Am. J. Hum. Genet. 60:1194–1201, 1997

The DiGeorge Syndrome Minimal Critical Region Contains a *Goosecoid*-like (*GSCL*) Homeobox Gene That Is Expressed Early in Human Development

Shoshanna Gottlieb,[1] Beverly S. Emanuel,[1,2] Deborah A. Driscoll,[1,3] Beatrice Sellinger,[1] Zhili Wang,[4,*] Bruce Roe,[4] and Marcia L. Budarf[1,2]

[1]Division of Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia, and Departments of [2]Pediatrics and [3]Obstetrics and Gynecology, University of Pennsylvania School of Medicine, Philadelphia; and [4]Departments of Chemistry and Biochemistry, University of Oklahoma, Norman

Summary

The majority of patients with DiGeorge syndrome (DGS) and velocardiofacial syndrome (VCFS) have deletions of chromosomal region 22q11.2. The abnormalities observed in these patients include conotruncal cardiac defects, thymic hypoplasia or aplasia, hypocalcemia, and characteristic facial features. To understand the genetic basis of these disorders, we have characterized genes within the region that is most consistently deleted in patients with DGS/VCFS, the minimal DiGeorge critical region (MDGCR). In this report, we present the identification and characterization of a novel gene, *GSCL*, in the MDGCR, with homology to the homeodomain family of transcription factors. Further, we provide evidence that this gene is expressed in a limited number of adult tissues as well as in early human development. The identification of *GSCL* required a genomic sequence–based approach because of its restricted expression and high GC content. The early expression, together with the known role of homeobox-containing proteins in development, make *GSCL* an outstanding candidate for some of the abnormalities seen in DGS/VCFS.

Introduction

Deletion of chromosomal region 22q11.2 is seen in association with a variety of human congenital abnormalities, including DiGeorge syndrome (DGS), velocardiofacial syndrome (VCFS), and isolated and familial congenital conotruncal heart defects (reviewed by Driscoll 1994). The structures affected in these genetic disorders—the skeletal and connective tissue of the head and neck, the thymus, the parathyroid glands, and the conotruncal region of the heart—are derived from the pharyngeal arches and pouches and thus have a common origin in development. The cranial neural crest cells that migrate from the neural fold and populate the pharyngeal arches are major contributors to the formation of these pharyngeal arch–derived structures. In addition, it has been shown that ablation of cranial neural crest causes defects in the conotruncal region of the avian heart that are similar to those seen in DiGeorge syndrome (Kirby et al. 1983). Taken together, these observations have led to the suggestion that deletion or disruption of a gene(s) located in 22q11 may perturb the movement or differentiation of the cephalic neural crest cells that give rise to or influence the development of the structures defective in these syndromes (Kirby and Bockman 1984; Lammer and Opitz 1986). A number of genes in mouse are known to play a role in the development of neural crest–derived tissues. For example, a mouse knockout of the transcription factor *AP-2* leads to defects in craniofacial development (Schorle et al. 1996; Zhang et al. 1996), while a knockout of the gene *endothelin-1* leads to defects in craniofacial development as well as cardiac defects similar to those observed in DGS/VCFS (Kurihara et al. 1994, 1995). Inactivation of the mouse *hoxa-3* gene also causes several of the features observed in DGS, such as absent thymus and parathyroid glands (Chisaka and Capecchi 1991). However, none of these mouse genes that influence neural crest cell–derived structures has a human homologue that maps to 22q11.

Our efforts to identify a candidate gene(s) for the 22q11 deletion syndrome have focused on a 250-kb area, the minimal DiGeorge critical region (MDGCR). The MDGCR maps to the proximal part of an ~2-Mb region that is deleted in the majority of DGS/VCFS patients (Gong et al. 1996). We have used multiple approaches to identify and characterize genes in this region, including the analysis of the complete genomic sequence of the region. While most of the genes pre-

Received December 2, 1996; accepted for publication February 21, 1997.

Address for correspondence and reprints: Dr. Marcia Budarf, Children's Hospital of Philadelphia, 34th and Civic Center Boulevard, 1004 Abramson Research Building, Philadelphia, PA 19104. E-mail: budarf@cbil.humgen.upenn.edu

*Present affiliation: Children's Hospital of Philadelphia, Philadelphia.

0002-9297/97/6005-0021$02.00

dicted by computational methods also were identified by experimental approaches, there was a notable exception. Database searches using the genomic sequence of cosmid 79h12 identified what appeared to be a functional gene with similarity to the homeodomain family of transcription factors. This gene has highest identity to *goosecoid* (*gsc*), a gene that is expressed in neural crest–derived tissues (Gaunt et al. 1993) and required for proper craniofacial development in mice (Rivera-Perez et al. 1995; Yamada et al. 1995). Here we present the genomic structure of this gene, *goosecoid-like* (*GSCL*), and show by reverse transcriptase (RT)–PCR experiments that it is expressed during early fetal development. This early expression, together with the known role of homeobox-containing proteins in embryogenesis, make *GSCL* an outstanding candidate for some of the developmental abnormalities seen in DGS/VCFS.

