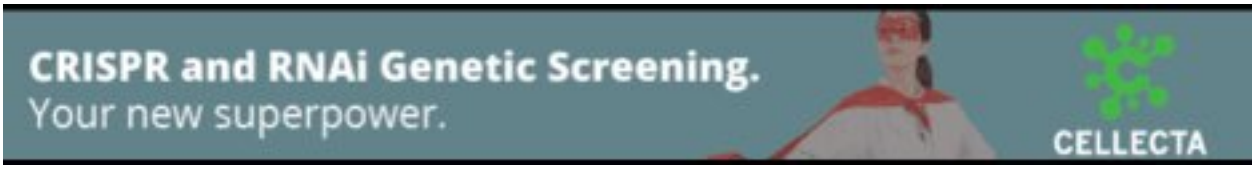
Additional Complexity on Human Chromosome 15q: Identification of a Set of Newly Recognized Duplicons (LCR15) on 15q11–q13, 15q24, and 15q26

Miguel Angel Pujana, Marga Nadal, Mònica Gratacòs, et al.

Genome Res. 2001 11: 98-111
Access the most recent version at doi:10.1101/gr.155601



Letter

Additional Complexity on Human Chromosome 15q: Identification of a Set of Newly Recognized Duplicons (LCR15) on 15q11–q13, 15q24, and 15q26

Miguel Angel Pujana,[1] Marga Nadal,[1] Mònica Gratacòs,[1] Belén Peral,[1,4] Katalin Csiszar,[2] Rogelio González-Sarmiento,[3] Lauro Sumoy,[1] and Xavier Estivill[1,5]

[1]Medical and Molecular Genetics Centre-Institut de Recerca Oncologica, Hospital Duran i Reynals, Barcelona, Spain; [2]Pacific Biomedical Research Center, University of Hawaii, Honolulu, Hawaii, USA; [3]Unidad de Medicina Molecular, Departamento de Medicina, Universidad de Salamanca, Salamanca, Spain

Several cytogenetic alterations affect the distal part of the long arm of human chromosome 15, including recurrent rearrangements between 12p13 and 15q25, which cause congenital fibrosarcoma (CFS). We present here the construction of a BAC/PAC contig map that spans 2 Mb from the neurotrophin-3 receptor (*NTRK3*) gene region on 15q25.3 to the proximal end of the Bloom's syndrome region on 15q26.1, and the identification of a set of new chromosome 15 duplicons. The contig reveals the existence of several regions of sequence similarity with other chromosomes (6q, 7p, and 12p) and with other 15q cytogenetic bands (15q11–q13 and 15q24). One region of similarity maps on 15q11–q13, close to the Prader-Willi/Angelman syndromes (PWS/AS) imprinting center. The 12p similar sequence maps on 12p13, at a distance to the ets variant 6 (*ETV6*) gene that is equivalent on 15q26.1 to the distance to the *NTRK3* gene. These two genes are the targets of the CFS recurrent translocations, suggesting that misalignments between these two chromosomes regions could facilitate recombination. The most striking similarity identified is based on a low copy repeat sequence, mainly present on human chromosome 15 (LCR15), which could be considered a newly recognized duplicon. At least 10 copies of this duplicon are present on chromosome 15, mainly on 15q24 and 15q26. One copy is located close to a *HERC2* sequence on the distal end of the PWS/AS region, three around the lysyl oxidase-like (*LOXL1*) gene on 15q24, and three on 15q26, one of which close to the IQ motif containing GTPase-activating protein 1 (*IQGAP1*) gene on 15q26.1. These LCR15 span between 13 and 22 kb and contain high identities with the golgin-like protein (*GLP*) and the SH3 domain-containing protein (*SH3P18*) gene sequences and have the characteristics of duplicons. Because duplicons flank chromosome regions that are rearranged in human genomic disorders, the LCR15 described here could represent new elements of rearrangements affecting different regions of human chromosome 15q.

[The sequence data described in this paper have been submitted to EMBL GenBank with the accession nos. AJ272070, AJ276448, AJ276449, AJ286892-AJ286929, AJ400620, AJ400621, AJ400817-AJ400820, AJ277869-AJ277874.]

