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

D. A. Hinds[1], C. Tian[1], A. K. Kiefer[1], J. L. Mountain[1], N. Eriksson[1], J. Y. Tung[1].

[1]23andMe, Inc., Mountain View, CA.

Introduction

Asthma and allergic disease are common, chronic conditions with substantial public health burdens. 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 primarily 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 sources, including a general medical history, a survey on allergies and asthma, and a survey on asthma symptoms (Table 1).

	Cases	Controls
Total N	20377	98172
Male	9915	56938
Female	10462	41234
Age (0,30]	3679	15652
(30,45]	6539	29801
(45,60]	5071	25070
(60,Inf]	5088	27649

Table 1. Demographics of the GWAS cohort.

Age of diagnosis was provided by 8143 participants, with 5059 reporting a diagnosis at age 18 or younger, and 3027 reporting age > 18.

The allergies and asthma survey and asthma symptoms survey both include questions about asthma triggers (total N=9449 responses), with common responses including air pollution (N=2555); cigarette smoke (N=2959); cold, flu, or sinus infection (N=4503); dust (N=3953); exercise (N=4389); seasonal changes (N=3882); and weather changes (N=2353).

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 combined results for the four questions to give an overall score for severity.

Results

In the GWAS of any asthma diagnosis, we successfully tested a total of 7.4 million variants, with a genomic control inflation factor of 1.11. We identified 17 loci at a significance level of P<5e-8 (Fig. 1, Table 2).

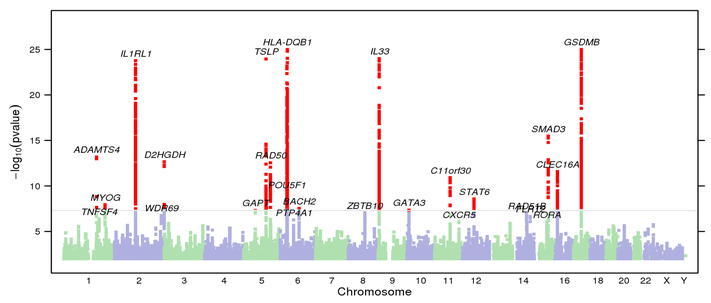


Figure 1. Manhattan plot of asthma GWAS results.

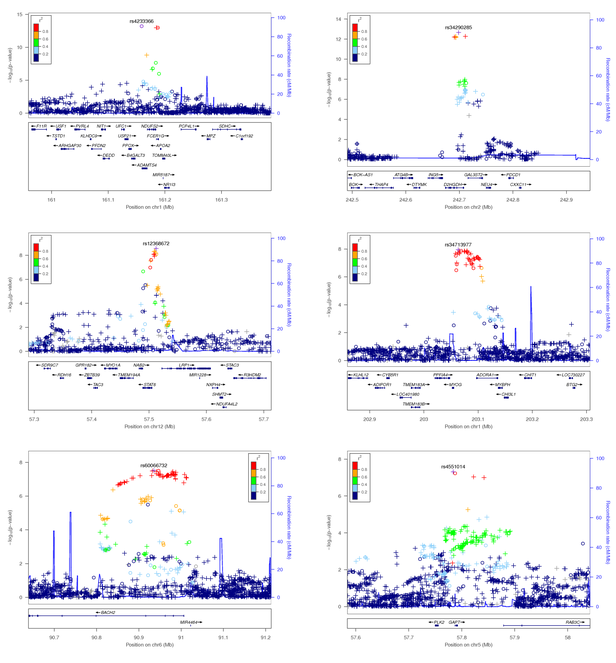
band	SNP	alleles	OR	95% CI	P	gene context
17q12	rs869402	T/C	1.195	[1.168,1.223]	1.0e-53	[GSDMB]
6p21.32	rs28407950	T/C	1.242	[1.200,1.286]	9.3e-35	HLA-DQA1--[]HLA-DQB1
9p24.1	rs1342326	A/C	1.169	[1.135,1.204]	9.9e-25	RANBP6---[]--IL33
5q22.1	rs1837253	T/C	1.152	[1.121,1.184]	1.1e-24	SLC25A46---[]-TSLP
2q12.1	rs60227565	A/G	1.191	[1.152,1.233]	1.7e-24	IL1RL2--[]--IL1RL1
15q22.33	rs56375023	G/A	1.117	[1.088,1.147]	3.4e-16	[SMAD3]
1q23.3	rs4233366	C/T	1.103	[1.075,1.131]	6.4e-14	B4GALT3--[]ADAMTS4
2q37.3	rs34290285	A/G	1.118	[1.085,1.152]	2.2e-13	[D2HGDH]
5q31.1	rs9687749	G/T	1.106	[1.077,1.136]	2.8e-13	[RAD50]
16p13.13	rs35441874	A/T	1.098	[1.070,1.127]	2.7e-12	[CLEC16A]
11q13.5	rs2155219	G/T	1.081	[1.057,1.106]	1.3e-11	C11orf30--[]--LRRC32
6p21.33	rs1128175	A/G	1.091	[1.061,1.121]	4.9e-10	POU5F1--[]--HLA-C
12q13.3	rs12368672	C/G	1.075	[1.050,1.101]	2.7e-09	STAT6-[]-LRP1
1q32.1	rs34713977	A/C	1.074	[1.048,1.101]	1.1e-08	MYOG-[]--ADORA1
6q15	rs60066732	G/A	1.070	[1.045,1.096]	3.1e-08	[BACH2]
10p14	rs1663693	T/C	1.081	[1.051,1.112]	4.3e-08	GATA3---[]
5q11.2	rs4551014	A/C	1.065	[1.041,1.090]	4.8e-08	PLK2--[]-GAPT

Table 2. Lead SNPs in the asthma GWAS, with new associations in red.

