Asthma genetics 2003

Scott T. Weiss* and Benjamin A. Raby

Channing Laboratory, Brigham and Women’s Hospital, 181 Longwood Avenue, Boston, MA 02115, USA

Received December 23, 2003; Revised and Accepted January 9, 2004

The use of positional cloning for the identification of complex trait susceptibility genes has gained momentum with the completion of the human genome project. The approach involves the collection of well-phenotyped cohorts (either family-based or case–control designs), the generation of high-density single-nucleotide polymorphism linkage disequilibrium maps, and the application of powerful statistical methods to localize narrow regions of genetic association with disease. In 2003, two novel genes relating to asthma were identified using this approach, PHF11 and DPP10, neither of which had previously been implicated in the pathobiology of either asthma or allergy. In addition, further support for ADAM33 (the first asthma susceptibility gene identified by positional cloning) as an asthma gene was presented, although with mixed results. These discoveries open new avenues for research in asthma and allergy, and highlight the power (and limitations) of positional cloning for the identification of asthma genes, and complex trait genes in general.

INTRODUCTION

In April 2003, coinciding with the 50th anniversary of the characterization of the double helix, the scientific world was presented with the completed sequence of the human genome and the promise of a new era in human genetics (1). Indeed, subsequent to the release of even the earliest draft sequences, positional cloning of disease genes, particularly for monogenic traits, has become both feasible and increasingly rewarding. Emphasis has now turned to the search for genes that contribute to the development of so-called complex traits—diseases thought to be caused by interactions between multiple genes of small to modest effect and equally important environmental factors. Progress has been made across a wide spectrum of complex diseases, including cardiovascular disease, diabetes, rheumatic disease, and inflammatory bowel disease (2–6).

Among the many exciting developments in 2003, there has been considerable progress in the field of asthma and allergy, with two publications describing the positional cloning of novel loci related to asthma and an important marker of allergic diathesis—total serum immunoglobulin E (IgE) levels (7,8). The methods of gene localization employed represent the state-of-the-art in complex trait positional cloning: large, well-phenotyped family-based populations were collected, genome-wide linkage analysis was performed, and regions of significant linkage were fine-mapped by laborious characterization of the genetic variation across the region of linkage using high-throughput sequencing, single-nucleotide polymorphism (SNP) genotyping and linkage disequilibrium (LD) mapping (Fig. 1). This strategy has been used to identify several complex trait genes, including the first cloned asthma gene, ADAM33, in 2002 (9). This brief review discusses these three papers and will illustrate some of their strengths and pitfalls. In doing so, we hope not only to present an update of the current state of asthma genetics, but also shed light on its future, and that of complex trait genetics in general.

ASTHMA GENETICS

Asthma is a syndrome characterized by intermittent narrowing of the small airways of the lung, with subsequent airflow obstruction and symptoms of wheeze, cough and breathlessness. Most asthma is diagnosed in childhood, with a prevalence of 6% (10,11). An important characteristic of asthma is airways hyper-responsiveness, which is the exaggerated narrowing of the airways in response to provocative agents. The majority of asthmatics are also atopic, with manifestation of allergic diathesis including clinical allergy to Aeroallergens and foods, or subclinical allergy manifest by skin test reactivity to allergen or elevated serum IgE. Asthma and atopy are considered complex traits, with evidence of both heritable and environmental factors contributing to their pathogenesis. Heritability estimates of asthma vary between 36 and 79% (12–15). Importantly, there is evidence that genetic liability for asthma, airways responsiveness and allergic traits are regulated through distinct loci, although there is likely some shared overlap as well (16). Multiple genome-wide linkage

*To whom correspondence should be addressed. Email: scott.weiss@channing.harvard.edu
studies for asthma and allergy have been performed, and have identified no fewer than 20 distinct chromosomal regions with linkage to asthma or related traits (17,18–25). Several of these linkages have been observed in more than one study, of which the most frequently reproduced include regions 6p, 5q, 12q and 13q. In addition, convincing evidence of linkage on chromosomes 14q and 7p has been observed in founder populations from Iceland and Finland, respectively (22,23). Most of these regions are large, spanning 10–30 Mb and harboring hundreds of genes. Until recently, follow-up mapping of these broad
regions has been difficult, owing to the lack of availability of high-resolution linkage disequilibrium maps. The situation began to change with the identification of ADAM33 on chromosome 20p.