The long arm of human chromosome 15 is characterized by the relatively frequent appearance of cytogenetic alterations. Among them there are deletions that cause the Prader-Willi/Angelman syndromes (PWS/AS: PWS [MIM 176270]; AS [MIM 105830]) (Khan and Wood 1999); pericentromeric inversions and duplications known as inv dup(15) or derivative/dicentric chromosome (Webb 1995); other types of duplications, such as those found in some cases of autism (Cook et al. 1997); and interstitial triplications (Schinzel et al. 1994). The existence of at least three to five copies of transcriptionally active large repeat units, also known as duplicons (Eichler 1998), which facilitate nonhomologous recombination events, provides the molecular basis for 15q11–q13 deletions observed in PWS/AS (Amos-Landgraf et al. 1999; Christian et al. 1999). This type of repeat regions has also been found in other genomic disorders (Lupski 1998), such as the Williams-Beuren syndrome on 7q11.23 (WBS [MIM 194050]) (Peoples et al. 2000), the Smith-Magenis syndrome deletion and the corresponding duplication (SMS [MIM 182290]) (Potocki et al. 2000), Charcot-Marie-Tooth 1A and hereditary neuropathy with liability to pressure palsies on 17p12 (CMT1A [MIM 118220]; HNPP [MIM 162500]) (Chen et al. 1997), and the DiGeorge/velo-cardio-facial syndrome on 22q11 (VCFS [MIM 192430]) (Edelmann et al. 1999).

Although most of 15q alterations are concentrated at the pericentromeric region, specifically on 15q11–

[4]Present address: Instituto de Investigaciones Biomédicas, CSIC, Madrid, Spain.
[5]Corresponding author.
E-MAIL estivill@iro.es; FAX 34-93-2607776.
Article and publication are at www.genome.org/cgi/doi/10.1101/gr.155601.

11:98–111 ©2001 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/01 $5.00; www.genome.org

q13, other regions on the long arm of human chromosome 15 also show rearrangements associated with human disease traits. These include translocations and deletions involving the 15q24 band (Bettelheim et al. 1998; Jewett et al. 1998); tetrasomies from 15q23–q24→qter (Blennow et al. 1994); partial monosomies from 15q26.1→qter (Chen et al. 1998); interstitial deletions of 15q25 (Verma et al. 1996); interstitial duplications (Han et al. 1999; Browne et al. 2000); and recurrent translocations between 15q25 and chromosome 12p13 present in congenital (or infantile) fibrosarcoma (CFS [MIM 600618]; [MIM 191316]). The recurrent t(12;15)(p13;q25) rearrangement in CFS fuses the *ets variant 6* gene (*ETV6* or *TEL* oncogene) on 12p13 to the *neurotrophin-3 receptor* gene (*NTRK3* or *TRKC*) on 15q25 (Knezevich et al. 1998).

To characterize the molecular basis of rearrangements involving the 15q25–q26 region we have constructed a bacterial-clone-based contig from 15q25.3 to 15q26.1. This contig spans ~2 Mb, from the *NTRK3* gene to the proximal end of the Bloom's syndrome region, and contains similarities with other chromosomes (6q, 7p, and 12p) and with other regions of 15q (15q11–q13 and 15q24). A low copy repeat sequence of 13–22 kb is present on 15q26.1, 15q24, and 15q11–q13 chromosome regions (LCR15–1, LCR15–2, and LCR15–3, respectively). At least 10 copies of similar LCR15 elements are present on 15q. We propose that these regions of similarity could be involved in chromosome rearrangements affecting the distal portion of human chromosome 15q.

