Genetic Linkage Studies in Alopecia Areata

Amalia Martinez-Mir,# Abraham Zlotogorski,¶ Jurg Ott,¥ Derek Gordon,¥
Angela M. Christiano,**

#Department of Dermatology and *Department of Genetics and Development, Columbia University, New York, New York 10032, USA, ¥Laboratory of Statistical Genetics, Rockefeller University, New York, New York 10021, USA; ¶Department of Dermatology, Hadassah Medical Center, Jerusalem, 91120, Israel.

Corresponding author:
Angela M. Christiano, Ph.D.
Associate Professor
Department of Dermatology
Columbia University
College of Physicians & Surgeons
630 West 168th Street VC-1526
New York, New York 10032

Phone: 212-305-9565
Fax: 212-305-7391
E-mail: amc65@columbia.edu

Abbreviations: AA, alopecia areata; AT, alopecia totalis; AU, alopecia universalis
Keywords: alopecia areata, complex inheritance, genome-wide scan, linkage analysis
ABSTRACT

Alopecia areata (AA) affects approximately 4.6 million individuals in the United States alone. It is typified by patchy hair loss on the scalp that can progress to cover the entire scalp (alopecia totalis), and eventually the entire body (alopecia universalis). Despite its high incidence, the genetic basis of AA is largely unknown. It is now generally accepted that AA fits the paradigm of a complex trait, in which a combination of genetic and environmental factors result in the final phenotype. Genetic studies have been limited thus far to association analyses, which suggest that a permissive HLA status may potentiate the development of AA. However, a systematic screen for identifying the primary genetic mechanisms underlying this disorder has never before been undertaken. Here, we discuss our approach towards the identification of susceptibility genes for AA. In particular, we have recently initiated a comprehensive genetic analysis of AA by performing a genome-wide scan in a collection of AA families with multiple affected family members. There are currently a number of examples of complex diseases of the skin, such as psoriasis and atopic dermatitis, in which genetic studies are being undertaken that substantiate the timeliness of this approach. We anticipate that these studies will lead to the identification of AA susceptibility genes, and provide a foundation for understanding the interactions of these genes with each other and with other variables such as the immune system and environmental factors.
INTRODUCTION

Alopecia areata (AA) is typified by patchy hair loss on the scalp that can progress to cover the entire scalp (alopecia totalis, AT), and eventually the entire body (alopecia universalis, AU). AA affects approximately 4.6 million individuals in the U.S. alone, including males and females of all ages and ethnic groups (Price 1991; Sawaya and Hordinsky 1992; Schwartz and Janniger 1997; Madani and Shapiro 2000). Despite its high incidence, the pathomechanisms underlying AA are largely unknown. An autoimmune etiology has been suggested for many years, but an autoantigen has never been identified. It is now generally accepted that AA fits the paradigm of a complex or multifactorial trait, in which genetic and environmental factors combine to result in the final phenotype (Green and Sinclair 2000; McElwee et al 2001; McDonagh and Tazi-Ahnini 2002; McElwee and Hoffmann 2002). However, genetic studies have been restricted to association analyses, which suggest that a permissive HLA status may potentiate the development of AA. A systematic screen for identifying the primary genetic mechanisms underlying this disorder has never before been undertaken.

Discovery of genes directly implicated in the pathogenesis of AA would have far-reaching implications for basic science, and importantly, for affected individuals. The pathology of AA extends far beyond the physical aspects of hair loss, and can have a deeply disturbing psychological impact. The losses pertain not only to hair, but profoundly impact the quality of life, the ability to function in society and the preservation of self-esteem. There are few diseases as prevalent as AA in which a complete lack of understanding precludes prediction of disease course or even a widely
effective treatment. With the completion of the Human Genome Project, we are now well-positioned to begin a comprehensive genetic analysis of AA. Here, we describe our approach towards the identification of susceptibility genes for AA. We are performing a genome-wide scan in a large cohort of AA families with multiple affected family members. These studies will systematically pinpoint candidate susceptibility genes for AA, and potentially illuminate therapeutic targets for AA patients in the future.

ALOPECIA AREATA: THE DISEASE

It is estimated that AA affects between 2-4% of patients within average dermatology practices (Price 1991). It usually begins as one or several oval patches of nonscarring hair loss in the scalp. They can appear suddenly or more gradually over several days or weeks. The hair loss may regress, or on the contrary, the patches can coalesce and progress to cover the entire scalp (AT) and eventually the entire body (AU). AA is sometimes accompanied by nail changes in the form of pitting, brittleness and splitting. In all, the prognosis of AA is unpredictable and there is no definitive treatment.

