



A region of mouse chromosome 16 is syntenic to the DiGeorge, velocardiofacial syndrome minimal critical region.

N Galili, H S Baldwin, J Lund, et al.

Genome Res. 1997 7: 17-26

Access the most recent version at doi:[10.1101/gr.7.1.17](https://doi.org/10.1101/gr.7.1.17)

References

This article cites 37 articles, 17 of which can be accessed free at:
<http://genome.cshlp.org/content/7/1/17.refs.html>

Article cited in:

<http://genome.cshlp.org/content/7/1/17#related-urls>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

Correction

A correction has been published for this article. The contents of the [correction](#) have been appended to the original article in this reprint. The correction is also available online at:

<http://genome.cshlp.org/content/7/4/399.full.html>

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>

RESEARCH

A Region of Mouse Chromosome 16 Is Syntenic to the DiGeorge, Velocardiofacial Syndrome Minimal Critical Region

Naomi Galili,^{1,6} H. Scott Baldwin,^{1,2} Jim Lund,⁴ Roger Reeves,⁴
Weilong Gong,³ Zhili Wang,⁵ Bruce A. Roe,⁵ Beverly S. Emanuel,³
Sudhir Nayak,¹ Craig Mickanin,² Marcia L. Budarf,³ and
Clayton A. Buck^{1,2}

¹Wistar Institute, Philadelphia, Pennsylvania; ²Divisions of Pediatric Cardiology and ³Human Genetics and Molecular Biology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; ⁴Department of Physiology, Johns Hopkins University Medical School, Baltimore, Maryland; ⁵Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma

DGS and VCFS, haploinsufficiencies characterized by multiple craniofacial and cardiac abnormalities, are associated with a microdeletion of chromosome 22q11.2. Here we document synteny between a 150-kb region on mouse chromosome 16 and the most commonly deleted portion of 22q11.2. Seven genes, all of which are transcribed in the early mouse embryo, have been identified. Of particular interest are two serine/threonine kinase genes and a novel *gooseoid*-like homeobox gene (*Gscf*). Comparative sequence analysis of a 38-kb segment reveals similarities in gene content, order, exon composition, and transcriptional direction. Therefore, if deletion of these genes results in DGS/VCFS in humans, then haploinsufficiencies involving this region of chromosome 16 should recapitulate the developmental field defects characteristic of this syndrome.

[The sequence data described in this paper have been submitted to GenBank under accession no. U70231.]

DiGeorge syndrome (DGS) is a developmental field defect of the third and fourth branchial arches (Kirby and Bockman 1984; Lammer and Opitz 1986). The resulting phenotypes in the neonate include hypoplasia of the parathyroid gland and thymus, abnormalities of the cardiac outflow tract, and mild craniofacial dysmorphism (Conley et al. 1979). Velocardiofacial syndrome (VCFS) shares phenotypic overlap with DGS. Nearly 90% of DGS patients, 85% of VCFS patients, and 10–30% of patients with isolated conotruncal defects carry a microdeletion on the long arm of one copy of chromosome 22 at band q11.2 (Driscoll et al. 1993; Goldmuntz et al. 1993), suggesting that these syndromes are related etiologically (Stevens et al. 1990). The deletion at 22q11.2 segregates as an autosomal dominant genetic disorder that results in a haploinsufficiency of products encoded by genes within the deleted region. Although the majority of deletions include a >1.4-Mb segment of 22q11.2, a number of patients carry deletions involving only some por-

tion of this domain (Driscoll et al. 1992; Budarf et al. 1995). However, regardless of the size of the individual deletion, nearly all share a 250-kb minimal deleted region designated previously as the minimal DiGeorge critical region (MDGCR) (Budarf et al. 1995; Gong et al. 1996; M. Li, M.L. Budarf, B. Selinger, M. Jaquez, R. Matalon, S. Ball, R.A. Pagon, S.S. Rosengren, B.S. Emanuel, and D.A. Driscoll, pers. comm.). Hence, it is likely that early cardiac and craniofacial development is sensitive to the dosage of a product encoded by one or a combination of genes within the MDGCR.

Despite the fact that numerous candidate genes have been identified, none has been shown to be directly involved in either DGS or VCFS (Halford et al. 1993; Demczuk et al. 1995, 1996; Heisterkamp et al. 1995; Wadey et al. 1995; Goldmuntz et al. 1996; Kedra et al. 1996; Lindsay et al. 1996; Sirotkin et al. 1996; Sutherland et al. 1996). It has been suggested that a gene disrupted by a balanced translocation, t(2;22), from a patient designated as ADU, would be a particularly good candidate gene (Augousseau et al. 1986). Unfortunately, sequence analysis across the ADU breakpoint failed to implicate unequivocally a

Corresponding author.
E-MAIL galili@wista.wistar.upenn.edu; FAX (215) 898-0663.

specific gene (Budarf et al. 1995; Sutherland et al. 1996). In addition, a DGS patient in which the ADU domain remains intact has been reported recently (Levy et al. 1995), suggesting that this region does not encode a single gene critical to the DGS/VCSF phenotype. Thus, at the present time, there is no evidence that a single gene within the MDGCR is responsible for this complex phenotype.

Here we describe an alternative approach for identifying candidate genes. Advantage is taken of two facts: (1) A human cosmid contig of the MDGCR recently has been constructed and used to assemble a transcription map including 11 different genes (Gong et al. 1996). (2) A region of mouse Chr 16, encompassing 4.2 cM on a composite map, shows conserved synteny with 22q11.1 (Cabin et al. 1996). In order to determine the extent of synteny and the degree of conservation between this portion Chr 16 and the MDGCR, a detailed analysis was undertaken. PCR-based simple sequence length polymorphism (SSLP) markers and landmark human genes were used to screen murine yeast artificial chromosome (YAC) and bacterial artificial chromosome (BAC) libraries. A region corresponding to much of the human 250-kb MDGCR domain, including the ADU translocation breakpoint, was identified and used for sequence comparisons with the contig and transcription map assembled from the human MDGCR.

The power of comparative sequencing to elucidate subtle features of genomic regions has been illustrated clearly in the analysis of human and mouse α/γ T-cell receptors (Koop and Hood 1994). Sequence comparisons provide a powerful tool for finding genes, identifying gene components that diverge from consensus signals used by sequence analysis algorithms, highlighting conserved non-coding regions that may have important functions (gene regulatory domains or sites affecting local DNA structure), and differentiation of pseudogenes

and “junk” DNA from sequences whose functions cannot be predicted based upon current knowledge of structure/function relationships at the genome level. Comparative physical mapping and sequencing reported here demonstrate a high level of conserved synteny between mouse Chr 16 and the MDGCR on human Chr 22. The comparison illuminates several features of the region that were not evident from human sequence alone and establishes the genetic basis for creating long-range deletions of this region in the mouse (Ramirez-Solis et al. 1995).

