LETTER TO JMG

Heterozygous truncating mutation in the human homeobox gene *GSH2* has no discernable phenotypic effect

J G Dauwerse, C E M de Die-Smulders, E Bakker, M H Breuning, D J M Peters

J Med Genet 2002;**39**:686–688

Mutations in transcription factors with homeobox domains have been identified in a number of developmental disorders, for instance, mutations in *PAX6* have been identified in patients with aniridia, mutations in *HOX13* in patients with synpolydactyly, and mutations in *MSX2* in patients with Boston-type craniosynostosis.[1]

The mouse *GSH2* gene, like the related *GSH1* gene, encodes a homeodomain containing gene that is homologous to the *Drosophila* intermediate neuroblasts defective *(ind)* gene.[2] In situ hybridisation of *GSH2* showed a dynamic, developmentally regulated, spatial and temporal expression pattern.[3] Transcripts are particularly abundant in the hindbrain and in the ventral domain of the forebrain.[3] *GSH2* is a downstream target of sonic hedgehog (SHH) and is probably a key regulator in downstream SHH patterning in the ventral forebrain. Mice lacking *GSH2* show profound defects in telencephalon development.[4] Comparing mice lacking functional alleles of either *GSH2* or *PAX6* indicated complementary roles for these two genes in dorsoventral patterning of the telencephalon.[5]

METHODS

In a search for a developmental gene on chromosome 4q12, we identified the human homologue of the *GSH2* gene. To do so, we performed inter-Alu PCR on yeast artificial chromosome clone 303b3 (CEPH megaYAC[6]) that by FISH was mapped to the chromosome 4q12 region. With these PCR products, we isolated cosmid 232G12 from a chromosome 4 cosmid library[7] and PAC clones pDJ194i7 and pDJ200G9 from the RPC1 PAC library.[8] The sequences initially generated on fragments of the cosmid showed homology with BAC clone RP11-56d20, a clone partially sequenced by the Whitehead Institute/MITcenter. The NIX program at www.hgmp.mrc.ac.uk, which is a tool to identify unknown nucleic acid sequences, was used to analyse the partial sequences of the BAC clone, suggesting the presence of (part of) a gene homologous to mouse GSH2 (swissprot P31316) and GSH1 proteins, and identified the human platelet derived growth factor receptor α (*PDGFRA*) gene (GenBank XM011186). The mouse *GSH2* gene (GenBank S79041) is located in the syntenic region on mouse chromosome 5. Blast searches at the NCBI (http://www.ncbi.nlm.nih.gov) showed homology with a large number of small ESTs (for example, A1005406, A1360912, A1797226, AA485189) and, indeed, these clones hybridised to cosmid 232G12. Subclones of this cosmid were used to obtain the complete genomic sequence, which was submitted to GenBank (AF439445). We sequenced the gene four times, including sequences obtained on genomic DNA. The *GSH2* gene contains a predicted ORF of 915 bp encoding a protein of 304 amino acids. The gene consists of two coding exons with a 651 bp intron in between. Exon 1 consists of 575 bp coding sequence and the size of the 5'UTR is not known, exon 2 consists of 340 bp coding sequence and 239 bp 3'UTR. Recently, a coding sequence has also been published in GenBank (AH010253). The putative translation start site, 5'-CUCGAC*AUG*U-3', at position 310 in our sequence is not accompanied by a typical Kozak sequence. A consensus polyadenylation signal (5'-AATAAA-3') was observed 553 bp after the stop codon. In contrast to mouse *GSH2*, which contains two polyadenylation signals close to each other, the human gene has only one.

RESULTS

The deduced amino acid sequence of human GSH2 and the sequences of mouse GSH2 and GSH1 are presented in fig 1. Sequence comparison with mouse GSH2 showed that the genes are highly conserved between mouse and human, with 89% identity and 91% similarity over the entire ORF. Most of the non-identical amino acids are located between residues 75 and 100, a region that is absent in mouse GSH1. In addition, the protein is homologous to mouse GSH1 (GenBank NM_008178). The 57 amino acid sized homeodomain, which is involved in DNA binding, is entirely conserved.

DISCUSSION

In an effort to identify a gene involved in acrodysostosis and short stature, we performed single strand conformation polymorphism analysis (SSCA) in a set of short stature patients and controls. Primers were selected to amplify the exons with adjacent splice sites and branch sites in six overlapping fragments (table 1). The products were separated on 10% non-denaturing polyacrylamide gels (49:1, acryamide:bis acrylamide) with and without glycerol and on 0.5 * sequagel MD gels (7-15 W for ±16 hours at room temperature)

Key points

- The mouse *GSH2* (genomic screened homeobox-2) gene is a homeobox domain containing transcription factor with a developmentally regulated expression pattern. Mice lacking *GSH2* show profound defects in telencephalon development.
- We describe the cloning, identification, and sequencing of the human homologue of the *GSH2* gene. The gene consists of two exons and the 915 bp open reading frame (ORF) encodes a protein of 304 amino acids. The human GSH2 protein is highly conserved with 89% identity and 91% similarity with mouse GSH2.
- In a father and daughter without a discernable developmentally aberrant phenotype, we identified a truncating mutation in the homeobox domain. These data indicate that, unlike most disease associated transcription factors, loss of one functional copy of *GSH2* does not result in a disease phenotype.

