Association of HLA-DRB1*09:01 with tIgE levels among African ancestry individuals with asthma

Nicolas Vince, Sophie Limou, Michelle Daya, Wataru Morii, Nicholas Rafaels, Estelle Geffard, Venceslas Douillard, Alexandre Walencik, Meher Preethi Boorgula, Sameer Chavan, et al.

Association of *HLA-DRB1*09:01* with tIgE levels among African ancestry individuals with asthma

Nicolas Vince, PhD 1, Sophie Limou, PhD 1, Michelle Daya, PhD 2, Wataru Morii, PhD 3, Nicholas Rafaels, MS 2, Estelle Geffard, MS 1, Venceslas Douillard, MS 1, Alexandre Walencik, PharmD 1, Meher Preethi Boorgula, MS 2, Sameer Chavan, MS 2, Candelaria Vergara, MD, PhD 4, Victor E. Ortega, MD, PhD 5, James G. Wilson, MD 6, Leslie A. Lange, PhD 2, Harold Watson, MD 7, Dan L. Nicolae, PhD 8, Deborah A. Meyers, PhD 9, Nadia N. Hansel, MD, MPH 4, Jean G. Ford, MD 10, Mezbah U. Faruque, MD, PhD 11, Eugene R. Bleecker, MD 9, Monica Campbell, MS 2, Terri H. Beaty, PhD 12, Ingo Ruczinski, PhD 13, Rasika A. Mathias, ScD 4,12, Margaret A. Taub, PhD 13, Carole Ober, PhD 14, Emiko Noguchi, MD, PhD 3, Kathleen C. Barnes, PhD 2 on behalf of CAAPA, Dara Torgerson, PhD 15, Pierre-Antoine Gourraud, PhD, MPH 1

1. Nantes Université, Centrale Nantes, CHU Nantes, Inserm, Centre de Recherche en Transplantation et Immunologie, UMR 1064, ITUN, F-44000 Nantes, France
2. Department of Medicine, University of Colorado Denver, Aurora, CO 80045 USA
3. Department of Medical Genetics, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan
4. Department of Medicine, Johns Hopkins University, Baltimore, MD 21224 USA
5. Department of Internal Medicine, Section on Pulmonary, Critical Care, Allergy and Immunologic Diseases, Center for Precision Medicine, Wake Forest School of Medicine, Winston-Salem, NC 27157 USA
6. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216 USA
7. Faculty of Medical Sciences Cave Hill Campus, The University of the West Indies, Bridgetown BB11000 Barbados
8. Department of Medicine, University of Chicago, Chicago, IL 60637 USA
9. Department of Medicine, University of Arizona College of Medicine, Tucson, AZ 85724 USA
10. Department of Medicine, Einstein Medical Center, Philadelphia, PA 19141 USA
11. National Human Genome Center, Howard University College of Medicine, Washington, DC 20059 USA

12. Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205 USA
13. Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205 USA
14. Department of Human Genetics, University of Chicago, Chicago, IL 60637 USA
15. McGill University and Genome Quebec Innovation Centre, Montreal, Canada

Corresponding author: Pierre-Antoine Gourraud, PhD, MPH
#ATIP-Avenir Team 5 "Translational Immunogenomics of Transplantation and Autoimmunity"
ITUN - CRTI - UMR Inserm 1064 - CHU de Nantes
30 bvd Jean Monnet - 2eme étage 44093 NANTES CEDEX 1, FRANCE
Email: pierre-antoine.gourraud@univ-nantes.fr Tel: +33 (0)2 447 68 234

The co-authors have disclosed no conflict of interest.

Abstract

Background: Asthma is a complex chronic inflammatory disease of the airways. Association studies between *HLA* and asthma were first reported in the 1970's, and yet, the precise role of *HLA* alleles in asthma is not fully understood. Numerous genome-wide association studies were recently conducted on asthma, but were always limited to simple genetic markers (SNPs) and not complex *HLA* gene polymorphisms (alleles/haplotypes), therefore not capturing the biological relevance of this complex locus for asthma pathogenesis.

Objective: To run the first *HLA*-centric association study with asthma and specific asthma-related phenotypes in a large cohort of African ancestry individuals.

Methods: We collected high-density genomics data for the CAAPA participants (Consortium on Asthma among African-ancestry Populations in the Americas, N=4,993). Using computer-intensive machine-learning attribute bagging methods to infer *HLA* alleles, and Easy-HLA to infer *HLA* 5-gene haplotypes, we conducted a high-throughput HLA-centric association study of asthma susceptibility and total serum IgE levels (tIgE) in subjects with and without asthma.

Results: Among the 1,607 individuals with asthma, 972 had available tIgE levels with a mean tIgE level of 198.7 $IU.ml^{-1}$. We could not identify any association with asthma susceptibility. However, we showed that *HLA-DRB1*09:01* was associated with increased tIgE levels ($P=8.5x10^{-4}$, weighted effect size 0.51 [0.15-0.87]).

Conclusions: We identified for the first time an *HLA* allele associated with tIgE levels in African ancestry individuals with asthma. Our report emphasizes that by leveraging powerful computational machine-learning methods, specific/extreme phenotypes, and population diversity, we can explore *HLA* gene polymorphisms in depth and reveal the full extent of complex disease associations.

Key messages:

- Using a machine learning-based method and a matching reference population of African ancestry, we were able to impute HLA alleles from SNP data with good accuracy in the CAAPA dataset. The reference population was instrumental to reach high prediction accuracy.
- The *HLA-DRB1*09:01* allele significantly associated with an increase of total serum IgE levels in asthmatics of African ancestry.

Keywords: Asthma, HLA, tIgE levels, Atopy, CAAPA, Imputation, Admixture.

Abbreviations:

ADPC: African Diaspora Power Chip

CAAPA: Consortium on Asthma among African-ancestry Populations in the Americas

CPU: Central Processing Unit

GWAS: Genome-Wide Association Study

HLA: Human Leukocyte Antigen

IRB: Institutional Review Boards

IQR: Inter Quartile Range

NMDP: National Marrow Donor Program

MHC: Major Histocompatibility Complex

QQplot: Quantile Quantile plot

SNP: Single Nucleotide Polymorphism

tIgE: total serum Immunoglobulin E levels

WGS: whole-genome sequencing

Capsule summary

By leveraging powerful computational machine-learning methods, specific/extreme phenotypes, and population diversity, we identified *HLA-DRB1*09:01* associated with total serum IgE levels in African ancestry individuals with asthma revealing key aspects of complex disease associations.

Introduction

Asthma is a complex chronic inflammatory disease of the airways presenting a large variety of phenotypes which can be divided between Th2-high (e.g. allergic asthma) and non-Th2 (very late-onset) [1,2]. The Human Leukocyte Antigen (HLA) molecules play a central role in the initiation and regulation of innate and acquired immune responses [3,4], and, as such, have been extensively studied for its potential links with numerous diseases. According to the GWAS catalog, discoveries inside the HLA genomic region account for 22% of all diseases and traits [5], emphasizing the crucial role played by HLA in a large number of immune-related pathologies [6]. The study of potential associations between *HLA* and asthma was first reported in the 1970's [7,8], and yet, the precise role of *HLA* alleles in asthma is not fully understood [1]. Difficulties for replication of associations include variability in asthma etiologies and biological characteristics [2], small sample sizes [1,9], population heterogeneity, the cost of *HLA* molecular typing, and the challenge of interpreting associations due to high allelic and structural variations and complex linkage disequilibrium patterns [10]. Overall, class II *HLA* alleles appear to be involved in late-onset and allergic asthma in European and Hispanic populations [1,11–15].

