Validating Genetic Associations of Asthma in The Consortium of Asthma among African-ancestry Populations in the Americas

Hrafnhildur Bjarnadóttir

Thesis for B.S. degree in Medicine
University of Iceland
School of Health Sciences
Validating Genetic Associations of Asthma in The Consortium of Asthma among African-ancestry Populations in the Americas

Hrafnhildur Bjarnadóttir

Thesis for B.S. degree in Medicine
Supervisors: Kathleen C. Barnes\(^2\), Michelle Daya\(^2\)

\(^1\)Department of Medicine, University of Iceland, \(^2\)Division of Biomedical Informatics and Personalized Medicine, University of Colorado Anschutz Medical Campus

Department of Medicine
School of Health Sciences
May 2017
Abstract

Introduction. Asthma is a chronic inflammatory disorder of the respiratory tract estimated to affect as many as 300 million people worldwide. Asthma comprises a range of phenotypes with the risk factors for each phenotype depending on both genetic and environmental factors as well as their interactions. Asthma affects people of all ethnic backgrounds but ethnic minorities, such as African Americans in the United States, are affected disproportionally. At present, it is not clear to what extent genetic variation contributes to these ethnic disparities. Genome-wide association studies (GWAS) have identified many risk variants relating to asthma but in these studies, populations of African ancestry have only been a small fraction of studied subjects. The Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) performed the largest meta-analysis of asthma GWAS in individuals of African ancestry to date. The study yielded a number of loci that may contribute to risk of asthma specifically in individuals of African ancestry. The aim of this study is to replicate the findings of the CAAPA Consortium in an independent European ancestry population, and thereby test the generalizability of associations in non-African ancestry populations.

Materials and methods. The six single nucleotide polymorphisms (SNPs) that showed strongest association with asthma in the CAAPA study were selected for replication. Genotyping was performed in 633 European ancestry samples previously collected for The Collaborative Studies on the Genetics of Asthma (CSGA). DNA extraction was performed using an AutoGen FLEX STAR. Genotyping was performed using a TaqMan ABI 7900. Hardy Weinberg testing and tests for Mendelian inconsistencies were performed using PLINK. Genetic association was tested using the generalized estimating equations (GEE).

Results. None of the six SNPs showed significant association with asthma in the replication population. However, the direction of effect for five of the six SNPs was the same as in the discovery sample.

Discussion. The CSGA population was selected, partly, due to the availability of the samples for genotyping. As the individuals of African ancestry were not available for a replication study it was decided to replicate the findings of the CAAPA Consortium in the portion of the CSGA population that was of European ancestry. Replication of the CAAPA findings in a population of European ancestry was of interest for a number of reasons. The gender distribution and mean age of subject in the CSGA and CAAPA populations were similar. The LD patterns for the SNPs on chromosome 2 were different between the two populations. It is, however, unlikely that differing LD patterns contributed to the insignificant results of this study. Due to the fact that a significant portion of the CSGA subjects had incomplete phenotype data, the proportion of cases and controls is the least comparable factor between study populations. It is likely that the use of GEE as this studies analytical method, in a small sample population with such a large portion of missing phenotype data, contributed to the insignificant findings of the study. It is possible that some of the genotyped SNPs have no association with asthma in populations of European ancestry. This study alone can, however, not confirm this and the role of these SNPs in asthma in European populations requires further investigation.
**Ágrip**


**Efnívidur og aðferðir:** Ákveðið var að rannsaka þá eins basapara breytileikar sem sýndu sterkustu tengslin við astma í CAAPA rannsókninni með aрfgerðargreiningu. Einangrun á DNA var framkvæmd með notkun AutoGen FLEX STAR. Aρfgerðargreining var framkvæmd með TaqMan ABI 7900 aðferðinni. Hardy Weinberg próf og prófun fyrir Mendelísku misræmi voru framkvæmd með PLINK aðferðinni. Tölfræðigreining var framkvæmd með GEE aðferðinni.

**Niðurstöður:** Eins basapara breytileikarinnir sem voru prófaðir sýndu ekki tölfræðilega marktæk tengsl við astma í rannsóknarþyðinu. Niðurstöður fylgdu hins vegar sómu stefnu og niðurstöðurnar í upphaflegu CAAPA rannsókninni hjá fimm af sex rónnsókuðum eins basapara breytileikum.

**Alyktaðir:** CSGA rannsóknarþyðið var valið, að hluta til, vegna tilltæki sýnanna. Þar sem að ekki var í bóði að núta einstalinga af afrískum uppruna úr CSGA þyðinu í þessa rannsókn var ákeðið að rannsóknarþyðið skildi skipa af einstaklingum í CSGA þyðinu sem voru af evrópskum uppruna. Vert var að athuga niðurstöður CAAPA rannsóknarinnar í evrópsku þyði af nokkrum ástæðum. Kynjadeifing og meðalaldur einstaklinga í CAAPA þyðum voru svipaður. Tengslaþjafnvægi á milli eins basapara breytileikanna á litningi 2 var ólíkt í CSGA og CAAPA þyðunum. Það er hins vegar ólíkt að mismunandi tengslaþjaðvægi í milli þyðanna hafi stuðlað að ómarktæki niðurstaða þessarar rannsóknar. Þátturinn sem var minnst sambærilegur á milli CSGA og CAAPA þyðanna var hlutfallið á milli tilfella og viðmiða. Þetta er vegna þess að upplýsingar um svipgerð vantaði fyrir hluta einstaklinga í CSGA þyðinu sem skektu hlutfall tilfellu og viðmiða í því þyði. Það er líktleg að notkun GEE tölfræðigreiningar í þyði þar sem að miklri af upplýsingum um svipgerð vantaði hafi stuðlað að ómarktæki niðurstaða í þessari rannsókn. Það er hugsanlegt að einhverjar eins basapara breytileikanna sem voru prófaðir í þessari rannsókn hafi ekki tengsl við astma í evrópskum þyðum en til að staðfestu það er þörð á frekari rannsóknurn.
Acknowledgements

First, I would like to thank my supervisor, Kathleen Barnes, for allowing me to come to Colorado to work in her lab. The experience was I both highly educational and enjoyable. I would also like to thank Þórunn Rafnar for connecting me with Kathleen in the first place and Bergljót Sigriður Rafnar Karlsdóttir for encouraging me to go abroad. Thank you to everyone in the Barnes Lab who helped me with my project; Nicholas Rafaels, Michelle Daya, Maher Boorgula, and special thanks to Monica Campbell for her guidance and support throughout my stay. Thank you to my parents, Sigrún Þorgeirsdóttir and Bjarni Óssurarson, for their love and support and to Sóley Bogadóttir for being a good friend in Denver. Lastly, thank you to Ingvi Þór Hermannsson for staying up late so I could call him every night.
Table of Contents

