ELSEVIER

Gene 225 (1998) 9–16

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Glypican 3 and glypican 4 are juxtaposed in Xq26.1

Reid Huber [a,*], Richard Mazzarella [b], Chun-Nan Chen [c], Ellson Chen [c], Maggie Ireland [d], Susan Lindsay [d], Giuseppe Pilia [e], Laura Crisponi [e]

[a] *Laboratory of Genetics, National Institute on Aging, NIH, Baltimore, MD 21224, USA*
[b] *Institute for Biomedical Computing and Center for Genetics in Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA*
[c] *Applied Biosystems Division, Advanced Center for Genetic Technology, Foster City, CA 94404, USA*
[d] *Department of Human Genetics, University of Newcastle upon Tyne, Newcastle upon Tyne, NE1 7RU, UK*
[e] *IRTAM CNR, Universita' di Cagliari, Ospedale Microcitemico, Via Jenner, Cagliari, Italy*

Received 16 August 1998; received in revised form 26 October 1998; accepted 26 October 1998; Received by V. Larionov

Abstract

Recently, we have shown that mutations in the X-linked glypican 3 (*GPC3*) gene cause the Simpson–Golabi–Behmel overgrowth syndrome (SGBS; Pilia et al., 1996). The next centromeric gene detected is another glypican, glypican 4 (*GPC4*), with its 5′ end 120 763 bp downstream of the 3′ terminus of *GPC3*. One recovered *GPC4* cDNA with an open reading frame of 1668 nt encodes a putative protein containing three heparan sulfate glycosylation signals and the 14 signature cysteines of the glypican family. This protein is 94.3% identical to mouse GPC4 and 26% identical to human GPC3. In contrast to *GPC3*, which produces a single transcript of 2.3 kb and is stringently restricted in expression to predominantly mesoderm-derived tissues, Northern analyses show that *GPC4* produces two transcripts, 3.4 and 4.6 kb, which are very widely expressed (though at a much higher level in fetal lung and kidney). Interestingly, of 20 SGBS patients who showed deletions in *GPC3*, one was also deleted for part of *GPC4*. Thus, *GPC4* is not required for human viability, even in the absence of *GPC3*. This patient shows a complex phenotype, including the unusual feature of hydrocephalus; but because an uncle with SGBS is less affected, it remains unclear whether the *GPC4* deletion itself contributes to the phenotype.

Keywords: X chromosome; Proteoglycan; Simpson–Golabi–Behmel Syndrome

1. Introduction

The fundamental mechanisms of genome evolution often involve the duplication of a sequence at a nearby site (Ohno, 1970). The copies of the sequences then diverge, but may be coregulated in complex ways (Mazzarella and Schlessinger, 1998). In some instances, multiple copies are involved in genetic pathology, as in the case of two sulfate transporters on chromosome 7, lesions that cause, respectively, chloride diarrhea (Hoglund et al., 1996) and Pendred's syndrome (Everett et al., 1997). Thus, similar juxtaposed genes can be of considerable interest in the analysis of disease etiology.

Recently, we have cloned a gene, glypican 3 (*GPC3*), mutations in which cause the X-linked Simpson–Golabi–Behmel overgrowth syndrome (SGBS) (Pilia et al., 1996). This gene encodes a member of the glypican-related integral membrane proteoglycan (GRIPS) family [reviewed by David (1993)]. GRIPS are distinguished from other proteoglycans by their attachment to the cell membrane by a glycosylphosphatidylinositol anchor, which is added to the C-terminus of the GRIP in the endoplasmic reticulum.

Currently, there are five members of the GRIPS family of proteoglycans, GPC1-5 (David et al., 1990; Watanabe et al., 1995; Pilia et al., 1996; Veugelers et al., 1997; Huber et al., 1997) that have been isolated from various species. GPC3, however, is the only member of the family thus far implicated in the etiology of an inherited disorder. Although the exact mechanism of

* Corresponding author. Present address: Applied Biotechnology, DuPont Pharmaceuticals Company, PO Box 80336, Experimental Station, E336/225, Wilmington, DE 19880, USA.
E-mail: Reid.M.Huber@dupontpharma.com

Abbreviations: GPC3, glypican 3; GPC4, glypican 4; SGBS, Simpson–Golabi–Behmel Syndrome.

0378-1119/98/$ – see front matter
PII: S0378-1119(98)00549-6

GPC3 involvement in overgrowth remains unclear, several recent studies suggest that GPC3 may be able to bind to and sequester insulin-like growth factor-2 (Pilia et al., 1996; Xu et al., 1998) or fibroblast growth factor-2 (Song et al., 1997). Additionally, there are indications that overexpression of *GPC3* in some cell lines may be able to induce apoptosis (Gonzalez et al., 1998; Huber, unpublished results).

In a search for the genes neighboring glypican 3 (*GPC3*) in Xq26, we have reported the sequence of a bacterial artificial chromosome [GenBank Accession No. AC002420 (bWXD9); Huber et al., 1997] centromeric to *GPC3*, which clearly shows another glypican-encoding gene, *GPC4*. Here, we report the characterization of human *GPC4* transcripts, compare features of its sequence and expression to *GPC3*, and comment on an SGBS patient and his uncle who both have deletions of *GPC4* as well as a portion of *GPC3*.

2. Materials and methods

2.1. Materials

EST clone AA046130 was obtained from Genome Systems (St. Louis, MO). MCF-7 cells (ATCC#HTB-22) were routinely cultured in DMEM containing 2 mM L-glutamine and 10% fetal calf serum.

