The EMBO Journal vol.9 no.2 pp.505-514, 1990

An interspersed repeated sequence specific for human subtelomeric regions

F.Rouyer, A.de la Chapelle[1], M.Andersson[1] and J.Weissenbach

Unité de Recombinaison et Expression Génétique, INSERM U163, CNRS UA271, Institut Pasteur, 28 rue du Dr Roux, 75015 Paris, France and [1]University of Helsinki, Department of Medical Genetics, Haartmaninkatu 3, 00290 Helsinki, Finland

Communicated by J.Weissenbach

A family of DNA loci (DNF28) from the pseudoautosomal region of the human sex chromosomes is characterized by a repeated element (STIR: subtelomeric interspersed repeat) which detects homologous sequences in the telomeric regions of human autosomes by *in situ* hybridization. Several STIR elements from both the pseudoautosomal region and terminal parts of autosomes were cloned and sequenced. A conserved 350 bp sequence and some characteristic structural differences between the autosomal and pseudoautosomal STIRs were observed. Screening of the DNA sequence databases with a consensus sequence revealed the presence of STIRs in several human loci localized in the terminal parts of different chromosomes. We mapped single copy probes flanking the cloned autosomal STIRs to the subtelomeric parts of six different chromosomes by *in situ* hybridization and genetic linkage analysis. The linkage data show a greatly increased recombination frequency in the subtelomeric regions of the chromosomes, especially in male meiosis. The STIR elements, specifically located in subtelomeric regions, could play a role in the peculiar recombination properties of these chromosomal regions, e.g. by promoting initiation of pairing at meiosis.

Key words: genetic recombination/human telomeres/meiotic pairing/pseudoautosomal region/repeated DNA sequences

Introduction

Only few repeated sequences have been associated with biological functions in eukaryotes, but repeated sequences of generally unknown function have been mapped to specific chromosomal regions like telomeres and centromeres. These repeats usually consist of very high numbers of copies of simple sequences arranged in tandem arrays as the alphoid sequences from the centromeric regions of human chromosomes (Manuelidis, 1976). Best characterized are the (G+T)-rich repeated sequences at the telomeres of various organisms from lower eukaryotes (Blackburn, 1984) to man (Allshire *et al.*, 1988; Moyzis *et al.*, 1988), apparently used for telomere replication (reviewed in Weiner, 1988). Some more complex subtelomeric repeats are found adjacent to the (G+T)-rich sequences in yeast (Chan and Tye, 1983) or at the ends and in the centromeric regions of *Drosophila melanogaster* chromosomes (Young *et al.*, 1983). It has been suggested that these structures could play a role in the telomere-telomere or telomere-nuclear membrane interactions of meiosis. Recently, some subtelomeric repeats have been proposed to be the site of unequal crossovers responsible for chromosome length polymorphisms in *Plasmodium falciparum* (Corcoran *et al.*, 1988).

We previously reported the occurrence of repeated DNA sequences in the human pseudoautosomal region (Simmler *et al.*, 1985), which forms the terminal part of the sex chromosomes short arms. These repeats were previously called DXYZ2 and define a family of repeats hereafter called STIR. Both *in situ* and Southern blot hybridization with a pseudoautosomal STIR sequence show a localization of the repeats at the tip of the short arms of the X and Y chromosomes. On Southern blots, this probe detects under stringent conditions 30-50 hybridization bands on the Y chromosome short arm and in the terminal part (Xp22.3) of the X chromosome short arm (Simmler *et al.*, 1985; Petit *et al.*, 1990). Several clones containing STIR elements have been mapped to different parts of the pseudoautosomal region (Rouyer *et al.*, 1986a,b; Petit *et al.*, 1983), indicating that the STIR repeats are interspersed. However, other STIR repeats have been found more recently in different parts of the Y chromosome and in the non-pseudoautosomal part of Xp22.3 (Petit *et al.*, 1990). All these loci define the DNF28 family.

In this paper, we report the detection of other STIR elements at the tip of many human autosomes. Seven non-pseudoautosomal STIRs have been cloned and were assigned to the terminal cytogenetic band of six different chromosome arms by *in situ* hybridization. Computer screening of DNA sequence databases showed the presence of STIRs in the terminal part of four human chromosomes and one in an unmapped bovine locus. Genetic mapping of five of these human loci revealed a large excess of recombination in the terminal regions of the chromosomes. Furthermore, as in the pseudoautosomal and other subtelomeric regions, no female excess or an excess of male over female recombination can be observed in three subtelomeric regions analyzed. The pseudoautosomal and autosomal STIRs might be involved in similar functions related to their common location in the subtelomeric parts of human chromosomes.

Results

Detection of autosomal STIRs

Probe 708 is a 580 bp DNA fragment from the pseudoautosomal region and detects a complex hybridization pattern of STIRs on Southern blots (Simmler *et al.*, 1985). *In situ* hybridization of probe 708 on metaphase chromosome spreads shows a very high density of STIRs at the tips of the short arms of the X and Y chromosomes (Simmler *et al.*, 1985). In addition, weaker hybridization signals occur on the terminal bands of numerous autosomes. In the experiment shown in Figure 1 the most obvious peaks on autosomal

telomeric regions are seen on 12q, 8p, 9p, 22q, 19p, 8q, 1p, 2q, 4p, 5q, 6q, 13q and 15q. It might be noted that there are no peaks on the short arms of the acrocentric chromosomes. This observation suggested that a given type of repetitive sequence was specifically mapping to terminal parts of human chromosomes. In order to define this DNA sequence and to test for its specific location, we isolated and mapped autosomal STIRs and compared their sequence to pseudoautosomal elements isolated previously.

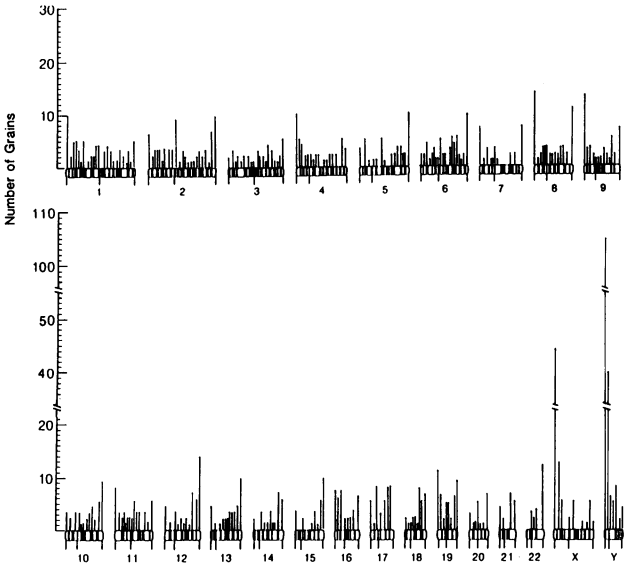

Fig. 1. *In situ* hybridization of probe 708 on metaphase spreads from a normal male. The histogram shows the distribution of a total of 1187 grains on a total of 175 metaphases.

Non-pseudoautosomal clones containing STIRs were obtained from three genomic DNA libraries (see Materials and methods) screened with either of two pseudoautosomal STIR probes, 113F (Simmler *et al.*, 1985) and 68B (Rouyer *et al.*, 1986b). After washes at moderate stringency (2 × SSC), ~110 clones were isolated. The 27 clones giving the strongest hybridization signals with STIR were characterized further by hybridization dosage on Southern blots of DNA from 49,XYYYY and 48,XXXX cell lines and with hybrids containing different parts of the human X chromosome (data not shown). Seventeen of these 27 clones turned out to be autosomal and one was from Xq. These 18 clones are designated as 'autosomal' hereafter. The nine other clones (DNF28 loci) were from the non-pseudoautosomal parts of Xp22.3 and Y (Petit *et al.*, 1990).

