

HLA and Disease

EDITED BY

Jean Dausset

Institut de recherches sur les maladies du sang,
Hôpital Saint-Louis, 2 Paris X^e

Arne Svejgaard

Tissue Typing Laboratory of the Blood Grouping Department,
State University Hospital of Copenhagen

Munksgaard

Copenhagen

North and South America:

Williams & Wilkins Co.

Baltimore

Genetic Analysis of HLA-Associated Diseases: The »Illness Susceptible« Gene Frequency and Sex Ratio in Ankylosing Spondylitis

K. K. Kidd¹, D. Bernoco², A. O. Carbonara³, V. Daneo⁴,
U. Steiger⁵ & R. Ceppellini²

Introduction

Diseases which show familial concentration but do not show simple Mendelian patterns are difficult to describe unambiguously in genetic terms (Kidd & Cavalli-Sforza (1), Kidd & Spence (2)). In fact, familial concentration is not even proof that genetic factors are relevant to the disease. On the other hand, provided that one is willing to consider population stratification to be a genetic (albeit population genetic/social) phenomenon, a population study can provide evidence for a genetic component. A significant association of any disease with a good genetic marker, such as an HLA antigen, is *per se* evidence that a genetic element exists in the etiopathogenesis of that

disease. However, unless the marker itself is clearly of etiologic significance, the interpretation of the association in genetic terms will not be clear. With HLA associated diseases both of the above complications exist; diseases being studied usually do not show simple Mendelian patterns and there are, at present, no known pathological alleles at the HLA loci (with the single exception of the C2 and C4 amorph mutants (3, 4) that can explain the significant, but not complete, association of the illness with particular allele and/or haplotypes. Since HLA markers (antigens) are co dominant traits with full penetrance for their serological expression, several reasons for the incomplete association of the disease with the antigen can be given. One explanation is linkage to (or hitchhiking within) the HLA region of a gene directly responsible for the disease (that is, an "Illness susceptibility" allele, hereafter abbreviated to *Is*). At the moment the best candidates may be the hypothesized Ir (immune response) genes, but other possible

¹) Department of Human Genetics, Yale University School of Medicine, Connecticut, U.S.A.

²) Basel Institute for Immunology, Basel, Switzerland.

³) Istituto di Genetica Medica, Torino, Italy.

⁴) Centro di Reumatologia, Ospedale S. Giovanni Battista, Torino, Italy.

⁵) Rheumatology Department, University Clinics of Basel, Basel, Switzerland.

polymorphisms, such as hormone receptors, should also be considered (5). A second explanation is epistasis, as in the typical case of antigen Lewis b (6). Ragweed pollen allergy is possibly due to epistasis between the ability to recognize a pollen determinant, due to an HLA linked Ir gene, and the high level phenotype for IgE (7-10), although conflicting evidence exists (11, 12). Another explanation, not necessarily in contrast with the others, is “incomplete penetrance” (lack of the phenotype expression expected for a given genotype) of the HLA antigen allele in regard to illness susceptibility (if directly responsible) or of the linked *Is* allele.

Because the diseases that are being considered are already known to show no simple Mendelian pattern, incomplete penetrance in some form is a likely and almost necessary component in any genetic explanation. Such incomplete entrance may be caused by a variety of factors, both genetic and environmental. On the genetic side, besides the simplest case mentioned above of epistatic interaction with a single major locus, polygenic inheritance, with only one of the many loci linked to the HLA system, is another purely genetic situation. Environmental explanations would include, for instance, a triggering infectious disease and also some “familial” factors, such as cohabitation with relatives sharing not only a similar genotype but also similar conditions of housing and habits of living.

Existence of environmental variation dilutes out the degree to which genetic factors determine the variation in the population for presence or absence of a trait or illness. The relative amount of genetic variation is frequently expressed as heritability (h^2), the ratio of the genetic variance in the population to the total variance for the trait. The authors feel that heritability ignores some typical and most

interesting features of human genetics, for example, possible modes of gene action and the difficult distinction between biological and cultural inheritance. Thus, when applied to human diseases, *heritability* is often a misnomer. Therefore, the authors will not use it. A more detailed discussion of the inappropriateness of heritability in human genetics is given in Matthyse & Kidd (13).

