Cytogenet Cell Genet 91:224–230 (2000)

Cytogenetics and Cell Genetics

Elevated DNA sequence diversity in the genomic region of the phosphatase PPP2R3L gene in the human pseudoautosomal region

K. Schiebel,[a] J. Meder,[a] A. Rump,[b] A. Rosenthal,[b] M. Winkelmann,[a] C. Fischer,[a] T. Bonk,[c] A. Humeny[c] and G.A. Rappold[a]

[a] Institute of Human Genetics, Heidelberg University, Heidelberg;
[b] Institute of Molecular Biotechnology, Department of Genome Analysis, Jena;
[c] Institute for Biochemistry, University of Erlangen-Nürnberg, Erlangen (Germany)

Dedicated to Professor Dr. Ulrich Wolf on the occasion of his retirement.

Abstract. The evolution, inheritance and recombination rate of genes located in the pseudoautosomal region 1 (PAR1) is exceptional within the human genome. Pseudoautosomal genes are identical on X and Y chromosomes and are not inherited in a sex linked manner. Due to an obligatory recombination event in male meiosis, pseudoautosomal genes are exchanged frequently between X and Y chromosomes. During the isolation, characterization and sequencing of a novel gene PPP2R3L, which was classified by sequence homology as a novel member of the protein phosphatase regulatory subunit families, it became apparent that cosmids of different origin harboring this gene are highly polymorphic between individuals, both at the nucleotide level and in the number and sequence of tandem repeats.

The final stage of the human genome sequencing project is accompanied by efforts to identify polymorphic sites within the genome which are useful for genome-wide screening. It has been suggested that the identification of genes with low penetrance or modest effects on human traits, such as those generating susceptibility or resistance to common disorders, may require the use of genetic maps containing 100,000 to 300,000 randomly spaced markers (Kruglyak, 1997; Wang et al., 1998a). Single nucleotide polymorphisms (SNPs) are considered to represent the most common type of DNA sequence variation and are estimated to occur once in every 500–1,000 bp when any two chromosomes are compared (Cooper et al., 1985; Li and Sadler, 1991; Harding et al., 1997). Furthermore, a positive correlation between recombination rate and the degree of polymorphism was observed (Charlesworth et al., 1993; Nachman et al., 1998).

The human pseudoautosomal region 1 (PAR1) located at the tips of the short arms of the human X and Y chromosomes is a genomic region with exceptionally high recombination rates. It has been suggested that an obligatory crossing over event occurs in every male meiosis which leads to a recombination rate of up to 50 % within the 2.6–2.9 Mbp of DNA (Rouyer et al., 1986; Lien et al., 2000).

The isolation and characterization of a novel regulatory subunit gene of a protein phosphatase within the pseudoautosomal region led to the identification of several new polymorphic sites, VNTRs and SNPs by comparative sequencing of three cosmid clones from two different libraries. Protein kinases and protein phosphatases are key regulators of cellular functions from metabolism and signal transduction to cell division and memory. Protein kinases and phosphatases belong to large gene families which can be subdivided into subfamilies according to their specificity or sensitivity to inhibitors, their expression

Supported by the Deutsche Forschungsgemeinschaft Ra380/9-4 and the Johannes and Frieda Marohn-Stiftung.

Received 13 June 2000; manuscript accepted 10 July 2000.

Request reprints from Gudrun A. Rappold, Institute of Human Genetics, Heidelberg University, Im Neuenheimer Feld 328, D–69120 Heidelberg (Germany); telephone: 49-6221-565059; fax: 49-6221-565332; e-mail:gudrun_rappold@med.uni-heidelberg.de

Present address of K.S.: Institute for Biochemistry, University of Erlangen-Nürnberg, Fahrstr. 17, 91054 Erlangen (Germany).

KARGER

Fax + 41 61 306 12 34
E-mail karger@karger.ch
www.karger.com

0301–0171/00/0914–0224$17.50/0

Accessible online at:
www.karger.com/journals/ccg

pattern, and sequence similarities (for review see Shenolikar, 1994; Wera and Hemmings, 1995; Barford, 1996; Cohen, 1997).

Members of the phosphatase subfamily PPP2 (formerly PP2A; for new nomenclature see http://wwww.gene.ucl.ac.uk/nomenclature/) are classified as Ser/Thr phosphatases which can be inhibited by okadaic acid (Wera and Hemming, 1995). PPP2 is a trimeric holoenzyme, consisting of a 36-kDa catalytic subunit (PPP2C or PP2Ac) and a constant 65-kDa regulatory subunit (PPP2R1, PR65 or A subunit) which form together the dimeric core enzyme and an additional variable regulatory subunit B (PPP2R2–PPP2R5). With the exception of PPP2R3, all regulatory B subunits isolated to date represent apparently unrelated gene (sub-)families each encoding several sequence-related subunits (Mayer et al., 1991; Hendrix et al., 1993; McCright and Virshup, 1995; McCright et al., 1996; Tehrani et al., 1996; Tanabe et al., 1996).

We describe the isolation and characterization of a novel protein phosphatase regulatory subunit gene PPP2R3L, exhibiting high homology to PPP2R3. Similar to PPP2R3, PPP2R3L is highly expressed in skeletal muscle and heart. Comparative sequencing of two distinct alleles of the PPP2R3L gene leads to the identification of new polymorphic sites within the telomeric part of the human pseudoautosomal region PAR1.