2.2. Sequence analysis

Sequence analysis of BACs was carried out according to Huber et al. (1997). Briefly, repetitive sequences were masked using CENSOR (Jurka et al., 1996), and candidate exons were identified using GRAIL2 (Uberbacher and Mural, 1991). Candidate exons were then analyzed using BLAST (Altschul et al., 1990). Two potential exons in bWXD9 (nt 164207–164366 and nt 239782–239940; Accession No. AC002420) were found to have a strong homology to mouse K-glypican (Watanabe et al., 1995) and were studied further. The status of all sequencing in the region, including that telomeric of *GPC3* toward HPRT, can be found at the Center for Genetics in Medicine, Washington University web site at http://www.ibc.wustl.edu/cgm/cgm.html.

2.3. cDNA amplification

The human *GPC4* cDNA was obtained by RT-PCR on total RNA from MCF-7 cells. Briefly, RNA was isolated and reverse-transcribed using the RT system according to the manufacturer's (Promega) instructions. PCR amplification was then carried out using a forward primer from the predicted human *GPC4* exon 2 in bWXD9 (5′-CTGGTTTGCTGGTGTCAAC) and a reverse primer developed from the 3′ EST sequence (5′-TCGCCTCTTCCACCAACTC). Amplification was performed with 35 cycles of 94°C for 60 s, 58°C for 2 min, and 72°C for 3 min. Amplification products were purified by agarose gel electrophoresis and sequenced using Thermosequenase radiolabeled terminator cycle sequencing kit (Amersham). The human *GPC4* cDNA sequence and its putative protein translation have been submitted to GenBank (Accession No. AF064826).

2.4. Physical mapping

A primer pair derived from the EST AA046130 sequence was used to screen YACs in the bWXD9 region. The primer pair sequence is forward: 5′-CAGAGGTCCAGGTTGACAC (nt 1–19 of the GenBank AA046130 sequence); reverse: 5′-CACTTCCACTTCCTTCTCC (nt. 159–141 of GenBank AA046130 sequence). PCR was performed on colony isolated YAC clones, and products were resolved by agarose gel electrophoresis followed by ethidium bromide staining.

2.5. Northern analysis

Northern analysis was performed using multiple tissue Northern blots from Clontech. Hybridizations were carried out according to the manufacturer's instructions using the 800-bp EST AA046130 as a probe. For β-actin controls, membranes were stripped and reprobed with the β-actin cDNA, according to the manufacturer's instructions.

2.6. Mutation analysis

Mutation analysis was performed on patient DNA using the following primer pairs: *GPC3* exon 6 (product size: 70 bp): forward: 5′-CATGAGCTGAAAATGAAGG, reverse: 5′-GGTTAATGTGCTTCAGTTTG; *GPC3* exon 8 (product size: 71 bp): forward: 5′-ATCTGGATGTGGATGATG, reverse: 5′-AAAGGTGCTTATCTCGTTG; *GPC4* 5′ region (GenBank #AF064826; product size: 82 bp): forward: 5′-GGTGATCATTTGAAGATCTG, reverse: 5′-CTTTACTTTGCAGGCTGTAC; *GPC4* 3′ region (GenBank #AF064826; product size: 117 bp): forward: 5′-GAGTGCCAATGAGAAAGC, reverse: 5′-TGAGAATTATCTCCACTCTC. Control reactions were performed with a primer pair from exon 2 of the somatostatin gene (chr. 3; GenBank Accession No. J00306): forward: 5′-TCGGCAGCTGTAAAAACTGG, reverse: 5′-GGGATCAGAGGTCTGATATG. All PCR reactions were 35 cycles of 30 s at 94°C, 45 s at 55°C, and 45 s at 72°C according to the manufacturer's (Perkin-Elmer) recommendations.

3. Results

3.1. GPC4 lies just centromeric of GPC3

Fig. 1 schematizes the approximate location of features of *GPC3* in genomic DNA. The 5′ and 3′ exons of *GPC3* are precisely positioned in bacterial artificial chromosomes (BACs bWXD8 and bWXD9; Huber et al., 1997) The finished sequence of bWXD9 (244 kb; GenBank Accession No. AC002420) contained two sequences that were almost identical to 5′ end sequence tracts in mouse *GPC4* ('K-glypican'; Watanabe et al., 1995). These provided two 5′ exons (the former beginning at the initiator methionine) and the intervening intron of a putative human *GPC4* (Fig. 2a). The precise positions are annotated in the reported sequence of bWXD9 in GenBank, with a putative translational initiation site that begins about 120.7 kb centromeric of the polyadenylation signal of *GPC3* (Fig. 1).

The sequence of the 3′ end of the human *GPC4* transcript was cued by an EST (GenBank Accession No. AA046130) that shows essential identity to the 3′ terminus of mouse *GPC4*. A primer pair/PCR product inferred from the 3′ EST sequence mapped approximately 150 kb centromeric of the first two predicted exons in yeast artificial chromosomes spanning the region (Fig. 1). In particular, this 3′ *GPC4* STS was contained in y2704, y6858, y808, and y440. Based on these data as well as the STS content of YACs in the region, human *GPC4* is juxtaposed centromerically to *GPC3*, spans approximately 150 kb, and, like *GPC3*, is transcribed in the telomeric to centromeric direction. This genomic region in human is currently being sequenced (at the Sanger Centre, Cambridge, UK) and, upon completion, should confirm and complete the partial genomic organization data, including all exon/intron junctions.

With the localization of *GPC4*, the full extent of *GPC3* can be refined. The zone between bWXD8 and bWXD9, containing exons 2–7 of *GPC3*, is still not completely sequenced, but the STS content, fingerprint patterns, and sizing of overlapping large-insert YAC clones indicate that the gene spans approximately 600 kb of genomic DNA (data not shown; Fig. 1).