Family data on diseases are rarely not insufficient to be informative; population association data of a disease and a marker locus are often not readily interpretable in genetic terms. The authors wish to show that family data on both a disease and an associated (HLA) marker can be used to gain a clearer understanding of the genetics

TABLE I

Criteria followed in Basel and Torino for diagnosis of Ankylosing Spondylitis¹

Subjective and functional
(a) limitation of motion of lumbar spine ²
(b) history of persistent pain
(c) chest expansion < 3 cm (measured at fourth intercostal space)
X-ray grading (each side graded separately)
0: normal
1: doubtful changes
2: minimal definite change
3: moderate or advanced sacroiliitis with erosion, evidence of sclerosis, widening, narrowing or partial ankylosis
4: total ankylosing
Definite AS: grade 3-4 X ray, bilateral, plus at least one clinical sign; grade 3-4 unilateral or grade 2 bilateral plus a+c or b+c
Probable AS: grade 3-4 bilateral in absence of any clinical sign
To maximize penetrance: all individuals ≤ 20 years of age were excluded
To maximize homogeneity: all cases with complications were excluded

1) As recommended by Gofton *et al* (18)

2) As recommended by Macrae & Wright (19)

of the disease. Ankylosing spondylitis (AS) is a paradigmatic model because AS is the disease that shows the highest relative risk with an HLA antigen (14, 15). Family data have been collected in Torino (TO) (16) and Basel (BS) (17) on 53 index or primary cases who had at least one first degree relative that could be examined as completely as the proband, according to a rigid application of the criteria in Table I (18, 19). Two groups of normal controls were also examined with the same diagnostic criteria: (a) a selected sample of

B27-positive individuals who had not reported subjective symptoms nor had been in any way aware of being affected by AS; and (b) a selected sample of B27 negative individuals, mostly chosen for being B7 positive because of the well known crossreactivity between these two B antigens. All data in the two studies (BS and TO) were collected without communication between the rheumatologists, the HLA typists, and the statisticians until the work had been completed.

Results and Discussion

The diagnostic criteria listed in Table I were followed in both the Basel and Torino studies. Table 11 presents the data for the primary cases (proband) and population controls. A preponderance of males among ankylosing spondylitis patients is found, as has been reported in other studies. The preponderance of males in the normal control populations reflects the male preponderance among employees of Hoffman LaRoche in Basel and in the blood donor panel in Torino. These controls give the B27-antigen frequency in the population. Through some difference exists between Basel and Torino, the increased frequency of B27 among AS patients is spectacular and highly

significant, as reported in all previous studies (see 14, 15). From the data on first degree relatives, presented in Table 111, the familiarity of ankylosing spondylitis is obvious. The two control populations included in Table III were examined using the same diagnostic criteria. These data confirm the low incidence of AS in a non-B27 population and the high frequency of affected individuals (7 out of 43 = 16 %) among the "normal" B27 positive controls, using a more rigorous diagnostic examination than in the report by Calin & Fries (20). Only one of the affected first degree relatives, a mother, did not have B27; however, in that family the proband's *Is B27* haplotype came from the father,

TABLE II

Number studied, sex, and B27 phenotype frequency in patients (primary cases) and controls studied in Basel (BS) and Torino (TO)

Number studied	Sex		Class	Area	B27 frequency
	M	F			
58	48	10	AS patients	BS	0.93
623	459	164	Normal controls ¹	BS	0.100
21	20	1	AS patients	TO	0.90
1428	980	448	Normal controls ²	TO	0.043

