doi: 10.1111/j.1365-2222.2009.03352.x

Clinical & Experimental Allergy, **39**, 1453–1455

© 2009 Blackwell Publishing Ltd

EDITORIAL

Clinical & Experimental Allergy

Replication of genes for asthma: the importance of well-designed studies

This editorial discusses the findings of the paper by Vergara et al. [4] pp. 1558–68.

A. A. Litonjua[1,2,3,4] and A. Murphy[1,3,4]

[1]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA, [2]Department of Medicine, Pulmonary and Critical Care Medicine Division, Brigham and Women's Hospital, Boston, MA, USA, [3]Harvard Medical School, Boston, MA, USA and [4]Center for Genomic Medicine, Brigham and Women's Hospital, Boston, MA, USA

In the 13 years since the first genome-wide linkage analysis for asthma was published [1], research has uncovered many genes associated with asthma and asthma-related traits. Over 100 asthma susceptibility genes have been identified using various methods, from positional cloning using linkage analysis to candidate gene association studies [2], and more recently, a genome-wide association study was published [3]. However, only a handful of these genes have been replicated in subsequent studies. There are many reasons for why genes from initial studies do not replicate in subsequent studies, including the fact that the gene is not a true susceptibility gene, which is compounded by the failure to adequately adjust for multiple comparisons. More importantly, though, the asthma literature is replete with studies that have not paid rigorous attention to study design issues that can lead to unsuccessful replication, such as (but not limited to) inadequately powered studies, population stratification, variability in subject ascertainment/phenotype measurement, to list a few. As genotyping costs decrease, it is likely that more gene association studies will be attempted.

In this issue of the journal, Vergara et al. [4] investigate single nucleotide polymorphisms (SNPs) in the *GPR154* gene and find associations for asthma and total IgE. This gene has several aliases and the HUman Genome Organization (HUGO), which attempts to assign a unique and meaningful name for every known human gene, has assigned it the official name of neuropeptide S receptor 1 (official HUGO symbol, *NPSR1*). Nevertheless, to be consistent with the article, we will use the *GPR154* alias for the rest of this discussion. This gene resides on chromosome 7p and was discovered by positional cloning [5, 6]. As the authors correctly note, replication of original studies is fundamental in determining real from chance findings in human genetics. This is not the first replication study of this gene. Indeed, this gene has been replicated in studies of Caucasian [6–10] and Chinese subjects [11], but was not replicated in a Mexican cohort of childhood asthmatics [12]. No doubt, there will be more attempts at replication of variants in this gene, and the study by Vergara and colleagues can be used to highlight the issues of study design that readers should keep in mind in evaluating these studies.

For the study of binary traits, such as asthma affection status, there are two basic types of genetic association studies – population-based (i.e. case–control), where the frequency of the risk allele in cases is compared with the frequency of the risk allele in the controls; and family-based studies, where the number of times the risk allele is transmitted to the affected offspring from heterozygous parents is compared with the expected number (assuming Mendelian transmission) if there is no association. The authors of the present study use both types of designs in independent samples (i.e. there is no overlap of subjects in either analysis) to conduct their replication study. This is a unique aspect because most groups use population-based designs only. The fact that they use independent samples is another strength in the design of their study. In using either case–control or family-based design, power is an important consideration for successful replication of a true susceptibility locus. This is because most of the initial effects of discovered genes are likely greater than in other populations (due to stringent requirements required for a gene variant to survive an initial genome-wide scan) and subsequent studies must expect more modest effects [13]. Most replication studies and, for that matter, most genetic association studies are designed as case–control studies. Case–control genetic association studies that show a

Correspondence:
Augusto A. Litonjua, Channing Laboratory, 181 Longwood Avenue, Boston, Massachusetts 02115, USA.
E-mail: augusto.litonjua@channing.harvard.edu

Cite this as: A. A. Litonjua and A. Murphy, *Clinical & Experimental Allergy*, 2009 (39) 1453–1455.

significant association have a greater likelihood of successful replication in other studies if the number of cases and controls are each at least 150 [14], and replication studies should, therefore, also have at least this number of cases and controls to have the minimum power to detect the association. The study by Vergara and colleagues had more than twice this minimum number for their case–control analysis, and despite this, only one SNP survived correction for multiple comparisons. On the other hand, even though power for a family-based study of N trios is comparable to a study of N case–control pairs [15], the power for the family-based study conducted by Vergara and colleagues was smaller (in comparison to their case–control study), as there were only 116 trios (two parents and one offspring). Additionally, as Vergara and colleagues noted, the minor allele frequency for some of the SNPs of interest was low, which reduces the number of heterozygous parents, and subsequently the number of informative families contributing to the family-based analysis. However, despite these constraints, the findings from the family-based analysis were consistent with those from the population-based analysis, giving more confidence that these results are not spurious.

