Brief papers

Linkage disequilibrium between four intragenic polymorphic microsatellites of the NF1 gene and its implications for genetic counselling

M Carmen Valero, Eladio Velasco, Ana Valero, Felipe Moreno, Concepción Hernández-Chico

Abstract

Four intragenic polymorphic microsatellite markers, AAAT *Alu* repeat, IVS27AC28.4, ACI27.2, and IVS38GT53.0, located along a 65 kb DNA region of the NF1 gene, were used to genotype 64 Spanish families with neurofibromatosis type 1 (NF1). Linkage disequilibrium between each pair of markers was evaluated. Three of these markers, AAAT *Alu* repeat, ACI27.2, and IVS38GT53.0, exhibit linkage disequilibrium between each other. Analysis of extended haplotypes provides further evidence of the disequilibrium within this region since only 11 haplotypes account for 52% of the total chromosomes. Because of linkage disequilibrium, the informativeness of marker combinations for genotyping of NF1 families is diminished. There was no difference in the overall distribution of alleles between affected and normal chromosomes. An at risk haplotype was not found, as expected for a disease with at least 50% of cases being sporadic.

(*J Med Genet* 1996;33:590–593)

Key words: neurofibromatosis 1; linkage disequilibrium; microsatellites.

Unidad de Genética Molecular, Hospital Ramón y Cajal, Carretera de Colmenar Km 9, 28034 Madrid, Spain
M C Valero
E Velasco
A Valero
F Moreno
C Hernández-Chico

Correspondence to:
Dr Hernández-Chico.

Received 11 October 1995
Revised version accepted for publication 13 March 1996

Neurofibromatosis type 1 (NF1) is an autosomal dominant disorder which is found in all populations, with an incidence of about 1 in 3000 births. NF1 is characterised by café au lait spots, cutaneous neurofibromas, axillary or inguinal freckling, and Lisch nodules.[1] Approximately one third of affected subjects display assorted medical complications, from mental retardation and learning disabilities to an increased risk of malignancy. The disease exhibits pleiotropic clinical manifestations even among family members who carry the same disease mutation.

The NF1 gene has been mapped to the 17q11.2 chromosomal region. This gene spans more than 300 kb of genomic DNA, and its coding sequence is composed of 59 exons. The 13 kb transcript encodes neurofibromin, a 2818 amino acid polypeptide.[2] The NF1 mutation rate, 1×10^{-4}/gamete/generation, is about 100-fold higher than the usual rate for a single locus. In fact, approximately 50% of cases are the result of de novo mutations.

Even now that the complete sequence for the NF1 gene is known, the detection of the molecular defect in patients is hampered by the apparent lack of a mutation "hot spot" in this large and highly mutable gene. Therefore, genetic diagnosis remains based mainly on the analysis of cosegregating DNA markers.

Linkage disequilibrium among RFLPs in the neurofibromatosis 1 region has been previously reported.[3,4] We have also investigated this issue using much more informative intragenic markers, those associated with STRs (short tandem repeats). Three of them are CA repeats (IVS27AC28.4,[5] ACI27.2 (GenBank accession No L03727), and IVS38GT53.0[6]), and another one is the AAAT repeat of an *Alu* sequence (AAAT *Alu* repeat).[7] These polymorphisms are distributed along a 65 kb DNA region of the NF1 gene. The *Alu* polymorphism is the most proximal of this set of markers, being located at the 5′ end of intron 27b, close to the translocation breakpoint t(1;17). The repeats IVS27AC28.4, ACI27.2 (both at intron 27b), and IVS38GT53.0 (intron 38) are 40, 45, and 65 kb, respectively, distal to the AAAT *Alu* repeat marker. Other STRs located upstream of intron 27 or downstream of intron 38 are not available.

Linkage disequilibrium analysis could provide insight into the usefulness of these markers for genetic counselling. Furthermore, since recombination and disequilibrium are closely related, the pattern of allelic association between these physically related markers could be useful to detect a recombinational "hot spot".

Linkage disequilibrium between these multiallelic marker loci was tested by the likelihood based method[8] and extended haplotypes were also constructed to evaluate this linkage disequilibrium.

