A Large Scale Genome Wide Association Study of Asthma in the 23andMe Cohort


Introduction
Asthma and allergic disease are common, chronic conditions with substantial public health burden. We carried out a genome-wide association study of self-reported asthma in the 23andMe participant cohort, including more than 20,000 cases and 98,000 controls with European ancestry, imputed against 1000 Genomes reference haplotypes.

We identified a total of 17 associations at the P<5e-8 level, including replications of established asthma GWAS findings, loci that have been associated with other forms of atopic or autoimmune disease, and loci that have not previously been linked to an immunological phenotype.

Most genetic associations with asthma and allergy are shared, with similar effect sizes, and identifying risk factors that distinguish between different allergic disease types seems more challenging. We further explore this issue with analyses of the susceptibility loci stratified by disease subgroups, determined by age of onset, triggers, and severity.

Methods
23andMe participants were genotyped on custom versions of either the Illumina HumanHap550 or the OmniExpress, then imputed against the August 2010 release of 1000 Genomes haplotypes, using Beagle for phasing and Minimac for imputation. Analyses were restricted to a subset of participants with primary European ancestry, and close relatives were excluded. For the GWAS, association tests were performed using logistic regression with covariates for age, gender, and the top five principal components, and results were adjusted using genomic control.

We identified participants who self-reported ever having been diagnosed with asthma through web-based surveys. We combined asthma data from several public health registries, including an opt-out (No+OptOut) cigarette smoke (N=2959), cold, flu, or sinus infection (N=1930), dust (N=3093); exercise (N=4888); seasonal changes (N=2887); and workplace (N=2933). The asthma symptom survey included four questions about quality of life, which were answered by 3300 participants:

- “In the past year, how often did your asthma keep you from getting as much as you would like done at work, school, or home?”
- “During the past year, how often have you had shortness of breath? (Not at all, less than once a week, once or twice a week, three to six times a week, once a day, or more than once a day)”
- “During the past year, how often did your asthma symptoms (wheezing, coughing, shortness of breath, chest tightness, or pain) wake you up at night or earlier than usual in the morning”?
- “During the past year, how often have you used your rescue inhaler or a nebulizer (mist inhaler) medication such as albuterol?”

We also performed the results for 12 loci reported in our recent GWAS of allergy, excluding loci that were genome-wide significant in the asthma GWAS. No loci had effects in the expected direction, and 9 of the 12 were significant at the P<0.01 level (Table 3).

We also conducted an eQTL analysis using data from 23andMe participants in the 23andMe cohort, including more than 20,000 cases and 98,000 controls with European ancestry, imputed against 1000 Genomes reference haplotypes.

In an analysis comparing adult cases versus controls, we see modest evidence for association (P(0.01) at the HLA-DQB1, HLA-F1, and D2HGDH loci.

In the analyses of quality-of-life measures in asthma cases, no associations passed a threshold for significance adjusted for multiple comparisons (P<0.05/17=0.0003). The strongest signals were for IL33 rs1342562 with overall severity (P=0.0003) and rescue inhaler frequency (P=0.005). Similarly, for association with individual triggers, no tests were significant (P<0.05*(17)=0.0004).

Conclusion
Asthma and allergic disease seem to be particularly good phenotypes for the 23andMe phenotypic data collection model. We have collected a large number of these phenotypes, and have been very successful in replicating and extending past work on the genetics of susceptibility to these conditions.

Our attempts to detect differences in disease characteristics associated with individual loci have been less successful. This is almost certainly due in part to the smaller sample sizes available to us for these measures; however, it is also likely that GWAS hits are more likely to be loci that contribute to susceptibility broadly, and are less likely to show strong heterogeneity across disease subtypes. These phenotypes may also be more difficult to accurately self-report than simple questions about disease diagnosis.

Acknowledgments
We thank the customers of 23andMe who answered surveys, as well as the employees of 23andMe, who together made this research possible. This work was supported in part by the National Heart Lung and Blood Institute of the National Institutes of Health under grant number 1R43HL115873-01.