Sequence analysis of STIRs

The STIR fragments of four pseudoautosomal (68B, 362F, U7F and 113F) and four non-pseudoautosomal probes (IP1U12B, IP12U28, IP23K193B and IP5J201) were sequenced and compared. Dot matrix homology analyses revealed a common DNA element of ~350 bp shared by all these fragments.

Pseudoautosomal STIRs. In the four pseudoautosomal loci analyzed, the 350 bp STIR element is tandemly repeated two (dimer) or four times (tetramer) in a head-to-tail orientation. Differences at the nucleotide level, deduced from sequence alignments, allow us to distinguish two major types of STIR monomers (Figure 2). These two types of monomers (A and B) differ in their first 80 bp. Monomer A is found in three of the four pseudoautosomal loci analyzed and monomer B in all of them. 3′ to this leader sequence, a 55% (G+C)-rich core of ~270 bp is shared by both types of monomers.

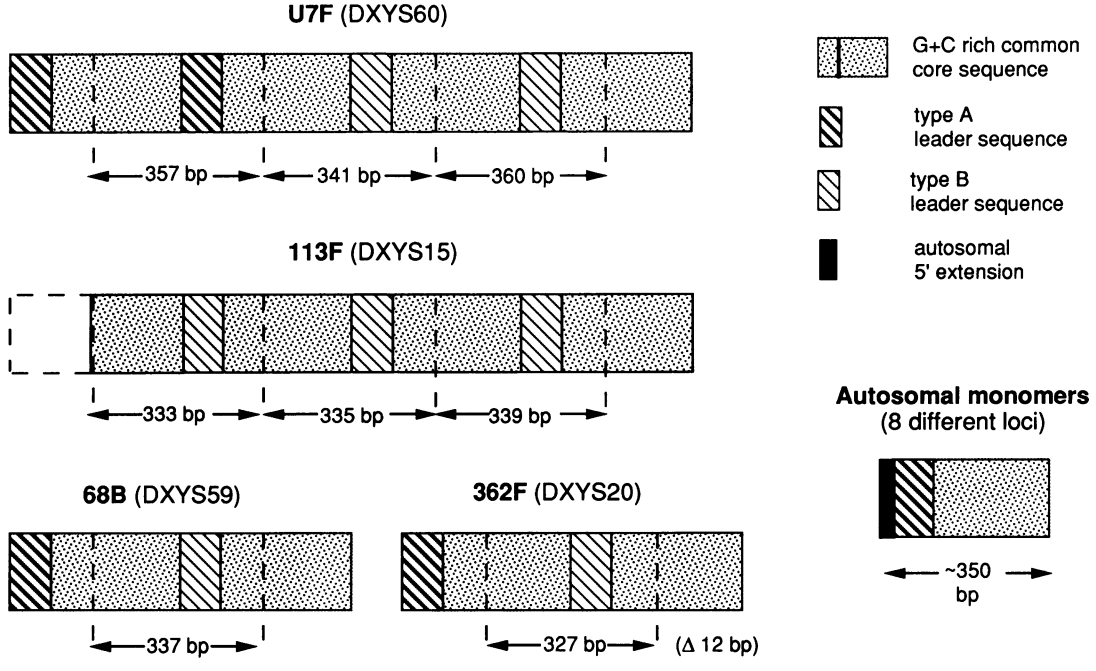

Fig. 2. Schematic representation of the pseudoautosomal and autosomal STIR sequences. Names of the DNA fragments are bold typed and corresponding loci are bracketed. Vertical dashed lines correspond to a motive arbitrarily chosen in the monomer sequence and used to define the length of each unit as indicated below between arrows. A 12 bp deletion indicated for 362F is clearly visible in Figure 3 and accounts for the shorter length of this fragment. Because the 5′ part of the first monomer of 113F is deleted (see Figure 4), the type (A or B) of this unit remains undefined. Autosomal elements are shown depicted with the 5′ extension of their leader sequence (in black).

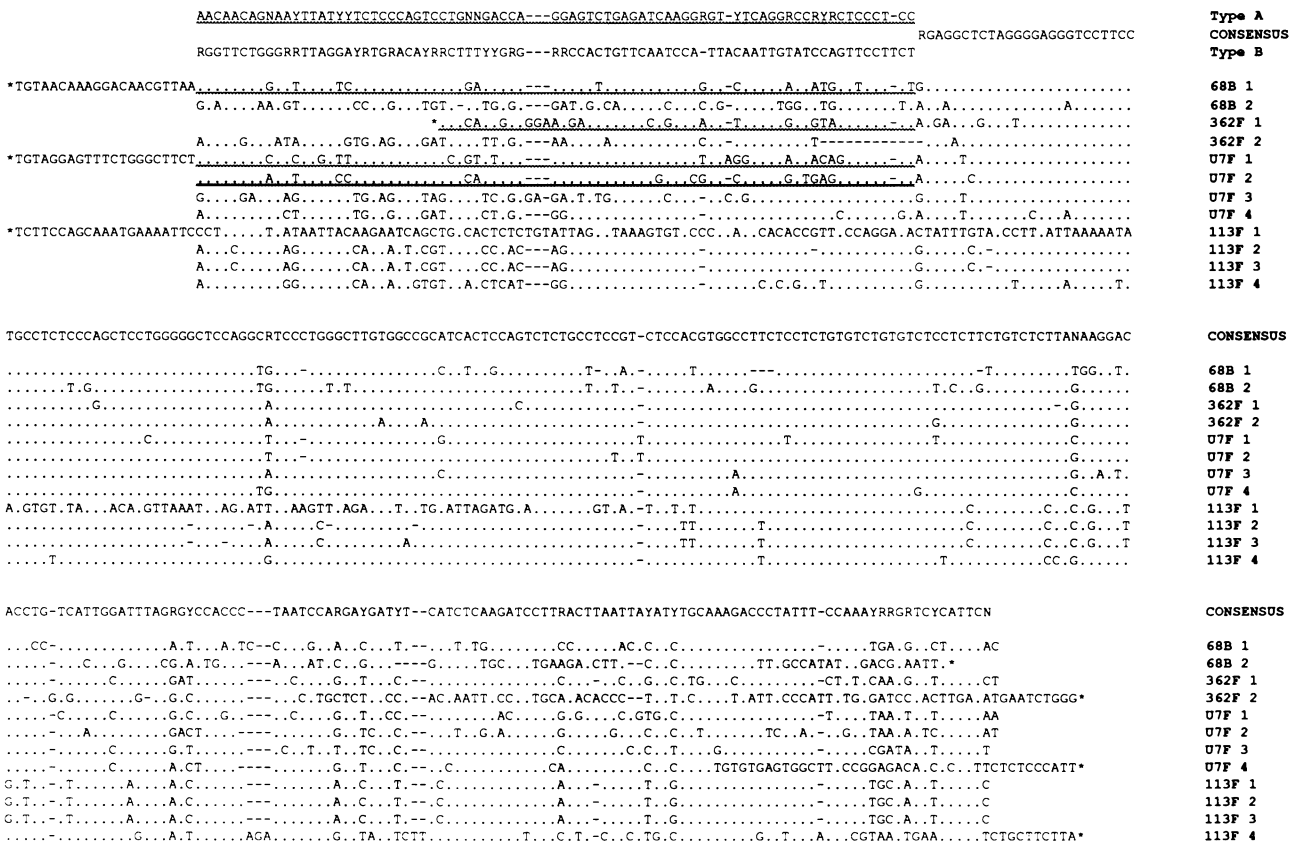

Fig. 3. Alignments of the pseudoautosomal STIR sequences. Nucleotides identical to those of the consensus sequence are replaced by dots and each dash represents a mononucleotide gap in the sequence. The multimeric loci are divided into two or four monomers and their sequences are continuous from the first to the last line. A consensus sequence is shown above and is divided in two different consensus for the 80 bp of the 5′ leader, representing the two types of STIR monomers. Type A leader sequences and corresponding consensus are underlined. Other leader sequences are of type B except for the first monomer of 113F in which the 5′ half has been deleted. The asterisks indicate the 5′ and 3′ ends of a multimeric fragment.