3.2. Comparison of predicted human GPC4 with mouse GPC4 and human GPC3

Using oligonucleotides from the 5′ exon and the 3′ EST for human *GPC4*, the complete coding sequence was amplified as a 1.9-kb fragment by RT-PCR from mRNA of MCF-7 cells. The sequence of the fragment was determined (GenBank Accession No. AF064826), and is illustrated in Fig. 2b. The open reading frame encodes a predicted protein of 556 amino acids. This predicted protein is 94.3% identical to the 557-amino-acid mouse GPC4 (Fig. 3), and both contain the heparan sulfate glycosylation signals (boldfaced) and the 14 cysteine residues (boxed) that are strictly conserved throughout the glypican family.

Comparisons with human GPC3 are much less striking with an overall amino acid identity of 26% (data

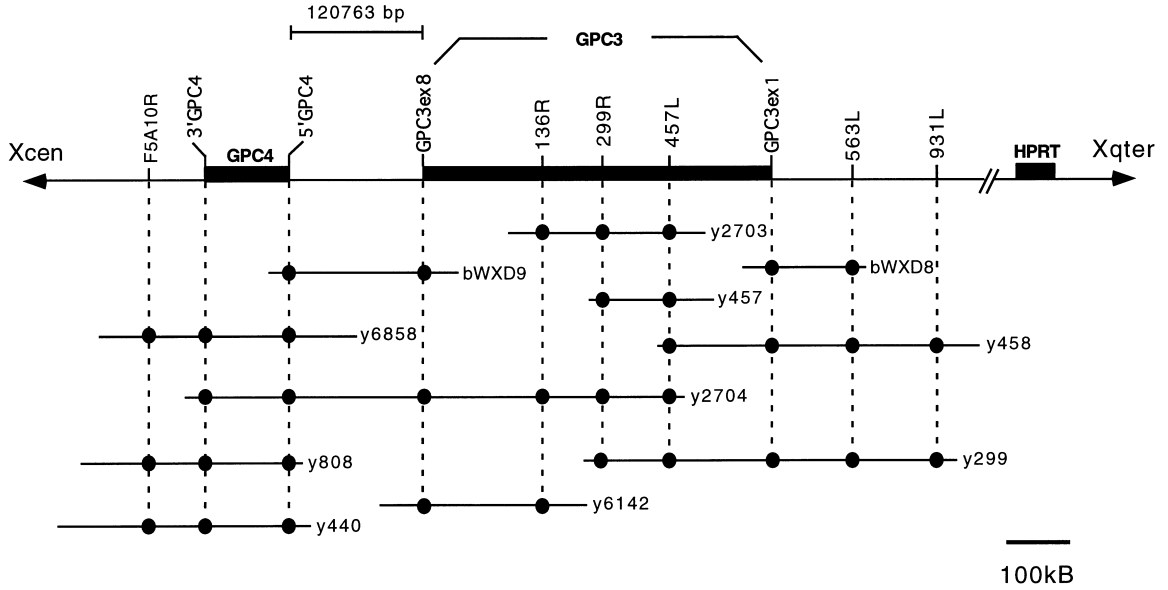

Fig. 1. Genomic region surrounding the human *GPC3* gene. Physical map of the Xq26.1 region containing *GPC3*, *GPC4*, and *HPRT*. Large insert clones are indicated as straight lines and numbered accordingly. STS content is indicated by dashed lines and circles. Both *GPC3* and *GPC4* are transcribed in the telomeric to centromeric direction. The distance between the polyadenylation signal of *GPC3* and the putative translation start site of *GPC4* (see text) was derived from the bWXD9 sequence and is indicated above the map.

A.

```
EXON 1
164207-164366

ATG GCA CGG TTC GGC TTG CCC GCG CTT CTC TGC ACC CTG GCA GTG CTC AGC GCC GCG CTG CTG
 M   A   R   F   G   L   P   V   L   L   C   T   L   A   V   L   S   A   V   L   L

GCT GCC GAG CTC AAG TCG AAA AGT TGC TCG GAA GTG CGA CGT CTT TAC GTG TCC AAA GGC TTC
 A   A   E   L   K   S   K   S   C   S   E   V   R   R   L   Y   V   S   K   G   F

AAC AAG AAC GAT GCC CCC CTC CAC GAG ATC AAC G
 N   K   N   D   A   P   L   H   E   I   N


EXON 2
239782-239940

 GT GAT CAT TTG AAG ATC TGT CCC CAG GGT TCT ACC TGC TGC TCT CAA GAG ATG GAG GAG
 G   D   H   L   K   I   C   P   Q   G   S   T   C   C   S   Q   E   M   E   E

AAG TAC AGC CTG CAA AGT AAA GAT GAT TTC AAA AGT GTG GTC AGC GAA CAG TGC AAT CAT
 K   Y   S   L   Q   S   K   D   D   F   K   S   V   V   S   E   Q   C   N   H

TTG CAA GCT GTC TTT GCT TCA CGT TAC AAG AAG TTT GAT G
 L   Q   A   V   F   A   S   R   Y   K   K   F   D
```

