J Hum Genet (2004) 49:97–101
DOI 10.1007/s10038-003-0117-0

ORIGINAL ARTICLE

Kazuaki Nagao · Katsunori Fujii · Masao Yamada
Toshiyuki Miyashita

Identification of a novel polymorphism involving a CGG repeat in the *PTCH* gene and a genome-wide screening of CGG-containing genes

Received: 2 September 2003 / Accepted: 20 November 2003 / Published online: 21 January 2004

Abstract Mutations in the human homologue of the *Drosophila patched* gene (*PTCH*) are responsible for the hereditary disorder called nevoid basal cell carcinoma syndrome (NBCCS). *PTCH* has a CGG triplet repeat located 4 bp upstream of the first methionine codon. Here we report a novel polymorphism involving the number of the CGG-repeat. The major allele (86.3%) contained a repeat size of seven, whereas the minor allele contained eight. No significant difference in the distributions of genotypes was observed between normal and NBCCS individuals. However, when the repeat was inserted between a heterologous promoter and the luciferase gene, the longer repeats tended to induce higher luciferase activities, suggesting that the repeat length potentially affects the levels of gene expression. A genome-wide screening revealed that 68 and 146 genes contained a CGG/CCG repeat in the coding region and in the 5'-untranslated region (5'-UTR), respectively. None of the genes had this repeat in 3'-UTR. Interestingly, the number of genes with a CGG repeat in the 5'-UTR was significantly higher than that with a CCG repeat in the 5'-UTR. The localization of a CGG/CCG repeat in *PTCH* is quite unique in that only four other genes have been found in which the repeat is localized up to 4 bp upstream of the first methionine.

Keywords PTCH · Nevoid basal cell carcinoma syndrome · Gorlin syndrome · Polymorphism · Triplet repeat

K. Nagao · M. Yamada · T. Miyashita (✉)
Department of Genetics,
National Research Institute for Child Health and Development,
3-35-31 Taishido, Setagaya-ku,
Tokyo 154-8567, Japan
E-mail: tmiyashita@nch.go.jp
Tel.: +81-3-34140181
Fax: +81-3-34143208

K. Fujii
Department of Pediatrics,
Graduate School of Medicine,
Chiba University, Chiba, Japan

Introduction

Mutations in the *PTCH* gene are responsible for the hereditary disorder called nevoid basal cell carcinoma syndrome (NBCCS; MIM# 109400) (Hahn et al., 1996; Johnson et al., 1996). NBCCS, also called Gorlin syndrome, is an autosomal dominant neurocutaneous disorder characterized by developmental abnormalities and tumorigenesis such as palmar and plantar pits, jaw cysts, calcification of the falx cerebri, skeletal anomalies, basal cell carcinoma, ovarian fibroma, and medulloblastoma (Gorlin, 1987). *PTCH* (MIM # 601309) is a human homologue of the *Drosophila* segment polarity gene *patched*. It has been mapped to 9q22.3-q31 and consists of 23 exons encoding a protein with 1,447 amino acid residues. The PTCH protein is a receptor for a secreted molecule Sonic hedgehog and has twelve transmembrane domains. At least two forms of PTCH protein are known to exist, reflecting the use of alternative exon 1a versus 1b (Hahn et al., 1996; Wicking et al. 1997a). Mutations in exon 1b have not been investigated so far due to, at least in part, the extreme GC-rich sequence (Wicking et al. 1997a; Fujii et al. 2003a). In the course of analyzing mutations in exon 1b, using a new set of primers and a PCR condition, we discovered a novel polymorphism involving a CGG trinucleotide repeat immediately upstream of the first in-frame methionine codon. We compared allele frequencies between healthy individuals and NBCCS patients. We also investigated the effect of the repeat length on the gene expression using a heterologous reporter gene. In addition, the results of a genome-wide screening of CGG/CCG-containing genes are demonstrated.