RESULTS

Genetic Mapping of Potential Markers of Synteny

The MDGCR is expected to map to mouse Chr 16 based on the known conserved synteny between Chr 22 and mouse Chr 16 (Nadeau et al. 1992; Bucan et al. 1993; Cabin et al. 1996). Mouse homologs of two genes in or near the MDGCR, *Tsk1* (Bielke et al. 1994; Gong et al. 1996) and gene fragment N41 (Emanuel et al. 1993), were mapped in the mouse to refine the comparative map of this region. Two large intersubspecific backcrosses, BXB and BMXB, were characterized previously to identify all DNAs with recombination between *Prm2* and *Igl1* on mouse chromosome 16 (Reeves et al. 1991). Restriction fragment length variants (RFLVs) between the founder strains, BALB/c (*Mus musculus domesticus*) and MOLD/Rk (*M.m. molossinus*), were identified for several markers from human Chr 22 (*Igl5*, *Comt*, and *N41*) and informative DNAs from the backcross panel were typed to determine the gene order (Fig. 1). The three SSLP markers were included to link this map to that of the Whitehead Center for Genome Research (Cambridge, MA). Not all recombinant DNAs were typed for all markers, giving rise to “fractional mice” in this interval mapping procedure as described by Irving et al. (1993). The genetic distances (recombination fractions) were: *D16Mit88*–*Igl5*– 1.9 ± 0.7 cM (7/376.8)–*Tsk1*, *Tsk2*, *Mit143*– 0.3 ± 0.3 cM (1/376.8)–*Tsk1*, *Tsk2*, *Mit143*– 0.5 ± 0.4 cM (2/391)–*Comt*, *N41*, *Igl1*– 0.3 ± 0.2 cM (1/407)–*Mit144*. Because *D16Mit143* showed no recombination with *Tsk1*, the homolog of which is in the middle of the MDGCR, either probe could be used to screen mouse YAC and BAC libraries for genomic segments possibly syntenic to the MDGCR.

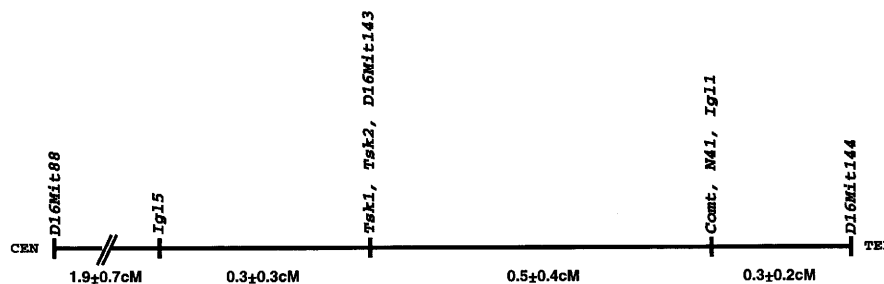


Figure 1 Composite genetic map of the region of mouse Chr 16 syntenic to the human DGS locus. Numbers refer to sex averaged distances.

Homologs of Human MDGCR Genes Are Located within a 150-kb Segment of Chr 16

Screening a mouse YAC library (Research Genetics) with *D16Mit143* resulted in the selection of a single 410-kb YAC (097C12) that was subsequently sub-cloned into cosmids. The presence of the *Tsk1* gene within this YAC was confirmed using PCR primers

(Table 1) specific to the unique carboxyl terminus of the gene. An independent screening of a BAC library (Research Genetics) using *Tsk1* PCR primers identified a single 125-kb BAC 40-O20 (designated BAC-O20) (Fig. 2). Two additional BACs (98D3 and 146J20, designated D3 and J20) were isolated with PCR primers from m595, a mouse genomic fragment mapping centromeric of *Lan* (see Methods).

Table 1. PCR Primers and Annealing Temperatures

| A. Primers used for YAC and BAC library screens | | | |
|---|---|-----------------------|---------------|
| Name | Sequence (5'-3') | Annealing temperature | Fragment size |
| <i>Mit143</i> | F GAATCCCTCCTCTGGACCTC R ACAGTTTTTGTAGGCTACGTATCG | 55 | 192 |
| <i>Tsk1</i> | F CCTGGTGCTGATAAGAAGTC R GGGGGGAGAAAGAAAGTGAC | 57 | 333 |
| <i>m595</i> | F ACGTCTGTGGGAACGACC R CTTTGCTTCTGGCTTTACGG | 56 | 229 |
| A. Primers used for RT-PCR of total RNA from 12 day mouse embryos | | | |
| Name | Sequence (5'-3') | Annealing temperature | Fragment size |
| <i>Lan</i> | F CTTGCCGCGGCCAGCTGGAT R CGGCTCGGTGACAGTGAGAAC | 64 | 210 |
| <i>Tsk1</i> | 1. F CAGCCAGATGTCAACCGC R GGGGGGAGAAAGAAAGTGAC | 57 | 477 |
| | 2. F CCTGGTGCTGATAAGAAGTC R GCTCTGCCTCATTCCCTG | 57 | 238 |
| <i>Tsk2</i> | 1. F CCATGTCTTCTGCCTCCTTCAAG R CCACTAGGTACTTGCTTTCTCCAC | 57 | 240 |
| | 2. F GCAAGTTGGATACTCGACCAGG R TTTAGCTCTGGAAGTCTCAGCCAG | 58 | 162 |
| <i>Dgsh</i> | 1. F TGGATGAGTCCTAGAGAAAAAGAC R TTCGTCAATTAACCACTATGTTG | 64 | 154 |
| | 2. F TCCCCACTGGCACTCTCAAT R TACGTTCTATTAGCTTTCTTCCAC | 64 | 102 |
| <i>Dgsi</i> | F TTCCAGGAGATCATGGAGGTGG R TGCTCAAAGCCTGACTGAAGGG | 60 | 291 |
| <i>Gscl</i> | 1. F GGATTAGCTCTGAGGACTGGCGACGGGATT R CTTGTGCTCTACTTCATAAAGCCAGATAAACT | 70 | 798 |
| | 2. F CTTCAGCGAGGAGCAGCTGC R TTGGGAGTTTTCTTAGTCCC | 57 | 207 |
| <i>Dgsj</i> | F GCGTCCTGGGCCTGTACCGCG R CCAGTTGCCTAGGGAGGTCATG | 64 | 389 |

GALILI ET AL.

