF9, THAP1, TRUB2, UBN2, ZMYM6, ZZZ3
	VOL	ARNT, BAG6, BAZ1B, BEND4, BRCA1, CCNT2, CLIC1, COG8, CSPP1, DDAH2, DHX9, DOCK3, DUT, EIF3L, ENAH, ERP29, FBN1, FBXW7, GKAP1, GMEB1, HIF1A, KDM3A, LAMC1, LAMC2, LBR, LMAN2, MCM4, MCM6, MYEF2, NAA16, NINL, NSD1, OPA1, ORC2, PPARA, PRKDC, RAB24, RAB34, RBM4, RBM4B, RCC1, RFC2, RGS14, RNF11, SCAI, SIK3, SND1, SNTB2, SRPK2, TAF12, TBL2, TERF2, TRAF4, TTC38, UBR5, VPS4A
	SOA	ACKR3, ADNP, AGO1, AIP, ANXA1, BAG6, BEND4, CLCC1, CLIC1, COQ5, DDAH2, DDX54, DOCK3, DPM1, FUS, FYB, GATC, GLRB, GRSF1, GSTP1, HIF1A, MACF1, MOB1B, MTOR, MYEF2, NCDN, NELL2, NUMA1, ORC1, ORC2, PANK4, PDS5A, PIAS1, PPARA, RNF10, RUFY3, SRSF9, STRBP, STX4, TIPIN, TNPO1, TTC38, VPS39, VWA9, ZMYM6
	WSI	AGO1, AP1G1, ATR, CALD1, CDK8, CRTC2, CSPP1, DOCK2, DOCK3, ELP4, IFT81, KRR1, MANF, MDN1, NASP, NMB, OAZ2, PAWR, PDS5B, PLS1, PUM1, RAB34, RTF1, SDK1, SNX14, SRPK2, STAU2, TAOK3, TPM3, TRAF4, TRHDE, UBE2W, VPRBP, XRCC4, YAF2
	SSI	ACTN1, AHSA1, BAG4, BAG6, BBX, BEND4, CEP63, CLIC1, CNBP, COPG1, COX8A, CPSF3, DDAH2, DDX55, DHRS7, DOCK3, EMC10, EXD2, EXOC5, FZD2, GOSR1, GTF2I, HAT1, HRG, IMPA1, ISY1, KCNH3, KDM3A, MARK2, MCRS1, MYH14, NINJ1, NSA2, OAZ2, ORC2, PANK4, PHTF1, PMVK, PPM1A, PYGO2, RBM4, RBM4B, RSBN1, SHC1, SPIB, SPNS1, STAU1, STAU2, STRBP, TANK, TMED2, UBE2W, WNK2
	CSI	AGO1, BICD2, CDK8, COG7, DDX56, DIP2B, DMAP1, DOCK3, DOCK7, EXOC6, FUT8, GATM, GGA2, HAT1, HINT3, IPPK, LARP4, MCM4, MED1, MYLK, MYO6, NACA, NCOA7, NCOR1, ORC2, ORC3, PANK4, PAPD5, PRKDC, RIMS2, SIK3, SMC5, STAU1, TIPIN, TMED4, TMED9, TNPO1, TRIO, TRPM7, UFL1, UPP1, UVRAG, VWA9, ZW10
	NSI	AIP, ALMS1, APMAP, CAMLG, CDK19, CLN6, COG8, CPT1A, DDX46, DOCK3, DZIP3, EGLN1, ELP4, EXOC8, FAM3C, FUT8, GNPAT, GSK3B, GSTP1, ITM2C, LAMC1, LAMC2, LRP5, MANF, MCM5, MED1, NIPBL, ORC2, PDS5B, PGAM1, PIAS1, POTEE, PPIC, PPIL4, RBM4, RBM4B, RBM14, RFC1, RIMS2, RPN1, RRP12, SLMAP, SMAD3, SNTB2, SNX4, TAOK3, TERF2, TFB1M, TNPO1, UBR5, VPRBP, VPS4A, WDR26, WNT5A, XPC, ZZZ3
	COL	ATP5H, AZI2, BAD, BRMS1, CCL28, CCND1, CD248, CDR2L, CINP, CLIC4, CMC1, CPT2, DAG1, DHX29, FKBP2, GMEB1, GSTA4, HMGA1, HMGB2, HYOU1, IFT81, IQCB1, ISCU, KCTD2, KLC2, MATR3, MDN1, MEN1, NMB, OAZ2, PACS1, PAIP1, PAIP2, PILRA, PLCB3, PRDX5, PRKDC, PYGM, RAB1B, RAB34, RALY, RBMS1, RBX1, RCC1, REXO4, RHOA, RIN1, SART3, SIL1, SMC5, SNX14, SPNS1, SUMO2, SURF4, SURF6, SYVN1, TAF12, TCTA, TIPIN, TRAF4, VEGFB, VPS11, VPS51, XPO1, ZFPL1, ZWINT

Table B.5 Genes from each significantly enriched immune category and population combination from the iHS test, part 3

B.1.3 Tajima's D enrichments

	Pop.	Genes
GO.Bact	WAA	FGR, HIST2H2BE, IL23A, PPM1D, TEKT2
	SWE	FGR, IL23A, PPM1D, ROMO1
	VOL	FGR, PPM1D, ROMO1
	SOA	DEFB121, DEFB123, DEFB124, HIST2H2BE, IL23A, NEMF, PPM1D, TEKT2
	CSI	FGR, NCF1, NEMF, ROMO1
	COL	CD36, FGR, GPX1, PPM1D
	SEA	GPX1, NCF1, NEMF, PPM1D
GO.Virus	AFR	ANKRD17, CAMLG, CBX5, FBXW7, HNRNPA1, ITGA5, LARP1B, MAGI3, MNAT1, PSMA1, PSMC2, PSMC5, PSMD6, RTN3, SNRNP200, THOC7, TMEM173
	WAA	ATP6V1H, CAMLG, CFLAR, CPSF4, DUOX2, ERVMER34-1, FBXW7, FGR, HDAC1, IL23A, LCK, NELFA, PARD6A, PSMA8, PSMB2, STAT2, TOP2A
	SWE	CPSF4, DUOX2, ERVW-1, FGR, IL23A, MNAT1, NELFA, PARD6A, PSMA8, PSMB2, RBM15B, STAT2, TOP2A, VPRBP
	ENE	ATP6V1H, CAMLG, FBXW7, FGR, MNAT1, NELFA, PARD6A, PSMB2, PSMC4, RBM15B, RPL6, RPL29, TOP2A, VPRBP
	VOL	BUB1, CAMLG, CPSF4, ERVMER34-1, FBXW7, FGR, MNAT1, NELFA, PARD6A, RB1, TMEM173
	COL	DAG1, FGR, GRB2, MNAT1, NUP85, NUPL2, NXF1, POLR2G, PSMD5, RBX1, RHOA, RTN3, SATB1, STAT3, TNPO1
GO.Tcell	ENE	PAWR, PPP3CB, PRKDC, PTPN11, SLA2, THEMIS2, ZEB1
GO.Bcell	AFR	POU2F2, PRKDC, PTEN, PTPN22, SLA2
	WAA	CLCF1, LCK, PRKDC, TXLNA
	ENE	CLCF1, PAWR, POU2F2, PRKDC, SLA2
GO.Innate	WAA	ADRBK1, AGO1, AGO3, AGO4, FGFR3, FGR, IL23A, LCK, POLR1C, PPP3CB, PRKDC, RPS6KB2, WASF2, WIPF2
	ENE	ADRBK1, ARPC1A, ARPC3, DUSP7, FGFR3, FGR, GSK3A, PPP3CB, PRKDC, PTPN11, RPS6KB2, WASF2, WIPF2
	VOL	ADRBK1, ARPC1A, ARPC3, FGFR3, FGR, PPP3CB, PRKDC, RPS6KB2, TMEM173, UBE2D2, WASF2
	SOA	ADRBK1, AGO1, AGO3, AGO4, ARPC3, BCL2L1, FGFR3, IL23A, MAPK10, MAPKAPK3, NCKIPSD, PRKAR2A, RPS6KB2, WIPF2
	SSI	ARPC1A, ARPC3, BCL2L1, CD247, FGFR3, FGR, GRB2, NCKAP1, PRKDC, WASF2
	CSI	ARPC1A, ARPC3, ERBB4, FGR, GRB2, HSP90AA1, NCF1, NCKAP1, POLR1C, SHC1, WASF2
	SEA	BCL2L1, BRK1, FGFR3, HSP90AA1, NCF1, NCK1, NFATC3, PDPK1, PIK3CB, PRKAR2A
GO.APP	AFR	DCTN1, KIF2A, PSMA1, PSMC2, PSMC5, PSMD6, SEC24A, SPTBN2
	SWE	DCTN1, PSMA8, PSMB2
	COL	CD36, PSMD5, SEC31A
	SEA	CTSE, DYNLL2, NCF1

Table B.6 Genes from each significantly enriched immune category and population combination from the Tajima's D test (positive selection), part 1

	Pop.	Genes
HP.Bact	AFR	AMY2B, AUP1, BFAR, BIRC6, CAMLG, CBX5, CIAO1, CIC, COG5, COPZ1, DDX42, DDX46, EHBP1, ELF1, FBXL3, FNTA, FRYL, HTRA2, LACE1, MSH6, NDRG3, NFE2, PTEN, RBM4, RTKN, RTN3, SPOPL, TAOK3, ZEB2
	WAA	AHDC1, ANXA7, ASH2L, CAMLG, CFLAR, CNPY2, CTCF, CYTIP, DDX46, DLG2, ENKD1, FGR, HDAC1, KDM2A, LCK, LRBA, PAN2, PATL1, POLH, RNF11, RNF41, RTTN, SFI1, SFPQ, SGK3, SSFA2, STAT2, STX3, SYCP2, TJAP1, TMEM2, TXLNA, WIPF2, XPO5, ZMYM4
	SWE	AHDC1, APLP2, AUP1, CNPY2, CTCF, CYTIP, DLG2, ECD, ENKD1, FBXL3, FGR, FNTA, HTRA2, KAT6B, KDM2A, LIPE, LRBA, MEGF8, MOB1B, NBEA, NDRG3, NETO2, PAN2, PATL1, RBM12, RBM39, RNF11, RNF41, RTKN, RTTN, RUFY3, SFPQ, SGK3, SSFA2, STAT2, STX3, SYCP2, SYNC, THSD1, WIPF2, ZMYM4
	ENE	AHDC1, ANXA7, APLP2, ARPC3, AUP1, BSDC1, CAMLG, CIC, CTCF, DDX46, ECD, ENKD1, FCGBP, FGR, HTRA2, KAT6B, KDM2A, LIPE, MEGF8, MOB1B, NDRG3, NETO2, RNF11, RTKN, RTTN, RUFY3, SFPQ, SGK3, SSFA2, SYCP2, VPS29, WIPF2, ZEB1, ZMYM4
	VOL	AHDC1, ARPC3, CAMLG, CDPF1, COPG1, CTCF, CYTIP, DDX46, ENKD1, FAF1, FGR, ITM2B, KDM2A, LPAR6, MOB1B, PATL1, RBM12, RBM39, RITA1, RNF11, RTTN, RUFY3, SGK3, SSFA2, STX3, TMCC1, TMEM2, VPS29, ZEB1
	SOA	ARF6, ARPC3, BSDC1, CDPF1, CNPY2, CYTIP, DPYD, EPS8, EVI5, FNTA, KDM2A, MOB1B, NDRG3, NETO2, PAN2, PATL1, RNF41, RTTN, RUFY3, SGK3, STAT2, STX3, SYCP2, SYNC, TMEM2, VPS29, WDR6, WIPF2
	WSI	ARPC3, BSDC1, CAMLG, CCNH, CDPF1, CNPY2, COPG1, CTCF, DDX46, DTNA, ENKD1, EVI5, FAF1, LIPE, LRP5, MCRS1, MEGF8, MOB1B, MSL2, NBEA, PAN2, PLAG1, POLH, QKI, RASA1, RNF11, RNF41, RTTN, RUFY3, SMC6, SNRK, SRSF5, STAG1, STAT2, TJAP1, TLK1, TMTC2, VPS29, XPO5, ZEB1
	SSI	AHDC1, ARF6, ARPC3, BSDC1, CDPF1, CFLAR, COPG1, CTCF, DPYD, DTNA, ENKD1, EPS8, ERP44, EVI5, FGR, FNTA, GRB2, HMCN1, LIPE, MCRS1, MEGF8, MOB1B, NANP, NBEA, NETO2, PATL1, PPIB, RGS3, RNF11, RTTN, RUFY3, SFI1, SGK3, SNX1, SRSF1, STX3, TRIP4, VPS29
	CSI	AGFG2, AHDC1, ARF6, ARPC3, ASXL2, ATG10, BSDC1, CCNH, CFLAR, CTCF, CYTIP, DTNA, EVI5, FAF1, FGR, FUS, GIT2, GRB2, GTF2I, ITSN2, LRCH4, MSL2, PATL1, PDE4D, POLH, RASA1, RBM12, RBM39, RGS3, RNF11, RTTN, SMC6, SRSF1, STAG1, STX3, TCAIM, TCHP, TJAP1, TMEM2, VPS29, XPO5
	NSI	AGFG2, AHDC1, ARF6, BSDC1, CNPY2, CPT1A, CTCF, DTNA, ELMO1, ENKD1, ERP44, EVI5, FAF1, FGR, GATM, GRB2, HMCN1, ITSN2, LIPE, LRBA, LRCH4, LRP5, MARK2, MEGF8, MSL2, OTUB1, PAN2, RNF11, RNF41, RTF1, RTTN, SRSF1, SRSF5, STAG1, STAT2, TMTC2, TRIP4
	COL	AGFG2, AHDC1, BSCL2, CD36, CLCN3, DAG1, DLG2, EXOC4, FAF1, FGR, GRB2, LRCH4, LRP5, MCM9, MCRS1, MED21, MUC1, NCOA7, NELL2, NUPL2, NXF1, RHOA, RITA1, RNF11, RTN3, SATB1, ST5, STAT3, THADA, TNPO1, TRIP4, TTC9C, WDR6, ZZZ3
	SEM	AGFG2, ARF6, ARPC3, CDPF1, CFLAR, CTCF, DDIT4, DTNA, ENKD1, EVI5, FAF1, FOXP1, GTF2I, LRCH4, MCRS1, MLH1, MSL2, NETO2, PATL1, RGS3, RNF11, RTN3, RTTN, SGK3, SRSF1, STAG1, STX3, TMEM2, TRIP4, VPS29, WDR6
	SEA	ARF6, BRK1, BSDC1, CAV1, CDPF1, COG8, CTCF, CTIF, CYTIP, DDX28, DPEP2, DTNA, EHBP1, ENKD1, EPS8, ERP44, GTF2I, MLH1, NCK1, NETO2, RHOA, RNF11, RTF1, RTN3, RTTN, SCYL2, SRSF1, STAG1, TRIP4, VHL, WDR6, WWC2, ZHX1

Table B.7 Genes from each significantly enriched immune category and population combination from the Tajima's D test (positive selection), part 2

	Pop.	Genes
HP.Virus	SWE	AGO1, APLP2, AUP1, CDC6, CNPY2, CPSF4, CSPP1, DCTN1, DOCK3, DOK1, ECD, ENKD1, GRSF1, HTRA2, LETM1, LPP, MANF, MCM4, MEGF8, MOB1B, MOGS, MYOZ1, NCDN, ORC1, P4HA1, PRKDC, PTCD1, PTPRK, RARA, RBBP4, RBM39, RNF11, RNF41, ROMO1, RUFY3, SFPQ, SSFA2, STAT2, STX12, SYNC, THAP1, TOP2A, TRADD, UVRAG, VPRBP, WIPF2, ZMYM6
	SSI	C2CD5, CNBP, COPG1, COX6C, CSPP1, ENKD1, ERP44, GGA3, GPN3, GRB2, GRSF1, IFT81, ISY1, KCNH3, LETM1, MCM4, MCRS1, MEGF8, MOB1B, NEMF, NUDC, NUP85, OAZ2, ORC2, PPARA, PPIB, PRKDC, PTPRK, RNF11, RUFY3, SFI1, SRSF1, TTC38, VPS29
	NSI	CCNY, CHD1, CNPY2, COX8A, CPSF4, CPT1A, CREM, DHX29, ELMO1, ENKD1, ERP44, EXOC6, GATM, GGA3, GRB2, ITSN2, LAMA2, LAYN, LRP5, MARK2, MEGF8, MEPCE, NEMF, NUDC, NUP85, OAZ2, PCCB, PTCD1, RNF11, RNF41, RTF1, SRSF1, SRSF5, STAT2, STX12, TNKS2, TRADD
	COL	AKIP1, ARMC8, BSCL2, CHD1, DAG1, DDX54, DPM3, EXOC4, GANAB, GGA3, GLRB, GPX1, GRB2, HINT3, IP6K2, KCNH3, KLHL7, LAMA2, LAMB2, LAYN, LRP5, MCRS1, MEPCE, MTX1, NCOA7, NEK1, NELL2, NUP85, NUPL2, NXF1, OAZ2, P4HTM, PIGG, PSMD5, PTPRK, PTRF, RBX1, RHOA, RNF11, RTN3, SATB1, STAT3, STX5, TAF6L, TCTA, THADA, TNPO1, UBXN1, WDR6, ZZZ3
	SEA	APMAP, BBX, BEND4, C2CD5, CAV1, CAV2, COG8, DDX28, ENKD1, ERP44, GPX1, GTF2I, KCTD5, LAMB2, LAYN, LETM1, MLH1, NCK1, NEMF, NGLY1, NUDC, OTX1, OXSM, P4HTM, PDPK1, PPARA, PTPRK, RHOA, RNF11, ROBO1, RTF1, RTN3, SCYL2, SNTB2, SRSF1, TERF2, TFDP2, TTC38, VHL, VPS4A, WDR6, WWC2

Table B.8 Genes from each significantly enriched immune category and population combination from the Tajima's D test (positive selection), part 3

B.1.4 d_i enrichments

	Pop.	Genes
GO.Bact	SWE	DNAH2, IL6, STAT1
	VOL	EPPIN, EPPIN-WFDC6, FCER1G, P2RX7, PLA2G2A, TLR1
	SEM	FZD5, IL6, TNFSF8
GO.Virus	AFR	APOBEC3H, CPSF4, DDX5, DUOX2, FKBP8, GRB2, IFNL1, NUP85, RPL13A
	WSI	ACY3, CXCL9, KCTD5, NEDD4L, NFKBIA, PDCD6IP, RAVER2, UBC, UNC93B1
	SEM	C9orf156, HYAL1, HYAL2, HYAL3, IL6, KARS, LTBR, POLR3H, RPL6, RPLP0, XRCC6, ZNF175
GO.Tcell	WAA	ADAM17, IFNG, LILRB2, PATZ1, PMAIP1
	SWE	CCND3, IL4, IL6, LAG3, LAX1, PVRL2, SELK, SPNS2, TNFSF18
	SEM	FZD5, HLA-DPA1, HLA-DPB1, IL4, IL6, PTPN11, RAB27A, TNFSF8
GO.Bcell	SWE	IL4, IL6, LAX1, SPNS2
	ENE	IL6, LAX1, SPNS2
	NSI	CD180, PLEKHA1, RAG1
GO.Innate	ENE	AGO1, AGO4, EIF2AK2, IRAK2, MAPK7, TLR1, TLR6, TLR10
	SOA	ADCY9, C1QB, CD247, FCN3, FGR, ITGAM, ITPR1, MAP2K2, NLRC4, PIK3CD, TRIM32
	WSI	CORO1A, KLRD1, MAP4K2, MYO10, NFKBIA, UBC, UNC93B1
	SSI	ADCY8, AGO1, AGO3, AGO4, C1QBP, HCK, LILRA5, MAPK3, NRG1, TAB1
	COL	ABI1, ADCY3, ADRBK1, C1RL, C1S, NLRX1, RPS6KB2, SYK, TYRO3
	SEM	IL4, ITPR3, POLR3H, PTPN11, TRAFD1, XRCC6
GO.APP	WAA	AP1S3, IFNG, KIF3C, LILRB2, PSMB2, SEC24C
	ENE	AP1M2, KIF3C, PSMB2, PSMB4
	SSI	IFNG, KIF3C, LILRB2, PSMB2
	NSI	CTSF, HLA-DQB1, HLA-DRA, PSMA7, SPTBN2
	COL	AZGP1, PSME1, SPTBN2, ULBP2
	SEA	HLA-DPA1, HLA-DPB1, HLA-F
GO.Adapt	WAA	IFNG
	VOL	DUSP10
	SOA	PIK3CD
	SSI	IFNG
	NSI	RAG1
	COL	SYK
HP.Bact	AFR	ANXA5, ATG13, DDX5, DGKZ, DLG2, FKBP8, FOXP1, GATA6, GRB2, HMGB1, PAF1, RILP, RNF41, RTTN, SPIN1, TRPS1, TXNL1
HP.Virus	AFR	ALMS1, ANXA5, BACE2, BEND4, CAP2, CLPP, CPSF4, DDX5, EXOC5, FABP5, FKBP8, FOXP1, GGA3, GRB2, HM13, HMGB1, ID1, MYLK2, NUP85, PAF1, PITX1, PMVK, RNF41, ROBO1, RRBP1, SPIN1, TEAD4, TRPS1, TTC38, VAV2

Table B.9 Genes from each significantly enriched immune category and population combination from the d_i test

B.2 Balancing selection

B.2.1 HKA enrichments

	Pop.	Genes
GO.Bact	SOA	DEFB1, DMBT1, PGLYRP4, TLR6
	SSI	DEFB1, DMBT1, PGLYRP4, TLR6
	SEM	DEFB1, DMBT1, P2RX7, PGLYRP4
GO.Virus	AFR	ACY3, HLA-A, HLA-B, HLA-C, NUP54, POLR2A, PSMB9, RPSA, TAP1, TAP2
	ENE	APOBEC3H, HLA-A, HLA-B, HLA-C, OPRK1, PLSCR1, PSMB9, RPS2, SMARCB1, TAP1
	WSI	APOBEC3H, BAX, DMBT1, HLA-A, HLA-B, HLA-C, NUP54, PLSCR1, PSMB9, RPSA, TAP1
	CSI	DMBT1, HLA-A, HLA-B, HLA-C, NUP54, OPRK1, PSMB9, RPS14, RPSA, SMARCB1, TAP1
	NSI	DMBT1, HLA-A, HLA-B, HLA-C, NUP54, OPRK1, PLSCR1, PSMB9, RPS14, TAP1, TNIP1, TRIM22
	SEA	APOBEC3H, DMBT1, HLA-A, HLA-B, HLA-C, IFIT3, NUP54, PLSCR1, RPS14, TAP2
GO.Tcell	AFR	HLA-A, HLA-B, HLA-C, HLA-DQA2, HLA-G, TAP1, TAP2, TESPA1
	WAA	HLA-A, HLA-B, HLA-C, HLA-DOA, HLA-DQA2, HLA-DRA, HLA-F, HLA-G, LGALS8, MICA, TAP1, TAP2, TESPA1, UBASH3A
	SWE	CLC, HLA-A, HLA-B, HLA-C, HLA-DQA2, HLA-DRA, HLA-F, HLA-G, INPP5D, LGALS8, MICA, P2RX7, TAP1, TESPA1
	ENE	BTN3A2, CLC, HLA-A, HLA-B, HLA-C, HLA-DOA, HLA-DQA2, HLA-DRA, HLA-F, HLA-G, INPP5D, LGALS8, MICA, TAP1, TESPA1
	VOL	CLC, HLA-A, HLA-B, HLA-C, HLA-DQA2, HLA-DRA, HLA-F, HLA-G, LGALS8, MICA, TAP1, TESPA1
	SOA	BTN3A2, CLC, HLA-A, HLA-B, HLA-C, HLA-DOA, HLA-DQA2, HLA-DRA, HLA-F, HLA-G, INPP5D, MICA, TAP1, TESPA1
	WSI	BAX, CLC, HLA-A, HLA-B, HLA-C, HLA-DQA2, HLA-DRA, HLA-G, LGALS8, MICA, TAP1, TESPA1
	SSI	HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DQA2, HLA-DRA, HLA-F, HLA-G, INPP5D, LGALS8, MICA, TAP1, TESPA1
	CSI	CLC, HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DQA2, HLA-DQB2, HLA-DRA, HLA-F, HLA-G, LGALS8, MICA, P2RX7, TAP1, TESPA1
	NSI	CLC, HLA-A, HLA-B, HLA-C, HLA-DQA2, HLA-DQB2, HLA-DRA, HLA-F, HLA-G, LGALS8, MICA, SIGLEC1, TAP1, TESPA1
	COL	CLC, HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DQA2, HLA-DQB2, HLA-DRA, HLA-F, HLA-G, LGALS8, TESPA1
	SEM	HLA-A, HLA-B, HLA-C, HLA-DOA, HLA-DPA1, HLA-DQA2, HLA-DRA, HLA-G, LGALS8, MICA, P2RX7, TESPA1
	SEA	HLA-A, HLA-B, HLA-C, HLA-DOA, HLA-DQA2, HLA-DRA, HLA-G, INPP5D, TAP2, TESPA1
GO.Innate	SWE	ARPC5, DEFB1, DOCK1, IGLL5, P2RX7, PGLYRP4, PLA2G6, PLSCR1
	VOL	APOBEC3H, ARPC5, DEFB1, DMBT1, DOCK1, IFI16, LY86, PGLYRP4, PLSCR1
	SEA	APOBEC3H, ARPC5, C1QC, DMBT1, IRGM, PGLYRP4, PLSCR1
GO.APP	AFR	HLA-A, HLA-B, HLA-C, HLA-DQA2, HLA-G, PSMB9, TAP1, TAP2
	WAA	HLA-A, HLA-B, HLA-C, HLA-DOA, HLA-DQA2, HLA-DRA, HLA-F, HLA-G, MICA, PSMB9, TAP1, TAP2
	SWE	HLA-A, HLA-B, HLA-C, HLA-DQA2, HLA-DRA, HLA-F, HLA-G, MICA, PSMB9, TAP1
	ENE	HLA-A, HLA-B, HLA-C, HLA-DOA, HLA-DQA2, HLA-DRA, HLA-F, HLA-G, MICA, PSMB9, TAP1
	VOL	HLA-A, HLA-B, HLA-C, HLA-DQA2, HLA-DRA, HLA-F, HLA-G, MICA, PSMB9, TAP1
	SOA	HLA-A, HLA-B, HLA-C, HLA-DOA, HLA-DQA2, HLA-DRA, HLA-F, HLA-G, MICA, PSMB9, TAP1
	WSI	HLA-A, HLA-B, HLA-C, HLA-DQA2, HLA-DRA, HLA-G, MICA, PSMB9, TAP1
	SSI	HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DQA2, HLA-DRA, HLA-F, HLA-G, MICA, PSMB9, TAP1
	CSI	HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DQA2, HLA-DQB2, HLA-DRA, HLA-F, HLA-G, MICA, PSMB9, TAP1
	NSI	HLA-A, HLA-B, HLA-C, HLA-DQA2, HLA-DQB2, HLA-DRA, HLA-F, HLA-G, MICA, PSMB9, TAP1
	COL	HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DQA2, HLA-DQB2, HLA-DRA, HLA-F, HLA-G
	SEM	HLA-A, HLA-B, HLA-C, HLA-DOA, HLA-DPA1, HLA-DQA2, HLA-DRA, HLA-G, MICA
	SEA	HLA-A, HLA-B, HLA-C, HLA-DOA, HLA-DQA2, HLA-DRA, HLA-G, TAP2
GO.Adapt	COL	TNFRSF11A

Table B.10 Genes from each significantly enriched immune category and population combination from the HKA test