Materials and methods

DNA samples

After informed consent was obtained from 51 healthy, unrelated individuals and 14 patients with NBCCS, total genomic DNAs were isolated from peripheral leukocytes by the standard phenol/

chloroform extraction method. Patients were diagnosed as having NBCCS according to the clinical criteria (Kimonis et al. 1997). All studies were approved by the local ethnic committee. Among 14 patients with NBCCS, *PTCH* mutations were found in 11. Some of the mutations have already been reported (Fujii et al. 1999, Fujii et al. 2003a, Fujii et al. 2003b) and some will be reported elsewhere.

Polymerase chain reaction and sequencing

The genomic region of *PTCH* including the 5'-untranslated region (5'-UTR) and exon 1b was amplified by using the forward primer, 5'-CGCGCAATGTGGCAATGGAA-3', and the reverse primer, 5'-AGAGGAGGGAAGAGAAAGTG-3'. The polymerase chain reaction (PCR) was carried out in a 20 μl reaction volume by using LA Taq with GC Buffer (TaKaRa) according to the manufacturer's instruction. PCR was run for 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 3 min on a Program Temp Control System PC-800 (ASTEC, Fukuoka, Japan). Both the sense and antisense strands of the PCR products were directly sequenced by using the same primers as described above. PCR products purified by a QIAquick PCR Purification Kit (QIAGEN) were used as the template DNA for cycle sequencing with a CEQ DTCS Quick-Start Kit (Beckman Coulter). Sequencing analysis was performed on a CEQ 8000 Genetic Analysis System (Beckman Coulter) according to the manufacturer's instructions.

Plasmid construction

Luciferase constructs containing the sequence of *PTCH* 5'-UTR were generated by a PCR-mediated method described previously (Imai et al. 1991) using pGV-P2 (Wako Chemicals, Osaka, Japan) as a template. The authenticity of all constructs was confirmed by sequencing.

Luciferase assay

The human embryonic kidney cell line 293 growing on six-well culture plates were cotransfected using Effectene reagent (QIAGEN) with 0.5 μg of luciferase plasmid and 0.5 μg of pCMVβGal. The cells were harvested at 24 h after the transfection and used for a luciferase assay. Luciferase activities were measured as described previously and normalized for transfection efficiency based on β-galactosidase activities (Shikama et al. 2001).

Real-time quantitative RT-PCR

Total RNA was extracted from the transfected cells described above using TRIzol reagent (Invitrogen). One-step RT-PCR was performed with a 7700 ABI PRISM Sequence Detector System (Perkin Elmer-Applied Biosystems) using primers 5'-TCTGGATCTACTGGTCTGCCTAA-3' and 5'-GCGCACTTTGAATCTTGTAATCCTG-3'. To normalize the expression of luciferase, the *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) housekeeping gene was also amplified, using primers 5'-GAAGGTGAAGGTCGGAGT-3' and 5'-GAAGATGGTGATGGGATTTC-3'. Fluorogenic probes 5'-CAAATCATTCCGGATACTGC-3' and 5'-CAAGCTTCCCGTTCTCAGCC-3' carrying 5' 6-carboxy-fluorescein as a reporter dye and 3' 6-carboxy-tetramethyl-rhodamine as a quencher dye were used to detect the PCR product of luciferase and *GAPDH*, respectively. In every experiment, *GAPDH* was amplified using a series of dilutions of a known amount of the standard RNA supplied by Perkin Elmer to prepare a standard curve.

Computational screen

Human mRNA sequences that contain more than seven repeats of CGG were downloaded from NCBI nucleotide databases using the search program termed as "Search for short, nearly exact sequences" with $(CGG)_7$ as a query (http://www.ncbi.nlm.nih.gov/BLAST/). A full screening of the genes of interest was confirmed because genes with exact matches (bit score 42) were followed by the genes with partial matches (bit score less than 42). Genes for unidentified coding sequences were excluded from further study.

Statistical analysis

Genotype distributions and allele frequencies of CGG repeat numbers were compared between cases and controls by means of the $\varkappa^2$ test. Odds ratios (OR) and 95% confidence intervals (95% CI) were calculated by Wolf's method.

