*Am. J. Hum. Genet. 65:111–124, 1999*

# Campomelic Dysplasia Translocation Breakpoints Are Scattered over 1 Mb Proximal to *SOX9:* Evidence for an Extended Control Region

Dietmar Pfeifer,[1] Ralf Kist,[1,2] Ken Dewar,[3] Keri Devon,[3] Eric S. Lander,[3] Bruce Birren,[3] Lech Korniszewski,[4] Elke Back,[1] and Gerd Scherer[1]

[1]Institute of Human Genetics and Anthropology and [2]Faculty for Biology, University of Freiburg, Freiburg, Germany; [3]Whitehead Institute/MIT Center for Genome Research, Cambridge, MA; and [4]Clinic of Infants and Metabolic Diseases, Medical Academy of Warsaw, Warsaw

## Summary

**Campomelic dysplasia (CD), a skeletal malformation syndrome with or without XY sex reversal, is usually caused by mutations within the *SOX9* gene on distal 17q. Several CD translocation and inversion cases have been described with breakpoints outside the coding region, mapping to locations >130 kb proximal to *SOX9*. Such cases are generally less severely affected than cases with *SOX9* coding-region mutations, as is borne out by three new translocation cases that we present. We have cloned the region extending 1.2 Mb upstream of the *SOX9* gene in overlapping bacterial-artificial-chromosome and P1-artificial-chromosome clones and have established a restriction map with rare-cutter enzymes. With sequence-tagged-site–content mapping in somatic-cell hybrids, as well as with FISH, we have precisely mapped the breakpoints of the three new and of three previously described CD cases. The six CD breakpoints map to an interval that is 140–950 kb proximal to the *SOX9* gene. With exon trapping, we could isolate five potential exons from the YAC 946E12 that spans the region, four of which could be placed in the contig in the vicinity of the breakpoints. They show the same transcriptional orientation, but only two have an open reading frame (ORF). We failed to detect expression of these fragments in several human and mouse cDNA libraries, as well as on northern blots. Genomic sequence totaling 1,063 kb from the *SOX9* 5′-flanking region was determined and was analyzed by the gene-prediction program GENSCAN and by a search of dbEST and other databases. No genes or transcripts could be identified. Together, these data suggest that the chromosomal rearrangements most likely remove one or more *cis*-regulatory elements from an extended *SOX9* control region.**

Received January 15, 1999; accepted for publication May 6, 1999; electronically published June 8, 1999.

Address for correspondence and reprints: Dr. Gerd Scherer, Institute of Human Genetics and Anthropology, Breisacherstrasse 33, D-79106 Freiburg, Germany. E-mail: scherer@humangenetik.ukl.uni-freiburg.de


0002-9297/99/6501-0016$02.00

## Introduction

Campomelic dysplasia (CD [MIM 114290]) is a rare autosomal dominant syndrome characterized by skeletal anomalies including bowed femora and tibiae, hypoplastic scapulae, 11 pairs of ribs, pelvic malformations, and clubbed feet. Malformations such as micrognathia, retroglossia, cleft palate, narrow airways caused by tracheobronchial cartilage defects, hypoplastic lungs, and a bell-shaped thorax are usually present and are the major cause of the severe respiratory problems that arise soon after birth, leading to death mostly during the neonatal period (Houston et al. 1983; Mansour et al. 1995). Male-to-female sex reversal of variable degree is present in approximately two-thirds of the XY cases (Houston et al. 1983; Mansour et al. 1995).

De novo chromosomal rearrangements involving distal 17q in several patients with CD (Maraia et al. 1991; Young et al. 1992; Tommerup et al. 1993) first allowed assignment of the CD/sex reversal locus to 17q24.3-q25.1 (Tommerup et al. 1993) and subsequently led to the identification of the *SOX9* gene in the vicinity of CD translocation breakpoints (Foster et al. 1994; Wagner et al. 1994). De novo heterozygous loss-of-function mutations in *SOX9* in cytologically normal patients with CD who have XY sex reversal identified haploinsufficiency for SOX9 as the cause of both phenotypes (Foster et al. 1994; Wagner et al. 1994). More than 25 mutations in *SOX9* that result in CD have been described so far, but no genotype/phenotype correlation is apparent (Foster et al. 1994; Wagner et al. 1994; Kwok et al. 1995; Cameron et al. 1996; Meyer et al. 1997; Goji et al. 1998; Hageman et al. 1998), indicating variable expressivity of both disease severity and the presence of XY sex reversal. Consistent with defects in skeletal and gonadal development seen in patients with CD, expression of the mouse *Sox9* gene has been documented at sites of ensuing cartilage deposition throughout embryogenesis (Wright et al. 1995; Ng et al. 1997) and in the developing testis (Kent et al. 1996; Morais da Silva et al. 1996). Both in vitro and in vivo studies have identified the *Col2a1* collagen gene as a target for SOX9 during chon-

**Table 1**

**Clinical Symptoms and Radiological Findings in Three New CD Translocation Cases**

| | Status in[a] | | |
|---|---|---|---|
| | Case 8, t(10;17) | Case 9, t(5;17) | Case 11, t(17;22) |
| Clinical symptoms: | | | |
| Macrocephaly | + | + | + |
| High forehead | + | + | + |
| Epicanthic folds | + | + | + |
| Flat nasal bridge | + | + | + |
| Low-set ears | + | ND | − |
| Cleft palate | − | + | + |
| Micrognathia | − | + | − |
| Respiratory distress | + | + | + |
| Congenital dislocation of hips | ND | ND | + |
| Bowed femora | + | − | − |
| Bowed tibiae | (+) | + | − |
| Pretibial skin dimples | + | + | ND |
| Clubfeet | + | + | + |
| Genitalia | Male, hypospadias | Normal female | Male, small penis |
| Radiological findings: | | | |
| Hypoplastic scapulae | + | + | + |
| Small chest | + | + | Bell shaped |
| 11 Ribs | − | + | − |
| Nonmineralized thoracic pedicles | + | ND | − |
| Kyphosis or scoliosis | + | + | − |
| Vertically narrow iliac bones | − | + | − |
| Abnormal ischial bones | + | + | ND |
| Poorly developed pubic bones | + | ND | + |
| Bowed femora | + | − | − |
| Bowed tibiae | + | + | ND |
| Hypoplastic fibulae | + | + | ND |
| Bowed humeri | ND | − | − |
| Bowed radii or ulnae | ND | − | − |
| Short first metacarpal | ND | ND | − |

[a] Case numbers are as in table 2. A plus sign (+) denotes presence of the symptom (the parentheses around the plus sign for "bowed tibiae" denotes that the bowing was mild), and a minus sign (−) denotes absence of the symptom; ND = not described.

drogenesis (Bell et al. 1997; Lefebvre et al. 1997; Ng et al. 1997). Recently, evidence for the involvement of SOX9 in the expression of the *AMH* gene, an early marker of Sertoli-cell differentiation, has been presented (de Santa Barbara et al. 1998).

*SOX9* is a member of the growing *SOX*-gene family, which is related by homology to the HMG-box region of the testis-determining gene *SRY* ("SOX" is an acronym for "SRY-related HMG-box") (Sinclair et al. 1990; Pevny and Lovell-Badge 1997). The HMG box encodes a DNA-binding domain encompassing 80 amino acids that is present in a large number of transcription factors (Grosschedl et al. 1994). In addition to its HMG domain, which has been shown to recognize typical SOX binding sequences (Lefebvre et al. 1997; Meyer et al. 1997; Ng et al. 1997), SOX9 has a second domain essential for its function, a C-terminal transcription-activation domain (Südbeck et al. 1996).

Until now, 11 CD cases have been described in which a chromosomal rearrangement seems to cause the disease (Maraia et al. 1991; Young et al. 1992; Tommerup et al. 1993; Mansour et al. 1995; Ninomiya et al. 1995; Wirth et al. 1996; Savarirayan and Bankier 1998; Wunderle et al. 1998); in 8 of these cases, the respective translocation or inversion breakpoints have been mapped to positions from 50 kb to possibly 350 kb proximal to *SOX9* (Foster et al. 1994; Wagner et al. 1994; Ninomiya et al. 1996; Wirth et al. 1996; Wunderle et al. 1998). It has been argued that these translocations or inversions either may act by exerting a long-range position effect by removing one or more *cis*-regulatory elements essential for correct *SOX9* expression or, alternatively, may disrupt a second CD gene (Foster et al. 1994; Wagner et al. 1994; Wirth et al. 1996). To shed light on these alternatives and to define in greater detail the critical region upstream of *SOX9,* we have more precisely mapped the translocation breakpoints of three CD cases described elsewhere (Wirth et al. 1996), and

we also present clinical and mapping data for three novel CD translocation cases. For this, we established a contig of bacterial artificial chromosome (BAC) and P1 artificial chromosome (PAC) clones that incorporates a total of 50 sequence-tagged site (STS) markers and spans ~1,200 kb of DNA upstream of *SOX9*. We show that translocations with breakpoints ⩽890–950 kb proximal to *SOX9* result in the CD phenotype. Extensive DNA sequencing of this region fails to reveal clear evidence of a second CD gene, suggesting that the translocations disrupt *cis*-acting regulation.