RESULTS

Contig Assembly

As an initial step to construct a bacterial-clone-based contig on 15q25.3–q26.1 we screened BAC and PAC libraries with multipoint STSs spanning ~1.7 Mb according to the Whitehead Institute Radiation Hybrid Map (http://carbon.wi.mit.edu:8000/cgi-bin/contig/phys_map). Eight sequences obtained by PCR were gel purified and used as radioactive probes for this initial purpose: *NTRK3* 3′ end (*WI-30075*), *D15S116*, *D15S202*, *CHLC.GCT14A01*, *D15S736*, *WI-7222*, *SHGC-34665*, and *D15S1082*. Positive clones were colony isolated and confirmed by PCR and hybridization analysis with each probe. From this initial manually assembled contig, different clones were selected to generate new STSs from end sequences to cover the contig gaps. To achieve this, DNA preparations from 22 BAC/PAC clones were end-sequenced as described in the Methods section. Twenty-six of these new sequences (Table 1) were used to assemble the contig shown in Figure 1A. BAC/PAC end sequences were also obtained from public databases (RPCI-11–139l4; GenBank AQ384390 and AQ384392). We also generated two additional STSs from the arms of YAC CEPH-802b4 to confirm the orientation of the initial contig within the *NTRK3* gene region. A total of 44 STSs were analyzed to complete the contig assembly, with 89 BAC/PAC clones and 5 YAC clones (Fig. 1A). Assembly of BAC and PAC clones was also ascertained by hybridization.

The minimal overlap set was of 17 BAC/PAC clones with a gap at the distal end of the contig, between markers *330n12-SP6* and *WI-20237*. This gap is located just before a region showing multiple similarities with other chromosome regions (Fig. 1A). On the basis of the radiation hybrid map and the distance covered by BAC/PAC clones on this region, the size of this gap should be <50 kb. Evidence on the difficulty in cloning this region is the fact that three RPCI-11 clones (416g15, 1069h13, and 3197p17) share the same sequence at the distal end of the gap (GenBank AJ400621, AQ155977, and AQ684334), whereas they span different regions on the opposite end. Furthermore, this gap is still uncovered by the public contig at Washington University (URL: http://genome.wustl.edu/gsc/index.shtml), which is only based on fingerprinting analysis and it presents clones that contain chromosome 6 markers such as RPCI-11–427m11. The analyzed region spans ~2.1 Mb, 1.9 Mb of which are formed by at least four *Not*I fragments (Fig. 1A). To these *Not*I distances we should add more than 160 kb at the proximal end of the contig (marked by BAC RG-1) giving rise to the total distance of ~2.1 Mb. This distance is close to the 1.7 Mb estimated from the Whitehead Institute Radiation Hybrid Map, using markers that are within the physically mapped region (334 cR to 341 cR; 1 cR ~240 kb).

15q25.3–q26.1 Transcriptional Map

Through the study of region 15q25.3–q26.1, 10 genes and 3 UniGene clusters have been localized (Fig. 1A). These genes are: *neurotrophic tyrosine kinase receptor type 3* (*NTRK3*) gene, *adaptor-related protein complex 3 sigma 3 subunit* (*AP3S3*) gene, *alanyl aminopeptidase* (*ANPEP*) gene, *IRO42138* (mRNA full-length codifying for a protein of unknown function), *IRO25206* (mRNA full-length codifying for a protein of unknown function), *DNA polymerase gamma* (*POLG*) gene, *retinaldehyde-binding protein* (*CRALBP*) gene, *osteoblast hyaluronan-binding protein* (*OE-HABP*) gene, *aggrecan 1* (*AGC1*) gene, and *IQ motif containing GTPase-activating protein 1* (*IQGAP1*) gene. The UniGene clusters are: *Hs.230269* (located at marker *576n19-R*), *Hs.9598* (located at marker *WI-20237* and similar to *semaphorin C* gene from *Mus musculus*) and *Hs.177472* (located at marker *SHGC-34665*).

BLASTN search on genomic sequences from the McDermott Center (pDJ10k5, pDJ105i19, and pDJ68d5 clones) revealed a series of similarities with cDNAs,