B.

```
   1 ccttctccct ccagctccac tcgctagtcc ccgactccgc cagccctcgg cccgctgccg
  61 tagcgccgct tcccgtccgg tcccaaaggt gggaacgtgt ccgccccggc ccgcaccATG
 121 GCACGGTTCG GCTTGCCCGC GCTTCTCTGC ACCCTGGCAG TGCTCAGCGC CGCGCTGCTG
 181 GCTGCCGAGC TCAAGTCGAA AAGTTGCTCG GAAGTGCGAC GTCTTTACGT GTCCAAAGGC
 241 TTCAACAAGA ACGATGCCCC CCTCCACGAG ATCAACGGTG ATCATTTGAA GATCTGTCCC
 301 CAGGGTTCTA CCTGCTGCTC TCAAGAGATG GAGGAGAAGT ACAGCCTGCA AAGTAAAGAT
 361 GATTTCAAAA GTGTGGTCAG CGAACAGTGC AATCATTTGC AAGCTGTCTT TGCTTCACGT
 421 TACAAGAAGT TTGATGAATT CTTCAAAGAA CTACTTGAAA ATGCAGAGAA ATCCCTGAAT
 481 GATATGTTTG TGAAGACATA TGGCCATTTA TACATGCAAA ATTCTGAGCT ATTTAAAGAT
 541 CTCTTCGTAG AGTTGAAACG TTACTACGTG GTGGGAAATG TGAACCTGGA AGAAATGCTA
 601 AATGACTTCT GGGCTCGCCT CCTGGAGCGG ATGTTCCGCC TGGTGAACTC CCAGTACCAC
 661 TTTACAGATG AGTATCTGGA ATGTGTGAGC AAGTATACGG AGCAGCTGAA GCCCTTCGGA
 721 GATGTCCCTC GCAAATTGAA GCTCCAGGTT ACTCGTGCTT TTGTAGCAGC CCGTACTTTC
 781 GCTCAAGGCT TAGCGGTTGC GGGAGATGTC GTGAGCAAGG TCTCCGTGGT AAACCCCACA
 841 GCCCAGTGTA CCCATGCCCT GTTGAAGATG ATCTACTGCT CCCACTGCCG GGGTCTCGTG
 901 ACTGTGAAGC CATGTTACAA CTACTGCTCA AACATCATGA GAGGCTGTTT GGCCAACCAA
 961 GGGGATCTCG ATTTTGAATG GAACAATTTC ATAGATGCTA TGCTGATGGT GGCAGAGAGG
1021 CTAGAGGGTC CTTTCAACAT TGAATCGGTC ATGGATCCCA TCGATGTGAA GATTTCTGAT
1081 GCTATTATGA ACATGCAGGA TAATAGTGTT CAAGTGTCTC AGAAGGTTTT CCAGGGATGT
1141 GGACCCCCCA AGCCCCTCCC AGCTGGACGA ATTTCTCGTT CCATCTCTGA AAGTGCCTTC
1201 AGTGCTCGCT TCAGACCACA TCACCCCGAG GAACGCCCAA CCACAGCAGC TGGCACTAGT
1261 TTGGACCGAC TGGTTACTGA TGTCAAGGAT AAACTGAAAC AGGCCAAGAA ATTCTGGTCC
1321 TCCCTTCCGA GCAACGTTTG CAACGATGAG AGGATGGCTG CAGGAAACGG CAATGAGGAT
1381 GACTGTTGGA ATGGGAAAGG CAAAAGCAGG TACCTGTTTG CAGTGACAGG AAATGGATTA
1441 GCCAACCAGG GCAACAACCC AGAGGTCCAG GTTGACACCA GCAAACCAGA CATACTGATC
1501 CTTCGTCAAA TCATGGCTCT TCGAGTGATG ACCAGCAAGA TGAAGAATGC ATACAATGGG
1561 AACGACGTGG ACTTCTTTGA TATCAGTGAT GAAAGTAGTG GAGAAGGAAG TGGAAGTGGC
1621 TGTGAGTATC AGCAGTGCCC TTCAGAGTTT GACTACAATG CCACTGACCA TGCTGGGAAG
1681 AGTGCCAATG AGAAAGCCGA CAGTGCTGGT GTCCGTCCTG GGGCACAGGC CTACCTCCTC
1741 ACTGTCTTCT GCATCTTGTT CCTGGTTATG CAGAGAGAGT GGAGATAAtt ctcaaactct
1801 gagaaaaagt gttcatcaaa aagttaaaag gcaccagtta tcacttttct accatcctag
1861 tgactttgct ttttaaatga atggacaacc atgtacagtt tttactatgt ggccactggg
1921 tttaagaa
```

Fig. 2. Partial exon structure and sequence of human *GPC4*. (**A**) The predicted exon 1 and exon 2 of *GPC4* are illustrated with their putative protein translation products. Nucleotide numbers correspond to the location of the sequences in bWXD9 (GenBank Accession No. AC002420). (**B**) The *GPC4* cDNA sequence obtained from an RT-PCR product from MCF-7 RNA is illustrated. Untranslated regions are in lower-case letters, and the putative open reading frame is in upper-case letters.

not shown); only the position and number of the 14 signature cysteines and the rough position of heparan sulfate glycosylation sites are conserved between the two proteins. A weak similarity like that between GPC3 and GPC4 has been observed between other glypicans [i.e. GPC1 (David et al., 1990) and GPC5 (Veugelers et al., 1997)] and provides a clue as to the molecular evolution of the family (see Section 4).

3.3. Expressed transcripts of GPC4

Northern analyses were carried out to determine the tissue distribution of *GPC4* expression. Using an 800-bp fragment of EST AA046130 from the 3′ end of *GPC4* as a probe, two transcripts of 3.4 and 4.6 kb were identified (Fig. 4). Expression appears to be widespread, and is particularly high in fetal kidney and fetal lung.