Materials and methods

Primers

The following primers were used for sequencing, RT-PCR and RACE experiments: **JM1rev:** 5′-ACC TGC GCC GAG CTG CGG-3′; **JM2rev:** 5′-GAA GGG AAG ATC AGC TAT GCC G-3′; **JM3rev:** 5′-TGG GAG AAG TAC GCG GCC G-3′; **JM1Buni:** 5′-GAA GAA GGG CCT CTG GGC C-3′; **JM2uni:** 5′-CCG TCC CTG AGC AGG GAG-3′; **pp1rev:** 5′-CGT CGT CGT GGC AGT TCT GG-3′; **pp2rev:** 5′-CCA CCA GGC CCA TGT CGT CC-3′; **pp1for:** 5′-ACC GGC CAC GAG CCA AAG C-3′; **ppe1for:** 5′-GCC AGG ACC CCG ACC TGC G-3′; **pp5′for:** 5′-CGC CCT CCT CAG CCT GAG G-3′.

Cosmid isolation

Cosmids were isolated from the ICRF X chromosome-specific cosmid library (ICRFc104), the Lawrence Livermore X chromosome-specific cosmid library (LLNLc110) and the Y-chromosome specific cosmid library (LLNOYC03′M′). The ethnic background of the DNA used for cloning is not known. Cosmids were isolated as described by Rao et al. (1997).

Southern blot and *Northern blot hybridization*

Southern blot hybridization was carried out under high stringency conditions in Church buffer (0.5 M NaPi, pH 7.2, 7 % SDS, 1 mM EDTA) at 65 °C and washed in 40 mM NaPi, 1 % SDS at 65 °C. A multiple-tissue Northern blot (Clontech) with RNA of adult human tissues was hybridized with a PCR-fragment (ppe1for-JM1Buni) of the PPP2R3L cDNA in formamide buffer at 42 °C and washed under high stringency conditions (0.1 × SSC, 0.1 % SDS, 65 °C) as described by the manufacturer.

cDNA isolation and 5′RACE

A 1.1-kbp *Eco*RI/*Not*I fragment (6475/12) of cosmid LLNLc110H032 was used to screen filters of a fetal brain cDNA library (Stratagene #936206; 5.5 × 10^5 pfu). After rescreening, one positive clone was isolated and sequenced. To further extend the 5′ end of this cDNA clone, 5′ RACE was used to amplify a skeletal muscle cDNA library (Rao et al., 1997) using a PPP2R3L specific primer, JM1Buni, and the vector primers λgt10for or λgt10rev. For nested PCR, the primer JM2uni was used.

RT-PCR

RT-PCRs of PCR amplified cDNA libraries (Rao et al., 1997) were used to show low level expression of the PPP2R3L gene in several tissues and to verify the expression of the first exon of PPP2R3L, which was identified by genomic sequencing. To characterize the X-inactivation behavior of the PPP2R3L gene, RNA isolated from three different somatic cell hybrids containing either the active human X chromosome, the inactive X or the human Y chromosome were reverse transcribed (Schiebel et al., 1993) and amplified with primers JM2rev and JM1Buni at 62 °C annealing temperature. PCR products were cloned into pMOS*Blue* using the pMOS*Blue*T-vector Kit (Amersham-Pharmacia Biotech).

Genomic sequencing

Sonicated fragments of three different overlapping cosmids (LLNLc110P1837, LLNLc110H032 and LLNOYC03′M′56G10) were subcloned separately into M13mp18 vector. At least 1,000 plaques were selected from each cosmid sublibrary, M13 DNA was prepared and sequenced with dye-terminators, ThermoSequenase (Amersham) and universal M13-primer (MWG-BioTech). The gels were run on ABI-377 sequencers and data were assembled and edited using the GAP4 program (Staden). Several genomic DNA sequence analysis programs (GENSCAN, GRAIL2, MZEF and XPOUND) were used to predict exons. With the exception of exon 12, all exons were recognized by at least one program. Not a single exon was detected by all programs in its correct length.

GenBank accession numbers

The PPP2R3L cDNA sequence is available from the sequence data base under accession number Y13629, the genomic sequence of cosmids LLNLc110P1837 and LLNLc110H032 as AF215839; and LLNOYC03′M′56G10 as AF215840.

Results

Isolation of PPP2R3L cDNA

Rare cutter restriction mapping of cosmids residing in the distal part of the PAR1 (Rao et al., 1997; Gianfrancesco et al., 1998) identified a cluster of restriction sites in cosmid LLNLc110H032, indicating a potential CpG island (Fig. 1). A 1.1-kbp *Eco*RI/*Not*I fragment (6475/12) of cosmid LLNLc110H032 was used to screen a fetal brain cDNA library, resulting in the isolation of a single 1.3-kbp cDNA (FB). 5′ RACE experiments using a skeletal muscle cDNA library as template further extended this cDNA by 0.3 kbp (SM-5′RACE clone). To identify the full length sequence, a 1.9-kbp *Not*I/*Eco*RI fragment of cosmid H032 was sequenced, revealing the missing first exon. Exon 1 overlaps by 62 bp with the previously isolated 5′RACE-PCR-product (Fig. 1). A 1.9-kbp consensus cDNA was assembled and confirmed by RT-PCR. This sequence contains an ORF with the potential to encode proteins of 510 or 548 amino acids depending on which of the two different start codons located at position 94 and 208, respectively, is used (Fig. 1B). Molecular weight of these proteins is predicted to be 58.6 and 62.6 kDa, respectively.

