European Journal of Human Genetics (1998) 6, 80–88

ORIGINAL PAPER

Modifier genes in humans: strategies for identification

RS Houlston and IPM Tomlinson

Institute of Cancer Research, Sutton, Surrey, UK

A number of genetic disorders exhibit inter- and intra-familial variability. Understanding the factors that control the expression of disease genes should provide insight into the fundamental disease processes and will have implications for counselling patients. Different mechanisms can account for this variability, including environmental factors, genotype–phenotype correlations and imprinting. There is also evidence that, in a number of genetic diseases, gene expression is under the control of modifier loci. In cases where the biological basis of the genetic disease is understood, any genes involved in the pathogenic process represent candidate modifier genes which can easily be evaluated. Alternatively, modifiers can be identified through approaches such as mouse models. Since modifier genes will generally be common and because of confounding environmental influences, linkage analyses in humans will generally be based upon affected or discordant sib pairs. Discordant sib pairs represent an attractive option for linkage studies, because recurrence rates are high and the reduced survival characteristics associated with severe phenotypes will make the likelihood of obtaining clinical material from two living cases difficult. Furthermore, the use of discordant siblings will select for those siblings which possess sufficient dissimilarity at the modifier locus to overcome any shared environmental influence.

Keywords: modifier loci; candidates' genes; linkage; mouse models; sib pairs

Introduction

A large number of mendelian and non-mendelian genetic disorders display considerable inter- and intra-familial variability in phenotypic expression. There are a number of explanations to account for such variability (Table 1).[1–31] In some disorders, it is clear that environmental factors (such as the effect of smoking in individuals with familial hypercholesterolaemia[8] account for marked differences in clinical expression between individuals harbouring the same mutation. Other mendelian disorders show strong genotype–phenotype correlations. Mutations in specific regions of the gene or specific types of mutation (e.g. mis-sense versus truncating) may be characterised by certain phenotypic features; in the *BRCA2* gene, for example, mutations at the 3' end of the gene are characterised by an increased risk of ovarian cancer compared with mutations in the 5' region.[4] Alternatively, genotype–phenotype correlation can result from gene modification, such as in the case of trinucleotide repeat

Correspondence: RS Houlston, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK
Received 11 August 1997; revised 6 October 1997; accepted 15 October 1997

expansion causing myotonic dystrophy and Huntington's chorea.[12] For X-linked disorders, differences in X-inactivation between carriers are another specific mechanism for phenotypic variability.

In recent years it has also been recognised that the route of inheritance of a gene, that is the parental origin, can influence its phenotypic effects and tissue expression. Non-disjunction during meiotic segregation in translocation heterozygotes can lead to transmission from one parent only, resulting in uniparental disomy or monosomy. For certain loci, uniparental inheritance has a detrimental effect, the nature and severity of which depends on whether the inheritance is maternal or paternal. Two of the most convincing examples of such imprinting in human disease are the Angelman and Prader-Willi syndromes, which have different phenotypes but both involve deletions of 15q11-13.[10] In addition, distortion of segregation ratios, associated with diabetes,[16] neural tube defects[15] and atopic pedigrees[14] may also reflect a parent of origin effect.

It is, however, important to distinguish between inter- and intra-familial variability in ascribing possible mechanisms to account for phenotypic variability in disease gene carriers. Differences in environmental factors and different mutations can easily be seen to underlie a proportion of inter-familial manifestations. However, intra-familial variability, especially in siblings, cannot intuitively be so readily accounted for by these types of mechanisms. There is now increasing evidence that the manifestations of many genetic disorders are influenced by so-called 'modifying' genes distinct from the disease locus. For some conditions, there is indirect support for the notion of modifying genes from studies undertaken in mice. In general, mice with specific genetic backgrounds often express the same genetic disease very differently. Specific modifying loci have also been identified. For example, two or more modifiers of the phenotype of the Multiple intestinal neoplasia (Min) mouse model of FAP are known to exist.[32,33] One (*Mom-1*) maps to equivalent of human 1p35-p36 and a candidate for *Mom-1* is the secretory phospholipase *A2* gene.[32,33] In a number of human disorders, moreover, there is direct evidence for modifying loci (for example in the case of *BRCA1* and H-*ras*.[20] In other cases, the evidence is indirect: for example, only one in three individuals with insulin dependent diabetes mellitus (IDDM) develops diabetic nephropathy,[34] although the relative risk of this complication is increased two-fold in relatives of IDDM patients with nephropathy.[28] Thus, it is likely that a modifying locus is acting if there is intra-familial variability in phenotype in the absence of known or probable environmental factors, and/or there is increased familial risk of a specific phenotypic feature within families.

