Published in final edited form as:

Sci Transl Med. 2017 May 10; 9(389): . doi:10.1126/scitranslmed.aai8708.

Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants

Jason H Karnes^{1,*}, Lisa Bastarache^{2,*}, Christian M Shaffer³, Silvana Gaudieri^{4,5,6}, Yaomin Xu^{7,8}, Andrew Glazer³, Jonathan D Mosley³, Shilin Zhao⁸, Soumya Raychaudhuri^{9,10,11,12,13,14}, Simon Mallal^{5,6,15}, Zhan Ye¹⁶, John G Mayer¹⁶, Murray H Brilliant¹⁷, Scott J Hebring¹⁷, Dan M Roden^{2,3,18}, Elizabeth J Phillips^{3,15}, and Joshua C Denny^{2,3,*}

¹Department of Pharmacy Practice and Science, University of Arizona College of Pharmacy, Tucson, AZ

²Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN

³Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN

⁴School of Anatomy, Physiology and Human Biology, University of Western Australia, Nedlands, Western Australia, Australia

⁵Division of Infectious Diseases, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN

⁶Institute for Immunology and Infectious Diseases, Murdoch University, Murdoch, Western Australia, Australia

⁷Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN

⁸Center for Quantitative Sciences, Vanderbilt University School of Medicine, Nashville, TN

⁹Division of Rheumatology, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, USA

*2525 West End Avenue #6049, Nashville, TN 37203, Telephone: (615) 936-5034; josh.denny@vanderbilt.edu.

*Authors contributed equally

Overline: Genomics

Related Resources: None

Supplementary Materials:

Materials and Methods(48–66)

Figure S1. PheWAS plot of all four digit HLA allele associations

Figure S2. QQ plot of associations from four digit HLA allele PheWAS analysis

Figure S3. QQ plot of associations from two digit HLA allele PheWAS analysis

Figure S4. QQ plot of associations from non-synonymous HLA variant PheWAS analysis

Table. S1. Most significant results from Trait-based Association Test (TATES)

Author contributions: J.H.K., L.B., S.R., D.M.R., E.J.P. and J.C.D. designed the study and interpreted the results. L.B., C.M.S., D.M.R., and J.C.D. coordinated and managed genotype and phenotype data ascertainment in the BioVU discovery cohort. Z.Y., J.G.M., M.H.B., and S.J.H. coordinated and managed genotype and phenotype data ascertainment in the PRMP replication cohort. S.G., S.M., and E.J.P. performed HLA typing and analysis of HLA data. L.B., C.M.S., Y.Z., A.G., J.D.M. and S. Z. performed statistical and bioinformatic analyses. J.H.K. drafted the manuscript. All authors contributed to the final version of the manuscript.

Competing interests: The authors declare no competing financial interests.

Data and materials availability: The results are freely available as a web catalog of HLA risk loci and their associated phenotypes at www.phewascatalog.org. These data are also available in the database of Genotypes and Phenotypes (dbGaP).

¹⁰Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, USA

¹¹Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, USA

¹²Partners Center for Personalized Genetic Medicine, Boston, USA

¹³Institute of Inflammation and Repair, University of Manchester, Manchester, UK

¹⁴Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden

¹⁵Department of Pathology, Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, TN

¹⁶Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, Wisconsin, USA

¹⁷Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, Wisconsin, USA

¹⁸Department of Pharmacology, Vanderbilt University School of Medicine, Nashville, TN

Abstract

While many phenotypes have been associated with variants in human leukocyte antigen (HLA) genes, the full phenotypic impact of HLA variants across all diseases is unknown. We imputed HLA genomic variation from two populations of 28,839 and 8,431 European ancestry individuals and tested association of HLA variation with 1,368 phenotypes. A total of 104 four-digit and 92 two-digit HLA allele-phenotype associations were significant in both discovery and replication cohorts, the strongest being *HLA-DQB1*03:02* and type 1 diabetes. Four previously unidentified associations were identified across the spectrum of disease with two and four digit HLA alleles and ten with non-synonymous variants. Some conditions associated with multiple HLA variants and stronger associations with more severe disease manifestations were identified. A comprehensive, publicly-available catalog of clinical phenotypes associated HLA variation is provided. Examining HLA variant disease associations in this large dataset allows comprehensive definition of disease associations to drive further mechanistic insights.

Introduction

The major histocompatibility complex (MHC) and human leukocyte antigen (HLA) genes have been extensively studied due to their key role in immune responses(1). Despite containing only a small fraction of all human genes, the MHC contains the greatest number of significant associations from genome-wide association studies (GWAS) relative to any other region in the genome(2). In GWAS alone, single nucleotide polymorphisms (SNPs) in the MHC have been associated with over 100 different phenotypes according to the National Human Genome Research Institute's GWAS catalogue(3). Four digit HLA alleles, which characterize MHC proteins at the amino acid level, and key HLA amino acid residues are often found to have critical roles in immune-mediated disease. This HLA variation can be

now imputed with high accuracy from SNP-level data(4–6). Previous investigations including GWAS have utilized a phenotype to genotype approach, which involves identifying individuals with a specific phenotype first and then determining genetic associations with that phenotype. The broader question of the influence of HLA variation in these genes across multiple human disease phenotypes has not been comprehensively investigated(7).

Here we report the results of a large phenome-wide association study (PheWAS) that systematically examined relationships between variants in the MHC and a broad set of human disease phenotypes. PheWAS using electronic health record (EHR) systems have replicated hundreds of previous GWAS associations (including many associations with MHC SNPs), identified new associations, and have been used in other applications, such as an evaluation of the phenotypic legacy of Neandertals in modern humans(7–14). We conducted a PheWAS on European ancestry individuals genotyped on the Illumina® HumanExome BeadChip and tested association of imputed HLA variants with 1,368 EHR phenotypes. Populations were derived from two institutions with DNA biobanks coupled to de-identified EHRs: Vanderbilt University Medical Center's BioVU (n=28,839) and the Marshfield Clinic's Personalized Medicine Research Project (PMRP, n=8,431). The results are freely available as a web catalog of HLA risk loci and their associated phenotypes at www.phewascatalog.org.

Results

Primary PheWAS Analysis

A total of 99 four digit HLA alleles, 77 two digit alleles, and 988 non-synonymous variants were imputed using SNP2HLA(6) and included in the analysis (Table 1). In a subset of 3,152 BioVU individuals of European ancestry who also underwent direct HLA sequencing, the overall concordance rate and call rate for SNP2HLA-imputed four digit alleles compared to sequencing were 96.93% and 99.98% respectively. Using a false discovery rate (FDR) adjustment of $q < 0.10$ in the discovery cohort (BioVU) and $p < 0.05$ in the replication cohort (PMRP), a total of 104 significant associations with the same direction of effect were identified between four digit HLA alleles and the 1,368 phenotypes tested (Figure 1 and Figure S1). A total of 284 four digit allele associations were significant in BioVU before replication in PMRP. The 104 significant associations were observed with 29 distinct phenotypes and 29 unique alleles. After meta-analysis of associations in BioVU and PMRP, there were 12 four-digit allelic associations with a meta-analysis P value 1.9×10^{-25} (Table 2; other associations are presented at www.phewascatalog.org). The strongest was observed between type 1 diabetes (T1D) and *HLA-DQB1*03:02* (OR 3.62, meta-analysis $p = 1.04 \times 10^{-54}$). In the analysis of two digit alleles, 92 significant associations were observed with 24 distinct phenotypes and 25 unique alleles. A total of 205 two digit allele associations were significant in BioVU before replication in PMRP. There were 9 meta-analysis associations with $P < 1.9 \times 10^{-25}$ (Table 2; other associations are presented at www.phewascatalog.org), and the strongest was a between *HLA-B*27* and ankylosing spondylitis (AS, OR 32.48, meta-analysis $p = 3.06 \times 10^{-45}$). In the analysis of non-synonymous variants, a total of 1,759 phenotype- non-synonymous variant associations were

observed with 53 distinct phenotypes and 467 unique amino acid substitutions. A total of 3,862 non-synonymous variant associations were significant in BioVU before replication in PMRP. The strongest association observed was between AS and an asparagine substitution at position 97 of *HLA-B* (OR 33.25, meta-analysis $p=1.17\times 10^{-45}$). QQ plots for association analyses are presented in the Supplementary Materials (Figure S2–S4). Results acquired using the additive genetic model were similar to results acquired using the dominant model.

Table 3 compares major HLA genetic associations for common autoimmune/inflammatory diseases based on previous literature and the association results from the PheWAS analysis in BioVU. The major association for each disease was highly significant with effect sizes similar to those previously published, suggesting the validity of the approach and results. Further evaluation of PheWAS results compared to previous associations is presented in Supplementary Materials.

The results also provide supportive evidence for associations previously reported but not yet widely-accepted. For instance, one of the strongest associations observed with multiple sclerosis (MS) was for a non-synonymous variant at position 71 in *HLA-DRB1*, (OR 2.66, meta-analysis $p=3.54\times 10^{-16}$) consistent with previous reports(15, 16). Similarly, cervical cancer has been linked with class II HLA alleles and class II HLA haplotypes(17, 18) and we observed a significant association between ‘Cervical cancer and dysplasia’ and non-synonymous variation at position 37 of *HLA-DRB1* (OR 1.43, meta-analysis $p=7.22\times 10^{-7}$).