Table 1. Amplimers Developed from Direct Sequencing of BAC, PAC, and YAC Ends

Sequence	GenBank acc. no.	F (5′-3′)	R (5′-3′)
RG1-F	AJ286911	AGTTTCCTCACAGTGCATGC	CAGGCTCAGATGCAGGTAC
RG1-R	AJ286912	TTCACTTTTCCTGGAGCTGC	AGACCTAGGACACAAGGATG
10d9-R	AJ277870	CTGAAAGATATCAGCACCTTG	TGGTCATAGCCATCCATGTAG
16o12-T7	AJ400818	GAGAGATATTGGTCAAAGGG	CCTCTCAGCAAATTTAAGCATC
30c6-F	AJ286924	CTTATGTAACACAGTATCTGTGG	GGTCTTATAGTCTTCAGTGG
30c6-R	AJ286925	TCAGAGTCACCCATCTAAGG	GGGAGCCACTTTAAGCTCC
104F6-F	AJ400820	GGGCAGTAACAGGAACAGG	TTTATCTGCGTGGCCTTTGC
104F6-R	AJ400819	TTACTCTGAATTCTAAATTTCC	TTGCCCAGCTGGGTTCTGG
167c4-R	AJ286928	CATCTCCAGCTCAACCTCATAGG	ATGGGAGACAGTGCCAGG
309j8-T7	AJ400817	TAGCAGGTGACAGCTCCATG	GTTAAAAACTAAGACGCAAATGC
330n12-T7	AJ276449	CGAAATGAGACTGGAAGTTGGC	CTTGTTCAGGAACAAACCAGCC
330n12-SP6	AJ276448	TCAGAACTGAAGTGTATGGATC	AGTTATGTTCCTCACTTGATGC
463e12-F	AJ286921	TCACTGTGACTTCTCACTGC	GGAATAGGAAGTCATATCAGG
463e12-R	AJ286922	GCCCTGCTCCATGACCTG	CTCTGGCATTCTCCTCAAGC
REP471	AJ286913	GCACTTAAAGCTCTTTACTGG	TGGTTTTCAATGATTGCAAGTC
567n19-F	AJ286926	CAGGACTGCAATTCCTCCAG	ATACTGGAATTGCTCAGAGG
576n19-R	AJ286927	GACTTCTGGATGGGGTATGG	GAGACAGATCTACTAGGTCC
10k5-T7	AJ286894	GCACACTGCAGCAGAGGT	GTGTCCTGGGACTGCCAAG
10k5-SP6	AJ286895	CTTCTCAGGCAACCCTACC	CTCATCTCCTGGCTCACTG
68e5-T7	AJ286896	TTAGGACTAGGTTGTTCTCAC	TTGGGTACCTGTCACTCTG
68e5-SP6	AJ286897	CATTGAACAACATGTACAAGG	GCCCTTGGTAAAAGAGTGAG
76a14-T7	AJ286899	ACAATTCTGATGCCAGTGTTC	GAGGAGTCATAGCAGCCAC
76a14-SP6	AJ286898	AGTTCTACAAGAGCAGGGAC	AATCTTCTCTAGCTACTCAG
80n21-T7	AJ286907	ATAGAAACTTCAGGTCCTTGC	TTGCCTCCATCCTTAGAAGC
80n21-SP6	AJ286906	CTTACAGTAGTCTTCTGGATC	CCAAGTTCAGCTGAGGCTG
103k15-T7	AJ286903	TACTGAGAGGTGACATGCC	CTGAAGGCACCAACAGATGC
103k15-SP6	AJ286902	TCGCCCAGTCTCCTCATG	TCCCCTGAACCTATAAGC
126h9-T7	AJ277872	GAAGCTCTCCTGAACCCTC	GTTCATTTAATCCTCACCATCAC
142g11-T7	Designed from pDJ443n8 (AC004587)	GTTCCACTTCAAACTTCATGG	AGACAGGGTTAACAGCTACG
142g11-SP6	AJ286900	TTACTGAAATCAATTACTCAC	CTACTTTGGGGACACTATCC
216i14-T7	AJ286905	ATTCTCATCCAAATCCACCTG	GTGTTGTGAGAGACCACGTG
216i14-SP6	AJ286904	GTATCCCAAATCCCGTCTTG	GGAAAATCTTTCTTGGAGATC
223e10-SP6	AJ286916	AATGGAATTAGAAATTGCATAGAG	GTAAGTTGTAAACTTAGGTTGATG
252a23-T7	AJ286909	ACCAAAGGGGCAAAGAGCAG	CGATTTTCTTAGGGTGTGTATC
252a23-SP6	AJ286910	ACCTTGGAAGCCCTTAATCC	TCCTTATATGAGGCTGGTCC
802_b_4-Left	AJ286893	GTCTTGCTAGAGCTGAGG	CATGTTGACCACGACCAG
802_b_4-Right	AJ286892	GAATTCTCTGGCCCAGCTGAC	CCTTCACTTAAAAACAGCC