	Pop.	Genes
GO.Bact	SWE	DEFB1, P2RX7, PGLYRP4
	VOL	DEFB1, DMBT1, PGLYRP4
	SOA	DEFB1, DMBT1, PGLYRP4, TLR6
	WSI	DEFB1, DMBT1, PGLYRP4
	SSI	DEFB1, DMBT1, PGLYRP4, TLR6
	CSI	DMBT1, P2RX7, PGLYRP4, TLR6
	NSI	DEFB1, DMBT1, PGLYRP4, TLR6
	COL	DEFB1, DMBT1, PGLYRP4
	SEM	DEFB1, DMBT1, P2RX7, PGLYRP4
	SEA	DEFB1, DMBT1, PGLYRP4
GO.Tcell	WAA	CLC, LGALS8, TESPA1, UBASH3A
	SWE	CLC, INPP5D, LGALS8, P2RX7, TESPA1
	ENE	BTN3A2, CLC, INPP5D, LGALS8, TESPA1
	SOA	BTN3A2, CLC, INPP5D, TESPA1
	WSI	BAX, CLC, LGALS8, TESPA1
	SSI	BAX, CLC, INPP5D, LGALS8, TESPA1
	CSI	CLC, LGALS8, P2RX7, TESPA1
	NSI	CLC, LGALS8, SIGLEC1, TESPA1
	SEM	CLC, LGALS8, P2RX7, TESPA1
GO.Bcell	SSI	BAX, INPP5D
GO.Innate	WAA	APOBEC3H, ARPC5, DEFB1, DOCK1, LY86, PGLYRP4
	SWE	APOBEC3H, ARPC5, DEFB1, DOCK1, IGLL5, P2RX7, PGLYRP4, PLA2G6, PLSCR1
	ENE	APOBEC3H, DEFB1, DOCK1, LY86, PGLYRP4, PLA2G6, PLSCR1
	VOL	APOBEC3H, ARPC5, DEFB1, DMBT1, DOCK1, IFI16, LY86, PGLYRP4, PLSCR1
	SOA	APOBEC3H, ARPC5, DEFB1, DMBT1, IGLL5, PGLYRP4, TLR6
	WSI	APOBEC3H, ARPC5, DEFB1, DMBT1, PGLYRP4, PLSCR1
	CSI	ARPC5, C1QC, DMBT1, LY86, P2RX7, PGLYRP4, TLR6, VAV3
	COL	DEFB1, DMBT1, LY86, PGLYRP4, PLA2G6, PLSCR1
	SEM	ARPC5, C1QC, DEFB1, DMBT1, IGLL5, P2RX7, PGLYRP4, PLSCR1
	SEA	APOBEC3H, ARPC5, C1QC, DEFB1, DMBT1, IRGM, PGLYRP4, PLSCR1
GO.Adapt	COL	TNFRSF11A

Table B.11 Genes from each significantly enriched immune category and population combination from the HKA test, excluding MHC genes

B.2.2 β enrichments

	Pop.	Genes
Go.Bact	AFR	ANXA3, CHIT1, DEFB1, DMBT1, EPPIN, EPPIN-WFDC6
	ENE	DEFB1, DMBT1, PGLYRP4
	SOA	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6, PGLYRP4
	WSI	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6, GNLY, PGLYRP4
	SSI	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6, PGLYRP4
	CSI	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6, PGLYRP4
	NSI	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6, PGLYRP4
	COL	DMBT1, EPPIN, EPPIN-WFDC6
	SEM	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6
	SEA	ANXA3, DEFB1, DMBT1, EPPIN, EPPIN-WFDC6, GNLY
GO.Tcell	AFR	CD209, CLC, HLA-DPA1, HLA-DQA2, HLA-DQB2, IRF1, SPINK5, TESPA1
	WAA	ADAM17, CD209, HLA-DPA1, HLA-DQA2, IRF1, LGALS8, MICA, PLCG2, TESPA1, TRPM4
	SWE	CLC, HLA-DPA1, IRF1, LGALS8, MICA, SFTPD, TESPA1, TRPM4
	ENE	CLC, HLA-DPA1, HLA-DQA2, INPP5D, IRF1, LGALS8, MICA, PMAIP1, TESPA1
	VOL	HLA-DPA1, HLA-DQA2, IRF1, LGALS8, TESPA1, TRPM4
	SOA	EBI3, HLA-DPA1, HLA-DQA2, IRF1, LGALS8, MICA, TESPA1
	WSI	CLC, HLA-DPA1, HLA-G, IRF1, LGALS8, MICA, TESPA1, TRPM4
	SSI	HLA-DPA1, HLA-DQA2, IRF1, LGALS8, MICA, PLCG2, SFTPD, TESPA1, VAV1
	CSI	FLT3, HLA-DPA1, MICA, SFTPD, SPINK5, TESPA1
	NSI	HLA-DQA2, HLA-G, IRF1, MICA, PLCG2, SFTPD, TESPA1
	COL	HLA-DPA1, HLA-DQA2, HLA-DQB2, IRF1, MICA, SFTPD, TESPA1
	SEA	FLT3, HLA-DPA1, HLA-DQA2, HLA-G, LGALS8, MICA, SPINK5, TESPA1, TRPM4
GO.Bcell	WAA	ADAM17, CD38, PLCG2
GO.Innate	WAA	APOL1, ARPC5, CD209, DEFB1, DMBT1, IFI16, IRF1, PGLYRP4, PLCG2, ZC3HAV1
	SWE	APOL1, ARPC5, DEFB1, DMBT1, IRF1, PGLYRP4, SFTPD
	VOL	ADCY3, APOBEC3H, APOL1, ARPC5, DEFB1, DMBT1, IRF1, PDE1C, PGLYRP4
	SSI	C3, DEFB1, DMBT1, IRF1, ITPR1, KL, PDE1C, PGLYRP4, PLCG2, SFTPD, VAV1
	CSI	APOBEC3H, ARPC5, C3, DEFB1, DMBT1, KL, PGLYRP4, SFTPD
	SEM	APOBEC3H, APOL1, ARPC5, C3, CLEC6A, DEFB1, DMBT1, IRF1, KL, PDE1C, SFTPD
GO.APP	AFR	AP1M2, CD209, HLA-DPA1, HLA-DQA2, HLA-DQB2
	WAA	AP1M2, CD209, HLA-DPA1, HLA-DQA2, MICA
	ENE	HLA-DPA1, HLA-DQA2, MICA
	WSI	AP1M2, HLA-DPA1, HLA-G, MICA
	CSI	AP1M2, HLA-DPA1, MICA
	NSI	AP1M2, HLA-DQA2, HLA-G, MICA
	COL	AP1M2, HLA-DPA1, HLA-DQA2, HLA-DQB2, MICA
	SEA	HLA-DPA1, HLA-DQA2, HLA-G, MICA
GO.Adapt	AFR	IRF1
	SWE	IRF1
	ENE	IRF1
	VOL	IRF1
	SOA	IRF1
	SSI	IRF1
	NSI	IRF1
	COL	IRF1
	SEM	IRF1

Table B.12 Genes from each significantly enriched immune category and population combination from the β test

	Pop.	Genes
GO.Bact	AFR	ANXA3, CHIT1, DEFB1, DMBT1, EPPIN, EPPIN-WFDC6
	SWE	DEFB1, DMBT1, PGLYRP4
	ENE	DEFB1, DMBT1, PGLYRP4
	VOL	DEFB1, DMBT1, PGLYRP4
	SOA	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6, PGLYRP4
	WSI	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6, GNLY, PGLYRP4
	SSI	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6, PGLYRP4
	CSI	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6, PGLYRP4
	NSI	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6, PGLYRP4
	COL	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6
	SEM	DEFB1, DMBT1, EPPIN, EPPIN-WFDC6
	SEA	ANXA3, DEFB1, DMBT1, EPPIN, EPPIN-WFDC6, GNLY
GO.Tcell	AFR	CD209, CLC, IRF1, SFTPD, SPINK5, TESPA1
	WAA	ADAM17, CD209, IRF1, LGALS8, PLCG2, TESPA1, TRPM4
	SWE	CLC, IRF1, LGALS8, SFTPD, TESPA1, TRPM4
	ENE	CLC, INPP5D, IRF1, LGALS8, PMAIP1, TESPA1
	WSI	CLC, IRF1, LGALS8, SFTPD, TESPA1, TRPM4
	SSI	IRF1, LGALS8, PLCG2, SFTPD, TESPA1, VAV1
	CSI	FLT3, IRF1, SFTPD, SPINK5, TESPA1
	NSI	IRF1, PLCG2, SFTPD, TESPA1
	SEA	FLT3, IRF1, LGALS8, SPINK5, TESPA1, TRPM4
GO.Bcell	WAA	ADAM17, CD38, PLCG2
GO.Innate	AFR	ARPC5, CD209, CLEC6A, DEFB1, DMBT1, IRF1, SFTPD
	WAA	APOL1, ARPC5, CD209, CLEC6A, DEFB1, DMBT1, IFI16, IRF1, PGLYRP4, PLCG2, ZC3HAV1
	SWE	APOL1, ARPC5, DEFB1, DMBT1, IRF1, PGLYRP4, SFTPD
	VOL	ADCY3, APOBEC3H, APOL1, ARPC5, DEFB1, DMBT1, IRF1, PDE1C, PGLYRP4
	SOA	ARPC5, DEFB1, DMBT1, IRF1, PDE1C, PGLYRP4, ZC3HAV1
	WSI	APOL1, ARPC5, DEFB1, DMBT1, IFI16, IRF1, KL, PDE1C, PGLYRP4, SFTPD
	SSI	C3, DEFB1, DMBT1, IFI16, IRF1, ITPR1, KL, PDE1C, PGLYRP4, PLCG2, SFTPD, VAV1
	CSI	APOBEC3H, ARPC5, C3, CLEC6A, DEFB1, DMBT1, IRF1, KL, PGLYRP4, SFTPD
	NSI	APOBEC3H, DEFB1, DMBT1, IFI16, IRF1, PDE1C, PGLYRP4, PLCG2, SFTPD
	SEM	APOBEC3H, APOL1, ARPC5, C3, CLEC6A, DEFB1, DMBT1, IRF1, KL, PDE1C, SFTPD
	SEA	APOBEC3H, APOL1, ARPC5, CLEC6A, DEFB1, DMBT1, IRF1
GO.Adapt	AFR	IRF1
	WAA	IRF1
	SWE	IRF1
	ENE	IRF1
	VOL	IRF1
	SOA	IRF1
	SSI	IRF1
	CSI	IRF1
	NSI	IRF1
	COL	IRF1
	SEM	IRF1
	SEA	IRF1

Table B.13 Genes from each significantly enriched immune category and population combination from the β test, excluding MHC genes

B.2.3 Tajima's D enrichments

	Pop.	Genes
GO.Bact	AFR	DYNC2H1, HLA-DRB1, PGLYRP2
	VOL	DYNC2H1, NFKB1, TLR2
GO.Virus	WSI	AP1S3, CCDC130, DDX1, DUOX2, HNRNPH3, NEDD4L, NUP43, RNASEL, SPEN, TRIM23
GO.Tcell	AFR	AMICA1, CD3E, HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-F, HLA-G, LGALS8, MICA, MPZL2, PRNP, RPS27A, SPINK5, TAP2
GO.APP	AFR	DYNC2H1, HLA-A, HLA-B, HLA-C, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-F, HLA-G, MICA, RPS27A, TAP2
	VOL	DYNC2H1, HLA-F, HLA-G, PSMD3
	SOA	DYNC2H1, HLA-A, HLA-DPA1, HLA-DPB1, PSMC1, PSMD3
GO.Adapt	COL	NLRP10
HP.Bact	COL	BICD2, DAAM1, DSCR4, EMC2, ETFA, G2E3, LDB2, MADD, MLH1, MTBP, MYO1G, NASP, NBAS, NFAT5, NOB1, NR1H3, PATL1, PI4KA, PPWD1, RTN1, RUFY2, SAAL1, SPI1, STX3, SYNC, ZBTB5

Table B.14 Genes from each significantly enriched immune category and population combination from the Tajima's D test (balancing selection)

	Pop.	Genes
GO.Bact	VOL	DYNC2H1, NFKB1, TLR2
GO.Virus	WSI	AP1S3, CCDC130, DDX1, DUOX2, HNRNPH3, NEDD4L, NUP43, RNASEL, SPEN, TRIM23
GO.APP	SOA	DYNC2H1, PSMC1, PSMD3
GO.Adapt	COL	NLRP10
HP.Bact	COL	BICD2, DAAM1, DSCR4, EMC2, ETFA, G2E3, LDB2, MADD, MLH1, MTBP, MYO1G, NASP, NBAS, NFAT5, NOB1, NR1H3, PATL1, PI4KA, PPWD1, RTN1, RUFY2, SAAL1, SPI1, STX3, SYNC, ZBTB5

Table B.15 Genes from each significantly enriched immune category and population combination from the Tajima's D test (balancing selection), excluding MHC genes

Appendix C

Full lists of significant SNPs from the window-based positive selection analyses

The following tables are the full lists per population of all SNPs that were significant after being in the top 1% of test results, at least 3 standard deviations above the neutral DIND cutoff, and having a CADD score of greater than or equal to 15.

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_106.6	nSL, iHS	1:106740553	22.3	IG	-	0.71	0.77	0.86	0.92	0.91	0.93	0.94	0.96	0.90	0.94	0.34	0.95	0.98
1_113.8	TD	1:113841736	17.21	IG	-	0.88	0.63	0.63	0.42	0.37	0.32	0.41	0.53	0.47	0.48	0.21	0.47	0.33
1_113.8	TD	1:113993588	21.1	I,N	MAGI3	0.83	0.60	0.48	0.48	0.46	0.68	0.24	0.18	0.18	0.22	0.34	0.29	0.31
1_118.6	iHS	1:118732429	20.3	U	SPAG17	0.83	0.52	0.36	0.29	0.37	0.43	0.47	0.66	0.66	0.74	0.61	0.55	0.33
1_178.8	nSL	1:178986624	16.09	IG,R	-	0.44	0.54	0.42	0.27	0.35	0.29	0.29	0.40	0.31	0.34	0.39	0.36	0.29
1_82.6	nSL	1:82642406	18.95	IG	-	0.73	0.81	0.75	0.74	0.67	0.50	0.62	0.40	0.24	0.32	0.55	0.47	0.61
10_10.2	nSL, iHS	10:10334668	20.7	IG	-	0.50	0.37	0.52	0.42	0.67	0.52	0.68	0.65	0.66	0.74	0.47	0.59	0.68
10_74	nSL	10:74058511	18.01	IG,R	-	0.56	0.25	0.23	0.39	0.57	0.68	0.91	0.82	0.97	0.96	0.58	0.91	0.92
10_74	nSL	10:74094172	15.61	R	-	0.71	0.92	0.97	0.97	0.98	0.91	0.91	0.91	0.95	0.88	0.97	0.91	0.82
10_74	nSL	10:74094172	15.61	3,D	DNAJB12	0.71	0.92	0.97	0.97	0.98	0.91	0.91	0.91	0.95	0.88	0.97	0.91	0.82
10_75.2	nSL	10:75385332	16.6	R	-	0.63	0.92	0.94	0.94	0.87	0.82	0.59	0.74	0.60	0.64	0.92	0.69	0.69
10_75.2	nSL	10:75385332	16.6	I,N,5	USP54	0.63	0.92	0.94	0.94	0.87	0.82	0.59	0.74	0.60	0.64	0.92	0.69	0.69
11_110.4	nSL, iHS	11:110575279	17.67	I	ARHGAP20	0.73	0.92	0.92	0.97	0.89	0.96	1.00	0.99	0.98	0.98	0.97	0.98	1.00
11_110.6	nSL, iHS	11:110686788	16.32	IG	-	0.56	0.46	0.31	0.35	0.35	0.18	0.44	0.25	0.16	0.12	0.00	0.22	0.22
11_38.4	nSL	11:38441080	19.37	IG	-	0.54	0.21	0.11	0.15	0.30	0.34	0.26	0.43	0.37	0.18	0.21	0.33	0.51
11_38.4	nSL	11:38487436	17.63	IG	-	0.65	0.60	0.47	0.44	0.76	0.79	0.79	0.87	0.98	1.00	0.97	0.95	1.00
11_55.8	TD	11:55834132	17.98	U	OR5J1P	0.48	0.40	0.23	0.42	0.35	0.61	0.53	0.50	0.52	0.60	0.71	0.62	0.60
11_63.4	TD	11:63452100	16.76	I,N,NMD	RTN3	0.40	0.67	0.81	0.86	0.80	0.66	0.97	0.88	0.95	1.00	0.97	0.97	0.96
12_111.4	nSL	12:111558554	16.67	I	CUX2	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
12_111.4	nSL	12:111518088	15.57	I	CUX2	0.67	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
12_54.6	TD	12:54749637	19.47	D	COPZ1	0.71	0.94	0.91	0.85	0.83	0.89	0.71	0.56	0.45	0.50	0.58	0.78	0.88
12_79.4	nSL	12:79562731	17.53	I,N	SYT1	0.83	0.96	0.98	0.91	0.89	0.91	0.82	0.82	0.82	0.96	1.00	0.98	1.00
12_88	iHS	12:88112132	17.8	IG	-	0.87	0.62	0.52	0.65	0.74	0.73	0.79	0.82	0.95	0.96	0.37	0.88	0.84
12_88	iHS	12:88177722	16.25	NC,U	MKRN9P	0.73	0.69	0.75	0.75	0.72	0.86	0.76	0.90	0.89	0.92	0.76	0.93	0.99
13_72	TD	13:72148746	17.55	I	DACH1	0.46	0.79	0.94	0.89	0.70	0.43	0.68	0.56	0.29	0.74	0.24	0.45	0.73
14_47.6	nSL	14:47774501	17.01	I,NMD,NC	MDGA2	0.50	0.08	0.14	0.14	0.20	0.14	0.18	0.25	0.40	0.26	0.03	0.48	0.46
14_48.2	nSL	14:48302450	20.8	IG	-	0.54	0.13	0.16	0.25	0.13	0.02	0.29	0.22	0.08	0.14	0.00	0.16	0.38
14_48.2	nSL	14:48333747	18.83	IG	-	0.58	0.17	0.13	0.23	0.17	0.13	0.41	0.37	0.32	0.44	0.05	0.34	0.47
14_48.2	nSL	14:48285397	15.67	IG	-	0.63	0.25	0.44	0.57	0.54	0.45	0.79	0.68	0.71	0.76	0.92	0.40	0.56
14_61	iHS	14:61085205	18.68	IG	-	0.62	0.88	0.86	0.93	0.98	0.70	0.97	0.84	0.97	1.00	0.92	0.81	0.82
14_61	iHS	14:61098683	17.89	IG	-	0.73	0.88	0.86	0.93	0.98	0.70	0.97	0.84	0.97	1.00	0.92	0.81	0.82
14_61	iHS	14:61153794	17.12	IG	-	0.67	0.73	0.80	0.75	0.78	0.70	0.59	0.51	0.47	0.70	0.87	0.78	0.66
14_61	iHS	14:61020617	16.7	R	-	0.58	0.88	0.88	0.95	0.98	0.71	0.97	0.82	0.97	1.00	0.92	0.81	0.82
14_61	iHS	14:61063771	16.36	IG	-	0.54	0.88	0.86	0.93	0.98	0.64	0.94	0.78	0.94	0.84	0.92	0.74	0.80
14_61	iHS	14:61062890	15.61	IG	-	0.54	0.88	0.86	0.93	0.98	0.63	0.94	0.78	0.94	0.84	0.92	0.74	0.80
14_61	iHS	14:61101089	15.51	IG	-	0.73	0.88	0.86	0.93	0.98	0.70	0.97	0.84	0.97	1.00	0.92	0.81	0.82
14_61	iHS	14:61180657	23.6	3,NMD,M	SIX4	0.60	0.73	0.80	0.75	0.76	0.70	0.53	0.49	0.45	0.70	0.87	0.79	0.66
14_61.2	TD	14:61322571	15.28	I,N	MNAT1	0.65	0.71	0.80	0.75	0.76	0.63	0.56	0.63	0.48	0.70	0.89	0.78	0.76
15_44.8	iHS	15:44943757	23.1	M	SPG11	0.48	0.35	0.56	0.58	0.54	0.38	0.47	0.43	0.45	0.54	0.74	0.50	0.53
15_73.8	nSL	15:73981701	15.81	I,NMD	CD276	0.50	0.62	0.55	0.43	0.52	0.50	0.44	0.41	0.45	0.52	0.42	0.50	0.54
16_14.6	TD	16:14658272	15.85	R	-	0.44	0.25	0.19	0.14	0.24	0.41	0.29	0.56	0.45	0.56	0.26	0.53	0.42
16_14.6	TD	16:14658272	15.85	I,N	PARN	0.44	0.25	0.19	0.14	0.24	0.41	0.29	0.56	0.45	0.56	0.26	0.53	0.42
16_23	nSL, iHS	16:23051713	15.9	IG	-	0.85	0.73	0.78	0.68	0.67	0.84	0.59	0.74	0.74	0.48	0.03	0.81	0.73
17_26.4	TD	17:26457614	21.4	I,NMD	NLK	0.77	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
17_3.6	nSL, iHS	17:3627619	15.87	R	-	0.46	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17_3.6	nSL, iHS	17:3627619	15.87	S	GSG2	0.46	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17_3.6	nSL, iHS	17:3657159	20.6	NC,M	ITGAE	0.79	0.58	0.67	0.70	0.52	0.38	0.32	0.21	0.03	0.10	0.37	0.28	0.23
17_3.6	nSL, iHS	17:3627619	15.87	I,N,U,D,NC	ITGAE	0.46	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17_61.8	TD	17:61804723	17.44	I,NMD,N,D,U	STRADA	0.54	0.27	0.28	0.34	0.20	0.13	0.06	0.10	0.06	0.14	0.16	0.10	0.07
18_20.4	iHS	18:20472563	16.98	I	RBBP8	0.56	0.08	0.19	0.25	0.41	0.34	0.35	0.57	0.45	0.20	0.37	0.53	0.66
18_20.4	iHS	18:20464171	15.56	I	RBBP8	0.56	0.10	0.22	0.25	0.41	0.39	0.38	0.57	0.45	0.22	0.37	0.53	0.66

Table C.1 Part one of the full list of significant SNPs from the window-based tests in the AFR population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
18_32	nSL	18:32139198	18.56	I	DTNA	0.50	0.77	0.86	0.81	0.96	0.82	1.00	0.96	1.00	1.00	1.00	0.98	0.92
18_32	nSL	18:32123883	17.85	I	DTNA	0.50	0.77	0.88	0.81	0.96	0.82	1.00	0.96	1.00	1.00	1.00	0.98	0.92
18_32	nSL	18:32170700	16.4	U,I	DTNA	0.77	0.98	0.89	0.90	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.93
18_32	nSL	18:32113901	15.46	I	DTNA	0.83	0.85	0.88	0.79	0.96	0.82	1.00	0.96	0.98	1.00	1.00	0.97	0.93
18_32	nSL	18:32178405	15.2	I	DTNA	0.65	0.87	0.84	0.80	0.91	0.82	0.97	0.94	1.00	1.00	1.00	0.97	0.93
18_32.2	nSL	18:32363229	16.42	I,N	DTNA	0.79	0.90	0.83	0.88	0.91	0.86	0.94	0.99	0.94	0.98	0.95	0.97	0.98
19_42.4	TD	19:42463447	17.72	R	-	0.71	0.42	0.53	0.75	0.50	0.70	0.74	0.65	0.79	0.72	0.68	0.78	0.76
19_42.4	TD	19:42549934	18.67	I,NMD,U	GRIK5	0.58	0.77	0.78	0.92	0.89	0.77	0.88	0.85	0.76	0.86	0.89	0.86	0.80
19_42.4	TD	19:42463447	17.72	NC,U,5	RABAC1	0.71	0.42	0.53	0.75	0.50	0.70	0.74	0.65	0.79	0.72	0.68	0.78	0.76
2_125.8	iHS	2:125875195	16.79	IG	-	0.48	0.52	0.36	0.50	0.52	0.50	0.21	0.22	0.11	0.18	0.26	0.16	0.14
2_150.4	nSL	2:150432976	16.11	S	MMADHC	0.65	0.12	0.05	0.09	0.07	0.13	0.15	0.13	0.10	0.08	0.24	0.12	0.09
2_156	TD	2:156023165	19.23	IG	-	0.63	0.27	0.48	0.38	0.48	0.54	0.50	0.51	0.53	0.56	0.61	0.55	0.63
2_17.6	TD	2:17775828	15.55	I	VSNL1	0.87	0.90	0.94	0.92	0.96	0.93	0.91	0.90	0.94	0.92	0.92	0.81	0.91
2_193.8	nSL, iHS	2:193810226	20.2	IG	-	0.52	0.54	0.45	0.57	0.57	0.66	0.76	0.65	0.60	0.40	0.21	0.50	0.37
2_194.8	nSL, iHS	2:194849951	19.24	IG	-	0.62	0.90	0.97	0.92	0.89	0.93	0.91	0.87	0.87	0.84	1.00	0.91	0.86
2_195	nSL, iHS	2:195072430	15.12	IG	-	0.73	0.96	1.00	0.99	0.96	1.00	0.94	0.88	0.95	0.94	1.00	0.98	0.98
2_195	nSL, iHS	2:195047879	20.1	U	HNRNPA1P47	0.83	0.98	1.00	0.99	0.96	1.00	0.94	0.88	0.95	0.94	1.00	0.98	0.98
2_199.8	nSL	2:199945178	16.07	IG	-	0.63	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2_199.8	nSL	2:199982854	15.94	IG	-	0.56	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2_201.8	iHS	2:201928083	16.69	I,NMD,N	FAM126B	0.42	0.12	0.19	0.17	0.26	0.09	0.15	0.01	0.02	0.02	0.00	0.02	0.01
2_201.8	iHS	2:201928083	16.69	NC	HNRNPA1P35	0.42	0.12	0.19	0.17	0.26	0.09	0.15	0.01	0.02	0.02	0.00	0.02	0.01
2_48	TD	2:48173866	19.65	R	-	0.83	0.56	0.61	0.51	0.72	0.61	0.71	0.91	0.89	0.86	0.84	0.93	0.74
2_63	TD	2:63195179	16.34	I,N	EHBP1	0.42	0.38	0.69	0.61	0.67	0.68	0.59	0.63	0.47	0.66	0.37	0.62	0.84
2_95.6	TD	2:95720630	15.79	D	MAL	0.73	0.75	0.77	0.73	0.74	0.63	0.56	0.66	0.56	0.70	0.97	0.72	0.63
20_32	TD	20:32088983	15.93	I,N	CBFA2T2	0.65	0.96	1.00	1.00	1.00	0.95	0.97	0.94	0.98	0.82	0.55	0.95	0.99
20_32	TD	20:32119377	15.73	I,N	CBFA2T2	0.69	0.96	1.00	1.00	1.00	0.95	0.97	0.94	0.98	0.82	0.55	0.95	0.99
20_35.2	nSL, TD	20:35347508	17.74	I	NDRG3	0.87	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20_37.4	nSL, iHS	20:37440046	17.74	I	PPP1R16B	0.69	0.56	0.58	0.65	0.61	0.71	0.68	0.50	0.58	0.42	0.66	0.74	0.73
20_37.4	nSL, iHS	20:37506313	17.55	I,N	PPP1R16B	0.44	0.37	0.44	0.61	0.63	0.82	0.62	0.53	0.37	0.44	0.53	0.60	0.48
20_37.4	nSL, iHS	20:37506313	17.55	D	RN7SL116P	0.44	0.37	0.44	0.61	0.63	0.82	0.62	0.53	0.37	0.44	0.53	0.60	0.48
3_115.6	nSL, iHS	3:115624440	18.67	I,N	LSAMP	0.79	0.83	0.80	0.89	0.63	0.80	0.71	0.87	0.69	0.86	1.00	0.84	0.67
3_115.6	nSL, iHS	3:115624856	17.01	I,N	LSAMP	0.77	0.83	0.80	0.89	0.63	0.80	0.71	0.87	0.69	0.86	1.00	0.84	0.67
3_115.6	nSL, iHS	3:115608931	15.46	I,N	LSAMP	0.50	0.56	0.63	0.58	0.46	0.59	0.47	0.47	0.48	0.64	0.89	0.41	0.39
3_19	nSL, iHS	3:19189694	15.92	R	-	0.60	0.98	0.98	0.97	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00	0.99
3_19	nSL, iHS	3:19189694	15.92	U	KCNH8	0.60	0.98	0.98	0.97	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00	0.99
4_1	nSL	4:1110735	20.3	R	-	0.44	0.04	0.08	0.13	0.07	0.04	0.06	0.00	0.00	0.02	0.00	0.00	0.00
4_1	nSL	4:1110735	20.3	U	RNF212	0.44	0.04	0.08	0.13	0.07	0.04	0.06	0.00	0.00	0.02	0.00	0.00	0.00
4_1	nSL	4:1110735	20.3	U	SNORA48	0.44	0.04	0.08	0.13	0.07	0.04	0.06	0.00	0.00	0.02	0.00	0.00	0.00
4_1	nSL	4:1110735	20.3	I,N	TMED11P	0.44	0.04	0.08	0.13	0.07	0.04	0.06	0.00	0.00	0.02	0.00	0.00	0.00
4_106.8	nSL	4:106910958	16.81	R	-	0.69	0.71	0.78	0.81	0.72	0.79	0.88	0.57	0.52	0.70	0.16	0.53	0.80
4_106.8	nSL	4:106891531	21.5	I,D,3	NPNT	0.67	0.48	0.42	0.49	0.39	0.52	0.53	0.35	0.37	0.46	0.61	0.21	0.18
4_106.8	nSL	4:106910958	16.81	I	NPNT	0.69	0.71	0.78	0.81	0.72	0.79	0.88	0.57	0.52	0.70	0.16	0.53	0.80
4_106.8	nSL	4:106959289	21.6	D	TBCK	0.71	0.71	0.64	0.61	0.76	0.68	0.62	0.78	0.79	0.64	0.89	0.83	0.74
4_107.6	nSL, iHS	4:107690286	17.62	IG	-	0.81	0.92	0.83	0.85	0.93	1.00	0.94	0.99	1.00	1.00	1.00	1.00	1.00
4_107.6	nSL, iHS	4:107647748	17.31	IG	-	0.81	0.92	0.81	0.83	0.91	0.96	0.91	0.99	1.00	1.00	0.97	1.00	0.99
4_107.6	nSL, iHS	4:107637162	16.99	IG	-	0.81	0.92	0.81	0.84	0.91	0.96	0.91	0.99	1.00	1.00	0.97	1.00	0.99
4_107.6	nSL, iHS	4:107620115	16.28	IG	-	0.65	0.92	0.75	0.81	0.89	0.89	0.91	0.96	0.97	0.98	0.97	0.91	0.94
4_107.6	nSL, iHS	4:107625430	15.42	IG	-	0.65	0.94	0.75	0.81	0.89	0.89	0.91	0.96	0.97	0.98	0.97	0.91	0.94
4_145.8	iHS, TD	4:145889835	19.25	I,N	ANAPC10	0.44	0.25	0.19	0.16	0.15	0.27	0.41	0.40	0.48	0.24	0.61	0.55	0.59