With the generalization of genomic studies, an important number of genome-wide association studies (GWAS) were conducted on asthma in the past few years [11,12,16–26], and notably reported several associations for SNPs within the *MHC* (Major Histocompatibility Complex) region that includes the *HLA* genes. In particular, SNPs in class II *HLA* genes were associated with asthma in European and Japanese populations [11,12,20,21,23–25]. Variations in *HLA-DQ* appear to be the main asthma contributors identified through GWAS, as signals near this gene were found in European American, African American, and Latino populations; but with different SNPs identified in each ethnic group [22]. In addition, genetic ancestry at the *HLA* locus has been associated with both asthma and allergen-specific IgE levels in Latinos [17,21], and other *HLA* variations also seem to play a role (e.g. *HLA-G* [27] and *HLA-DRA* [18]). However, all these studies only focused on investigations at the SNP level, which do not convey HLA biological complexity and functional relevance. To fully capture the complex information related to antigen presentation and interactions with other immune-related molecules in the context of asthma, it is necessary to examine *HLA* alleles, defined by *HLA* gene sequence (www.ebi.ac.uk/ipd/imgt/hla/) [10] or SNP haplotypes, and *HLA* 5-genes haplotypes (*HLA-A~HLA-B~HLA-C~HLA-DQB1~HLA-DRB1*).

Due to the high degree of diversity in the *HLA* region, *HLA* alleles exhibit relatively low frequencies and only the top 10 most frequent 5-genes haplotypes are in the 1% frequency range [28]; therefore, a very large sample size or a large effect size is necessary for adequate statistical power to detect an association between an *HLA* allele and asthma. However, collecting a very large asthmatic cohort often implies a mix of diverse phenotypes, which may introduce heterogeneity of genetic causes and lead to false negative signals. Focusing on specific or extreme phenotypes may then contribute to find specific associations [29,30], even though the working sample size is reduced.

Here, we leveraged genomic data generated by CAAPA, the Consortium on Asthma among African-ancestry Populations in the Americas, to explore for the first time the role of *HLA* alleles on asthma and specific asthma-related phenotypes in a large cohort of African ancestry individuals (N = 4,993) [31]. CAAPA recently published a GWAS that found strong evidence for association at four previously reported asthma loci (whose discovery was driven largely by non-African populations), and also identified two novel loci that may be specific to asthma risk in African ancestry populations [31]. Here, our aims were to deliver the first high-throughput HLA-centric study of asthma outcomes by performing a case/control analysis of asthma susceptibility and a tIgE levels quantitative analysis in subjects with asthma to explore atopy. Our hypothesis is that *HLA* alleles are associated with asthma phenotypes. We also hypothesize that a classical GWAS design cannot identify these associations as GWAS focus on simple genetics markers (SNPs) that do not fully recapitulate the complexity of the HLA genomic region.

Methods

CAAPA

The CAAPA multi-centre participants were previously described [31]. A total of 1,607 asthmatic cases and 3,365 controls of African ancestry (USA and Barbados) were recruited (see table S1, below and in Daya et al. [31] for a full description). Briefly, 8 of 17 of the CAAPA investigators contributed data to this study. The distributions of age, sex and age of asthma onset for each cohort are summarized in Table S1. Participants in these studies were unrelated except for BAGS and HUFS which included families; however, we only included the founder's individuals to perform our statistical analyses on unrelated subjects. Childhood onset asthma is defined as having been diagnosed with asthma before 16 years of age. Studentized residuals of log10 transformed total serum IgE (tIgE, adjusted for age and sex as previously described [32]) was available from 4/8 CAAPA studies (see table S1 for mean and standard error of tIgE before transformation and figure S8 for population distribution). We defined atopy as tIgE level > 80 KU/L. Previous work has shown that the genetic structure of the Barbados population is similar to that of African Americans, with subjects from Barbados having on average higher proportions of African ancestry compared to African Americans [19,31]. Asthma cases were defined by reported and documented histories of current or past physician-diagnosed asthma, while controls reported a negative history of asthma. All participants provided written informed consent. This study is an initiative of CAAPA, which was funded by the NIH (R01 HL104608). NIH guidelines for conducting human genetic research were followed. The Institutional Review Boards (IRB) of Johns Hopkins University (for the GRAAD, BAGS and BRIDGE cohorts), Howard University (for the GRAAD and HUFS cohorts), Wake Forest University (for the SARP cohort), the University of Chicago (for the CAG cohort), University of the West Indies, Cave Hill Campus, Barbados (for the BAGS cohort), and University of Mississippi Medical Center (for the JHS and ARIC cohorts) all reviewed and approved this study.

Genotyping and SNP imputation

Details of genotyping and SNP imputation can be found in the GWAS paper [31]. Briefly, each CAAPA study was separately genotyped on a variety of GWAS chips, as well as genotyped on an African-ancestry specific chip (African Diaspora Power Chip or ADPC) [33]. All these genomic datasets were imputed using the CAAPA whole-genome sequence reference panel on the Michigan imputation server (https://imputationserver.sph.umich.edu) [34].

HLA allele SNP-based imputation and haplotyping

The HIBAG R package is designed to impute *HLA* alleles from SNP genotypes using a reference panel (or bagging set) built with an attribute bagging machine-learning technique [35,36]. Appropriate reference panels matching the population of interest are crucial to obtain accurate imputation. Several reference panels are available publicly (http://www.biostat.washington.edu/~bsweir/HIBAG/) but are mostly suitable for European ancestry imputation. Due to the complexity and diversity of African ancestry genomes, large reference panels are necessary and are yet to be collected; the currently available panel being derived from only ~150 African-ancestry individuals. To maximize imputation quality, we created our own African ancestry reference panel from a subset of 917 CAAPA individuals from whom we had high-resolution genotyped *HLA* alleles (second field or four-digit), and *MHC* SNP genotypes (accessed from dbGAP, phs001123.v1.p1). The *HLA* alleles were called with the Omixon software (Budapest, Hungary) from whole-genome sequencing (WGS) data, and we selected 29,970 *MHC* SNPs overlapping between the WGS data and the GWAS data. From this large HLA+SNP dataset, we used HIBAG to train statistical models called bags [35] which will serve as a reference for *HLA* allele imputation from SNP data. With this strategy, we obtained 5 bagging sets corresponding to the *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DQB1* and *HLA-DRB1* genes (available upon request). The bagging step is computationally very demanding as it required 30,000 CPU-hours on 700 CPUs to build a full reference panel on the 5 HLA genes (table S2). We then imputed *HLA* alleles in all available CAAPA samples (N = 4,993, table S1), and selected a post-probability threshold ≥ 0.5 as the recommended compromise between good accuracy imputation and call rate from Zheng et al. [35] (table S2 and S3). Choosing a higher post-probability threshold results in marginal accuracy improvements but in a large call rate drop [35].