## Cases and Methods

### *Case Reports*

*Patient YS* (*case 8 in tables 1 and 2*).—This male proband was the second live-born child of a 28-year-old mother and a 30-year-old father. The parents were phenotypically normal, as was the elder sister. The pregnancy was normal. The child was delivered at term, by cesarean section, because of pathological cardiotocogram. Birth weight was 3,540 g, length was 43 cm, and head circumference was 38 cm. At birth, the boy displayed severe asphyxia and therefore needed intensive care. He had dysmorphic features, which are summarized in table 1. Moreover, the boy demonstrated the following additional clinical signs: muscular hypotonia, flat face, big eyes, nose with anteverted nostrils and nevus flammeus on its tip, broad philtrum, high-arched palate, dysmorphic ears with hypoplastic tragus, retrognathia, short and narrow chest, decreased mobility of the upper extremities, and pes adductus on the left side. The iliac bones were dysplastic and wide apart. The external genitalia were male with hypospadias. He did not show either excessive skinfolds at the neck or osteoporosis. Because of the dysmorphic signs, chromosomal analysis was performed, which revealed, by GTG-banding, a male karyotype with a balanced reciprocal translocation of the long arms of chromosomes 10 and 17: 46,XY,t(10;17)(q24;q23). The karyotypes of the parents were normal.

*Patient MJ* (*case 9 in tables 1 and 2*).—This female proband was the second live-born child of 30-year-old parents who were phenotypically normal, as was the elder sister. Pregnancy was normal; only weak fetal movements were reported. The child was delivered during the 40th wk of gestation, by cesarean section, because of uterine insufficiency during the second stage of labor. Birth weight was 3,930 g, length was 50 cm, head circumference was 37 cm, and chest circumference was 30 cm. Physical examination revealed a hypotonic female infant with multiple dysmorphies, which are summarized in table 1. In addition, the child displayed the following clinical signs: muscular hypotonia, dolichocephaly, short and down-slanting palpebral fissures, small nose with upturned nostrils, long and wide philtrum, short neck with excessive skinfolds, and slight contractures of elbow, hip, and knee joints. The external genitalia were unambiguously female. During the first months of life, the baby was hypotonic and required tube feeding. Because of respiratory distress, supplemental oxygen by nasal cannula was indispensable. Later, the girl developed growth retardation and was unable to walk. Despite severe deficits in motor development, her intellectual development was within normal limits; she attends a regular public school. She is now >12 years of age.

Because of the multiple anomalies found, chromosomal analysis was performed during the 3d wk of life and revealed, by GTG-banding and FISH, a female karyotype with a balanced reciprocal translocation between the long arms of chromosomes 5 and 17: 46,XX,t(5;17)(q13.3;q24.2). The karyotypes of the parents were normal.

*Patient MS* (*case 11, tables 1 and 2*).—This male proband was the second live-born child of a 28-year-old mother and a 31-year-old father. The parents were unrelated and phenotypically normal, as was the elder sister. The family history was unremarkable. The pregnancy was normal. The child was delivered during the 38th wk of gestation, by cesarean section, because of reduced fetal heart beats. Birth weight was 3,130 g, length was 48 cm, and head circumference was 37.5 cm. Perinatally, the boy displayed a severe asphyxia and therefore needed nasal intubation. He had the dysmorphic signs summarized in table 1. The child demonstrated the following additional clinical features not listed in table 1: generalized edema, flat face, small and short nose with anteverted nostrils, long and flat philtrum, coarse-shaped ears, excessive skinfolds at the neck, bell-shaped chest, short and clumsy hands, retarded bone age, and osteoporosis. The external genitalia were those of a normal male with a small penis. The boy suffered from severe respiratory distress caused by tracheomalacia due to hypoplasia of the bronchial cartilage; bilobectomy had to be performed at age 14 d, because of infantile lobar emphysema of the upper and middle lobe of the right lung. Malformations of other inner organs were not demonstrable. Later, the boy developed statomotor and psychomotor retardation. He died at age 6 years.

Chromosomal analysis was done on whole-blood cultures during the neonatal period and was repeated when the child was 2 years old. To obtain prometaphase chromosomes, the cultures were treated with methotrexate (0.05 mg/ml) after 48 h incubation; bromodeoxyuridine (20 mg/ml) and fluorodeoxyuridine (1 mg/ml) were

added $6\frac{1}{2}$ h before being harvested. RBA-, QFQ-, GTG-, and NOR-banding revealed a male karyotype with a balanced reciprocal translocation between the long arm of chromosome 17 and the short arm of chromosome 22: 46,XY,t(17;22)(q25.1;p11.2). The karyotypes of the parents were normal.

### *Somatic-Cell Hybridization*

Human-mouse and human-hamster somatic-cell hybrids, with (*a*) either skin-fibroblast or lymphoblastoid cell lines from patients with CD and (*b*) either mouse RAG cells ($Hprt^-$) or hamster A3 cells ($Tk^-$), were established as described elsewhere (Wagner et al. 1997).

### *Construction of BAC/PAC and Phage Contigs*

To isolate BACs and PACs, a variety of probes—including exon-trap fragments, STS markers generated by Généthon or the Whitehead Institute, and sequences from an existing cosmid contig (Wirth et al. 1996)—were used to screen, by PCR, the human RPCI1 3-5 DNA Pools PAC library (provided by the German Human Genome Project Resource Center, Berlin) and the CITB Human BAC DNA Pools Release II (Research Genetics). DNA from BAC and PAC clones was purified by use of the Qiagen tip-100 kit, according to the manufacturer's protocols. Clones were end-sequenced by use of T7 and Sp6 vector primers. New STS markers were derived from single-copy sequences to rescreen the libraries for overlapping clones. PCR with these markers was used on DNA from human-rodent somatic-cell hybrids containing one of the derivative chromosomes from lymphoblastoid or fibroblast cell lines of CD translocation patients, to determine the orientation of subcontigs spanning the breakpoints in these hybrids.

To generate the phage contig, CEPH Mega-YAC 946E12 was first subcloned into λGET, a phage λ-based exon-trap vector that permits conversion of the phage molecule into a multicopy plasmid by cre-lox-mediated recombination (Nehls et al. 1994). The YAC was subcloned, without prior purification from the yeast background, and recombinants with human inserts were selected by plaque hybridization with radioactively labeled human *Cot-1* DNA (Gibco BRL). The plasmid versions of 540 phages were gridded and dot-blotted onto nylon filters. Two cosmids that had been shown to span several translocation breakpoints in the 150–300-kb region *5′* to *SOX9* were used as probes to isolate phage from the YAC sublibrary. Inter-*Alu* PCR and vector-*Alu* PCR (Lengauer et al. 1992), with these subclones as templates, were used to generate probes that were hybridized back onto the sublibrary, to isolate overlapping clones.

BACs from the RPCI-11 library (BACPAC Resources, Roswell Park Cancer Institute; Pieter de Jong, personal communication) were isolated by hybridization to high-density colony filters, by use of pools of overlapping oligo probes (J. McPherson, personal communication). Contigs based on marker content were generated by hybridizing each clone identified during the library screen with the individual markers. Paths of clones for sequencing were selected on the basis of marker content and fingerprints (Marra et al. 1997).

### *PCR Typing of Microsatellite and STS Markers*

Primer sequences used either already have been published (*D17S970*, *D17S1350*, and *GATA63G01* [GenBank]; and WI-5830, WI-2860, and WI-7760 [Whitehead Institute for Biomedical Research/MIT Center for Genome Research]) or are available on request from the authors (DP and RK markers). PCR reactions were done in a total volume of 30 μl with either 10 ng of BAC/PAC DNA or 200–400 ng of genomic DNA from somatic-cell hybrids and with human and mouse cells as controls, with 5 pmol of each primer in 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1–1.5 mM $MgCl_2$, 0.25% Nonidet P-40, 0.25 mg BSA/ml, 200 μM of each dNTP, and 0.5 U of *Taq* polymerase.

