Asthma and Genetics of Admixture in Puerto Ricans

by

Israel Rivera

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
BIOLOGY

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2015

Approved by:

_____	_____
Gloribell Ortiz-Rios Graduate representative	Date
_____	_____
Juan C. Martínez-Cruzado, PhD Member, Graduate Committee	Date
_____	_____
Edu Suarez, PhD Member, Graduate Committee	Date
_____	_____
Taras K. Oleksyk, PhD President, Graduate Committee	Date
_____	_____
Nannette Diffoot, PhD Chairperson of the Department	Date

ABSTRACT

Asthma is one of the most recognized complex human disorders, characterized by an airway obstruction due to an exacerbated immune response. Currently, it affects around 300 million people worldwide, showing an increased prevalence over the past 2 decades particularly in developed countries, and has been recognized as one of the major public health problems worldwide. However, there is an uncertainty to the contribution of genetic and environmental factors to the disease. In Puerto Rico, asthma prevalence rose by 15% between 2000 and 2007, reaching an alarming 20% in 2003, making Puerto Rico a population with one of the highest prevalence worldwide. Since the condition often shows increased frequencies in the populations with African ancestry around 12-14% (Afro-Americans), and since Puerto Ricans share on average a 21.2% of African ancestry, it has been suggested that a genetic factor of African origin might be involved. In this thesis, we used a two-tier experimental strategy: (1) a candidate gene approach to confirm genetic associations and (2) an admixture approach to test for associations of the asthma phenotype with one of the three common ancestries on the island (European, African, and Native American). Using the candidate gene approach, we confirmed a positive association between *ORMDL3* (rs8076131) and asthma morbidity, showing a higher SNP call rate for the minor allele in Puerto Rican individuals with asthma. However, contrary to the original expectation, but in line with the current literature, our results also suggested an association with a higher European ancestry. In conclusion, our results confirmed the role of ancestry in the genetics of asthma among Puerto Ricans, and supported the importance of population genetic characterization in order to fully capture the elements involved in asthma pathogenesis.

RESUMEN

Asma es una de las enfermedades complejas más reconocidas, caracterizada por una obstrucción de las vías respiratorias debido a una respuesta inmune exacerbada. Hoy día, afecta alrededor de 300millones de personas alrededor del mundo, incrementando su prevalencia por las pasadas 2 décadas, particularmente en países desarrollados. No obstante, hay un alto grado de incertidumbre entre la contribución genética y la contribución ambiental. En Puerto Rico, la prevalencia en asma a aumentado un 15% durante el 2000 y 2007, hasta llegar a un 20% en el 2003, siendo Puerto Rico uno de los países con la más altas tasa de prevalencia a nivel mundial. Dado que se han observado una alta prevalencia en poblaciones de origen Africano (12%-14%), y Puerto Rico comparte en promedio un 21.2% de origen Africano, se ha sugerido que un factor genético puede ser responsable de la observada prevalencia en la isla. En esta tesis, utilizamos dos estrategias experimentales: 1) observamos la asociación entre mestizaje y prevalencia utilizando genes candidatos comparados simultáneamente con marcadores informativos genéticos que permiten calcular el promedio de mestizaje en los individuos de interés. Utilizando este método, confirmamos una asociación positiva entre *ORMDL3* (rs8076131) y la prevalencia de asma, observando una mayor frecuencia alélica para este polimorfismo en individuos con asma. Por otro lado, contrario a lo esperado originalmente, pero concordante con literatura previa, nuestros resultados sugieren una asociación con individuos que presentan un mayor origen ancestral Europeo. En conclusión, nuestros resultados confirman el rol de mestizaje en la genética de asma en la isla entre los Puertorriqueños, y denota la importancia de caracterizar a la población a manera de capturar los elementos envueltos en la patogenicidad de asma.

To my family and Friends I hope to make you proud.

To my kids, the greatest gift ever received.

ACKNOWLEDGEMENTS

First, I want to thank my faculty committee for their unconditional support. Dr. Edu Suarez, Dr. Juan Carlos Martinez-Cruzado and Dr. Taras Oleksyk, have been fundamental in my academic success and personal growth. Second, I want to thank my family, Daria Colon (mother), Israel Rivera (father), Emmanuel Rivera (brother), and Xiomara Rivera (girlfriend), they have encouraged and supported me in every step of the way. Also, I want to thank my beautiful kids Xionelys Gabriela Rivera and Israel Gabriel Rivera, their mere presence warms my heart with indescribable joy and love.

I want to extent my special thanks to my colleagues, Yashira Afanador, Stefanie Cosme, Ingrid Rivera, Frances Marín, Priscila Rodríguez, Nikole Ayala, Angelia Caro and Wilfried Guiblet. It was an honor working alongside with you, thanks you for the fun moments and the shared knowledge.

Finally, I want to address my special thanks for the person that made all this possible, Dr. Taras Oleksyk. Thank you for giving me this wonderful opportunity and believing in my potential. To you I own a great debt.

Table of Contents

List of Tables

List of Figures

Introduction

Asthma is one of the most recognized complex human disorders characterized by an airway obstruction due to an exacerbated immune response. Currently, it affects around 300 million people worldwide, showing an increasing prevalence over the past two decades particularly in developed countries, and has been recognized as one of the major public health problems worldwide (Barnett *et al.*, 2011). According to recent estimates, the total annual cost of asthma in the U.S. was $56 billion dollars (Sullivan *et al. 2011*). The relative contribution of environmental and genetics factors to the asthma prevalence has been discussed: the levels of asthma are highest among socioeconomically disadvantaged populations, but its impact is also differentially distributed among ethnic groups (Thakur *et al.*, 2013). Specifically, there is a difference in the prevalence rate among African Americans and the Latino population in comparison to the population of European ancestry (Galanter *et al.*, 2011; Torgerson *et al.*, 2011). From 2001 to 2003 the prevalence of asthma in the U.S. varied among ethnic groups with rates ranging from 7.7% - 7.8% in European Americans, to 9.5% - 12.5% in African Americans, 3.9% in Mexicans and 14.2% - 14.5% in Puerto Ricans (Sullivan *et al.*, 2011).

Statistical data obtained from the Puerto Rico's Department of Health shows that, at the time of the study in 2007, approximately 172,651 Puerto Ricans (out of nearly 4.0M) had asthma, and 439,773 Puerto Ricans were reported as diagnosed with asthma at some point in their lives by a health care professional (Puerto Rico Asthma Surveillance Report). Data from the same study demonstrated higher lifetime prevalence (18.8%) in comparison to the U.S. lifetime prevalence (13%), with mortality rates of 280 deaths per million, the highest asthma morbidity and mortality rates of all states and territories. The ethnically admixed genetic background of Puerto Ricans has

been proposed as one of the main reasons for the observed health disparity (Choudhry *et al.*, 2008 and Galanter *et al.*, 2011)

Previous studies estimated the average ancestry values for the Puerto Rican population to be 15.2%, 21.2%, and 63.7% for the Native American, African, and European contributions respectively (Via *et al.*, 2011). Since asthma is more frequently seen in populations of African ancestry, and Puerto Ricans share an average of 21.2 % African ancestry, it could be assumed that an admixture factor might be involved in the observed disparity among Puerto Ricans. These ancestral genetic contributions may be responsible for the observed differences in asthma morbidity, since they may carry risk alleles associated with the continental genome ancestries: Native American, African, and European (Via *et al.*, 2011).

Until recently, research has mainly been focused on the comparative association at a genomic scale by performing case-control comparative analyses. Genome-wide association studies (GWAS) have been used successfully to identify hundreds of common genetic variants associated with complex human diseases and traits, and provided valuable insights into their genetic architecture (Bossé *et al.*, 2007; Choudhry *et al.*, 2008; Daley *et al.*, 2012 and Ferreira *et al.*, 2011). On the other hand, although some of the variants identified through the use of GWAS analysis does increase the disease risk prevalence in the assayed sampled cohorts, this approach has been somewhat limited by the lack of reproducibility achieved in follow-up studies (Sharma *et al.*, 2005 and Deindl *et al.*, 2004). Furthermore, due to the Puerto Rican unique historical heritage, common variations found through GWAS analysis have been either non-informative or did not address the protective form of allele presence (Galanter *et al.*, 2011). In that matter, with a difficulty to reach a clear consensus, this approach alone cannot completely address all the elements involved in the disease. To address these limitations, genome-wide association testing for admixture mapping has

been used as an alternative to the traditional genome-wide association studies, a model that is theoretically more suitable for studying complex disease genetics in admixed populations (Manolio *et al.*, 2009). Moreover, implementation of candidate genes sequencing and genotyping techniques across highly admixed populations allow for a more robust identification of the factors that might be associated in the morbidity of the disease and are not considered with current GWAS. Puerto Ricans are a suitable model to understand this relationship as they have acquired characteristics from three ancestral origins, allowing for a deeper understanding of the inheritance of complex diseases.

In this work, we selected candidate gene polymorphisms associated with asthma morbidity from across different ethnic groups. These genes have been chosen among the ones reported to affect different populations (including Europeans, Africans, as well as Native Americans) in order to access genetic contribution of historic ancestry among the Puerto Rico's highly admixed population. Also, given that asthma prevalence rate is prominent in the population with a high African ancestry**,** we followed the association tests with the study of admixture to evaluate the possibility of differential ancestral genetic contributions to asthma. Our working hypotheses are that (1) variants present in the genes previously associated with asthma contribute to morbidity among Puerto Ricans, and (2) that one of the three ancestries in the admixture is elevated among the people diagnosed with this disease.

Literature review

Genes and Asthma

Asthma is an inflammatory disorder of the airways characterized by a chronic airway inflammation and airway hyper-reactivity (Kim *et al.*, 2010). The study of asthma has been characterized by a high degree of difficulty based on the complexity associated with the disease, mostly due to the genetic heterogeneity and the lack of a clear phenotype stratification of the disease (Ferreira *et al.*, 2011; Galanter *et al.*, 2011; Ming-liang *et al.*, 2011; Nguyen *et al.*, 2012 and Torgerson *et al.*, 2011). Several phenotypes such as allergic asthma, severe resistant asthma and asthma induced by exposure to air pollution, makes the studies of this disease a real challenge (Kim *et al.*, 2010). Furthermore, the lack of gene heterogeneity present in different assayed cohorts (mostly based on differences in ethnic contributions) has not provided a clear definition neither for the disease genetic characteristics nor for the specific molecular pathways that play a role in the immune system response (Ferreira *et al.*, 2010; Galanter *et al.*, 2008, Gordon *et al.*, 2010; Holtzman *et al.*, 2012 and Hsu *et al.*, 2013)

A significant number of candidate gene variants have been reported to be associated with asthma (Bossé *et al.*, 2007; Choudhry *et al.*, 2008; Daley *et al.*, 2012; Gaudieri *et al.*, 2012 and Kathleen *et al.*, 2010). Most of the findings are based on common SNP's polymorphisms, but many studies also point out genetic variants such as: Copy Number Variants (CNV) (Murphy *et al.*, 2010 and Craddock *et al.*, 2010) and rare SNPs variants (Torgerson *et al.*, 2012). However, although most of the previous works agree that the main role in the asthma pathology belongs to the exacerbated immune response, there is no clear consensus about the specific molecular pathways. This is not surprising, since the immune system activation is complex and studies suggest that pathways involved in inflammatory response are cross-activated by multiple molecular factors and

components (Freeman *et al.*, 2011; Gaudieri *et al.*, 2012 and Moffatt *et al.*, 2010). Nonetheless, based on previous finding of associated candidate genes, several models of immune system activation have been proposed (Moffatt *et al.*, 2010; Rogers *et al.*, 2009; and Torgerson *et al.*, 2011).

The main accepted model involves the activation pathway of the Type 2 helper T cells (Th2) inflammation in response to epithelial damages (Moffatt *et al.*, 2010). This pathway is supported by the high levels of IgE commonly found in asthmatic individuals, associated with the close relation seen between Th2 inflammatory response and atopy (Moffatt *et al.*, 2010; Torgerson *et al.* 2012; Holtzman *et al.*, 2012 and Kim *et al.*, 2010). This model supports that Th2 activation regulates inflammation related cytokines such as: Interleukin 4 (*IL-4*) Interleukin 3 (*IL-31*), Interleukin 33 (*IL-33*), Interleukin 12 (*IL-12*), *SMAD3, ORMDL3, SDMB*, eosinophil recruitment (*IL-5*), the recruitment and growth of mast cells (*IL-9*) and several interleukin receptors such as: *IL1RL1* and *IL18R1* among other significant important effectors of the inflammatory immune response. Most of cytokines and regulation factors involved in the homeostasis of the inflammatory response have been documented to present some sort of genetic variant that might be responsible for exacerbation in the inflammatory response observed in asthma individuals from different ethnic ancestries (Galanter *et al.*, 2011).

For instance, cytokine IL-33, constitutively expressed in endothelial and epithelial cells, drives production of Th2 associated cytokines and activates a nuclear factor κβ (NF- κβ) and mitogen-activated protein (MAP) kinases, alerting the immune system of epithelial damage during infection (Holtzman *et al.*, 2012; Kim *et al.*, 2010; and Moffatt *et al.*, 2010). IL-33 has been linked to asthma in Latino and African American, as well as European populations (Galanter *et al.*, 2011). Another widely supported effector involved with asthma among Europeans, African Americans

and Latinos is ORM1-like protein 3 (*ORDML3*). This gene encodes for a transmembrane protein localized in the endoplasmic reticulum that may be involved in endoplasmic reticulum stress and inflammation (Galanter *et al.*, 2008). Furthermore, there are instances in which replicated results have not been capable to provide a physiological explanation for the involvement of an observed SNP in a target population. For example, *RAD50* encodes for a constitutively low expressed protein involved in DNA double-strand break repair and has been replicated in several asthma cohort population although there is not any direct function related to asthma (Xingnan *et al.*, 2009 and Murk *et al.*, 2011). However, Rad50 is located on chromosome 5q31, which is a Th2-cytokine locus that contains several asthma-implicated genes, such as *IL4*, *IL13* and *IL5* (Murk *et al.*, 2011), this denotes the importance of chromosomal position and linkage disequilibrium values of the assayed SNP with other asthma related genes.

It is relevant to point out the importance of gene variant selection as a function of admixture, since many reported variants that might be responsible for the disease prognosis in one population could be non-informative in another ethnic populations with different local adaptations and demographic histories. As an example, *PYH1N1* encodes a protein-protein interaction domain and is only found to be associated in Puerto Ricans as opposed to other populations with a distinct ethnic background (Galanter *et al.*, 2011). Another example is a SNP in the histamine-N-methiltransferase (*HNMT*), an important mediator of the allergy immune response, that has been shown to reduce its enzymatic activity and to represent a risk factor for asthma, by lowering the enzyme activity thus raising histamine levels, increasing bronchoconstriction and asthma susceptibility (Yan *et al.*, 2010; and Yan *et al.*, 2012). Nonetheless, in the follow-up studies researchers were unable to reproduce the same result in different population cohorts. For instance, in two independent studies, cohorts from German (Deindl *et al*,. 2005) and Indian populations

(Sharma *et al.*, 2005) did not confirm any associations with asthma within the *HNMT* gene. On the other hand, a different study using admixture mapping found a strong association between the *HNMT* region and asthma in a Puerto Rican cohort (Galanter *et al*,. 2011). These results demonstrate the lack of consensus and the high level of disease heterogeneity, showing the importance of population characterization especially in admixed individuals. It also indicates that asthma related variants may be related to the local ancestry. Thus, there is a reason to validate the GWAS results in local, ethnically distinct and admixed populations.