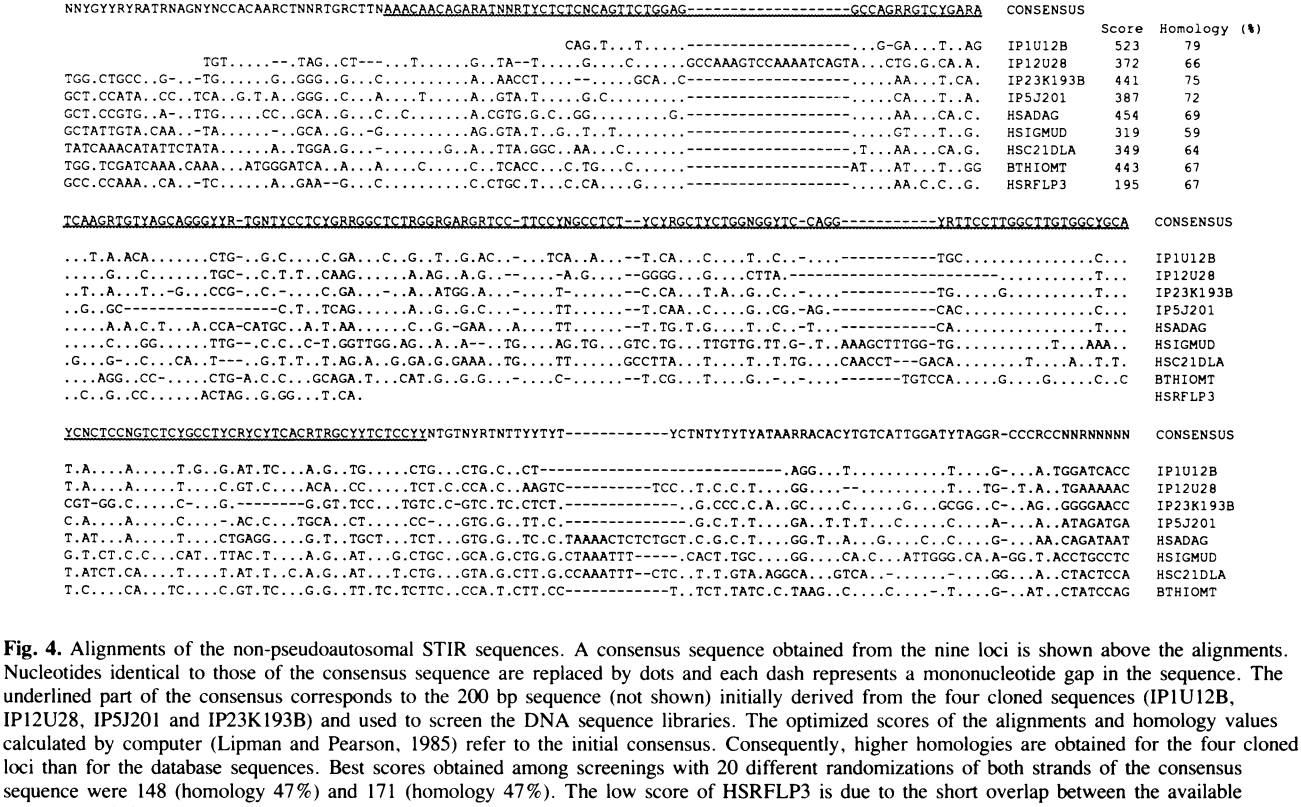

Fig. 4. Alignments of the non-pseudoautosomal STIR sequences. A consensus sequence obtained from the nine loci is shown above the alignments. Nucleotides identical to those of the consensus sequence are replaced by dots and each dash represents a mononucleotide gap in the sequence. The underlined part of the consensus corresponds to the 200 bp sequence (not shown) initially derived from the four cloned sequences (IP1U12B, IP12U28, IP5J201 and IP23K193B) and used to screen the DNA sequence libraries. The optimized scores of the alignments and homology values calculated by computer (Lipman and Pearson, 1985) refer to the initial consensus. Consequently, higher homologies are obtained for the four cloned loci than for the database sequences. Best scores obtained among screenings with 20 different randomizations of both strands of the consensus sequence were 148 (homology 47%) and 171 (homology 47%). The low score of HSRFLP3 is due to the short overlap between the available sequence and the consensus.

Comparison of the monomers within dimers shows that both dimeric loci (DXYS20 and 59) contain a STIR of type A and one of type B with cores of 80% sequence homology (Figure 3). This degree of homology is similar to that observed between cores of different pseudoautosomal loci. In all known instances A monomers occur 5′ to B monomers (Figure 2). The AABB tetramer of locus DXYS60 seems to have arisen from the duplication of each element of a former AB dimer, since the DNA sequence of the core shows a better conservation between type A monomers or type B monomers, than between an A and a B monomer (Figure 3). The four elements of locus DXYS15 are very similar, suggesting the amplification of a single type B element. Sequences at both ends of the dimeric or tetrameric structures (5′ end leader and 3′ end of 'rightmost' core) are clearly more diverged than the edges of the internal repeats (Figure 3), suggesting an internal homogenization as noted previously for minisatellite sequences (Simmler *et al.*, 1987).

Non-pseudoautosomal STIRs. In contrast to the pseudoautosomal repeats, autosomal STIRs appear monomeric (Figure 2). In each case, sequencing of the complete restriction fragment hybridizing to a STIR probe showed the presence of a single monomer related to type A. However, homology between autosomal repeats can extend up to 20 bp 5′ to the leader portion defined by the pseudoautosomal elements (Figures 2 and 4). The sequence conservation between these repeats is weaker than between the pseudoautosomal STIRs, ranging from 55 to 70%.

The four non-pseudoautosomal STIRs were aligned and a 200 bp consensus sequence corresponding to the best conserved part of the element was generated (Figure 4). This consensus sequence and its complementary strand were used to screen the EMBL and Los Alamos DNA sequence libraries with a computer program derived from Lipman and Pearson (1985). Five sequences giving significant matches were obtained in this search (Figure 4). Two-by-two sequence comparisons did not show any differences between the repeats found in the database and those obtained by hybridization. Thus, taken together, they constitute a homogeneous class of STIR sequences. The same matches were obtained with the pseudoautosomal consensus sequences derived from the complete type A or B monomers as well as the 80 bp leader sequence of type A (Figure 3). No significant matches were found with the type B pseudoautosomal leader sequence. Using the same program, the autosomal consensus sequence was aligned with the sequences picked up by hybridization and scores corresponding to the eight alignments are given in Figure 4. The significance of the alignments was evaluated by computer screening of the library with 20 randomized sequences from both strands of the autosomal consensus sequence used for the initial search (see legend of Figure 4).

Three of the five sequences obtained in the databases belong to genes but the matching regions are in non-coding parts of these genes. The STIR sequences are found in the first intron of the human adenosine deaminase gene (HSADAG), the $\mu-\delta$ intron of the human immunoglobulin heavy-chain genes (HSIGMUD) and the non-transcribed 3′ region of the bovine pineal hydroxindole *O*-methyl transferase gene (BTHIOMT). The two other STIR-containing sequences are part of the DNA probes PW231C (HSC21DLA) and G8 (HSRFLP3). This latter is genetically linked to the locus of Huntington disease. As for the autosomal probes IP1U12B, IP12U28, IP5J201 and probe IP23K193B from Xq, the sequences HSADAG, HSIGMUD, HSC21DLA and BTHIOMT contain a single type A STIR.