The pathogenetic basis and etiology of AA are largely unknown. In the initial stages, the number of hair follicles appears to remain the same, however, in the more advance stages, this number decreases and miniaturization of the anagen hair follicles is observed (McDonagh and Messenger 1996). A cardinal feature of AA is the presence of a lymphocytic infiltrate in the scalp biopsies of affected patients. This, together with the response of patients to steroid treatment, has led to the suggestion of an autoimmune mechanism for many years (Khoury et al 1988; Welsh et al 1994). Recently, it has been hypothesized that melanocyte-associated antigens could induce hair loss in AA (Gilhar et
However, an autoantigen has never been identified. On the other hand, haplotype association studies suggest that a permissive HLA status may potentiate the development of the AA phenotype (Kianto et al 1977; Kuntz et al 1977; Hacham-Zadeh et al 1981; Morling et al 1991; Welsh et al 1994; Colombe et al 1995; de Andrade et al 1999). While the studies describing a positive association between HLA and AA are numerous, there is a lack of biological data implicating specific alleles in the disease phenotype. Moreover, association between HLA and AA has been excluded in some familial cases of AA (Zlotogorski et al 1990).

ALOPECIA AREATA: A COMPLEX TRAIT

The term "complex trait" is used to describe any phenotype that does not exhibit classic Mendelian inheritance attributable to a single gene locus, but does have a genetic component as demonstrated by twin, adoption and epidemiological studies (Lander and Schork 1994).

**Polygenic Inheritance of Alopecia Areata**

No single resource of evidence for polygenic inheritance exists for AA, however, it is becoming increasingly evident that a significant genetic predisposition underlies the AA phenotype (Green and Sinclair 2000; McElwee et al 2001; McDonagh and Tazi-Ahnini 2002; McElwee and Hoffmann 2002). There are several independent lines of evidence in favor of polygenic inheritance of AA (Aita and Christiano 2001). Among them, i) the high prevalence of the trait, typical of complex traits, for which the predisposing alleles are more common than the relatively rare mutations identified for
Mendelian disorders; ii) the Gaussian curve of distribution, for both the stages of disease progression and the distribution of AA, with a threshold effect that could be lowered by the presence of a particular HLA haplotype or autoimmune susceptibility, for example (Welsh et al 1994); iii) the heritability as defined by both the frequency of affected family members, ranging from 3 to 42% (Green and Sinclair 2000), and concordance in twins (Jackow et al 1998); and finally, iv) the presence of congenital AA, strongly suggesting the contribution of genetic factors (de Viragh et al 1997; Bardazzi et al 1999; Bereket et al 2001; Crowder et al 2002). In addition, Sundberg and coworkers (J.P.S., in this issue) have reported the identification of potential susceptibility loci underlying the AA phenotype in the C3H/HeJ mouse model for AA.

**Association Studies in AA**

As a first attempt to identify the genetic basis of AA, a number of association studies for candidate genes have been conducted. There are numerous reports that indicate a significant association between AA and the HLA alleles DQB1*0301 (for severe AA) and DRB1*1104 (Morling et al 1991; Welsh et al 1994). In a family-based study, DQB1*03 alleles were shown to be present in 85% of AA patients as compared to 46% of controls (de Andrade et al 1999). Several authors have also suggested association with the IL-1 receptor antagonist gene (IL1RN) and IL-1 receptor antagonist homologue IL-1L1 (Tarlow et al 1994; Barahamani et al 2002; Tazi-Ahnini et al 2002), as well as with the MXI gene on chromosome 21 (Tazi-Ahnini et al 2000). Finally, strong association between autoimmune polyglandular syndrome type 1 (APS1), caused by mutations in the
AIRE gene on chromosome 21, and AA has also been reported (Betterle et al 1998). In their study, Betterle et al (1998) observed AA in 37% of their APS1 patients.