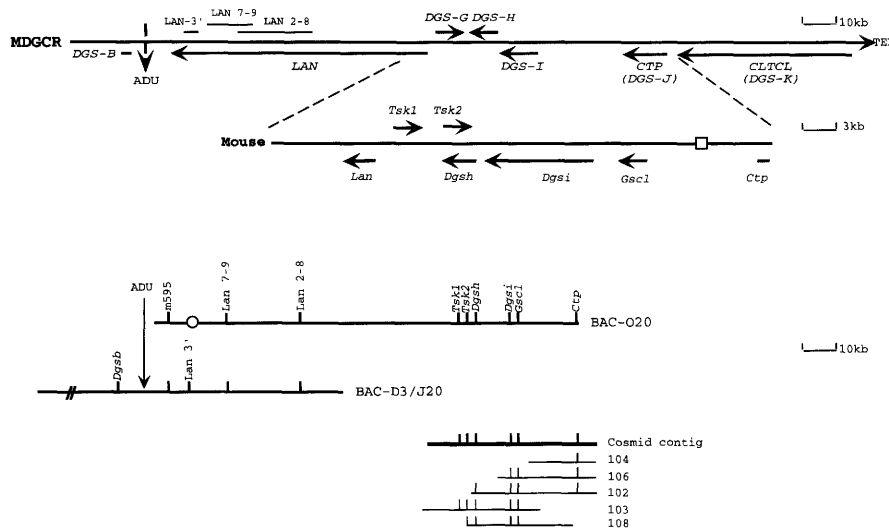


Figure 2 Comparison of the human MDGCR with the homologous region in the mouse by Southern blotting and sequence analysis. (*Top*) Schematic representation of the human MDGCR showing the positions of transcribed genes (Gong et al. 1996) as well as the location of partial cDNA probes (LAN-3', LAN 7-9, Lan 2-8) within the *LAN* gene (W. Gong and M.L. Budarf, in prep.). The position of the ADU balanced translocation is indicated by the broken vertical arrow. The direction of transcription is noted by the horizontal arrows. The enlarged insert represents the portion of the mouse syntenic region identified within the cosmids. This includes only the first exon of *Lan*. The exact distance distal to the open square is unknown. Grouped in the lower half of the figure is a schematic summary of the BACs and cosmids. The genes identified on the BACs by hybridization are noted. The cosmid contig was assembled by the hybridization pattern of 150 cosmids with the noted probes. Five representative cosmids are arrayed below the contig. Cosmid 103 (c143-103) was sequenced (see text). The open circle indicates a deletion. The 120 kb size of the BACs D3/J20 is off scale as indicated by the break.

cDNA probes designed to detect the presence of additional genes found within the human MDGCR were used to further characterize the YAC-derived cosmids and the three BACs. A comparison of the three BACs and the cosmids with the human MDGCR is shown in Figure 2. Three probes (Lan-3', Lan 7-9, Lan 2-8) representing different regions of the 85-kb human gene *LAN* (W. Gong and M.L. Budarf, in prep.) were detected in two of the three BACs. BAC-O20 was negative for Lan 3', suggesting a deletion of this region. However, the deletion cannot exceed 20 kb because probes m595 and Lan 7-9 are both present on a 20-kb lambda clone isolated from a mouse genomic library (see Methods). The cDNA probes designed from the human genes *DGS-H*, *DGS-I*, and *CTP* (Gong et al. 1996) hybridized with BAC-O20, but not with BACs D3 and J20. In the human, *DGS-I* occupies 11 kb of genomic DNA distal to *DGS-H*, and *CTP*, a tricarboxylate transport protein gene, is located telomeric to *DGS-I*. *CTP* was

the most distal gene represented within the BACs and cosmids. Restriction analysis of D3 and J20 with the *LAN* probes revealed homology with a telomeric 60-kb *NotI* fragment from these BACs. D3 and J20 failed to hybridize with the human ADU translocation breakpoint probe. However, preliminary sequence analysis of D3 suggests the presence of a region with 59% nucleotide identity to the ADU translocation breakpoint. Both BACs were positive for the *DGS-B* probe, which represents a gene mapped 5 kb proximal to the ADU breakpoint in the human MDGCR (Gong et al. 1996). *DGS-B* was the most proximal of the human gene homologs found within these BACs even though they carried an additional 40–50 kb of DNA presumably extending proximal to *Dgsh*.

Over 150 cosmids were constructed from YAC 097C12 for more detailed comparisons. Various combinations of *DGS-H*, *DGS-I*, and *CTP* were detected on each of the cosmids by Southern blotting and

hybridization. The restriction digest fragment sizes for the cosmids were the same as for the YAC, indicating that the cosmids were intact. No cosmids were detected using probes mapping centromeric to *DGS-G* or telomeric to *CTP*, suggesting that only a part of YAC 097C12 was subcloned. These data, combined with those of the BACs, account for a segment of mouse Chr 16 that appears syntenic to 150 kb of the MDGCR starting 5 kb proximal to the ADU breakpoint and extending distally to *CTP*.

Comparative Sequence Analysis Indicates Highly Conserved Genomic Synteny

The cosmid c143-103, likely to contain the greatest number of genes (Fig. 2), was selected for sequencing (GenBank accession no. U70231) and more detailed comparisons with the human MDGCR (Fig. 3). Five complete genes and an exon of one other were identified within this cosmid. These genes and

