Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome

Ken-Shiung Chen[1], Prasad Manian[1,2], Thearith Koeuth[1], Lorraine Potocki[1,3], Qi Zhao[1], A. Craig Chinault[1,4], Cheng Chi Lee[1,4] & James R. Lupski[1,3,4]

Smith-Magenis syndrome (SMS), caused by del(17)p11.2, represents one of the most frequently observed human microdeletion syndromes. We have identified three copies of a low-copy-number repeat (SMS-REPs) located within and flanking the SMS common deletion region and show that SMS-REP represents a repeated gene cluster. We have isolated a corresponding cDNA clone that identifies a novel junction fragment from 29 unrelated SMS patients and a different-sized junction fragment from a patient with dup(17)p11.2. Our results suggest that homologous recombination of a flanking repeat gene cluster is a mechanism for this common microdeletion syndrome.

DNA deletion is a common mutational event. The size of deletion can vary from a single base to megabases (Mb) of DNA. Small DNA deletions (under 20 bp) may be mediated by short repeats[1]. In such cases, the deletion frequency is proportional to the size of the repeat and to the extent of the homology, but is inversely proportional to the distance between the repeats[2–4]. Factors affecting DNA deletion over megabase distances are not known; however, it is observed that DNA deletion, which results in loss of megabases of DNA, occurs with a higher frequency in specific regions of the human genome and leads to various genetic disorders. Whether these specific deletions represent preferential survival of rearrangements that do not contain haplolethal genes, or reflect the effects of specific genome structural features that lend themselves to DNA rearrangement, remains an open question. Studying recombination events in these disorders may provide insight into the molecular mechanism which leads to these deletion syndromes.

Contiguous-gene syndromes (CGS) comprise a group of disorders associated with chromosomal microdeletions or microduplications. Deletion or duplication of functionally unrelated but physically contiguous dosage-sensitive genes can lead to complex phenotypic abnormalities. The Smith-Magenis syndrome (SMS) is thought to be a contiguous-gene deletion syndrome associated with an interstitial deletion of the short arm of chromosome 17 in band p11.2. The birth prevalence of SMS has been estimated to be 1/25,000 (ref. 5). It is generally believed, however, that the frequency of this disorder is underestimated because the clinical features may be subtle and the cytogenetic abnormality is small and may not be detected by routine cytogenetic analysis. The clinical findings of SMS include mental retardation, neuro-behavioral abnormalities, sleep disturbances, short stature, minor craniofacial and skeletal anomalies, congenital heart defects and renal anomalies[5–9]. The wide spectrum of phenotypic variation could arise from variation in the size of the DNA deletion; however, molecular studies[5,10,11] demonstrate that most patients have the same genomic markers deleted, suggesting that there is a common deletion estimated to span approximately 5 Mb[12,13]. It is likely that haplo-insufficiency of one or more genes in the commonly deleted region underlies this disorder.

To define the molecular basis of SMS and to investigate the hypothesis that specific genome structural features may be responsible for the deletion rearrangement associated with this disorder, we have constructed a physical map of the 17p11.2–12 region in the form of a yeast artificial chromosome (YAC) contig covering 8 Mb. We have identified repetitive sequences both within and surrounding the SMS common deletion region, including three copies of a repeat gene cluster (designated SMS-REPs). Using a probe to flanking SMS-REP sequence, we have identified a patient-specific junction fragment in more than 90% of SMS individuals tested (29 of 31). Our data suggest that recombination between flanking repeat gene clusters is a frequent cause of the SMS common deletion.

Low copy repeats in 17p11.2–12

We constructed a YAC contig for 17p11.2–12 spanning the SMS and the adjacent Charcot-Marie-Tooth disease type-1A (CMT1A) region (Fig. 1). A number of DNA markers had been previously mapped to the SMS critical region[5]. Sequence-tagged sites (STSs) derived from these markers[10,14] were used to screen the CEPH 'mega-YAC'[15] or CEPH Mark I (ref. 16) YAC libraries. We obtained additional STS markers used for screening from the public databases and published genetic linkage maps[17]. We also used seven STSs for genes (*ADORA2B, FLI, snU3, MFAP4, ALDH3, SREBP1* and *cSHMT*) and two STSs (R191 and R193) derived from YAC ends (354G11 and B11C4, respectively) for YAC screening.

The map constructed using an STS-content mapping approach[18–21] contains three overlapping sets of YACs and two gaps. The most telomeric contig extends from the 1,600-kb CEPH mega-YAC 961F10, which contains the *D17S124* locus, to a 1,200 kb CEPH mega YAC 901C9, which contains locus *D17S71* (marker 1041). This region consists of 30 YACs, which include the CMT1A critical region and distal SMS-REP (SMS-REPD). The second contig consists of two YACs, 717F7 and 349A2, which contain *SREBP1*. The third contig spans the markers *FLI* and *D17S58* and contains 55 YACs. This region contains the middle SMS-REP (SMS-REPM) and proximal SMS-REP (SMS-REPP). Although the precise size of the two gaps could not be determined, it is likely that the region encompassing markers 1041-*SREBP1*-*FLI* is less than 1 Mb, as the estimated size of the SMS common deletion is 5 Mb, and the exist-

[1]*Department of Molecular and Human Genetics,* [2]*Department of Medicine,* [3]*Department of Pediatrics and Texas Children's Hospital,* [4]*Human Genome Center, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. Correspondence should be addressed to J.R.L. e-mail: jlupski@bcm.tmc.edu*