¹) random employees of Hoffman La Roche

²) random blood donors

TABLE III*Familiarity of Ankylosing Spondylitis*

Primary case			First-Degree Relatives		Normal		Affected	
Area	No.	Type	Type	No.	B27 +	B27-	B27+	B27-
BS	7	Parent	Child	13	4	7	2	0
BS	15	Sib	Sib	32	13	13	6	0
BS	18	Child	<u>Parent</u>	<u>27</u>	<u>10</u>	<u>9</u>	<u>8</u>	<u>0</u>
BS			All	72	27	29	16	0
TO	3	Parent	Child	5	4	1	0	0
TO	11	Sib	Sib	15	6	4	5	0
TO	9	Child	<u>Parent</u>	<u>16</u>	<u>6</u>	<u>7</u>	<u>2</u>	<u>1</u>
TO			All	36	16	12	7	1
BS + TO combined			All	108	43	41	23	1
BS	147	Non-B27 controls			-	146	-	1
BS	43	"Normal" B27 controls			36	-	7	-

who was also AS affected. Also worth noting is that one affected individual (B7+) was found among the non B27 controls. Thus, to the best diagnostic criteria accepted today, ankylosing spondylitis can occur, albeit rarely, in the absence of the antigen B27.

Tables IV and V show that the preponderance of affected males among probands neither occurs among the "normal" B27 + controls nor among the first degree relatives; Calin & Fries (20) also found essentially equal frequencies of affected males and females in their study.

TABLE IV*Frequency of Females (ff) in Ankylosing Spondylitis*

Area	Class of Data	Total	Number		
			Males	Females	ff
BS	Primary cases	58	48	10	0.17
BS	Secondary cases	72	8/33 ¹	8/39 ¹	0.46 ²
BS	"Normal" B27+	43	5/34 ¹	2/9 ¹	0.60 ²
TO	Primary cases	21	20	1	0.05
TO	Secondary cases	36	4/19 ¹	4/17 ¹	0.53 ²

¹ Ratio of number affected by AS to total number studied

² ff corrected for a 1:1 sex ratio in sample

TABLE V*Frequency of Affected Relatives according to Sex¹*

		Primary Cases	
		Males (n = 33)	Females (n = 8)
All First-degree relatives	Males	9/41 = 0.22 ± 0.06	3/11 = 0.27 ± 0.13
	Females	11/42 = 0.26 ± 0.07	1/14 = 0.07 ± 0.07

¹ Values given as number affected/total = mean ± standard error

χ^2 (3 d. f.) = 2.3 gP = 0.50

The family data also show that AS is as frequent, within the statistical limits of the small sample, among the relatives of female probands as of male probands, suggesting that there is no "hereditary" component to any hypothesized sex limitation or greater male susceptibility. An obvious conclusion is that the male preponderance among AS patients is not a reflection of a truly greater incidence among males. Possible explanations for this could be greater severity in males, a greater tendency for the diagnosis to be made in males, and a greater tendency for males to seek treatment. Whatever the explanation, the "fact" that males are more frequently affected with AS than females is probably not true in an absolute sense.

Genetic Analyses

Analyses of family data normally involve testing the goodness of fit of different genetic hypotheses (models) to the observed data. The two models most commonly used in human genetics are the simple extremes of single major locus (SML) and multifactorial polygenic (MF) inheritance (see Matthyse & Kidd (12) for a review of these two models). One major problem with these models is that they are statistically indistinguishable and agreement cannot be tested when family incidence data are used without additional sources of genetic information (1, 21). Nonetheless, both models have been applied to the AS family data to see whether the results would be biologically reasonable.

The single parameter of the MF model needing estimation is the genetic correlation among relatives. This correlation is often translated into an heritability (h^2) value; however, the correlation or heritability does not apply to the measured trait, in this case ankylosing

