Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Samuel L Parks Date

HLA typing using genome wide data reveals susceptibility types for infections in a psychiatric disease enriched Ashkenazi Jewish population.

By

Samuel L Parks

Master of Science in Public Health

Epidemiology

_____

Bradley Pearce, PhD

Committee Chair

HLA Typing using genome wide data reveals susceptibility types for infections in a psychiatric disease enriched Ashkenazi population.

By

Samuel L Parks

B.S.P.H.

The George Washington University

2014

Thesis Committee Chair: Bradley Pearce, PhD.

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

In partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Epidemiology

2016

Abstract

HLA typing using genome wide data reveals susceptibility types for infections in a psychiatric disease enriched Ashkenazi Jewish population.

By Samuel L Parks

Introduction:

The infections *Toxoplasma gondii (T. gondii)*, cytomegalovirus, and herpes simplex 1 (HSV1) are common persistent infections and have been associated with schizophrenia and bipolar disorder. The major histocompatibility complex MHC (termed HLA in humans) region has been implicated in these infections and these mental illnesses. The interplay of MHC genetics, mental illness, and infection has not been systematically examined in previous research.

Methods:

In a cohort of 1636 Ashkenazi Jewish individuals, we imputed 7 HLA types (A, B, C, DRB1, DQA1, DQB1, DPB1) utilizing the HIBAG software package for R. HLA types were combined with pre-existing serology data. Logistic regression modeling stratified by mental disease status was used to assess the association between these HLA alleles and the infections, controlling for age and sex. Additional stratified logistic models assessed associations between homozygosity at the imputed HLA loci and the infections controlling for age and sex.

Results:

Several alleles had significant effect on infection risk after Bonferroni correction for multiple comparisons. Moreover, certain HLA haplotypes were consistently associated with increased or decreased risk of infections. The haplotypes HLA DRB*03:01~HLA DQA*05:01~HLA DQB*02:01 and HLA B*08:01~HLA C*07:01 displayed risk increasing effects for *T. gondii* and HSV1 that were more pronounced in mental illness free groups. Haplotypes B*38:01~HLA C*12:03 displayed protective effects against CMV and HSV1, which was also more pronounced in mental illness free groups. Homozygosity at HLA DQA correlated with decreased risk for CMV and HSV1 in the mentally ill group while homozygosity at HLA DRB correlated with increased CMV and HSV1 risk in the mentally ill group.

Conclusions:

This first epidemiologic analysis of HLA types, mental illness, and selected infections identified three haplotypes that potentially have implications for generally increasing or decreasing disease risk across multiple conditions as well as several individual alleles of interest for each infection. The fact we detected the most significant effects in the mental illness free groups may indicate that the effect of HLA alleles on infection risk may be modified in schizophrenia or bipolar disorder. Additionally HLA homozygosity appears to be a contributing factor in an individual's infection risk.

HLA typing using genome wide data reveals susceptibility types for infections in a psychiatric disease enriched Ashkenazi population.

By

Samuel L Parks

B.S.P.H.

The George Washington University

2014

Thesis Committee Chair: Bradley Pearce, PhD.

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

In partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Epidemiology

2016

Table of Contents

Abbreviations

AJ – Ashkenazi Jew

CDC – The Centers for Disease Control and Prevention

CIL – 95% Confidence Interval (lower bound)

CIU – 95% Confidence Interval (upper bound)

CMV – Cytomegalovirus

EpiGen – Epidemiology and Genetics in Psychiatry program

HLA – Human Leucocyte Antigens

HSV1 – Herpes Simplex Virus 1

GWAS- Genome Wide Association Study

IgG – Immunoglobulin G

LD – Linkage Disequilibrium

MHC – Major Histocompatibility Complex

NIMH- National Institute of Mental Health

OR – Odds Ratio

SAS – Statistical Analysis Software

SNPs- Single Nucleotide Polymorphisms

T. gondii – Toxoplasma gondii

Introduction

HLA and MHC

The Human Leucocyte antigen (HLA) region of chromosome 6 (the major histocompatibility complex in humans) encompasses roughly 4.5 million base pairs containing more than 200 genes, many of them involved in the immune system through coding for HLA proteins (Brucato, Guadalupe, Franke, Fisher, & Francks, 2015). This region (and the genes within it) can be divided into several classes, the most important of which are MHC I and MHC II. The Class I MHC genes (HLA A,B,C) code for transmembrane glycoproteins present on the surface of nucleated cells (Delves, 2014). These molecules present foreign antigens from inside the cells to CD8 T cells, which signals the CD8 killer T cells that a cell is infected and must be destroyed (Delves, 2014). Class II MHC molecules (HLA DP, DQ, DR, DM) on the other hand present antigens from outside the cell, are typically only present on specific cells (B cells, macrophages, dendritic cells) and do so primarily to stimulate the division and maturation of T-Helper cells (CD4) (Delves, 2014). The entire HLA region is known to be extremely genetically diverse from person to person (high polymorphism); biologically this helps the immune system to have a greater ability to react to a wide array of pathogens (Corvin & Morris, 2014). Additionally much of the region is known to be in linkage disequilibrium (LD) (Corvin & Morris, 2014). Due to LD there can exist common HLA haplotypes in certain groups, some have been identified in the Ashkenazi population primarily along for the HLA loci A~B~DRB (Bonné-Tamir et al., 1978; Klitz et al., 2010).

Due to the fact that the human MHC region has become highly polymorphic through rapid evolution in order to respond to large numbers of variable antigens, it has also been hypothesized that individuals who are homozygotic at one or more loci in the region would be at increased risk for disease as they would have the ability to immunologically respond to

fewer unique antigens (Corvin & Morris, 2014; Keller et al., 2012). Conversely heterozygotes would have a significant immunological advantage compared to their homozygote counterparts as their more polymorphic MHC region would confer the ability to successfully respond to more pathogens (Corvin & Morris, 2014; Keller et al., 2012). This "heterozygote advantage" hypothesis for the MHC region has been supported through both epidemiological studies and animal models for a wide array of infectious diseases including Hepatitis B, HTLV-1, HIV and malaria though not specifically mental illnesses (Carrington et al., 1999; Hill et al., 1991; Jeffery et al., 2000; Paterson, Wilson, & Pemberton, 1998; Thursz, Thomas, Greenwood, & Hill, 1997).

In addition to the MHC region's known immunologic function it has been previously associated with neuropsychiatric illness (namely schizophrenia and bipolar disorder) in several genome wide association studies, and this has been of particular interest as it hints at potential immunologic components of these diseases' pathogenesis (Corvin & Morris, 2014) (Brucato et al., 2015) (Debnath, Cannon, & Venkatasubramanian, 2013). These previous studies have largely focused on specific SNPs and relatively few have utilized imputed HLA types until the last several years and have also been largely limited to populations of European descent (Corvin & Morris, 2014).

Toxoplasma gondii

Toxoplasma gondii is a single-celled protozoan parasite that can cause an illness known as toxoplasmosis; A CDC sponsored study using NHANES data estimates that roughly 22.5% of the over 12 population of the United States is infected with *T. gondii* (Jones et al., 2001). The infection is not spread person to person except in rare instances of congenital transmission and blood transfusions (Jones et al., 2001). The most typical vectors of transmission are foodborne and zoonotic-- since house cats (and other members of the

family *felidae*) are the primary carriers of *T. gondii* (Jones et al., 2001). Despite the very high prevalence estimates very few people infected with the parasite typically exhibit symptoms; the most common serious sequella are toxoplasmosis in pregnant women and individuals with compromised immune systems (as in HIV patients) (Jones et al., 2001).

A great array of studies and subsequent reviews have shown a consistent and significant association between *T. gondii* infection (often assessed as presences of antibodies) and schizophrenia (Torrey, Bartko, Lun, & Yolken, 2007; Torrey & Robert, 2003). More recent studies have found that *T. gondii* has not been associated with mood disorders like depression but has displayed a significant association with bipolar disorder (Brad D. Pearce, Deanna Kruszon-Moran, & Jeffrey L. Jones, 2012). Given that bipolar disorder and schizophrenia have been previously suggested to share genetic underpinnings and mechanisms this strengthens the case (Corvin & Morris, 2014; Lichtenstein et al.).

The literature is much less consistent on the subject of *T. gondii* and HLA. A recent study found that HLA alleles associated with accelerated progression from HIV infection to AIDS were associated with resistance to toxoplasmic encephalitis (Rodrigues et al., 2016). A few studies have found specific alleles in the MHC Class II region to be associated with either *T. gondii* infection or worse outcomes of that infection (Mack et al., 1999; Shimokawa et al., 2016). Altogether, however there is not a single consensus indicating the association between HLA and *Toxoplasma gondii*, and there has not been a large systematic study of this relationship even in healthy populations.

Cytomegalovirus

Cytomegalovirus (CMV) is a member of the herpesvirus family that like *T. gondii* is very prevalent in the population but typically asymptomatic (Cannon, Schmid, & Hyde,

2010). The virus can be spread person to person via fluid transmission or congenitally (Cannon et al., 2010). Seroprevalence in the US varies considerably state to state but generally falls in the range of 50-80% in adult populations with women and nonwhites generally having higher prevalence (Cannon et al., 2010).

Previous studies have found that seroprevalence of CMV in individuals with schizophrenia and bipolar disorder to be significantly higher than in controls (Tedla et al., 2011; Torrey et al., 2012) (Leweke et al.). Several of these studies also found that anti-psychotic treatment lowers presence of CMV antibodies which can make infection more difficult to detect (Leweke et al.) (Torrey et al., 2012). This evidence has led many to hypothesize that CMV could play a role in schizophrenia's etiology.

Previous research has also implicated HLA as having a role in CMV, or even specifically genes previously associated with schizophrenia as also being associated with CMV (C.J. Carter, 2009) (Avramopoulos et al., 2015; Borglum et al., 2014; Torrey et al., 2012). This research has indicated that particular genes associated with schizophrenia are involved in the life cycles of pathogens including CMV and the strength of association observed between a particular gene and schizophrenia may in fact be conditional on the presence of CMV (or a similar pathogen such as HSV1)(C.J. Carter, 2009).

Herpes Simplex Virus 1

Herpes Simplex Virus 1 (HSV1) is a member of the herpesvirus family with a seroprevalence of 53.9% in the United States as of 2010 (Bradley, Markowitz, Gibson, & McQuillan, 2013). As with the previously discussed pathogens, most HSV-1 infections are subclinical, and when symptoms do occur they most frequently manifest as episodic ulcerative lesions in either the genital or orolabial regions (Bradley et al., 2013). HSV1 can

be transmitted both sexually and perinatally mother to child, the latter causing potentially fatal neonatal infection (Bradley et al., 2013).

Previous research on the association of HSV1 with HLA has not been as consistent, and many studies focused on HSV2 rather than HSV1. One study also identified a significant association between HSV1 seropositivity and a few specific MHC Class I alleles (B*35, C*15) but not any of the loci generally (Moraru et al., 2012). Several different studies have suggested associations between HSV1 and non-HLA genes implicated in schizophrenia risk (C. J. Carter, 2011; R. H. Yolken & Torrey, 2008).

The literature is also not clear on HSV1's relationship with schizophrenia and psychotic disease generally. Multiple studies that attempted to identify an association between the virus and schizophrenia or bipolar disorder returned null results (Leweke et al.; Tedla et al., 2011; R. Yolken, 2004) (Torrey et al., 2012). However, research has also shown HSV1 to be associated with significantly worse health outcomes for individuals with schizophrenia including cognitive deficits and gray matter loss (Dickerson et al., 2003; Prasad et al., 2011).

Schizophrenia and Bipolar Disorder

Schizophrenia is a chronic and debilitating mental disorder that is typically characterized by deficits in thought and emotional processes or perceptions; the National Institute of Mental Health currently estimates schizophrenia prevalence to be 1.1% of the US Adult population (Regier et al., 1993). Schizophrenia's onset typically occurs between ages 16-30 but is now generally regarded in a developmental perspective with a spectrum of severity and full psychosis regarded as the later more severe aspect of the syndrome (Health,

2016). Schizophrenia is also known to be highly heritable but also highly polygenic in that it lacks even a handful of specific very strongly associated genes (Debnath et al., 2013).

Schizophrenia (and family history of schizophrenia) has been associated with increased risk of auto-immune disease (Benros et al., 2014). Additionally previous research has identified that autozygosity (a specific type of homozygosity where two identical chromosomal segments are inherited from a common ancestor) across the genome is associated with increased schizophrenia risk (Keller et al., 2012). Though this prior research did not assess MHC specifically, it strengthens the hypothesis that homozygosity in the MHC region could potentially increase risk for schizophrenia as well as infection.

Bipolar disorder is characterized by dramatic changes in energy and mood that are debilitating to day to day life; prevalence is currently estimated to be 2.6% of the US adult population (Kessler, Chiu, Demler, & Walters, 2005). Bipolar disorder, like schizophrenia, is known to have high heritability but also be highly polygenic; many recent studies including GWAS have hypothesized this is no coincidence but indicates that the diseases share common genetic profiles (Avramopoulos et al., 2015; Cardno & Owen, 2014) (Fallin et al., 2005). Additionally infections have also been associated with bipolar disorder, most notably *Toxoplasma gondii* and Cytomegalovirus (though with less consistency than schizophrenia) (Hamdani et al., 2013; B. D. Pearce, D. Kruszon-Moran, & J. L. Jones, 2012; Rizzo et al., 2013). Given these similarities studying the diseases together (or compared to one another) is prudent.

Multiple genetic, environmental, and immunologic factors have been implicated in schizophrenia as well as the pathogenesis of bipolar disorder but no definitive neurobiological mechanism has yet been established (Debnath et al., 2013). However, many recent studies (including GWAS) have strongly implicated the MHC region as being

determinants of one's schizophrenia risk (Debnath et al., 2013). Since these genes are themselves known to be involved in the immune system and in many cases also associated with infections which are known risk factors for schizophrenia, examining all three is concert would be informative (Debnath et al., 2013).

Given the aforementioned fact that both schizophrenia and bipolar disorder have significant genetic and heritable components it then follows that family members of diseased individuals would share genetics that could be associated with the above infections whether these individuals are symptomatic of schizophrenia/bipolar disorder or not.

Current Study

The current study seeks to study the interplay between genetics, mental illness, and infection in a single analysis via analyzing HLA types, schizophrenia and bipolar disorder, and several infections of interest in one cohort. Most previous analyses have focused on only one aspect of this relationship, leaving larger questions on how these aspects interact. Using a cohort with available data on mental health, infection and immune markers, and genetics we carried out an analysis to identify the associations of specific HLA Alleles with the infections *Toxoplasma gondii*, Cytomegalovirus, and Herpes simplex 1 and determine whether or not these associations are modified by an individual's schizophrenia or bipolar disorder disease status. This analysis should provide additional concrete data to substantiate the suspected relationship between these factors.

Additionally with the available information, we are able to examine not only specific alleles but whether or not homozygosity at any particular HLA loci is associated with seropositivity in any of the above infections. This relationship is one which has substantial biological plausibility but has not been examined in a large cohort. By also stratifying this

analysis by mental disease status we hope to illuminate if this homozygosity is important in this relationship and deserving of further examination.

The availability of data on mental disease free family members of cases also presents an interesting opportunity since these individuals will share the genetics of cases but lack disease symptoms. This will allow for the assessment of whether a given allele is associated with an increased risk of infection both in the presence and absence of mental disease symptoms.

Methods

Subjects

The original cohort utilized in this study included research participants with a diagnosis of schizophrenia (including schizoaffective disorder) or bipolar disorder recruited over a 15-year period (1996–2011) through advertisements, talks, letters to leaders and service providers of the Jewish community, and a web site (Avramopoulos et al., 2015). Controls (termed EpiGen controls here) were recruited over a four-year period (2003–2007) at Jewish community professional meetings, community centers and synagogues (Avramopoulos et al., 2015). All participants reported four grandparents of Ashkenazi ethnicity. Examiners were blinded to the subject's diagnosis. Most of the subjects were seen in their homes. Blood for DNA and plasma from cases parents and controls were collected and kept frozen at -18°C (Avramopoulos et al., 2015).