Abstract ........................................................................................................................................... i
Ágrip ................................................................................................................................................ ii
Acknowledgements ........................................................................................................................ iii
List of abbreviations ......................................................................................................................... 1
List of figures ..................................................................................................................................... 2
List of tables ...................................................................................................................................... 2
1 Introduction .................................................................................................................................... 3
  1.1 Asthma .................................................................................................................................... 3
    1.1.1 Definition and diagnosis of asthma .................................................................................... 3
    1.1.2 Treatment of asthma ....................................................................................................... 3
    1.1.3 Burden of asthma ............................................................................................................ 4
    1.1.4 Etiology of asthma ........................................................................................................... 4
    1.1.5 Asthma in individuals of African descent ......................................................................... 4
  1.2 Genome-wide association studies of asthma ............................................................................ 4
    1.2.1 GWAS technology .......................................................................................................... 5
    1.2.2 Advantages of GWAS .................................................................................................... 5
    1.2.3 Limitations of GWAS ..................................................................................................... 6
    1.2.4 Meta-analysis of asthma GWAS ..................................................................................... 7
    1.2.5 Genotype imputation ...................................................................................................... 7
    1.2.6 GWAS across populations of diverse ancestry ................................................................. 7
  1.3 The CAAPA Consortium and the African Diaspora Power Chip ............................................. 8
  1.4 Replication studies of GWAS ................................................................................................ 8
    1.4.1 Replication of CAAPA study .......................................................................................... 9
    1.4.2. SNPs chosen for replication .......................................................................................... 9
2 Aim of study .................................................................................................................................. 10
3 Materials and methods ................................................................................................................ 11
  3.1 Study population .................................................................................................................... 11
    3.1.1 The CSGA population .................................................................................................... 11
    3.1.2 Sample collection ........................................................................................................... 11
    3.1.3 Identification of asthma ................................................................................................ 11
  3.2 Genotyping ............................................................................................................................... 12
3.2.1 DNA extraction ........................................................................................................ 12
3.2.2 The Taqman method ............................................................................................ 12
3.2.3 Genotyping quality control ................................................................................. 12
3.3 Statistical methods ................................................................................................... 13

4 Results .......................................................................................................................... 14

4.1 Clinical characteristics of subjects .......................................................................... 14
  4.1.1 Clinical characteristics of analyzed CSGA subjects ............................................. 14
  4.1.2 Clinical characteristics of CAAPA subjects ......................................................... 14
4.2 Summary of analyzed SNPs ....................................................................................... 15
4.3 GEE analysis of genotyped SNPs ............................................................................. 17
4.4 Allelic discrimination plots of genotyped SNPs ....................................................... 18

5 Discussion .................................................................................................................... 24

5.1 Population selection ................................................................................................ 24
5.2 Clinical characteristics of CSGA subjects compared to CAAPA subjects .......... 25
5.3 Comparison of LD patterns in SNPs on chromosome 2 ........................................ 25
5.4 Analytical methods .................................................................................................. 26
5.5 Direction of effect of genotyped SNPs .................................................................... 27
5.6 Conclusion ................................................................................................................ 27
5.7 Future studies .......................................................................................................... 28

References ..................................................................................................................... 29
List of abbreviations

ADPC: African Diaspora Power Chip
CAAPA: Consortium on Asthma among African-ancestry Populations in the Americas
Chr: Chromosome
CSGA: The Collaborative Studies on the Genetics of Asthma
DNA: Deoxyribonucleic acid
FEV1: forced expiratory rate at one second
GEE: Generalized Estimating Equations
GWAS: Genome-wide association study
HapMap: Haplotype map
HWE: Hardy Weinberg Equilibrium
IgE: Immunoglobulin E
LD: Linkage Disequilibrium
N: Number of subjects
SD: Standard deviation
SEM: Standard error of the mean
SNP: Single nucleotide polymorphism
List of figures

Figure 1: LD pattern of genotyped SNPs on Chr. 2 in CSGA population ........................................ 16
Figure 2: LD pattern of genotyped SNPs on Chr. 2 in SAGE population ...................................... 16
Figure 3: Genotyping call plot for rs10927221 ........................................................................ 18
Figure 4: Genotyping call plot for rs10927221 ........................................................................ 18
Figure 5: Genotyping call plot for rs17834780 ....................................................................... 19
Figure 6: Genotyping call plot for rs17834780 ....................................................................... 19
Figure 7: Genotyping call plot for rs17834780 ....................................................................... 19
Figure 8: Genotyping call plot for rs787160 ........................................................................... 20
Figure 9: Genotyping call plot for rs787160 ........................................................................... 20
Figure 10: Genotyping call plot for rs787160 .......................................................................... 20
Figure 11: Genotyping call plot for rs787151 .......................................................................... 21
Figure 12: Genotyping call plot for rs787151 .......................................................................... 21
Figure 13: Genotyping call plot for rs74532409 ..................................................................... 22
Figure 14: Genotyping call plot for rs74532409 ..................................................................... 22
Figure 15: Genotyping call plot for rs9300316 ....................................................................... 23
Figure 16: Genotyping call plot for rs9300316 ....................................................................... 23
Figure 17: Genotyping call plot for rs9300316 ....................................................................... 23

List of tables

Table 1: Clinical characteristics of analyzed CSGA subjects ......................................................... 14
Table 2: Clinical characteristics of CAAPA subjects .................................................................... 14
Table 3: Summary of analyzed SNPs .......................................................................................... 15
Table 4: GEE analysis of genotyped SNPs ................................................................................. 17
1 Introduction

1.1 Asthma

1.1.1 Definition and diagnosis of asthma

Asthma is a chronic inflammatory disorder of the respiratory tract characterized by the presence of underlying inflammation of the airways, bronchial hyper-responsiveness and reversible airflow obstruction. Asthma presents itself clinically as recurrent episodes of wheezing, breathlessness, chest tightness, and coughing.\textsuperscript{1,2} There are no known biomarkers or tests that can diagnose asthma with complete certainty and therefore, the diagnosis of asthma is based on the presence of clinical symptoms and the exclusion of alternative causes of those symptoms.\textsuperscript{1,3}

