Hum Genet (2000) 107 : 127–131
Digital Object Identifier (DOI) 10.1007/s004390000343

ORIGINAL INVESTIGATION

**Michel Guipponi · Marie-Laure Yaspo
Lisa Riesselman · Haiming Chen · Albertina De Sario
Gérard Roizès · Stylianos E. Antonarakis**

Genomic structure of a copy of the human TPTE gene which encompasses 87 kb on the short arm of chromosome 21

Received: 4 May 2000 / Accepted: 9 June 2000 / Published online: 1 August 2000

Abstract The testis-expressed human TPTE is a putative transmembrane tyrosine phosphatase, probably involved in signal transduction pathways of the endocrine and/or the spermatogenetic function of the testis. TPTE was mapped to the pericentromeric region of human chromosomes 21 and 13, and to chromosomes 15, 22, and Y. It is unknown which of the TPTE copies are transcribed, contain intronic sequences, and/or have open reading frames. Here, in silico analysis of the genomic sequence of human chromosome 21 allowed the determination of the genomic structure of a copy of the TPTE gene. This copy consists of 24 exons and spans approximately 87 kb. The mapping position of this copy of TPTE on the short arm of chromosome 21 was confirmed by FISH using the BAC 15L0C0 clone as a probe that contains almost the entire TPTE gene. This is the first description of the genomic sequence of a non-RNR gene on the short arm of human acrocentric chromosomes.

Introduction

The short arms of the five human acrocentric chromosomes, 13p, 14p, 15p, 21p, and 22p are mainly composed of different types of repetitive elements and RNR gene cluster (Choo et al. 1989; Greig et al. 1993; Henderson et al. 1973; Vissel et al. 1992). These short arms share extensive homology among the different acrocentric chromosomes (Van Camp et al. 1992). It is believed that 21p and the p-arms of all other acrocentrics do not contain single copy expressed genes other than RNR.

We have recently identified a cDNA, named TPTE (Transmembrane Phosphatase with Tensin homology), encoding a predicted protein of 551 amino acids with potential transmembrane domains, a tyrosine phosphatase motif and sequence homology to several phosphatases including the tumor suppresser PTEN (Chen et al. 1999). The gene is strongly expressed exclusively in testis. There are copies of TPTE sequences on the pericentromeric regions of chromosomes 21, 13, 15, 22, and the Y chromosome (Chen et al. 1999). The total copy number of the TPTE genes and potential pseudogenes is polymorphic but unknown. Furthermore, it is unknown how many of TPTE gene copies are transcribed and if there is chromosome-specific expression. Finally, it is not clear whether all TPTE copies contain introns, and their reading frames are all open. One copy of TPTE has been assigned by physical mapping to chromosome 21p in a YAC contig between markers D21S237 and D21Z4 (Wang et al. 1999). The advances of the sequencing of human chromosomes will provide a better understanding of the different TPTE copies and help to distinguish expressed genes from the pseudogenes.

We report here the genomic structure, the sequence, and the confirmation of the mapping of a copy of the human TPTE gene on the short arm of chromosome 21 that is likely to be expressed.

M. Guipponi · H. Chen · S. E. Antonarakis (✉)
Division of Medical Genetics,
University of Geneva Medical School, Geneva, Switzerland
e-mail: Stylianos.Antonarakis@medecine.unige.ch,
Tel.: +41-22-7025708, Fax: +41-22-7025706

S. E. Antonarakis
University and Cantonal Hospital,
1 rue Michel Servet 1211 Geneva 4, Switzerland

M.-L. Yaspo · L. Riesselman
Max-Planck-Institut fur Molekulare Genetik,
D-14195 Berlin-Dahlem, Germany

A. De Sario · G. Roizès
Séquences Répétées et Centromères Humains,
CNRS UPR 1142, Institut de Biologie, Montpellier, France