MOUSE CHR16 SYNTENY AND DGS

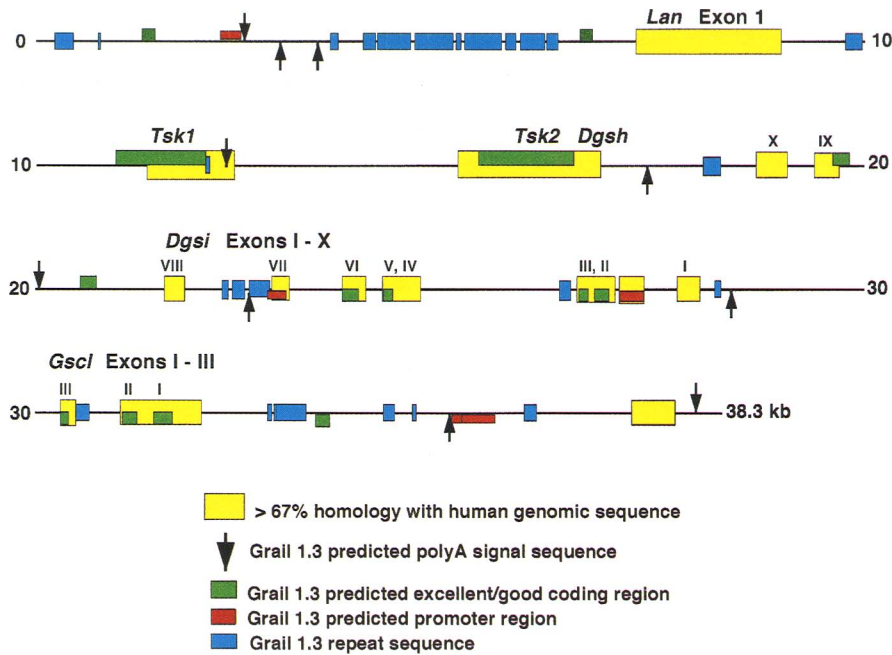


Figure 3 Sequence comparisons of the mouse cosmid c143-103 with a homologous portion of the human MDGCR. All comparisons are based on sequence analysis of cosmid c143-103 (accession no. U70231) and the homologous region of the human MDGCR (Gong et al. 1996). The regions of greater than 67% homology between the mouse and human sequences were determined using the Genetics Computer Group CG COMPARE and DOTPLOT programs. The Grail 1.3 software was used. Gene names are noted above the exons. Roman numerals identify exons. Grail 1.3 repeat sequences (blue box) identify mouse interspersed repeat sequence families (B1, B2, L1, and MT). Other color codes are noted on the figure.

gene fragments accounted for 17 exons, all of which were within regions showing >67% homology to genes within the human 22q11.2 detected by cDNA selection (Gong et al. 1996) (Fig. 3, yellow boxes). Interestingly, only nine of these exons were predicted by Grail 1.3 (Genetics Computer Group, Madison, WI) (Fig. 3, green boxes). Grail 1.3 also failed to accurately identify the promoter and poly(A) signal sequences. However, Grail 1.3 was able to identify most of the mouse repetitive sequences (Fig. 3, blue boxes) found by sequence analysis. The order of these 17 exons is identical to that of the corresponding fragment of human 22q11.2. Specifically, the proximal end of cosmid c143-103 commenced within the first intron of *Lan* and extended 8 kb distally through the first *Lan* exon. The coding region of this exon showed a 93% identity, at the amino acid level, to that of the human. Interestingly, two serine/threonine kinase genes, *Tsk1* and *Tsk2*, were identified starting 2.9 kb distal to the *Lan* exon. They are separated by 3.3 kb

and consist of single exons (Figs. 3 and 4A). *Tsk1* is predicted to encode a protein containing 364 amino acids, whereas *Tsk2* encodes a protein of 357 amino acids. The two kinase domains share 87% amino acid identity. However, the carboxy-terminal regions are highly divergent, allowing the two genes to be distinguished by Southern blot and PCR analysis. *Tsk1* is the previously described mouse testicular serine-threonine kinase gene (Bielke et al. 1994). *Tsk2*, on the other hand, most closely resembles *DGS-G*, identified previously in the human by cDNA selection (Gong et al. 1996). These two genes share a 95% amino acid identity within the kinase domain and a 83% amino acid identity within the variable carboxyl terminus. As with *DGS-G*, transcription for the two mouse genes is predicted to be in a centromeric to telomeric direction. Immediately distal to *Tsk2* is *Dgsh*, which appears to be similar to *DGS-H* in the MDGCR. This gene is represented by a 1145-

bp intronless cDNA of unknown function (Gong et al. 1996). However, sequence analysis shows significant homology (75%) between the mouse and human gene only within the last 240 bp. Distal to *Dgsh* is *Dgsi*, a second gene of unknown function, which is similar to a gene in *Caenorhabditis elegans* (accession no. P34420). Mouse *Dgsi* shows 92% amino acid identity to the human *DGS-I* and, like the human gene, consists of 10 exons with a predicted telomeric to centromeric direction of transcription. Its last exon is 2.0 kb distal to *Tsk2*. Finally, a Grail-predicted exon containing a homeobox motif was identified. This exon is part of a gene located 3.8 kb distal to *Dgsi*. As seen in Figure 4B, the homeobox is highly homologous to that of mouse *gooseoid* (*Gsc*) (Blum et al. 1992) and, like *Gsc*, consists of three exons. The mouse *Gsc* and the *Gsc*-like (*Gscl*) gene identified here show 73% identity within the predicted homeobox. In both cases, the second helical domain of the homeobox is split between two exons (Scott et al. 1989; Blum et al. 1992). The re-

GALILI ET AL.

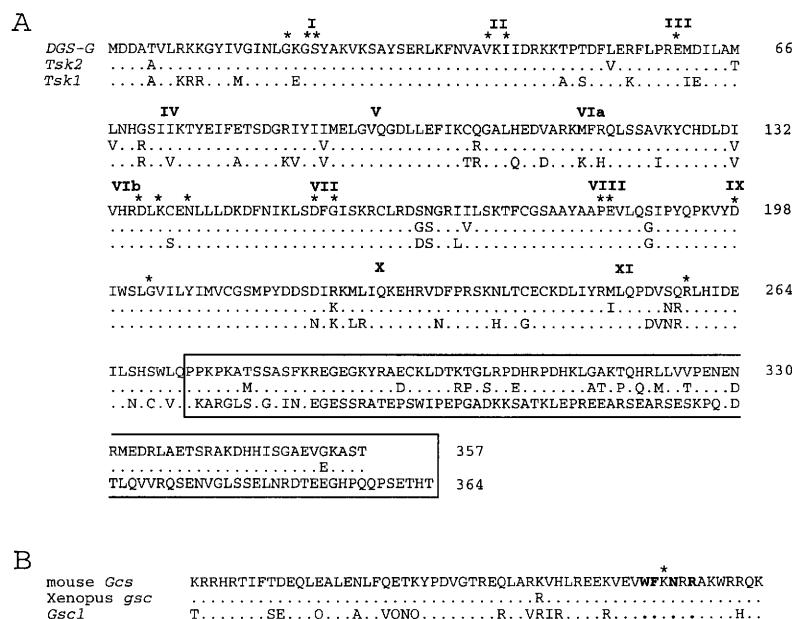


Figure 4 Amino acid sequence comparison of *Tsk1*, *Tsk2*, and *Gscl* with previously reported homologs. (A) Comparison of the human and mouse serine–threonine kinases. The unique carboxy-terminal region of the protein is boxed. Dots indicate amino acid identity. Asterisks indicate the amino acids conserved in serine–threonine kinases. Roman numerals identify consensus subdomains (Hanks and Hunter 1995). (B) Comparison of the homeobox domains of *Gscl* to those of *Xenopus* and mouse *Gsc* (Blum et al. 1992). The four invariant amino acids typical of most homeodomains (Scott et al. 1989) are shown in bold. The position of the conserved exon boundaries are indicated by an arrow. The asterisk indicates the lysine in *Gsc* and *Gscl* that replaces the glutamine found in other homeodomains (see text).

maining cosmid sequence extends 6.6 kb distally from the first *Gscl* exon, and although a 500-bp region with >67% homology to a similar segment of the MDGCR was noted, no additional genes were identified.