Other chronic obstructive lung diseases such as emphysema and chronic bronchitis can mimic the symptoms of asthma but these diseases, unlike asthma, result in a permanent change in patient spirometry.\textsuperscript{1,3} Using spirometry to demonstrate at least partial reversibility of the airflow obstruction asthma patients experience can greatly enhance diagnostic confidence. The term reversibility refers to changes in symptoms as well as in airflow obstruction that occur spontaneously or in response to treatment. An example of reversibility of lung function limitations is rapid improvements in FEV1 (forced expiratory rate at one second) measured within minutes after inhalation of a rapid-acting bronchodilator. Another example would be sustained improvement over days or weeks after the introduction of effective treatment, such as inhaled corticosteroids.\textsuperscript{4}

1.1.2 Treatment of asthma

The treatment of asthma is subject to the severity of the condition. The specific measures used to determine asthma severity are: symptoms, use of Short-Acting Beta Agonists for quick relief of symptoms, limitations to normal activities because of asthma, pulmonary function, and exacerbations. Appropriate treatment is then determined based on asthma severity and the age of the patient in question.\textsuperscript{3}

Medications that are commonly used to treat asthma are: Short-Acting Beta Agonists, Long-Acting Beta Agonists, Inhaled Corticosteroids in a range of doses and Oral Corticosteroids. Because asthma is a chronic inflammatory disorder, persistent asthma is most effectively controlled with daily long-term control medication directed toward suppressing said inflammation. Inhaled Corticosteroids are used to treat persistent asthma and is the most consistently effective anti-inflammatory therapy for all age groups, at all stages of severity. Once treatment is determined asthma control should be monitored regularly. Therapy can then be increased as necessary and decreased when possible if asthma is well controlled.\textsuperscript{3} Presently, a patient’s genetic factors are not taken in to consideration when treatment is selected but could hopefully be an option in the future.
1.1.3 Burden of asthma
Asthma is one of the most common chronic diseases in the world and represents 1% of the total global disease burden. Asthma prevalence is increasing globally, especially among children. It is estimated that as many as 300 million people worldwide suffer from asthma and that by the year 2025 the number is expected to increase to 400 million. The burden of asthma is experienced by the affected individual not only in terms of health care costs but also as lost productivity, time lost from work and reduced participation in family life. It is estimated that asthma accounts for 1 in every 250 deaths worldwide - 250,000 deaths each year. Given that patients receive optimal long-term medical care or timely help during the final attack it is believed that many of these deaths are preventable.

1.1.4 Etiology of asthma
Although the pathophysiology of asthma is rather well understood, the exact etiology is not. Asthma comprises a range of phenotypes with the risk factors for each phenotype depending on both genetic and environmental factors as well as their interactions. Heritability estimates vary considerably between studies and study populations, ranging between 35% and 95%. Asthma affects people of all ages and ethnic backgrounds but ethnic minorities, such as African Americans and Hispanics in the United States, are affected disproportionately.

1.1.5 Asthma in individuals of African descent
Asthma patients of African descent have more severe asthma, more severe clinical symptoms, higher IgE (immunoglobulin E) levels and a higher degree of steroid dependency, than individuals of European descent with asthma. Asthma morbidity and mortality are also disproportionately high among African Americans and they continue to increase. In 2004, African Americans had emergency department visit rates of 350%, hospitalization rates of 240%, and mortality rates of over 200% higher than individuals of European descent. (CDC report, https://www.cdc.gov/nchs/data/hestat/asthma03-05/asthma03-05.htm)

These striking ethnic disparities cannot be explained entirely by environmental, social, cultural, or economic factors, which along with genetic variation, all attribute to the complex risk architecture of the disease. Although it has been well established that genetic factors greatly affect susceptibility to asthma, it is less clear to what extent genetic variation contributes to the ethnic disparities observed.

1.2 Genome-wide association studies of asthma
Genetic studies offer a means of understanding the causes of diseases as well as identifying targets that can potentially be used to treat them. Genome-wide association studies (GWAS) of adequate power are currently the method of choice for identifying genes that influence complex disease such as
asthma. In GWAS, 1-2 million single nucleotide polymorphisms (SNPs) covering the genome are investigated for association with certain phenotypes in large samples of both cases and control subjects. The number of investigated SNPs can then be multiplied to ten times that number through imputation. With the application of GWAS the number of identified asthma susceptibility genes has increased rapidly over the last 10 years.

1.2.1 GWAS technology

GWAS have been made possible by the availability of chip-based microarray technology. DNA (deoxyribonucleic acid) microarrays are a collection of DNA probes attached to a solid surface, or chip. These chips are used to measure the degree of hybridization that occurs between the probes of the chip and SNPs of a genome. In this way it is possible to measure the variation of up to several million SNPs of a specific genome.

It is important to note that chip-based microarrays can only measure the variation of SNPs assembled on the microarray. This can be important depending on the specific human population being studied as linkage disequilibrium can differ greatly between populations of different ancestry.

Linkage disequilibrium (LD) refers to the fact that particular alleles can co-occur on the same haplotype more often than would be expected by chance. The rate of LD is dependent on numerous factors, including population size, the number of founding chromosomes in the population, and the number of generations for which the population has existed. For these reason, it is consistently observed that populations of African ancestry have more genetic variance and decreased LD compared to populations of non-African ancestry.

Although chip-based microarrays only measure the variation of SNPs assembled on the chip itself, by selecting the correct SNPs to study, LD between selected SNPs can be leveraged to maximize the genetic variation information that can be acquired from a single chip with a limited number of DNA probes. SNPs in high LD with one another are very frequently inherited together and in those cases represent a single genetic signal. This way, the SNPs assembled on a microarray can provide information about other SNPs that are not assembled on the microarray, but are in strong LD with the selected SNPs.

Keeping this in mind it is clear that when using chip-based microarrays to study populations of African ancestry it is important to use a microarray that has more SNPs with better overall genomic coverage then one would need for a study of, for example, Europeans.