UniGene clusters, or ESTs that could be mapped on the analyzed region. These similarities range from 70% to 100% of identity. Identical sequences, representing probably real putative genes, with similarities at different positions on analyzed pDJ genomic sequences, representing therefore different exons, were: *Hs.136313*; *Hs.99364* (*HS1–2 putative transmembrane protein* gene); and *Hs.6673* (Fig. 1A). Lower sequence similarities were for 6 UniGene clusters and 10 ESTs within pDJ10k5 sequence; 4 UniGene clusters and 2 ESTs within pDJ105i19 sequence; and 1 gene (*NDUFA3*), 6 UniGene clusters, and 5 ESTs similarities within pDJ68d5 sequence (not shown on Fig. 1A and available from the authors).

Similarities Between Sequences of Chromosome 15q25.3–q26.1 and Chromosomes 6, 7, and 12

To evaluate the putative involvement of the 15q25.3–q26.1 region in chromosome rearrangements, we first analyzed all the clones on the contig for the presence of mariner elements (*Hsmar2*), as these sequences have been found associated to chromosomal reorganizations (Reiter et al. 1999). No positive clones were revealed by hybridization using a *Hsmar2* consensus primer sequence.

We also studied BAC, PAC, and YAC clones of the STS contig by FISH. These results are shown in Figure 1A, with nine BAC/PAC/YAC clones (shown in yellow) mapping only on 15q25.3–q26.1, five PAC clones (HGMP-5c5, 173b6, 217b13, 251c1, and 288m22; shown in a red rectangle) hybridizing only on 6q12–13, and two PAC clones (HGMP-142g11 and 143g18; shown in a green rectangle) hybridizing to two regions on the long arm of chromosome 15, bands q11–13 and q26.1.