**Linkage Analysis in Complex Traits**

Genome-wide scan on large cohorts of patients is currently the strategy more frequently applied with success for the study of complex traits. As an example, a number of groups have performed genome-wide scans in affected families with psoriasis, and have identified at least three predisposing genetic loci on chromosomes 4, 6 and 17 (Bhalerao and Bowcock 1998). Atopic dermatitis, another complex skin disease, has also been the subject of genetic studies (Lee et al 2000; Söderhäll et al 2001; Bradley et al 2002). Lee et al (2000) have reported the identification of a major susceptibility locus on chromosome 3. It is anticipated that this type of study will lead to the identification of the pathomechanisms for these common diseases, as has been the case for Crohn's disease, with the identification of the actual alleles conferring susceptibility in the NOD2 gene (Hugot et al 2001; Ogura et al 2001).

**IDENTIFYING SUSCEPTIBILITY GENES FOR ALOPECIA AREATA**

**Genome-wide Scan**

As discussed above, the genome-wide linkage strategy has been widely applied for the study of complex traits. The design of a complex trait study is dependent on a few variables: *i*) the collection of families; *ii*) the number and spacing of genetic markers; and *iii*) the statistical power to identify a locus as a function of these choices. These three considerations of the study design are addressed in the following sections.
I. Diagnostic Criteria and Ascertainment of Alopecia Areata Families

Critical to the success of any genome-wide initiative for a complex disease, a patient collection initiative must fulfill a number of requirements: i) accurate and uniform diagnosis; ii) a large size for the results to be significant; iii) a sample in which the contribution of genetic factors is enriched; and iv) a sample amenable to be analyzed as a single group, or as smaller, more homogeneous subgroups.

Although AA presents with some cardinal morphological and histological features, its diagnosis can be complicated by the variability in the hair loss and characteristic waxing and waning nature of the disease. To achieve homogeneity in the collection of families, the ascertainment, diagnosis and collection of the families included in our study is being undertaken by dermatologists with longstanding expertise in AA and using the diagnostic questionnaire developed by the NIH AA Registry. The sources of the AA families recruited for this study comprise a large group of families from the Israeli Registry ascertained and diagnosed personally by A. Z., as well as families from the NIH AA Registry and different physicians in the U.S.

An example of the patient collection used in our study is shown in Figure 1 and summarized in Table I, showing a family history in 35% of the patients examined. The genetic dissection of complex traits has traditionally been focused in large collections of small families (affected sib pairs, for example). However, it has recently been shown that a small sample of larger pedigrees can potentially derive better results (Tomfohrde et al 1994; Hugot et al 1996; Matthews et al 1996). The rationale for this is the following: first, the use of a small sample reduces the level of genetic heterogeneity ("noise") among
the pedigrees, one of the hallmarks of complex traits; second, a major contribution of
genetic factors is suspected in those pedigrees with several affected family members (i.e.,
such families are "enriched" for predisposing alleles). Finally, knowing the relationship
between the individuals in a large pedigree results in higher statistical power (Terwilliger
and Goring 2000). In order to maximize the success of the gene-mapping study for AA,
we have focused our initial efforts of the genome-wide scan in pedigrees with three or
more affected family members. Importantly, such a collection of families is amenable to
the establishment of different subsets, based on severity of the disease (patchy AA, AT or
AU) or ethnic origin, among other criteria.

II. Genetic Markers

In order to achieve statistically significant results, a sufficient number of highly
polymorphic markers must be genotyped. There are some reports on successful results
with a low number of markers, such as the identification of the \textit{PSOR1} and \textit{IBD1} loci,
conferring susceptibility to psoriasis and Crohn's disease, respectively (Matthews \textit{et al}
1996; Trembath \textit{et al} 1997; Hugot \textit{et al} 2001). In our study, however, we have chosen to
perform a genome-wide scan using a panel of 324 microsatellite markers, with an average
marker spacing of 10 cM and a semi-automated fluorescence-based genotyping system.
Most of the markers are chosen from version 8.0 of the Marshfield fluorescence-labeled
genome screening set. This approach has been described in detail elsewhere (Aita \textit{et al}
1999; Liu \textit{et al} 2001; Liu \textit{et al} 2002) and utilized by our group in collaboration with the
Columbia University Genome Center (Ahmad \textit{et al} 1998a; Ahmad \textit{et al} 1998b; Martinez-
Mir \textit{et al} 2002).
III. Statistical Analysis

Due to the inherent nature of complex traits, it is expected that a number of genetic components will be contributing to the final presentation of the phenotype. However, different combinations of genetic factors can be contributing to the disease in each family. For this reason, once the data-set from the genome-wide scan is obtained, it is extremely critical to perform a thorough and exhaustive statistical analysis to extract all possible information. These data-sets are amenable to be analyzed with a large number and wide variety of statistical tests.