1.2.2 Advantages of GWAS

Until about ten years ago, the primary two approaches of discovering the genetics associated with inherited diseases were genome-wide linkage studies and candidate-gene association studies. The
main advantage of genome-wide linkage studies is their ability to identify novel genes and pathways whereas candidate-gene association studies, that focus on a selection of genes that are known to be in some way related to a disease, can identify genes that contribute to disease risk with modest effect. 37

Genome-wide association studies have the combined advantages of both of these approaches. GWAS can, through large sample size, identify genes that contribute to disease risk with modest effect and, by considering all regions of the genome without any previous hypotheses about the location of genetic variants that contribute most to risk, identify novel genes and pathways. 36-38

1.2.3 Limitations of GWAS

Typically GWAS test for association with a million or more SNPs across the genome. This large number of test performed results in a great multiple testing correction burden and a requirement for very large samples sizes to achieve statistical significance. Furthermore, when dealing with variants that have a small effect on disease risk, which is typically the case for complex disease, an even larger sample size is required. 36 This is also the case for asthma, and it has been suggested that the total number of genes contributing to risk of disease may exceed 100, with small individual gene effects. 37,38 Consequently, when studying asthma by means of a GWAS, very large sample size is crucial.

The fact that each of the many genes contributing to risk of asthma has small effect, also limits the application of GWAS by reducing the power of GWAS to detect risk variants. Most genotyping platforms used in GWAS include mostly common variants and as a result GWAS primarily detect common risk variants that are represented on those platforms. In addition, GWAS are not powered to detect SNPs with low minor allele frequencies. In fact, GWASs to date, covering multiple diseases including asthma, suggest that the risk variants identified by this approach account for a very small proportion of the genetic risk. A possible explanation for this is that rare variants have an overall larger effect than common ones on risk of complex disease, including asthma, and that these variants may not be detectable by GWASs. 36,39

Yet another limitation of GWAS is that once variants contributing to risk are discovered, the relationship between the variants and pathogenesis is not always obvious, and it can be difficult to select SNPs for further investigation. 6,36 This is particularly true when SNPs are in high LD with each other. 33,38 As mentioned above, when SNPs are in high LD they are very frequently inherited together and in those cases represent a single genetic signal. A single causative locus may drag with it other non-causative variants but with GWAS alone it is may be unclear which variants contribute to risk of disease and which do not. 33,36,38
1.2.4 Meta-analysis of asthma GWAS

One of the limitations of GWAS is the requirement for very large samples sizes to achieve statistical significance. The simplest solution for increasing power is to perform a meta-analysis of GWAS where data from many studies is pooled together. Two such meta-analyses of asthma GWAS have been completed in the last 10 years. The GABRIEL Consortium of European Investigators pooled together studies covering European populations and the EVE Consortium of U.S. Investigators combined results from studies covering the three major ethnic groups in the U.S., European Americans, African American/African Caribbeans, and Latino individuals. These studies produced remarkably similar results. SNPs in or near seven loci were associated with asthma in both studies and in both studies four of these loci had P-values at or near genome-wide levels of significance. The EVE Consortium also identified at least one locus that may contribute to risk only in populations of African descent.

1.2.5 Genotype imputation

When combining data from multiple studies, as is done in a meta-analysis, the data must be comparable. In a perfect world, studies pooled together would all have the same SNPs genotyped using the same genotyping platform but unfortunately, that is not the case. A way to facilitate the combination of data across multiple studies is to use genotype imputation, a process of estimating or imputing genotypes that are not directly assayed in a sample of individuals. That is to say, genotype imputation is a way of accurately predicting the data that is missing. This way, if different cohorts have used different genotyping platforms, as is often the case, imputation can be used to equate the set of SNPs in each study. As a result, all the pooled data becomes comparable and can be analyzed together in a meta-analysis. This method can also boost the number of SNPs that can be tested for association.

1.2.6 GWAS across populations of diverse ancestry

It has been shown that some asthma susceptibility loci differ between populations of different ancestry, and that only studies of diverse populations will allow for their discovery. Furthermore, due to the different patterns of LD between racial groups, studying populations of diverse ancestral backgrounds can increase the resolution of susceptible asthma regions of the genome. In other words, studying the same SNP in multiple populations with varying patterns of LD may differentiate between the causative SNP and the non-causative variants that are inherited together.

In spite of the benefits of studying populations of diverse ancestry most GWAS to date have been conducted in samples of European-ancestry. If we take asthma as an example the largest asthma GWAS in African ancestry populations only included 908 asthma cases whereas the largest asthma GWAS in European ancestry populations included 10,365 asthma cases. Another example is The
Trans-National Asthma Genetic Consortium, which is currently conducting a meta-analysis of more than 100,000 subjects, of which only 2,149 are cases of African ancestry.\textsuperscript{46}

1.3 The CAAPA Consortium and the African Diaspora Power Chip

In order to address the research disparities between different racial groups, the Consortium on Asthma among African-ancestry Populations in the Americas (from now on referred to as the CAAPA Consortium) was established with the main goal of discovering genes that contribute to risk of asthma in populations of African descent.\textsuperscript{46}

One of the obstacles facing researchers wishing to discover the genetic risk factors for asthma in African ancestry populations has been inadequately designed genotyping platforms for non-European populations that do not provide adequate coverage of LD patterns for diverse populations.\textsuperscript{22} To address this problem the CAAPA Consortium develop the African Diaspora Power Chip (ADPC), a gene-centric SNP genotyping array designed to complement commercially available genome-wide chips. The data used to develop the chip was acquired through whole genome sequencing of 642 samples from individuals of African ancestry from 19 North, Central and South American and Caribbean populations, as well as Yoruba-speaking individuals from Ibadan, Nigeria.\textsuperscript{46}

After developing the ADPC the CAAPA program proceeded to utilize the chip to genotype subjects from 9 CAAPA studies. These data were then combined in a genome-wide meta-analysis of asthma. This meta-analysis included a total of 4,827 asthma cases and 5,397 controls and is the largest GWAS of asthma in individuals of African ancestry to date. The study yielded a number of loci that may contribute to risk of asthma specifically in individuals of African ancestry. These signals warrant further investigation in the form of replication studies.\textsuperscript{46}

1.4 Replication studies of GWAS

In modern science, the test of experiment’s reliability has always been reproducibility. The same goes for GWAS and the gold standard for validating a genotype-phenotype association observed in a GWAS is to replicate the association in an independent sample of sufficient size.\textsuperscript{36,45,47} This is done to rule out the possibility that the observed association is a chance finding or due to uncontrolled bias. For a replication study to be accurate it must use the same phenotype, test the same marker and use the same analytical methods as the original study. In addition, initial replication studies should focus on populations with similar genetic ancestry as the sample in the original study. The reason for this is that a variant may be strongly associated with a trait in one population and at the same time have no association in another.\textsuperscript{45} Overall, the strategy for a replication study is to repeat the ascertainment and design of the GWAS as closely as possible, but examine only specific genetic effects found significant in the GWAS. Effects that are consistent across both GWAS and replication study can be labeled replicated effects.\textsuperscript{34}
1.4.1 Replication of CAAPA study

In order to authenticate the findings of the CAAPA study the SNPs that showed greatest association with asthma are to be genotyped in independent samples. At present there are ongoing replication studies with more to follow in the future.