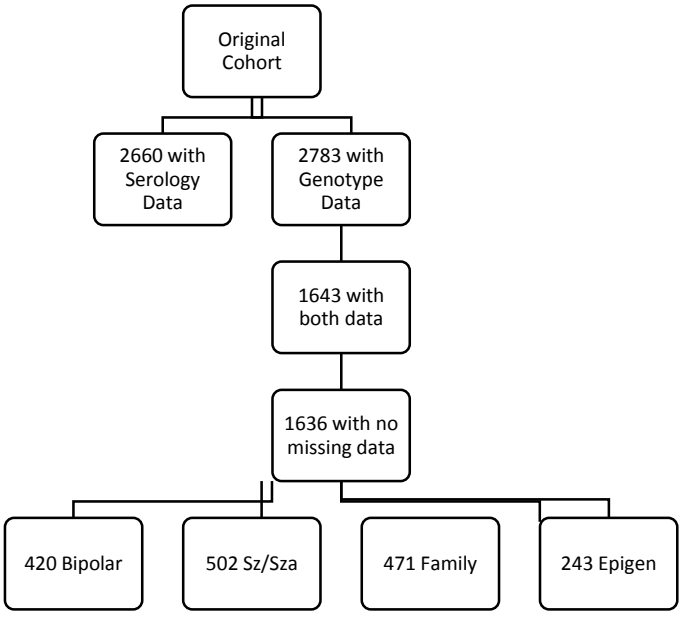

Figure 1: Subject Flow Diagram

The Johns Hopkins institutional review board approved the recruitment methods, protocols and informed consent documents and all subjected provided written informed consent (Avramopoulos et al., 2015). For the current study we possessed serology data on 2660 subjects and genotype data on 2783 subjects; of these 1643 subjects had both serology and genotype data. This led

after data cleaning to a final value of 1636 subjects. (Figure 1) The Controls included two groups, disease free EpiGen controls and mental disease free immediate family members of cases. Family members were recruited and examined in the same manner as cases: EpiGen controls were screened, non-mentally diseased, non-related to cases Ashkenazi individuals who had their data collected by the Epidemiology-Genetics program in psychiatry at Johns Hopkins University (Avramopoulos et al., 2015).

Immunoassay and Genotyping Measurements and cleaning

As shown in the figure, data for analysis included 2660 subjects' plasma sample assay results. Plasma IgG class anti-HSV1, anti-CMV, anti-TOXO, were tested using previously described immunoassay methods (Avramopoulos et al., 2015). A previous study had identified a bimodal distribution for these three variables of interest and determined a cut-point for each to convert these to dichotomous variables; additionally these values had been previously adjusted for the assay plate and storage years of each sample (Avramopoulos et al., 2015). DNA from blood was extracted with the Gentra Puregene Kit or the QIAGEN DNeasy Blood and Tissue Kit. Genotyping was performed with the Affymetrix Human Genome-Wide SNP Array 6.0 at Emory University and has been previously described (Mulle et al., 2010). Genotypes were called using the corrected robust linear mixture model (CRLMM), an algorithm for preprocessing and genotype calling of Affymetrix SNP array data (Scharpf, Irizarry, Ritchie, Carvalho, & Ruczinski, 2011). Genotype data cleaning was performed using the software package PLINK using previously described methods (Purcell et al., 2007).

HLA Imputation

Every subject with available genotype data had 7 HLA types imputed (A, B, C, DRB1, DQA1, DQB1, DPB1) utilizing the HIBAG software package for R (Zheng et al.,

2014). This software package allows for imputation without the use of large training datasets by using published parameter estimates for given populations; for this study published estimates for subjects of European descent run on Affymetrix Human Genome-Wide SNP Array 6.0 were utilized (Zheng et al., 2014). The program utilizes the concept of attribute bagging, an ensemble classifier method, with haplotype inference for SNPs and HLA types. Attribute bagging is a technique which improves the accuracy and stability of classifier ensembles deduced using bootstrap aggregating and random variable selection, further information on the computational methods are available (Zheng et al., 2014). By utilizing this program in conjunction with the genotype data we were able to impute both alleles at 7 HLA loci by utilizing SNPs within the xMHC region of chromosome 6. Through this method we obtained two imputed 4 digit alleles (as XX:XX) for each of the seven HLA types for each subject.

HIBAG provides not only alleles but also Bayesian posterior probabilities as output, the original authors of the program discuss the possibility of exclusion of observations with low values but also caution that while doing so can possibly increase prediction accuracy it can also drastically reduce call rates (Zheng et al., 2014). The vast majority of published literature using this software do not establish such exclusions (Abraham, Rohmer, Tye-Din, & Inouye, 2015; Chang et al., 2015; Khankhanian, Gourraud, Lizee, & Goodin, 2015; Nunes et al., 2016; Ollila et al., 2015; Parham et al., 2016). Given this and the high average posterior probabilities of our data (see A1) we utilized the full breadth of available data.

Statistical Analyses

The newly imputed HLA types were combined with the pre-existing serology data to form a combined set of 1643 subjects. Data cleaning and all additional analyses were conducted using SAS version 9.4. Data cleaning revealed 7 controls with missing data for

age who were excluded leaving a final count of 1636 subjects. Additional variables were generated for analysis purposes, these notably included a multi-infection index (0-3) indicating how many of the infections of interest a given subject was positive for, and homozygosity variables. Table 1 describes the sample and relevant variables. The homozygosity variables included dichotomous variables for each HLA type (e.g. A, B, C) as well as a dichotomous variable indicating if a subject is homozygotic at any of the 7 possible loci (0 if heterozygous at all 7 HLA loci) and finally an index variable indicating how many (0-7) of the HLA loci an individual was homozygotic at.

Utilizing SAS proc freq the imputed alleles at each HLA loci were assessed and the most common in the sample were identified. (Tables 2, 3). A cutoff of 3.75% prevalence in the full sample was utilized to determine a "common" allele. For each of these common alleles we additionally examined their percent prevalence in the seropositive portions of the sample for each of the three infections of interest as well as the family controls.

All analyses were conducted in a stratified fashion due to the presence of genetically related individuals in the cohort. In this data many subjects without mental illness were family members of cohort members with mental illness so comparison of these individuals in standard statistical models would be a fallacy as their genetics (the predictors of interest) are not independent. However, by stratifying on mental disease status into four strata (schizophrenia, bipolar, family non-disease, and EpiGen non-disease) not only were we able to solve the above independence issue but also observe if the effect of genetics is modified by mental disease status.

We generated dichotomous variables for each identified common allele indicating if an individual possessed one or more copies; this variable did not differentiate between heterozygotes and homozygotes. Each of these dichotomous common allele variables was

used as a predictor in a logistic model using SAS 9.4 proc logistic for each of the three infections of interest as well as the multi-infection index controlling for age and sex. The models for the multi-infection index utilized this variable as an ordinal outcome and were run as proportional odds models. These models were additionally run stratified on neuropsychiatric disease status. There were four strata, schizophrenia (including schizoaffective disorder), bipolar disorder, EpiGen controls, and family of cases. Models were also run using only two strata (mentally ill versus non-mentally ill) which are available in the appendix (A4-A15).

A second analysis was conducted concerning the effects of homozygosity at each HLA loci. A dichotomous variable indicating whether a subject was homozygotic or not at each loci (A, B, C, DRB, DQA, DQB, DPB) was used as a predictor for each infection of interest as well as the multi-infection index controlling for age and sex. As in the previous analysis the models regressing on the multi-infection index were ordinal logistic regressions using the proportional odds model. Additional models were run for each of these four dependent infection variables with a variable indicating if an individual was homozygous at any of the seven HLA loci as well as the total number of homozygotic loci variable as predictors. Finally two additional models for each dependent infection variable were run grouping the HLA types into their appropriate MHC classes. The first model contained the dichotomous homozygosity variables for HLA A, B, and C as predictors for infection controlling for sex and age, and the second model containing the dichotomous homozygosity variables for HLA DRB, DQA, DQB, and DPB as predictors for infection controlling for sex and age. Given the similar biological purposes of the HLA alleles within each MHC class, examining whether homozygosity at a given HLA locus is a factor for infection controlling for homozygosity in the other HLA loci of that class in addition to independently would

reveal if any observed effects are driven by homozygosity at that particular HLA locus or simply homozygosity in that MHC class as a whole.

Since each analysis was run using a separate model there was potential for multiple comparison issues, and to counteract this a Bonferroni correction was applied to alpha, results that remained significant are noted with an asterisk in the following tables. For the common allele analyses this correction was 0.05/48 resulting yielding an alpha of 0.00104 with 48 being the number of alleles tested. For the homozygosity analysis this correction was 0.05/11 resulting in a cutoff of 0.0045 with 11 being the number of models run for each analysis.

Table 1. Sample Characteristics

N=1636	N (%) or Mean (Std)
Demographics	
Age at Plasma	53.1 (15.2)
Sex (Female)	777 (47.49%)
Mental Illness	
Sz/Sza*	502 (30.68%)
Bipolar**	420 (25.67%)
Family	471 (28.79%)
EpiGen	234 (14.85%)
Infections	
T. gondii Positive	295 (18.03%)
Cytomegalovirus Positive	589 (36.00%)
Herpes Simplex Virus 1 Positive	638 (39.00%)
Multi-Infection Index	
3 Infections	89 (5.44%)
2 Infections	331 (20.23%)
1 Infection	593 (36.25%)
Homozygosity	
Homozygous HLA A	160 (9.78%)
Homozygous HLA B	136 (8.31%)
Homozygous HLA C	225 (13.75%)
Homozygous HLA DRB	160 (9.78%)
Homozygous HLA DQA	244 (14.91%)
Homozygous HLA DQB	229 (14.00%)
Homozygous HLA DPB	427 (26.71%)
Homozygous Any	799 (48.84%)
Homozygous Total	
7 HLAs	21 (1.28%)
6 HLAs	20 (1.22%)
5 HLAs	26 (1.59%)
4 HLAs	37 (2.26%)
3 HLAs	109 (6.66%)
2 HLAs	133 (8.13%)
1 HLAs	452 (27.69%)

*schizophrenia and schizoaffective disorder were classified together for this analysis.
**There were no subjects with both Sz/Sza and bipolar disorder

Results

Sample Characteristics

The characteristics of the sample are displayed in Table 1. Sample characteristics by strata are available in appendix tables A1 and A2. The average age of the sample at the time plasma was taken was 53.1 years old with a standard deviation of 15.2 years and 47.49% of the sample was female. In this dataset those diagnosed with schizophrenia and schizoaffective disorder were grouped and analyzed together, there were a total of 502 of these individuals in the sample. This grouped condition is denoted as Sz/Sza on all following tables and figures. There were 420 individuals diagnosed with bipolar disorder in the sample, there were no individuals who were both bipolar and Sz/Sza positive in the sample. Family of Sz/Sza/BP cases made up 28.79% of the sample, the remaining 14.85% were AJ EpiGen controls. The prevalence of the infections of interest in this sample varied from 18.03% for *T. gondii*. to 39.00% for Herpes Simplex Virus 1. Notably there was a fair amount of coinfection with

Table 2: Common Alleles, MHC Class I

HLA A	Percent Total Population	Family	TOXO+	CMV+	HSV1+
01:01	12.69	12.84	13.63	13.3	14.17
11:01	5.17	4.88	4.88	5.05	5.84
02:01	16.59	15.18	13.63	15.91	15.73
02:05	3.93	3.82	4.71	4.29	3.82
24:02	9.13	7.96	7.74	8.84	8.8
26:01	14.97	17.83	15.15	13.8	15.34
3:01	9.03	8.6	9.26	8.25	8.33
33:01	3.8	3.18	3.7	3.28	3.58
Other	24.69	25.71	27.3	27.28	24.39

HLA B	Percent Total Population	Family	TOXO+	CMV+	HSV1+
14:02	11.87	9.87	11.45	11.95	11.06
35:01	5.33	5.1	5.89	5.39	5.45
35:02	8.67	9.02	7.91	9.26	8.49
35:03	3.9	3.29	3.87	4.71	3.35
38:01	17.83	20.06	17.17	15.66	16.82
52:01	5.05	4.67	5.22	4.55	4.98
57:01	4.44	4.24	6.06	5.39	5.06
08:01	4.08	3.93	6.06	3.7	5.06
Other	38.83	39.82	36.37	39.39	39.73

HLA C	Percent Total Population	Family	TOXO+	CMV+	HSV1+
12:02	5.05	4.67	5.22	4.63	4.98
12:03	20.97	23.04	20.03	19.36	19.24
04:01	19.54	19.11	18.86	20.79	19
06:02	11.23	10.93	12.96	12.46	11.6
07:01	8.19	9.24	11.62	7.91	10.2
08:02	11.9	10.08	11.62	11.87	11.06
Other	23.12	22.93	19.69	22.98	23.92

25.67% of the sample having 2 or 3 of these infections and 5.44% having all 3. Prevalence of homozygosity at the imputed HLA loci was predictably fairly high given the subjects were drawn from an Ashkenazi cohort, ranging from 8.31% at HLA B to 26.71% at HLA DPB. Notably 48.84% of the sample was homozygotic at least one of the 7 imputed HLA loci.

Common Alleles

The alleles which had a prevalence of at least 3.75% in the full sample were then chosen to be further analyzed. Tables 2 and 3 display the most common alleles for each HLA loci in MHC class I and class II respectively in the total population as well as the prevalence of those alleles in the seropositive for each infection of interest and in family members of cases.

Table 3: Common Alleles, MHC Class II

HLA DRB	Percent Total Population	Family	TOXO+	CMV+	HSV1+
01:02	10.1	8.49	9.76	9.93	9.74
11:01	12.87	12.63	13.97	13.13	13.24
11:04	5.23	4.35	4.55	5.55	4.83
13:01	7.21	7.43	7.41	6.82	6.85
13:02	4.08	4.25	3.87	3.03	4.75
15:02	3.77	3.82	3.7	3.54	3.43
03:01	5.72	5.52	7.07	5.81	7.09
04:02	15.82	17.3	16.33	15.91	15.19
07:01	14.36	15.61	14.65	15.32	15.11
Other	20.84	20.6	18.69	20.96	19.77

HLA DQA	Percent Total Population	Family	TOXO+	CMV+	HSV1+
01:01	11.78	9.87	10.77	11.53	10.98
01:02	8.46	8.6	6.9	7.07	9.11
01:03	11.26	11.46	11.45	10.61	10.9
02:01	14.46	15.5	15.15	15.24	15.19
03:01	17.5	18.9	18.35	17.26	17.06
05:01	5.75	5.41	7.24	5.72	7.01
05:05	22.76	21.66	22.9	23.57	22.27
Other	8.03	8.6	7.24	9	7.48

HLA DQB	Percent Total Population	Family	TOXO+	CMV+	HSV1+
02:01	5.72	5.31	7.24	5.64	6.93
02:02	12.54	13.38	12.79	13.05	12.85
03:01	24.35	23.25	23.91	25.84	24.45
03:02	17.56	19.11	18.18	17.17	16.59
05:01	13.27	11.78	12.63	13.72	12.54
06:03	7.36	7.64	7.74	7.07	7.32
Other	19.2	19.53	17.51	17.51	19.32

HLA DPB	Percent Total Population	Family	TOXO+	CMV+	HSV1+
104:01	5.23	4.35	4.55	5.98	5.61
02:01	25.32	24.84	25.42	24.83	25.86
04:01	43.12	44.8	44.61	42.09	42.29
04:02	9.16	8.28	9.43	10.02	9.27
Other	17.17	17.73	15.99	17.08	16.97

The prevalence of each allele remains relatively consistent between the total sample, each infection, and family. Some alleles were particularly prevalent in the total sample such as HLA C*12:03, HLA DQA*05:05, HLA DQB*03:01, and HLA DPB*04:01 which each had over 20% prevalence in the sample. HLA DPB*04:01 in particular had a prevalence of 43.12% in the sample. The family group was of particular interest; due to bias concerns we examined the prevalence of these alleles in this group to observe whether this group was having a significant effect on observed allele prevalence in the total sample and therefore altering our selection of common alleles. This did not appear to be the case and therefore our selection of common alleles was not skewed by the inclusion of this group.

Regression analysis testing the associations of HLA type predictor variables with infection status outcomes are shown in Tables 4-7, which show those alleles for which at least one statistically significant result was obtained in stratified logistic modeling. Larger versions of these tables stratified as mentally ill vs controls with all alleles tested can be found in Appendix A (A4-A11).

Table 4 shows those alleles with statistically significant results for *T. gondii*. In the schizophrenia group there were many alleles had significance at alpha=0.05 but none maintained their significance after Bonferroni corrections. In this group HLA B*08:01, HLA C*07:01, HLA DRB*03:01, HLA DQA*05:01, and HLA DQB*02:01 were significantly ($p<=0.05$) associated with increased odds of *T. gondii* seropositive status and HLA C*04:01 was significantly associated with decreased odds of *T. gondii* seropositivity. In the bipolar disorder group only one allele was significant at $p<=0.05$; HLA DQA*01:01 was associated with increased odds of *T. gondii* seropositivity. In the EpiGen controls there were several alleles significant at alpha=0.05 but none maintained this after Bonferroni correction. At $p<=0.05$, HLA B*08:01, HLA C*04:01, HLA DRB*03:01, HLA DQA*05:01, and HLA DQB*02:01 were all significantly associated with increased odds of *T. gondii* seropositivity. For the Family of cases groups there were two alleles significant at $p<=0.05$; HLA DQA*01:01 and HLA DPB*04:02 were both associated with decreased odds of seropositivity in the family group.