spondylitis, but to a hypothetical underlying continuous and normally distributed variable called "liability to disease". The general incidence of the illness and the frequency of affected first-degree relatives are the necessary data for estimating the correlation. The generally accepted population prevalence for AS is about 0.5 % for males and considerably less for females. The average incidence among first degree relatives is 0.22 ± 0.04 . These values give an estimate of the tetrachoric correlation (1) of about 0.65, a nonsensical result according to the MF model, for which the theoretical maximum is 0.50, i.e. $h^2 = 100$ %. However, the general incidence estimate is based on the frequency of patients seeking treatment and not upon a complete diagnostic survey of the population using both functional tests and X rays. In fact, primary and secondary cases are usually ascertained according to widely different criteria. The authors' finding that 16 % of a sample of B27+ individuals were affected suggests that the population prevalence may be about 1.5 % for both males and females. Using this value for the incidence in the population and 0.22 for the frequency of affected first degree relatives, the correlation coefficient is lower but still above 0.50 the heritability (h^2) is > 100 %. Such high correlations are often interpreted as suggesting that a locus of major effect is involved.

Methods for using family data to estimate the parameters of the SML model are extensively discussed elsewhere (1, 2, 13, 21, 22). To save space the model and methods will not be discussed here. Analyses show that no solution is possible for an incidence of the disease in the population of less than 1.0 %. However, the closest solution to the data (using the method of Suarez *et al.* (22)) predicts a population incidence for AS of about

TABLE VI

Genetic Analysis of Ankylosing Spondylitis by the Method of Thomson & Bodmer (23)

Basic Values:

$q = \text{frequency of HLA-B27 allele (B27)} = 0.048$

$f(M | Ill) = \text{frequency of the marker (B27) phenotype among ill (AS) patients} = \frac{73}{79} = 0.924 \pm 0.030$

$k_{rec.} = 1 - \sqrt{1 - f(M | Ill)} = 0.724$

$k_{dom.} = (1 - (1 - f(M | Ill)) / (1 - q)) = 0.920$

Observed genotype distribution in AS patients (primary cases only)

	B27/B27	B27/-	-/-
BS	3	51	4
TO	<u>0</u>	<u>19</u>	<u>2</u>
Total	3	70	6

Expected genotype distributions

Recessive hypothesis	41.5	31.5	6	$X^2_1 = 82.5$ $p < 0.0001$
Dominant hypothesis	3.5	69.5	6	$X^2_1 = 0.075$ $p = 0.78$

1.6 % and frequencies of AS among first degree relatives of about 22 %, essentially the same as the “adjusted” values used for the MF model. This SML “solution” predicts a gene frequency of 1.8 % for the Is allele and penetrances (that is, risks of actually contracting the illness) of < 0.001, 0.43, and > 0.995 for the homozygous “normal” (*is/is*), heterozygous (*is/Is*), and homozygous “susceptible” (*Is/Is*) genotypes, respectively. The predicted genotypic composition of the population of affected individuals is about 3 % homozygous “normal” (i.e. phenocopies), 95 % heterozygous, and 2 % homozygous “susceptible”. Although there is no way to test this model statistically using the available data, the SML results, in combination with the unrealistically high correlations of the MF model, strongly suggest that a single locus hypothesis is the more likely one. This “conclusion” is independent of the association of AS with B27.

Thomson & Bodmer (23) (largely reprinted in this volume) have developed a method for testing two simple models that assumes linkage disequilibrium with a co

dominant marker. Their method is based only on the genotype distribution among affected probands and does not utilize family data. This makes it useful in practice since, unfortunately, family data are at present not generally available for HLA associated diseases. In Table VI data are organized appropriately and the results of the analysis by the Thomson & Bodmer method are presented. The analysis strongly rejects the recessive hypothesis and is in almost exact agreement with the dominant hypothesis. Thus, AS appears to be due to an illness susceptibility allele at a locus closely linked to the HLA B locus and susceptibility seems to be inherited in a dominant fashion. The linkage disequilibrium estimate, measured in this analysis as k (= 0.928), is a very high value, providing indirect proof that tight linkage is responsible for the strong population association.

