Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Genetics & Development

Asthma genetics and genomics 2009

Scott T Weiss, Benjamin A Raby and Angela Rogers

Asthma Genetic Association studies have been plagued by methodologic problems that are common in all studies of complex traits: small sample size, lack of replication, and lack of control of population stratification. Despite this, the field has identified 43 replicated genes from association studies. The most frequently replicated are: TNF alpha, IL4, FCERB, Adam 33, and GSTP1. Several genes have been identified by linkage and fine mapping (ADAM33, DPP10, GPR154, and PHF11) and one gene has been identified by GWAS (ORMD3). The major issue is that these genes have been looked at one at a time rather than in some more holistic manner where epistasis is considered. For asthma genetics to begin to have an impact on clinical medicine we need to consider epistatic interaction.

Address
Harvard Medical School, Channing Laboratory, 181 Longwood Avenue, Boston, MA 02115, USA

Corresponding author: Weiss, Scott T (scott.weiss@channing.harvard.edu)

Current Opinion in Genetics & Development 2009, **19**:279–282

This review comes from a themed issue on
Genetics of disease
Edited by Kay Davies and Anthony Wynshaw-Boris

0959-437X/$ – see front matter
Published by Elsevier Ltd.

DOI 10.1016/j.gde.2009.05.001

Introduction

Since the completion of the human genome project in 2000 there have been remarkable advances in complex trait genetics. The field has evolved from a difficult to use, and incomplete, map of the human genome to a much more nuanced approach embracing all of the possibilities for genome variation that are being revealed. We are now moving forward more rapidly with a variety of different approaches for understanding the biologic implications of genome diversity. Although much has been discovered, much remains to be learned about how the genome influences disease risk and disease pathobiology. Many are impatient for these advances to have a real impact on translational medicine and human health. In this brief review we will explain where we are with gene identification for asthma, how that work is being accomplished, and what the implications are for the future and for clinical medicine.

Genetic association studies

The basic approach in human genetics is to relate allele frequency in a collection of cases to a set of controls and to see if specific alleles in the cases are more common than in the control group. The measure of effect is the ratio of the allele frequency in the cases to that in the controls and that approximates the conventional concept of an odds ratio for risk in the disease or case group versus the control group. This study design is so simple that it is easy to do poorly and yet doing it well can be remarkably difficult. The conventional problems seen in the literature with this approach are many: small sample size, failure to collect and phenotype the cases and controls in the same manner, failure to replicate the initial results, failure to account for population stratification, and failure to account for linkage disequilibrium. It is worth discussing each of these concerns in a bit more detail, as they are so common in the published literature. Many studies are simply too small to give a meaningful result; case samples of less than 150–200 cases are underpowered and hence not able to detect effect sizes of 1.1–1.4 which is the common range seen in complex traits. Positive studies in this sample size range need to be repeated because of the tendency for false positive results. This design error and a related one, failure to collect and phenotype the controls in exactly the same way as the cases are the two most common errors in the literature of genetic association. Allele frequency is also an important factor influencing power in genetic association studies and small numbers of cases and controls will only be powered to detect genetic associations at allele frequency of greater than 10–15%. Replication is also an essential feature of the well designed genetic association study. Replication insures that the observed effects are more likely to be real, and hence believable. As already noted this is especially true for smaller studies. Finally the issue of population stratification or spurious differences in allele frequency between the cases and the controls is an issue that is specific to this type of study design. Its importance has varied in the literature, but given the small effect sizes seen in most studies, even a small degree of stratification could cause a spurious result. Clearly stratifying toward homogeneity of genetic origins in population samples and using one of the many approaches to control for population admixture (another name for stratification) is desirable. Finally, linkage disequilibrium or nonrandom association of alleles in the genome is always a possible explanation for a positive result. There is no way short of additional data external to the study to rule out this possibility. In addition linkage disequilibrium can contribute to nonreplication since