Cytogenetic mapping of autosomal STIRs

Seven independent subclones derived from a subset of the 19 clones characterized above were used as probes for *in situ* hybridization. Six of the seven probes highly significantly label the telomeric region of one autosome arm, while the seventh probe preferentially labels the end of Xq. Table I summarizes the relevant grain counts. A histogram of grain distribution for each probe is shown in Figure 5. It is notable that two of the probes showing preferential autosomal labelling (IP1U12, IP12U28) also label the short arm of one or both sex chromosomes to some extent. Probe IP12U28 still contains an STIR probably accounting for the labelling of the pseudoautosomal region on Xp and Yp. Moreover, several probes, especially IP1U12 and IP21K443, appear to label several telomeric regions slightly but significantly more strongly than other regions. This might indicate the existence of other repeated sequence families in these telomeric regions.

The chromosomal localization of the four human sequences detected by computer screening had been reported by others and, as the above probes were selected at random, they are located in the terminal intervals 4p16-pter (HSRFLP3 in locus D4S10) (Magenis *et al.*, 1986), 14q32.3-qter (HSIGMUD in locus IGH) (Kirsch *et al.*, 1982), 20q13.1-qter (HSADAG in locus ADA) (Philip *et al.*, 1980; Jhanwar *et al.*, 1987) and 21q22.3-qter (HSC21DLA in locus D21S3) (see Tanzi *et al.*, 1988). These results tend to show that the presence of a STIR in a DNA sequence

Table I. Grain counts on normal human male metaphases after in situ hybridization of probes from STIR-containing loci

			No. of cells with at least one grain on chromosomes						No. of grains on chromosomes							
Probe	Locus	No. of mitoses	1	5	12	20	21	X	1	5	12	20	21	X	Y	Total
IP1U12	ND	81	49						68					31	9	574
IP5J201	D5S109	59		27						41				11	6	339
IP12U28	D12S37	72			39						48			13	13	275
IP20K061	D20S25	97				55						78		7	3	362
IP20K09	D20S26	48				33						46		2	2	225
IP21K443	D21S154	94					41						53	24	4	540
IP23K193	ND	47						19						21	1	313

ND: not defined.

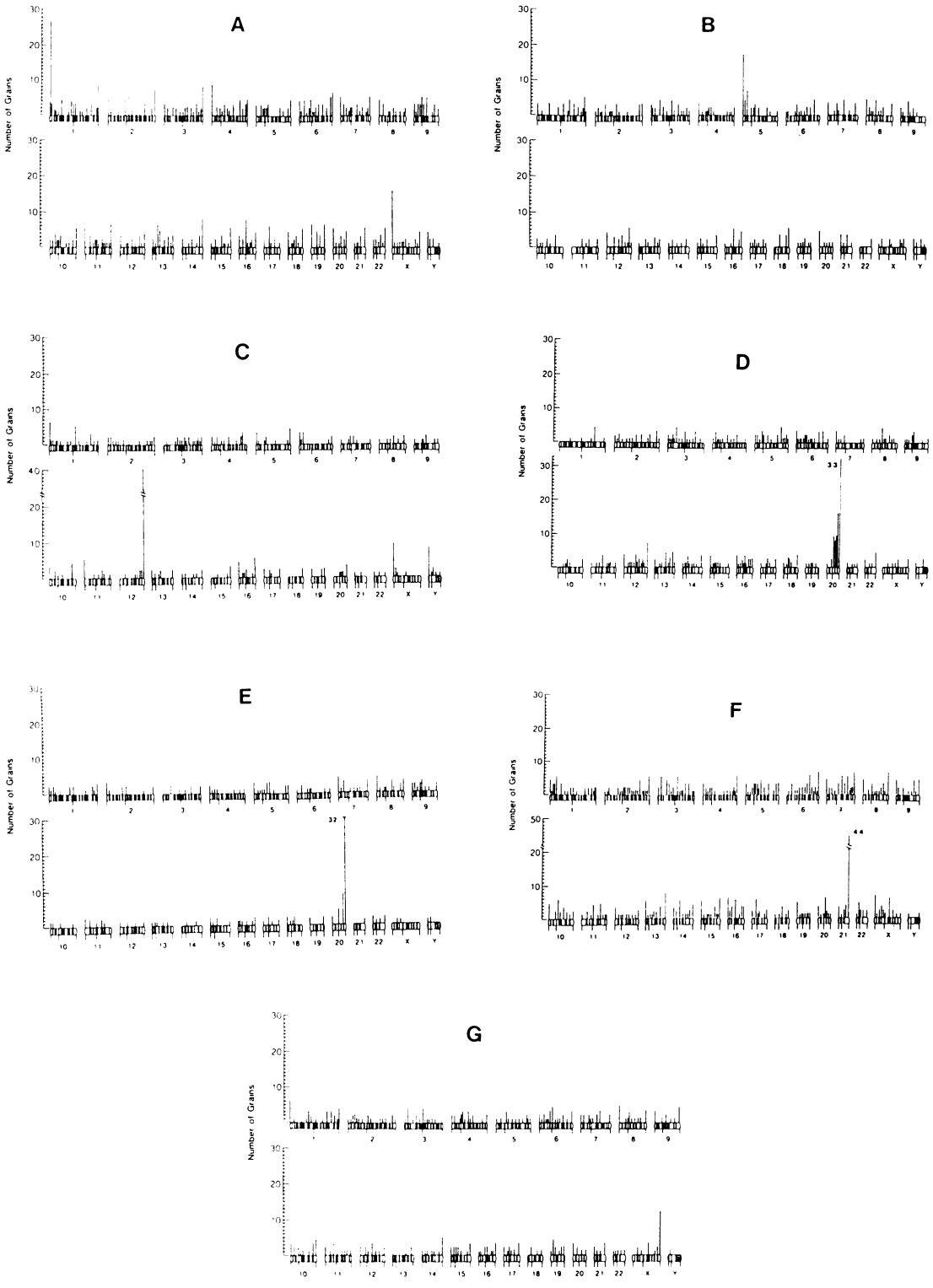

Fig. 5. Histograms of *in situ* hybridization showing the grain distributions corresponding to the data in Table I. (**A**) Probe IP1U12; (**B**) IP5J201; (**C**) probe IP12U28; (**D**) probe IP20K061; (**E**) probe IP20K09; (**F**) probe IP21K443; (**G**) probe IP23K193.

predicts a chromosomal localization in the terminal part of a chromosome. To improve the regional localization based on *in situ* hybridization we mapped the STIR-containing autosomal loci with respect to telomere by *Bal*31 digestion, and relative to other subtelomeric loci, by linkage analysis.

Bal31 exonuclease digestion

Controlled digestions with *Bal*31 exonuclease were performed on high mol. wt DNA and followed by *Bam*HI restriction digestions. After blotting, the DNAs were hybridized with six of the seven probes used for the *in situ*