Results and discussion

The *PTCH* gene has two alternative first exons—exon 1a and exon 1b (Fig. 1A). Exon 1b contains the first in-frame methionine codon, while exon 1a is a noncoding exon. We noticed a CGG trinucleotide repeat located 4 bp upstream of the first methionine codon. Although exon 1b is a coding exon, mutations in this exon have not been reported. In the course of analyzing mutations in exon 1b using samples from NBCCS individuals that do not have a mutation elsewhere in *PTCH*, we discovered a novel polymorphism involving the CGG trinucleotide repeat (Fig. 1B). The major allele contained

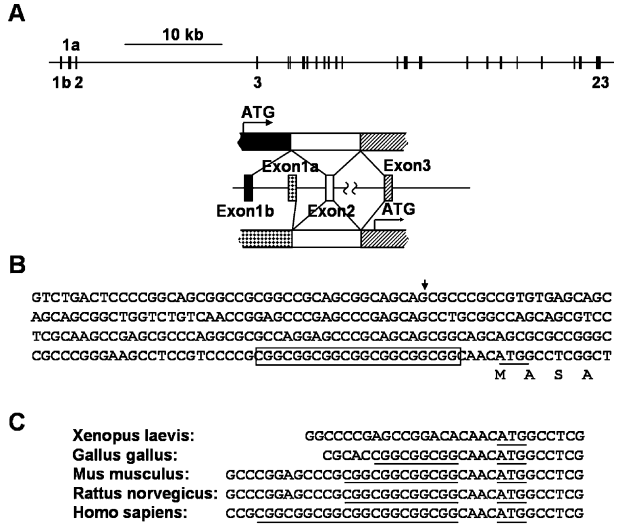

Fig. 1 **A** Genomic organization of human *PTCH*. The *PTCH* locus based on the sequence AL161729 is shown at the *top*. Two cDNA sequences, GenBank U43148 and U59464, are generated by alternative splicing using exon 1a and 1b, respectively, as schematically depicted at the *bottom*. **B** Nucleotide sequence of the human *PTCH* gene including 5'-UTR and exon 1b. The first methionine codon is *underlined*. Polymorphic CGG repeat is *boxed*. The putative transcription start site is indicated by an *arrowhead*. **C** Nucleotide sequence alignment of the *PTCH* genes. CGG repeat and the first methionine codon are underlined

seven repeats of CGG, while the minor one contained eight (Table 1). As far as we examined, we did not find repeat numbers other than seven or eight. The repeat is conserved among vertebrates, since chicken, mouse, and rat *PTCH* contain four or five repeats of CGG (Fig. 1C). However, the repeat has not been found in Xenopus *PTCH*, indicating it is not conserved in amphibians.

Abnormal expansion of the CGG triplet repeat in the 5'-UTR of the *fragile X mental retardation-1* (*FMR1*) gene is responsible for fragile X syndrome, in which the repeat is abnormally hypermethylated, resulting in the silence of the *FMR1* (reviewed by Jin et al. 2000). Since CGG repeat in *PTCH* is immediately upstream of the first in-frame methionine codon, the repeat number may influence the efficiency of translation as well as of transcription. To address this issue, various lengths of $(CGG)_nCAAC$ were subcloned into the luciferase plasmid pGV-P2 between the SV40 promoter and the coding sequence for luciferase (Fig. 2A), and luciferase assays were performed. Luciferase activities gradually increased with the number of CGG repeats, at least within the range we examined. The highest level of luciferase activity was obtained when cells were transfected with the plasmid pGV-$(CGG)_{19}CG(CGG)_6$, which was generated by chance during PCR reaction (Fig. 2B). These results suggest that individuals with $(CGG)_8/(CGG)_8$ have higher levels of PTCH protein expression than those with $(CGG)_7/(CGG)_7$. This is contradictory to the case of *FMR1*. However, it should be noted that in fragile X syndrome, the repeat is massively expanded over 230, and the repeat is located more than 50 bp upstream of the first methionine codon.