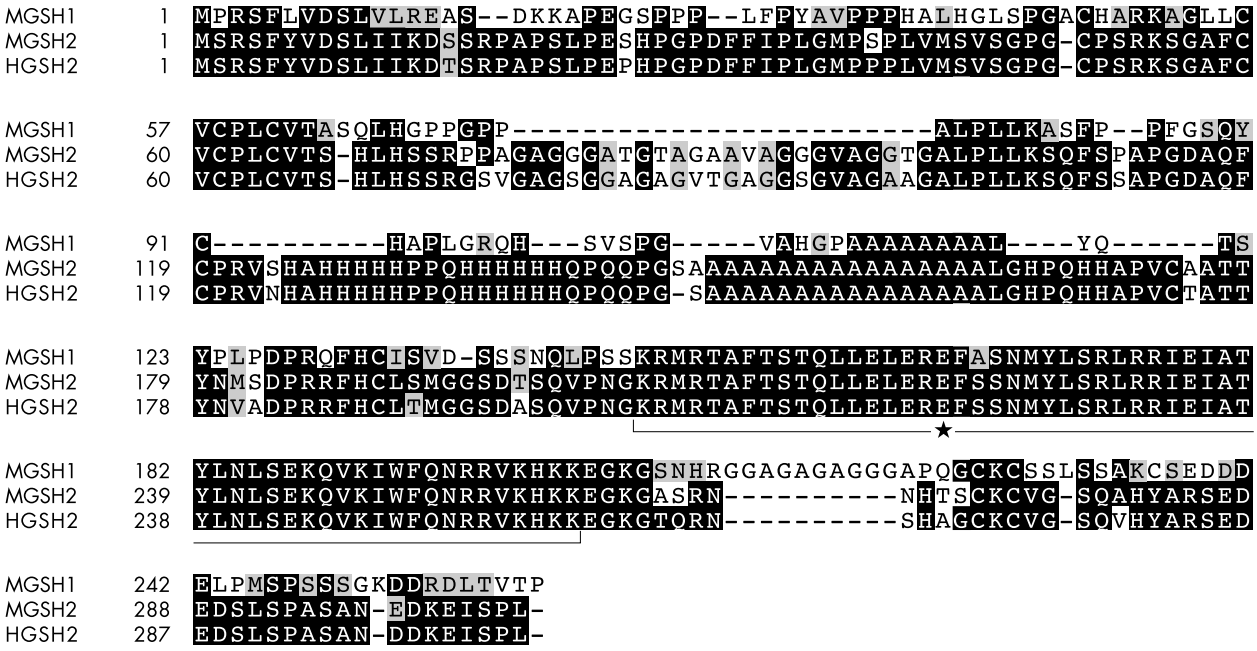

Figure 1 Alignment of the amino acid sequences of mouse GSH1, GSH2, and human GSH2. Human GSH2 (GenBank AF439445) is aligned to mouse GSH2 (SwissProt P31316) and mouse GSH1 (GenBank NM_008178). Most differing amino acids between human and mouse GSH2 are between AA 75 and 100. The homeobox domain is indicated by an underline. The nonsense mutation, Glu220X, is indicated by an asterisk. The mutation is designated according to the recommendations of Antonarakis *et al*.[10]

Table 1 Primer pairs for SSCP analyses

Primer name		Sequence (5′-3′)	Position	Fragment size	Annealing temp	PCR buffer
Gsh12	F	GAGCACCTTGCCCGAGCCTTACC	1496–1518	364 bp	56	A
	R	CTCCTTGTCATCGTTGGCTGAGG	1838–1860			
Gsh13	F	TCCGAGGATGAGGACTCCCTGTC	1810–1832	370 bp	56	A
	R	CCCCATACAGTGTTAAACTTAA	2159–2180			
Gsh14	F	CAACTCATTCTTGTCGTATAAC	2117–2138	374 bp	55	B
	R	GCAGGAGAAGATAGGGAGAGCAA	2159–2180			
Gsh15	F	AGGGCAGAGCTTAGAACACTAGA	234–256	286 bp	56	A*
	R	AGTGCAGGTGCGAAGTGAC	502–520			
Gsh16	F	CCATTGGTGATGTCCGTGT	427–445	302 bp	ND	ND
	R	CTGATGATGGTGATGGTGGT	710–729			
Gsh17	F	GTGAACCATGCGCATCATC	673–691	249 bp	56	A*
	R	AAAGGCGCGGAGCGCAGGTG	903–922			

Buffer A: Red *Taq* buffer (Sigma, Zwijndrecht, The Netherlands); buffer A* is Red *Taq* buffer with final concentration of 1.2 mol/l Betain; buffer B: super*taq* buffer (Sphaero Q, Leiden, The Netherlands).

