GWAS Identifies Classical HLA Alleles Associated with Susceptibility to Infectious Diseases

Chao Tian[1], Joanna L. Mountain[1], Nicholas Eriksson[1], Joyce Y. Tung[1], Amy K. Kiefer[1], David A. Hinds[1]
[1]23andMe, Inc, Mountain View, CA;

Introduction

Human Leukocyte Antigen (HLA) system, the major histocompatibility complex(MHC) in humans, has long been known to play an important role in susceptibility and resistance to many infectious diseases and responsiveness to pathogens or vaccines. Most studies by far have tested only candidate loci in small samples, and the associations between HLA alleles and many infectious diseases are not well studied. We conducted genome wide association studies on 23 infectious diseases. We also imputed classic HLA alleles using HIBAG[1] over all 23andMe customers to explore the relationship between individual HLA alleles and susceptibility to infectious diseases. We identified single nucleotide polymorphisms in the MHC region and HLA alleles that are significantly associated with chicken pox, cold sores, childhood ear infection, plantar warts, mumps, pneumonia, positive TB test, scarlet fever, shingles and tonsillectomy. For most of these diseases, the most significantly associated SNP is highly correlated with the top association HLA allele (r2>0.95). The reconciliation of SNP and HLA allele associations implicate the function of specific HLA peptides as the major factors modulating infectious diseases. We did not detect associations between HLA loci and bladder infection, hepatitis, chronic sinus infection, colds last year, gingivitis, strep throat, urinary tract infection, measles, intestinal parasite, myringotomy, mononucleosis, rubella or rheumatic fever, some of which are probably due to small cohort sizes.

Methods

We conducted genome-wide association study (GWAs) on 23 infectious diseases. All participants were drawn from the customer base of 23andMe, Inc., a personal genetics company. All participants were of primarily European ancestry, and no pair was more closely related than at the level of first cousins. To explore the relationship between individual HLA alleles and susceptibility to infectious diseases we used HIBAG[1], a statistical method combining a large database of individuals with known HLA alleles and SNP variation within the MHC region, to impute HLA alleles at key class I and class II loci over all 23andMe customers of European ancestry. We identified the primary association signal (lowest Pvalue) for each disease and performed iterative conditional regression to identify other independent signals in MHC region. All analysis controlled for sex and five principal components of genetic ancestry. P-values were calculated using likelihood ratio test.

Phenotypic Data

Has HLA hits					
	Chicken pox	Cold sores	Childhood ear infection	Plantar warts	Mumps
Cases	81547	24004	27187	qtl > 14000	29688
Control	14869	61471	45328	29078	52776
	Pneumonia	Positive TB	Scarlet fever	Shingles	Tonsillectomy
Cases	20789	6279	5029	11538	39011
Control	65002	81428	90763	88667	79225

No NO HLA hits						
	Myringotomy	Bladder infection	Hepatitis	Chronic sinus infection	Colds last year	Gingivitis
Cases	3938	25553	3094	3091	qtl >25000	14786
Control	82327	61062	84640	48586	9422	38266
	Intestinal parasites	Mononucleosis	Strep throat	Uti frequency	Measles	Rheumatic fever	Rubella
Cases	7732	10252	qtl > 25000	qtl > 18000	36399	1071	11421
Control	46291	41651	12170	29486	46177	85107	69329

Table 1 Our discovery cohorts drawn from the more than 350,000 genotyped customers who reported via a web-based questionnaire whether they had been diagnosed with certain kind of infectious diseases. Sex and ancestry were determined based on genetic data; for the GWAS, participants were of European ancestry.

Genotyping

Participants were genotyped for 586,916 to 1,008,949 on one of three illumine –based beadchips. An additional 7,356,559 imputed SNPs were included in the analysis.

Results

We identified single nucleotide polymorphisms in the MHC region that are significantly (p<5x10-8) associated with 10 of the 23 infectious diseases. We also identified the either a susceptibility or resistance of specific HLA alleles (p<1x10-5) to those infectious diseases(Table 2, Figure1). The HLA alleles were imputed in all samples using HIBAG. Similar to the SNP imputation, we measured the imputation quality using the ratio of the empirically observed variance of the allele dosage to the expected binomial variance p(1-p) at Hardy-Weinberg equilibrium, where p is the observed allele frequency. In our data, all the top association HLA has high imputation quality(Table 2).
To determine whether the association detected for the top SNP and the top HLA allele is related or independent, we performed reciprocal regression tests and also examined the correlation (r2) between the SNP and the HLA allele and generally found consistency between them. The top SNP association is stronger than the top HLA association except scarlet fever (Table 2). For scarlet fever and pneumonia, their top SNP and top HLA allele signals are localized in different genes (Figure 1). Next we conditioned each disease dataset on the most significant association to identify other signals independent of the top association. For many of the diseases, there was evidence of multiple additional association signals within MHC region (Figure 1).