### *Exon Trapping and Expression Analysis*

Exon trapping was performed with the λGET vector system, as described elsewhere (Nehls et al. 1994). The same sublibrary of 540 λGET phage clones established from YAC 946E12 for phage contig construction, as outlined above, was used. The 540 phage clones were subdivided into nine pools of 60 clones each, for electroporation into COS-7 cells. The trapped fragments from each individual pool were blunt-end cloned into *Eco*RV-cut pBluescript KS vector (Stratagene) and were sequenced from both strands, with vector primers.

For expression analysis of the trapped fragments, the following cDNA libraries were screened: human fetal brain (German Human Genome Project Resource Center, Berlin), human amniocyte (a gift of H. Leffers, Denmark), mouse total embryo 9.5 dpc, and mouse adult brain (German Human Genome Project Resource Center, Berlin). For northern blot analysis, commercial multiple-tissue northern blots from Clontech were used.

### *FISH Analysis*

BAC and PAC DNAs were labeled by nick translation (Gibco BRL Nick translation kit) with 16-biotin-dUTP (Boehringer Mannheim). Biotinylated probes were used for chromosome in situ suppression hybridizations as described by Toder et al. (1995). Biotinylated human centromere-specific probes for chromosomes 10 and 17 were purchased from Oncor and were cohybridized with the labeled BAC and PAC DNAs.

### *Pulse-Field Gel Electrophoresis (PFGE) Analysis*

The preparation of large DNA fragments from lymphoblastoid cell lines and from yeast was performed as described by Scherer and Tsui (1991). Cell numbers were adjusted such that $2 \times 10^6$ lymphoblasts or $5 \times 10^7$ yeast cells were present per agarose plug. Digestion of genomic DNA in agarose plugs with rare-cutter restriction enzymes was performed overnight in a volume of 300 μl with 6 U of enzyme/μg DNA.

For the sizing of BAC and PAC clones, the protocol of Jones and Chi (in "DNA isolation in agarose plugs for sizing and characterization of bacterial artificial chromosome (BAC) clones" [BioMedNet, Technical Tips Online]) was used. One-fifth of a plug was digested overnight with 50 U of *Not*I, to release the insert. After PFGE and Southern blotting, filters were hybridized with radioactively labeled human *Cot-1* DNA (Gibco BRL), to evaluate the size of the human inserts.

High-molecular-weight DNA fragments were separated by PFGE, as described elsewhere (Wirth et al. 1996). Yeast chromosomes (strain YP148) and phage-lambda oligomers (Biorad Laboratories) were used as size standards. A Waltzer-type electrophoresis unit was used, with a switch-time between 20 s, for the sizing of BACs and PACs, and 60 s, for the separation of digested yeast and human genomic DNA. The 1.5 % FastLane (FMC Bioproducts) agarose gel was run at 15°C with a field strength of 6 V/cm in 0.5 × Tris-acetate EDTA running buffer, for 24 h.

### *DNA Sequencing*

Small-scale DNA sequencing of the various clone DNAs was performed either with specific primers or with vector primers. For the direct sequencing of BACs and PACs, 2–3 μg of DNA were sequenced with vector primers and fluorescent dye terminator technology (BigDye™ Terminator Cycle Sequencing Ready Reaction Kit; Perkin Elmer/Applied Biosystems), under the following PCR conditions: 97°C for 5 min, followed by 35 cycles of 96°C for 1 min, 50°C for 1 min, and 60°C for 4 min. The reactions were analyzed on ABI 373 or ABI 310 sequencers. Large-scale shotgun sequencing and assembly of BACs and PACs was performed as described elsewhere (Whitehead Institute for Biomedical Research/MIT Center for Genome Research).

### *Computer-Assisted Sequence Analysis*

The programs GENSCAN and GRAIL II (Grail) were used to predict potential genes and exons in genomic sequences. Alignments of predicted peptides against various databases were made by use of the NCBI's BLAST server (BLAST, National Center for Biotechnology Information) and the programs BLASTP and TBLASTN. Genomic sequences were also analyzed in blocks of 20 kb of sequence, by use of the BLASTN program against the dbEST database and by use of the BLASTX program against a nonredundant version of several protein databases.

## Results

### *Construction of a BAC/PAC Contig, Rare-Cutter Map, and Sequence Analysis*

To more precisely map the positions of three previously published CD translocation breakpoints (Wagner et al. 1994; Wirth et al. 1996), as well as those of three new cases, we set out to extend an existing cosmid contig that ended 130 kb upstream of the *SOX9* gene. First, the nonchimeric YAC 946E12 of ~2 Mb, spanning the *SOX9* gene and several CD translocation breakpoints (Wirth et al. 1996), was subcloned into the phage-vector λGET (Nehls et al. 1994). A total of 540 phage clones from this sublibrary were then used to assemble a 120-kb phage contig extending from STS marker DP 2 to STS marker DP 8 and spanning three translocation breakpoints (fig. 1*A*, not shown in detail).

As starting points for the construction of a BAC/PAC contig, we used the markers DP 2 (fig. 1*A*, end of cosmid contig as described by Wirth et al. [1996]), *D17S970, D17S1350,* WI-7760, and the exon-trap fragment 38.14 (see the Exon Trapping subsection, below). Contig building was performed by end sequencing of newly identified clones, to establish new STS markers for further rounds of screening. The resulting contig (fig. 1*A*) extends as much as ~1,200 kb proximal to *SOX9*.

In parallel, we used all the STS markers as hybridization probes, to establish a long-range restriction map on the basis of the YAC 946E12. Because YACs are known to frequently delete variable parts of their inserts (Nehls et al. 1995), we compared the restriction fragments obtained with five different rare-cutter enzymes after digestion of the YAC DNA versus those of genomic DNA from several human lymphoblastoid cell lines, probing with different markers from the contig. An *Sfi*I fragment of 940 kb in both YAC and lymphoblastoid-cell DNA could be detected (fig. 2), suggesting that the YAC faithfully represents the genomic sequence in the contig region. (The *Sfi*I site, which delimits the telomeric end of this 940-kb fragment, is located 5 kb 5′ of the *SOX9* gene and was considered to be the 0-kb coordinate in fig. 1*B*). Complete methylation of the CpG dinucleotides present in the recognition sites of *Bss*HII, *Mlu*I, and *Eag*I was observed in DNA from lymphoblastoid cells (not shown), whereas *Nru*I showed partial methylation of internal *Nru*I sites, resulting in a 650-kb fragment (Wirth et al. 1996). The corresponding sites