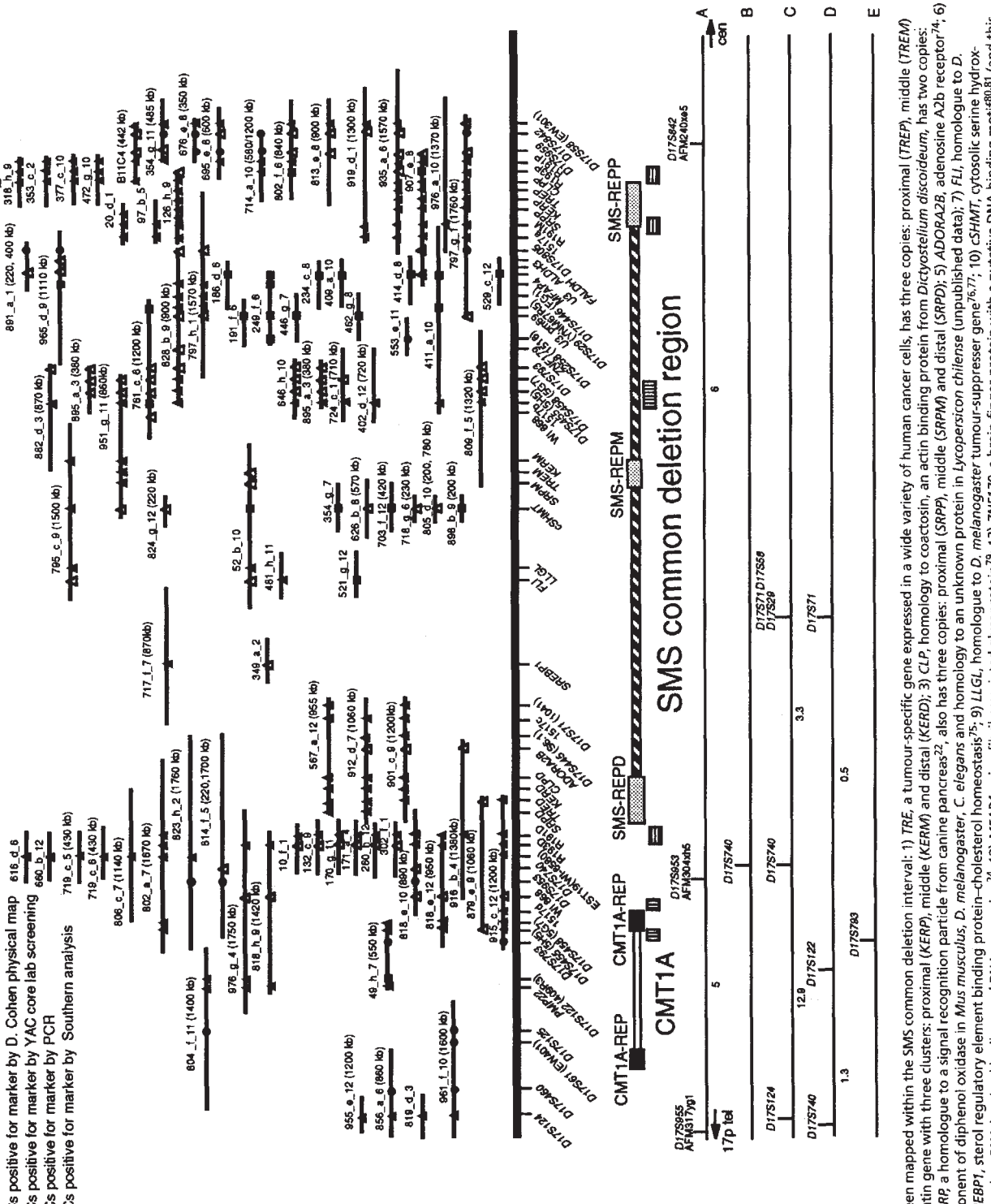

Fig. 1 YAC contig of the 17p11.2 region. Horizontal bars represent 87 YAC clones that map within or adjacent to the CMT1A and SMS critical region. The vertical bars under the bold horizontal line at the middle of the figure represent the physical ordering of markers. The cross-hatched bar represents the SMS common deletion region, which is flanked by and has contained within it a repetitive sequence (SMS-REP) represented by dotted rectangles. The rectangles with vertical lines represent a repeated sequence, one copy of which is present within the SMS critical deletion interval and one copy of which is interrupted by the proximal CMT1A-REP and located adjacent to the proximal CMT1A-REP. The rectangles with horizontal lines represent yet another repeat sequence, known as 'R191', which is present in three copies, one of which is contained within the common deletion. The horizontal lines below the cross-hatched bar represent the current published genetic linkage maps assigned for the characterization of this region: A, the genetic linkage map published by Génénethon[70]; B, Matise[71]; C, Weissenbach[72]; and D, Buetow[73]; E, the current genetic linkage map in the database of Cooperative Human Linkage Center. Numbers under each line represent distances between markers in centimorgans (cM). Two gaps exist in this contig between *D17S71* and *SREBP1*, and between *SREBP1* and *FLI*. Several genes have been mapped within the SMS common deletion interval: 1) *TRE*, a tumour-specific gene expressed in a wide variety of human cancer cells, has three copies: proximal (*TREP*), middle (*TREM*) and distal (*TRED*)[23]; 2) *KER*, type-I keratin gene with three clusters: proximal (*KERP*), middle (*KERM*) and distal (*KERD*); 3) *CLP*, homology to coactosin, an actin binding protein from *Dictyostelium discoideum*, has two copies: proximal (*CLPP*) and distal (*CLPD*); 4) *SRP*, a homologue to a signal recognition particle from canine pancreas[22], also has three copies: proximal (*SRPP*), middle (*SRPM*) and distal (*SRPD*); 5) *ADORA2B*, adenosine A2b receptor[74]; 6) 1041 (*DOXA2*), homology to A2 component of diphenol oxidase in *Mus musculus*, *D. melanogaster*, *C. elegans* and homology to an unknown protein in *Lycopersicon chilense* (unpublished data); 7) *FLI*, homologue to *D. melanogaster* flightless-I gene[66]; 8) *SREBP1*, sterol regulatory element binding protein–cholesterol homeostasis[75]; 9) *LLGL*, homologue to *D. melanogaster* tumour-suppresser gene[76,77]; 10) *cSHMT*, cytosolic serine hydroxymethyltransferase[78]; 11) *snU3*, small nuclear RNA involved in ribosomal RNA processing[24]; 12) *MFAP4*, microfibril-associated glycoprotein[79]; 13) *ZNF179*, a brain finger protein with a putative DNA binding motif[80,81] (and this work); 14) *ALDH3*, aldehyde dehydrogenase 3 (ref. 82); and 15) *FALDH*, fatty aldehyde dehydrogenase associated with autosomal recessive Sjögren-Larsson syndrome[83].

Table 1 • Methods and materials used to study copy number of repeat markers in Fig. 1

Gene/Locus	Copies	Genomic reagents	Methods
D17S793	2	★	❑
5H5 (*D17S455*)	2	★	❑
5G7 (*D17S458*)	2	★	❑
WI868	2	★	❑
1517	4	★	▲
R193	2	✚★✔	▲
R191	3	✚★✔	▲
SRP	3	★✔	▲
TRE	3	★✔	▲
KER cluster	3	★✔	▲
CLP	2	✚★✔	▲
snU3	at least 2	★✔	❑

✚Somatic cell hybrids. ★YAC. ✔Cosmids. ❑PCR. ▲Southern analysis. The genes in the shaded area represent the SMS-REP repeat gene cluster. The *CLP* gene is present in both the distal repeat gene cluster (SMS-REPD) and proximal repeat gene cluster (SMS-REPP) but not in the middle repeat gene cluster (SMS-REPM), whereas *SRP*, *TRE* and *KER* clusters are present in SMS-REPD, SMS-REPM and SMS-REPP.

ing YACs within the SMS region cover at least 4 Mb.

This physical map, in addition to somatic cell hybrid mapping (data not shown), not only confirms the presence of each marker in the SMS region but also identifies several low-copy-number repeats (Table 1). Among the 39 DNA markers derived from the YAC contig, twelve markers — *D17S793*, *D17S445*, *D17S458*, *D17S259* (1517), WI868, R193, R191, *SRP* (signal recognition particle)[22], *TRE* (tumour-specific gene)[23], *KER* (keratin gene), *CLP* (coactosin-like protein) and *snU3* (ref. 24) — were mapped to more than one locus outside, within or flanking the SMS common deletion region. The locations of each repeat were confirmed by YACs that carry adjacent unique markers, which reveal the same repeat sequence in otherwise distinct YACs. Each repeat copy was distinguished either by polymerase chain reaction (PCR)–based analysis or by Southern analysis on a somatic-cell hybrid panel. Each repeat sequence identified is found in normal individuals and also in a monochromosomal (17) somatic-cell hybrid cell line, suggesting that they are present in multiple copies and are not allelic polymorphisms.

CLP flanks the common deletion

Further study of repeat R193 indicates that the proximal copy of R193 (R193P) maps between markers R191P and R191M. R193 identified twenty cosmid clones from a flow-sorted human chromosome 17 cosmid library (LA17NCO1; ref. 25), including c145G12. This cosmid was also identified by reciprocal probing[26] of the same LA17NCO1 library with clones from a human placenta cDNA library. Cosmid c145G12 was mapped by fluorescence *in situ* hybridization (FISH) to 17p11.2, and was found to hybridize by Southern blotting with a 1.4-kb cDNA clone 41G7A. When clone 41G7A was used to probe a northern blot, it identified a single mRNA band of approximately 2 kb expressed predominantly in placenta, lung, liver and kidney (Fig. 2*a*). Re-screening of the placenta cDNA library with 41G7A identified several cDNAs of 1.85 kb. The sequence revealed an open reading frame of 576 bp with a potential start codon (ATG) at position 150 and stop codon (TAA) at nucleotide 576 (Fig. 2*b*) and identified more than twenty EST clones in the genome database. The deduced amino-acid sequence shows significant peptide homology to coactosin (P_{BLAST}=5.4e-25), an actin-binding protein from *Dictyostelium discoideum*, with 74.9% of the amino acids being identical or conserved (Fig. 2*c*; ref. 27). We therefore designated this gene coactosin-like protein (*CLP*). Although the genomic structure of this gene is not yet fully characterized, the 3′ untranslated region shows more than 95% sequence identity over 210 bp to a human cDNA that contains a sequence with autonomously replicating activity (Fig. 2*b*; ref. 28).

To further localize the genomic region of *CLP*, we performed Southern analysis using a somatic-cell hybrid-mapping panel of the 17p region. A 1.1-kb *Hin*dIII fragment from the cDNA clone 41G7A, which contains the 3′ end of the coding region and part of the 3′ untranslated region, was used as the *CLP* probe. These studies mapped *CLP* to each end of the SMS common deletion region (Fig. 3). The presence of 8.1-, 5.0- and 0.9-kb *Hin*dIII bands in total human genomic DNA and in the monochromosomal hybrid for chromosome 17 (MH22-6) (Fig. 3*a*, lanes 1 and 4, respectively) indicates that these bands are specific for chromosome 17 and are not due to allelic polymorphism. The 8.1-kb band is absent in hybrids 254-1D and 357-2D, indicating that this band corresponds to the distal copy of *CLP* (Fig. 3*a*, lanes 5 and 6, respectively). Similarly, the 5.0-kb and 0.9-kb bands are absent in hybrid 484-2D and 88H5 (Fig. 3*a*, lanes 7 and 8, respectively) but present in 254-1D, indicating the presence of the other copy of *CLP* in proximal 17p11.2 (*CLPP*). Consistent with these results, the chromosome 17–specific bands (8.1, 5.0 and 0.9 kb) are absent

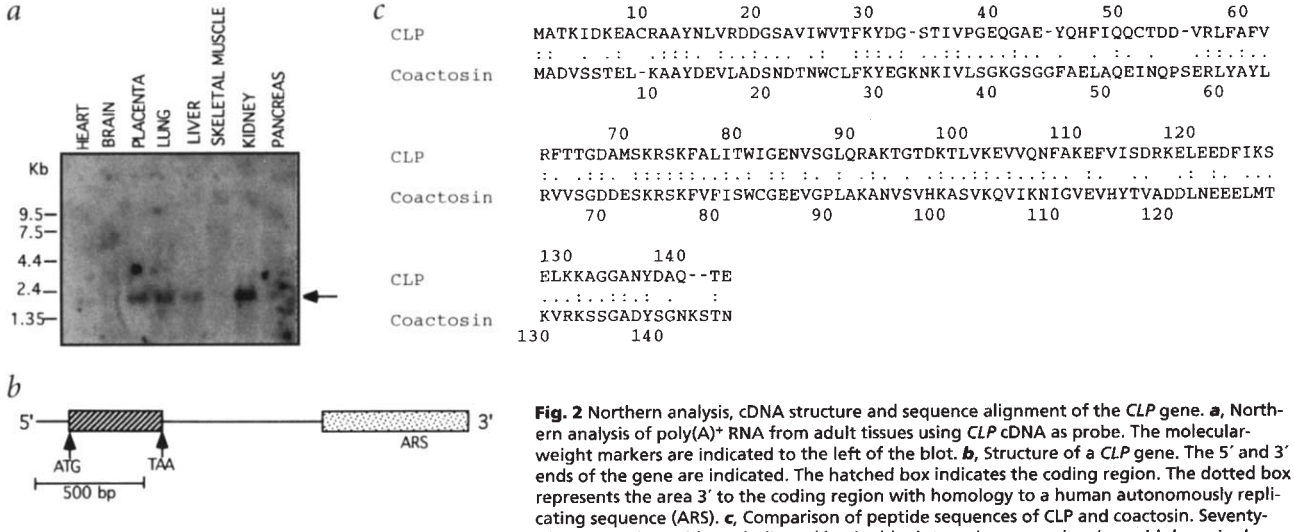

Fig. 2 Northern analysis, cDNA structure and sequence alignment of the *CLP* gene. ***a***, Northern analysis of poly(A)⁺ RNA from adult tissues using *CLP* cDNA as probe. The molecular-weight markers are indicated to the left of the blot. ***b***, Structure of a *CLP* gene. The 5′ and 3′ ends of the gene are indicated. The hatched box indicates the coding region. The dotted box represents the area 3′ to the coding region with homology to a human autonomously replicating sequence (ARS). ***c***, Comparison of peptide sequences of CLP and coactosin. Seventy-five percent of amino acids are homologous; 33.3% are identical. Identical amino acids are indicated by double dots and conserved amino acids by a single dot. Dashes indicate gaps that have been introduced to achieve optimal sequence alignment.

in hybrids LS-1 and 484-2D. The presence of a 1.6-kb band in human genomic DNA and somatic-cell hybrids 254-1D and 357-2D, which was not detected in MH22-6, suggests homologous sequences elsewhere in the genome. Southern analysis of a rodent somatic-cell hybrid panel retaining individual human chromosomes suggests the presence of homologous sequences on chromosomes 15, 16 and Y (data not shown).

