

Letter to the Editor

Assessing the reproducibility of asthma genome-wide association studies in a general clinical population

To the Editor:

Recently, genome-wide association (GWA) studies of asthma have found that variants in or near several genes, including ORM1-like 3 (*S cerevisiae*) / gasdermin B (*ORMDL3/GSDML*) (MIM 610075, 611221),¹ chitinase 3-like 1 (cartilage glycoprotein-39) (*CHI3L1*) (MIM 601525),² phosphodiesterase 4D, cAMP-specific (*PDE4D*) (MIM 600129),³ crumbs homolog 1 (*Drosophila*) / DENN/MADD domain containing 1B (*CRB1/DENND1B*) (MIM 604210, 613292),⁴ RAD50 homolog (*S cerevisiae*) / interleukin 13 (*RAD50/IL13*) (MIM 604040, 147683),⁵ and the *HLA-DR/DQ* region on chromosome 6p21.3,⁵ contribute to the risk of asthma. The heterogeneity of top GWA hits is thought to be a result of a combination of (1) variability of study population composition, including asthma definitions, (2) differences in statistical power among studies, and (3) unaccounted-for environmental differences. Nonetheless, the single nucleotide polymorphisms (SNPs) that have been reported in these GWA studies have been replicated more successfully in independent populations than had most older candidate gene associations.⁶ Most notably, SNPs in the *ORMDL3/GSDML* region have been replicated in over 10 independent populations from around the world.

Electronic medical records (EMRs) have been widely heralded for their potential to improve the quality of patient care. Most recently, data extracted from EMRs has been used to identify large numbers of subjects for genomics studies quickly and cost-effectively.⁷ We attempted to replicate some of the reported top asthma GWA hits in a cohort of subjects identified via EMR data to determine whether the results of such studies apply to a general clinical population.

The i2b2 Crimson Asthma Project (iCAP) consists of Partners Healthcare System, Inc (Boston, Mass) patients who were selected by using tools developed by the National Center for Biomedical Computing entitled "Informatics for Integrating Biology to the Bedside" (<http://www.i2b2.org>), which have facilitated extraction of de-identified demographic and clinical information from EMRs.⁸ To conduct genomic studies of these patients, clinical samples that are routinely collected at health care visits were obtained via the Crimson Project (<http://www.crimsonproject.org>), which identifies discarded Partners Healthcare clinical samples that are ordered for routine clinical tests and prospectively collects the samples that have been requested by an approved study. Cases were identified on the basis of presence of inpatient and/or outpatient International Classification of Diseases, Ninth Revision (ICD-9) codes for asthma (ie, those beginning with 493*) recorded between January 1, 1980, and November 30, 2006, within Partners Healthcare's EMRs. Controls were identified on the basis of absence of asthma ICD-9 codes recorded between January 1, 1980, and August 9, 2007, being between 17 and 40 years of age, and having been seen as a Brigham and Women's Hospital outpatient between May 1, 2004, and May 1, 2007. The last requirement was to help ensure that patients would be available for Crimson Project sample collection. No minimum observation period thresholds were enforced in controls, but these patients were selected to have total numbers of

observations (ie, health care events) similar to cases. Recruitment criteria for the collection of blood from cases included a restriction of age to be <50 years. Currently, DNA has been successfully collected for 3610 white (1731 cases, 1879 controls) and 368 black (183 cases, 185 controls) patients. The sex composition of the cohort is predominantly female: among white subjects, 77.8% (n = 1346) of cases and 84.0% (n = 1578) of controls were female; among black subjects, 80.9% (n = 148) of cases and 91.4% (n = 169) of controls were female. The mean age among white subjects was 35.8 (SD, 8.7; range, 5-48) for cases and 33.5 (SD, 5.5; range, 18-43) for controls. The mean age among black subjects was 35.8 (SD, 7.3; range, 7-48) for cases and 30.3 (SD, 6.0; range, 18-42) for controls.