The imputed *HLA* alleles were then uploaded in our in-house web application, Easy-HLA (http://hla.univ-nantes.fr/) [37], to infer the *HLA* 5-gene haplotypes (*HLA-A~HLA-B~HLA-C~HLA-DQB1~HLA-DRB1*) by extending the method of Gourraud *et al.* [38]. We identified a total of 4,077 unique 5-gene haplotypes, including 2,698 haplotypes with only one occurrence (frequency = 1.1×10^{-4}) up to one haplotype (30:01~42:01~17:01~03:02~04:02) with 149 occurrences (frequency = 0.017). To increase power of detecting an association with *HLA* haplotypes, we chose to restrict the regression analyses to 27 haplotypes with at least 20 occurrences representing a frequency of 0.2% in the pooled CAAPA individuals.

A complete workflow of the analyses is available as figure S1.

Ancestry estimation

Details of local ancestry inference can be found in the GWAS paper [31]. Briefly, after performing SNP phasing using ShapeIT [39], the 4,993 genomes were merged with African and European genomes from the 1000 Genomes project [40]. RFMix [41] inferred 15,824 local ancestry segments across the CAAPA genomes, including 20 segments in the *MHC* region on chromosome 6. The mean proportion of African ancestry was calculated from the RFMix estimation both genome-wide and in the *MHC* region. The local ancestry estimate was further used to correct for population differences in statistical analyses.

Statistical analysis

We performed association analyses on *HLA* alleles with good imputation accuracy (post-probability threshold ≥ 0.5) that were exhibiting an overall (case + controls) frequency ≥ 2% (54 *HLA* alleles in total, table S4). For each CAAPA study and each *HLA* allele, we performed a case-control (overall: 1,607 cases vs. 3,365 controls, table S1, fig S1) analysis in R [42] using logistic regression models to test for association with asthma susceptibility. From RFMix inference (see Ancestry estimation above), we have the information of local ancestry within *MHC*. To account for ancestry, we applied 2 strategies: 1) stratification on local ancestry only from African origin and 2) local ancestry as covariate in the regressions. The asthmatic cases were further divided between childhood onset (N = 804) and adult onset asthma (N = 445) and compared to controls.

Studentized residuals of log10 transformed total serum IgE (tIgE, adjusted for age and sex as described previously [32]) was available from 4/8 CAAPA studies. We further transformed this adjusted tIgE level into a z-score [43] to allow comparison between cohorts (Fig S8). Each *HLA* allele was then tested for association with tIgE levels separately in asthmatics (N = 972, table S1) and non-asthmatics (N = 816) using a linear regression model. We accounted for ancestry as described for case-control analyses. Atopy was defined as tIgE level > 80 KU/L and was only explored in asthma cases (atopic individuals = 725, and non-atopic individuals = 247).

We then performed a meta-analysis using METAL [44] for both case-control and tIgE statistical tests. We used the default sample size-based method from the METAL tool. Z scores in tables S5 and S6 are a weighted sum of Z score across studies; each Z score reflecting the direction of effect and P-value [44]. We also computed a weighted effect size using the inverse variance based method from the METAL tool for the most significant

alleles. We established a stringent Bonferroni threshold of significance at $P = 9.3x10^{-4}$ (0.05/54) accounting for the number of alleles tested [45,46].

Results

Description of the study population

In our study, we had access to a total of 4,993 individuals of African ancestry (USA and Barbados) from 8 CAAPA studies (Table S1). Overall, our patients were 39.3% male and were recruited at 38.7 years of age on average (for a description of the demographics in each cohort, see Table S1). Our analyses only focused on unrelated individuals with a total of 1,607 subjects with asthma and 3,365 controls (BAGS and HUFS cohorts included some family members, these individuals were excluded from our dataset). Finally, tIgE data was available for 972 subjects with asthma and 816 controls. In this subset group, median (IQR) tIgE was 214.0 (69.0-590.7) IU ml^{-1} for cases and 66.4 (25.0-212.2) IU ml^{-1} for controls respectively (for a description in each cohort, see Figure S8 and Supplementary Table S1). The workflow of the analysis is described in figure S1.

Admixture in the *MHC* region

Due to the recent history of admixture between African and European populations, individuals living in the Americas with African ancestry show a high proportion of European ancestry (on average 20% of European ancestry across the genome for African Americans and African Caribbean) [47]. The specific admixture structure within the *MHC* genomic region has not been studied in detail before. To understand the local ancestry structure in the *MHC* and the sensitivity of association results to differences in local ancestry, we inferred the local ancestry for the unrelated CAAPA individuals (full dataset, N = 4,993) both across the whole genome and specifically within the *MHC* region, and compared their admixture structure (figure 1A).

The whole genome admixture observed in CAAPA was typical of African ancestry individuals with a history of European admixture (figure 1A, top panel), and represents 81.9% on average with a median of 83.9% [inter-quartile range 77.3-89.0] of African ancestry across the genome for CAAPA individuals (figure 1B), as previously described in Mathias *et al.* [47]. As expected, the *MHC* region showed a non-normally distributed admixture among CAAPA individuals with a particular tri-modal distribution (figure 1A, bottom panel) that reflects the diploid origin of a small portion of the genome (1/1000). Indeed, we obtained similar tri-modal distribution patterns with similarly sized chromosome 6 segments (5Mb, figure 1C). Within the *MHC*, we observed European/European ancestry (100% light gray, 2.3% of the population), African/African ancestry (100% dark gray, 64.1% of the population), European/African ancestry (50/50, 12.6% of the population), and admixed individuals

carrying a mosaic of European and African *MHC* fragments emphasizing the recombination events in that locus (21.0% of the population, figure 1A, bottom panel). Importantly, similar distributions were observed across the individual CAAPA studies (Fig S3), indicating their homogeneity regarding admixture structure.

While the distribution patterns looked different, the whole genome African ancestry proportion correlated significantly with the *MHC* region African ancestry proportion (figure 1D; Pearson correlation: R = 0.39 [95% confidence interval 0.37-0.41], $P = 4.8 \times 10^{-180}$). This result indicates that the *MHC* admixture distribution correlates to the whole-genome distribution: individuals above whole-genome African ancestry proportion median (83.9%) are carrying a homozygous *MHC* region segment of African origin for 77.6% of them compared to 50.8% for individuals below median (fold increase of 1.5). However, it is important to emphasize that this very significant correlation does not preclude specific extreme cases such as the following two cases: 1) the red triangle on figure 1D represents an individual with 100% African ancestry within MHC region but 41.5% across whole genome; 2) at the opposite, the green diamond on figure 1B represents an individual with 0% African ancestry within MHC region but 89.9% across whole genome.

HLA alleles imputation

The imputation of *HLA* alleles is a powerful computational technique to infer *HLA* alleles only from genotyped SNPs of the *MHC* using a reference panel [35]. The accuracy of the imputation depends on the reference panel size and matching with the population of interest. Since the publicly available African ancestry reference panel was small (~150) and did not yield accurate *HLA* allele imputation, and because the allelic diversity is larger in African ancestry individuals compared to Europeans, we developed our own reference panel from a subset of 917 CAAPA individuals for whom we had genotyped *HLA* alleles in addition to the *MHC* SNP genotypes. Our large matching reference panel provided a much-improved quality of imputation (table S3) with an overall call rate of 81% (compared to 42% with the public embedded reference panel). This result emphasizes the need to develop and share large and diverse reference panels to improve *HLA* imputation accuracy (our panel is available upon request). This innovative approach therefore generated accurate *HLA* alleles for all CAAPA individuals, including 54 common *HLA* alleles with a frequency ≥ 2% that were further tested for association with asthma-related phenotypes.