Disease	Type	ID	Position	OR	Pvalue	Reciprocal Pvalue	R2 in cohort	Imputation quality
Chicken Pox	SNP	rs471120	29987184	1.10	6.50E-10	2.84E-03	0.90	
	HLA	A*02:01		1.09	2.76E-08	3.09E-01		0.98
Cold sores	SNP	rs4360170	31430359	0.86	5.57E-12	2.52E-05	0.45	
	HLA	B*18:01		0.86	3.59E-08	2.00E-01		0.91
Childhood ear infection	SNP	rs9271112	32576717	0.92	9.43E-08	3.26E-04	0.40	
	HLA	DQB1*06:02		0.93	1.06E-05	2.65E-01		0.98
Plantar warts	SNP	rs660895	32577380	0.064(effect)	5.71E-22	6.67E-08	0.50	
	HLA	DQA1*03:01		0.078(effect)	3.98E-17	3.62E-02		0.87
Mumps	SNP	rs115443396	29820300	0.43	2.98E-15	1.30E-01	0.85	
	HLA	A*02:05		0.58	7.55E-16	7.48E-02		0.95
Pneumonia	SNP	rs972254	32603098	0.92	4.21E-08	2.71E-06	0.06	
	HLA	B*08:01		0.92	1.59E-05	5.72E-03		1.00
Positive TB	SNP	rs9268882	32431623	0.75	6.14E-29	1.41E-14	0.23	
	HLA	DQA1*01:02		0.75	3.19E-21	4.49E-06		0.97
Scarlet fever	SNP	rs204990	32161430	0.85	3.87E-10	3.16E-07	0.04	
	HLA	DQB1*03:01		0.83	2.36E-12	4.67E-09		0.97
Shingles	SNP	rs2523591	31326960	1.14	5.22E-19	1.44E-10	0.14	
	HLA	B*44:02		0.81	9.99E-17	1.13E-06		0.98
Tonsillectomy	SNP	rs115846244	31430060	1.21	6.70E-14	3.04E-02	0.85	
	HLA	B*57:01		1.22	1.19E-15	2.98E-01		0.98

Table 2. Top association SNP and HLA signals.

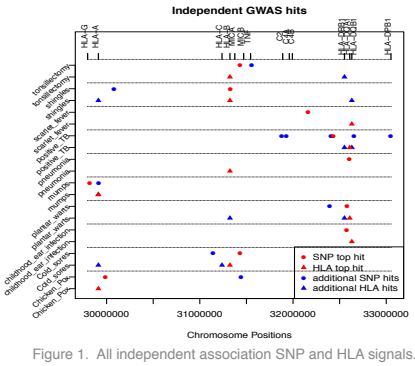

Figure 1. All independent association SNP and HLA signals.

We also detected signals outside MHC region for Childhood ear infection, mumps, scarlet fever, tonsillectomy and myringotomy (Figure 2). Variants in FUT2 are highly associated with susceptibility to childhood ear infection and mumps. TBX1 locus is highly associated with susceptibility to childhood ear infection, myringotomy and tonsillectomy. The shared effects highlighted potentially shared functional pathways.

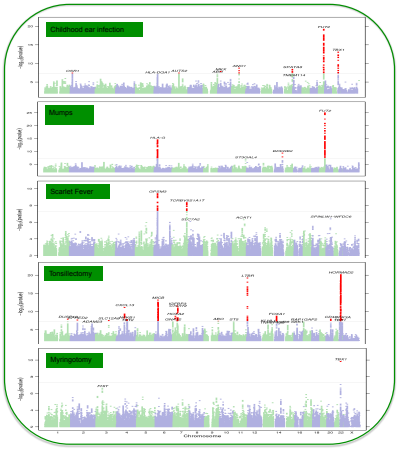

Figure 2. GWAS signals outside MHC region.

Conclusion

- The conciliation of current high density SNP analysis and association test with imputed HLA alleles confirms that the predominant signals in many infectious diseases map to specific HLA regions and certain HLA peptides play important roles in modulating infectious diseases.
- There seem to be multiple independent HLA genes involves in the susceptibility to tonsillectomy, shingles, scarlet fever, positive TB test, pneumonia, plantar warts, cold sores and chicken pox by conditional analysis.
- The top SNP associations seem all stronger (lower Pvalue) than HLA allele associations except scarlet fever. The top associated SNP might be a proxy for multiple functional HLA alleles which contribute combinedly in viral/bacterial peptide interaction.
- Shingles and chicken pox are related infectious diseases, the result from the current study, strongly implicated a shared association with HLA-A*02:01 allele.
- A more detail analysis of specific amino acids changes in HLA peptide of the association HLA alleles[2] will further help to define the host genetic effects on the outcome of the corresponding infectious diseases.

Acknowledgments

We thank the customers of 23andMe for participating in this research and the employees of 23andMe for their contributions to this work.

References

1. Zheng X, Shen J, Cox C, Wakefield J, Ehm M, Nelson M, Weir BS. HIBAG – HLA Genotype Imputation with Attribute Bagging. Pharmacogenomics Journal (2013).
2. International HIV Controllers Study, Pereyra F and Zhao M. et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science. 2010 Dec 10;330(6010):1551-7.