Previously Unidentified Associations

This PheWAS also identified associations between immune-mediated diseases and non-synonymous variants where only a four or two digit allele association was previously reported. For instance, several HLA alleles have been associated with increased risk for Sicca syndrome (SS) in European populations, including *HLA-DQB1*02:01*(19, 20), but the association of non-synonymous variants with SS is largely unknown(21). This PheWAS identified *HLA-DQB1*02:01* as the strongest HLA risk factor in our population and also implicated non-synonymous variation at position 77 in *HLA-DRB1* with SS susceptibility. Similarly, uveitis is known to be associated with *HLA-B*27*(22) and the PheWAS now implicates non-synonymous variation at position 114 of *HLA-B* with uveitis susceptibility. Notably, this non-synonymous variant association with uveitis was observed to have a higher odds ratio and stronger association than the four digit allele. These results may clarify the driving association for a given phenotype is a two or four digit allele, or a non-synonymous variant, based on the strength of association and the work supports further attempts to identify critical and more shared peptide residues among autoimmune diseases.

The PheWAS also identified associations with diseases for which evidence for predisposing HLA variation is minimal or has not been previously reported. These were primarily with non-synonymous variation (Table 4), and were observed across organ systems (neurologic, gastrointestinal, genitourinary, endocrine, infectious, oncologic, neurologic, cardiologic, and respiratory conditions). The large numbers of subjects studied also allowed evaluation of “subphenotypes” such as those indicating specific T1D complications or manifestations. In this analysis, a stronger OR was consistently observed for the T1D sub-phenotype than for the T1D phenotype (Table 5). For instance, the OR for the association between *HLA-*

*DQB1*0302* and T1D (OR 4.42) was reduced compared to that between *HLA-DQB1*0302* and T1D with ketoacidosis (OR 7.09), T1D with neurological manifestations (OR 6.26), T1D with ophthalmic manifestations (OR 7.33), and T1D with renal manifestations (OR 6.66). Similar significant associations were observed for phenotypes indicating more severe manifestations of disease, including non-synonymous variants associations with psoriatic arthropathy compared to psoriasis (Table 5).

Pleiotropic associations

For four and two digit HLA alleles, we found consistent evidence of association with multiple phenotypes. Figure 2 represents a heatmap of four digit HLA allele associations. In order to determine cross-phenotype associations potentially driven by pleiotropic effects, we implemented the Trait-based Association Test (TATES)(23). TATES was implemented for all phenotypes that had over 40 cases and crossed the FDR significance threshold for any four digit HLA allele (MAF>1%) in BioVU. We found 60 HLA alleles with significant ($p<0.05$) evidence for cross-phenotypic effects with the strongest evidence for *HLA-B*27:05* (TATES p value= 7.76×10^{-68}), which was associated with AS and uveitis (Supplementary Table 1). Among four digit alleles associated with multiple phenotypes, *HLA-DQA1*01:02* and *HLA-DQB1*02:01* were significantly associated with 8 phenotypes, the largest number of associations with any allele in our dataset. The most significant results of the TATES analysis and the HLA alleles with multiple phenotypic associations are presented in the Supplemental Materials (Supplementary Table 1). Shared HLA signatures have been previously observed for T1D(24–26) and T1D was associated with most of the alleles that had the strongest evidence of cross-phenotypic associations. The *HLA-DQB1*02:01* and *DRB1*03:01* alleles shared associations with T1D, SS, systemic lupus erythematosus, and celiac disease. Shared associations were also common among the non-synonymous variants. For instance, substitutions at position 26 of *HLA-DRB1* were significantly associated with an increased risk of T1D, RA, celiac disease, SS, and diffuse disease of connective tissue.

In order to further refine our models of pleiotropy, we conducted conditional analyses of phenotypes in a pairwise fashion. The purpose of the conditional analysis was to determine whether association of multiple phenotypes with the same allele were due to co-occurrence of phenotypes rather than independent association of both phenotypes with the allele. This analysis showed that HLA associations with T1D, rheumatoid arthritis (RA), MS, SS, and celiac disease were independent. Notably, the HLA associations with RA and polymyalgia rheumatica appeared to be independent in conditional analyses, as did HLA associations with AS and uveitis in addition to T1D and diabetic retinopathy. However, conditional analyses suggested that the optic neuritis, ‘other demyelinating diseases of central nervous system,’ and ‘lack of coordination’ phenotypes are more strongly associated with MS than with HLA variation, i.e. these associations were driven by an association with MS.

To further understand the relationships between HLA loci, alleles, and phenotypes, we conducted a network analysis using nodes to represent PheWAS codes, HLA genes, and HLA alleles with links representing associations that were significant at $p<1\times 10^{-5}$. HLA genes and their alleles are further organized into a hierarchical structure by linking alleles to HLA genes. A total of 135 PheWAS analysis associations between four-digit HLA alleles

and PheWAS codes were observed in the BioVU dataset. The majority of the most significant associations (at $p < 1 \times 10^{-15}$) are among T1D with *HLA-DQB1*02:01*, *-DRB1*03:01* and *-DRB1*03:01*, T1D/RA with *HLA-DQA1*03:01* and *-DRB1*01:02*, and T1D/MS with *HLA-DQB1*06:02* and *-DRB1*04:01* but with the opposite (protective) effect on MS. Figure 3 presents an undirected graph with nodes representing PheWAS codes, HLA genes, and four digit HLA alleles. Nodes with more significant associations and that share more connections are topologically clustered together. The layout is intuitive to visualize the complex PheWAS associations where clusters of PheWAS codes stay closer due to shared genetic associations, and vice versa. The genes *HLA-DRB1*, *-DQB1*, and *-DQA1* cluster together strongly with autoimmune diseases with T1D and T1D sub-phenotypes at the focus. An interactive version of Figure 3 is available at <http://phewascatalog.org>.

Discussion

We conducted a systematic test of phenomic associations with HLA variation. Significant associations with large effect sizes were replicated in autoimmune conditions well represented in the population, such as T1D, AS, MS, and RA, providing strong evidence of the validity of the PheWAS approach and HLA imputation. We find evidence for several previously undescribed associations with MS and cervical cancer. Previously unidentified associations were also observed between HLA variation and multiple phenotypes, including gastrointestinal hemorrhage, nodular lymphoma, cancer of the brain and nervous system, cervical cancer & dysplasia, kidney transplant, partial epilepsy, and atherosclerosis of the extremities. We found consistent evidence for pleiotropy between HLA variation and common autoimmune diseases, especially for class II HLA alleles associated with T1D and other autoimmune diseases. Our analysis confirms the ability of PheWAS to identify multiple associations from well-powered studies. In each of the common diseases studied, the most significant known HLA contributor to disease was identified as a risk factor in this study, suggesting the validity of the results. Our results also provide a publicly available, comprehensive set of HLA risk loci for validation of previous findings and hypothesis generation for future research.

The strongest observed associations in our study were with T1D, in keeping with the fact that T1D is a common, dependably diagnosed disease and 40–50% of T1D heritability has been attributed to HLA variability(27). Genetic influences on T1D risk include a hierarchy of predisposing and protective HLA haplotypes, with the greatest risk conferred by the DR3-DQ2 and DR4-DQ7/8 haplotypes, especially in patients heterozygous for both haplotypes(28, 29). These haplotypes are well represented among our observed associations. The strongest association observed overall was for a non-synonymous variant at position 57 of *HLA-DQB1*, a locus which was identified early as a susceptibility locus for T1D(30, 31). Our data support this observation and the notion that class II HLA allelic associations for T1D are a reflection of underlying amino acid substitutions involved directly in autoimmune etiology. In this study, the non-synonymous 57 *HLA-DQB1* variant was a stronger association than any other HLA variation for T1D, suggesting that the non-synonymous 57 *HLA-DQB1* variant is the driver of the association, which is in line with previous observations(30). Our observations were consistent with prior disease models for a number

of other autoimmune diseases, such as RA, AS, MS, celiac disease, psoriasis, primary biliary cirrhosis, and SS.

This is one of the first systematic studies of clinical phenotypes with HLA variation in a single population. Previous studies investigating phenotype associations with HLA variation have identified similar associations through recruitment of specific case control cohorts. The availability of extensive phenotype data enables an in depth investigation of pleiotropy. This extensive phenotype data also allows investigation of sub-phenotypes, such as T1D with various manifestations of the disease. A larger OR was consistently observed for the T1D sub-phenotype than for the T1D phenotype. This might indicate a stronger HLA association with T1D disease severity or potentially a phenotype with less misclassification. Finally, we consider the paucity of observed associations for non-autoimmune phenotypes to be interesting and informative. The relatively small number of associations with infectious diseases, psychiatric disorders, or oncologic disorders could be due to the lack of a significant biological effect of HLA variation on these disorders or a lack of power for these relatively uncommon phenotypes within our dataset in the presence of multiple comparisons.