Table 1 Mechanisms for the modification of genetic diseases

Mechanism	*Example*
Genotype-phenotype correlations	i. *Gene expansion*
	Huntingdon's disease[1]
	Myotonic dystrophy[2]
	ii. *Mutation site*
	Adenomatous polyposis coli[3]
	BRCA2[4]
Skewed X inactivation	Acardi syndrome[5]
	Wiskott Aldrich[6]
	Lesch Nyhan[7]
Environmental factors	Familial hypercholesterolaemia[8]
	Phenylketonuria[9]
Imprinting	Prader-Willi syndrome[10]
	Angelman syndrome[10]
	Albright's osteodystrophy[11]
	Huntington's disease[12]
	Myotonic dystrophy[12]
	Neurofibromatosis type 1[12]
	Neurofibromatosis type 2[12]
	Retinoblastoma[13]
	Spinocerebellar ataxia type 1[12]
	Atopy[14]
	Neural tube defects[15]
	Diabetes[16]
	Adult polycystic disease[17]
Mosaicism	Hypohidrotic ectodermal dysplasia[18]
	Machado Joseph disease[19]
Modifying gene	Familial hypercholesterolaemia[20]
	BRCA1[21]
	Familial polyposis coli[22]
	Alpha thalassaemia[23,24]
	Neurofibromatosis type 1[25]
	Fragile X[26]
	Sickle cell anaemia[27]
	IDDM[28]
	Graft-versus-host disease[29–31]

What is a Modifier Locus?

For a Mendelian disease, a modifying locus can be defined as inherited genetic variation (distinct from the disease locus) that leads to a quantitative or qualitative difference in any aspect of the disease phenotype. The definition thus includes genes that lead to non-penetrance of the disease allele. For complex genetic diseases, the definition of a modifying locus is the same, although genes which contribute to the risk of disease are better classed as disease-predisposing loci, rather than as modifiers, even though they may have some additional effect on the phenotype. Thus, modifiers of complex disease do not have a qualitative effect on the

basic disease phenotype, although they have qualitative effects on sub-phenotypes, such as the development of nephropathy in IDDM.

Modifying genes can act in many different ways. The major possibilities are as follows:

(i) cell-autonomous, tissue specific or systemic action;

(ii) qualitative or quantitative effects, both at the level of the protein and the disease phenotype;

(iii) specific action on disease pathways, or non-specific action;

(iv) direct action, or requiring activation through somatic mutation or a specific environment.

Methods

Evaluation and Identification of Modifier Genes

Where there is an understanding of the pathological basis of the disease, other genes involved in the disease process provide candidates for modifier loci. The possible participation of such loci can easily be examined by determining their relative frequency according to disease severity or some other clinical variable. Familial hypercholesterolaemia, for example, is caused by a defective *LDL* receptor gene. It is likely that variation in genes for other elements of the lipoprotein pathway, such as Apo-a or Apo-E which are ligands for the receptor, leads to phenotypic variation in familial hypercholesterolaemia.[20] The possible role and effect of variation in any putative candidate modifier gene can be determined either using a case-control approach or, where parental DNAs are available, the transmission disequilibrium test.[35]

It is clear, however, that for most genetic diseases where the underlying basis of the disorder has not been established, this approach will not be viable. Identification of modifier loci will generally depend upon linkage. This can be based on mouse models, generally utilising an intercross between strains with discordant disease features, followed by a backcross or production of F_2 hybrids which can be used for linkage analysis. The Min mouse provides a model for familial colon cancer: it carries a mutant mouse *APC* gene and develops many intestinal adenomas. Using an intercross/backcross approach, a modifier locus *Mom-1* (Modifier of *Min-1*) was mapped to mouse chromosome 14, syntenic with human chromosome 1p35-p36.[33] Variation at this locus accounted for about 50% of the observed genetic variation in tumour number. Subsequently, the secretory phospholipase A2 locus was suggested as a candidate for *Mom-1*. Although experiments suggested an effect of a locus on 1p35-p36 on human FAP, no variation was identified in human secretory phospholipase A2 (PLA2G2A) that could account for this effect. In other cases, although the sites of modifier loci have been determined, no candidate genes have been identified. Mohlke *et al* analysed a cross between RIIIS/J and CASA/Rk, two strains of mice that exhibit a 20-fold difference in plasma von Willebrand factor (vWF) levels.[36] DNA samples from F_2 progeny with very high or very low plasma vWF levels were pooled and genotyped at 41 markers throughout the genome. A locus accounting for 63% of the total variance in vWF level was mapped to distal mouse chromosome 11. Using survival as a phenotype, Rozmahel *et al* showed that a major modifier locus of the cystic fibrosis transmembrane conductance regulator maps near the centromere of mouse chromosome 7.[37] Iakoubova *et al* found evidence for modifiers of mouse juvenile polycystic kidney disease on chromosomes 10 and 1.[38]