As explained by Kidd & Ceppellini (24), it is possible to modify the Thomson & Bodmer method to incorporate estimates of the frequency of the disease among individuals with the marker and of the “penetrance”, that is, the frequency of the

illness among those with the susceptibility allele. Utilizing whichever model fits the data, it is then possible to estimate all four haplotype frequencies. Though more exact treatments are possible, the low frequencies of the marker allele (B27) and the illness (AS) allow a simplified calculation for the dominant model, as outlined in Table VII. The estimates for the penetrance and the haplotype frequencies agree well with the SML solution obtained without consideration of the HLA association.

Some caveats about the penetrance estimate are very important in evaluating these results. It is impossible to generalize from this penetrance estimate to families not ascertained through a proband. The value of 0.38 is not significantly less than 0.5, the approximate value expected in families if an allele at an independent locus were also required for development of the illness. Probands, being affected, would be carriers for both alleles; because of independent segregation, only about half of the first-degree relatives with the *I_s* allele at the main locus would also have the necessary allele at the second. Thus, the family estimate of penetrance will be an

overestimate for an unrelated group of individuals. The accuracy of the estimate depends on the degree of epistasis and, if epistasis exists, on the frequencies of the interacting alleles.

Another caveat about the family estimate of penetrance is that it includes recombination that separates the marker from the susceptibility allele.

Even an exact calculation of the penetrance from family data would include corrections for the age of onset distribution (only roughly done in this study by limiting consideration to adult relatives) and for the probability of ascertainment, which is determined by the number of affected individuals in a family. In spite of these potential biases, not all of which operate in the same direction, the penetrance estimate is roughly consistent with the population and control data. From what is known about AS, it is very unlikely to be as high as 100 %, and the “normal” B27+ control data show that it cannot be below 16 %.

The estimated frequency of the haplotype *m I_s* is too low (0.002) to explain the 4 % of AS patients who lack the B27 marker, even if the “penetrance”

TABLE VII

Summary of a simplified Bayesian probability analysis for Ankylosing Spondylitis

M = marker allele for B27 antigen; frequency of M = 0.048 ± 0.01

I = allele for illness susceptibility

A. observed frequency of M among AS patients = $0.92 \pm 0.03 = \Pr(M|I) = \Pr(M|AS)$

B. observed frequency of AS among M + (B27 +) population = $0.16 \pm 0.06 = \Pr(AS|M)$

C. observed frequency of AS among M+ (B27+)

first degree relatives of probands = “penetrance” = $0.38 + 0.06 = \Pr(AS|I)$

Therefore, $\Pr(I|M) = 0.16/0.38 = 0.42$

and $\Pr(I) = \Pr(I|M) \cdot \Pr(M) / \Pr(M|I) = 0.022$

It follows that:

$x_1 = \text{freq. (MI haplotype)} = \Pr(I|M) \cdot \Pr(M) = 0.020$

$x_2 = \text{freq. (Mi)} = \Pr(M) - x_1 = 0.028$

$x_3 = \text{freq. (mI)} = \Pr(I) - x_1 = 0.002$

$x_4 = \text{freq. (mi)} = 1 - x_1 - x_2 - x_3 = 0.950$

$\Delta = x_1x_4 - x_2x_3 = 0.019$

for that haplotype were 100 %. Sampling error in the estimate of the penetrance or in the frequency of affected individuals who are B27 negative might have produced this

apparent discrepancy. However, the SML solution discussed earlier predicted about 3 % phenocopies, a value in close agreement.

Conclusions

Though mathematical genetics has many limitations when applied to familial traits that are not obviously Mendelian, it becomes a powerful tool when a marker locus can also be studied in those families. The authors have shown that the oft mentioned preponderance of affected males is found neither in families of AS patients nor in population surveys and one must conclude that the apparent sex limitation is mainly due to ascertainment. Susceptibility to AS is inherited as an autosomal dominant. The locus seems to be tightly linked to HLA B. According to this model, the susceptibility allele has a gene frequency of about 2 %. Also, it is in

strong linkage disequilibrium with the allele determining HLA B27: nearly all (93 %) of the haplotypes with the *is* allele also carry the *B27* allele. The majority (62 %) of the carriers of the *Is* allele do not develop AS. As repeatedly mentioned, this incomplete penetrance may be due to environmental factors or to epistasis with an allele at an independent locus.