Sequence homologies of PPP2R3L

By data base analysis the new cDNA was found to encode a member of the regulatory subunit family of protein phosphatase PPP2 (2A). The highest DNA homology of 72 % over 1,058 bp was found to a protein phosphatase 2A regulatory subunit PR59 mRNA in *Mus musculus* (Acc. No. AF050165; Voorhoeve et al., 1999). Several human ESTs and the protein phosphatase PPP2R3 (72-kDa regulatory subunit) revealed the

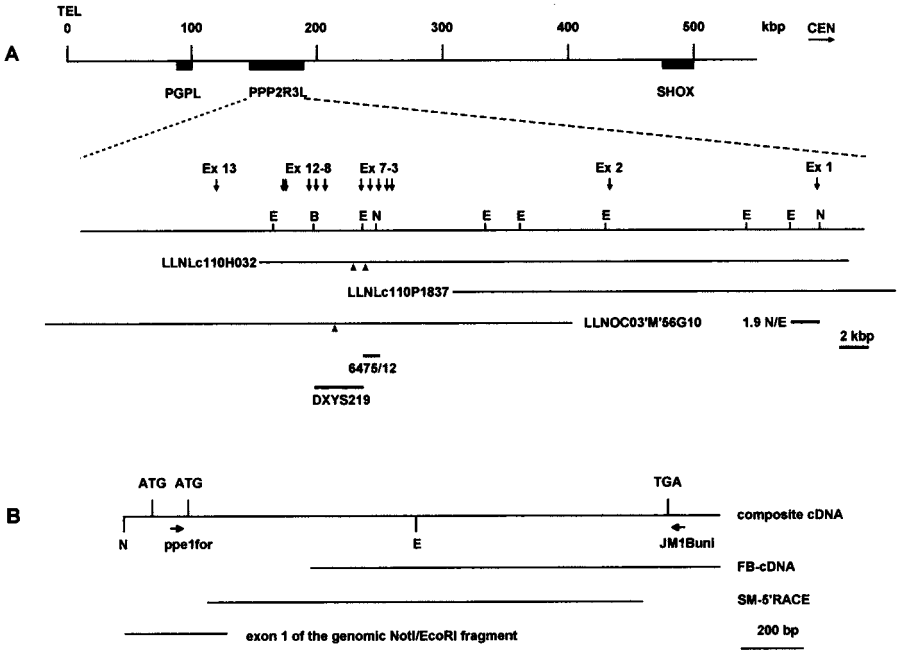

Fig. 1. Localization of the PPP2R3L gene in the telomeric part of the PAR1. (**A**) The PPP2R3L gene-containing cosmids are part of the cosmid contig isolated by Rao et al. (1997). The *Eco*RI restriction map of the cosmids containing the cDNA is depicted (according to sequence data). Some of the *Not*I (N) and *Bss*HII (B) sites are shown because of their relevance for the primary cDNA isolation (1.1-kbp *Eco*RI/*Not*I fragment 6475/12), the isolation of the first exon (1.9-kbp *Not*I/*Eco*RI fragment) and the localization of the polymorphic probe DXYS219 (Ried et al., 1995). Exon positions are marked by arrows. Arrowheads indicate small sequencing gaps (<3 kbp). The *Eco*RI/*Bss*HII fragment containing exons 7–9 is variable in length and is depicted in the minimal length as sequenced. (**B**) cDNA subfragments isolated by cDNA library screening and 5′RACE. The exon portion of the 1.9-kbp *Not*I/*Eco*RI fragment was used to complete the coding region.