Mouse models provide certain advantages, such as the availability of large numbers of experimental matings and the ability to minimise variation in the environmental influences on disease. There are, however, potential problems with mouse models. First, in Mendelian disease, mouse and human disease phenotypes are often dissimilar, even if the underlying mutation is the same. Second, different genetic and cellular pathways may underlie apparently similar diseases in mouse and humans. Third, selective pressures and chance effects mean that the type and extent of genetic variation varies between different strains of laboratory mice and between mice and humans; hence, there may be no genetic variation at some potential modifying loci in one species whilst there is variation in another species. Fourth, the inbred nature of the laboratory mouse means that as a result of linkage disequilibrium, it is often difficult to identify modifier genes once they have been mapped; creation of a 'knockout' mouse may be necessary to test the effects of any putative modifier.

Owing to the lack of good candidate genes and potential problems with mouse models, the identification of many modifier genes will often primarily be based on linkage analysis in human pedigrees. Mapping modifier genes is, however, likely to be much less straightforward than identifying the primary disease

gene loci. Detection of linkage will probably depend upon the analysis of sibling pairs for a number of reasons. Firstly, although the sibling recurrence risk for a specific manifestation of the disease may be high, the overall risk will also be high; thus, relative risks will be at most modest, making the value of more distant relatives poor. Secondly, the population frequency of a modifier allele has to be high in order to co-segregate with the primary disease; thus, new modifying alleles are likely to derive from pedigree founders and the analysis of distant relatives will therefore provide no information on linkage or will be against linkage. Thirdly, if a particular disease manifestation is also influenced by environmental factors, affected individuals outside the nuclear family are obviously less attractive for study. Finally, many phenotypic variables are highly age-dependent and, although adjustment can be made for this, siblings are generally of similar ages.

Sibling Pair Methods for Identifying Modifier Genes

Linkage studies utilising sibling pairs can be based either on the analysis of affected sib pairs (ASPs), or on discordant sib pairs (DSPs). For a linked marker, the ASP and DSP allele probabilities of sharing 0, 1 or 2 alleles IBD (Z_i and Y_i respectively) will deviate in opposite directions from the null expectation (α_i.[39, 40]

In a sib-pair study, the power to demonstrate linkage depends upon the deviation of allele sharing between siblings from the probability of identity by descent (IBD) under the null expectation. This depends on the contribution that the locus makes to the genetic variation in the trait and is generally measured in terms of the increased risk to relatives of affected probands compared with the population risk (Appendix 1). Compared with ASPs, DSPs are generally not the most powerful approach to mapping disease genes characterised by low recurrence risks. However, to detect modifying loci, DSPs can be superior to ASPs, both in terms of power and because discordant sibs are *a priori* more likely to be segregating a modifying allele.

In order to compare the power of the two sib-pair methods in the identification of modifier loci, consider a hypothetical situation in which the a specific qualitative clinical feature of a genetic disease is present in 40% of cases and is wholly due to the co-inheritance of a modifier gene. If the modifier phenotype were dominantly inherited, the offspring recurrence risk will be 66% and the sibling recurrence risk 68%. Alternatively if the gene were to act recessively, then the offspring risk would be 63% and the sibling recurrence risk would be 66%. Figure 1 shows the power of DSPs and ASPs to demonstrate a lod score of 3.0 for varying numbers of sibling pairs if the gene were to act recessively. Because both the offspring and sibling recurrence risks are similar for both gene models the power to demonstrate linkage under the dominant model will be almost identical. Figure 1 clearly shows the DSP approach would be far superior to the ASP.

The power to detect linkage depends on the degree of deviation that the sharing IBD has from its null expectation. This depends on the contribution that the locus makes to the genetic variation in the trait and is generally measured in terms of the increased risk to relatives of affected probands compared with the population risk. However, the sibling recurrence risk is the most important parameter determining the relative efficiency of DSPs and ASPs. Figure 2 shows the absolute deviations of Z_0 and Y_0 from α_0 for ASPs and DSPs for a range of sibling recurrence risks under dominant and recessive modes of inheritance for a modifier gene. In both cases, when the sibling recurrence risk is low, there will be much less power in using DSPs than ASPs; when the sibling recurrence risk is high, however, the optimal method for detecting linkage will be by use of DSPs. This is because the magnitude of the deviations for DSPs compared with ASPs is given by $K_s/(1\text{-}K_s)$ and will be less than unity provided $K_s < 0.5$.[39]

The inevitability of reduced penetrance leads to a reduced power to detect linkage. Figure 3 shows the