Presumptive Transcription Products for Shared Genes are Detected in the Early Mouse Embryo

Total RNA was extracted from 12 day post conception (dpc) mouse embryos, and screened by RT–PCR for the presence of mRNA transcribed from each of these genes. The primers used are shown in Table 1. Care was taken to eliminate and control for genomic DNA contamination, as several of these genes consist of a single exon. All samples were digested with DNase I, and control PCR reactions were run without prior reverse transcription. Where possible, primers were designed across predicted introns. As shown in Figure 5, presumptive transcripts from *Lan*, *Tsk1*, *Tsk2*, *Dgsi*, *Gscl*, *Dgsh*, and *Ctp* were

detected in 12-dpc mouse embryos. Southern blot analysis with internal oligonucleotides confirmed the specificity of each product (data not shown).

DISCUSSION

The etiology of congenital defects associated with DGS and VCFS is unknown. Although the current consensus is that many of these abnormalities may be the result of an interruption in developmental processes involving the cranial neural crest (Kirby and Bockman 1984; Lammer and Opitz 1986), the molecular events leading to the disruption of normal cardiac and craniofacial development have not been elucidated. At the genetic level, several possibilities are suggested. These include an autosomal dominant mutation or deletion involving a single gene within the MDGCR of 22q11.2, a dominant cumulative effect on embryonic development resulting from the loss of multiple genes within this region, or position effects resulting in aberrant regulation of genes near the site of chromosome breakage and reunion. To test these possibilities, it is necessary to develop a model in which genetic changes similar to those associated

with DGS might be induced. The mouse is the obvious animal in which to develop such a model. However, there is no single mouse chromosome homologous to human 22. Thus, the suitability of the mouse as a model depends upon the existence of a genomic segment syntenic to the 250-kb MDGCR on chromosome 22q11.2 (Budarf et al. 1995; Gong et al. 1996).

The results reported here demonstrate that the 150-kb segment of mouse Chr 16 carried by the isolated BACs and YAC- derived cosmids is syntenic to a large portion of the human MDGCR. Southern blot analysis confirms the presence of homologs to six of the 11 genes within the MDGCR identified by transcriptional analysis (Gong et al. 1996). Sequence comparisons of four of the six mouse genes with their human counterparts reveal that they are present in the same order, consist of the same number of exons, and are transcribed in the same direction. Further, RT–PCR analysis of 12-dpc embryos provides an initial indication of the

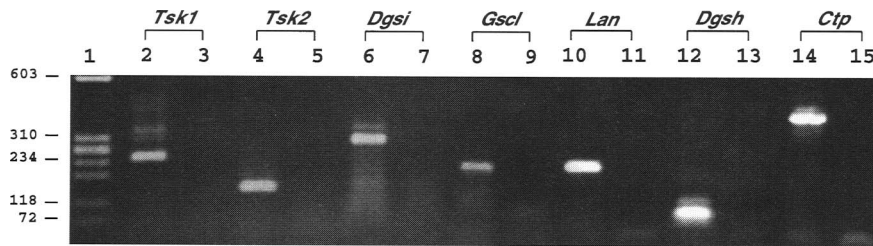


Figure 5 RT-PCR products corresponding to mouse homologs of genes within the human MDGCR are detected in the developing embryo. RT-PCR products from total RNA isolated from 12-dpc mouse embryos are shown in lanes 2–15. The corresponding genes are noted *above* each lane. All total RNA samples were treated with DNase I, and reverse transcriptase was omitted from control reactions shown in lanes 3,5,7,9,11,13, and 15. The positions of (ϕ X174 markers are noted in lane 1.

presence of transcripts from the genes identified by comparative sequencing. Assessment of their biological activity and their relevance to DGS will require a detailed analysis of expression throughout development.

Although the similarities between these corresponding regions of the mouse and human genomes are striking, there are several interesting differences. Two genes predicted to encode serine-threonine kinases, *Tsk1* and *Tsk2*, were identified from the cosmid sequence. They are similar through the kinase domain, but diverge at the carboxyl terminus (Fig. 4A). Strong homology throughout, including the carboxyl terminus, shows *Tsk2* to be the homolog of the human *DGS-G* gene (Gong et al. 1996). The second more proximal mouse serine-threonine kinase is identical to the *Tsk1* gene identified as a transcript in mouse testis but not mapped previously (Bielke et al. 1994). *Tsk1* shows significant homology to a Grail-predicted exon sequence from the corresponding region of human Chr 22. Although the human *TSK1* has several in frame stop codons (E. Goldmuntz, B.A. Roe, and M.L. Budarf, pers. comm.), the mouse sequence predicts a functional gene, and a *Tsk1*-specific transcript was detected in 12-dpc mouse embryos.

Another interesting difference identified by comparative sequencing is divergence in *DGS-H* and *Dgsh*. Although these genes reside at corresponding positions in mouse and human, they share strong homology only within the last 240 bp of the 1.2-kb predicted exon. Neither sequence contains an open reading frame, bringing into question whether these sequences represent true genes. Their conservation through evolution, however, suggests that these sequences might be functionally important.

We have been unable to identify sequences close

to the ADU breakpoint that are conserved in mouse by hybridization based approaches. This is consistent with the idea that the genomic segments used as probes to date do not represent coding regions. Given the demonstrated ability of sequence comparison to pick up homologies missed by hybridization, completion of the comparative sequence analysis currently under way, which already has revealed regions with at least 59% identity, should be very useful in identifying conserved features around this important site.

A gene of particular interest, *Gscl*, is located distal to the first exon of *Dgsi*. A human homolog, *GSC1*, has been identified in the genomic sequence of the MDGCR (S. Gottlieb, B.S Emanuel, D.A. Driscoll, B. Sellinger, Z. Wang, B.A. Roe, and M.L. Budarf, in prep.). Both genes are composed of three exons and show a >70% identity at the amino acid level to the mouse *Gsc* gene (Blum et al. 1992). Like *Gsc*, the homeobox domain of *Gscl* is encoded in the last two exons, with a splice site in the second helix. As with *Gsc* and the related *Drosophila* gene *bicoid*, *Gscl* encodes a lysine residue in the homeodomain at a position occupied by a glutamine in most other homeobox genes (Scott et al. 1989). Interestingly, whereas mice with two inactivated copies of the *Gsc* gene develop to term, the pups are born with craniofacial and rib abnormalities (Rivera-Perez et al. 1995). However, heterozygous litter mates (haploid equivalents) with a single functional *Gsc* gene show no developmental defects.