Asthma susceptibility

For each *HLA* allele, we conducted a logistic regression analysis comparing asthmatics and non-asthmatics while stratifying by 2 African ancestry chromosomes (resulting in 1,178 cases and 2,421 controls).

First, we assessed the association of *HLA* alleles under an allelic model with asthma susceptibility in each CAAPA cohort using logistic regression. The results obtained in each study (table S1) were subsequently combined by meta-analysis (figure 2). The QQplot showed no inflation of P-values, demonstrating the fitness of our statistical model (lambda 0.93, META in figure S4). No *HLA* allele passed the Bonferroni threshold of significance ($P \leq 9.3x10^{-4}$) and the best association was identified for *HLA-B*15:16* (P = 0.04, weighted effect size 0.43 [0.04-0.83], figure 2). In this setting, we expect a statistical power > 90% to detect HLA associations with an allele frequency of 5% and an effect size of 1.5. Running the analysis by correcting for local ancestry (Fig S6) or under a dominant model (Fig S5 and S7) did not reveal any significant signals (full results presented in table S5).

We then inferred *HLA* 5-gene haplotypes (*HLA-A~HLA-B~HLA-C~HLA-DQB1~HLA-DRB1*) for each CAAPA individual using the local version of an HLA oriented web application (http://hla.univ-nantes.fr/). No statistically significant associations were observed.

Next, we investigated age-at-onset by running a stratified case-control association analysis (N = 804 childhood onset and N = 445 adult onset, respectively) but no significant associations were observed.

Finally, we focused specifically on the *HLA* alleles that were previously associated with asthma-related phenotypes in European and Asian populations [1,9,13–15] (Table 1), but did not identify any significant associations.

Total IgE levels in individuals with asthma

We next explored *HLA* associations with asthma atopy by focusing on the tIgE levels (available in 4/8 CAAPA studies) among subjects with asthma. tIgE levels are higher in asthmatics compared to non-asthmatics (Fig S8), and high tIgE levels associate with more severe asthma and lower baseline lung function [48]. For each *HLA* allele, we conducted a linear regression analysis in asthmatics using tIgE as a continuous variable while stratifying for 2 African ancestry chromosomes (resulting in 724 individuals). In the meta-analysis, *HLA-DRB1*09:01* was significantly associated with elevated tIgE levels under an allelic model ($P = 8.5x10^{-4}$, weighted effect size 0.51 [0.15-0.87], moderate heterogeneity across studies P = 0.04, figure 2 and 3, Fig S9), indicating that this allele is associated with degree of atopy in individuals with asthma. Association effects were in the same direction across all cohorts

and *HLA-DRB1*09:01* reached nominal significance in two cohorts (P = 0.02 in CAG and P = 0.02 in SARP). We obtained similar results by correcting for local ancestry (Fig S11) or using a dominant model (Fig S10 and S12, see table S6 for full results). This association is specific to the asthmatic population, as no association was found between *HLA-DRB1*09:01* (or any other *HLA* allele) and tIgE in the controls (N = 816). *HLA-DRB1*09:01* had a frequency of 3.8% in the overall CAAPA study (Table S4), 3.0% in the tIgE CAAPA subgroup (Table S6), 3.1% in the *HLA* genotyped CAAPA subset (N = 917), and 3.0% in a reference African American dataset (http://www.allelefrequencies.net, USA NMDP African American) [49], confirming the quality of our imputation and the confidence in this association signal. To test for possible biases between cohorts with and without tIgE availability, we performed an asthma case/control analysis focusing on the 4 cohorts with tIgE levels data. The results did not significantly differ from the whole data analysis and there was still no significant signal identified (Table S7). In a sensitivity analysis, we increased the post-probability threshold to test if our signal would hold. The direction and strength of the effect is conserved up to a > 0.7 and > 0.6 threshold, respectively; higher thresholds result in a large call rate drop (71% of HLA-DRB1 alleles reach a post-prob > 0.5 and 30% a post-prob > 0.8; see Table S8).

Finally, the *HLA* alleles that were previously associated with asthma-related phenotypes in European and Asian populations [1,9,13–15] were not associated with tIgE levels in our population (Table 1). Using tIgE levels, we also stratified our cohort into atopic (N = 725) and non-atopic (N = 247) individuals. Performing atopic vs. non-atopic analysis did not reveal any significant results, likely due to a lack of statistical power.

Discussion

Our study focused for the first time on *HLA* allele associations with asthma susceptibility and atopy in a large African ancestry cohort of individuals with asthma (CAAPA, N = 4,993). We identified the *HLA-DRB1*09:01* allele associated with elevated tIgE levels in individuals with asthma.

Thus far, most *HLA* studies in asthma either had small sample sizes or assessed associations only at the SNP level, therefore not fully capturing the complexity and biological relevance of the *HLA* genes in the *MHC* region. Moreover, since previous studies focused on populations of European and Asian descent, studying a large population of African ancestry is of great interest as it can shed new light on the relationship between HLA and asthma. In the recently published CAAPA GWAS (of the same data) [31], Daya et al. did not identified any significant associations for SNPs inside the *HLA* region (best SNP: rs9272346, P = 0.03). This reinforced our strategy to study *HLA* alleles. To investigate the specific role of HLA in asthma, we used a powerful machine-learning based computational technique to infer *HLA* alleles. We developed a new reference panel matching individuals of African ancestry in order to increase imputation accuracy in this (genetically diverse) population. We inferred *HLA* alleles with at least 86% accuracy for the *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DQB1* and *HLA-DRB1* genes in all CAAPA studies (table S2).

As the African American and Barbados African Caribbean genomes are a mosaic of European and African ancestral genomes, we first explored the specific admixture structure within the *MHC* genomic region. As *HLA* genes are under strong selection pressure [50], one could have expected differences between the *MHC* region and the rest of the genome. Our analyses revealed a similar proportion of admixture as in the rest of the genome (around 20% of European ancestry), and an admixture pattern that is comparable to similarly sized segments on chromosome 6 (figure 1C). This indicates that the *MHC* does not seem to be under ancestry-specific selective pressure compared to the rest of the genome and that we can analyze this genomic region as any other part of the genome. Yet, we must acknowledge the fairly recent history of admixture in African Americans and African Caribbean, having occurred only in the last few centuries, which does not exclude the possibility of detecting ancestry-specific selection in the future, after many more generations. In our analyses to account for admixture, we applied 2 strategies: 1) stratification on local ancestry only from African origin and 2) local ancestry as covariate in the regressions (see methods).

Our case-control analyses (1,607 vs. 3,365) did not reveal any significant *HLA* associations with asthma susceptibility. Focusing on tIgE levels to study atopy in individuals with asthma, we restricted our sample size to only 972 asthmatics from 4 studies (724 after African local ancestry stratification) and found *HLA-DRB1*09:01* associated with an increased level of tIgE in blood ($P = 8.5x10^{-4}$). This P-value reaches our Bonferroni threshold ($9.3x10^{-4}$) of significance calculated by dividing 0.05 by the number of tested alleles (N = 54). This association was not found in control individuals, indicating the specificity of the signal for atopy in individuals with asthma. These results illustrate how an extreme phenotype (high tIgE levels associated with increased asthma severity [48]) such as atopy can maximize statistical power of detection by limiting the noise due to the diversity of asthmatic phenotypes.