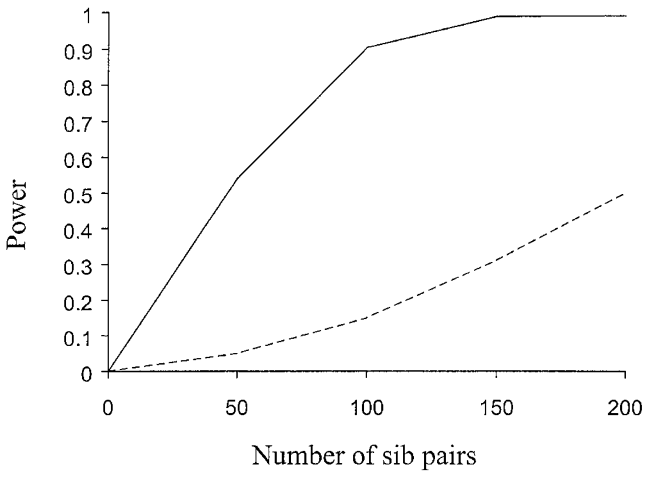

Figure 1 *Power comparison of ASPs (– – – –) and DSPs (——) to demonstrate a lod score of 3.0 assuming average marker spacing of 10 cM, infinite number of marker alleles and availability of parental data (for determination of power see Appendix 2)*

effect of incomplete penetrance on the absolute deviations from the null expectation of zero sharing under dominant and recessive models. Provided recurrence risk is high, DSPs still offer the most efficient method of demonstrating linkage.

The prospects for detecting linkage also depend upon the number of loci involved. Two or more genes can act either multiplicitively (i.e. the penetrance of the disease is the product of the penetrances contributed by two or more loci) or additively (i.e. the penetrance of the disease is represented by the sum of the penetrances contributed by two or more loci). Risch[40] showed that the ASP probabilities for a multiplicative model are identical to a single-locus formula. Hence, the DSP probabilities will be specified accordingly. For an additive model or if genetic heterogeneity exists, the situation is quite different, and for both ASPs and DSPs the power to detect linkage depends on the total relative risk and not solely on the relative risk attributable to the one locus. The existence of multiple contributing loci obviously greatly reduces the power of detecting a given locus, but the deterioration of power is usually greater for an additive model than for a multiplicative. The effect of genetic heterogeneity on the comparative power of the ASPs and DSPs is also shown. Again, the magnitude of deviations for DSPs compared with ASPs is given by $K_s/(1\text{-}K_s)$, making the DSPs a more favourable approach to mapping modifiers if the recurrence risk is high (Appendix 1).

If some of the disease features can be ascribed to phenocopies, DSPs are less likely to be discordant and ASPs less likely to be concordant at a marker which is linked to the trait locus, leading to a reduction in power for both approaches.

Irrespective of the relative merits of DSPs versus ASPs the power of any proposed study for detecting a modifier gene will depend upon the absolute deviations of Z_i or Y_i from the null expectations, and in many cases where the overall frequency of the condition is high the numbers of cases required to have any hope of

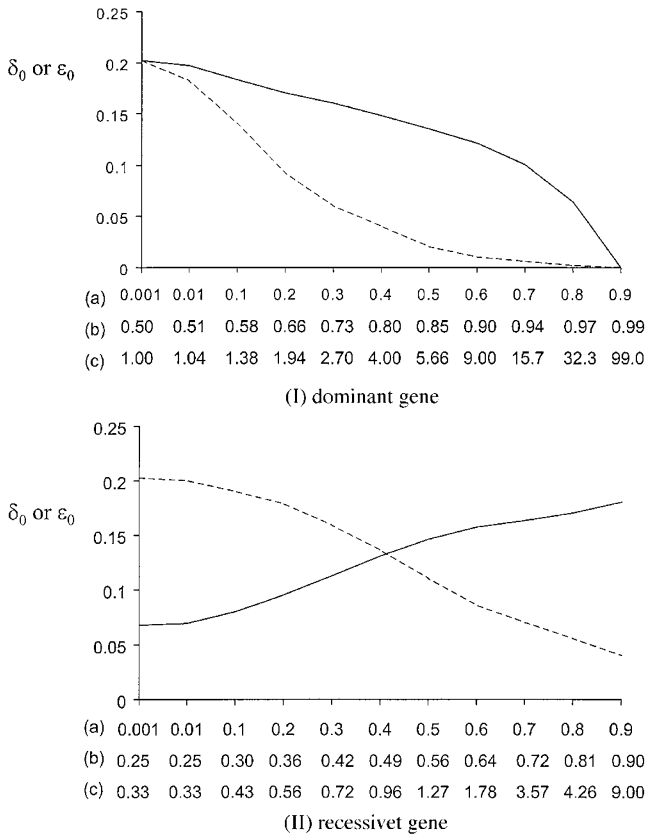

Figure 2 *Effect of (a) frequency of disease gene (b) sibling recurrence risk and (c) $K_s/(1\text{-}K_s)$ on deviations of Z_0 and Y_0 from α_0 (δ_i and ε_i respectively)* **(i)** *dominant gene ASP– – – – DSP——* **(ii)** *recessive gene ASP– – – – DSP——*