Material and Methods

Cosmid Isolation and Sequencing

A gridded chromosome 22–only cosmid library (LL22NC03) was screened by colony hybridization using a unique, genomic 2.5-kb *Hind*III fragment that had been subcloned from a *Not*I linking clone, N25. Cosmid 79h12 was selected for sequence analysis. The cosmid was sequenced by a double-stranded random shotgun approach (Bodenteich et al. 1993). The non-*Alu* sequences were searched against the databases (Brookhaven Protein Data Bank, GenBank, EMBL) with the BLASTN program (Altschul et al. 1990).

cDNA Synthesis

cDNA was synthesized in a 50-μl reaction with 100 ng of polyA$^+$ mRNA from various tissues. The RNA was heated with oligo(dT) primers at 65°C for 5 min and allowed to cool at room temperature for 10 min. RT was performed at 37°C for 1 h after adding 5 μl 10× first strand buffer (Stratagene), 40 U RNase inhibitor (Stratagene), 2 μl 0.1 M dNTPs, and 50 U Maloney murine leukemia virus RT. The cDNA mixture was then heated at 90°C for 5 min. For PCR amplification, 1 μl of cDNA was used per 20 μl reaction as described below.

PCR

Because of the GC-rich nature of the gene, it was necessary to modify standard PCR reaction protocols by the addition of 15% glycerol or by use of a high GC protocol (Dutton et al. 1993) with a reaction buffer containing 15% glycerol and no potassium (Woodford et al. 1995). In almost all cases, it was necessary to perform a nested PCR reaction or reamplification with the same primers to observe a product. This is probably due to a combination of two factors: the low abundance of the transcript and the reduced efficiency of the PCR reaction because of the high GC content of the region.

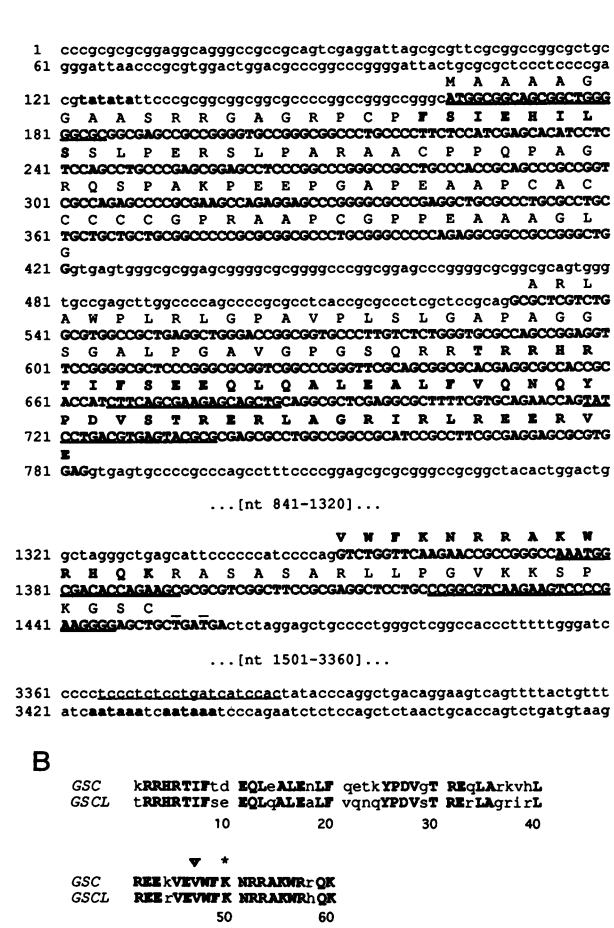

Figure 1 Genomic structure and amino acid sequence of *GSCL*. *A*, Genomic sequence of *GSCL*, showing the amino acid translation of the predicted protein product. The amino acids corresponding to the octapeptide and to the homeodomain are shown in bold. A TATA-like element and two potential polyadenylation signals are shown in bold. The locations of primers used for PCR are underlined. The direction of transcription is from telomere to centromere. *B*, Comparison of human *GSC* and *GSCL* in the homeodomain. The numbers refer to the homeodomain and correspond to amino acids 125–185 of the entire GSCL protein. The arrowhead denotes the conserved splice site found in all members of the *goosecoid* family as well as other members of the prd-like class. The lysine at position 50 is marked by an asterisk (K*). Identities are shown as bold, uppercase letters.

The locations and sequences of the primers used are shown in figure 1*A*. Amplification of the entire coding region was carried out in a 20-μl reaction volume using Vent polymerase (New England Biolabs) and a two-step PCR protocol (98°C 1 min, 70°C 5 min) in 15% glycerol and 75 mM Tris-Cl (pH 8.8), 20 mM $(NH_4)_2SO_4$, 0.1% Tween20, and 1.5 mM $MgSO_4$ (DNAmp). The entire reaction was analyzed on a 1% low-melt agarose gel, the resulting product was excised, and 1 μl of the gel slice was reamplified using the same set of primers. Amplification between exons 2 and 3 was carried out using *Taq* polymerase in *Taq* buffer (Boehringer-Mannheim) in 15% glycerol at 94°C 30 s, 60°C 30 s, and 72°C 1

Table 1

PCR Primers Used to Amplify Genomic Sequence from Exons 2 and 3 of *GSCL*

Exon	Sequence (5′–3′)
5′ end of exon 2	F AGT GGG TGC CGA GCT TGG CCC CA
	R GCT CGC GCG TAC TCA CGT CAG GA
3′ end of exon 2	F AGG CGC CAC CGC ACC ATC TTC AG
	R AGG CCG TCT TTG CAA AGG GCG C
Exon3:[a]	
1	F CCA CAT CCC TGT TGC GAA
	R GGG TAG ACC TGT CTT CTC
2	F GAA GCT CCC CTC TCG GTC
	R CCA ACT CCA AAG ATC CCA AA

[a] Exon 3 was amplified using a nested set of primers.

min. The reaction products were then diluted 1/50, and 1 μl was reamplified using a nested set of primers. To generate the product extending from the second exon to close to the polyadenylation site, amplification was carried out using the Expand™ High Fidelity PCR System protocol (Boehringer-Mannheim) in 15% glycerol. The reaction products were diluted 1/50, and 1 μl was reamplified using a nested primer at the 5′ end and the original primer at the 3′ end.