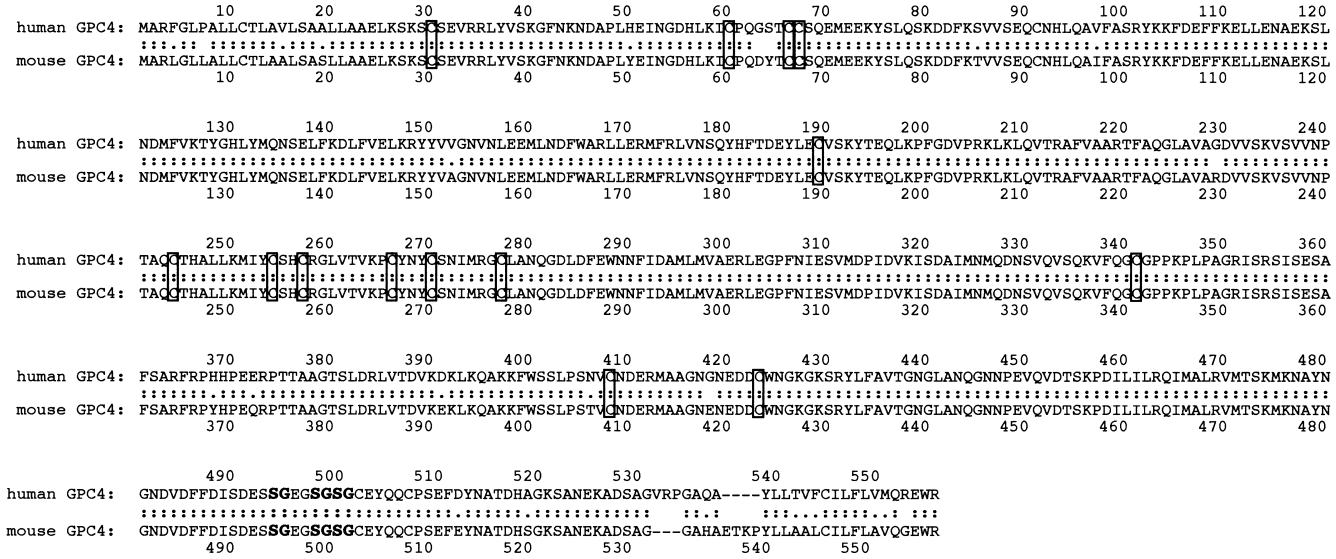

Fig. 3. Amino acid alignment of mouse and human GPC4. The predicted protein sequence encoded by human GPC4 is shown aligned with that of mouse GPC4 (Watanabe et al., 1995). Gapped regions are indicated by dashes. The 14 cysteines that are strictly conserved throughout the glypican family are boxed, and the three putative heparan sulfate glycosylation signals (Ser–Gly) are shown in bold.

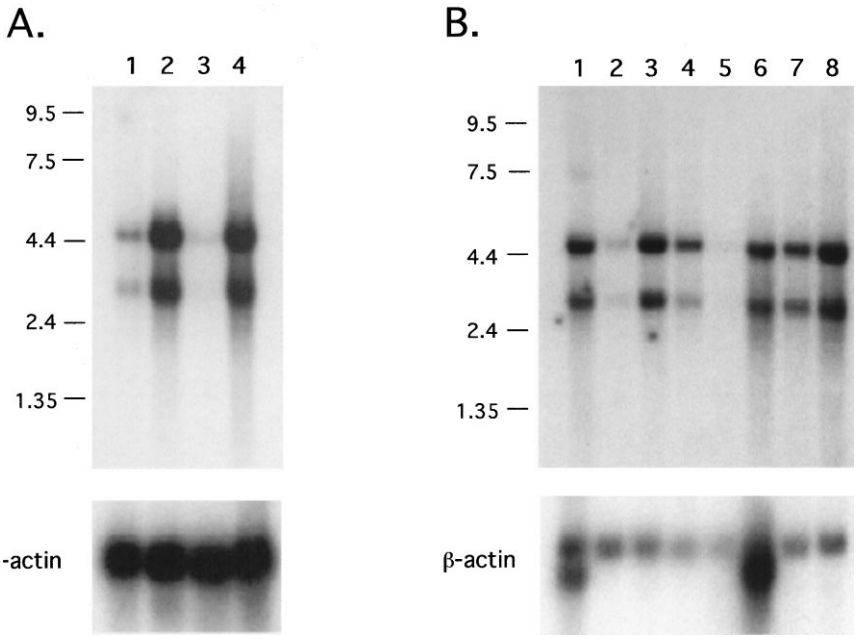

Fig. 4. Expression analysis of human *GPC4*. The 800-bp EST AA046130 corresponding to the 3′ end of the *GPC4* cDNA was hybridized against multiple tissue Northern blots (Clontech) prepared from fetal (**A**) or adult (**B**) poly-A RNA. (**A**) Lanes: 1, brain; 2, lung; 3, liver; 4, kidney. (**B**) Lanes: 1, heart; 2, brain; 3, placenta; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; 8, pancreas. To control for the integrity and equal loading of RNA, *β-Actin* cDNA was hybridized to the same blots (bottom panel).

Expression in liver is very low throughout development and can only be visualized upon long exposures (data not shown). Additionally, there appears to be no tissue-specific increase in either transcript.

The discrepancy in size between the 1.9 kb sequenced cDNA for human *GPC4* and the two mRNA species visualized in Northern blots can be explained by the presence of long untranslated sequences. Since the putative upstream region of *GPC4*, extending 1 kb 5′ of the translation initiation site, lies in a CpG island containing 74% G+C (12% CpG; data not shown), it appears unlikely that human *GPC4* 5′ untranslated sequences contribute significantly to the overall size of the transcripts; rather, it is more likely that long 3′ untranslated sequences are present. This has previously been reported for the corresponding 3.4-kb transcript in mouse, which

encodes a GPC4 protein of a similar size (see Fig. 3; Watanabe et al., 1995). Thus, based on homologies to mouse GPC4 and to all other known members of the glypican family, it appears that the complete open reading frame of *GPC4* is illustrated in Fig. 2B, and the larger species detected in Northern blots result from alternative splicing of additional 3′ untranslated exonic sequences. Sequencing of the *GPC4* genomic region (currently underway at the Sanger Centre) should provide the necessary sequences to further characterize such exons.