1.4.2. SNPs chosen for replication

SNPs that passed the genome-wide significance threshold of $p < 5 \times 10^{-8}$ in the CAAPA study are to be genotyped in independent samples. The genome-wide significance threshold is defined as the P-value that can, with true significance, distinguish between true positive associations and false positive associations. SNPs that borderline on genome-wide significance, here defined as having $p < 5 \times 10^{-6}$ are also to be replicated.\textsuperscript{46}

In the CAAPA study, two SNPs on chromosome 2q22.3, rs787160 and rs17834780, passed the genome-wide significance threshold of $p < 5 \times 10^{-8}$. SNP rs787151, also located on chromosome 2q22.3, showed borderline genome-wide significance with $p = 5.26 \times 10^{-7}$. SNPs rs10927221, rs74532409 and rs9300316 reached borderline genome-wide significance with $p < 5 \times 10^{-6}$.\textsuperscript{46}
2 Aim of study

The aim of this study was to replicate, and thereby validate, the findings of the Consortium on Asthma among African-ancestry Populations in the Americas in an additional independent European ancestry sample to test for replication and generalizability of CAAPA findings. SNPs that passed the genome-wide significance threshold (rs787160 and rs17834780) along with SNPs with borderline genome-wide significance (rs10927221, rs787151, rs74532409 and rs9300316) were replicated using the method of SNP genotyping.
3 Materials and methods

3.1 Study population

This study was performed using samples and data previously collected for The Collaborative Studies on the Genetics of Asthma (CSGA), which was competed in August 2002. Subjects with available blood or DNA samples and available phenotype data were included.

3.1.1 The CSGA population

The CSGA study was composed of five centers: Johns Hopkins University, University of Chicago, University of Maryland, University of Minnesota, and a data coordinating center at Wake Forest. The original study population was composed of 839 individuals from 111 families from Baltimore and Washington, DC. The families were ascertained at CSGA centers, through two siblings with asthma, and then extended to include other affected relatives. These extensions were made through asthmatic relatives or one unaffected relative. A second round of collection was carried out in Baltimore and included 348 individuals from 116 trios as well as 205 unrelated controls. The total CSGA population, cases and controls, comprised of 1392 subjects. The subjects represented all three major ethnic groups in the U.S. (European Americans, African American/African Caribbeans, and Latino individuals).^{31} (https://clinicaltrials.gov/ct2/show/NCT00005500)

3.1.2 Sample collection

Samples were collected from 1992-2002. DNA was extracted from whole blood, cell pellets or lymphoblastoid lines by standard methods at each CSGA center, excluding the data-coordinating center at Wake Forest. Willing participants were asked to give blood for storage and future genotyping. Adults participating in the study provided written consent or gave verbal consent that was recorded by a witness. Children gave verbal consent and written consent was obtained from a parent or guardian. Children under the age of six, along with children who had a birth weight of less than 2 kilograms or had three or more years of cigarette exposure were excluded from the study.

3.1.3 Identification of asthma

Each subject was characterized through the use of spirometry, bronchial responsiveness to methocholine or reversibility testing, and questionnaire data. Baseline spirometry was performed according to American Thoracic Society criteria.^{48} FEV1 was increased 12% for African-Americans to adjust for racial differences^{49} and the standard ATS questionnaire was modified to include questions about the frequency, severity and duration of symptoms of asthma.^{41}

To be classified asthmatic the siblings in each family had to meet certain criteria. Each child showed either a fall in baseline FEV1 by ≥20% at ≥25 mg/ml methocholine or ≥15% increase in baseline FEV1
after bronchodilator use. All children had two or more of the defined symptoms of asthma (cough, wheeze, dyspnea) and had no other conflicting pulmonary diagnosis.41

Other subjects were classified as asthmatic if they reported a past or current history of at least two of the three determined symptoms of asthma (cough, wheeze, dyspnea), had been previously diagnosed with asthma by a doctor and demonstrated either bronchial hyper-responsiveness or airway reversibility.41

3.2 Genotyping

3.2.1 DNA extraction
For subjects who did not have available extracted DNA samples, DNA was extracted from stored whole blood samples (7-mL) using an AutoGen FLEX STAR (Autogen, Massachusetts, USA). Lysis buffer (FG1) was used to break open the cell membrane, and cell nuclei and mitochondria were pelleted by centrifugation. Proteins and contaminants were removed by incubating the pellet in a denaturation buffer (FG2), which contained a chaotropic salt and QIAGEN Protease. DNA was precipitated by addition of isopropanol. The DNA was then recovered by centrifugation, washed in 70% ethanol and dried. The DNA was finally re-suspended in a hydration buffer (FG3). Buffers FG1, FG2, FG3 and QIAGEN Protease were manufactured by AutoGen.

3.2.2 The Taqman method
Six SNPs were genotyped on 633 DNA samples using a TaqMan ABI 7900HT (Applied Biosystems, Foster City, CA, USA). PCR was conducted using 2.5µl TaqMan Genotyping Master Mix (Applied Biosystems, Part No. 4351374), a TaqMan-validated assay manufactured by Applied Biosystems for each SNP, and 10 ng DNA in a total volume of 5µl per well in 384 micro-plates. Thermal cycling conditions were as follows: 95°C for 10 minutes, followed by 42 cycles of 95°C for 15 seconds, followed by 60°C for 1 minute and an extension step of 60°C for 5 minutes.

Allelic discrimination was detected by ABI 7900 HT Sequence Detection System using software SDS2.4 (Applied Biosystems, Foster City, CA, USA).