Table C.2 Part two of the full list of significant SNPs from the window-based tests in the AFR population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
4_147	nSL	4:147075530	16.27	IG	-	0.73	0.96	0.98	0.98	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4_48.6	TD	4:48789269	22.4	IG	-	0.75	0.58	0.44	0.58	0.54	0.79	0.44	0.63	0.65	0.34	0.58	0.72	0.63
4_74	TD	4:74074624	16.01	I,N	ANKRD17	0.67	0.63	0.61	0.49	0.57	0.64	0.53	0.43	0.45	0.54	0.32	0.50	0.32
5_143.6	nSL	5:143769614	16.07	I	KCTD16	0.54	0.65	0.66	0.59	0.76	0.91	0.76	0.74	0.69	0.80	0.92	0.84	0.78
5_143.6	nSL	5:143621324	15.63	I	KCTD16	0.88	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
5_15	nSL, iHS	5:15118045	20.3	R	-	0.77	0.85	0.84	0.88	0.91	0.95	0.97	0.94	1.00	1.00	0.97	1.00	1.00
5_15.4	nSL, iHS	5:15499642	16.53	R	-	0.69	0.96	0.98	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
5_15.4	nSL, iHS	5:15499642	16.53	U	FBXL7	0.69	0.96	0.98	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
5_23	nSL, iHS	5:23013484	18.63	R	-	0.63	0.96	0.73	0.80	0.83	0.70	0.62	0.66	0.63	0.34	0.68	0.47	0.40
5_43.8	nSL	5:43834933	15.94	IG	-	0.67	0.87	0.92	0.92	0.85	0.84	0.74	0.76	0.68	0.70	0.97	0.81	0.71
5_65.6	iHS	5:65702974	17.38	IG	-	0.87	0.88	0.86	0.80	0.85	0.68	0.88	0.97	0.89	0.98	1.00	0.91	0.91
5_65.6	iHS	5:65690103	16.33	IG	-	0.87	0.92	0.86	0.82	0.87	0.68	0.88	0.97	0.89	0.98	1.00	0.91	0.91
5_74.6	TD	5:74637711	16.16	I,N	HMGCR	0.73	0.60	0.61	0.56	0.46	0.43	0.38	0.47	0.60	0.82	0.42	0.48	0.52
5_93.2	TD	5:93365716	15.35	I,NMD,N	FAM172A	0.46	0.75	0.67	0.87	0.80	0.88	0.79	0.84	0.82	0.94	0.97	0.84	0.86
5_93.4	TD	5:93568101	16.63	I	KIAA0825	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5_97	nSL	5:97050539	20.4	IG	-	0.79	0.92	0.91	0.92	0.96	0.95	1.00	1.00	1.00	1.00	0.95	1.00	1.00
6_108.8	TD	6:108975464	16.09	R	-	0.77	0.85	0.88	0.83	0.78	0.91	0.71	0.84	0.89	0.94	0.97	0.86	0.77
6_108.8	TD	6:108975464	16.09	I,U	FOXO3	0.77	0.85	0.88	0.83	0.78	0.91	0.71	0.84	0.89	0.94	0.97	0.86	0.77
6_111.6	TD	6:111620758	15.23	3,NMD,NC	REV3L	0.48	0.13	0.20	0.19	0.26	0.16	0.18	0.16	0.10	0.08	0.11	0.17	0.20
6_158.8	nSL	6:158932817	15.13	D	CACYBPP3	0.67	0.52	0.48	0.51	0.61	0.57	0.56	0.47	0.40	0.38	0.37	0.45	0.50
6_158.8	nSL	6:158886564	15.52	I	TULP4	0.83	0.44	0.39	0.36	0.54	0.43	0.50	0.50	0.40	0.56	0.29	0.45	0.43
6_158.8	nSL	6:158932817	15.13	D,3	TULP4	0.67	0.52	0.48	0.51	0.61	0.57	0.56	0.47	0.40	0.38	0.37	0.45	0.50
6_40.4	nSL	6:40574859	21.1	IG	-	0.50	0.54	0.39	0.44	0.46	0.48	0.26	0.31	0.37	0.34	0.00	0.21	0.22
6_40.4	nSL	6:40510655	17.55	I	LRFN2	0.58	0.56	0.36	0.43	0.33	0.43	0.29	0.31	0.39	0.28	0.05	0.17	0.18
6_55.8	iHS	6:55973669	15.26	I,NMD	COL21A1	0.50	0.48	0.52	0.52	0.50	0.64	0.76	0.76	0.68	0.44	0.34	0.74	0.74
6_56.8	nSL, iHS	6:56917538	17.19	M,NC	KIAA1586	0.75	0.87	0.72	0.74	0.91	0.84	0.82	0.93	0.82	0.62	0.95	0.91	0.76
6_64.4	nSL	6:64523696	16.64	I	EYS	0.88	0.98	0.94	0.88	0.93	0.91	0.88	0.88	0.87	0.92	0.84	0.84	0.92
7_111	nSL	7:111186693	18.99	I,N	IMMP2L	0.50	0.87	0.88	0.88	0.91	0.89	0.91	1.00	0.98	1.00	1.00	0.97	0.92
7_111	nSL	7:111108645	16.34	I	IMMP2L	0.58	0.90	0.91	0.91	0.93	0.96	0.97	1.00	1.00	1.00	1.00	1.00	0.98
7_130.2	iHS	7:130327837	15.13	I	COPG2	0.50	0.12	0.03	0.06	0.07	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
7_29	nSL, iHS	7:29130712	21.6	I	CPVL	0.79	0.50	0.48	0.56	0.61	0.52	0.41	0.41	0.39	0.32	0.63	0.28	0.31
7_40.8	iHS	7:40844463	15.79	R	-	0.83	0.44	0.59	0.47	0.41	0.48	0.62	0.50	0.50	0.64	0.82	0.52	0.47
7_40.8	iHS	7:40842145	19.97	I,N	SUGCT	0.83	0.62	0.64	0.58	0.52	0.57	0.65	0.53	0.55	0.66	0.82	0.52	0.51
7_40.8	iHS	7:40885658	15.87	I,N	SUGCT	0.88	0.75	0.83	0.78	0.57	0.75	0.59	0.44	0.37	0.40	0.47	0.17	0.11
7_40.8	iHS	7:40844463	15.79	I,N,U	SUGCT	0.83	0.44	0.59	0.47	0.41	0.48	0.62	0.50	0.50	0.64	0.82	0.52	0.47
7_40.8	iHS	7:40889040	15.01	I,N	SUGCT	0.88	0.75	0.83	0.78	0.57	0.75	0.59	0.44	0.37	0.40	0.47	0.17	0.13
7_66.2	TD	7:66212597	19.2	I,NMD	KCTD7	0.60	0.37	0.36	0.42	0.28	0.29	0.32	0.29	0.06	0.14	0.71	0.29	0.48
7_66.2	TD	7:66212597	19.2	I,NMD	RABGEF1	0.60	0.37	0.36	0.42	0.28	0.29	0.32	0.29	0.06	0.14	0.71	0.29	0.48
8_117	iHS	8:117185988	17.08	I,N	LINC00536	0.56	0.75	0.88	0.89	0.80	0.75	1.00	0.94	0.98	1.00	0.95	0.93	0.91
8_37.8	nSL, iHS	8:37849831	15.4	IG	-	0.77	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
8_38	nSL, iHS	8:38149404	16.71	I,U,D	WHSC1L1	0.73	0.88	0.75	0.84	0.87	0.88	0.79	0.72	0.81	0.76	0.74	0.78	0.77
8_67.4	iHS	8:67592249	15.46	R	-	0.54	1.00	0.98	0.96	0.98	0.96	0.85	0.97	0.95	0.86	1.00	0.98	0.90
8_67.4	iHS	8:67592249	15.46	3,NC,D	C8orf44	0.54	1.00	0.98	0.96	0.98	0.96	0.85	0.97	0.95	0.86	1.00	0.98	0.90
8_67.4	iHS	8:67592249	15.46	5,NC	C8orf44-SGK3	0.54	1.00	0.98	0.96	0.98	0.96	0.85	0.97	0.95	0.86	1.00	0.98	0.90
8_71	nSL	8:71100634	15.54	I,NMD	NCOA2	0.56	0.71	0.94	0.92	0.83	0.54	0.76	0.63	0.69	0.80	0.76	0.48	0.40
8_73.4	nSL	8:73591578	15.93	R	-	0.71	0.13	0.17	0.11	0.20	0.21	0.38	0.40	0.50	0.34	0.00	0.33	0.16
8_73.4	nSL	8:73587628	19.09	I	KCNB2	0.71	0.13	0.17	0.11	0.20	0.21	0.38	0.40	0.50	0.34	0.03	0.33	0.16
8_73.4	nSL	8:73591578	15.93	I	KCNB2	0.71	0.13	0.17	0.11	0.20	0.21	0.38	0.40	0.50	0.34	0.00	0.33	0.16
8_73.6	nSL, iHS	8:73603360	20.2	I	KCNB2	0.69	0.13	0.17	0.11	0.20	0.21	0.38	0.40	0.50	0.26	0.00	0.33	0.14
8_73.6	nSL, iHS	8:73660251	19.26	I	KCNB2	0.50	0.27	0.39	0.30	0.41	0.14	0.38	0.37	0.40	0.24	0.00	0.36	0.19
8_86.2	TD	8:86361082	17.16	3	CA3	0.65	0.65	0.56	0.61	0.72	0.46	0.62	0.51	0.71	0.72	0.47	0.40	0.33
8_9.4	nSL, iHS	8:9573284	16.69	I,U	TNKS	0.60	0.65	0.66	0.66	0.63	0.52	0.47	0.59	0.58	0.48	0.34	0.43	0.42
8_99.8	TD	8:99925628	17.19	I,N	STK3	0.44	0.71	0.58	0.69	0.67	0.77	0.53	0.53	0.65	0.56	0.18	0.55	0.54
9_100.6	nSL, iHS	9:100700654	16.98	R	-	0.73	0.54	0.56	0.57	0.67	0.86	0.88	0.72	0.87	0.94	0.76	0.83	0.91
9_100.6	nSL, iHS	9:100620412	17.97	D	FOXE1	0.62	0.52	0.48	0.67	0.59	0.75	0.79	0.78	0.81	0.60	0.66	0.90	0.97
9_100.6	nSL, iHS	9:100700654	16.98	I	HEMGN	0.73	0.54	0.56	0.57	0.67	0.86	0.88	0.72	0.87	0.94	0.76	0.83	0.91
9_116.2	nSL	9:116293017	18.25	I,N	RGS3	0.83	0.88	0.92	0.80	0.89	0.89	0.94	0.99	0.98	1.00	1.00	1.00	1.00
9_38	nSL	9:38079516	15.05	R,IG	-	0.88	0.98	0.98	1.00	0.91	1.00	0.88	0.99	0.98	1.00	1.00	1.00	0.99
9_84.6	TD	9:84732557	15.7	IG	-	0.79	0.90	0.92	0.81	0.93	0.98	1.00	0.94	0.95	0.94	0.97	0.93	0.96

Table C.3 Part three of the full list of significant SNPs from the window-based tests in the AFR population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_214.2	nSL	1:214356500	16.14	IG	-	0.08	0.65	0.64	0.62	0.52	0.39	0.18	0.12	0.10	0.00	0.16	0.03	0.02
1_36	iHS, TD	1:36194819	18.64	I	CLSPN	0.33	0.88	0.89	0.85	0.63	0.82	0.41	0.24	0.15	0.36	0.32	0.47	0.33
1_36.2	iHS, TD	1:36267349	16.07	IG	-	0.21	0.88	0.89	0.85	0.63	0.84	0.41	0.24	0.15	0.36	0.34	0.45	0.34
1_46.4	iHS	1:46493460	19.03	M,D,U	MAST2	0.31	0.50	0.45	0.41	0.50	0.32	0.62	0.60	0.71	0.50	0.42	0.69	0.52
10_76.6	nSL, iHS	10:76719679	16.44	I	KAT6B	0.17	0.73	0.86	0.85	0.80	0.36	0.74	0.56	0.68	0.54	0.71	0.62	0.67
10_83.6	nSL	10:83694001	18.57	I,NMD	NRG3	0.00	0.63	0.78	0.77	0.52	0.57	0.29	0.25	0.21	0.06	0.03	0.19	0.29
10_83.6	nSL	10:83715438	16.03	I,NMD	NRG3	0.06	0.63	0.80	0.79	0.57	0.57	0.32	0.29	0.24	0.10	0.05	0.29	0.29
11_85	TD	11:85079419	15.23	I	DLG2	0.44	0.73	0.66	0.62	0.43	0.50	0.50	0.51	0.47	0.26	0.50	0.50	0.57
12_81.6	nSL	12:81655861	15.18	I,N,NMD	PPFIA2	0.38	0.42	0.59	0.54	0.41	0.46	0.21	0.26	0.27	0.32	0.24	0.14	0.14
14_63.4	nSL	14:63568154	16.38	I	KCNH5	0.25	0.63	0.64	0.58	0.61	0.63	0.62	0.47	0.60	0.78	0.76	0.33	0.27
14_64	TD	14:64068806	19.18	I	WDR89	0.04	0.73	0.66	0.63	0.80	0.64	0.68	0.56	0.73	0.30	0.47	0.34	0.22
14_64	TD	14:64068806	19.18	D	HSPE1P2	0.04	0.73	0.66	0.63	0.80	0.64	0.68	0.56	0.73	0.30	0.47	0.34	0.22
15_48.4	nSL, iHS	15:48591267	15.31	I,N	SLC12A1	0.50	0.71	0.81	0.80	0.80	0.66	0.82	0.59	0.35	0.52	0.53	0.50	0.70
15_48.6	nSL, iHS	15:48653760	21.3	IG	-	0.27	0.73	0.86	0.84	0.80	0.57	0.68	0.54	0.26	0.38	0.42	0.33	0.51
15_48.6	nSL, iHS	15:48772936	16.19	I,NMD	FBN1	0.35	0.69	0.81	0.68	0.63	0.54	0.47	0.41	0.26	0.64	0.68	0.47	0.41
16_67.4	TD	16:67581883	15.96	D	FAM65A	0.15	0.83	0.86	0.85	0.96	0.71	0.85	0.91	0.89	0.96	1.00	0.97	0.98
16_67.4	TD	16:67581883	15.96	R	-	0.15	0.83	0.86	0.85	0.96	0.71	0.85	0.91	0.89	0.96	1.00	0.97	0.98
2_109.4	nSL, iHS	2:109501404	15.62	I,NC	CCDC138	0.33	0.88	0.88	0.86	0.85	0.68	0.88	0.97	0.98	1.00	0.95	0.91	0.94
2_136.2	iHS	2:136228887	19.85	I,NMD,N	ZRANB3	0.13	0.81	0.75	0.78	0.83	0.54	0.79	0.87	0.73	0.84	0.50	0.81	0.77
2_152.4	iHS	2:152536498	23.2	M	NEB	0.42	0.81	0.83	0.91	0.76	0.66	0.53	0.47	0.34	0.42	0.97	0.52	0.29
2_152.4	iHS	2:152587814	19.8	I	NEB	0.33	0.85	0.83	0.91	0.78	0.64	0.50	0.53	0.37	0.42	0.92	0.59	0.34
2_178.4	nSL, iHS	2:178461560	16.92	IG	-	0.27	0.67	0.80	0.78	0.61	0.57	0.59	0.50	0.45	0.68	0.95	0.57	0.64
20_20.4	iHS	20:20485588	16.07	I	RALGAPA2	0.38	0.65	0.67	0.74	0.54	0.80	0.47	0.44	0.23	0.56	0.82	0.41	0.41
20_53.6	nSL	20:53670219	19.22	IG	-	0.29	0.75	0.53	0.75	0.83	0.80	0.79	0.74	0.81	0.86	0.97	0.76	0.71
4_34.4	iHS	4:34421742	18.85	IG	-	0.15	0.83	0.83	0.88	0.87	0.64	0.76	0.79	0.76	0.74	0.47	0.76	0.70
4_34.4	iHS	4:34490761	16.65	IG	-	0.25	0.87	0.88	0.88	0.87	0.63	0.76	0.79	0.74	0.66	0.89	0.79	0.70
4_34.4	iHS	4:34421782	15.86	IG	-	0.15	0.83	0.83	0.88	0.87	0.64	0.76	0.79	0.76	0.70	0.47	0.76	0.70
4_41.4	nSL, iHS	4:41526749	16.3	I,N	LIMCH1	0.13	0.56	0.47	0.63	0.72	0.41	0.82	0.68	0.89	0.84	0.76	0.67	0.59
5_112.6	iHS	5:112725480	15.1	I	MCC	0.00	0.73	0.48	0.48	0.67	0.63	0.65	0.69	0.76	0.70	0.32	0.69	0.46
5_43.8	TD	5:43834933	15.94	IG	-	0.67	0.87	0.92	0.92	0.85	0.84	0.74	0.76	0.68	0.70	0.97	0.81	0.71
6_108.2	iHS	6:108375468	15.26	I,U	OSTM1	0.04	0.73	0.63	0.66	0.59	0.66	0.59	0.46	0.34	0.24	0.13	0.52	0.41
6_35.2	iHS	6:35260530	16.3	3,NMD,S,D,NC,U	ZNF76	0.27	0.67	0.80	0.84	0.78	0.77	0.85	0.79	0.79	0.84	0.82	0.66	0.84
6_35.2	iHS	6:35261990	15.73	3,NMD,I,N,D	ZNF76	0.27	0.67	0.80	0.84	0.78	0.77	0.85	0.79	0.77	0.84	0.82	0.66	0.84
6_35.2	iHS	6:35261990	15.73	U	DEF6	0.27	0.67	0.80	0.84	0.78	0.77	0.85	0.79	0.77	0.84	0.82	0.66	0.84
7_97.8	iHS	7:97807882	20.4	I	LMTK2	0.02	0.71	0.53	0.50	0.37	0.20	0.06	0.19	0.02	0.06	0.08	0.17	0.13
9_108	TD	9:108037717	15.06	I,NMD	SLC44A1	0.27	0.85	0.86	0.85	0.78	0.79	0.68	0.72	0.66	0.50	0.58	0.88	0.87
9_126.2	nSL	9:126397618	16.38	I,N	DENND1A	0.21	0.73	0.64	0.53	0.52	0.71	0.74	0.78	0.77	0.76	0.66	0.76	0.72
9_37.8	iHS	9:37820416	16.18	I,N	DCAF10	0.23	0.56	0.38	0.36	0.43	0.41	0.32	0.13	0.08	0.20	0.16	0.16	0.14

Table C.4 Full list of significant SNPs from the window-based tests in the WAA population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_36	iHS, TD	1:36194819	18.64	I	CLSPN	0.33	0.88	0.89	0.85	0.63	0.82	0.41	0.24	0.15	0.36	0.32	0.47	0.33
1_36.2	TD	1:36267349	16.07	IG	-	0.21	0.88	0.89	0.85	0.63	0.84	0.41	0.24	0.15	0.36	0.34	0.45	0.34
10_117.6	nSL, iHS	10:117784322	18.23	IG	-	0.29	0.67	0.77	0.80	0.76	0.61	0.74	0.72	0.69	0.40	0.53	0.43	0.42
10_65.8	nSL, iHS	10:65827306	18.5	IG	-	0.52	0.94	0.78	0.83	0.87	0.82	0.94	0.85	0.76	0.82	0.92	0.81	0.83
10_65.8	nSL, iHS	10:65837750	18.43	IG	-	0.23	0.94	0.78	0.79	0.85	0.82	0.91	0.84	0.74	0.78	0.89	0.81	0.87
10_76.6	TD	10:76719679	16.44	I	KAT6B	0.17	0.73	0.86	0.85	0.80	0.36	0.74	0.56	0.68	0.54	0.71	0.62	0.67
10_83.6	nSL	10:83694001	18.57	I,NMD	NRG3	0.00	0.63	0.78	0.77	0.52	0.57	0.29	0.25	0.21	0.06	0.03	0.19	0.29
10_84	iHS	10:84115932	15.52	I,NMD	NRG3	0.08	0.77	0.86	0.84	0.83	0.88	0.74	0.59	0.60	0.52	0.47	0.53	0.36
12_111.6	nSL, iHS	12:111664404	17.08	I,N	CUX2	0.52	0.67	0.78	0.85	0.80	0.50	0.82	0.54	0.39	0.78	0.95	0.40	0.49
12_80.2	TD	12:80392551	17.06	IG	-	0.17	0.83	0.86	0.87	0.89	0.61	0.79	0.72	0.69	0.94	0.95	0.53	0.48
13_77.6	TD	13:77727365	15.17	I,U	MYCBP2	0.00	0.52	0.67	0.59	0.59	0.30	0.35	0.34	0.32	0.36	0.26	0.26	0.16
14_67.4	TD	14:67527028	16.31	I,N,D	GPHN	0.13	0.85	0.89	0.86	0.83	0.59	0.71	0.81	0.66	0.58	0.61	0.60	0.60
15_34.4	nSL	15:34519563	15.14	I,NMD,N,U,D,NC	EMC4	0.35	0.42	0.73	0.62	0.67	0.66	0.88	0.72	0.71	0.94	0.76	0.62	0.51
15_48.4	nSL	15:48591267	15.31	I,N	SLC12A1	0.50	0.71	0.81	0.80	0.80	0.66	0.82	0.59	0.35	0.52	0.53	0.50	0.70
15_48.6	nSL	15:48653760	21.3	IG	-	0.27	0.73	0.86	0.84	0.80	0.57	0.68	0.54	0.26	0.38	0.42	0.33	0.51
15_48.6	nSL	15:48629884	19.44	I,NMD,N,U	DUT	0.33	0.75	0.86	0.85	0.80	0.57	0.76	0.56	0.26	0.48	0.47	0.33	0.51
15_48.6	nSL	15:48772936	16.19	I,NMD	FBN1	0.35	0.69	0.81	0.68	0.63	0.54	0.47	0.41	0.26	0.64	0.68	0.47	0.41
15_48.6	nSL	15:48668532	15.12	IG	-	0.15	0.73	0.86	0.87	0.85	0.68	0.85	0.84	0.85	0.88	0.61	0.67	0.78
15_64.6	nSL, iHS	15:64706259	18.72	I,NMD,D	TRIP4	0.29	0.69	0.86	0.90	0.85	0.50	1.00	0.96	0.92	0.98	0.95	0.95	0.94
15_64.6	nSL, iHS	15:64793617	17.79	I,3,U	ZNF609	0.31	0.69	0.86	0.89	0.85	0.52	1.00	0.96	0.92	0.98	0.95	0.95	0.94
16_67.4	TD	16:67581883	15.96	R	-	0.15	0.83	0.86	0.85	0.96	0.71	0.85	0.91	0.89	0.96	1.00	0.97	0.98
16_67.4	TD	16:67581883	15.96	D	FAM65A	0.15	0.83	0.86	0.85	0.96	0.71	0.85	0.91	0.89	0.96	1.00	0.97	0.98
18_7.4	nSL, iHS	18:7548731	18.59	IG	-	0.67	0.60	0.88	0.82	0.80	0.61	0.74	0.35	0.39	0.70	0.34	0.31	0.29
2_109.4	iHS	2:109501404	15.62	I,NC	CCDC138	0.33	0.88	0.88	0.86	0.85	0.68	0.88	0.97	0.98	1.00	0.95	0.91	0.94
2_136.2	nSL, iHS	2:136228887	19.85	I,NMD,N	ZRANB3	0.13	0.81	0.75	0.78	0.83	0.54	0.79	0.87	0.73	0.84	0.50	0.81	0.77
2_136.4	nSL, iHS	2:136407479	21.1	I,N,M,D	R3HDM1	0.02	0.06	0.48	0.52	0.41	0.14	0.12	0.09	0.05	0.06	0.05	0.03	0.00
2_136.4	nSL, iHS	2:136576577	16.01	I	LCT	0.10	0.25	0.56	0.58	0.59	0.34	0.59	0.50	0.50	0.60	0.26	0.36	0.49
2_136.6	nSL, iHS	2:136608646	17.65	I,N	MCM6	0.02	0.04	0.48	0.45	0.30	0.11	0.06	0.01	0.03	0.06	0.00	0.02	0.00
2_136.6	nSL, iHS	2:136740900	15.48	I,NMD,N,U	DARS	0.04	0.25	0.59	0.56	0.57	0.29	0.62	0.38	0.48	0.62	0.24	0.36	0.50
2_136.8	nSL	2:136899859	17.56	IG	-	0.27	0.13	0.41	0.50	0.30	0.18	0.12	0.19	0.10	0.04	0.18	0.16	0.20
2_152.4	iHS	2:152587814	19.8	I	NEB	0.33	0.85	0.83	0.91	0.78	0.64	0.50	0.53	0.37	0.42	0.92	0.59	0.34
2_178.4	nSL, iHS	2:178461560	16.92	IG	-	0.27	0.67	0.80	0.78	0.61	0.57	0.59	0.50	0.45	0.68	0.95	0.57	0.64
2_28	iHS	2:28019106	16.44	R	-	0.69	0.77	0.86	0.75	0.65	0.73	0.74	0.38	0.45	0.44	0.29	0.45	0.51
2_28	iHS	2:28019106	16.44	I,NMD	MRPL33	0.69	0.77	0.86	0.75	0.65	0.73	0.74	0.38	0.45	0.44	0.29	0.45	0.51
2_28	iHS	2:28019106	16.44	I,NMD	RBKS	0.69	0.77	0.86	0.75	0.65	0.73	0.74	0.38	0.45	0.44	0.29	0.45	0.51
20_34.2	TD	20:34252654	20.2	I	RBM12	0.40	0.62	0.89	0.72	0.85	0.59	0.85	0.82	0.94	0.90	0.89	0.83	0.72
20_34.2	TD	20:34252654	20.2	R	-	0.40	0.62	0.89	0.72	0.85	0.59	0.85	0.82	0.94	0.90	0.89	0.83	0.72
20_34.2	TD	20:34252654	20.2	I,NMD,N,U,NC	CPNE1	0.40	0.62	0.89	0.72	0.85	0.59	0.85	0.82	0.94	0.90	0.89	0.83	0.72
20_34.2	TD	20:34252654	20.2	D	NFS1	0.40	0.62	0.89	0.72	0.85	0.59	0.85	0.82	0.94	0.90	0.89	0.83	0.72
3_130.2	nSL, iHS	3:130284284	25.2	M	COL6A6	0.06	0.81	0.89	0.88	0.83	0.59	0.68	0.59	0.65	0.60	0.92	0.50	0.44
3_130.2	nSL, iHS	3:130271485	17.06	IG	-	0.15	0.81	0.89	0.88	0.80	0.61	0.53	0.49	0.56	0.58	0.92	0.48	0.44
3_32	nSL	3:32130343	15.04	IG	-	0.02	0.35	0.55	0.47	0.43	0.30	0.47	0.18	0.24	0.38	0.29	0.02	0.11
4_140.2	iHS	4:140359958	17.09	U	ACA64	0.21	0.60	0.52	0.54	0.50	0.43	0.65	0.47	0.66	0.52	0.95	0.57	0.37
4_141.2	nSL	4:141265921	19.12	I	SCOC	0.23	0.60	0.70	0.65	0.59	0.50	0.88	0.57	0.73	0.74	0.45	0.52	0.64
4_148	nSL, iHS	4:148174084	15.68	IG	-	0.42	0.62	0.77	0.75	0.59	0.48	0.50	0.51	0.52	0.28	0.50	0.45	0.14
5_145	nSL	5:145072132	17.53	I,N	PRELID2	0.48	0.94	0.89	0.92	0.93	0.80	0.97	0.81	0.87	0.72	0.32	0.83	0.88
6_130.6	nSL	6:130663546	18.37	I	SAMD3	0.06	0.60	0.75	0.65	0.63	0.50	0.76	0.51	0.55	0.60	0.55	0.59	0.54
7_120.2	nSL	7:120344218	15.4	I	KCND2	0.37	0.79	0.77	0.74	0.61	0.75	0.53	0.50	0.40	0.32	0.32	0.57	0.72
7_132.6	iHS	7:132624847	19.59	I,NMD,N	CHCHD3	0.37	0.37	0.45	0.39	0.28	0.27	0.12	0.19	0.15	0.20	0.00	0.36	0.54
8_16.2	nSL	8:16369596	19.39	I,N,U	MSR1	0.50	0.87	0.88	0.92	1.00	0.91	0.97	0.99	0.95	1.00	0.95	0.97	0.97
9_108	iHS, TD	9:108037717	15.06	I,NMD	SLC44A1	0.27	0.85	0.86	0.85	0.78	0.79	0.68	0.72	0.66	0.50	0.58	0.88	0.87
9_12.6	nSL	9:12612529	16.46	IG	-	0.35	0.40	0.70	0.72	0.46	0.27	0.18	0.10	0.00	0.04	0.00	0.05	0.17