We have to acknowledge the difficulty of finding additional replication cohorts, with CAAPA standing as an exceptionally large dataset. The *HLA-DRB1*09:01* allele is relatively rare in African and European populations (1-3%) but frequent in East Asian populations (15-25%) [28]. In a validating cohort of 544 Japanese subjects including 103 asthmatics with tIgE [51], we did not identify any significant association (tIgE association with *HLA-DRB1*09:01*, P = 0.49, frequency = 11.8%). This is probably due to lack of power, but also could be explained by genetic background difference between African ancestry and Asian ancestry individuals or dissimilar environmental effects between the Japanese islands and the Americas.

Previous reports showed *HLA-DRB1*09:01* associated with several autoimmune diseases (rheumatoid arthritis, systemic lupus erythematosus [52], type 1 diabetes -especially late-onset- [53] and dermatomyositis [54]), especially in Asian populations where the allele is more frequent. This suggests that HLA-DRB1*09:01 carriers are at increased risk for self-peptide presentation and/or exacerbated chronic inflammation. This is in line with the pleiotropic effects observed in autoimmune diseases where some factors, especially genetics, share pathophysiological mechanisms across a span of autoimmune diseases [55].

Interestingly, *HLA-DRB1*09:01* was identified as a susceptibility allele to food allergy (peach allergy) [56]. This correlates with our findings of increased tIgE levels for *HLA-DRB1*09:01* carriers. Finally, in the immune epitope database (iedb.org) [57], CD4+ T cell epitopes against common environmental allergens were identified using Tetramer Guided Epitope Mapping. Among 86 epitopes, 4 Derp2 overlapping epitopes showed evidence of specific binding with HLA-DRB1*09:01 tetramers (direct submission to the immune epitope database, reference ID: 1024966) [57,58]. Derp2 is a well-known allergen from house dust mite

and a strong inducer of allergic asthma in susceptible individuals, which can lead to the immune system activation and elevation of tIgE level [59,60]. In the context of HLA-DRB1*09:01, this represents another lead in the possible mechanism of action of this allele in asthma atopy. The allele could be responsible of an increased activation of CD4+ T cells compared to other alleles, by presentation of specific dust mite peptides, which would trigger an amplified inflammation environment and result in an increased production of IgE by B cells. Further experiments are needed to confirm the association and define the functional relationship between HLA-DRB1*09:01 allele and tIgE.

We have to acknowledge some limitations in our study. Even if this is the largest ever African ancestry oriented study of HLA alleles association with asthma, this study remains restricted to a relatively small sample size, especially for samples with tIgE levels available (N = 972 individuals with asthma). However, this study conveys the ambitious will of gathering data from several research groups across multiple cities and countries which appears to be the future of genetics studies. The large diversity of cohorts is also a limitation both in terms of clinical definitions (different clinicians, definitions evolve through time) and in genetics and environmental backgrounds. These effects were limited here by using a meta-analysis technique and corrections from genetically-computed ancestry. The main difficulty remains to find an appropriate replication cohort when the goal of CAAPA was already to assemble as many cohorts as possible.

In conclusion, we identified for the first time *HLA-DRB1*09:01* allele association with tIgE levels in African ancestry individuals with asthma, potentially through a mechanism involving specific peptide presentation and/or increasing inflammation.

Acknowledgments

The authors thank Labex IGO (ANR-11-LABX-0016-01) and IHU-CESTI for their support. Nicolas Vince has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 846520.

We gratefully acknowledge the contributions of Pissamai and Trevor Maul, Paul Levett, Anselm Hennis, P. Michele Lashley, Raana Naidu, Malcolm Howitt, and Timothy Roach (BAGS), Audrey Grant, Eduardo Viera Ponte, Alvaro A. Cruz, and Edgar Carvalho (BIAS), Susan Balcer-Whaley, Maria Stockton-Porter, and Mao Yang (GRAAD), Delmy-Aracely Mejía-Mejía, Mario Meraz, Jaime Nuñez, Eileen Fabiani Herrera Mejía (HONDAS), Trevor Ferguson, Deanna Ashley (JAAS), Silvia Jimenez, Nathalie Acevedo, Dilia Mercado (PGCA), Ann Jedlicka (REACH), Addison K. May, Caroline Gilmore, Patricia Minton (Vanderbilt University), Qun Niu (University of Chicago), Adeyinka Falusi, Abayomi Odetunde (University of Ibadan, Nigeria). The authors also acknowledge the support of John Jay Shannon (Cook County Health Systems) and Kevin Weiss (Northwestern University), Regina Miranda and the Indians Zenues guards (San Basilio de Palenque, Bolivar, Colombia), Ulysse Ateba Ngoa (Leiden University), and Charles Rotimi, Adeyemo Adebowale, Floyd J Malveaux, and Elena Reece (Howard University). We thank the numerous health care providers and community clinics and co-investigators who assisted in the phenotyping and collection of DNA samples, and the families and patients for generously donating DNA samples to BAGS, BIAS, BREATHE, CAG, GRAAD, HONDAS, Jackson Heart Study, REACH, VALID, SARP, COPDGene, JAAS, PGCA, AEGS, and the asthma studies in Gabon and Palenque, Colombia. Special thanks to community leaders, teachers, doctors and personnel from health centers at the Garifuna communities for organizing the medical brigades and to the medical students at Universidad Católica de Honduras, Campus San Pedro y San Pablo for their participation in the fieldwork related to HONDAS; study coordinator Sandra Salazar; and health liaisons and public health officers of the main Conde office, Adaliudes Conceição, Luciana Quintela, Ivanice Santos, Analú Lima, Benivaldo Valber Oliveira Silva, and Iraci Santos Araujo, and students from the Federal University of Bahia who assisted in data collection in BIAS: Rafael Santana, Roberta Barbosa, Ana Paula Santana, Charlton Barros, Marcele Brandão, Ludmila Almeida, Thiago Cardoso and Daniela Costa. We are grateful for the support from the international state governments and universities from Honduras, Colombia, Brazil, Gabon, Nigeria, Netherlands, Jamaica,

Barbados and the United States who made this work possible. Funding for this study was provided in part by NIH grants R01-HL129239 (KCB) and R01HL104608 (KCB).

References

1. Kontakioti E, Domvri K, Papakosta D, Daniilidis M. HLA and asthma phenotypes/endotypes: A review. Hum Immunol. 2014;75:930–9.

2. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. Nat Med. 2012;18:716–25.

3. Hansen TH, Bouvier M. MHC class I antigen presentation: learning from viral evasion strategies. Nat Rev Immunol. 2009;9:503–13.

4. Jones EY, Fugger L, Strominger JL, Siebold C. MHC class II proteins and disease: a structural perspective. Nat Rev Immunol. 2006;6:271–82.

5. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014;42:D1001–6.

6. Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. Annu Rev Genomics Hum Genet. 2013;14:301–23.

7. Brostoff J, Mowbray JF, Kapoor A, Hollowell SJ, Rudolf M, Saunders KB. 80% of patients with intrinsic asthma are homozygous for HLA W6. Is intrinsic asthma a recessive disease? Lancet Lond Engl. 1976;2:872–3.

8. Bruce CA, Bias WB, Norman PS, Lightenstein LM, Marsh DG. Studies of HLA antigen frequencies, IgE levels, and specific allergic sensitivities in patients having ragweed hayfever, with and without asthma. Clin Exp Immunol. 1976;25:67–72.