Population structure

Because ancestry contribution plays a major role in the discovery and reproducibility of important variants, studies of diverse populations will allow for identification of variants that contribute to the disease prevalence across different ethnic groups (Torgerson *et al.*, 2012). Previous results have demonstrated the power of local ancestry association (or admixture mapping) in Latinos by finding susceptibility differences using previously reported risk gene variants, even including African admixture components associated with the morbidity of the disease (Nguyen *et al.*, 2012; Torgerson *et al.*, 2011 and Torgerson *et al.*, 2012). Furthermore, different studies on Puerto Rican admixture have demonstrated the impact of genetic ancestry and risk disease prevalence (Smith & O'Brien, 2005 and Martinez-Cruzado *et al.*, 2005)

In 2011, Galanter *et al.* reported different patterns of genetic association contributing to disease risk between two Latino populations (Mexicans and Puerto Ricans). In this work, only a minority of previously asthma related genes were shared among the two, showing discrete differences in the genetics factors associated with asthma that may account for the clear disparities. They suggested that studies attempting to analyze different Latino populations jointly should design the experiment so that there is no significant between-group heterogeneity and caution

should be exercised in applying the results from genetic studies in one Latino population to others (Galanter *et al.*, 2011).

Given the importance of genetic homogeneity when establishing risk disease factors proper population characterization should be carefully applied. In 2011, Via *et al.* suggested that genetic ancestry may confer different risks for asthma among Puerto Ricans in low versus high socioeconomic groups as a result of interactions between genetic and social factors. Using a set of ancestry informative markers (AIMs) they assessed the genetic admixture component of a census-based sample of Puerto Ricans, thus estimating the admixture proportions of the tri-hybrid admixture Puerto Rican population. In the current study, we aim to validate candidate gene association, and test for correlation between genome-wide admixture proportions and asthma morbidity by comparing the average ancestral admixture component between cases and reference samples. We apply STRUCTURE, a Bayesian Cluster Algorithm (Pritchard *et al.*, 2000), using a total of 93 AIMs and three ancestral populations (K=3 for European, African and Native American ancestral components), similar to the earlier studies, to determine the mean admixture proportions of both assayed samples (asthma cases and references) and allowing to test for any possible correlations between admixture and asthma disease prevalence on the target samples (Via *et al.*, 2011).

In conclusion, we believe that a robust analysis based on the unique genetic characteristics of the Puerto Rican population, can lead to the identification of locally relevant, specific genetics variants associated with asthma. If new variants were discovered, this would contribute to the better understanding of the genetics of the disease in admixed populations, especially in Puerto Rico.

Methods

Recruitment and Collection

A case-reference approach was performed with 80 asthma phenotype individuals (provided by Edu Suárez from the University of Puerto Rico at Ponce, see Appendix Table 2). In addition, a total of 259 samples from the Local Genome Diversity Studies (LGDS) project representing 23 municipalities were selected as part of our reference panel cohort. Municipalities represented in our study are shown in Figure 1. Sample id description is shown in Appendix

Figure 1. Map of Puerto Rico showing the outlined of its 78 municipalities (districts). For the purpose of our study, a total of 23 municipalities were selected (in green) to represent different regions of the island

DNA Extraction

Genomic DNA was extracted from saliva mouthwash as well as from the whole blood samples. DNA from our case cohort (asthma samples) was extracted from blood samples using QIAamp® Blood Mini Kit protocol, as recommended by the manufacturer (Qiagen 2007).

Genomic DNA for the reference (LGDS cohort) samples (Puerto Rico Genotyping Project) was extracted from saliva samples collected in a 20% ethanol solution. The three-day extraction protocol used DNAzol® reagent as the lysis buffer and ethanol for precipitation steps. The saliva samples were centrifuged at 5,000 rpm for 10 minutes and the supernatant was discarded. Next, DNAzol® and proteinase K were added to the pellet, and left at room temperature overnight. In the second day the solution was centrifuged at 13,600 rpm and the supernatant was transferred to a new 1.5mL microcentrifuge tube, and 100% solution of ethanol was added to make it 70% ethanol. After a 6 to 10 hour incubation period at -20°C, the solution was centrifuged again at 13,600 rpm, and the supernatant was discarded. The remaining pellet was washed twice with 70% solution of ethanol and left to dry overnight inside a biological hood. The extracted DNA was re-suspended in 1X TE and stored at -20ºC until needed. Samples reaching more than 50ng/ul and a 1.7-1.9 in the 260/280 ratio were used for genotyping.

Candidate SNPs genotyping

A total of 5 candidate genes were selected based on previous literature reports linking their respective SNPs variants to asthma. A summary of selected SNPs is shown in Table 1. Out of the 259 samples originally chosen we selected a total of 93 LGDS samples as our reference panel (Appendix Table 3) and 80 samples from our case cohort (see Appendix Table 2 for case cohort samples details).

Table 1. Summary of the Candidate SNPs for asthma. The SNPs were chosen based on previous literature reports showing association with asthma in Hispanic populations and in populations of European ancestry.

Gene	Alleles	rs#	Population	Reference

RAD50	G/T	rs2706347	Mexican Americans, Puerto Rican	Galanter *et al.*, 2011
ORMDL3	A/G	rs8076131	European, Mexican Americans, Puerto Rican	Galanter *et al.*, 2011, Ferreira *et al*,.2008, Moffatt *et al*,.2010
IL33	A/C	rs2381416	European, Mexican Americans, Puerto Rican	Galanter *et al*,.2011, Moffatt *et al*,.2010, Ferreiraet *et al*,.2011
HNMT	G/A	rs3100725	Puerto Ricans	Galanter *et al.*, 2011
GSDMB	C/T	rs2305479	European, Mexican Americans, Puerto Rican	Galanter *et al.*, 2011, Moffatt *et al*,.2010, Ferreiraet *et al*,.2011

Genotyping protocol

Genotyping was carried out using the TaqMan® Real Time ViiA7™ Genotyping System. The workflow required thermal cycling and fluorescent imaging of TaqMan® Assays that were set up online. Genotyping assays were ordered in the Applied Biosystems webpage (*https://www.lifetechnologies.com*). Each individual reaction was carried out on a 384 well plate, with 2.25µL of DNA (15ng/µL), 2.5µL of TaqMan® Universal PCR Master Mix (2X), and 0.25 µL of 20X Working stock of SNP assay for a total of 5.0µL. A real-time PCR was run with the following specifications: 10 min at 95°C followed by 50 cycles of 92°C for 15 sec and 60°C for 1.5 min, for denaturing and annealing/extension steps, respectively. The results were analyzed using the Sequence Detection System (SDS) software, provided by the company (ABI, Foster City, California). SDS uses the fluorescence measurements made during the plate read to plot fluorescence (Rn) values based on the signals from each well. The plotted fluorescence signals indicate which alleles are in each sample.

Statistical analyses of genetic associations:

Each SNP was tested for association to asthma correcting for the presence of all other SNPs as covariates. Each SNP or SNP combination was compared between those with and without asthma by performing minor allele frequency test, allelic association test, genotypic test and allele dominant associations. All statistical tests were performed in SAS (SAS 9.1, 2014, Cary N.C.).

Ancestry Informative Markers (AIMs)

For testing local ancestry, standard assays incorporating AIMs were screened on 24 samples simultaneously genotyping 128 unique probes on a single plate on the OpenArray® NT Imager (Life Technologies Inc.), providing high-throughput results for 72 samples at a time. AIMs allow estimation of Sub-Saharan African, European and Native American ancestry proportions of each sample. Of the 128 probes, we ran 96 SNPs previously used by Via *et al.* to estimate the ancestral contribution of our case-reference cohorts. A complete list of the SNPs and their description is provided in Appendix Table 4. In addition, to compare the ancestral admixture values in our samples, we re-analyzed two data sets from Via *et al.*, 2011 using STRUCTURE. The first set contained 108 non-admixed (contained less than 10% overlapping admixture) individuals in order to train the STRUCTURE analysis. The second data set compromises a total of 60 census-based individuals with an average ancestral value of 11.7%, 21.9%, and 66.4% for the Native American, African, and European contributions, respectively.

Statistical analyses of genetic associations with Individual ancestral estimates (IAEs)

Individual ancestral estimates (IAEs) were calculated using Bayesian clustering algorithms within STRUCTURE. This algorithm assumes a model in which there are K ancestral populations

(where K may be unknown), each of which is characterized by a set of allele frequencies at each locus. Individuals are assigned to populations on the basis of their genotypes, while population allele frequencies are estimated simultaneously. We used an admixture model for 20,000 burn-ins and 10,000 further iterations, assuming three ancestral populations (K=3, for African, Native American, and European), and assumed that the current population is observed 15 generations since the admixture event took place (450 years ago) as suggested earlier (Via *et al.*, 2010; Gravel *et al.*, 2013).

For analysis, we used the genotype of a total of 508 samples: 80 asthma samples, 259 reference population samples (LGDS cohort), 109 non-admixed and 60 admixed reference samples in order to train the STRUCTURE software program (37 samples were used as our African reference, 42 samples were used as European reference and 30 were used as Native American reference). The non-admixed ancestral reference individuals had less than 10% of other overlapping ancestral populations.

To test for differences in IAE between the case and control cohorts, we performed Type III General Linear Models (GLM) Analysis of Variance (ANOVA) using SAS® 9.1 statistical software (SAS 9.1, 2014, Cary N.C.). We compared the distributions of observed admixture proportions between individuals sampled as cases and as a reference under the null hypothesis that assumed no differences between the two samples.

Results

Part 1. Candidate gene genotyping results

(i) Statistical analyses of minor allele frequency

First, the minor allele frequencies were calculated for each SNP using the complete set of samples (Table 2). For all 5 SNPs none of the observed *p*-values showed significant differences between the cases and the references, and no positive association was observed for the assayed SNPs.

We then calculated the minor allele frequencies values of each SNP based on sample municipality origin. However, none of the SNPs reached statistical relevance between the case and reference cohort ($p > 0.05$), failing to reach any association in our sampled cohorts (Table 3). Under the premise that number of samples per municipality was not allowing us to establish proper statistical analysis due to the low number of individuals, we assigned a zone location for each municipality (East, West, South and North) and tested each SNP with the new groupings in order to raise the overall number of individuals per assayed group (Figure 2 and Table 4). Again, no significant differences were observed between the case-reference cohorts, as the p-value was higher than 0.05 for each of the assayed SNPs.

(ii) Allelic association tests

A single SNP located in *ORMDL3* showed a significant statistical difference between cases and references ($p=0.05$; Table 5). We did not confirm positive associations with asthma phenotypes for *RAD50* (rs2706347), *GSDMB* (rs2305479), *HNMT* (rs3100725) and *IL33* (rs2381416). Figure 5 shows the odds ratio and confidence interval values for all 5 assayed SNPs.

(iii) Genotypic association test results

In the genotypic association test *ORMDL3* contained the only SNP rs8076131 with a positive asthma association (p=0.029). AA genotype was present in 32 individual cases, AG in 39 individuals and only 5 individuals had GG genotype. In comparison the reference cohort there were 52, 26 and 6 individuals with AA, AG and GG genotypes respectively. *HNMT* (p-value=0.32), and other candidate gene SNPs (*RAD50* (rs2706347), *GSDMB* (rs2305479) and *IL33* (rs2381416) did not demonstrate significant differences between cases and references (see Table 6 and Table 7).

(iv) Minor and major allele dominant association tests

If the dominant model of inheritance was assumed for the minor allele (G), there was a significant difference between genotypes for *ORMDL3* rs807613 (p= 0.01): 42% (32 individuals) of the cases had AA genotype and 58% (44 individuals) had AG or GG genotype, whereas in reference 62% (52 individuals) of the references had AA genotype and 38% (32 individuals) had AG or GG genotype in the reference cohort. Other gene polymorphisms in RAD50, GSDMB, HNMT and IL33, did not show any statistical significance (p-values of 0.8959, 0.7823, 1.5037 and 0.9583 respectively (Table 7). On the other hand, when a major allele dominant model is assumed (using A as a dominant allele) there are no significant differences observed between the case and reference cohort.

(v) Ancestry associations

From our original AIMS panel a total of seven SNPs (rs2253624, rs9302185, rs798887, rs798887, rs10515919, rs10501474, rs423025, and rs9320808) failed to reach a 70% call-rate threshold (Figure 5) and were removed from the final analysis. Our final analysis consisted of 93-7 = 86 SNPs.

Our power analysis showed power for detecting association between ancestry proportions and asthma for two out of the three ancestries. With our current amount of samples power is sufficient ($\boldsymbol{\beta} \geq 0.80$ level) for the European sample set and marginally sufficient for the African set, but not for Native American sample set (Figure 9).

Complete Q1, Min, Median and Max values for each of the assigned clusters can be observed in the Appendix Figures 3-5, with no apparent differences between the case and reference cohorts.

However, using Analysis of Variance (ANOVA) we demonstrated that two of the ancestral populations have significantly different distributions of European and African ancestral components between asthma cases and references. Figure 6 shows STRUCTURE assigned a 4% higher European ancestry and a 2% lower African ancestry to asthma samples relative to reference samples. Although these differences seem meaningless in box plots (Figure 7), histogram plots suggest low European ancestry and high African ancestry are protective against asthma (Figure 8). The obtained p-values were highly significan (<0.0001 and <0.0056 for European and African ancestry, respectively. There were no statistical significant differences observed for the Native American component between our cases and references panels.

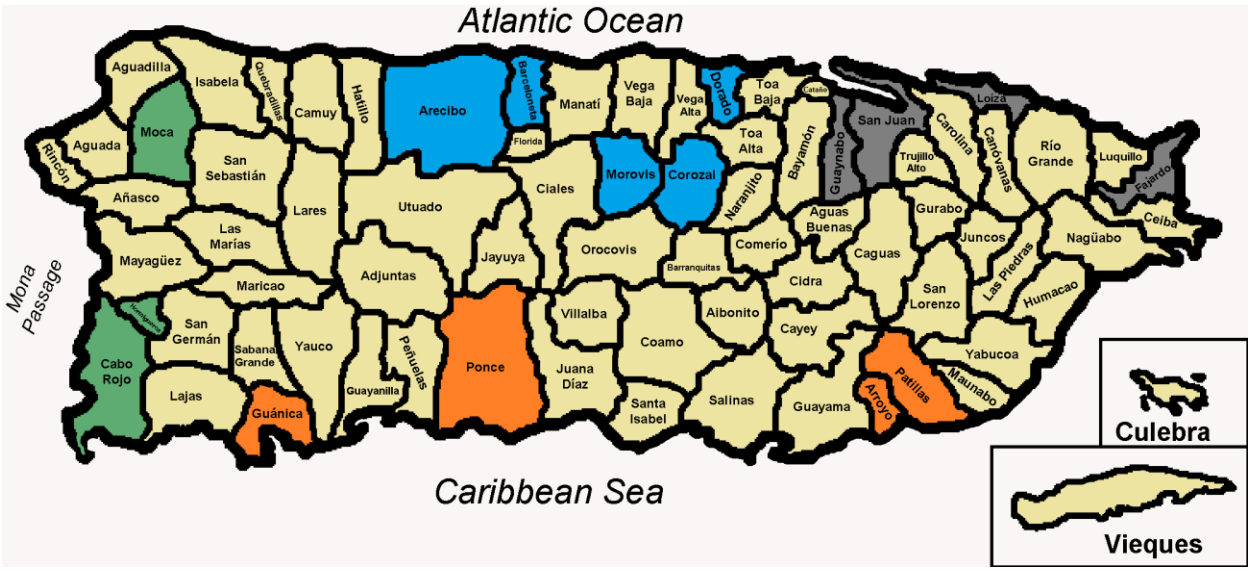
Table 2. Candidate allele frequencies values for each SNP in cases (n=80) and reference samples (n=93)

SNP	Chr.	Position* (bp)	Gene	MAF		X^2	p-value
				Reference	Cases		
rs8076131	17	39,924,659	*ORMDL3*	0.23	**0.32**	0.0408	0.840
rs2706347	5	132,569,425	*RAD50*	0.28	**0.28**	0.0001	0.991
rs2305479	17	39,905,964	*GSDMB*	0.33	**0.33**	1.2E-05	0.997
rs3100725	2	137,985,596	*HNMT*	0.32	**0.24**	0.018	0.894
rs2381416	9	6,193,455	*IL33*	0.35	**0.34**	1.2E-05	0.997

*SNP positions are from NCBI build 38

Table 3. Candidate gene minor allele frequency (MAF) results for each SNP by municipality for cases (n=80) and reference (93). Cases numeric values are represented in bold font.