Table 1

Gene	Ref seq.	Position	Total populations showing SNP association with asthma
ADAM33	chr20	3596621–3610738	9
ADRB2	chr5	148186349–148188381	5
CCL11	chr17	29636800–29639312	3
CCL24	chr7	75279052–75280969	2
CCL5	chr17	31222610–31231490	3
CD14	chr5	139991509–139993194	4
CHI3L1	chr1	201414682–201422545	3
CTLA4	chr2	204440754–204446928	2
CX3CR1	chr3	39279990–39296531	2
CYSLTR2	chr13	48178952–48181499	3
DPP10	chr2	114916369–116318406	2
EDN1	chr6	12398645–12404761	3
FCER1B	chr11	59612713–59622590	9
GPR154	chr7	34664422–34856115	3
GSTP1	chr11	67107862–67110699	8
HAVCR1	chr5	156389015–156418548	1
IFN gamma	chr12	66834817–66839788	2
IL10	chr1	205007571–205012462	4
IL12b	chr5	158674369–158690059	2
IL13	chr5	132021764–132024700	8
IL4	chr5	132037272–132046267	11
IL4 R	chr16	27232752–27283599	7
INPP4A	chr2	98427845–98564716	2
IRAK-3	chr12	64869284–64928652	2
ITGB3	chr17	42686207–42745075	3
LTA	chr6	31648072–31650077	3
MYLK	chr3	124813835–125085839	2
NAT2	chr8	18293035–18303003	3
NOD1	chr7	30430675–30484790	4
NOS3	chr7	150319080–150342608	1
NPPA	chr1	11828363–11830422	2
ORMDL3	chr17	35330822–35337380	8
PAFAH	chr6	46780238–46811055	3
PHF11	chr13	48967802–49001117	4
PTGDR	chr14	51804181–51813191	5
TBXA2R	chr19	3545504–3557658	2
TGFB1	chr19	46528491–46551656	2
TLR4	chr9	119506431–119519587	2
TLR9	chr3	52230138–52235219	3
TLR10	chr4	38450647–38460984	2
TNF	chr6	31651329–31654089	17
UGB (CC10)	chr11	61943099–61947242	4
VDR	chr12	46521587–46585081	3

LD operates as an imperfect proxy for the true disease or causal variant [1•].

The major alternative to the case–control study is the family based study design. Most of the problems associated with a case–control study apply here as well except for a few notable exceptions. Here population stratification is not an issue as the controls are internal to the trio (parents and affected proband or offspring). Indeed, the odds ratio created here, is not comparable to the odds ratio in the case–control study. Here the odds ratio is the ratio of the transmitted to untransmitted alleles in the heterozygous parents and the affected offspring. This is completely not comparable to the odds ratio in the case–control setting, which as stated reflects the allele frequency in cases versus the controls.

Types of study designs to identify asthma genes

There are three types of studies that have been used to identify asthma genes: association studies of candidate genes, linkage and then fine mapping of a linked region, or Genome Wide Association Studies. Each of these approaches has advantages and disadvantages and each will be discussed in turn to present the state of the field.

Summary of Asthma Genetic Association studies

Table 1 presents a list of the 43 genes and their chromosomal location that have been associated with the asthma phenotype in at least one study of samples of greater than a total of 300 subjects (150 cases and 150 controls) [2•]. Additional criteria for inclusion included replication in at

least one other population and replication with the same SNP. As can be seen from the table there a relatively small number of genes with many replications TNF alpha lead the list with 17 total replication studies followed by IL4, FCERB1, Adam33, and GSTP1. The supplementary table provides more details on the populations and the study design of these investigations and demonstrates that most of these genes have been replicated in multiple ethnic groups [2•]. Several of these genes were first identified by linkage and then fine mapping (ADAM33, DPP10, GPR154, and PHF11) or by GWAS (ORMDL3) [3•]. The advantage of this approach is that no prior knowledge of asthma pathobiology was used to identify the genes and they most likely represent genes that point to novel biology. Most of the genes were simply candidates that were identified by association but came from known asthma pathobiology and hence are of less interest. Although individual studies in the supplementary table can be criticized the overwhelming likelihood is that most of these genes play at least some role in disease pathogenesis. Again the effect estimates (odds ratios) are mostly between 0.5 and 1.5 indicating that none of these genes are genes of large effect.

How much of explained phenotypic variance of the asthma phenotype comes from these 43 genes?

The answer to the above question is unknown because no one has actually done the study to genotype these 43 genes and to relate their polymorphism to asthma. However, a similar exercise has been performed for height with instructive results.

Height has extremely high genetic heritability being about at or about 0.8 or 0.9, substantially greater than asthma. Weedon *et al.* performed a genome wide association study on 13 665 individuals and found 39 variants predictive of adult height and then genotyped an additional 16 482 individuals [4•]. They found 20 variants that had *P* values below 5×10^{-7}. Collectively these 20 variants explained only 3.8% of the variability of adult height.