Obviously, this is a formal explanation with a strong heuristic component. Tests of these conclusions will have to be based on additional family studies of B27 negative AS probands, of female AS probands and of relatives of different degrees.

Acknowledgement

The part of the work done in Torino was supported by C.N.R. grant "Centro di Studio per l'Istocompatibilità a

l'Immunogenetica". The analyses were supported in part by N.I.H. grant NS 11786.

References

1. Kidd, K. K. & Cavalli Sforza, L. L. (1973) An analysis of the genetics of schizophrenia. *Soc. Biol.* **20**, 254-265.
2. Kidd, K. K. & Spence, M. A. (1976) Genetic analyses of pyloric stenosis suggesting a specific maternal effect. *J. med. Genetics* **13**, 290-294.
3. Agnello, V. M., de Bracco, M. E. & Kunkel, H. J. (1972) Hereditary C2 deficiency with some manifestations of systemic lupus erythematosus. *J. Immunol.* **108**, 873-881.
4. Hauptmann, G., Grosshause, E., Heid, E., Mayer, S. & Basset, A. (1976) Lupous érythémateux aigu avec déficit complet de la fraction C4 du complément. *La Nouv. Presse Méd.* **3**, 881-885.
5. Ceppellini, R. (1971) Old and new facts and speculations about transplantation antigens of man. In: *Progress in Immunology I*, ed. Amos, D. B., p. 973. Academic Press Inc., New York.
6. Ceppellini, R. (1959) Physiological genetics of human blood factors. In: *Ciba Foundation symposium on Biochemistry of Human Genetics*, eds. Wolstenholme, G. E. W. & O'Connor, C. M., Churchill, London.
7. Blumenthal, M. N., Amos, D. B., Noreen, H., Mendell, N. R. & Yunis, E. J. (1974) Genetic mapping of Ir locus in man:

- Linkage to second locus of HL-A. *Science* **184**, 1301-1303.
8. Levine, B. B., Stember, R. H. & Fotino, M. (1972) Ragweed hay fever: Genetic control and linkage to HL-A haplotypes. *Science* **178**, 1201-1203.
 9. Marsh, D. G. & Bias, W. B. (1974) Control of specific allergic response in man: HL A associated and IgE regulating genes. *Fed. Proc.* **33**, 774.
 10. Marsh, D. G., Bias, W. B. & Ishizaka, K. (1974) Genetic control of basal serum immunoglobulin E level and its effect on specific reaginic sensitivity. *Proc. nat. Acad. Sci. (Wash.)* **71**, 3588-3592.
 11. Bias, W. B. & Marsh, D. G. (1975) HL-A linked antigen E immune response genes: An unproven hypothesis. *Science* **188**, 375-377.
 12. Black, P. L., Marsh, D. G., Jarrett, E., Delespesse, G. J. & Bias, W. B. (1976) Family studies of association between HLA and specific immune responses to highly purified pollen allergens. *Immunogenetics* **3**, 349-368.
 13. Matthyse, S. W. & Kidd, K. K. (1976) Estimating the genetic contribution to schizophrenia. *Amer. J. Psychiatry* **133**, 185-191.
 14. Brewerton, D. A., Caffrey, M., Hart, F. D., James, D. C. O., Nicholls, A. & Sturrock, R. D. (1973) Ankylosing spondylitis and HL-A 27. *Lancet* **301**, 904-907.
 15. Schlosstein, L., Terasaki, P. I., Bluestone, R. & Pearson, C. M. (1973) High association of an HL A antigen, W27, with ankylosing spondylitis. *New. Eng. J. Med.* **288**, 704.
 16. Daneco, V., Migone, N., Modeno, V., Bianchi, S. D., Alfieri, G., Diotallevi, P., Carbona, A. O. & Piazza, A. (1976) Family studies and HLA typing in ankylosing spondylitis and sacroiliitis. *J. Rheumatol.*, (in press).
 17. Truog, P., Steiger, U., Contu, L., Galfre, G., Trucco, M., Bernoco, D., Bernoco, M., Birgen, I., Dolivo, P. & Ceppellini, R. (1975) Ankylosing spondylitis (AS): A population and family study using HL-A serology and MLR. In: *Histocompatibility Testing 1975*, ed. Kissmeyer-Nielsen, F., p. 788-796. Munksgaard, Copenhagen.
 18. Gofton, I. P., Bennett, P. H., Bremmer, I. M., Bywaters, E. G. L., Calabro, J. I., McEwan, C. & Martel, W. (1966) Population studies of rheumatic diseases, p. 456. *Proc. 3rd International Symposium*, New York.
 19. Macrae, I. F. & Wright, U. (1969) Measurement of back movement. *Ann. rheum. Dis.* **28**, 584.
 20. Calin, A. & Fries, J. F. (1975) Striking prevalence of ankylosing spondylitis in "healthy" W27 positive males and females. *New Eng. J. Med.* **293**, 835-839.
 21. Reich, T., James, J. W. & Morris, C. A. (1972) The use of multiple thresholds in determining the mode of transmission of semi continuous traits. *Ann. Hum. Genetics* **36**, 163-184.
 22. Suarez, B. K., Reich, T. & Trost, J. (1976) Limits of the general two allele single locus model in genetic explanations. *Ann. Hum. Genetics*, (in press).
 23. Thomson, G. & Bodmer, W. F. (1977) The genetics of HLA and disease association. In: *Measuring Selection In Natural Populations*, eds. Barndorff-Nielsen, O., Christiansen, F. B. & Fenchel, T. Munksgaard, Copenhagen, (in press).
 24. Kidd, K. K. & Ceppellini, R. (1977) A bayesian method for estimation of HLA-associated illness susceptibility (Is) allele frequencies. I. Dominant susceptibility. In: *HLA and Disease*, eds. Dausset, J. & Svejgaard, A., p. 81, Munksgaard, Copenhagen.