		MAF					
		rs8076131 (*ORMDL3*)	**rs2706347 (*RAD50*)**	**rs2305479 (*GSDMB*)**	**rs3100725 (*HNMT*)**	**rs2381416 (*IL33*)**	**Number of samples**
Municipalities	Arecibo	0.25	0.17	0.17	0.17	0.40	3
	Arroyo	0.1	0.4	0.6	0.5	0.5	5
	Barceloneta	0.17	0.17	0.5	0.5	0.31	9
	Cabo Rojo	0.1	0.42	0.25	0.3	0.4	5
	Corozal	0.25	0	1	0.75	0.5	2
	Dorado	0.38	0.63	0.13	0.13	0.25	4
	Fajardo	0.5	0.1	0.4	0.4	0.3	5
	Guanica	0.25	0.39	0.28	0.28	0.11	9
	Guaynabo	0.33	0.1	0	0	0.5	6
	Hormigueros	0.5	0	0.6	0.6	0.75	5
	Loiza	0.06	0.13	0.25	0.25	0.44	8
	Moca	0.22	0.5	0.39	0.33	0.31	9
	Morovis	0.5	0.33	0.17	0.17	0.67	3
	Patillas	0.25	0.08	0.33	0.33	0.25	6
	Ponce	0	1	0.17	0.17	0	3
	San Juan	0.17	0.5	0.38	0.38	0.17	4
	Vieques	0.17	0.25	0	0	0.33	3
Reference Total		**0.23**	**0.28**	**0.33**	**0.32**	**0.35**	89
Asthma Cases		**0.32**	**0.28**	**0.33**	**0.24**	**0.34**	



Municipality designation	No. of samples	Minor allele frequencies				
		rs8076131 (*ORMDL3*)	rs2706347 (*RAD50*)	rs2305479 (*GSDMB*)	rs3100725 (*HNMT*)	rs2381416 (*IL33*)
North	20	0.28	0.26	0.32	0.29	0.39
East	21	0.19	0.33	0.35	0.33	0.17
West	20	0.22	0.35	0.42	0.39	0.44
South	21	0.25	0.17	0.24	0.15	0.38
Total Reference	**82**	**0.24**	**0.28**	**0.33**	**0.29**	**0.35**
Asthma Cases	**80**	**0.32**	**0.28**	**0.33**	**0.24**	**0.34**

Figure 2. Map of sample municipalities of Puerto Rico. Map shows assigned municipality designation for each of the 4 zones (Green=West, Orange=South, Blue=North and Grey=East). A total of 16 municipalities were used to choose the reference samples. Minor allele frequencies per zone for each SNPs are shown.

Table 4. Case vs. control association test results for each SNP by assigned municipality zone location (Figure 2).

Municipality location	rs8076131 (*ORMDL3*)		rs2706347 (*RAD50*)		rs2305479 (*GSDMB*)		rs3100725 (*HNMT*)		rs2381416 (*IL33*)	
	X^2	*p*-value	X^2	*p*-value	X^2	*p*-value	X^2	*p*-value	X^2	*p*-value
North	0.006	0.940	0.002	0.969	0.000	0.986	0.009	0.926	0.006	0.936
East	0.089	0.766	0.008	0.931	0.001	0.973	0.025	0.876	0.170	0.680
West	0.045	0.831	0.014	0.906	0.019	0.890	0.058	0.810	0.023	0.880
South	0.020	0.889	0.071	0.790	0.034	0.854	0.054	0.816	0.004	0.950

Table 5. Allelic Association test for all five candidate listed loci in our asthma cohort vs. general population in PR. X^2 test with *p*-values with less than 0.05 are highlighted in bold font.

	Observed	Genotype		X^2	p-value	O.R.	95% C.I.
rs8076131(*ORMDL3*)		A	G	3.729	**0.053**	0.6	0.374 -1.009
	Case	103	49				
	Reference	130	38				
rs2706347 (*RAD50*)		**G**	**T**	0.013	0.909	0.97	0.599 -1.579
	Case	112	44				
	Reference	123	47				
rs2305479 (*GSDMB*)		**A**	**G**	0.002	0.964	0.99	0.618- 1.582
	Case	97	47				
	Reference	119	57				
rs3100725 (*HNMT*)		**A**	**G**	1.483	0.223	1.35	0.831 - 2.206
	Case	114	36				
	Reference	131	56				
rs2381416 (*IL33*)		**A**	**C**	0.001	0.975	1.01	0.654- 1.547
	Case	139	73				
	Reference	106	56				

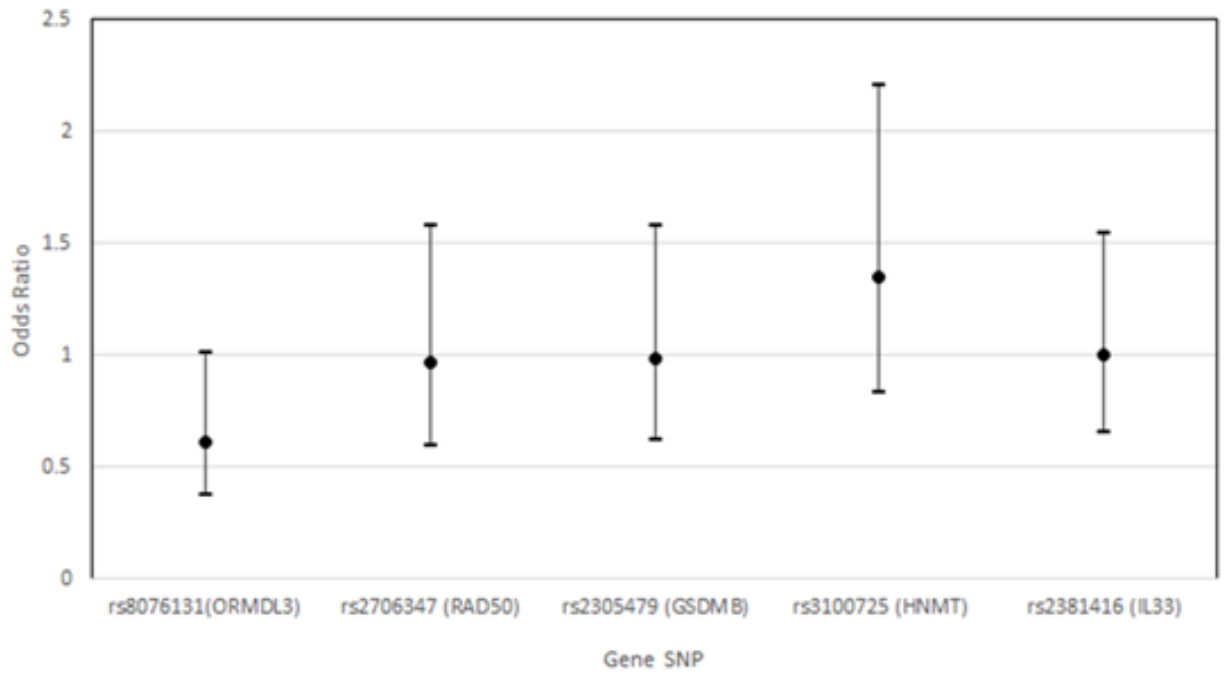


Figure 3. SNP-based replication showing ORs (95% CI) for the assayed samples. OR values are as follow: *ORMDL3*=0.614, *RAD50*=0.972, *GSDMB*=0.989, *HNMT*=1.35 and *IL33*=1.01.

Table 6. Genotypic tests for all five listed candidate loci in our asthma cohort vs general population in Puerto Rico. X^2 test with *p*-values with less than 0.05 are shown in bold font. Percentage values for each genotype is shown next to the number of calls for cases and controls

	Genotype	**Cases**	**References**	***X²***	***P-value***
rs8076131(*ORMDL3*)	A/A A/G G/G	32 (42%) 39 (51%) 5 (7%)	52 (62%) 26 (31%) 6 (7%)	7.07	**0.03**
rs2706347 (*RAD50*)	G/G G/T T/T	41 (53%) 30 (38%) 7 (9%)	47 (55%) 29 (34%) 9 (11%)	0.376	0.83
rs2305479 (*GSDMB*)	A/A A/G G/G	30 (42%) 37 (51%) 5 (7%)	42 (48%) 35 (40%) 11 (13%)	2.73	0.26
rs3100725 (*HNMT*)	C/C C/T T/T	43 (57%) 28 (37%) 4 (5%)	42 (47%) 38 (43%) 9 (10%)	2.271	0.32
rs2381416 (*IL33*)	A/A A/C C/C	46 (43%) 47 (44%) 13 (12%)	36 (44%) 34 (42%) 11 (14%)	0.133	0.94

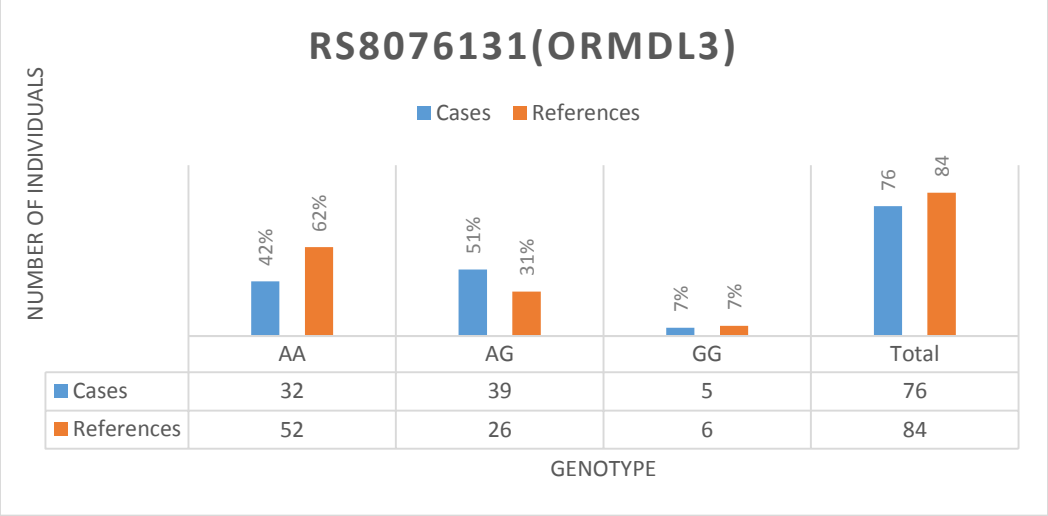


Figure 4. Genotype frequencies result for rs8076131 (*ORMDL3*) involved in asthma in Puerto Ricans. (Blue = cases; orange = controls). $p = 0.02916$

Table 7. Minor allele (dominant) tests for candidate listed loci in asthma cohort vs general population. X^2 test with *p*-values with less than 0.05 are shown in bold font. Percentage values for selected genotypes are shown next to the numbers of calls for cases and controls.

	Genotype	Cases	Controls	X^2	*p*-value	O.R.	95% C.I.
rs8076131(*ORMDL3*)	A/A A/G + G/G	32 (42%) 44 (58%)	52 (62%) 32 (38%)	6.272	**0.012**	0.448	0.028-0.843
rs2706347 (*RAD50*)	G/G G/T + T/T	41 (53%) 37 (47%)	47 (55%) 38 (45%)	0.122	0.727	0.896	0.484-1.66
rs2305479 (*GSDMB*)	A/A A/G + G/G	30 (42%) 42 (58%)	42 (48%) 46 (52%)	0.588	0.443	0.782	0.417-1.466
rs3100725 (*HNMT*)	C/C C/T + T/T	43 (57%) 32 (43%)	42 (47%) 47 (53%)	1.677	0.195	1.504	0.810-2.791
rs2381416 (*IL33*)	A/A A/C + C/C	46 (43%) 60 (57%)	36 (44%) 45 (56%)	0.02	0.888	0.958	0.535-1.716

Table 8. Dominant model test (major allele) for candidate loci association in asthma cases vs general population. X^2 test with p-values with less than 0.05 are shown in bold font.

	Genotype	**Cases**	**Controls**	***X²***	***P value***	**OR**	**95% CI**
rs8076131(*ORMDL3*)	A/A + A/G G/G	71 (93%) 5 (7%)	78 (93%) 6 (7%)	0.02	0.88754	1.0923	0.3194 -3.7356
rs2706347 (*RAD50*)	G/G + G/T T/T	71 (89%) 7 (9%)	76 (89%) 9 (11%)	0.12	0.72903	1.2011	0.4248 - 3.3961
rs2305479 (*GSDMB*)	A/A + A/G G/G	67 (93%) 5 (7%)	77 (88%) 11 (13%)	1.358	0.24388	1.9143	0.633 - 5.7895
rs3100725 (*HNMT*)	C/C + C/T T/T	71 (95%) 4 (5%)	80 (90%) 9 (10%)	1.274	0.25902	1.9969	0.5893 - 6.7665
rs2381416 (*IL33*)	A/A +A/C C/C	93 (88%) 13 (12%)	70 (86%) 11 (14%)	0.071	0.78989	1.1242	0.4753 - 2.6587