The obvious question is what explains the missing heritability? One speculation is that other polymorphism in the human genome is responsible for copy number variation, epigenetic variation, variation in gene expression, could all be contributing and that SNPs are simply not that important in explaining human height variability. Although that is certainly true it does not present the whole picture. Another possibility is that linkage disequilibrium, which is commonly used in genetic association studies, is imprecise in finding the SNPs and genes associated with a disease trait. This is also true but can hardly be the whole story or even the major aspect of the explanation for the height results. Certainly we know that individual SNPs and individual genes are not that important because almost all genes that have an effect in asthma and other complex traits have effect sizes of between 1.1 and 1.5 for most complex traits. What we have not accounted for is that genes (and SNPs) work in networks and hence we need to put the gene networks together to get the true meaning of how genes function epistatically. The implications of this latter possibility are profound because it means that for clinical relevance either for pathobiology or for prediction, we will need to put genes together in the appropriate networks to be able to get to high levels of prediction that will be necessary to translate genetic association data into clinical relevance.

How well do current GWAS chips cover the human genome?

An alternative, not fully competing, explanation to the epistasis or gene–gene interaction argument to explain the Weedon results are contained in the recent paper of Hao *et al.* [5•]. These investigators clearly show that actual genome coverage with current GWAS chips is closer to 50% that the 80% that was claimed for them. Thus lack of coverage is another potential reason all genes are not being found and identified. Rogers *et al.* showed that even candidate genes with multiple replications as noted in Table 1 are poorly detected by adequately powered studies using conventional GWAS chips [2•]. Thus GWAS is not a good way to confirm known associations as the coverage of any individual gene and the genome as a whole is not as good as initially advertised.

What will be the path forward to enhance the translational potential of genetics for asthma and other complex traits?

First, all sources of genetic polymorphism of relevance will need to be identified and characterized. This will include: copy number variation, *cis*-acting and *trans*-acting regulatory variation, epigenetic marks, and coding SNPs. The true functional variants will need to be identified and linked in epistatic networks. It is quite likely that this will be approached incrementally with epistatic networks being built with LD SNPs rather than causal SNPs as a first pass. It is likely that machine learning approaches will be used to integrate multi-SNP signals and provide hypothesis free inference as to how the SNPs will be linked together. Complex trait genetics is still in its infancy, but significant insights are to be had even at this stage of the development of the field.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gde.2009.05.001.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Nielsen DM, Suchindran S, Smith CP: **Does strong linkage**
• **disequilibrium guarantee redundant association results?** *Genet Epidemiol* 2008, **32(6)**:546-552.
This article explains how, with linkage disequilbrium that is similar, but not exactly identical, in two different populations you get lack of replication of results. In addition to different clinical phenotypes, and different environmental exposures this is a major reason for lack of replication in genetic association studies.

2. Rogers AJ, Raby BA, Lasky-Su J, Murphy A, Lazarus R,
• Klanderman BJ, Sylvia JS, Ziniti JP, Lange C, Celedon JC *et al.*: **Assessing the reproducibility of asthma candidate gene associations using genome-wide data**. *Am J Respir Crit Care Med* 2009. [epub ahead of print].
This paper provides an explanation for why GWAS genotyping is not adequate to look at candidate genes. It also provides a comprehensive summary of genetic association studies in asthma and all of the references to replication results for the genes identified in the supplementary table of this review.

3. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S,
• Depner M, von Berg A, Bufe A, Rietschel E *et al.*: **Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma**. *Nature* 2007, **448(7152)**:470-473 [epub 2007 July 4].
This paper is the first asthma GWAS study ever published. It is worth reading for its methods alone.

4. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM,
• Mangino M, Freathy RM, Perry JR, Stevens S, Hall AS *et al.*: **Genome-wide association analysis identifies 20 loci that influence adult height**. *Nat Genet* 2008, **40(5)**:575-583.
This paper has excited a great deal of discussion in the scientific literature about the cause of the missing variation. These authors found 20 replicated height SNPS from GWAS and put them in a regression model and could only explain 3.8% of the variability of height a phenotype known to be highly heritable. This has sparked controversy as to know the reason for the missing variability. As described above my hypothesis is that this is due to the inability to model the epistasis directly.

5. Hao K, Schadt EE, Storey JD: **Calibrating the performance of**
• **SNP arrays for whole-genome association studies**. *PLoS Genet* 2008, **4(6)**:e1000109.
This paper demonstrates that the current estimates of genome coverage from GWAS chips is overestimated and probably these chips only cover about 60% of the genome not 80–90%.