Direct sequence comparisons of DNA from the MDGCR and the region of putative mouse synteny was essential for evaluating gene homology, gene order, and direction of transcription. Thus, whereas Grail analysis of the mouse sequence predicted only nine exons, direct sequence comparison identified seventeen, and showed them to be organized in the same order as in the MDGCR. In addition, Grail predicted two other highly conserved regions between exons I and II of *Dgsi*. One was identified as a promoter, but the function of the second is unknown. Clearly, comparative sequencing, as well as Grail analysis, is required to facilitate the identification of subtle differences and pinpoint cryptic conserved genomic segments and thus enable accurate structural and functional comparisons between genomic fragments.

Finally, the data presented here show that although there are some differences, there are striking similarities in gene content, order, exon composition, and direction of transcription between a 150-kb segment of mouse Chr 16 and the human MDGCR. These similarities are sufficient to suggest that the genetic perturbations of this region of the mouse chromosome will provide insight into the origins of the molecular basis of the DGS and VCFS phenotypes.

METHODS

Recombinational Mapping

Markers were mapped on (BALB/cj × MOLD/Rk)F1 × BALB/cj and BALB/cj × (BALB/cj × MOLD/Rk)F1 backcrosses by Southern blotting or PCR as described (Reeves et al. 1990). Mapping of *Igf1* on this panel was reported previously (Reeves et al. 1990). Probes and RFLVs used: *Igf5* was mapped using pλ5 (Miller et al. 1993) with an *EcoRV* RFLV (19.7-, 6.3-, 3.6-kb BALB/cj, 16.0-, 8.2-kb MOLD/Rk); a human DGS-G cDNA probe was used to map *Tsk2* with a *HincII* RFLV (4.6-, 1.7-, 0.5-, 0.3-kb BALB/cj, 6.7-, 4.5-, 0.3-kb MOLD/Rk) and *Tsk1* with a *BamHI* RFLV (8.3-, 3.4-kb BALB/cj, 8.4-, 4.1-kb MOLD/Rk); *Comt* was mapped using a rat *Comt* cDNA probe (Grossman et al. 1992) with a *BglI* RFLV (4.3-, 2.6-kb BALB/cj, 5.0-, 2.4-kb MOLD/Rk); and *N41* was mapped using a human cDNA probe (Emanuel et al. 1993) with a *BamHI* RFLV (12.6-, 7.3-, 5.1-, 2.7-, 1.7-kb BALB/cj, 14.4-, 7.3-, 5.1-, 2.7-kb MOLD/Rk). *D16Mit88*, *143*, and *144* primers were obtained from Research Genetics (Dietrich et al. 1992) and mapped using a nonradioactive protocol and PCR conditions as described previously (Irving et al. 1994) with the modification that PCR was performed for 45 cycles.

Isolation of Murine YAC and BAC Clones

The YAC clone 097C12 was isolated by PCR screening of the mouse YAC DNA pools of Research Genetics using primers for *D16Mit143*. The BAC 40-O20 clone was isolated by PCR screening of the mouse BAC DNA pools, Release I of Research Genetics, with primers for *Tsk1*, whereas BACs 98D3 and 146J20 were isolated from Release II of the Mouse BAC DNA pools with the primers for m595. All primers are listed in Table 1.

Cosmid Library Construction

Total yeast DNA containing the YAC clone was partially digested with *Sau3A* and subcloned into the SuperCos cosmid vector (Stratagene). Several thousand cosmids were plated and probed with radiolabeled total mouse genomic DNA. A total of 150 mouse clones were isolated and analyzed by Southern blot.

Electrophoresis and Southern Blots

The size and stability of the YAC and BAC DNA was analyzed in a contour-clamped homogeneous electric-field electropho-

resis gel system (Bio-Rad CHEF-DRIII) as described (Ausubel et al. 1995). Southern blots of *EcoRI* digests of BAC and cosmid DNA were carried out as described (Sambrook et al. 1989). High-stringency final washes of $0.1 \times$ SSC, 0.1% SDS at 64°C were used with murine probes, while low-stringency final washes of $0.5\text{--}1.0 \times$ SSC, 0.1% SDS at 50°C were used with human probes.

Probes

Human probes with the DGS designation are partial cDNAs and are described in detail elsewhere (Gong et al. 1996). The human probes for the *LAN* gene are derived from the cDNA as described in W. Gong and M.L. Budarf (in prep.) and refer to the 3' untranslated region (*LAN-3'*), exons 2–8 (*LAN 2–8*), and exons 7–9 (*LAN 7–9*). The human ADU breakpoint spanning probe (ADU) is described as *pf13-mex40* in Budarf et al. (1995). The murine probe m595 was isolated from a lambda phage mouse genomic library screened with the *LAN-3'* probe as a 0.6-kb *SstI-HindIII* fragment mapping 8 kb proximal to *LAN*. This fragment was subcloned and partially sequenced, and primers were derived for PCR (Table 1). Probes for the unique carboxyl terminus of both of the murine serine–threonine kinase genes were generated by PCR of either total genomic mouse DNA or isolated cosmid DNA. The primers (Table 1) were designed from sequence entirely within the unique carboxy-terminal region to prevent any possible cross-reactivity with other serine–threonine kinases. Mapping probes are described under recombinational mapping.

PCR Primers

Primer pairs are listed in Table 1. Nested PCR primers for RT-PCR are listed as the second set where appropriate. Primers for *Dgsi*, *Gscl*, and *Ctp* cross at least one intron.

PCR and RT-PCR

YAC and BAC DNA pools were screened by PCR according to supplier recommendations. Annealing temperatures used for all primers are noted in Table 1. Thermocycling was performed in either a Perkin Elmer Cetus thermal cycler or a Perkin Elmer Cetus Geneamp system 9600 using *Taq* polymerase (Boehringer-Mannheim) in standard buffer conditions except where noted below. For RT-PCR, total RNA, isolated from 12-dpc embryos using Ultraspec RNA (Biotecx) according to supplier instructions, was treated with RNase-free DNase I (Boehringer-Mannheim) at 1 U/μg RNA. Approximately 5 μg of DNase I-treated RNA was used for oligo(dT)-primed reverse transcription (RT) using Super Script RNase H- reverse transcriptase (GIBCO-BRL). Controls were processed identically with the omission of the reverse transcriptase. One-tenth of the RT reaction was used for PCR amplification. *Dgsi*, *Lan*, and *Ctp* gene transcripts were readily detected after one round of amplification of 35 cycles. Nested PCR was required in order to readily detect transcripts from *Tsk1*, *Tsk2*, *Dgsh*, and *Gscl*. The first amplification consisted of 30 cycles and the nested amplification consisted of 12 cycles. Because of the extremely high GC content of the *Gscl* gene, a two-step PCR program (98°C for 1 min, 70°C for 5 min) using native Pfu polymerase (Stratagene) was used for the first round of amplification (Dutton et al. 1993). The nested amplification of 12 cycles was performed with *Taq* polymerase (Boehringer-Mannheim) in the presence of 5% formamide.

