Genetics of Allergy and Related Phenotypes in Participant Driven and Cross Sectional Cohorts

D.A. Hinds[1], G. McMahon[2], A.K. Kiefer[1], C.B. Do[1], N. Eriksson[1], M. Curran[3], M. Loza[3], D. Talantov[3], N.J. Timpson[2], D.M. Evans[2], B. StPourcain[4], S.M. Ring[4], K.C. Nadeau[5], D. Miralles[3], G. Davey-Smith[2], J.Y. Tung[1].

[1]23andMe, Inc, Mountain View, CA; [2]MRC CAiTE Centre/School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; [3]Pharmaceutical R&D, Johnson & Johnson, San Diego, CA, USA; [4]School of Social and Community Medicine, Oakfield House, Oakfield Grove, Bristol, United Kingdom; [5]Division of Allergy and Clinical Immunology, Department of Pediatrics, Stanford University, Stanford, CA, USA.

Introduction

We have performed a meta-analysis of genome wide association data for self reported allergy across participants in the 23andMe Personal Genome Service®, and mothers from the Avon Longitudinal Study of Parents and Children (ALSPAC). We examined four phenotypes (pollen allergy, dust mite allergy, cat allergy, and any allergy), representing a total of more than 20,000 cases. In total, we identified 8 independent loci with P<5E-8 for one or more allergens. All these associations are in or near genes that have previously been implicated as having roles in immune function, with previously reported associations with phenotypes such as asthma, atopic dermatitis, and autoimmune disease. We also replicate many known associations with asthma, with effect sizes consistent with the original reports. Our results support the existence of a shared genetic etiology for these conditions, as well as suggesting distinct patterns of association across phenotypes. Our findings also demonstrate that self report is an effective method for collecting phenotypic information for genetic analysis of allergy and related phenotypes.

Methods

Self reported allergy and asthma phenotypes were assessed in the 23andMe participant cohort (Eriksson et al., 2010) and the ALSPAC mothers (Golding et al., 2010) using web based and paper surveys, respectively. For 23andMe, allergies were assessed mainly through an "Allergies and Asthma" survey. For genome wide analyses, we focused on four binary allergy phenotypes with close agreement between the 23andMe and ALSPAC surveys: cat, dust mite, pollen, and any allergy. Pollen allergy was constructed as the union of grass, tree, and weed allergies in the 23andMe survey. We also assessed asthma (5194 cases, 25771 controls), using the 23andMe general health survey. All participants provided informed consent for use of their data for research.

	23andMe		ALSPAC	
	Case	Control	Case	Control
Any	17348	9749	3150	4035
Pollen	7549	9462	1201	6015
Dust mite	4522	9575	704	6512
Cat	4901	9560	964	6252

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. The 23andMe analyses were restricted to a subset of participants with primarily European ancestry, and close relatives were excluded. ALSPAC participants were genotyped on the Illumina HumanHap550, then imputed against HapMap2 rel22. The UCSC LiftOver tool was used to remap the ALSPAC imputed data from NCBI build 36 to build 37. We used METAL to perform a fixed-effects inverse variance-weighted meta-analysis across the 23andMe and ALSPAC imputed datasets. We performed genomic control corrections on the individual datasets as well as the meta-analysis results. Results are shown for NCBI Build 37 (+) strand.

Results

Meta-analyses of cat, dust mite, pollen, and any allergy yielded broadly similar results, though the significance of individual associations varied. Eight loci were significant at the P<5E-8 level for at least one allergen.

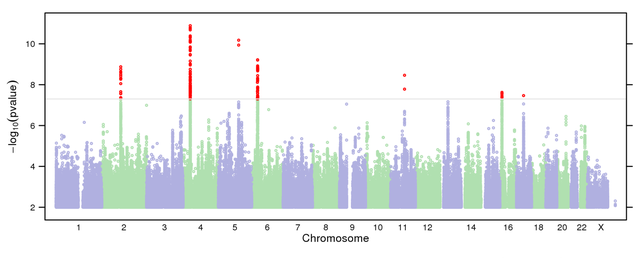

Figure 1. Manhattan plot of superimposed GWAS results for cat, dust mite, pollen, and any allergy.

All of the observed associations are in loci that have been previously associated with related phenotypes such as asthma or atopic dermatitis, but have not been specifically associated with allergy.

Region	Position	SNP	Alleles	Freq1	Pmin	GWAS	Context
4p14	38845198	rs7673348	G/A	0.248	1.3E-11	Cat	*TLR1, TLR6*
5q22.1	110470137	rs6594499	C/A	0.497	6.6E-11	Cat	*TSLP*
6p21.32	32412435	rs3177928	G/A	0.137	6.0E-10	Cat	*HLA-DRA*
2q12.1	102966549	rs10197862	A/G	0.847	3.1E-09	Pollen	*IL1RL1*
11q13.5	76299194	rs2155219	T/G	0.486	3.5E-09	Dust	*LRRC32*
6p21.33	31325270	rs9266227	G/C	0.439	7.7E-09	Pollen	*HLA-C*
16p13.13	11197390	rs76896814	A/T	0.567	2.4E-08	Dust	*CLEC16A*
17q21.1	38122680	rs7212938	G/T	0.484	3.4E-08	Any	*GSDMA*

At the 4p14 locus, the lead SNP is in high linkage disequilibrium with rs4833095, a nsSNP in *TLR1*. Effect sizes for loci that were genome-wide-significant in any individual analysis were generally similar across allergens, though there appears to be heterogeneity at *TLR1* as well as several loci in the HLA region. However, this consistency is in large part due to use of shared controls in the 23andMe cohort.