essentially according to the protocols described by Orita *et al*.[9] In a family where the index case had short stature and mental retardation, we identified in the father and sister with "normal" phenotypes, a G to T transition at position 658, changing a glutamine into a stop. This mutation, which was not present in the index patient, is predicted to truncate the protein at amino acid 220, within the homeobox domain of the gene (fig 1). We do not know whether this mutation generates a stable truncated protein. If it does, the truncated protein does not have a dominant negative affect on the function of the full length protein, probably because the mutation is in the beginning of the homeodomain. Apparently, loss of one functional copy of *GSH2* has no discernable phenotypic effect. For most disease associated transcription factors, haploinsufficiency results in the disease phenotype, indicating that gene dosage is critical at specific time points in development. This high degree of dosage sensitivity often appears to affect only a subset of tissues that express the gene.[1] Our data are in agreement with the observation that heterozygous *GSH2*$^{+/-}$ mice seem to be indistinguishable from wild type mice and have a normal life span.[3]

GSH2 contains a polyalanine stretch of 16 residues. Polyalanine tracts are common in homeobox genes and in other transcription factors. For several of these genes, inactivation by polyalanine tract expansion has been shown to cause a disease phenotype as, for example, in HOXA13 involved in the hand-foot-genital syndrome[11] or HOXD13 involved in synpolydactyly.[12] In these genes the expansions are short, are meiotically stable, and are probably caused by unequal crossing over during replication, thus being different from other unstable pathological trinucleotide expansions.[11] In GSH2, the polyalanine tract is not encoded by a perfect triplet repeat and, indeed, in 200 unrelated chromosomes we did not observe a repeat expansion.

It can be concluded that *GSH2* is a highly conserved gene and that a heterozygous truncating mutation does not cause a discernable phenotype.

ACKNOWLEDGEMENTS

We thank P Taschner (LUMC) for assistance in NIX analyses, M Losekoot for DNA samples, and M van de Berg (Leiden Genome Technology Centre) for screening the PAC library.

Authors' affiliations

J G Dauwerse, E Bakker, M H Breuning, D J M Peters, Center of Human and Clinical Genetics, Leiden University Medical Centre, The Netherlands
C E M de Die-Smulders, Department of Clinical Genetics, Academic Hospital Maastricht, The Netherlands

Correspondence to: Dr D J M Peters, Center of Human and Clinical Genetics, Leiden University Medical Centre, Sylvius Laboratory, Wassenaarseweg 72, 2333AL Leiden, The Netherlands; d.j.m.peters@lumc.nl

REFERENCES

1 **Engelkamp D**, van Heyningen V. Transcription factors in disease. *Curr Opin Genet Dev* 1996;**6**:334-42.
2 **Weiss JB**, Von Ohlen T, Mellerick DM, Dressler G, Doe CQ, Scott MP. Dorsoventral patterning in the Drosophila central nervous system: the intermediate neuroblasts defective homeobox gene specifies intermediate column identity. *Genes Dev* 1998;1**2**:3591-602.
3 **Szucsik JC**, Witte DP, Li H, Pixley SK, Small KM, Potter SS. Altered forebrain and hindbrain development in mice mutant for the Gsh-2 homeobox gene. *Dev Biol* 1997;**191**:230-42.
4 **Corbin JG**, Gaiano N, Machold RP, Langston A, Fishell G. The Gsh2 homeodomain gene controls multiple aspects of telencephalic development. *Development* 2000;**127**:5007-20.
5 **Yun K**, Potter S, Rubenstein JL. Gsh2 and Pax6 play complementary roles in dorsoventral patterning of the mammalian telencephalon. *Development* 2001;**28**:193-205.
6 **Cohen D**, Chumakov I, Weissenbach J. A first-generation physical map of the human genome. *Nature* 1993;**366**:698-701.
7 **Riess O**, Siedlaczck I, Kredtke S, Melmer G, Epplen JT, Deaven LL. Characterization of a human chromosome 4 flow-sorted cosmid library. *Cytogenet Cell Genet* 1994;**65**:238-42.
8 **Ioannou PA**, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet* 1994;**6**:84-9.
9 **Orita M**, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci USA* 1989;**86**:2766-70.
10 **Antonarakis SE**. Recommendations for a nomenclature system for human gene mutations. Nomenclature Working Group. *Hum Mutat* 1998;**11**:1-3.
11 **Goodman FR**, Bacchelli C, Brady AF, Brueton LA, Fryns JP, Mortlock DP, Innis JW, Holmes LB, Donnenfeld AE, Feingold M, Beemer FA, Hennekam RC, Scambler PJ. Novel HOXA13 mutations and the phenotypic spectrum of hand-foot-genital syndrome. *Am J Hum Genet* 2000;**67**:197-202.
12 **Muragaki Y**, Mundlos S, Upton J, Olsen BR. Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* 1996;**272**:548-51.