Table C.5 Full list of significant SNPs from the window-based tests in the SWE population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_10	TD	1:10093457	15.61	R	-	0.06	0.81	0.92	0.80	0.70	0.50	0.62	0.65	0.53	0.68	0.97	0.71	0.64
1_10	TD	1:10093457	15.61	5,U	UBE4B	0.06	0.81	0.92	0.80	0.70	0.50	0.62	0.65	0.53	0.68	0.97	0.71	0.64
1_10	TD	1:10037468	15.37	I,NMD,D,U	NMNAT1	0.13	0.79	0.92	0.82	0.76	0.48	0.65	0.63	0.53	0.66	0.66	0.71	0.59
1_215.2	nSL, iHS	1:215308203	15.05	I,NMD	KCNK2	0.33	0.48	0.64	0.78	0.67	0.29	0.68	0.71	0.71	0.72	0.74	0.57	0.49
1_219.2	iHS	1:219210952	16.6	IG	-	0.23	0.31	0.52	0.58	0.63	0.50	0.38	0.26	0.27	0.32	0.42	0.07	0.09
1_36	TD	1:36194819	18.64	I	CLSPN	0.33	0.88	0.89	0.85	0.63	0.82	0.41	0.24	0.15	0.36	0.32	0.47	0.33
10_21.2	nSL	10:21390298	15.94	I	NEBL	0.10	0.67	0.58	0.57	0.35	0.43	0.50	0.32	0.29	0.22	0.39	0.28	0.30
10_64	nSL, iHS	10:64024417	15.98	I	RTKN2	0.21	0.52	0.59	0.59	0.67	0.54	0.85	0.93	0.94	0.94	0.97	0.84	0.86
10_64	nSL, iHS	10:64024417	15.98	R	-	0.21	0.52	0.59	0.59	0.67	0.54	0.85	0.93	0.94	0.94	0.97	0.84	0.86
10_65.8	nSL, iHS, TD	10:65827306	18.5	IG	-	0.52	0.94	0.78	0.83	0.87	0.82	0.94	0.85	0.76	0.82	0.92	0.81	0.83
10_65.8	nSL, iHS, TD	10:65837750	18.43	IG	-	0.23	0.94	0.78	0.79	0.85	0.82	0.91	0.84	0.74	0.78	0.89	0.81	0.87
10_65.8	nSL, iHS, TD	10:65815577	16.92	IG	-	0.10	0.92	0.78	0.83	0.87	0.73	0.94	0.85	0.76	0.78	0.92	0.81	0.81
10_76.6	nSL, TD	10:76719679	16.44	I	KAT6B	0.17	0.73	0.86	0.85	0.80	0.36	0.74	0.56	0.68	0.54	0.71	0.62	0.67
10_83.6	nSL	10:83694001	18.57	I,NMD	NRG3	0.00	0.63	0.78	0.77	0.52	0.57	0.29	0.25	0.21	0.06	0.03	0.19	0.29
11_41.8	nSL	11:41900027	16.01	R	-	0.25	0.67	0.73	0.77	0.67	0.55	0.74	0.74	0.76	0.84	0.89	0.69	0.68
11_67	TD	11:67164133	19.52	I,N,NMD,D	RAD9A	0.15	0.90	0.84	0.90	0.91	0.86	0.65	0.66	0.81	0.74	0.29	0.62	0.69
11_67	TD	11:67164133	19.52	U	RNU6-1238P	0.15	0.90	0.84	0.90	0.91	0.86	0.65	0.66	0.81	0.74	0.29	0.62	0.69
11_67	TD	11:67164133	19.52	D	PPP1CA	0.15	0.90	0.84	0.90	0.91	0.86	0.65	0.66	0.81	0.74	0.29	0.62	0.69
12_111.6	nSL, iHS	12:111664404	17.08	I,N	CUX2	0.52	0.67	0.78	0.85	0.80	0.50	0.82	0.54	0.39	0.78	0.95	0.40	0.49
12_2.6	nSL	12:2765860	15.51	I,D	CACNA1C	0.15	0.63	0.63	0.78	0.72	0.75	0.76	0.71	0.82	0.96	0.61	0.69	0.51
13_26.8	iHS	13:26906088	15.92	I	CDK8	0.12	0.63	0.58	0.42	0.52	0.57	0.29	0.28	0.50	0.50	0.47	0.33	0.22
13_75.6	nSL, iHS	13:75729189	16.49	IG	-	0.27	0.60	0.56	0.63	0.39	0.48	0.47	0.50	0.63	0.60	0.24	0.50	0.58
13_75.6	nSL, iHS	13:75717402	15.07	IG	-	0.31	0.60	0.56	0.64	0.35	0.46	0.38	0.51	0.61	0.56	0.24	0.55	0.61
14_67.4	iHS, TD	14:67527028	16.31	I,N,D	GPHN	0.13	0.85	0.89	0.86	0.83	0.59	0.71	0.81	0.66	0.58	0.61	0.60	0.60
14_67.6	iHS, TD	14:67705668	18.72	U	MPP5	0.58	0.85	0.91	0.90	0.85	0.68	0.74	0.85	0.81	0.70	0.68	0.62	0.63
15_48.6	nSL, iHS	15:48629884	19.44	I,NMD,N,U	DUT	0.33	0.75	0.86	0.85	0.80	0.57	0.76	0.56	0.26	0.48	0.47	0.33	0.51
15_48.6	nSL, iHS	15:48772936	16.19	I,NMD	FBN1	0.35	0.69	0.81	0.68	0.63	0.54	0.47	0.41	0.26	0.64	0.68	0.47	0.41
15_48.6	nSL, iHS	15:48668532	15.12	IG	-	0.15	0.73	0.86	0.87	0.85	0.68	0.85	0.84	0.85	0.88	0.61	0.67	0.78
15_59.2	nSL	15:59323048	22.9	M,S,U	RNF111	0.02	0.60	0.44	0.49	0.35	0.25	0.24	0.26	0.15	0.10	0.11	0.16	0.14
15_59.2	nSL	15:59347929	15.52	S	RNF111	0.23	0.77	0.59	0.64	0.70	0.84	0.68	0.63	0.56	0.54	0.63	0.60	0.54
16_67.4	TD	16:67581883	15.96	R	-	0.15	0.83	0.86	0.85	0.96	0.71	0.85	0.91	0.89	0.96	1.00	0.97	0.98
16_67.4	TD	16:67581883	15.96	D	FAM65A	0.15	0.83	0.86	0.85	0.96	0.71	0.85	0.91	0.89	0.96	1.00	0.97	0.98
17_58.2	TD	17:58226259	17.34	U	CA4	0.27	0.73	0.80	0.80	0.85	0.89	1.00	0.94	0.97	0.84	0.76	0.81	0.72
18_34.6	iHS	18:34664093	18.03	3,NMD,I,M,NC	KIAA1328	0.21	0.85	0.61	0.64	0.63	0.71	0.85	0.88	0.81	0.58	0.74	0.79	0.83
18_7.4	nSL, iHS	18:7548731	18.59	IG	-	0.67	0.60	0.88	0.82	0.80	0.61	0.74	0.35	0.39	0.70	0.34	0.31	0.29
2_123.4	nSL	2:123441331	18.22	IG	-	0.08	0.63	0.78	0.70	0.78	0.75	0.82	0.81	0.65	0.84	0.82	0.74	0.67
2_123.4	nSL	2:123440075	17.27	IG	-	0.12	0.63	0.78	0.70	0.78	0.75	0.82	0.81	0.65	0.84	0.82	0.74	0.67
2_136.2	nSL, iHS	2:136228887	19.85	I,NMD,N	ZRANB3	0.13	0.81	0.75	0.78	0.83	0.54	0.79	0.87	0.73	0.84	0.50	0.81	0.77
2_136.4	nSL, iHS	2:136407479	21.1	I,N,M,D	R3HDM1	0.02	0.06	0.48	0.52	0.41	0.14	0.12	0.09	0.05	0.06	0.05	0.03	0.00
2_136.4	nSL, iHS	2:136576577	16.01	I	LCT	0.10	0.25	0.56	0.58	0.59	0.34	0.59	0.50	0.50	0.60	0.26	0.36	0.49
2_136.6	nSL, iHS	2:136608646	17.65	I,N	MCM6	0.02	0.04	0.48	0.45	0.30	0.11	0.06	0.01	0.03	0.06	0.00	0.02	0.00
2_136.6	nSL, iHS	2:136740900	15.48	I,NMD,N,U	DARS	0.04	0.25	0.59	0.56	0.57	0.29	0.62	0.38	0.48	0.62	0.24	0.36	0.50
2_136.8	nSL	2:136899859	17.56	IG	-	0.27	0.13	0.41	0.50	0.30	0.18	0.12	0.19	0.10	0.04	0.18	0.16	0.20
2_137.6	iHS	2:137687492	21.1	I,NMD	THSD7B	0.08	0.42	0.44	0.54	0.33	0.29	0.21	0.10	0.19	0.18	0.24	0.05	0.12
2_137.6	iHS	2:137731295	18.71	I,NMD	THSD7B	0.00	0.35	0.42	0.52	0.33	0.27	0.18	0.07	0.08	0.14	0.11	0.02	0.00
20_53.6	nSL	20:53670219	19.22	IG	-	0.29	0.75	0.53	0.75	0.83	0.80	0.79	0.74	0.81	0.86	0.97	0.76	0.71
20_53.6	nSL	20:53706622	16.24	IG	-	0.23	0.75	0.58	0.80	0.85	0.86	0.79	0.68	0.79	0.86	0.97	0.76	0.72
3_103.2	nSL, iHS	3:103358697	15.02	IG	-	0.56	0.52	0.66	0.70	0.74	0.36	0.56	0.38	0.53	0.62	0.58	0.38	0.39
3_114.4	iHS	3:114594139	15.28	I,N	ZBTB20	0.33	0.90	0.81	0.90	0.91	0.91	0.97	0.93	0.94	1.00	0.89	0.98	0.84
3_114.4	iHS	3:114594139	15.28	I,N	ZBTB20-AS3	0.33	0.90	0.81	0.90	0.91	0.91	0.97	0.93	0.94	1.00	0.89	0.98	0.84
3_129.4	iHS	3:129574472	16.68	I,N,D	TMCC1	0.12	0.81	0.86	0.85	0.89	0.77	0.79	0.71	0.63	0.78	0.71	0.47	0.38
3_129.4	iHS	3:129574472	16.68	R	-	0.12	0.81	0.86	0.85	0.89	0.77	0.79	0.71	0.63	0.78	0.71	0.47	0.38
3_38.4	nSL, iHS	3:38496193	15.22	I,N	ACVR2B	0.02	0.44	0.61	0.46	0.37	0.61	0.62	0.32	0.40	0.44	0.39	0.31	0.18
3_38.4	nSL, iHS	3:38496193	15.22	NC	ACVR2B-AS1	0.02	0.44	0.61	0.46	0.37	0.61	0.62	0.32	0.40	0.44	0.39	0.31	0.18
3_38.4	nSL, iHS	3:38496193	15.22	R	-	0.02	0.44	0.61	0.46	0.37	0.61	0.62	0.32	0.40	0.44	0.39	0.31	0.18
4_140.2	iHS	4:140359958	17.09	U	ACA64	0.21	0.60	0.52	0.54	0.50	0.43	0.65	0.47	0.66	0.52	0.95	0.57	0.37
4_141.2	nSL	4:141265921	19.12	I	SCOC	0.23	0.60	0.70	0.65	0.59	0.50	0.88	0.57	0.73	0.74	0.45	0.52	0.64
4_151.6	iHS	4:151781297	16.74	I	LRBA	0.08	0.90	0.95	0.89	0.80	0.71	0.62	0.72	0.73	0.94	0.66	0.64	0.56
4_2	iHS	4:2176454	23.1	M,NC	POLN	0.35	0.81	0.86	0.90	0.78	0.71	0.76	0.78	0.82	0.76	0.74	0.64	0.57
4_34.4	nSL	4:34421742	18.85	IG	-	0.15	0.83	0.83	0.88	0.87	0.64	0.76	0.79	0.76	0.74	0.47	0.76	0.70
4_34.4	nSL	4:34421782	15.86	IG	-	0.15	0.83	0.83	0.88	0.87	0.64	0.76	0.79	0.76	0.70	0.47	0.76	0.70
5_21.8	nSL	5:21906641	19.24	I,N	CDH12	0.12	0.75	0.77	0.78	0.65	0.59	0.62	0.56	0.52	0.48	0.39	0.57	0.59
5_64.8	iHS	5:64965757	19.77	D	TRAPPC13	0.50	0.37	0.38	0.47	0.46	0.46	0.35	0.41	0.60	0.50	0.55	0.47	0.52
5_64.8	iHS	5:64965757	19.77	3	SGTB	0.50	0.37	0.38	0.47	0.46	0.46	0.35	0.41	0.60	0.50	0.55	0.47	0.52
5_92.8	TD	5:92995013	17.39	I,NMD	FAM172A	0.42	0.71	0.70	0.90	0.83	0.91	0.79	0.71	0.87	0.94	0.92	0.74	0.67
5_93.2	TD	5:93365716	15.35	I,NMD,N	FAM172A	0.46	0.75	0.67	0.87	0.80	0.88	0.79	0.84	0.82	0.94	0.97	0.84	0.86
7_33.6	nSL	7:33671662	16.47	IG	-	0.21	0.56	0.70	0.75	0.61	0.43	0.56	0.22	0.29	0.44	0.34	0.28	0.30
7_33.6	nSL	7:33627588	16.28	I,NMD,N	BBS9	0.06	0.60	0.75	0.75	0.72	0.61	0.85	0.50	0.60	0.80	0.97	0.57	0.46
9_12.6	nSL, iHS	9:12612529	16.46	IG	-	0.35	0.40	0.70	0.72	0.46	0.27	0.18	0.10	0.00	0.04	0.00	0.05	0.17
9_130.6	nSL	9:130699917	16.51	R	-	0.54	0.77	0.88	0.87	0.76	0.71	0.47	0.31	0.35	0.58	0.76	0.34	0.42
9_130.6	nSL	9:130699917	16.51	I,N,NC	DPM2	0.54	0.77	0.88	0.87	0.76	0.71	0.47	0.31	0.35	0.58	0.76	0.34	0.42
9_130.6	nSL	9:130699917	16.51	D	FAM102A	0.54	0.77	0.88	0.87	0.76	0.71	0.47	0.31	0.35	0.58	0.76	0.34	0.42
9_130.6	nSL	9:130725441	15.01	I,N	FAM102A	0.29	0.63	0.75	0.80	0.67	0.55	0.26	0.22	0.23	0.46	0.66	0.17	0.30
9_87.2	nSL	9:87330255	16.55	I	NTRK2	0.44	0.46	0.52	0.64	0.46	0.57	0.50	0.47	0.50	0.38	0.71	0.55	0.59

Table C.6 Full list of significant SNPs from the window-based tests in the ENE population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_10	TD	1:10093457	15.61	R	-	0.06	0.81	0.92	0.80	0.70	0.50	0.62	0.65	0.53	0.68	0.97	0.71	0.64
1_10	TD	1:10093457	15.61	5,U	UBE4B	0.06	0.81	0.92	0.80	0.70	0.50	0.62	0.65	0.53	0.68	0.97	0.71	0.64
1_10	TD	1:10037468	15.37	I,NMD,D,U	NMNAT1	0.13	0.79	0.92	0.82	0.76	0.48	0.65	0.63	0.53	0.66	0.66	0.71	0.59
1_113	iHS	1:113106633	19.63	I,NMD,N	ST7L	0.04	0.19	0.25	0.20	0.41	0.21	0.26	0.56	0.40	0.70	0.50	0.43	0.39
1_183	nSL, iHS	1:183060072	19.06	I	LAMC1	0.62	0.40	0.42	0.43	0.61	0.34	0.38	0.44	0.61	0.38	0.24	0.41	0.32
1_183	nSL, iHS	1:183013988	15.77	R	-	0.62	0.46	0.42	0.45	0.61	0.34	0.41	0.44	0.63	0.38	0.24	0.43	0.32
1_183	nSL, iHS	1:183013988	15.77	I	LAMC1	0.62	0.46	0.42	0.45	0.61	0.34	0.41	0.44	0.63	0.38	0.24	0.43	0.32
1_193.6	nSL	1:193635400	16.22	IG	-	0.31	0.58	0.53	0.41	0.57	0.70	0.53	0.68	0.68	0.76	0.53	0.83	0.73
1_219.2	nSL	1:219210952	16.6	IG	-	0.23	0.31	0.52	0.58	0.63	0.50	0.38	0.26	0.27	0.32	0.42	0.07	0.09
1_28.8	iHS	1:28826587	15.39	3,D	PHACTR4	0.38	0.69	0.63	0.77	0.70	0.89	0.91	0.87	0.92	0.94	0.76	0.86	0.86
1_32.4	TD	1:32403089	15.2	I,N	PTP4A2	0.17	0.75	0.78	0.81	0.89	0.63	0.82	0.81	0.74	0.64	0.47	0.64	0.48
1_32.4	TD	1:32403089	15.2	R	-	0.17	0.75	0.78	0.81	0.89	0.63	0.82	0.81	0.74	0.64	0.47	0.64	0.48
1_87.4	nSL	1:87488813	17.54	I	HS2ST1	0.42	0.83	0.70	0.73	0.78	0.82	0.85	0.94	0.95	0.94	1.00	0.98	0.91
10_125	iHS	10:125090813	20.2	IG	-	0.38	0.71	0.66	0.64	0.83	0.79	0.88	0.76	0.87	0.78	0.95	0.81	0.84
10_125	iHS	10:125065215	16.4	IG	-	0.33	0.73	0.67	0.66	0.83	0.77	0.88	0.76	0.87	0.78	0.95	0.78	0.82
10_99.6	nSL	10:99682028	18.86	I	CRTAC1	0.48	0.75	0.84	0.70	0.83	0.64	0.91	0.88	0.95	1.00	0.97	0.71	0.82
12_44.4	TD	12:44488039	19.32	I,NMD,N	TMEM117	0.27	0.73	0.89	0.89	0.89	0.84	1.00	0.85	0.92	0.96	0.97	0.90	0.88
12_44.4	TD	12:44532975	19.2	I,NMD,N	TMEM117	0.29	0.73	0.89	0.89	0.89	0.84	1.00	0.85	0.92	0.96	0.95	0.90	0.88
12_44.4	TD	12:44419592	18.75	I,NMD	TMEM117	0.58	0.67	0.83	0.93	0.89	0.93	1.00	0.96	0.95	1.00	0.97	0.95	0.98
12_85.4	nSL	12:85517642	17.52	I,N	LRRIQ1	0.04	0.50	0.56	0.62	0.61	0.71	0.59	0.57	0.56	0.46	0.45	0.57	0.37
13_60.4	iHS	13:60406912	16.17	I,N	DIAPH3	0.52	0.67	0.63	0.70	0.78	0.46	0.88	0.69	0.81	0.86	0.84	0.76	0.73
14_61.2	TD	14:61322571	15.28	I,N	MNAT1	0.65	0.71	0.80	0.75	0.76	0.63	0.56	0.63	0.48	0.70	0.89	0.78	0.76
14_62	nSL, iHS	14:62149156	19.86	I,N	HIF1A-AS1	0.21	0.88	0.94	0.91	0.89	0.88	0.97	0.82	0.94	0.76	1.00	0.88	0.91
14_67.4	TD	14:67527028	16.31	I,N,D	GPHN	0.13	0.85	0.89	0.86	0.83	0.59	0.71	0.81	0.66	0.58	0.61	0.60	0.60
14_68.4	iHS, TD	14:68577892	16.69	R	-	0.40	0.73	0.72	0.87	0.89	0.84	0.91	0.84	0.89	0.86	0.92	0.91	0.88
14_68.4	iHS, TD	14:68577892	16.69	I,N,NMD	RAD51B	0.40	0.73	0.72	0.87	0.89	0.84	0.91	0.84	0.89	0.86	0.92	0.91	0.88
14_68.4	iHS, TD	14:68478139	15.96	I,N,NMD	RAD51B	0.48	0.71	0.72	0.88	0.87	0.84	0.85	0.88	0.89	0.86	0.95	0.95	0.87
15_48.6	iHS	15:48653760	21.3	IG	-	0.27	0.73	0.86	0.84	0.80	0.57	0.68	0.54	0.26	0.38	0.42	0.33	0.51
15_48.6	iHS	15:48629884	19.44	I,NMD,N,U	DUT	0.33	0.75	0.86	0.85	0.80	0.57	0.76	0.56	0.26	0.48	0.47	0.33	0.51
15_48.6	iHS	15:48772936	16.19	I,NMD	FBN1	0.35	0.69	0.81	0.68	0.63	0.54	0.47	0.41	0.26	0.64	0.68	0.47	0.41
15_83.8	iHS	15:83927857	17.15	I	BNC1	0.27	0.73	0.75	0.71	0.83	0.79	0.74	0.87	0.92	0.66	0.79	0.78	0.87
17_75.8	nSL	17:75920984	17.38	R	-	0.19	0.63	0.67	0.69	0.76	0.50	0.68	0.74	0.58	0.50	0.55	0.60	0.62
17_75.8	nSL	17:75920984	17.38	D	RNU1-80P	0.19	0.63	0.67	0.69	0.76	0.50	0.68	0.74	0.58	0.50	0.55	0.60	0.62
19_19.4	nSL	19:19431963	16.89	U	SUGP1	0.35	0.75	0.69	0.66	0.76	0.41	0.68	0.65	0.87	0.58	0.42	0.66	0.61
19_19.4	nSL	19:19431963	16.89	R	-	0.35	0.75	0.69	0.66	0.76	0.41	0.68	0.65	0.87	0.58	0.42	0.66	0.61
19_19.4	nSL	19:19431963	16.89	I,U	MAU2	0.35	0.75	0.69	0.66	0.76	0.41	0.68	0.65	0.87	0.58	0.42	0.66	0.61
2_136.4	nSL, iHS	2:136407479	21.1	I,N,M,D	R3HDM1	0.02	0.06	0.48	0.52	0.41	0.14	0.12	0.09	0.05	0.06	0.05	0.03	0.00
2_136.6	iHS	2:136740900	15.48	I,NMD,N,U	DARS	0.04	0.25	0.59	0.56	0.57	0.29	0.62	0.38	0.48	0.62	0.24	0.36	0.50
2_194.8	TD	2:194849951	19.24	IG	-	0.62	0.90	0.97	0.92	0.89	0.93	0.91	0.87	0.87	0.84	1.00	0.91	0.86
2_213	nSL	2:213096239	21.1	I,N	ERBB4	0.37	0.60	0.70	0.49	0.50	0.46	0.68	0.78	0.85	0.82	0.63	0.78	0.70
2_213	nSL	2:213083282	17.27	I,N	ERBB4	0.06	0.48	0.58	0.42	0.50	0.36	0.65	0.76	0.85	0.80	0.63	0.79	0.69
2_29.8	nSL	2:29865419	17.41	I	ALK	0.42	0.62	0.66	0.73	0.80	0.71	0.79	0.85	0.87	0.80	0.58	0.81	0.87
2_86.6	nSL, iHS	2:86743708	16.76	U	RNU6-640P	0.08	0.42	0.47	0.49	0.61	0.48	0.65	0.76	0.73	0.68	0.58	0.67	0.56
2_86.6	nSL, iHS	2:86743708	16.76	I	RNF103-CHMP3	0.08	0.42	0.47	0.49	0.61	0.48	0.65	0.76	0.73	0.68	0.58	0.67	0.56
2_86.6	nSL, iHS	2:86743708	16.76	I,N,U	CHMP3	0.08	0.42	0.47	0.49	0.61	0.48	0.65	0.76	0.73	0.68	0.58	0.67	0.56
2_86.6	nSL, iHS	2:86732576	15.06	D	RNF103-CHMP3	0.06	0.42	0.47	0.49	0.61	0.48	0.65	0.76	0.73	0.68	0.58	0.67	0.56
2_86.6	nSL, iHS	2:86732576	15.06	3,D	CHMP3	0.06	0.42	0.47	0.49	0.61	0.48	0.65	0.76	0.73	0.68	0.58	0.67	0.56
20_34	nSL, iHS	20:34023962	15.07	D	GDF5OS	0.04	0.52	0.59	0.58	0.70	0.38	0.76	0.56	0.53	0.96	0.89	0.62	0.62
20_34	nSL, iHS	20:34023962	15.07	R	-	0.04	0.52	0.59	0.58	0.70	0.38	0.76	0.56	0.53	0.96	0.89	0.62	0.62
20_34	nSL, iHS	20:34023962	15.07	I	GDF5	0.04	0.52	0.59	0.58	0.70	0.38	0.76	0.56	0.53	0.96	0.89	0.62	0.62
20_34.2	TD	20:34252654	20.2	D	NFS1	0.40	0.62	0.89	0.72	0.85	0.59	0.85	0.82	0.94	0.90	0.89	0.83	0.72
20_34.2	TD	20:34252654	20.2	I,NMD,N,U,NC	CPNE1	0.40	0.62	0.89	0.72	0.85	0.59	0.85	0.82	0.94	0.90	0.89	0.83	0.72
20_34.2	TD	20:34252654	20.2	R	-	0.40	0.62	0.89	0.72	0.85	0.59	0.85	0.82	0.94	0.90	0.89	0.83	0.72
20_34.2	TD	20:34252654	20.2	I	RBM12	0.40	0.62	0.89	0.72	0.85	0.59	0.85	0.82	0.94	0.90	0.89	0.83	0.72
3_103.2	nSL, iHS	3:103358697	15.02	IG	-	0.56	0.52	0.66	0.70	0.74	0.36	0.56	0.38	0.53	0.62	0.58	0.38	0.39
3_129.4	iHS, TD	3:129574472	16.68	I,N,D	TMCC1	0.12	0.81	0.86	0.85	0.89	0.77	0.79	0.71	0.63	0.78	0.71	0.47	0.38
3_129.4	iHS, TD	3:129574472	16.68	R	-	0.12	0.81	0.86	0.85	0.89	0.77	0.79	0.71	0.63	0.78	0.71	0.47	0.38
3_29.6	iHS	3:29652486	19.41	I,N	RBMS3	0.44	0.67	0.86	0.72	0.87	0.59	0.71	0.60	0.73	0.78	0.74	0.59	0.66
3_29.6	iHS	3:29652486	19.41	D	RBMS3-AS2	0.44	0.67	0.86	0.72	0.87	0.59	0.71	0.60	0.73	0.78	0.74	0.59	0.66
4_106.6	TD	4:106799485	15.52	I	INTS12	0.06	0.67	0.75	0.84	0.83	0.75	0.88	0.87	0.79	0.92	0.84	0.90	0.90
4_107.6	TD	4:107611250	16.58	IG	-	0.75	0.81	0.72	0.76	0.89	0.86	0.91	0.96	0.97	0.98	0.97	0.91	0.94
4_107.6	TD	4:107620115	16.28	IG	-	0.65	0.92	0.75	0.81	0.89	0.89	0.91	0.96	0.97	0.98	0.97	0.91	0.94
4_107.6	TD	4:107625430	15.42	IG	-	0.65	0.94	0.75	0.81	0.89	0.89	0.91	0.96	0.97	0.98	0.97	0.91	0.94
4_2	iHS	4:2176454	23.1	M,NC	POLN	0.35	0.81	0.86	0.90	0.78	0.71	0.76	0.78	0.82	0.76	0.74	0.64	0.57
4_34.4	nSL, iHS	4:34421742	18.85	IG	-	0.15	0.83	0.83	0.88	0.87	0.64	0.76	0.79	0.76	0.74	0.47	0.76	0.70
4_34.4	nSL, iHS	4:34490761	16.65	IG	-	0.25	0.87	0.88	0.88	0.87	0.63	0.76	0.79	0.74	0.66	0.89	0.79	0.70
4_34.4	nSL, iHS	4:34421782	15.86	IG	-	0.15	0.83	0.83	0.88	0.87	0.64	0.76	0.79	0.76	0.70	0.47	0.76	0.70
4_81.6	TD	4:81623151	16.05	I,N,NMD	C4orf22	0.35	0.75	0.88	0.95	0.89	0.71	0.76	0.76	0.81	0.56	0.76	0.72	0.76
4_81.8	TD	4:81960465	20.5	I	BMP3	0.08	0.63	0.72	0.84	0.89	0.64	0.76	0.66	0.85	0.98	0.76	0.79	0.64
5_108.4	nSL	5:108478799	16.01	I,NMD	FER	0.52	0.60	0.53	0.60	0.78	0.70	0.85	0.79	0.92	0.88	0.97	0.74	0.71
5_109.6	TD	5:109695143	18.64	I,N	TMEM232	0.35	0.83	0.86	0.87	0.89	0.71	0.56	0.65	0.66	0.68	0.79	0.64	0.48
5_133.6	iHS	5:133741881	16.62	D,I	CDKN2AIPNL	0.17	0.62	0.77	0.68	0.52	0.63	0.56	0.54	0.39	0.56	0.16	0.45	0.52
5_152.6	iHS	5:152699481	22	IG	-	0.13	0.87	0.77	0.75	0.67	0.84	0.82	0.65	0.71	0.88	0.97	0.67	0.71
5_152.6	iHS	5:152688178	18.26	IG	-	0.13	0.87	0.77	0.75	0.67	0.84	0.82	0.65	0.71	0.88	0.97	0.64	0.72
5_92.8	TD	5:92995013	17.39	I,NMD	FAM172A	0.42	0.71	0.70	0.90	0.83	0.91	0.79	0.71	0.87	0.94	0.92	0.74	0.67
6_37.8	iHS	6:37972970	16.82	I	ZFAND3	0.25	0.33	0.28	0.36	0.52	0.48	0.38	0.66	0.73	0.64	0.61	0.55	0.51
7_102.6	iHS	7:102704135	15.91	I,NMD,N	FBXL13	0.21	0.87	0.66	0.76	0.76	0.45	0.53	0.43	0.40	0.46	0.66	0.38	0.32