**ADAM33**

In 2002, Van Eerdewegh and coworkers reported the fine mapping of the ADAM33 as an asthma and airway hyper-responsiveness gene on chromosome 20p13 (9). These investigators initially performed a genome-wide scan on 460 Caucasian families and identified a locus on chromosome 20p13 that was linked to asthma, with an LOD score of 2.94. The investigators more precisely localized this by refining their phenotype to a more stringent definition of asthma that included airways hyper-responsiveness, a hallmark of asthma. Despite the reduced sample size due to more stringent phenotypic criteria, the LOD score increased, reaching genome-wide significance (LOD score = 3.93 near marker D20S482). The linked region, including a 1-LOD-support interval around marker D20S482, spans 4.28 cm, with a corresponding physical distance of 2.5 Mb. The investigators were fortunate in this regard—most linkages observed in asthma have been considerably larger, spanning more than 20–30 Mb in some cases. A gene-oriented physical map was constructed using a variety of methods. More than 40 genes were localized to this region, and SNP discovery performed using a combination of single-stranded conformational polymorphisms screening and subsequent sequencing. One-hundred and thirty-five SNPs in 23 genes were then selected for genotyping in a case–control association study composed of probands that contributed to the LOD score from the linkage study and ‘hypernormal’ controls. Using this approach, the investigators identified a cluster of SNPs in the ADAM33 gene that demonstrated significant associations with asthma. Although SNPs in other nearby genes also demonstrated evidence for association, the large number of positively associated SNPs in ADAM33 led the investigators to conclude that ADAM33 variants explained their initial findings of linkage. Additional SNPs in ADAM33 were subsequently tested and found to be associated with asthma. Family-based association analysis and pairwise haplotype analysis using the original pedigrees lent further evidence that the associations seen were not due to population admixture. Expression analysis revealed that ADAM33 is expressed in multiple tissues (including the lung and lymph node) and confirmed that, within the lung, ADAM33 expression was highest in bronchial smooth muscle and lung fibroblasts (two cells critical to airway hyper-responsiveness and remodeling). These features suggest that ADAM33 variants may directly impact lung architecture and function. However, because other ADAM proteins (ADAM10 and ADAM17) appear to interact with inflammatory cytokines (including TNF-α), it has been speculated that ADAM33 may also have important cytokine-stimulating effects, and may contribute to asthma susceptibility through these pathways (26).

The identification of ADAM33 as an asthma-susceptibility gene was received with great enthusiasm, as it marked a new phase in asthma research, and a new target for drug development. However, several features of the initial report raise questions regarding the generalizability of the reported results, and subsequent attempts to replicate the findings of linkage, have yielded mixed results. Three features of the original report deserve emphasis. First, although the Van Eerdewegh group observed significant evidence for linkage of 20p13, asthma linkage to this region had not been observed in most previous genome screens. This raises a question of the generalizability of the linkage results for this region. Second, the initial publication did not have a truly independent replication. Two Caucasian populations of asthmatic families were utilized in the report of Van Eerdewegh and coworkers—one in the UK, the other in the USA. Although single SNP and haplotype associations were observed in each of the populations (suggesting replication), these two populations were pooled for the initial genome screen, and therefore should not be considered independent. In addition, no single SNP demonstrated significant associations in both populations. A third concern is that the haplotype based analysis (which gave the most statistically significant results—with P-values of $10^{-5}$) consisted of testing all pairwise SNP comparisons for association, without regard for the LD pattern across the gene. This somewhat arbitrary approach ignores the LD within the gene and leads to the important statistical problem of multiple comparisons and inflation of type I error. A more appropriate approach would have been to first characterize the haplotype structure of the gene, and only perform tests that reflect that genetic architecture. Lastly, no functional data regarding the role of the associated variants to the development of the asthma phenotypes were presented. Most of the associated polymorphisms do not reside within coding regions, but are imbedded deep with introns. It is conceivable that these variants alter gene expression or splicing profile, but to date there is no data to support this.