3.2.3 Genotyping quality control
Non-template negative and genotyping-positive controls along with three HapMap controls (mother, father and proband) were included in each genotyping plate. Automatic calling was performed with a quality value of > 99%.
3.3 Statistical methods

Hardy Weinberg testing and tests for Mendelian inconsistencies were performed using PLINK. Statistical analysis was done using the generalized estimating equations (GEE) method with an exchangeable covariance matrix to incorporate related individuals into the model.
4 Results

4.1 Clinical characteristics of subjects

4.1.1 Clinical characteristics of analyzed CSGA subjects

Table 1 shows the clinical characteristics of the analyzed subjects in the CSGA population. All analyzed subjects were of European ancestry. Founders of the population, defined as unrelated individuals, were mostly parents of the asthmatic children ascertained at CSGA centers. Therefore, this group is older and less asthmatic than the total population. 29.38% of subjects had missing phenotype information and were therefore neither defined as asthmatics or non-asthmatics. SD = Standard deviation. N = number of subjects.

<table>
<thead>
<tr>
<th>Trait</th>
<th>All subjects</th>
<th>Founders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Asthmatics</td>
</tr>
<tr>
<td>N</td>
<td>633</td>
<td>201</td>
</tr>
<tr>
<td>Males (N; %)</td>
<td>283 (45.1%)</td>
<td>90 (44.8%)</td>
</tr>
<tr>
<td>Age (Mean; SD)</td>
<td>35.3 (18.0)</td>
<td>21.5 (13.0)</td>
</tr>
</tbody>
</table>

4.1.2 Clinical characteristics of CAAPA subjects

Table 2 shows the clinical characteristics of the CAAPA subjects. The information summarized in this table was not available for all subjects included in the association analysis; therefore the number of subjects (N) used to calculate these statistics is given in brackets. SEM = standard error of the mean.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Total</th>
<th>Asthmatics</th>
<th>Non-Asthmatics</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>10,287</td>
<td>4,827</td>
<td>5,397</td>
</tr>
<tr>
<td>Males (N; %)</td>
<td>4,398</td>
<td>45.2% (N=4,750)</td>
<td>40.3% (N=5,332)</td>
</tr>
<tr>
<td>Age (Mean; ± SEM)</td>
<td>21.735 (5.038)</td>
<td>20.64 (4.745)</td>
<td>22.83 (5.331)</td>
</tr>
</tbody>
</table>
4.2 Summary of analyzed SNPs

Table 3 provides information on the genotyped SNPs. It shows their chromosomal positions, the major and minor alleles, the gene nearest to the SNPs, the genomic region the SNPs are located in and their P-values in the CAAPA study. SNPs rs17834780, rs787160 and rs787151 are all located on chromosome 2q22.3. SNPs rs787160 and rs17834780 had CAAPA P-values lower than $5 \times 10^{-8}$ and therefore reached genome-wide significance. The other four SNPs had CAAPA p-values lower than $5 \times 10^{-6}$ and therefore reached borderline genome-wide significance.

Table 3: Summary of analyzed SNPs

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chromosome: hg19 position</th>
<th>Major allele</th>
<th>Minor allele</th>
<th>Nearest gene</th>
<th>Region</th>
<th>CAAPA P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10927221</td>
<td>1:244532963</td>
<td>T</td>
<td>C</td>
<td>C1orf100</td>
<td>Intronic</td>
<td>$8.91 \times 10^{-7}$</td>
</tr>
<tr>
<td>rs17834780</td>
<td>2:144638809</td>
<td>G</td>
<td>A</td>
<td>GTDC1</td>
<td>Intergenic</td>
<td>$2.97 \times 10^{-8}$</td>
</tr>
<tr>
<td>rs787160</td>
<td>2:144650140</td>
<td>T</td>
<td>A</td>
<td>GTDC1</td>
<td>Intergenic</td>
<td>$6.51 \times 10^{-9}$</td>
</tr>
<tr>
<td>rs787151</td>
<td>2:144655836</td>
<td>G</td>
<td>A</td>
<td>GTDC1</td>
<td>Intergenic</td>
<td>$5.26 \times 10^{-7}$</td>
</tr>
<tr>
<td>rs74532409</td>
<td>10:105730220</td>
<td>G</td>
<td>A</td>
<td>SLK</td>
<td>Intronic</td>
<td>$9.52 \times 10^{-7}$</td>
</tr>
<tr>
<td>rs9300316</td>
<td>12:111196204</td>
<td>T</td>
<td>C</td>
<td>PPP1CC</td>
<td>Intergenic</td>
<td>$8.83 \times 10^{-7}$</td>
</tr>
</tbody>
</table>
Figure 1 shows a LD plot for the three genotyped SNPs located on chromosome 2q22.3 in the CSGA population. The three SNPs are all a part of the same haplotype block and are frequently inherited together. SNP rs787160 and rs787151 are almost in complete LD.

Figure 1: LD pattern of genotyped SNPs on Chr. 2 in CSGA population

Figure 2 shows a LD plot for the three genotyped SNPs located on chromosome 2q22.3 in the SAGE study, one of the studies included in the CAAPA meta-analysis. Since the SAGE study is the largest African American study within CAAPA, with only unrelated subjects, it is the best representative study of African Americans in the CAAPA study. The three SNPs are all a part of the same haplotype block but are inherited together less frequently than in the CSGA population.

Figure 2: LD pattern of genotyped SNPs on Chr. 2 in SAGE population
4.3 GEE analysis of genotyped SNPs

Table 4 shows the results of the GEE analysis of the genotyped SNPs. Analysis was performed adjusting for batch. The table shows Hardy Weinberg equilibrium (HWE) as well as association between genotyped SNPs and asthma in individuals in the CSGA population. All SNPs were in Hardy Weinberg equilibrium and were therefore included in the analysis. None of the genotyped SNPs showed significant association with asthma. However, the direction of effect was the same as in the original study for all SNPs excluding rs74532409. Statistical analysis could not be performed on SNP rs74532409 as the SNP was monomorphic in the CSGA population.

Table 4: GEE analysis of genotyped SNPs

<table>
<thead>
<tr>
<th>Marker</th>
<th>HWE</th>
<th>Estimation</th>
<th>Standard error</th>
<th>Log(P)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10927221</td>
<td>0.1166</td>
<td>-0.0160</td>
<td>0.0349</td>
<td>0.8228</td>
<td>0.6464</td>
</tr>
<tr>
<td>rs17834780</td>
<td>0.8191</td>
<td>-0.0404</td>
<td>0.0317</td>
<td>0.7180</td>
<td>0.2016</td>
</tr>
<tr>
<td>rs787160</td>
<td>0.8519</td>
<td>-0.0136</td>
<td>0.0722</td>
<td>0.4690</td>
<td>0.8505</td>
</tr>
<tr>
<td>rs787151</td>
<td>0.8578</td>
<td>-0.0235</td>
<td>0.0459</td>
<td>0.2750</td>
<td>0.6085</td>
</tr>
<tr>
<td>rs74532409</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs9300316</td>
<td>1</td>
<td>0.0056</td>
<td>0.0767</td>
<td>0.6920</td>
<td>0.9416</td>
</tr>
</tbody>
</table>
4.4 Allelic discrimination plots of genotyped SNPs

Figures 3 and 4 show the allelic discrimination plots of SNP rs10927221. The two plots show the results of genotyping of the 633 samples in two 384 well plates. Individuals homozygous for the major allele T are shown in blue. Individuals homozygous for the minor allele C are shown in red. Heterozygous individuals are shown in green. Undetermined subjects are shown with a black X. The pattern of effect was the same as in the original study.