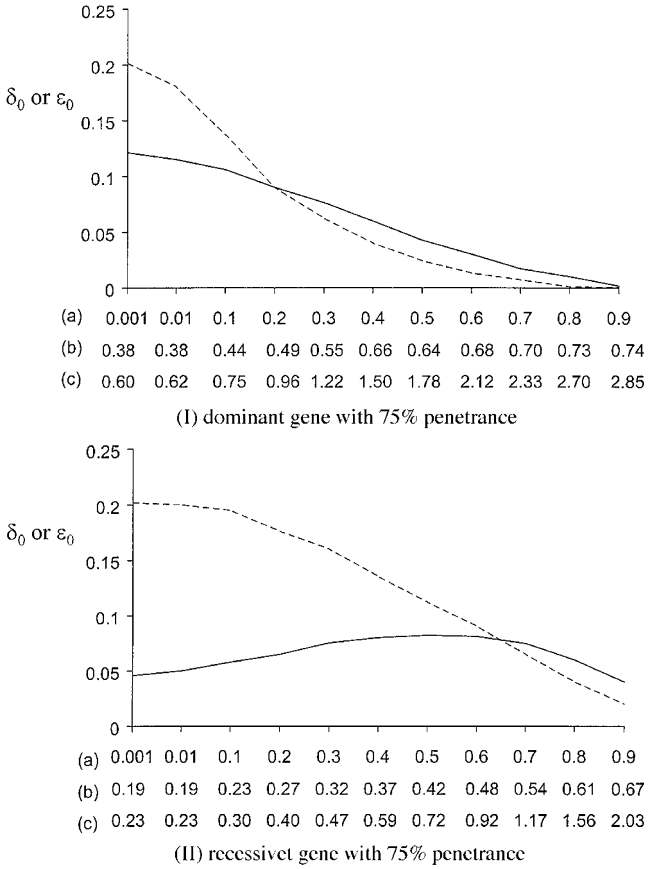

Figure 3 *Effect of (a) frequency of disease gene (b) sibling recurrence risk and (c) $K_s/(1\text{-}K_s)$ on deviations of Z_0 and Y_0 from α_0 (°$_i$ and ε_i respectively)* **(i)** *dominant gene with 75% penetrance ASP – – – – DSP——* **(ii)** *recessive gene with 75% penetrance ASP– – – – DSP——*

detecting linkage will be prohibitively large. For example if 27% of individuals carry and manifest the effects of the modifier allele and the sibling recurrence risk is 50%, 200 sibling pairs will provide 90% power to demonstrate linkage using either ASPs or DSPs; but if 56% of individuals carry and manifest the modifier allele and the sibling recurrence risk is 64%, 500 siblings pairs will provide only 75% power to demonstrate linkage using DSPs and only 13% using ASPs.

Alternative Strategies to Identify Modifier Genes

To identify modifier genes in most cases will be reliant on either ASPs or DSPs, although some novel approaches can be applied to certain situations, for instance if a modifying gene acts as a tumour suppressor. According to the two-hit hypothesis, both copies of a tumour suppressor gene have to be abnormal for tumour development. Patients with an inherited predisposition carry a defective allele of a tumour suppressor gene in the germ line. In tumours from these patients, the second allele is inactivated by a somatic event, such as a deletion. This can be assessed by loss of heterozygosity (LOH) studies. LOH in tumours can be treated as an additional observation on disease phenotype and used to enhance the power to detect linkage.[41] The use of LOH can be extended and used to map predisposition loci directly and it is therefore conceivable that this approach could be adopted for the detection of modifying loci in certain Mendelian conditions characterised by variable tumour expression, such as macrocephaly in neurofibromatosis or desmoids in FAP. It is, however, arguable that any modifier locus that undergoes allele loss will also undergo somatic mutation plus allele loss in some tumours which do not have the modifying allele in the germ line. Thus far, study of tumour suppressor genes has failed to identify any loci which are good candidates for modifying loci in humans.

Another slightly different approach to linkage could be applied to the situation of graft-versus-host (GVH) reaction. The greatest determinant of GVH reaction in tissue transplants is the human leukocyte antigen (HLA) system. It is, however, recognised that even with HLA matching between siblings, patients can have very different outcomes in terms of GVH reaction. A proportion of transplanted individuals requires little immunosuppression, and a proportion experiences GVH reaction despite immunosuppression, thus suggesting that other genes influence this process. It is very probable that a recipient of a transplant with no sequelae will be more likely to be concordant with his or her donor sibling at the modifier locus as compared to those with a poor course. Minor histocompatibility antigens have been shown to segregate in families[30] and immunodominant minor histocompatibility antigens identified in humans and the mouse.[28,29] Taking sibling pairs from both ends of the clinical spectrum offers an opportunity to undertake a combined discordant and concordant type of analysis in order to identify further modifier loci.