DNA Sequencing

The sequence of the entire mouse genomic DNA cosmid clone was determined at a level of 4.5-fold redundancy via the previously described, double-stranded, shotgun-based approach (Bodenteich et al. 1993) using ABI-PRISM fluorescent-labeled terminators and either forward or reverse universal primers. Regions corresponding to the sequencing vector (pUC-18) were removed from the individual sequence reads, and the resulting data was assembled into contiguous fragments using the TED and XGAP programs (Dear and Staden 1991). The individual contigs were joined into a final, unique sequence using custom, synthetic primers and Taq DNA polymerase cycle sequencing with fluorescent terminators.

ACKNOWLEDGMENTS

We thank Dr. Maja Bucan for the Comt probe. This work was supported in part by Public Health Service grants HG 00405 (R.H.R.), 1 P50 HL 51533, and CA 10815 (C.A.B.), 1 PO1 DC 02027, and 1 P50 HG 00425 (B.S.E.), and National Center for Human Genome Research grant HG 00313 (B.A.R.) and a grant from the Philadelphia Heart Institute (C.A.B.).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Augusseau, S., S. Jouk, P. Jalbert, and M. Prieur. 1986.
- DiGeorge syndrome and 22q11 rearrangements. *Hum. Mol. Genet.* **74**: 206.
- Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. 1995. *Current protocols in molecular biology*. Wiley & Sons, New York, NY.
- Bielke, W., R.J. Blaschke, G.C. Miescher, G. Zurcher, A.C. Andres, and A. Ziemiecki. 1994. Characterization of a novel murine testis-specific serine/threonine kinase. *Gene* **139**: 235–239.
- Blum, M., S.T. Gaunt, K.W.Y. Cho, H. Steinbeisser, B. Blumberg, D. Bittner, and E.M. De Robertis. 1992. Gastrulation in the mouse: The role of the homeobox gene *gooseoid*. *Cell* **69**: 1097–1106.
- Bodenteich, A., S. Chissoe, Y.F. Wang, and B.A. Roe. 1993. Shotgun cloning as the strategy of choice to generate templates for high-throughput dideoxynucleotide sequencing. In *Automated DNA sequencing and analysis techniques* (ed. J.C. Venter), pp. 42–50. Academic Press, London, UK.
- Bucan, M., B. Gatalica, P. Nolan, A. Chung, A. Leroux, M.H. Grossman, J.H. Nadeau, B.S. Emanuel, and M. Budarf. 1993. Comparative mapping of 9 human chromosome 22q loci in the laboratory mouse. *Hum. Mol. Genet.* **2**: 1245–1252.
- Budarf, M.L., J. Collins, W. Gong, B. Roe, Z. Wang, B. Sellinger, D. Michaud, D. Driscoll, and B.S. Emanuel. 1995. Cloning a balanced translocation associated with DiGeorge syndrome and identification of a disrupted candidate gene. *Nature Genet.* **10**: 269–288.
- Cabin, D.E., M.P. Citron, J. McKee-Johnson, A.E. Mjaatvedt, and R.H. Reeves. 1996. Mouse chromosome 16. *Mamm. Genome* **7** (in press).
- Conley, M.E., J.B. Beckwith, J.F.K. Mancer, and L. Tenckhoff. 1979. The spectrum of the DiGeorge syndrome. *J. Pediatr.* **94**: 883–890.
- Dear, S. and R. Staden. 1991. A sequence assembly and editing program for efficient management of large scale DNA sequencing projects. *Nucleic Acids Res.* **19**: 3907–3911.
- Demczuk, S., R. Aledo, J. Zucman, O. Delattre, C. Desmaze, L. Dauphinot, P. Jalbert, G.A. Rouleau, G. Thomas, and A. Aurias. 1995. Cloning of a balanced translocation breakpoint in the DiGeorge syndrome critical region and isolation of a novel potential adhesion receptor gene in its vicinity. *Hum. Mol. Genet.* **4**: 551–558.
- Demczuk, S., G. Thomas, and A. Aurias. 1996. Isolation of a novel gene from the DiGeorge syndrome critical region with homology to *Drosophila gdl* and to human *LAMC1* genes. *Hum. Mol. Genet.* **5**: 633–638.
- Dietrich, W., H. Katz, S.E. Lincoln, and H.S. Shin. 1992. A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131**: 423–447.
- Driscoll, D.A., M.L. Budarf, and B.S. Emanuel. 1992. A genetic etiology for DiGeorge syndrome: Consistent deletions and microdeletions of 22q11. *Am. J. Hum. Genet.* **50**: 924–933.
- Driscoll, D.A., J. Salvin, B. Sellinger, M.L. Budarf, D.M. McDonald-McGinn, E.H. Zackai, and B.S. Emanuel. 1993. Prevalence of 22q11 microdeletions in DiGeorge and velocardiofacial syndromes: Implications for genetic counseling and prenatal diagnosis. *J. Med. Genet.* **30**: 813–817.
- Dutton, C.M., C. Paynton, and S.S. Sommer. 1993. General method for amplifying regions of very high G + C content. *Nucleic Acids Res.* **21**: 2953–2954.
- Emanuel, B.S., D. Driscoll, E. Goldmuntz, S. Baldwin, J. Biegel, E.H. Zackai, D. McDonald-McGinn, B. Sellinger, N. Gorman, S. Williams, and M.L. Budarf. 1993. Molecular and phenotypic analysis of the chromosome 22 microdeletion syndromes. In *The phenotypic mapping of Down syndrome and other aneuploid conditions* (ed. C. Epstein), pp. 207–224. Wiley-Liss, New York, NY.
- Goldmuntz, E., D. Driscoll, M.L. Budarf, E.H. Zackai, D.M. McDonald-McGinn, J.A. Biegel, and B.S. Emanuel. 1993. Microdeletions of chromosomal region 22q11 in patients with congenital conotruncal cardiac defects. *J. Med. Genet.* **30**: 807–812.
- Goldmuntz, E., Z. Wang, B.A. Roe, and M.L. Budarf. 1996. Cloning, genomic organization and chromosomal localization of human citrate transport protein to the DiGeorge/velocardiofacial syndrome minimal critical region. *Genomics* **33**: 271–276.

GALILI ET AL.