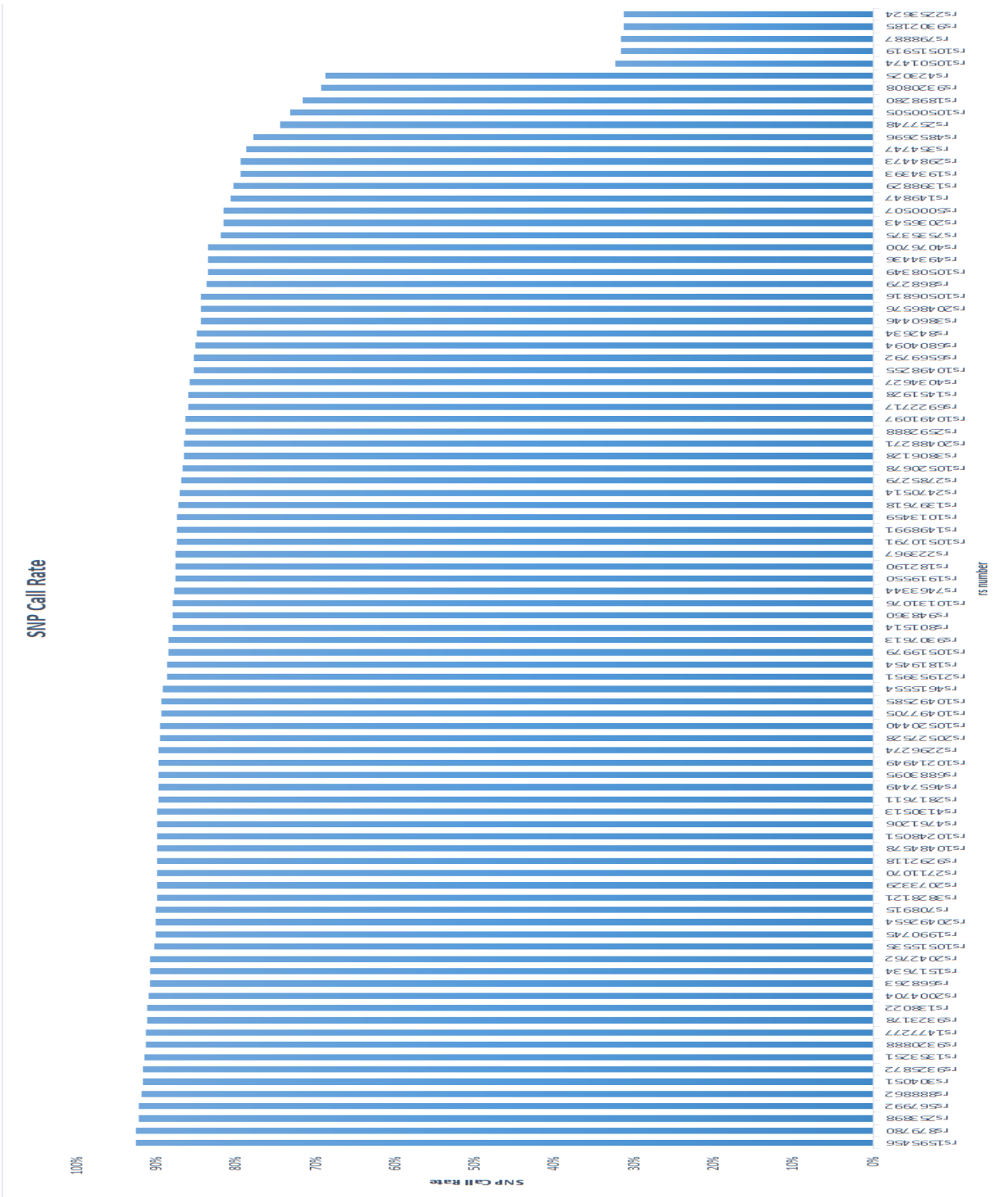
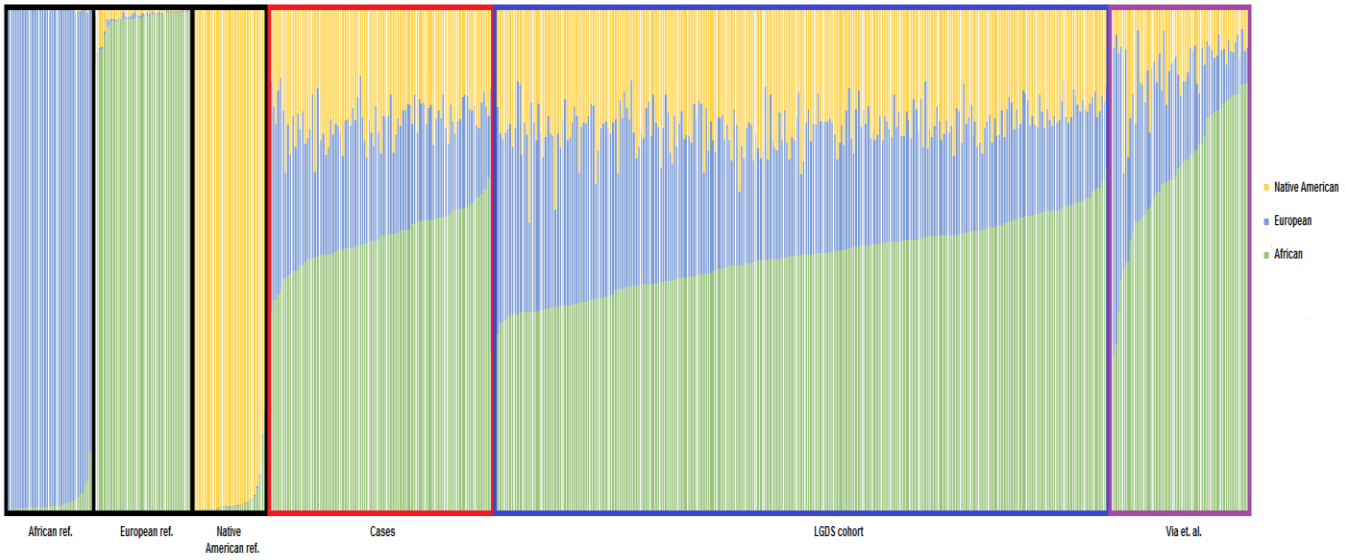


Figure 5. List of SNPs call rate distribution in the STRUCTURE run. A total of 93 SNPs were used in our ancestry panel. Seven variants resulted in values lower than 70 % call rate. The followings SNPs did not achieved a call rate higher than 70% and thus were discarded in the final analysis: rs9320808, rs423025, rs105919, rs1051474, rs798887, rs9302185 and rs2253624



Sample set	European	African	Native American
African reference*	2%	98%	0%
European reference*	97%	1%	2%
Native American reference*	3%	0%	97%
Asthma cases	54%	24%	22%
Reference samples	50%	26%	23%
Puerto Ricans*	66%	22%	12%

* from Via at al, 2011

Figure 6. STRUCTURE bar plot representing admixture fractions in different population datasets. Black line frame set for African, European and Native American reference ancestral panel. Red, blue and purple line frame set are for cases, reference individuals and Via *et al.* (2011) cohorts

respectively. Average values for each population are shown in the Table above.

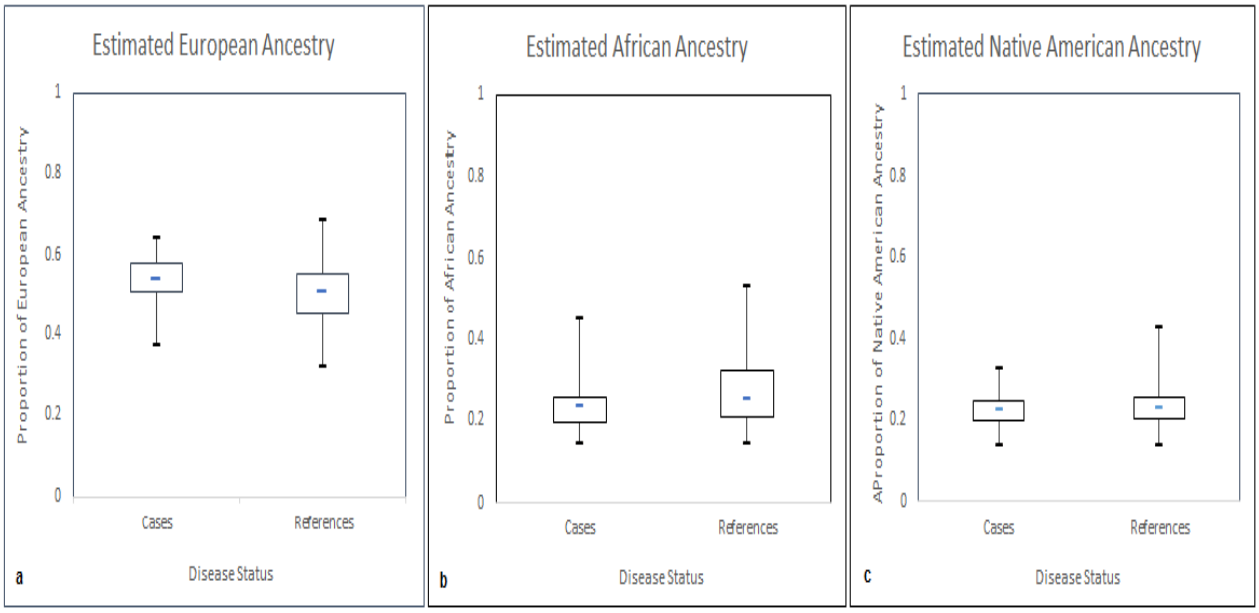


Figure 7. Box plots for the ancestry estimates a) European ancestry component. b) African ancestry. c) Native American ancestral component. Observed *p*-values for European, African and Native American ancestral estimates were 0.9612, 0.9608 and 0.9849 respectively.

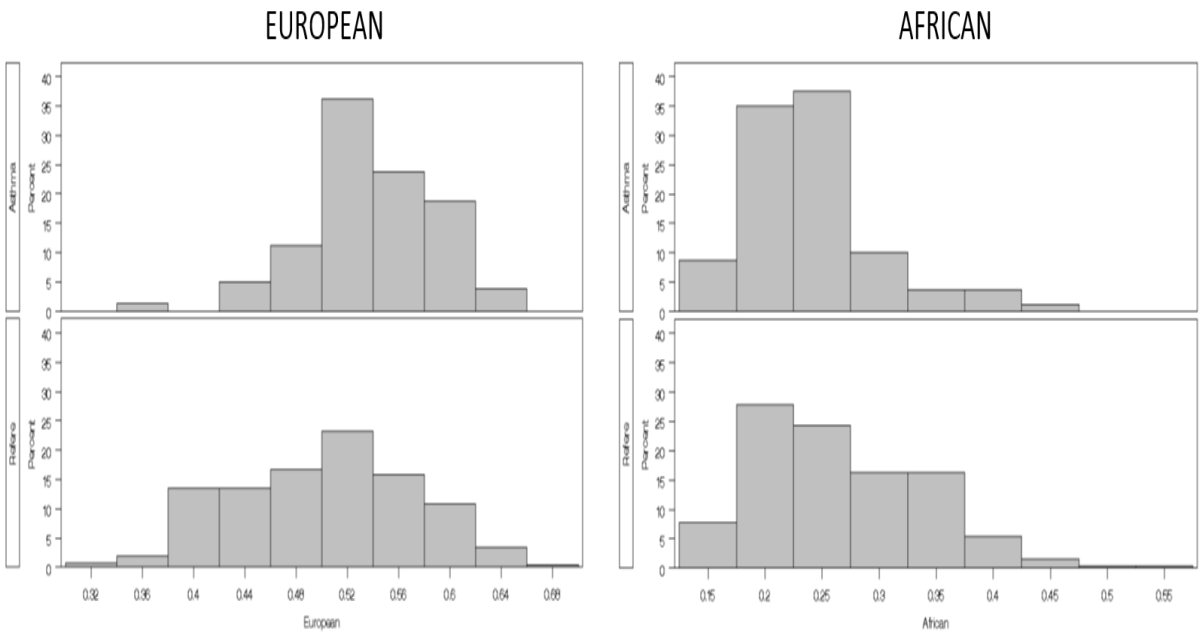
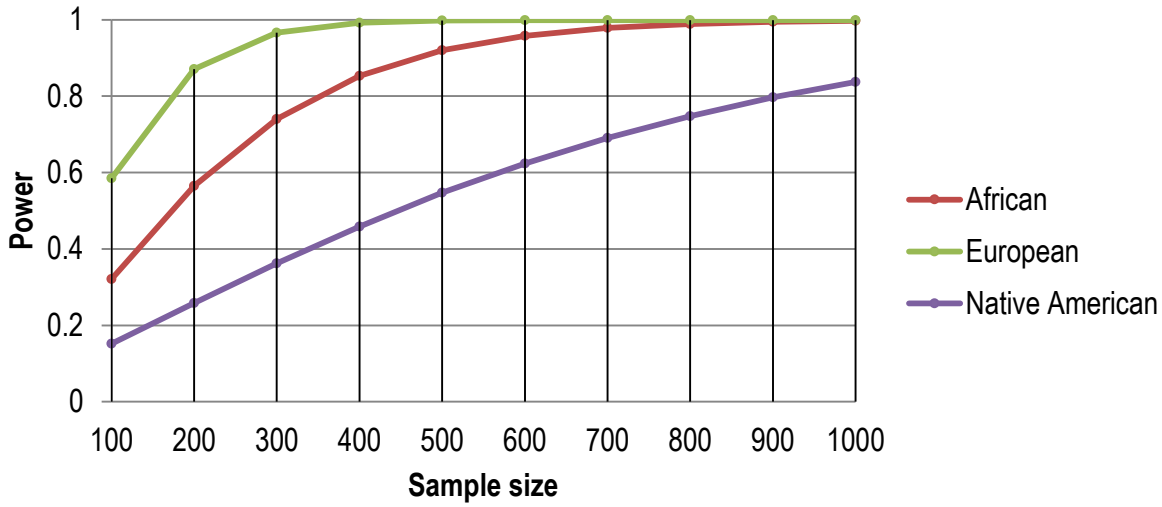


Figure 8. Histogram distributions showing the observed European and African ancestral contributions in asthma cases and the reference population (LGDS cohort). Mean values for African Asthma cases and African references are 0.24 and 0.26 respectively. Mean values for European Asthma cases and European references are 0.54 and 0.50 respectively. Standard Deviation values for African Asthma cases and African references 0.06 and 0.07 respectively. Standard deviation values for European Asthma cases and European references are 0.53 and 0.07 respectively.



POWER ANALYSIS: Power for the current test

Sample Size	African	European	Native American
341	0.793	0.981	0.403

Figure 9. Power analysis for detecting association between ancestry proportions and asthma for three ancestries. Current power is sufficient ($\beta>0.80$) for African and European samples, but not for Native American sample ($\beta=0.4$)

Discussion

In this study, we used a locally collected set of Puerto Rican samples to test the most common candidate gene variants for association with asthma. Out of the five candidate variants, only rs8076131 in the *ORMDL3* gene showed significant association with asthma phenotype compared to the geographically collected reference (LGDS cohort, Figure 1). rs8076131 is found in the intron that precedes the starting codon exon in *ORMDL3.* This observation suggests that rs8076131 might play a role in intron removal, consequently affecting the effective exportation of the *ORMDL3* messenger from the nucleus. Alternatively, it could be affecting the transcription promoter of the gene. In addition, after a closer look, we realized that rs8076131 is found >2,500 base pairs upstream to an uncharacterized gene. This proximity could influence a regulatory region for the unknown gene expression, and thus we further suggest a closer future examination.

While, in our allelic association test *ORMDL3* resulted in a suggestive, but marginally significant association with asthma (p = 0.05), in the genotypic association test, we observed significant differences between the asthma case and the reference samples (p = 0.029; Table 6). In addition, assuming that a protective AA genotype is recessive (AA vs. AG+GG), the effect seemed to increase (p = 0.012, Table 7). Assuming dominance in this case (AA+AG vs. GG) would render the difference insignificant (p = 088; Table 8). Therefore, we assumed that the most likely model scenario for risk allele exposure is one in which the minor allele is present as heterozygous and homozygous (dominant model, Table 7).

The rest of the assayed SNPs did not show any associations, and thus no differences were observed between the case and reference panel cohorts. Our results also demonstrate a higher European ancestral component in asthma samples in comparison to the reference (Figures 9 and 10). Since it has been observed that rs8076131 - G is more frequent in Europeans than in other

worldwide populations (Appendix Figure 6), this observation is consistent with a true association between the rs8076131-G allele and asthma.

Previous ethnic-specific studies showed *ORMDL3* on chromosome 17q12 a genetic risk factor for Latino (Puerto Rican and Mexican) and Australian populations (Galanter *et al.,* 2011 and Ferreira *et al.,* 2011). Furthermore, *ORMDL3* has been previously linked with an increased risk for childhood asthma and early asthma onset (Ono *et al.,* 2013). The protein encoded by *ORMDL3* has been previously reported (Cantero-Recasens *et al.,* 2009 and Ferreira *et al.,* 2011) to be associated to asthma. Increased *ORMDL3* expression alters Endoplasmic Reticulum-mediated Ca(2+) homeostasis and facilitates unfolded-protein response, a process that is known to induce endogenous inflammation and thus highly associated with pro-inflammatory diseases, making it a strong candidate and an asthma genetic risk factor in the Puerto Rican population.