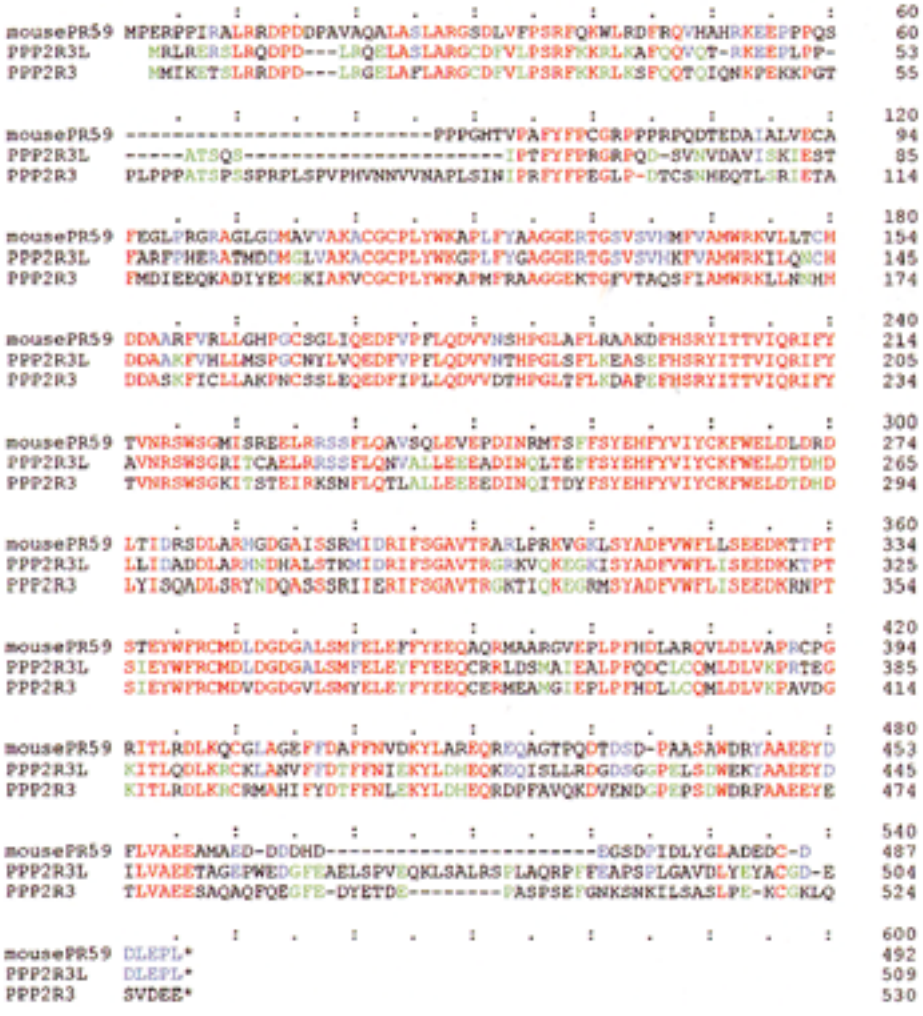

Fig. 2. Comparison of protein sequences of mouse PR59, human PPP2R3L and PPP2R3. Identical amino acids in all three proteins are highlighted in red. Blue indicates identity of PR59 and PPP2R3L and green indicates identity of the two human proteins PPP2R3 and PPP2R3L.

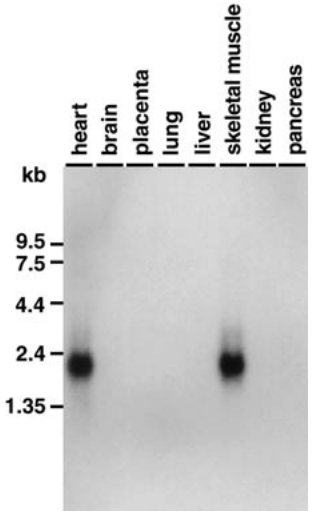

Fig. 3. Expression of PPP2R3L cDNA. A Northern blot of human adult tissues was hybridized with the PCR fragment ppe1for-JM1Buni generated from the PPP2R3L cDNA. Strong expression was detected after 6 h exposure (– 80 °C) in skeletal muscle and heart.

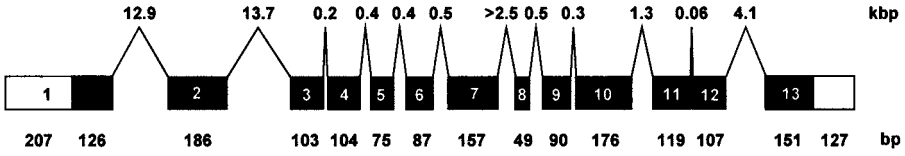

Fig. 4. Genomic organization of the PPP2R3L gene. The cDNA is composed of 13 exons. Due to two sequencing gaps and a VNTR polymorphism, the size of intron 7 is not exactly known. Exon length is given in bp, intron length in kbp.

highest homology scores in humans (Acc. No. L12146; 66% over 1,217 bp; Hendrix et al., 1993). PPP2R3L has a high similarity to the human PPP2R3 72-kDa protein (77.7 % over 491 aa), higher than to the mouse PR59 protein (72.7% over 453 aa) although it exhibits the same level of identity (67%) (Fig. 2). Therefore, the new cDNA was termed PPP2R3-like (or PPP2R3L). Sequence comparison of PPP2R3 and PPP2R3L genes revealed that the sequence homology upstream of the second potential translation start codon at position 208 is very weak without any homology on the amino acid level, suggesting that this ATG/AUG represents the translation start point in PPP2R3L (Fig. 1B).

Expression of PPP2R3L

The PPP2R3L cDNA was isolated from a fetal brain cDNA library (one clone out of 5.5×10^5 pfu). Northern blot hybridization of a PCR-fragment (ppe1for-JM1Buni) of the cDNA clone (see Fig. 1) revealed strong signals in skeletal muscle and heart after 3–6 h of exposure (Fig. 3). Weaker signals were detectable in all adult tissues investigated after longer exposures of 1–3 days. To verify that the signals are not due to cross-hybridization with other protein phosphatases, 13 different cDNA libraries were analyzed by RT-PCR with PPP2R3L-specific primers. Expression was found in fetal brain, fetal liver, skeletal muscle, ovary, kidney and chondrocytes. Amplification products of brain stem, bone marrow fibroblasts, hypophyse and fibroblasts were only detectable after hybridization (data not shown).

X-Inactivation behavior of PPP2R3L

To test the X-inactivation behavior and transcriptional activity of X and Y copies of PPP2R3L, expression was analyzed by RT-PCR (primers JM2rev and JM1Buni) using RNA isolated from somatic cell hybrid lines bearing either the human Y chromosome or the active or the inactive X chromosome as templates. No amplification was found in somatic cell hybrids probably due to the low expression of the PPP2R3L gene within these cell lines (data not shown).

Therefore, the methylation pattern of the CpG rich region at the 5′end of the PPP2R3L gene was examined by rare cutter restriction analysis of male and female genomic DNA. Hybridization of the 1.9-kbp *Not*I/*Eco*RI fragment (Fig. 1) containing the first exon of the PPP2R3L gene revealed no differences between male and female DNA and all rare cutter enzymes tested such as *Bss*HII, *Ecl*XI, *Not*I, *Nar*I, *Sac*II and *Sma*I cleaved the 10-kbp *Eco*RI fragment into smaller fragments (data not shown). These findings indicate that PPP2R3L has an unmethylated CpG island in male and female DNA and therefore represents a gene escaping X-inactivation.