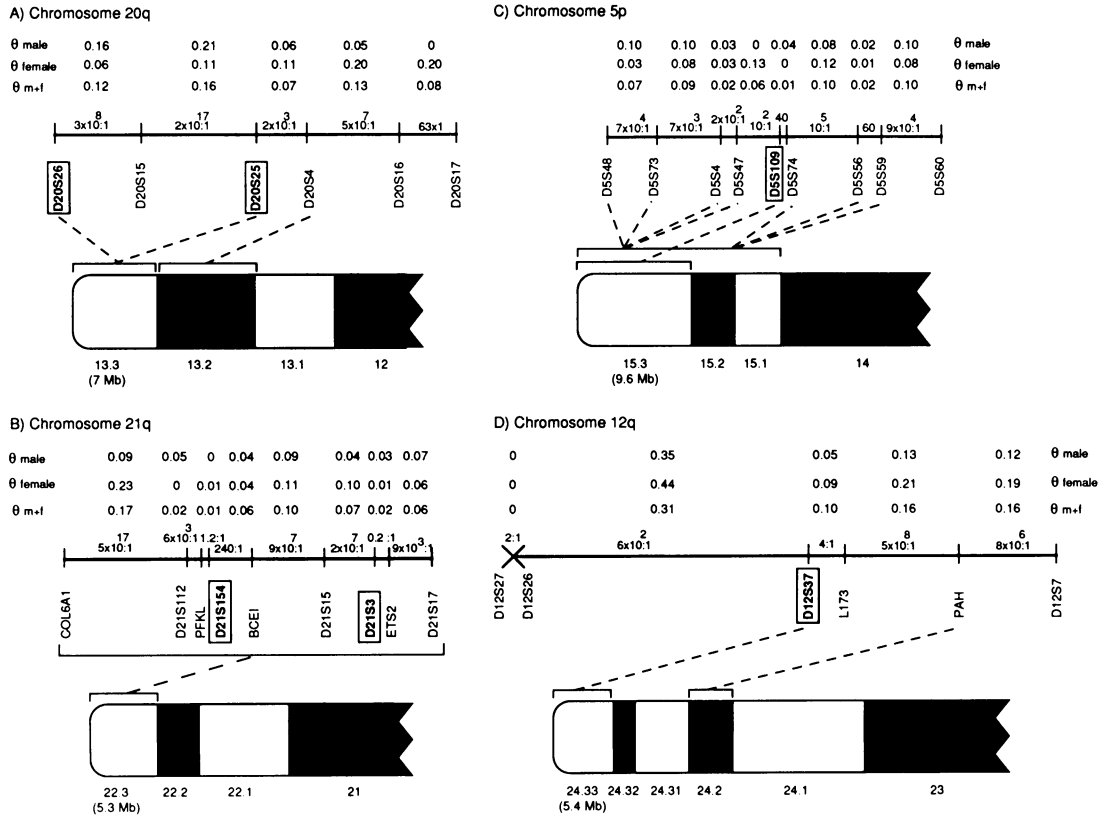

Fig. 6. Genetic and cytogenetic maps around the STIR-containing loci. Loci are represented by vertical bars on the genetic maps and the STIR-containing ones are boxed. The genetic maps are drawn according to the recombination fractions measured in male and female meioses (θ_{m+f}). For each interval, the odds corresponding to an inverted order of flanking markers are given above the map. Known physical localizations are indicated with dashed lines on the cytogenetic maps (see Materials and methods for references) and the corresponding band numbers are indicated below. A tentative size of the terminal bands in megabases was expressed as the product of the DNA content of the chromosome by the relative length of a chromosome band. DNA content of chromosomes based on the incorporation of DNA precursors was estimated by Korenberg and Engels (1978). The length ratio of the distal chromosomal band to the entire chromosome was derived from the ISCN karyograms.

hybridization experiments. As a positive control, the blots were probed with the DNA fragment 29C1 (DXYS14) which is located ~20 kb from the telomeres of the short arms of the X and Y chromosomes and sensitive to *Bal*31 digestion (Cooke *et al.*, 1985). None of the probes from the six STIR-containing loci detected any degradation of the *Bam*HI restriction fragment even after 2 h of incubation with the highest concentration of exonuclease (0.1 U/mg DNA, IBI fast form) (data not shown). Under the same digestion conditions, the *Bam*HI fragment encompassing 29C1 was completely degraded. This indicates that the STIR-containing loci tested here are >20 kb from their respective telomeres.

Genetic mapping of the autosomal STIR-containing loci

Five STIR-containing loci were found to be polymorphic (data presented in a poster at HGM10, Rouyer *et al.*, 1989) and mapped by linkage studies with markers of chromosomes 5, 12, 20 and 21 including the locus D21S3 (see Materials and methods). In each case, the STIR-containing locus was linked by two-point analyses to the distal markers of one chromosomal arm, in agreement with the *in situ* hybridization results.

The two loci from chromosome 20, D20S25 and D20S26 map distally to D20S4 assigned to 20q13.2 (see Donis-Keller *et al.*, 1987), in agreement with the *in situ* mapping (Figure 6A). D20S26 is the most distal locus on chromosome 20q known at present. Multi-point analysis gives a recombination fraction of 0.28 between the two STIR-containing loci, both assigned to band q13.3, and suggests that they are probably localized far apart on the physical map. On chromosome 21, recombination data map locus D21S154 close to PFKL and distal to D21S3, both located in the terminal band 21q22.3 (Tanzi *et al.*, 1988; Warren *et al.*, 1989a,b), confirming the *in situ* hybridization data. Figure 6(B) shows the most likely localization of D21S154 on the genetic map of 21q22.3 at 20 cM of the most distal marker COL6A1 and 23 cM distal to D21S3. The sum of recombination fractions between loci COL6A1 and D21S17 is 0.51 and represents the minimum genetic length of the terminal band 21q22.3.

Figure 6(C) and (D) show the linkage data in the terminal parts of chromosomes 5p and 12q respectively. Locus D5S109 maps between loci previously located in the region 5p15-pter (Weiffenbach *et al.*, 1989). These results would indicate that the group of four loci distal to D5S109 map to the terminal band 5p15.3 and give a minimum genetic length of 24 cM for band 5p15.3, from the telomere to J201. On chromosome 12, locus D12S37 maps distally to PAH located in 12q24.2 (see O'Connell *et al.*, 1987), in agreement with the physical mapping, and 31 cM proximally to the most distal markers. This gives a minimum genetic length of 31 cM for the terminal band 12q24.33.

Comparison of recombination fractions in male and female

Table II. Sex-specific genetic distances in the terminal parts of human chromosomes

Chromosomal region	5p15.3	9q34	10p	10q	11p15.5	12p13	12q24.33	14q32.3	16p13	17q13-qter	17q23-pter	19p13.2	20q13.3	21q22.3	pseudoautosomal region
No. of markers	5	10	3	3	3	5	3	2	4	4	6	3	3	4	4
$\Sigma\ \theta$ male	0.2	0.46	0.3	0.18	0.17	0.28	0.35	0.15	0.17	0.29	0.49	0.35	0.37	0.41	0.49
$\Sigma\ \theta$ female	0.11	0.4	0.11	0.03	0.05	0.12	0.44	0.06	0.02	0.13	0.13	0.1	0.17	0.56	0.03
θ_f/θ_m[a]	1	0.9	0.4	0.2	0.3	0.4	1	0.4	0.1	0.5	0.3	0.3	0.5	1.3	0.06
Same ratio for the whole chromosome	1.8	1.4		1.4	1.3	2.5	1.4[b]	1	1.4	1.3		1.4	1.7[b]	2.2[b]	–

The most distal loci available from each map have been taken into account in this table. The number of markers from the telomere is arbitrarly defined and $\Sigma\ \theta$ is the sum of the recombination fractions in male and female meioses for the defined interval. Linkage data are drawn from the study for chromosomes 5, 12, 20 and 21 and from Lathrop *et al.* (1988), Nakamura *et al.* (1988a), Holm *et al.* (1987), Nakamura *et al.* (1989), Reeders *et al.* (1988), Nakamura *et al.* (1988b), Nakamura *et al.* (1988c) and Rouyer *et al.* (1986b) for the chromosomes 9, 10, 14, 16, 17, 19 and sex chromosomes respectively.

[a]The θ_f/θ_m ratios are given by CILINK for 5p, 12q, 20q and 21q and thus do not always reflect the ratio of $\Sigma\ \theta$ indicated above. For the other chromosomes, the ratios are calculated by hand with the $\Sigma\ \theta$ values of the table. The whole chromosome ratios are derived from Donis-Keller *et al.* (1987) for chromosomes 5, 11, 14, 16 and 20 and from O'Connell *et al.* (1987) for chromosomes 12.