Materials and methods

In silico analysis of human chromosome 21 genomic DNA sequence

The chromosome 21 sequencing group in the Department of Genome Analysis of the GBF Institute in Braunschweig Germany (http://genome.gbf.de/seqproj.html#Human21) determined the nucleotide sequences of BACs B7L1C4 (106.7 kb; Genbank No. AL078476) and B15L0C0 (80.6 kb; Genbank No. AL078471), and a contig (Genbank No. NT_002217) of approximately 144.3 kb

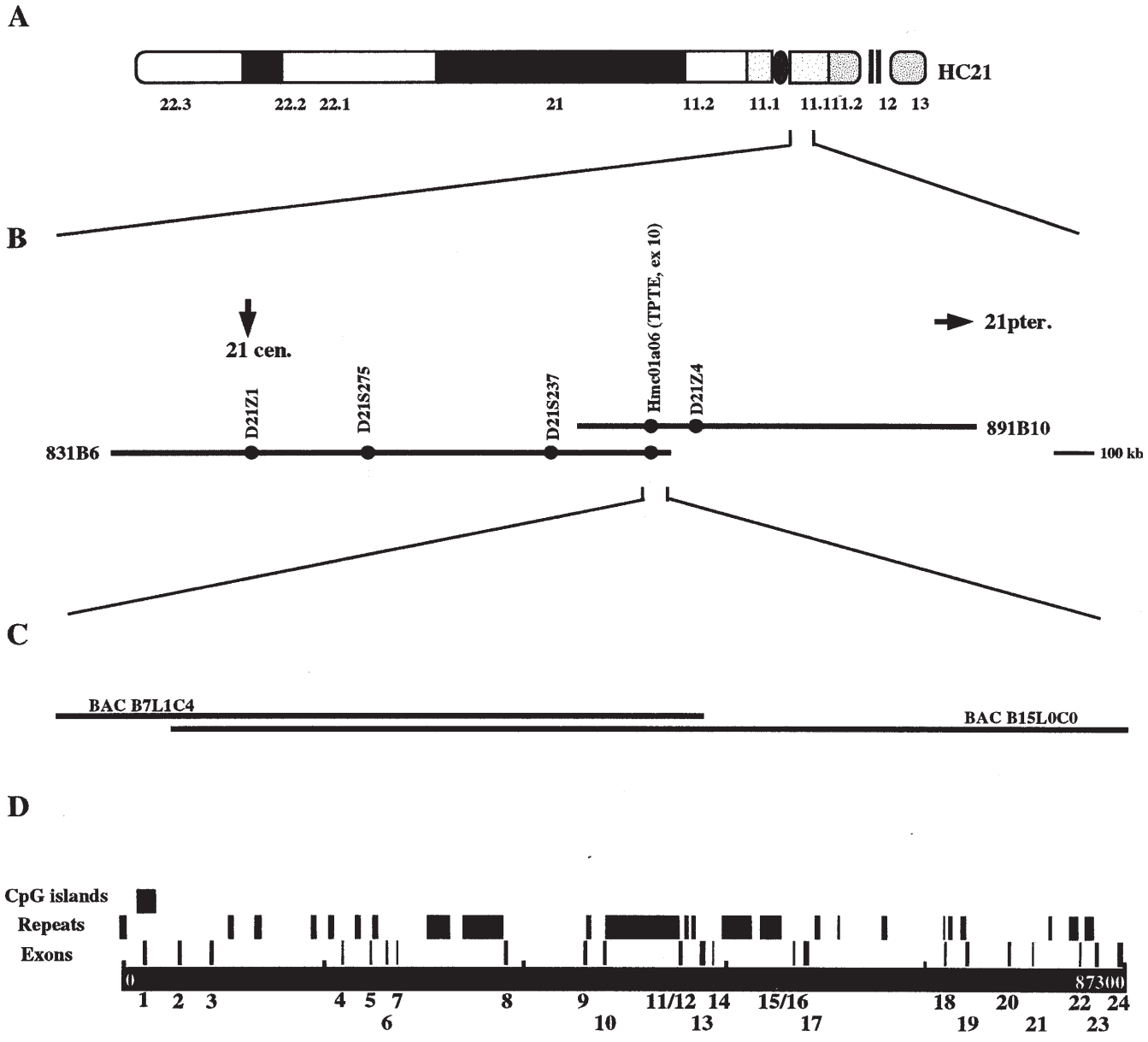

Fig. 1A–D Schematic representation of the TPTE locus on chromosome 21p. **A** Ideogram of human chromosome 21. **B** The chromosome 21p YAC contig that defines the position of the TPTE locus (Wang et al. 1999). The positions of the markers and the hmc01a06 trapped exon (exon 10 of TPTE) are shown. **C** BAC clones that contain the TPTE gene. **D** The structure and some features of the TPTE gene are shown. TPTE exons are in *black* and *numbered underneath*. The presence of repeats (*Alu*s, L1s, MERs, MIRs, etc.) and CpG islands are also indicated. *Numbers within the thick black line* are nucleotide numbers. The orientation of the transcription is from the centromere to the telomere

was constructed. These BACs were cloned from the WAV17 mouse human somatic cell hybrid that contains only human chromosome 21 (Kozak et al. 1977). Detailed analysis of the nucleotide sequences of the contig was performed using numerous computer programs (Williams et al. 1998) interfaced by NIX. This program integrates most gene-identification software (GRAIL, Fex, Hexon, MZEF, Genemark, Genefinder, FGene, Polyah, RepeatMasker, tRNAscan), Blast searches (against many databases), and also allows personal annotations with a graphical interface.

The genomic structure of the TPTE gene was determined using the "EST_genome" molecular biology software available at HGMP (http://www.hgmp.mrc.ac.uk/Registered/Option/est_genome.html). This software predicts the genomic organization of genes by comparing the genomic to the cDNA sequence. The TPTE cDNA (Genbank Nos. NM_013315 or AF007118) was used for sequence comparisons.