Discussion

Understanding the factors that control phenotypic expression in genetic diseases is important for understanding the disease process and for patient counselling. The considerable inter- and intra-familial variability in a number of genetic diseases is likely to result from the action of modifier genes. These modifier loci may provide targets for therapy that are more useful than the genes primarily involved in causing the disease. One approach to locate modifier genes has been to use mouse models. However, this indirect approach is unlikely to obviate the need to undertake linkage searches to identify novel modifier genes in humans. Where there is an understanding of the biological basis of the disease, other genes involved in the disease process can be evaluated as candidates for disease modification. However, it is likely that for the foreseeable future the identification of modifier genes will depend principally upon linkage analysis, either in the form of ASPs or DSPs. Extended pedigrees are unlikely to be useful, essentially because of the risk of phenocopies and the probable high frequency of modifier alleles in the general population, and hence in pedigree founders. In contrast to primary mapping of disease genes where ASPs are generally preferable, DSPs represent an attractive option, because of the high recurrence rates seen, and also because of the reduced survival characteristics associated with severe phenotypes which make the likelihood of ascertaining and obtaining clinical material from living cases less likely. Furthermore, as has been pointed out by others,[42] concordant siblings include not only pairs sharing at the locus of interest, but also those in whom the features are phenocopies due to common environment; in contrast discordant siblings will select for those siblings which possess sufficient dissimilarity at the locus to the any environmental influence. We therefore favour the use of DSPs for identifying modifying loci in

both qualitative and quantitative human genetic diseases.