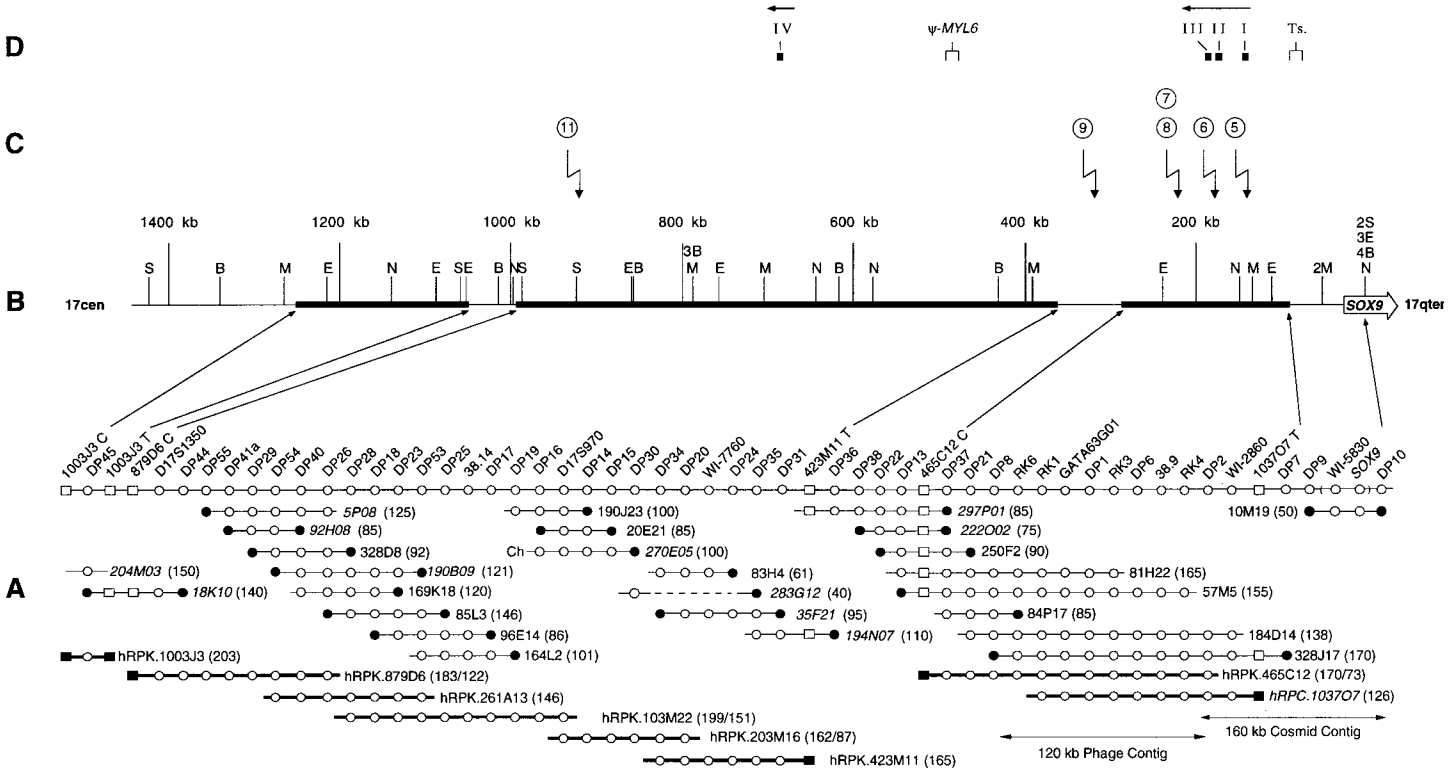


**Figure 1** Map of 1.4-Mb region proximal to *SOX9*. *A*, BAC, PAC, cosmid, and phage contigs proximal to the *SOX9* gene. The extent of the cosmid (Wirth et al. 1996) and phage contigs is shown by a double arrow. BAC and PAC clones are drawn as horizontal lines (not to scale), with the clone identifiers given, on the right side, in regular print and italics, respectively. The insert sizes of the clones are given in parentheses (in kb). Markers present on each clone are represented by circles, with blackened circles representing end sequence–derived STS markers used for subsequent library screening. Note that PAC 283G12 has an internal deletion, as indicated by the dashed line. The order of all the markers present in the clone contig is shown above, with each circle representing one marker. The exact order is not known for the markers *SOX9* and WI-5830, given in parentheses. Clones with the prefix "hRPK or "hRPC" were chosen for sequencing and are indicated by the thicker lines. The ends of sequence contigs are shown as blackened squares on the clone from which they derive and as unblackened squares on the clones on which they are present. Submitted sequences from overlapping clones have been trimmed to a 2-kb overlap; the corresponding clones thus show two numbers in parentheses, the first referring to the insert size and the second referring to the amount of sequence, in the database, derived from this clone. The GenBank accession numbers for the sequenced clones are AC005181 (hRPK.1003J3), AC005273 (hRPK.879D6), AC005138 (hRPK.261A13), AC005279 (hRPK.103M22), AC006473 (hRPK.203M16), AC006448 (hRPK.423M11), AC005144 (hRPK.465C12), and AC005152 (hRPC.1037O7). Ch = chimeric end. *B*, Positions of rare-cutter restriction-enzyme sites (drawn to scale). B = *Bss*HII; E = *Eag*I; M = *Mlu*I; N = *Nru*I; S = *Sfi*I. Sites present within a few kilobases, as in the *SOX9* CpG island, are shown together, with the frequency of occurrence of restriction-enzyme sites being represented by the numbers preceding the enzyme abbreviations. The accuracy of the position of the rare-cutter sites is estimated to be within ±10 kb for the sites not present in the sequenced regions, reflecting the resolution of PFGE. The direction of transcription of the 5.4-kb *SOX9* gene is given by the outlined arrow. The thicker lines represent the regions that have been completely sequenced. *C*, Position of translocation breakpoints. The circled numbers correspond to the case numbers shown in table 2, with the zigzag arrows pointing to the middle of the interval where the breakpoints have been localized. The size of the interval as well as the karyotype of the cases is given in table 2. *D*, Position of trapped exons, putative genes, and pseudogenes. "Ts." denotes the cDNA described by Ninomiya et al. (1996), which was mapped 74–88 kb upstream of *SOX9*. "I"–"IV" denote positions of exon-trap fragments: I = ET 38.9 (142 kb), II = ET 82.12 (172 kb), III = ET 82.14 (185 kb), and IV = ET 38.14 (686 kb); for details of the putative exons, see figure 4. The arrows above the trapped fragments indicate the orientation of the flanking splice-donor and -acceptor sites. Ψ-*MYL6* is a *MYL6* pseudogene published by Lenz et al. (1989) and located 457 kb centromeric to *SOX9*.

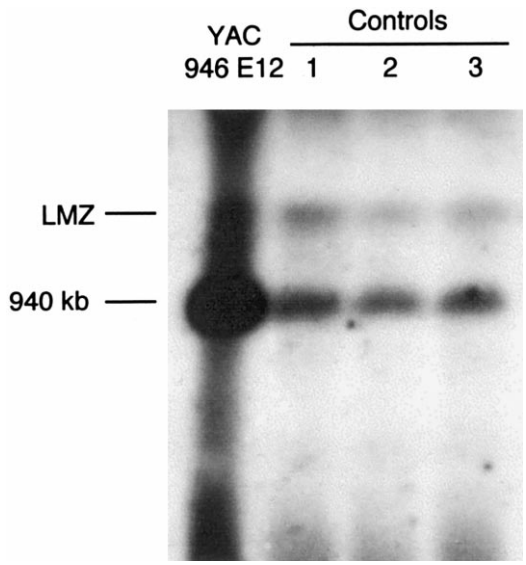


**Figure 2** PFGE analysis of YAC 946E12 and genomic controls. Autoradiograph of a Southern blot from a pulse-field gel with *Sfi*I-digested DNA from YAC 946E12 and three lymphoblastoid cell lines (controls 1–3), hybridized with a pool of STS fragments between DP 29 and *SOX9* (fig. 1*A*). Note that the genomic 940-kb *Sfi*I fragment seen in the controls is also present in the YAC 946E12. LMZ = limiting mobility zone. The gel was run at 6 V/cm, switch time 60 s, for 24 h.

were all readily cut in the YAC DNA, because of the lack of methylation in yeast.

The long-range map established from YAC 946E12 by use of the five rare-cutter enzymes used singly and in combination is shown in figure 1*B*. The map extends from a cluster of rare-cutter sites around the *SOX9* gene to ~1,400 kb in the centromeric direction. In total, 50 STS markers are spread over a distance of 1,200 kb, resulting in a mean marker density of 1 marker/24 kb. All markers hybridized to fragments on the YAC rare-cutter filters, providing further evidence for the integrity of the YAC in the region studied.

The STS markers derived during contig building were used to isolate additional BACs and PACs, to establish an optimal path for genomic sequencing. The eight BACs or PACs sequenced to date are indicated by thicker lines in figure 1*A*, with squares denoting the centromeric or telomeric ends of the individual sequence contigs. Specifically, the following regions, indicated by the thicker black line in fig. 1*B*, have been completely sequenced: 88–285 kb, 332–995 kb, and 1,050–1,253 kb, all proximal to *SOX9*. These 1,063 kb of sequence unequivocally define the order of the STS markers and refine the rare-cutter map established by PFGE.

### *Fine-Mapping of CD Translocation Breakpoints*

In addition to the three CD translocation cases for which we had previously mapped the breakpoints to >130 kb upstream of *SOX9* (Wirth et al. 1996; cases 5–7 in table 2), two new de novo translocation cases (cases 8 and 9 in table 1; also see the Case Reports subsection, above) with many features characteristic of CD were made available to us. We also reanalyzed a t(17;22) translocation case that we had published in the context of generating a somatic cell–hybrid panel for distal 17q but which, at the time, we had not recognized as being a likely CD case (Wagner et al. 1997). After additional, more-detailed clinical data were obtained, it became apparent that this patient (case 11 in table 1; also see the Case Reports subsection, above) had many features typical of CD. We sequenced the *SOX9* ORF, as well as the four exon/intron junctions, in this and the two new cases and did not find any mutation.