FISH analysis

Metaphase chromosomes were prepared from human peripheral blood lymphocytes. Standard FISH protocols were followed (Ward et al. 1995). BAC 15L0C0 and the alpha satellite were labeled by nick translation. The alpha satellite probe was Z13/21 (Vissel and Choo 1991). Labeled BAC DNA was coprecipitated with human Cot-1 competitor DNA (GIBCO-BRL) and herring sperm carrier DNA to block repetitive elements. Biotinylated probes were detected by fluorescein isothiocyanate (FITC)-conjugated avidin (Vector) and digoxygenated probes by Cy3-conjugated anti-digoxygenin antibody (Dianova). Chromosomes were counterstained with 4,6-diamidino-2-phenylindole (DAPI). Images

were taken with a Zeiss epifluorescence microscope equipped with a thermoelectronically cooled CCD camera (Photometrics CH250).

Results and Discussion

Figure 1 schematically represents the mapping and genomic organization of the chromosome 21 copy of TPTE (we designate this copy as TPTE-21). The gene spans 87 kb and is divided into 24 exons. There is 100% identity between the cDNA sequence determined by Chen et al. (1999) and the genomic sequence of contig NT_002217. Only two nucleotide substitutions were found, a c.1495G>A and a c.1748C>T (nucleotide numbers are from the Genbank entry with accession number NM_013315 or AF007118). Only the second change results in an amino acid substitution, P470L. Table 1 lists the size of each exon and intron and the sequences of intron-exon junctions. Exons range from 51 to 162 nucleotides; intron sizes range from 98 (intron 16) to 11,856 nucleotides (intron 17). All splice junction sequences conform to the GT-AG rule. The codon ATG for the initiator methionine is on exon 4; there are therefore at least three untranslated 5´-exons. There is no experimental evidence for alternative splicing; however, this possibility needs to be investigated in testicular RNA once the genomic organization of the TPTE gene family is elucidated. There is a predicted CpG island in exon 1 and 5´-UTR, implying that it may represent part of the promoter sequences in this copy of TPTE.

