

A genome screen of families at high risk for Hodgkin lymphoma: evidence for a susceptibility gene on chromosome 4

L R Goldin, M L McMaster, M Ter-Minassian, S Saddlemire, B Harmsen, G Lalonde, M A Tucker

J Med Genet 2005;42:595–601. doi: 10.1136/jmg.2004.027433

Hodgkin's disease was recently designated Hodgkin lymphoma (HL) in the World Health Organization Classification.¹ The National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) population based registries estimate that 7900 new cases are diagnosed annually in the USA.² Clues to its aetiology have been suggested by the bimodal age distribution; higher risks in males, in people with higher socioeconomic status, and in smaller families; and occurrence of Epstein-Barr virus in HL tumour cells.³ The importance of genetic factors is indicated by reports of multiply affected families from case series,^{4–6} a twin study,⁷ a case-control study,⁸ and population registry studies carried out in Utah,⁹ Denmark,¹⁰ Israel,¹¹ and Sweden.^{12–14} We recently analysed data from registries in Sweden and Denmark and found significant familial aggregation of HL and other lymphoproliferative tumours.¹⁵ The relative risk for HL among first degree relatives of cases compared with controls was 3.1. Relative risks were higher in males compared with females, and in siblings of cases compared with parents and offspring. Relatives of earlier onset cases were at higher risk for HL and for all lymphoproliferative tumours and were also at higher risk for developing early onset tumours themselves. These findings are consistent with those seen from earlier case series studies but have the advantage of being from large, population based samples.

It is not known whether or how extrinsic risk factors interact with genetic susceptibility. Identifying inherited susceptibility genes is an important step towards defining the pathway(s) leading to development of HL and understanding its complex aetiology. There have been many studies of somatic mutations in HL tumour cells, but although there are associations with HLA types, specific germline genes causing susceptibility have not yet been identified. Early studies of HLA Class I alleles in familial HL showed increased haplotype sharing among affected sibling pairs.^{16–17} We have previously conducted studies of HLA Class II loci in 16 high risk HL families and found that alleles reported to be associated in case-control studies (such as DRB1*1501 and DQB1*0602) were also associated with familial HL using a family based analytical approach.¹⁸ There have been no comprehensive searches of the genome for HL genes, largely due to the difficulty in assembling informative samples. Even though this tumour is strongly familial, the proportion of cases with a family history is small, and affected families typically have very few cases.

We studied 44 informative high risk HL families and applied a whole genome search using densely spaced microsatellite markers in order to localise susceptibility genes.

METHODS

Ascertainment of HL pedigrees

The Genetic Epidemiology Branch (Division of Cancer Epidemiology and Genetics) has been recruiting families

Key points

- Hodgkin lymphoma (HL) has a strong familial component but no genes have yet been identified.
- We performed a genomewide linkage screen in 44 high risk HL families with a total of 254 individuals with DNA samples. Among these families, there were 95 HL cases and four cases of non-Hodgkin lymphoma (NHL) who were informative for linkage. The cases were characterised by a young age at diagnosis and an even gender ratio. In two-thirds of the families, the cases were siblings or cousins.
- We genotyped 1058 microsatellite markers with an average spacing of 3.5 cM, analysed the data using both non-parametric and parametric linkage analysis, and computed both two point and multi-point linkage statistics.
- The strongest linkage finding was on chromosome 4p near the marker D4S394. The lod score calculated by Genehunter Plus was 2.6 (nominal $p=0.0002$) when both HL and NHL individuals were considered affected. The mean identity by descent sharing among 35 affected sibling pairs was 72% in this region (nominal $p=0.00007$).
- The results are consistent with recessive inheritance. Other locations suggestive of linkage were found on chromosomes 2 and 11. The number of independent regions identified is more than expected by chance, although no one region met genomewide significance levels.
- These linkage findings represent the first step towards identifying one or more loci leading to susceptibility to HL and understanding its complex aetiology.

with two or more living cases of HL since 1970. This study was approved by an institutional review board, and informed consent was obtained on all subjects in this report. At the NIH clinical centre or on field trips, we evaluated all available affected individuals and first degree relatives of those affected, and obtained biospecimens. We also obtained original pathology material and reports for all HL and NHL cases where possible, and these were reviewed by the National Cancer Institute Laboratory of Pathology. Of the families investigated, 44 were judged to be informative for linkage studies, based on the number of available DNA samples (total 254) from affected and unaffected individuals. Sixteen of these families had been included in an earlier study of linkage and association with the HLA region.¹⁸