One limitation of the work is that the observed strength of association depends on the accuracy of ICD9 code diagnosis, accuracy of imputation, and frequency of both diagnosis and allele in our population. Any misclassification error should underestimate the magnitude of effect size estimates, thus leading to a bias toward a null result. PheWAS has been previously shown to accurately replicate diverse phenotype-genotype associations(12, 14), and data presented here replicate many previous reports of HLA associations, as noted above. Similarly, our results depend on the accuracy of prediction of HLA alleles and non-synonymous variants based on SNP data from the HumanExome BeadChip. Previous literature suggests HLA alleles can be imputed accurately(4–6, 32) and alleles were imputed with high accuracy compared to sequence data in a subset of our population. Rare alleles were not included here and the study does have limited power to detect significant associations for phenotypes with limited sample sizes in our population. While the total number of individuals within our study is large, the number of individuals with specific phenotypes was relatively small, which limits our ability to detect associations with modest effect sizes. Additional diagnostic information that would add specificity, such as anti-cyclic citrullinated peptides (CCP) antibodies in RA, was not employed in this study. Finally, we encountered challenges in the selection of appropriate methods for multiple comparisons adjustment that minimized false positive results while accounting for strong linkage disequilibrium in the HLA region. Such challenges in defining statistical significance with extensive correlations in the genotypes and phenotypes have likely resulted in some false positive associations.

We provide large-scale replication of phenotypic associations with variation in the HLA region. Our observations are valuable since they add evidence to known and uncertain HLA-disease associations and may help uncover HLA variation that drives a given immune-mediated phenotype and identify key residues that provide mechanistic insights. The associated type of HLA variation (i.e. four digit allele versus non-synonymous variant) gives clues to the molecular and structural basis of disease. For some diseases, variation in HLA may be necessary but not sufficient and in some cases HLA alleles may be protective with

effects that outweigh risk alleles. A two versus four digit association may point to the serotype, the HLA protein which is expressed from the respective HLA gene and is detected by the respective antibody, more broadly associated with general phenotypes of diseases and may relate to shared peptide binding specificities. Locations of non-synonymous variants in peptide binding domains or killer cell immunoglobulin-like receptor (KIR) interaction sites may also inform disease mechanisms. Our results are freely available and provide a comprehensive set of HLA risk loci for additional research, allowing an in-depth evaluation of the underlying genetic architectures for a range of diseases.

Materials and Methods

Study Design

The objective of our study was to determine associations between variation in the HLA region and a broad set of phenotypes. We employed a PheWAS approach, which tests the association of genomic variation with multiple phenotypes (8). Multiple phenotypes for each individual were derived from de-identified EHR data. HLA variation was imputed using SNP-level data from GWAS platforms using two large biobanks that link DNA extracted from discarded blood samples to de-identified EHRs. The discovery analysis was performed using BioVU, the Vanderbilt DNA databank, and the replication analysis was performed in Marshfield Clinic's PMRP. All individuals with sufficient EHR and genotype data (described below) from the replication and discovery cohorts were included in a non-randomized and non-blinded fashion.

Patient Populations

The discovery cohort included a population from BioVU, the Vanderbilt DNA databank that links DNA extracted from discarded blood samples to de-identified EHRs(33). BioVU accrues DNA samples extracted from blood drawn for routine clinical testing. A full description of this resource and its associated ethical, privacy and other protections has been previously published(33). At the time of this study, BioVU contained approximately 210,000 patients with DNA and 28,839 self-identified white/non-Hispanic patients with available Illumina HumanCoreExome BeadChip genotyping were included. The mean age of the individuals was 55.4 years and 53% were female. BioVU operated on an opt-out basis until January 2015 and on an opt-in basis since(33). The phenotypic data in BioVU are all de-identified and the study was designated "non-human subjects" research by the Vanderbilt Institutional Review Board.

The replication cohort consisted of self-identified white/non-Hispanic Marshfield Clinic patients recruited into the PMRP as described previously(34). PMRP represents a homogenous population with 77% of participants claiming German ancestry. In this PheWAS, 8,431 patients were included for independent validation of BioVU signals. PMRP Discovery set participants were all over age 40 with over 30 years of EHR data on average, and have been genotyped by Illumina® HumanExome BeadChip.

Genotyping and HLA imputation

All samples were genotyped using the Illumina® HumanExome BeadChip, which contains 2,061 HLA tag SNPs. Before imputation, SNP data were cleaned using the quality control (QC) pipeline developed by the eMERGE Genomics Working Group(35, 36). Classical two and four digit HLA alleles and HLA non-synonymous variants were imputed from SNP data on the HumanExome BeadChip using SNP2HLA(6) with the Type 1 Diabetes Genetics Consortium (T1DGC) reference panel, a marker window size of 1000, and a posterior probability (gprob) threshold of 0.5. We restricted our analysis to individuals of at least 90% European descent, as defined using 3,242 ancestry informative markers (AIMs) from the HumanExome BeadChip input into STRUCTURE using HapMap reference populations(37). Identity-by-descent (IBD) analysis was performed and related individuals (parent-child and full sibling) were removed from the dataset. Principal components (PC) analysis was performed on AIMs using the EIGENSTRAT method implemented in PLINK(38, 39). To minimize confounding by population stratification, association analyses were adjusted for PCs 1 and 2. We imputed individual dosages for classical four-digit alleles, two-digit alleles and polymorphic non-synonymous variants at *HLA-A*, *-B*, *-C*, *-DPA1*, *-DPB1*, *-DQA1*, *-DQB1* and *-DRB1* from 28,839 individuals of genetically-determined European ancestry. All HLA alleles and non-synonymous variants with minor allele frequencies greater than 0.01 were included in the analysis.

In order to confirm accuracy of HLA imputation for four digit alleles, high resolution, four-digit HLA sequencing was performed in 3,152 subjects from the Vanderbilt Electronic Systems for Pharmacogenomic Assessment (VESPA) cohort(40) for *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DR*, *HLA-DP*, and *HLA-DQ* at the Institute for Immunology and Infectious Diseases (IIID) at Murdoch University in Perth, Australia. The IIID is accredited by the American Society for Histocompatibility and Immunogenetics (ASHI) and the National Association of Testing Authorities (NATA) and is an established center for sequencing the HLA loci and participating in large, multicenter trials which use HLA sequence data(41, 42). All samples that were successfully typed were included in the study population. Concordance rates and call rates for four digit HLA alleles were then calculated for SNP2HLA compared to direct HLA sequencing results. A detailed description of HLA sequencing methods are available in the Supplemental Materials.

PheWAS Analysis

Phenotypes were defined by patient EHR data as previously described(11, 13, 14). The ICD9 coding system describes diseases, signs and symptoms, injuries, poisonings, procedures and screening codes(43). We translated the ICD9 data into the custom 'case phenotypes' (phecodes) and their algorithm-defined control groups using version 1.2 of the phecode mappings (available from <http://phewascatalog.org>)(14). Phecodes aggregate codes so that their granularity is appropriate for research purposes, including large-scale genomic studies(11). We used a minimum code count threshold of 2 distinct days to define a case for a phenotype. We tested only phenotypes with at least 40 cases, including a total of 1,368 phenotypes, and used a dominant genetic model adjusted for age, gender, and PCs one and two.

In the primary discovery PheWAS analysis, a total of 99 four-digit HLA alleles, 77 two-digit HLA alleles, and 988 non-synonymous variants were tested across 1,368 phenotypes that had at least 40 cases. Phenotypes reflected diagnoses extracted from the EHR and included a broad range of diseases grouped into the following categories: infectious, neoplastic, endocrine and metabolic, hematopoietic, psychiatric, neurologic, cardiovascular, pulmonary, digestive, genitourinary, dermatologic, musculoskeletal, symptoms and signs, and injuries.

In the discovery cohort, we employed the FDR method to correct for multiple comparisons in association tests by variant type (two or four digit HLA allele, non-synonymous variant) (44). We declared statistical significance at a q value cutoff of 0.10 in the discovery cohort (BioVU) combined with $p < 0.05$ in the replication cohort (PMRP). All predicted HLA alleles and non-synonymous variants were tested for association with phenotypes derived from the EHR in a dominant model adjusted for age, gender, and PCs one and two. Although the dominant genetic model was used for the primary analysis, an additive model was also performed for comparison. We conducted association tests using logistic regression to predict case/control status based on the HLA allele, age, sex, and a categorical ascertainment variable based on the experiment under which the genotyping was originally performed. Fixed effects meta-analysis was employed to determine overall effects of HLA variation in both discovery and replication cohorts. The analysis was conducted using PLINK and Perl scripts and graphics were generated using the R PheWAS package(11, 45).

Pleiotropy Assessment and Network Visualization of PheWAS Associations

In order to assess pleiotropic or cross-phenotypic associations for four digit HLA alleles in the discovery cohort, we first calculated the number of phenotypes significantly associated with each allele. We then performed conditional analysis when multiple significant associations with the same allele were present, adjusting for the second phenotype in a logistic regression to predict the first phenotype. The purpose of the conditional analysis was to determine whether association of multiple phenotypes with the same allele were due to co-occurrence of phenotypes rather than independent association of both phenotypes with the allele. For instance, the association of an allele with both multiple sclerosis and the 'lack of coordination' phenotype might suggest that co-occurrence of 'lack of coordination' and MS is driving this association. In order to adjust for this effect, we performed regressions adjusted for the second phenotype whenever an allele was significantly associated with multiple phenotypes. If an allelic association became non-significant after this adjustment, we concluded that the allelic association was due to co-occurrence with another phenotype, rather than a true association of the allele.