Genomic organization and localization of PPP2R3L

The complete alignment between the cDNAs and the sequence of cosmids LLNOYC03′M′56G10, LLNLc110H032 and LLNLc110P1837 revealed the genomic structure of the PPP2R3L gene. PPP2R3L consists of 13 exons each 49–327 bp in length extending over a genomic region of at least 36 kbp (Fig. 1 and 4). Intron 7 is highly polymorphic due to a 27-bp VNTR that complicates correct alignment of shotgun sequencing fragments. The previously described 4.8-kbp polymorphic marker DXYS219 (Ried et al., 1995), starting at the *Eco*RI-site in exon 7 and ending at the *Bss*HII-site in intron 9, includes the 27-bp VNTR of intron 7. Due to various VNTR polymorphisms, the cosmids exhibit slightly different restriction patterns. PPP2R3L is transcribed (5′ to 3′) from centromere to telomere. Data from the available cosmid contig maps PPP2R3L less than 50 kbp proximal to the PGPL (Gianfrancesco et al., 1998) and 300 kbp distal to the SHOX gene (Ellison et al., 1997; Rao et al., 1997). Southern blot hybridization, fluorescence in-situ-hybridization as well as the presence of the X/Y specific marker DXYS219 unequivocally show that all cosmids are derived from the same genomic locus (data not shown).

Analysis of genomic sequences

Sequence analysis of approximately 43 kbp of the cosmids carrying the PPP2R3L gene were noteworthy. The GC level of the analyzed region is 59.3% in total. Local GC values are up to 77% and the largest predicted CpG core region of 885 bp (nt 2568–3452) surrounds the first exon of the PPP2R3L gene (nt 2917–3249) which was shown by rare cutter restriction analysis

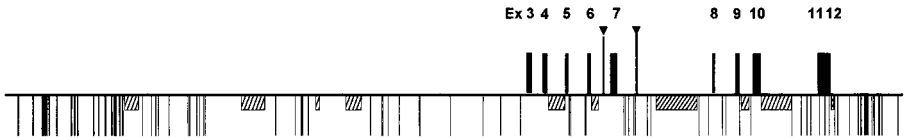

Fig. 5. Localization of polymorphic sites in the genomic region of PPP2R3L gene. SNPs are depicted as vertical lines, hatched rectangles mark VNTRs. Exons are depicted as black rectangles and arrows indicate sequencing gaps (see Fig. 1).

to represent an unmethylated CpG island (see above). Identification of repetitive elements by the Repeat Masker program showed that 32.7% of the DNA in this region consists of repetitive DNA such as ALU (16.5%) and LINE1 (13.2%) elements, as well as elements of low complexity or simple repeats (2%). The repeat content may even be underestimated due to three (small) gaps near exon 7.

Significant differences were found in comparisons of three different cosmids representing two different alleles of the PPP2R3L gene, isolated from two libraries (LLNOYC03′M′ or LLNLc110). Approximately 20 kbp of overlapping sequences contain ten different VNTRs ranging from 4 to 71 bp in length which are responsible for the length polymorphisms seen by comparative restriction mapping of cosmids. In addition, the alignment of cosmid DNA revealed 108 SNPs (one single base insertion/deletion or base exchange) in the non-VNTR DNA and a further 66 SNPs within VNTRs (Fig. 5). Three further SNPs located in the translated region were determined by database analysis of ESTs. One SNP located at nt 1565 of the cDNA sequence (exon 12) also leads to a protein variability by an amino acid exchange of Ala → Val. Analysis of four variable nucleotides located within the coding region in 32 Caucasian individuals was subsequently carried out and verified all as SNPs.

Discussion

Protein phosphatases PPP2 are trimeric enzyme complexes consisting of a constant catalytic subunit, a constant regulatory subunit as well as a variable regulatory subunit. As one of the regulatory subunits is variable, the functional relevance of phosphatases is diverse. The isolation of a novel human regulatory subunit highly homologous to PPP2R3 demonstrates that all regulatory subunits of PPP2 (PPP2R1–PPP2R5) isolated to date have closely related family members within the human genome (http://wwww.gene.ucl.ac.uk/nomenclature/; Mayer et al., 1991; Hendrix et al., 1993; McCright and Virshup, 1995; Tehrani et al., 1996; Tanabe et al., 1996). Sequence homology on both the DNA and the protein levels is found within a regulatory subfamily but is not shared with other regulatory subunit families. In contrast to PPP2R3 (Hendrix et al., 1993), PPP2R3L seems not to be expressed in different splice variants as cDNA analysis, Northern blot hybridization, 5′RACE and sequencing of the genomic locus did not reveal any evidence of additional transcripts.

Besides intraspecies homology, the highest sequence homology of PPP2R3L was found in the murine protein phosphatase 2A regulatory subunit PR59 mRNA (Voorhoeve et al., 1999). Despite the high homology of PPP2R3L to PR59 at the DNA level, we hypothesize that PR59 is not the ortholog of PPP2R3L. This is supported by the ubiquitous expression of PR59 in mouse in contrast to the restricted predominant expression of PPP2R3 (PR72; Hendrix et al., 1993) and PPP2R3L in skeletal muscle and heart. FISH mapping of the PR59 gene in mouse has also shown that PR59 is not localized to the mouse X chromosome (unpublished results). These findings are consistent with the observation that none of the human pseudoautosomal genes examined to date have been shown to reside on the mouse X chromosome (Blaschke and Rappold, 1997).