[b]For chromosomes 12, 20 and 21 this ratio is based for the whole chromosome except the terminal bands 12p13, 20q13.3 and 21q22.3 respectively.

meioses were made for the terminal parts of chromosomes 5p, 20q and 21q for which a sufficient number of markers was available. For the three loci assigned to 20q13.3, the model assuming a constant ratio (0.5) of female/male map distances was more likely than a model assuming no sex differences ($0.5 < P < 0.1$). Conversely, the genetic map of the rest of the chromosome, from D20S4 (20q13.2) to D20S5 on 20p (see Donis−Keller *et al.*, 1987) shows a female/male constant ratio of 6.8 ($P < 0.001$). For chromosome 21, no constant sex difference in map distances can be assessed in the terminal portion, whereas a 2.2 ratio had been reported for the other part of chromosome 21, from D21S82 to D21S13 (Warren *et al.*, 1989a). The same results are observed for the distal band 5p15.3 (from D5S109 to the telomere), whereas a constant female/male ratio of 1.6 has been found for the entire genetic map of chromosome 5 (Leppert *et al.*, 1987). One chromosome 12q, no sex differences can be assessed to the most distal band from D12S37 to D12S27. A 1.4 female/male ratio has been estimated for the rest of the genetic map, excluding the tip of the short arm (O'Connell *et al.*, 1987). A comparison between male and female genetic distances is given in Table II, which includes data for chromosomes 9, 10, 11, 12, 14, 16, 17 and 19 from the literature.

Discussion

The sequences of four pseudoautosomal and nine autosomal STIRs have been determined and show structural differences. The multimeric organization of pseudoautosomal STIRs contrasts with the monomeric form of autosomal repeats. Whereas two different types of STIR monomers (A and B) have been found in pseudoautosomal loci, the eight complete autosomal STIR sequences are all constituted of a slightly longer monomer, related to type A exclusively. The pseudoautosomal STIRs represent a more complex class characterized by the presence of at least one B monomer, in each of the four loci analyzed. However, additional sequence analyses are required to extend this observation to pseudoautosomal STIRs in general and to the non-pseudoautosomal repeats from Xp22.3. Indeed it is highly likely that these latter are remnants of a larger ancestral pseudoautosomal region which included more proximal loci of Xp22.3 (see Weissenbach *et al.*, 1989). It is interesting to note that probe IP23K193B from Xq28-qter contains an autosomal type STIR, indicating that the long arm telomere of the X chromosome is structurally closer to an autosomal telomere than to that of Xp. Altogether these observations suggest that pseudoautosomal elements might be involved in a different or additional function compared to the autosomal repeats.

In situ hybridization of the probes from the seven STIR-containing loci shows significant labelling of the telomeric part of a chromosomal arm. Similarly, the four human STIRs found in the DNA sequence databases are also located in terminal parts of autosome arms. The presence of a STIR element in any human DNA sequence might predict a terminal chromosomal location of this DNA sequence. Therefore we favour the most distal of the two conflicting localizations proposed on chromosome 20q for the ADA locus (Philip *et al.*, 1980; Jhanwar *et al.*, 1987). The mapping deduced from the linkage studies of the five polymorphic STIR-containing loci are in complete agreement with the *in situ* results. Linkage analysis localizes five of the six STIR-containing loci proximal to one or several other mapped loci, which is consistent with a subtelomeric localization of the STIR sequences. Moreover, the *Bal*31 digestion experiments clearly exclude the autosomal STIRs from the very tip of the chromosomes.

Though no physical maps are available to give an accurate sizing of the distal chromosomal bands, a rough estimate of their size in kilobases can be made by measuring their length relative to whole chromosomes on quinacrine-stained preparations (see legend of Figure 6). On the other hand, our linkage studies indicate genetic distances of 24, 31, 28 and 50 cM between the most distant markers in bands 5p15.3, 12q24.33, 20q13.3 and 21q22.3 respectively. Estimated physical distance of the terminal bands have been compared to the above linkage distances (Figure 6). Assuming a mean value of 1 cM for 1000 kb in the human genome, it appears that bands 5p15.3, 12q24.33, 20q13.3 and 21q22.3 show expansions of their genetic map of about 2.5-, 5.7, 4- and 9.6-fold respectively. Moreover, the genetic distances of these terminal regions are underestimates,

subject to increase as new distal markers are isolated. A genetic map expansion has also been observed in other terminal regions of chromosomes, namely 21q22.3 (Tanzi *et al.*, 1988; Warren *et al.*, 1989a) and 12q32 (Nakamura *et al.*, 1989). Altogether, these observations show that recombination excess in terminal regions of chromosomes is a general feature of human meiosis.

Electron microscopy studies of human male meiosis show that both recombination nodules (Rasmussen and Holm, 1978; Solari, 1980) and chiasmata (Laurie and Hultén, 1985) occur preferentially in the terminal regions of the chromosomes. Though an overall excess of female recombination of 1.9 has been noted for the human gemone (Donis-Keller *et al.*, 1987), the distal parts of 5p, 12q and 21q do not show any sex differences in this study and a male/female recombination excess of 2 is observed in the terminal region of 20q. Published genetic linkage maps show expansions in male meiosis for the terminal regions of 10p and 10q (Nakamura *et al.*, 1988a), 11p (Holm *et al.*, 1987), 12p (O'Connell *et al.*, 1987), 14q (Nakamura *et al.*, 1989), 16p (Reeders *et al.*, 1988), 17p/q (Nakamura *et al.*, 1988b) and 19p (Nakamura *et al.*, 1988c). An even more dramatic increase has been observed in the pseudoautosomal region, where male and female genetic lengths are respectively 50 and 3 cM (Goodfellow *et al.*, 1986; Rouyer *et al.*, 1986a,b; Page *et al.*, 1987). In contrast with other chromosomal ends, the pseudoautosomal region displays no map expansion in female meiosis (Brown, 1988; Petit *et al.*, 1988).

We propose that the excess of recombination observed in the terminal regions of human chromosomes could be a simple consequence of the pairing process of homologs at meiosis. In spermatocytes, pairing and formation of the central component of the synaptonemal complex during zygotene stage of prophase starts from the chromosome telomeres attached to the inner nuclear membrane and extends along the bivalent. The recombination nodules assumed to be the sites of cross-overs appear very early, as the chromosomal pairing progresses from the ends towards the centromere (Rasmussen and Holm, 1978). If recombination takes place as soon as these structures are formed, a cross-over excess will be observed in the terminal regions of chromosomes as a consequence of their very early pairing. In oocytes pairing of homologs also starts from the telomeres but interstitial pairing occurs frequently unlike male meiosis (Bojko, 1983). The more random initiation of pairing could explain the lower genetic map expansion observed in the distal ends of chromosomes in female meiosis compared to male meiosis. The primer role of chromosomal ends in the recognition between homologs and initiation of pairing suggests that they are directly involved in these meiotic processes. The STIR sequences may be involved in such mechanisms. Interestingly we did not detect any hybridization of STIR probes to the terminal parts of short arms of the acrocentric chromosomes. These short arms carry the nucleolus organizer and do not pair from their own telomeres but by extension of the pairing initiated at the end of the long arms (Rasmussen and Holm, 1978). Furthermore, mouse chromosomes which do not contain STIR sequences initiate pairing interstitially in a preferential manner (Guitart *et al.*, 1985).

If STIR sequences are involved in recombination, a difference in recombination excess between two subtelomeric regions should be paralleled by a difference in density of STIRs in these same regions. Estimates of copy numbers of STIRs in the subtelomeric regions cannot be inferred from Southern blot or *in situ* hybridizations because of the sequence divergence. Three distinct elements have been observed for the terminal part of 20q and two for the terminal band 21q22.3, indicating the occurrence of multiple STIRs on a single chromosomal end. Blots of DNA digested with rare cutter restriction enzymes were hybridized to probes from two STIR-containing loci of 20q13.3. None of the large restriction fragments detected (ranging from 100 to 600 kb) was common to both probes, suggesting a rather large distance between these loci. Similarly, a 1500 kb *Not*I fragment is recognized by D21S15 and not by D21S3 and BCEI (Gardiner *et al.*, 1988), providing a minimum estimation of the distance separating the two STIR loci D21S3 and D21S154 on chromosome 21q22.3. These data are consistent with the genetic linkage results, which support a distribution of the repeats over several thousand kilobases from the 20q telomere. It should be noted that STIRs are interspersed in the pseudoautosomal region (Simmler *et al.*, 1985) within a terminal DNA segment of 2600 kb (Brown, 1988; Petit *et al.*, 1988).