To locate the breakpoints in these six CD translocation cases, we isolated the derivative autosome of each case from the der(17) and the normal chromosome 17, in human/rodent somatic-cell hybrids. DNA from these hybrids was then used for typing of STS markers from our contig. Figure 1*C* shows the position of the various breakpoints thus determined along the chromosome, with the distances from *SOX9* being listed in table 2. As can be seen, the breakpoints of cases 5, 6, and 7 mapped, respectively, 134–142 kb (between RK 4 and 38.9), 173–179 kb (between RK 3 and DP 1), and 212–224 kb (between RK 1 and RK 6) centromeric to *SOX9*. With positions at 228–229 kb (between DP 8 and DP 21) and 288–319 kb (between DP 13 and DP 22), the breakpoints of the two new cases, 8 and 9, are also a considerable distance upstream of the *SOX9* coding region. Remarkably, the breakpoint of case 11 mapped to the interval between DP 29 and DP 41a, ~890–950 kb upstream of *SOX9*. DNA from the somatic-cell hybrid with the der(22) of this case typed positive for all STS markers from *SOX9*, up to and including DP 29. The single exception was marker DP 21, which was negative in PCR but was shown to be present by Southern blot hybridization, which detected the same fragment in both genomic DNA and in DNA from the hybrid (not shown). The negative PCR result is therefore most likely caused by a sequence polymorphism affecting a primer-binding site.

### *FISH Analysis of CD Translocation Cases*

The STS content mapping on the somatic-cell hybrids showed that the breakpoints are scattered over a considerable distance upstream of *SOX9*. To rule out the presence of cloning artifacts during the preparation of the hybrids, which could have resulted in erroneous PCR results, we used BACs and PACs that were expected to span the breakpoints as probes for FISH on metaphase spreads from the patients' cell lines. Figure 3 shows the results for the three most distant cases—8, 9, and

## Table 2

**CD Translocation and Inversion Cases**

| Case | Karyotype | Sex[a] | Campomelia[b] | Survival Time | Distance to *SOX9* (kb) | Reference(s) |
|---|---|---|---|---|---|---|
| 1 | 46,XY,t(7;17)(q34;q25.1) | f* | Severe | >6 years | 50 | Tommerup et al. (1993); Wagner et al. (1994) |
| 2 | 46,XY,t(12;17)(q21.32;q24.3-q25.1) | f* | None | 11 mo | 74–88[c] | Ninomiya et al. (1995, 1996) |
| 3 | 46,XY,t(2;17)(q35;q23-q24) | f* | Severe | Abortion | 88 | Young et al. (1992); Foster et al. (1994) |
| 4 | 46,XY,t(9;17) | m | Severe | >3 years | 110–140 | Wunderle et al. (1998) |
| 5 | 46,XY,t(13;17)(q22;q25.1) | f* | None | >30 years | 134–142[c] | Tommerup et al. (1993); Wirth et al. (1996); present study |
| 6 | 46,XX,t(1;17)(q42.13;q24.3-q25.1) | f | Severe | >6 years | 173–179[c] | Tommerup et al. (1993); Wirth et al. (1996); present study |
| 7 | 46,XY,t(6;17)(q14;q24) | f* | Mild | >3 years | 212–224[c] | Wirth et al. (1996); present study |
| 8 | 46,XY,t(10;17)(q24;q23) | m | Severe | >1 year | 228–229[c] | Present study |
| 9 | 46,XX,t(5;17)(q13.3;q24.2) | f | Severe | >12 years | 288–319[c] | Present study |
| 10 | 46,XY,inv(17)(q11.2;q24.3-q25.1) | f* | Mild | >2 years | 70–350 | Mansour et al. (1995); Wunderle et al (1998) |
| 11 | 46,XY,t(17;22)(q25.1;p11.2) | m | None | 6 years | 890–950[c] | Wagner et al. (1997); present study |
| 12 | 46,XX,inv(17)(q12;q25) | f | Severe | 3 mo | ? | Maraia et al. (1991) |
| 13 | 46,XX,t(4;17)(q21.3;q23.3) | f | Mild/none | >2 years | ? | Mansour et al. (1995) |
| 14 | 46,XX,t(5;17)(q15;q25.1) | f | Mild | 11 d | ? | Savarirayan and Bankier (1998) |

[a] f = XX female; f* = sex-reversed XY female; m = XY male.

[b] Severe = overt bending of long bones; Mild = only mild bowing of long bones; None = straight long bones.

[c] Accurate within ±10 kb—for the absolute position of the interval relative to *SOX9*, not for the size of the interval itself (see legend to fig. 1*B*).

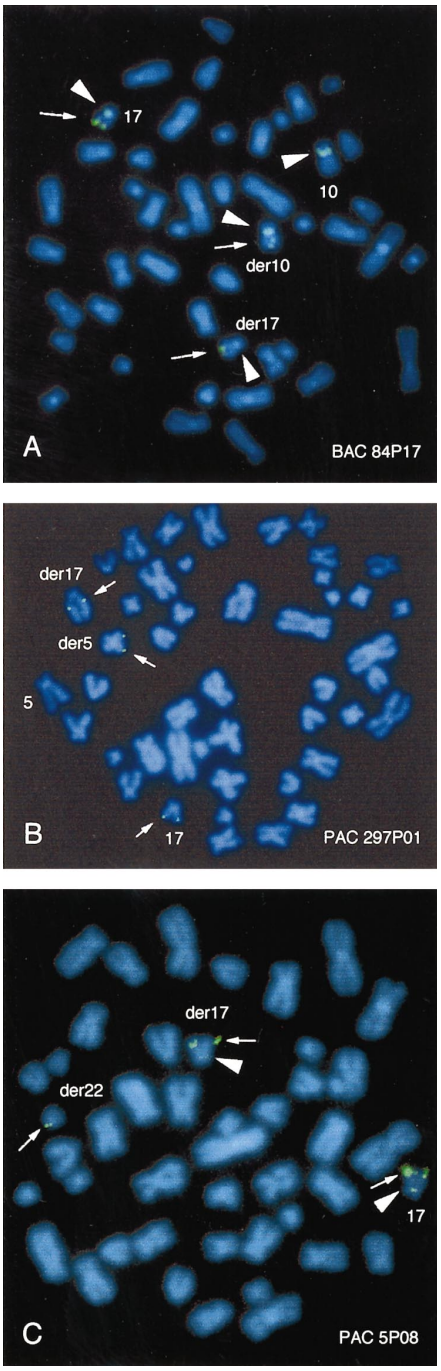


**Figure 3** FISH with BACs and PACs on metaphase spreads of CD translocation cases. *A*, t(10;17) translocation (case 8; table 2) and BAC 84P17 as probe (fig. 1*A*), showing overlapping hybridization signals on der(10) and der (17). *B*, t(5;17) translocation (case 9; table 2) and PAC 297P01as probe (fig. 1*A*), showing overlapping signals on der(5) and der(17). *C*, t(17;22) translocation (case 11; table 2) and PAC 5P08 as probe (fig. 1*A*), showing overlapping signals on der(17) and der(22). Centromere-specific probes for chromosomes 17 (*A* and *C*) and 10 (*A*) were used. Centromeres of the chromosomes are indicated by large arrowheads (*A* and C), and signals from the BAC and PAC clones are indicated by small arrows.

11—with the clones 84P17 (fig. 3*A*), 297P01 (fig. 3*B*), and 5P08 (fig. 3C), respectively. In all three cases, signals could be seen on the normal chromosome 17, as well as on the two derivative chromosomes, indicating that all clones span the corresponding breakpoints on the patients' chromosomes. The same kind of analysis was also done with CD translocation cases 5–7, with clones from the 120-kb phage contig, again confirming the results from the somatic cell–hybrid analysis (data not shown).

### *Analysis of Genomic Sequences*

The 1,063 kb of genomic sequence established so far (indicated by the thicker lines in fig. 1*B*) were searched for potential genes and pseudogenes, by analysis of the sequence by use of the GENSCAN gene-prediction program (Burge and Karlin 1997) and by comparison of the genomic sequence against the dbEST database by use of the BLASTN program (Altschul et al. 1990). The peptides predicted by GENSCAN were compared against a nonredundant version of several protein databases, by the BLASTP program. In addition, we used the TBLASTN program, to compare the peptides against a six-frame translation of the dbEST database. Neither strategy revealed any significant matches to known genes in the 1,063 kb of sequence analyzed.

### *Mapping of Putative Genes and Pseudogenes*

Of all translocation breakpoints mapped so far (table 2), only one has been shown to be in close proximity to a gene or pseudogene. Ninomiya et al. (1996) isolated a testis cDNA with a genomic fragment that contained the breakpoint from a t(12;17) translocation case (case 2 in table 2). We mapped this transcript—and, consequently, the breakpoint—to 74–88 kb proximal to *SOX9* (data not shown). The genomic sequence of this potential gene is colinear with the published cDNA sequence, as determined by PCR with overlapping primer pairs. This transcript could be derived from a pseudogene, since it contains an *Alu* element at its 5′ end and has no match to entries in the database, no introns, and no obvious coding potential.