Table 5 shows those alleles with statistically significant results for Cytomegalovirus. In the schizophrenia group there was one allele significant at alpha=0.05; HLA B*57:01 was significantly associated with increased odds of CMV seropositivity. In the bipolar disorder group there were no alleles significant at alpha=0.05. In the EpiGen controls there were two alleles significant at alpha=0.05; HLA DRB*13:02 and HLA DQA*01:02 were significantly

associated with decreased odds of CMV seropositivity. For the Family of cases group there were two alleles significant at 0.05; HLA B*38:01 and HLA DRB*13:02 were both associated with decreased odds of seropositivity in the family group.

Table 6 shows those alleles with statistically significant results for Herpes Simplex Virus 1. In the schizophrenia group there were no alleles significant at alpha=0.05. In the bipolar disorder group there were no alleles significant at alpha=0.05. In the EpiGen controls there were many alleles significant at alpha=0.05 and one significant after Bonferroni correction; HLA C*07:01 was significantly associated with increased odds (OR 3.4, 95% CI 1.7-6.9) of HSV1 seropositivity after Bonferroni correction. Among EpiGen controls HLA B*08:01, HLA DRB*03:01, HLA DQA*05:01, and HLA DQB*02:01 were significantly ($p<=0.05$) associated with increased odds of HSV1 seropositivity; HLA B*38:01 and HLA C*12:03 were associated with decreased odds of HSV1 seropositivity. In the Family of cases group HLA A*01:01 and HLA C*07:01 were significantly ($p<=0.05$) associated with increased odds of HSV1 seropositivity. HLA C*12:03 (p-value 0.001, significant after Bonferroni correction) and HLA B*38:01 (p-value 0.03) were associated with decreased odds of HSV1 seropositivity.

Table 7 displays those alleles which had significant effects for models where the dependent variable was the multi-infection index. These models were run as ordinal logistic regressions using the proportional odds assumption, therefore the displayed odds ratios indicate the ratio of odds for number of infections being one higher between a person with the allele and without it. In the schizophrenia group there were no alleles significant at alpha=0.05. In the bipolar disorder group there was one allele significant at alpha=0.05; HLA DPB*04:01 was associated with decreased odds of number of infections. In the EpiGen controls there were many significant alleles including several after Bonferroni corrections.

HLA B*38:01 and HLA C*12:03 were significantly (after Bonferroni) associated with decreased odds of increased number of infections among the EpiGen controls. HLA B*08:01, HLA C*07:01, HLA DRB*03:01, HLA DQA*05:01, and HLA DQB*02:01 were significantly (p-value<=0.05) associated with increased odds of increased number of infections among the EpiGen group. The Family of cases group had several alleles significant at alpha=0.05 but none significant after Bonferroni corrections. HLA A*01:01, HLA B*57:01, HLA C*07:01 were significantly (p-value<=0.05) associated with increased odds of increased number of infections among the Family group. HLA B*38:01 and HLA C*12:03 were significantly (0.05) associated with decreased odds of increased number of infections among the Family group.

Homozygosity Analysis

Given the information available we were able to conduct an additional analysis concerning homozygosity. Tables 8-11 display p-values and odds ratios from logistic models where whether an individual was homozygotic at the various HLA loci was used as a predictor for the infections of interest as well as the infection index. Appendix A contains tables stratified as mentally ill vs controls as well as tables using total number of homozygotic loci as a categorical variable (A12-A19).

Table 8 displays the results of the homozygosity analysis for *Toxoplasma gondii*. The only statistically significant results occurred in controls. In the EpiGen group there were statistically significant results for the MHC Class I model which contained the homozygosity variables for HLA A, B, and C controlling for sex and age. In this model homozygosity at HLA B was significantly associated with decreased odds of *T. gondii* seropositivity and homozygosity at HLA C was associated (0.05) with increased odds of *T. gondii* seropositivity. These results indicate that homozygosity HLA B and HLA C have opposite

effects when controlling for one another and that homozygosity in MHC class I is not consistent in its effect on HSV1 seropositivity risk. In the Family of cases group homozygosity at HLA DPB was significantly (0.05) associated with increased odds of *T. gondii* seropositivity, this association was also significant in the Class II model which included the homozygosity variables for HLA DRB, DQA, DQB, and DPB controlling for age and sex. This significance in the class II model indicates the observed effect is driven by homozygosity at HLA DPB and not other homozygosity in MHC class II since it remained significant after controlling for homozygosity at the other MHC class II HLA loci.

Table 9 displays the results of the homozygosity analysis for Cytomegalovirus. There were no statistically significant results at alpha=0.05 in the schizophrenia group, though homozygosity at HLA DPB was borderline both alone and in the Class II model. In the bipolar disorder group there was one statistically significant at alpha=0.05 result. In the Class II model homozygosity at HLA DRB was significantly associated with increased odds of CMV seroprevalence, additionally homozygosity at HLA DQA was borderline (p=0.078) associated with decreased odds of CMV seroprevalence. In the EpiGen group homozygosity at HLA DPB was associated with decreased odds of CMV seroprevalence both alone and in the Class II model; additionally homozygosity at HLA DQB was borderline (p-value=0.055) associated with increased odds of CMV seroprevalence in the Class II model. There were no statistically significant results at alpha=0.05 in the Family of cases group.

Table 10 displays the results of the homozygosity analysis for Herpes Simplex Virus 1. In the schizophrenia group there were two statistically significant at alpha=0.05 results. In the Class II model homozygosity at HLA DRB was associated with increased odds of HSV1 seropositivity and homozygosity at HLA DQA was significantly associated with decreased odds of HSV1 seropositivity. In the bipolar disorder group homozygosity at HLA DQA was

associated with decreased odds of HSV1 seropositivity alone however the association was not significant in the Class II model (p=0.083). In the EpiGen group being homozygous at any HLA loci was associated with increased odds of HSV1 seropositivity compared to those who were homozygous at no HLA loci. In the Family of cases group there were no statistically significant at alpha=0.05 results.

Table 11 displays the results of the homozygosity analysis for the Multi-Infection Index. In the schizophrenia group there were no statistically significant at alpha=0.05 results, however homozygosity at HLA DPB was borderline associated with increased odds of increased number of infections individually (p=0.0575) as well as in the Class II model (p=0.08). In the bipolar disorder group there were statistically significant results in the Class II model. Homozygosity at HLA DRB was significantly (after Bonferroni) associated with increased odds of increased number of infections and homozygosity at HLA DQA was significantly (p-value<=0.05) associated with decreased odds of increased number of infections. There were no statistically significantly results at alpha=0.05 in the EpiGen group. In the Family of cases group homozygosity at HLA C was significantly associated (p-value <=0.05) with increased odds of increased number of infections in the Class I model. Additionally homozygosity at HLA B was borderline (0.07) associated with decreased odds of increased number of infections in the Class I model.

Results Tables

Table 4: Results of Stratified Logistic models for statistically significant alleles predicting dichotomous *T. gondii* outcome controlling for age and sex

	Sz/Sza				Bipolar				EpiGen				Family			
HLA B	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
08:01	0.0149	3.267	1.259	8.476	0.7076	1.213	0.441	3.335	0.0104	3.423	1.335	8.779	0.2783	1.477	0.73	2.988
HLA C	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
04:01	0.0162	0.443	0.228	0.86	0.2944	0.689	0.343	1.383	0.0044	2.628	1.353	5.107	0.4806	0.86	0.565	1.308
07:01	0.0268	2.177	1.093	4.337	0.1027	1.926	0.876	4.231	0.0751	2.109	0.927	4.796	0.3264	1.289	0.776	2.142
HLA DRB	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
03:01	0.0401	2.319	1.039	5.176	0.896	0.936	0.345	2.536	0.0362	2.526	1.061	6.014	0.6678	1.15	0.607	2.182
HLA DQA	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
01:01	0.6382	1.165	0.616	2.206	0.0451	2.025	1.016	4.038	0.5493	0.774	0.335	1.789	0.0324	0.54	0.307	0.949
05:01	0.0401	2.319	1.039	5.176	0.8619	0.915	0.338	2.478	0.0135	2.923	1.248	6.842	0.5672	1.207	0.634	2.296
HLA DQB	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
02:01	0.0401	2.319	1.039	5.176	0.8619	0.915	0.338	2.478	0.0135	2.923	1.248	6.842	0.4985	1.25	0.655	2.386
HLA DPB	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
04:02	0.3444	1.389	0.703	2.748	0.0455	2.087	1.015	4.292	0.4108	1.41	0.622	3.198	0.0065	0.413	0.218	0.781

Table 5: Results of Stratified Logistic models for statistically significant alleles predicting dichotomous CMV outcome controlling for age and sex

	Sz/Sza				Bipolar				EpiGen				Family			
HLA A	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
3:01	0.6829	1.131	0.627	2.041	0.9283	0.975	0.559	1.699	0.1554	0.592	0.287	1.22	0.1803	0.708	0.428	1.173
HLA B	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
35:03	0.9474	0.972	0.423	2.237	0.4406	1.324	0.649	2.702	0.0981	2.288	0.858	6.104	0.0605	2.198	0.966	5
38:01	0.5503	0.863	0.533	1.399	0.4957	0.85	0.532	1.357	0.1377	0.628	0.34	1.161	0.0135	0.611	0.414	0.903
57:01	0.0413	2.057	1.029	4.114	0.3702	1.371	0.687	2.735	0.711	0.812	0.27	2.441	0.3799	1.361	0.684	2.711
HLA DRB	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
13:02	0.8339	1.092	0.481	2.479	0.8216	0.908	0.394	2.094	0.0138	0.151	0.034	0.68	0.0394	0.477	0.236	0.965
HLA DQA	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
01:02	0.2879	1.374	0.765	2.466	0.4419	0.786	0.425	1.453	0.0083	0.309	0.129	0.739	0.2468	0.74	0.444	1.232

Table 6: Results of Stratified Logistic models for statistically significant alleles predicting dichotomous HSV1 outcome controlling for age and sex

	Sz/Sza				Bipolar				EpiGen				Family			
HLA A	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
01:01	0.4146	0.817	0.503	1.327	0.3879	1.226	0.772	1.947	0.1186	1.654	0.879	3.113	0.008	1.81	1.168	2.806
HLA B	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
38:01	0.1317	1.379	0.908	2.095	0.9754	0.993	0.63	1.564	0.03	0.517	0.285	0.938	0.0225	0.641	0.437	0.939
08:01	0.3843	1.423	0.642	3.154	0.9521	0.979	0.491	1.953	0.0079	3.323	1.369	8.064	0.1223	1.755	0.86	3.58
HLA C	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
12:03	0.5533	1.131	0.753	1.697	0.91	1.026	0.662	1.588	0.005	0.443	0.251	0.783	0.001*	0.529	0.363	0.772
07:01	0.5847	1.167	0.671	2.031	0.6037	1.165	0.655	2.07	0.0007*	3.387	1.668	6.878	0.0486	1.651	1.003	2.717
HLA DRB	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
03:01	0.1655	1.551	0.834	2.882	0.7656	1.099	0.591	2.044	0.0146	2.626	1.21	5.698	0.1004	1.683	0.904	3.131
HLA DQA	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
05:01	0.1655	1.551	0.834	2.882	0.837	1.067	0.576	1.977	0.0238	2.403	1.123	5.138	0.1174	1.648	0.882	3.078
HLA DQB	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
02:01	0.1655	1.551	0.834	2.882	0.837	1.067	0.576	1.977	0.0238	2.403	1.123	5.138	0.1561	1.576	0.841	2.954
HLA DPB	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
04:01	0.4685	0.856	0.562	1.303	0.0882	0.687	0.447	1.058	0.9309	0.976	0.559	1.703	0.1512	0.749	0.505	1.111

*Significant after Bonferroni correction for multiple comparisons

Table 7: Results of Stratified Logistic models for statistically significant alleles predicting ordinal multi-infection outcome controlling for age and sex

	Sz/Sza				Bipolar				EpiGen				Family			
HLA A	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
01:01	0.4306	0.845	0.555	1.285	0.2488	1.274	0.844	1.922	0.4735	1.235	0.693	2.202	0.0042	1.762	1.195	2.599
33:01	0.2666	1.427	0.762	2.671	0.8747	1.058	0.526	2.129	0.1077	0.489	0.205	1.169	0.2698	0.681	0.344	1.348
HLA B	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
35:03	0.3529	0.727	0.371	1.424	0.743	0.898	0.472	1.71	0.0964	2.079	0.877	4.928	0.1379	1.661	0.85	3.248
38:01	0.0867	1.385	0.954	2.01	0.3493	0.825	0.552	1.234	0.0003*	0.371	0.217	0.634	0.0028	0.587	0.414	0.832
57:01	0.0649	1.743	0.966	3.145	0.2094	1.48	0.802	2.731	0.7135	1.187	0.476	2.957	0.0319	1.927	1.059	3.508
08:01	0.5287	1.264	0.61	2.617	0.3299	0.734	0.394	1.367	0.0037	3.235	1.465	7.142	0.0876	1.716	0.924	3.188
HLA C	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
12:03	0.179	1.279	0.893	1.83	0.2425	0.793	0.538	1.17	0.0006*	0.416	0.252	0.688	0.0016	0.576	0.409	0.811
07:01	0.4223	1.223	0.748	1.999	0.5654	0.859	0.511	1.443	0.0053	2.469	1.309	4.658	0.015	1.728	1.112	2.686
HLA DRB	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
03:01	0.3143	1.339	0.758	2.364	0.8838	1.042	0.6	1.811	0.0016	3.122	1.54	6.332	0.1551	1.486	0.861	2.563
HLA DQA	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
05:01	0.3143	1.339	0.758	2.364	0.9943	1.002	0.579	1.735	0.0016	3.069	1.528	6.165	0.1553	1.491	0.859	2.587
HLA DQB	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
02:01	0.3143	1.339	0.758	2.364	0.9943	1.002	0.579	1.735	0.0016	3.069	1.528	6.165	0.1985	1.44	0.826	2.51
HLA DPB	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
04:01	0.9136	1.021	0.702	1.486	0.0389	0.668	0.456	0.98	0.6341	0.885	0.535	1.463	0.3353	0.84	0.589	1.198

*Significant after Bonferroni correction for multiple comparisons

Table 8: Results of Stratified Logistic models for homozygosity variables predicting dichotomous *T. gondii* outcome controlling for age and sex

	Sz/Sza				Bipolar				EpiGen				Family			
	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU
Homozy HLA A	0.759	1.158	0.453	2.966	0.918	1.059	0.351	3.196	0.867	1.082	0.427	2.742	0.632	0.838	0.406	1.728
Homozy HLA B	0.620	1.266	0.499	3.214	0.929	1.059	0.300	3.736	0.296	0.442	0.096	2.041	0.758	1.127	0.528	2.407
Homozy HLA C	0.869	0.933	0.412	2.116	0.802	1.137	0.416	3.108	0.120	1.931	0.842	4.427	0.597	1.165	0.662	2.048
Homozy HLA DRB	0.810	1.121	0.443	2.836	0.347	1.572	0.612	4.035	0.666	1.305	0.390	4.366	0.697	0.874	0.443	1.723
Homozy HLA DQA	0.810	0.905	0.403	2.036	0.488	1.365	0.567	3.284	0.101	1.987	0.874	4.516	0.717	1.110	0.630	1.957
Homozy HLA DQB	0.933	0.966	0.429	2.174	0.595	1.289	0.506	3.279	0.183	1.820	0.754	4.390	0.810	0.930	0.513	1.684
Homozy HLA DPB	0.318	1.371	0.738	2.544	0.839	1.080	0.515	2.262	0.671	0.853	0.409	1.777	0.044	1.547	1.012	2.367
Homozy Any	0.785	1.082	0.614	1.909	0.625	1.176	0.615	2.248	0.786	1.093	0.574	2.084	0.134	1.355	0.911	2.017
Total # Homozy	0.711	1.034	0.865	1.236	0.487	1.084	0.863	1.362	0.343	1.112	0.893	1.386	0.483	1.051	0.914	1.210
CLASS I HLA A	0.814	1.122	0.430	2.928	0.921	1.058	0.348	3.221	0.621	1.279	0.483	3.387	0.557	0.797	0.373	1.702
CLASS I HLA B	0.356	2.196	0.413	11.690	0.888	0.877	0.143	5.390	0.017	0.109	0.017	0.677	0.929	1.049	0.371	2.966
CLASS I HLA C	0.411	0.541	0.125	2.339	0.785	1.222	0.290	5.152	0.005	4.726	1.612	13.853	0.672	1.176	0.555	2.494
CLASS II HLA DRB	0.568	1.674	0.286	9.795	0.517	1.888	0.276	12.927	0.502	0.583	0.121	2.812	0.355	0.586	0.189	1.816
CLASS II HLA DQA	0.591	0.583	0.081	4.173	0.946	1.078	0.122	9.544	0.292	2.241	0.500	10.044	0.121	2.716	0.769	9.595
CLASS II HLA DQB	0.939	1.073	0.176	6.530	0.769	0.740	0.099	5.534	0.873	1.146	0.216	6.077	0.303	0.517	0.148	1.811
CLASS II HLA DPB	0.340	1.354	0.726	2.523	0.821	1.089	0.519	2.286	0.725	0.876	0.420	1.830	0.036	1.582	1.031	2.427