The functional relevance of PPP2R3L may be deduced from two identical but partial cDNA sequences recently published in the database (Acc. Nos. AF135016 and AF155098). AF155098 was isolated by screening with autologous antibodies in patients with renal-cell carcinoma (Scalan et al., 1999). AF135016 (PR48) has been shown to form functional holoenzyme complexes with the A and C subunits of PPP2 (Yan et al., 2000). The authors further demonstrate that PR48 specifically interacts with cdc6, an essential component of prereplicative complexes and that PR48 overexpression dysregulates the cell cycle. PR48 cDNA starts at nt 382 in exon 2 of the PPP2R3L cDNA. The deduced protein consequently is 95 amino acids shorter than PPP2R3L protein. These two findings together with the observation of an alteration of the PPP2R1B gene in human cancer (Wang et al., 1998b) may argue for an involvement of the PPP2R3L gene in carcinogenesis. Loupart et al. (1995) have demonstrated that loss of heterozygosity for markers located in the distal 500 kbp of the PAR1 are frequently found in patients with breast cancer. Recently, a locus predisposing for Hodgkin disease was shown to be linked to the human pseudoautosomal region 1 (Horwitz and Wiernik, 1999). Therefore an analysis of PPP2R3L gene in different types of tumors is warranted to determine the relevance in human tumorigenesis.

Compared to the genome average, our finding of one nucleotide exchange (SNP) in 120–180 bp of DNA in two cosmids of different origin revealed an elevated rate of sequence diversity in the PPP2R3L gene region. Initial estimates of the frequency of SNPs within the genome varied from "one in every one hundred bp" (Jeffreys, 1979) to 1 in 1 kbp (Cooper et al., 1985) based on RFLP analysis. Recent data of four different genomic regions were gained from comparative sequencing and are depicted on Table 1. Calculating the nucleotide diversity π according to Nei and Li (1979) is a prerequisite to compare these data. Nucleotide diversity is defined as the rate of variable sites found comparing two randomly selected homologous chromosomes. The value of all SNPs identified in a distinct region is expected to be much higher, depending on the heterozygosity of different alleles. Although only two alleles were compared, the highest values for nucleotide diversity were found in the pseudoautosomal PPP2R3L gene region on Xp/Yp with a value of 0.0054 (1 SNP in 185 bp), followed by the

Table 1. Nucleotide diversity in different regions of the human genome

Reference	Gene(s)	Genomic localization	Nucleotide diversity π[a]	Length of analyzed sequence (kbp)	No. of chromosomes
Nickerson et al., 1998	Lipoprotein lipase	8p22	0.0020	4.7	142
Halushka et al., 1999	75 blood pressure candidate genes	not specified	0.00085	190	148
Cargill et al., 1999	106 genes	not specified	0.00054	196.2	114
this paper	PPP2R3L	Xp22.3; Yp11.3	0.0054	20	2

[a] Nucleotide diversity π is defined according to Nei and Li (1979).

lipoprotein lipase gene on chromosome 8p22 (LPL; Nickerson et al., 1998) (Table 1). In contrast to the investigation of Halushka et al. (1999) and Cargill et al. (1999), the LPL (Nickerson et al., 1998) and the PPP2R3L regions are predominantly non-coding (90% and 95%, respectively). It was estimated by Halushka et al. (1999) that nucleotide diversity in coding and non-coding regions are almost identical. The elevated rate of nucleotide diversity in the LPL and PPP2R3L region, however, suggest that differences may indeed exist. In addition, the low substitution rate found in a 10-kbp non-coding region of Xq13.3 (Kaessmann et al., 1999) and gene-to-gene differences observed by Halushka et al. (1999) suggest that also regional effects may influence the rate of nucleotide diversity. This was recently shown by the high frequency of SNPs (1 in 65 bp) in a 450-bp sequence within 20 kbp of the PAR1 telomere (Baird et al., 1995). A potential influence of chromosomal position on evolutionary sequence diversity was previously discussed for the *Fxy* gene in mice. The *Fxy* gene spans the pseudoautosomal boundary and exhibits an estimated 170-fold higher diversity in the pseudoautosomal compared to the X-linked part arguing for an accelerated divergence in the PAR (Perry and Ashworth, 1999).

What is the cause of this high polymorphism rate within the telomeric part of the human pseudoautosomal region? Three major characteristics of the PAR1 may contribute to this variability. First, the available 43 kbp of genomic sequence in this part of the PAR1 is GC-rich with a GC content of approximately 60%, normally only found in CpG islands. Due to the prevalence of CpG methylation and frequent deamination of 5-methylcytosine, the transition C→T is a common mutation in the human genome (Vogel, 1972) and an 8.5-fold increased mutation rate of CpG nucleotides has been described (Cooper et al., 1995). In the PPP2R3L gene region approximately half of the SNPs (51 of 108) are C→T transitions in CpG dinucleotides, compared to only 31% in the study of Halushka et al. (1999). Second, a positive correlation of the local recombination rate and an elevated DNA variability was recently reported for X-linked genes (Nachman et al., 1998) and regional clustering of SNPs in a region regarded as a hot spot of recombination in the β-globin locus (Harding et al., 1997). Strathern et al. (1995) suggested that recombinational events per se are mutagenic because of errors in double-stranded DNA repair. Within the PAR1, a 20–60-fold elevated rate of recombination is the result of an obligate crossover during male meiosis (Rouyer et al., 1986; Henke et al., 1993, Lien et al., 2000). On average, a 14-fold increased recombination rate was observed for the distal 450-kbp region of the PAR1 by sperm typing (Lien et al., 2000). Sperm typing of different individuals has also demonstrated that there is individual variation in recombination rate within the PAR1 (Lien et al., 2000). It is therefore possible that the elevated rate of polymorphisms observed in the PPP2R3L gene region may also be influenced by individual variation and random choice of chromosomes.