However, it is worth mentioning that although our results point to an *ORMDL3*-asthma association, precaution should be employed. Given that our reference panel showed a discrepancy with Via. *et al.* data set for the European ancestral component (50% vs. 64% respectively), caution should be taken when comparing the two data sets (asthma cases vs. references). There is still a chance that the observed differences are the results of sporadic association as a result of sample bias in our reference cohort. Consequently, our asthma individuals could potentially be showing a European-asthma association due to the lack of European component representation in our reference panel. On the other hand, our results suggest a reduction in European ancestry in asthmatic individuals, in turn favored by an increase in Native American ancestry (15% Via *et al.* vs 22% in our case cohort). However, this scenario is unlikely given the robust sample set employed in Via *et al.* in comparison to our modest sample size. In order the further analyze this scenario, we evaluated the allelic frequency of the alternative allele in all 3 ancestral populations

for the five candidate SNPs using Ensembl database for population genetics. However, we could not draw any conclusion since the major allelic frequency observed for the candidate genes many of the time overlaps with the ancestral population. For example, re2706347 which showed similar distribution in both cohorts are quite similar in the European and American population (20% and 19% respectively). For rs2305479 (*GSDMB*), the obtained allelic results seem to favor an African ancestral component since the observed allele is more prevalent in African individuals (87% Africans, 62% Americans, and 49% European). On the other hand, rs3100725 (*HNMT*) is slightly more common in American individuals in comparison to African and Europeans (31%, 30% and 22% respectively). Whereas rs2381416 (*IL33*) is more common in Americans in comparison to Europeans (77% vs 76% respectively). And finally, for rs8076131 (*ORMDL3*) the observed major allele frequency in our samples seems to favor a European component whereas in our references the observed frequency favors the observed frequency in African individuals. This information suggests the possible scenario in which the obtained results will closely resemble the population with a similar allele frequency for the observed allele. In this sense the observed allele frequency in our asthma samples for *ORMDL3* could match either a population from European or Native American origin. Whereas the reference panel suggest an African population. Again, these results should be taken with great care, since it is difficult to infer average values given the modest set of candidate SNPs assayed.

In addition to *ORMDL3*, it is worth mentioning that among the reported variants, some showed slight trends that could be explored given a larger sample size. For instance, in our work *HNMT* (rs3100725) showed a trend among assayed SNPs with *p*-values of 0.223 and 0.321 for our allelic and genotypic test respectively. *HNMT* have been reported in multiple occasion as a possible candidate SNP in several asthma studies (Szczepankiewicz *et al.*, 2010) including an ethnic

specific study performed in a Puerto Rican population reported in Galanter *et al.*, 2011. Since our observation could have been the result of the limited number of samples, it is possible that the increased number of samples could reveal *HNMT* association with asthma. In addition, it is important to point out that our reference panel is composed of random unknown samples that represent the overall Puerto Rican population, and we do not know the asthma diagnosis status (if any) of these individuals. There is a high chance that approximately 14% of the assayed reference individuals would have an asthma phenotype. Therefore, there is a significant chance that we are underestimating the association of any of the SNPs assayed and asthma. For these reason we recommend further testing increasing the number of reference individuals or establish a non-asthma control panel in order to increase analysis power.

To determine the average ancestral contribution in our Puerto Rican asthma and control cohort we used a Bayesian statistical approach in STRUCTURE (Pritchard *et al.*, 2000). This analysis calculates admixture components of the assayed populations given the known number of sources (or K)(Figure 6). The average percentages for the three main ancestral components for our asthma samples were 54%, 24% and 22% for the European, African and Native American respectively, whereas in our reference cohort the observed ancestral components was 50%, 26% and 23% (Figure 6). The analysis of variance demonstrates a higher European ancestral component trend in asthma individuals in comparison to the reference cohort with average European ancestral values of 54% for the case cohort and 50% for the reference cohort. The opposite pattern was observed for the African ancestral component: a significantly higher contribution was observed for the reference cohort in comparison to the case cohort with average ancestral values of 26% for the reference cohort and 24% for the case cohort (Figures 8, 9 and 10). However, in case of the Native American ancestral component we could not detect significant differences between the

groups, as more than 900 samples would be required to achieve sufficient statistical power (Figure 9).

It is important to mention that our results are consistent with the previously reported by Choudhry *et al.*, (2008), in which the observed European ancestral contribution was higher in their case cohort than in their control cohort. Furthermore, our candidate gene results for the *ORMDL3* variant coincide with an increased European ancestry contribution, since the observed genotype is more likely observed in European populations (see Ensembl population data for rs8076131 from 1000 Genomes data; Appendix Figure 6). Our results suggest that *ORMDL3* is a significant contributor to the asthma phenotype in Puerto Ricans.

It is interesting that CDC reports lower levels of asthma for whites than for Hispanics in Puerto Rico (3.6 vs 5.2%) (Behavioral Risk Factor Surveillance System (BRFSS), 2008. *http://www.cdc.gov/asthma/stateprofiles/asthma_in_pr.pdf*). Since there is hardly a defined difference between the two groups on the island, this is probably due to the uncertainty in the ethnicity definition, and may be due primarily to the socioeconomic status (Via *et.al.,* 2011).

As a result our work suggests a positive link between European ancestry contribution and asthma for the Puerto Rican population. It is worth mentioning that although our dataset was somewhat modest, our power analysis indicates positive statistical power for the current dataset, achieving >80% power for the European and close to 80% power for the African case-reference comparisons. On the other hand, for the Native American component comparison we were unable to achieve statistical power since our power analysis indicates that it will take more than 900 samples to have seen the statistically significant effects for the Native American contribution.

While the ancestry difference estimates are robust for the cases vs. reference comparison, since they were genotyped on the same OpenArray plates, the overall ancestry proportions in our study seem to overestimate the Native American component. We compared the ancestral estimated values of our case-reference samples by independently calculating ancestries in Via *et al.* (2011): 66%, 22%, and 12% vs. 63.7%, 21.2%, and 15.2 for the European, African and Native American respectively. The observed differences between the estimates are not likely to be due to the difference in calculations, but can be explained by the lower call rate threshold of 70% in our study. It is possible that the SNPs that failed to amplify contained information that would improve our estimate. While this does not affect the aims established for this study (since cases and reference are genotyped using the same loci), if this data is used to report the ancestry proportions in Puerto Rican population, and if additional funds are available, the seven failed assays should be individually genotyped and incorporated in the analysis.

It is important to mention that although we could find differences in the ancestral components between asthma cases and reference population, it is still possible that increasing the sample size in this study could lead to the discovery of more robust associations possibly missed in our study. In addition, even though there were significant differences between cases and the references, the estimates used to determine ancestral contribution were average values. These values do not represent regional variation that might harbor asthma related loci and that are being contributed by different ancestral populations in the unique local admixture context. This possibly could have been missed in our study, and having a misleading effect on the risk loci ancestral contribution determination. In order to account for this possibility, we suggest that a larger study is performed using a higher number of cases and controls, each representing different parts of the

island and substantially increasing the number of genetic markers as this would allow to target ancestral contribution loci specific sites.

Conclusions

In this work we investigated the association of candidate gene variants and explored the ancestral admixture background component to evaluate a role of genetic component in asthma morbidity in Puerto Rico. Our results demonstrate a positive association between *ORMDL3* (rs8076131) and asthma, showing a higher SNP call rate for *ORMDL3* in Puerto Rican asthma individuals. Furthermore, there was strong indication of higher European ancestral contribution in asthma cases, consistent with previous reported literature.

Our results were consistent in validating a clear relation between *ORMDL3* (rs8076131) and asthma prevalence rate among Puerto Ricans, and also further established an association between European ancestral contribution and asthma morbidity among Puerto Ricans. We did not have sufficient power to observe a substantial Native American ancestral differences between our case (asthma phenotype individuals) and our reference panel. Our work underlines the importance of population genetic characterization in order to fully capture the elements involved in asthma pathology on the island.

Recommendations

In order to assess the exact ancestral contribution at a gene resolution level it is imperative to perform ancestry association test with a significant higher number of markers in order to fully capture the specific gene contribution. Furthermore, to obtain a better representation of the average admixture values average it is important to increase the total number of samples. In addition, it is worth mention that the number of candidate SNPs used in this study is very modest, thus increasing the number of asthma candidate genes could allow for a more robust association between the assayed samples. Given that our results are consistent with an association between *ORMDL3* and asthma, we suggest further characterization of this gene. Finally, since asthma is characterized for being highly heterogeneous, we strongly recommend future studies that takes in consideration confounding effects such as social economic status, geography, etc. This evaluation could help explain the observed asthma heterogeneity.

References

1. Amy Murphy, Sungho Won, Angela Rogers, et. al, (2010). On the genome-wide analysis of copy number variants in family based designs: Methods for combining family-based and population based information for testing dichotomous or quantitative traits, or completely ascertained samples. Genet Epidemiol, 34(6), 582–590. doi:10.1002/gepi.20515.On
2. Barnett, S. B. L., & Nurmagambetov, T., (2011). Costs of asthma in the United States: 2002-2007. The Journal of allergy and clinical immunology, 127(1), 145–52. doi:10.1016/j.jaci.2010.10.020
3. Bloom, D.E., Cafiero, E.T., Jané-Llopis, E., et. al, (2011). The Global Economic Burden of Noncommunicable Diseases. Geneva: World Economic Forum. A report by the World Economic Forum and the Harvard School of Public Health, (September).
4. Bossé, Y., & Hudson, T. J. (2007). Toward a comprehensive set of asthma susceptibility genes. Annual review of medicine, 58, 171–84. doi:10.1146/annurev.med.58.071105.111738
5. Bousquet, J., Bousquet, P. J., Godard, P., & Daures, J., (2005). Public Health Reviews The public health implications of asthma. Bulletin of the World Health Organization, 83(7), 548–554.
6. Cantero-Recasens, G., Fandos, C., Rubio-Moscardo, F., Valverde, M. a, & Vicente, R. (2010). The asthma-associated ORMDL3 gene product regulates endoplasmic reticulum-mediated calcium signaling and cellular stress. Human Molecular Genetics, 19(1), 111–21. doi:10.1093/hmg/ddp471
7. Choudhry, S., Taub, M., Mei, R., Rodriguez-santana, J., et. al, (2008). Genome-wide Screen for Asthma in Puerto Ricans: Evidence for Association with 5q23 Region. Hum Genet, 123(5), 455–468. doi:10.1007/s00439-008-0495-7.Genome-wide
8. Craddock, N., Hurles, M. E., Cardin, N., Pearson, R. D., et. al, (2010). Genome-wide association study of copy number variation in 16,000 cases of eight common diseases and 3,000 shared controls. Nature, 464(7289), 713–720. doi:10.1038/nature08979.Genome-wide
9. Daley, D., Park, J. E., He, J.-Q., Yan, J., Akhabir, L., et. al, (2012). Associations and interactions of genetic polymorphisms in innate immunity genes with early viral infections and susceptibility to asthma and asthma-related phenotypes. The Journal of allergy and clinical immunology, 130(6), 1284–93. doi:10.1016/j.jaci.2012.07.051
10. Daley, D., Park, J. E., He, J.-Q., Yan, J., Akhabir, L., Stefanowicz, D., … Sandford, A. J. (2012). Associations and interactions of genetic polymorphisms in innate immunity genes with early viral infections and susceptibility to asthma and asthma-related phenotypes. The Journal of Allergy and Clinical Immunology, 130(6), 1284–93. doi:10.1016/j.jaci.2012.07.051
11. Deindl, P., Peri-Jerkan, S., Deichmann, K., Niggemann, B., Lau, S., Sommerfeld, C., Sengler, C., Müller, S., Wahn, U., Nickel, R., Heinzmann, A. and the German Multicenter Atopy Study Group (2005), No association of histamine- N-

methyltransferase polymorphism with asthma or bronchial hyperresponsiveness in two German pediatric populations. Pediatric Allergy and Immunology, 16: 40–42. doi: 10.1111/j.1399-3038.2005.00218.x

12. Ferreira, M. a R., McRae, A. F., Medland, S. E., Nyholt, et. al, (2011). Association between ORMDL3, IL1RL1 and a deletion on chromosome 17q21 with asthma risk in Australia. European journal of human genetics : EJHG, 19(4), 458–64. doi:10.1038/ejhg.2010.191
13. Galanter, J. M., Torgerson, D., Gignoux, C. R., Sen, S., Roth, L., *et al.*, (2011). Cosmopolitan and ethnic-specific replication of genetic risk factors for asthma in 2 Latino populations. The Journal of allergy and clinical immunology, 128(1), 37–43.e12. doi:10.1016/j.jaci.2011.03.050
14. Galanter, J., Choudhry, S., Eng, C., Nazario, S., et. al, (2008). ORMDL3 gene is associated with asthma in three ethnically diverse populations. American journal of respiratory and critical care medicine, 177(11), 1194–200. doi:10.1164/rccm.200711-1644OC
15. Gaudieri, S., Lucas, M., Lucas, A., McKinnon, E., Albloushi, et. al, (2012). Genetic variations in IL28B and allergic disease in children. PloS one, 7(1), e30607. doi:10.1371/journal.pone.0030607
16. Gaudieri, S., Lucas, M., Lucas, A., McKinnon, E., Albloushi, H., Rauch, A., … Tulic, M. K. (2012). Genetic variations in IL28B and allergic disease in children. PloS One, 7(1), e30607. doi:10.1371/journal.pone.0030607
17. Genovese, G., Friedman, D. J., Ross, M. D., Lecordier, L., et. al, (2010). Association of Trypanolytic ApoL1 Variants with Kidney Disease in African-Americans. Science, 329(5993), 841–845. doi:10.1126/science.1193032.Association
18. Gordon J. Freeman1, Jose M. Casasnovas2, Dale T. Umetsu3, and R. H. D. (2010). TIM genes: a family of cell surface phosphatidylserine receptors that regulate innate and adaptive immunity. Immunol Rev, 235(1), 172–189. doi:10.1111/j.0105-2896.2010.00903.x.TIM
19. Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J. K., Muzzio, M., Rodriguez-Flores, J. L., … Bustamante, C. D. (2013). Reconstructing Native American migrations from whole-genome and whole-exome data. PLoS Genetics, 9(12), e1004023. doi:10.1371/journal.pgen.1004023
20. Holtzman, M. J. (2012). Review series Asthma as a chronic disease of the innate and adaptive immune systems responding to viruses and allergens. The Journal of Clinical Investigation, 122(8), 2741–2748. doi:10.1172/JCI60325.is
21. Hsu, K. J., & Turvey, S. E. (2013). Functional analysis of the impact of ORMDL3 expression on inflammation and activation of the unfolded protein response in human airway epithelial cells. Allergy, asthma, and clinical immunology : official journal of the

Canadian Society of Allergy and Clinical Immunology, 9(1), 4. doi:10.1186/1710-1492-9-4
22. Jennie G. Ono, Tilla S. Worgall & Stefan Worgall. 17q21 locus and ORMDL3: an increased risk for childhood asthma. Pediatric Research (2014) 75, 165–170 doi:10.1038/pr.2013.186.
23. Kathleen C. Barnes, P. (2010). Genomewide association studies in allergy and the influence of ethnicity. Curr Opin Allergy Clin Immunol, 10(5), 427–433. doi:10.1097/ACI.0b013e32833de6ee.Genomewide
24. Kim, H. Y., DeKruyff, R. H., & Umetsu, D. T. (2010). The many paths to asthma: phenotype shaped by innate and adaptive immunity. Nature immunology, 11(7), 577–84. doi:10.1038/ni.1892
25. Knowler, W. C., Pettitt, D. J., Savage, P. J., & Bennett, P. H. (1981). Diabetes incidence in Pima indians: contributions of obesity and parental diabetes. Am J Epidemiol, 113(2), 144-156.
26. Lazarus, R., Raby, B. a, Lange, C., Silverman, E. K., Kwiatkowski, D. J., et. al, (2004). TOLL-like receptor 10 genetic variation is associated with asthma in two independent samples. American journal of respiratory and critical care medicine, 170(6), 594–600. doi:10.1164/rccm.200404-491OC
27. Manolio, T. a, Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L., et. al, (2009). Finding the missing heritability of complex diseases. Nature, 461(7265), 747–53. doi:10.1038/nature08494
28. Martinez-Cruzado JC, Toro-Labrador G, Viera-Vera J, Rivera-Vega MY, Startek J, Latorre-Esteves M, Roman-Colon A, Rivera-Torres R, Navarro-Millan IY, Gomez-Sanchez E *et al.*: Reconstructing the population history of Puerto Rico by means of mtDNA phylogeographic analysis. Am J Phys Anthropol 2005, 128(1):131-155
29. Ming-liang, G. U., Jing, Z., Hapmap, I., Yokouchi, S., & Caucasian, S. (2011). Mapping and localization of susceptible genes in asthma. Chin Med, 124(1), 132–143. doi:10.3760/cma.j.issn.0366-6999.2011.01.024
30. Moffatt, M. F., Gut, I. G., Demenais, F., Strachan, D. P., Bouzigon, E., et. al, (2010). A large-scale, consortium-based genomewide association study of asthma. The New England journal of medicine, 363(13), 1211–21. doi:10.1056/NEJMoa0906312
31. Nguyen, E. a, & Burchard, E. G. (2012). Asthma Research for All of the United States. Pediatric allergy, immunology, and pulmonology, 25(3), 128–131. doi:10.1089/ped.2012.0173
32. Puerto Rico Asthma Project (2009). Puerto Rico Asthma Surveillance Report.
33. Rogers, A. J., Raby, B. a, Lasky-Su, J. a, Murphy, A., Lazarus, R., et. al, (2009). Assessing the reproducibility of asthma candidate gene associations, using genome-wide data. American journal of respiratory and critical care medicine, 179(12), 1084–90. doi:10.1164/rccm.200812-1860OC