Table 1 shows the distribution of the number of affected individuals per family and the relationships among the HL cases. A total of 106 individuals in the families have been diagnosed with HL. Of these, DNA samples were available for 89, and genotypes for an additional six cases could be inferred from other family members. Of eight cases of NHL, four had either DNA samples or inferable genotypes. The level of diagnostic certainty for the HL cases was high, with 85% confirmed by either an outside pathology report, a slide reviewed at the NCI Laboratory of Pathology, or both. All cases were considered "affected" for linkage analysis. The mean age at diagnosis of HL was 26.8 years, which is much lower than that in the population, where the median age at diagnosis is 37 years.² Over 90% of our cases would be considered as having onset in childhood or young adulthood (earlier than 45 years of age). There was no difference in age at diagnosis among families with only two HL cases compared with those with more than two cases. Among those cases who could be classified into subtypes (75% of the total), there was a predominance (80%) of the nodular sclerosis (NS) subtype, with nearly all of the remaining having the mixed cellularity subtype, consistent with the young age distribution. There was a slight female predominance. As can be seen in table 1, two thirds of the families had cases among siblings and/or cousins; the remainder showed parent-offspring configurations with or without siblings or other relatives.

Genotyping

DNA was extracted from cryopreserved lymphocytes using standard methods. Genotyping was conducted under contract to deCODE Genetics using their screening set of 1058 microsatellite markers containing markers from the ABI linkage marker (version 2) screening and intercalating sets, and 500 custom made markers with known allele size distributions. Marker positions were obtained from the deCODE genetic map.¹⁹ PCR reactions were set up in multiplex reactions with fluorescently labelled primer pairs selected to amplify highly informative two, three, and four microsatellite loci. Following PCR amplification, DNA samples were loaded into ABI 3730 capillary sequencers. In each 96 well DNA plate, 93 DNA samples and three CEPH controls (family 1347-2) were run. Alleles were automatically classified using deCODE Allele Caller software,²⁰ which provides consistently >99.7% accuracy of genotyping calls compared with manual procedures. The samples were barcoded and tracked at each step, and profiled with sufficient markers for unique identification. Sample identities were checked for accuracy based on the pedigree structures in order to identify sample duplications and exchanges. The extensive use of robotics and automation at all steps in the process provided a high degree of reliability,

and reduced sample handling errors. A multistage data analysis approach was used to minimise errors in genotyping. After initial genotype identification was made, analyses were conducted to detect non-Mendelian transmission of genotypes from parents to offspring. These errors were then checked by re-analysing the results from the ABI sequencer or by re-typing the samples.

Power analysis

Because the true genetic model for HL is not known, we computed power assuming a rare gene with either dominant or recessive inheritance and heterogeneity. We estimated the power to detect linkage using the program SLINK.^{21, 22} We conducted simulations assuming both dominant and recessive inheritance models and penetrances of both 50% and 80% for the at risk genotypes and 0.1% for the normal genotype (a total of four models). Allele frequencies were set at values that kept the lifetime risk constant at 0.24 as estimated from SEER data.³ We assumed close linkage ($\theta = 0.001$) of the disease locus to a marker locus with eight alleles, which is reasonable given the dense spacing and high information content of the real genotypes. We generated 200 replicates under each model to compute the average lod scores and power of detecting lod scores of 1–3.

Linkage analysis

The genotype data were stored in a Microsoft Access database. Formatting changes needed for specific programs were made using MEGA2.²³ The genotype data were first checked for Mendelian consistency using the program PEDCHECK.²⁴ The RECODE program was used to prepare the data files for analysis and to estimate allele frequencies from all founders in the pedigrees. We checked for the presence of additional genotype errors using the mistyping option of Simwalk2 (version 2.89),^{25–27} and eliminated genotypes that had probabilities of ≥ 0.25 of being errors. In total, only a very small number of genotypes (<0.5%) was eliminated because of either Mendelian inconsistencies or high mistyping probability.

We first screened the 1058 markers using two point analyses with the MLINK program from the FASTLINK package.^{28–31} We calculated lod scores assuming the same models as described above for the power simulations. In addition, we assumed that penetrance increased with age, using age incidence rates in the population to construct liability classes. Multipoint analyses were conducted using Genehunter³² to compute both parametric lod scores, assuming heterogeneity (Hlod, with $\alpha =$ proportion of linked families) and non-parametric linkage (NPL) scores (z scores). Genehunter Plus lod scores were also calculated, because this method has been shown to give less conservative estimates of p values than does the original Genehunter method.³³ As there were only a few NHL cases in this familial cohort, only individuals with HL were classified as affected for initial linkage analyses across the genome, and all other individuals were considered unaffected. Regions of the genome with nominal p values ≤ 0.01 by any analysis method were followed up with additional analyses, including broadening the affection status to include NHL and calculation of mean IBD sharing in affected sibling pairs (ASPs) using the program Sibpal in SAGE (version 4.5).³⁴

RESULTS

Power of linkage detection

The simulations showed that if $\geq 75\%$ of the families were linked to the same locus, then the probability of obtaining a lod score of 3.0 or more was at least 75%. The power to find a recessive gene was always higher than for a dominant gene; this is not surprising, as two thirds of the families have

Table 1 Description of HL families

	No. of families
Family types	
2 HL cases	29
3 HL cases	12
4 HL cases	3
HL families with at least 1 NHL* case	8
Relationships among HL cases	
Siblings	22
Siblings + parent	5
Siblings + cousin or avuncular	6
Parent-offspring	8
Parent-offspring + cousin or avuncular	1
Cousins	2

*5/8 NHL cases diagnosed with diffuse large cell lymphoma.