To address the question of whether the difference in luciferase activity is transcriptional or translational, the levels of luciferase RNA expression were quantified by a real-time RT-PCR. As shown in Fig. 2C, in contrast to the activities of luciferase, no significant difference in luciferase transcription was observed. Moreover, unexpectedly, the cells transfected with the plasmid pGV-$(CGG)_{19}CG(CGG)_6$ expressed significantly lower levels of luciferase RNA. Therefore, the increase in luciferase activities with the expansion of the CGG repeat is due to the increased efficiency of translation.

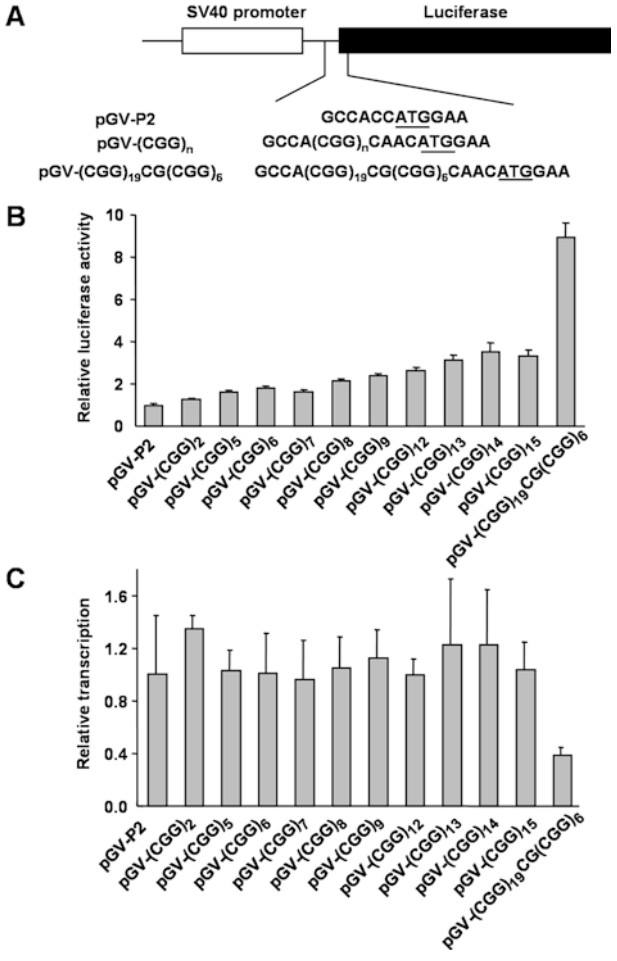

Fig. 2 **A** Schematic depiction of reporter gene constructs used for a luciferase assay. Nucleotide sequences inserted between SV40 promoter and the luciferase gene are indicated at the *bottom*. The first methionine codon of the luciferase gene is *underlined*. **B** The effect of the repeat length on luciferase activities; 293 cells transfected with plasmids indicated at the *bottom* were harvested 24 h after the transfection and subjected to a luciferase assay. **C** The effect of the repeat length on luciferase transcriptions. Total RNA was extracted from 293 cells transfected with plasmids indicated at the *bottom* and subjected to a real-time RT-PCR. Luciferase transcriptions were normalized by those of *GAPDH*

Table 1 Genotype data of a $(CGG)_n$ on the *PTCH* gene

	Controls	NBCCS
Major allele (repeat number: 7) [%]	88 [86.3]	25 [89.3]
Minor allele (repeat number: 8) [%]	14 [13.7]	3 [10.7]
Total	102 [100.00]	28 [100.00]
Major homozygous [%]	39 [76.5]	12 [85.7]
Heterozygous [%]	10 [19.6]	1 [7.1]
Minor homozygous [%]	2 [3.9]	1 [7.1]
Total	51 [100.0]	14 [100.0]
χ^2 [*P*]		
Genotype frequency (2×3 table)	1.38 [0.50]	
Allele frequency (major versus minor)	0.18 [0.68]	
Major homozygous versus others	0.56 [0.46]	
Minor homozygous versus others	0.26 [0.61]	
Odds ratio [95% CI]		
Major homozygous versus others	1.85 [0.36–9.43]	
Minor homozygous versus others	1.88 [0.16–22.44]	
Major allele versus minor allele	0.75 [0.20–2.83]	