Figure 3: Genotyping call plot for rs10927221

Figure 4: Genotyping call plot for rs10927221
Figures 5-7 show the allelic discrimination plots of SNP rs17834780. The three plots show the results of genotyping of the 633 samples in three 384 well plates. Individuals homozygous for the major allele G are shown in blue. Individuals homozygous for the minor allele A are shown in red. Heterozygous individuals are shown in green. Undetermined subjects are shown with a black X. The pattern of effect was the same as in the original study.
Figures 8-10 show the allelic discrimination plots of SNP rs787160. The three plots show the results of genotyping of the 633 samples in three 384 well plates. Individuals homozygous for the major allele T are shown in red. Individuals homozygous for the minor allele A are shown in blue. Heterozygous individuals are shown in green. Undetermined subjects are shown with a black X. The pattern of effect was the same as in the original study.
Figures 11-13 show the allelic discrimination plots of SNP rs787151. The three plots show the results of genotyping of the 633 samples in three 384 well plates. Individuals homozygous for the major allele G are shown in blue. Individuals homozygous for the minor allele A are shown in red. Heterozygous individuals are shown in green. Undetermined subjects are shown with a black X. The pattern of effect was the same as in the original study.
**Figures 14 and 15** show the allelic discrimination plots of SNP rs74532409. The two plots show the results of genotyping of the 633 samples in two 384 well plates. The SNP was monomorphic in the CSGA population and therefore all individuals are homozygous for the major allele G and are shown in red. Undetermined subjects are shown with a black X.

**Figure 14:** Genotyping call plot for rs74532409

**Figure 15:** Genotyping call plot for rs74532409
Figures 16 and 17 show the allelic discrimination plots of SNP rs9300316. The two plots show the results of genotyping of the 633 samples in two 384 well plates. Individuals homozygous for the major allele T are shown in blue/red. Individuals homozygous for the minor allele C are shown in red/blue. Heterozygous individuals are shown in green. Undetermined subjects are shown with a black X. The pattern of effect was the same as in the original study.

Figure 16: Genotyping call plot for rs9300316

Figure 17: Genotyping call plot for rs9300316
5 Discussion

5.1 Population selection

The aim of this study was to replicate, and thereby validate, the findings of the CAAPA study in an independent European ancestry sample. The CSGA population was selected, partly, due to the availability of the samples for genotyping. A portion of the CSGA population, specifically individuals of African ancestry in the CSGA population, had already been genotyped on the ADPC as part of the CAAPA study. Those individuals were thereby no longer available for use in replication of the CAAPA study, as they were no longer an independent sample population. As a result the European ancestry portion of the CSGA population was selected as the independent sample population for this replication study. Although initial replication studies generally focus on populations with similar genetic ancestry to the original study population, replication of the CAAPA findings in a population of European ancestry was of interest for a number of reasons. It is also important to note that there are currently ongoing CAAPA replication studies that focus on populations of African ancestry.

SNPs rs787160 and rs17834780, located on chromosome 2q22.3, were the only SNPs to reach genome wide significance in the CAAPA study. To our knowledge, associations in this region have only been reported once, in a GWAS of diisocyanate-induced occupational asthma in individuals of European ancestry. Therefore, the replication of these SNPs in a population of European ancestry was of interest.

Another reason for replicating rs787160 in a population of European ancestry is the fact that the stratified local ancestry models, used in the CAAPA study, did not identify the SNP as having any African specific effect. This suggests that the association may exist in populations other than those of African ancestry. Even though rs787160 was the top SNP in the CAAPA study, it has never before been identified by a GWAS. This may be due to the fact that it is located one base pair away from an insertion-deletion polymorphism and was therefore not included as part of the reference panel in the 1000 Genomes Project, the most detailed catalogue of human variation and genotype data available.

SNP rs787151 is in fairly strong LD with SNPs rs787160 and rs17834780, and very nearly reached the threshold for genome-wide significance, and was therefore also of interest to replicate in a population of European ancestry. The other three SNPs (rs10927221, rs74532409 and rs9300316) have, to our knowledge, never before been identified by and asthma GWAS. Therefore, we had no prior assumptions regarding whether or not they would show association with asthma in populations other than those of African ancestry, but we were interested in finding out.
5.2 Clinical characteristics of CSGA subjects compared to CAAPA subjects

For a replication study to be accurate it is ideal to repeat the design of the original study as closely as possible. Having the replication population as similar as possible to the CAAPA population was therefore a concern.

The distribution of gender in the CSGA and CAAPA populations is similar. Asthmatics are 44.8% male in the CSGA population and 45.2% male in the CAAPA population. Non-asthmatics are 42.5% male in the CSGA population and 40.3% male in the CAAPA population.

The mean age of asthmatics is similar in the CSGA and CAAPA populations, 21.5 years and 20.64 years respectively. The mean age of total subjects in the CSGA population is, on average, 14.4 years higher than in the CAAPA population. The mean age of non-asthmatics in the CSGA population is, on average, 14.5 years higher in the CAAPA population. This is a result of the CSGA population being significantly smaller that the CAAPA population and the older and more non-asthmatic founders in the CSGA populations representing a proportionately larger fraction of the CSGA population that they do in the CAAPA population. However, the large age difference in controls is not a concern given that we are still comparing two groups of adults.

The proportion of cases and controls is the least comparable factor between study populations. In the CSGA population asthmatics were 31.75% compared to 46.96% asthmatics in the CAAPA population. Non-asthmatics were 38.86% in the CSGA population compared to 52.46% in the CAAPA population. 29.38% of individuals in the CSGA population had incomplete phenotype information compared to only 0.60% in the CAAPA population. The CSGA population was a family-based cohort, so many individuals part of those families were included in the cohort, even though their asthma status was not ascertained. As a result, we have many individuals with incomplete phenotype data, which skews the case-control proportions. The CAAPA cohorts were mostly part of case-control designs were individuals with incomplete phenotype data were not included. This case of missing phenotype data is less of a concern in terms of comparability of the populations and more of a concern in terms of the analytical methods available and statistical power of the study. This is explained further below.