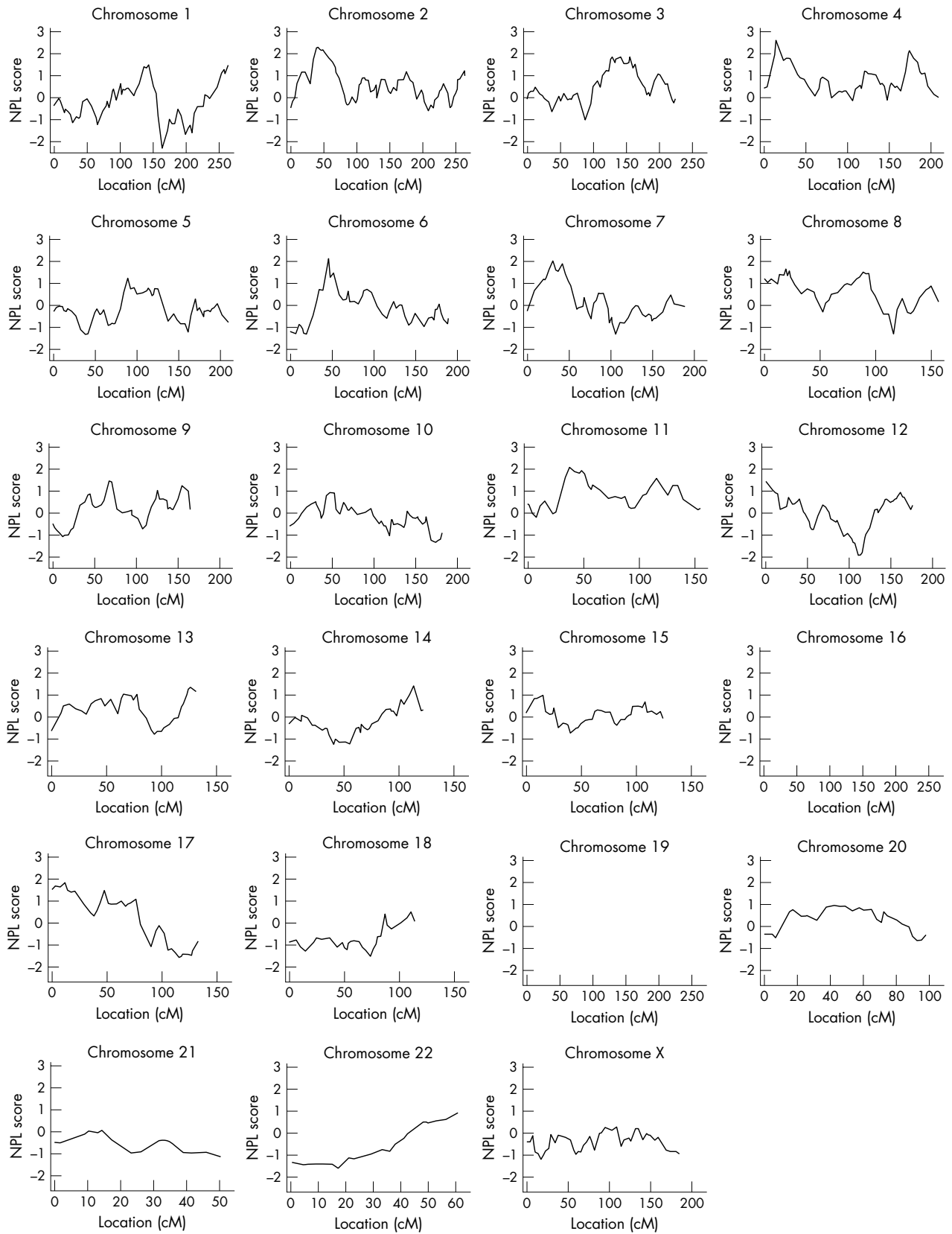


Figure 1 Multipoint non-parametric linkage (NPL) statistics calculated by Genehunter for each chromosome assuming only HL cases affected.