One of the markers used for contig building was the EST-derived marker WI-7760, also known as "*D17S1993*." This marker, derived from the 3′ UTR of the alkali myosin light-chain 6 gene (*MYL6*), has been mapped to both chromosome 12 and chromosome 17, by radiation-hybrid and STS content mapping (Genome Database accession numbers 9767349, 1075901, and 9206886). Analysis of the genomic sequence locates this marker to 457 kb proximal to *SOX9*, as part of a *MYL6* pseudogene, a partial sequence of which has been described elsewhere (Lenz et al. 1989).

### *Exon Trapping*

For exon trapping, we subcloned YAC 946E12 into λGET, a phage vector designed for large-scale exon trapping (Nehls et al. 1994). A total of 540 phages with a mean insert length of 14 kb of human DNA were analyzed, in nine pools. In addition, two pools of 20 clones each, representing the phage contig shown in figure 1*A*, were analyzed. Only five different fragments were recovered (see fig. 1*D*, I–IV; fragment V is located proximal to the contig). Whereas this was not an exhaustive analysis, because the redundancy of this sublibrary was only approximately threefold, the small number of trapped fragments is likely to reflect a low gene density in this region.

Sequencing of the flanking regions of the exon-trap fragments revealed canonical splice-donor and -acceptor sites (see fig. 4). Only exon-trap fragments ET 38.9 and ET 82.14 have an ORF, whereas ET 82.12, ET 38.14, and ET 38.15 contain several stop codons that close all three reading frames. By PCR, ET 38.15 was shown to be present on YAC 946E12 and in DNA from a chromosome 17–only hybrid (Wagner et al. 1997), but absent from the hybrids of the CD translocations, placing it proximal to the contig shown in figure 1. Interestingly, the fragments ET 38.9, ET 82.12, ET 82.14, and ET 38.14 (I-IV in fig. 1*D*) all have the same orientation regarding their splice-donor and -acceptor sites, with their inferred transcriptional orientation in the opposite direction to that of the *SOX9* gene.

The trapped fragments were hybridized as a pool to cDNA libraries prepared from human fetal brain, human amniocytes, mouse 9-day total embryo, and mouse adult brain. No positive clones were detected. Northern blots probed with the same pool of trapped fragments did not reveal any signal in various human fetal and adult tissues. Furthermore, database searches with the sequences of the trapped fragments showed no matches to known genes or transcripts, and none of the four trapped exons I–IV were predicted by either the GENSCAN or the Grail exon-prediction programs.

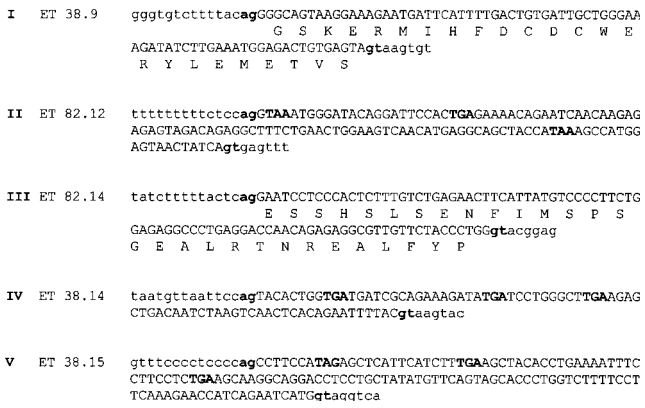


**Figure 4** Exon fragments trapped from YAC 946E12. "I"–"IV" denote the exon-trap fragments shown in figure 1*D*. Putative intron sequences are given in lowercase, exon sequences in uppercase; the invariant ag and gt nucleotides at the splice-acceptor and -donor sites, respectively, are in boldface, as are the stop codons that close all reading frames in ET 82.12, ET 38.14, and ET 38.15. The conceptual translation of the ORFs from ET 38.9 and ET 82.14 is given below the corresponding codons.

## Discussion

Up to now, 14 CD cases with translocations or inversions involving distal 17q have been described, including the three new cases from this study. Table 1 summarizes the clinical and radiological features of these new cases. Patients number 8 and 9 show most of the clinical manifestations typically seen in patients with CD (Mansour et al. 1995). Both are still alive at >1 and 12 years of age, respectively (table 2). The third patient with the t(17;22) translocation died at the age of 6 years. This reflects the tendency of patients with CD translocation to have a longer life expectancy than patients with mutations in the *SOX9* coding region (Mansour et al. 1995; Wirth et al. 1996). Excluding the abortion case, 9 of the 13 patients with translocation or inversion survived for >2 years, and one patient (case 8) is just <2 years of age (table 2). In contrast, only 1 of 28 patients with CD without a chromosomal rearrangement survived >2 years (when terminated pregnancies are excluded), in a study by Mansour et al. (1995). Likewise, of 22 patients with CD with documented *SOX9* coding-region mutations and known survival status, only 5 survived >2 years (when terminated pregnancies are excluded) (Wagner et al. 1994; Kwok et al. 1995; Cameron et al. 1996; Meyer et al. 1997; Goji et al. 1998; Hageman et al. 1998). Three of these five long-term survivors had the same Y440X stop-codon mutation (Wagner et al. 1994; Meyer et al. 1997; Hageman et al. 1998), with the resulting truncated SOX9 protein retaining some transactivation potential (Meyer et al. 1997).

Another difference between the two classes of CD mutations is the presence or absence of the eponymous feature, campomelia. In contrast to the 22 patients with *SOX9* ORF mutations who have been mentioned in the preceding paragraph, only 1 of whom showed acampomelic CD (Meyer et al. 1997), 7 of the 14 cases with rearrangements outside the *SOX9* gene show only mild bowing or a complete lack of bending or bowing of the upper or lower limbs (table 2). Likewise, in the study by Mansour et al. (1995), only 2 of 36 cases showed no or only mild campomelia, and these 2 (cases 10 and 13 in table 2) were patients with chromosomal rearrangements outside the coding region.

When the positions of the various breakpoints listed in table 2 are compared, it is obvious that there is no apparent correlation between the position of a breakpoint and the phenotype of the patient. There are two patients with breakpoints 110–142 kb upstream of *SOX9,* one showing sex reversal and no bending of the long bones (acampomelia; case 5) and the other presenting with severe campomelia but without sex reversal (case 4). Likewise, the survival time of the patients cannot be correlated with the positions of the breakpoints, since the patient with CD who has the breakpoint closest to *SOX9* (case 1 in table 2) is still alive, whereas a patient with a similar breakpoint (case 2 in table 2) died at age 11 mo. This lack of a genotype-phenotype correlation parallels our previous results from the mutation analysis of patients with CD with mutations in the coding region of *SOX9* (Meyer et al. 1997).

Table 2 summarizes our results and the data from the literature regarding the distances of the various breakpoints from the *SOX9* gene. They all lie at considerable distance proximal to *SOX9,* with the majority of the cases having their breakpoints within an interval 50–350 kb upstream of the gene (also see fig. 1). The only exception is case 11, which has its breakpoint as far as 890–950 kb proximal to *SOX9*. This enormous distance between a breakpoint and its corresponding gene is not without precedent. de Kok et al. (1996) described a deletion hot-spot region 900 kb proximal to the *POU3F4* gene in Xq21.1, where several deletions overlap in an 8-kb region in patients with X-linked deafness type 3 (DFN3). Several other examples, in human and mouse, have been described in which chromosomal rearrangements exert position effects on somewhat distant genes (Bedell et al. 1996). For example, translocation breakpoints have been mapped ⩾85 kb 3′ to the *PAX6* gene in human aniridia (Fantes et al. 1995). Likewise, at the murine *Steel* locus, chromosomal rearrangements in two mutant alleles with breakpoints 115 kb and 195 kb proximal to the *Mgf* gene have been shown to alter the expression level of the *Mgf* gene (Bedell et al. 1995).

Different mutational mechanisms that could account for the CD phenotype shown by the translocation cases have been discussed elsewhere (Wirth et al. 1996). One possibility is the disruption of alternative transcripts from one or more upstream promotors. The existence of a cartilage-specific and/or gonadal-specific promoter still cannot be ruled out, because the *SOX9* transcriptional start site defined by us was found in fetal brain tissue (Wagner et al. 1994). However, the fact that a translocation breakpoint >890 kb from *SOX9* leads to a CD phenotype would imply an intron of unprecedented length.