Table C.7 Full list of significant SNPs from the window-based tests in the VOL population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_11.2	iHS	1:11205058	18.23	S,NC	MTOR	0.02	0.71	0.69	0.75	0.72	0.59	0.62	0.82	0.82	0.76	0.74	0.84	0.72
1_11.2	iHS	1:11205058	18.23	I,N	MTOR-AS1	0.02	0.71	0.69	0.75	0.72	0.59	0.62	0.82	0.82	0.76	0.74	0.84	0.72
1_11.2	iHS	1:11322565	17.3	U	MTOR	0.08	0.71	0.64	0.71	0.67	0.61	0.59	0.79	0.81	0.76	0.71	0.78	0.79
1_11.2	iHS	1:11322565	17.3	R	-	0.08	0.71	0.64	0.71	0.67	0.61	0.59	0.79	0.81	0.76	0.71	0.78	0.79
1_193.6	nSL	1:193635400	16.22	IG	-	0.31	0.58	0.53	0.41	0.57	0.70	0.53	0.68	0.68	0.76	0.53	0.83	0.73
1_28.8	TD	1:28826587	15.39	3,D	PHACTR4	0.38	0.69	0.63	0.77	0.70	0.89	0.91	0.87	0.92	0.94	0.76	0.86	0.86
1_35.6	nSL	1:35648316	15.1	I,NMD,D,U	SFPQ	0.65	0.96	1.00	0.96	0.80	0.80	0.53	0.40	0.39	0.48	0.24	0.47	0.46
1_36.2	iHS, TD	1:36267349	16.07	IG	-	0.21	0.88	0.89	0.85	0.63	0.84	0.41	0.24	0.15	0.36	0.34	0.45	0.34
1_39.6	iHS	1:39714238	17.81	I,N	MACF1	0.42	0.67	0.69	0.70	0.80	0.77	0.68	0.62	0.60	0.66	0.71	0.59	0.68
1_39.6	iHS	1:39643464	16.54	R	-	0.15	0.63	0.70	0.64	0.80	0.70	0.68	0.62	0.60	0.66	0.71	0.59	0.66
1_39.6	iHS	1:39643464	16.54	I,N	MACF1	0.15	0.63	0.70	0.64	0.80	0.70	0.68	0.62	0.60	0.66	0.71	0.59	0.66
1_92.8	TD	1:92970209	15.42	D	EVI5	0.29	0.62	0.70	0.67	0.76	0.82	0.94	0.90	0.95	0.94	0.82	0.93	0.77
1_92.8	TD	1:92875071	15.07	IG	-	0.38	0.48	0.47	0.49	0.59	0.77	0.79	0.84	0.90	0.74	0.50	0.95	0.86
1_93	TD	1:93115127	16.46	I,N	EVI5	0.48	0.62	0.72	0.70	0.76	0.82	0.94	0.90	0.95	0.94	0.87	0.95	0.77
1_93	TD	1:93115127	16.46	U	HMGB3P9	0.48	0.62	0.72	0.70	0.76	0.82	0.94	0.90	0.95	0.94	0.87	0.95	0.77
1_93	TD	1:93165207	15.84	D	RNU4-59P	0.29	0.62	0.72	0.69	0.76	0.79	0.94	0.90	0.95	0.94	0.87	0.93	0.76
1_93	TD	1:93165207	15.84	I,D	EVI5	0.29	0.62	0.72	0.69	0.76	0.79	0.94	0.90	0.95	0.94	0.87	0.93	0.76
1_93	TD	1:93192134	15.11	I,N	EVI5	0.29	0.62	0.72	0.69	0.76	0.77	0.94	0.90	0.95	0.94	0.87	0.93	0.76
10_65.8	nSL, iHS	10:65815577	16.92	IG	-	0.10	0.92	0.78	0.83	0.87	0.73	0.94	0.85	0.76	0.78	0.92	0.81	0.81
10_65.8	nSL, iHS	10:65803370	15.56	IG	-	0.37	0.96	0.88	0.86	0.89	0.73	0.94	0.85	0.76	0.78	0.92	0.81	0.81
11_60.8	nSL	11:60893235	31	M	CD5	0.56	0.42	0.53	0.58	0.59	0.73	0.71	0.88	0.85	1.00	0.89	0.98	0.96
11_60.8	nSL	11:60893235	31	D	VPS37C	0.56	0.42	0.53	0.58	0.59	0.73	0.71	0.88	0.85	1.00	0.89	0.98	0.96
11_67	nSL, iHS, TD	11:67164133	19.52	I,N,NMD,D	RAD9A	0.15	0.90	0.84	0.90	0.91	0.86	0.65	0.66	0.81	0.74	0.29	0.62	0.69
11_67	nSL, iHS, TD	11:67164133	19.52	D	PPP1CA	0.15	0.90	0.84	0.90	0.91	0.86	0.65	0.66	0.81	0.74	0.29	0.62	0.69
11_67	nSL, iHS, TD	11:67164133	19.52	U	RNU6-1238P	0.15	0.90	0.84	0.90	0.91	0.86	0.65	0.66	0.81	0.74	0.29	0.62	0.69
11_71.6	iHS	11:71734211	18.55	I,N,D	NUMA1	0.67	0.96	0.91	0.91	0.98	0.88	0.91	0.74	0.73	0.58	1.00	0.76	0.87
11_71.6	iHS	11:71648966	17.9	I,NMD,N	RNF121	0.46	0.94	0.91	0.91	0.98	0.88	0.91	0.74	0.73	0.58	1.00	0.76	0.87
11_71.6	iHS	11:71740521	17.24	I,N,U	NUMA1	0.67	0.96	0.91	0.91	0.98	0.88	0.91	0.74	0.73	0.58	1.00	0.76	0.87
11_71.6	iHS	11:71753582	16.98	I,N,U	NUMA1	0.52	0.96	0.91	0.91	0.98	0.88	0.91	0.74	0.73	0.58	1.00	0.76	0.87
11_71.6	iHS	11:71783056	16.02	D	MIR3165	0.46	0.96	0.91	0.91	0.98	0.88	0.91	0.74	0.73	0.58	1.00	0.76	0.87
11_71.6	iHS	11:71783056	16.02	I,N,U	NUMA1	0.46	0.96	0.91	0.91	0.98	0.88	0.91	0.74	0.73	0.58	1.00	0.76	0.87
11_71.6	iHS	11:71783056	16.02	R	-	0.46	0.96	0.91	0.91	0.98	0.88	0.91	0.74	0.73	0.58	1.00	0.76	0.87
12_44.4	iHS	12:44488039	19.32	I,NMD,N	TMEM117	0.27	0.73	0.89	0.89	0.89	0.84	1.00	0.85	0.92	0.96	0.97	0.90	0.88
12_44.4	iHS	12:44532975	19.2	I,NMD,N	TMEM117	0.29	0.73	0.89	0.89	0.89	0.84	1.00	0.85	0.92	0.96	0.95	0.90	0.88
13_65.4	iHS	13:65512989	21.5	IG	-	0.06	0.29	0.13	0.20	0.33	0.52	0.44	0.37	0.37	0.52	0.42	0.59	0.53
15_48.6	nSL	15:48629884	19.44	I,NMD,N,U	DUT	0.33	0.75	0.86	0.85	0.80	0.57	0.76	0.56	0.26	0.48	0.47	0.33	0.51
15_48.6	nSL	15:48772936	16.19	I,NMD	FBN1	0.35	0.69	0.81	0.68	0.63	0.54	0.47	0.41	0.26	0.64	0.68	0.47	0.41
16_27.6	TD	16:27687430	17.63	I,NMD,N	KIAA0556	0.54	0.87	0.88	0.89	0.91	0.86	0.91	0.91	0.92	0.76	0.89	0.90	0.92
16_31	iHS	16:31099000	26.2	M,NC	PRSS53	0.00	0.00	0.00	0.00	0.00	0.57	0.00	0.00	0.00	0.00	0.00	0.05	0.00
16_31	iHS	16:31099000	26.2	D	VKORC1	0.00	0.00	0.00	0.00	0.00	0.57	0.00	0.00	0.00	0.00	0.00	0.05	0.00
16_31	iHS	16:31099000	26.2	D	ZNF646	0.00	0.00	0.00	0.00	0.00	0.57	0.00	0.00	0.00	0.00	0.00	0.05	0.00
17_59.2	nSL, iHS	17:59253284	15.81	I,N	BCAS3	0.23	0.73	0.75	0.82	0.83	0.89	0.76	0.81	0.65	0.50	0.92	0.69	0.67
18_23.8	iHS	18:23877175	15.03	I,NMD,D	TAF4B	0.06	0.48	0.39	0.42	0.59	0.70	0.79	0.75	0.68	0.72	0.50	0.79	0.81
18_23.8	iHS	18:23877175	15.03	R	-	0.06	0.48	0.39	0.42	0.59	0.70	0.79	0.75	0.68	0.72	0.50	0.79	0.81
18_23.8	iHS	18:23877175	15.03	D	U3	0.06	0.48	0.39	0.42	0.59	0.70	0.79	0.75	0.68	0.72	0.50	0.79	0.81
18_67.6	TD	18:67733234	15.83	I,NMD	RTTN	0.31	0.98	0.92	0.91	0.93	0.88	0.97	0.96	0.98	1.00	0.66	0.98	0.93
2_158.4	nSL, iHS	2:158574077	21.7	IG	-	0.17	0.81	0.67	0.77	0.72	0.86	0.56	0.68	0.61	0.76	0.95	0.76	0.72
2_158.4	nSL, iHS	2:158503739	15.02	IG	-	0.38	0.81	0.83	0.87	0.87	0.86	0.94	0.90	1.00	1.00	0.89	0.98	0.89
2_178	nSL	2:178073075	17.52	U	MIR4444-2	0.29	0.75	0.80	0.86	0.78	0.86	0.62	0.72	0.79	0.74	0.92	0.83	0.73
2_178	nSL	2:178073075	17.52	U	HNRNPA3	0.29	0.75	0.80	0.86	0.78	0.86	0.62	0.72	0.79	0.74	0.92	0.83	0.73
2_178	nSL	2:178066599	17.24	D	KRT8P40	0.29	0.75	0.80	0.86	0.78	0.86	0.62	0.72	0.79	0.74	0.92	0.83	0.73
2_206.2	iHS	2:206227030	20.5	I,N	PARD3B	0.15	0.71	0.73	0.75	0.76	0.82	0.79	0.78	0.76	0.80	1.00	0.59	0.56
2_206.2	iHS	2:206227030	20.5	I,N	PARD3B	0.15	0.71	0.73	0.75	0.76	0.82	0.79	0.78	0.76	0.80	1.00	0.59	0.56
2_206.2	iHS	2:206208049	16.17	I,U	PARD3B	0.10	0.69	0.66	0.72	0.70	0.82	0.79	0.76	0.77	0.80	1.00	0.59	0.57
2_206.2	iHS	2:206217855	15.04	I,N	PARD3B	0.15	0.71	0.73	0.75	0.76	0.82	0.79	0.78	0.77	0.80	1.00	0.59	0.57
3_109.8	nSL, iHS	3:109858111	20.9	IG	-	0.19	0.75	0.53	0.75	0.80	0.82	0.85	0.76	0.84	0.82	0.79	0.78	0.89
3_142.2	nSL	3:142316443	20.6	I	PLS1	0.00	0.42	0.41	0.49	0.63	0.59	0.62	0.51	0.60	0.64	0.79	0.55	0.59
3_167.4	nSL	3:167454126	15.66	U	PDCD10	0.08	0.58	0.61	0.52	0.54	0.75	0.62	0.62	0.60	0.52	0.84	0.55	0.53
3_167.4	nSL	3:167454126	15.66	R	-	0.08	0.58	0.61	0.52	0.54	0.75	0.62	0.62	0.60	0.52	0.84	0.55	0.53
3_167.4	nSL	3:167454126	15.66	I	SERPINI1	0.08	0.58	0.61	0.52	0.54	0.75	0.62	0.62	0.60	0.52	0.84	0.55	0.53
3_38.4	nSL, iHS	3:38529825	16.7	3,D	ACVR2B	0.02	0.48	0.64	0.56	0.41	0.61	0.62	0.34	0.40	0.44	0.39	0.31	0.19
3_38.4	nSL, iHS	3:38496193	15.22	R	-	0.02	0.44	0.61	0.46	0.37	0.61	0.62	0.32	0.40	0.44	0.39	0.31	0.18
3_38.4	nSL, iHS	3:38496193	15.22	NC	ACVR2B-AS1	0.02	0.44	0.61	0.46	0.37	0.61	0.62	0.32	0.40	0.44	0.39	0.31	0.18
3_38.4	nSL, iHS	3:38496193	15.22	I,N	ACVR2B	0.02	0.44	0.61	0.46	0.37	0.61	0.62	0.32	0.40	0.44	0.39	0.31	0.18
4_107.6	TD	4:107611250	16.58	IG	-	0.75	0.81	0.72	0.76	0.89	0.86	0.91	0.96	0.97	0.98	0.97	0.91	0.94
4_107.6	TD	4:107620115	16.28	IG	-	0.65	0.92	0.75	0.81	0.89	0.89	0.91	0.96	0.97	0.98	0.97	0.91	0.94
4_107.6	TD	4:107625430	15.42	IG	-	0.65	0.94	0.75	0.81	0.89	0.89	0.91	0.96	0.97	0.98	0.97	0.91	0.94
4_71.6	nSL, iHS, TD	4:71619607	20.7	I	RUFY3	0.40	0.90	0.92	0.96	0.96	0.88	0.94	0.90	0.85	0.44	0.84	0.83	0.82
4_95.4	nSL	4:95438714	17.41	I,N	PDLIM5	0.35	0.71	0.66	0.71	0.65	0.80	0.62	0.66	0.73	0.90	0.95	0.76	0.64
5_117.6	nSL, iHS	5:117679670	18.68	IG	-	0.23	0.48	0.50	0.72	0.63	0.80	0.82	0.62	0.77	0.96	0.89	0.84	0.78
5_121.2	nSL, iHS	5:121249324	22.1	IG	-	0.31	0.63	0.69	0.69	0.63	0.71	0.76	0.88	0.87	0.76	0.92	0.79	0.76
9_108	TD	9:108037717	15.06	I,NMD	SLC44A1	0.27	0.85	0.86	0.85	0.78	0.79	0.68	0.72	0.66	0.50	0.58	0.88	0.87
9_126.4	nSL, iHS	9:126434103	18.34	I,N	DENND1A	0.08	0.73	0.66	0.53	0.48	0.73	0.68	0.76	0.76	0.76	0.71	0.76	0.74
9_131.8	nSL	9:131877474	20.3	I,U	PPP2R4	0.12	0.65	0.75	0.67	0.65	0.86	0.68	0.66	0.68	0.72	0.68	0.57	0.56
9_131.8	nSL	9:131877474	20.3	U	CRAT	0.12	0.65	0.75	0.67	0.65	0.86	0.68	0.66	0.68	0.72	0.68	0.57	0.56
9_24.2	TD	9:24219902	22.5	IG	-	0.69	0.73	0.80	0.73	0.74	0.89	0.62	0.69	0.61	0.76	0.39	0.72	0.73

Table C.8 Full list of significant SNPs from the window-based tests in the SOA population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_100.4	iHS, TD	1:100547994	20.7	D	SASS6	0.02	0.25	0.44	0.42	0.59	0.21	0.65	0.57	0.74	0.80	0.76	0.60	0.52
1_100.4	iHS, TD	1:100547994	20.7	3	HIAT1	0.02	0.25	0.44	0.42	0.59	0.21	0.65	0.57	0.74	0.80	0.76	0.60	0.52
1_100.4	iHS, TD	1:100575933	18.62	M,NC	SASS6	0.02	0.25	0.44	0.42	0.59	0.21	0.65	0.57	0.74	0.80	0.76	0.60	0.52
1_31.4	iHS	1:31504162	15.21	I,NMD,N	PUM1	0.31	0.15	0.17	0.27	0.22	0.32	0.47	0.35	0.66	0.48	0.00	0.59	0.39
1_73.4	iHS	1:73498942	18.1	IG	-	0.08	0.42	0.67	0.52	0.50	0.30	0.44	0.26	0.19	0.58	0.66	0.21	0.17
11_27.2	nSL	11:27337345	18.24	IG	-	0.19	0.29	0.23	0.25	0.43	0.39	0.47	0.60	0.79	0.76	0.74	0.33	0.36
12_2.6	nSL	12:2759141	16.2	I,N,U	CACNA1C	0.21	0.62	0.66	0.77	0.72	0.71	0.79	0.69	0.82	0.96	0.61	0.72	0.52
12_89	TD	12:89018883	15.61	R,IG	-	0.15	0.94	0.95	0.98	0.87	0.93	0.82	0.59	0.60	0.66	1.00	0.67	0.69
12_90.8	nSL	12:90821635	16.78	IG	-	0.33	0.19	0.23	0.17	0.37	0.18	0.53	0.28	0.53	0.32	0.11	0.43	0.49
13_39.6	nSL	13:39687431	15.33	IG	-	0.15	0.58	0.73	0.72	0.70	0.57	0.79	0.56	0.53	0.76	0.89	0.60	0.63
14_31.8	nSL	14:31858209	22.1	M,SR,U	HEATR5A	0.06	0.29	0.44	0.43	0.48	0.29	0.71	0.49	0.50	0.42	0.21	0.29	0.21
16_71.6	iHS	16:71727571	15.97	D	SNORA70D	0.29	0.27	0.34	0.35	0.39	0.29	0.62	0.69	0.66	0.74	0.63	0.74	0.71
16_71.6	iHS	16:71727571	15.97	I,N	PHLPP2	0.29	0.27	0.34	0.35	0.39	0.29	0.62	0.69	0.66	0.74	0.63	0.74	0.71
16_71.6	iHS	16:71660310	15.75	R	-	0.13	0.27	0.33	0.35	0.39	0.27	0.65	0.68	0.66	0.74	0.63	0.71	0.68
16_71.6	iHS	16:71660310	15.75	M,U	MARVELD3	0.13	0.27	0.33	0.35	0.39	0.27	0.65	0.68	0.66	0.74	0.63	0.71	0.68
16_71.6	iHS	16:71760988	15.7	D	AP1G1	0.13	0.27	0.34	0.34	0.39	0.27	0.65	0.68	0.66	0.72	0.63	0.74	0.71
16_71.6	iHS	16:71760988	15.7	U	PHLPP2	0.13	0.27	0.34	0.34	0.39	0.27	0.65	0.68	0.66	0.72	0.63	0.74	0.71
16_71.6	iHS	16:71760988	15.7	R	-	0.13	0.27	0.34	0.34	0.39	0.27	0.65	0.68	0.66	0.72	0.63	0.74	0.71
16_71.6	iHS	16:71775080	15.37	I,NMD,D	AP1G1	0.13	0.27	0.34	0.35	0.39	0.27	0.65	0.68	0.66	0.74	0.63	0.74	0.71
19_32.2	TD	19:32263763	15.63	IG	-	0.69	0.81	0.81	0.86	0.74	0.77	0.88	0.75	0.76	0.74	0.92	0.67	0.64
2_160.4	nSL, iHS	2:160493125	16.95	IG	-	0.46	0.31	0.25	0.39	0.43	0.41	0.62	0.49	0.56	0.72	0.13	0.45	0.21
2_160.4	nSL, iHS	2:160515935	15.06	IG	-	0.48	0.31	0.25	0.39	0.43	0.41	0.62	0.49	0.56	0.72	0.13	0.45	0.21
3_103.2	nSL	3:103358697	15.02	IG	-	0.56	0.52	0.66	0.70	0.74	0.36	0.56	0.38	0.53	0.62	0.58	0.38	0.39
3_142.2	iHS	3:142316443	20.6	I	PLS1	0.00	0.42	0.41	0.49	0.63	0.59	0.62	0.51	0.60	0.64	0.79	0.55	0.59
3_142.2	iHS	3:142384363	19.45	I,N	PLS1	0.02	0.42	0.34	0.40	0.57	0.63	0.65	0.51	0.66	0.76	0.87	0.69	0.62
3_164.2	nSL	3:164274245	21.8	IG	-	0.02	0.56	0.38	0.37	0.46	0.43	0.76	0.71	0.73	0.66	0.63	0.55	0.49
3_184.6	iHS	3:184727093	16.9	I,N	VPS8	0.00	0.21	0.31	0.26	0.46	0.45	0.44	0.46	0.55	0.48	0.55	0.66	0.63
3_43.2	TD	3:43325835	17.54	U	SNRK	0.12	0.44	0.47	0.64	0.61	0.61	0.88	0.78	0.89	0.66	0.34	0.79	0.84
4_168.6	iHS	4:168771488	15.03	IG	-	0.29	0.48	0.53	0.63	0.61	0.59	0.79	0.81	0.97	0.92	0.84	0.79	0.80
4_41.4	iHS	4:41526749	16.3	I,N	LIMCH1	0.13	0.56	0.47	0.63	0.72	0.41	0.82	0.68	0.89	0.84	0.76	0.67	0.59
5_109.2	nSL	5:109288645	15.66	IG	-	0.56	0.94	0.95	0.90	0.74	0.86	0.79	0.60	0.65	0.88	0.84	0.76	0.76
5_176.4	nSL	5:176402401	15.1	S,NMD,NC,D	UIMC1	0.13	0.46	0.50	0.46	0.63	0.57	0.56	0.44	0.63	0.68	0.63	0.43	0.57
6_18.8	nSL, iHS	6:18962766	21.3	IG	-	0.06	0.40	0.48	0.49	0.67	0.30	0.79	0.68	0.87	0.60	0.37	0.62	0.56
6_18.8	nSL, iHS	6:18865424	19.71	IG	-	0.15	0.44	0.58	0.54	0.72	0.38	0.82	0.69	0.87	0.60	0.39	0.66	0.62
6_18.8	nSL, iHS	6:18946415	19.37	IG	-	0.27	0.44	0.58	0.53	0.72	0.39	0.82	0.69	0.87	0.60	0.39	0.66	0.59
6_18.8	nSL, iHS	6:18930434	15.8	IG	-	0.23	0.44	0.58	0.53	0.72	0.39	0.82	0.69	0.87	0.60	0.39	0.66	0.59
6_18.8	nSL, iHS	6:18906867	15.73	IG	-	0.23	0.44	0.58	0.53	0.72	0.39	0.82	0.69	0.87	0.60	0.39	0.66	0.59
6_53.6	nSL, iHS	6:53700093	17.17	I,NMD	LRRC1	0.40	0.25	0.14	0.33	0.33	0.34	0.44	0.16	0.21	0.24	0.21	0.31	0.23
6_86.6	nSL, iHS	6:86729103	21.1	IG	-	0.21	0.23	0.33	0.37	0.50	0.14	0.44	0.28	0.34	0.26	0.08	0.21	0.50
7_144	nSL	7:144179061	20.2	I,NMD,N	TPK1	0.19	0.29	0.19	0.25	0.57	0.48	0.79	0.79	0.89	0.96	0.82	0.83	0.73
7_144	nSL	7:144159670	17.75	I,NMD,N	TPK1	0.46	0.31	0.20	0.25	0.57	0.48	0.79	0.79	0.89	0.96	0.82	0.83	0.74
7_86.4	TD	7:86407567	18.42	I	GRM3	0.15	0.54	0.72	0.70	0.63	0.71	0.82	0.72	0.69	0.82	0.92	0.76	0.86
7_86.4	TD	7:86471913	15.23	I	GRM3	0.29	0.71	0.72	0.66	0.65	0.71	0.82	0.72	0.74	0.92	0.92	0.76	0.81
8_67.6	iHS	8:67609366	15.79	I,N	C8orf44-SGK3	0.02	0.42	0.55	0.58	0.61	0.63	0.65	0.78	0.66	0.80	0.84	0.78	0.62
8_67.8	iHS	8:67946941	20.9	I,N	PPP1R42	0.52	1.00	0.98	0.94	0.96	0.93	0.82	0.96	0.89	0.86	0.87	0.93	0.91
9_13.8	nSL	9:13904860	15.41	IG	-	0.27	0.50	0.41	0.58	0.72	0.71	0.88	0.84	0.89	0.96	1.00	0.86	0.76