Southern analysis of the YAC clones carrying unique markers further confirmed the presence of two copies of *CLP*, one at each end of the SMS critical region. DNA from the YAC clones spanning the SMS common deletion junctions were digested with *Hin*dIII, and Southern blots were then hybridized with the *CLP* probe. This probe detects an 8.1-kb band in distal YACs 912D7 and 567A12 (Fig. 3*b*, middle panel, lanes 1 and 2), confirming that this band corresponds to the distal copy of SMS-REP. In contrast, both 5.0-kb and 0.9-kb bands are identified within 797G1 and 907E8 (Fig. 3*b*, middle panel, lanes 3 and 4), which span the proximal junction of the SMS critical region. Southern analysis using the same *CLP* probe identified cosmid clones corresponding to proximal (*CLPP*) and distal (*CLPD*) copies of this gene. Human chromosome 17–specific cosmid library (LA17NCO1) was screened with R193 and *CLP* probes, respectively. DNA from putatively positive cosmid clones was purified, and subsequent Southern-blot analysis of selected clones indicated that three of them —c99C12, c80C2 and c52H1—contain the 8.1-kb *Hin*dIII fragment and thus map to the distal SMS-REP (Fig. 3*b*, right panel, lanes 1, 2 and 3), while c116H12, c100H4 and c77E6 contain the 5.0-kb and 0.9-kb *Hin*dIII fragments and thus map to the proximal copy of SMS-REP (Fig. 3*b*, right panel, lanes 4, 5 and 6).

SMS-REP is a complex repeated gene cluster

To further characterize SMS-REP, we used a ~400-kb CEPH Mark I YAC clone (126H9; Fig. 1) spanning the proximal SMS common deletion junction to screen an arrayed chromosome 17–specific cosmid library[25] and identified 149 cosmids. The unusually high number of cosmid clones identified (five to ten times the expected frequency) suggests that this sequence has multiple copies on chromosome 17. Cosmid clones were digested with *Eco*RI, and subsequent Southern analysis and restriction mapping allowed us to assign 41 of them into three regions. Group 1, between c99C12 and c5F2 with one gap, contains twelve cosmids from the distal SMS-REP interval. Group 2, between c37A5 and c28G10 with one gap, contains eight cosmids from the middle SMS-REP. Group 3, between c145G12 and c31C8 with no gap, contains 21 cosmids from the proximal SMS-REP (Fig. 4). Among these cosmids, c99C12, c80C2 and c52H1 have been shown to associate with *CLPD*, while c116H12, c100H4 and c77E6 have been shown to associate with *CLPP* (Figs 3*b*,4). These cosmids were therefore used as 'anchors' for subsequent cosmid walking.

DNA sequence obtained from the SMS-REP cosmids by sequencing from the T3 and T7 ends of the cosmid vector identified homology to three genes. Sequence analysis of the T3 end of c155B11 and c37A5 using BLASTN reveals significant homology to *SRP* (encoding a signal recognition particle from canine pancreas) ($P=9.5e^{-16}$) and *KER* (type-I keratin) ($P=9.6e^{-14}$), respectively. Moreover, the T7 end of c37A5 shows significant homology to human *TRE* oncogene (tumour-specific antigen) ($P=4.9e^{-92}$). The T7 end of c37A5 also hybridizes with cosmid c77E6, which contains *CLPP*, suggesting that a copy of *TRE* is adjacent to *CLPP*. Subsequent characterization demonstrated that SMS-REPP is a gene cluster containing at least one copy of *TRE*, *SRP* and a *KER* gene cluster in addition to *CLP*. The size of proximal SMS-REPP is at least 200 kb, with the *KER* cluster and *TRE* located between *CLP* and *SRP*. The SMS-REP repeats appear to be highly homologous throughout their entire length on the basis of i) cross-hybridization by Southern analysis under stringent conditions using independent

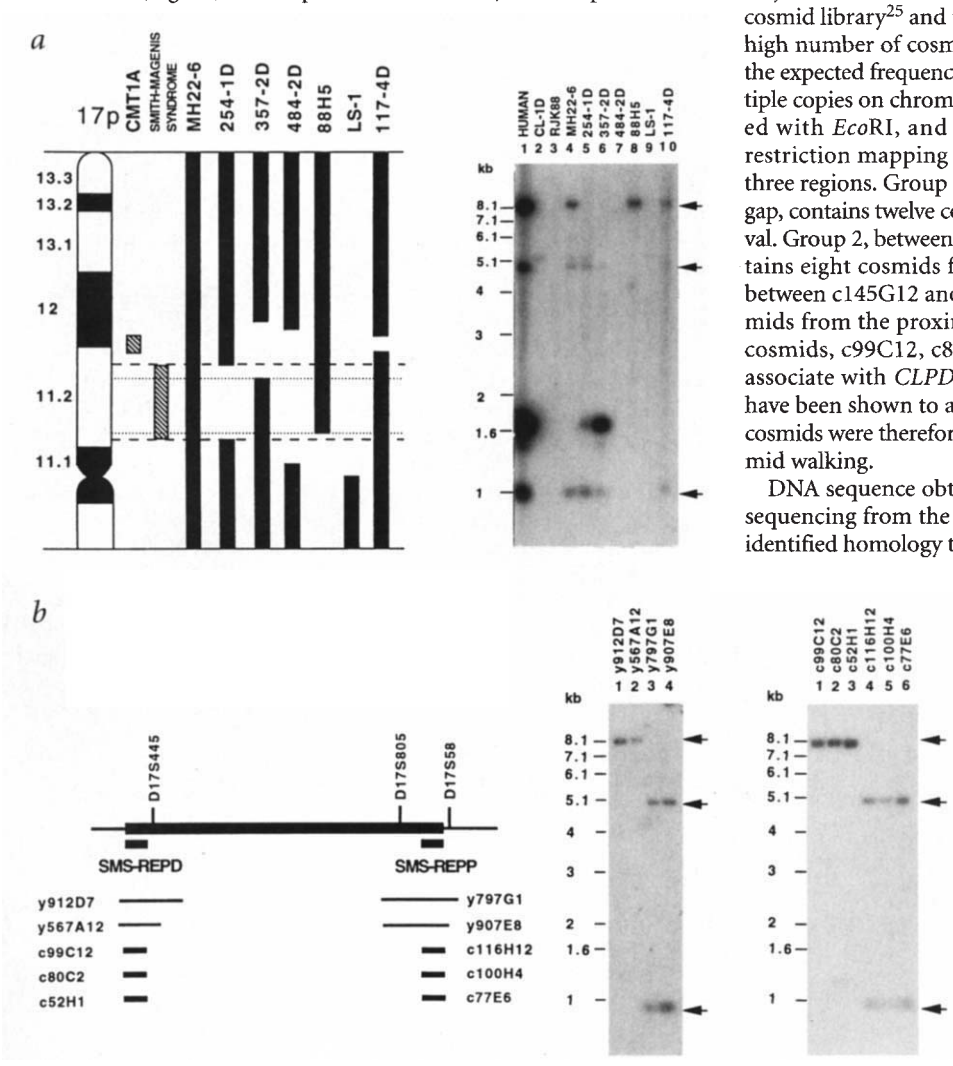

Fig. 3 Genomic localization of *CLPP* and *CLPD* in 17p11.2. ***a***, To the left is an ideogram of 17p and deletions represented in our deletion hybrid mapping panel. The hatched rectangles represent the CMT1A and SMS critical regions. Solid bars indicate the 17p region retained in each hybrid. The dashed lines delineate the SMS common deletion region. The dotted lines delineate the proximal breakpoint of 357-2D and translocation breakpoint of 88H5. To the right is the Southern analysis identifying two copies of *CLP* in chromosome 17p11.2. DNA samples were digested with *Hin*dIII. Arrows indicate the human chromosome 17–specific DNA fragments that hybridize with a 1.1-kb *CLP* cDNA probe. The 8.1-kb *Hin*dIII band maps to the distal SMS-REP as seen in hybrids 88H5 and 117-4D (lanes 8 and 10). Two bands, 5.0 kb and 0.9 kb, map to the proximal SMS-REP in hybrids 254-1D, 357-2D and 117-4D (lanes 5, 6 and 10). The 5.1-kb band (lanes 2, 4, 5 and 9) represents cross-hybridization of this probe to the mouse chromosome. ***b***, To the left is a map of the SMS critical region showing the location of proximal and distal SMS-REP and the associated YAC and cosmid clones. In the middle is the Southern blot demonstrating YACs associated with SMS-REP. To the right is the Southern blot showing identification of cosmids with the *CLP* probe.

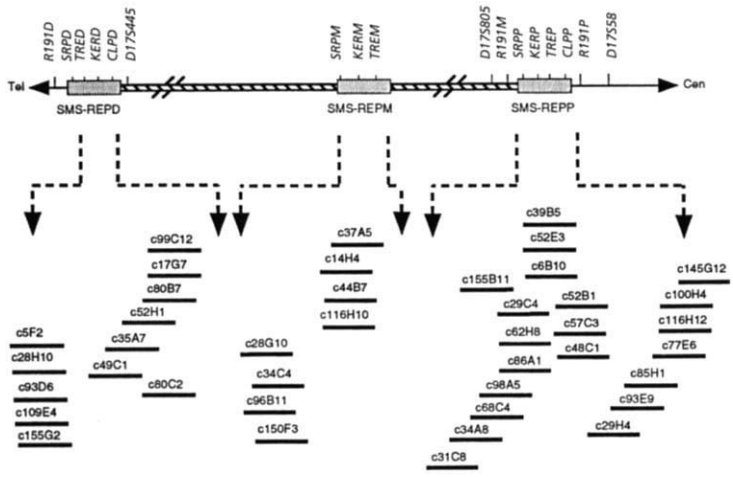

Fig. 4 Partial cosmid contigs of the SMS-REP region. The cross-hatched bar represents the SMS common deletion region. The dotted rectangles represent the SMS-REP that maps to each end and within the common deletion region. The precise order and orientation of genes within SMS-REPD and SMS-REPM cannot be determined on the basis of current information. To build these contigs, a flow-sorted human chromosome 17 cosmid library was screened with YAC 126H9 that encompasses the proximal SMS-REP. A total of 149 cosmids were obtained and analysed. Horizontal bars beneath each SMS-REP represent cosmids mapped to that region.

probes to repeat specific clones, ii) nucleotide sequence data of selected clones from different repeats and iii) restriction mapping of repeat-specific cosmids showing highly homologous patterns.