A nearly complete linkage disequilibrium, which creates only a few diverged haplotypes in the subtelomeric region (Baird et al., 1995; Baird et al., 2000), is somewhat contradictory to the observed high recombination rate but may be due to a suppression of recombination within the telomere repeats and the 1-kbp telomere-adjacent DNA. However, apparent linkage disequilibrium can arise not only by suppression of recombination but also by admixture, genetic drift and selection. To test whether the strong linkage disequilibrium is restricted to the subtelomeric region directly adjacent to the telomeric repeats or whether it involves the PPP2RL region as well, the haplotypes of individuals of different ethnic backgrounds will need to be analyzed. Finally, the high prevalence of VNTRs within the PAR1 in general and within the PPP2R3L gene region in specific must be considered. Slippage and illegitimate allelic recombination may lead to a variability in the length of tandem repeats. Single nucleotide mutations, as a consequence of recombination and C→T transitions, further augment the variability within these repeats.

In conclusion, we demonstrate the isolation of a novel gene, PPP2R3L, in the human PAR1. By homology, PPP2R3L was characterized as a novel member of the protein phosphatase regulatory subunit family with a potential involvement in carcinogenic processes. Comparative sequencing of cosmids from different libraries revealed a high rate of sequence variability in the PPP2R3L gene region with 1 SNP every 115–185 bp (108–174 SNPs in 20 kbp of genomic DNA). All tested SNPs occurring in the coding region confirm the data found in cosmids. Therefore, an increased rate of SNPs is not restricted to the region directly adjacent to the telomere but also occurs in regions more proximal to it.

Acknowledgements

We thank Pieter de Jong for cosmid libraries (constructed at the HGC, LLNC, Livermore, CA94550 under the auspices of the National Laboratory Gene Library project sponsored by U.S. DOE. We thank G. Zehetner, Hans Lehrach and the resource center team in Berlin for providing the libraries. We also thank A. Mertz, M.-C. Meffert and D. Lagemann for technical assistance. We are grateful to H. Kaessmann for critical reading of the manuscript.

References

Baird DM, Coleman J, Rosser ZH Royle NJ: High levels of sequence polymorphism and linkage disequilibrium at the telomere of 12q: implications for telomere biology and human evolution. Am J hum Genet 66:235–250 (2000).

Baird DM, Jeffreys AJ, Royle NJ: Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere. EMBO J 14:5433–5443 (1995).

Barford D: Molecular mechanisms of the protein serine/threonine phosphatases. Trends Biochem Sci 21:407–412 (1996).

Blaschke RJ, Rappold GA: Man to mouse – lessons learned from the distal end of the human X chromosome. Genome Res 7:1114–1117 (1997).

Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES: Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nature Genet 22:231–238 (1999).

Charlesworth B, Morgan MT, Charlesworth D: The effect of deleterious mutations on the neutral molecular variation. Genetics 134:1289–1303 (1993).

Cohen PTW: Novel protein serine/threonine phosphatases: variety is the spice of life. Trends Biochem Sci 22:245–250 (1997).

Cooper DN, Krawczak M, Antonorakis SE: in Scriver CR, Beaudet AL, Sly WS, Valle D (eds): The Metabolic and Molecular Basis of Inherited Disease, 7th ed, pp 259–291 (McGraw-Hill, New York 1995).

Cooper DN, Smith BA, Cooke HJ, Niemann S, Schmidtke J: An estimate of unique DNA sequence heterozygosity in the human genome. Hum Genet 69:201–205 (1985).

Ellison JW, Wardak Z, Young MF, Robey PG, Laig-Webster M, Choing W: PHOG, a candidate gene for involvement in the short stature of Turner syndrome. Hum Mol Genet 6:1341–1347 (1997).

Gianfrancesco F, Esposito T, Montanini L, Ciccodicola A, Mumm S, Mazzarella R, Rao E, Rappold GA, Forabosco A: A novel pseudoautosomal gene encoding a putative GTP-binding protein resides in the vicinity of the Xp/Yp telomere. Hum Mol Genet 7:407–414 (1998).

Haluskha MK, Fan J-B, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A: Patterns of single nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nature Genet: 22:239–247 (1999).

Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, Clegg JB: Archaic African and Asian lineages in the genetic ancestry of modern humans. Am J hum Genet 60:772–789 (1997).

Hendrix P, Mayer-Jaekel RE, Cron P, Goris J, Hofsteenge J, Merlevede W, Hemmings BA: Structure and expression of a 72-kDa regulatory subunit of protein phosphatase 2A. J biol Chem 268:15267–15276 (1993).

Henke A, Fischer C, Rappold GA: Genetic map of the human pseudoautosomal region reveals a high rate of recombination in female meiosis at the Xp telomere. Genomics 18:478–485 (1993).