Table C.9 Full list of significant SNPs from the window-based tests in the WSI population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_93	TD	1:93115127	16.46	I,N	EVI5	0.48	0.62	0.72	0.70	0.76	0.82	0.94	0.90	0.95	0.94	0.87	0.95	0.77
1_93	TD	1:93115127	16.46	U	HMGB3P9	0.48	0.62	0.72	0.70	0.76	0.82	0.94	0.90	0.95	0.94	0.87	0.95	0.77
1_93	TD	1:93106871	15.88	I,N	EVI5	0.83	0.65	0.72	0.69	0.76	0.93	0.94	0.90	0.95	0.94	0.87	0.95	0.86
1_93	TD	1:93165207	15.84	D	RNU4-59P	0.29	0.62	0.72	0.69	0.76	0.79	0.94	0.90	0.95	0.94	0.87	0.93	0.76
1_93	TD	1:93165207	15.84	I,D	EVI5	0.29	0.62	0.72	0.69	0.76	0.79	0.94	0.90	0.95	0.94	0.87	0.93	0.76
1_93	TD	1:93192134	15.11	I,N	EVI5	0.29	0.62	0.72	0.69	0.76	0.77	0.94	0.90	0.95	0.94	0.87	0.93	0.76
10_95	nSL	10:95022185	15.53	R,IG	-	0.37	0.10	0.28	0.29	0.46	0.36	0.71	0.66	0.79	0.84	0.68	0.74	0.57
10_95	nSL	10:95022185	15.53	R,IG	-	0.37	0.10	0.28	0.29	0.46	0.36	0.71	0.66	0.79	0.84	0.68	0.74	0.57
11_17.6	TD	11:17709767	16.28	R,IG	-	0.13	0.56	0.44	0.51	0.52	0.34	0.59	0.63	0.69	0.62	0.37	0.83	0.80
11_60.8	nSL	11:60893235	31	M	CD5	0.56	0.42	0.53	0.58	0.59	0.73	0.71	0.88	0.85	1.00	0.89	0.98	0.96
11_60.8	nSL	11:60893235	31	D	VPS37C	0.56	0.42	0.53	0.58	0.59	0.73	0.71	0.88	0.85	1.00	0.89	0.98	0.96
12_124	iHS	12:124082240	17.66	U	DDX55	0.35	0.46	0.59	0.62	0.61	0.59	0.94	0.79	0.85	0.98	0.79	0.84	0.77
12_124	iHS	12:124082240	17.66	3,D	TMED2	0.35	0.46	0.59	0.62	0.61	0.59	0.94	0.79	0.85	0.98	0.79	0.84	0.77
13_105	nSL	13:105183214	19.44	IG	-	0.15	0.33	0.39	0.45	0.57	0.55	0.71	0.75	0.76	0.58	0.76	0.60	0.56
13_35.6	TD	13:35784929	17.35	I	NBEA	0.90	0.75	0.88	0.85	0.87	0.79	0.94	0.90	1.00	0.98	1.00	0.95	0.89
13_63.2	nSL, iHS	13:63326573	16.99	IG	-	0.13	0.04	0.14	0.11	0.13	0.21	0.53	0.66	0.66	0.80	0.39	0.60	0.51
13_72	TD	13:72031460	21.2	I	DACH1	0.33	0.77	0.89	0.92	0.91	0.96	0.97	0.88	0.95	0.84	0.95	0.88	0.76
14_69.6	iHS	14:69641807	18.2	IG	-	0.02	0.27	0.23	0.30	0.43	0.54	0.74	0.76	0.90	0.80	0.47	0.76	0.88
14_69.6	iHS	14:69617972	15.85	R	-	0.10	0.37	0.30	0.41	0.50	0.61	0.74	0.85	0.90	0.94	0.92	0.86	0.88
14_69.6	iHS	14:69617972	15.85	I,N,NMD	DCAF5	0.10	0.37	0.30	0.41	0.50	0.61	0.74	0.85	0.90	0.94	0.92	0.86	0.88
15_63.8	iHS	15:63953153	16.03	I	HERC1	0.13	0.12	0.17	0.19	0.39	0.32	0.59	0.85	0.92	0.80	0.79	0.83	0.86
15_64	iHS	15:64012859	15.69	I,N	HERC1	0.13	0.23	0.27	0.22	0.43	0.43	0.59	0.85	0.92	0.80	0.79	0.83	0.86
16_65.6	nSL	16:65737243	18.83	IG	-	0.35	0.46	0.55	0.49	0.54	0.57	0.59	0.90	0.84	0.86	0.97	0.93	0.98
17_20	nSL, iHS	17:20038273	16.2	I,N	SPECC1	0.31	0.38	0.47	0.42	0.52	0.75	0.62	0.81	0.84	0.78	0.61	0.78	0.81
17_43.8	TD	17:43880047	21.8	I,NMD,N,U	CRHR1	0.27	0.42	0.41	0.46	0.59	0.61	0.94	0.84	0.95	0.92	0.82	0.90	0.82
17_44.2	TD	17:44240181	16.9	I,N	KANSL1	0.00	0.06	0.00	0.06	0.28	0.05	0.44	0.49	0.52	0.78	0.58	0.38	0.29
2_109	iHS	2:109196763	17.87	R	-	0.08	0.33	0.33	0.45	0.52	0.39	0.50	0.78	0.85	0.88	0.95	0.79	0.80
2_109	iHS	2:109196763	17.87	I	LIMS1	0.08	0.33	0.33	0.45	0.52	0.39	0.50	0.78	0.85	0.88	0.95	0.79	0.80
2_109.4	nSL, iHS	2:109513601	19.61	M	EDAR	0.00	0.02	0.00	0.02	0.22	0.00	0.59	0.71	0.97	0.92	0.95	0.76	0.72
2_17.4	nSL	2:17409594	16.1	IG	-	0.00	0.06	0.03	0.08	0.13	0.27	0.44	0.53	0.65	0.56	0.47	0.60	0.74
2_178.8	iHS	2:178874907	17.77	I	PDE11A	0.02	0.37	0.38	0.39	0.35	0.32	0.47	0.72	0.73	0.70	1.00	0.53	0.61
2_86.4	nSL	2:86400824	24.4	M,NMD,U,NC	IMMT	0.02	0.37	0.44	0.33	0.48	0.61	0.47	0.66	0.48	0.72	0.61	0.50	0.37
2_86.4	nSL	2:86558376	18.43	I,N	REEP1	0.06	0.40	0.41	0.39	0.52	0.54	0.62	0.76	0.82	0.76	0.76	0.48	0.38
2_86.4	nSL	2:86553250	16.18	I,N	REEP1	0.06	0.40	0.45	0.44	0.57	0.54	0.65	0.76	0.82	0.76	0.76	0.48	0.38
2_86.6	nSL, iHS	2:86635502	19.75	IG	-	0.29	0.52	0.59	0.62	0.63	0.43	0.74	0.81	0.81	0.80	0.76	0.76	0.49
2_86.6	nSL, iHS	2:86743708	16.76	I,N,U	CHMP3	0.08	0.42	0.47	0.49	0.61	0.48	0.65	0.76	0.73	0.68	0.58	0.67	0.56
2_86.6	nSL, iHS	2:86743708	16.76	U	RNU6-640P	0.08	0.42	0.47	0.49	0.61	0.48	0.65	0.76	0.73	0.68	0.58	0.67	0.56
2_86.6	nSL, iHS	2:86743708	16.76	I	RNF103-CHMP3	0.08	0.42	0.47	0.49	0.61	0.48	0.65	0.76	0.73	0.68	0.58	0.67	0.56
2_86.6	nSL, iHS	2:86732576	15.06	3,D	CHMP3	0.06	0.42	0.47	0.49	0.61	0.48	0.65	0.76	0.73	0.68	0.58	0.67	0.56
2_86.6	nSL, iHS	2:86732576	15.06	D	RNF103-CHMP3	0.06	0.42	0.47	0.49	0.61	0.48	0.65	0.76	0.73	0.68	0.58	0.67	0.56
3_134.2	iHS	3:134278270	20.5	SG,SR,M,I,N,D	CEP63	0.02	0.29	0.34	0.38	0.50	0.34	0.50	0.69	0.73	0.72	0.92	0.53	0.54
4_13.2	TD	4:13211708	15.17	IG	-	0.56	0.60	0.64	0.73	0.67	0.54	0.82	0.87	0.97	1.00	0.84	0.97	0.98
4_140.2	iHS	4:140359958	17.09	U	ACA64	0.21	0.60	0.52	0.54	0.50	0.43	0.65	0.47	0.66	0.52	0.95	0.57	0.37
4_160.4	nSL, iHS	4:160483950	17.47	IG	-	0.46	0.60	0.38	0.41	0.50	0.46	0.65	0.74	0.68	0.72	0.92	0.78	0.71
4_41.8	nSL, iHS	4:41951250	19.24	I,NMD	TMEM33	0.54	0.87	0.83	0.90	0.91	0.89	0.97	0.90	1.00	0.98	0.97	0.95	0.91
4_42	nSL, iHS	4:42073472	15.41	I,N,NMD	SLC30A9	0.15	0.79	0.78	0.79	0.85	0.93	0.91	0.90	0.98	0.94	0.97	0.95	0.94
4_71.6	TD	4:71619607	20.7	I	RUFY3	0.40	0.90	0.92	0.96	0.96	0.88	0.94	0.90	0.85	0.44	0.84	0.83	0.82
5_103	nSL	5:103137013	18.55	IG	-	0.10	0.13	0.16	0.10	0.28	0.32	0.38	0.59	0.63	0.36	0.08	0.67	0.47
5_103	nSL	5:103137013	18.55	IG	-	0.10	0.13	0.16	0.10	0.28	0.32	0.38	0.59	0.63	0.36	0.08	0.67	0.47
5_112.2	nSL	5:112214093	16.56	D,3,NC	REEP5	0.00	0.40	0.23	0.25	0.35	0.45	0.44	0.59	0.76	0.64	0.79	0.50	0.59
5_121.2	TD	5:121249324	22.1	IG	-	0.31	0.63	0.69	0.69	0.63	0.71	0.76	0.88	0.87	0.76	0.92	0.79	0.76
5_134.6	nSL	5:134665082	17.7	D	H2AFY	0.17	0.77	0.80	0.86	0.83	0.77	0.91	0.90	0.94	0.92	0.97	0.86	0.64
5_134.6	nSL	5:134665082	17.7	I,N	C5orf66	0.17	0.77	0.80	0.86	0.83	0.77	0.91	0.90	0.94	0.92	0.97	0.86	0.64
5_138	iHS	5:138162129	16.62	I,NMD,N,U,D	CTNNA1	0.35	0.67	0.67	0.65	0.65	0.77	0.74	0.90	0.97	0.88	0.68	0.93	0.89
5_8.8	iHS	5:8977512	20.4	IG	-	0.46	0.73	0.72	0.71	0.80	0.75	0.94	0.90	0.84	0.98	0.95	0.93	0.79
5_97	nSL	5:97182715	15.55	IG	-	0.85	0.87	0.77	0.87	0.87	0.80	0.85	0.90	0.92	0.98	0.95	0.91	0.81
6_129.2	nSL	6:129350436	21.7	I,N	LAMA2	0.08	0.69	0.70	0.74	0.78	0.70	0.74	0.90	0.56	0.94	0.97	0.98	0.98
7_148.8	iHS	7:148827909	16.84	U	RN7SL521P	0.13	0.62	0.58	0.67	0.74	0.84	0.82	0.84	0.87	0.80	0.84	0.84	0.83
7_148.8	iHS	7:148827909	16.84	U	ZNF425	0.13	0.62	0.58	0.67	0.74	0.84	0.82	0.84	0.87	0.80	0.84	0.84	0.83
7_148.8	iHS	7:148827909	16.84	I	ZNF398	0.13	0.62	0.58	0.67	0.74	0.84	0.82	0.84	0.87	0.80	0.84	0.84	0.83
7_98.8	nSL, TD	7:98850491	16.05	I,N,NC	MYH16	0.15	0.83	0.88	0.93	0.91	0.54	0.85	0.84	0.90	0.94	0.82	0.79	0.69
9_126.4	nSL	9:126434103	18.34	I,N	DENND1A	0.08	0.73	0.66	0.53	0.48	0.73	0.68	0.76	0.76	0.76	0.71	0.76	0.74
9_126.4	nSL	9:126558810	18.16	I,N	DENND1A	0.23	0.88	0.83	0.69	0.67	0.79	0.82	0.90	0.82	0.82	0.84	0.93	0.98
9_126.4	nSL	9:126558843	18.06	I,N	DENND1A	0.23	0.88	0.83	0.69	0.67	0.79	0.82	0.90	0.82	0.82	0.84	0.93	0.98
9_126.4	nSL	9:126529097	16.47	I,N	DENND1A	0.48	0.88	0.86	0.71	0.74	0.89	0.79	0.84	0.76	0.82	0.84	0.93	0.98

Table C.10 Full list of significant SNPs from the window-based tests in the SSI population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_181.2	nSL, iHS	1:181232399	22.2	IG	-	0.19	0.37	0.52	0.43	0.61	0.48	0.79	0.69	0.85	0.92	0.63	0.71	0.69
1_181.2	nSL, iHS	1:181278022	15.65	IG	-	0.15	0.37	0.52	0.43	0.61	0.46	0.79	0.69	0.85	0.92	0.61	0.71	0.69
10_22.2	iHS	10:22241212	17.76	I,NMD	DNAJC1	0.42	0.58	0.59	0.71	0.63	0.82	0.65	0.85	0.85	1.00	0.74	0.97	0.87
10_78.4	iHS	10:78403288	16.45	R,IG	-	0.58	0.35	0.48	0.47	0.43	0.34	0.41	0.40	0.40	0.26	0.24	0.47	0.48
10_95	nSL	10:95030756	20.3	IG	-	0.37	0.10	0.28	0.31	0.48	0.36	0.71	0.68	0.79	0.84	0.68	0.74	0.60
10_95	nSL	10:95022185	15.53	R,IG	-	0.37	0.10	0.28	0.29	0.46	0.36	0.71	0.66	0.79	0.84	0.68	0.74	0.57
10_95	nSL	10:95022185	15.53	R,IG	-	0.37	0.10	0.28	0.29	0.46	0.36	0.71	0.66	0.79	0.84	0.68	0.74	0.57
12_100	iHS	12:100103164	18.51	I,U	ANKS1B	0.37	0.44	0.17	0.25	0.35	0.25	0.44	0.46	0.63	0.36	0.32	0.38	0.37
12_44.2	nSL	12:44204068	19.75	U	TWF1	0.00	0.27	0.34	0.28	0.26	0.39	0.41	0.56	0.73	0.50	0.13	0.57	0.46
12_51	nSL	12:51053515	20.7	I,N	DIP2B	0.33	0.17	0.20	0.34	0.48	0.41	0.44	0.32	0.45	0.34	0.42	0.36	0.51
12_51	nSL	12:51053515	20.7	D	RNU6-238P	0.33	0.17	0.20	0.34	0.48	0.41	0.44	0.32	0.45	0.34	0.42	0.36	0.51
13_26.8	iHS	13:26906088	15.92	I	CDK8	0.12	0.63	0.58	0.42	0.52	0.57	0.29	0.28	0.50	0.50	0.47	0.33	0.22
13_42	nSL	13:42182736	17.02	I	VWA8	0.00	0.15	0.08	0.08	0.20	0.27	0.35	0.31	0.47	0.16	0.58	0.21	0.16
13_63.6	nSL, iHS	13:63682130	17.36	IG	-	0.65	0.33	0.34	0.34	0.39	0.45	0.59	0.78	0.89	0.84	0.63	0.83	0.87
13_75.6	nSL	13:75717402	15.07	IG	-	0.31	0.60	0.56	0.64	0.35	0.46	0.38	0.51	0.61	0.56	0.24	0.55	0.61
15_65.8	iHS	15:65898134	15.71	I,NMD,N	VWA9	0.10	0.15	0.17	0.32	0.30	0.36	0.32	0.43	0.53	0.58	0.24	0.74	0.64
15_65.8	iHS	15:65897863	15.19	I,N,NMD	VWA9	0.10	0.15	0.17	0.32	0.30	0.36	0.35	0.43	0.53	0.58	0.24	0.74	0.64
16_23.4	iHS	16:23563501	16.77	3,NMD,S,NC,D,U	EARS2	0.42	0.79	0.80	0.70	0.76	0.71	0.65	0.56	0.76	0.82	0.55	0.66	0.72
16_23.4	iHS	16:23563501	16.77	U	UBFD1	0.42	0.79	0.80	0.70	0.76	0.71	0.65	0.56	0.76	0.82	0.55	0.66	0.72
18_9.2	iHS	18:9255982	22.9	3,NMD,M,D	ANKRD12	0.38	0.71	0.66	0.63	0.52	0.80	0.50	0.53	0.61	0.56	0.79	0.67	0.76
2_17.6	TD	2:17775840	15.19	I	VSNL1	0.06	0.21	0.19	0.29	0.46	0.41	0.76	0.81	0.89	0.92	0.79	0.74	0.78
2_179	iHS	2:179124117	18.85	I,N	OSBPL6	0.08	0.33	0.45	0.39	0.30	0.30	0.56	0.74	0.82	0.88	0.82	0.60	0.43
2_179	iHS	2:179049910	15.98	IG	-	0.25	0.27	0.41	0.38	0.43	0.34	0.53	0.71	0.77	0.86	0.74	0.59	0.51
2_190.2	TD	2:190344985	17.25	D	WDR75	0.23	0.67	0.66	0.61	0.76	0.55	0.85	0.75	0.85	0.70	0.63	0.60	0.66
2_214.6	nSL, iHS	2:214726989	15.71	I,NMD	SPAG16	0.12	0.44	0.44	0.40	0.43	0.34	0.53	0.69	0.85	0.54	0.53	0.45	0.38
3_135.8	TD	3:135980673	15.38	I,N,NMD	PCCB	0.02	0.29	0.20	0.30	0.33	0.39	0.62	0.57	0.73	0.64	0.50	0.59	0.46
4_159.2	TD	4:159375206	15.01	I	RXFP1	0.02	0.23	0.33	0.39	0.46	0.32	0.79	0.63	0.77	0.68	0.45	0.69	0.81
4_64.8	nSL, iHS	4:64831920	15.3	IG	-	0.27	0.31	0.30	0.38	0.26	0.29	0.53	0.31	0.53	0.46	0.68	0.16	0.12
5_121.2	TD	5:121249324	22.1	IG	-	0.31	0.63	0.69	0.69	0.63	0.71	0.76	0.88	0.87	0.76	0.92	0.79	0.76
5_176.4	nSL	5:176402401	15.1	S,NMD,NC,D	UIMC1	0.13	0.46	0.50	0.46	0.63	0.57	0.56	0.44	0.63	0.68	0.63	0.43	0.57
5_97.8	nSL	5:97981935	15.37	IG	-	0.10	0.46	0.44	0.53	0.59	0.25	0.74	0.66	0.84	0.92	0.87	0.62	0.70
6_123.6	nSL	6:123702499	15.49	I	TRDN	0.23	0.33	0.38	0.45	0.50	0.77	0.62	0.74	0.85	0.82	0.42	0.74	0.50
6_134.6	nSL	6:134791075	19.28	I,N	LINC01010	0.42	0.21	0.20	0.22	0.26	0.20	0.50	0.51	0.81	0.56	0.58	0.52	0.44
6_37.2	nSL	6:37252210	15.46	S	TBC1D22B	0.42	0.67	0.61	0.64	0.80	0.82	0.65	0.76	0.89	0.74	0.74	0.83	0.76
6_87.2	nSL	6:87211767	15.16	IG	-	0.13	0.29	0.36	0.30	0.41	0.32	0.47	0.53	0.60	0.14	0.21	0.40	0.46
7_100.8	nSL	7:100977861	16.47	IG	-	0.33	0.06	0.19	0.13	0.15	0.16	0.41	0.40	0.69	0.36	0.11	0.36	0.32
7_34	nSL	7:34149593	16.23	I,N	BMPER	0.38	0.71	0.70	0.60	0.78	0.80	0.76	0.81	0.87	0.96	0.95	0.83	0.91
7_34	nSL	7:34147159	15.02	I,N	BMPER	0.19	0.69	0.70	0.59	0.78	0.68	0.76	0.78	0.85	0.86	0.95	0.76	0.87
7_34	nSL	7:34147159	15.02	R	-	0.19	0.69	0.70	0.59	0.78	0.68	0.76	0.78	0.85	0.86	0.95	0.76	0.87

Table C.11 Full list of significant SNPs from the window-based tests in the CSI population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_119.4	nSL, iHS, TD	1:119579276	16.11	I,N	WARS2	0.00	0.06	0.14	0.23	0.26	0.04	0.44	0.43	0.60	0.78	0.84	0.31	0.08
1_174.6	iHS	1:174729671	15.59	U	BANF1P4	0.08	0.06	0.06	0.07	0.22	0.04	0.29	0.29	0.58	0.64	0.00	0.26	0.38
1_174.6	iHS	1:174729671	15.59	I,N,NMD	RABGAP1L	0.08	0.06	0.06	0.07	0.22	0.04	0.29	0.29	0.58	0.64	0.00	0.26	0.38
1_216.2	nSL	1:216297936	16.35	I	USH2A	0.13	0.37	0.36	0.36	0.39	0.59	0.71	0.68	0.81	0.88	0.71	0.66	0.73
1_216.2	nSL	1:216295188	16.07	I	USH2A	0.15	0.37	0.36	0.36	0.39	0.59	0.71	0.68	0.81	0.88	0.71	0.66	0.73
1_39	nSL	1:39135474	18.27	IG	-	0.29	0.62	0.70	0.72	0.59	0.52	0.32	0.47	0.53	0.84	0.47	0.67	0.63
10_93.4	TD	10:93564315	16.84	I	TNKS2	0.31	0.37	0.36	0.42	0.37	0.59	0.62	0.47	0.79	0.90	0.63	0.66	0.42
10_93.4	TD	10:93564315	16.84	D	SRP9P1	0.31	0.37	0.36	0.42	0.37	0.59	0.62	0.47	0.79	0.90	0.63	0.66	0.42
11_121.4	nSL	11:121593385	17.14	R,IG	-	0.10	0.67	0.77	0.84	0.65	0.71	0.82	0.87	0.90	0.90	0.71	0.78	0.73
11_121.4	nSL	11:121568262	16.56	IG	-	0.10	0.40	0.61	0.58	0.48	0.59	0.76	0.82	0.94	0.90	0.71	0.81	0.80
11_121.4	nSL	11:121570520	16.39	IG	-	0.00	0.65	0.73	0.79	0.63	0.68	0.82	0.85	0.94	0.90	0.71	0.81	0.80
11_121.6	nSL	11:121661507	17.23	IG	-	0.10	0.63	0.72	0.85	0.61	0.68	0.65	0.82	0.82	0.82	0.68	0.71	0.79
11_121.6	nSL	11:121641790	17.04	IG	-	0.10	0.63	0.72	0.85	0.61	0.68	0.62	0.81	0.81	0.86	0.71	0.71	0.77
11_67.2	iHS	11:67220015	20.2	M	GPR152	0.00	0.00	0.02	0.00	0.09	0.00	0.24	0.13	0.13	0.40	0.03	0.17	0.10
11_67.2	iHS	11:67220015	20.2	R	-	0.00	0.00	0.02	0.00	0.09	0.00	0.24	0.13	0.13	0.40	0.03	0.17	0.10
11_67.2	iHS	11:67220015	20.2	U,NC,5	CABP4	0.00	0.00	0.02	0.00	0.09	0.00	0.24	0.13	0.13	0.40	0.03	0.17	0.10
11_67.6	nSL, iHS	11:67798336	15.68	I,N,NMD,5,U	NDUFS8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.56	0.00	0.00	0.00
11_67.6	nSL, iHS	11:67798336	15.68	U	MIR4691	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.56	0.00	0.00	0.00
11_67.6	nSL, iHS	11:67798336	15.68	D	ALDH3B1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.56	0.00	0.00	0.00
11_67.6	nSL, iHS	11:67798336	15.68	R	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.56	0.00	0.00	0.00
15_86.8	nSL	15:86952892	19.1	I	AGBL1	0.00	0.04	0.09	0.04	0.22	0.14	0.29	0.28	0.45	0.60	0.47	0.45	0.46
16_28.4	iHS	16:28513403	23.5	D	APOBR	0.06	0.19	0.25	0.35	0.41	0.21	0.56	0.37	0.35	0.56	0.74	0.10	0.04
16_28.4	iHS	16:28513403	23.5	5,M	IL27	0.06	0.19	0.25	0.35	0.41	0.21	0.56	0.37	0.35	0.56	0.74	0.10	0.04
2_162.6	TD	2:162730312	17.75	I,N,NMD,NC	SLC4A10	0.40	0.62	0.53	0.59	0.63	0.57	0.71	0.87	0.97	0.84	0.42	0.86	0.69
3_135.8	TD	3:135980673	15.38	I,N,NMD	PCCB	0.02	0.29	0.20	0.30	0.33	0.39	0.62	0.57	0.73	0.64	0.50	0.59	0.46
3_38.4	nSL	3:38529825	16.7	3,D	ACVR2B	0.02	0.48	0.64	0.56	0.41	0.61	0.62	0.34	0.40	0.44	0.39	0.31	0.19
3_38.4	nSL	3:38496193	15.22	NC	ACVR2B-AS1	0.02	0.44	0.61	0.46	0.37	0.61	0.62	0.32	0.40	0.44	0.39	0.31	0.18
3_38.4	nSL	3:38496193	15.22	I,N	ACVR2B	0.02	0.44	0.61	0.46	0.37	0.61	0.62	0.32	0.40	0.44	0.39	0.31	0.18
3_38.4	nSL	3:38496193	15.22	R	-	0.02	0.44	0.61	0.46	0.37	0.61	0.62	0.32	0.40	0.44	0.39	0.31	0.18
5_54.6	TD	5:54734788	16.24	I,D	PPAP2A	0.42	0.77	0.72	0.63	0.50	0.66	0.53	0.49	0.76	0.74	0.63	0.43	0.53
5_54.6	TD	5:54734788	16.24	R	-	0.42	0.77	0.72	0.63	0.50	0.66	0.53	0.49	0.76	0.74	0.63	0.43	0.53
5_58.2	nSL	5:58331812	22.7	I,NMD,N	PDE4D	0.10	0.08	0.06	0.06	0.09	0.13	0.24	0.32	0.50	0.54	0.18	0.43	0.39
5_58.2	nSL	5:58297703	16.38	I,NMD,N,U	PDE4D	0.12	0.08	0.05	0.07	0.09	0.13	0.21	0.31	0.48	0.54	0.18	0.45	0.40
5_60.6	nSL	5:60624958	19.21	U	ZSWIM6	0.02	0.50	0.52	0.50	0.61	0.34	0.47	0.51	0.60	0.68	0.34	0.72	0.46
5_98	TD	5:98107341	18.03	R	-	0.00	0.40	0.47	0.54	0.57	0.34	0.65	0.68	0.74	0.88	0.87	0.66	0.73
5_98	TD	5:98107341	18.03	I,N,D,U	RGMB	0.00	0.40	0.47	0.54	0.57	0.34	0.65	0.68	0.74	0.88	0.87	0.66	0.73
5_98	TD	5:98107341	18.03	I,N	RGMB-AS1	0.00	0.40	0.47	0.54	0.57	0.34	0.65	0.68	0.74	0.88	0.87	0.66	0.73
6_55.4	iHS	6:55531009	18.19	IG	-	0.27	0.52	0.55	0.37	0.61	0.23	0.44	0.44	0.40	0.68	0.66	0.41	0.23
6_55.4	iHS	6:55531375	17.06	IG	-	0.27	0.52	0.55	0.37	0.61	0.23	0.44	0.44	0.40	0.68	0.66	0.41	0.23
8_104.6	iHS	8:104748336	17.83	I	RIMS2	0.31	0.10	0.20	0.28	0.24	0.09	0.15	0.07	0.05	0.44	0.00	0.03	0.00