Northern Blot Analysis

Multiple tissue northern blots (Clontech) were hybridized to a radiolabeled purified insert from a cloned PCR product (see Results) at 65°C in hybridization buffer as described by Church and Gilbert (1984), for 16–24 h. Filters were washed twice in 2× SSC, 0.1% SDS at room temperature for 5 min and then twice in 0.1× SSC, 0.1% SDS at 65°C for 15 min.

Cloning and Sequencing of PCR Products

PCR products were cloned using the Original TA Cloning Kit from Invitrogen. PCR products were isolated from 1% low-melting-temperature agarose and purified using the QIAquick gel purification system. Sequencing of plasmids and direct sequencing of PCR products was done on an ABI373A Sequencer.

Mutational Analysis

Genomic DNA from nondeleted DGS/VCFS patients was screened for point mutations and small insertions and deletions by direct sequencing of PCR products (for patient descriptions, see Results). Exons 2 and 3 were amplified using the primers listed in table 1 and sequenced as described above. The PCR conditions to amplify exon 2 used the high-GC protocol also described above. Exon 3 was amplified using standard reaction conditions with the addition of glycerol.

To assay for deletions, insertions, and rearrangements within the *GSCL* gene, we analyzed genomic DNA from the nondeleted patients by Southern blotting. Genomic DNA was digested with the restriction endonuclease *Pst*I, separated on a 0.8% agarose gel, and transferred to Hybond N^+ (Amersham). In all cases, the genomic fragments used as probes were purified from a plasmid and radiolabeled and hybridized at 65°C in hybridization buffer as described by Church and Gilbert (1984), for 16–24 h. Filters were washed twice in 2× SSC, 0.1% SDS at room temperature for 5 min and then twice in 0.5× SSC, 0.1% SDS at 65°C or, in the case of the probe corresponding to the 5′ end of the gene, in 0.2× SSC, 0.1% SDS at 65°C, for 15 min.

Results

The cosmid 79h12 was chosen for sequencing as part of the minimal overlap of cosmids that represent the MDGCR. When the sequence of 79h12 was subjected to database searches (Altschul et al. 1990), matches to homeodomain-containing proteins were noted with the greatest similarity to *gsc* (Blumberg et al. 1991) (see below). On the basis of this result, further characterization of this *GSCL* gene was undertaken. By combining the results from database homology searches with GRAIL analysis (Uberbacher and Mural 1991), the predicted *GSCL* gene structure shown in figure 1*A* was assembled. The coding region is contained within three exons and encodes a predicted protein of 205 amino acids. *GSCL* is a member of the paired-like class (prd-like) of homeobox genes. Similar to *gsc* and other members of the prd-like class of genes, the homeodomain is split by an intron between amino acids 46 and 47 (arrowhead in fig. 1*B*). Within the 60–amino acid homeodomain, there is 72% identity to human *GSC* (Blum et al. 1994) (fig. 1*B*) and 71% identity to *Drosophila gsc* (*D-gsc*) (Goriely et al. 1996). By contrast, the vertebrate *gsc* genes have 98%–100% identity to each other in this region and 76% identity to *D-gsc*. On the basis of these comparisons alone, it is not possible to predict whether *D-gsc* represents a common ancestor to both *GSC* and *GSCL* or whether there is a second gene in *Drosophila*. The next-most-related member of the prd-like class is *otx1*, which has 55% identity to *GSCL*. The seven most-conserved amino acids in homeodomains are present in the MDGCR gene, as are six additional highly conserved amino acids, indicating that *GSCL* very likely encodes a functional DNA-binding protein (Bürglin 1994). Like *gsc*, the amino acid at position 50 of the homeodomain, a residue that has been shown to be important in determining the specificity of DNA binding (Hanes and Brent 1989; Treisman et al. 1989) is a lysine (K* in fig. 1*A*). Although the vast majority of homeodomains have a glutamine at this position, several homeobox genes, including *gsc*, *bicoid*, and *orthodenticle/otx*, have a lysine at position 50. In general, the genes with a lysine at this position are expressed in the anterior region of the embryo.

By use of an enhanced version of the National Center for Biotechnology Information's BLAST search tool, BEAUTY (Worley et al. 1995), a second region of homology in the first exon was found between *GSCL* and *gsc* (boldface in fig. 1*A*). This region also is conserved in *D-gsc* (Goriely et al. 1996), as well as other classes of homeoproteins, and is thought to have a role in transcriptional repression (Smith and Jaynes 1996). The extreme N-terminus of *GSCL* is rich in alanine, a feature that has been reported for other homeobox genes. Near the end of the first exon is a stretch of five cysteines. Although homopolymeric amino acid stretches are common in homeobox proteins, a stretch of Cys is novel. Further, the homopolymeric amino acid stretches are not well conserved, and their function is not clear (Bürglin 1994). The predicted start codon is contained within a good consensus for translation initiation (Kozak 1986), and a TATA-like sequence is found 40 nt upstream of the predicted start site of translation. In addition, there are polyadenylation signals 1,967 and 1,975 nt beyond the predicted stop codon. Therefore, the size of the inferred transcript, including 5′ and 3′ UTRs, is 2.6 kb.