Horwitz M, Wiernik PH: Pseudoautosomal linkage of Hodgkin disease. Am J hum Genet 65:1413–1422 (1999).

Jeffreys AJ: DNA sequence variants in the $^{G}\gamma$, $^{A}\gamma$, δ- and β-globin genes of man. Cell 18:1–10 (1979).

Kaessmann H, Heißig F, von Haesler A, Pääbo S: DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nature Genet 22:78–81 (1999).

Kruglyak L: The use of a genetic map of biallelic markers in linkage studies. Nature Genet 17:442–447 (1997).

Li W-H, Sadler LA: Low nucleotide diversity in man. Genetics 129:513–523 (1991).

Lien S, Szyda J, Schechinger B, Rappold GA, Arnheim N. Heterogeneity in rates of recombination in the human pseudoautosomal region: high resolution analysis by sperm typing and radiation hybrid mapping. Am J hum Genet 66:557–566 (2000).

Loupart M-L, Adams S, Armour JAL, Walker R, Brammer W, Varley J: Loss of heterozygosity on the X chromosome in human breast cancer. Genes Chrom Cancer 13:229–238 (1995).

Mayer RE, Hendrix P, Cron P, Matthies R, Stone SR, Goris J, Merlevede W, Hofsteenge J, Hemmings B: Structure of the 55-kDa regulatory subunit of protein phosphatase 2A: Evidence for a neuronal-specific isoform. Biochemistry 30:3589–3597 (1991).

McCright B, Virshup DM: Identification of a new family of protein phosphatase 2A regulatory subunits. J biol Chem 270:26123–26128 (1995).

McCright B, Brothman AR, Virshup DM: Assignment of human protein phosphatase 2A regulatory subunit genes B56α, B56β, B56γ, B56δ, and B56ε (PPP2R5A-PPP2R5E), highly expressed in muscle and brain, to chromosome regions 1q41, 11q12, 3p21, 6p21.1 and 7p11.2→12. Genomics 36:168–170 (1996).

Nachman MW, Bauer VL, Crowell SL, Aquadro CF: DNA variability and recombination rates at X-linked loci in humans. Genetics 150:1133–1141 (1998).

Nei M, Li W-H: Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc natl Acad Sci, USA 76:5269–5273 (1979).

Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF: DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nature Genet 19:233–240 (1998).

Perry J, Ashworth A: Evolutionary rate of a gene affected by chromosomal position. Curr Biol 9:987–989 (1999).

Rao E, Weiss B, Fukami M, Rump A, Niesler B, Mertz A, Muroya K, Binder G, Kirsch S, Winkelmann M, Nordsiek G, Heinrich U, Breuning MH, Ranke MB, Rosenthal A, Ogata T, Rappold GA: Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. Nature Genet 16:54–63 (1997).

Ried K, Mertz A, Nagaraja R, Trusgnich M, Riley JH, Anand R, Lehrach H, Page D, Ellison JW, Rappold G: Characterization of a YAC contig spanning the pseudoautosomal region. Genomics 29:787–792 (1995).

Rouyer F, Simmler M-C, Johnsson C, Vergnaud G, Cooke HJ, Weissenbach J: A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. Nature 319:291–295 (1986).

Scanlan MJ, Gordan JD, Williamson B, Stockert E, Bander NH, Jongeneel V, Gure AO, Jager D, Jager E, Knuth A, Chen Y-T, Old LJ: Antigens recognized by autologous antibody in patients with renal-cell carcinoma. Int J Cancer 83:456–464 (1999).

Schiebel K, Weiss B, Wöhrle D, Rappold GA: A human pseudoautosomal gene, ADP/ATP translocase escapes X-inactivation whereas a homologue on Xq is subject to X-inactivation. Nature Genet 3:82–85 (1993).

Shenolikar S: Protein serine/threonine phosphatases – new avenues for cell regulation. A Rev Cell Biol 10:55–86 (1994).

Strathern JN, Shafer BK, McGill CB: DNA synthesis errors associated with double-strand-break repair. Genetics 140:965–972 (1995).

Tanabe O, Nagase T, Murakami T, Nozaki H, Usui H, Nishito Y, Hayashi H, Kagamiyama H, Takedo M: Molecular cloning of a 74-kDa regulatory subunit (B′′ or δ) of human protein phosphatase 2A. FEBS Letters 379:107–111 (1996).

Tehrani MA, Mumby MC, Kamibayashi C: Identification of a novel protein phosphatase 2A regulatory subunit highly expressed in muscle. J biol Chem 271:5164–5170 (1996).

Vogel F: Non-randomness of base replacement in point mutation. J Mol Evol 1:334–367 (1972).

Voorhoeve PM, Hijmans EM, Bernards R: Functional interaction between a novel protein phosphatase 2A regulatory subunit, PR59, and the retinoblastom-related p107 protein. Oncogene 18:515–524 (1999).

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbel E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES: Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–1082 (1998a).

Wang SS, Esplin ED, Li JL, Huang L, Gadzar A, Minna J, Evans GA: Alterations of the PPP2R1B gene in human lung and colon cancer. Science 282:284–287 (1998b).

Wera S, Hemmings BA: Serine/threonine protein phosphatases. Biochem J 311:17–29 (1995).

Yan Z, Fedorov SA, Mumby MC, Williams S: PR48, a novel regulatory subunit of protein phosphatase 2A, interacts with cdc6 and modulates DNA replication in human cells. Mol Cell Biol 20:1021–1029 (2000).