9. Lara-Marquez ML, Yunis JJ, Layrisse Z, Ortega F, Carvallo-Gil E, Montagnani S, et al. Immunogenetics of atopic asthma: association of DRB1*1101 DQA1*0501 DQB1*0301 haplotype with Dermatophagoides spp.-sensitive asthma in a sample of the Venezuelan population. Clin Exp Allergy J Br Soc Allergy Clin Immunol. 1999;29:60–71.

10. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 2015;43:D423–31.

11. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, et al. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. Nat Genet. 2011;43:893–6.

12. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med.

2010;363:1211–1221.

13. Ivković-Jureković I, Zunec R, Balog V, Grubić Z. The distribution of HLA alleles among children with atopic asthma in Croatia. Coll Antropol. 2011;35:1243–9.

14. Munthe-Kaas MC, Carlsen KL, Carlsen KH, Egeland T, Håland G, Devulapalli CS, et al. HLA Dr-Dq haplotypes and the TNFA-308 polymorphism: associations with asthma and allergy. Allergy. 2007;62:991–8.

15. Torío A, Sánchez-Guerrero I, Muro M, Villar LM, Minguela A, Marín L, et al. HLA class II genotypic frequencies in atopic asthma: association of DRB1*01-DQB1*0501 genotype with Artemisia vulgaris allergic asthma. Hum Immunol. 2003;64:811–5.

16. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmüller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat Genet. 2018;50:42–53.

17. Galanter JM, Gignoux CR, Torgerson DG, Roth LA, Eng C, Oh SS, et al. Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: The Genes-environments & Admixture in Latino Americans study. J Allergy Clin Immunol. 2014;134:295–305.

18. Li X, Ampleford EJ, Howard TD, Moore WC, Torgerson DG, Li H, et al. Genome-wide association studies of asthma indicate opposite immunopathogenesis direction from autoimmune diseases. J Allergy Clin Immunol. 2012;130:861-868.e7.

19. Mathias RA, Grant AV, Rafaels N, Hand T, Gao L, Vergara C, et al. A genome-wide association study on African-ancestry populations for asthma. J Allergy Clin Immunol. 2010;125:336-346.e4.

20. Noguchi E, Sakamoto H, Hirota T, Ochiai K, Imoto Y, Sakashita M, et al. Genome-Wide Association Study Identifies HLA-DP as a Susceptibility Gene for Pediatric Asthma in Asian Populations. McCarthy MI, editor. PLoS Genet. 2011;7:e1002170.

21. Pino-Yanes M, Gignoux CR, Galanter JM, Levin AM, Campbell CD, Eng C, et al. Genome-wide association study and admixture mapping reveal new loci associated with total IgE levels in Latinos. J Allergy Clin Immunol. 2015;135:1502–1510.

22. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat Genet. 2011;43:887–92.

23. Lasky-Su J, Himes BE, Raby BA, Klanderman BJ, Sylvia JS, Lange C, et al. HLA-DQ strikes again: Genome-wide association study further confirms *HLA-DQ* in the diagnosis of asthma among adults. Clin Exp Allergy. 2012;42:1724–33.

24. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet. 2017;49:1752–7.

25. Shrine N, Portelli MA, John C, Soler Artigas M, Bennett N, Hall R, et al. Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study. Lancet Respir Med. 2019;7:20–34.

26. Kim KW, Ober C. Lessons Learned From GWAS of Asthma. Allergy Asthma Immunol Res. 2019;11:170–87.

27. Nicodemus-Johnson J, Laxman B, Stern RK, Sudi J, Tierney CN, Norwick L, et al. Maternal asthma and microRNA regulation of soluble HLA-G in the airway. J Allergy Clin Immunol. 2013;131:1496-1503.e4.

28. Gourraud P-A, Khankhanian P, Cereb N, Yang SY, Feolo M, Maiers M, et al. HLA diversity in the 1000 genomes dataset. PloS One. 2014;9:e97282.

29. Limou S, Coulonges C, Herbeck JT, van Manen D, An P, Le Clerc S, et al. Multiple-Cohort Genetic Association Study Reveals CXCR6 as a New Chemokine Receptor Involved in Long-Term Nonprogression to AIDS. J Infect Dis. 2010;202:908–15.

30. Limou S, Zagury J-F. Immunogenetics: Genome-Wide Association of Non-Progressive HIV and Viral Load Control: HLA Genes and Beyond. Front Immunol. 2013;4:118.

31. Daya M, Rafaels N, Brunetti TM, Chavan S, Levin AM, Shetty A, et al. Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. Nat Commun. 2019;10:880.

32. Levin AM, Mathias RA, Huang L, Roth LA, Daley D, Myers RA, et al. A meta-analysis of genome-wide association studies for serum total IgE in diverse study populations. J Allergy Clin Immunol. 2013;131:1176–84.

33. Johnston HR, Hu Y-J, Gao J, O'Connor TD, Abecasis GR, Wojcik GL, et al. Identifying tagging SNPs for African specific genetic variation from the African Diaspora Genome. Sci Rep. 2017;7:46398.

34. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48:1279–83.

35. Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR, et al. HIBAG--HLA genotype imputation with attribute bagging. Pharmacogenomics J. 2014;14:192–200.

36. Pappas DJ, Lizee A, Paunic V, Beutner KR, Motyer A, Vukcevic D, et al.

Significant variation between SNP-based HLA imputations in diverse populations: the last mile is the hardest. Pharmacogenomics J. 2018;18:367–76.

37. Geffard E, Limou S, Walencik A, Daya M, Watson H, Torgerson D, et al. Easy-HLA, a validated web application suite to reveal the full details of HLA typing. Bioinforma Oxf Engl. 2019;

38. Gourraud P-A, Lamiraux P, El-Kadhi N, Raffoux C, Cambon-Thomsen A. Inferred HLA haplotype information for donors from hematopoietic stem cells donor registries. Hum Immunol. 2005;66:563–70.

39. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013;10:5–6.

40. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.

41. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am J Hum Genet. 2013;93:278–88.

42. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2017. Available from: https://www.R-project.org/

43. Kreyszig E. Advanced engineering mathematics. 4. ed. New York: Wiley; 1979. 939 p.

44. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinforma Oxf Engl. 2010;26:2190–1.

45. Vince N, Bashirova AA, Lied A, Gao X, Dorrell L, McLaren PJ, et al. HLA class I and KIR genes do not protect against HIV type 1 infection in highly exposed uninfected individuals with hemophilia A. J Infect Dis. 2014;210:1047–51.

46. Dunn OJ. Multiple Comparisons among Means. J Am Stat Assoc. 1961;56:52–64.

47. Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD, et al. A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. Nat Commun. 2016;7:12522.

48. Naqvi M, Choudhry S, Tsai H-J, Thyne S, Navarro D, Nazario S, et al. Association between IgE levels and asthma severity among African American, Mexican, and Puerto Rican patients with asthma. J Allergy Clin Immunol. 2007;120:137–43.

49. Maiers M, Gragert L, Klitz W. High-resolution HLA alleles and haplotypes in the United States population. Hum Immunol. 2007;68:779–88.

50. Meyer D, C Aguiar VR, Bitarello BD, C Brandt DY, Nunes K. A genomic perspective on HLA evolution. Immunogenetics. 2017;

51. Morii W, Sakai A, Ninomiya T, Kidoguchi M, Sumazaki R, Fujieda S, et al. Association of Japanese cedar pollinosis and sensitization with HLA-DPB1 in the Japanese adolescent. Allergol Int Off J Jpn Soc Allergol. 2018;67:61–6.