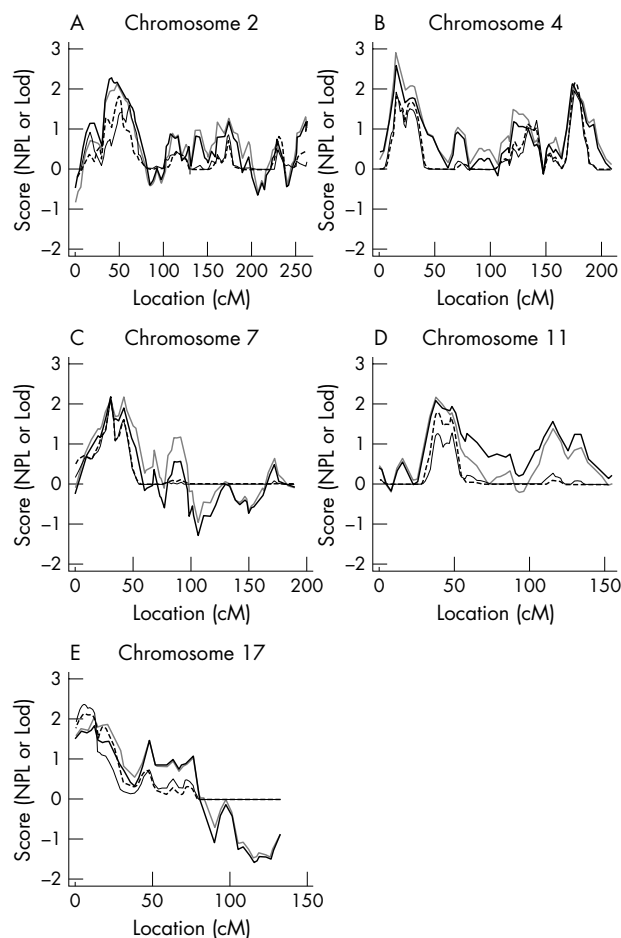


Figure 2 (A–E) Multipoint NPL and heterogeneity lod scores (Hlod) for chromosomes with suggestive results. Thick black line, NPL (HL affected); thin black line, Hlod (HL affected); thick grey line, NPL (HL or NHL affected); broken line, Hlod (HL or NHL affected). Model assumed for Hlod score was recessive with maximum penetrance of 50% for chromosomes 2, 4, 7, and 11, and dominant with maximum penetrance of 50% for chromosome 17.

affected sibling or cousin configurations and are thus consistent with recessive inheritance. If only half of the families were linked to a single locus, then the power to detect linkage was modest (expected lod score = 2.3 under recessive inheritance, 1.9 under dominant inheritance). If only 25% of the families were linked, we would have minimal power to detect a susceptibility gene.

Linkage results

Two point lod scores revealed several regions of the genome with evidence for linkage to HL. The strongest findings were in regions on chromosomes 2, 3, and 4. Table 2 shows loci that had two point lod scores ≥ 2.0 under any one of the four

inheritance models tested. These regions, on chromosomes 2, 3, and 4, showed clusters of consecutive markers with positive scores. Another region on chromosome 4 and a region on chromosome 11 also showed clusters of consecutive loci with positive scores, with maximum scores between 1.5 and 2.0 (not shown). Multipoint NPL statistics of all of the chromosomes as calculated by Genehunter are shown in fig. 1. Parametric lod scores with and without heterogeneity were also calculated but are not shown. The densely spaced markers resulted in a high level of informativeness, with information content calculated by Genehunter averaging 0.75 throughout the genome. Fig 2 (A–E) shows regions on chromosomes 2, 4, 7, 11, and 17 where either multipoint NPL or Hlod scores had nominal p values < 0.01 . These figures show NPL and parametric statistics for both narrow (HL only) and broad (including NHL) affection status models. The parametric lod scores were calculated assuming heterogeneity using the inheritance model (dominant or recessive) that gave the highest two point lod scores. As seen in table 2, one marker on chromosome 3 had a two point lod score of 4.0 under dominant inheritance. The flanking markers had lods of 1.7 and 1.0 (not shown), but multipoint statistics in this region were substantially lower (highest NPL score was < 2.0 and highest Hlod score was 1.4) than the two point results; thus no additional graphs are shown. Tables 3 and 4 summarise the linkage statistics, locations, and marker names for these six regions.

The strongest evidence for linkage occurred on chromosome 4 where the peak NPL or Hlod was found at 14 cM, flanked by markers D4S2935 and D4S394 (fig 2B). Strong evidence for linkage was seen under both affection status models. Under the broader model (in which NHL cases counted as affected), the peak NPL score was 2.9 and peak Genehunter Plus lod score was 2.6 ($p = 0.0002$), which is strongly suggestive of linkage (table 3). The locations of the peak linkage scores were consistent among all analyses performed. The Hlod score was highest under the recessive model and the proportion of families linked was estimated at 43%. Table 4 shows that the mean IBD sharing among ASPs as calculated by Sibpal ($> 70\%$ and highly significant) gave results consistent with the other methods.