A Bayesian Method for Estimation of HLA Associated Illness Susceptibility (Is) Allele Frequencies. I. Dominant Susceptibility

Kenneth K. Kidd¹ & Ruggero Ceppellini²

Introduction

Thomson & Bodmer (1) present two types of genetic analysis for disease phenotypes associated with a genetic marker. Their first method utilizes the marker phenotype frequency in a population of unrelated affected individuals; the second is derived from more general models by Day & Simons (2) and utilizes the frequencies of the number of marker chromosomes (haplotypes) shared by affected sib pairs. Both methods involve tests for specific models of inheritance of illness susceptibility (Is) recessive or rare dominant. Green & Woodrow (3) have also produced an analysis utilizing frequencies of shared chromosomes among affected siblings to test the null hypothesis of there being no association between marker locus and illness. None of these papers considers direct estimates of the haplotype frequencies, though Thomson &

Bodmer (1) consider indirect estimates for their recessive model applied to data on pairs of affected siblings. Direct estimates are possible when certain types of data are available: (a) the frequency of illness among individuals with the associated marker, $f(\text{Ill} | \text{M})$; and (b) the “penetrance” or frequency of the illness among those with the susceptibility genotype.

The general importance of these types of data, particularly estimating penetrance from family data, has been recognized in the design of the study of ankylosing spondylitis in Basel and Torino (Ceppellini, personal communication). Suitable data are consequently already available for ankylosing spondylitis; they may soon be collected for other HLA associated illnesses. $f(\text{Ill} | \text{M})$ can be collected using the same criteria for diagnosis as in probands and relatives and should be less subject to bias and uncertainty than estimates of the frequency of the illness in the general population. The penetrance can be estimated in many ways, but the simplest and most common is the proportion of affected individuals

¹) Department of Human Genetics, Yale University School of Medicine, Connecticut, U.S.A.