Materials and methods

FAMILIES

A total of 64 NF1 families, 25 familial cases (67 affected and 64 unaffected members) and 39 sporadic cases (39 affected and 175 unaffected members) were analysed. Clinical

diagnosis was established according to the NIH consensus criteria.[1] All the families were unrelated and came from different regions of the country, representing an apparently random sample of the Spanish population.

DETECTION OF POLYMORPHISMS

DNA was extracted from peripheral blood lymphocytes by standard extraction procedures. Family members were typed with four intragenic polymorphic microsatellites: AAAT *Alu* repeat, IVS27AC28.4, ACI27.2, and IVS38GT53.0.

Analysis of AAAT *Alu* repeat was performed as described by Xu *et al*,[7] except that no radioactive nucleotide was added in the PCR reaction and the PCR products were analysed by denaturing gradient gel electrophoresis (DGGE).[9] Samples were electrophoresed onto 5% polyacrylamide gels with a linear increasing denaturant gradient 10 to 60% (100% denaturant = 7 mol/l urea and 40% formamide; acrylamide/bisacrylamide = 37·5/1). The electrophoresis was performed at constant temperature (60°C) at 160 V for three and a half hours and bands were visualised by ethidium bromide staining.

The IVS27AC28.4 microsatellite was amplified with primers previously described.[5] PCR samples contained 150 ng genomic DNA, 200 µmol/l each dNTP, 15 pmol of each primer, 2 mmol/l $MgCl_2$ and 1 unit *Taq* polymerase (Perkin-Elmer) in a final volume of 25 µl. After initial denaturation the reaction mixtures were subjected to 30 cycles of PCR consisting of one minute at 94°C, one minute at 55°C, and one minute at 72°C, with a final extension of five minutes at 72°C.

ACI27.2 has formerly been cited (GenBank accession number L03723). For PCR amplification of this microsatellite, we designed the following pair of primers: ACI27.2A 5'GTGGAACTGCAGCAATTATTT 3' and ACI27.2B 5' AAACTTGAGGTGATGACAGGA 3'. PCR conditions were identical to those for IVS27AC28.4 but 1·5 mmol/l $MgCl_2$ was added.

The amplification products of IVS27AC28.4 and ACI27.2 were analysed on native polyacrylamide gels (5·5% or 7% respectively) and bands were visualised by ethidium bromide staining.

The analysis of IVS38GT53.0 was performed under conditions previously described.[6]

STATISTICAL ANALYSIS

We used the likelihood based approach to estimate the allele and haplotype frequencies. This approach is independent of the number of marker alleles and also independent of the number of marker loci considered.

All microsatellites were tested to determine whether the observed genotypic distributions conformed to those expected under Hardy–Weinberg equilibrium (HWE) by using the HWE program (Linkage utility programs).[10] To test and estimate linkage disequilibrium between alleles of these four markers, we used the EH program (version 1.1) (Linkage utility programs).[10] To run this program without any complex modification of the source code, it has to perform the test between two loci each having a maximum number of seven alleles. Therefore, we omitted those unrelated subjects carrying low frequency alleles of ACI27.2 (alleles 2 and 3) and IVS38GT53.0 (alleles 3 and 8). For multiple comparisons the significance levels were corrected by using the Bonferroni procedure.

We tested whether there was a difference in the overall distribution of alleles on disease bearing and normal chromosomes. This involves a χ^2 test with $2 \times n$ contingency table, where n is the number of alleles at a particular locus. We also tested whether there is a single allele with a significantly higher frequency on disease bearing chromosomes than on normal chromosomes. Thus, we examined each allele separately to look for an increase (using a standard one sided test).

Table 1 Allele sizes, frequency distributions, and estimated heterozygosity for each microsatellite

Microsatellite	*Allele*	*Size (bp)*	*Observed frequency*	*Estimated heterozygosity*
AAAT *Alu* repeat				0·5201
	1	407	0·1042	
	2	403	0·6500	
	3	399	0·2125	
	4	395	0·0333	
IVS27AC28.4				0·6289
	1	219	0·0333	
	2	217	0·0458	
	3	215	0·0500	
	4	213	0·1042	
	5	211	0·5792	
	6	209	0·1208	
	7	207	0·0667	
ACI27.2				0·7364
	1	173	0·0122	
	2	171	0·0040	
	3	165	0·0120	
	4	161	0·0161	
	5	159	0·2419	
	6	157	0·0122	
	7	155	0·0847	
	8	151	0·3668	
	9	149	0·2500	
IVS38GT53.0				0·8123
	1	187	0·0420	
	2	185	0·2773	
	3	183	0·0210	
	4	181	0·2353	
	5	179	0·1681	
	6	177	0·0840	
	7	175	0·0252	
	8	173	0·0168	
	9	171	0·1303	