Fig 2 (A–E) and tables 3 and 4 show five other locations that had positive non-parametric linkage or Hlod scores. For each region, the positive findings from Genehunter were supported by increased mean IBD sharing among ASPs (table 4). The second most significant region was on chromosome 2 (Genehunter Plus lod score was 2.4, $p = 0.0004$), although the location varied from 41 to 62 cM (fig 2A). On chromosome 11, there was a peak at location 37 to 39 cM (fig 2D), depending on the model, which had a maximum Genehunter Plus lod of 2.2 (rounded from 2.18) and $p = 0.0007$. Regions on 4q, 7, and 17 also showed positive results. There was a second positive peak on chromosome 4 at location 173–176 cM (fig 2B), a peak on chromosome 7 (fig 2C), and one on chromosome 17, in which the peak was at the p telomeric region (fig 2E). There was also a modest signal (NPL score of ~ 2.0) on chromosome 6 at the marker

Table 2 Summary of two point lods ≥ 2.0

Chrom	Locus	Location (cM)	Recessive 50% penetrance lod (theta)	Dominant 50% penetrance lod (theta)
2	D2S2168	49.4	2.52 (0.10)	1.41 (0.10)
3	D3S1558	128.9	2.52 (0.10)	4.03 (0.01)
4	D4S2935	13.3	2.09 (0.05)	0.91 (0.10)
4	D4S394	15.3	2.39 (0.10)	0.15 (0.25)

Chrom, chromosome number.

Table 3 Linkage signals from multipoint analysis

Chrom.	Affected HL only			Affected HL or NHL		
	Statistic	Location (cM)	Markers	Statistic	Location (cM)	Markers
2						
NPL	2.20	40	D2S2346	2.11	47	D2S305–D2S2168
GHP lod	2.44	41	D2S2346	2.10	45	D2S305–D2S2168
Hlod (α)	1.72 (0.48)	61	D2S2230	1.81 (0.46)	50	D2S2168
4p						
NPL	2.59	14	D4S2935–D4S394	2.91	14	D4S2935–D4S394
GHP lod	2.51	14	D4S2935–D4S394	2.63	14	D4S2935–D4S394
Hlod (α)	2.15 (0.50)	14	D4S2935–D4S394	1.78 (0.45)	15	D4S2935–D4S394
4q						
NPL	2.10	174	D4S3030	2.05	174	D4S3030
GHP lod	1.37	174	D4S3030	1.27	174	D4S3030
Hlod (α)	2.20 (0.46)	176	D4S3030–D4S415	2.20 (0.43)	176	D4S3030–D4S415
7						
NPL	2.10	30	D7S2557–D7S2495	2.10	30	D7S2557–D7S2495
GHP lod	1.33	30	D7S2557–D7S2495	1.35	41	D7S2463–D7S516
Hlod (α)	2.43 (0.45)	30	D7S2557–D7S2495	2.27 (0.43)	30	D7S2557–D7S2495
11						
NPL	2.09	37	D11S928–D11S1359	2.18	37	D11S928–D11S1359
GHP lod	2.18	38	D11S1359	2.16	39	D11S1359
Hlod (α)	1.34 (0.43)	38	D11S1359	1.78 (0.43)	39	D11S1359
17						
NPL	1.83	12	D17S1876	2.03	12	D17S1876
GHP lod	0.90	6	D17S1529–D17S831	1.06	12	D17S1876
Hlod (α)	2.37 (0.64)	6	D17S1529–D17S831	2.14 (0.53)	5	D17S1529–D17S831

Chrom, chromosome number; GHP, Genehunter Plus.

Table 4 Maximum IBD sharing among affected sibling pairs (and p values) in six regions suggestive of linkage

Chrom.	Affected HL only (32 affected sibling pairs)				Affected HL or NHL (35 affected sibling pairs)		
	Marker (cM)	IBD	p	Marker (cM)	IBD	p	
2	D2S305 (43)	0.67	0.00029	D2S2168 (49)	0.69	0.00021	
4p	D4S394 (15)	0.73	0.00002	D4S394 (15)	0.72	0.000073	
4q	D4S3030 (175)	0.66	0.007	D4S3030 (175)	0.63	0.017	
7	D7S2495 (34)	0.66	0.007	D7S2495 (34)	0.64	0.0099	
11	D11S1359 (39)	0.67	0.004	D11S1359 (39)	0.68	0.0013	
17	D17S960 (21)	0.64	0.005	D17S960 (21)	0.66	0.001	

Chrom, chromosome number.