Another notable aspect of this study is that the population differs from all previous populations studied for this gene. The subjects in this study come from an admixed population in the South American country of Colombia, with a high proportion of African and Amerindian ancestry, i.e. an admixed population [16]. This brings up two points. The first is that it is important to replicate findings in diverse populations. Most initial reports of associations for novel genes have been made in Caucasian populations; thus, demonstrating an association in a population with different racial and ethnic backgrounds, and potentially different environmental exposures is important because this suggests that the association may be real and robust. Second, in performing case–control genetic association studies in admixed populations, one of the biggest threats to the validity of the findings is population stratification, where there is an imbalance of population subtypes between cases and controls, which can subsequently lead to spurious associations. There are techniques for evaluating and addressing potential population stratification in case–control studies [17], using hundreds to thousands of single-nucleotide polymorphisms scattered throughout the genome or, alternatively, selected panels of ancestry informative markers (a set of polymorphisms which exhibits substantially different frequencies between populations from different geographical regions). Vergara and colleagues chose to use ancestry informative markers in their study, and adjusted for the stratification that was detected. Additionally, the authors replicated many of their findings in an independent family-based cohort. A considerable strength of family-based designs is robustness against population stratification. Again, the fact that the findings in the family-based design were consistent with the findings in the population-based design helps substantiate the association.

A final aspect of study design that this study demonstrates concerns the issue of variation in subject ascertainment or phenotypic measurement across studies. To avoid these issues, the authors used the same locations to recruit subjects for both the case–control and family-based study. Identical criteria (based on internationally developed guidelines), for determining asthma affection status also was applied to both the case–control and family-based populations. Finally, a lack of both personal and family history of asthma was used to identify control subjects. While these issues in study design are more subtle than the ones that were previously discussed, they can be a major concern in non-replication of studies. It is important to note that Vergara and colleagues were able to replicate findings in a study of adolescent subjects, when their initial population comprised adult subjects. Careful control of other factors in study design may have increased the power to detect associations in *GPR154* across adult and young adult populations.

What, then, have we learned from the efforts of Vergara and colleagues? First, given successful replication of the association of variants in this gene with asthma in various populations, it is now highly likely that *GPR154* is a true asthma susceptibility gene. The next step for this gene is to perform studies that fully define function in attempts to translate the findings to clinical use. This, of course is easier said than done. Second, performing replication studies of genetic associations is not easy to do well. Careful thought needs to be placed on issues, from defining phenotypes to analyzing data, in order to minimize the chances of both false positive and false negative findings. Vergara and colleagues have done a commendable amount of work in attending to these issues. We have highlighted some of these issues in order to emphasize their importance. Only with well-designed gene association studies will we be able to wade through the large amount of information that is being generated.

References

1 Daniels S, Bhattacharrya S, James A *et al*. A genome-wide search for quantitative trait loci underlying asthma. *Nature* 1996; **383**: 247–50.

2 Ober C, Hoffjan S Asthma genetics 2006: the long and winding road to gene discovery. *Genes Immun* 2006; **7**:95–100.

3 Moffatt MF, Kabesch M, Liang L *et al*. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007; **448**:470–28.

4 Vergara C, Jiménez S, Acevedo N *et al*. Association of *G-protein-coupled receptor 154* with asthma and total IgE in a population of the Caribbean coast of Colombia. *Clin Exp Allergy* 2009; **39**:1558–68.

5 Laitinen T, Daly MJ, Rioux JD *et al*. A susceptibility locus for asthma-related traits on chromosome 7 revealed by genome-wide scan in a founder population. *Nat Genet* 2001; **28**:87–91.
6 Laitinen T, Polvi A, Rydman P *et al*. Characterization of a common susceptibility locus for asthma-related traits. *Science* 2004; **304**:300–4.
7 Kormann MS, Carr D, Klopp N *et al*. G-Protein-coupled receptor polymorphisms are associated with asthma in a large German population. *Am J Respir Crit Care Med* 2005; **171**:1358–62.
8 Melen E, Bruce S, Doekes G *et al*. Haplotypes of G protein-coupled receptor 154 are associated with childhood allergy and asthma. *Am J Respir Crit Care Med* 2005; **171**:1089–95.
9 Malerba G, Lindgren CM, Xumerle L *et al*. Chromosome 7p linkage and GPR154 gene association in Italian families with allergic asthma. *Clin Exp Allergy* 2007; **37**:83–9.
10 Hersh CP, Raby BA, Soto-Quiros ME *et al*. Comprehensive testing of positionally cloned asthma genes in two populations. *Am J Respir Crit Care Med* 2007; **176**:849–57.
11 Feng Y, Hong X, Wang L *et al*. G protein-coupled receptor 154 gene polymorphism is associated with airway hyperresponsiveness to methacholine in a Chinese population. *J Allergy Clin Immunol* 2006; **117**:612–7.
12 Wu H, Romieu I, Sienra-Monge JJ *et al*. Lack of association between genetic variation in G-protein-coupled receptor for asthma susceptibility and childhood asthma and atopy. *Genes Immun* 2008; **9**:224–30.
13 Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008; **19**:640–8.
14 Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet* 2001; **29**:306–9.
15 Laird NM, Lange C. Family-based designs in the age of large-scale gene-association studies. *Nat Rev Genet* 2006; **7**: 385–94.
16 Vergara C, Caraballo L, Mercado D *et al*. African ancestry is associated with risk of asthma and high total serum IgE in a population from the Caribbean Coast of Colombia. *Hum Genet* 2009; **125**:565–79.
17 Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003; **361**:598–604.