As a first approach, we will apply the following test statistics to the data-set obtained in the genome-wide scan: i) the heterogeneity LOD score (Smith 1963; Ott 1999), maximized over four settings of the penetrance parameters (MAXHLOD); and ii) the mean test for affected sib-pairs, as implemented in the ANALYZE program (Terwilliger and Ott 1994) (ASP); a test of allele sharing that uses all sibs (Terwilliger and Ott 1994) (ALLSIBS); and a likelihood version of the transmission disequilibrium test (Spielman et al 1993), as developed by Terwilliger (Terwilliger 1995) (TDT-LIKE).

With the MAXHLOD calculations, we apply a model-based linkage analysis in which the LOD score is calculated under both autosomal dominant and autosomal recessive patterns of inheritance. For both models, two different values of penetrance are considered, 50 and 80%. ASP, ALLSIBS and TDT-LIKE tests, on the other hand, are chosen because they all are genetic model-free tests (Elston 1989; Hodge and Elston 1994), in the sense that they do not require a specification of the genetic model parameters (penetrance and disease allele frequency). MAXHLOD is used because it has
been shown that it is at least as powerful in localizing disease loci as tests like ASP and ALLSIBS (Abreu et al 1999), and is, under certain circumstances, a more precise indicator of the location of a disease locus than statistics like ASP and ALLSIBS (Finch et al 2001). Finally, TDT-LIKE is chosen because it has been shown that it may be more powerful than linkage tests (MAXHLOD, ASP, ALLSIBS) when there is linkage and linkage disequilibrium between a disease and a marker locus. Some of the AA families collected by A.Z. in Israel are Ashkenazi Jewish (AJ). The AJ population is considered a genetically isolated population and the extent of linkage disequilibrium is thought to extend over larger regions of the human genome (Ostrer 2001). The strategy used here is similar to exploratory methods employed by other researchers analyzing genome-scan data for complex traits (Wise et al 1999; Wise and Lewis 1999), with the exception that we will study a whole chromosome at a time, rather than subdividing each chromosome into bins. We consider the whole chromosome as the unit of measure because in simulated data-sets it has been observed that the methods we plan to employ are more powerful at determining the correct chromosome, rather than a particular region of a chromosome harboring a disease susceptibility locus (Gordon et al 2001).

Identification of Genes and DNA Variants Conferring Genetic Susceptibility for Alopecia Areata

Once a complex trait has been mapped to one or several susceptibility loci, the task of identifying the specific alleles conferring susceptibility for the phenotype of interest still remains. First, a small interval of amenable size for positional cloning is rarely identified in initial linkage studies. Second, the nature of a complex disease implies that
the alleles predisposing to the final phenotype can be numerous. Finally, common polymorphisms, rather than nonsense or frameshift mutations, are the expected gene variants contributing to the etiology of complex traits. For this reason, multiple approaches are utilized towards the identification of candidate susceptibility genes and alleles for AA.

I. Fine-Mapping of Susceptibility Loci for Alopecia Areata

The results from the first stage of the genome-wide scan, as described above, are expected to indicate the most likely location for the susceptibility genes for AA. In order to confirm the results from this first stage, as well as to exclude false positive results, a second stage of analysis is performed. A dense map of polymorphic markers in the regions of interest is developed using both microsatellite markers and SNPs. These are available in the public databases, such as UCSC, NCBI, Ensembl, and deCODE Genetics (Kong et al 2002), for map and sequence information on microsatellites and SNPs; the Marshfield genetic map (Broman et al 1998), for map information on microsatellites; and the SNP consortium database (Sachidanandam et al 2001), for SNP selection. The entire collection of AA families, together with newly collected pedigrees, can then be genotyped in this fine-mapping stage. In the case SNP analysis is required for a large genomic region, and in order to optimize the genotyping effort in this refinement stage, the presence of haplotype blocks can be determined (Daly et al 2001; Gabriel et al 2002). The use of haplotype blocks can considerably reduce the cost of genotyping, since a few SNPs are typed which represent the entire haplotype. Similar statistical analyses as those
II. Identification of Candidate Genes