D6S1571, which is close to the HLA region (fig 1); however, no other markers on this chromosome gave positive results. This is consistent with a small HLA effect in HL and with previous data showing an association with HLA types within a subset of these families.¹⁸

DISCUSSION

Our unique sample of high risk HL families has allowed us to conduct a genomewide scan using a dense set of markers. The strongest finding in this study was on chromosome 4p. The inheritance is likely to be recessive, given that the highest Hlod scores were found assuming a recessive model and affected sibling pairs had a mean IBD >70%. The sharing in ASPs was highly significant although the sample size was not large. The likelihood of recessive inheritance is also supported by previous studies showing higher risks in siblings.^{4–15} The strength of linkage to this region increased slightly when the few individuals with NHL were considered affected. This is consistent with our population study showing that both HL and NHL were found more frequently in relatives of HL cases.¹⁵ In fact, the cases of NHL in our families had a much earlier age of onset (mean 49 years) than the population (median age of diagnosis is 67 years²). This location on chromosome 4p is a high priority for follow up with

additional families. The regions on chromosomes 2 and 11, with Genehunter Plus lod scores of 2.4 and 2.2, are also strong candidates for containing HL susceptibility genes. The three other findings on chromosomes 4q, 7, and 17 have lower significance levels but warrant further follow up. The promising finding on chromosome 3 based on two point lod scores did not hold up under multipoint analyses. Consistent with other data, we also have evidence from this study that the HLA region may play a role in familial risk (fig 1). It is possible that applying more complex modelling of gene effects (such as multilocus models) to the genomewide linkage data would lead to more definitive detection of a susceptibility gene or genes.

There is some disagreement about how to interpret significance levels when conducting a genomewide scan.³⁵ Wiltshire *et al*³⁶ suggested an approach for evaluating the significance of linkage findings. They pointed out that for complex diseases, several genes may be involved and therefore it is less likely that any single region will reach a high level of significance. They propose counting the number of independent regions of linkage detected and comparing this to chance expectations. They pointed out that the lod score thresholds for “significant” (lod = 3.6) or “suggestive” (lod = 2.2) linkage, as defined by Lander and Kruglyak,³⁷

are often too stringent, as the thresholds for any one study depend on sample size, marker density, and marker informativeness. Using simulations, Wiltshire *et al*³⁶ found that for a 5 cM scan with 100 ASP families, a Genehunter Plus lod score of 1.78 would predict one linkage finding in a genomewide scan by chance and a more stringent lod score threshold of 2.2 would predict 0.37 linkage findings by chance. Our three strongest findings on chromosomes 4p, 2p, and 11p meet the more stringent threshold of 2.2 (table 3). Thus, we have identified more regions than expected by chance, which strongly suggests that there are one or more true loci causing susceptibility to HL among the locations identified.

There are several regions of the genome where recurrent cytogenetic changes are found in HL cells,^{38,39} including amplifications of regions of 2p, 4p, 4q, and 9p, and deletions of chromosome 6q25.⁴⁰ The regions we identified by linkage do not appear to overlap with the cytogenetic regions. For example, duplication of the *c-REL-BCL11A* region on 2p was reported in HL cells, but these loci are about 30–50 cM centromeric from the peak we found. Similarly, amplifications on 4p may involve the fibroblast growth factor 3 gene,³⁹ which is about 10 Mb from the peak we identified. The linkage peak on the telomere of chromosome 17 is near the *p53* gene, but somatic mutations of this gene are not frequently found in HL cells,⁴¹ and one study found no germline mutations in familial HL cases from our sample.⁴² Locations determined from linkage analysis of complex diseases are imprecise owing to uncertainty about the underlying model, so it is possible that one or more of the regions identified by cytogenetic studies overlaps with regions we identified by linkage.

There are some limitations to this study. The highly selected families in our sample are not representative of HL in the population. Consistent with clinical descriptions of familial HL in the literature,⁴³ referrals to our group are mostly families with cases that have early onset and NS subtype. There are a few later onset cases in these families, but even these are found within families having early onset in other members. In terms of histological subtype, we also find mixed cellularity cases in the same families with NS subtype. Thus, it is not possible for us to analyse linkage to subgroups based on age at onset or histological characteristics of the tumour. In addition, most of our families have HL in siblings and/or cousins, which makes it difficult to detect a dominant susceptibility locus. Future studies applied to more families and a broader representation of clinical types will lead to more robust conclusions about the effects of susceptibility genes, genetic heterogeneity of HL, and the range of phenotypic expression of specific susceptibility genes.

The findings presented here are the first step in the discovery of germ line susceptibility gene(s) and delineation of the pathways involved in development of HL. Even though these susceptibility loci are being discovered in high risk families, they may also play a role in development of sporadic HL. Defining these pathways and determining their interactions with environmental factors may lead to more effective treatment and prevention, which could have a great impact on patients, many of who are young and lose years of life/productivity to disease or treatment related morbidity.