Hybridization screening of the RPCI-1 PAC library

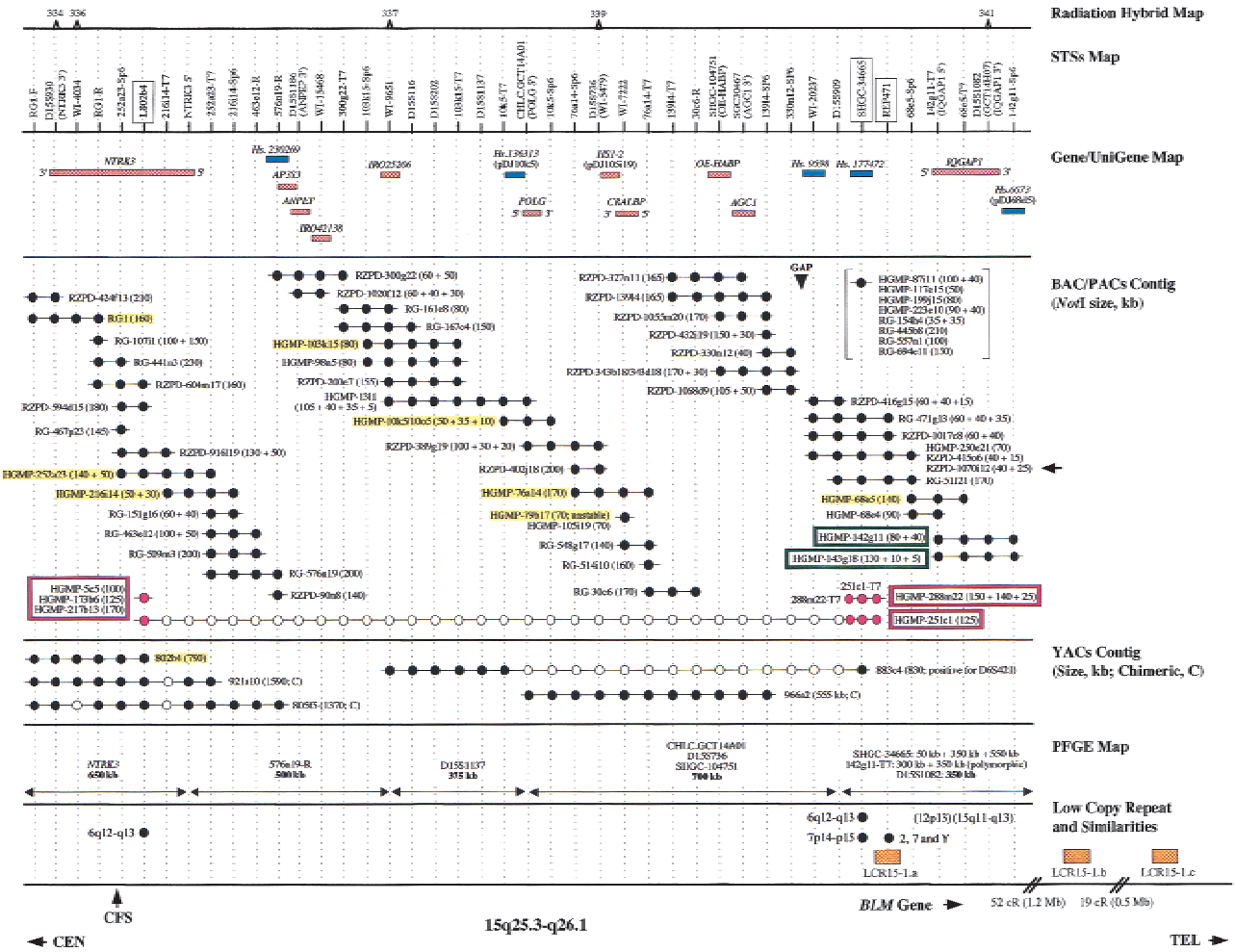


Figure 1 (*A*) Genomic contig of the 15q25.3–q26.1 region. The radiation hybrid map corresponds to the Genebridge 4 panel (Whitehead Institute for Biomedical Research/MIT Center for Genome Research), from 334 cR to 341 cR (1 cR ~240 kb). STSs L802b4, *SHGC-34665* and REP471 are included in rectangles. Genes are shown in red and UniGene clusters in blue. UniGene clusters derived from pDJ clones (10k5, 105i19, and 68d5) are oriented in relation to the 5′ or 3′ ends of genes *POLG*, *CRALBP*, and *IQGAP1*, respectively. Filled and unfilled circles correspond to the presence or absence of markers, respectively. Clones shown in yellow were mapped by FISH on 15q25.3–q26.1; clones in green rectangles were mapped on 15q11–q13 and 15q25.3–q26.1; and clones in red rectangles were mapped on 6q12–q13. An arrow marks clone RZPD-1070i12 (LCR15–1) shown in Figure 4. HGMP, UK Human Genome Mapping Project Resource Centre (RPCI-1 PAC library); RG, Research Genetics (CITB BAC library); and RZPD, Resource Center within the German Human Genome Project (RPCI-11 BAC library). Only the largest BAC/PAC clones and clones mentioned in the text are shown; information about other positive clones for markers of this region is available from the authors. The relative position of the gap within the contig is marked by a filled triangle. A group of clones positive for *SHGC-34665* but probably not mapping to 15q25.3–q26.1 is shown in brackets. PFGE *Not*I restriction fragments detected with different probes are shown as bidirectional arrows. The relative location of the low copy repeats on 15q26.1 (LCR15–1.a/b/c) is shown as an orange box. The LCR15–1.b and LCR15–1.c are located according to the marker content of the respective RPCI-11 clones (groups 7 and 8 in Fig. 5C; *D15S901* for clone 286b10; *WI-14285* and *D15S642* for clones 341b7 and 95f11, respectively). The markers and regions containing sequence similarities with other chromosomes or chromosome 15q regions are shown. The location of the centromere (CEN) and telomere (TEL), the region involved in the congenital fibrosarcoma (CFS) translocation and the orientation of Bloom's disease (*BLM*) gene are indicated. (*B*) Partial genomic contig of chromosome 15q24 and identification of low copy repeat sequences. Clones within boxes were positive for REP471 (LCR15) and are grouped in three nonoverlapping blocks (a, b, and c). An arrow marks clone RZPD-161c1 shown in Figure 4. The relative location of LCR15–2 (a, b, and c) on 15q24 are shown. (Figure continues on following page.)