²) Basel Institute for Immunology, Basel, Switzerland.

among relatives who carry the same marker chromosome (for the dominant model) or marker chromosomes (recessive model) as the proband. These estimates, combined with the marker allele frequency and the frequency of the marker phenotype among affected individuals, can be used to

calculate haplotype frequencies. These types of data should become available before sufficient pairs of affected sibs have been studied for the other types of analysis, especially if the association is strong and illness susceptibility is dominant.

Method

A more exact and general treatment of the dominant and recessive models, utilizing the additional types of data and considering more precise estimates of penetrance (involving correction for age of onset variation and for ascertainment biases) will be presented elsewhere. Only the simplified dominant model, as can be used for ankylosing spondylitis, is presented here. This approximate analysis is possible whenever both the marker allele and the *Is* allele have a low frequency in the population. The general method is based on the Bayesian probability.

$$\Pr(Is) = \frac{\Pr(Is | M) \cdot \Pr(M)}{\Pr(M | Is)}, \quad (1)$$

i.e., that the frequency or probability of the *Is* genotype is equal to the frequency of the *Is* genotype among individuals with the marker phenotype multiplied by the frequency of the marker phenotype, and the product divided by the frequency of the marker among affected individuals. This assumes that illness occurs among some fraction of those carrying the *Is* genotype but is otherwise independent of the presence of the marker.

The frequency of the marker, $\Pr(M)$, is estimated from population surveys independent of studies of disease association. The frequency of the marker among affected persons, $\Pr(M | Ill) = \Pr(M | Is) = f(M | Ill)$, is the estimate used to

show association and used in the initial tests of Thomson & Bodmer (their “FAD”). The third quantity in Equation 1, $\Pr(Is | M)$, can be estimated from

$$\Pr(Is | M) = \frac{\Pr(Illness | M)}{\Pr(Illness | Is)}, \quad (2)$$

i.e., from the frequency of illness among those with the marker, $f(Ill | M)$ divided by the penetrance. Thus, the frequency of the *Is* genotype is a function of estimable values.

For the dominant model with the additional simplifying assumption that the phenotype frequency is twice the allele frequency, for both the marker and *Is* allele, Equation 1 roughly estimates the *Is* allele frequency by using the marker allele frequency for $\Pr(M)$. The four haplotype frequencies are then estimated as follows:

$$\begin{aligned} x_1 &= \text{freq. } (M \text{ } Is \text{ haplotype}) = \Pr(Is | M) \cdot \Pr(M) \\ x_2 &= \text{freq. } (M \text{ } is \text{ haplotype}) = \Pr(M) - x_1 \\ x_3 &= \text{freq. } (m \text{ } Is \text{ haplotype}) = \Pr(Is) - x_1 \\ x_4 &= \text{freq. } (m \text{ } is \text{ haplotype}) = 1 - x_1 - x_2 - x_3 \end{aligned}$$

This simplified form of the analysis is used by Kidd *et al.* (4) on data in ankylosing spondylitis. The error, when compared with a more exact analysis using the same data, is not detectable to three decimal places.

Acknowledgement

The work was supported in part by grant NS 11786 from the N.I.H., U.S. Public Health Service.

References

1. Thomson, G. & Bodmer, W. F. (1976) The genetics of HLA and disease association. In: *Measuring Selection in Natural Populations*, eds. Barndorff-Nielsen, O., Christiansen, F. B. & Fenchel, T. Munksgaard, Copenhagen, (in press).
2. Day, N. E. & Simons, M. J. (1976) Disease susceptibility genes their identification by multiple case family studies. *Tissue Antigens* **8**, 109-119.
3. Green, J. R. & Woodrow, J. C. (1976) Sibling method for detecting HLA linked genes in disease. (Submitted for publication.)
4. Kidd, K. K., Bernoco, D., Carbonara, A.O., Daneo, V, Steiger, U. & Ceppellini, R. (1977) Genetic analysis of HLA associated diseases: the "illness susceptible" gene frequency and sex ratio in ankylosing spondylitis. In: *HLA and Disease*, eds. Dausset, J. & Svejgaard, A. Munksgaard, Copenhagen, p. 72.