Initial attempts to examine the expression of *GSCL* with a small probe derived from the genomic sequence failed to detect a signal on commercial northern blots (data not shown). Therefore, a RT-PCR–based assay, which provides greater sensitivity, was performed. With a forward primer beginning at the predicted start codon and a reverse primer beginning three amino acids before the stop codon (fig. 1*A*), we were able to amplify a product of the expected size, using polyA$^+$ mRNA from adult testis as starting material. This tissue was chosen because it has been noted to express many genes that would not necesssarily be predicted to be expressed during spermatogenesis (Hecht 1995). Direct sequencing of this PCR product verified the predicted gene structure. Because of the GC-rich nature of the gene (76% across the entire coding region and 82% across the first exon), it was difficult to amplify consistently the entire coding region. Therefore, we chose a different set of primers to assay expression of *GSCL* in other tissues. Using primers spanning the second and third exons (fig. 1*A*), we were able to detect expression in a number of tissues, including adult testis and pituitary, and 9–10-wk fetal tissue (thorax) (fig. 2*A*). The PCR products obtained from testis and 9–10-wk fetal tissue were sequenced directly, which confirms that the RT-PCR products represented properly spliced *GSCL* transcripts. Further, to verify the use of the predicted stop codon, we amplified and sequenced a 2,050-bp product from testis polyA$^+$ mRNA that extended from the second exon to 40 bp 5′ of the first predicted polyadenylation signal. The sequence of this correctly spliced product indicates that the predicted stop codon (fig. 1*A*) is used.

This 2,050-bp PCR product was cloned and used to reprobe commercial northern blots of fetal and adult tissues. It detected an ~2.6-kb signal in adult testis only (fig. 2*B*). Expression in testis is consistent with the RT-PCR results, since the amplified product was most abundant in testis. As might be predicted on the basis of the RT-PCR results, no signal was observed in any other tissue, even after a long exposure, suggesting that if *GSCL* is expressed in the other tissues represented on these blots it must be of low abundance. The absence of a signal in 20–26-wk fetal heart (the developmental stage of mRNA used for the commercial fetal northern blot) is not inconsistent with *GSCL* playing a role in the heart defects associated with DGS/VCFS, because neural crest cell migration and aorticopulmonary septation would be completed by 7 wk. The apparent low abundance of the message in all tissues tested, including those tested by RT-PCR, may be due to the fact that we have not yet studied those tissues in which *GSCL* is normally highly expressed. In addition, the expression in testis does not necessarily exclude a role for *GSCL* in other tissues. There are several examples of genes that are expressed in testis but are known to be expressed in and have a function in other tissues and at other stages of development (e.g., Rubin et al. 1986; Shackleford and Varmus 1987). Taken together, these results demonstrate expression of *GSCL* in multiple tissues, albeit at a low level, and, notably, that *GSCL* is expressed during early fetal development. Further, preliminary results from RT-PCR of RNA from 10.5-d mouse embryos indicate that the mouse homologue of *GSCL* is expressed in the anterior portion of the embryo (S. Gottlieb and N. Galili, unpublished data).

Although the vast majority of patients diagnosed with DGS/VCFS have deletions on chromosome 22, a small number of patients have the phenotypic features but do not have a deletion, as determined by FISH analysis using the probe N25 (for description of this probe, see Driscoll et al. 1993; Budarf et al. 1995). Genomic DNA was available from 14 patients referred to our laboratory with a diagnosis of DGS/VCFS and no detectable deletions in 22q11. Clinical summaries were provided by either a referring geneticist and/or immunologist. All 14 had two or more of the major clinical features seen in these disorders, including congenital heart defect, absent thymus or a history of frequent infections, palatal abnormalities and/or speech difficulties, facial dysmorphia, and a history of learning disabilities or developmental delay. Since 12 of the 14 cases are sporadic and the remaining 2 are from small families, it is not possible to determine by linkage whether the phenotype maps to 22q11 or a different chromosomal locus.

To detect point mutations and small insertions and deletions, genomic DNA from these patients was screened by direct sequencing of PCR products. Exons 2 and 3 were amplified separately using the primers listed in table 1. The first exon and promotor region,