34. SAS Institute Inc. 2011. Base SAS® 9.3 Procedures Guide. Cary, NC: SAS Institute Inc. Base SAS® 9.3 Procedures Guide Copyright © 2011, SAS Institute Inc., Cary, NC, USA ISBN 978-1-60764-895-6 All rights reserved. Produced in the United States of America.
35. Sasaki Y1, Ihara K, Ahmed S, Yamawaki K, Kusuhara K, Nakayama H, Nishima S, Hara T. Lack of association between atopic asthma and polymorphisms of the histamine H1 receptor, histamine H2 receptor, and histamine N-methyltransferase genes. Immunogenetics. 2000 Mar;51(3):238-40.
36. Sharma S, Mann D.,, Singh T.P, Ghosh B. Lack of association of histamine-N-methyltransferase (HNMT) polymorphism with asthma in the Indian population. J Hum Genet (2005) 50:611-617
37. Smith, M. W., & O'Brien, S. J. (2005). Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. Nature reviews. Genetics, 6(8), 623–32. doi:10.1038/nrg1657
38. Sullivan, P. W., Ghushchyan, V. H., Slejko, J. F., Belozeroff, V., Globe, D. R., & Lin, S.-L. (2011). The burden of adult asthma in the United States: evidence from the Medical Expenditure Panel Survey. The Journal of allergy and clinical immunology, 127(2), 363–369.e1–3. doi:10.1016/j.jaci.2010.10.042
39. Szczepankiewicz A, Bręborowicz A, Sobkowiak P, Popiel A. Polymorphisms of two histamine-metabolizing enzymes genes and childhood allergic asthma: a case control study. Clin Mol Allergy. 2010 Nov 1;8:14. doi: 10.1186/1476-7961-8-14.

40. Neeta Thakur, Sam S. Oh, Elizabeth A. Nguyen, Melissa Martin, Lindsey A. Roth, Joshua Galanter, Christopher R. Gignoux, Celeste Eng, Adam Davis, Kelley Meade, Michael A. LeNoir, Pedro C. Avila, Harold J. Farber, Denise Serebrisky, Emerita Brigino-Buenaventura, William Rodriguez-Cintron, Rajesh Kumar, L. Keoki Williams, Kirsten Bibbins-Domingo, Shannon Thyne, Saunak Sen, Jose R. Rodriguez-Santana, Luisa N. Borrell, and Esteban G. Burchard. (2013) **Socioeconomic Status and Childhood Asthma in Urban Minority Youths. The GALA II and SAGE II Studies.** American Journal of Respiratory and Critical Care Medicine 188:10, 1202-1209

41. Torgerson, D. G., Ampleford, E. J., Chiu, G. Y., Gauderman, W. J., Gignoux, C. R., Graves, et. al, (2011). Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nature genetics, 43(9), 887–92. doi:10.1038/ng.888
42. Torgerson, D. G., Capurso, D., Mathias, R. a, Graves, P. E., Hernandez, R. D., et. al, (2012). Resequencing candidate genes implicates rare variants in asthma susceptibility. American journal of human genetics, 90(2), 273–81. doi:10.1016/j.ajhg.2012.01.008
43. Vercelli, D. (2003). Genetic polymorphism in allergy and asthma. Current Opinion in Immunology, 15(6), 609–613. doi:10.1016/j.coi.2003.09.005
44. Via M, Gignoux CR, Roth LA, Fejerman L, Galanter J, Choudhry S, Toro-Labrador G, Viera-Vera J, Oleksyk TK, Beckman K *et al.*: History shaped the geographic distribution of genomic admixture on the island of Puerto Rico. PLoS ONE 2011, 6(1):e16513.

45. Xingnan Li, Timothy D. Howard, Siqun L. Zheng, Tmirah Haselkorn, Stephen P. Peters, Deborah A. Meyers, Eugene R. Bleecker. Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. J Allergy Clin Immunol. Author manuscript; available in PMC 2011 February 1. J Allergy Clin Immunol. 2010 February; 125(2): 328–335.e11. doi: 10.1016/j.jaci.2009.11.018 PMCID: PMC2824608

Appendix

Table 1. List of samples selected from Local Genome Diversities Studies cohort (LGDS) and their

Sample id	Municipality	Sample id	Municipality	Sample id	Municipality	Sample id	Municipality	Sample id	Municipality	Sample id	Municipality
an_11a	Anasco	cg24	Caguas	gb6	Guaynabo	mc9	Moca	po86	Ponce	sj_77b	San Juan
an_18a	Anasco	cg30	Caguas	gb9	Guaynabo	mv10	Morovis	po93	Ponce	sj_80b	San Juan
an_2a	Anasco	cg31	Caguas	gn63	Guanica	mv12	Morovis	rg_13a	Rio Grande	sj_82b	San Juan
an_4a	Anasco	cg4	Caguas	gn64	Guanica	mv14	Morovis	rg_16b	Rio Grande	sj_83b	San Juan
an_5a	Anasco	cg8	Caguas	gn65	Guanica	mv16	Morovis	rg_18a	Rio Grande	sj_86b	San Juan
an_65a	Anasco	cr1	Carolina	gn66	Guanica	mv5	Morovis	rg_19b	Rio Grande	sj13	San Juan
an_66b	Anasco	cr14	Carolina	gn67	Guanica	mv5	Morovis	rg_33b	Rio Grande	sj15	San Juan
an_67a	Anasco	cr18	Carolina	gn68	Guanica	mv6	Morovis	rg_36b	Rio Grande	sj21	San Juan
an_6a	Anasco	cr2	Carolina	gn69	Guanica	my_22b	Mayaguez	rg_39b	Rio Grande	sj27	San Juan
an_77a	Anasco	cr27	Carolina	gn71	Guanica	my_25a	Mayaguez	rg_51a	Rio Grande	sj34	San Juan
an_83b	Anasco	cr3	Carolina	gn73	Guanica	my_26b	Mayaguez	rg_52a	Rio Grande	sj41	San Juan
an_85b	Anasco	cr37	Carolina	ho62	Hormigueros	my_27a	Mayaguez	rg_54b	Rio Grande	sj50	San Juan
an_91a	Anasco	cr83	Carolina	ho67	Hormigueros	my_28a	Mayaguez	rg_58a	Rio Grande	sj51	San Juan
an_94a	Anasco	cz71	Corozal	ho76	Hormigueros	my_29a	Mayaguez	rg_61a	Rio Grande	sj64	San Juan
an_96b	Anasco	cz74	Corozal	ho81	Hormigueros	my_30a	Mayaguez	rg_69b	Rio Grande	vq_10b	Vieques
an_9b	Anasco	cz76	Corozal	ho92	Hormigueros	my_32b	Mayaguez	rg_70b	Rio Grande	vq_11b	Vieques
ao51	Arroyo	cz81	Corozal	lo104	Loiza	my_53a	Mayaguez	rg_72a	Rio Grande	vq_14b	Vieques
ao53	Arroyo	do18	Dorado	lo106	Loiza	my_55a	Mayaguez	rg_74a	Rio Grande	vq_16b	Vieques
ao54	Arroyo	do25	Dorado	lo108	Loiza	my_61a	Mayaguez	si1	Santa Isabel	vq_18b	Vieques
ao55	Arroyo	do31	Dorado	lo37	Loiza	my_65a	Mayaguez	si10	Santa Isabel	vq_1b	Vieques
ao56	Arroyo	do53	Dorado	lo38	Loiza	my_67a	Mayaguez	si11	Santa Isabel	vq_20b	Vieques
ao60	Arroyo	do54	Dorado	lo46	Loiza	my_71a	Mayaguez	si13	Santa Isabel	vq_23b	Vieques
ar24	Arroyo	do66	Dorado	lo49	Loiza	my_72a	Mayaguez	si14	Santa Isabel	vq_27b	Vieques
ar29	Arroyo	do73	Dorado	lo51	Loiza	my_79a	Mayaguez	si15	Santa Isabel	vq_28b	Vieques
ar60	Arroyo	do74	Dorado	lo70	Loiza	pa1	Patillas	si16	Santa Isabel	vq_31b	Vieques
ar61	Arroyo	do76	Dorado	lo71	Loiza	pa18	Patillas	si19	Santa Isabel	vq_33b	Vieques
bc16	barceloneta	do9	Dorado	lr_10b	Lares	pa25	Patillas	si2	Santa Isabel	vq_5b	Vieques
bc19	barceloneta	fa10	Fajardo	lr_11b	Lares	pa28	Patillas	si3	Santa Isabel	vq_6b	Vieques
bc2	barceloneta	fa11	Fajardo	lr_12b	Lares	pa30	Patillas	si5	Santa Isabel	vq_8b	Vieques
bc3	barceloneta	fa12	Fajardo	lr_13b	Lares	pa32	Patillas	si6	Santa Isabel	vq10	Vieques
bc71	barceloneta	fa2	Fajardo	lr_17b	Lares	pa35	Patillas	si7	Santa Isabel	vq13	Vieques
bc72	barceloneta	fa3	Fajardo	lr_18b	Lares	pa37	Patillas	si8	Santa Isabel	vq14	Vieques
bc73	barceloneta	fa5	Fajardo	lr_19b	Lares	pa37a	Patillas	si9	Santa Isabel	vq17	Vieques
bc75	barceloneta	fa6	Fajardo	lr_21b	Lares	pa4	Patillas	sj_10b	San Juan	vq19	Vieques
bc76	barceloneta	fa7	Fajardo	mc13	Moca	pa68	Patillas	sj_15b	San Juan	vq19b	Vieques
bc82	barceloneta	fa8	Fajardo	mc14	Moca	pa8	Patillas	sj_27b	San Juan	vq24	Vieques
bc83	barceloneta	fa9	Fajardo	mc17	Moca	po35	Ponce	sj_30b	San Juan	vq33	Vieques
bc84	barceloneta	gb1	Guaynabo	mc2	Moca	po50	Ponce	sj_35b	San Juan	vq35	Vieques
bc9	barceloneta	gb11	Guaynabo	mc25	Moca	po57	Ponce	sj_50b	San Juan	vq41	Vieques
cg12	Caguas	gb14	Guaynabo	mc28	Moca	po58	Ponce	sj_57b	San Juan		
cg13	Caguas	gb15	Guaynabo	mc3	Moca	po61	Ponce	sj_64b	San Juan		
cg14	Caguas	gb2	Guaynabo	mc36	Moca	po79	Ponce	sj_68b	San Juan		
cg19	Caguas	gb3	Guaynabo	mc5	Moca	po80	Ponce	sj_73b	San Juan		
cg20	Caguas	gb5	Guaynabo	mc7	Moca	po82	Ponce	sj_76b	San Juan		

respective municipalities. A total of 259 individuals were selected from 25 different municipalities.

Table 2. List of asthma phenotype individuals used under our study. A total of 80 samples were selected as our asthma case cohort.

Sample id	Municipality	Sample id	Municipality	Sample id	Municipality	Sample id	Municipality
0	N/A	21	N/A	44	N/A	74	N/A
1	N/A	22	N/A	45	N/A	75	N/A
3	N/A	23	N/A	46	N/A	76	N/A
4	N/A	24	N/A	47	N/A	77	N/A
5	N/A	25	N/A	48	N/A	78	N/A
5a	N/A	26	N/A	49	N/A	80	N/A
6	N/A	29	N/A	57	N/A	83	N/A
7	N/A	30	N/A	59	N/A	84	N/A
8	N/A	31	N/A	60	N/A	87	N/A
9	N/A	31	N/A	61	N/A	88	N/A
10	N/A	32	N/A	62	N/A	89	N/A
11	N/A	33	N/A	64	N/A	92	N/A
12	N/A	34	N/A	65	N/A	94	N/A
14	N/A	35	N/A	66	N/A	95	N/A
15	N/A	36	N/A	67	N/A	96	N/A
16	N/A	37	N/A	68	N/A	97	N/A
17	N/A	38	N/A	70	N/A	98	N/A
18	N/A	39	N/A	71	N/A	100	N/A
19	N/A	40	N/A	72	N/A	103A	N/A
20	N/A	43	N/A	73	N/A	103B	N/A

Table 3. Sample description Table. A total of 93 samples were used for our genotyping assay.

Sample id	Municipality	Sample id	Municipality	Sample id	Municipality
AR 24	Arecibo	FA 03	Fajardo	LO 70	Loiza
AR 29	Arecibo	FA 05	Fajardo	LO 71	Loiza
AR 60	Arecibo	FA 06	Fajardo	MC 02	Moca
AO 51	Arroyo	FA 07	Fajardo	MC 03	Moca
AO 53	Arroyo	FA 08	Fajardo	MC 05	Moca
AO 55	Arroyo	GN 63	Guanica	MC 07	Moca
AO 56	Arroyo	GN 64	Guanica	MC 09	Moca
AO 60	Arroyo	GN 65	Guanica	MC 13	Moca
BC 19	Barceloneta	GN 66	Guanica	MC 17	Moca
BC 2	Barceloneta	GN 67	Guanica	MC 28	Moca
BC 3	Barceloneta	GN 68	Guanica	MC 36	Moca
BC 71	Barceloneta	GN 69	Guanica	MC14	Moca
BC 72	Barceloneta	GN 71	Guanica	MV 16	Morovis
BC 73	Barceloneta	GN 73	Guanica	MV 5	Morovis
BC 75	Barceloneta	GB 11	Guaynabo	MV 6	Morovis
BC 76	Barceloneta	GB 14	Guaynabo	PA 18	Patillas
BC 83	Barceloneta	GB 15	Guaynabo	PA 28	Patillas
CR 1	Cabo Rojo	GB 3	Guaynabo	PA 30	Patillas
CR 14	Cabo Rojo	GB 6	Guaynabo	PA 32	Patillas
CR 18	Cabo Rojo	GB 9	Guaynabo	PA 35	Patillas
CR 2	Cabo Rojo	HO 62	Hormigueros	PA 37	Patillas
CR 27	Cabo Rojo	HO 67	Hormigueros	PO 58	Ponce
CR 3	Cabo Rojo	HO 76	Hormigueros	PO 86	Ponce
CR 37	Cabo Rojo	HO 81	Hormigueros	PO 93	Ponce
CR 83	Cabo Rojo	HO 92	Hormigueros	SJ 21	San Juan
CZ 76	Corozal	LO 104	Loiza	SJ 41	San Juan
CZ 81	Corozal	LO 37	Loiza	SJ 50	San Juan
DO 18	Dorado	LO 38	Loiza	SJ 64	San Juan
DO 25	Dorado	LO 46	Loiza	VQ 10	Vieques
DO 53	Dorado	LO 49	Loiza	VQ 14	Vieques
DO 54	Dorado	LO 51	Loiza	VQ 35	Vieques

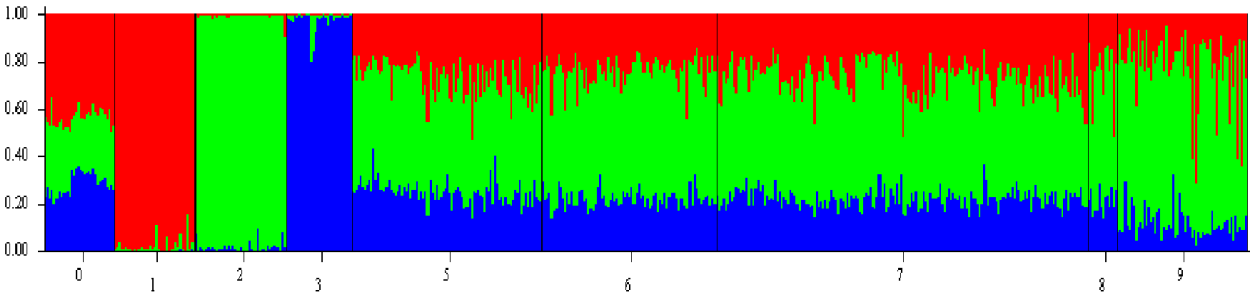
Table 4. List description of all 93 SNPs implemented in our Ancestry Informative Panel. Description includes chromosome position, gene description, location on NCBI assembly and SNP type.