with the left arm sequence of YAC CEPH-802b4 (L802b4), previously shown to map on 15q25.3 and not being chimeric, yielded five positive clones (5c5, 173b6, 217b13, 251c1, and 252a23), all containing the mentioned STS. FISH analysis showed hybridization only on 6q12–q13 for four of these PACs, whereas 252a23 only mapped on 15q25.3–q26.1. Later, through the screening of the same library with a distal marker (*SHGC-34665*), two positive clones (251c1 and 288m22) were detected, one of them also being L802b4 positive (PAC 251c1). Again, this new PAC clone (288m22) was mapped on 6q12–q13 by FISH, with no

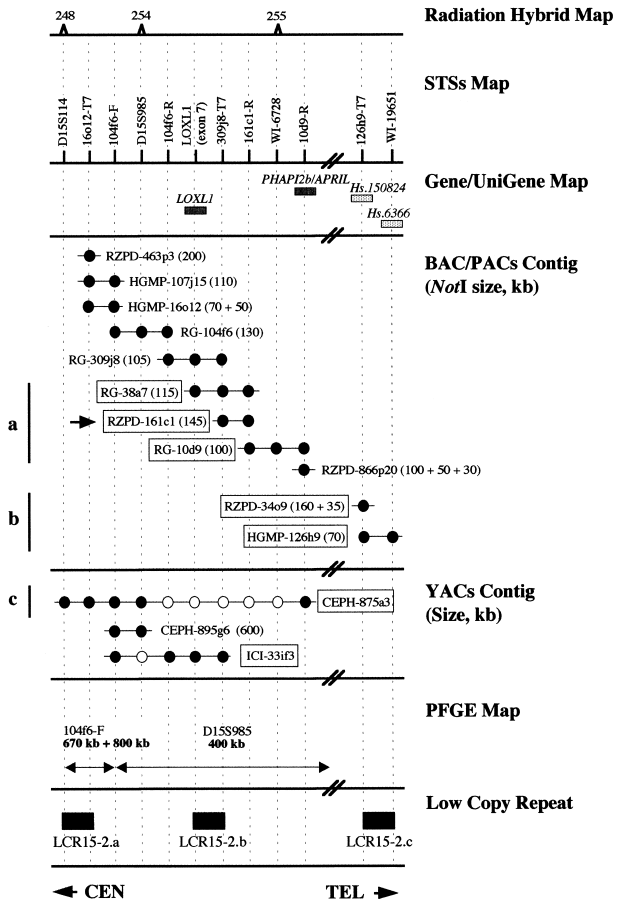


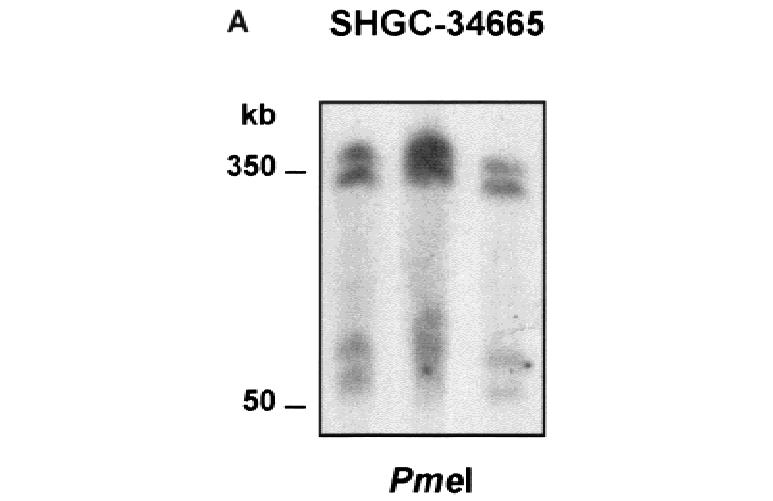
Figure 1 (*Continued*)

