ARTICLE

Combined segregation and linkage analysis of 59 Hodgkin's disease families indicates the role of HLA determinants

Yin Yao Shugart[1,2] and Andrew Collins[3]

[1]Center for Inherited Disease Research and [2]Department of Pediatrics, Johns Hopkins University, Baltimore, MD, USA; [3]Human Genetics, Southampton University, Southampton, UK

Through a literature search, we identified a total of 59 nuclear families with Hodgkin's disease (HD). The affected individuals in these families were HLA haplotyped. To evaluate the role of HLA-linked determinants of HD and investigate the mode of inheritance of familial HD, a combined segregation and linkage analysis was performed, together with a semi-parametric analysis of sib pairs. Amongst single locus models, a general single locus model showed no improvement over an additive model. However, a two locus additive model gave the best overall fit and a lod score of 3.55. We conclude that HD is most likely to be determined by both an HLA-associated major gene and other non-HLA genetic factors together with environmental effects. *European Journal of Human Genetics* (2000) **8**, 460–463.

Introduction

Hodgkin's disease (HD) is a malignancy of the immune system. Diagnosis of HD is based on the presence of the giant polynuclear Reed-Sternberg cell (RSC). Information from the surveillance, epidemiology, and end results (SEER) program of the NCI showed that the age adjusted incidence rate (1986–90) is 2.8 per 100000 person-years.[1] The mortality rates are 0.7/10000 in men and 0.4/10000 in women. Two incidence peaks occur between the ages of 15 and 34 and after age 45. Even though *EBV* gene expression has been detected in Reed-Sternberg cells and the latency membrane protein-1 (LMP1) is expressed in EBV-associated HD, a genetic etiology has been suggested. We will focus on evidence for familial predisposition to HD based upon previous investigations. Razis *et al*[2] reported that first-degree relatives of HD patients have a three-fold increased risk of carrying the disease. Grufferman *et al*[3] found a seven-fold increased risk among the siblings of a proband, with siblings of the same sex being at higher risk than siblings of discordant sexes. The occurrence of multiple HD cases within families has been reported.[4] More recently, Mack *et al*[5] provided a more solid basis for a genetic component which underlines HD through a twin study. They found that 10 out of 179 monozygotic twins became concordant for HD, whilst none of the 187 dizygotic twins became concordant for the disease.

The HLA system has been a usual suspect as a causative factor in HD since the 1960s. Hors and Dausset[6] reviewed a number of studies concerning the HLA (class I) haplotype sharing among the affected relatives, and demonstrated the distorted segregation of HLA haplotypes among affected siblings ($P < 0.005$) and among first cousin pairs ($P < 0.05$). Based on 19 multiplex families, they performed a lod score analysis, and reported a maximum lod score of 2.18 with a 10% penetrance under a recessive model, and of 2.11 with a 60% penetrance under a dominant model. Their finding suggested a weak genetic factor for familial HD. Risch[7] re-examined the data reported by Hors and Dausset using a non-parametric approach based on an estimation of IBD sharing among sibs. From that limited data set, Risch concluded that HLA alone can account for the familial aggregation for HD.

It is of interest to study the mode of inheritance and re-examine the relationship between HLA haplotypes and HD cases which appear to have a familial pattern. Since the

Correspondence: YY Shugart, Center for Inherited Disease Research, Johns Hopkins University, 333 Cassell Drive/Suite 2000, Baltimore, MD 21224, USA. Tel: +1 410 550 7118; Fax: +1 410 550 3559; E-mail: yyao@cidr.jhmi.edu

mode of inheritance of HD is not clearly defined, we performed a joint segregation and linkage analysis. We also compare this parametric approach as a method for detecting genes underlying complex traits to a semi-parametric approach as a method for detecting genes underlying complex traits to a semi-parametric alternative implemented in the BETA program.[8] The results are compared with the studies conducted by Hors and Dausset using a lod score method but without estimating any segregation parameters.

Subjects and methods
Subjects
Through a literature search, we identified a total of 59 nuclear families from 33 pedigrees affected with Hodgkin's disease (HD) who were HLA haplotyped. The data sources are as listed in Table 1. The number of relative pairs is also specified for each study.