Table 9: Results of Stratified Logistic models for homozygosity variables predicting dichotomous CMV outcome controlling for age and sex

	Sz/Sza				Bipolar				EpiGen				Family			
	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU
Homozy HLA A	0.486	1.295	0.627	2.675	0.385	0.711	0.329	1.536	0.313	1.486	0.689	3.205	0.755	0.901	0.468	1.734
Homozy HLA B	0.716	1.146	0.550	2.388	0.770	1.121	0.522	2.405	0.776	0.862	0.309	2.401	0.371	0.715	0.342	1.492
Homozy HLA C	0.242	1.422	0.788	2.567	0.717	1.125	0.596	2.122	0.785	0.897	0.410	1.960	0.626	1.148	0.659	2.001
Homozy HLA DRB	0.507	0.768	0.352	1.674	0.095	1.732	0.908	3.302	0.310	0.545	0.169	1.759	0.149	0.628	0.334	1.180
Homozy HLA DQA	0.533	0.819	0.438	1.534	0.685	1.131	0.625	2.045	0.938	1.030	0.485	2.190	0.318	0.758	0.440	1.305
Homozy HLA DQB	0.649	0.864	0.462	1.618	0.328	1.353	0.739	2.479	0.394	1.407	0.642	3.082	0.813	0.935	0.534	1.635
Homozy HLA DPB	0.065	1.572	0.972	2.543	0.860	1.045	0.640	1.708	0.007	0.395	0.201	0.778	0.418	0.843	0.558	1.274
Homozy Any	0.510	1.160	0.747	1.802	0.717	0.925	0.607	1.410	0.371	0.779	0.451	1.346	0.689	0.926	0.637	1.348
Total # Homozy	0.545	1.043	0.910	1.195	0.456	1.059	0.911	1.232	0.478	0.929	0.757	1.139	0.318	0.933	0.814	1.069
CLASS I HLA A	0.512	1.287	0.605	2.735	0.357	0.694	0.318	1.512	0.285	1.532	0.701	3.352	0.969	0.986	0.498	1.954
CLASS I HLA B	0.376	0.603	0.197	1.848	0.892	1.082	0.349	3.354	0.774	0.818	0.207	3.226	0.109	0.424	0.148	1.212
CLASS I HLA C	0.165	1.879	0.772	4.574	0.828	1.109	0.435	2.827	0.930	0.954	0.337	2.703	0.147	1.797	0.814	3.965
CLASS II HLA DRB	0.592	0.714	0.209	2.443	0.043	4.289	1.051	17.510	0.088	0.249	0.051	1.228	0.161	0.442	0.141	1.384
CLASS II HLA DQA	0.813	0.852	0.225	3.225	0.078	0.197	0.033	1.197	0.317	0.415	0.074	2.326	0.370	0.545	0.144	2.056
CLASS II HLA DQB	0.810	1.167	0.331	4.113	0.357	2.002	0.457	8.767	0.055	6.337	0.958	41.921	0.106	2.907	0.796	10.609
CLASS II HLA DPB	0.054	1.614	0.993	2.624	0.944	1.018	0.620	1.670	0.009	0.397	0.199	0.791	0.390	0.833	0.550	1.263

Table 10: Results of Stratified Logistic models for homozygosity variables predicting dichotomous HSV1 outcome controlling for age and sex

	Sz/Sza				Bipolar				EpiGen				Family			
	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU
Homozy HLA A	0.451	1.287	0.668	2.480	0.273	0.655	0.307	1.395	0.255	1.531	0.736	3.185	0.190	0.648	0.340	1.238
Homozy HLA B	0.661	1.161	0.597	2.256	0.931	0.966	0.449	2.082	0.726	1.179	0.470	2.959	0.291	0.678	0.330	1.395
Homozy HLA C	0.580	1.168	0.674	2.022	0.605	0.844	0.443	1.608	0.461	1.310	0.639	2.687	0.557	1.176	0.685	2.020
Homozy HLA DRB	0.400	1.315	0.695	2.486	0.509	0.795	0.402	1.572	0.287	1.665	0.651	4.256	0.549	0.828	0.448	1.533
Homozy HLA DQA	0.447	0.804	0.459	1.409	0.038	0.503	0.263	0.962	0.310	1.436	0.714	2.886	0.972	0.990	0.582	1.686
Homozy HLA DQB	0.945	0.981	0.567	1.696	0.068	0.535	0.274	1.046	0.494	1.296	0.617	2.721	0.568	0.853	0.494	1.472
Homozy HLA DPB	0.649	1.109	0.711	1.731	0.129	1.447	0.898	2.330	0.616	1.158	0.652	2.057	0.868	1.035	0.691	1.551
Homozy Any	0.429	0.852	0.573	1.267	0.682	0.917	0.606	1.387	0.017	1.901	1.121	3.224	0.880	0.972	0.674	1.403
Total # Homozy	0.659	1.028	0.908	1.164	0.297	0.919	0.785	1.077	0.193	1.126	0.942	1.346	0.584	0.964	0.843	1.101
CLASS I HLA A	0.503	1.261	0.641	2.481	0.271	0.651	0.303	1.399	0.268	1.524	0.723	3.213	0.306	0.702	0.357	1.382
CLASS I HLA B	0.933	0.957	0.338	2.704	0.586	1.378	0.435	4.364	0.734	0.806	0.231	2.803	0.095	0.415	0.148	1.167
CLASS I HLA C	0.729	1.161	0.499	2.705	0.505	0.721	0.276	1.885	0.486	1.407	0.538	3.676	0.089	1.961	0.902	4.261
CLASS II HLA DRB	0.045	4.013	1.031	15.616	0.056	5.012	0.958	26.218	0.544	1.486	0.413	5.342	0.475	0.674	0.228	1.992
CLASS II HLA DQA	0.043	0.218	0.050	0.949	0.083	0.185	0.027	1.250	0.504	1.606	0.400	6.444	0.249	2.089	0.598	7.298
CLASS II HLA DQB	0.523	1.438	0.472	4.379	0.675	0.718	0.153	3.368	0.622	0.684	0.151	3.096	0.363	0.574	0.174	1.896
CLASS II HLA DPB	0.787	1.064	0.678	1.671	0.173	1.398	0.863	2.263	0.708	1.117	0.625	1.997	0.799	1.054	0.702	1.583

*Significant after Bonferroni correction for multiple comparisons

Table 11: Results of Stratified Logistic models for homozygosity variables predicting ordinal multi-infection outcome controlling for age and sex

	Sz/Sza				Bipolar				EpiGen				Family			
	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU
Homozy HLA A	0.225	1.435	0.800	2.572	0.286	0.702	0.366	1.345	0.228	1.511	0.773	2.954	0.241	0.705	0.392	1.265
Homozy HLA B	0.314	1.350	0.753	2.423	0.815	1.082	0.557	2.101	0.712	0.852	0.364	1.992	0.336	0.726	0.378	1.393
Homozy HLA C	0.229	1.347	0.829	2.188	0.881	1.043	0.599	1.818	0.574	1.207	0.627	2.326	0.497	1.183	0.728	1.923
Homozy HLA DRB	0.631	1.152	0.646	2.055	0.249	1.408	0.787	2.518	0.747	1.153	0.485	2.745	0.201	0.694	0.396	1.215
Homozy HLA DQA	0.440	0.826	0.509	1.341	0.521	0.842	0.498	1.424	0.179	1.551	0.818	2.938	0.680	0.904	0.558	1.463
Homozy HLA DQB	0.789	0.936	0.578	1.516	0.871	0.956	0.556	1.644	0.187	1.577	0.801	3.104	0.605	0.878	0.535	1.439
Homozy HLA DPB	0.058	1.460	0.988	2.158	0.292	1.256	0.822	1.920	0.157	0.683	0.403	1.157	0.558	1.115	0.774	1.608
Homozy Any	0.718	1.066	0.753	1.511	0.775	0.948	0.658	1.367	0.368	1.240	0.776	1.982	0.762	1.053	0.755	1.467
Total # Homozy	0.267	1.064	0.954	1.187	0.816	1.016	0.889	1.161	0.521	1.055	0.896	1.241	0.566	0.965	0.856	1.089
CLASS I HLA A	0.306	1.369	0.750	2.498	0.266	0.689	0.357	1.329	0.184	1.589	0.803	3.142	0.392	0.765	0.415	1.412
CLASS I HLA B	0.975	0.985	0.393	2.471	0.763	1.166	0.430	3.161	0.211	0.482	0.153	1.515	0.070	0.433	0.175	1.070
CLASS I HLA C	0.492	1.302	0.614	2.761	0.961	0.979	0.427	2.246	0.228	1.712	0.714	4.106	0.046	1.949	1.011	3.760
CLASS II HLA DRB	0.191	1.934	0.720	5.196	0.004*	5.809	1.739	19.407	0.560	0.704	0.216	2.294	0.121	0.465	0.177	1.223
CLASS II HLA DQA	0.137	0.446	0.154	1.293	0.019	0.174	0.040	0.754	0.538	1.486	0.422	5.236	0.410	1.582	0.532	4.705
CLASS II HLA DQB	0.692	1.212	0.469	3.128	0.713	1.270	0.355	4.539	0.683	1.327	0.341	5.155	0.944	0.963	0.339	2.736
CLASS II HLA DPB	0.080	1.422	0.958	2.110	0.361	1.220	0.796	1.871	0.167	0.688	0.404	1.170	0.545	1.120	0.776	1.617

Table 12: Haplotype Frequencies across Strata

HLA DRB*03:01~HLA DQA*05:01~HLA DQB*02:01	N (%)
Total Sample (N=1636)	181 (11.06%)
Sz/Sza (N=502)	51 (10.16%)
Bipolar (N=420)	52 (12.38%)
EpiGen (N=243)	31 (12.76%)
Family (N=471)	47 (9.98%)
HLA B*08:01~HLA C*07:01	**N (%)**
Total Sample (N=1636)	131 (8.01%)
Sz/Sza (N=502)	30 (5.98%)
Bipolar (N=420)	41 (9.76%)
EpiGen (N=243)	23 (9.47%)
Family (N=471)	37 (7.86%)
HLA B*38:01~HLA C*12:03	**N (%)**
Total Sample (N=1636)	520 (31.78%)
Sz/Sza (N=502)	155 (30.88%)
Bipolar (N=420)	124 (29.52%)
EpiGen (N=243)	74 (30.45%)
Family (N=471)	167 (35.46%)

Discussion

These analyses identified several previously unidentified HLA alleles associated with each infection as well as observing differential effects by mental disease status in several cases. Additionally, homozygosity at HLA loci in MHC class II was indicated to be potential factors in CMV and HSV1 seropositivity.

We consider our results in the context of haplotypes discussed in the literature. As displayed in Table 12 there were several common haplotypes that were present in the sample. HLA DRB*03:01, HLA DQA*05:01, and HLA DQB*02:01 co-occurred perfectly in the schizophrenic and EpiGen groups and very closely in the bipolar disorder and family groups. This produced nearly identical results in the common allele analysis and may indicate that this is a common haplotype in our population that could also be in linkage disequilibrium given the proximity of these alleles (Debnath et al., 2013). The Alleles HLA B*08:01~HLA C*07:01 and HLA B*38:01~HLA C*12:03 also co-occurred to a high level and had very similar effects to a slightly lesser extent. The Haplotypes HLA DR*03~HLA DQ*02 and

HLA B*8~HLA C*7 have both been previously been identified as commonly occurring in a south Tunisian population but not in an Ashkenazi population (Mahfoudh et al., 2013).

The alleles HLA B*8 and HLA DRB*03 have been previously associated with rapid progression from HIV to AIDS (Rodrigues et al., 2016). DRB*03 has previously been shown to be protective against Crohns Disease in a French population (Mahdi, 2015). Though the allele HLA B*38 has been previously identified as commonly occurring in AJ populations the rest of these haplotypes have not been identified as commonly occurring.(Bonné-Tamir et al., 1978; Klitz et al., 2010).

In the analyses of *Toxoplasma gondii* there were several findings of import. In the common allele analysis the alleles HLA B*08:01, HLA C*07:01, HLA DRB*03:01, HLA DQA*05:01 and HLA DQB 02:01 were all identified as having a risk increasing effect across all 4 groups with varying levels of significance in each group. This is consistent with potentially representing two haplotypes as described above. However, HLA C*04:01, HLA DQA*01:01, and HLA DPB*04:02 all exhibited inverse effects between at least one group of cases and controls. HLA C*04:01 displayed an infection risk decreasing effect among all groups save EpiGen controls, indicating that this allele may increase risk for *T. gondii* seropositivity in disease free individuals unrelated to mentally diseased individuals but not other groups. HLA DQA*01:01 and HLA DPB*04:02 displayed a risk increasing effect in both mental illness groups and a risk decreasing effect in both control groups which illustrates a possible interaction of mental disease and HLA genetics on *T. gondii* risk. All of these alleles bear further investigation particular those which displayed differential effects across mental disease status as they could reveal clues concerning etiology. A prior study found HLA DRB*03 may confer resistance against toxoplasmic encephalitis (though not *T. gondii* seroprevalence) among AIDS patients, but no other alleles here have been previously

identified as having a significant effect on *Toxoplasma gondii* infection (Mack et al., 1999; Rodrigues et al., 2016; Shimokawa et al., 2016).

The homozygosity analysis of *Toxoplasma gondii* had almost entirely null results across all indicating that this is likely not an important factor generally. The only significant results indicate that being homozygous at HLA DPB is a potentially risk increasing factor for family of cases, and among EpiGen controls being homozygous at HLA C is indicated as increasing risk of *T. gondii* seropositivity and being homozygous at HLA B is indicated to decrease risk. These relationships bear further investigation as homozygosity decreasing risk of infection is directly contrary to the heterozygote advantage hypothesis, however this is not necessarily biologically implausible since this analysis contained only 3 pathogens and individuals could be homozygotic for alleles which protect against those pathogens.

In the analyses of Cytomegalovirus very few individual alleles had any significant results or trends to observe. HLA B*38:01 and HLA DRB*13:02 both displayed a consistent protective effect against CMV seropositivity across all four groups with several significant at 0.05 results. HLA B*57:01 showed a risk increasing effect in all groups save EpiGen controls, indicating it may be a risk factor for infection among the mentally ill and their relatives. Lastly HLA DQA*01:02 showed a protective effect in both control and the bipolar disorder groups (in EpiGen this was significant after Bonferroni) though not in the schizophrenic group. Even so this allele could be implicated as having a strong protective effect against CMV particularly in disease free individuals. Previous research into CMV and HLA has concerned individual SNPs in GWAS or larger antigen groups and so none of these individual alleles have been previously identified as having significant effects (Borglum et al., 2014; O'Grady et al., 1988).

The homozygosity analysis of CMV revealed that homozygosity at HLA DPB was significantly associated with decreased risk of CMV in the EpiGen controls, this relationship is again contrary to the pre-existing heterozygote advantage hypotheses and has a plausible alternative hypothesis and so merits further research. Additionally homozygosity at HLA DRB appeared to increase risk of CMV among the bipolar disorder group.

The analyses of HSV1 produced several interesting results. HLA DRB*03:01, HLA DQA*05:01, and HLA DQB*02:01 again all appeared to be consistently risk increasing; more strongly in the disease free groups (OR=1.58-2.63) than in the mentally ill groups (OR=1.07-1.55). As stated previously these alleles were co-occurring so it can be interpreted that this particular haplotype is risk increasing for HSV1 particularly among mental disease free individuals. HLA C*07:01 was also consistently risk increasing and was significant after Bonferroni correction in the EpiGen controls. Interestingly HLA B*38:01 and HLA C*12:03 both displayed strong and significant protective effects in the mental disease free groups (OR=0.44-0.64) and null effects in the schizophrenia and bipolar disorder groups (OR=0.99-1.38), showing a possible interaction with mental disease status. Previous research had identified several MHC Class I alleles as associated either positively or negatively with HSV prevalence (HLA A*24, HLA B*27, HLA B*35, HLA B*44, HLA B*53, HLA*58 and HLA C*15) (Moraru et al., 2012; Samandary et al., 2014). We observed a significant effect in none of these alleles, though it is important to note that many of these alleles were found to be associated with HSV1 and HSV2 prevalence combined and in non-European populations so this is not unsurprising. Given the limited amount of research on HSV1 and HLA none of the alleles we observed as having significant effects had been previously identified as having significant effects either (Moraru et al., 2012; Samandary et al., 2014).