Assay ID	Gene Symbol	Chromosome	NCBI Gene Reference	NCBI SNP Reference	SNP Type	Location on NCBI Assembly
C___1962691_1_	BEND5;AGBL4	1	NM_024603.2;NM_032785.3	rs1934393	INTRON	49208618
C__11624297_10		1		rs2592888	INTERGENIC/UNKNOWN	159585573
C__26194598_10	LYST	1	NM_000081.2	rs7535375	INTRON	235913159
C___2508040_10		1		rs2817611	INTERGENIC/UNKNOWN	11613163
C__26670165_10	BCL9	1		rs3806218	INTERGENIC/UNKNOWN	147011783
C__27497242_10	LPHN2	1	NM_012302.2	rs3828121	INTRON	82422200
C___2895558_10	LOC400794	1		rs4657449	INTERGENIC/UNKNOWN	165465281
C___2989477_10		1		rs6684063	INTERGENIC/UNKNOWN	30699340
C___7479863_10	NCKAP5	2	NM_207363.2;NM_207481.3	rs1036543	INTRON	133676214
C__28991571_20		2		rs3860446	INTERGENIC/UNKNOWN	104489351
C__27994514_10		2		rs4852696	INTERGENIC/UNKNOWN	83151641
C__29872311_20		2		rs10497705	INTERGENIC/UNKNOWN	190492014
C__30169537_20	CAB39	2	NM_001130849.1;NM_001130850.1;NM_016289.3	rs10498255	INTRON	231612230
C___3219728_10	FLJ16341	2		rs842634	INTERGENIC/UNKNOWN	61091222
C___7900010_10	LOC375295	2		rs868179	INTERGENIC/UNKNOWN	177549497
C____637315_10	ALK	2	NM_004304.4	rs1073319	INTRON	29440454
C___3228183_10		2		rs1470524	INTERGENIC/UNKNOWN	45129515
C___8721135_10		2		rs1517634	INTERGENIC/UNKNOWN	224183485
C__15921070_10	PKP4	2	NM_001005476.1;NM_003628.3	rs2711070	INTRON	159502533
C___1930284_10		3		rs1984473	INTERGENIC/UNKNOWN	155811284
C___1930637_10	ITPR1	3	NM_001099952.2;NM_001168272.1;NM_002222.5	rs304051	INTRON	4578306
C__29113781_20		3		rs6804094	INTERGENIC/UNKNOWN	187057970
C__25594306_10	APPL1;ASB14	3	NM_012096.2	rs10510791	INTRON	57294085
C__30354117_20		3		rs9310888	INTERGENIC/UNKNOWN	29286762
C___8767693_10		3		rs1498991	INTERGENIC/UNKNOWN	20900132
C__11905631_10	HCLS1	3	NM_005335.4	rs1919550	INTRON	121364173
C__30238451_20		4		rs10517518	INTERGENIC/UNKNOWN	61795416
C__29789523_20		4		rs10519979	INTERGENIC/UNKNOWN	149634951
C__29784870_20		4		rs10520440	INTERGENIC/UNKNOWN	180799005
C___7462898_10		4		rs1398829	INTERGENIC/UNKNOWN	22023275
C___7778042_10		4		rs9307613	INTERGENIC/UNKNOWN	130357404
C____675372_10	FBXL7	5	NM_012304.3	rs257748	INTRON	15819615
C__29887779_20		5		rs10515535	INTERGENIC/UNKNOWN	143516142
C___3020716_10		5		rs6883095	INTERGENIC/UNKNOWN	79891047
C__42854506_10	IL7R	5	NM_002185.2	rs1353251	INTRON	35857207
C___3167333_10	SNORD95;TRIM52;SNORD96A;GNB2L1	5		rs1477277	INTERGENIC/UNKNOWN	180675022
C___2840547_20	MCTP1	5	NM_001002796.2;NM_024717.4	rs153898	INTRON	94188622
C___1384222_10		5		rs1990745	INTERGENIC/UNKNOWN	103381922
C___2907748_10	SERAC1	6	NM_032861.3	rs9295316	INTRON	158567066
C__11635757_10	C6orf170	6	NM_152730.4	rs9320808	INTRON	121654596
C___3278240_10	ZNF76	6	NM_003427.3	rs10484578	INTRON	35246319
C__29433174_20	MOXD1	6	NM_015529.2	rs6569792	INTRON	132694751
C___3107030_10		6		rs6911727	INTERGENIC/UNKNOWN	9116398
C___2580493_10	MAGI2	7	NM_012301.3	rs10214949	INTRON	79048593
C__29779124_10	COBL	7	NM_015198.3	rs10248051	INTRON	51119353
C___2621385_10	JAZF1	7	NM_175061.3	rs10486576	INTRON	28119143
C__29612207_20	EXOC4	7	NM_021807.3	rs10488172	INTRON	133335176
C___1220570_10	CNTNAP2	7	NM_014141.5	rs802524	INTRON	145951642
C__12111235_10		8		rs1898280	INTERGENIC/UNKNOWN	116074460
C__27991782_10		8		rs4130405	INTERGENIC/UNKNOWN	99420775
C__29256197_20		8		rs7463344	INTERGENIC/UNKNOWN	33863527
C__29818677_20		8		rs9325872	INTERGENIC/UNKNOWN	20480271
C___2172116_10		9		rs10491654	INTERGENIC/UNKNOWN	102139527
C__15815793_10	BNC2	9	NM_017637.5	rs2840290	INTRON	16733957
C__11770073_10		9		rs4013967	INTERGENIC/UNKNOWN	76897070
C__29964593_20		10		rs10508349	INTERGENIC/UNKNOWN	8298964
C__12127268_10	EIF3A	10	NM_003750.2	rs1397618	INTRON	120832675
C___7467860_10		10		rs2785279	INTERGENIC/UNKNOWN	33709876
C__27966795_10	FAS	10		rs4934436	INTERGENIC/UNKNOWN	90783320
C___9602437_20	OR10A2	11		rs2595456	INTERGENIC/UNKNOWN	6884763
C____768677_10		11		rs567992	INTERGENIC/UNKNOWN	106262397
C___1673694_10	APLP2	11	)1142276.1;NM_001142277.1;NM_001142278.1;NM_00	rs879780	INTRON	130008104
C___8374539_10	RIN1;BRMS1;B3GNT1	11	NM_001024957.1;NM_015399.3	rs948360	INTRON	66106725
C__30493927_20		12		rs10506816	INTERGENIC/UNKNOWN	79924857
C___2876583_10		12		rs249847	INTERGENIC/UNKNOWN	98867716
C__27930805_10		12		rs4034627	INTERGENIC/UNKNOWN	128397472
C__25751940_10	FBXW8	12	NM_012174.1;NM_153348.2	rs4076700	MIS-SENSE MUTATION	117383320
C__26956649_10	CCND2	12		rs4625554	INTERGENIC/UNKNOWN	4416304
C__26024638_10		12		rs4762106	INTERGENIC/UNKNOWN	66018473
C__29172327_20		13		rs5000507	INTERGENIC/UNKNOWN	82088954
C__30164860_20		13		rs10492585	INTERGENIC/UNKNOWN	105386176
C__30104882_20		14		rs10131076	INTERGENIC/UNKNOWN	80774385
C___7580005_10		14		rs1451928	INTERGENIC/UNKNOWN	48340741
C__15752328_10	PRKCH	14	NM_006255.3	rs2296274	INTRON	61917178
C___3141697_20		14		rs9323178	INTERGENIC/UNKNOWN	23113646
C__11676142_20		15		rs10520678	INTERGENIC/UNKNOWN	88937283
C__29773186_20		16		rs10500505	INTERGENIC/UNKNOWN	64943026
C__25654196_20	WWOX	16	NM_016373.2	rs4130513	INTRON	78458750
C___1431083_10		16		rs1004704	INTERGENIC/UNKNOWN	48537421
C___2560150_10		17		rs10491097	INTERGENIC/UNKNOWN	19361211
C__27006283_10	GNAL	18		rs1013459	INTERGENIC/UNKNOWN	11700534
C__26869828_10	RTTN	18	NM_173630.3	rs12953952	INTRON	67737927
C__12086674_10		18		rs2042762	INTERGENIC/UNKNOWN	35277622
C___2883673_20		19		rs888861	INTERGENIC/UNKNOWN	35381852
C____778619_10		20		rs354747	INTERGENIC/UNKNOWN	58912660
C___2506154_30	PLCB1	20	NM_015192.2;NM_182734.1	rs708915	INTRON	8400667
C___2449352_10		21		rs2829454	INTERGENIC/UNKNOWN	26273071
C_____13219_1_	TNRC6B	22	NM_001024843.1;NM_001162501.1;NM_015088.2	rs138022	INTRON	40613036
		2		rs10515919		75540596
		11		rs10501474		80400647
		15		rs9302185		54954864
		17		rs2253624		69732081
		19		rs798887		54793188

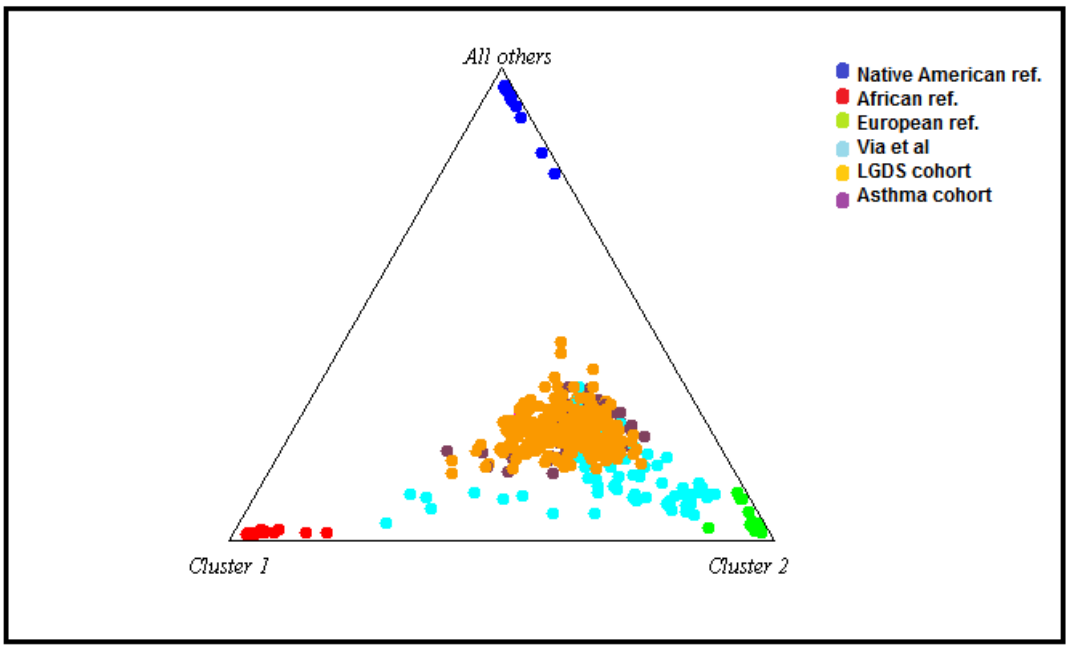
Table 5. Results for each SNP assay by municipality. VIC and FAM stand for the fluorescent probes dyes for each allele call. A FAM/FAM results implies homozygous for allele 2, VIC/FAM equals heterozygous and VIC/VIC implies homozygous for allele 1.

			Allele 2		Allele 1	
Gene	**Municipality**		**FAM/FAM**	**VIC/FAM**	**VIC/VIC**	**NO AMPLIFICATION**
RAD50	Arecibo	AR		1	2	
RAD50	Arroyo	AO		4	1	
RAD50	Barceloneta	BC		3	6	
RAD50	Cabo Rojo	CR		5	1	2
RAD50	Corozal	CZ			2	
RAD50	Dorado	DO	1	3		
RAD50	Fajardo	FA		1	4	
RAD50	Guanica	GN	2	3	4	
RAD50	Guaynabo	GB		1	4	1
RAD50	Hormigueros	HO			5	
RAD50	Loiza	LO		2	6	
RAD50	Moca	MC	4	1	4	1
RAD50	Morovis	MV		2	1	
RAD50	Patillas	PA		1	5	
RAD50	Ponce	PO	1			2
RAD50	San Juan	SJ	1	1	1	1
RAD50	Vieques	VQ		1	1	1
ORMDL3	Arecibo	AR		1	1	1
ORMDL3	Arroyo	AO		1	4	
ORMDL3	Barceloneta	BC		3	6	
ORMDL3	Cabo Rojo	CR		1	4	3
ORMDL3	Corozal	CZ		1	1	
ORMDL3	Dorado	DO		3	1	
ORMDL3	Fajardo	FA	2	1	2	
ORMDL3	Guanica	GN	1	2	5	1
ORMDL3	Guaynabo	GB	1	2	3	
ORMDL3	Hormigueros	HO		2		1
ORMDL3	Loiza	LO		1	7	
ORMDL3	Moca	MC		4	5	1
ORMDL3	Morovis	MV	1	1	1	
ORMDL3	Patillas	PA	1	1	4	
ORMDL3	Ponce	PO			2	1
ORMDL3	San Juan	SJ		1	2	1
ORMDL3	Vieques	VQ		1	2	
IL33	Arecibo	AR	1	2	2	
IL33	Arroyo	AO		1		2
IL33	Barceloneta	BC	1	3	4	1
IL33	Cabo Rojo	CR	1	2	2	3
IL33	Corozal	CZ		2		
IL33	Dorado	DO		2	2	
IL33	Fajardo	FA		3	2	
IL33	Guanica	GN	1		8	