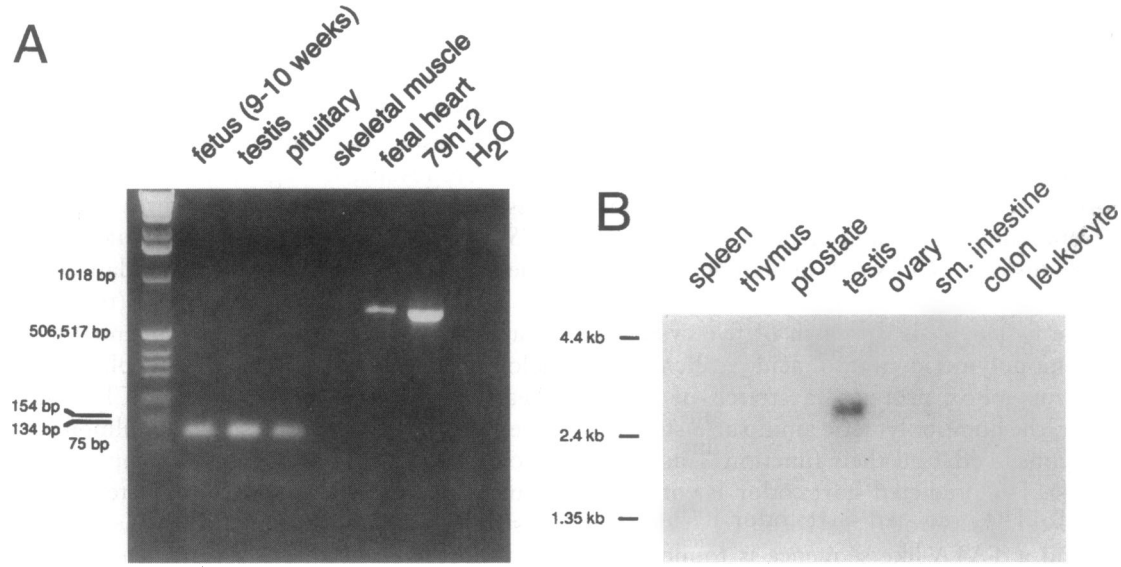

Figure 2 Expression of *GSCL*. *A*, Agarose gel electrophoresis of the products of PCR amplification of cDNA from various human tissue sources and the cosmid 79h12, using nested sets of primers between the second and third exons (see fig. 1*A*). The figure shows that amplification of cDNA and genomic DNA yields products of 109 and 676 bp, respectively. The band seen in fetal heart is due to genomic DNA contamination. PolyA$^+$ mRNA from adult testis, pituitary and skeletal muscle, and cDNA from fetal heart (20–26 wk) were purchased from Clontech; polyA$^+$ mRNA from 9–10-wk fetal tissue (thorax) was purified using the FastTrack 2.0 mRNA isolation kit from Invitrogen. *B*, Multiple adult tissue northern blot analysis (Clontech), using a cloned, purified 2,050-bp cDNA probe extending from within the second exon to close to the predicted polyadenylation site. An ~2.6-kb message is detected only in testis even after a 2-wk exposure. In addition to the blot shown here, another adult tissue northern blot (heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas), as well as a fetal tissue northern blot (brain, lung, liver, and kidney), were hybridized, and no signal was observed.

which are >82% GC, are recalcitrant to PCR and still require analysis in all patients. While no changes were detected in the coding sequence, a polymorphism in intron 2 of a single patient that was also present in the patient's mother, as well as in 2/48 control individuals, was found.

To assay for larger deletions, insertions, and rearrangements within the *GSCL* gene, we analyzed genomic DNA from the nondeleted patients by Southern blotting (see Material and Methods). This analysis would detect some types of rearrangements or deletions within the 5′ end of the gene, the region not amenable to PCR and sequence analysis. Blots were hybridized serially with three different cloned probes covering the entire *GSCL* gene including the predicted 3′ UTR. In addition, a second probe, pH20 (D22S41; Budarf et al. 1996), that maps to a more telomeric region on chromosome 22 was added to each hybridization as an internal control. To determine whether there was a deletion of the locus, the relative intensities of the bands from the two loci were compared.

Although no rearrangments or deletions of the locus have been detected using this approach, a *Pst*I polymorphism in the 3′ UTR was identified. This polymorphism was present in 3/14 patients analyzed and 1/14 control individuals. In addition, this polymorphism was not present in the affected parent of one of the patients. Sequence analysis indicated that the polymorphism is due to a base change in the *Pst*I recognition site. Further analysis of these patients is required to determine whether there are mutations present that have not been detected by the methods described above.

Discussion

The experiments presented here clearly demonstrate that an expressed homeobox gene termed *GSCL* is encoded within the MDGCR. Identification of this gene required a sequence-driven approach, since it was not detected by any of the previous experimental methods we used, including cDNA selection, exon amplification, and direct screening of cDNA libraries. Including this gene, we now have characterized the complete open reading frames of six genes in the minimal critical region (Goldmuntz et al. 1996; Gong et al. 1996, 1997; Holmes et al. 1997; present report). On the basis of analysis of the complete genomic sequence of the MDGCR, it is unlikely that additional genes will be found. These genes and their directions of transcription are shown on the bottom lines in figure 3. Approximately 90% of all patients diagnosed with DGS/VCFS have large deletions of chromosomal region 22q11.2 that include all of these genes (Carey et al. 1992; Driscoll et al. 1993). Although the size of the deletion is generally large, 1–2 Mb, it has been possible to define a smallest region of deletion overlap by mapping the deletion endpoints and positions