A Trait-based Association Test (TATES)(23) was conducted using R (www.R-project.org) for all phenotypes that had over 40 cases and crossed the FDR significance threshold for any four digit HLA allele ($MAF > 1\%$) in BioVU. TATES combines the p values obtained in single marker associations to determine a global p value while correcting for observed correlation structure among phenotypes(23, 46). The TATES procedure affords improved power to detect a combined effect for a group of correlated phenotypes. Thereby, we examined the significance of TATES result to indicate a pleiotropic effect. All pleiotropy analyses were performed on a dataset pruned of redundant phenotypes, including phecode

“subphenotypes” that indicated specific T1D complications such as T1D with ketoacidosis, neurological manifestations, ophthalmic manifestations, or renal manifestations. To minimize co-occurrence of related phenotypes in the TATES analysis, we removed redundant phenotypes related to complications (e.g. “Type 1 diabetes with renal manifestations”) and broad versus narrow phenotype classifications (e.g. “Hypotension” was pruned while “Orthostatic hypotension” was preserved). We also eliminated sex specific phenotypes such as “Dysplasia of female genital organs.”

To visualize the complex many-to-many relationships between phenotypes and HLA genes from the PheWAS analysis, we generated both a heatmap and a network map for four digit allelic associations. The heatmap was generated using the odd ratios of associations between 4-digit HLA alleles and phenotypes. Heatmaps were generated with the heatmap.2 function in the R package gplots and colors were chosen from the R package RColorBrewer. All phenotypes and HLA alleles with at least one association with $p < 1 \times 10^{-5}$ were included in the heatmap. Dendrograms of phenotypes and HLA alleles clustering were generated with heatmap.2 and used Euclidean distance. To determine which clusters represented actual structure in the dataset, a bootstrap analysis was performed. For both the HLA allele and phenotype clustering, 1,000 simulations were performed, sampling with replacement from the actual dataset for each simulation. Clusters that were present in over 90% of the simulations were identified and indicated in the heatmap.

For the network map, we used R package igraph 1.0.0 and constructed a graph representation that maps associations between phenotypes and HLA genes/alleles(47). In this network map, we use nodes to represent either PheWAS codes or HLA genes/alleles and links between HLA alleles and PheWAS codes represent the significant associations between four-digit HLA alleles and PheWAS codes that are considered significant at $\alpha = 1 \times 10^{-5}$. In addition, HLA genes and their alleles are further organized into a hierarchical structure by linking all alleles to its HLA genes. A Force-directed layout is then applied to the graph to stimulate the network graph as a physical system so links act as springs that attract connected nodes closer together and are weighted by p-values so more significantly associated nodes have stronger attraction forces to stay closer. As a result, nodes with more significant associations and share more connections are topologically clustered together. The layout is intuitive to visualize the complex PheWAS associations where clusters of PheWAS codes stay closer due to shared genetic associations, and vice versa.

Acknowledgments

Funding: The dataset used in the analyses described were obtained from Vanderbilt University Medical Centers BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant ULTR000445 from NCATS/NIH. HumanExome BeadChip genotyping was supported institutionally. PheWAS method and website deployment is supported by R01 LM010685. BioVU is supported by institutional funding and by the Vanderbilt CTSA grant UL1 TR000445 from NCATS/NIH. Marshfield investigators and the PMRP were supported by the generous donors of the Marshfield Clinic, National Institutes of Health National Center for Advancing Translations Sciences grant number UL1TR000427, National Human Genome Research Institute grant number 1U01HG006389, National Institute of General Medical Science grant number 1R01GM114128 and National Library of Medicine grant number 1K22LM011938. This work was supported by the National Institutes of Health, including P50 GM115305, U19 HL65962, P30 AI110527, 1R01AI103348, and 5T32AI007474. JHK receives support from the American Heart Association (16SDG29090005 and 15POST22660017), the VUMC Clinical Pharmacology Training grant (T32 GM07569), and the American College of Clinical Pharmacy Research Institute (Futures Grant). SR is

supported by 5U01GM092691-04 and 1R01AR062886-01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References and Notes

1. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet.* 2013; 14:661–673. [PubMed: 23917628]
2. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, Lush MJ, Povey S, Talbot CC Jr, Wright MW, Wain HM, Trowsdale J, Ziegler A, Beck S. Gene map of the extended human MHC. *Nat Rev Genet.* 2004; 5:889–899. [PubMed: 15573121]
3. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014; 42:D1001–1006. [PubMed: 24316577]
4. Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR, Weir BS. HIBAG--HLA genotype imputation with attribute bagging. *Pharmacogenomics J.* 2014; 14:192–200. [PubMed: 23712092]
5. Dilthey AT, Moutsianas L, Leslie S, McVean G. HLA*IMP--an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics.* 2011; 27:968–972. [PubMed: 21300701]
6. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, Raychaudhuri S, de Bakker PI. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One.* 2013; 8:e64683. [PubMed: 23762245]
7. Liu J, Ye Z, Mayer JG, Hoch BA, Green C, Rolak L, Cold C, Khor SS, Zheng X, Miyagawa T, Tokunaga K, Brilliant MH, Hebring SJ. Phenome-wide association study maps new diseases to the human major histocompatibility complex region. *J Med Genet.* 2016
8. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Field JR, Pulley JM, Ramirez AH, Bowton E, Basford MA, Carrell DS, Peissig PL, Kho AN, Pacheco JA, Rasmussen LV, Crosslin DR, Crane PK, Pathak J, Bielinski SJ, Pendergrass SA, Xu H, Hindorff LA, Li R, Manolio TA, Chute CG, Chisholm RL, Larson EB, Jarvik GP, Brilliant MH, McCarty CA, Kullo IJ, Haines JL, Crawford DC, Masys DR, Roden DM. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol.* 2013; 31:1102–1110. [PubMed: 24270849]
9. Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat Rev Genet.* 2016; 17:129–145. [PubMed: 26875678]
10. Hall MA, Verma A, Brown-Gentry KD, Goodloe R, Boston J, Wilson S, McClellan B, Sutcliffe C, Dilks HH, Gillani NB, Jin H, Mayo P, Allen M, Schnetz-Boutaud N, Crawford DC, Ritchie MD, Pendergrass SA. Detection of pleiotropy through a Phenome-wide association study (PheWAS) of epidemiologic data as part of the Environmental Architecture for Genes Linked to Environment (EAGLE) study. *PLoS Genet.* 2014; 10:e1004678. [PubMed: 25474351]
11. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics.* 2010; 26:1205–1210. [PubMed: 20335276]
12. Hebring SJ, Schrodi SJ, Ye Z, Zhou Z, Page D, Brilliant MH. A PheWAS approach in studying HLA-DRB1*1501. *Genes Immun.* 2013; 14:187–191. [PubMed: 23392276]
13. Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, Chisholm RL, Crosslin DR, Hebring SJ, Jarvik GP, Kullo IJ, Li R, Pathak J, Ritchie MD, Roden DM, Verma SS, Tromp G, Prato JD, Bush WS, Akey JM, Denny JC, Capra JA. The phenotypic legacy of admixture between modern humans and Neandertals. *Science.* 2016; 351:737–741. [PubMed: 26912863]
14. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Field JR, Pulley JM, Ramirez AH, Bowton E, Basford MA, Carrell DS, Peissig PL, Kho AN, Pacheco JA, Rasmussen LV, Crosslin DR, Crane PK, Pathak J, Bielinski SJ, Pendergrass SA, Xu H, Hindorff LA, Li R, Manolio TA, Chute CG, Chisholm RL, Larson EB, Jarvik GP, Brilliant MH, McCarty CA, Kullo IJ, Haines JL, Crawford DC, Masys DR, Roden DM. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nature biotechnology.* 2013; 31:1102–1111.

15. Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Duijn CM, Noble JA, Raj T, Gourraud PA, Stranger BE, Oksenberg J, Olsson T, Taylor BV, Sawcer S, Hafler DA, Carrington M, De Jager PL, de Bakker PI. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet.* 2013; 9:e1003926. [PubMed: 24278027]
16. Zipp F, Windemuth C, Pankow H, Dichgans J, Wienker T, Martin R, Muller C. Multiple sclerosis associated amino acids of polymorphic regions relevant for the HLA antigen binding are confined to HLA-DR2. *Hum Immunol.* 2000; 61:1021–1030. [PubMed: 11082515]
17. Madeleine MM, Brumback B, Cushing-Haugen KL, Schwartz SM, Daling JR, Smith AG, Nelson JL, Porter P, Shera KA, McDougall JK, Galloway DA. Human leukocyte antigen class II and cervical cancer risk: a population-based study. *J Infect Dis.* 2002; 186:1565–1574. [PubMed: 12447731]
18. Zoodmsa M, Nolte IM, Schipper M, Oosterom E, van der Steege G, de Vries EG, te Meerman GJ, van der Zee AG. Analysis of the entire HLA region in susceptibility for cervical cancer: a comprehensive study. *J Med Genet.* 2005; 42:e49. [PubMed: 16061555]
19. Mori M, Yamada R, Kobayashi K, Kawaida R, Yamamoto K. Ethnic differences in allele frequency of autoimmune-disease-associated SNPs. *J Hum Genet.* 2005; 50:264–266. [PubMed: 15883854]
20. Cruz-Tapias P, Rojas-Villarraga A, Maier-Moore S, Anaya JM. HLA Sjogren's syndrome susceptibility. A meta-analysis of worldwide studies. *Autoimmun Rev.* 2012; 11:281–287. [PubMed: 22001416]
21. Huang R, Yin J, Chen Y, Deng F, Chen J, Gao X, Liu Z, Yu X, Zheng J. The amino acid variation within the binding pocket 7 and 9 of HLA-DRB1 molecules are associated with primary Sjogren's syndrome. *J Autoimmun.* 2015
22. Verhagen FH, Brouwer AH, Kuiper JJ, Ossewaarde-van Norel J, Ten Dam-van Loon NH, de Boer JH. Potential Predictors of Poor Visual Outcome in Human Leukocyte Antigen-B27-Associated Uveitis. *Am J Ophthalmol.* 2016; 165:179–187. [PubMed: 26940162]
23. van der Sluis S, Posthuma D, V C. Dolan, TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genet.* 2013; 9:e1003235. [PubMed: 23359524]
24. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, Abecasis GR, Barrett JC, Behrens T, Cho J, De Jager PL, Elder JT, Graham RR, Gregersen P, Klareskog L, Siminovitch KA, van Heel DA, Wijmenga C, Worthington J, Todd JA, Hafler DA, Rich SS, Daly MJ. Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.* 2011; 7:e1002254. [PubMed: 21852963]
25. Smigoc Schweiger D, Mendez A, Kunilo Jamnik S, Bratanic N, Bratina N, Battelino T, Brecelj J, Vidan-Jeras B. High-risk genotypes HLA-DR3-DQ2/DR3-DQ2 and DR3-DQ2/DR4-DQ8 in co-occurrence of type 1 diabetes and celiac disease. *Autoimmunity.* 2016; 49:240–247. [PubMed: 27138053]
26. Hu X, Deutsch AJ, Lenz TL, Onengut-Gumuscu S, Han B, Chen WM, Howson JM, Todd JA, de Bakker PI, Rich SS, Raychaudhuri S. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet.* 2015; 47:898–905. [PubMed: 26168013]
27. Lambert AP, Gillespie KM, Thomson G, Cordell HJ, Todd JA, Gale EA, Bingley PJ. Absolute risk of childhood-onset type 1 diabetes defined by human leukocyte antigen class II genotype: a population-based study in the United Kingdom. *J Clin Endocrinol Metab.* 2004; 89:4037–4043. [PubMed: 15292346]
28. Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, Concannon P, Mychaleckyj JC, Todd JA, Bonella P, Fear AL, Lavant E, Louey A, Moonsamy PA. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. *Diabetes.* 2008; 57:1084–1092. [PubMed: 18252895]
29. Thomson G, Robinson WP, Kuhner MK, Joe S, MacDonald MJ, Gottschall JL, Barbosa J, Rich SS, Bertrams J, Baur MP, et al. Genetic heterogeneity, modes of inheritance, and risk estimates for a joint study of Caucasians with insulin-dependent diabetes mellitus. *Am J Hum Genet.* 1988; 43:799–816. [PubMed: 3057885]

30. Todd JA, Bell JI, McDevitt HO. HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature*. 1987; 329:599–604.
31. Horn GT, Bugawan TL, Long CM, Erlich HA. Allelic sequence variation of the HLA-DQ loci: relationship to serology and to insulin-dependent diabetes susceptibility. *Proc Natl Acad Sci U S A*. 1988; 85:6012–6016. [PubMed: 2842756]
32. Karnes JH, Shaffer CM, Bastarache L, Gaudieri S, Glazer AM, Steiner HE, Mosley JD, Mallal S, Denny JC, Phillips EJ, Roden DM. Comparison of HLA allelic imputation programs. *PLoS One*. 2017; 12:e0172444. [PubMed: 28207879]
33. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, Masys DR. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther*. 2008; 84:362–369. [PubMed: 18500243]
34. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, Li R, Masys DR, Ritchie MD, Roden DM, Struwing JP, Wolf WA. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics*. 2011; 4:13. [PubMed: 21269473]
35. Zuvich RL, Armstrong LL, Bielinski SJ, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT, de Andrade M, Doheny KF, Haines JL, Hayes MG, Jarvik GP, Jiang L, Kullo IJ, Li R, Ling H, Manolio TA, Matsumoto ME, McCarty CA, McDavid AN, Mirel DB, Olson LM, Paschall JE, Pugh EW, Rasmussen LV, Rasmussen-Torvik LJ, Turner SD, Wilke RA, Ritchie DM. Pitfalls of merging GWAS data: lessons learned in the eMERGE network and quality control procedures to maintain high data quality. *Genet Epidemiol*. 2011; 35:887–898. [PubMed: 22125226]
36. Guo Y, He J, Zhao S, Wu H, Zhong X, Sheng Q, Samuels DC, Shyr Y, Long J. Illumina human exome genotyping array clustering and quality control. *Nat Protoc*. 2014; 9:2643–2662. [PubMed: 25321409]
37. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155:945–959. [PubMed: 10835412]
38. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38:904–909. [PubMed: 16862161]
39. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–575. [PubMed: 17701901]
40. Bowton E, Field JR, Wang S, Schildcrout JS, Van Driest SL, Delaney JT, Cowan J, Weeke P, Mosley JD, Wells QS, Karnes JH, Shaffer C, Peterson JF, Denny JC, Roden DM, M J. Pulley, Biobanks and electronic medical records: enabling cost-effective research. *Sci Transl Med*. 2014; 6:234cm233.
41. Mallal S, Nolan D, Witt C, Masel G, Martin AM, Moore C, Sayer D, Castley A, Mamotte C, Maxwell D, James I, Christiansen FT. Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet*. 2002; 359:727–732. [PubMed: 11888582]
42. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J, Jagel-Guedes E, Rugina S, Kozyrev O, Cid JF, Hay P, Nolan D, Hughes S, Hughes A, Ryan S, Fitch N, Thorborn D, Benbow A. HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med*. 2008; 358:568–579. [PubMed: 18256392]
43. Israel RA. The International Classification of Disease. Two hundred years of development. *Public Health Rep*. 1978; 93:150–152. [PubMed: 635089]
44. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 1990; 9:811–818. [PubMed: 2218183]
45. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics*. 2014; 30:2375–2376. [PubMed: 24733291]
46. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet*. 2013; 14:483–495. [PubMed: 23752797]

47. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet.* 2011; 12:56–68. [PubMed: 21164525]
48. Middleton D, Menchaca L, Rood H, Komerofsky R. New allele frequency database: <http://www.allelefreqencies.net>. *Tissue Antigens.* 2003; 61:403–407. <http://www.allelefreqencies.net>. [PubMed: 12753660]
49. Reveille JD, Maganti RM. Subtypes of HLA-B27: history and implications in the pathogenesis of ankylosing spondylitis. *Adv Exp Med Biol.* 2009; 649:159–176. [PubMed: 19731628]
50. Kim K, Bang SY, Lee S, Lee HS, Shim SC, Kang YM, Suh CH, Sun C, Nath SK, Bae SC, Kim TH. An HLA-C amino-acid variant in addition to HLA-B*27 confers risk for ankylosing spondylitis in the Korean population. *Arthritis Res Ther.* 2015; 17:342. [PubMed: 26613595]
51. Han B, Diogo D, Eyre S, Kallberg H, Zhernakova A, Bowes J, Padyukov L, Okada Y, Gonzalez-Gay MA, Rantapaa-Dahlqvist S, Martin J, Huizinga TW, Plenge RM, Worthington J, Gregersen PK, Klareskog L, de Bakker PI, Raychaudhuri S. Fine mapping seronegative and seropositive rheumatoid arthritis to shared and distinct HLA alleles by adjusting for the effects of heterogeneity. *Am J Hum Genet.* 2014; 94:522–532. [PubMed: 24656864]
52. Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, Zhernakova A, Stahl E, Viatte S, McAllister K, Amos CI, Padyukov L, Toes RE, Huizinga TW, Wijmenga C, Trynka G, Franke L, Westra HJ, Alfredsson L, Hu X, Sandor C, de Bakker PI, Davila S, Khor CC, Heng KK, Andrews R, Edkins S, Hunt SE, Langford C, Symmons D, Concannon P, Onengut-Gumuscu S, Rich SS, Deloukas P, Gonzalez-Gay MA, Rodriguez-Rodriguez L, Arlsetig L, Martin J, Rantapaa-Dahlqvist S, Plenge RM, Raychaudhuri S, Klareskog L, Gregersen PK, Worthington J. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet.* 2012; 44:1336–1340. [PubMed: 23143596]
53. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, Alfredsson L, Padyukov L, Klareskog L, Worthington J, Siminovitch KA, Bae SC, Plenge RM, Gregersen PK, de Bakker PI. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet.* 2012; 44:291–296. [PubMed: 22286218]
54. Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. *Nat Rev Rheumatol.* 2013; 9:141–153. [PubMed: 23381558]
55. Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat Genet.* 2015
56. Sawcer S, Franklin RJ, Ban M. Multiple sclerosis genetics. *Lancet Neurol.* 2014; 13:700–709. [PubMed: 24852507]
57. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, Freeman C, Hunt SE, Edkins S, Gray E, Booth DR, Potter SC, Goris A, Band G, Oturai AB, Strange A, Saarela J, Bellenguez C, Fontaine B, Gillman M, Hemmer B, Gwilliam R, Zipp F, Jayakumar A, Martin R, Leslie S, Hawkins S, Giannoulatos E, D'Alfonso S, Blackburn H, Martinelli Boneschi F, Liddle J, Harbo HF, Perez ML, Spurkland A, Waller MJ, Mycko MP, Ricketts M, Comabella M, Hammond N, Kockum I, McCann OT, Ban M, Whittaker P, Kempainen A, Weston P, Hawkins C, Widaa S, Zajicek J, Dronov S, Robertson N, Bumpstead SJ, Barcellos LF, Ravindrarajah R, Abraham R, Alfredsson L, Ardlie K, Aubin C, Baker A, Baker K, Baranzini SE, Bergamaschi L, Bergamaschi R, Bernstein A, Berthele A, Boggild M, Bradfield JP, Brassat D, Broadley SA, Buck D. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature.* 2011; 476:214–219. [PubMed: 21833088]
58. Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, Vyse TJ, Rioux JD. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet.* 2008; 4:e1000024. [PubMed: 18437207]
59. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, Bakker SF, Bardella MT, Bhaw-Rosun L, Castillejo G, de la Concha EG, de Almeida RC, Dias KR, van Diemen CC, Dubois PC, Duerr RH, Edkins S, Franke L, Fransen K, Gutierrez J, Heap GA, Hrdlickova B, Hunt S, Plaza Izurieta L, Izzo V, Joosten LA, Langford C, Mazzilli MC, Mein CA, Midah V, Mitrovic M, Mora B, Morelli M, Nutland S, Nunez C, Onengut-Gumuscu S, Pearce K, Platteel M, Polanco I, Potter S, Ribes-Koninckx C, Ricano-Ponce I, Rich SS, Rybak A, Santiago JL, Senapati S, Sood A, Szajewska H, Troncone R, Varade J, Wallace C, Wolters VM, Zhernakova A, Thelma BK, Cukrowska B, Urcelay E, Bilbao JR, Mearin ML, Barisani D, Barrett JC, Plagnol V, Deloukas P, Wijmenga C, van Heel DA. Dense genotyping identifies and localizes multiple common and rare

variant association signals in celiac disease. *Nat Genet.* 2011; 43:1193–1201. [PubMed: 22057235]

60. Mustalahti K, Catassi C, Reunanen A, Fabiani E, Heier M, McMillan S, Murray L, Metzger MH, Gasparin M, Bravi E, MakiBravi M. The prevalence of celiac disease in Europe: results of a centralized, international mass screening project. *Ann Med.* 2010; 42:587–595. [PubMed: 21070098]
61. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, Wapenaar MC, Barnardo MC, Bethel G, Holmes GK, Feighery C, Jewell D, Kelleher D, Kumar P, Travis S, Walters JR, Sanders DS, Howdle P, Swift J, Playford RJ, McLaren WM, Mearin ML, Mulder CJ, McManus R, McGinnis R, Cardon LR, Deloukas P, Wijmenga C. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet.* 2007; 39:827–829. [PubMed: 17558408]
62. Tsoi LC, Spain SL, Knight J, Ellinghaus E, Stuart PE, Capon F, Ding J, Li Y, Tejasvi T, Gudjonsson JE, Kang HM, Allen MH, McManus R, Novelli G, Samuelsson L, Schalkwijk J, Stahle M, Burden AD, Smith CH, Cork MJ, Estivill X, Bowcock AM, Krueger GG, Weger W, Worthington J, Tazi-Ahnini R, Nestle FO, Hayday A, Hoffmann P, Winkelmann J, Wijmenga C, Langford C, Edkins S, Andrews R, Blackburn H, Strange A, Band G, Pearson RD, Vukcevic D, Spencer CC, Deloukas P, Mrowietz U, Schreiber S, Weidinger S, Koks S, Kingo K, Esko T, Metspalu A, Lim HW, Voorhees JJ, Weichenthal M, Wichmann HE, Chandran V, Rosen CF, Rahman P, Gladman DD, Griffiths CE, Reis A, Kere J, Nair RP, Franke A, Barker JN, Abecasis GR, Elder JT, Trembath RC. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat Genet.* 2012; 44:1341–1348. [PubMed: 23143594]
63. Liu JZ, Almarri MA, Gaffney DJ, Mells GF, Jostins L, Cordell HJ, Ducker SJ, Day DB, Heneghan MA, Neuberger JM, Donaldson PT, Bathgate AJ, Burroughs A, Davies MH, Jones DE, Alexander GJ, Barrett JC, Sandford RN, Anderson CA. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat Genet.* 2012; 44:1137–1141. [PubMed: 22961000]
64. Invernizzi P, Ransom M, Raychaudhuri S, Kosoy R, Lleo A, Shigeta R, Franke A, Bossa F, Amos CI, Gregersen PK, Siminovitch KA, Cusi D, de Bakker PI, Podda M, Gershwin ME, Seldin MF. Classical HLA-DRB1 and DPB1 alleles account for HLA associations with primary biliary cirrhosis. *Genes Immun.* 2012; 13:461–468. [PubMed: 22573116]
65. Bronstad I, Skiningsrud B, Bratland E, Lovas K, Undlien D, Sverre Husebye E, Wolff AS. CYP21A2 polymorphisms in patients with autoimmune Addison's disease, and linkage disequilibrium to HLA risk alleles. *Eur J Endocrinol.* 2014; 171:743–750. [PubMed: 25249698]
66. Fleischnick E, Awdeh ZL, Raum D, Granados J, Alosco SM, Crigler JF Jr, Gerald PS, Giles CM, Yunis EJ, Alper CA. Extended MHC haplotypes in 21-hydroxylase-deficiency congenital adrenal hyperplasia: shared genotypes in unrelated patients. *Lancet.* 1983; 1:152–156. [PubMed: 6130199]

Figure 1.

Association of all imputed four digit HLA alleles with PheWAS phenotypes by HLA locus. HLA alleles were imputed from HumanExome BeadChip using SNP2HLA. Odds ratios, 95% confidence intervals and p values were generated using a χ^2 distribution or Fisher's Exact Test. Strength of the association is plotted along the y axis as $-\log(p)$ and phenotypes are represented on the x axis categorized by PheWAS code group. AS indicates Ankylosing spondylitis; Acute bronchitis and bronchiolitis (Bronch.); Celiac disease (CeD); Cholangitis (Cholang.); Circulatory disease (CircDis); Degenerative and vascular disorders of ear (EarDis); Dermatomyositis (Derm); Diabetes mellitus (DM); Diabetic retinopathy (DR); Diffuse diseases of connective tissue (CTDis); Disorders of other cranial nerves (CNDIs); Graves' disease (GD); Hypothyroidism (HypoTh); Ileostomy status (Ileost.); Inflammatory bowel disease and other gastroenteritis and colitis (IBD); Juvenile rheumatoid arthritis (JuvRA); Lack of coordination (LOC); Loss of teeth or edentulism (Tooth loss); Lupus, localized and systemic (Lupus); Multiple sclerosis (MS); Other demyelinating diseases of central nervous system (CNSDis); Other endocrine disorders (Endocrine); Other inflammatory spondylopathies (Spondyl.); Polymyalgia Rheumatica (PR); Posttraumatic wound infection (Wound inf.); Primary biliary cirrhosis (PBC); Psoriasis and related disorders (Psoriasis*); Psoriasis vulgaris (PsoriasisV); Rheumatoid arthritis (RA); Rheumatoid arthritis and other inflammatory polyarthropathies (RA*); Secondary malignant neoplasm of digestive systems (2o neopl.); Sicca syndrome (SS); Systemic Lupus erythematosus (SLE); Type 1 diabetes (T1D); Type 1 diabetes with ketoacidosis (T1DK); Type 1 diabetes with neurological manifestations (T1DN); Type 1 diabetes with ophthalmic manifestations (T1DO); Type 1 diabetes with renal manifestations (T1DR); Thyrotoxicosis with or without goiter (Thyrotox.); Uveitis, noninfectious or not otherwise specified (Uveitis).

Figure 2.

Heatmap of four digit HLA allelic associations with $p < 1 \times 10^{-5}$. Red indicates an increased odds ratio and blue indicates a decreased odd ratio. Phenotypes and HLA alleles are clustered by similarity based on association results. Carats indicate groupings that were present in over 90% of 1,000 bootstrap simulations, calculated from sampling with replacement from the actual association results.

Figure 3.