FISH analysis using a TPTE containing BAC clone from chromosome 21 clearly showed a hybridization signal on 21p (Fig. 2). This confirms the previous mapping of TPTE in a YAC contig of 21p (Wang et al. 1999). A hybridization signal on 13p was also observed, as expected.

Database searches revealed the existence of unordered pieces of sequences from the human chromosome 13 clone RP11–95E16 (Genbank No. AL139386, first draft sequence from the Sanger Center) containing all except the last TPTE exon. Analysis of the sequences indicates that chromosome 13 also contains at least one copy of TPTE. There was complete identity among the exons of the chromosome 21 and 13 TPTEs. The introns were also highly conserved (98–100% identity). The coding regions of the chromosome 13 TPTE sequence only show one nucleotide difference from the TPTE cDNA. This is c.1784A>C, resulting in a Y482S amino acid substitution. However, the chromosomal mapping of this clone by FISH showed signal on both 21p and 13p (http://www.sanger.ac.uk/cgi-bin/humace/searcher.cgi). It is not clear that this clone originates from chromosome 13 and not 21.

Blast searches identified additional genomic sequences with high homology to TPTE. These sequences are "working draft" quality, unordered and not precisely mapped to specific human chromosomes. For example, a homologous draft sequence with mapping position to chromosome Y (BAC clone RP11–428D10; Genbank No. AC019099) has also been mapped by the Sanger Center to chromosome 13 by FISH (http://www.sanger.ac.uk/cgi-bin/humace/searcher.cgi). Because of the preliminary nature

Table 1 Intron-exon boundaries and exon and intron sizes of the human chromosome 21 copy of TPTE

Exon	Size (bp)	5´-Splice donor	Intron	Size (bp)	3´-Splice acceptor
1	118[a]	CCCGAGgtgagc	1	2886	cttagAGAATG
2	109[a]	GTATAGgtaagc	2	2666	ggtagAGTTAT
3	58[a]	CAATGGgtgagt	3	11268	tctagTCCACC
4	54[b]	TGAAAGgtgagt	4	2377	tttagTCCTGA
5	54	TGACAGgtgagt	5	1229	actagTCCACA
6	54	AGAAAGgtgagc	6	880	tctagCCCACA
7	54	TGAAAGgtgagt	7	9274	taaagTGTGTT
8	60	ATATGAgtaagt	8	6777	tgcagCAGCAA
9	51	ATTTGGgtatgt	9	1485	tttagACTATT
10	162	AGAAAGgtaagt	10	6478	tacagGAGACA
11	120	TCCCAGgtatga	11	1647	cttagATGGAC
12	100	AGGCGGgtaagt	12	146	gtcagGTTTCA
13	64	TTACAGgtttgt	13	738	tttagAACGTA
14	65	ATCAAGgtaagg	14	6910	tttagGAAGTT
15	61	TATGCAgtatgt	15	816	tttagGTGAAA
16	82	TCTACAgtaagt	16	98	atcagTCAGAT
17	89	GCACAGgtaata	17	11856	tctagATAGAA
18	62	GCAAAGgtatga	18	1769	tgtagGAAAGC
19	79	TCTCAGgtaagt	19	3608	ttcagCGAAAA
20	106	TTCCTCgtaagt	20	1927	aacagGTTATG
21	80	TGTTCGgtaagg	21	3963	catagGTACTT
22	91	TATTCGgtgagt	22	1411	tttagAATCTT
23	71	TAACAGgtatga	23	1784	cttagGCTTTA
24	290[c,d]				

[a]Untranslated, contains longest cDNA sequence known
[b]Forty-three nucleotides to start codon
[c]One hundred and thirty-six nucleotides to stop codon
[d]Two hundred and seventy-two nucleotides to the polyadenylation signal