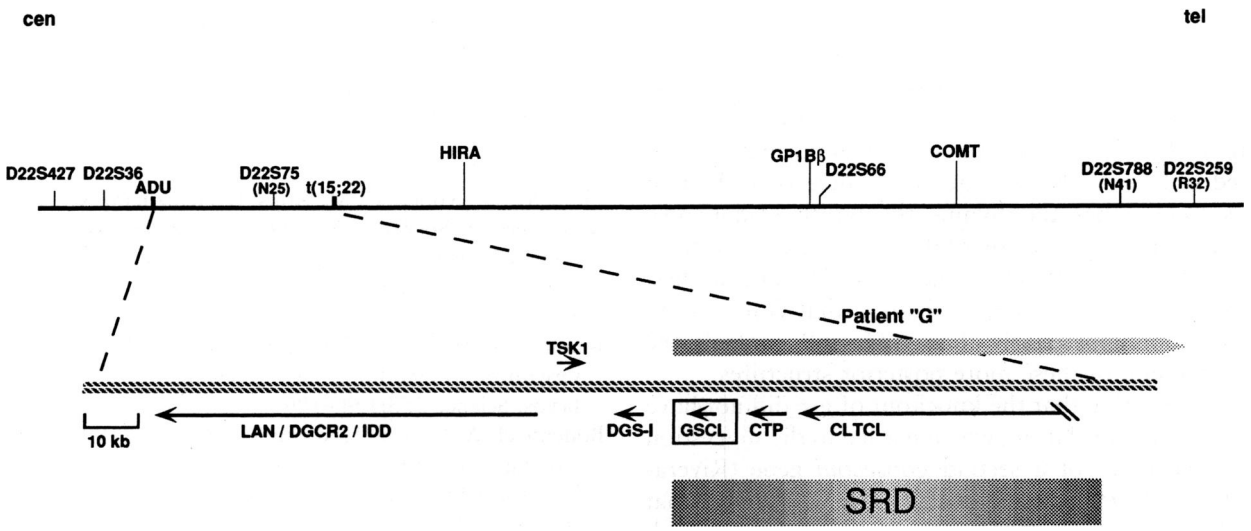

Figure 3 DiGeorge chromosomal region. The top part of the figure corresponds to the region on chromosome 22 that is commonly deleted in DGS/VCFS patients. The expanded region corresponds to the MDGCR. The positions and directions of transcription of genes in the MDGCR are indicated by arrows. For a description of patient G and the t(15;22) translocation, see Discussion.

of translocation breakpoints in a series of DGS and VCFS patients (fig. 3). In all but one case (Kurahashi et al. 1996), the deletions analyzed result in haploinsufficiency of *GSCL,* including a DGS patient who has an atypical proximal deletion boundary (patient G in the study by Levy et al. [1995]).

The centromeric deletion boundary of the interstitial deletion in patient G recently has been positioned immediately distal to *DGSI* (fig. 3) (Rizzu et al. 1996). Therefore, patient G narrows the centromeric boundary of the MDGCR and appears to exclude the breakpoint region of the only known balanced translocation in a patient with DGS (ADU; Augusseau et al. 1986). Perhaps the ADU translocation effects expression of downstream genes through a position effect. Our distal deletion boundary shown in figure 3 is defined by an unbalanced 15;22 translocation in a patient with classic features of VCFS (Jaquez et al., in press; D. Driscoll, M. Li, and M. Budarf, unpublished data). If the breakpoints of these two patients (G and t(15;22)) can be combined to define the "smallest region of deletion" (SRD) for DGS/VCFS, then only *CLTCL* (clathrin heavy chain–like), *GSCL,* and *CTP* (citrate transport protein) reside within the new SRD. The genomic location, together with the potential role of *GSCL* as a developmental control gene, make *GSCL* a strong candidate for playing a causative role in this disorder. In addition, it is known that several diseases in humans are caused by mutations in one allele of a homeobox-containing gene (for example, see Noll 1993; Brunelli et al. 1996). However, it is also clear from recent studies that disease-related genes are not necessarily located in the "minimal critical region" and can be located at a considerable distance from the translocation breakpoint(s) of patients with the disease phenotype (see, for example, Foster et al. 1994; Wagner et al. 1994; Belloni et al. 1996). Therefore, although the chromosomal location of *GSCL* suggests that this gene may be important, it is possible that a gene outside the SRD shown in figure 3 is involved in the disease phenotype.

Among the small group of nondeleted patients we have analyzed, we have not yet detected mutations in *GSCL.* However, the failure to find mutations in these patients does not necessarily preclude *GSCL* from having an important role in the disorder. As stated previously, our analysis did not include the promotor region or first exon of the gene. In addition, the etiology of DGS/VCFS is heterogeneous, including exposure to teratogens, maternal diabetes, and other chromosomal loci (Lammer and Opitz 1986). Since it is not possible to perform linkage analysis on the nondeleted patients we have studied, we do not know whether their disease phenotype is linked to chromosome 22. In fact, in one nondeleted familial case we have studied, it was possible to exclude the involvement of 22q11 by linkage (Gong et al. 1997). Finally, while *GSCL* may have an important role in DGS/VCFS, haploinsufficiency for multiple genes may be required to observe the disease phenotype. In this latter case, one would not expect to find point mutations in *GSCL* in patients with DGS/VCFS. Further analysis is required to distinguish between these possibilities.

Similarity comparisons indicate that *GSCL* is the closest known homologue of *gsc.* Expression studies of *gsc* in mouse embryos demonstrate an early, transient phase of expression in the primitive streak (6.4–6.8 d) (Blum et al. 1992) and a later phase, beginning at 10.5 d in the craniofacial region, in derivatives of cephalic neural crest cells (Gaunt et al. 1993). The *gsc*-null knockout

mouse has numerous craniofacial defects (Rivera-Perez et al. 1995; Yamada et al. 1995) primarily in structures derived from the first and second pharyngeal arches, the regions corresponding to highest *gsc* expression (Gaunt et al. 1993). While the craniofacial abnormalities observed in DGS/VCFS may be due to defects in the first and second arches, the thymus, the parathyroid glands, and the conotruncal region of the heart are derived from the third and fourth pharyngeal pouches and arches. Therefore, an attractive hypothesis is that *GSCL* is expressed more posteriorly than *GSC* and controls the differentiation of these more posterior structures.