Combined segregation and linkage (parametric) analysis
The data were analysed using COMDS[13] for combined segregation and linkage analysis, which implements a two-gene model that includes a linked major locus and an unlinked modifier. COMDS requires pedigree data to be decomposed into nuclear families with pointers, who are relatives through whom the nuclear family was ascertained. Liability classes are used to correct for covariates such as age and sex, which are assumed to be independent of a genetic aetiology. The following parameters describe the major locus: d, dominance, t, displacement between the two homozygote means on the underlying liability scale, q, gene frequency and recombination fraction θ between a marker locus and the major gene. Additional phenotypic information in the form of severity classes among affecteds, and diathesis classes among normals, can be incorporated. Dominance ranges from 0 for a recessive gene to 1 for a dominant gene. The second or modifier locus has the parameters qm, dm, and tm

which are analogous to d, t and q associated with the major locus. This modifier locus does not allow for a linked marker. For the current analysis, single and two-locus models were fitted using COMDS.

Most nuclear families had identified probands but a small number (10), obtained from larger pedigrees, were assigned single probands based on the assumption that pedigrees are highly selected. As age at observation was not available for many individuals liability classes were constructed based on sex with liability class frequencies of 0.000033 for males and 0.000025 for females corresponding to population risks.[1] However, since age at onset data was available three severity classes were formed (Table 2). Severity classes should ideally be corrected for situational variables but given the incomplete data on age at observation this was not possible. This, together with the inability to form age-specific liability classes inevitably influences the segregation model parameters but the linkage test is considered to lose little power under an inappropriate segregation model[14] and the parametric model provides a baseline for future studies.

Classes were ordered by increasing severity suggested by earlier onset. Severity class population frequencies were estimated from the sample (Table 2). An ascertainment frequency Π of 0.01 was assumed but results were found to be insensitive to varying Π (data not shown).

Semi-parametric linkage analysis
The program BETA (http://cedar.genetics.soton.ac.uk/public — html/) obtains allele-sharing proportions in a pair of sibs as a function of a single parameter (β). The model assumes that, in the case of an affected sib pair, the probability of sharing k alleles identical by decent (where k takes values 0, 1, 2) is proportional to $e^{\beta k}$. The allele-sharing properties for 0, 1, 2 alleles identical by descent (Z_0, Z_1 and Z_2) are:

$$Z_0 = 1/(1 + e^{\beta})^2$$

$$Z_1 = 2e^{\beta}/(1 + e^{\beta})^2$$

$$Z_2 = e^{2\beta}/(1 + e^{\beta})^2$$

This method is in contrast to the 'possible triangle' approach of Holmans which confines estimates to biological possibility but has two independent (constrained) parameters.[15]

Table 1 Data source

Authors	Year of publication	No of pedigrees	Relative pairs
Conte et al[9]	1983	4	2 affected sib pairs, 2 affected parent–child pairs.
Hafez et al[10]	1985	3	3 affected sub pairs with typed parents and unaffected sibs.
Lynch et al[11]	1976	1	1 affected sib pair; 2 second cousin pairs.
Hors and Dausset[6]	1983	19	29 sib pairs with typed parents and unaffected sibs; 2 parent–child pairs; 4 uncle–niece pairs; 4 first cousin pairs.
Kalidi et al[12]	1989	4	6 sib pairs with typed parents; 2 uncle–nephew pairs; 1 uncle–niece pair.

Table 2 Severity classes used by COMDS

Age at onset	N	SV frequency[a]	SV class[b]
30+	28	0.3146	1
15–30	45	0.5056	2
1–15	16	0.1798	3

[a]Assumed population frequency in severity (SV) class; [b]Severity class ordered by increasing severity (= earlier onset).

Results

Table 3 presents the results from the COMDS analysis. Among single locus models the general single locus (model 9) shows no improvement over the additive model (model 8). Comparing the additive model excluding severity (model 6) with the equivalent recessive model (model 3) there is a marginal improvement by the AIC criteria but a dramatic improvement over the dominant model (266 versus 356). Amongst single locus models the additive model is therefore the most favoured. On the linkage test (model 8 versus model 7) the χ_1^2 is $259.5 - 248.7 = 10.6$ corresponding to a lod of 2.3 at a recombination fraction of 0.194. However, two-locus models give a markedly improved fit. Comparing a two-locus additive model (model 15) with single-locus additive (model 8) the χ_2^2 is $248.7 - 186.3 = 62.4$ ($P < 0.001$). Compared to a general two-locus model (model 11) the two-locus additive is the more parsimonious by AIC (AIC = 198.3 versus 201.6). In the two-locus version the recombination fraction is markedly lower ($\theta = 0.072$) and the evidence for linkage much increased (model 15 vs 14; $202.7 - 186.3 = 16.348$) corresponding to a lod score of 3.55. It is evident that the severity information (as age of onset) contributes significantly to the overall model. This is true whether θ is 0.5 (model 14 vs model 12) $\chi_1^2 = 8.3$ or estimated (model 15 vs model 13) $\chi_1^2 = 7.9$. The magnitude of the severity parameter is variable and it is not clear that there is sufficient information to determine the relationship of an HLA linked susceptibility gene to age at onset.