52. Shimane K, Kochi Y, Suzuki A, Okada Y, Ishii T, Horita T, et al. An association analysis of HLA-DRB1 with systemic lupus erythematosus and rheumatoid arthritis in a Japanese population: effects of *09:01 allele on disease phenotypes. Rheumatology. 2013;52:1172–82.

53. Murao S, Makino H, Kaino Y, Konoue E, Ohashi J, Kida K, et al. Differences in the contribution of HLA-DR and -DQ haplotypes to susceptibility to adult- and childhood-onset type 1 diabetes in Japanese patients. Diabetes. 2004;53:2684–90.

54. Lin JM, Zhang YB, Peng QL, Yang HB, Shi JL, Gu ML, et al. Genetic association of HLA-DRB1 multiple polymorphisms with dermatomyositis in Chinese population. HLA. 2017;90:354–9.

55. Anaya J-M. Common mechanisms of autoimmune diseases (the autoimmune tautology). Autoimmun Rev. 2012;11:781–4.

56. Khor S-S, Morino R, Nakazono K, Kamitsuji S, Akita M, Kawajiri M, et al. Genome-wide association study of self-reported food reactions in Japanese identifies shrimp and peach specific loci in the HLA-DR/DQ gene region. Sci Rep. 2018;8:1069.

57. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. Nucleic Acids Res. 2015;43:D405-412.

58. Wambre E, James EA, Kwok WW. Characterization of CD4+ T cell subsets in allergy. Curr Opin Immunol. 2012;24:700–6.

59. Wang X, Yang Q, Wang P, Luo L, Chen Z, Liao B, et al. Derp2-mutant gene vaccine inhibits airway inflammation and up-regulates Toll-like receptor 9 in an allergic asthmatic mouse model. Asian Pac J Allergy Immunol. 2010;28:287–93.

60. Genc S, Eroglu H, Kucuksezer UC, Aktas-Cetin E, Gelincik A, Ustyol-Aycan E, et al. The decreased CD4+CD25+ FoxP3+ T cells in nonstimulated allergic rhinitis patients sensitized to house dust mites. J Asthma Off J Assoc Care Asthma. 2012;49:569–74.

Figure 1: Ancestry proportion are equivalent between whole genome and *MHC* genomic region. (A) CAAPA individual's ancestry proportion (full dataset, N = 4,993). X axis: each individual. Y axis: ancestry proportion. Black horizontal line: 80% African ancestry. The *MHC* only represents a thousandth of the genome (5Mb on chromosome 6). Mean local ancestry: light gray for European, dark gray for African. (B) Violin plot comparing African ancestry proportion. Median: white dots. Interquartile range: black boxes. Black lines: 95% confidence interval. (C) Density plot comparing African ancestry proportion between whole genome (red), MHC region (blue) and similarly-sized chromosome 6 segments (gray, N = 42 segments). (D) The graph combines 1) a scatter plot comparing African ancestry proportion across all CAAPA individuals and 2) a density plot showing the proportion of individuals with a specific level of African ancestry in whole genome (right) and *MHC* genomic region (top). The blue line was drawn with the correlation coefficient. Pearson correlation: $P = 4.8x10^{-180}$, R = 0.39 [95% confidence interval 0.37-0.41]. Two extreme individuals are represented: red triangle (100% African ancestry MHC, 41.5% whole genome), green diamond (0% African ancestry MHC, 89.9% whole genome).

Figure 2: *HLA-DRB1*09:01* allele associates with asthma atopy. Manhattan plot of each *HLA* allele association for case-control analysis (blue) and tIgE analysis (red). P-values are meta-analysis from regression analyses stratified for African ancestry chromosomes and represented as -log10 of P-values. Red horizontal line represents Bonferroni threshold of significance ($P = 9.3x10^{-4}$). Black horizontal line represents nominal threshold of significance (P = 0.05). Each *HLA* allele tested are represented.

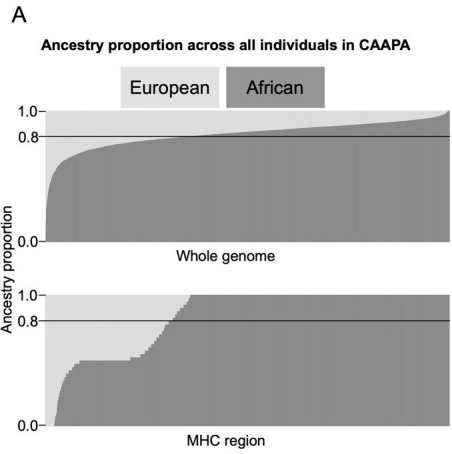
Figure 3: *HLA-DRB1*09:01* allele associates with higher tIgE. (A) Box plot comparing tIgE level between Asthmatics with 1 or 2 *HLA-DRB1*09:01* allele (+/- and +/+) and Asthmatics without HLA-DRB1*09:01 allele (-/-). tIgE level is expressed in z-score. The N representing the number of individuals in each group is depicted on the graph, these numbers exclude *HLA-DRB1* alleles with a post-probability < 0.5 and individuals with 2 chromosomes from African ancestry (Figure S2). (B) Forest plot representing the association between HLA-DRB1*09:01 and tIgE in the different CAAPA cohorts where tIgE is available. weighted effect size 0.51 [0.15-0.87]. BAGS effect size 0.07 [-0.41-0.56]. CAG effect size 1.48 [0.31-2.64]. GRAAD effect size 0.80 [0.00-1.60]. SARP effect size 1.23 [0.24-2.21].

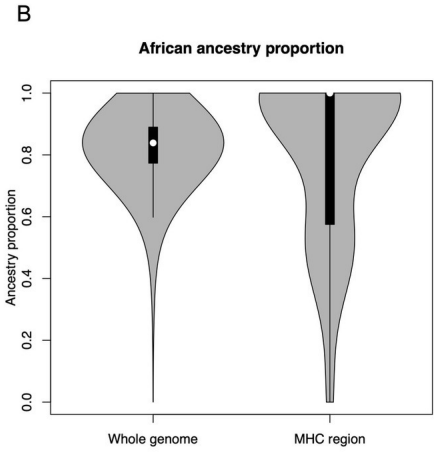
HLA	**Case vs. Control meta-analysis P-**	**tIgE level meta-analysis**	**Population**	**Pubmed ID**

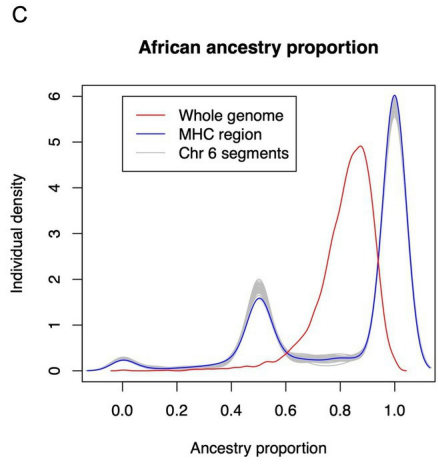
	value	**P-value**		
DQB1*05:01	0.82	0.55	Hispanic, European	10051703 [9], 17686102 [14]
DRB1*01:02	0.95	0.66	Hispanic, European	12878360 [15], 22397267 [13], 17686102 [14]
DRB1*13:01	0.09	0.09	European	17686102 [14]

Table 1: P-values of previously found asthma associated *HLA* alleles tested in our study. P-values are from the meta-analysis on case-control and tIgE levels in CAAPA individuals with asthma using an allelic model and stratifying for 2 African ancestry chromosomes.