The homozygosity analysis of HSV1 did not reveal many clear new insights. Being homozygous at any HLA loci was significantly associated with increased odds of HSV1 among the EpiGen controls which fit previous hypotheses. However, homozygosity at HLA DQA was associated with decreased risk among the bipolar disorder group.

In the analysis of the multi-infection index the haplotype HLA DRB*03:01~HLA DQA*05:01~HLA DQB*02:01 was again consistently risk increasing with strongest effects in the control groups (particularly EpiGen). This is not surprising given the aforementioned relationship these alleles had with HSV1. Additionally HLA B*08:01 and HLA C*07:01 displayed strong risk increasing effects in the control group that were not present in the mentally disease groups supporting a differential effect by mental disease status as well as a possible haplotype effect. On the other hand HLA B*38:01 and HLA C*12:03 both displayed strong and significant protective effects in the control groups as well as in the bipolar disorder group (non-significantly) but not the schizophrenia group support the same conclusions drawn above. These alleles displayed strong evidence for a protective effect in disease free individuals but not mentally ill individuals.

The final homozygosity analysis for the multi-infection index was also not very informative; the most important result was a very significant association of homozygosity at HLA DRB with increased infection risk in the bipolar disorder group along with homozygosity at HLA DQA being associated with decreased infection risk in the same group.

This analysis had several limitations the most notable of which were the use of several stratifications reducing sample size in each model as well as the large number of models run leaving vulnerabilities to multiple comparison issues. Additionally this study did not run any analyses that group HLA alleles by antigen (such as including both HLA

DPB*04:01 and HLA DPB*04:02 as a single HLA DPB*04 term). While this may have increased strata sizes this categorization can also mask significant effects as alleles within antigen types can have conflicting risk increasing and protective effects (Gragert et al., 2014). Despite these limitations these analyses still produced several significant results as well as general trends worth noting.

Conclusions and Public Health Implications

This study was the first systematic epidemiological analysis to consider the effects of HLA alleles on *T. gondii*, CMV, and HSV1. This study also accounted for mental disease and was in several cases was able to identify alleles which exhibited a differential effect in mentally ill individuals versus non-mentally ill individuals. These results have identified several individual HLA alleles of interest for each infection as well as several haplotypes which could have broader implications. Additionally the fact we detected the most significant allele effects in the mental disease free groups despite sample size in those strata being equal or lesser is intriguing and may indicate that the effect of HLA alleles on infection risk may be mitigated or modified by having schizophrenia or bipolar disorder. The role of structurally diverse alleles of the complement component 4 (C4) genes needs to be further considered in this differential finding between individuals with schizophrenia and the control groups. This locus is within the MHC class III region, which is positioned between MHC class 1 and Class II regions in the human genome, and has been strongly linked to schizophrenia risk. Hence the apparent disengagement in mentally ill groups of the connection between HLA-types and some infections could be a reflection of differences between these mentally ill patients and controls in their C4 genotype. In this way, an individual's C4 genotype could represent an effect modifier of the relationship between HLA-type specific antigen presentation and resistance to infection.

The presence of several statistically significant results that indicate homozygosity as decreasing risk of infection do not fit with the hypothesis of heterozygote advantage; however these results are not implausible since an individual homozygotic for alleles which protect against the few pathogens we analyzed could achieve this result. In terms of trends homozygosity at HLA DQA appeared to decrease risk in CMV and HSV1 particularly in the mentally ill group while homozygosity at HLA DRB appeared to increase infection risk for CMV and HSV1 also the mentally ill group; all other observed results were not consistent between infections and the index. Future analyses of which specific alleles at each HLA loci promote an increase or decrease in disease risk when homozygotic will be prudent to further this course of study.

In addition to individual alleles pertinent to a single infection; the possible haplotype HLA B*38:01~HLA C*12:03 displayed consistent protective effects against infection stronger in mental disease free groups (see above), and the possible haplotype HLA B*08:01~HLA C*07:01 which displayed consistent risk increasing effects across infections and strata though again generally stronger in mental disease free groups. The haplotype HLA DRB*03:01~HLA DQA*05:01~HLA DQB*02:01 appeared consistently risk increasing across all strata and diseases. Though individual alleles uncovered in our analysis are potentially important these three haplotypes potentially have cross cutting implications for generally increasing or decreasing disease risk across multiple conditions. Further analysis of the haplotypes as well as the individually identified alleles is a pertinent future course of study.

This study's discovery of new alleles and haplotypes that could influence infection risk with differential effects in mentally ill versus non-mentally ill groups has not only scientific but public health implications. Given that the studied infections are all highly

prevalent with potentially severe complications identifying alleles that may genetically predisposed individuals to contracting them could have implications for screening and prevention. Additionally since genetics and infection are both implicated in the etiology of the mental illnesses schizophrenia and bipolar disorder, prevention of infections and a better understanding of a role for HLA types could lead to new paths for preventing or delaying onset of these mental illnesses. This research has also added to the growing body of knowledge of the interplay between genetics, the immune system, and mental illness which could lead to a better understanding of the complex etiology of these disease and improved screening, prevention, and pharmacological strategies.

Appendix:

Table A1: Average Imputation posterior probabilities for each HLA Loci

n=1636

HLA TYPE	mean (Std)
HLA A	0.8764 (0.1419)
HLA B	0.8035 (0.1955)
HLA C	0.9283 (0.1445)
HLA DRB	0.7129 (0.2461)
HLA DQA	0.8790 (0.1552)
HLA DQB	0.9039 (0.1409)
HLA DPB	0.7069 (0.1727)
Total	0.8301 (0.0911)

Table A2: Stratified Sample Characteristics

	Total (N=1636)	**Sz/Sza* (N=502)**	**BP** (N=420)**	**EpiGen (N=243)**	**Family (N=471)**
	N (%) or Mean (Std)	**N (%) or Mean (Std)**	**N (%) or Mean (Std)**	**N (%) or Mean (Std)**	**N (%) or Mean (Std)**
Demographics					
Age at Plasma	53.1 (15.2)	47.2 (12.1)	43.2 (15.0)	57.2 (11.1)	66.0 (9.3)
Sex (Female)	777 (47.49%)	178 (35.46%)	218 (51.90%)	148 (60.91%)	233 (50.53%)
Infections					
Toxoplasma Positive	295 (18.03%)	60 (11.95%)	42 (10.00%)	50 (20.58%)	143 (30.36%)
Cytomegalovirus Positive	589 (36.00%)	114 (22.71%)	134 (31.90%)	85 (34.95%)	256 (54.35%)
Herpes Simplex Virus 1 Positive	638 (39.00%)	140 (27.89%	152 (36.19%)	94 (38.68%)	252 (53.50%)
Multi-Infection Index					
3 Infections	89 (5.44%)	11 (2.19%)	10 (2.38%)	14 (5.76%)	54 (11.46%)
2 Infections	331 (20.23%)	58 (11.55%)	73 (17.38%)	46 (18.93%)	154 (32.70%)
1 Infection	593 (36.25%)	165 (32.87%)	152 (36.19%)	95 (39.09%)	181 (38.43%)

*Schizophrenia and schizoaffective Disorder were classified together for this analysis.

**There were no subjects with both Sz/Sza and bipolar disorder

Table A3: Stratified Sample Characteristics cont.

	Total (N=1636)	Sz/Sza* (N=502)	Bipolar** (N=420)	EpiGen (N=243)	Family (N=471)
	N (%) or Mean (Std)	N (%) or Mean (Std)	N (%) or Mean (Std)	N (%) or Mean (Std)	N (%) or Mean (Std)
Homozygosity					
Homozygous HLA A	160 (9.78%)	46 (9.16%)	38 (9.05%)	34 (13.99%)	42 (8.92%)
Homozygous HLA B	136 (8.31%)	46 (9.16%)	36 (8.57%)	21 (8.64%)	33 (7.01%)
Homozygous HLA C	225 (13.75%)	72 (14.34%)	53 (12.62%)	36 (14.81%)	64 (13.59%)
Homozygous HLA DRB	160 (9.78%)	49 (9.75%)	45 (10.71%)	20 (8.23%)	46 (9.77%)
Homozygous HLA DQA	244 (14.91%)	80 (15.94%)	60 (14.29%)	39 (16.05%)	65 (13.80%)
Homozygous HLA DQB	229 (14.00%)	78 (15.54%)	56 (13.33%)	34 (13.99%)	61 (12.95%)
Homozygous HLA DPB	427 (26.71%)	130 (25.90%)	102 (24.29%)	68 (27.98%)	137 (29.09%)
Homozygous Any	799 (48.84%)	240 (47.81%)	198 (47.14%)	126 (51.85%)	235 (49.80%)
Homozygous Total					
7 HLAs	21 (1.28%)	9 (1.79%)	3 (0.71%)	4 (1.65%)	5 (1.06%)
6 HLAs	20 (1.22%)	11 (2.19%)	5 (1.19%)	0 (0%)	4 (0.85%)
5 HLAs	26 (1.59%)	9 (1.79%)	6 (1.43%)	5 (2.06%)	6 (1.27%)
4 HLAs	37 (2.26%)	4 (0.80%)	11 (2.62%)	9 (3.70%)	13 (2.76%)
3 HLAs	109 (6.66%)	34 (6.77%)	27 (6.43%)	16 (6.58%)	32 (6.79%)
2 HLAs	133 (8.13%)	36 (7.17%)	38 (9.05%)	23 (9.47%)	36 (7.64%)
1 HLAs	452 (27.69%)	137 (27.29%)	108 (25.71%)	69 (28.40%)	139 (29.51%)

*Schizophrenia and schizoaffective Disorder were classified together for this analysis.

**There were no subjects with both Sz/Sza and bipolar disorder

Table A4: Results of Stratified Logistic models for statistically significant alleles predicting dichotomous T. gondii outcome controlling for age and sex, MHC Class I

Sz/Sza+BP					Controls				
HLA A	**p-value**	**OR**	**CIL**	**CIU**	**HLA A**	**p-value**	**OR**	**CIL**	**CIU**
01:01	0.7102	1.096	0.675	1.782	01:01	0.3377	1.21	0.82	1.785
11:01	0.784	1.102	0.551	2.202	11:01	0.4238	0.793	0.449	1.401
02:01	0.0511	0.607	0.367	1.002	02:01	0.7052	0.929	0.636	1.358
02:05	0.7725	1.117	0.527	2.367	02:05	0.3004	1.382	0.749	2.547
24:02	0.5367	0.838	0.478	1.469	24:02	0.2928	0.772	0.478	1.25
26:01	0.602	1.135	0.705	1.826	26:01	0.4018	0.852	0.585	1.239
3:01	0.3868	0.765	0.416	1.404	3:01	0.3712	1.219	0.79	1.881
33:01	0.2491	1.512	0.749	3.051	33:01	0.2012	0.623	0.302	1.287
HLA B	**p-value**	**OR**	**CIL**	**CIU**	**HLA B**	**p-value**	**OR**	**CIL**	**CIU**
14:02	0.1395	1.429	0.89	2.294	14:02	0.1808	0.738	0.473	1.151
35:01	0.4287	0.743	0.355	1.552	35:01	0.1006	1.562	0.917	2.661
35:02	0.2007	0.657	0.345	1.25	35:02	0.6095	0.89	0.569	1.393
35:03	0.1077	0.426	0.15	1.205	35:03	0.0709	1.756	0.953	3.236
38:01	0.6758	1.102	0.698	1.741	38:01	0.169	0.776	0.541	1.114
52:01	0.6191	1.184	0.608	2.306	52:01	0.61	0.864	0.493	1.515
57:01	0.1813	1.547	0.816	2.931	57:01	0.0518	1.766	0.996	3.132
08:01	0.0754	1.855	0.939	3.667	08:01	0.0226	1.927	1.096	3.388
HLA C	**p-value**	**OR**	**CIL**	**CIU**	**HLA C**	**p-value**	**OR**	**CIL**	**CIU**
12:02	0.6068	1.191	0.612	2.319	12:02	0.61	0.864	0.493	1.515
12:03	0.5686	1.136	0.733	1.761	12:03	0.1497	0.773	0.544	1.097
04:01	0.0103	0.535	0.332	0.863	04:01	0.3443	1.185	0.834	1.684
06:02	0.9082	1.03	0.619	1.715	06:02	0.1459	1.35	0.901	2.024
07:01	0.0046	2.101	1.257	3.51	07:01	0.0958	1.441	0.937	2.215
08:02	0.1179	1.46	0.909	2.345	08:02	0.2832	0.788	0.509	1.218

Table A5: Results of Stratified Logistic models for statistically significant alleles predicting dichotomous T. gondii outcome controlling for age and sex, MHC Class II

Sz/Sza+BP					Controls				
HLA DRB	**p-value**	**OR**	**CIL**	**CIU**	**HLA DRB**	**p-value**	**OR**	**CIL**	**CIU**
01:02	0.0807	1.535	0.949	2.483	01:02	0.0998	0.665	0.409	1.081
11:01	0.7143	1.093	0.68	1.756	11:01	0.6823	0.921	0.62	1.368
11:04	0.1415	0.52	0.218	1.243	11:04	0.9771	0.992	0.57	1.726
13:01	0.893	1.043	0.563	1.931	13:01	0.7072	1.095	0.682	1.759
13:02	0.9426	0.971	0.439	2.147	13:02	0.6915	0.882	0.473	1.642
15:02	0.3639	1.404	0.675	2.92	15:02	0.252	0.675	0.344	1.323
03:01	0.1944	1.499	0.813	2.761	03:01	0.1395	1.473	0.881	2.463
04:02	0.5393	1.155	0.728	1.833	04:02	0.7383	1.064	0.738	1.536
07:01	0.1235	0.662	0.392	1.119	07:01	0.3514	1.195	0.822	1.738
HLA DQA	**p-value**	**OR**	**CIL**	**CIU**	**HLA DQA**	**p-value**	**OR**	**CIL**	**CIU**
01:01	0.1001	1.477	0.928	2.351	01:01	0.0343	0.603	0.378	0.963
01:02	0.7834	0.919	0.505	1.673	01:02	0.1558	0.704	0.434	1.143
01:03	0.3675	1.261	0.762	2.087	01:03	0.591	0.893	0.591	1.349
02:01	0.0985	0.642	0.38	1.086	02:01	0.2355	1.254	0.863	1.823
03:01	0.5841	1.134	0.723	1.778	03:01	0.6179	1.095	0.767	1.563
05:01	0.2045	1.485	0.806	2.734	05:01	0.0674	1.611	0.966	2.685
05:05	0.4363	0.842	0.546	1.298	05:05	0.725	0.94	0.666	1.327
HLA DQB	**p-value**	**OR**	**CIL**	**CIU**	**HLA DQB**	**p-value**	**OR**	**CIL**	**CIU**
02:01	0.2045	1.485	0.806	2.734	02:01	0.0556	1.65	0.988	2.754
02:02	0.1122	0.637	0.365	1.111	02:02	0.2787	1.24	0.84	1.829
03:01	0.2139	0.76	0.493	1.171	03:01	0.758	0.948	0.673	1.334
03:02	0.8053	1.059	0.671	1.671	03:02	0.4984	1.13	0.793	1.61
05:01	0.096	1.469	0.934	2.31	05:01	0.1199	0.712	0.464	1.092
06:03	0.8347	1.068	0.577	1.977	06:03	0.6487	1.115	0.698	1.781
HLA DPB	**p-value**	**OR**	**CIL**	**CIU**	**HLA DPB**	**p-value**	**OR**	**CIL**	**CIU**
104:01	0.8539	0.938	0.473	1.858	104:01	0.509	0.811	0.436	1.51
02:01	0.1105	0.703	0.455	1.084	02:01	0.4187	1.15	0.819	1.615
04:01	0.9486	1.015	0.646	1.595	04:01	0.9519	0.989	0.689	1.419
04:02	0.0379	1.684	1.029	2.755	04:02	0.0502	0.609	0.37	1

Table A6: Results of Stratified Logistic models for statistically significant alleles predicting dichotomous CMV outcome controlling for age and sex, MHC Class I