3.4. Deletions in patient DNAs

In a screening of X-linked Simpson–Golabi–Behmel patients, deletions were detected in *GPC3* in 20 families (Lindsay et al., 1997). In one instance, a proband and an affected uncle both showed a deletion that extended from exons 7 and 8 of *GPC3* through the 3′ exonic sequence of *GPC4* [based on the physical map of the region (Nagaraja et al., 1998), this deletion is estimated to span approximately 350 kb]. Fig. 5 shows diagnostic PCR tests for exons 6 and 8 of *GPC3*, as well as 5′ and 3′ primer pairs of *GPC4*. The deletions in *GPC3* are evident in exon 8 of patient 27 (lane 2 of Fig. 5) and in the two members of family SGB 7; patient SGB 12 is not deleted for these regions. The *GPC4* deletions in the two SGB 7 patients (the proband III-1 and his uncle, II-2) are also shown.

Clinical data were then compared for patients with these lesions. In general, deletions and other lesions in *GPC3* are loss-of-function mutations (Pilia et al., 1996; Lindsay et al., 1997), but SGBS patients show a range of severity and features. The phenotype of the propositus (III-1; SGB7) is particularly severe, with heart defects as well as central nervous system abnormalities such as hydrocephalus. To investigate the possibility that the lack of *GPC4* could be related to the phenotype, we compared three additional patients (see Table 1 for a summary of findings).

Although the paternal uncle (II-2) has the same deletion, he does not show the same severe CNS malformations. This would argue against the notion of critical *GPC4* involvement in the severe phenotype. The only other SGBS patient with a terminal deletion in *GPC3* (SGB 27) had no deletion of *GPC4*, but showed other defects (kidney abnormalities) in keeping with the range of features associated with cases of SGBS. Finally, we investigated the only other one of the 50 available males with SGBS, SGB 12, who also had hydrocephalus. This patient is particularly interesting, since no point mutation or deletion in *GPC3* has been found in his DNA (Chiappe et al., unpublished results). However, as indicated in Fig. 5 and Table 1, as of yet, no gross deletion in *GPC4* has been found (see Section 4).

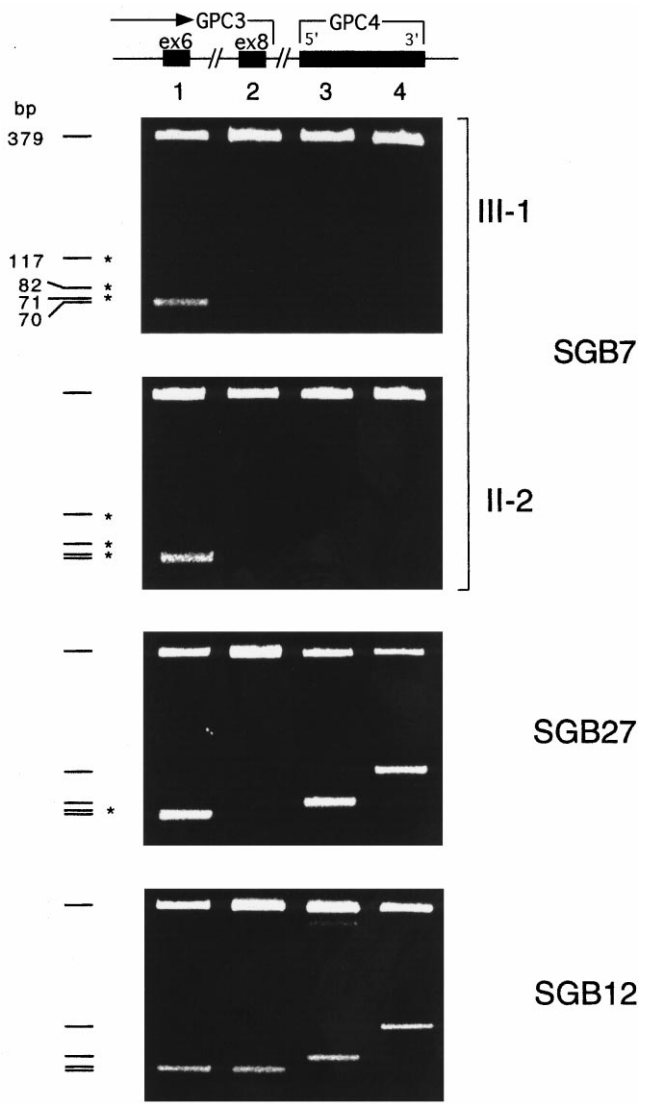

Fig. 5. Deletion analysis of *GPC3* and *GPC4* in SGBS patients. Mutation analysis was carried out on the indicated patient DNA using primer pairs from *GPC3* exon 6 (lane 1; 70 bp), *GPC3* exon 8 (lane 2; 71 bp), *GPC4* 5′ exon 2 (lane 3; 82 bp), and *GPC4* 3′ end (lane 4, 117 bp). Control reactions were performed in multiplex with a primer pair from exon 2 of the somatostatin gene (379 bp; the additional 350-bp band in lane 3 is an irreproducible artifact). Products that are absent from the patient DNA are indicated with an asterisk.