Once the candidate regions cannot be further refined, it is necessary to proceed with the analysis of the genes therein by both in silico and experimental procedures. We now have access to continuous updated sequence data of the human genome at three main databases, UCSC, NCBI and Ensembl. Based on the information available from these databases, a detailed physical map of the region of interest is built in order to prioritize the candidate genes to be analyzed. This usually includes previously known genes, novel genes, ESTs and predicted genes. As a second but parallel approach in the case of AA, we can cross-reference the positional data from the genome-wide scan with expression data. In particular, Carroll et al (2002) have already performed expression studies in human AA patients and in a mouse model for AA. As a result, they have established between two groups of differentially expressed genes for the initial and late stages of disease development. Finally, several animal models for AA have been described. Among them, the C3H/HeJ mouse and the DEBR rat have been extensively characterized (McElwee and Hoffmann 2002). Recently, Sundberg and coworkers (J.P.S., in this issue) have reported the first attempt for the identification of genes underlying the AA phenotype in the C3H/HeJ mouse model. As a result, several suggestive loci have been identified on chromosomes 9 and 17. The study of the syntenic regions containing AA loci in mouse and human can help in defining candidate genes.
**III. Gene Characterization and Detection of DNA Variants**

Once a candidate gene has been identified the final goal is the characterization of the actual alleles conferring susceptibility to the phenotype of interest. With this aim, the candidate gene is sequenced in a collection of patients and control individuals.

Since genomic DNA is the material used in a genome-wide scan, it is necessary to know the genomic structure of the candidate gene for PCR primer design. The strategy followed will depend on the available information on each gene. For those genes fully sequenced as part of the Human Genome Project public efforts, the intron/exon boundaries are known or can otherwise be determined by cDNA and genomic sequence comparison. For those genes partially sequenced, a PCR-based cloning strategy can be followed. Finally, in the case of predicted genes, the predicted gene structure must be verified by RT-PCR.

For each candidate gene, intron/exon boundaries will eventually be identified, as well as flanking intron sequences for the design of exon-specific PCR amplification primers for the detection of DNA variants. For each gene both coding and non-coding regions, mainly the promoter region, can then be analyzed. The first samples to be sequenced will be usually those corresponding to the pedigrees linked to a particular chromosomal region. Once DNA variants are identified, both in coding and non-coding sequences, it is necessary to test whether they are significantly associated to the disease phenotype. With this goal, their frequency can be determined in the complete collection of AA families, isolated AA cases and a control-matched population. If needed, the cohort of AA pedigrees can be further subdivided according to severity and ethnic origin.
to test for associations between the variants identified and particular subgroups of the patient collection.

We anticipate that this type of studies will provide a foundation for understanding the interactions of these genes with each other and with other variables such as the immune system and environmental triggers. Ultimately, it is expected that they will help to define therapeutic targets for the future, and eventually provide an effective treatment for this psychologically devastating dermatologic disorder.
ACKNOWLEDGEMENTS

The authors are grateful to the NIH AA Registry for its invaluable contribution to this study. This study was supported in part by grants from the National Alopecia Areata Foundation (A.M.C. and A.Z.), by the General Administrator, Israel (A.Z.), and NIH grants N01-AR-0-2249 (NIH AA Registry), K01-HG00055 (D.G.), MH59492 (J.O.) and HG00008 (J.O.).
ELECTRONIC-DATABASE INFORMATION

The URLs for data presented herein are as follows:

Center for Medical Genetics, Marshfield Medical Research Foundation,
http://research.marshfieldclinic.org/genetics/

Ensembl Genome Server, http://us.ensembl.org/

UCSC Human Genome Project Working Draft ("Golden Path"),
http://genome.ucsc.edu/index.html


TDT-AE software, http://linkage.rockefeller.edu/soft/list4.html#tdtae
REFERENCES


Khoury EL, Price VH, Greenspan JS: HLA-DR expression by hair follicle keratinocytes in alopecia areata: Evidence that it is secondary to the lymphoid infiltration. *J Invest Dermatol* 90:193-200, 1988


Terwilliger JD: A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am J Hum Genet* 56:777-787, 1995


Welsh EA, Clark HH, Epstein SZ, Reveille JD, Duvic M: Human leukocyte antigen-DQB1*03 alleles are associated with alopecia areata. *J Invest Dermatol* 103:758-763, 1994


Zlotogorski A, Weinrauch L, Brautbar C: Familial alopecia areata: No linkage with HLA.
*Tissue Antigens* 36:40-41, 1990
Figure 1. Representative example of AA pedigrees available for this study.
**Table I.** Summary of AA patients examined

<table>
<thead>
<tr>
<th>Number of affected family members</th>
<th>Number of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (no family history)</td>
<td>91</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>