Three copies of SMS-REP in the 17p11.2 region

PCR amplification of thirteen representative YACs spanning the SMS region suggests that *SRP*, *KER* and *TRE* are present in SMS-REPP, SMS-REPM and SMS-REPD, whereas *CLP* maps to SMS-REPP and SMS-REPD but not SMS-REPM (data not shown). Three groups of YACs appear to associate with *TRE*, *KER* and *SRP*. The presence of *TRE*, *KER* and *SRP* in YAC clones 567A12 and 912D7, which were shown to map to the distal SMS region with the *CLP* probe (Fig. 3), indicates that 567A12 and 912D7 associate with the SMS-REPD. Similarly, the presence of *TRE*, *KER* and *SRP* and the unique maker *D17S959* in YAC clones 797G1 and 907E8, which map to proximal SMS region (Fig. 3), indicates that they are associated with SMS-REPP. Finally, presence of these three markers and a unique marker *cSHMT* in non-overlapping YAC clones identified SMS-REPM.

Southern analysis of *Eco*RI-digested YAC DNA from the distal (912D7 and 567A12), middle (795C9 and 951G11) and proximal (797G1 and 126H9) SMS regions with *SRP*, *KER*, *TRE* and *CLP* probes further confirmed the presence of three copies of SMS-REP (Fig. 5). The probes used for the Southern analysis to detect *SRP*, *KER*, *TRE* and *CLP* were derived from the T3 end of c155B11, the T3 end of c37A5, the 3′ end of *TRE* and the 1.1-kb *Hin*dIII fragment of the cDNA clone 41G7A, respectively. The *SRP* probe gave a 12-kb hybridization signal, and the *TRE* probe gave two hybridization signals of 7.2 kb and 7.3 kb with each of the YACs except 795C9. The *KER* probe hybridized to each YAC at a band of approximately 12.5 kb. By contrast, the *CLP* probe gave no hybridization to two middle YACs, while giving a 10-kb hybridization signal to two distal YACs and two hybridization signals (23.1 and 5 kb) to two proximal YACs. These results allowed three copies of SMS-REPs to be distinguished.

Identification of a new junction fragment

To test the hypothesis that the repeat-sequence SMS-REP might mediate large-scale chromosomal recombination and lead to recurrent SMS deletion, we performed Southern analysis using the *CLP* probe to hybridize to *Not*I-digested genomic DNA from three unrelated SMS families. After pulsed field gel electrophoresis (PFGE), we identified a 1.2-Mb *de novo* junction fragment (Fig. 6*a*) in all three SMS patients, but not in any of their parents. These results suggest that a homologous recombination event mediated by the flanking SMS-REP elements is the molecular basis for the SMS deletion. To determine the frequency of this junction fragment among SMS patients, we performed Southern analysis on samples from 28 additional unrelated SMS patients and a negative control. Representative results (Fig. 6*b*) show that the *CLP* probe identified the 1.2-Mb junction fragment in every SMS patient tested except for patients 266 and 536 and the negative control 242. Patients 266 and 536 are known to carry atypical deletions that extend beyond the common SMS distal breakpoint[11].

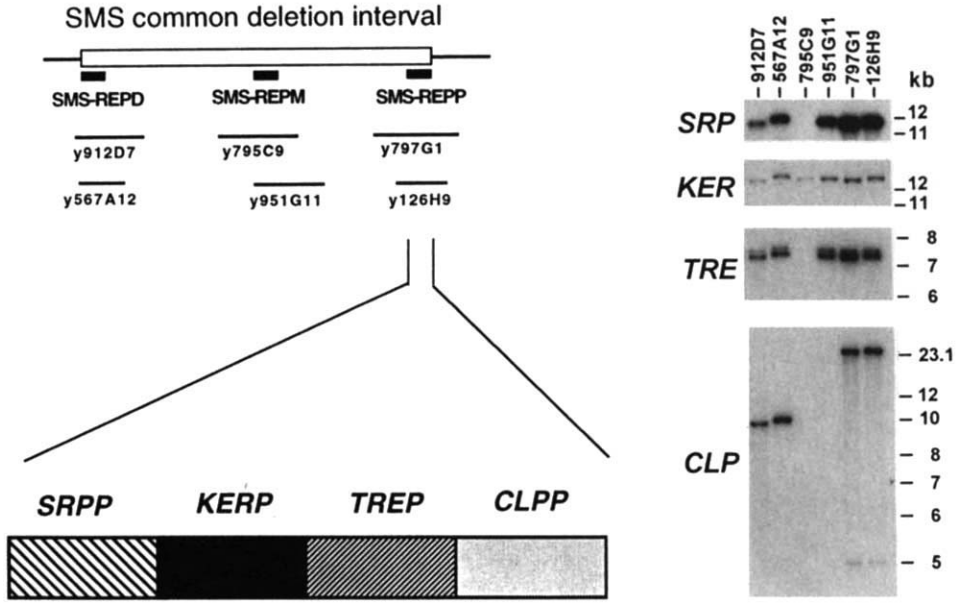

Fig. 5 Identification of a repeated-gene cluster in proximal, middle and distal SMS-REPs. To the left is a map of the SMS common deletion region showing the location of proximal, middle and distal SMS-REP and the associated YACs. Four genes—*SRP*, *KER*, *TRE* and *CLP*—have been mapped to the SMS-REPP as illustrated. To the right are Southern-blot results, demonstrating that *SRP*, *KER* and *TRE* are present in SMS-REPP, SMS-REPM and SMS-REPD, whereas *CLP* maps to SMS-REPP and SMS-REPD but not SMS-REPM. DNA from YAC clones were digested with *Eco*RI. The probes used for the Southern analysis to detect *SRP*, *KER*, *TRE* and *CLP* were derived from the T3 end of c155B11, the T3 end of c37A5, the PCR product amplified from 3′ end of *TRE* and the 1.1 kb *Hin*dIII fragment of cDNA clone 41G7A, respectively.

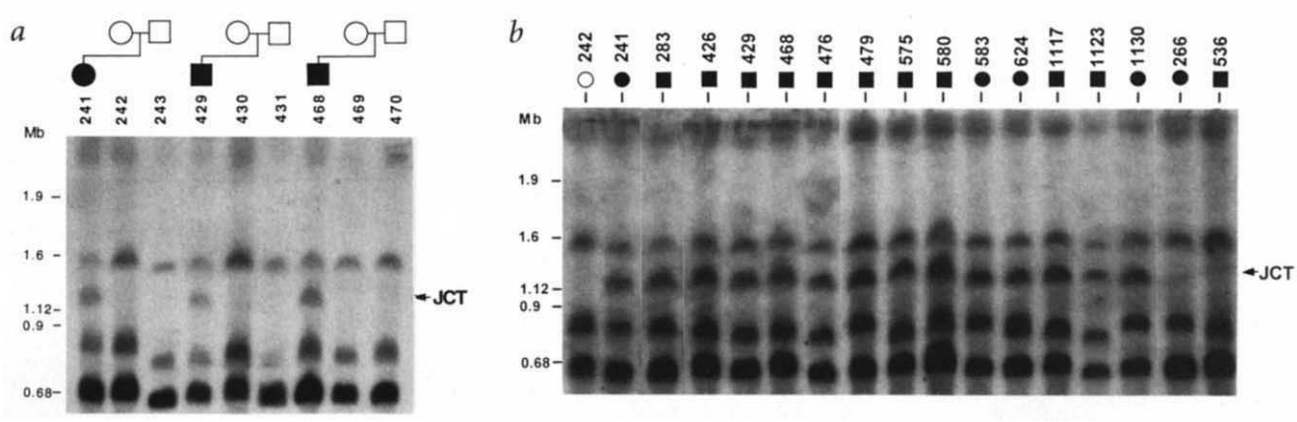

Fig. 6 PFGE analysis of an SMS junction fragment. ***a***, PFGE analysis to detect the SMS deletion junction fragment in three unrelated SMS families. Lymphoblastoid DNA from three SMS patients and their parents were used for preparation of agarose plugs[48], which were digested with *Not*I and fractionated by PFGE. The Southern blot was hybridized with a *CLP* cDNA probe, which identified a 1.2-Mb junction fragment (arrow, JCT) that is present in SMS patients 241, 429 and 468, but not their unaffected parents (242, 243, 430, 431, 469 and 470). ***b***, Detection of the 1.2-Mb *de novo* junction fragment from unrelated SMS patients. PFGE plugs were prepared from lymphoblast cell lines derived form SMS patients and an unaffected parent (242). Note the presence of the 1.2-Mb band or junction fragment in every SMS patient except patients 266, 536 (patients with larger deletions, see text) and negative control (242). Total genomic yeast DNA from *Saccharomyces cerevisiae* was used for a size marker.

We have thus identified a specific SMS deletion junction fragment in a total of 29 *de novo* unrelated patients using the *CLP* probe. This junction fragment appears to be identical in size (1.2 Mb), within the limits of resolution of PFGE analysis.

Mapping SMS common deletion breakpoints

To further define the breakpoints of the SMS common deletion, we performed Southern analysis with somatic-cell hybrids retaining the del(17)p11.2 chromosome from six patients with the common deletion (244, 246, 254, 263, 280 and 424; ref. 10) using the *CLP* cDNA probe. The result showed an absence of the 8.1-kb *CLPD* hybridization but positive 5.0- and 0.9-kb *CLPP* hybridizations in each case examined (Fig. 7*a*), suggesting that the distal and proximal breakpoints among these patients are telomeric to *CLPD* and telomeric to *CLPP* cross-hybridizing fragments, respectively.