References

1 Andrew SE *et al*: The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nat Genet* 1993; **4**: 398–403.
2 Harley HG *et al*: Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* 1992; **355**: 545–547.
3 Cama A *et al*: Multiplex PCR analysis and genotype phenotype correlations of frequent *APC* mutations. *Hum Mutat* 1995; **5**: 144–152.
4 Gayther SA *et al*: Variation of risks of breast and ovarian cancer associated with different germline mutations of the *BRCA2* gene. *Nat Genet* 1997; **15**: 103–105.
5 Neidich JA *et al*: Heterogeneity of clinical severity and molecular lesions in Aicardi syndrome. *J Pediatr* 1990; **116**: 911–917.
6 De Saint Basile G *et al*: X-linked thrombocytopenia and Wiskott Aldrich syndrome: similar regional assignment but distinct X inactivation pattern in carriers. *Ann Hematol* 1991; **63**: 107–110.
7 Aral B *et al*: Novel nonsense mutation in the hypoxanthine guanine phosphoribosyltransferase gene and nonrandom *X* inactivation causing Lesch Nyhan syndrome in a female patient. *Hum Mutat* 1996; **7**: 52–58.
8 Beaumont V, Jacotot B, Beaumont JL: Ischaemic disease in men and women with familial hypercholesterolaemia and xanthomatosis. A comparative study of genetic and environmental factors in 274 heterozygous cases. *Atherosclerosis* 1976; 24: 441–450.
9 Schulz B, Bremer HJ: Nutrient intake and food consumption of adolescents and young adults with phenylketonuria. *Acta Paediatr* 1995; **84**: 743–748.
10 Buiting K *et al*: Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome. *Nat Genet* 1995; **9**: 395–400.
11 Wilson LC *et al*: Parental origin of *Gs* alpha gene mutations in Albright's hereditary osteodystrophy. *J Med Genet* 1994; **31**: 357–359.
12 Chatkupt S, Antonowicz M, Johnson WG: Parents do matter: genomic imprinting and parental sex effects in neurological disorders. *J Neurol Sci* 1995; **130**: 1–10.
13 Leach RJ *et al*: Preferential retention of paternal alleles in human retinoblastoma: evidence for genomic imprinting. Cell Growth Differ 1990; 1: 401–406.
14 Cookson WO *et al*: Maternal inheritance of atopic IgE responsiveness on chromosome IIg. *Lancet* 1992; **340**: 381–384.
15 Byrne J *et al*: Multigeneration maternal transmission in Italian families with neural tube defects. *Am J Med Genet* 1996; 66: 303–310.
16 Margaritte Jeannin P, Clerget Darpoux F, Hors J, Deschamps I: Testing parental imprinting in insulin dependent diabetes mellitus by the marker association segregation chi 2 method. *Am J Hum Genet* 1995; **56**: 1080–1087.
17 Zerres K, Rudnik Schoneborn S, Deget F: Childhood onset autosomal dominant polycystic kidney disease in sibs: clinical picture and recurrence risk. *J Med Genet* 1993; **30**: 583–588.
18 Bartstra HL *et al*: Mosaic expression of hypohidrotic ectodermal dysplasia in an isolated affected female child. *Arch Dermatol* 1994; **130**: 1421–4.
19 Tanaka F *et al*: Differential pattern in tissue specific somatic mosaicism of expanded CAG trinucleotide repeat in dentatorubral pallidoluysian atrophy, Machado Joseph disease, and X-linked recessive spinal and bulbar muscular atrophy. *J Neurol Sci* 1996; **135**: 43–50.
20 Lindahl G, Mailly F, Humphries S, Seed M: Apolipoprotein E phenotype and ipoprotein (a) in familial hypercholesterolaemia: implication for lipoprotein (a) metabolism. *Clin Investig* 1994; **72**: 631–638.
21 Phelan CM *et al*: Ovarian cancer risk in *BRCA1* carriers is modified by the *HRAS1* variable number of tandem repeat (VNTR) locus. *Nat Genet* 1996; **12**: 309–311.
22 Tomlinson IPM *et al*: A modifying locus for familial adenomatous polyposis may be present on chromosome 1p35-p36. *J Med Genet* 1996; **33**: 268–273.
23 Kulozik AE: Beta thalassemia: molecular pathogenesis and clinical variability. *Klin Padiatr* 1991; **203**: 276–283.
24 Cao A, Murru S: Molecular pathology and detection of beta thalassemias. *Prog Clin Biol Res* 1989; **309**: 3–11.
25 Easton DF, Ponder MA, Huson SM, Ponder BA: An analysis of variation in expression of neurofibromatosis (NF) type 1 (NF1): evidence for modifying genes. *Am J Hum Genet* 1993; **53**: 305–313.
26 Steinbach P: Mental impairment in Martin Bell syndrome is probably determined by interaction of several genes: simple explanation of phenotypic differences between unaffected and affected males with the same X chromosome. *Hum Genet* 1986; **72**: 248–252.
27 Embury SH: Alpha thalassemia. A modifier of sickle cell disease. *Ann N Y Acad Sci* 1989; **565**: 213–221.
28 Quinn M, Angelico MC, Warram JH, Krolewski AS: Familial factors determine the development of diabetic nephropathy in patients with IDDM. *Diabetologia* 1996; **39**: 940–945.
29 den Haan JM *et al*: Identification of a graft-versus-host disease associated human minor histocompatibility antigen. *Science* 1995; **268**: 1476–1480.
30 Perreault C *et al*: Identification of an immunodominant mouse minor histocompatibility antigen (MiHA). T cell response to a single dominant MiHA causes graft-versus-host disease. *J Clin Invest* 1996; **98**: 622–628.
31 Nishimura M *et al*: Family segregation analysis of inheritance of human minor histocompatibility antigen. *Cell Immunol* 1995; **163**: 319–322.
32 Moser AR *et al*: *ApcMin*: a mouse model for intestinal and mammary tumorigenesis. *Eur J Cancer* 1995; **31A**: 1061–1064.
33 Dietrich WF *et al*: Genetic identification of *Mom-1*, a major modifier locus affecting Min-induced intestinal neoplasia in the mouse. *Cell* 1993; **75**: 631–639.
34 Krolewski AS *et al*: The changing natural history of nephropathy in type I diabetes. *Am J Med* 1985; **78**: 785–794.
35 Ewens WJ, Spielman RS: The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet* 1995; **57**: 556–564.

36 Mohlke KL *et al*: A novel modifier gene for plasma von Willebrand factor level maps to distal mouse chromosome 11. *Proc Natl Acad Sci USA* 1996; **93**: 15352–15327.
37 Rozmahel R *et al*: Modulation of disease severity in cystic fibrosis transmembrane conductance regulator-deficient mice by a secondary genetic factor. *Nat Genet* 1996; **12**: 280–287.
38 Iakoubova OA, Dushkin H, Beier DR: Localization of a murine recessive polycystic kidney disease mutation and modifying loci that affect disease severity. *Genomics* 1995; **26**: 107–114.
39 Risch N: Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 1990; **46**: 222–228.
40 Risch N: Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 1990; **46**: 229–241.
41 Teare MD, Rohde K, Santibanez Koref MF: The use of loss of constitutional heterozygosity data to ascertain the location of predisposing genes in cancer families. *J Med Genet* 1994; **31**: 448–52.
42 Rogus JJ, Krolewski AS: Using discordant sib pairs to map loci for qualitative traits with high sibling recurrence risk. *Am J Hum Genet* 1996; **59**: 1579–81.