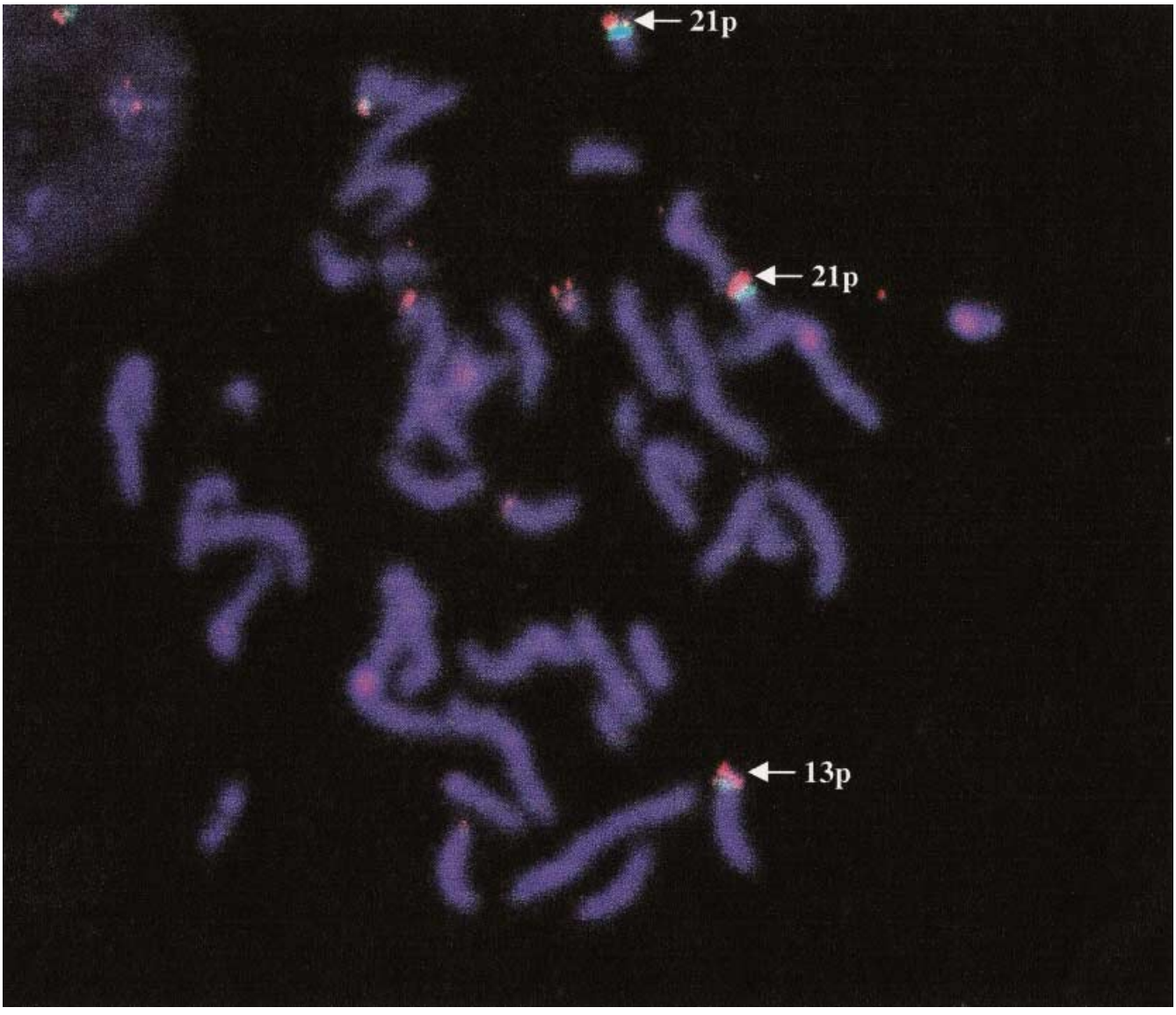

Fig. 2 FISH mapping of BAC 15L0C0 containing part of the human TPTE gene on the short arm of human chromosome 21. Chromosomes were stained with DAPI. The BAC 15L0C0 signal is shown in *red* and the alpha satellite signal is shown in *green*. On both chromosomes 21 the BAC hybridization is clearly on the short arms; labeling on chromosome 13 is also observed because of the known sequence sharing between chromosome 21 and 13 in the centromeres and short arms

of these draft sequences, we did not extend our analysis to TPTE copies in other chromosomes.