We performed similar experiments using the R191 probe. The locations of repeat R191, derived from the end of a YAC clone (354G11), were initially determined by YAC contig mapping (Fig. 1) and by somatic-cell hybrid mapping (Fig. 7*b*). Probe R191 appears to map to three loci in the SMS region. The presence of 9.5-, 2.8- and 2.1-kb *Eco*RI bands in total human genomic DNA and in the monochromosomal hybrid for chromosome 17 (MH22-6) (Fig. 7*b*, lanes 1 and 4, respectively) indicates that these bands are specific for chromosome 17 and not due to allelic polymorphism. The 9.5 kb band is absent in hybrids 254-1D and 88H5 (Fig. 7*b*, lanes 5 and 8, respectively) but is present in 357-2D (lane 6), indicating that this

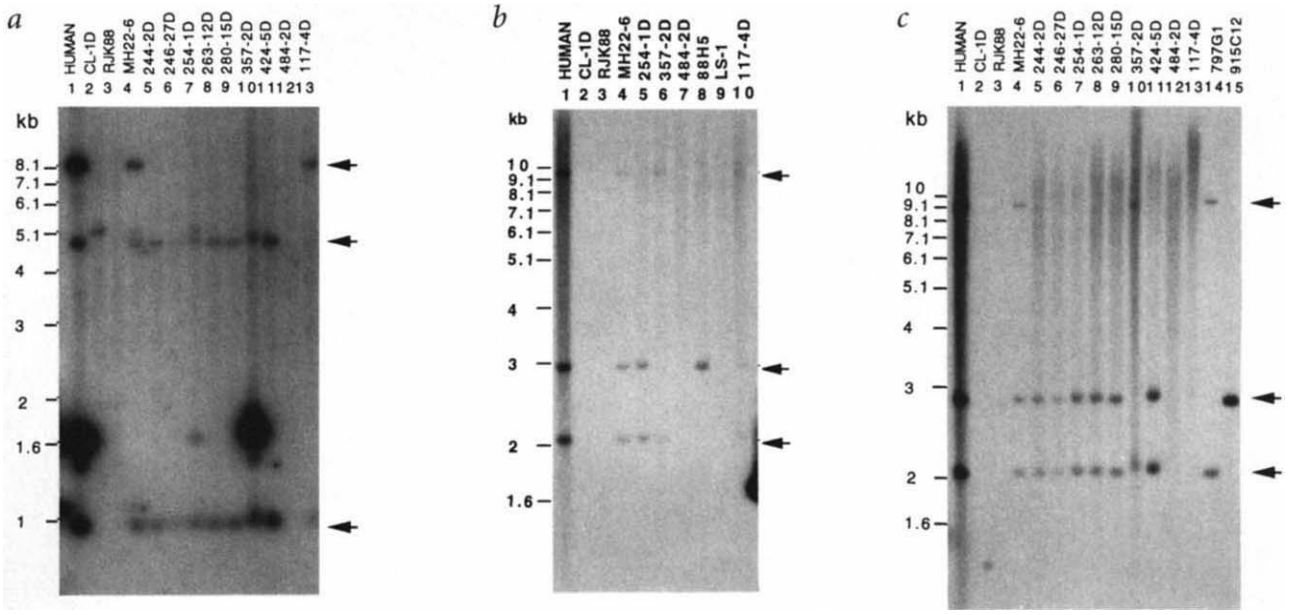

Fig. 7 Deletion of *CLP* and R191 cross-hybridizing fragments in SMS chromosomes. ***a***, DNA samples were digested with *Hin*dIII and hybridized to the *CLP* cDNA probe (1.1-kb *Hin*dIII fragment). The arrows indicate the genomic fragments derived from chromosome 17 that hybridize with the probe. The 1-kb DNA ladder was used as size marker. ***b***, Identification of three copies of R191 within and flanking the SMS deletion at the 17p11.2 region by Southern analysis (see text for detail). ***c***, Southern analysis using probe R191 was performed with *Eco*RI-digested DNA samples arranged in the same order as in *a* but with two additional YAC clones, 797G1 and 915c12 (lanes 14 and 15). The 9.5-kb and 2.1-kb bands map proximal to 17p11.2, as seen in 357-2D (lane 10) and YAC 797G1 (lane 14; see also Fig. 1). A 2.8-kb band maps distal to SMS region, as seen in YAC 915C12 (lane 15; see also Fig. 1). The 9.5-kb band is absent in every SMS patient tested (lanes 5–9 and 11,12).

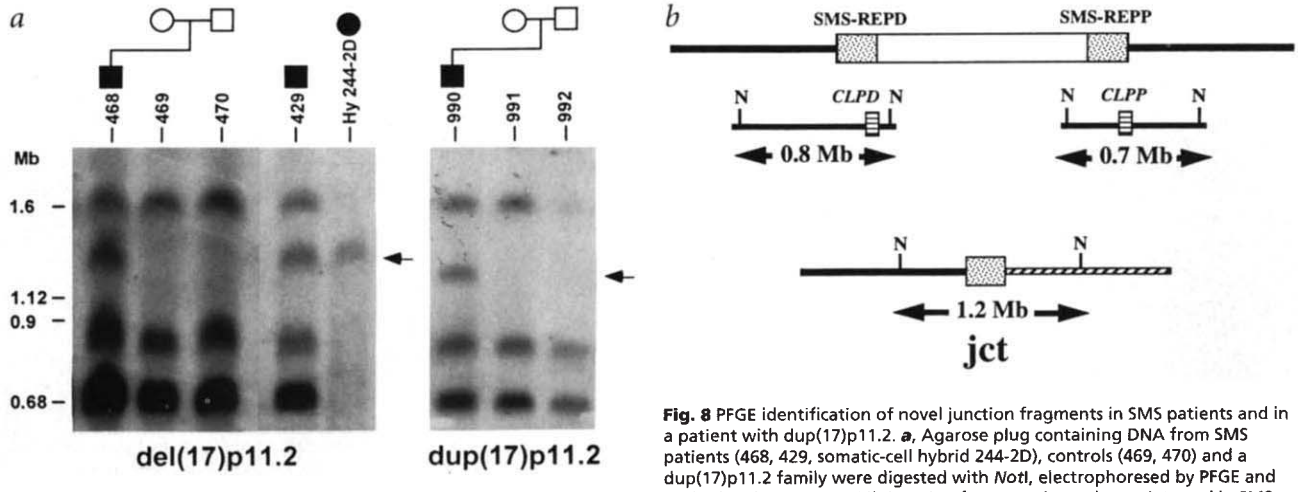

Fig. 8 PFGE identification of novel junction fragments in SMS patients and in a patient with dup(17)p11.2. ***a***, Agarose plug containing DNA from SMS patients (468, 429, somatic-cell hybrid 244-2D), controls (469, 470) and a dup(17)p11.2 family were digested with *Not*I, electrophoresed by PFGE and probed with *CLP*. A 1.2-Mb junction fragment (arrow) was detected in SMS patients 468, 429 and 244. Similarly, a novel junction fragment (arrow) was detected in patient 990, who has a dup(17)p11.2. ***b***, The recombination between SMS-REPD and SMS-REPP that gives rise to the 1.2-Mb *de novo* fragment is summarized in this illustration. The open box and dotted boxes show the SMS common deletion region and SMS-REP, respectively. The vertical lines represent *Not*I (N) sites, and the numbers below the lines refer to the size of *Not*I fragment. The jct depicts the 1.2-Mb observed junction fragment representing a recombinant chromosome from an interchromosomal exchange between chromosome homologues (bold and slashed horizontal lines).

band maps within the SMS common deletion region and is thereby designated R191M. The 2.8-kb band is absent in hybrid 357-2D (lane 6) but present in 254-1D and 88H5, indicating the presence of another copy in distal 17p11.2 (R191D). The 2.0-kb band is present in 254-1D and 357-2D (lanes 5 and 6, respectively) but absent in 88H5 (lane 8), indicating the presence of a copy in proximal 17p11.2 (R191P). Consistent with these results, the chromosome 17–specific bands (9.5, 2.8 and 2.1 kb) are not present in either hybrid LS-1, where 17p is deleted, or 484-2D [del(17)p11.1p12]. Presence of these three bands in hybrid 117-4D excludes them from the CMT1A region. Probe R191 was used to analyse somatic-cell hybrids retaining the del(17)p11.2 chromosome as described with the *CLP* cDNA probe. The results show an absence of the 9.5-kb (R191M) band and a positive hybridization to the 2.8-kb (R191D) band and the 2.1-kb (R191P) band in each case examined (Fig. 7*c*), and suggest that the DNA deletion in these patients extends centromeric to R191D and telomeric to R191P.

These results suggest that the proximal breakpoints in SMS patients with the common deletion all fall within the region between R191M and *CLPP*, and the distal breakpoints all fall between R191D and *CLPD*. These results are consistent with our YAC physical mapping data.

Junction fragment from a dup(17)p11.2 patient

To determine whether the common deletion defined by somatic-cell hybrid mapping with probes R191 and *CLP* also gives rise to a 1.2-Mb junction fragment, we performed PFGE on the rodent somatic-cell hybrid 244-2D, which retains a del(17)p11.2 chromosome from an SMS patient (Fig. 8). Indeed, both 244-2D and 429 (SMS patient sample) showed the same 1.2-Mb junction fragment. In addition, three bands with the size of 1.6, 0.8 and 0.7 Mb are absent from the somatic-cell hybrid. Similar analyses using other somatic-cell hybrids indicate that 0.8-Mb and 0.7-Mb bands are derived from *Not*I fragments containing *CLPD* and *CLPP*, respectively (data not shown), while the 1.6-Mb band is probably derived from a *CLP* homologue from another chromosome. The recombination between the 0.8-Mb *Not*I fragment containing SMS-REPD and the 0.7-Mb *Not*I fragment containing SMS-REPP gives rise to the 1.2-Mb junction fragment.

To determine whether recombination events mediated by SMS-REP repeats could lead to duplication of the SMS region, we subjected a DNA sample from a *de novo* dup(17)p11.2 patient (#990) to the same analysis. The dup(17)p11.2 was identified by G-banding and confirmed by FISH[29], showing duplication for *FLI* and not for *PMP22*. PFGE analysis shows a novel fragment in this patient when compared to his parents (Fig. 8*a*), which is apparently a different size from that found in SMS patients with the common deletion. The identification of such a novel fragment in *de novo* dup(17)p11.2 suggests that recombination mediated by SMS-REP not only results in del(17)p11.2 but also may cause duplication in proximal 17p11.2 (Fig. 8*b*).

Discussion

An interesting feature of the physical map in the SMS region, which has implications regarding the rearrangements in this region, is that 12 of 40 markers are repeated. Among them, *D17S259* (1517) appears to have four copies, five markers (R191, *SRP*, *TRE*, *KER* gene cluster and *snU3*) appear to have three copies and six markers (WI868, R193, *D17S793*, *D17S455*, *D17S458* and *CLP*) appear to have two copies. R191, R193 and *CLP* were distinguished by a combination of YAC physical mapping and Southern analysis on somatic-cell hybrid panels, whereas the rest of the markers were distinguished by YACs that carry adjacent unique markers. These low-copy-number repeats are problematic in YAC contig construction using PCR-based STS-content analysis. YACs derived from different loci containing clusters of low-copy repeats can easily result in a false YAC contig. Resolution of the physical map may therefore require Southern analysis of YACs and somatic-cell hybrid panel mapping to distinguish each locus. In fact, the present physical map from this region (http://www.genome.wi.mit.edu:80/cgi-bin/contig/lookup_contig?contig=WC17.3&database=release) has a number of discrepancies that can be resolved only with independent methods. In addition, repetitive sequences within YACs may lead to YAC instability by recombination between these repeats. In fact, we have observed that mega-YACs from the SMS region are unstable. This SMS critical region overlaps with a breakpoint cluster region associated with primitive neuro-ectodermal tumours, suggesting that it is prone to DNA rearrangements in somatic cells[30,31].