Fig.1


A
Ancestry proportion across all individuals in CAAPA
European
African
1.0
0.8
0.0
Ancestry proportion
Whole genome
MHC region


B
African ancestry proportion
Ancestry proportion
0.0
0.2
0.4
0.6
0.8
1.0
Whole genome
MHC region


C
African ancestry proportion
Whole genome
MHC region
Chr 6 segments
Individual density
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
Ancestry proportion

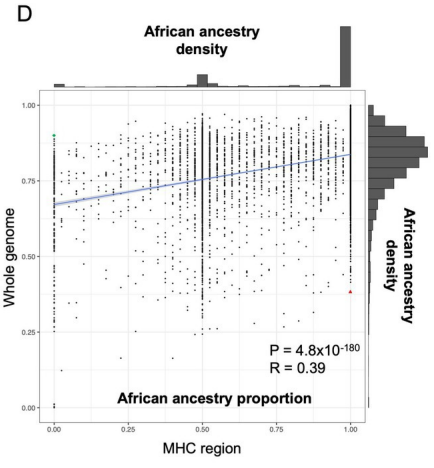

D
African ancestry density
Whole genome
African ancestry density
P = 4.8x10-180
R = 0.39
African ancestry proportion
0.00
0.25
0.50
0.75
1.00
MHC region

Fig.2

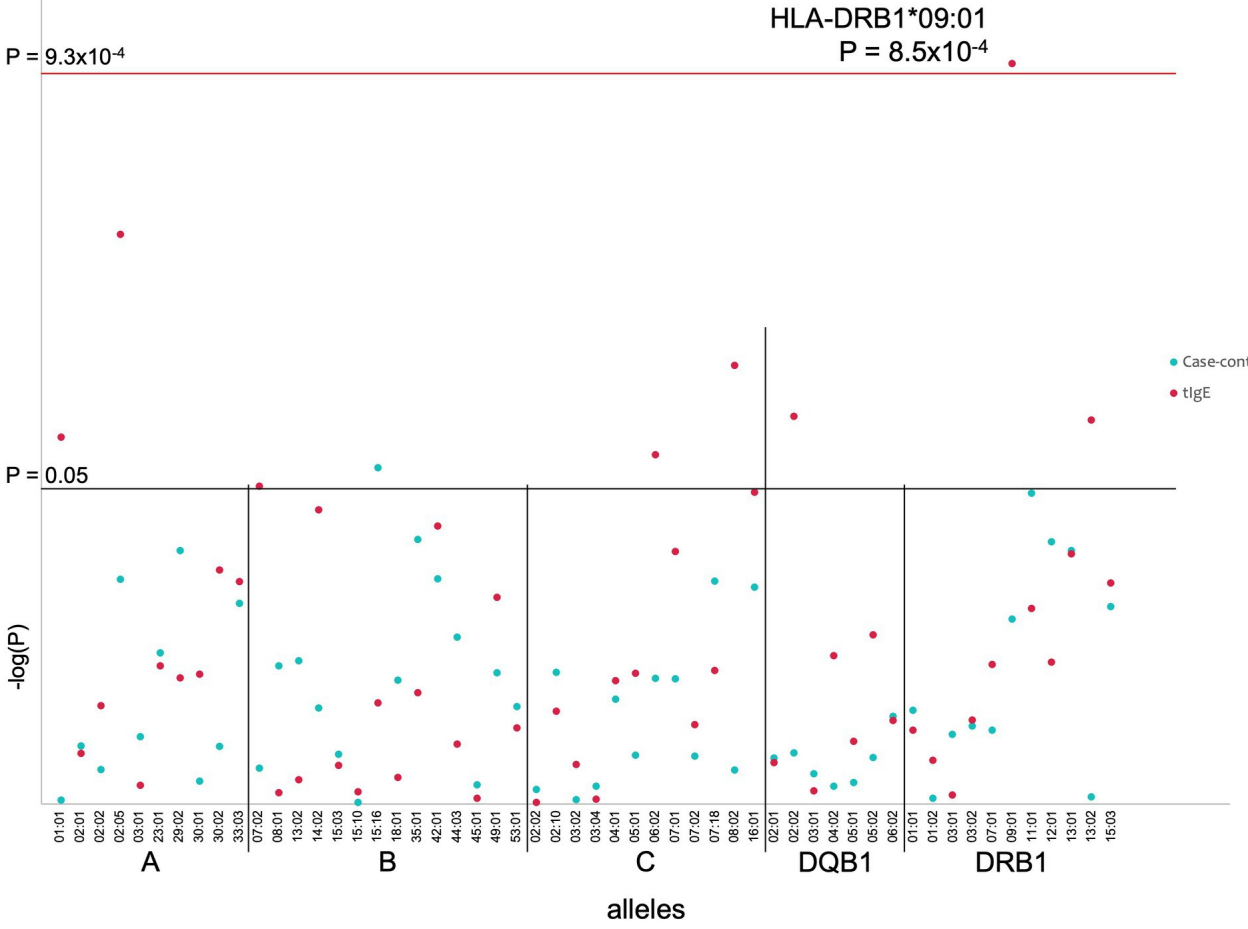
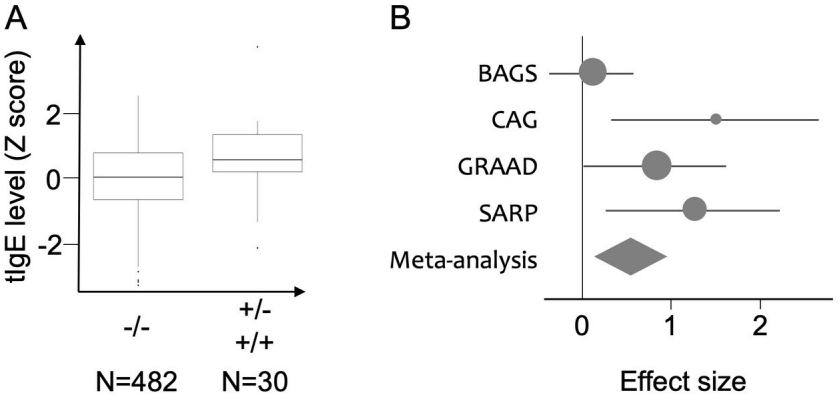

HLA-DRB1*09:01
P = 8.5x10-4
P = 9.3x10-4
P = 0.05
-log(P)
Case-control
tIgE
01:01
02:01
02:02
02:05
03:01
23:01
29:02
30:01
30:02
33:03
A
07:02
08:01
13:02
14:02
15:03
15:10
15:16
18:01
35:01
42:01
44:03
45:01
49:01
53:01
B
02:02
02:10
03:02
03:04
04:01
05:01
06:02
07:01
07:02
07:18
08:02
16:01
C
02:01
02:02
03:01
04:02
05:01
05:02
06:02
DQB1
01:01
01:02
03:01
03:02
07:01
09:01
11:01
12:01
13:01
13:02
15:03
DRB1
alleles

Fig.3


A
tIgE level (Z score)
2
0
-2
-/-
+/-
+/+
N=482
N=30
B
BAGS
CAG
GRAAD
SARP
Meta-analysis
0
1
2
Effect size