Using conditional logistic regression, we see evidence for multiple independent associations at the *IL1RL1* locus on chromosome 2, the *TSLP* locus on chromosome 5, and the *IL33* locus on chromosome 9.

New associations with P<5e-08:

- 1q23.3: near *FCER1G*, gamma subunit of high affinity IgE receptor; rs4233366 is also reported to be an eQTL for *FCER1G* in monocytes.
- 2q37.3: *D2HGDH,* or mitochondrial D-2-hydroxyglutarate dehydro-genase, has no clear connection with immunological processes. Nearby gene *GAL3ST2*, or galactose-3-O-sulfotransferase, is involved in biosynthesis of sulfated mucins in goblet cells.
- 16p13.13: variation in *CLEC16A* has been associated with a variety of autoimmune diseases, as well as allergy and allergic sensitization.
- 12q13.3: near *STAT6*, signal transducer and activator of transcription 6: has been associated with allergy and allergic sensitization.
- 1q32.1: near *ADORA1*, adenosine A1 receptor. There is a reported association between *ADORA1* variants and aspirin-intolerant asthma.
- 6q15: in *BACH2*, basic leucine zipper transcription factor 2. Variants in moderate LD (r2>0.5) have been reported to be associated with celiac disease, vitiligo, T1D, Crohn's disease, and multiple sclerosis.
- 5q11.2: lead SNP rs4551014 is in moderate LD (r2>0.5) with eQTLs for *GAPT*, or transmembrane GRB2-binding adapter protein.



We also looked up test results for 12 loci reported in our recent GWAS of allergy, excluding loci that were genome-wide significant in the asthma GWAS. All loci had effects in the expected direction, and 9 of the 12 were significant at the P<0.01 level (Table 3).

Band	SNP	alleles	OR	95% CI	P	gene context
4p14	rs7653908	C/G	1.048	[1.018,1.078]	0.0013	[TLR10]
5p13.1	rs7720838	G/T	1.037	[1.013,1.061]	0.0020	[]---PTGER4
2q33.1	rs10497813	T/G	1.037	[1.014,1.061]	0.0016	[PLCL1]
3q28	rs9860547	G/A	1.050	[1.026,1.075]	3.3e-05	[LPP]
20q13.2	rs6021270	C/T	1.045	[0.995,1.099]	0.077	[NFATC2]
4q27	rs17388568	G/A	1.060	[1.034,1.087]	6.2e-06	[ADAD1]
14q21.1	rs1998359	C/G	1.028	[0.999,1.058]	0.058	FOXA1--[]---TTC6
8q21.13	rs6473223	C/T	1.067	[1.042,1.092]	9.1e-08	TPD52---[]---ZBTB10
2p25.1	rs10174949	A/G	1.056	[1.029,1.083]	3.0e-05	[]---ID2
16p12.1	rs2107357	G/A	1.072	[1.038,1.108]	2.8e-05	IL4R--[]--IL21R
1p36.22	rs2056417	A/G	1.041	[1.016,1.067]	0.0013	[PEX14]
11q24.3	rs10893845	T/G	1.019	[0.997,1.043]	0.096	[]---ETS1

Table 3. Results for allergy loci, excluding loci in Table 2.

We tested the lead SNPs for the 17 genome-wide-significant loci in Table 2 for association with secondary phenotypes including age of onset, quality of life metrics, and triggers.

We performed a case-only analysis to assess association between log-transformed age of onset and SNP genotype. Generally, the higher risk allele for asthma was associated with lower age of onset (Table 4).

	SNP	effect	stderr	P	gene context
1	rs869402	-0.092	0.012	2.2e-15	[GSDMB]
2	rs28407950	-0.046	0.018	0.010	HLA-DQA1--[]HLA-DQB1
3	rs1342326	-0.069	0.015	4.9e-06	RANBP6---[]--IL33
4	rs1837253	-0.067	0.014	2.1e-06	SLC25A46---[]-TSLP
5	rs60227565	-0.040	0.018	0.023	IL1RL2--[]--IL1RL1
6	rs56375023	-0.030	0.014	0.029	[SMAD3]
7	rs4233366	-0.038	0.013	0.0038	B4GALT3--[]ADAMTS4
8	rs34290285	-0.017	0.016	0.30	[D2HGDH]
9	rs9687749	-0.033	0.014	0.018	[RAD50]
10	rs35441874	-0.028	0.014	0.036	[CLEC16A]
11	rs2155219	-0.038	0.012	0.0013	C11orf30--[]--LRRC32
12	rs1128175	-0.027	0.014	0.061	POU5F1--[]--HLA-C
13	rs12368672	-0.025	0.012	0.042	STAT6-[]-LRP1
14	rs34713977	0.013	0.013	0.31	MYOG-[]--ADORA1
15	rs60066732	-0.019	0.012	0.12	[BACH2]
16	rs1663693	-0.022	0.014	0.11	GATA3---[]
17	rs4551014	0.004	0.012	0.75	PLK2--[]-GAPT

Table 4. Case-only tests for effects on age of onset.

In an analysis comparing just adult cases versus controls, we see modest evidence for association (P<0.01) at the *HLA-DQB1*, *IL1RL1*, 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/(5*17) = 0.0006). The strongest signals were for IL33 rs1342326 with overall severity (P=0.003) and rescue inhaler frequency (P=0.005). Similarly, for association with individual triggers, no tests were significant (P < 0.05*(8*17) = 0.0004).

Conclusion

Asthma and allergic disease seem to be particularly good phenotypes for the 23andMe phenotypic data collection model. We have collected large sample sizes for 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 true 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 diagnoses.

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.