Results

GENOTYPING OF NF1 FAMILIES

A total of 345 subjects from 64 NF1 families were genotyped with the four microsatellites described above. The allele sizes, frequency distributions, and estimated heterozygosity for each microsatellite were obtained after analysing 125 unrelated subjects (25 affected and 100 unaffected) and are summarised in table 1. No recombination or mutation events were observed in a total of 318 meioses.

The allele frequencies for AAAT *Alu* repeat, IVS27AC28.4, and IVS38GT53.0 do not differ significantly from the original descriptions of these microsatellites. ACI27.2 has formerly

Table 2 Frequency of extended haplotypes

Haplotype*	No of chromosomes	%
2582	32	13·5
2659	15	6·4
2584	13	5·5
2594	13	5·5
1594	11	4·7
3575	10	4·2
3555	7	3·0
2794	6	2·5
2795	6	2·5
2382	5	2·1
2282	5	2·1
	(Σ 123	52·0)
74 others	113	48·0

*The numbers indicate the alleles of AAAT *Alu* repeat, IVS27AC28.4, ACI27.2, and IVS38GT53.0, respectively.

been cited but there is no description of the number of alleles and their frequencies. We detected nine alleles for this microsatellite, and heterozygosity of 73% was estimated (table 1).

LINKAGE DISEQUILIBRIUM BETWEEN MARKERS IN THE NF1 GENE

For all the microsatellites analysed, the genotypic distributions conform to their respective Hardy–Weinberg equilibrium expectations in the group of 125 unrelated subjects.

We tested linkage disequilibrium between alleles of each pair of markers by using the EH program (version 1.1). For all the marker pairs except AAAT *Alu* repeat/IVS27AC28.4 we could reject the hypothesis of no association (Hardy–Weinberg equilibrium), that is to say all the marker pairs with the exception of the AAAT *Alu* repeat/IVS27AC28.4 pair show linkage disequilibrium at the significance level $p<0·001$.

As described in Methods, we studied the possible disequilibrium between NF1 and the four markers. Our sample of 25 affected chromosomes identified in familial cases and 25 normal chromosomes carried by these patients were compared. The 39 sporadic affected subjects were not included in the study since we could not identify which chromosome was the affected one. In the overall distribution of alleles on affected and normal chromosomes, there was no significant difference. As expected for a disease in which at least 50% of cases are sporadic, no disequilibrium was observed between the disease locus and any of the alleles of these markers (data not shown).

ANALYSIS OF EXTENDED HAPLOTYPES

To investigate the linkage disequilibrium further, extended haplotypes with the four microsatellite markers were constructed. This analysis included 236 unrelated parental chromosomes (211 normal and 25 affected) for which unambiguous haplotypes were established. Among the possible 2268 haplotypes, only 85 were observed. Eleven haplotypes, with frequencies higher than 2%, represent 52% of the total chromosomes (table 2). Ten more haplotypes with frequencies between 1 and 2% represent 15% of the total chromosomes. The rest of the haplotypes (64) have frequencies of less than 1% and most of them (54) appeared just once (frequency = 0·4).

The finding of 11 haplotypes that account for 52% of the total chromosomes is consistent with the allelic association that we have detected between these microsatellites.

For genetic counselling purposes, it is relevant to evaluate the consequences of allelic association on the incidence of homozygotes for this set of polymorphic microsatellites. It is likely that the frequency of homozygotes is increased since only 11 haplotypes account for 52% of the total chromosomes. Analysis of homozygosity was carried out on unaffected subjects to avoid confusion between putative hemizygosity (apparent homozygotes), owing to deletions causing disease, and real homozygosity.