Table C.12 Full list of significant SNPs from the window-based tests in the NSI population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_161.8	iHS	1:161876448	19.92	I	ATF6	0.29	0.08	0.06	0.15	0.15	0.09	0.35	0.34	0.60	0.62	0.76	0.29	0.33
1_161.8	iHS	1:161825850	18.35	I,N	ATF6	0.29	0.08	0.06	0.15	0.15	0.07	0.35	0.34	0.60	0.62	0.76	0.29	0.33
1_161.8	iHS	1:161806356	15.35	I,U	ATF6	0.12	0.04	0.06	0.14	0.13	0.07	0.35	0.32	0.60	0.62	0.76	0.29	0.33
1_28.8	iHS	1:28826587	15.39	3,D	PHACTR4	0.38	0.69	0.63	0.77	0.70	0.89	0.91	0.87	0.92	0.94	0.76	0.86	0.86
1_36.6	nSL	1:36638206	33	M,U,NC,D	MAP7D1	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.61	0.00	0.00
1_41.8	nSL	1:41910825	16.53	IG	-	0.27	0.29	0.36	0.41	0.35	0.36	0.18	0.19	0.24	0.32	0.47	0.43	0.36
1_50.8	TD	1:50851923	15.18	IG	-	0.17	0.42	0.45	0.62	0.67	0.61	0.76	0.75	0.90	0.96	0.89	0.84	0.67
1_53.4	nSL, iHS	1:53527132	22.6	U	PODN	0.35	0.17	0.03	0.08	0.24	0.11	0.29	0.25	0.44	0.14	0.61	0.41	0.29
1_53.4	nSL, iHS	1:53525644	20.4	U	PODN	0.00	0.02	0.03	0.03	0.02	0.05	0.03	0.07	0.15	0.06	0.61	0.14	0.19
1_53.4	nSL, iHS	1:53517863	17.15	D	SCP2	0.15	0.02	0.03	0.03	0.02	0.05	0.03	0.07	0.15	0.06	0.61	0.14	0.21
11_118.8	iHS	11:118845813	15.24	I,NMD,U	FOXR1	0.04	0.17	0.28	0.25	0.22	0.05	0.06	0.24	0.16	0.10	0.76	0.03	0.07
11_118.8	iHS	11:118845813	15.24	D	Y_RNA	0.04	0.17	0.28	0.25	0.22	0.05	0.06	0.24	0.16	0.10	0.76	0.03	0.07
11_64.4	nSL, iHS	11:64450479	19.31	I,N	NRXN2	0.12	0.25	0.13	0.13	0.07	0.16	0.21	0.21	0.15	0.32	0.74	0.21	0.28
11_64.8	iHS	11:64902002	21.3	R,TF	-	0.02	0.29	0.17	0.15	0.22	0.13	0.15	0.04	0.05	0.18	0.71	0.22	0.23
11_64.8	iHS	11:64902002	21.3	5,U	SYVN1	0.02	0.29	0.17	0.15	0.22	0.13	0.15	0.04	0.05	0.18	0.71	0.22	0.23
11_67.2	nSL	11:67276158	19.33	R	-	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.03	0.02	0.74	0.00	0.09
11_67.2	nSL	11:67276158	19.33	U	CDK2AP2	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.03	0.02	0.74	0.00	0.09
11_67.2	nSL	11:67276158	19.33	U	PITPNM1	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.03	0.02	0.74	0.00	0.09
12_39.8	TD	12:39822917	16.36	I	KIF21A	0.40	0.81	0.75	0.84	0.87	0.68	0.91	0.71	0.87	0.70	0.87	0.53	0.56
14_47.6	iHS	14:47770258	16.63	I,N,NMD	MDGA2	0.00	0.04	0.14	0.13	0.13	0.16	0.15	0.09	0.08	0.36	0.84	0.00	0.01
14_61.2	TD	14:61322571	15.28	I,N	MNAT1	0.65	0.71	0.80	0.75	0.76	0.63	0.56	0.63	0.48	0.70	0.89	0.78	0.76
15_66.6	iHS	15:66726369	17.63	I,N	MAP2K1	0.06	0.19	0.23	0.26	0.20	0.21	0.29	0.16	0.15	0.40	0.58	0.17	0.03
16_11.2	nSL	16:11337788	15.78	D	HNRNPCP4	0.10	0.04	0.03	0.02	0.11	0.25	0.35	0.34	0.52	0.44	0.89	0.59	0.36
17_33.2	iHS	17:33210737	15.11	IG	-	0.27	0.44	0.28	0.17	0.24	0.20	0.18	0.16	0.13	0.42	0.68	0.07	0.16
2_41.2	nSL, iHS	2:41294721	15.94	IG	-	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.03	0.10	0.61	0.00	0.00
2_86.6	nSL	2:86635502	19.75	IG	-	0.29	0.52	0.59	0.62	0.63	0.43	0.74	0.81	0.81	0.80	0.76	0.76	0.49
2_86.6	nSL	2:86607189	15.06	IG	-	0.52	0.42	0.52	0.48	0.63	0.55	0.82	0.90	0.87	0.86	0.76	0.81	0.59
20_2.8	nSL, iHS	20:2889720	17.6	I	PTPRA	0.13	0.00	0.00	0.00	0.02	0.04	0.12	0.06	0.15	0.16	0.58	0.03	0.10
20_2.8	nSL, iHS	20:2860087	15.69	I	PTPRA	0.00	0.00	0.00	0.00	0.02	0.00	0.12	0.06	0.15	0.16	0.58	0.03	0.10
3_121.4	iHS	3:121500699	22	3,NMD,M	IQCB1	0.19	0.21	0.20	0.25	0.33	0.45	0.24	0.18	0.24	0.22	0.89	0.17	0.22
3_121.8	nSL	3:121908434	16.53	I	CASR	0.00	0.19	0.27	0.16	0.17	0.18	0.29	0.29	0.21	0.36	0.82	0.34	0.20
3_138	TD	3:138191232	23.6	3,NMD,M,U,D,NC	ESYT3	0.02	0.42	0.48	0.37	0.43	0.54	0.44	0.56	0.60	0.58	0.87	0.43	0.49
3_18.2	TD	3:18201464	20.3	I,N	TBC1D5	0.00	0.13	0.31	0.26	0.30	0.36	0.41	0.47	0.53	0.54	0.84	0.48	0.48
5_103.2	nSL	5:103279166	21.8	IG	-	0.00	0.15	0.14	0.25	0.35	0.41	0.56	0.72	0.71	0.70	0.79	0.52	0.48
5_126.2	iHS	5:126365228	20.1	I,N	MAR3	0.02	0.12	0.20	0.17	0.22	0.09	0.44	0.37	0.29	0.48	0.63	0.33	0.31
5_126.2	iHS	5:126365228	20.1	R	-	0.02	0.12	0.20	0.17	0.22	0.09	0.44	0.37	0.29	0.48	0.63	0.33	0.31
5_40.6	iHS	5:40704952	15.1	IG	-	0.13	0.42	0.39	0.49	0.28	0.36	0.44	0.24	0.11	0.26	0.84	0.14	0.19
5_40.6	iHS	5:40734626	15.63	I,N	TTC33	0.15	0.42	0.39	0.49	0.28	0.36	0.44	0.24	0.11	0.26	0.84	0.14	0.19
6_119.2	TD	6:119215902	18.07	R	-	0.04	0.23	0.22	0.22	0.26	0.36	0.38	0.37	0.61	0.58	0.82	0.33	0.34
6_119.2	TD	6:119215402	15.34	R	-	0.15	0.44	0.38	0.35	0.39	0.52	0.53	0.62	0.73	0.64	0.82	0.62	0.53
6_119.2	TD	6:119215902	18.07	I,N	ASF1A	0.04	0.23	0.22	0.22	0.26	0.36	0.38	0.37	0.61	0.58	0.82	0.33	0.34
6_119.2	TD	6:119215402	15.34	5,NC	ASF1A	0.15	0.44	0.38	0.35	0.39	0.52	0.53	0.62	0.73	0.64	0.82	0.62	0.53
6_119.2	TD	6:119215902	18.07	I,U	MCM9	0.04	0.23	0.22	0.22	0.26	0.36	0.38	0.37	0.61	0.58	0.82	0.33	0.34
6_119.2	TD	6:119215402	15.34	I,NC	MCM9	0.15	0.44	0.38	0.35	0.39	0.52	0.53	0.62	0.73	0.64	0.82	0.62	0.53
6_159	nSL, iHS	6:159178345	24.7	M	SYTL3	0.21	0.54	0.36	0.36	0.41	0.36	0.62	0.76	0.73	0.98	0.89	0.72	0.68
6_50.2	iHS	6:50201951	18.63	IG	-	0.06	0.08	0.20	0.16	0.37	0.16	0.71	0.54	0.71	0.68	0.74	0.45	0.41
6_50.2	iHS	6:50283373	18.53	IG	-	0.02	0.00	0.08	0.11	0.09	0.00	0.21	0.09	0.08	0.16	0.76	0.00	0.00
6_50.2	iHS	6:50378717	18.04	IG	-	0.02	0.02	0.06	0.12	0.11	0.00	0.18	0.09	0.05	0.14	0.66	0.02	0.00
6_51.6	iHS	6:51746757	19.15	I	PKHD1	0.21	0.27	0.28	0.49	0.41	0.52	0.41	0.68	0.66	0.60	0.89	0.72	0.76
6_51.6	iHS	6:51628094	15.81	I	PKHD1	0.02	0.33	0.41	0.43	0.39	0.52	0.53	0.60	0.66	0.70	0.71	0.81	0.67
6_51.6	iHS	6:51691632	15.66	I	PKHD1	0.27	0.38	0.47	0.49	0.43	0.63	0.47	0.68	0.74	0.74	0.87	0.79	0.68
6_51.6	iHS	6:51758405	15.38	I	PKHD1	0.15	0.23	0.28	0.49	0.41	0.52	0.41	0.68	0.66	0.60	0.89	0.72	0.76
6_51.8	nSL	6:51914956	18.79	R	-	0.15	0.44	0.53	0.58	0.57	0.34	0.41	0.37	0.34	0.38	0.89	0.31	0.48
6_51.8	nSL	6:51914956	18.79	M,SR	PKHD1	0.15	0.44	0.53	0.58	0.57	0.34	0.41	0.37	0.34	0.38	0.89	0.31	0.48
7_132.6	nSL, iHS	7:132683047	18.26	I,NMD,N	CHCHD3	0.23	0.44	0.47	0.41	0.30	0.36	0.15	0.35	0.40	0.42	0.66	0.57	0.71
7_132.6	nSL, iHS	7:132693536	18.08	I,NMD,N	CHCHD3	0.27	0.35	0.39	0.34	0.24	0.16	0.12	0.32	0.39	0.42	0.66	0.38	0.34
7_132.6	nSL, iHS	7:132666863	15.74	I,NMD,N	CHCHD3	0.21	0.44	0.47	0.41	0.30	0.34	0.15	0.35	0.40	0.42	0.66	0.57	0.71
8_79.2	TD	8:79284507	18.15	IG	-	0.00	0.33	0.36	0.33	0.37	0.41	0.29	0.32	0.23	0.24	0.74	0.50	0.33

Table C.13 Full list of significant SNPs from the window-based tests in the COL population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_16.4	nSL, iHS	1:16518748	15.92	IG	-	0.13	0.35	0.42	0.50	0.50	0.38	0.47	0.62	0.44	0.62	0.42	0.72	0.74
1_16.6	nSL, iHS	1:16641899	15.22	S,U	FBXO42	0.06	0.38	0.36	0.40	0.35	0.36	0.47	0.60	0.42	0.32	0.34	0.74	0.62
1_179.8	nSL	1:179989742	22.4	M	CEP350	0.29	0.48	0.67	0.61	0.65	0.70	0.82	0.69	0.89	0.98	1.00	0.67	0.62
1_193.6	iHS	1:193635400	16.22	IG	-	0.31	0.58	0.53	0.41	0.57	0.70	0.53	0.68	0.68	0.76	0.53	0.83	0.73
1_243.8	iHS	1:243868398	19.12	I,N	AKT3	0.21	0.73	0.73	0.63	0.72	0.48	0.24	0.35	0.11	0.18	0.53	0.53	0.46
1_52.2	TD	1:52364576	16.05	R	-	0.35	0.58	0.61	0.55	0.70	0.46	0.65	0.76	0.66	0.74	0.92	0.86	0.80
1_73.4	iHS	1:73408437	17.45	IG	-	0.35	0.48	0.39	0.50	0.39	0.70	0.38	0.71	0.66	0.36	0.45	0.72	0.71
1_73.4	iHS	1:73573941	16.56	D	KRT8P21	0.19	0.42	0.30	0.44	0.35	0.66	0.38	0.72	0.66	0.36	0.29	0.74	0.71
1_73.4	iHS	1:73531160	16.34	IG	-	0.35	0.44	0.31	0.46	0.35	0.64	0.38	0.72	0.66	0.34	0.29	0.72	0.71
11_92	nSL, iHS	11:92109106	15.88	I,N	FAT3	0.25	0.79	0.83	0.82	0.89	0.77	0.65	0.81	0.56	0.68	0.95	0.88	0.80
12_112.8	iHS	12:112991832	18.73	IG	-	0.10	0.08	0.06	0.08	0.20	0.11	0.12	0.50	0.47	0.44	0.29	0.76	0.78
12_33.2	nSL, iHS	12:33354199	16.97	IG	-	0.13	0.15	0.06	0.04	0.17	0.36	0.12	0.62	0.32	0.24	0.21	0.78	0.67
12_44.4	iHS, TD	12:44488039	19.32	I,NMD,N	TMEM117	0.27	0.73	0.89	0.89	0.89	0.84	1.00	0.85	0.92	0.96	0.97	0.90	0.88
12_44.4	iHS, TD	12:44532975	19.2	I,NMD,N	TMEM117	0.29	0.73	0.89	0.89	0.89	0.84	1.00	0.85	0.92	0.96	0.95	0.90	0.88
13_34.8	nSL	13:34880795	15.78	IG	-	0.10	0.15	0.02	0.02	0.11	0.27	0.53	0.59	0.73	0.64	0.89	0.71	0.56
13_86.4	iHS	13:86424426	18.18	IG	-	0.08	0.19	0.13	0.20	0.20	0.46	0.50	0.41	0.47	0.34	0.66	0.71	0.71
13_86.4	iHS	13:86430443	17.23	IG	-	0.19	0.19	0.14	0.23	0.20	0.46	0.50	0.43	0.47	0.34	0.66	0.71	0.71
17_4	iHS	17:4010412	15.13	I,N	ZZEF1	0.10	0.62	0.63	0.68	0.65	0.70	0.50	0.56	0.56	0.48	0.71	0.78	0.61
17_63.8	nSL	17:63996645	19.65	I	CEP112	0.27	0.54	0.41	0.31	0.52	0.59	0.65	0.75	0.79	0.86	0.95	0.88	0.89
17_63.8	nSL	17:63964425	17.61	I	CEP112	0.23	0.52	0.41	0.31	0.52	0.59	0.65	0.75	0.81	0.86	0.95	0.88	0.89
17_64	nSL, iHS	17:64052744	18.75	I	CEP112	0.27	0.54	0.41	0.31	0.52	0.59	0.65	0.75	0.79	0.86	0.95	0.88	0.89
18_23.8	nSL, iHS	18:23877175	15.03	D	U3	0.06	0.48	0.39	0.42	0.59	0.70	0.79	0.75	0.68	0.72	0.50	0.79	0.81
18_23.8	nSL, iHS	18:23877175	15.03	I,NMD,D	TAF4B	0.06	0.48	0.39	0.42	0.59	0.70	0.79	0.75	0.68	0.72	0.50	0.79	0.81
18_23.8	nSL, iHS	18:23877175	15.03	R	-	0.06	0.48	0.39	0.42	0.59	0.70	0.79	0.75	0.68	0.72	0.50	0.79	0.81
19_38	iHS	19:38091932	18.41	M,I,NC	ZNF540	0.35	0.12	0.11	0.20	0.22	0.41	0.26	0.50	0.40	0.40	0.34	0.69	0.46
19_38.2	iHS	19:38377773	22.9	M	WDR87	0.38	0.12	0.13	0.21	0.22	0.38	0.26	0.49	0.40	0.40	0.34	0.62	0.38
19_38.2	iHS	19:38375290	16.55	D	WDR87	0.42	0.12	0.13	0.21	0.22	0.38	0.26	0.49	0.40	0.40	0.34	0.62	0.38
2_109.4	nSL, iHS	2:109513601	19.61	M	EDAR	0.00	0.02	0.00	0.02	0.22	0.00	0.59	0.71	0.97	0.92	0.95	0.76	0.72
2_126.4	nSL	2:126410056	19.16	IG	-	0.10	0.33	0.20	0.17	0.37	0.46	0.41	0.57	0.69	0.40	0.42	0.78	0.57
2_159	nSL	2:159092241	23.3	I,NMD	CCDC148	0.04	0.31	0.33	0.38	0.43	0.71	0.32	0.38	0.42	0.32	0.21	0.57	0.49
2_159	nSL	2:159092241	23.3	SA,N	CCDC148-AS1	0.04	0.31	0.33	0.38	0.43	0.71	0.32	0.38	0.42	0.32	0.21	0.57	0.49
2_197.8	nSL, iHS	2:197845449	19.14	I	ANKRD44	0.17	0.77	0.69	0.68	0.65	0.71	0.71	0.90	0.92	0.72	0.92	0.84	0.89
2_219.4	iHS, TD	2:219517088	17.93	I,NMD,D	ZNF142	0.02	0.56	0.55	0.58	0.74	0.64	0.74	0.78	0.82	0.86	0.71	0.90	0.84
2_219.4	iHS, TD	2:219426972	15.89	I	USP37	0.02	0.56	0.55	0.58	0.74	0.63	0.74	0.78	0.81	0.86	0.71	0.90	0.77
3_121.6	nSL	3:121635230	17.75	I,N,D	SLC15A2	0.52	0.35	0.48	0.34	0.46	0.32	0.44	0.49	0.58	0.34	0.05	0.71	0.52
3_121.6	nSL	3:121664112	15.82	D	SLC15A2	0.15	0.35	0.48	0.33	0.46	0.32	0.44	0.49	0.56	0.34	0.05	0.72	0.61
3_121.6	nSL	3:121643804	15.49	M,U,D	SLC15A2	0.52	0.35	0.48	0.33	0.46	0.32	0.44	0.49	0.56	0.34	0.05	0.72	0.52
3_132.6	nSL	3:132773737	16.82	I,N	TMEM108	0.12	0.60	0.47	0.49	0.54	0.36	0.53	0.56	0.53	0.62	0.42	0.55	0.54
3_65.4	nSL	3:65521324	19.42	I,N,U	MAGI1	0.04	0.13	0.42	0.50	0.39	0.41	0.32	0.54	0.45	0.28	0.32	0.66	0.57
3_65.4	nSL	3:65536944	16.4	I,N,U	MAGI1	0.06	0.15	0.47	0.49	0.52	0.38	0.41	0.59	0.52	0.34	0.34	0.69	0.51
3_65.4	nSL	3:65464712	15.33	I,N,U	MAGI1	0.06	0.17	0.47	0.50	0.41	0.38	0.44	0.57	0.50	0.68	0.42	0.72	0.61
4_160.4	nSL, iHS	4:160483950	17.47	IG	-	0.46	0.60	0.38	0.41	0.50	0.46	0.65	0.74	0.68	0.72	0.92	0.78	0.71
5_121.2	nSL, TD	5:121249324	22.1	IG	-	0.31	0.63	0.69	0.69	0.63	0.71	0.76	0.88	0.87	0.76	0.92	0.79	0.76
6_47.6	nSL	6:47670269	16.71	I	GPR115	0.13	0.23	0.27	0.28	0.37	0.14	0.53	0.68	0.77	0.70	0.82	0.64	0.59
6_47.6	nSL	6:47670269	16.71	D	GPR111	0.13	0.23	0.27	0.28	0.37	0.14	0.53	0.68	0.77	0.70	0.82	0.64	0.59
8_120.8	nSL	8:120904009	21.9	I	DEPTOR	0.25	0.46	0.58	0.47	0.65	0.43	0.62	0.81	0.84	0.92	0.87	0.88	0.84
8_120.8	nSL	8:120904009	21.9	U	RNA5SP277	0.25	0.46	0.58	0.47	0.65	0.43	0.62	0.81	0.84	0.92	0.87	0.88	0.84
8_120.8	nSL	8:120925083	18.79	I	DEPTOR	0.25	0.46	0.58	0.49	0.65	0.43	0.62	0.81	0.84	0.92	0.87	0.88	0.83
8_129.8	nSL, iHS	8:129884159	22.1	IG	-	0.12	0.06	0.27	0.20	0.26	0.14	0.41	0.57	0.65	0.74	0.74	0.79	0.64
8_129.8	nSL, iHS	8:129844527	19.87	IG	-	0.10	0.06	0.27	0.19	0.24	0.13	0.41	0.59	0.65	0.74	0.74	0.81	0.66
9_126.2	nSL, iHS	9:126328539	20.2	I,N,D	DENND1A	0.31	0.73	0.66	0.51	0.41	0.63	0.74	0.65	0.69	0.68	0.55	0.74	0.67
9_126.2	nSL, iHS	9:126397618	16.38	I,N	DENND1A	0.21	0.73	0.64	0.53	0.52	0.71	0.74	0.78	0.77	0.76	0.66	0.76	0.72

Table C.14 Full list of significant SNPs from the window-based tests in the SEM population