Table 2 CGG/CCG-containing genes. *UTR* untranslated region

	$(CGG)_n$	$(CCG)_n$	Total
CDS	39	29	68
5' UTR	95[a]	51[a]	146
Total	134	80	214

Genes with a repeat of seven or more CGG/CCG were downloaded from NCBI nucleotide databases:
[a]The number of CGG-containing genes are significantly higher than CCG-containing genes (χ^2 test, $P = 0.00027$)

Table 3 $(CGG)_n/(CCG)_n$-containing genes in which the triplet repeat is located immediately upstream of the first in-frame methionine codon

Accession no.[a]	Gene[b]	Nucleotides between $(CGG)_n/(CCG)_n$ and ATG	CGG or CCG	n
BC029158	Clone MGC:34313 IMAGE:5198758	CCCCCGGGGC	CCG	10
NM_025075	Hypothetical protein FLJ23445	CGCACGCC	CCG	8
BC015930	Clone MGC:8881 IMAGE:3920963	CGGGGGCC	CCG	8
NM_013396	Ubiquitin specific protease 25 (USP25)	CGGGGGCC	CCG	8
NM_013272	Solute carrier family 21 (organic anion transporter), member 11 (SLC21A11)	GGGAAGG	CGG	8
NM_005360	Transcription factor C-MAF (c-maf)	CAGGAGA	CGG	7
NM_004699	DNA segment on chromosome X (unique) 9928 expressed sequence (DXS9928E) (XAP-5)	CTGCC	CCG	9
NM_000264	PTCH	CAAC	CGG	7
NM_145296	TSLC1-like 2 (TSLL2)	CACC	CGG	7
NM_002958	RYK receptor-like tyrosine kinase (RYK)	CCC	CGG	7
NM_017811	Ubiquitin-conjugating enzyme E2R 2 (UBE2R2)	CG	CCG	7
NM_173054	Reelin (RELN)	C	CGG	8–10[c]

[a]One representative accession number for one gene
[b]Genes that have less than ten nucleotides between triplet repeat and the first methionine
[c]Repeat number varies depending on deposited sequences

The distributions of genotypes that we observed in NBCCS patients and controls did not differ from the expected frequencies under the assumption of Hardy-Weinberg equilibrium (data not shown), nor were significant associations with NBCCS observed (Table 1). Thus far, no genotype-phenotype correlation between the position of mutations and major clinical features of NBCCS is evident (Wicking et al. 1997b). Since developmental defects associated with the disorder are most likely due to haploinsufficiency, and the repeat length potentially alter the expression levels of PTCH, the repeat number may have an effect on the severity of the disease. It would also be interesting to examine the association of the repeat number with sporadic or non-inherited basal cell carcinoma or medulloblastoma, since PTCH acts as a tumor suppressor in these tumors (reviewed by Hunter 1997).

In order to find other genes with CGG repeats, we next performed a genome-wide screening of CGG/CCG-containing genes from NCBI nucleotide databases. A total of 214 genes having seven or more of the repeat number were downloaded. A complete list of the CGG/CCG-containing genes can be obtained from our Web site, http://genetics.nch.go.jp/supplements.htm. Of those 214 genes, 146 (68.2%) contained the repeat in the 5'-UTR (Table 2). Interestingly, significantly more genes have CGG repeats than CCG repeats (65.1% versus 34.9%, $P = 0.00027$). More significantly, none of the downloaded genes contained repeats in the 3'-UTR. The genes containing CGG/CCG repeats in close proximity to their first methionine codons are listed in Table 3. Only five genes including *PTCH* have intervening sequences of up to 4 bp between $(CGG)_n/(CCG)_n$ and ATG. In this regard, *PTCH* is quite unique in terms of the location of the repeat. Considering our results, polymorphisms of the repeat number that might exist in these genes potentially affect their expression levels.

Acknowledgements We would like to thank the patients and family members as well as the clinicians who contributed to this study. We gratefully acknowledge assistance from Mami U, Kaori Inoue, and Kayoko Saito. This study was supported in part by grants for Brain Research, Pediatric Research, Cancer Research, and Genome Research from the Ministry of Health and Welfare, and a Grant-in-Aid for Scientific Research and the Budget for Nuclear Research from the Ministry of Education, Culture, Sports, Science, and Technology.

References

Fujii K, Miyashita T, Takanashi J, Sugita K, Kohno Y, Nishie H, Yasumoto S, Furue M, Yamada M (1999) γ-irradiation deregulates cell cycle control and apoptosis in nevoid basal cell carcinoma syndrome-derived cells. Jpn J Cancer Res 90:1351–1357

Fujii K, Kohno Y, Sugita K, Nakamura M, Moroi Y, Urabe K, Furue M, Yamada M, Miyashita T (2003a). Mutations in the human homologue of *Drosophila patched* in Japanese nevoid basal cell carcinoma syndrome patients. Hum Mutat 21:451–452

Fujii K, Miyashita T, Omata T, Kobayashi K, Takanashi J, Kouchi K, Yamada M, Kohno Y (2003b) Gorlin syndrome with ulcerative colitis in a Japanese girl. Am J Med Genet 121:65–68

Gorlin RJ (1987) Nevoid basal-cell carcinoma syndrome. Medicine 66:98–113

Hahn H, Wicking C, Zaphiropoulous PG, Gailani MR, Shanley S, Chidambaram A, Vorechovsky I, Holmberg E, Unden AB, Gillies S, Negus K, Smyth I, Pressman C, Leffell DJ, Gerrard B, Goldstein AM, Dean M, Toftgard R, Chenevix-Trench G, Wainwright B, Bale AE (1996) Mutations of the human homolog of Drosophila *patched* in the nevoid basal cell carcinoma syndrome. Cell 85:841–851

Hunter T (1997) Oncoprotein networks. Cell 88, 333–346

Imai Y, Matsushima Y, Sugimura T, Terada M (1991) A simple and rapid method for generating a deletion by PCR. Nucleic Acids Res 19:2785

Jin P, Warren ST (2000) Understanding the molecular basis of fragile X syndrome. Hum Mol Genet 9:901–908

Johnson RL, Rothman AL, Xie J, Goodrich LV, Bare JW, Bonifas JM, Quinn AG, Myers RM, Cox DR, Epstein EH, Jr, Scott MP (1996). Human homolog of *patched*, a candidate gene for the basal cell nevus syndrome. Science 272:1668–1671

Kimonis VE, Goldstein AM, Pastakia B, Yang ML, Kase R, DiGiovanna JJ, Bale AE, Bale SJ (1997) Clinical manifestations in 105 persons with nevoid basal cell carcinoma syndrome. Am J Med Genet 69:299–308

Shikama Y, U M, Miyashita T, Yamada M (2001) Comprehensive studies on subcellular localizations and cell death-Inducing activities of eight GFP-tagged apoptosis-related caspases. Exp Cell Res 264:315–325

Wicking C, Gillies S, Smyth I, Shanley S, Fowles L, Ratcliffe J, Wainwright B, Chenevix-Trench G (1997a) De novo mutations of the *patched* gene in nevoid basal cell carcinoma syndrome help to define the clinical phenotype. Am J Med Genet 73:304–307

Wicking C, Shanley S, Smyth I, Gillies S, Negus K, Graham S, Suthers G, Haites N, Edwards M, Wainwright B, Chenevix-Trench G (1997b). Most germ-line mutations in the nevoid basal cell carcinoma syndrome lead to a premature termination of the PATCHED protein, and no genotype-phenotype correlations are evident. Am J Hum Genet 60:21–26