We observed an increase (about five times higher than expected) in the number of homozygotes for the four markers ($\chi^2 = 12·32$, $p = 0·00045$, $df = 1$). The distribution of homozygotes for all possible combinations of three markers is shown in table 3. A significant increase in homozygosity was found for the triads AAAT *Alu* repeat/ACI27.2/IVS38GT53.0 and IVS27AC28.4/ACI27.2/IVS38GT53.0, but it was not found for the triads in which AAAT *Alu* repeat and IVS27AC28.4 were included together. Finally, we achieved a similar effective heterozygosity by genotyping either the combination of more informative markers (IVS27AC28.4/ACI27.2/IVS38GT53.0) or a combination of less informative ones. These results are in agreement with the pattern of disequilibrium found by using the EH program, which did not detect significant disequilibrium between AAAT *Alu* repeat and IVS27AC28.4.

Table 3 Frequency of unaffected homozygotes for the different triads of markers

Triad	E (%)	O (%)	χ^2	p value
AAAT *Alu* repeat/IVS27AC28.4/ACI27.2	5·0	12·0	6·05	NS
AAAT *Alu* repeat/ACI27.2/IVS38GT53.0	2·5	10·0	21·12	4×10^{-6}
AAAT *Alu* repeat/IVS27AC28.4/IVS38GT53.0	3·3	7·5	4·08	NS
IVS27AC28.4/ACI27.2/IVS38GT53.0	2·0	7·5	10·12	1×10^{-3}

Corrected significance level $\alpha' = 0·0127$.
E = expected. O = observed.
NS = not significant.

Discussion

Our results provide further evidence of linkage disequilibrium in the neurofibromatosis type 1 (NF1) region. Previous reports have shown such disequilibrium among different intragenic and extragenic RFLP markers.[3 4 11] Messiaen *et al*[4] found that the AAAT *Alu* repeat marker segregates independently of five intragenic RFLPs and they recommended genotyping this *Alu* polymorphism. We tested such linkage disequilibrium between four intragenic microsatellites, including the AAAT *Alu* repeat marker, which in our view provides a more insightful reading of linkage disequilibrium in the NF1 region.

Our estimation of allelic association between the different marker pairs showed significant disequilibrium, except for the AAAT *Alu* repeat/IVS27AC28.4 pair. Several explanations

for this non-significant allelic association can be considered. First, this result may be a simple consequence of their allele frequencies. Both markers show one allele with a frequency higher than 0·50 and act like diallelic systems. However, our sample (236 haplotypes) is statistically large enough to allow us to detect even weak disequilibrium if it exists.[12] Second, a non-uniform disequilibrium pattern across a relatively small genomic region could be the result of an uneven recombinational rate.[13 14] The existence of a "hot spot" of recombination between AAAT *Alu* repeat and IVS27AC28.4, or the possible recombinant nature of the *Alu* sequence seem to be unlikely since then the alleles of distal markers would reflect the putative "crossings over". Third, non-significant disequilibrium between the AAAT *Alu* repeat and the IVS27AC28.4 markers could be the result of a higher sequence instability of these microsatellites. For the AAAT *Alu* repeat marker, a high mutation rate (0·36% per allele) has been previously reported.[15] However, we did not detect any new allele from a total of 318 meioses. Perhaps the mutation rate was overestimated because most of the new alleles were detected in that study by using DNA from transformed lymphoblastoid cell lines of CEPH families. It has been reported that the majority of events occurred during or after establishment of the in vitro lymphoblastoid cell lines.[16] The mutation rate of this AAAT *Alu* repeat marker could be calculated by combining our data with those of Lázaro *et al*[15] in NF1 families (not obtained from lymphoblastoid cell lines of CEPH families), giving an accurate mutation rate of $1{\cdot}6 \times 10^{-3}$/gamete/generation. This value is similar to the average rate calculated for tetra- and dinucleotide repeats.[16] On the other hand, we have not detected any mutation event of marker IVS27AC28.4. However, the number of meioses analysed (318) is insufficient to calculate an accurate mutation rate and further segregation studies are required to consider a greater instability of this microsatellite.