Strikingly, we found that two STIR-containing loci were closely associated to the breakpoints of two chromosomal rearrangements. The pseudoautosomal locus DXYS60 is the site on the X chromosome of an abnormal X−Y interchange occurring in a human XX male (Rouyer *et al.*, 1987). Locus D21S3 is the junction site on the distal long arm of a ring chromosome 21 (Wong *et al.*, 1989). The STIR sequences are in the same orientation, at a distance of 500 and 800 bp from the breakpoints, but towards the centromere of the X chromosome and the telomere of 21q respectively. This could suggest that STIR sequences can act as recombination enhancers. It should be noted that both many STIR elements and breakpoints resulting from abnormal X−Y interchanges occur in Xp22.3 (Petit *et al.*, 1990).

The occurrence of a family of interspersed repeated elements localized in specific chromosomal portions raises the question of their origin and spreading. STIR sequences have been found by Southern blotting in most mammals (B. Weber and W.Schempp, personal communication) with the notable exception of rodents (Simmler *et al.*, 1985). No STIR can be detected in the murine homolog of the μ-δ intron of the immunoglobulin heavy-chain gene. Conservation of STIRs among a number of mammals argues in favor of a biological function. This hypothesis could be strengthened if mapping studies could establish a similar localization on terminal parts of chromosomes in other mammals. The STIR sequence present in the bovine HIOMT gene allows such a subtelomeric localization in this species to be tested for.

Materials and methods

Genomic cloning

Three genomic DNA libraries were screened for autosomal STIR sequences. The first one was from a 46,XX human DNA digested with *Hin*dIII, sized on sucrose gradient and ligated in lambda 47.1 *Hin*dIII arms. The two other libraries were constructed with DNA from two human/hamster somatic hybrid cell lines KICH6 and JUCH1 (Petit *et al.*, 1990). In addition to a non-characterized set of human autosomes, the human sex chromosomes of these somatic hybrids were lost or deleted for their pseudoautosomal region (Petit *et al.*, 1990). DNAs were digested with *Sau*3AI and ligated with lambda EMBL3 *Bam*HI arms. *In vitro* packaging was done with Gigapack (Stratagene) or Amersham kits, phages were plated on *Escherichia coli* strains LA101 or P2392 and transferred onto Hybond-N filters (Amersham). Probe

labelling and hybridizations were made as for the Southern blots (see below). About 5×10^5 phages were screened with probe 113F, and 10^6 with probe 68B, giving a total of 110 STIR positive clones.

DNA probes

Probes were isolated by subcloning various restriction fragments of the lambda inserts free of human repetitive DNA. Probes IP1U12 and IP1U12B are *Hin*dIII−*Eco*RI and *Eco*RI fragments of 5 and 2 kb respectively. Probe IP12U28 is a 5 kb *Hin*dIII fragment. These three probes are subcloned in pUC8. The five other probes are subcloned in pBS or pBS+ vectors (Stratagene). IP5J201, IP20K09, IP23K193 and IP23K193B are *Sal*I fragments of 5, 12, 13 and 3 kb respectively. IP20K061 is a 0.8 kb *Hpa*II fragment and IP21K443 is a 0.9 kb *Sau*3A fragment.

In situ hybridization

In situ hybridization was carried out as previously described (Andersson *et al.*, 1988). Metaphase cells were obtained from phytohemagglutinin-stimulated 3 day cultures of whole blood from normal males. Slides were incubated with ^{3}H-labelled probe DNA (sp. ac. $0.3-1.5 \times 10^7$ c.p.m./μg; concentration 30−100 ng/ml) at 42°C in 50% formamide, 2 × SSC, 10 × Denhardt's solution, 0.01% denatured salmon sperm DNA and 10% dextran sulfate. The slides were washed three times for 2 min each at 39°C in 50% formamide, 2 × SSC and developed after 6−50 days of exposure. Metaphase chromosomes were first stained with 0.25% Wright's stain and photographed. Destaining, trypsinization and G-banding were then performed (Andersson *et al.*, 1988).

Southern blot hybridizations

Genomic DNAs were obtained from the Centre d'Etude du Polymorphisme Humain (CEPH) in Paris. DNAs were digested, electrophoresed, blotted and hybridized according to standard procedures. Restriction enzymes were from Appligène. The nylon membrane used was Hybond-N (Amersham). Hybridization probes were labelled by nick-translation or random priming with ^{32}P-labelled nucleotides and labelling kits from Amersham. Blots were washed at 68°C at stringencies varying from 2 × SSC to 0.1 × SSC (according to the probe) with 0.1% SDS, and were exposed with two intensifying screens for 1−8 days at −70°C.

Linkage studies

The segregation analysis was done with the large kindreds of the CEPH for each of the five probes which have been included in the chromosomal sets of markers of the CEPH database version 3. Genetic linkage analyses were performed with the program package LINKAGE (Lathrop *et al.*, 1985) version 4.9. First two-point analyses with various markers of each chromosome were performed with CLODSCORE, allowing subset of markers linked to the STIR-containing loci to be determined. Multi-point analyses were then performed to construct genetic maps of the terminal regions of chromosomes 5p, 12q, 20q and 21q. For chromosomes 20q (loci D20S25 and D20S26) and 21q (locus D21S154) the loci were added with the CMAP program to previously published maps (Donis-Keller *et al.*, 1987; Warren *et al.*, 1989b respectively). The recombination fractions between ordered loci of the new maps were estimated with CILINK. For chromosome 5p, the two most distal loci of Leppert *et al.* (1987) (D5S4 and D5S74) and the locus D5S109 were added to the map established by Donis-Keller *et al.* (1987). For chromosome 12q, the two most distal loci of Donis-Keller *et al.* (1987) (D12S26 and D12S27) and the locus D12S37 were added to the chromosome 12q map established by O'Connell *et al.* (1987). The occurrence of sex differences in recombination was tested with the CILINK program.

Acknowledgements

We thank the Centre d'Etude du Polymorphisme Humain (CEPH) for the family DNA samples; P.Cartwright, C.Julier, M.Lathrop and W.Saurin for their help in the computer analyses; P.Stapleton for providing the unpublished part of the HSRFLP3 sequence; C.Johnsson and S.Compain for technical assistance; and P.Tiollais for support and encouragement. This work was supported by the Ministère de la Recherche (grant 88.C.0061), the Association Française contre les Myopathies, the Sigried Juselius Foundation, the Academy of Finland and the Folkhälsan Institute of Genetics.

References

Allshire,R.C., Gosden,J.R., Cross,S.H., Cranston,G., Rout,D., Sugawara,N., Szostak,J.W., Fantes,P.A. and Hastie,N.D. (1988) *Nature*, **332**, 656−659.

Andersson,M., Page,D.C., Pettay,D., Subrt,I., Turleau,C., de Grouchy,J. and de la Chapelle,A. (1988) *Hum. Genet.*, **79**, 2−7.

Blackburn,E.H. (1984) *Cell*, **37**, 7−8.

Bojko,M. (1983) *Carlsberg Res. Commun.*, **48**, 457−483.

Brown,W.R.A. (1988) *EMBO J.*, **7**, 2377−2385.

Chan,C.S.M. and Tye,B.K. (1983) *Cell*, **33**, 563−573.

Cooke,H.J., Brown,W.R.A. and Rappold,G.A. (1985) *Nature*, **317**, 687−692.

Corcoran,L.M., Thompson,J.K., Walliker,D. and Kemp,D.J. (1988) *Cell*, **53**, 807−813.