The BETA program gave a lod score of 3.098 at $\beta = 0.918$ from 41 affected sib pairs derived from 32 families, where the magnitude of β represents the effect of the disease gene on the phenotype. The relative recurrence risks to sibs is $(1 + e^{\beta})^2/4$, and is therefore 3.1.

Discussion

Mode of inheritance

Harris et al[16] proposed a recessive mode of inheritance through a segregation analysis based on 11 multiplex families. The best-fitting one-locus model using COMDS is an additive model, where in males the lifetime penetrance are 1.0, 0.006 and 0 for genotypes aa, Aa, and AA, where 'a' represents the susceptibility allele, and in females, 0.93, 0.0 and 0.0. Evidently, our analysis also indicates a 'recessive-like' model. However, the best overall model is a two-locus additive model with an estimated recombination fraction with the HLA region of 0.072.

Power to detect linkage

Hors and Dausset[6] performed a linkage analysis[17,18] and reported a maximum lod score of 2.119 and 2.181 with a 60% and 10% penetrance respectively in dominant and recessive models. Our data set contains more pedigrees and the analysis incorporates ascertainment schemes through probands, estimates genetic parameters such as disease allele frequencies, penetrances and recombination fraction, and allows for an additive two-locus model. The new results suggest that there is consistent evidence for a genetic factor responsible for HD, but the genetic background of HD appears to be heterogeneous.

Proportion of HD contributed by HLA determinants

Risch[7] stated that HLA may account for all familial HD. Our finding contradicts Risch's conclusion that HD is most likely due to HLA-related genetic susceptibility but not to other factors such as viral infections. His analysis was based on the data reviewed by Hors and Dausset which apparently underrepresented the cases who had zero IBD sharing. Our current results support the hypothesis that, in addition to genetic factors, other factors such as non-HLA determinants and/or common viral infections may also play a role in HD aetiology. The recurrence risk to sibs suggested by the β model is 3.1, which is consistent with the finding of Razis et al[2] (3-fold increase) but is less than that reported by Grufferman[3] (7-fold increase). This again suggests the existence of additional factors other than HLA determinants.

It is worth noting that several biases are contained in this study:

Table 3 Single and two-locus models under COMDS (conditional likelihood) (= 0.01)

Model	d	t	q	s	th	dm	tm	qm	−2LnL	AIC
1	(1)	38.5	0.0002	(0)	(0.5)				352.3	356.3
2	(1)	5.3	0.0002	(0)	0.077				339.6	345.6
3	(0)	9.5	0.0058	(0)	(0.5)				262.3	266.3
4	(0)	9.9	0.0057	0.100	(0.5)				262.3	268.3
5	(0)	9.7	0.0057	0.099	0.220				254.3	262.3
6	(0.5)	17.9	0.0056	(0)	(0.5)				262.1	266.1
7	(0.5)	13.3	0.0053	0.837	(0.5)				259.5	265.5
8	(0.5)	19.6	0.0052	0.383	0.194				248.7	256.7
9	0.23	16.2	0.0052	1.0	0.194				248.7	258.7
10	0.77	16.3	0.0028	0.358	(0.5)	0.21	20.3	0.0019	202.0	216.0
11	0.77	16.1	0.0029	0.358	0.074	0.22	19.7	0.0019	185.6	201.6
12	(0.5)	21.7	0.0030	(0)	(0.5)	(0.5)	16.9	0.0020	211.0	219.0
13	(0.5)	14.9	0.0022	(0)	0.000	(0.5)	15.9	0.0025	194.2	204.2
14	(0.5)	14.0	0.0027	0.231	(0.5)	(0.5)	17.1	0.0020	202.7	212.7
15	(0.5)	13.8	0.0028	0.239	0.072	(0.5)	16.7	0.0020	186.3	198.3

(1) Researchers tend to report positive findings in HLA studies. The estimated proportion of familial HD cases may be biased in the upward direction;

(2) The data were pooled from several different sources and the techniques used in each study may have different levels of sensitivity of HLA haplotyping.