Linkage disequilibrium data obtained in this outbred Spanish population sample allow one to make a number of observations with regard to the use of these intragenic markers for genotyping NF1 families. It must be stressed that, while linkage disequilibrium diminishes the informativeness of intragenic microsatellite combinations, this situation is not as significant as reported for five intragenic RFLP markers for which 65% of chromosomes had a specific haplotype.[4] To obtain conclusive information in familial segregation analysis, genotyping of at least three of these intragenic microsatellites might be necessary. To achieve the greatest amount of information two alternative combinations of this set of markers should be used (AAAT *Alu* repeat/IVS27AC28.4/IVS38GT53.0 or IVS27AC28.4/ACI27.2/IVS38GT53.0) which yield an effective heterozygosity of up to 92·5%. For genetic counselling, testing of additional extragenic polymorphisms must be considered. Recombination in the region between the centromere and the NF1 locus appears to be very rare, and tight linkage disequilibrium among proximal RFLP markers in the pericentromeric region is observed.[11] Therefore, testing of proximal markers might provide no useful information. Genotyping of distal markers is more advisable, in particular the p11-1F10 polymorphism which is only 70 kb from the 3′ end of the NF1 gene, and exhibits low disequilibrium values with several RFLP markers in the NF1 gene.[3] Identification of new microsatellites in the NF1 region should improve the genetic analysis of NF1 families.

We thank the NF1 families and the clinicians who collaborated in this study, particularly I Pascual-Castroviejo. We are also grateful to J Ott for his advice on linkage disequilibrium and for providing us with the computer programs used in this work, V Allmand and J S Beckmann for their constructive comments, and J L San Millán and I del Castillo for their critical reading of this manuscript.

1 National Institutes of Health Consensus Development Conference. Neurofibromatosis Conference Statement. *Arch Neurol* 1988;**45**:575–8.
2 Li Y, O'Connell P, Huntsman Breidenbach H, *et al.* Genomic organization of the neurofibromatosis 1 gene (NF1). *Genomics* 1995;**25**:9–18.
3 Jorde LB, Watkins WS, Viskochil D, O'Connell P, Ward K. Linkage disequilibrium in the neurofibromatosis I (NFI) region: implications for gene mapping. *Am J Hum Genet* 1993;**53**:1038–50.
4 Messiaen L, De Bie S, Moens T, Van den Enden A, Leroy J. Lack of independence between five DNA polymorphisms in the NF1 gene. *Hum Mol Genet* 1993;**2**:485.
5 Lázaro C, Gaona A, Estivill X. Two CA/GT repeat polymorphisms in intron 27 of the human neurofibromatosis type 1 (NF1) gene. *Hum Genet* 1994;**93**:351–2.
6 Lázaro C, Gaona A, Xu G, Weiss R, Estivill X. A highly informative CA/GT repeat polymorphism in intron 38 of the human neurofibromatosis type 1 (NF1) gene. *Hum Genet* 1993;**92**:429–30.
7 Xu G, Nelson L, O'Connell P, White R. An Alu polymorphism intragenic to the neurofibromatosis type 1 gene (NF1). *Nucleic Acids Res* 1991;**19**:3764.
8 Ott J. *Analysis of human genetic linkage.* Revised edition. Baltimore: The Johns Hopkins University Press, 1991.
9 Myers RM, Maniatis T, Lerman LS. Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol* 1987;**155**:501–27.
10 Terwilliger JD, Ott J. *Handbook of human genetic linkage.* Baltimore: The Johns Hopkins University Press, 1994.
11 Ward K, O'Connell P, Carey JC, *et al.* Diagnosis of neurofibromatosis I by using tightly linked, flanking DNA markers. *Am J Hum Genet* 1990;**46**:943–9.
12 Brown AHD. Sample sizes required to detect linkage disequilibrium analyses and restriction between two or three loci. *Theor Popul Biol* 1975;**8**:184–201.
13 Chakravarti A, Buetow KH, Antonarakis SE, *et al.* Non recombination within the human β-globin gene cluster. *Am J Hum Genet* 1984;**36**:1239–58.
14 Miserez AR, Schuster H, Chiodetti N, Keller U. Polymorphic haplotypes and recombination rates at the LDL receptor gene locus in subjects with and without familial hypercholesterolemia who are from different populations. *Am J Hum Genet* 1993;**52**:808–26.
15 Lázaro C, Gaona A, Ravella A, *et al.* Novel alleles, hemizygosity and deletions at an *Alu*-repeat within the neurofibromatosis type 1 (NF1) gene. *Hum Mol Genet* 1993;**2**:725–30.
16 Weber JL, Wong C. Mutation of human short tandem repeats. *Hum Mol Genet* 1993;**2**:1123–8.