5.3 Comparison of LD patterns in SNPs on chromosome 2

When we compare the LD patterns of the three SNPs on chromosome 2q22.3 (rs17834780, rs787160, rs787151) in the two study populations, we see that the patterns adhere to the general rule of LD being less in the African ancestry CAAPA population compared to the CSGA population of European descent. Different patterns of LD between populations of different ancestry can have an affect on study results. However, it is unlikely that this is the reason we did not observe significant association between the genotyped SNPs on chromosome 2q22.3 and asthma, in this particular replication study.
As the CSGA population is of European descent, we can assume that the population has less genomic variance than an African ancestry population, and that patterns of LD stretch over larger regions of the genome. In other words, we can assume that the CSGA population has larger haplotype blocks than the CAAPA population. Therefore, if, as the CAAPA study suggests, the SNPs on chromosome 2q22.3 were haplotype tags for a causal asthma variant in the CAAPA population, they would also be tags for a causal asthma variant in the CSGA population.

For these reasons, we cannot argue that different patterns of LD observed between the replication study population and the original study population attributed to the insignificant findings of the study.

5.4 Analytical methods

For a replication study to be as accurate as possible it is ideal to use the same analytical methods as were used in the original study. A number of analytical models were used in the CAAPA study depending on the characteristics of the cohort in question.

The GENESIS package in R uses logistic mixed effect models for cluster analysis of microarray data to visualize and analyze whole set of gene expression experiments. The GENESIS package was used for association testing in one of the CAAPA cohorts. In this cohort, there was a mixture of related and unrelated individuals, similar to the CSGA population. The GENESIS analysis method would have been the analytical method of choice for this study except for one important limitation. The GENESIS test uses a kinship matrix that only encapsulates relatedness and excludes other sources of variance. This matrix, however, can only be used if detailed genetic information of the subjects, obtained through whole-genome sequencing, is available. As whole-genome sequencing had not been performed on the CSGA subjects, the GENESIS analysis method was not suitable for this study.

The MQLS test was designed for genome-wide association testing of binary traits where a portion of the analyzed individuals are related with known pedigrees. MQLS allows both linkage and association to contribute to the test statistic, and also leverages genetic information from subjects with unknown phenotypes to maximize statistical power. The MQLS test was used for association testing in a CAAPA cohort with similar characteristics to the CSGA cohort. The particular CAAPA cohort was comprised of families ascertained through asthmatic cases and included subjects with unknown phenotypes. The MQLS method would therefore have been suitable for this study except for the fact that the MQLS method assumes that the known pedigrees within the population are accurate. As previously stated, whole-genome sequencing had not been performed on the CSGA subjects and therefore it was not possible to confirm the pedigrees of the subjects through genetics. When researching affected individuals with affected relatives, the prevalence of disease alleles will be higher than in the general population. It is therefore of vital importance to take pedigrees into account when analyzing study results in order to obtain accurate data. Although the MQLS analytical method would
have maximized the statistical power of the study, it could also have compromise the accuracy of the acquired data.

The GEE method, like the MQLS method, was designed to be used for association testing in populations where a portion of the individuals are related. What makes GEE different from MQLS is the fact that GEE can handle possible unknown correlation within families. That is, GEE can handle the unconfirmed pedigrees of our study population, which eliminates the requirement for whole-genome sequencing of the subjects. By using the GEE method we secure the most accurate data we can obtain without having the more detailed genetic information the other analytical methods require. Unlike MQLS, The GEE method can only use information from subjects with known phenotypes in the analysis. Therefore, by using the GEE analytical method we secure accurate data, but lose statistical power.

Ultimately, although not used in the CAAPA study, the GEE analytical method was selected, with the goal of obtaining as accurate data as possible. However, by using GEE, statistical power of the study was diminished. Because the sample population was not large to begin with, it is likely that this choice of analytical method contributed to the insignificant results of the study.

5.5 Direction of effect of genotyped SNPs

The genotyped SNPs did not show significant association with asthma in the CSGA population. However, in both the CSGA and CAAPA populations, the same allele was present more often in the asthmatic cases compared to controls for five of the six SNPs. As the number of CSGA cases with this allele was no different than what would be expected by chance, the results of this study were insignificant. However, for these five SNPs we can say that the direction of effect was the same as in the original study population. There is therefore a possibility, that in a study with more statistical power, that some of these SNPs could show association with asthma in populations of European ancestry. This would require further investigation.

For SNP rs74532409, we did not observe the same pattern of effect in both the CSGA and CAAPA populations. In the CSGA population this SNP was monomorphic, meaning all the individuals of the population, asthmatics and non-asthmatics, had the same allele. It is therefore likely that this SNP is not associated with asthma in populations of European ancestry.

5.6 Conclusion

The genotyped SNPs in this study did not show significant association with asthma in the European ancestry study population. However, the direction of effect was the same as in the original study for five of the six SNPs. It is likely that small sample size combined with less than optimal analytical methods contributed to the insignificance of these results. It is possible that in a study with more statistical power that some of the genotyped SNPs, excluding rs74532409, could show association
with asthma in a European population. This would require further investigation. One of the genotyped SNPs, rs74532409, was monomorphic in the European ancestry study population. It is likely that this SNP is not associated with asthma in populations of European ancestry.

5.7 Future studies

The CAAPA study was established with the main goal of discovering genes that contribute to risk of asthma in populations of African descent. Future studies will therefore focus on increasing the information already obtained on the genetics of asthma among African-ancestry populations. Further studies are underway and the genotyping of an additional ~5000 case and control samples of African descent has already begun. These samples represent subjects from the United States, Jamaica, Brazil, Colombia, Honduras, and Peru and will hopefully give more insight into the genetic variations of asthma specific to diverse group of similar ancestry.

Another project of interest for the CAAPA study is the analysis of asthma in 23andMe data. 23andMe is a direct to consumer genetics testing company with more than two million genotyped customers around the world. (23andMe website, https://www.23andme.com) 23andMe data could prove to be a valuable source of accessible genetic information from a large group of subjects.
References

32. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.