A second possibility is a classic position effect causing altered *SOX9* expression by translocating the gene into a heterochromatic environment. The only case in which such a mechanism could be operating is case 11, with its breakpoint within band 22p11.2 (table 2), placing the *SOX9* gene rather close to the centromere and <1 Mb from a chromosomal region that is generally assumed to consist of highly repetitive sequences.

Third, the breakpoints could disrupt a second CD-causing gene on 17q. From the analysis of the 1,063 kb of sequence available, it does not seem that the chromosomal rearrangements seen in our cases disrupt any protein-coding gene. In addition, we failed to detect expression of the fragments recovered by exon trapping, nor did we find matches in the dbEST database. However, because *SOX9* is also expressed early in development, we may have missed the appropriate tissue and/or time of expression. Interestingly, the four fragments have the same putative transcriptional orientation and could derive from the same noncoding transcriptional unit. Noncoding RNAs such as the human *XIST* transcript have been described (Brown et al. 1992) that exert their function at the level of the RNA molecule. The transcript described by Ninomiya et al. (1996) is not very likely to represent such a regulatory RNA, because, in the majority of the CD translocation cases, it is not separated from *SOX9*. Nevertheless, several transcription units present in the region could act together to open up the chromatin and to allow transcription of *SOX9*. Such a mechanism has been proposed by Sutherland et al. (1996), who found noncoding transcripts disrupted by a translocation event in a patient with DiGeorge syndrome.

The fourth and most likely mutational mechanism is the removal of one or more *cis*-regulatory elements from an extended *SOX9* control region. Support for this notion comes from recent work by Wunderle et al. (1998), which describes mice transgenic for truncated versions of the human YAC 946E12 carrying a *SOX9-lacZ* fusion gene. Comparing the expression of the *SOX9-lacZ* gene versus endogenous *Sox9,* they observed highly similar expression patterns during skeletal development when the transgenic YAC contained 350 kb of *SOX9* upstream sequence but a drastic reduction of expression in all tissues except the neuroectoderm when the transgenic YAC contained only 75 kb of *SOX9* 5′-flanking sequence. Neither YAC transgene showed expression in the developing gonads. These data strongly argue for the presence of multiple *cis*-acting elements scattered over a region ⩾350 kb upstream of *SOX9*. Given the incomplete expression pattern of the larger construct, one has to assume either species-specific elements not working in the mouse or the presence of additional elements >350 kb from *SOX9*. Such additional regulatory elements could perhaps be >890 kb from *SOX9,* as indicated by the t(17;22) breakpoint, unless we invoke the heterochromatin effect discussed above. In this context, it must be stressed that, although extensive STS content map-

ping and PFGE analysis of the isolated der(22) chromosome provided no evidence that a deletion between the breakpoint and *SOX9* may have occurred concomitantly with the translocation, a small deletion or an inversion cannot entirely be ruled out by these data. This caveat also applies to the other translocation cases.

The assumption that the translocations disrupt *cis*-acting regulation of *SOX9* implies an altered expression of the *SOX9* allele on the translocation chromosome. By exploiting a *SOX9* coding region–sequence polymorphism, we have had previously found (and reported elsewhere [Wirth et al. 1996]) a similar expression level for both *SOX9* alleles in the t(13;17) translocation case. This result, however, was obtained from a low-expression lymphoblastoid cell line and thus may reflect basal promoter activity rather than the regulated expression of *SOX9* during embryonic development (Wright et al. 1995). *SOX9* expression in embryonic tissue from CD translocation cases could not be analyzed, since such tissue was not available to us.

The most accurate cytogenetic measurements place the CD translocation breakpoints at the border between the G band, at 17q24.3, and the R band, at 17q25.1 (Tommerup et al. 1993). G bands are known to be generally AT rich and gene poor, whereas R bands have a higher GC content and a higher gene density (Bernardi 1995). We note that the GC content of the three sequence contigs in the 1.3-Mb region 5′ to *SOX9* (thicker lines in fig. 1*B*) slightly increases from the centromeric side: from 37.5% to 37.7% to 40.3%, finally reaching 52.0% in an 8-kb-sequence contig that includes the 5.4-kb *SOX9* gene (G.S., unpublished data). The present data indicate that, on the basis of its GC content, the region 5′ to *SOX9* belongs to the GC-poor isochores of the L1 and L2 families that are mainly present in G bands but also in R bands, whereas the *SOX9* gene and surrounding sequences belong to the GC-rich H2 isochore found exclusively in R bands (Bernardi 1995). The picture thus emerging from the cytogenetic and sequence data is that *SOX9* may in fact be located at the transition zone between a G band and an R band. This makes it partly understandable why the region centromeric to *SOX9* is devoid of protein-coding genes—and why a small gene such as *SOX9* may have the luxury of distributing its regulatory elements over such an enormous distance. Interestingly, our recent data show that the 5′-flanking region of the *SOX9* orthologue in the pufferfish *Fugu rubripes* extends over ~65 kb (Pfeifer et al. 1998), an unusually large intergenic region for the otherwise compact *Fugu* genome. This observation could indicate the need for an extended *SOX9* control region in vertebrates in general. Once the sequencing of both the region proximal to *SOX9* in the human and the corresponding region in the mouse, a project that is now under way, is completed, we will be able to compare these sequences versus the more distantly related *Fugu* sequence, to identify conserved elements that can subsequently be tested in transgenic mouse experiments, for their potential to drive the expression of a reporter gene toward the right tissue and time point during development.

## Acknowledgments

We wish to thank the Resource Center/Primary Database (RZPD) of the German Human Genome Project, Berlin, for providing us with PAC and cDNA libraries; Oncor Appligene, for partial funding of the publication costs; Henrik Leffers, for the human amniocyte cDNA library; Werner Schempp, Birgitta Gläser, and Roland Toder, for assistance in FISH analysis and for helpful discussions; Jobst Meyer and Thomas Wagner, for cosmid clones; Jürgen Zimmer, for expert technical assistance in cell-culture work; Veronique Wunderle, for contributing patient data; Friedhelm Heitmann and Ulrike Kriwet-Barz, for contributing case YS and clinical data; Pawel Stankiewicz, for making us aware of case MJ and for establishing contacts; Deborah Morris-Rosendahl, for helpful comments on the manuscript, and all members of the Whitehead/MIT CGR sequencing group. L.K. expresses his thanks to Elzbieta Chmarzynska, for karyotyping case MJ. D.P. would like to thank Diana Pfeifer, for her support, and the Institute of Virology, University of Freiburg, for access to their ABI 373. This work was supported by Deutsche Forschungsgemeinschaft grants Sche 194/11-1 and /11-2 (to G.S.) and National Institutes of Health grants (to E.S.L.).

## Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

BACPAC Resources, Roswell Park Cancer Institute, http://bacpac.med.buffalo.edu (for BACs)

BLAST, National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/BLAST (for BLASTN, BLASTP, BLASTX, and TBLASTN programs)

GenBank, http://www.ncbi.nlm.nih.gov/Web/Genbank/index.html (for markers *D17S970, D17S1350,* and *GATA63G01* and for sequenced clones hRPK.1003J3 [AC005181], hRPK.879D6 [AC005273], hRPK.261A13 [AC005138], hRPK.103M22 [AC005279], hRPK.203M16 [AC006473], hRPK.423M11 [AC006448], hRPK.465C12 [AC005144], and hRPC.1037O7 [AC005152])

Genome Database, The, http://www.gdb.org (for WI-7760 [9767349, 1075901, and 9206886])

GENSCAN, http://CCR-081.mit.edu/GENSCAN.html

Grail, http://avalon.epm.ornl.gov/Grail-bin/EmptyGrailForm (for GRAIL II program)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for CD [MIM 114290])

Technical Tips Online, BioMedNet, http://www.biomednet.com/db/tto

Whitehead Institute for Biomedical Research/MIT Center for

Genome Research, http://www-genome.wi.mit.edu (for markers WI-2860, WI-5830, and WI-7760)

## References


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

Bedell MA, Brannan CI, Evans EP, Copeland NG, Jenkins NA, Donovan PJ (1995) DNA rearrangements located over 100 kb 5′ of the *Steel* (*Sl*)-coding region in *Steel-panda* and *Steel-contrasted* mice deregulate *Sl* expression and cause female sterility by disrupting ovarian follicle development. Genes Dev 9:455–470

Bedell MA, Jenkins NA, Copeland NG (1996) Good genes in bad neighbourhoods. Nat Genet 12:229–232