Three lines of evidence support the conclusion that three

copies of SMS-REP are present in the SMS common deletion region. First, the combined use of YAC contig, somatic-cell hybrid and cosmid mapping enables SMS-REPP, SMS-REPM and SMS-REPD to be distinguished (Figs 1,3,4). Second, PCR analysis of the YACs with unique associated markers spanning the SMS region identified non-overlapping YACs containing SMS-REPP, SMS-REPM and SMS-REPD (data not shown). Third, Southern analysis of YACs with unique associated markers further confirmed that three copies of the SMS-REP are present in the SMS region (Fig. 5). At least four genes have been identified in SMS-REPP and SMS-REPD, but it is not certain whether these represent functional copies or pseudogenes. Regardless, an extensive region of homology (apparently at least 200 kb) is provided by these low-copy repeat-gene clusters.

We demonstrate that the deletion breakpoint in most SMS patients occurs within a repeated gene cluster, SMS-REP, which flanks the commonly deleted region. Moreover, the entire approximately 200 kb of SMS-REPs appears to be homologous by virtue of cross-hybridization on Southern analysis with probes on repeat specific clones. Furthermore, an SMS-REP repeat-sequence probe detects a novel junction fragment that is patient specific, deletion chromosome specific and of similar size in more than 90% of SMS individuals. These observations suggest that homologous recombination occurs within the flanking SMS-REP repeat gene cluster.

Low-copy repeats have been implicated as the cause of DNA re-arrangement associated with several human diseases, including hypertension secondary to glucocorticoid-remediable aldosteronism[32], spinal muscular atrophy[33–35], X-linked ichthyosis[36–38], colour-blindness[39], haemoglobinopathies[40] and haemophilia[41]. The presence of repetitive sequences in the 17p11.2 region can hypothetically facilitate homologous recombination and lead to chromosomal deletion and SMS. If so, the size of the deletion might be expected to vary according to the distance between each misaligned repeat. However, most SMS patients seem to have a similar region deleted, although a few of them reportedly have smaller[5,42,43] and larger[44] 17p11.2 deletions[12]. The identification of a *de novo* patient-specific junction fragment of similar size, within the limits of resolution of PFGE analysis, in several SMS patients suggests that a precise recombination mechanism mediates the SMS common deletion. The general uniformity of the deletion size suggests that only one set of the repeats flanking the SMS common-deletion region is responsible for most of the recombinants. Indeed, our data demonstrate that recombination occurs at SMS-REPP and SMS-REPD, located 5 Mb apart on 17p11.2, in more than 90% of the SMS deletion patients examined. Given the repetitive sequences found in the SMS region, it is conceivable that repeats other than SMS-REPs may also lead to deletion in this region. Consistent with this hypothesis is the fact that the *CLP* probe did not detect a junction fragment in patients 266 and 536. One can speculate that genomic repeat selection favouring recombination may relate to i) the size of the repeat, ii) the extent of the homology, iii) the distance between the repeats, iv) the proximity to promoters or open transcription complexes[45,46], v) potential Z-DNA elements[47], vi) higher-order chromosome structural features or yet unknown mechanisms.

Considering the distance between SMS-REPs, it is intriguing to find that recombination in SMS patients almost always occurs between more remotely spaced repeats, SMS-REPD and SMS-REPP, rather than between more closely spaced repeats, SMS-REPP and SMS-REPM or SMS-REPD and SMS-REPM. Mapping *CLP* to both ends of the SMS common-deletion region in SMS-REPD and SMS-REPP but not in SMS-REPM and the identification of a specific junction fragment in 29 of the 31 unrelated SMS patients suggest the preferential involvement of the larger homologous regions and potentially *CLP* itself in the recombination event leading to SMS deletion. The presence of a DNA sequence with an autonomous replication activity identified by database searches in the 3′ end of *CLP* has potential implications for replication initiation and recombination. Initiation of DNA replication from an autonomously replicating sequence or active transcription through the region may be sufficient to unwind DNA and render a localized region of DNA (which includes *TRE*, the *KER* cluster and *SRP*) more accessible to recombinase-promoted homologous recombination.

To our knowledge, only three mechanisms have been proposed to explain large chromosomal DNA deletions in humans. One is based on an interchromosomal rearrangement, another is based on intrachromosomal rearrangement and the third is based on a fragile site associated with an expanded trinucleotide repeat. The inherited neuropathies CMT1A and hereditary neuropathy with liability to pressure palsies (HNPP), which result from a reciprocal duplication/deletion, represent one of the best-characterized examples of human genetic disease resulting from homologous interchromosomal recombination between widely spaced (1.5 Mb) flanking low-copy repeats[48–51]. An intrachromosomal rearrangement has been observed in a sporadic HNPP patient with *de novo* deletion[52]. A recently proposed mechanism for terminal 11q deletions in Jacobson syndrome patients is the expansion of $p(CCG)_n$ repeats at a folate sensitive site[53]. A $p(CCG)_n$ repeat is located at 5′ of *CBL2*, a known proto-oncogene on 11q. Expansion of this trinucleotide repeat causes methylation of the adjacent CpG island and deletion of the distal chromosomal segment; however, the deletion breakpoint is not at the fragile site. In the case of SMS deletion, intrachromosomal recombination may occur and cause the 17p11.2 deletion. The fact that a *CLP* probe identified an apparent junction in a dup(17)p11.2 patient suggests that interchromosomal recombination can also be a mechanism for SMS deletion. Two patients with duplication of 17p11.2 have been suggested to represent the reciprocal of the SMS deletion, but clearly other dup(17p) can vary in size[54], and rigorous physical analysis is required to provide supportive evidence for a reciprocal recombination mechanism.

So far, two distinct regions on 17p are known to be potentially active in unequal crossover events that lead to duplication/deletion events and cause distinct clinical syndromes. These regions are the CMT1A/HNPP region in 17p12 (refs 48–51) and SMS region in 17p11.2 (this study). The discoveries of repeat-gene clusters in the SMS-REP flanking the SMS region and the *COX10* gene in the CMT1A-REP flanking the CMT1A/HNPP region[55] suggest that extensive homology within the repeats is needed for efficient recombination. These chromosome-specific repetitive structures appear to predispose the human genome to frequent microdeletion or microduplication that causes gross change of genomic DNA and can be considered genomic disorders caused by an alteration of the genome rather than a mutation or disruption of a putative coding sequence. From an evolutionary perspective, genomic disorders are simply the outcome of selective processes that reflect the dynamics of genome evolution.

Our data indicate that homologous recombination between flanking SMS-REP repeat-gene clusters is the molecular basis for the SMS common deletion. Low-copy repeats have been identified in the genomic region implicated in DiGeorge/velo-cardiofacial [del(22)q11] (ref. 56), Prader-Willi/ Angelman [del(15)q11–13] (refs 57,58) and Williams syndromes [del(7)q11.23] (ref. 59). Whether other human contiguous-gene syndromes also reflect unique genome structural features of the affected loci remains to be seen.

Methods

PCR. PCRs were performed with extracted YAC DNA (10 ng) using each of the following primer sets (5′→3′): R191 (CTTCTTGGTGGTGGCTTTTACC/GCCCTGACTTTTGAGTGCTCCC), R193 (GGCAGCTCAGGGTGAGCTCTTC/TATTGGCCTTAAATGCATCTCA), *CLP* (TCTGTAAACTGTCTGAGTGCAGAG/CGTCTGCACCACACAATCAAAAGG), *3′TRE* (ACAGGTAGCACAATCTACTAA/TTCTGTGTTTACTTGTATGAGG), *KER* (CTCTGCTCTGACCCTCTA/AGCCCTGATCCTTGGGGTCCAG), *SRP* (GGAAGCACTTGCTGTCATCC/GCCCAGGCCAAATGGCCCTGG), *FLI* (GGCAGACAGTGGGTGGCACCT/GTCACTTAAGTTACTGAGATT), *LLGL* (AGCGGCATCGCTTCGTGCGT/TTAGCCTGGCTCAGCTTGGG), *ZNF179* (ACCTTTCCCCAAGTTGGAGCTA/GAGCTTCATCTTCTCGCCCAGG), *cSHMT* (AGGAGCGGGCCCACTCTGGA/AAGAAACCCCCGCAAAGGGC), 5H5 (AAAGCACAACGCTAGGAGACTGG/CCTCCCCAGTCAGTCCATCAGC), 5G7 (GATAGAGGACTCCACACAGGAAG/GCTCTGGGTTATCTACTGATCCCT), *D17S793* (TGTTGGAGTTAATGTGCCAT/TCTTTGACCCAGACCTCTAA), *snU3* (GCTGTCATTCAGTATTATGCTAAT/AATGATCCCTGAAAGTATAGTCTT), *MFAP* (CTGAATGAACTGGCTTCATACACAC/GAGGCCTGTGAAATACAAGGTTCC), *D17S805* (ATCACTTGAACCTGAGGGG/AATGAGATACCGATCCATGC), *D17S959* (TCAGATTGAACTCTCGGTAT/GCTGACACAGGCAATG) and EW301 (GTCTTCTCTCCCACCACTAATAC/GCGTTTGAACCATACTCATCC). PCR reactions were performed in 30-µl volumes using 1 U of Ampli-*Taq* polymerase in 1 µM of each primer; 250 µM each of dATP, dTTP, dGTP and dCTP; and 1×GeneAmp PCR buffer (10 mM Tris-HCl, pH8.3 [at 25 °C]; 50 mM KCl; 1.5 mM $MgCl_2$; 0.001% [w/v] gelatin) under 30 µl of mineral oil. PCR condition was 95 °C for 4 min, 35 cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 1 min, followed by 72 °C for 6 min. The products were visualized on a 2% agarose gel.

Sequencing and sequence analysis. Sequences of inserts in cosmids were performed with the ABI 373A automated sequencer using T7 and T3 primers. Searches of the database were performed with BLAST on the NCBI file server[60].