- Gong, W., B.S. Emanuel, J. Collins, D.H. Kim, Z. Wang, F. Chen, G. Zhang, B. Roe, and M.L. Budarf. 1996. A transcription map of the DiGeorge and velo-cardio-facial syndrome minimal critical region on 22q11. *Hum. Mol. Genet.* **5**: 789–800.
- Grossman, M.H., B.S. Emanuel, and M.L. Budarf. 1992. Chromosomal mapping of the human catechol-O-methyltransferase gene to 22q11.1–q11.2. *Genomics* **12**: 822–825.
- Halford, S., R. Wadey, C. Roberts, S.C.M. Daw, J.A. Whiting, H. O'Donnell, I. Dunham, D. Bentley, E.A. Lindsay, A. Baldini, F. Francis, H. Lehrach, R. Williamson, D.I. Wilson, J. Goodship, I. Cross, J. Burn, and P.J. Scambler. 1993. Isolation of a putative transcriptional regulator from the region of 22q11 deleted in DiGeorge syndrome, Shprintzen syndrome and familial congenital heart disease. *Hum. Mol. Genet.* **2**: 2099–2107.
- Hanks, S.K. and T. Hunter. 1995. The eukaryotic protein kinase superfamily: Kinase (catalytic) domain structure and classification. *FASEBJ.* **9**: 576–596.
- Heisterkamp, N., M.P. Mulder, A. Langeveld, J.T. Hoeve, Z. Wang, B.A. Roe, and J. Groffen. 1995. Localization of the human mitochondrial citrate transporter protein gene to chromosome 22q11 in the DiGeorge syndrome critical region. *Genomics* **29**: 451–456.
- Irving, N.G., M.P. Citron, and R.H. Reeves. 1993. The positions of twelve simple sequence repeat markers relative to reference markers on mouse chromosome 16. *Mamm. Genome* **4**: 364–367.
- Kedra, D., M. Peyrard, I. Fransson, J.E. Collins, I. Dunham, B.A. Roe, and J.P. Dumanski. 1996. Characterization of a second human clathrin heavy chain polypeptide gene (CLH-22) from chromosome 22q11. *Hum. Mol. Genet.* **5**: 625–631.
- Kirby, M.L. and D.E. Bockman. 1984. Neural crest and normal development: A new perspective. *Anat. Rec.* **209**: 1–6.
- Koop, B.F. and L. Hood. 1994. Striking sequence similarity over almost 100 kilobases of human and mouse T cell receptor DNA. *Nature Genet.* **7**: 48–53.
- Lammer, E.J. and J.M. Opitz. 1986. The DiGeorge anomaly as a developmental field defect. *Am. J. Med. Genet.* **29**: 113–127.
- Levy, A., S. Demczuk, A. Aurias, D. Depetris, M. Mattei, and W. Philip. 1995. Interstitial 22q11 microdeletion excluding the ADU breakpoint in a patient with DiGeorge syndrome. *Hum. Mol. Genet.* **4**: 2417–2419.
- Lindsay, E.A., P. Rizzu, R. Antonacci, V. Jurecic, J. Delmas-Mata, C. Lee, U. Kim, P.J. Scambler, and A. Baldini. 1996. A transcription map in the CATCH22 critical region: Identification, mapping and ordering of four novel transcripts expressed in heart. *Genomics* **32**: 104–112.
- Miller, R.D., J.H. Ozaki, and R. Riblet. 1993. The mouse severe combined immune deficiency (scid) mutation is closely linked to the B-cell-specific developmental genes VpreB and lambda 5. *Genomics* **16**: 740–744.
- Nadeau, J., M.T. Davisson, D.P. Doolittle, P. Grant, A.L. Hillyard, M.R. Kosowsky, and T.H. Roderick. 1992. Comparative map for mice and humans. *Mamm. Genome* **3**: 480–536.
- Ramirez-Solis, R., P. Liu, and A. Bradley. 1995. Chromosome engineering in mice. *Nature* **378**: 720–724.
- Reeves, R.H., M.R. Crowley, B.F. O'Hara, and J.D. Gearhart. 1990. Sex strain and species differences affect recombination across an evolutionary conserved segment of mouse Chromosome 16. *Genomics* **8**: 141–148.
- Reeves, R.H., M.R. Crowley, W.S. Mosely, and M.F. Seldin. 1991. Comparison of interspecific to intersubspecific backcrosses demonstrates sex and species differences in recombination on mouse chromosome 16. *Mamm. Genome* **1**: 158–164.
- Rivera-Perez, J.A., M. Mallo, M. Gendron-Maguire, T. Gridley, and R.R. Behringer. 1995. *gooseoid* is not an essential component of the mouse gastrula organizer but is required for craniofacial and rib development. *Development* **121**: 3005–3012.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Scott, M.P., J.W. Tamkun, and G.W. Hartzell III. 1989. The structure and function of the homeodomain. *Biochem. Biophys. Acta* **989**: 25–37.
- Sirotkin, H.A., B.A. Morrow, R. DasGupta, R. Goldberg, S.R. Patanjali, G. Shi, L. Cannizzaro, R. Shprintzen, S.M. Weissman, and R. Kucherlapati. 1996. Isolation of a new clathrin heavy chain gene with muscle-specific expression from the region commonly deleted in velo-cardio-facial syndrome. *Hum. Mol. Genet.* **5**: 617–624.
- Stevens, C.A., J.C. Carey, and A.O. Shigeoka. 1990. DiGeorge anomaly and velo-cardio-facial syndrome. *Pediatrics* **85**: 526–530.
- Sutherland, H.F., R. Wadey, J.M. McKie, C. Taylor, U. Atif, K.A. Johnstone, S. Halford, U.-J. Kim, J. Goodship, A. Baldini, and P.J. Scambler. 1996. Identification of a novel transcript disrupted by a balanced translocation associated with DiGeorge syndrome. *Am. J. Hum. Genet.* **59**: 23–31.
- Wadey, R., S. Daw, C. Taylor, U. Atif, S. Kamath, S. Halford, H. O'Donnell, D. Wilson, J. Goodship, J. Burn, and P. Scambler. 1995. Isolation of a gene encoding an integral membrane protein from the vicinity of a balanced translocation breakpoint associated with DiGeorge syndrome. *Hum. Mol. Genet.* **4**: 551–558.

Received August 13, 1996; accepted in revised form November 15, 1996.

Erratum

Genome Research 7: 17–26 (1997)

A Region of Mouse Chromosome 16 Is Syntenic to the DiGeorge, Velocardiofacial Syndrome Minimal Critical Region

Naomi Galili, H. Scott Baldwin, Jim Lund, Roger Reeves, Weilong Gong, Zhili Wang, Bruce A. Roe, Beverly S. Emanuel, Sudhir Nayak, Craig Mickanin, Marcia L. Budarf, and Clayton A. Buck

An incorrect forward primer sequence was given for the first set of primers used in the RT-PCR of the *Gsc1* gene. The correct set should read

1. FATGGCGACTGCAGGCAGCGCGCC
RCTTGTGCTCTAC TTCATAAAGCCAGATAAACT