ACKNOWLEDGEMENTS

We acknowledge the contributions of A Goldstein, L Harty, A Lin, J Whitehouse, and G Shaw for family ascertainment, and of E Jaffe for pathology review. The genotyping was conducted under NCI contract N02-CP-01108 with a subcontract BRC-1108-35 to deCODE Genetics. Some of the results of this paper were obtained by using the program package SAGE, which is supported by a US Public Health Service

Resource Grant (RR03655) from the National Center for Research Resources.

Authors' affiliations

L R Goldin, M L McMaster, M Ter-Minassian, S Saddlemire, M A Tucker, Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

B Harmsen, G Lalonde, Westat Research Inc., Rockville, MD, USA

Competing interests: none declared

Correspondence to: Dr L R Goldin, Genetic Epidemiology Branch, DCEG, NCI, 6120 Executive Blvd, 7236 MSC, Bethesda, MD 20892-7236, USA; goldin@mail.nih.gov

Received 22 September 2004

Revised 4 November 2004

Accepted 4 November 2004

REFERENCES

- Jaffe ES, Harris NL, Stein H, Vardiman JW, eds. *World Health Organization classification of tumours. Pathology and genetics of tumours of haematopoietic and lymphoid tissues*. Lyon, France: IARC Press, 2001.
- Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, Mariotto A, Feuer EJ, Edwards BK, eds. *SEER cancer statistics review, 1975–2001*. Bethesda, MD: National Cancer Institute, 2004. Available at: http://seer.cancer.gov/csr/1975_2001/ Accessed September 21, 2004.
- Mueller NE. Hodgkin's Disease. In: Schottenfield D, Fraumeni JF Jr, eds. *Cancer epidemiology and prevention*, 2nd ed. New York: Oxford University Press, 1996:893–919.
- Grufferman S, Cole P, Smith PG, Lukes RJ. Hodgkin's disease in siblings. *New Engl J Med* 1977;**296**:248–50.
- Haim N, Cohen Y, Robinson E. Malignant lymphoma in first-degree blood relatives. *Cancer* 1982;**49**:2197–2200.
- Kerzin-Storarr L, Faed MJW, MacGillivray JB, Smith PG. Incidence of familial Hodgkin's disease. *Br J Cancer* 1983;**47**:707–12.
- Mack TM, Cozen W, Shigbata DK, Weiss LM, Nathwani BN, Hernandez AM, Taylor CR, Hamilton AS, Deapen DM, Rappaport EB. Concordance for Hodgkin's Disease in identical twins suggesting genetic susceptibility to the young-adult form of the disease. *N Eng J Med* 1995;**332**:413–18.
- Bernard SM, Cartwright RA, Darwin CM, Richards IDG, Roberts B, O'Brien C, Bird CC. Hodgkin's disease: Case control epidemiological study in Yorkshire. *Br J Cancer* 1987;**55**:85–90.
- Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst* 1994;**86**:1600–8.
- Westergaard T, Melbye M, Pedersen JB, Frisch M, Olsen JH, Andersen PK. Birth order, sibship size and risk of Hodgkin's disease in children and young adults: A population-based study of 31 million person-years. *Int J Cancer* 1997;**2**:977–81.
- Paltiel O, Schmit T, Adler B, Rachmilevitz EA, Polliak A, Cohen A, Haim N, Ben Shachar M, Epelbaum R, Barchana M, Cohen R, Ben Yehuda D. The incidence of lymphoma in first-degree relatives of patients with Hodgkin disease and non-hodgkin lymphoma. *Cancer* 2000;**88**:2357–66.
- Shugar YY, Hemminki K, Vaitinen P, Kingman A, Dong C. A genetic study of Hodgkin's lymphoma: an estimate of heritability and anticipation based on the familial cancer database in Sweden. *Hum Genet* 2000;**106**:553–6.
- Lindelof B, Eklund G. Analysis of hereditary component of cancer by use of a familial index by site. *Lancet* 2001;**358**:1695–8.
- Hemminki K, Li X, Czene. Familial risk of cancer: data for clinical counseling and cancer genetics. *Int J Cancer* 2004;**108**:109–14.
- Goldin LR, Pfeiffer RM, Gridley G, Gail MH, Li X, Møller-Jensen L, Olsen JH, Hemminki K, Linet MS. Familial aggregation of Hodgkin lymphoma and related tumors. *Cancer* 2004;**100**:1902–8.
- Greene MH, McKeen EA, Li FP, Blattner WA, Fraumeni JF Jr. HLA antigens in familial Hodgkin's disease. *Int J Cancer* 1979;**23**:777–80.
- Hors J, Dausset J. HLA and susceptibility to Hodgkin's disease. *Immunol Rev* 1983;**70**:167–92.
- Harty LC, Lin AY, Goldstein AM, Jaffe ES, Carrington M, Tucker MA, Modi WS. HLA-DR, HLA-DQ, and TAP genes in familial Hodgkin disease. *Blood* 2002;**99**:690–3.
- Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K. A high-resolution recombination map of the human genome. *Nat Genet* 2002;**31**:241–7.
- Fjalldal JG, Benediktsson K, Sigurdsson J, Ellingsen LM. Automated genotyping: combining neural networks and decision trees to perform robust allele calling. *Proc Int Joint Conf Neural Networks* 2001:A1–6.
- Weeks DE, Ott J, Lathrop GM. SLINK: a general simulation program for linkage analysis. *Am J Hum Genet* 1990;**47**:A204.
- Ott J. Computer-simulation methods in human linkage analysis. *Proc Natl Acad Sci USA* 1989;**86**:4175–8.