It is interesting that the knockout of *gsc* did not have an effect on gastrulation, which has led to the suggestion of the existence of a second *goosecoid* gene (Rivera-Perez et al. 1995; Yamada et al. 1995). We propose that *GSCL* could represent this second locus, even though the sequence homology between *GSCL* and *GSC* is restricted primarily to the homeodomain. In support of this idea, recent experiments have shown that *D-gsc* is able to rescue UV-irradiated Xenopus embryos in a manner similar to *X-gsc* (Goriely et al. 1996), and, as is the case for *GSCL*, the sequence homology of *D-gsc* to vertebrate *gsc* is primarily confined to the homeodomain.

In summary, *GSCL* is a homeobox gene that maps within the smallest region of deletion of the MDGCR. Its expression in early human development and predicted function as a DNA-binding protein make it an outstanding candidate for some of the developmental defects associated with the 22q11.2 deletion syndrome. Studies of the expression of the murine or avian homologue of *GSCL* should help to illuminate the function of this gene.

Acknowledgments

The authors wish to thank S. Holmes and R. Oakey for critical reading of the manuscript, C. Montalvo and C. Wong for technical assistance, and W. Gong, B. Goldmuntz, N. Galili, and T. Bürglin for helpful discussions. These studies were supported in part by grants CA39926 (to B.S.E.), HG00425 (to B.S.E. and M.L.B.), HL51533 (to M.L.B., B.R., and B.S.E.), DC02027 (to M.L.B., B.S.E., and B.R.), and HG00313 (to B.R.) from the National Institutes of Health. The accession numbers for the genomic and cDNA entries are U30597 and GSDB:S:127445, respectively.

References

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

Augusseau S, Jouk S, Jalbert P, Prieur M (1986) DiGeorge syndrome and 22q11 rearrangements. Hum Genet 74:206

Belloni E, Muenke M, Roessler E, Traverso G, Siegel-Bartelt J, Frumkin A, Mitchell HF, et al (1996) Identification of *sonic hedgehog* as a candidate gene responsible for holoprosencephaly. Nat Genet 14:353–356

Blum M, Gaunt SJ, Cho KWY, Steinbeisser H, Blumberg B, Bittner D, De Robertis EM (1992) Gastrulation in the mouse: the role of the homeobox gene *goosecoid*. Cell 69: 1097–1106

Blum M, De Robertis EM, Kojis T, Heinzmann C, Klisak I, Geissert D, Sparkes RS (1994) Molecular cloning of the human homeobox gene *goosecoid* (*GSC*) and mapping of the gene to human chromosome 14q32.1. Genomics 21: 388–393

Blumberg B, Wright CVE, De Robertis EM, Cho KWY (1991) Organizer-specific homeobox genes in *Xenopus laevis* embryos. Science 253:194–196

Bodenteich A, Chissoe S, Wang YF, Roe BA (1993) Shotgun cloning as the strategy of choice to generate templates for high-throughput dideoxynucleotide sequencing. In: Venter JC (ed) Automated DNA sequencing and analysis techniques, Academic Press, London, pp 42–50

Brunelli S, Faiella A, Capra V, Nigro V, Simeone A, Cama A, Boncinelli E (1996) Germline mutations in the homeobox gene *EMX2* in patients with severe schizencephaly. Nat Genet 12:94–96

Budarf ML, Collins J, Gong W, Roe B, Wang Z, Bailey LC, Sellinger B, et al (1995) Cloning a balanced translocation associated with DiGeorge syndrome and identification of a disrupted candidate gene. Nat Genet 10:269–288

Budarf ML, Eckman B, Michaud D, McDonald T, Gavigan S, Buetow KH, Tatsumara Y, et al (1996) Regional localization of over 300 loci on human chromosome 22 using a somatic cell hybrid mapping panel. Genomics 35:275–288

Bürglin TR (1994) A comprehensive classification of homeobox genes. In: Duboule D (ed) Guidebook to the homeobox genes: a Sambrook and Tooze publication. Oxford University Press, New York, pp 27–71

Carey AH, Kelly D, Halford S, Wadey R, Wilson D, Goodship J, Burn J, et al (1992) Molecular genetic study of the frequency of monosomy 22q11 in DiGeorge syndrome. Am J Hum Genet 51:964–970

Chisaka O, Capecchi MR (1991) Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *hox-1.5*. Nature 350:473–479

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991–1995

Driscoll DA (1994) Genetic basis of DiGeorge and velocardiofacial syndromes. Curr Opin Pediatr 6:702–706

Driscoll DA, Salvin J, Sellinger B, Budarf ML, McDonald-McGinn DM, Zackai EH, Emanuel BS (1993) Prevalence of 22q11 microdeletions in DiGeorge and velocardiofacial syndromes: implications for genetic counselling and prenatal diagnosis. J Med Genet 30:813–817

Dutton CM, Paynton C, Sommer SS (1993) General method for amplifying regions of very high G + C content. Nucleic Acids Res 21:2953–2954

Foster JW, Dominguez-Steglich MA, Guioli S, Kwok C, Weller PA, Stevanovic M, Weissenbach J, et al (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. Nature 372:525–530

Gaunt SJ, Blum M, De Robertis EM (1993) Expression of the mouse *goosecoid* gene during mid-embryogenesis may mark

mesenchymal cell lineages in the developing head, limbs and body wall. Development 117:769–778

Goldmuntz E, Wang Z, Roe BA, Budarf ML (1996) Cloning, genomic organization and chromosomal localization of human citrate transport protein to the DiGeorge/Velocardial-facial syndrome minimal critical region. Genomics 33:271–276

Gong W, Emanuel BS, Collins J, Kim DH, Wang Z, Chen F, Zhang G, et al (1996) A transcription map of the DiGeorge and velo-cardio-facial syndrome minimal critical region on 22q11. Hum Mol Genet 5:789–800