Subsequent to the initial report, two other groups have tested ADAM33 SNPs in their asthma cohorts with mixed results. Howard and colleagues reported evidence of association of ADAM33 and asthma in four distinct cohorts: a US Caucasian cohort, a Dutch Caucasian cohort, an African-American cohort, and a Hispanic cohort (27). Several SNPs showed evidence of association in each of the cohorts with asthma, some conferring modest risk; however, there was no consistency of the associations across these populations and those seen in the Van Eerdewegh report. With the exception of one polymorphism (ST+7, an intronic polymorphism with no known functional effect), no single SNP or haplotype demonstrated consistent associations, either with respect to the magnitude or the direction of effect. In a second report, six ADAM33 SNPs were tested for association with asthma in 583 Hispanic (Puerto Rican and Mexican) trios, with no significant associations found to asthma or a variety of asthma phenotypes (28). Our group has extensively examined SNPs in this gene in a large population of North American asthma trios with null results (29). Using a family-based association study design, we genotyped 17 ADAM33 polymorphisms in 587 nuclear families ascertained through participation in a multicentered North American clinical trial. Despite assessing nine markers that demonstrated association with asthma in the original publication, no marker was associated with asthma, and only weak associations were observed with IgE levels and total...
blood eosinophilia. No associations with airways responsiveness were observed.

These findings leave one with mixed results in regard to the role that ADAM33 plays in asthma pathobiology. There appears to be some limited replication of the Van Eerdewegh findings, although it is weak at best. Not all populations show the association and, certainly, the magnitude and direction of the association is not consistent in those populations where significant effects were observed. Failure to replicate genetic associations in complex disease is a common occurrence (30,31). Commonly proposed causes for this failure include false discovery in the initial study (type I error due to multiple SNP testing), insufficient power in replication studies or spurious results due to population admixture. Equally important are the very real issues of genetic and phenotypic heterogeneity across study populations, and unrecognized differences in environmental exposures. Further work needs to be performed both in terms of replication of genetic association as well as attempts to determine gene function before any firm conclusions can be made regarding the importance of ADAM33 in asthma.

PHF11

The second positionally cloned gene for asthma is a locus for total immunoglobulin E levels located on chromosome 13q14 (8). Unlike the ADAM33 locus, 20p13, this genomic region has been linked with asthma in multiple genome screens for asthma and serum IgE levels (32–34). Linkage was initially detected in a set of Australian families with a history of atopy to a segment of chromosome 13q14 that spanned 7.5 cM. To narrow the region for subsequent study, the investigators genotyped additional microsatellite markers within this region, and assessed each marker for association with total serum IgE. Strong association was detected with one microsatellite—USAT24GI (35). Linkage disequilibrium around this marker extended approximately 100 kb, suggesting that the disease susceptibility locus nearby (36). The investigators then constructed a 1.5 Mb contig centering on this microsatellite, and mapped expressed sequences from available databases, and by 3’ and 5’ RACE. To identify polymorphism in this region, resequencing was performed in 10 individuals (five atopic subjects, five normal) and a pool of 32 unrelated individuals. Fifty-four common variants were then genotyped in 80 nuclear families. Based on this genotype data, linkage disequilibrium across the region appeared to be clustered in four islands (or blocks). Using a variance-components analysis, evidence for association with total IgE was identified with variants in two adjacent blocks, and centered on one gene—PHF11. Variants in two other genes that flanked PHF11 (that were also within the associated LD blocks) also associated with IgE level, but to a lesser degree than those in PHF11. Using a stepwise analytic procedure, the investigators demonstrated that three polymorphisms localizing to PHF11 carried the bulk of the association with IgE: two were intronic, and one was in the 3’-UTR of PHF11. This locus explained 11.5% of the variation in total serum IgE levels, more than 50% of which could be attributed to the three variants described above. They went on to attempt to replicate their findings in four additional atopic or asthmatic populations: a second set of Australian families with atopy, a British asthma family cohort, a cohort of families with atopic dermatitis, and a British asthma case–control panel. In three populations, significant associations were noted with one set of three-locus haplotypes (which included two of the PHF11 markers most strongly associated in the first population), including associations with IgE levels or atopic dermatitis. Evidence for single-SNP association replication was not reported. Associations with the asthma phenotype were observed, but were less impressive. Associations with a diagnosis for asthma with two SNPs ($P = 0.02$) were observed in the Australians only when both the original linkage families and the replication families were combined. In the British case–control panel, significant trends were noted when the asthmatics were stratified by severity. Considering that atopic status (and IgE level in particular), is an important determinant of asthma severity, it is likely that the observed associations with asthma are secondary to the variants’ atopy-related effects.