Network diagram of 171 PheWAS analysis associations between four-digit HLA alleles and PheWAS codes at $p < 1 \times 10^{-5}$. Large circular nodes indicate HLA genes, small circular nodes represent four digit HLA alleles and colored squares represent PheWAS phenotypes. Lines indicate an association and distance between nodes indicates the strength of the association by p value. Dashed lines represent a protective effect ($OR < 1$) and solid lines indicate a risk effect ($OR > 1$). The shade of the lines indicated three different levels of statistical significances based on P values, with the darkest level for the most significant associations at $p < 1 \times 10^{-15}$, the intermediate level for the significances at p value between 1×10^{-8} and 1×10^{-15} and the light level for those at $p < 10^{-5}$. An interactive version is available at <http://pheWAScatalog.org>. AS indicates Ankylosing spondylitis; Celiac disease (CeD); Circulatory Disease (CirDis); Degenerative and vascular disorders of ear (EarDis); Dermatomyositis (Derm); Diabetes mellitus (DM); Diabetic retinopathy (DR); Diffuse diseases of connective

tissue (CTDis); Disorders of other cranial nerves (CNDIs); Other endocrine disorders (EndoDis); Graves' disease (GD); Hypothyroidism (HypoTh); Hypothyroidism NOS (HypoTh*); Ileostomy Status (Ileostomy); Insulin pump user (T1Dins); Juvenile rheumatoid arthritis (JuvRA); Lack of coordination (LOC); Multiple sclerosis (MS); Noninflammatory disorders of cervix (CervixDis); Other demyelinating diseases of central nervous system (CNSDis); Other inflammatory spondylopathies (Spondyl.); Polymyalgia Rheumatica (PR); Primary biliary cirrhosis (PBC); Psoriasis and related disorders (Psoriasis*); Psoriasis vulgaris (PsoriasisV); Rheumatoid arthritis (RA); Rheumatoid arthritis and other inflammatory polyarthropathies (RA*); Sicca syndrome (SS); Type 1 diabetes (T1D); Type 1 diabetes with ketoacidosis (T1DK); Type 1 diabetes with neurological manifestations (T1DN); Type 1 diabetes with ophthalmic manifestations (T1DO); Type 1 diabetes with renal manifestations (T1DR); Ulcerative colitis (UC)

Table 1

Number of alleles or variants observed by HLA gene in BioVU

HLA Gene	2 digit alleles (n) ¹	NS (n) ¹	4 digit HLA alleles (n) ¹	Concordance Rate (%) ²	Call Rate (%) ²
<i>HLA-A</i>	13	150	13	98.54	99.98
<i>HLA-B</i>	17	220	21	97.33	100
<i>HLA-C</i>	12	125	15	98.76	99.98
<i>HLA-DPA1</i>	2	12	3	-	-
<i>HLA-DPB1</i>	11	41	12	95.36	100
<i>HLA-DQA1</i>	5	71	7	-	-
<i>HLA-DQB1</i>	5	134	12	98.26	100
<i>HLA-DRB1</i>	12	235	16	93.19	99.93
Total	77	988	99	96.93	99.98

NS indicates a non-synonymous variant; HLA, human leukocyte antigen.

¹Number of two and four digit HLA alleles and amino acid changing variants with minor allele frequency greater than 0.01 imputed from HumanExome BeadChip using SNP2HLA by classical HLA gene.

²Concordance rate and call rate calculated for four digit HLA alleles comparing SNP2HLA imputation to direct HLA sequencing in a subset of 3,152 samples. *HLA-DPA1* and *HLA-DQA1* genes were not sequenced

Table 2

Most significant associations in two and four digit HLA allele PheWAS analysis

HLA Allele	Phenotype	Discovery Cohort (BioVU)				Replication (PMRP)				Meta-Analysis ⁴			
		MAF (%) ¹	Cases (n)	Controls (n)	OR ²	p value ²	Cases (n)	Controls (n)	OR	p value ³	OR	p value	
Four digit HLA alleles													
<i>DQB1*03:02</i>	Type 1 diabetes	11.3	453	20,695	4.42	4.6E-54	192	3,749	2.05	9.8E-06	3.62	1.0E-54	
<i>B*27:05</i>	Ankylosing spondylitis	4.0	55	23,403	34.70	1.6E-30	32	5,191	31.09	6.5E-18	33.30	9.4E-47	
<i>DQA1*03:01</i>	Type 1 diabetes	20.5	453	20,695	3.18	2.3E-31	192	3,749	1.96	8.2E-06	2.75	3.9E-34	
<i>B*27:05</i>	Other inflammatory spondylopathies	4.0	223	23,403	3.61	2.7E-15	181	5,191	5.02	9.1E-21	4.21	5.1E-34	
<i>DRB1*04:01</i>	Type 1 diabetes	11.3	453	20,695	2.82	1.2E-26	192	3,749	2.37	1.8E-07	2.69	1.3E-32	
<i>DRB1*04:01</i>	Rheumatoid arthritis	11.3	1,072	23,403	2.02	2.1E-25	269	5,191	1.80	5.8E-05	1.98	8.2E-29	
<i>DQB1*06:02</i>	Multiple sclerosis	13.6	285	21,830	3.48	4.4E-25	54	5,936	2.70	3.0E-04	3.34	8.8E-28	
<i>DQB1*02:01</i>	Type 1 diabetes	13.2	453	20,695	2.93	2.5E-29	192	3,749	1.47	0.020	2.47	1.2E-27	
<i>DRB1*03:01</i>	Type 1 diabetes	13.1	453	20,695	2.94	2.2E-29	192	3,749	1.46	0.023	2.47	1.3E-27	
<i>DRB1*15:01</i>	Multiple sclerosis	14.1	285	21,830	3.36	8.0E-24	54	5,936	2.53	6.9E-04	3.21	3.9E-26	
<i>C*06:02</i>	Psoriasis vulgaris	9.2	289	22,973	2.50	2.4E-13	340	3,526	2.55	6.9E-14	2.53	1.1E-25	
<i>C*06:02</i>	Psoriasis	9.2	369	22,973	2.20	3.7E-12	353	43,526	2.61	4.8E-15	2.38	1.9E-25	
Two digit HLA alleles													
<i>B*27</i>	Ankylosing spondylitis	4.2	55	23,403	32.43	1.1E-29	32	5,191	32.56	3.2E-17	32.48	4.0E-45	
<i>DQA1*03</i>	Type 1 diabetes	20.5	453	20,695	3.18	2.3E-31	192	3,749	1.96	8.2E-06	2.75	3.9E-34	
<i>DRB1*04</i>	Type 1 diabetes	19.0	453	20,695	3.07	1.5E-30	192	3,749	2.01	4.4E-06	2.71	6.0E-34	
<i>B*27</i>	Other inflammatory spondylopathies	4.2	223	23,403	3.51	6.1E-15	181	5,191	4.67	1.3E-19	4.02	1.2E-32	
<i>DQA1*01</i>	Type 1 diabetes	40.2	453	20,695	0.31	7.7E-32	192	3,749	0.70	0.020	0.40	1.0E-28	
<i>DRB1*03</i>	Type 1 diabetes	13.1	453	20,695	2.93	2.9E-29	192	3,749	1.46	0.023	2.46	1.7E-27	
<i>DRB1*15</i>	Multiple sclerosis	14.8	285	21,830	3.38	5.6E-24	54	5,936	2.40	1.4E-03	3.19	6.6E-26	
<i>C*06</i>	Psoriasis vulgaris	9.2	289	22,973	2.50	2.4E-13	340	3,526	2.55	6.9E-14	2.53	1.1E-25	
<i>C*06</i>	Psoriasis	9.2	369	22,973	2.20	3.7E-12	353	3,526	2.61	4.8E-15	2.38	1.9E-25	

PMRP indicates Personalized Medicine Research Project; HLA, human leukocyte antigen; MAF, minor allele frequency; OR, odds ratio.

¹) Overall minor allele frequency of allele in the discovery cohort (BioVU)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

²⁾Odds ratios and p values generated using logistic regression adjusted for age, gender, and two principal components and considered significant in Discovery cohort (BioVU) at FDR-adjusted $q < 0.10$.

³⁾P values considered significant in Replication cohort (PMRP) at $p < 0.05$.

⁴⁾Meta-analysis performed in a fixed effects model adjusted for age, gender and two principal components.