4. Discussion

GenBank entries of sequences covering most of the DNA surrounding *GPC3* [a series of bacterial artificial chromosomes including bWXD8 at the 5′ (telomeric) end of *GPC3* and bWXD9 at the 3′ (centromeric) end of *GPC3*; see Section 2 and http://www.ibc.wustl.edu/cgm/cgm.html] contain largely a melange of repetitive sequences of a number of types. Multiple copies of some of the repetitive elements, including four large and almost identical L1 segments in a region of 87 kb (Huber et al., 1997; Mazzarella and Schlessinger, 1998), suggest

Table 1
Summary of SGBS patient phenotypes

	SGB7 III-1	SGB7 II-2	SGB27	SGB12
GPC3 status	Deletion exon 7+8	Deletion exon 7+8	Deletion exon 8	Nil found
GPC4 status	Deletion 5′ and 3′	Deletion 5′ and 3′	Nil found	Nil found
Facial features	✓	✓	✓	✓
Cleft palate±lip	✓	✓	✓	Not present
Hydrocephalus	✓	Not present	Not present	✓
Congenital heart abnormalities	VSD, ASD	VSD[a]	Not present	VSD, PS, PDA
Renal abnormalities	Not present	Not present	Large cystic kidneys	Not present

VSD, ventricular septal defect; ASD, atrial septal defect; PS, pulmonary stenosis; PDA, patient ductus arteriosus.
[a] Closed spontaneously.

that duplications of nongenic sequence tracts have occurred in this region, and the DNA surrounding *GPC3* may be quite dynamic.

Such a low local content of genes is compatible with the recovery of large local deletions that have no effect on the survival of either cultured cells (Fuscoe et al., 1994) or human peripheral blood cells (Nelson et al., 1995). It is striking, therefore, that two glypicans are juxtaposed in a region containing very few genes.

It is notable that the glypicans described here are of different classes. Based on similarities of protein sequences inferred from cDNAs (Veugelers et al., 1997), the human *GPC4* reported here fits into a homology group with GPC1 (David et al., 1990), whereas the neighboring GPC3 is in a second group with GPC5 (Veugelers et al., 1997). One speculative explanation regarding the molecular evolution of the neighboring glypican genes is that *GPC3* and *GPC4* may have come to be adjacent in the genome because a primordial glypican was duplicated, the copies diverged, and further copies of each were later transferred to one or more other sites in the genome. The only inference that can be made with certainty, however, is that the divergence in the glypican family is ancient. This and other notions may be distinguishable by studies of the numbers and relative locations of glypican family members sampled from various phylogenetic levels.

The development of glypican family members has been accompanied by the development of differential expression patterns. Northern analyses show that, in particular, *GPC4* is widely expressed, whereas *GPC3* is predominantly observed in mesodermal-derived tissues (Pilia et al., 1996).

The relatively restricted tissue distribution of *GPC3* is consistent with its physiological effects. It is the only one of the glypican family that has thus far been associated with defined phenotypes in human disease. In its absence, the overgrowth/gigantism of SGBS is observed (Pilia et al., 1996). GPC3 is thus implicated in a complex regulatory pathway determining overall size and growth rate. Other instances of comparable gigantism and overgrowth are correlated with translocations and mutations over several megabases in 11p15.5 (Beckwith–Wiedemann syndrome), and are also frequently associated with increases in IGF2 expression (Weksberg et al., 1993). GPC3 may itself act on the IGF2 pathway, because (1) there is evidence that it binds directly to IGF2 (Pilia et al., 1996; Xu et al., 1998); (2) IGF2 overexpression and IGF2-receptor knockout lead to equivalent overgrowth phenotypes in mice (Sun et al., 1997; Ludwig et al., 1996); and (3) the embryonic patterns of expression of IGF2 and GPC3 are superimposable (Pelligrini et al., 1999).

Particularly in light of the tendency of BWS 'genes' to cluster in 11p15.5, one can ask whether genes neighboring *GPC3* may also be related to overgrowth phenotypes in a comparable way. The finding of another glypican juxtaposed with *GPC3* is suggestive, and the further finding of a deletion in an SGBS family that extends into *GPC4* as well as *GPC3* sharpens the question of whether a lesion in *GPC4* contributes to phenotype in patients.

The current data on SGBS patients do not support such a possibility strongly, but neither do they exclude it. *GPC3* and *GPC4* are rather divergent, both in their sequence similarity and in their expression patterns; also, the phenotypic features of the two affected family members deficient in both genes (see Table 1 and text) are sufficiently different so that no 'contiguous deletion syndrome' more extensive than SGBS can be inferred, and as yet, there are no individuals with lesions only in *GPC4* to provide the necessary independent evidence for a contribution of that gene to the hydrocephalic phenotype. Nevertheless, it remains intriguing that the other 'SGBS' patient with hydrocephalus, SGB 12, has no lesion in any *GPC3* exon. We plan to continue our analysis of *GPC4* in SGB 12 and other males where we have been unable to find any mutations in *GPC3*.

Acknowledgement

We wish to thank D. Schlessinger for many helpful discussions and critical review of the manuscript. This

work was supported by Telethon grant n.E357 and Assessorato Igiene e Sanitá, Regione Sardegna, L.R. No. 11, 30.04.90.

References

Altschul, S.F., Gish, W., Miller, W., Meyers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

David, G., Lories, V., Decock, B., Marynen, P., Cassiman, J.J., van den Berghe, H., 1990. Molecular cloning of a phosphatidylinositol-anchored membrane heparan sulfate proteoglycan from human lung fibroblasts. J. Cell Biol. 111, 3165–3176.

David, G., 1993. Integral membrane heparan sulfate proteoglycans. FASEB J. 7, 1023–1030.