There is also a concern about ascertainment bias, even though we applied corrections based on the proband information. Ten out of 59 nuclear pedigrees did not contain information on probands. For such pedigrees, we made the assumption that the pedigrees are highly selected. Thus, we are not confident that the correction was entirely adequate. However, confirmation by a semi-parametric method suggests that the genetic model used in our analysis was robust. The robustness of combined segregation and linkage analysis was also suggested by the IDDM study by Lio and Morton.[19] They analysed 95 sib pairs affected with insulin dependent diabetes[20] and noted that COMDS is at least as powerful as semi-parametric alternatives such as BETA and MAPMAKER/SIBS. However, the current version of COMDS is only suitable to data in the format of nuclear families, and is limited to single-point linkage. A modified version which will accommodate more sophisticated pedigree structures should be considered.

Future goals
This study confirms an HLA-based genetic component for familial HD and indicates that HD is a heterogeneous disorder. The data source was various publications on HLA haplotyping in HD families. Today, we should be able to examine linkage between HLA determinants and HD using a microsatellite DNA technique given the success of this approach in mapping various genetic disorders. An ideal design would include a gene–environment component through a combination of DNA markers in both HLA and non HLA regions and EBV examination on the same set of familial HD cases. Statistical tools will be used to further assess the proportion of HLA and non-HLA determinants. A large, long term, international collaboration should be established for collecting informative pedigrees, genotyping, EBV typing, and development of suitable statistical methods for data analysis.

References
1 Mueller NE: Hodgkin's disease. In: Scottenfeld D, Fraumeni JF Jr. *Cancer Epidemiology and Prevention*. 2nd edn. Oxford University Press: New York, 1996.
2 Razis DV, Dimond HD, Craver LF: Familial Hodgkin's disease: its significance and implications. *Ann Intern Med* 1959; **51**: 933–971.
3 Grufferman S *et al*: Hodgkin's disease in siblings. *N Engl J Med* 1977; **296**: 248–253.
4 Robertson SJ, Lowman JT, Grufferman S *et al*: Familial Hodgkin's disease, a clinical and laboratory investigation. *Cancer* 1987; **59**: 1314–1319.
5 Mack *et al*: Concordance for Hodgkin's disease in identical twins suggesting genetic susceptibility to the young-adult form of the disease. *N Engl J Med* 1995; **16**: 332; 413–418.
6 Hors J, Dausset J: HLA and susceptibiity to Hodgkin's disease. *Immunol Rev* 1983; **70**: 167–192.
7 Risch N: Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet* 1987; **40**: 1–14
8 Morton NE: Logarithm of odds (lods) for linkage in complex inheritance. *Proc Natl Acad Sci USA* 1996; **93**: 3471–3476.
9 Conte R, Laria F, Zucchelli: HLA in familial Hodgkin's disease. *J Immunogenet* 1983; **10**: 251–255.
10 Hafez M, El-Tahau H, El-Moris Z *et al*: Genetic susceptibility in Hodgkin's lymphoma. In: Muller JH, Weber W (eds). *Familial Cancer*. Karger: Basel, 1985; pp 175–179.
11 Lynch HT, Saldivar VA, Guirgis HA *et al*: Familial Hodgkin's disease and associated cancer, a clinical-pathologic study. *Cancer* 1976; **38**: 2033–2041.
12 Kalidi I, Masset M, Gony J *et al*: MHC related genetic susceptibility to Hodgkin's disease. *Hématologie* 1989; **31**: 149–152.
13 Morton NE, Shields DC, Collins A: Genetic epidemiology of complex phenotypes. *Ann Hum Genet* 1991; **55**: 301–314.
14 Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J: Effects of mis-specifying genetic parameters in lod score analysis. *Biometrics* 1986; **42**: 393–399.
15 Holmans P: Asymptotic properties of affected sib-pair linkage analysis. *Am J Hum Genet* 1993; **52**: 362–374.
16 Harris EI, Bale SJ, Tucker MA, Greene MH: Complex segregation analysis of multiple case Hodgkin's disease families. *Am J Hum Genet* 1985; **37**: A197.
17 Morton NE: Sequential tests for the detection of linkage. *Am J Hum Genet* 1955; **7**: 277–318.
18 Ott J: *Analysis of Human Genetic Linkage* (rev. edn) Johns Hopkins University Press: Baltimore and London, 1991.
19 Lio P, Morton NE: Comparison of parametric and nonparametric methods to map oligogenes by linkage. *Proc Natl Acad Sci USA* 1997; **94**: 5344–5348.
20 Davies JL, Kawaguchi Y, Bennet JT *et al*: A genome-wide search for human type I diabetes susceptibility genes. *Nature* 1994; **371**: 130–135.