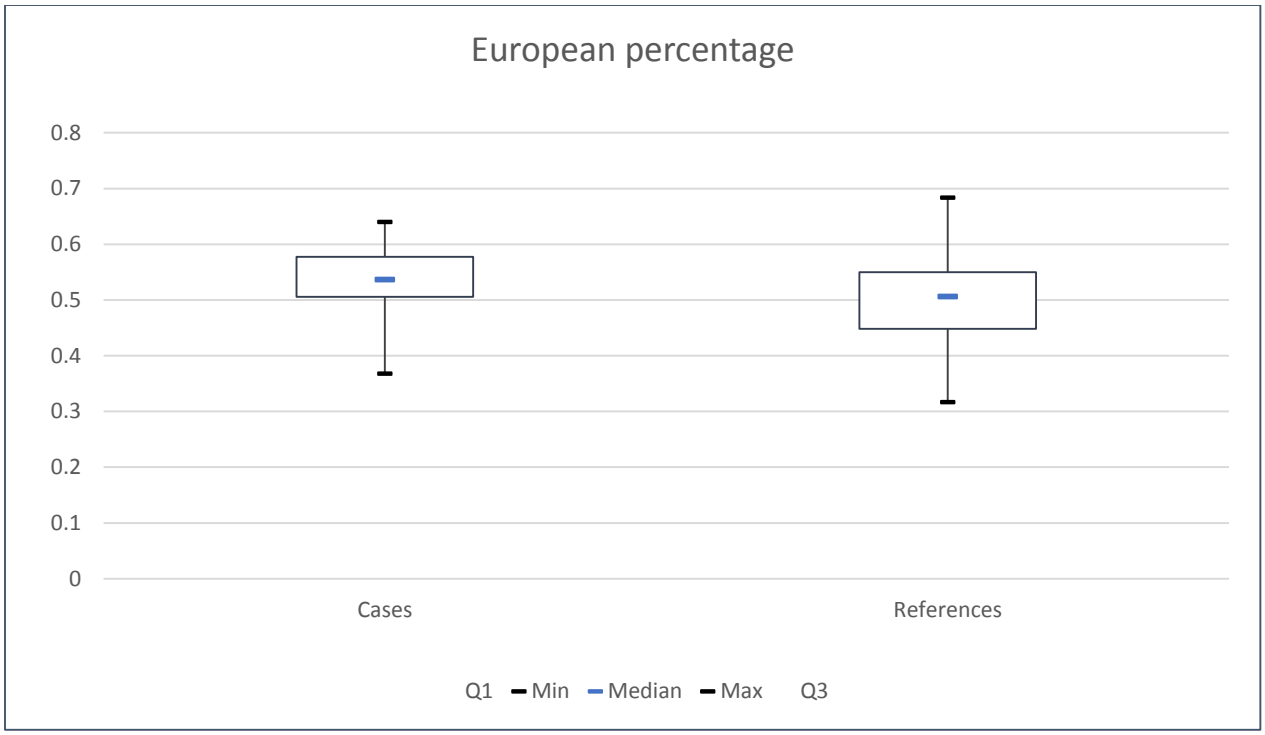
IL33	Guaynabo	GB	1	3	1	1
IL33	Hormigueros	HO	2	2		1
IL33	Loiza	LO	1	5	2	
IL33	Moca	MC	1	3	4	2
IL33	Morovis	MV	1	2		
IL33	Patillas	PA	1	1	4	
IL33	Ponce	PO			2	1
IL33	San Juan	SJ		1	2	1
IL33	Vieques	VQ		2	1	
HNMT	Arecibo	AR		1	2	
HNMT	Arroyo	AO	1	3	1	
HNMT	Barceloneta	BC	2	5	2	
HNMT	Cabo Rojo	CR		3	2	3
HNMT	Corozal	CZ	1	1		
HNMT	Dorado	DO		1	3	
HNMT	Fajardo	FA	1	2	2	
HNMT	Guanica	GN		5	4	
HNMT	Guaynabo	GB			6	
HNMT	Hormigueros	HO	2	2	1	
HNMT	Loiza	LO		4	4	
HNMT	Moca	MC	2	2	5	1
HNMT	Morovis	MV		1	2	
HNMT	Patillas	PA		4	2	
HNMT	Ponce	PO		1	2	
HNMT	San Juan	SJ		3	1	
HNMT	Vieques	VQ			3	
GSDMB	Arecibo	AR		1	2	
GSDMB	Arroyo	AO	1	4		
GSDMB	Barceloneta	BC	2	5	2	
GSDMB	Cabo Rojo	CR		2	2	3
GSDMB	Corozal	CZ	2			
GSDMB	Dorado	DO		1	3	
GSDMB	Fajardo	FA	1	2	2	
GSDMB	Guanica	GN	1	3	5	
GSDMB	Guaynabo	GB			6	
GSDMB	Hormigueros	HO	2	2	1	
GSDMB	Loiza	LO		4	4	
GSDMB	Moca	MC	2	3	4	1
GSDMB	Morovis	MV		1	2	
GSDMB	Patillas	PA		4	2	
GSDMB	Ponce	PO		1	2	
GSDMB	San Juan	SJ		3	1	
GSDMB	Vieques	VQ			3	
Totals			46	163	216	37



Appendix Figure 1. STRUCTURE bar plot assuming k=3 ancestry clusters (for European, African, and the Native American ancestry). Population data is designated from 0 to 9 where 0 = blank, 1 to 3 are the references for the inferred ancestral population clusters, designated as 1 = African, 2 = European and 3 = Native American. The average values for the reference ancestral populations are 98% for their respective represented population. Population data 5 and 7 stand for the reference population (LGDS cohort) with average values of 50% European, 26% African and 23% Native American. Population data 6 and 8 represent the asthma cases with average ancestral component values of 54% European, 24% African and 22% Native American. Population data 9 represent Via *et al*. dataset with values of 66% European, 22% African and 12% Native American

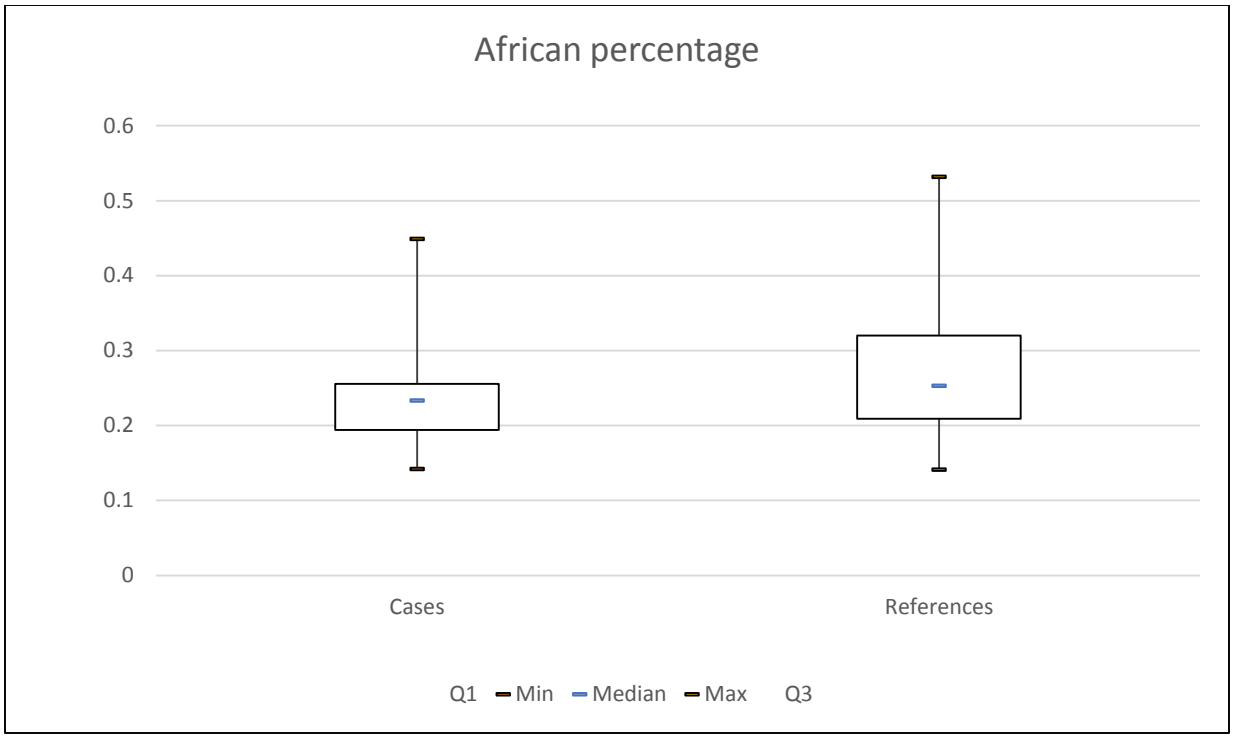


Appendix Figure 2. Triangle plot reveals spatial distribution of all three clusters for the asthma cases, reference cohort, as well as the dataset from Via *et al.* (2011) dataset



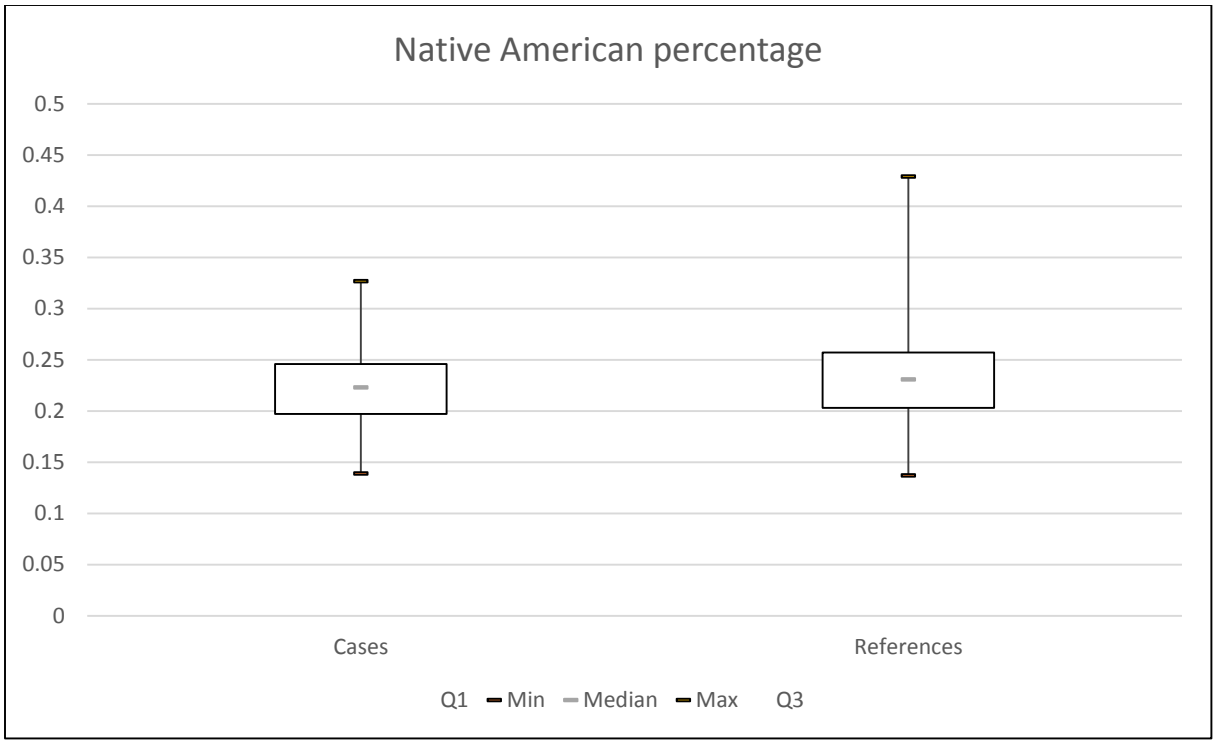
	Cases	References
Q1	0.5055	0.448
Min	0.368	0.317
Median	0.5365	0.506
Max	0.64	0.684
Q3	0.57775	0.55

Appendix Figure 3. Box plot graph for European average ancestral component in the case-control panel. Values of Q1, Min, Median, Max and Q3 are provided in the Table above



	Cases	References
Q1	0.194	0.209
Min	0.142	0.141
Median	0.2335	0.253
Max	0.449	0.532
Q3	0.25575	0.32

Appendix Figure 4. Box plot graph for African average ancestral component in the case-control panel. Values of Q1, Min, Median, Max and Q3 are provided in the Table above



	Cases	References
Q1	0.197	0.203
Min	0.139	0.137
Median	0.223	0.231
Max	0.327	0.429
Q3	0.24575	0.257

Appendix Figure 5. Box plot graph for Native American average ancestral component in the case-control panel. Values of Q1, Min, Median, Max and Q3 are provided in the Table above

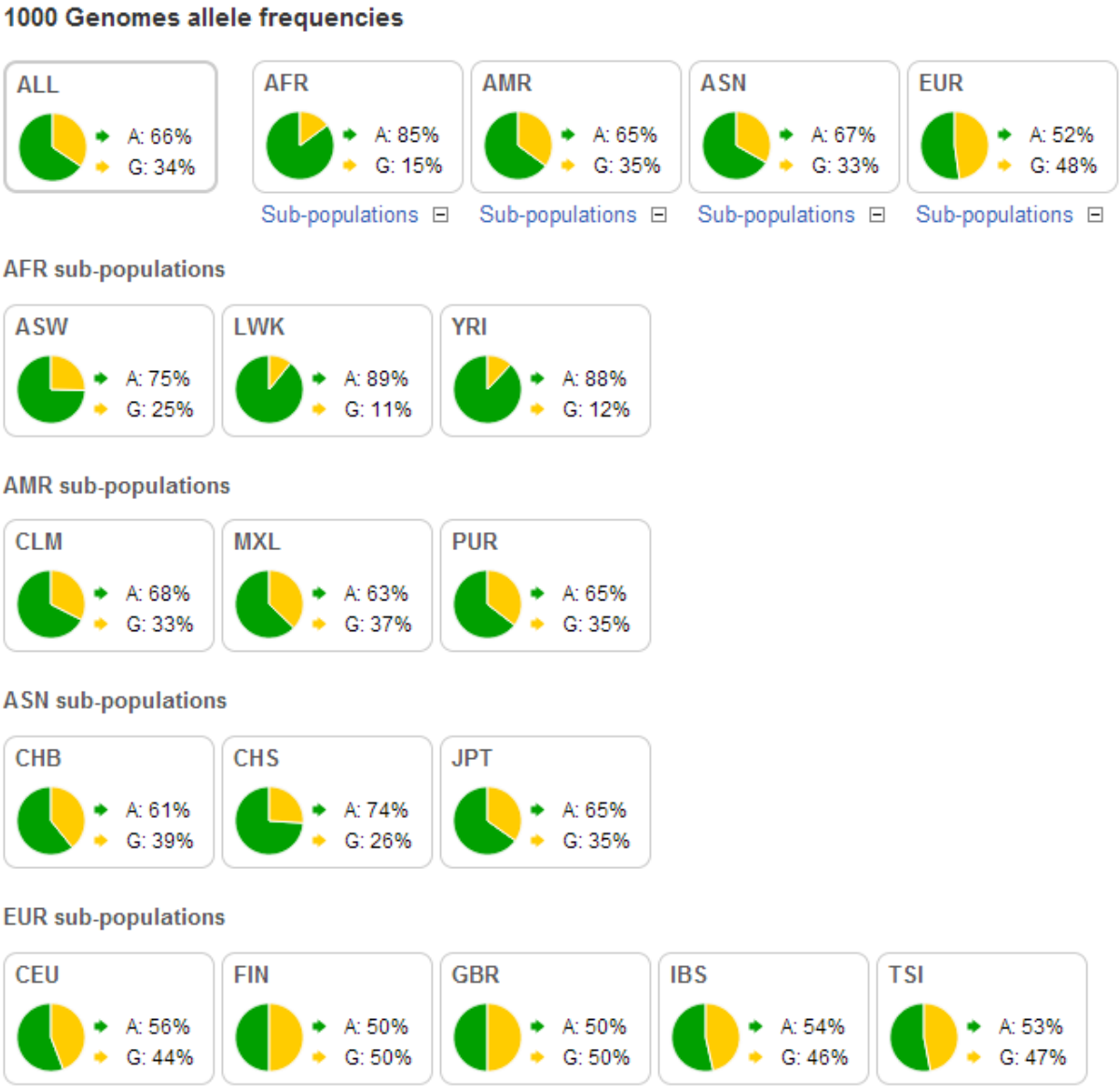


Figure 6. (ORMDL3) rs8076131 allele frequencies for 4 population and their respective sub-population from 1000 genomes data. Source. Ensembl population data