Gong W, Emanuel BS, Galili N, Kim DH, Roe B, Driscoll DA, Budarf ML (1997) Structural and mutational analysis of a conserved gene (*DGSI*) from the minimal DiGeorge syndrome critical region. Hum Mol Genet 6:267–276

Goriely A, Stella M, Coffinier C, Kessler D, Mailhos C, Dessain S, Desplan C (1996) A functional homologue of *goosecoid* in *Drosophila*. Development 122:1641–1660

Hanes S, Brent R (1989) DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. Cell 57:1275–1283

Hecht NB (1995) The making of a spermatozoon: a molecular perspective. Dev Genet 16:95–103

Holmes SE, Riazi MA, Gong W, McDermid HE, Sellinger BT, Hua A, Chen F, et al (1997) Disruption of the clathrin heavy chain-like gene (CLTCL) associated with features of DGS/VCFS: a balanced (21;22)(p12;q11) translocation. Hum Mol Genet 6:357–367

Jaquez M, Driscoll DA, Li M, Emanuel BS, Hernandez I, Jaquez F, Lembert N, et al. Unbalanced 15;22 translocation in a patient with features of both DiGeorge and velocardiofacial syndrome. Am J Med Genet (in press)

Kirby ML, Bockman DE (1984) Neural crest and normal development: a new perspective. Anat Rec 209:1–6

Kirby ML, Gale TF, Stewart DE (1983) Neural crest cells contribute to normal aorticopulmonary septation. Science 220:1059–1061

Kozak M (1986) Point mutations define a sequence flanking the AUG initiatior codon that modulates translation by eukaryotic ribosomes. Cell 44:283–292

Kurihara Y, Kurihara H, Oda H, Maemura K, Nagai R, Ishikawa T, Yazaki Y (1995) Aortic arch malformations and ventricular septal defect in mice deficient in endothelin-1. J Clin Invest 96:293–300

Kurihara Y, Kurihara H, Suzuki H, Kodama T, Maemura K, Nagai R, Oda H, et al (1994) Elevated blood pressure and craniofacial abnormalities in mice deficient in endothelin-1. Nature 368:703–710

Kurahashi H, Nakayama T, Osugi Y, Tsuda E, Masuno M, Imaizumi K, Kamiya T, et al (1996) Deletion mapping of 22q11 in CATCH22 syndrome: identification of a second critical region. Am J Hum Genet 58:1377–1381

Lammer EJ, Opitz JM (1986) The DiGeorge anomaly as a developmental field defect. Am J Med Genet 29:113–127

Levy A, Demczuk S, Aurias A, Depetris D, Mattei M, Philip N (1995) Interstitial 22q11 microdeletion excluding the ADU breakpoint in a patient with DiGeorge syndrome. Hum Mol Gen 4:2417–2419

Noll M (1993) Evolution and role of *Pax* genes. Curr Opin Genet Dev 3:595–605

Rivera-Perez JA, Mallo M, Gendron-Maguire M, Gridley T, Behringer RR (1995) *Goosecoid* is not an essential component of the mouse gastrula organizer but is required for craniofacial and rib development. Development 121:3005–3012

Rizzu P, Lindsay EA, Taylor C, O'Donnell H, Levy A, Scambler P, Baldini A (1996) Cloning and comparative mapping of a gene from the commonly deleted region of DiGeorge and velocardiofacial syndromes conserved in *C. elegans*. Mamm Genome 7:639–643

Rubin MR, Toth LE, Patel MD, D'Eustachio P, Nguyen-Huu MC (1986) A mouse homeo box gene is expressed in spermatocyes and embryos. Science 233:663–667

Schorle H, Meier P, Buchert M, Jaenisch R, Mitchell P (1996) Transcription factor AP-2 essential for cranial closure and craniofacial development. Nature 381:235–238

Shackleford GM, Varmus HE (1987) Expression of the proto-oncogene *int*-1 is restricted to postmeiotic male germ cells and the neural tube of mid-gestational embryos. Cell 50: 89–95

Smith ST, Jaynes JB (1996) A conserved region of engrailed, shared among all en-, gsc-, Nk1-, Nk2- and msh-class homeoproteins, mediates active transcriptional repression in vivo. Development 122:3141–3150

Treisman J, Gonczy P, Vashishtha M, Harris E, Desplan C (1989) A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell 59:553–562

Uberbacher EC, Mural RL (1991) Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. Proc Natl Acad Sci USA 88:11261–11265

Wagner T, Wirth T, Meyer J, Zabel B, Held M, Zimmer J, Pasantes J, et al (1994) Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. Cell 79:1111–1120

Woodford K, Weitzmann MN, Usdin K (1995) The use of K^+-free buffers eliminates a common cause of premature chain termination in PCR and PCR sequencing. Nucleic Acids Res 23:539

Worley KC, Wiese BA, Smith RF (1995) BEAUTY: an enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. Genome Res 5:173–187

Yamada G, Mansouri A, Torres M, Stuart ET, Blum M, Schultz M, De Robertis EM, et al (1995) Targeted mutation of the murine *goosecoid* gene results in craniofacial defects and neonatal death. Development 121:2917–2922

Zhang J, Hagopian-Donaldson S, Serbedzija G, Elsemore J, Plehn-Dujowich D, McMahon AP, Flavell RA, et al (1996) Neural tube, skeletal, and body wall defects in mice lacking transcription factor AP-2. Nature 381:238–241