Bell DM, Leung KKH, Wheatley SC, Ng LJ, Zhou S, Ling KW, Sham MH, et al (1997) SOX9 directly regulates the type-II collagen gene. Nat Genet 16:174–178

Bernardi G (1995) The human genome: organization and evolutionary history. Annu Rev Genet 29:445–476

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF (1992) The human *XIST* gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 71:527–542

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78–94

Cameron FJ, Hageman RM, Cooke-Yarborough C, Kwok C, Goodwin LL, Sillence DO, Sinclair AH (1996) A novel germ line mutation in *SOX9* causes familial campomelic dysplasia and sex reversal. Hum Mol Genet 5:1625–1630

de Kok YJM, Vossenaar ER, Cremers CWRJ, Dahl N, Laporte J, Hu LJ, Lacombe D, et al (1996) Identification of a hot spot for microdeletions in patients with X- linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene *POU3F4*. Hum Mol Genet 5:1229–1235

de Santa Barbara P, Bonneaud N, Boizet B, Desclozeaux M, Moniot B, Südbeck P, Scherer G, et al (1998) Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Müllerian hormone gene. Mol Cell Biol 18:6653–6665

Fantes J, Redeker B, Breen M, Boyle S, Brown J, Fletcher J, Jones S, et al (1995) Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. Hum Mol Genet 4:415–422

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

Goji K, Nishijima E, Tsugawa C, Nishio H, Pokharel RK, Matsuo M (1998) Novel missense mutation in the HMG box of *SOX9* gene in a Japanese XY male resulted in campomelic dysplasia and severe defect in masculinization. Hum Mutat Suppl 1:S114–S116

Grosschedl R, Giese K, Pagel J (1994) HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. Trends Genet 10:94–100

Hageman RM, Cameron FJ, Sinclair AH (1998) Mutation analysis of the *SOX9* gene in a patient with campomelic dysplasia. Hum Mutat Suppl 1:S112–S113

Houston CS, Opitz JM, Spranger JW, Macpherson RI, Reed MH, Gilbert EF, Herrmann J, et al (1983) The campomelic syndrome: review, report of 17 cases, and follow-up on the currently 17-year-old boy first reported by Maroteaux et al in 1971. Am J Med Genet 15:3–28

Kent J, Wheatley SC, Andrews JE, Sinclair AH, Koopman P (1996) A male-specific role for *SOX9* in vertebrate sex determination. Development 122:2813–2822

Kwok C, Weller PA, Guioli S, Foster JW, Mansour S, Zuffardi O, Punnett HH, et al (1995) Mutations in *SOX9*, the gene responsible for campomelic dysplasia and autosomal sex reversal. Am J Hum Genet 57:1028–1036

Lefebvre V, Huang W, Harley VR, Goodfellow PN, de Crombrugghe B (1997) SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. Mol Cell Biol 17:2336–2346

Lengauer C, Green ED, Cremer T (1992) Fluorescence in situ hybridization of YAC clones after Alu-PCR amplification. Genomics 13:826–828

Lenz S, Lohse P, Seidel U, Arnold HH (1989) The alkali light chains of human smooth and nonmuscle myosins are encoded by a single gene: tissue-specific expression by alternative splicing pathways. J Biol Chem 264:9009–9015

Mansour S, Hall CM, Pembrey ME, Young ID (1995) A clinical and genetic study of campomelic dysplasia. J Med Genet 32:415–420

Maraia R, Saal HM, Wangsa D (1991) A chromosome 17q *de novo* paracentric inversion in a patient with campomelic dysplasia: case report and etiologic hypothesis. Clin Genet 39:401–408

Marra MA, Kucuba T, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, et al (1997) High throughput fingerprinting analysis of large-insert clones. Genome Res 7: 1072–1084

Meyer J, Südbeck P, Held M, Wagner T, Schmitz ML, Bricarelli FD, Eggermont E, et al (1997) Mutational analysis of the *SOX9* gene in campomelic dysplasia and autosomal sex reversal: lack of genotype/phenotype correlations. Hum Mol Genet 6:91–98

Morais da Silva S, Hacker A, Harley V, Goodfellow P, Swain A, Lovell-Badge R (1996) *Sox9* expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds. Nat Genet 14:62–68

Nehls M, Lüno K, Schorpp M, Pfeifer D, Krause S, Matysiak-Scholze U, Dierbach H, et al (1995) YAC/P1 contigs defining the location of 56 microsatellite markers and several genes across a 3.4-cM interval on mouse chromosome 11. Mamm Genome 6:321–331

Nehls M, Pfeifer D, Boehm T (1994) Exon amplification from complete libraries of genomic DNA using a novel phage vector with automatic plasmid excision facility: application to the mouse neurofibromatosis-1 locus. Oncogene 9: 2169–2175

Ng LJ, Wheatley S, Muscat GEO, Conway-Campbell J, Bowles J, Wright E, Bell DM, et al (1997) SOX9 binds DNA, activates transcription, and coexpresses with type II collagen during chondrogenesis in the mouse. Dev Biol 183:108–121

Ninomiya S, Isomura M, Narahara K, Seino Y, Nakamura Y (1996) Isolation of a testis-specific cDNA on chromosome 17q from a region adjacent to the breakpoint of t(12;17)

observed in a patient with acampomelic campomelic dysplasia and sex reversal. Hum Mol Genet 5:69–72

Ninomiya S, Narahara K, Tsuji K, Yokoyama Y, Ito S, Seino Y (1995) Acampomelic campomelic syndrome and sex reversal associated with de novo t(12;17) translocation. Am J Med Genet 56:31–34

Pevny LH, Lovell-Badge R (1997) *Sox* genes find their feet. Curr Opin Genet Dev 7:338–344

Pfeifer D, Bagheri-Fam S, Scherer G (1998) Campomelic dysplasia and *SOX9:* comparative analysis of the *SOX9* region in *Fugu rubripes*. Paper presented at the 48th annual meeting of The American Society of Human Genetics. Denver, October 27–31

Savarirayan R, Bankier A (1998) Acampomelic campomelic dysplasia with de novo 5q;17q reciprocal translocation and severe phenotype. J Med Genet 35:597–599

Scherer S, Tsui L-C (1991) Cloning and analysis of large molecules. In: Adolph KW (ed) Advanced techniques in chromosome research. Marcel Dekker, New York, Basel, Hong Kong, pp 33–56

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, et al (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240–244

Südbeck P, Schmitz ML, Baeuerle PA, Scherer G (1996) Sex reversal by loss of the C-terminal transactivation domain of human SOX9. Nat Genet 13:230–232

Sutherland HF, Wadey R, McKie JM, Taylor C, Atif U, Johnstone KA, Halford S, et al (1996) Identification of a novel transcript disrupted by a balanced translocation associated with DiGeorge syndrome. Am J Hum Genet 59:23–31

Toder R, Rappold GA, Schiebel K, Schempp W (1995) ANT3 and STS are autosomal in prosimian lemurs: implications for the evolution of the pseudoautosomal region. Hum Genet 95:22–28

Tommerup N, Schempp W, Meinecke P, Pedersen S, Bolund L, Brandt C, Goodpasture C, et al (1993) Assignment of an autosomal sex reversal locus (SRA1) and campomelic dysplasia (CMPD1) to 17q24.3-q25.1. Nat Genet 4:170–174

Wagner T, Tommerup N, Wirth J, Leffers H, Zimmer J, Back E, Weissenbach J, et al (1997) A somatic cell hybrid panel for distal 17q: GDIA1 maps to 17q25.3. Cytogenet Cell Genet 76:172–175

Wagner T, Wirth J, 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

Wirth J, Wagner T, Meyer J, Pfeiffer RA, Tietze HU, Schempp W, Scherer G (1996) Translocation breakpoints in three patients with campomelic dysplasia and autosomal sex reversal map more than 130 kb from *SOX9*. Hum Genet 97: 186–193

Wright E, Hargrave MR, Christiansen J, Cooper L, Kun J, Evans T, Gangadharan U, et al (1995) The *Sry*-related gene *Sox9* is expressed during chondrogenesis in mouse embryos. Nat Genet 9:15–20

Wunderle VM, Critcher R, Hastie N, Goodfellow PN, Schedl A (1998) Deletion of long-range regulatory elements upstream of *SOX9* causes campomelic dysplasia. Proc Natl Acad Sci USA 95:10649–10654

Young ID, Zuccollo JM, Maltby EL, Broderick NJ (1992) Campomelic dysplasia associated with a de novo 2q;17q reciprocal translocation. J Med Genet 29:251–252