Sz/Sza+BP					Controls				
HLA A	**p-value**	**OR**	**CIL**	**CIU**	**HLA A**	**p-value**	**OR**	**CIL**	**CIU**
01:01	0.7581	1.056	0.746	1.496	01:01	0.3277	1.199	0.833	1.726
11:01	0.7787	0.929	0.558	1.549	11:01	0.3475	0.785	0.473	1.301
02:01	0.9924	1.002	0.721	1.391	02:01	0.6906	1.072	0.761	1.509
02:05	0.1306	1.502	0.886	2.544	02:05	0.6864	0.885	0.488	1.604
24:02	0.3543	0.831	0.561	1.23	24:02	0.7468	1.071	0.705	1.627
26:01	0.1394	0.768	0.541	1.09	26:01	0.2593	0.822	0.585	1.155
3:01	0.7932	1.055	0.706	1.576	3:01	0.0487	0.662	0.44	0.998
33:01	0.4596	0.802	0.448	1.438	33:01	0.1939	0.665	0.36	1.23
HLA B	**p-value**	**OR**	**CIL**	**CIU**	**HLA B**	**p-value**	**OR**	**CIL**	**CIU**
14:02	0.8858	0.974	0.683	1.39	14:02	0.7388	1.069	0.723	1.58
35:01	0.3288	1.258	0.794	1.993	35:01	0.1746	0.692	0.407	1.177
35:02	0.6878	0.919	0.608	1.389	35:02	0.2749	1.253	0.836	1.879
35:03	0.574	1.165	0.684	1.984	35:03	0.0158	2.163	1.156	4.049
38:01	0.3187	0.844	0.605	1.177	38:01	0.0062	0.634	0.457	0.879
52:01	0.5372	0.85	0.508	1.423	52:01	0.1153	0.662	0.396	1.106
57:01	0.0337	1.694	1.042	2.755	57:01	0.5561	1.186	0.673	2.09
08:01	0.2143	0.676	0.364	1.254	08:01	0.4431	1.242	0.714	2.163
HLA C	**p-value**	**OR**	**CIL**	**CIU**	**HLA C**	**p-value**	**OR**	**CIL**	**CIU**
12:02	0.7638	0.925	0.556	1.538	12:02	0.1153	0.662	0.396	1.106
12:03	0.2984	0.845	0.614	1.161	12:03	0.0758	0.751	0.548	1.03
04:01	0.5494	1.099	0.807	1.497	04:01	0.5953	1.092	0.789	1.511
06:02	0.1472	1.304	0.911	1.867	06:02	0.4152	1.171	0.801	1.711
07:01	0.0922	0.674	0.426	1.067	07:01	0.2494	1.271	0.845	1.913
08:02	0.9687	0.993	0.696	1.417	08:02	0.9647	1.009	0.685	1.486

Table A7: Results of Stratified Logistic models for statistically significant alleles predicting dichotomous CMV outcome controlling for age and sex, MHC Class II

Sz/Sza+BP					Controls				
HLA DRB	**p-value**	**OR**	**CIL**	**CIU**	**HLA DRB**	**p-value**	**OR**	**CIL**	**CIU**
01:02	0.575	0.898	0.618	1.306	01:02	0.7615	1.067	0.703	1.619
11:01	0.5862	0.908	0.64	1.287	11:01	0.2971	1.211	0.845	1.736
11:04	0.9867	1.004	0.606	1.664	11:04	0.8708	1.043	0.628	1.731
13:01	0.7377	0.927	0.595	1.444	13:01	0.5815	0.885	0.572	1.368
13:02	0.9665	0.988	0.553	1.763	13:02	0.0013	0.369	0.201	0.678
15:02	0.8308	0.938	0.524	1.681	15:02	0.3727	0.769	0.432	1.369
03:01	0.9861	1.004	0.621	1.624	03:01	0.2269	1.353	0.829	2.209
04:02	0.5907	1.095	0.786	1.525	04:02	0.4813	0.885	0.63	1.243
07:01	0.8227	1.04	0.738	1.465	07:01	0.0865	1.354	0.957	1.916
HLA DQA	**p-value**	**OR**	**CIL**	**CIU**	**HLA DQA**	**p-value**	**OR**	**CIL**	**CIU**
01:01	0.5313	0.893	0.627	1.272	01:01	0.6199	1.105	0.745	1.639
01:02	0.8509	1.041	0.685	1.583	01:02	0.01	0.572	0.375	0.875
01:03	0.5796	0.899	0.616	1.311	01:03	0.3706	0.844	0.582	1.223
02:01	0.9187	1.018	0.723	1.434	02:01	0.1024	1.336	0.944	1.89
03:01	0.8202	1.038	0.751	1.435	03:01	0.3213	0.846	0.608	1.177
05:01	0.978	0.993	0.615	1.605	05:01	0.2867	1.305	0.8	2.13
05:05	0.8652	0.974	0.718	1.322	05:05	0.198	1.229	0.898	1.684
HLA DQB	**p-value**	**OR**	**CIL**	**CIU**	**HLA DQB**	**p-value**	**OR**	**CIL**	**CIU**
02:01	0.978	0.993	0.615	1.605	02:01	0.3478	1.266	0.774	2.072
02:02	0.8082	0.956	0.667	1.371	02:02	0.2398	1.242	0.865	1.782
03:01	0.5245	1.103	0.815	1.492	03:01	0.1243	1.277	0.935	1.745
03:02	0.8222	0.963	0.694	1.336	03:02	0.4229	0.874	0.63	1.214
05:01	0.8928	0.977	0.696	1.372	05:01	0.3277	1.206	0.829	1.753
06:03	0.7669	0.935	0.6	1.457	06:03	0.5566	0.878	0.57	1.353
HLA DPB	**p-value**	**OR**	**CIL**	**CIU**	**HLA DPB**	**p-value**	**OR**	**CIL**	**CIU**
104:01	0.3397	1.253	0.788	1.993	104:01	0.794	1.076	0.619	1.871
02:01	0.315	0.856	0.632	1.159	02:01	0.9105	0.982	0.72	1.34
04:01	0.3866	0.869	0.633	1.194	04:01	0.6281	0.922	0.663	1.282
04:02	0.4502	1.158	0.791	1.697	04:02	0.0777	1.455	0.959	2.206

Table A8: Results of Stratified Logistic models for statistically significant alleles predicting dichotomous HSV1 outcome controlling for age and sex, MHC Class I

Sz/Sza+BP					Controls				
HLA A	**p-value**	**OR**	**CIL**	**CIU**	**HLA A**	**p-value**	**OR**	**CIL**	**CIU**
01:01	0.8465	1.033	0.744	1.435	01:01	0.0015	1.786	1.249	2.553
11:01	0.6186	1.127	0.704	1.804	11:01	0.2731	1.313	0.807	2.139
02:01	0.2466	0.831	0.608	1.136	02:01	0.8962	1.022	0.734	1.423
02:05	0.5667	0.855	0.499	1.463	02:05	0.9209	0.971	0.546	1.729
24:02	0.9351	1.015	0.708	1.456	24:02	0.5766	0.891	0.594	1.337
26:01	0.3809	1.152	0.84	1.581	26:01	0.479	0.888	0.638	1.234
3:01	0.7518	1.063	0.729	1.549	3:01	0.0975	0.716	0.482	1.063
33:01	0.3764	1.262	0.753	2.116	33:01	0.1834	0.668	0.369	1.21
HLA B	**p-value**	**OR**	**CIL**	**CIU**	**HLA B**	**p-value**	**OR**	**CIL**	**CIU**
14:02	0.3859	1.157	0.832	1.61	14:02	0.0629	0.696	0.475	1.02
35:01	0.7597	0.932	0.592	1.466	35:01	0.6198	1.137	0.685	1.889
35:02	0.4386	0.856	0.578	1.268	35:02	0.7919	1.054	0.712	1.562
35:03	0.1299	0.651	0.374	1.134	35:03	0.9221	0.971	0.542	1.741
38:01	0.3346	1.162	0.856	1.577	38:01	0.0029	0.617	0.449	0.847
52:01	0.8842	1.036	0.646	1.661	52:01	0.6917	0.905	0.554	1.479
57:01	0.5618	1.153	0.713	1.864	57:01	0.1685	1.478	0.848	2.577
08:01	0.417	1.239	0.739	2.077	08:01	0.0041	2.261	1.295	3.95
HLA C	**p-value**	**OR**	**CIL**	**CIU**	**HLA C**	**p-value**	**OR**	**CIL**	**CIU**
12:02	0.8509	1.046	0.653	1.677	12:02	0.6917	0.905	0.554	1.479
12:03	0.7136	1.057	0.787	1.42	12:03	<.0001	0.513	0.376	0.699
04:01	0.1618	0.81	0.602	1.088	04:01	0.9831	1.003	0.733	1.374
06:02	0.4673	0.877	0.617	1.248	06:02	0.3363	1.198	0.829	1.732
07:01	0.4391	1.17	0.787	1.739	07:01	0.0003	2.14	1.423	3.217
08:02	0.3333	1.177	0.846	1.638	08:02	0.0542	0.689	0.471	1.007

Table A9: Results of Stratified Logistic models for statistically significant alleles predicting dichotomous HSV1 outcome controlling for age and sex, MHC Class II

Sz/Sza+BP					Controls				
HLA DRB	**p-value**	**OR**	**CIL**	**CIU**	**HLA DRB**	**p-value**	**OR**	**CIL**	**CIU**
01:02	0.5659	1.107	0.783	1.563	01:02	0.2961	0.806	0.537	1.208
11:01	0.9155	1.018	0.735	1.41	11:01	0.8516	0.967	0.683	1.371
11:04	0.1895	0.714	0.432	1.181	11:04	0.9806	1.006	0.617	1.641
13:01	0.5909	0.892	0.587	1.355	13:01	0.8885	0.97	0.636	1.48
13:02	0.3692	1.272	0.752	2.15	13:02	0.2738	1.351	0.788	2.316
15:02	0.6401	0.876	0.503	1.526	15:02	0.4371	0.801	0.459	1.401
03:01	0.1717	1.355	0.876	2.096	03:01	0.0054	1.994	1.226	3.242
04:02	0.7747	1.047	0.765	1.432	04:02	0.1455	0.782	0.562	1.089
07:01	0.524	1.11	0.805	1.53	07:01	0.5772	1.1	0.786	1.539
HLA DQA	**p-value**	**OR**	**CIL**	**CIU**	**HLA DQA**	**p-value**	**OR**	**CIL**	**CIU**
01:01	0.6297	0.921	0.661	1.285	01:01	0.6423	0.914	0.624	1.338
01:02	0.3161	1.219	0.828	1.796	01:02	0.6961	1.083	0.727	1.612
01:03	0.8148	1.043	0.735	1.479	01:03	0.5881	0.906	0.633	1.296
02:01	0.4026	1.146	0.833	1.577	02:01	0.6407	1.083	0.774	1.517
03:01	0.6609	1.071	0.789	1.452	03:01	0.3273	0.852	0.619	1.174
05:01	0.1923	1.336	0.864	2.063	05:01	0.0093	1.9	1.171	3.083
05:05	0.29	0.856	0.641	1.142	05:05	0.6641	1.07	0.789	1.451
HLA DQB	**p-value**	**OR**	**CIL**	**CIU**	**HLA DQB**	**p-value**	**OR**	**CIL**	**CIU**
02:01	0.1923	1.336	0.864	2.063	02:01	0.0131	1.85	1.138	3.006
02:02	0.6751	1.074	0.769	1.501	02:02	0.851	1.034	0.728	1.468
03:01	0.5478	0.916	0.688	1.219	03:01	0.328	1.163	0.86	1.573
03:02	0.911	0.983	0.722	1.336	03:02	0.1035	0.767	0.557	1.056
05:01	0.6086	0.919	0.666	1.269	05:01	0.9139	0.98	0.682	1.408
06:03	0.8823	1.032	0.683	1.557	06:03	0.8132	0.951	0.626	1.444
HLA DPB	**p-value**	**OR**	**CIL**	**CIU**	**HLA DPB**	**p-value**	**OR**	**CIL**	**CIU**
104:01	0.9135	1.025	0.653	1.61	104:01	0.1665	1.463	0.853	2.508
02:01	0.6513	0.937	0.705	1.245	02:01	0.6662	1.068	0.791	1.443
04:01	0.0713	0.76	0.565	1.024	04:01	0.2148	0.817	0.594	1.124
04:02	0.5693	1.111	0.773	1.596	04:02	0.6537	0.912	0.611	1.362

Table A10: Results of Stratified Logistic models for statistically significant alleles predicting ordinal multi-infection outcome controlling for age and sex, MHC Class I

Sz/Sza+BP					Controls				
HLA A	**p-value**	**OR**	**CIL**	**CIU**	**HLA A**	**p-value**	**OR**	**CIL**	**CIU**
01:01	0.5885	1.084	0.81	1.449	01:01	0.0048	1.586	1.151	2.184
11:01	0.6765	1.094	0.718	1.666	11:01	0.9825	1.005	0.647	1.561
02:01	0.0855	0.786	0.597	1.034	02:01	0.8905	1.021	0.757	1.378
02:05	0.4862	1.177	0.744	1.864	02:05	0.7937	1.072	0.637	1.803
24:02	0.5	0.895	0.648	1.235	24:02	0.3923	0.852	0.59	1.23
26:01	0.8666	1.024	0.773	1.358	26:01	0.0571	0.748	0.554	1.009
3:01	0.9973	0.999	0.714	1.398	3:01	0.1238	0.756	0.53	1.079
33:01	0.3795	1.232	0.773	1.963	33:01	0.0475	0.584	0.343	0.994
HLA B	**p-value**	**OR**	**CIL**	**CIU**	**HLA B**	**p-value**	**OR**	**CIL**	**CIU**
14:02	0.3106	1.164	0.868	1.563	14:02	0.1587	0.781	0.554	1.101
35:01	0.9962	0.999	0.671	1.487	35:01	0.7911	1.064	0.673	1.682
35:02	0.2768	0.826	0.585	1.166	35:02	0.5988	1.1	0.771	1.571
35:03	0.4164	0.826	0.521	1.31	35:03	0.0416	1.733	1.021	2.94
38:01	0.5989	1.076	0.82	1.411	38:01	<.0001	0.53	0.397	0.707
52:01	0.848	0.96	0.629	1.464	52:01	0.1864	0.741	0.475	1.156
57:01	0.0248	1.625	1.064	2.482	57:01	0.0424	1.675	1.018	2.756
08:01	0.9383	0.982	0.613	1.571	08:01	0.0018	2.168	1.333	3.527
HLA C	**p-value**	**OR**	**CIL**	**CIU**	**HLA C**	**p-value**	**OR**	**CIL**	**CIU**
12:02	0.9171	1.023	0.671	1.558	12:02	0.1864	0.741	0.475	1.156
12:03	0.9567	1.007	0.775	1.308	12:03	<.0001	0.532	0.401	0.704
04:01	0.1746	0.835	0.644	1.083	04:01	0.4298	1.122	0.844	1.491
06:02	0.7318	1.055	0.776	1.435	06:02	0.0919	1.332	0.954	1.86
07:01	0.7745	1.053	0.738	1.504	07:01	0.0003	1.947	1.358	2.791
08:02	0.2451	1.191	0.887	1.599	08:02	0.1429	0.775	0.551	1.09

Table A11: Results of Stratified Logistic models for statistically significant alleles predicting ordinal multi-infection outcome controlling for age and sex, MHC Class II