Cosmid screening. A flow-sorted human chromosome 17 cosmid library using sCos1 as a vector was kindly provided by L. Deavan (Los Alamos National Laboratory)[25]. This gridded and arrayed library was screened with PFGE-purified YAC 126H9 DNA as the probe. DNA from each candidate cosmid clone was prepared and used for subsequent analysis. To construct the restriction maps, purified cosmid DNA was digested completely with the rare cutting enzyme *Not*I, followed by partial digestion with *Eco*RI, electrophoresed and transferred to a nylon membrane (Sureblot, Oncor). This membrane was subsequently hybridized to T3 and T7 promoter-specific oligonucleotides as described[61].

PFGE analysis. High-molecular-weight DNA was isolated in agarose plugs from peripheral blood samples, somatic-cell hybrid cell lines and Epstein-Barr virus–transformed lymphoblastoid cell lines established from controls and patients[48]. Plugs were rinsed four times for 30 min each in 20 vol of TE buffer. The plugs were equilibrated for 30 min at room temperature with *Not*I restriction buffer provided by the manufacturer (New England Biolabs). Finally, the plugs were incubated at 37 °C overnight with 20 U of enzyme. Separation of DNA fragments was achieved by using a CHEF MAPPER (BioRAD) for 26 h in 0.5×TBE running buffer with a pulse time of 24–204 s ramp at 6 v/cm. After treatment with 0.25 N HCl for 14 min, followed by 0.4 N NaOH for 40 min, gels were blotted onto nylon membrane (Oncor).

YAC library screening. An STS-based PCR screening approach[18–21] was used to screen 'mega-YAC'[15] or CEPH Mark I (ref. 16) YAC library. The mega-YAC library was screened with seven STS markers: 5H5, 5G7, 1517, WI6580, *ADORA2b*, 1041 (*D17S71*) and *MFAP4*. The CEPH library was screened with marker R193. Both libraries were screened with seven additional STS markers: R191, *SREBP1*, *FLI*, *cSHMT*, *snU3*, *ALDH3* and EW301. Individual YAC clones were grown in AHC media[62], and genomic DNA was prepared according to the standard method[18]. YAC end fragments were derived by inverse PCR using conditions described previously[63,64].

Southern blotting. DNA samples were digested with restriction enzymes and electrophoresed in a 1% agarose gel with 1×TAE buffer (40 mM Tris-HCl [pH 8.5], 40 mM sodium acetate, 2 mM EDTA) at 30 V for 16 h. The DNA was transferred to a nylon membrane (Sureblot, Oncor) and hybridized to DNA probes labelled by random primer incorporation of [α-^{32}P] dCTP[65]. Hybridizations were performed according to stringent conditions as described[66].

Cell lines. DNA from the following somatic-cell hybrids were used: MH22-6, a human–mouse somatic-cell hybrid retaining human chromosome 17 as the only human chromosome[67]; 254-1D, a human–mouse somatic hybrid containing the SMS del(17)p11.2p11.2 chromosome[10]; 244-2D, 246-27D, 263-12D, 280-15D, 283-15D, 424-5D and 429-6D, human–hamster somatic-cell hybrids containing the common SMS del(17)p11.2–11.2 chromosome[10]; 357-2D, a human–hamster somatic-hybrid from a del(17)p11.2–12 patient with an unique 17p deletion that retains most of the 17p11.2; 484-2D, a human–hamster somatic cell hybrid derived from a SMS patient with an unusual deletion, del(17)p11.1p12, that has one end point more proximal than that of 254-1D and the other end point extending distally beyond the CMTIA duplication[44]; 88H5, human–hamster somatic hybrid that retains 17pter-p11.2 (ref. 67); LS-1, human–mouse somatic hybrid that retains 17cen-qter[68]; 117-4D, human–hamster somatic hybrid derived from an HNPP patient containing human chromosome 17 with a 1.5-Mb deletion in 17p12 (ref. 69); the two rodent parental cell lines, CL-1D and RJK88, derived from mouse and hamster, respectively.

GenBank accession numbers. cDNA clone 41G7A, L54057; SMS-REP cosmids c155B11, AF012852; c37A5, AF012851, AF012852.

Acknowledgements

We thank the patients, families and their physicians and health-care providers, without whom this study would not have been possible. We appreciate the critical reviews of A. Beaudet, A. Bradley, L. Pentao and L.G. Shaffer. This research was supported in part by the Texas Children's Hospital General Clinical Research Center (M01 RR-00188), the Baylor College of Medicine Human Genome Center (HG00210), the Mental Retardation Research Center (HD24064), the Child Health Research Center (HD94021), Department of the Army (DAMD17-94-J-4484) and the Clayton Foundation for Research.

Received 19 May; accepted 24 August 1997.

1. Krawczak, M. & Cooper, D.N. Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. *Hum. Genet.* **86**, 425–441 (1991).
2. Peeters, B.P.H., de Boer, J.H., Bron, S. & Venema, G. Structural plasmid instability in *Bacillus subtilis*: effect of direct and inverted repeats. *Mol. Gen. Genet.* **212**, 450–458 (1988).
3. Singer, B.S. & Westlye, J. Deletion formation in bacteriophage T4. *J. Mol. Biol.* **202**, 233–243 (1988).
4. Whoriskey, S.K., Schofield, M.A. & Miller, J.H. Isolation and characterization of *Escherichia coli* mutants with altered rates of deletion formation. *Genetics* **127**, 21–30 (1991).
5. Greenberg, F. *et al.* Molecular analysis of the Smith-Magenis syndrome: a possible contiguous-gene syndrome associated with del(17)(p11.2). *Am. J. Hum. Genet.* **49**, 1207–1218 (1991).
6. Smith, A.C.M. *et al.* Interstitial deletion of (17)(p11.2p11.2) in nine patients. *Am. J. Med. Genet.* **24**, 393–414 (1986).
7. Stratton, R.F. *et al.* Interstitial deletion of (17)(p11.2p11.2): report of six additional patients with a new chromosome deletion syndrome. *Am. J. Med. Genet.* **24**, 421–432 (1986).
8. Greenberg, F. *et al.* Multi-disciplinary clinical study of Smith-Magenis syndrome (deletion 17p11.2). *Am. J. Med. Genet.* **62**, 247–254 (1996).
9. Chen, K.-S., Potocki, L. & Lupski, J.R. The Smith-Magenis syndrome [del(17)p11.2]: clinical review and molecular advances. *Ment. Retard. Dev. Disabil. Res. Rev.* **2**, 122–129 (1996).
10. Guzzetta, V. *et al.* Somatic cell hybrids, sequence-tagged sites, simple repeat polymorphisms, and yeast artificial chromosomes for physical and genetic mapping of proximal 17p. *Genomics* **13**, 551–559 (1992).
11. Juyal, R.C. *et al.* Molecular analyses of 17p11.2 deletions in 62 Smith-Magenis