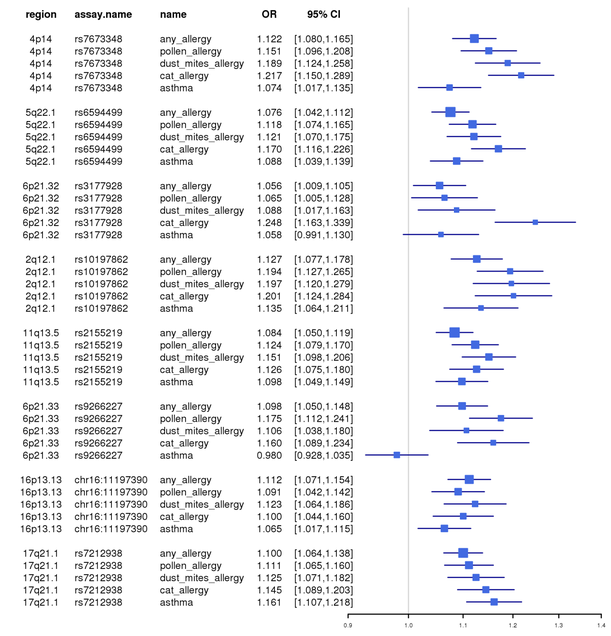

Figure 2. Forest plot of odds ratios for lead SNPs in each associated region, from the meta-analysis.

We computed genetic risk scores by counting risk alleles across the 8 loci passing the combined P<5.0E-08 threshold. We computed effect sizes for these risk scores on each of 29 allergens assessed in the 23andMe allergy survey. Confidence intervals for many allergens are wide, but the results suggested that there were systematic differences across allergen types.

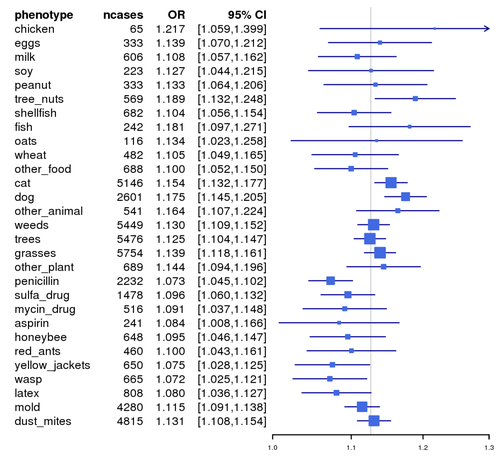

Figure 3. Forest plot of odds ratios for individual allergens in the 23andMe participant cohort

To account for the effect of correlations due to shared controls, we used bootstrap sampling to characterize the joint distribution of effect sizes for allergen classes. There were significant differences in the strength of association of the risk score across classes (animal versus plant: P=4.8E-5, plant versus drug: P=2.3E-05). Effects for foods and insects were less well determined, due to smaller sample sizes.

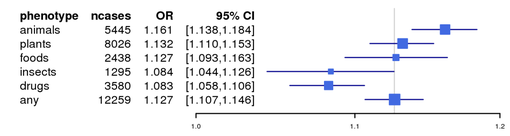

Figure 4. Forest plot of odds ratios for allergen classes in the 23andMe participant cohort.

We also evaluated our ability to replicate associations with asthma (Moffatt et al., 2010) in the 23andMe cohort. As shown below, we replicate all but one locus with P<0.05, and all effects were in the reported direction.

Region	Position	SNP	Alleles	Freq1	P	OR1	95% CI	Context
chr2	102986222	rs3771166	A/G	0.385	0.0011	1.075	[1.029,1.123]	*IL18R1*
chr6	32627714	rs1063355	T/G	0.414	2.7e-10	1.150	[1.101,1.201]	*HLA-DQB1*
chr9	6190076	rs1342326	A/C	0.836	0.00019	1.113	[1.053,1.178]	*IL33*
chr15	67446785	rs744910	A/G	0.512	0.0030	1.067	[1.022,1.113]	*SMAD3*
chr17	38062196	rs2305480	A/G	0.444	2.2e-14	1.182	[1.132,1.234]	*GSDMB*
chr17	38121993	rs3894194	G/A	0.546	2.5e-11	1.156	[1.108,1.206]	*GSDMA*
chr22	37534034	rs2284033	A/G	0.425	4.6e-05	1.093	[1.047,1.141]	*IL2RB*
chr5	131723288	rs2073643	C/T	0.547	0.19	1.029	[0.986,1.074]	*SLC22A5*
chr5	131995843	rs1295686	C/T	0.795	0.0040	1.080	[1.025,1.138]	*IL13*
chr15	61069988	rs11071559	T/C	0.137	0.00045	1.120	[1.051,1.194]	*RORA*

Discussion

Our high level finding of a substantial shared genetic susceptibility for allergy across a wide range of allergens is unsurprising. That said, we do see some significant differences in strength of associations across classes of allergens. These could reflect either differences in genetic etiology, or differences in accuracy of self assessment.

We have set aside some of the statistical issues raised by examination of multiple highly-correlated phenotypes, since all loci we identified have already been implicated in genetic studies of other phenotypes related to immune hypersensitivity. As more cohorts become more densely phenotyped, we think there will be a need for better methods for taking advantage of this phenotypic complexity.

Acknowledgments

We thank the participants in the 23andMe and ALSPAC cohorts for their contributions to this research.

References

Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, et al. (2010) Web-based, participant-driven studies yield novel genetic associations for common traits. PLoS Genet 6: e1000993.

Golding J, Pembrey M, Jones R, & the ALSPAC Study Team (2001) ALSPAC—the Avon Longitudinal Study of Parents and Children. I. Study methodology. Paediatr Perinat Epidemiol 15: 74-87.

Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, et al. (2010) A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 363: 1211-1211.