Sz/Sza+BP					Controls				
HLA DRB	**p-value**	**OR**	**CIL**	**CIU**	**HLA DRB**	**p-value**	**OR**	**CIL**	**CIU**
01:02	0.3129	1.171	0.862	1.592	01:02	0.2467	0.805	0.559	1.161
11:01	0.7951	0.962	0.72	1.286	11:01	0.8037	1.041	0.759	1.427
11:04	0.1893	0.749	0.487	1.153	11:04	0.9899	0.997	0.64	1.553
13:01	0.6186	0.911	0.631	1.315	13:01	0.8913	0.974	0.664	1.428
13:02	0.8591	1.044	0.649	1.681	13:02	0.141	0.692	0.424	1.13
15:02	0.8513	0.955	0.588	1.55	15:02	0.1118	0.664	0.401	1.1
03:01	0.3359	1.213	0.818	1.8	03:01	0.0027	1.933	1.256	2.973
04:02	0.3548	1.14	0.864	1.503	04:02	0.2347	0.834	0.619	1.125
07:01	0.9749	1.005	0.755	1.337	07:01	0.0798	1.313	0.968	1.78
HLA DQA	**p-value**	**OR**	**CIL**	**CIU**	**HLA DQA**	**p-value**	**OR**	**CIL**	**CIU**
01:01	0.8872	1.021	0.763	1.367	01:01	0.3957	0.861	0.609	1.216
01:02	0.8358	1.038	0.732	1.471	01:02	0.0619	0.707	0.492	1.017
01:03	0.9731	1.005	0.737	1.371	01:03	0.2751	0.834	0.603	1.155
02:01	0.9374	1.011	0.761	1.345	02:01	0.0744	1.32	0.973	1.791
03:01	0.4604	1.107	0.845	1.45	03:01	0.3586	0.873	0.654	1.166
05:01	0.3831	1.191	0.804	1.764	05:01	0.0027	1.932	1.256	2.973
05:05	0.3033	0.875	0.678	1.129	05:05	0.4169	1.121	0.851	1.477
HLA DQB	**p-value**	**OR**	**CIL**	**CIU**	**HLA DQB**	**p-value**	**OR**	**CIL**	**CIU**
02:01	0.3831	1.191	0.804	1.764	02:01	0.0039	1.893	1.228	2.92
02:02	0.6964	0.942	0.699	1.27	02:02	0.2207	1.219	0.888	1.675
03:01	0.7703	0.963	0.748	1.24	03:01	0.1983	1.197	0.91	1.574
03:02	0.9197	1.014	0.773	1.331	03:02	0.261	0.848	0.636	1.131
05:01	0.7798	1.041	0.785	1.381	05:01	0.8661	0.972	0.701	1.349
06:03	0.9416	0.986	0.684	1.422	06:03	0.8671	0.968	0.663	1.414
HLA DPB	**p-value**	**OR**	**CIL**	**CIU**	**HLA DPB**	**p-value**	**OR**	**CIL**	**CIU**
104:01	0.5287	1.137	0.763	1.693	104:01	0.549	1.159	0.715	1.878
02:01	0.1514	0.832	0.646	1.07	02:01	0.4893	1.101	0.838	1.446
04:01	0.1058	0.803	0.616	1.047	04:01	0.2764	0.852	0.638	1.137
04:02	0.1033	1.305	0.947	1.798	04:02	0.9948	0.999	0.695	1.435

Table A12: Results of Stratified Logistic models for homozygosity variables predicting dichotomous T. gondii outcome controlling for age and sex

	Sz/Sza+BP				Controls			
	p-value	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
Homozygous HLA A	0.8095	1.091	0.536	2.224	0.7367	0.907	0.514	1.602
Homozygous HLA B	0.6167	1.21	0.573	2.555	0.6931	0.875	0.452	1.696
Homozygous HLA C	0.9077	1.07	0.699	1.636	0.2267	1.333	0.837	2.123
Homozygous HLA DRB	0.4125	1.318	0.681	2.549	0.8698	0.952	0.527	1.719
Homozygous HLA DQA	0.785	1.086	0.6	1.967	0.2586	1.308	0.821	2.082
Homozygous HLA DQB	0.7773	1.092	0.593	2.012	0.6696	1.113	0.681	1.819
Homozygous HLA DPB	0.4244	1.212	0.756	1.943	0.1288	1.326	0.921	1.909
Homozygous Any	0.6769	1.094	0.716	1.672	0.1611	1.273	0.908	1.786
Total # Homozygous	0.4739	1.052	0.915	1.21	0.314	1.062	0.944	1.195
CLASS I HLA A	0.876	1.059	0.514	2.182	0.8295	0.937	0.521	1.688
CLASS I HLA B	0.5568	1.431	0.433	4.734	0.152	0.532	0.224	1.262
CLASS I HLA C	0.7043	0.823	0.301	2.251	0.0546	1.807	0.988	3.303
CLASS II HLA DRB	0.351	1.86	0.505	6.856	0.1904	0.553	0.228	1.342
CLASS II HLA DQA	0.7023	0.753	0.175	3.233	0.0594	2.492	0.964	6.442
CLASS II HLA DQB	0.879	0.902	0.24	3.392	0.4627	0.691	0.257	1.854
CLASS II HLA DPB	0.4627	1.194	0.744	1.916	0.1117	1.346	0.933	1.941

Table A13: Results of Stratified Logistic models for homozygosity variables predicting dichotomous CMV outcome controlling for age and sex

	Sz/Sza+BP				Controls			
	p-value	**OR**	**CIL**	**CIU**	**p-value**	**OR**	**CIL**	**CIU**
Homozygous HLA A	0.8996	0.967	0.572	1.634	0.7616	1.08	0.655	1.782
Homozygous HLA B	0.6781	1.118	0.661	1.891	0.3288	0.744	0.411	1.347
Homozygous HLA C	0.3179	1.244	0.811	1.909	0.8628	1.04	0.666	1.625
Homozygous HLA DRB	0.4094	1.224	0.757	1.98	0.0864	0.62	0.359	1.071
Homozygous HLA DQA	0.7984	0.946	0.619	1.446	0.403	0.829	0.533	1.287
Homozygous HLA DQB	0.8436	1.044	0.681	1.601	0.8087	1.058	0.671	1.668
Homozygous HLA DPB	0.1861	1.258	0.895	1.768	0.0302	0.683	0.484	0.964
Homozygous Any	0.8744	1.025	0.758	1.385	0.365	0.867	0.637	1.18
Total # Homozygous	0.407	1.043	0.944	1.152	0.1986	0.929	0.831	1.039
CLASS I HLA A	0.8378	0.945	0.553	1.617	0.5643	1.164	0.695	1.951
CLASS I HLA B	0.6533	0.835	0.38	1.835	0.123	0.526	0.232	1.19
CLASS I HLA C	0.3017	1.398	0.74	2.639	0.2647	1.414	0.769	2.598
CLASS II HLA DRB	0.1423	1.958	0.798	4.804	0.0322	0.383	0.159	0.922
CLASS II HLA DQA	0.1635	0.479	0.17	1.348	0.1767	0.493	0.176	1.376
CLASS II HLA DQB	0.6187	1.258	0.509	3.109	0.0144	3.779	1.303	10.961
CLASS II HLA DPB	0.228	1.234	0.877	1.736	0.0298	0.68	0.48	0.963

Table A14: Results of Stratified Logistic models for homozygosity variables predicting dichotomous HSV1 outcome controlling for age and sex

	Sz/Sza+BP				Controls			
	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU
Homozygous HLA A	0.8753	0.961	0.587	1.574	0.707	0.911	0.561	1.48
Homozygous HLA B	0.9177	1.027	0.622	1.694	0.5073	0.825	0.468	1.456
Homozygous HLA C	0.9034	0.975	0.642	1.478	0.3937	1.206	0.784	1.855
Homozygous HLA DRB	0.8708	1.039	0.653	1.654	0.8682	1.045	0.624	1.748
Homozygous HLA DQA	0.0422	0.645	0.422	0.985	0.5939	1.122	0.735	1.713
Homozygous HLA DQB	0.1604	0.739	0.484	1.127	0.9498	0.986	0.635	1.53
Homozygous HLA DPB	0.1819	1.245	0.903	1.716	0.6519	1.079	0.776	1.5
Homozygous Any	0.4243	0.891	0.671	1.183	0.2336	1.199	0.89	1.616
Total # Homozygous	0.6576	0.978	0.888	1.078	0.7624	1.017	0.914	1.131
CLASS I HLA A	0.855	0.954	0.577	1.578	0.8483	0.952	0.577	1.572
CLASS I HLA B	0.7514	1.133	0.524	2.449	0.1151	0.532	0.243	1.166
CLASS I HLA C	0.7628	0.907	0.48	1.712	0.0859	1.677	0.93	3.026
CLASS II HLA DRB	0.0053	4.427	1.554	12.608	0.9202	0.96	0.431	2.14
CLASS II HLA DQA	0.0072	0.2	0.062	0.647	0.2535	1.701	0.684	4.231
CLASS II HLA DQB	0.9157	1.05	0.425	2.595	0.3181	0.624	0.247	1.576
CLASS II HLA DPB	0.242	1.213	0.878	1.678	0.632	1.084	0.779	1.51

Table A15: Results of Stratified Logistic models for homozygosity variables predicting ordinal multi-infection index outcome controlling for age and sex

	Sz/Sza+BP				Controls			
	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU
Homozygous HLA A	0.8274	1.049	0.681	1.617	0.7963	0.944	0.608	1.465
Homozygous HLA B	0.4005	1.207	0.779	1.869	0.2838	0.755	0.451	1.263
Homozygous HLA C	0.3704	1.181	0.821	1.699	0.4452	1.164	0.789	1.717
Homozygous HLA DRB	0.2157	1.293	0.861	1.944	0.409	0.821	0.514	1.312
Homozygous HLA DQA	0.2748	0.821	0.576	1.17	0.7203	1.073	0.731	1.574
Homozygous HLA DQB	0.659	0.923	0.645	1.32	0.7901	1.056	0.709	1.572
Homozygous HLA DPB	0.0473	1.336	1.004	1.779	0.7418	0.951	0.705	1.282
Homozygous Any	0.991	0.999	0.777	1.283	0.5081	1.095	0.836	1.435
Total # Homozygous	0.3818	1.038	0.955	1.129	0.8354	0.99	0.899	1.09
CLASS I HLA A	0.9504	1.014	0.652	1.577	0.9338	1.019	0.648	1.605
CLASS I HLA B	0.8083	1.087	0.555	2.128	0.0229	0.442	0.219	0.893
CLASS I HLA C	0.6927	1.118	0.643	1.943	0.0248	1.821	1.079	3.072
CLASS II HLA DRB	0.0015	3.394	1.597	7.212	0.1006	0.544	0.263	1.125
CLASS II HLA DQA	0.0069	0.309	0.132	0.724	0.4124	1.406	0.623	3.173
CLASS II HLA DQB	0.748	1.131	0.534	2.392	0.7617	1.136	0.499	2.588
CLASS II HLA DPB	0.0853	1.288	0.965	1.717	0.7823	0.959	0.71	1.293

Table A16: T. gondii, Number of Homozygotic Loci Categorical, Controlling for age and sex

	Sz/Sza				Bipolar				EpiGen				Family			
	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU
Homozy 7 v 0	0.044	4.199	1.038	16.989	0.184	5.267	0.453	61.259	0.984	0.001	0.001	>999.9	0.358	0.421	0.067	2.657
Homozy 6 v 0	0.630	1.410	0.348	5.720	0.633	1.571	0.246	10.042	N/A	N/A	N/A	N/A	0.265	0.267	0.026	2.721
Homozy 5 v 0	0.986	0.001	0.001	>999.9	0.068	5.154	0.887	29.956	0.981	0.001	0.001	>999.9	0.530	1.792	0.290	11.072
Homozy 4 v 0	0.878	1.203	0.114	12.719	0.286	0.423	0.087	2.055	0.513	1.608	0.388	6.674	0.138	0.404	0.122	1.337
Homozy 3 v 0	0.721	1.176	0.484	2.857	0.830	1.099	0.463	2.608	0.505	1.459	0.481	4.428	0.761	0.888	0.412	1.912
Homozy 2 v 0	0.275	0.563	0.201	1.578	0.138	0.512	0.212	1.240	0.831	1.109	0.428	2.878	0.149	1.758	0.817	3.783
Homozy 1 v 0	0.248	1.355	0.810	2.268	0.803	0.938	0.566	1.554	0.170	0.629	0.325	1.220	0.606	0.892	0.577	1.378

*There were no individuals in the EpiGen group who were homozygotic at 6 HLA loci

Table A17: CMV, Number of Homozygotic Loci Categorical, Controlling for age and sex

	Sz/Sza				Bipolar				EpiGen				Family			
	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU
Homozy 7 v 0	0.044	4.199	1.038	16.989	0.184	5.267	0.453	61.259	0.984	0.001	0.001	>999.9	0.358	0.421	0.067	2.657
Homozy 6 v 0	0.630	1.410	0.348	5.720	0.633	1.571	0.246	10.042	N/A	N/A	N/A	N/A	0.265	0.267	0.026	2.721
Homozy 5 v 0	0.986	0.001	0.001	>999.9	0.068	5.154	0.887	29.956	0.981	0.001	0.001	>999.9	0.530	1.792	0.290	11.072
Homozy 4 v 0	0.878	1.203	0.114	12.719	0.286	0.423	0.087	2.055	0.513	1.608	0.388	6.674	0.138	0.404	0.122	1.337
Homozy 3 v 0	0.721	1.176	0.484	2.857	0.830	1.099	0.463	2.608	0.505	1.459	0.481	4.428	0.761	0.888	0.412	1.912
Homozy 2 v 0	0.275	0.563	0.201	1.578	0.138	0.512	0.212	1.240	0.831	1.109	0.428	2.878	0.149	1.758	0.817	3.783
Homozy 1 v 0	0.248	1.355	0.810	2.268	0.803	0.938	0.566	1.554	0.170	0.629	0.325	1.220	0.606	0.892	0.577	1.378

*There were no individuals in the EpiGen group who were homozygotic at 6 HLA loci

Table A18: HSV1, Number of Homozygotic Loci Categorical, Controlling for age and sex

	Sz/Sza				Bipolar				EpiGen				Family			
	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU
Homozy 7 v 0	0.044	4.199	1.038	16.989	0.184	5.267	0.453	61.259	0.984	0.001	0.001	>999.9	0.358	0.421	0.067	2.657
Homozy 6 v 0	0.630	1.410	0.348	5.720	0.633	1.571	0.246	10.042	N/A	N/A	N/A	N/A	0.265	0.267	0.026	2.721
Homozy 5 v 0	0.986	0.001	0.001	>999.9	0.068	5.154	0.887	29.956	0.981	0.001	0.001	>999.9	0.530	1.792	0.290	11.072
Homozy 4 v 0	0.878	1.203	0.114	12.719	0.286	0.423	0.087	2.055	0.513	1.608	0.388	6.674	0.138	0.404	0.122	1.337
Homozy 3 v 0	0.721	1.176	0.484	2.857	0.830	1.099	0.463	2.608	0.505	1.459	0.481	4.428	0.761	0.888	0.412	1.912
Homozy 2 v 0	0.275	0.563	0.201	1.578	0.138	0.512	0.212	1.240	0.831	1.109	0.428	2.878	0.149	1.758	0.817	3.783
Homozy 1 v 0	0.248	1.355	0.810	2.268	0.803	0.938	0.566	1.554	0.170	0.629	0.325	1.220	0.606	0.892	0.577	1.378

*There were no individuals in the EpiGen group who were homozygotic at 6 HLA loci

Table A19: Multi-infection index, Number of Homozygotic Loci Categorical, Controlling for age and sex

	Sz/Sza				Bipolar				EpiGen				Family			
	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU	p-value	OR	CIL	CIU
Homozy 7 v 0	0.044	4.199	1.038	16.989	0.184	5.267	0.453	61.259	0.984	0.001	0.001	>999.9	0.358	0.421	0.067	2.657
Homozy 6 v 0	0.630	1.410	0.348	5.720	0.633	1.571	0.246	10.042	N/A	N/A	N/A	N/A	0.265	0.267	0.026	2.721
Homozy 5 v 0	0.986	0.001	0.001	>999.9	0.068	5.154	0.887	29.956	0.981	0.001	0.001	>999.9	0.530	1.792	0.290	11.072
Homozy 4 v 0	0.878	1.203	0.114	12.719	0.286	0.423	0.087	2.055	0.513	1.608	0.388	6.674	0.138	0.404	0.122	1.337
Homozy 3 v 0	0.721	1.176	0.484	2.857	0.830	1.099	0.463	2.608	0.505	1.459	0.481	4.428	0.761	0.888	0.412	1.912
Homozy 2 v 0	0.275	0.563	0.201	1.578	0.138	0.512	0.212	1.240	0.831	1.109	0.428	2.878	0.149	1.758	0.817	3.783
Homozy 1 v 0	0.248	1.355	0.810	2.268	0.803	0.938	0.566	1.554	0.170	0.629	0.325	1.220	0.606	0.892	0.577	1.378

*There were no individuals in the EpiGen group who were homozygotic at 6 HLA loci

References

Abraham, G., Rohmer, A., Tye-Din, J. A., & Inouye, M. (2015). Genomic prediction of celiac disease targeting HLA-positive individuals. *Genome Medicine, 7*(1), 1-11. doi:10.1186/s13073-015-0196-5

Avramopoulos, D., Pearce, B. D., McGrath, J., Wolyniec, P., Wang, R., Eckart, N., . . . Pulver, A. E. (2015). Infection and inflammation in schizophrenia and bipolar disorder: a genome wide study for interactions with genetic variation. *PLoS One, 10*(3), e0116696. doi:10.1371/journal.pone.0116696

Benros, M. E., Pedersen, M. G., Rasmussen, H., Eaton, W. W., Nordentoft, M., & Mortensen, P. B. (2014). A Nationwide Study on the Risk of Autoimmune Diseases in Individuals With a Personal or a Family History of Schizophrenia and Related Psychosis. *American Journal of Psychiatry, 171*(2), 218-226. doi:doi:10.1176/appi.ajp.2013.13010086