syndrome patients. *Am. J. Hum. Genet.* **58**, 998–1007 (1996).
12. Trask, B.J. *et al.* Quantification by flow cytometry of chromosome-17 deletions in Smith-Magenis syndrome patients. *Hum. Genet.* **98**, 710–718 (1996).
13. Ledbetter, D.H. & Ballabio, A. Molecular cytogenetics of contiguous gene syndromes: mechanisms and consequences of gene dosage imbalance. in *The Metabolic and Molecular Bases of Inherited Disease,* 7th Ed., Vol. 1 (eds Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D.) 811–839 (McGraw-Hill, New York, 1995).
14. Guzzetta, V., Montes de Oca-Luna, R., Lupski, J.R. & Patel, P.I. Isolation of region-specific and polymorphic markers from chromosome 17 by restricted *Alu* polymerase chain reaction. *Genomics* **9**, 31–36 (1991).
15. Bellanné-Chantelot, C. *et al.* Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell* **70**, 1059–1068 (1992).
16. Albertsen, H.M. *et al.* Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl. Acad. Sci. USA* **87**, 4256–4260 (1990).
17. Murray, J.C. *et al.* A comprehensive human linkage map with centimorgan density. *Science* **265**, 2049–2054 (1994).
18. Green, E.D. & Olson, M.V. Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **87**, 1213–1217 (1990).
19. Kwiatkowski, T.J.J., Zoghbi, H.Y., Ledbetter, S.A., Ellison, K.A. & Chinault, A.C. Rapid identification of yeast artificial chromosome clones by matrix pooling and crude lysate PCR. *Nucleic Acids Res.* **18**, 7191–7192 (1990).
20. Chumakov, I.M. *et al.* Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. *Nature Genet.* **1**, 222–225 (1992).
21. Chumakov, I. *et al.* Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359**, 380–387 (1992).
22. Herz, J., Flint, N., Stanley, K., Frank, R. & Dobberstein, B. The 68 kDa protein of signal recognition particle contains a glycine-rich region also found in certain RNA-binding proteins. *FEBS Lett.* **276**, 103–107 (1990).
23. Nakamura, T. *et al.* A novel transcriptional unit of the *tre* oncogene widely expressed in human cancer cells. *Oncogene* **7**, 733–741 (1992).
24. Chevillard, C. *et al.* Relationship between Charcot–Marie–Tooth 1A and Smith–Magenis regions: *snU3* may be a candidate gene for the Smith–Magenis syndrome. *Hum. Mol. Genet.* **2**, 1235–1243 (1993).
25. Kallioniemi, O.-P. *et al.* Physical mapping of chromosome 17 cosmids by fluorescence *in situ* hybridization and digital image analysis. *Genomics* **20**, 125–128 (1994).
26. Lee, C.C. *et al.* Isolation of chromosome-specific genes by reciprocal probing of arrayed cDNA and cosmid libraries. *Hum. Mol. Genet.* **8**, 1373–1380 (1995).
27. de Hostos, E.L., Bradtke, B., Lottspeich, F. & Gerisch, G. Coactosin, a 17 kDa F-actin binding protein from *Dictyostelium discoideum*. *Cell. Motil. Cytoskeleton* **26**, 181–191 (1993).
28. Wu, C., Friedlander, P., Lamoureux, C., Zannis-Hadjopoulos, M. & Price, G.B. cDNA clones contain autonomous replication activity. *Biochim. Biophys. Acta* **1174**, 241–257 (1993).
29. Shaffer, L.G., Kennedy, G.M., Spikes, A.S. & Lupski, J.R. Diagnosis of CMT1A duplications and HNPP deletions by interphase FISH: implications for testing in the cytogenetics laboratory. *Am. J. Med. Genet.* **69**, 325–331 (1997).
30. Scheurlen, W.G. *et al.* High-resolution deletion mapping of chromosome arm 17p in childhood primitive neuroectodermal tumors reveals a common chromosomal disruption within the Smith-Magenis region, an unstable region in chromosome band 17p11.2. *Genes Chromosomes Cancer* **18**, 50–58 (1997).
31. Wilgenbus, K.K. *et al.* Molecular characterization of a genetically unstable region containing the SMS critical area and a breakpoint cluster for human PNETs. *Genomics* **42**, 1–10 (1997).
32. Lifton, R.P. *et al.* A chimaeric 11β-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. *Nature* **355**, 262–265 (1992).
33. Melki, J. *et al. De novo* and inherited deletions of the 5q13 region in spinal muscular atrophies. *Science* **264**, 1474–1477 (1994).
34. Roy, N. *et al.* The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. *Cell* **80**, 167–178 (1995).
35. Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80**, 155–165 (1995).
36. Ballabio, A., Bardoni, B., Guioli, S., Basler, E. & Camerino, G. Two families of low-copy-number repeats are interspersed on Xp22.3: implications for the high frequency of deletions in this region. *Genomics* **8**, 263–270 (1990).
37. Yen, P.H. *et al.* Frequent deletions of the human X chromosome distal short arm result form recombination between low copy repetitive elements. *Cell* **61**, 603–610 (1990).
38. Li, X.-M., Yen, P.H. & Shapiro, L.J. Characterization of a low copy repetitive element S232 involved in the generation of frequent deletions of the distal short arm of the human X chromosome. *Nucleic Acids Res.* **20**, 1117–1122 (1992).
39. Nathans, J., Piantanida, T.P., Eddy, R.L., Shows, T.B. & Hogness, D.S. Molecular genetics of inherited variation in human color vision. *Science* **232**, 203–210 (1986).
40. Maniatis, T., Fritsch, E.F., Lauer, J. & Lawn, R.M. The molecular genetics of human hemoglobins. *Annu. Rev. Genet.* **14**, 145–178 (1980).
41. Lakich, D., Kazazian, H.H., Jr., Antonarakis, S.E. & Gitschier, J. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nature Genet.* **5**, 236–241 (1993).
42. Juyal, R.C. *et al.* Apparent mosaicism for del(17)(p11.2) ruled out by fluorescence in situ hybridization in a Smith-Magenis syndrome patient. *Am. J. Med. Genet.* **59**, 406–407 (1995).
43. Juyal, R.C. *et al.* Smith-Magenis syndrome deletion: a case with equivocal cytogenetic findings resolved by fluorescence in situ hybridization. *Am. J. Med. Genet.* **58**, 286–291 (1995).
44. Zori, R.T. *et al.* Clinical, cytogenetic, and molecular evidence for an infant with Smith-Magenis syndrome born from a mother having a mosaic 17p11.2p12 deletion. *Am. J. Med. Genet.* **47**, 504–511 (1993).
45. Botto, M. *et al.* Homozygous hereditary C3 deficiency due to a partial gene deletion. *Proc. Natl. Acad. Sci. USA* **89**, 4957–4961 (1992).
46. Lehrman, M.A., Russell, D.W., Goldstein, J.L. & Brown, M.S. Alu-Alu recombination deletes splice acceptor sites and produces secreted low density lipoprotein receptor in a subject with familial hypercholesterolemia. *J. Biol. Chem.* **262**, 3354–3361 (1987).
47. Kmiec, E.B., Angelides, K.J. & Holloman, W.K. Left-handed DNA and the synaptic pairing reaction promoted by Ustilago Rec1 protein. *Cell* **40**, 139–145 (1985).
48. Pentao, L., Wise, C.A., Chinault, A.C., Patel, P.I. & Lupski, J.R. Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nature Genet.* **2**, 292–300 (1992).
49. Chance, P.F. *et al.* Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. *Hum. Mol. Genet.* **3**, 223–228 (1994).
50. Roa, B.B. & Lupski, J.R. Molecular genetics of Charcot-Marie-Tooth neuropathy. in *Advances in Human Genetics,* Vol. 22 (eds Harris, H. & Hirschhorn, K.) 117–152 (Plenum, New York, 1994).
51. Reiter, L.T. *et al.* A recombination hotspot responsible for two inherited peripheral neuropathies is located near a *mariner* transposon-like element. *Nature Genet.* **12**, 288–297 (1996).
52. LeGuern, E. *et al.* A *de novo* case of hereditary neuropathy with liability to pressure palsies (HNPP) of maternal origin: a new mechanism for deletion in 17p11.2? *Hum. Mol. Genet.* **5**, 103–106 (1996).
53. Jones, C. *et al.* Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene *CBL2*. *Nature* **376**, 145–149 (1995).
54. Roa, B.B. *et al.* Duplication of the *PMP22* gene in 17p partial trisomy patients with Charcot-Marie-Tooth type-1A neuropathy. *Hum. Genet.* **97**, 642–649 (1996).
55. Reiter, L.T., Murakami, T., Koeuth, T., Gibbs, R.A. & Lupski, J.R. The human *COX10* gene is disrupted during homologous recombination between the 24-Kb proximal and distal CMT1A-REPs. *Hum. Mol. Genet.* **6**, 1595–1603 (1997).
56. Halford, S. *et al.* Low-copy-number repeat sequences flank the DiGeorge/velo-cardio-facial syndrome loci at 22q11. *Hum. Mol. Genet.* **2**, 191–196 (1993).
57. Ji, Y. *et al.* An evolutionarily conserved gene associated with the common deletion breakpoint regions in the Prader-Willi/Angelman syndromes. *Am. J. Hum. Genet.* **59**, A158 (1996).
58. Christian, S.L. *et al.* Molecular characterization of two proximal deletion breakpoint regions in both Prader-Willi and Angelman syndrome patients. *Am. J. Hum. Genet.* **57**, 40–48 (1995).
59. Pérez Jurado, L.A., Peoples, R., Kaplan, P., Hamel, B.C.J. & Francke, U. Molecular definition of the chromosome 7 deletion in Williams syndrome and parent-of-origin effects on growth. *Am. J. Hum. Genet.* **59**, 781–792 (1996).
60. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
61. Evans, G.A., Lewis, K. & Rothenberg, B.E. High efficiency vectors for cosmid microcloning and genomic analysis. *Gene* **79**, 9–20 (1989).
62. Brownstein, B.H. *et al.* Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244**, 1348–1351 (1989).
63. Ochman, H., Gerber, A.S. & Hartl, D.L. Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621–623 (1988).
64. Silverman, G.A. *et al.* Yeast artificial chromosome cloning of a two-megabase-size contig within chromosomal band 18q21 establishes physical linkage between *BCL2* and plasminogen activator inhibitor type-2. *Genomics* **9**, 219–228 (1991).
65. Feinberg, A.P. & Vogelstein, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6–13 (1983).
66. Chen, K.-S. *et al.* The human homologue of the *Drosophila melanogaster* flightless-I gene (*fliI*) maps within the Smith-Magenis microdeletion critical region in 17p11.2. *Am. J. Hum. Genet.* **56**, 175–182 (1995).
67. van Tuinen, P., Rich, D.C., Summers, K.M. & Ledbetter, D.H. Regional mapping panel for human chromosome 17: application to neurofibromatosis type 1. *Genomics* **1**, 374–381 (1987).
68. Elder, F.F.B., Nichols, M.M., Hood, O.J. & Harrison, W.R, III. Unbalanced translocation (15;17)(q13;p13.3) with apparent Prader-Willi syndrome but without Miller-Dieker syndrome. *Am. J. Med. Genet.* **20**, 519–524 (1985).
69. Roa, B.B. *et al.* Evidence for a recessive *PMP22* point mutation in Charcot-Marie-Tooth disease type 1A. *Nature Genet.* **5**, 189–194 (1993).
70. Gyapay, G. *et al.* The 1993–94 Généthon human genetic linkage map. *Nature Genet.* **7**, 246–249 (1994).
71. Matise, T.C., Perlin, M. & Chakravarti, A. Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nature Genet.* **6**, 384–390 (1994).
72. Weissenbach, J. *et al.* A second-generation linkage map of the human genome. *Nature* **359**, 794–801 (1992).
73. Buetow, K.H. *et al.* Human genetic map; genome maps V: wall chart. *Science* **265**, 2055–2070 (1994).
74. Townsend-Nicholson, A., Baker, E., Sutherland, G.R. & Schofield, P.R. Localization of the adenosine A2b receptor subtype gene (ADORA2B) to chromosome 17p11.2–p12 by FISH and PCR screening of somatic cell hybrids. *Genomics* **25**, 605–607 (1995).
75. Hua, X., Wu, J., Goldstein, J.L., Brown, M.S. & Hobbs, H.H. Structure of the human gene encoding sterol regulatory element binding protein-1 (*SREBF1*) and localization of *SREBF1* and *SREBF2* to chromosomes 17p11.2 and 22q13. *Genomics* **25**, 667–673 (1995).
76. Campbell, H.D. *et al.* Genomic structure, evolution, and expression of human *FLII*, a gelsolin and leucine-rich-repeat family member: overlap with *LLGL*. *Genomics* **42**, 46–54 (1997).
77. Koyama, K. *et al.* The human homologue of the murine *Llglh* gene (*LLGL*) maps within the Smith-Magenis syndrome region in 17p11.2. *Cytogenet. Cell Genet.* **72**, 78–82 (1996).
78. Elsea, S.H. *et al.* Haploinsufficiency of cytosolic serine hydroxymethyltransferase in the Smith-Magenis syndrome. *Am. J. Hum. Genet.* **57**, 1342–1350 (1995).
79. Zhao, Z. *et al.* The gene for a human microfibril-associated glycoprotein is commonly deleted in Smith-Magenis syndrome patients. *Hum. Mol. Genet.* **4**, 589–597 (1995).
80. Matsuda, Y. *et al.* Chromosome mapping of human (*ZNF179*), mouse, and rat genes for brain finger protein (bfp), a member of the RING finger family. *Genomics* **33**, 325–327 (1996).
81. Kimura, T. *et al.* The brain finger protein gene (*ZNF179*), a member of the RING finger family, maps within the Smith-Magenis syndrome region at 17p11.2. *Am. J. Med. Genet.* **69**, 320–324 (1997).
82. Hiraoka, L.R., Hsu, L. & Hsieh, C.-L. Assignment of *ALDH3* to human chromosome 17p11.2 and *ALDH5* to human chromosome 9p13. *Genomics* **25**, 323–325 (1995).
83. De Laurenzi, V. *et al.* Sjögren–Larsson syndrome is caused by mutations in the fatty aldehyde dehydrogenase gene. *Nature Genet.* **12**, 52–57 (1996).