Everett, L.A., Glaser, B., Beck, J.C., Idol, J.R., Buchs, A., Heyman, M., Adawi, F., Hazani, E., Nassir, E., Baxevanis, A.D., Sheffield, V.C., Green, E.D., 1997. Pendred syndrome is caused by mutations in a putative sulphate transporter gene (PDS). Nature Genet. 17, 411–422.

Fuscoe, J.C., Nelsen, A.J., Pilia, G., 1994. Detection of deletion mutations extending beyond the HPRT gene by multiplex PCR analysis. Somat. Cell Mol. Genet. 20, 39–46.

Gonzalez, A.D., Kaya, M., Shi, W., Song, H., Testa, J.R., Penn, L.Z., Filmus, J., 1998. OCI-5/GPC3, a glypican encoded by a gene that is mutated in the Simpson–Golabi–Behmel overgrowth syndrome induces apoptosis in a cell line-specific manner. J. Cell Biol. 141, 1407–1414.

Hoglund, P., Haila, S., Socha, J., Tomaszewski, L., Saarialho-Kere, U., Karjalainen-Lindsberg, M.L., Airola, K., Holmberg, C., de la Chapelle, A., Kere, J., 1996. Mutations of the Down-regulated in adenoma (DRA) gene cause congenital chloride diarrhoea. Nature Genet. 14, 316–319.

Huber, R., Crisponi, L., Mazzarella, R., Chen, C.-N., Su, Y., Shizuya, H., Chen, E.Y., Cao, A., Pilia, G., 1997. Analysis of exon/intron structure and 400 kb of genomic sequence surrounding the 5′-promoter and 3′-terminal ends of the human glypican 3 (GPC3) gene. Genomics 45, 48–58.

Jurka, J., Klonowski, P., Dagman, V., Pelton, P., 1996. CENSOR-A program for the identification and elimination of repetitive sequences in the human genome. Nucleic Acids Res. 17, 4731–4738.

Lindsay, S., Ireland, M., O'Brien, O., Clayton-Smith, J., Hurst, J.A., Mann, J., Cole, T., Sampson, J., Slaney, S., Schlessinger, D., Burn, J., Pilia, G., 1997. Large scale deletions in the GPC3 gene may account for a minority of cases of Simpson–Golabi–Behmel syndrome. J. Med. Genet. 34, 480–483.

Ludwig, T., Eggenschwiler, J., Fisher, P., D'Ercole, A.J., Davenport, M.L., Efstratiadis, A., 1996. Mouse mutants lacking the type 2 IGF receptor (IGF2R) are rescued from perinatal lethality in Igf2 and Igf1r null backgrounds. Dev. Biol. 177, 517–535.

Mazzarella, R., Schlessinger, D., 1998. Duplication and distribution of repetitive elements and non-unique regions in the human genome. Gene 205, 29–38.

Nagaraja, R., MacMillan, S., Jones, C., Masisi, M., Pengue, G., Porta, G., miao, S., Casamassimi, A., D'Urso, M., Brownstein, B., Schlessinger, D., 1998. Integrated YAC/STS physical and genetic map of 22.5 Mb of human Xq46–q26 at 56 kb inter-STS resolution. Genomics 52, 247–266.

Nelson, S.L., Jones, I.M., Fuscoe, J.C., Burkhart-Schultz, K., Grosovsky, A.J., 1995. Mapping the end points of large deletions affecting the hprt locus in human peripheral blood cells and cell lines. Radiat. Res. 141, 2–10.

Ohno, S., 1970. Evolution by Gene Duplication. Springer, New York, 150 pp.

Pelligrini, M., Pilia, G., Pantano, S., Lucchini, F., Uda, M., Fumi, M., Cao, A., Schlessinger, D., Forabosco, A., 1999. Gpc3 expression correlates with the phenotype of the Simpson–Golabi–Behmel syndrome. Dev. Dyn., in press.

Pilia, G., Hughes-Benzie, R.M., MacKenzie, A., Baybayan, P., Chen, E.Y., Huber, R., Neri, G., Cao, A., Forabosco, A., Schlessinger, D., 1996. Mutations in GPC3, a glypican gene, cause the Simpson–Golabi–Behmel overgrowth syndrome. Nature Genet. 12, 241–247.

Song, H.H., Shi, W., Filmus, J., 1997. OCI-5/rat glypican-3 binds to fibroblast growth factor-2 but not to insulin-like growth factor-2. J. Biol. Chem. 272, 7574–7577.

Sun, F.L., Dean, W.L., Kelsey, G., Allen, N.D., Reik, W., 1997. Transactivation of Igf2 in a mouse model of Beckwith–Wiedemann syndrome. Nature 389, 809–815.

Uberbacher, E.C., Mural, R.J., 1991. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. Proc. Natl. Acad. Sci. USA 88, 11261–11265.

Veugelers, M., Vermeesch, J., Reekmans, G., Steinfeld, R., Marynen, P., David, G., 1997. Characterization of Glypican-5 and chromosomal localization of human GPC5, a new member of the glypican gene family. Genomics 40, 24–30.

Watanabe, K., Yamada, H., Yamaguchi, Y., 1995. K-Glypican: A novel GPI anchored proteoglycan that is highly expressed in developing brain and kidney. J. Cell Biol. 130, 1207–1218.

Weksberg, R., Shen, D.R., Fei, Y.-L., Song, Q.L., Squire, J., 1993. Disruption of insulin-like growth factor 2 imprinting in Beckwith–-Wiedemann syndrome. Nature Genet. 5, 143–150.

Xu, Y., Papageorgiou, A., Polychronakos, C., 1998. Developmental regulation of the soluble form of insulin-like growth factor-II/mannose 6-phosphate receptor in human serum and amniotic fluid. J. Clin. Endocrinol. Metab. 83, 437–442.