Donis-Keller,H., Green,P., Helms,C., Cartinhour,S., Weiffenbach,B., Stephens,K., Keith,T.P., Bowden,D.W., Smith,D.R., Lander,E.S., Botstein,D., Akots,G., Rediker,K.S., Gravius,T., Brown,V.A., Rising,M.B., Parker,C., Powers,J.A., Watt,D.E., Kauffman,E.R., Bricker,A., Phipps,P., Muller-Kahle,H., Fulton,T.R., Ng,S., Schumm,J.W., Braman,J.C., Knowlton,R.G., Barker,D.F., Crooks,S.M., Lincoln,S.E., Daly,M.J. and Abrahamson,J. (1987) *Cell*, **51**, 319−337.

Gardiner,K., Watkins,P., Münke,M., Drabkin,H., Jones, C. and Patterson,D. (1988) *Somat. Cell Genet.*, **14**, 623−638.

Goodfellow,P.J., Darling,S.M, Thomas,N.S. and Goodfellow,P.N. (1986) *Science*, **243**, 740−743.

Guitart,M., Coll,M.D., Ponsà,M. and Egozcue,J. (1985) *Genetica*, **67**, 21−30.

Holm,T., Lalouel,J.M., White,R., O'Connell,P., Leppert,M., Julier,C. and Lathrop,M. (1987) *Cytogenet. Cell Genet.*, **46**, 630.

Jhanwar,S.C., Berkvens,T.M., Meera Khan,P., Valerio,D. and Breukel,C. (1987) *Cytogenet. Cell Genet.*, **46**, 634.

Kirsch,I.R., Morton,C.C., Nakahara,K. and Leder,P. (1982) *Science*, **216**, 301−303.

Korenberg,J.R. and Engels,W.R. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 3382−3386.

Lathrop,G.M., Lalouel,J.-M., Julier,C. and Ott,J. (1985) *Am. J. Hum. Genet.*, **37**, 482−498.

Lathrop,M., Nakamura,Y., O'Connell,P., Leppert,M., Woodward,S., Lalouel,J.M. and White,R. (1988) *Genomics*, **3**, 361−366.

Laurie,D.A. and Hultén,M.A. (1985) *Ann. Hum. Genet.*, **49**, 203−214.

Leppert,M., Dobbs,M., Scambler,P., O'Connell,P., Nakamura,Y., Stauffer,D., Woodward,S., Burt,R., Hughes,J., Gardner,E., Lathrop,M., Wasmuth,J., Lalouel,J.-M. and White,R. (1987) *Science*, **238**, 1411−1413.

Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435−1441.

Magenis,R.E., Gusella,J., Weliky,K., Olson,S., Haight,G., Toth-Fejel,S. and Sheehy,R. (1986) *Am. J. Hum. Genet.*, **39**, 383−391.

Manuelidis,L. (1976) *Nucleic Acids Res.*, **3**, 3063−3076.

Moyzis,R.K, Buckingham,J.M., Cram,L.C., Dani,M., Deaven,L.L., Jones,M.D., Meyne,J., Ratliff,R.L. and Wu,J.R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6622−6626.

Nakamura,Y., Lathrop,M., Bragg,T., Leppert,M., O'Connell,P., Jones,C., Lalouel,J.-M. and White,R. (1988a) *Genomics*, **3**, 389−392.

Nakamura,Y., Lathrop,M., O'Connell,P., Leppert,M., Baker,D., Wright,E., Skolnick,M., Kondoleon,S., Litt,M., Lalouel,J.-M. and White,R. (1988b) *Genomics*, **3**, 302−309.

Nakamura,Y., Lathrop,M., O'Connell,P., Leppert,M., Lalouel,J.-M. and White,R. (1988c) *Genomics*, **3**, 67−71.

Nakamura,Y., Lathrop,M., O'Connell,P., Leppert,M., Kamboh,M.I., Lalouel,J.-M. and White,R. (1989) *Genomics*, **4**, 76−81.

O'Connell,P., Lathrop,G.M., Law,M., Leppert,M., Nakamura,Y., Hoff,M., Kumlin,E., Thomas,W., Elsner,T., Ballard,L., Goodman,P., Azen,E., Sadler,J., Cai,G., Lalouel,J.-M. and White,R. (1987) *Genomics*, **1**, 93−102.

Page,D.C., Bieker,K., Brown,L.G., Hinton,S., Leppert,M., Lalouel,J.M., Lathrop,M., Nystrom-Lahti,M., de la Chapelle,A. and White,R. (1987) *Genomics*, **1**, 243−256.

Petit,C., Levilliers,J. and Weissenbach,J. (1988) *EMBO J.*, **7**, 2369−2376.

Petit,C., Levilliers,J., Rouyer,F., Simmler,M.-C., Herouin,E. and Weissenbach,J. (1990) *Genomics*, in press.

Philip,T., Lenoir,G., Rolland,M.O., Phillip,I., Hamet,M., Lauras,B. and Fraisse,J. (1980) *Cytogenet. Cell Genet.*, **27**, 187−189.

Rasmussen,S.W. and Holm,P.B. (1978) *Carlsberg Res. Commun.*, **43**, 275−327.

Reeders,S.T., Keith,T., Green,P., Germino,G.G., Barton,N.J., Lehmann,O.J., Brown,V.A., Phipps,P., Morgan,J., Bear,J.C. and Parfrey,P. (1988) *Genomics*, **3**, 150−155.

Rouyer,F., Simmler,M.-C., Johnsson,C., Vergnaud,G., Cooke,H.J. and Weissenbach,J. (1986a) *Nature*, **319**, 291−295.

Rouyer,F., Simmler,M.-C., Vergnaud,G., Johnsson,C., Levilliers,J.,

Petit,C. and Weissenbach,J. (1986b) *Cold Spring Harbor Symp. Quant. Biol.*, **51**, 221–228.
Rouyer,F., Simmler,M-C., Page,D.C. and Weissenbach,J. (1987) *Cell*, **51**, 417–425.
Rouyer,F., Weissenbach,J., Andersson,M. and de la Chapelle,A. (1989) *Cytogenet. Cell Genet.*, **51**, 1070.
Simmler,M.-C., Rouyer,F., Vergnaud,G., Nyström-Lahti,M., Ngo,K.Y., de la Chapelle,A. and Weissenbach,J. (1985) *Nature*, **317**, 692–697.
Simmler,M.-C., Johnsson,C., Petit,C., Rouyer,F., Vergnaud,G. and Weissenbach,J. (1987) *EMBO J.*, **6**, 963–969.
Solari,A.J. (1980) *Chromosoma*, **81**, 307–314.
Tanzi,R.E., Haines,L.H., Watkins,P.C., Stewart,G.D., Wallace,M.R., Hallewell,R., Wong,C., Wexler,N.S., Conneally,P.M. and Gusella,J.F. (1988) *Genomics*, **3**, 129–136.
Warren,A.C., Slaugenhaupt,S.A., Lewis,J.G., Chakravarti,A. and Antonorakis,S.E. (1989a) *Genomics*, **4**, 579–591.
Warren,A.C., Slaugenhaupt,S.A., Lewis,J.G., Chakravarti,A. and Antonorakis,S.E. (1989b) *Cytogenet. Cell Genet.*, **51**, 1102.
Weiner,A.M. (1988) *Cell*, **52**, 155–157.
Weiffenbach,B., Falls,K., Green,P., Shute,N., Keith,T. and Donis-Keller,H. (1989) *Cytogenet. Cell Genet.*, **51**, 1104.
Weissenbach,J., Goodfellow,P.N. and Smith,K.D. (1989) *Cytogenet. Cell Genet.*, **51**, 438–449.
Wong,C., Kazazian,H.H.Jr, Stetten,G., Earnshow,W.C., Van Keuren,M.L. and Antonarakis,S.E. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1914–1918.
Young,B.S., Pession,A., Traverse,K.L., French,C. and Pardue,M.L. (1983) *Cell*, **34**, 85–94.

Received on October 27, 1989