Table 3

Major HLA genetic determinant specific autoimmune/inflammatory diseases in the BioVU PheWAS analysis

Disease/ Phenotype	Concordance rate MZ twins	Previous Literature		PheWAS Analysis		
		HLA allele/haplotype	Relative Risk (range)	Cases / Controls (n)	OR (95%CI)	p value
Ankylosing spondylitis	50–75	<i>B*27</i>	12	55/23,403	32.43 (17.75–59.25)	1.12×10 ⁻²⁹
Celiac disease	60–75	<i>DQB1*02:01-DQA1*05:01</i>	7	77/20,529	4.88 (3.07–7.75)	1.99×10 ⁻¹¹
Lupus	-	<i>DRB1*03:01</i>	2	448/24,065	1.71 (1.40–2.08)	1.15×10 ⁻⁷
Multiple Sclerosis	5.8–30.8	<i>DRB1*15:01</i>	3	285/21,830	3.36 (2.65–4.25)	8.03×10 ⁻²⁴
Primary Biliary Cirrhosis	77	<i>DRB1*08:01</i>	2.4–3.3	230/22,347	3.32 (2.24–4.93)	2.73×10 ⁻⁹
Psoriasis	35–64	<i>C*06</i>	5.7–20.8	289/22,973	2.50 (1.96–3.20)	2.39×10 ⁻¹³
RA	0–21	<i>DRB1*04:01</i>	4	1,072/23,403	2.02 (1.77–2.30)	2.14×10 ⁻²⁵
Type 1 diabetes	13–47.4	<i>DQB1*03:02-DRB1*04:01</i>	4.1–10.7	453/20,695	4.79 (3.92–5.83)	4.70×10 ⁻⁵⁴
Sicca syndrome	-	<i>DQB1*02:01</i>	10	151/23,906	2.51 (1.82–3.47)	2.60×10 ⁻⁸

Major HLA genetic determinants for specific autoimmune/inflammatory diseases and their association results from the PheWAS analysis. Odds ratios, 95% confidence intervals and p values were generated using logistic regression adjusted for age, gender, and two principal components. MZ indicates monozygotic; HLA, human leukocyte antigen; OR, odds ratio; 95%CI, 95% confidence interval.

Table 4

Previously unidentified associations with HLA variation

Phenotype	HLA Gene	Position (aa) or Allele	MAF (%) ¹	Discovery Cohort (BioVU)			Replication (PMRP)			Meta-Analysis			
				Cases (n)	Controls (n)	OR ²	p value ²	Cases (n)	OR	p value ³	OR ⁴	P value ⁴	
HLA Alleles													
Ileostomy status	<i>DRB1</i>	<i>DRB1*01:03</i>	1.2	229	20,452	3.24	8.3E-06	41	2,705	6.04	1.1E-05	3.63	5.8E-08
Thyrototoxicosis with or without goiter	<i>DQB1</i>	<i>DQB1*02:01</i>	13.1	470	22,637	1.56	9.0E-06	363	5,314	1.40	7.6E-03	1.49	2.8E-07
Other upper respiratory disease	<i>DQA1</i>	<i>DQA1*01:01</i>	14.3	886	20,162	1.34	6.7E-05	1,716	1,577	1.21	0.017	1.28	5.2E-06
Kidney replaced by transplant	<i>DQB1</i>	<i>DQB1*03:02</i>	11.3	864	21,329	1.37	5.8E-05	22	5,544	2.46	0.043	1.40	1.7E-05
Non-synonymous variants													
Thyrototoxicosis with or without goiter	<i>DRB1</i>	77 (-), 74 (r)	26.2	470	22,637	1.56	7.9E-06	363	5,314	1.40	7.6E-03	1.49	3.3E-07
Hemorrhage of gastrointestinal tract	<i>A</i>	73 (t)	3.8	774	22,928	1.58	1.1E-04	536	4,517	1.61	3.2E-03	1.59	1.1E-06
Nodular lymphoma	<i>DRB1</i>	13 (s/r)	49.5	227	26,093	0.57	4.8E-05	27	7,936	0.35	7.6E-03	0.54	2.3E-06
Atherosclerosis of the extremities	<i>C</i>	66 (-)	18.5	618	23,122	1.37	2.2E-04	298	5,747	1.34	0.029	1.36	1.7E-05
Cervical cancer & dysplasia	<i>DRB1</i>	37 (y)	31.4	365	11,648	1.51	2.1E-04	574	2,278	1.38	8.5E-04	1.43	7.2E-07
Other upper respiratory disease	<i>DQB1</i>	224 (q)	15.6	886	20,162	1.33	9.8E-05	1,716	1,577	1.18	0.032	1.26	1.7E-05
Optic neuritis/neuropathy ⁵	<i>DQB1</i>	9 (y/l)	16.2	131	24,855	2.20	7.4E-06	296	2,436	1.38	0.013	1.63	3.2E-06
Malignant neoplasm of female breast	<i>DRB1</i>	37 (y/l)	32.7	1,400	11,655	0.80	1.1E-04	345	2,232	0.74	0.013	0.79	5.2E-06
Kidney replaced by transplant	<i>DRB1</i>	71 (k)	25.3	864	21,329	1.32	8.8E-05	22	5,544	2.70	0.025	1.34	2.4E-05
Cancer of brain & nervous system	<i>DRB1</i>	74 (r/e)	18.5	599	27,004	0.71	1.8E-04	22	7,761	0.22	0.040	0.69	7.2E-05

PMRP indicates Personalized Medicine Research Project; aa, amino acid; HLA, human leukocyte antigen; MAF, minor allele frequency; OR, odds ratio.

¹) Overall minor allele frequency of allele in the discovery cohort (BioVU)²) Odds ratios and p values generated using logistic regression adjusted for age, gender, and two principal components and considered significant in Discovery cohort (BioVU) at FDR-adjusted q<0.10.³) P values considered significant in Replication cohort (PMRP) at p<0.05.⁴) Meta-analysis performed in a fixed effects model adjusted for age, gender, and two principal components.⁵) Conditional analyses suggest that the optic neuritis phenotype is more strongly associated with multiple sclerosis than with HLA variation

Table 5

Associations with type 1 diabetes and psoriasis and their disease complications

HLA Variant	Phenotype	Discovery Cohort (BioVU)				Replication (PMRP)				Meta-Analysis	
		Cases (n)	Controls (n)	OR ¹	p value ¹	Cases (n)	Controls (n)	OR	p value ²	OR ³	P value ³
<i>DRB1*04</i>	T1D	453	20,695	3.07	1.5E-30	192	3,749	2.01	4.0E-06	2.71	6.0E-34
	T1D w/ketoacidosis	85	20,695	5.52	1.4E-11	14	3,749	4.70	5.7E-03	5.15	2.9E-13
	T1D w/ neurological manif.	73	20,695	4.83	1.2E-09	39	3,749	4.21	1.5E-05	4.59	8.7E-14
	T1D w/ ophthalmic manif.	75	20,695	5.02	3.8E-10	58	3,749	4.21	1.5E-07	4.62	3.4E-16
	T1D w/ renal manif.	76	20,695	4.51	1.7E-09	24	3,749	9.98	5.1E-06	5.28	1.2E-13
<i>DQB1*03:02</i>	T1D	453	20,695	4.42	4.6E-54	192	3,749	2.05	9.8E-06	3.62	1.0E-54
	T1D w/ketoacidosis	85	20,695	7.09	1.0E-17	14	3,749	3.29	0.028	6.32	2.1E-18
	T1D w/ neurological manif.	73	20,695	6.26	3.0E-14	39	3,749	4.19	9.8E-06	5.42	2.5E-18
	T1D w/ ophthalmic manif.	75	20,695	7.33	2.7E-16	58	3,749	4.01	2.0E-07	5.57	1.3E-21
	T1D w/ renal manif.	76	20,695	6.66	1.9E-15	24	3,749	8.77	6.1E-07	7.10	7.4E-21
<i>DRB1</i> position 67 (i)	T1D	453	20,695	0.27	1.8E-38	192	3,749	0.54	3.6E-05	0.34	3.3E-39
	T1D w/ketoacidosis	85	20,695	0.18	4.7E-12	14	3,749	0.17	2.9E-03	0.18	5.2E-14
	T1D w/ neurological manif.	73	20,695	0.32	2.4E-06	39	3,749	0.23	1.6E-05	0.29	2.1E-10
	T1D w/ ophthalmic manif.	75	20,695	0.16	2.0E-11	58	3,749	0.23	1.1E-07	0.19	1.8E-17
	T1D w/ renal manif.	76	20,695	0.22	2.1E-09	24	3,749	0.17	1.1E-04	0.21	1.1E-12
B position 97 (n/v)	Psoriasis	369	22,973	1.58	3.0E-04	353	3,526	2.18	2.2E-09	1.85	1.3E-11
B position 70 (q/n)	Psoriatic arthropathy	131	22,973	2.03	3.7E-04	44	3,526	3.34	1.5E-04	2.33	4.9E-07
	Psoriasis	369	22,973	1.58	1.9E-04	353	3,526	1.92	1.9E-06	1.73	4.8E-10
C position 156 (l)	Psoriatic arthropathy	131	22,973	1.92	8.3E-04	44	3,526	2.84	1.5E-03	2.14	4.8E-07
	Psoriasis	369	22,973	0.59	9.0E-07	353	3,526	0.67	6.7E-04	0.62	3.3E-09
	Psoriatic arthropathy	131	22,973	0.49	6.0E-05	44	3,526	0.36	8.9E-04	0.45	2.7E-07

PMRP indicates Personalized Medicine Research Project; aa, amino acid; HLA, human leukocyte antigen; MAF, minor allele frequency; manif, manifestations; OR, odds ratio; T1D, type 1 diabetes.

^{1/}Odds ratios and p values generated using logistic regression adjusted for age, gender, and two principal components and considered significant in Discovery cohort (BioVU) at FDR-adjusted q<0.10.^{2/}P values considered significant in Replication cohort (PMRP) at p<0.05.

Meta-analysis performed in a fixed effects model adjusted for age, gender and two principal components.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript