Hum Genet (2000) 106 : 370–373
Digital Object Identifier (DOI) 10.1007/s004390000255

SHORT REPORT

Marjan Huizing · Yair Anikster · William A. Gahl

Characterization of a partial pseudogene homologous to the Hermansky-Pudlak syndrome gene *HPS-1*; relevance for mutation detection

Received: 22 November 1999 / Accepted: 18 January 2000 / Published online: 16 February 2000

Abstract The *HPS-1* gene is the first gene found to be responsible for the autosomal recessive disorder Hermansky-Pudlak syndrome (HPS). HPS is characterized by oculocutaneous albinism, a platelet storage pool deficiency, and ceroid lipofuscinosis. The *HPS-1* gene has been mapped to chromosome 10q23.1–23.3 and encodes a 79-kDa protein of unknown function with no homology to any known protein. A sequence database search has revealed that a portion of clone HS1119A7 shows high sequence similarity to *HPS-1* cDNA. By performing sequence alignments and PCR amplification of cDNA from several human tissues, we have shown that part of this clone consists of an unprocessed partial *HPS-1* pseudogene located on chromosome 22q12.2–12.3. The pseudogene contains several intact *HPS-1* exons and shows 95% sequence homology to the *HPS-1* cDNA. Exon 6 of the pseudogene has 100% sequence homology to exon 6 of *HPS-1* itself. In the pseudogene, this exon is surrounded by portions of both its normal flanking introns. These data provide the first characterization of an *HPS-1* pseudogene, called *HPS1-ψ1*. During amplification of exon 6 of the *HPS-1* gDNA for mutation identification, the pseudogene might also be amplified, leading to a false positive for mutation. In addition, amplification of specific parts of the *HPS-1* cDNA (e.g., exons 2–5) for mutation detection might lead to false positives for mutations, if the cDNA is contaminated with gDNA. This calls for caution when employing these screening approaches.

M. Huizing · Y. Anikster · W. A. Gahl
Section on Human Biochemical Genetics,
Heritable Disorders Branch,
National Institute of Child Health and Human Development,
National Institutes of Health, Bethesda, MD 20892, USA

W. A. Gahl (✉)
10 Center Drive, MSC 1830, Building 10, Room 9S-241,
NICHD, NIH, Bethesda, MD 20892–1830, USA
e-mail: bgahl@helix.nih.gov,
Tel.: +1 301 402 2739, Fax: +1 301 402 2740

Introduction

Hermansky-Pudlak syndrome (HPS; MIM no. 20330) is an autosomal recessive disorder characterized by oculocutaneous albinism, a platelet storage pool deficiency, and ceroid lipofuscinosis (Hermansky and Pudlak 1959; King et al. 1995). Although rare in the general population, HPS occurs in northwest Puerto Rico with a prevalence of 1 in 1800 (Witkop et al. 1990). *HPS-1* was the first gene found to be responsible for HPS (Oh et al. 1996). The gene, located on chromosome 10q23.1–23.3 (Fukai et al. 1995; Wildenberg et al. 1995), produces a 3.6-kb transcript that codes for a 700-amino-acid peptide and is ubiquitously expressed (Oh et al. 1996; Bailin et al. 1997). An alternate 1.5-kb transcript codes for a 324-amino-acid peptide and is expressed only in bone marrow and melanoma cells (Wildenberg et al. 1998). Several mutations in *HPS-1* have been reported, including a 16-bp duplication within exon 15 found exclusively in patients from northwest Puerto Rico (Oh et al. 1996, 1998; Gahl et al. 1998; Shotelersuk and Gahl 1998; Shotelersuk et al. 1998; Oetting and King 1999). A number of polymorphisms have also been reported for *HPS-1* (Oetting and King 1999).

The function of the HPS protein is unknown, and its sequence shows no homology to any known protein. A database search for sequence similarities of *HPS-1* has revealed a human clone with an extremely high sequence similarity. We have characterized this clone and found it to be an unprocessed partial pseudogene of *HPS-1*, *HPS1-ψ1*, located on chromosome 22q12.2–12.3 and containing several intact *HPS-1* exons. Recognition of the existence of this pseudogene may help prevent misinterpretation of mutation analyses in Hermansky-Pudlak syndrome patients.

Materials and methods

Database screening

Sequence homology searches were performed by means of BLAST (basic local alignment search tool) computer-based analyses, made available through the National Center for Biotechnology Information (Altschul et al. 1990, 1997).

Primers, polymerase chain reaction, and sequencing

The sequences of the primers used for polymerase chain reaction (PCR) amplifications were: Ex2F (nt 154–175 on *HPS-1*) 5'-TGCAGCCCTTTCTGAACCTCTG-3'; Ex5F (nt 464–484) 5'-TGGAGAATGCCTGTTCATTGC-3'; Ex5R (nt 489–470) 5'-TGATGGCAATGAACAGGCAT-3'; Ex6R (nt 673–654): 5'-TAGGTCCACAGCAGGCTCTG-3'. PCR amplifications were performed in 50-µl reactions containing 1.5 mM $MgCl_2$, 10 mM TRIS-HCl pH 8.3, 50 mM KCl, 200 µM each dNTP, 0.2 µM each primer, 50–100 ng DNA, and 2.5 U Platinum *Taq* DNA polymerase (Gibco BRL, Rockville, Md.). PCR cycling consisted of an initial denaturation step at 96°C for 4 min, followed by 35 cycles each with a denaturing step at 94°C for 40 s, an annealing step at 53°C for 40 s, and an extension step at 72°C for 2 min. Amplification cycles were followed by an elongation step at 72°C for 10 min. The PCR products were electrophoresed in 1% Agarose and were stained with ethidium bromide.

For sequencing, PCR products were purified from agarose gels by using the GeneClean II kit (Bio101, Vista, Calif.). Sequencing reactions were performed by the Applied Biosystems BigDye Terminator Cycle Sequencing Ready Reaction Kit according to the manufacturor's guidelines (PE Applied Biosystems, Foster City, Calif.). Reaction products were ethanol-precipitated, resuspended in formamide-loading buffer, and electrophoresed on an ABI 377 automated sequencer.

Tissue and chromosome screening

Multiple tissue cDNA screening was performed by using Human Multiple Tissue cDNA (MTC) Panel I (Clontech, Palo Alto, Calif.), containing eight normalized first-strand cDNA preparations derived from various adult human tissues. For screening the gDNA of separate chromosomes, we employed the NIGMS Human Somatic Cell Hybrid Panel no. 2 (Coriell Cell Repositories, Camden, N.J.). "Total" cDNA and gDNA consisted of control DNA that contained all the human chromosomes and was provided with the panels.

Results

A BLAST database search for sequence similarity to the *HPS-1* cDNA sequence (GenBank accession no. HSU65676) identified clone HS1119A7 (GenBank accession no. AL022313) with high sequence similarity to portions of *HPS-1*. This clone was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group (http://www.sanger.ac.uk/HGP/Chr22). Clone HS1119A7 was assigned by the Sanger Centre to chromosomal location 22q12.2–12.3.

As shown in Fig. 1, the part of the clone overlapping *HPS-1* is 3208 bp in length and starts with the last 10 bp of the intron 1 sequence, followed by a sequence similar to *HPS-1* cDNA extending from nt 102 in exon 2 through nt 542 in exon 5. This region is further referred to as stretch 1. After stretch 1, the clone has a 508-bp sequence that is similar to the last portion of *HPS-1* intron 5. The next region in the clone (referred to as stretch 2) contains the complete exon 6 sequence (nt 606–713). The sequence of the clone after stretch 2 is similar to the first 2137 bp of *HPS-1* intron 6.

In order to define the "breakpoints" in introns 1, 5, and 6 of the pseudogene, we completely sequenced these *HPS-1* introns. Intron 1 was 1396 bp long (GenBank accession no. AF200628). The pseudogene sequence contained the last 10 bp of the *HPS-1* intron 1 sequence (i.e., nt 1387–1396 of intron 1). Intron 5 appeared to be 1180 bp in length (GenBank accession no. AF200629). The last 509 bp of intron 5 (i.e., nt 673–1181 of intron 5) were pre-

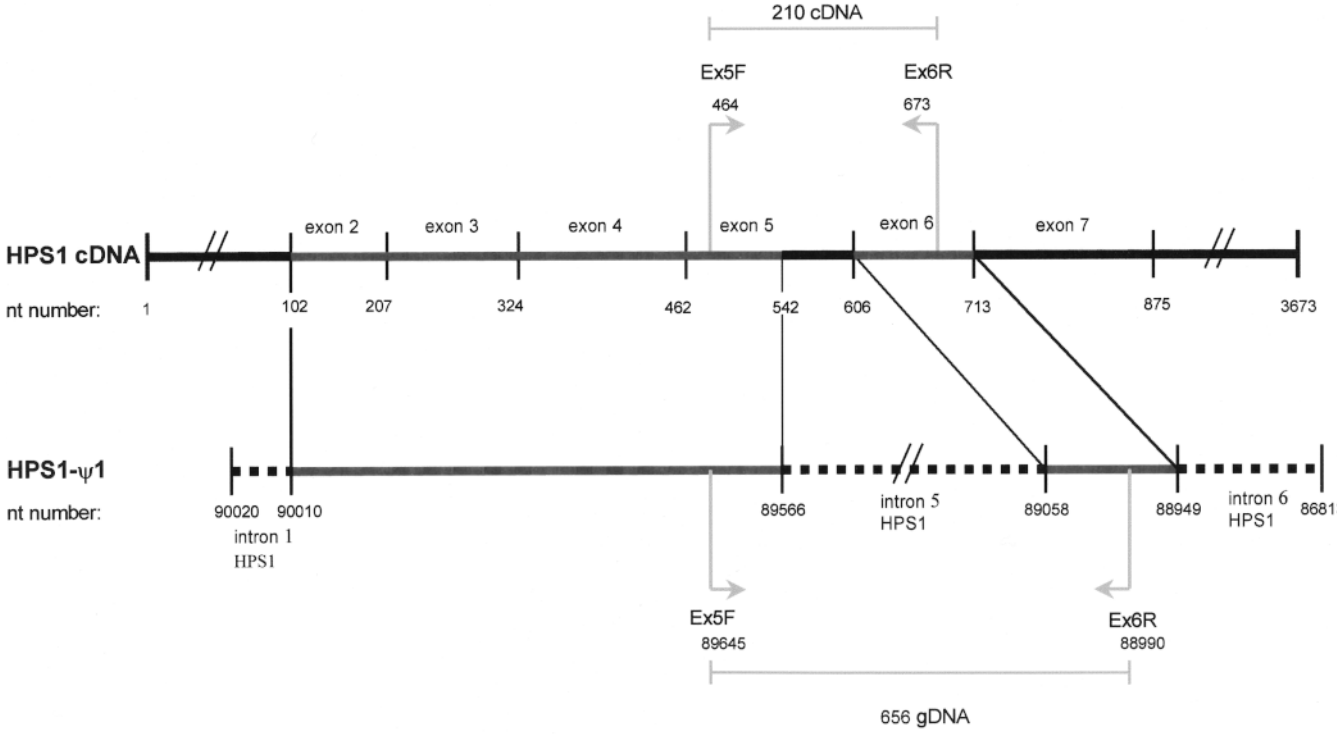


Fig. 1 Organization of the *HPS-1* gene and its pseudogene *HPS1-ψ1* (GenBank accession no. AF201372). PCR-primers Ex5F and Ex6R are indicated. The nucleotide numbering of the *HPS-1* cDNA is according to that for GenBank accession no. HSU65676. The nucleotide numbering of *HPS1-ψ1* is according to that for clone HS1119A7 (GenBank accession no. AL022313), to show the organization of the pseudogene within this clone

sent in the pseudogene. Intron 6 is 2688 bp long (GenBank accession no. AF200630). The first 2137 bp of this intron are present in the pseudogene. Table 1 lists all the nucleotide differences between *HPS-1* and the pseudogene in the coding region.

Table 1 Sequence differences between *HPS-1* (HSU 65676) and its partial pseudogene (*HPS1-ψ 1*) (AF201372)

nt(*HPS-1*→*HPS1-ψ 1*)	*HPS-1* exon	Altered amino acid
127 (C→G)	2	5'UTR
133 (C→G)	2	5'UTR
141 (C→A)	2	5'UTR
147 (G→T)	2	5'UTR
178 (C→T)	2	5'UTR
186 (T→C)	2	5'UTR
200 (insCCT)	2	5'UTR
216 G→A	3	V4I
229 C→T	3	T8I
237 G→A	3	A11T
244 T→C	3	V13A
260 A→G	3	T18T
264 C→G	3	Q20E
286 G→A	3	R27Q
335 G→A	4	L43L
336 G→C	4	E44Q
351 A→C	4	T49P
427 C→T	4	T74M
434 T→C	4	N76N
435 G→A	4	G77S
442 T→C	4	F79S
494 T→C	5	G96G
504 G→A	5	E100K
523 G→A	5	R106Q
536 T→C	5	Y110Y
537 G→A	5	V111M

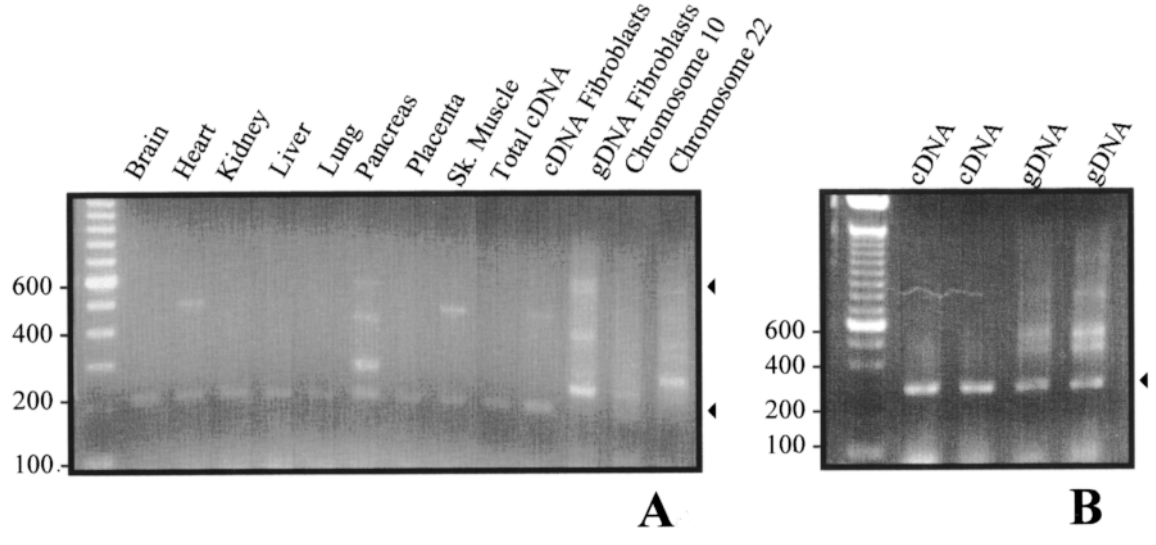
The presence or absence of the *HPS-1* pseudogene can be determined by performing PCR amplification with a forward primer in exon 5 (Ex5F) and a reverse primer in exon 6 (Ex6R). The presence of *HPS-1* cDNA results in a 210-bp band, whereas the presence of the pseudogene results in a 656-bp band (see Figs. 1, 2). The Ex5F-Ex6R PCR products were determined for a panel of eight normalized first-strand cDNA preparations derived from various adult human tissues, together with cDNA derived from fibroblasts and total cDNA (Fig. 2A). All revealed only the 210-bp *HPS-1* cDNA band, which was confirmed by sequencing. Therefore, only the *HPS-1* gene was expressed in these tissues; the pseudogene was not processed. In contrast, the Ex5F-Ex6R PCR products of human fibroblast gDNA and chromosome 22 gDNA included the 656-bp band of the pseudogene (Fig. 2A). The PCR product of chromosome 10 yielded no significant band, because the *HPS-1* gDNA between the primers was too large to be amplified. The extra bands amplified for pancreas and skeletal muscle are nonspecific PCR products, as determined by sequencing.

Figure 2B shows an example of misinterpreted mutation-detection results. The PCR amplification in Fig. 2B was performed on both cDNA and gDNA with primers Ex2F and Ex5R. With cDNA, a 336-bp band representing the *HPS-1* gene was produced. With gDNA, a 336-bp band was also produced, but this band represented the pseudogene sequence on chromosome 22, as confirmed by sequencing. The PCR product of the *HPS-1* gene itself is too large (~ 9.3 kb) to be amplified by using these primers.

Fig. 2A, B PCR amplification of *HPS-1* and its pseudogene *HPS1-ψ1*. **A** Products of PCR amplification with primers Ex5F-Ex6R, yielding a 210-bp product for the *HPS-1* cDNA and a 656-bp product for the pseudogene (*arrowheads*). Amplification products are shown for eight different human tissue cDNAs, and "total" cDNA and gDNA (i.e., DNA from cells containing all the human chromosomes), chromosome-10-specific gDNA, and chromosome-22-specific gDNA. **B** Products of PCR amplification of gDNA and cDNA with primers Ex2F-Ex5R, yielding a 336-bp product for both cDNA and gDNA (*arrowhead*)

Discussion

This study presents the structure of a partial unprocessed *HPS-1* pseudogene, which we denote as *HPS1-ψ1* (GenBank accession no. AF201372). *HPS1-ψ1* is not expressed in the nine different human tissues examined, has



sequence similarities to two stretches of the *HPS-1* cDNA, and contains three partial intron sequences.

The pseudogene spans portions of the *HPS-1* gene (Fig. 1). The portion spanning exons 2–4 and part of exon 5 (referred to as stretch 1) contains a sequence in the coding region that differs from the corresponding region of *HPS-1* (Table 1). Stretch 1 of the pseudogene, which is composed of 444 nucleotides, has apparently accumulated 28 random point mutations, resulting in 93.7% homology of stretch 1 to *HPS-1*. In contrast, stretch 2 in the pseudogene exhibits 100% sequence homology to *HPS-1* exon 6. This might be explained by the finding that exon 6 of the pseudogene is surrounded by substantial portions of its original introns, so that the fidelity of its sequence is protected by standard DNA repair mechanisms. In general, transcriptionally active genes are repaired preferentially compared with nontranscribed genes such as a pseudogene, whose repair rate resembles that of the rest of the genome (Leadon and Snowden 1988). Presumably, a combination of factors, including the RNA polymerase, type of repair apparatus, and genomic structure, contribute to determining the repair rate and fidelity of genomic sequences.

In searching for the mechanism of the formation of this pseudogene, we have looked for transposable elements flanking the pseudogene (Hernandez et al. 1999). An Alu repeat exists 160 bp upstream of the 5' border of the pseudogene, and another Alu repeat lies 157 bp downstream of the 3' border. These sequences, together with an Alu repeat and MER3, MER5, and MER33 repeats in intron 6, may play some undetermined role in the formation of this pseudogene.

To date, all reported *HPS-1* mutations have occurred in the gene itself, not in the pseudogene. This will continue to be the case as long as mutation detection for *HPS-1* is performed as described by Bailin et al. (1997), with PCR amplification of every exon separately and by using primers in the introns. With this method, the pseudogene will not be amplified. Particular attention should be paid when amplifying exon 6 with primers in its surrounding introns. The primers described by Bailin et al. (1997) involve a region with a high rate of mismatch between the pseudogene and the functional gene and allow the amplification of only the functional *HPS-1* exon without any "contaminating" sequences of the pseudogene. This should be kept in mind when designing custom primers for specific mutation detection of exon 6.

Mutation analysis at the cDNA level might be complicated by PCR coamplification of pseudogene sequences. When screening with a cDNA primer-pair spanning exons 2–5 (nt 102–542) or exon 6 (nt 606–713), and if the cDNA is contaminated with gDNA, the pseudogene will be amplified as well. This may lead to false positive results of mutation analysis for *HPS-1* and calls for caution when employing such screening approaches.

Acknowledgements Y. Anikster is a Howard Hughes Medical Institute Physician Postdoctoral Fellow. The authors appreciate the advice and assistance of V. Shotelersuk and E. Sidransky.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bailin T, Oh J, Feng GH, Fukai K, Spritz RA (1997) Organization and nucleotide sequence of the Hermansky-Pudlak syndrome (HPS) gene. J Invest Dermatol 108:923–927

Fukai K, Oh J, Frenk E, Almodovar C, Spritz RA (1995) Linkage disequilibrium mapping of the gene for Hermansky-Pudlak syndrome to chromosome 10q23.1–23.3. Hum Mol Genet 4:1665–1669

Gahl WA, Brantly M, Kaiser-Kupfer MI, Iwata F, Hazelwood S, Shotelersuk V, Duffy LF, Kuehl EM, Troendle J, Bernardini I (1998) Genetic defects and clinical characteristics of patients with a form of oculocutaneous albinism (Hermansky-Pudlak syndrome). N Engl J Med 338:1258–1264

Hermansky F, Pudlak P (1959) Albinism associated with hemorrhagic diathesis and unusual pigmented reticular cells in the bone marrow: report of two cases with histochemical studies. Blood 14:162–169

Hernandez JM, Hernandez CS, Giner CP, Donat V, Hernandez-Yago J (1999) Identification of ψ3Tom20, a novel processed pseudogene of the human Tom20 gene, and complete characterization of ψ1Tom20 and ψ2Tom20. Mol Gen Genet 262:207–211

King RA, Hearing VJ, Creel DJ, Oetting WS (1995) In: Scriver CR, Beaudet AL, Sly WS, Valle DL (eds) The metabolic and molecular bases of inherited disease, vol 3, 7th edn. McGraw-Hill, New York, pp 4353–4392

Leadon SA, Snowden MM (1988) Differential repair of DNA damage in the human metallothionein gene family. Mol Cell Biol 8:5331–5338

Oetting WS, King RA (1999) Molecular basis of albinism: mutations and polymorphisms of pigmentation genes associated with albinism. Hum Mutat 13:99–115

Oh J, Bailin T, Fukai K, Feng GH, Ho L, Mao J-i, Frenk E, Tamura N, Spritz RA (1996) Positional cloning of a gene for Hermansky-Pudlak syndrome, a disorder of cytoplasmic organelles. Nat Genet 14:300–306

Oh J, Ho L, Ala-Mello S, Amato D, Armstrong L, Bellucci S, Carakushansky G, Ellis JP, Fong C-T, Green JS, Heon E, Legius E, Levin AV, Nieuwenhuis HK, Pinckers A, Tamura N, Whiteford ML, Yamasaki H, Spritz RA (1998) Mutation analysis of patients with Hermansky-Pudlak syndrome: a frameshift hot spot in the HPS gene and apparent locus heterogeneity. Am J Hum Genet 62:593–598

Shotelersuk V, Gahl W (1998) Hermansky-Pudlak syndrome: models for intracellular vesicle formation. Mol Genet Metab 15:85–96

Shotelersuk V, Hazelwood S, Larson D, Iwata F, Kaiser-Kupfer MI, Kuehl E, Bernardini I, Gahl WA (1998) Three new mutations in a gene causing Hermansky-Pudlak syndrome: clinical correlations. Mol Genet Metab 64:99–107

Wildenberg SC, Oetting WS, Almodovar C, Krumwiede M, White JG, King RA (1995) A gene causing Hermansky-Pudlak syndrome in a Puerto Rican population maps to chromosome 10q2. Am J Hum Genet 57:755–765

Wildenberg SC, Fryer JP, Gardner JM, Oetting WS, Brilliant MH, King RA (1998) Identification of a novel transcript produced by the gene responsible for the Hermansky-Pudlak syndrome in Puerto Rico. J Invest Dermatol 110:777–781

Witkop CJ, Babcock MN, Rao GHR, Gaudier F, Summers CG, Shanahan F, Harmon KR, Townsend DW, Sedano HO, King RA, Cal SX, White JG (1990) Albinism and Hermansky-Pudlak syndrome in Puerto Rico. Bol Asoc Med P R Agosto 82:333–339