Bonné-Tamir, B., Bodmer, J. G., Bodmer, W. F., Pickbourne, P., Brautbar, C., Gazit, E., . . . Zamir, R. (1978). HLA Polymorphism in Israel: 9. An Overall Comparative Analysis. *Tissue Antigens, 11*(3), 235-250. doi:10.1111/j.1399-0039.1978.tb01255.x

Borglum, A. D., Demontis, D., Grove, J., Pallesen, J., Hollegaard, M. V., Pedersen, C. B., . . . Mors, O. (2014). Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Mol Psychiatry, 19*(3), 325-333. doi:10.1038/mp.2013.2

Bradley, H., Markowitz, L. E., Gibson, T., & McQuillan, G. M. (2013). Seroprevalence of Herpes Simplex Virus Types 1 and 2—United States, 1999–2010. *Journal of Infectious Diseases*. doi:10.1093/infdis/jit458

Brucato, N., Guadalupe, T., Franke, B., Fisher, S. E., & Francks, C. (2015). A schizophrenia-associated HLA locus affects thalamus volume and asymmetry. *Brain Behav Immun, 46*, 311-318. doi:10.1016/j.bbi.2015.02.021

Cannon, M. J., Schmid, D. S., & Hyde, T. B. (2010). Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. *Rev Med Virol, 20*(4), 202-213. doi:10.1002/rmv.655

Cardno, A. G., & Owen, M. J. (2014). Genetic Relationships Between Schizophrenia, Bipolar Disorder, and Schizoaffective Disorder. *Schizophrenia Bulletin*. doi:10.1093/schbul/sbu016

Carrington, M., Nelson, G. W., Martin, M. P., Kissner, T., Vlahov, D., Goedert, J. J., . . . O'Brien, S. J. (1999). HLA and HIV-1: Heterozygote Advantage and B*35-Cw*04 Disadvantage. *Science, 283*(5408), 1748-1752. doi:10.1126/science.283.5408.1748

Carter, C. J. (2009). Schizophrenia Susceptibility Genes Directly Implicated in the Life Cycles of Pathogens: Cytomegalovirus, Influenza, Herpes simplex, Rubella, and Toxoplasma gondii. *Schizophrenia Bulletin, 35*(6), 1163-1182. doi:10.1093/schbul/sbn054

Carter, C. J. (2011). Schizophrenia: A Pathogenetic Autoimmune Disease Caused by Viruses and Pathogens and Dependent on Genes. *Journal of Pathogens, 2011*. doi:10.4061/2011/128318

Chang, A. L. S., Raber, I., Xu, J., Li, R., Spitale, R., Chen, J., . . . Tung, J. Y. (2015). Assessment of the Genetic Basis of Rosacea by Genome-Wide Association Study. *Journal of Investigative Dermatology, 135*(6), 1548-1555. doi:http://dx.doi.org/10.1038/jid.2015.53

Corvin, A., & Morris, D. W. (2014). Genome-wide association studies: findings at the major histocompatibility complex locus in psychosis. *Biol Psychiatry, 75*(4), 276-283. doi:10.1016/j.biopsych.2013.09.018

Debnath, M., Cannon, D. M., & Venkatasubramanian, G. (2013). Variation in the major histocompatibility complex [MHC] gene family in schizophrenia: Associations and functional implications. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 42*, 49-62. doi:http://dx.doi.org/10.1016/j.pnpbp.2012.07.009

Delves, P. J. P. (2014). Human Leukocyte Antigen (HLA) System. *Merck Manual: Professional Version*.

Dickerson, F. B., Boronow, J. J., Stallings, C., Origoni, A. E., Ruslanova, I., & Yolken, R. H. (2003). ASsociation of serum antibodies to herpes simplex virus 1 with cognitive deficits in individuals with schizophrenia. *Archives of General Psychiatry, 60*(5), 466-472. doi:10.1001/archpsyc.60.5.466

Fallin, M. D., Lasseter, V. K., Avramopoulos, D., Nicodemus, K. K., Wolyniec, P. S., McGrath, J. A., . . . Pulver, A. E. (2005). Bipolar I disorder and schizophrenia: a 440-single-nucleotide polymorphism screen of 64 candidate genes among Ashkenazi Jewish case-parent trios. *Am J Hum Genet, 77*(6), 918-936. doi:10.1086/497703

Gragert, L., Fingerson, S., Albrecht, M., Maiers, M., Kalaycio, M., & Hill, B. T. (2014). Fine-mapping of HLA associations with chronic lymphocytic leukemia in US populations. *Blood, 124*(17), 2657-2665. doi:10.1182/blood-2014-02-558767

Hamdani, N., Daban-Huard, C., Lajnef, M., Richard, J. R., Delavest, M., Godin, O., . . . Leboyer, M. (2013). Relationship between Toxoplasma gondii infection and bipolar disorder in a French sample. *J Affect Disord, 148*(2-3), 444-448. doi:10.1016/j.jad.2012.11.034

Health, N. I. o. M. (2016). Schizophrenia. Retrieved from http://www.nimh.nih.gov/health/topics/schizophrenia/index.shtml

Hill, A. V. S., Allsopp, C. E. M., Kwiatkowski, D., Anstey, N. M., Twumasi, P., Rowe, P. A., . . . Greenwood, B. M. (1991). Common West African HLA antigens are associated with protection from severe malaria. *Nature, 352*(6336), 595-600. Retrieved from http://dx.doi.org/10.1038/352595a0

Jeffery, K. J. M., Siddiqui, A. A., Bunce, M., Lloyd, A. L., Vine, A. M., Witkover, A. D., . . . Bangham, C. R. M. (2000). The Influence of HLA Class I Alleles and Heterozygosity on the Outcome of Human T Cell Lymphotropic Virus Type I Infection. *The Journal of Immunology, 165*(12), 7278-7284. doi:10.4049/jimmunol.165.12.7278

Jones, J. L., Kruszon-Moran, D., Wilson, M., McQuillan, G., Navin, T., & McAuley, J. B. (2001). Toxoplasma gondii infection in the United States: seroprevalence and risk factors. *Am J Epidemiol, 154*(4), 357-365.

Keller, M. C., Simonson, M. A., Ripke, S., Neale, B. M., Gejman, P. V., Howrigan, D. P., . . . Sullivan, P. F. (2012). Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. *PLoS Genet, 8*(4), e1002656. doi:10.1371/journal.pgen.1002656

Kessler, R. C., Chiu, W., Demler, O., & Walters, E. E. (2005). PRevalence, severity, and comorbidity of 12-month dsm-iv disorders in the national comorbidity survey replication. *Archives of General Psychiatry, 62*(6), 617-627. doi:10.1001/archpsyc.62.6.617

Khankhanian, P., Gourraud, P.-A., Lizee, A., & Goodin, D. S. (2015). Haplotype-based approach to known MS-associated regions increases the amount of explained risk. *Journal of Medical Genetics, 52*(9), 587-594. doi:10.1136/jmedgenet-2015-103071

Klitz, W., Gragert, L., Maiers, M., Fernandez-Viña, M., Ben-Naeh, Y., Benedek, G., . . . Israel, S. (2010). Genetic differentiation of Jewish populations. *Tissue Antigens, 76*(6), 442-458. doi:10.1111/j.1399-0039.2010.01549.x

Leweke, F. M., Gerth, C. W., Koethe, D., Klosterkötter, J., Ruslanova, I., Krivogorsky, B., . . . Yolken, R. H. Antibodies to infectious agents inindividuals with recent onset schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience, 254*(1), 4-8. doi:10.1007/s00406-004-0481-6

Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *The Lancet, 373*(9659), 234-239. doi:http://dx.doi.org/10.1016/S0140-6736(09)60072-6

Mack, D. G., Johnson, J. J., Roberts, F., Roberts, C. W., Estes, R. G., David, C., . . . McLeod, R. (1999). HLA-class II genes modify outcome of Toxoplasma gondii infection. *Int J Parasitol, 29*(9), 1351-1358. doi:http://dx.doi.org/10.1016/S0020-7519(99)00152-6

Mahdi, B. M. (2015). Role of HLA typing on Crohn's disease pathogenesis. *Annals of Medicine and Surgery, 4*(3), 248-253. doi:http://dx.doi.org/10.1016/j.amsu.2015.07.020

Mahfoudh, N., Ayadi, I., Kamoun, A., Ammar, R., Mallek, B., Maalej, L., . . . Makni, H. (2013). Analysis of HLA-A, -B, -C, -DR, -DQ polymorphisms in the South Tunisian population and a comparison with other populations. *Annals of Human Biology, 40*(1), 41-47. doi:10.3109/03014460.2012.734334

Moraru, M., Cisneros, E., Gómez-Lozano, N., de Pablo, R., Portero, F., Cañizares, M., . . . Vilches, C. (2012). Host Genetic Factors in Susceptibility to Herpes Simplex Type 1 Virus Infection: Contribution of Polymorphic Genes at the Interface of Innate and Adaptive Immunity. *The Journal of Immunology, 188*(9), 4412-4420. doi:10.4049/jimmunol.1103434

Mulle, J. G., Dodd, A. F., McGrath, J. A., Wolyniec, P. S., Mitchell, A. A., Shetty, A. C., . . . Warren, S. T. (2010). Microdeletions of 3q29 confer high risk for schizophrenia. *Am J Hum Genet, 87*(2), 229-236. doi:10.1016/j.ajhg.2010.07.013

Nunes, K., Zheng, X., Torres, M., Moraes, M. E., Piovezan, B. Z., Pontes, G. N., . . . Meyer, D. (2016). HLA imputation in an admixed population: An assessment of the 1000 Genomes data as a training set. *Human Immunology, 77*(3), 307-312. doi:http://dx.doi.org/10.1016/j.humimm.2015.11.004

O'Grady, J., Sutherland, S., Harvey, F., Calne, R., Alexander, G. M., Donaldson, P., . . . Williams, R. (1988). Cytomegalovirus infection and donor/recipient HLA antigens: interdependent co-factors in pathogenesis of vanishing bileduct syndrome after liver transplantation. *The Lancet, 332*(8606), 302-305.

Ollila, Hanna M., Ravel, J.-M., Han, F., Faraco, J., Lin, L., Zheng, X., . . . Mignot, E. (2015). HLA-DPB1 and HLA Class I Confer Risk of and Protection from Narcolepsy. *The American Journal of Human Genetics, 96*(1), 136-146. doi:http://dx.doi.org/10.1016/j.ajhg.2014.12.010

Parham, L. R., Briley, L. P., Li, L., Shen, J., Newcombe, P. J., King, K. S., . . . Spraggs, C. F. (2016). Comprehensive genome-wide evaluation of lapatinib-induced liver injury yields a single genetic signal centered on known risk allele HLA-DRB1[ast]07:01. *Pharmacogenomics J, 16*(2), 180-185. doi:10.1038/tpj.2015.40

Paterson, S., Wilson, K., & Pemberton, J. M. (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (Ovis aries L.). *Proceedings of the National Academy of Sciences, 95*(7), 3714-3719. Retrieved from http://www.pnas.org/content/95/7/3714.abstract

Pearce, B. D., Kruszon-Moran, D., & Jones, J. L. (2012). The relationship between Toxoplasma gondii infection and mood disorders in the third National Health and Nutrition Survey. *Biol Psychiatry, 72*(4), 290-295. doi:10.1016/j.biopsych.2012.01.003

Pearce, B. D., Kruszon-Moran, D., & Jones, J. L. (2012). The Relationship Between Toxoplasma Gondii Infection and Mood Disorders in the Third National Health and Nutrition Survey. *Biol Psychiatry, 72*(4), 290-295. doi:http://dx.doi.org/10.1016/j.biopsych.2012.01.003

Prasad, K. M., Eack, S. M., Goradia, D., Pancholi, K. M., Keshavan, M. S., Yolken, R. H., & Nimgaonkar, V. L. (2011). Progressive Gray Matter Loss and Changes in Cognitive Functioning Associated With Exposure to Herpes Simplex Virus 1 in Schizophrenia: A Longitudinal Study. *American Journal of Psychiatry, 168*(8), 822-830. doi:doi:10.1176/appi.ajp.2011.10101423

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet, 81*(3), 559-575. doi:10.1086/519795

Regier, D. A., Narrow, W. E., Rae, D. S., Manderscheid, R. W., Locke, B. Z., & Goodwin, F. K. (1993). The de facto us mental and addictive disorders service system: Epidemiologic catchment

area prospective 1-year prevalence rates of disorders and services. *Archives of General Psychiatry, 50*(2), 85-94. doi:10.1001/archpsyc.1993.01820140007001

Rizzo, L. B., Do Prado, C. H., Grassi-Oliveira, R., Wieck, A., Correa, B. L., Teixeira, A. L., & Bauer, M. E. (2013). Immunosenescence is associated with human cytomegalovirus and shortened telomeres in type I bipolar disorder. *Bipolar Disord, 15*(8), 832-838. doi:10.1111/bdi.12121

Rodrigues, M. d. L., Deghaide, N. H., Figueiredo, J. F., de Menezes, M. B., Demarco, A. L., Donadi, E., & Fernandes, A. P. (2016). Toxoplasmic encephalitis: role of Human Leucocyte Antigens/alleles associated with rapid progression to Acquired Immunodeficiency Syndrome. *The Brazilian Journal of Infectious Diseases, 20*(2), 115-118. doi:http://dx.doi.org/10.1016/j.bjid.2015.10.010

Samandary, S., Kridane-Miledi, H., Sandoval, J. S., Choudhury, Z., Langa-Vives, F., Spencer, D., . . . BenMohamed, L. (2014). Associations of HLA-A, HLA-B and HLA-C alleles frequency with prevalence of herpes simplex virus infections and diseases across global populations: implication for the development of an universal CD8+ T-cell epitope-based vaccine. *Human Immunology, 75*(8), 715-729.

Scharpf, R. B., Irizarry, R. A., Ritchie, M. E., Carvalho, B., & Ruczinski, I. (2011). Using the R Package crlmm for Genotyping and Copy Number Estimation. *J Stat Softw, 40*(12), 1-32.

Shimokawa, P. T., Targa, L. S., Yamamoto, L., Rodrigues, J. C., Kanunfre, K. A., & Okay, T. S. (2016). HLA-DQA1/B1 alleles as putative susceptibility markers in congenital toxoplasmosis. *Virulence*, 1-9. doi:10.1080/21505594.2016.1150401

Tedla, Y., Shibre, T., Ali, O., Tadele, G., Woldeamanuel, Y., Asrat, D., . . . Habte, A. (2011). Serum antibodies to Toxoplasma gondii and Herpesvidae family viruses in individuals with schizophrenia and bipolar disorder: a case-control study. *Ethiopian medical journal, 49*(3), 211-220. Retrieved from http://europepmc.org/abstract/MED/21991754

Thursz, M. R., Thomas, H. C., Greenwood, B. M., & Hill, A. V. S. (1997). Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nat Genet, 17*(1), 11-12. Retrieved from http://dx.doi.org/10.1038/ng0997-11

Torrey, E. F., Bartko, J. J., Lun, Z.-R., & Yolken, R. H. (2007). Antibodies to Toxoplasma gondii in Patients With Schizophrenia: A Meta-Analysis. *Schizophrenia Bulletin, 33*(3), 729-736. doi:10.1093/schbul/sbl050

Torrey, E. F., Leweke, M. F., Schwarz, M. J., Mueller, N., Bachmann, S., Schroeder, J., . . . Yolken, R. H. (2012). Cytomegalovirus and Schizophrenia. *CNS Drugs, 20*(11), 879-885. doi:10.2165/00023210-200620110-00001

Torrey, E. F., & Robert, H. Y. (2003). Toxoplasma gondii and Schizophrenia. *Emerging Infectious Disease journal, 9*(11), 1375. doi:10.3201/eid0911.030143

Yolken, R. (2004). Viruses and schizophrenia: a focus on herpes simplex virus. *Herpes : the journal of the IHMF, 11 Suppl 2*, 83A-88A. Retrieved from http://europepmc.org/abstract/MED/15319094

Yolken, R. H., & Torrey, E. F. (2008). Are some cases of psychosis caused by microbial agents? A review of the evidence. *Mol Psychiatry, 13*(5), 470-479. Retrieved from http://dx.doi.org/10.1038/mp.2008.5

http://www.nature.com/mp/journal/v13/n5/pdf/mp20085a.pdf

Zheng, X., Shen, J., Cox, C., Wakefield, J. C., Ehm, M. G., Nelson, M. R., & Weir, B. S. (2014). HIBAG—HLA genotype imputation with attribute bagging. *The Pharmacogenomics Journal, 14*(2), 192-200. doi:10.1038/tpj.2013.18