Appendix 1: Allele Sharing Probabilities of ASPs and DSPs

Risch has previously derived the allele sharing probabilities of ASPs and DSPs.[39, 40] Adopting the same nomenclature:

$\alpha_i = P$ (sibs share marker alleles IBD)

$Z_i = P$ (sibs share i marker alleles IBD|ASP)

$Y_i = P$ (sibs share i marker alleles IBD|DSP)

For $I = 0, 1, 2$, α_i equals 1/4, 1/2 and 1/4 respectively. When a marker is unlinked, $Z_i = Y_i = \alpha_i$.

For ASPs the degree of deviation, δ_i that the probability of sharing IBD has from its null expectation is given by:[40]

$$\delta_i = Z_i - \alpha_i$$

Similarly the degree of deviation for DSPs, ε_i, is given by:[40]

$$\varepsilon_i = Y_i - \alpha_i$$

When $\theta = 0$, the ASP probabilities are given by:[40]

$$Z_0 = \alpha_0(1/\lambda_s)$$

$$Z_1 = \alpha_1(\lambda_0/\lambda_s)$$

$$Z_2 = \alpha_2(\lambda_m/\lambda_s)$$

and the DSP probabilities are given by:[40]

$$Y_{s0} = \alpha_0[(1-K)/(1-K_s)]$$

$$Y_{s1} = \alpha_1[(1-K_o)/(1-K_s)]$$

$$Y_{s2} = \alpha_2[(1-K_m)/(1-K_s)]$$

$\lambda_s = K_s/K$ and $\lambda_o = K_o/K$ where K_s and K_o are the sibling and offspring recurrence risks respectively, and K is the population risk.

These formulae hold true irrespective of the mode of inheritance at the disease locus any number of alleles and their frequencies, penetrance and population prevalence. The only requirement is that recombination be negligible.

Risch[40] showed that the ASP probabilities for a multiplicative model are identical to a single locus formulae except that λ_{1s} replaces λ_s. Hence the DSP probabilities will be according specified. The formulae for the genetic heterogeneity model is well approximated by the additive model. The ASP probabilities for two unlinked loci given by Risch[37] are:

$$Z_0 = \alpha_0 - \alpha_0(K_1/K_2).1/\lambda_s.(\lambda_{1s}-1)$$

$$Z_1 = \alpha_1 - \alpha_1(K_1/K_2).1/\lambda_s.(\lambda_{1o}-\lambda_{1s})$$

$$Z_2 = \alpha_2 + \alpha_2(K_1/K_2).1/\lambda_s.(\lambda_{1m}-\lambda_{1s})$$

where $K = K_1 + K_2$ and the overall sibling recurrence risk is given by:[39]

$$KK_R = K_1K_{1R} + K_2K_{2R} + 2K_1K_2$$

Through a conditional probability argument the DSP probabilities can be shown to be given by:

$$Y_0 = \alpha_0 + a_0.(K_1/K_2).K.(\lambda_{1s}-1)/(1-K_S)$$

$$Y_1 = \alpha_1 - \alpha_1.(K_1/K_2).K.(\lambda_{1o}-\lambda_{1s})/(1-K_S)$$

$$Y_2 = \alpha_2 - \alpha_2.(K_1/K_2).K.(\lambda_{1m}-\lambda_{1s})/(1-K_S)$$

For both ASPs and DSPs the power to detect linkage depends on the total value of λ and not solely on the λ value at the one locus.

Appendix 2. Power Calculations

The maximum likelihood statistics for ASPs and DSPs are defined as the maximised likelihood ratio divided by the likelihood when Z_i or Y_i equals α_i. Using the ASP and DSP probabilities of sharing 0, 1 and 2 alleles IBD, the number of affected sibling pairs who share 0, 1, or 2 alleles was simulated based upon these allele sharing probabilities for N sibling pairs. Incorporating the effect of recombination between the trait and the

marker loci Risch[40] showed that the ASPs probabilities are given by:

$$Z_0 = \alpha_0 - \alpha_0(2\Psi-1).1/\lambda_s.(\lambda_s-1)$$

$$Z_1 = \alpha_1 + \alpha_1(2\Psi-1).1/\lambda_s.(\lambda_s-\lambda_o)$$

$$Z_2 = \alpha_2 + \alpha_2(2\Psi-1).1/\lambda_s.(\lambda_m-\lambda_s)$$

where the parameter $\Psi = \theta^2 + (1-\theta)^2$and $Y_i = \alpha_i[1/(1-K_s)] - Z_i[K_s/(1-K_s)]$[42]

To reflect undertaking a genome-wide screen using markers spaced at 10 cM intervals, θ was set at 0.05. For a range of *N*, 5000 simulations were generated. The power was given by the percentage of simulations which exceeded a lod of 3.0.