Window	Tests	Chr:pos	CADD	Cons.	Gene	AFR	WAA	SWE	ENE	VOL	SOA	WSI	SSI	CSI	NSI	COL	SEM	SEA
1_216.2	nSL	1:216306127	18.46	I	USH2A	0.79	0.63	0.78	0.86	0.72	0.80	0.88	0.78	0.82	0.92	0.76	0.78	0.80
1_216.2	nSL	1:216297936	16.35	I	USH2A	0.13	0.37	0.36	0.36	0.39	0.59	0.71	0.68	0.81	0.88	0.71	0.66	0.73
1_216.2	nSL	1:216295188	16.07	I	USH2A	0.15	0.37	0.36	0.36	0.39	0.59	0.71	0.68	0.81	0.88	0.71	0.66	0.73
1_26.6	nSL	1:26797508	16.72	3,NMD,D	DHDDS	0.46	0.35	0.27	0.20	0.35	0.46	0.59	0.72	0.92	0.90	0.24	0.60	0.59
1_26.6	nSL	1:26797508	16.72	U	HMGN2	0.46	0.35	0.27	0.20	0.35	0.46	0.59	0.72	0.92	0.90	0.24	0.60	0.59
1_26.6	nSL	1:26786627	16.19	3,NMD,M,I	DHDDS	0.46	0.35	0.27	0.20	0.35	0.46	0.59	0.72	0.92	0.90	0.24	0.60	0.59
1_69	nSL, iHS	1:69069503	18.5	IG	-	0.12	0.17	0.25	0.37	0.33	0.57	0.32	0.54	0.55	0.56	0.61	0.59	0.76
1_69	nSL, iHS	1:69062741	15.8	IG	-	0.12	0.19	0.25	0.36	0.33	0.57	0.32	0.54	0.55	0.56	0.61	0.59	0.76
1_73	nSL	1:73095148	17.78	IG	-	0.13	0.37	0.13	0.32	0.39	0.63	0.32	0.68	0.89	0.42	0.39	0.66	0.80
1_73.4	nSL	1:73408437	17.45	IG	-	0.35	0.48	0.39	0.50	0.39	0.70	0.38	0.71	0.66	0.36	0.45	0.72	0.71
1_73.4	nSL	1:73573941	16.56	D	KRT8P21	0.19	0.42	0.30	0.44	0.35	0.66	0.38	0.72	0.66	0.36	0.29	0.74	0.71
1_73.4	nSL	1:73531160	16.34	IG	-	0.35	0.44	0.31	0.46	0.35	0.64	0.38	0.72	0.66	0.34	0.29	0.72	0.71
1_75.4	nSL	1:75541418	20.9	IG	-	0.10	0.12	0.20	0.29	0.35	0.36	0.47	0.82	0.71	0.84	0.58	0.72	0.89
1_75.4	nSL	1:75474961	19.47	IG	-	0.13	0.12	0.20	0.29	0.35	0.36	0.47	0.81	0.71	0.84	0.58	0.72	0.84
1_75.4	nSL	1:75461538	15.72	IG	-	0.10	0.15	0.22	0.36	0.28	0.39	0.35	0.69	0.69	0.84	0.50	0.71	0.82
1_75.4	nSL	1:75497178	15.23	IG	-	0.12	0.12	0.20	0.29	0.35	0.36	0.47	0.81	0.71	0.84	0.58	0.72	0.84
10_107.2	nSL, iHS, TD	10:107233183	19.05	IG	-	0.00	0.60	0.52	0.53	0.63	0.45	0.65	0.78	0.65	0.72	0.58	0.84	0.89
10_107.2	nSL, iHS, TD	10:107319562	18.3	IG	-	0.00	0.63	0.45	0.49	0.48	0.38	0.65	0.81	0.71	0.72	0.63	0.72	0.77
10_107.2	nSL, iHS, TD	10:107388758	16.79	IG	-	0.04	0.56	0.47	0.47	0.52	0.34	0.68	0.91	0.85	0.78	0.68	0.81	0.89
10_107.2	nSL, iHS, TD	10:107295182	16.03	D	RNU6-463P	0.02	0.63	0.42	0.50	0.48	0.38	0.65	0.81	0.71	0.72	0.63	0.78	0.84
10_59.6	nSL, TD	10:59631611	22.3	IG	-	0.17	0.40	0.58	0.68	0.78	0.68	0.82	0.94	0.97	1.00	0.87	0.93	0.82
11_121.4	nSL	11:121593385	17.14	R,IG	-	0.10	0.67	0.77	0.84	0.65	0.71	0.82	0.87	0.90	0.90	0.71	0.78	0.73
12_44.4	TD	12:44488039	19.32	I,NMD,N	TMEM117	0.27	0.73	0.89	0.89	0.89	0.84	1.00	0.85	0.92	0.96	0.97	0.90	0.88
12_44.4	TD	12:44532975	19.2	I,NMD,N	TMEM117	0.29	0.73	0.89	0.89	0.89	0.84	1.00	0.85	0.92	0.96	0.95	0.90	0.88
12_44.6	TD	12:44602376	17.98	I,NMD,N	TMEM117	0.38	0.77	0.89	0.89	0.93	0.95	1.00	0.85	0.92	0.96	0.95	0.91	0.90
14_23.4	nSL	14:23441179	15.98	3,D	AJUBA	0.15	0.60	0.58	0.63	0.59	0.80	0.62	0.72	0.76	0.78	0.68	0.88	0.84
14_90.6	nSL	14:90715014	15.35	IG	-	0.13	0.27	0.34	0.40	0.41	0.52	0.76	0.53	0.50	0.58	0.34	0.45	0.62
14_97.2	iHS	14:97272382	19.09	I,N	VRK1	0.06	0.12	0.05	0.10	0.22	0.16	0.53	0.60	0.76	0.60	0.84	0.76	0.81
14_97.2	iHS	14:97321689	17.47	S,U,D	VRK1	0.08	0.46	0.27	0.42	0.37	0.32	0.71	0.72	0.71	0.60	0.84	0.78	0.84
15_43.2	iHS	15:43398108	15.7	D	EPB42	0.08	0.50	0.63	0.64	0.70	0.55	0.62	0.63	0.60	0.58	0.87	0.60	0.40
15_43.2	iHS	15:43398108	15.7	R	-	0.08	0.50	0.63	0.64	0.70	0.55	0.62	0.63	0.60	0.58	0.87	0.60	0.40
15_43.2	iHS	15:43398108	15.7	I,N,NMD	UBR1	0.08	0.50	0.63	0.64	0.70	0.55	0.62	0.63	0.60	0.58	0.87	0.60	0.40
15_64.6	TD	15:64637091	16.95	I	CSNK1G1	0.02	0.06	0.06	0.11	0.33	0.11	0.35	0.49	0.50	0.38	0.11	0.64	0.73
15_64.6	TD	15:64637091	16.95	R	-	0.02	0.06	0.06	0.11	0.33	0.11	0.35	0.49	0.50	0.38	0.11	0.64	0.73
15_64.6	TD	15:64637091	16.95	D	SNORA48	0.02	0.06	0.06	0.11	0.33	0.11	0.35	0.49	0.50	0.38	0.11	0.64	0.73
15_64.6	TD	15:64759279	15.72	I	ZNF609	0.02	0.06	0.06	0.10	0.35	0.11	0.35	0.56	0.55	0.48	0.05	0.69	0.69
15_65.8	iHS	15:65959729	15.82	I,U	DENND4A	0.10	0.15	0.17	0.33	0.30	0.36	0.38	0.43	0.53	0.58	0.26	0.74	0.62
15_65.8	iHS	15:65898134	15.71	I,NMD,N	VWA9	0.10	0.15	0.17	0.32	0.30	0.36	0.32	0.43	0.53	0.58	0.24	0.74	0.64
15_65.8	iHS	15:65897863	15.19	I,N,NMD	VWA9	0.10	0.15	0.17	0.32	0.30	0.36	0.35	0.43	0.53	0.58	0.24	0.74	0.64
16_73.6	iHS	16:73714553	15.89	IG	-	0.15	0.54	0.52	0.46	0.50	0.57	0.71	0.66	0.73	0.82	0.79	0.79	0.80
2_109.4	nSL, iHS	2:109513601	19.61	M	EDAR	0.00	0.02	0.00	0.02	0.22	0.00	0.59	0.71	0.97	0.92	0.95	0.76	0.72
2_158.6	nSL	2:158603556	18.51	I	ACVR1	0.19	0.83	0.73	0.78	0.72	0.89	0.56	0.69	0.68	0.70	0.95	0.79	0.80
2_17.4	nSL	2:17409594	16.1	IG	-	0.00	0.06	0.03	0.08	0.13	0.27	0.44	0.53	0.65	0.56	0.47	0.60	0.74
2_17.6	nSL	2:17722654	15.2	I,N,5	VSNL1	0.10	0.23	0.23	0.30	0.46	0.36	0.79	0.81	0.92	0.98	0.76	0.79	0.86
2_17.6	nSL	2:17775840	15.19	I	VSNL1	0.06	0.21	0.19	0.29	0.46	0.41	0.76	0.81	0.89	0.92	0.79	0.74	0.78
2_177.6	nSL, iHS	2:177770595	18.62	IG	-	0.06	0.37	0.34	0.42	0.61	0.63	0.79	0.91	0.97	0.96	0.63	0.93	0.87
2_178.2	nSL	2:178306584	15.01	I,D	AGPS	0.29	0.71	0.80	0.76	0.61	0.61	0.56	0.59	0.52	0.64	0.95	0.57	0.63
2_178.4	nSL, iHS	2:178461560	16.92	IG	-	0.27	0.67	0.80	0.78	0.61	0.57	0.59	0.50	0.45	0.68	0.95	0.57	0.64
2_197.8	nSL	2:197845449	19.14	I	ANKRD44	0.17	0.77	0.69	0.68	0.65	0.71	0.71	0.90	0.92	0.72	0.92	0.84	0.89
2_44	nSL	2:44004010	15.71	S,NMD,M,I,N,5	DYNC2LI1	0.10	0.48	0.59	0.70	0.61	0.52	0.74	0.71	0.74	0.88	0.32	0.76	0.80
2_44	nSL	2:44004010	15.71	U	RN7SKP66	0.10	0.48	0.59	0.70	0.61	0.52	0.74	0.71	0.74	0.88	0.32	0.76	0.80
2_64	iHS	2:64041944	19.19	I,N	WDPCP	0.10	0.27	0.22	0.18	0.28	0.36	0.56	0.62	0.63	0.50	0.66	0.76	0.83
21_27.8	iHS	21:27988908	17.23	IG	-	0.10	0.63	0.61	0.67	0.59	0.68	0.56	0.40	0.60	0.70	0.47	0.74	0.74
3_142	iHS	3:142149212	16.75	I,N,NMD,D,U	XRN1	0.04	0.44	0.44	0.47	0.59	0.64	0.65	0.51	0.60	0.64	0.79	0.55	0.59
3_41.8	iHS	3:41996136	16.52	U	RPL36P20	0.27	0.73	0.77	0.82	0.74	0.86	0.76	0.76	0.87	0.96	0.97	0.69	0.81
3_41.8	iHS	3:41996136	16.52	M,NMD,I,N,NC	ULK4	0.27	0.73	0.77	0.82	0.74	0.86	0.76	0.76	0.87	0.96	0.97	0.69	0.81
3_41.8	iHS	3:41829397	15.41	I	ULK4	0.08	0.44	0.39	0.47	0.26	0.55	0.21	0.37	0.13	0.22	0.74	0.17	0.40
3_75	nSL	3:75062144	16.48	IG	-	0.06	0.17	0.27	0.25	0.41	0.36	0.47	0.63	0.60	0.66	0.82	0.50	0.69
4_105	nSL	4:105031750	22.2	IG	-	0.10	0.33	0.19	0.26	0.35	0.32	0.62	0.62	0.61	0.40	0.89	0.71	0.84
4_105	nSL	4:105085890	22.1	IG	-	0.37	0.44	0.38	0.40	0.50	0.54	0.68	0.63	0.55	0.50	0.68	0.64	0.83
4_159.2	TD	4:159375206	15.01	I	RXFP1	0.02	0.23	0.33	0.39	0.46	0.32	0.79	0.63	0.77	0.68	0.45	0.69	0.81
4_159.6	TD	4:159729794	18.64	I,N,NC	FNIP2	0.02	0.27	0.36	0.48	0.48	0.43	0.68	0.66	0.58	0.66	0.61	0.74	0.88
4_173.6	nSL	4:173602023	17.33	I	GALNTL6	0.13	0.40	0.48	0.56	0.54	0.48	0.62	0.57	0.63	0.68	0.29	0.71	0.70
4_173.6	nSL	4:173673064	15.73	I	GALNTL6	0.12	0.40	0.52	0.60	0.54	0.48	0.62	0.57	0.63	0.68	0.29	0.71	0.72
5_112	iHS	5:112095775	16.36	I,N,NMD	APC	0.04	0.63	0.42	0.40	0.59	0.59	0.53	0.66	0.82	0.70	0.87	0.59	0.64
5_117.6	nSL	5:117728643	16.87	IG	-	0.21	0.85	0.70	0.65	0.72	0.89	0.74	0.79	0.69	0.76	0.71	0.91	0.80
5_117.6	nSL	5:117726694	16.14	IG	-	0.19	0.92	0.75	0.64	0.72	0.89	0.74	0.79	0.69	0.76	0.71	0.91	0.80
5_133.6	iHS	5:133741881	16.62	D,I	CDKN2AIPNL	0.17	0.62	0.77	0.68	0.52	0.63	0.56	0.54	0.39	0.56	0.16	0.45	0.52
5_153	TD	5:153188930	18.91	I	GRIA1	0.29	0.13	0.09	0.27	0.41	0.27	0.56	0.69	0.85	0.62	0.74	0.81	0.83
5_45	nSL, iHS	5:45109961	15.53	IG	-	0.90	0.69	0.88	0.86	0.74	0.86	0.94	0.93	0.87	0.84	0.74	0.93	0.81
6_101	iHS	6:101112345	16.92	I,NMD	ASCC3	0.44	0.40	0.39	0.46	0.35	0.54	0.35	0.22	0.10	0.04	0.13	0.29	0.40
6_101	iHS	6:101112345	16.92	R	-	0.44	0.40	0.39	0.46	0.35	0.54	0.35	0.22	0.10	0.04	0.13	0.29	0.40
6_129	nSL	6:129045204	16.86	R,IG	-	0.10	0.06	0.03	0.06	0.17	0.07	0.38	0.37	0.27	0.54	0.45	0.55	0.58
6_29.6	nSL	6:29755073	17.87	U	HLA-V	0.13	0.02	0.09	0.04	0.09	0.16	0.26	0.31	0.52	0.36	0.00	0.26	0.57
6_29.6	nSL	6:29755073	17.87	D	HCG4	0.13	0.02	0.09	0.04	0.09	0.16	0.26	0.31	0.52	0.36	0.00	0.26	0.57
6_80.8	iHS	6:80973665	16.88	I,N	BCKDHB	0.15	0.56	0.45	0.45	0.48	0.38	0.65	0.47	0.56	0.24	0.79	0.41	0.71
8_129.8	nSL, iHS	8:129844527	19.87	IG	-	0.10	0.06	0.27	0.19	0.24	0.13	0.41	0.59	0.65	0.74	0.74	0.81	0.66
8_67.6	iHS	8:67609366	15.79	I,N	C8orf44-SGK3	0.02	0.42	0.55	0.58	0.61	0.63	0.65	0.78	0.66	0.80	0.84	0.78	0.62
8_68.4	iHS	8:68598907	19.4	I,NMD	CPA6	0.17	0.69	0.56	0.64	0.63	0.48	0.62	0.63	0.63	0.66	0.53	0.74	0.73
9_102.8	TD	9:102984967	19.77	I,N	INVS	0.33	0.56	0.50	0.42	0.67	0.73	0.85	0.91	0.90	0.98	1.00	0.84	0.88
9_108	TD	9:108037717	15.06	I,NMD	SLC44A1	0.27	0.85	0.86	0.85	0.78	0.79	0.68	0.72	0.66	0.50	0.58	0.88	0.87

Table C.15 Full list of significant SNPs from the window-based tests in the SEA population

Appendix D

Graphical representations of significant selection enrichments in each immune gene category and selection statistic

It is interesting to consider the selection history of each population using the combination of enrichment results for all seven selection tests used in this thesis. These results are combined into Tables D.1 through D.13. Note that balancing selection enrichments contain MHC genes.

West and Central Africa

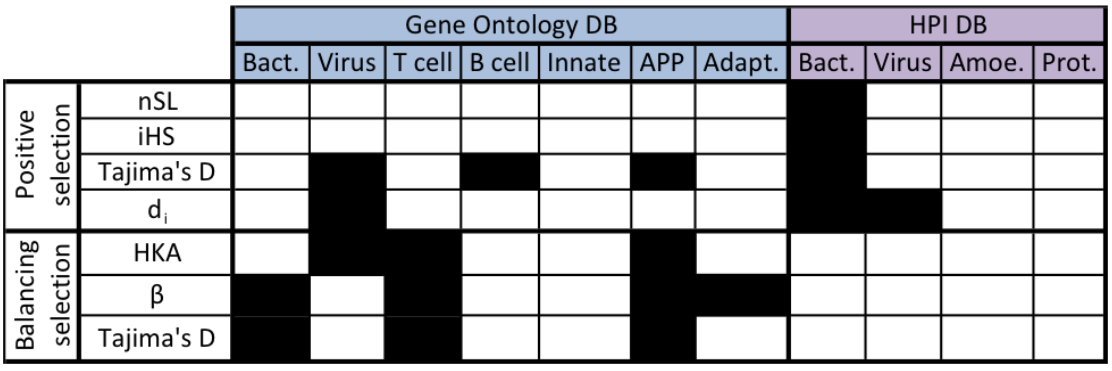


Table D.1 Summary of significant results in each class of genes and selection test for the AFR population. Squares filled in black indicate significant enrichment.

West Asia and Armenia

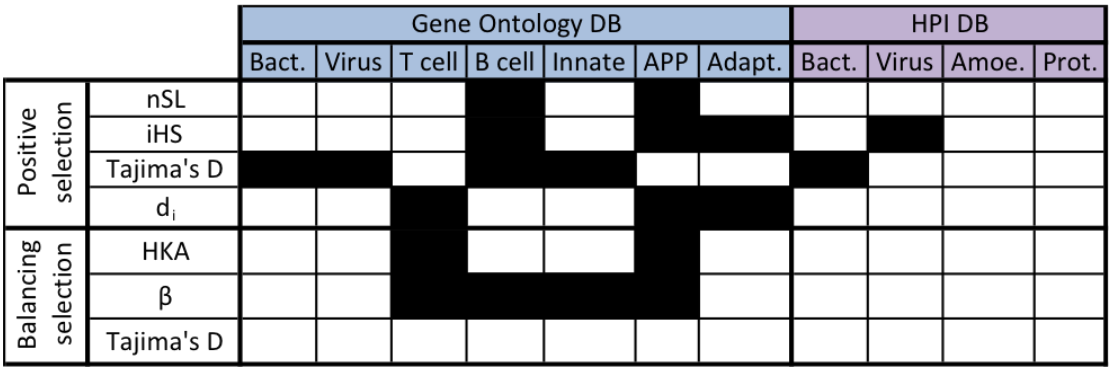


Table D.2 Summary of significant results in each class of genes and selection test for the WAA population. Squares filled in black indicate significant enrichment.

Southwest Europe

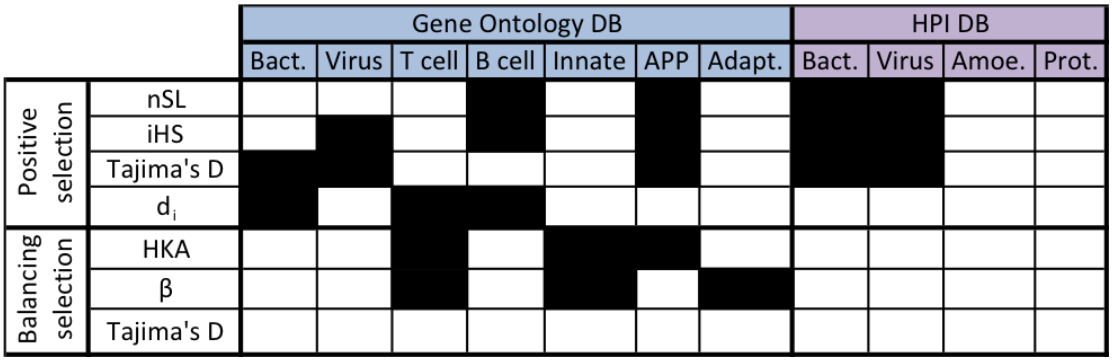


Table D.3 Summary of significant results in each class of genes and selection test for the SWE population. Squares filled in black indicate significant enrichment.

Northeast Europe

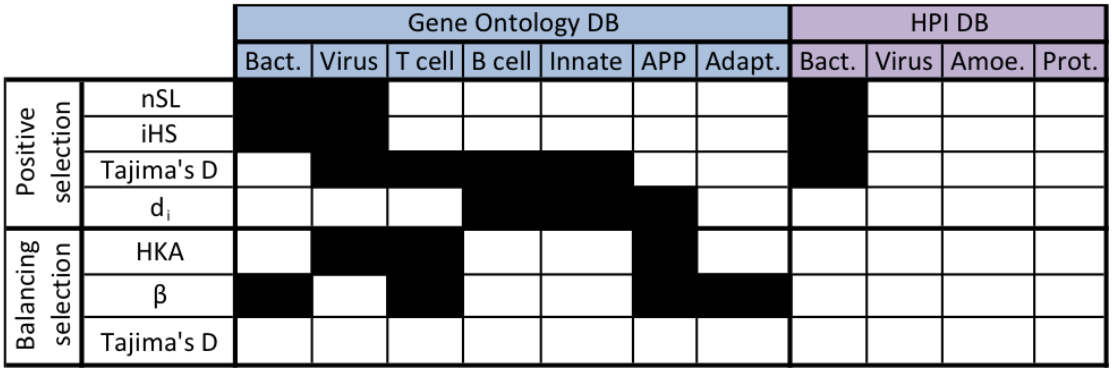


Table D.4 Summary of significant results in each class of genes and selection test for the ENE population. Squares filled in black indicate significant enrichment.

Volga Uralic

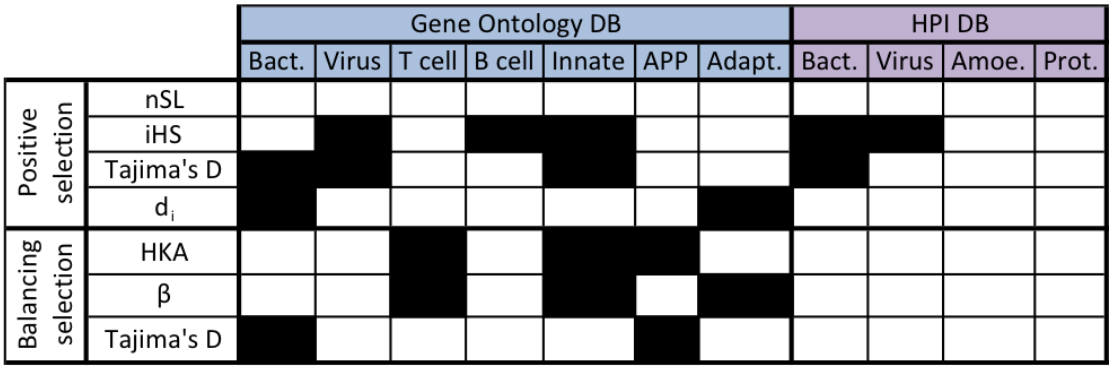


Table D.5 Summary of significant results in each class of genes and selection test for the VOL population. Squares filled in black indicate significant enrichment.

South Asia

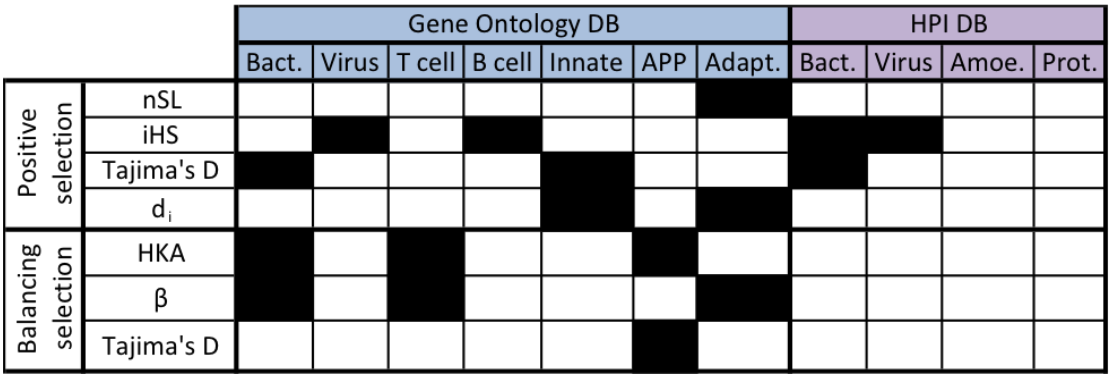


Table D.6 Summary of significant results in each class of genes and selection test for the SOA population. Squares filled in black indicate significant enrichment.

West Siberia

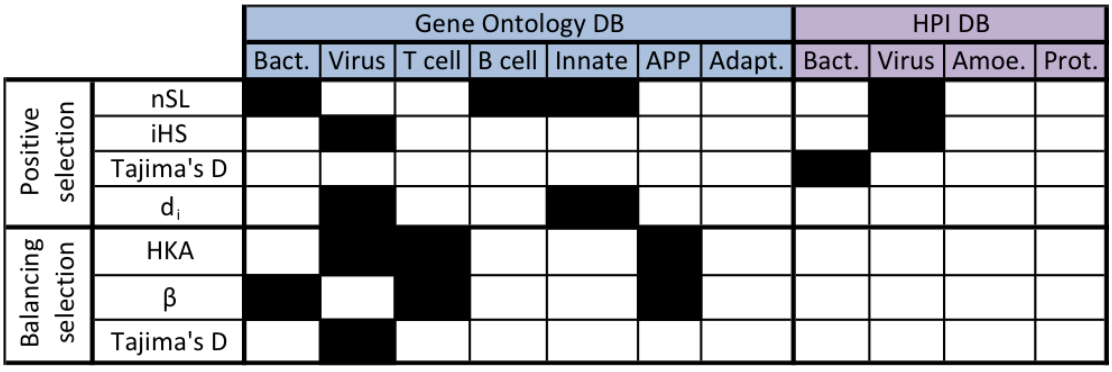


Table D.7 Summary of significant results in each class of genes and selection test for the WSI population. Squares filled in black indicate significant enrichment.

South Siberia and Mongolia

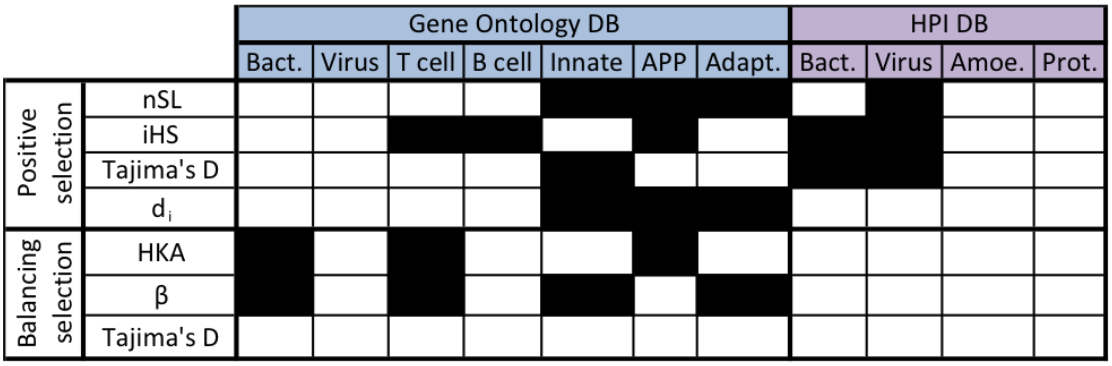


Table D.8 Summary of significant results in each class of genes and selection test for the SSI population. Squares filled in black indicate significant enrichment.

Central Siberia

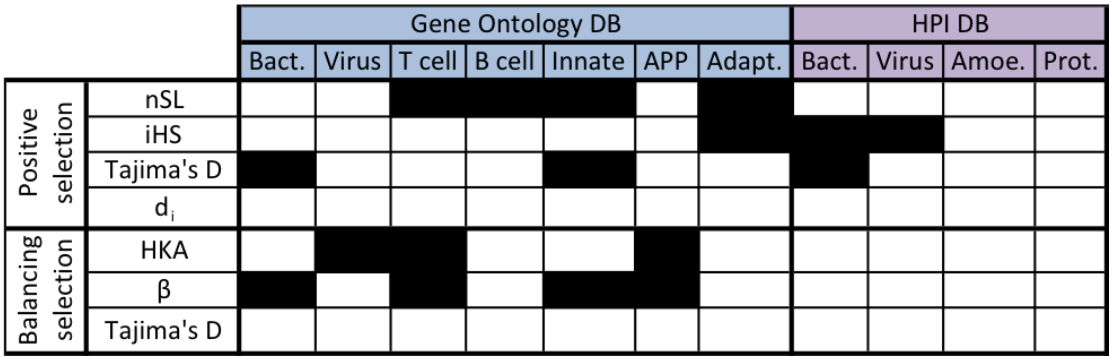


Table D.9 Summary of significant results in each class of genes and selection test for the CSI population. Squares filled in black indicate significant enrichment.

Northeast Siberia

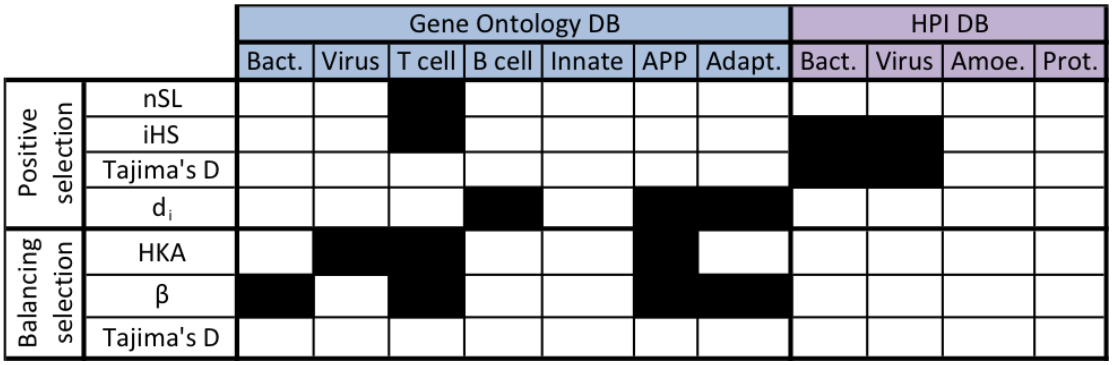


Table D.10 Summary of significant results in each class of genes and selection test for the NSI population. Squares filled in black indicate significant enrichment.

Colla

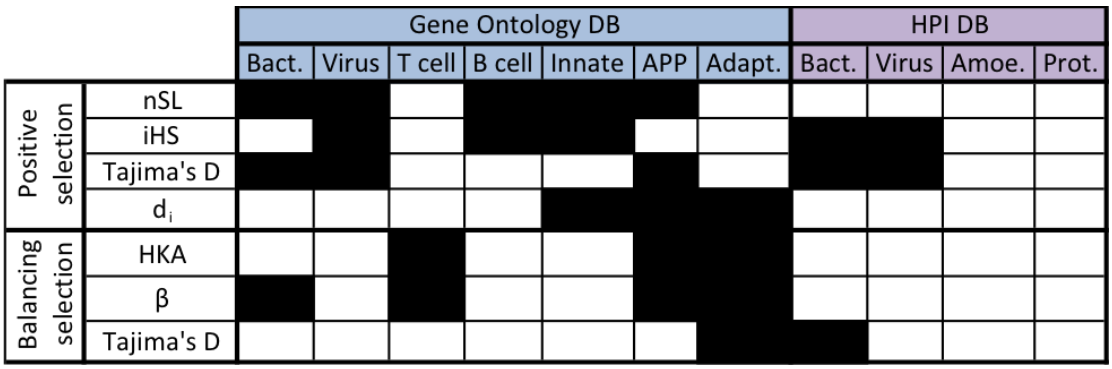


Table D.11 Summary of significant results in each class of genes and selection test for the COL population. Squares filled in black indicate significant enrichment.

Mainland Southeast Asia

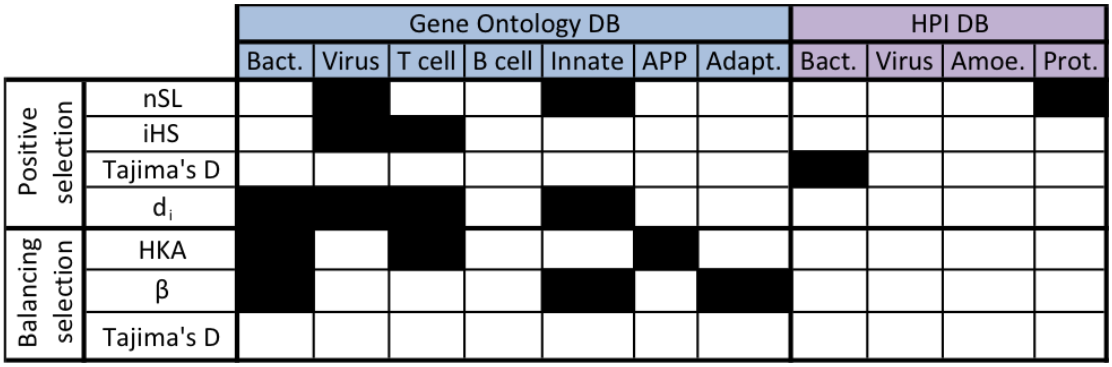


Table D.12 Summary of significant results in each class of genes and selection test for the SEM population. Squares filled in black indicate significant enrichment.

Island Southeast Asia

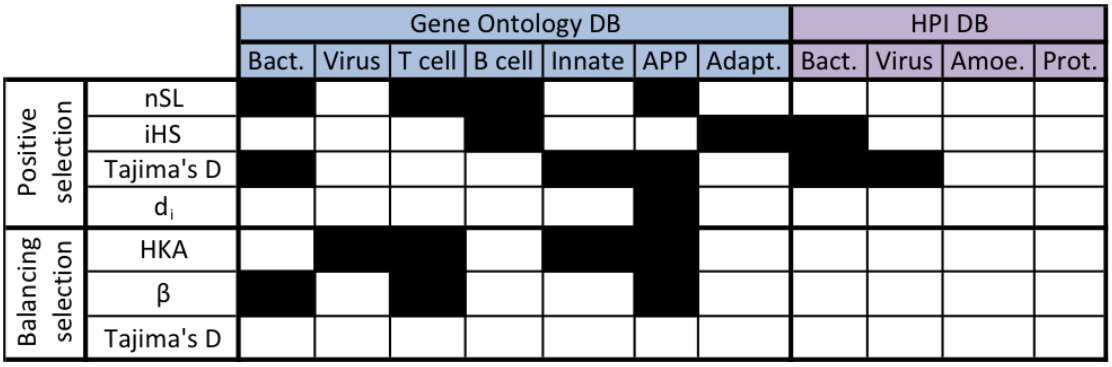


Table D.13 Summary of significant results in each class of genes and selection test for the SEA population. Squares filled in black indicate significant enrichment.