To our knowledge, the TPTE described here is the first gene besides the RNR gene family that maps the short arm of an acrocentric human chromosome. The determination and analysis of the genomic sequence of TPTE provide tools for mutation analyses in selected phenotypes and the elucidation of the evolutionary history of this family of genes and pseudogenes. The genomic organization of the additional members of the TPTE gene family is likely to be determined in the next couple of years. It is currently unknown which members of the TPTE gene family are expressed in the testis. The TPTE-21 has the potential of being expressed; however, its mapping position near the centromere may prevent its expression. It is well known that genes placed in the heterochromatin region of *Drosophila* are silenced (see Karpen 1994 for review). It is possible that similar mechanisms may operate in human chromosomes and prevent the expression of genes located near the centromeres.

Acknowledgements This study was supported by grants from the Swiss FNRS 31-57149.99, the European Union/OFES BMH4-CT98-3039, and funds from the University and Cantonal Hospital of Geneva.

References

Chen H, Rossier C, Morris MA, Scott HS, Gos A, Bairoch A, Antonarakis SE (1999) A testis-specific gene, TPTE, encodes a putative transmembrane tyrosine phosphatase and maps to the pericentromeric region of human chromosomes 21 and 13, and to chromosomes 15, 22, and Y. Hum Genet 105:399–409

Choo KH, Vissel B, Earle E (1989) Evolution of alpha-satellite DNA on human acrocentric chromosomes. Genomics 5:332–344

Greig GM, Warburton PE, Willard HF (1993) Organization and evolution of an alpha satellite DNA subset shared by human chromosomes 13 and 21. J Mol Evol 37:464–475

Henderson AS, Warburton D, Atwood KC (1973) Ribosomal DNA connectives between human acrocentric chromosomes. Nature 245:95–97

Karpen GH (1994) Position-effect variegation and the new biology of heterochromatin. Curr Opin Genet Dev 4:281–291

Kozak CA, Lawrence JB, Ruddle FH (1977) A sequential staining technique for the chromosomal analysis of the interspecific mouse/hamster and mouse/human somatic cell hybrids. Exp Cell Res 105:109–117

Van Camp G, Cruts M, Backhovens H, Wehnert A, Van Broeckhoven C (1992) Unique sequence homology in the pericentromeric regions of the long arms of chromosomes 13 and 21. Genomics 12:158–160

Vissel B, Choo KH (1991) Four distinct alpha satellite subfamilies shared by human chromosomes 13, 14 and 21. Nucleic Acids Res 25:271–277

Vissel B, Nagy A, Choo KH (1992) A satellite III sequence shared by human chromosomes 13, 14, and 21 that is contiguous with alpha satellite DNA. Cytogenet Cell Genet 61:81–86

Wang SY, Cruts M, Del-Favero J, Zhang Y, Tissir F, Potier MC, Patterson D, Nizetic D, Bosch A, Chen H, Bennett L, Estivill X, Kessling A, Antonarakis SE, Broeckhoven C van (1999) A high-resolution physical map of human chromosome 21p using yeast artificial chromosomes. Genome Res 9:1059–1073

Ward DC, Boyle A, Haaf T (1995) Fluorescence in situ hybridization techniques. Metaphase chromosomes, interphase nuclei, and extended chromatin fibers. In: Verma RS, Babu A (eds) Human chromosomes. Principles and techniques. MacGraw-Hill, New York, pp 184–192

Williams GW, Woollard PM, Hingamp P (1998) NIX: a nucleotide identification system at the HGMP-RC. http://www.hgmp.mrc.ac.uk/NIX/