- 23 **Mukhopadhyay N**, Almasy L, Schroeder M, Mulvihill WP, Weeks DE. Mega2, a data-handling program for facilitating genetic linkage and association analyses. *Am J Hum Genet* 1999;**65**:A436.
- 24 **O'Connell JR**, Weeks DE. PedCheck: A program for identifying genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998;**63**:259–66.
- 25 **Sobel E**, Lange K. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker sharing statistics. *Am J Hum Genet* 1996;**58**:1323–37.
- 26 **Sobel E**, Sengul H, Weeks DE. Multipoint estimation of identity-by-descent probabilities at arbitrary positions among marker loci on general pedigrees. *Hum Hered* 2001;**52**:121–31.
- 27 **Sobel E**, Papp JC, Lange K. Detection and integration of genotyping errors in statistical genetics. *Am J Hum Genet* 2002;**70**:496–508.
- 28 **Lathrop GM**, Lalouel JM, Julier C, Ott J. Strategies for multilocus analysis in humans. *PNAS* 1984;**81**:3443–6.
- 29 **Lathrop GM**, Lalouel JM. Easy calculations of lod scores and genetic risks on small computers. *Am J Hum Genet* 1984;**36**:460–5.
- 30 **Cottingham Jr RW**, Idury RM, Schäffer AA. Faster sequential genetic linkage computations. *Am J Hum Genet* 1993:252–63.
- 31 **Schäffer AA**, Gupta SK, Shiriram K, Cottingham RW Jr. Avoiding recomputation in linkage analysis. *Hum Hered* 1994;**44**:225–37.
- 32 **Kruglyak L**, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996;**58**:1347–63.
- 33 **Kong A**, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997;**61**:1179–88.
- 34 **Statistical Solutions Ltd**. SAGE: statistical analysis for genetic epidemiology. Cork: Statistical Solutions Ltd, 2003.
- 35 **Witte J**, Elston RC, Schork NJ. Genetic dissection of complex traits. *Nature Genet* 1996;**12**:355–6.
- 36 **Wiltshire S**, Cardon LR, McCarthy MI. Evaluating the results of genomewide linkage scans of complex traits by locus counting. *Am J Hum Genet* 2002;**71**:1175–82.
- 37 **Lander E**, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995;**11**:241–7.
- 38 **Joos S**, Kupper M, Ohl S, von Bonin F, Mechttersheimer G, Bentz M, Marynen P, Moller P, Pfreundschuh M, Trumper L, Lichter P. Genomic imbalances including amplification of the tyrosine kinase gene *JAK2* in CD30+ Hodgkin cells. *Cancer Res* 2002;**60**:549–52.
- 39 **Joos S**, Granzow M, Holtgreve-Grez H, Siebert R, Harder L, Martin-Subero JI, Wolf J, Adamowicz M, Barth TF, Lichter P, Jauch A. Hodgkin's lymphoma cell lines are characterized by frequent aberrations on chromosomes 2p and 9p including *REL* and *JAK2*. *Int J Cancer* 2003;**103**:489–95.
- 40 **Re D**, Starostik P, Massoudi N, Staratschek-Jox A, Dries V, Thomas RK, Diehl V, Wolf J. Allelic losses on chromosome 6q25 in Hodgkin and Reed Sternberg cells. *Cancer Res* 2003;**63**:2606–9.
- 41 **Maggio EM**, Stekelenburg E, Van den Berg A, Poppema S. TP53 gene mutations in Hodgkin lymphoma are infrequent and not associated with absence of Epstein-Barr virus. *Int J Cancer* 2001;**94**:60–6.
- 42 **Weintraub M**, Lin AY, Franklin J, Tucker MA, Magrath IT, Bhatia KG. Absence of germline p53 mutations in familial lymphoma. *Oncogene* 1996;**12**:687–91.
- 43 **Ferraris AM**, Racchi O, Repezzi D, Gaetani GF, Boffetta P. Familial Hodgkin's disease: a disease of young adulthood? *Ann Hematol* 1997;**74**:131–4.