We genotyped 16 SNPs corresponding to 9 genes (Table I) via Sequenom iPLEX assays on the MassARRAY system (Sequenom, Inc, San Diego, Calif) by using the manufacturer's recommended protocol at the Partners Center for Personalized Genomic Medicine core laboratory. Case-control associations were measured by using the allelic model; all *P* values reported are 2-sided. Among white iCAP subjects, the 3 *ORMDL3/GSDML* SNPs genotyped had nominally significant *P* values with effect directions consistent with the original *ORMDL3/GSDML* findings.¹ Two of these *P* values remain significant after correction for multiple comparisons (ie, are $\leq 0.05/32 = 1.6E-03$). Also among white iCAP subjects, the 3 *CHI3L1* SNPs genotyped approach a nominal significant threshold of .05. However, the direction of effect is opposite to that reported in the original *CHI3L1* article.² Of note, the iCAP effect direction is consistent with that observed in at least 1 other cohort in which *CHI3L1* had significant *P* values but a direction opposite that originally reported.⁹ The only other genotyped SNP that had a *P* value suggestive of replication was rs660895 near *HLA-DRB1*. However, this association was also in the opposite direction of that originally reported.⁵ Whether associations with effects in opposite directions across cohorts indicate that the low *P* values are a result of chance alone and hence are not biologically significant, or whether the effects are indicative of genes with biological significance in asthma whose SNP association effects are in opposite directions because of a factor that has not been accounted for, may be determined by further studies in other cohorts and more comprehensive studies within these genes.

Among black iCAP subjects, for which a much smaller number has been recruited thus far, only the 2 SNPs near *RAD50* had nominally significant *P* values. The association measured for rs2244012 is significant after correction for multiple comparisons (ie, is $\leq 0.05/32 = 1.6E-03$), and the effect direction of this association is consistent with that of the original report,⁵ even though the original finding was obtained in a cohort of white subjects with markedly different allelic frequencies. As Table I demonstrates, the minor allele among white iCAP subjects for rs2244012 was G (frequency, 0.22), whereas in black iCAP subjects, the minor allele was A (frequency, 0.32). In the original *RAD50* article,⁵ the rs2244012 minor allele was also G (frequency, 0.21), consistent with that among white iCAP subjects. Thus, the rs2244012 result could represent a true association that more significantly affects black than white patients by virtue of the different allelic frequencies in these populations. Further studies of this gene in larger black and white populations are

TABLE I. Association of reported asthma GWA study SNPs in iCAP subjects

SNP	CHR	BP	Gene	White iCAP subjects				Black iCAP subjects			
				Minor allele	MAF	P value	OR (95% CI)	Minor allele	MAF	P value	OR (95% CI)
rs2786098	1	195592531	<i>CRB1</i>	4	0.20	.90	0.99 (0.88-1.12)	4	0.09	.71	1.11 (0.64-1.91)
rs12134409	1	195658690	<i>CRB1</i>	4	0.20	.78	0.98 (0.88-1.11)	4	0.05	.62	0.85 (0.44-1.62)
rs1775456	1	195999678	<i>DENND1B</i>	3	0.20	.84	0.99 (0.88-1.11)	3	0.05	.74	1.13 (0.54-2.35)
rs4950928	1	201422505	<i>CHI3L1</i>	3	0.21	.10	1.10 (0.98-1.24)	3	0.31	.67	0.93 (0.68-1.28)
rs4950929	1	201426749	<i>CHI3L1</i>	3	0.20	.072	1.12 (0.99-1.26)	3	0.08	.95	1.02 (0.60-1.74)
rs946263	1	201432004	<i>CHI3L1</i>	3	0.17	.057	1.13 (1.00-1.28)	3	0.07	.38	1.32 (0.71-2.44)
rs1588265	5	59405551	<i>PDE4D</i>	3	0.30	.84	1.01 (0.91-1.12)	3	0.20	.99	1.00 (0.69-1.45)
rs1544791	5	59474839	<i>PDE4D</i>	4	0.31	.97	1.00 (0.91-1.11)	4	0.33	.87	0.97 (0.71-1.34)
rs2244012	5	131929124	<i>RAD50</i>	3	0.22	.77	1.02 (0.91-1.14)	1	0.32	6.8E-05	0.54 (0.39-0.73)
rs6871536	5	131997773	<i>RAD50</i>	2	0.21	.48	1.04 (0.93-1.17)	2	0.35	.030	1.43 (1.03-1.97)
rs20541	5	132023863	<i>IL13</i>	1	0.20	.65	0.97 (0.87-1.09)	1	0.19	.93	1.02 (0.69-1.49)
rs2395185	6	32541145	<i>HLA-DRA</i>	4	0.33	.94	1.00 (0.91-1.11)	4	0.24	.15	1.30 (0.91-1.87)
rs6660895	6	32685358	<i>HLA-DRB1</i>	3	0.19	.087	0.90 (0.80-1.02)	3	0.15	.90	1.03 (0.67-1.56)
rs9303277	17	35229995	<i>ORMDL3/GSDML</i>	4	0.45	3.6E-04	0.84 (0.77-0.93)	2	0.42	.74	1.05 (0.78-1.42)
rs11557467	17	35282160	<i>ORMDL3/GSDML</i>	4	0.44	2.4E-04	0.84 (0.76-0.92)	4	0.38	.63	0.93 (0.69-1.26)
rs2290400	17	35319766	<i>ORMDL3/GSDML</i>	2	0.46	3.9E-03	0.87 (0.79-0.96)	2	0.49	.82	0.97 (0.72-1.30)