PHF11—PDH Finger Protein 11—encodes NY-REN-34, a protein that was first described in patients with renal cell carcinoma (37). The precise function of this gene has not been determined, but the presence of two zinc finger motifs in the translated protein suggests a role in transcriptional regulation. The gene is expressed in most tissues, but Zhang and colleagues observed consistent expression in many immune-related tissues. Moreover, they identified multiple transcript isoforms, including variants expressed exclusively in the lung and in peripheral blood leukocytes. Because variation in this gene was strongly associated with serum IgE levels and, as described by Zhang, with circulating IgM immunoglobulin, and because the gene is expressed heavily in B-cells, the authors suggest that this locus may be an important regulator in immunoglobulin synthesis. Although the statistical evidence presented by Zhang and colleagues most consistently supports PHF11 as the IgE regulator in this region, the authors are appropriately cautious to propose the two genes that co-localize with PHF11 within the same LD blocks (SETDB2 and RCBTB1) may also be important, particularly SETDB2 which has a similar expression profile to PHF11 in immune-related cells. This possibility highlights one of the important limitations of LD mapping of disease genes: if strong linkage disequilibrium extends over a segment of genome that harbors multiple genes, it is difficult (if not impossible) to identify which gene within the cluster is the disease gene using LD mapping alone (38). Given that no functional data has been presented, the next major focus of investigation for this locus should include functional dissection of these three genes and the IgE-associated variants. Alternatively, it may be possible to use additional asthma populations that have (on average) shorter spanning linkage disequilibrium to attempt to more precisely define the true susceptibility gene. The use of African American populations, where the prevalence of asthma is high, may be of great use in this endeavor (39).

DPP10

Polymorphisms in the Interleukin 1 (IL1) gene cluster have been inconsistently associated with asthma (40,41), and linkage with asthma to the interval including this gene (chromosome
2q14–2q32) has been observed (42,43). Two mouse genome screens have also linked airways hyperresponsiveness to the region homologous to 2q14 in the human (34,43–46). To more precisely localize the variants responsible for asthma, Allen and colleagues (the group also primarily responsible for the identification of PHF11 as an asthma gene) fine-mapped this region using a hierarchical association approach nearly identical to that used for the identification of PHF11. Beginning with a population of 244 families in whom associations with IL1 variants had previously been identified, they found an association between asthma and microsatellite marker D2S308 (allele 3), located 800 kb distal to the IL1 gene cluster. A physical map was constructed around this marker, and 105 SNPs were identified, 82 of which were genotyped. Four distinct regions of LD were identified, and significant SNP associations with asthma were localized to within one LD block. The strongest associations were with SNP WTC122, in close proximity to D2S308. Weaker associations with SNPs at the other end of this LD block were found to be due to their association with WTC122 on a shared haplotype, which also included the asthma-associated D2S308*3 allele. Complicating matters was the observation that associations with IgE level were noted with a distinct set of polymorphisms in an adjacent LD block. The associated alleles in this block were not found on shared haplotypes, and multivariate analysis suggested that each associated allele in this region contributed independently to IgE variation. Together these findings suggest that multiple alleles across this region contribute independently to the development of asthma and atopy, and may also interact with one another. The associations were replicated in an independently ascertained cohort of German schoolchildren. In this second cohort, D2308*3 was positively associated with both asthma and atopy phenotypes, as was a multilocus haplotype including WTC122. However, the haplotypes were not identical to those associated in the original population, and single SNP associations with the WTC122 allele are not reported. Like the PHF11 locus, WTC122-D2S308 haplotype associations with severe asthma (but not all asthma) were observed in an asthma case-control panel. The authors conclude that the interaction between these loci is complex, and ‘likely interacts with other unidentified susceptibility alleles’.

Although adjacent to the IL1 cluster, no previously known genes had been localized to the region of association including D2S308 and WTC122. By screening a panel of cDNA libraries for all potential exons within this region, the investigators identified one expressed clone in a fetal brain library. Further characterization including northern- and zoo-blot analysis, and S’ and 3’ RACE, led to the identification of a novel gene approximately 3.6 kb in length that spans more than 1.4 Mb genomic DNA. The protein, named DPP10, shares features with members of the S9B family of DPP serine proteases, which includes DPP4, a widely expressed enzyme that plays a central role in chemokine processing as part of the innate immune system (47). The locus displays a complex pattern of transcript splicing, with eight alternate first exons; four of which localize within the LD block strongly associated with asthma. No coding variants were identified in DPP10. However, the WTC122 alters the sequence of a known promotor element (CdxA) adjacent to one of the alternative exons. The investigators presented some functional evaluation of this polymorphism, demonstrating that the CdxA promotor sequence differentially binds nuclear protein extract depending on which allele is present. The authors suggest that the WTC122 variant may therefore impact DPP10 expression (in either a quantitative or qualitative fashion), but no data is presented regarding WTC122 effects on DPP10 expression.

CONCLUSIONS AND FUTURE DIRECTIONS

In summary, 2003 has seen the identification of two novel asthma genes identified by positional cloning. One of these genes (PHF11) appears to primarily influence total IgE level, and impact asthma severity through this intermediate phenotype. The other, DPP10, is more strongly associated with asthma, although variation at adjacent loci may independently affect atopy-related phenotypes. In addition, weak replication was reported for the asthma associations with ADAM33 first identified in 2002. These three genes hopefully offer new insights into the pathobiology of asthma and allergy. However, it must be stressed that the functional variants, and the mechanisms by which they manifest phenotype have yet to be definitively characterized. Large-scale studies of these genes in additional populations are also needed to accurately quantify the disease risk conferred by these loci.

The papers detailed above demonstrate that asthma genes, like those contributing to other complex traits, can be found using the powerful approach of positional cloning, which does not require any understanding of the genes’ roles a priori. As a result, new pathways, which had previously never been considered important to asthma pathogenesis, are now identified, opening new avenues for asthma research. Not to be overlooked, however, is the enormity of the effort and cost that was required for these first fruits. At the time that the bulk of this work was carried out, the human genome was not as well characterized as it is today, mandating the manual construction of physical gene maps and high-density SNP marker sets. The completion of a high-quality physical map of the human genome should make the next wave of gene hunting more feasible. Additional infrastructure is being developed to support these efforts, including the development of an extensive catalog of the common genetic variation of the genome, less costly and higher-throughput genotyping technologies to enable large-scale fine-mapping efforts, and more powerful statistical methods to efficiently analyze the data. With a rapid acceleration in SNP identification, better description of the LD pattern across the human genome as a result of the HapMap project (48), as well as a decrease in cost of SNP genotyping, 2004 should see a marked increase in the number of genes identified by positional cloning for asthma and other complex traits. Hopefully, as the technical difficulties of positional cloning diminish, we will be able to place greater emphasis on translating these discoveries to improvements in disease classification, prevention and treatment.

REFERENCES