BP, Base pair; CHR, chromosome; MAF, minor allele frequency; OR, odds ratio.

required to help determine whether this result represents a true functionally significant association that modifies asthma risk.

Our results suggest that some of the reported asthma GWA study findings are directly relevant to a broadly selected general clinical population classified on the basis of ICD-9 codes alone. In particular, the *ORMDL3/GSDML* locus associations, which have been replicated in many GWA studies composed of carefully recruited subjects, can also be observed in a white population identified via extracted ICD-9 codes from EMRs. Despite the current small sample size of black subjects, it is evident from our data, and consistent with reported GWA studies, that the allelic frequency distribution among black subjects can be markedly different than that among white subjects at some of the loci that are most prominently reported in GWA studies. Thus, continuing to pursue asthma GWA studies in black subjects, as well as subjects of diverse ethnic backgrounds, will likely yield important insights into the genetics of asthma. Despite the disadvantages related to a more incomplete characterization of disease and environmental variables than that provided by traditional epidemiologic and clinical trial populations, iCAP and other general clinical populations are likely to provide a faster and more cost-effective means to conduct genomic studies of asthma. The insights gained through studies of these general clinical populations will complement those of traditional ones.

Blanca E. Himes, PhD^{a,b}
 Barbara Klanderman, PhD^a
 Isaac S. Kohane, MD, PhD^b
 Scott T. Weiss, MD, MS^a

From ^athe Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School; and ^bthe Children's Hospital Informatics Program and Harvard Medical School, Boston, Mass. E-mail: blanca.himes@channing.harvard.edu.

Supported by grants U01 HL65899 and P01 HL083069 from the National Heart, Lung, and Blood Institute, National Institutes of Health.

Disclosure of potential conflict of interest: B. Klanderman has received research support from the National Heart, Lung, and Blood Institute, National Institutes of Health. B. E. Himes, B. Klanderman, I. S. Kohane, and S. T. Weiss have received research support from the National Institutes of Health.

REFERENCES

- Moffatt MF, Kabisch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants regulating *ORMDL3* expression contribute to the risk of childhood asthma. *Nature* 2007;448:470-3.
- Ober C, Tan Z, Sun Y, Possick JD, Pan L, Nicolae R, et al. Effect of variation in *CHI3L1* on serum YKL-40 level, risk of asthma, and lung function. *N Engl J Med* 2008;358:1682-91.
- Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, et al. Genome-wide association analysis identifies *PDE4D* as an asthma-susceptibility gene. *Am J Hum Genet* 2009;84:581-93.
- Sleiman PM, Flory J, Imielinski M, Bradfield JP, Annaiah K, Willis-Owen SA, et al. Variants of *DENND1B* associated with asthma in children. *N Engl J Med* 2009;362:36-44.
- Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA, et al. Genome-wide association study of asthma identifies *RAD50-IL13* and *HLA-DR/DQ* regions. *J Allergy Clin Immunol* 2010;125:328-35.e11.
- Ober C, Hoffjan S. Asthma genetics 2006: the long and winding road to gene discovery. *Genes Immun* 2006;7:95-100.
- Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet* 2010;86:560-72.
- Himes BE, Lasky-Su J, Wu AC, Wilk JB, Hunninghake GM, Klanderman B, et al. Asthma-susceptibility variants identified using probands in case-control and family-based analyses. *BMC Med Genet* 2010;11:122.
- Rogers AJ, Raby BA, Lasky-Su JA, Murphy A, Lazarus R, Klanderman BJ, et al. Assessing the reproducibility of asthma candidate gene associations, using genome-wide data. *Am J Respir Crit Care Med* 2009;179:1084-90.

doi:10.1016/j.jaci.2010.12.007