signals on 15q25.3–q26.1. The presence of this similarity between 6q12–q13 and 15q25.3–q26.1 is not easily explained by chimerism of the clones: they only map on 6q12–13, with no signals on 15q25.3–q26.1; they do not have the same insert sizes and therefore, are not identical; and two of them overlap through their T7 ends, but these T7 end amplimers do not show positive amplification on clones really anchored on the 15q25.3–q26.1 contig (Fig. 1A). The presence of similar sequences between 6q12–q13 and 15q25.3–26.1 chromosome regions is also reflected in data at GDB showing amplification of marker *D6S421* at 6q12–q13 with CEPH-883c4 clone at 15q25.3–q26.1.

The similarities presented between chromosome 6 and chromosome 15 were also revealed by PCR analysis on somatic cell hybrids (Dubois and Naylor 1993). This approach confirmed that both chromosomes share the L802b4 sequence. In addition, PCR analysis of marker *SHGC-34665* also confirmed the existence of this sequence on chromosomes 6 and 15, but also on chromosome 7. This is in agreement with the observation of more than two copies of *SHGC-34665* in the human genome revealed by PFGE analysis. Thus, in a *Pme*I restriction blot of unrelated human DNA samples, this marker showed four different fragments between 50 and 350 kb (Fig. 2A). The evidence of different chromosome localizations for *SHGC-34665* was further confirmed by the existence of at least eight additional BAC/PAC clones obtained from two different libraries, none containing 15q25.3–26.1 markers, although they cover a relatively large DNA sequence (clones shown in brackets in Fig. 1A). With regard to the *SHGC-34665* sequence on chromosome 7, because forward arm of BAC clone RG-471g13 (further named REP471) is similar (90% identity) to a sequence belonging to a PAC clone deposited at GenBank that maps at 7p14–p15 (accession no. AC005154), this *SHGC-34665* copy could be mapped on this chromosome 7 region.

To evaluate the level of identity between the sequences on chromosomes 6, 7, and 15, we sequenced PCR products obtained from the positive somatic cell hybrid DNAs containing these human chromosomes and from the different clones mapping to chromosomes 6 and 15. The PCR products of L802b4 sequence were identical between chromosome 6 containing hybrid and clone HGMP-251c1 at 6q12–q13, and between chromosome 15 containing hybrid, clone HGMP-252a23 at 15q25.3 and CEPH-802b4 clone at 15q25.3, respectively. These two groups of L802b4 sequences, corresponding to chromosomes 6 and 15, were aligned with the CLUSTAL W program, giving rise to an identity of 43.64% between them. If these two different L802b4 sequences have a common origin, a relatively high number of base substitutions, insertions, and deletions have occurred since they diverged (not shown).

PCR sequences of *SHGC-34665* were identical when compared between chromosome 6 containing hybrid and clone HGMP-288m22 at 6q12–q13, and between chromosome 15 containing hybrid and clone HGMP-250e21 at 15q25.3–q26.1, respectively. Both *SHGC-34665* sequences were different from that derived from chromosome 7-containing hybrid. The calculated identity between the three *SHGC-34665* sequences was 79.56% (83% between chromosomes 6 and 7; 85% between chromosomes 6 and 15; and 89% between chromosomes 7 and 15). On the contrary to L802b4 sequences, relatively few changes have occurred on *SHGC-34665* sequences since they diverged (Fig. 2B). If we assume that *SHGC-34665* sequences represent pseudogenes, the time since chromosomal interchanges occurred could be calculated by the Kimura two-parameter model (Kimura 1980) with an estimated substitution rate for pseudogenes of 2.2×10^{-9} substitutions/bp per year (Eichler et al. 1999) (the relatively high number of deletions or insertions between chromosomes 6 and 15 L802b4 sequences does not allow this calculation). In that sense, chromosomes 7 and 15 *SHGC-34665* sequences could have diverged 24




```
                10        20        30        40        50        60
                |         |         |         |         |         |
Hsa7    GCCTGGATTTCTGCGTATAGATAAGCCCCCTTGGAATAGAAATAAGTGATGAGCCTGGAC
Hsa15   GCCTGGATTTCTGCATATGGATAAGCCACCTTGGAATAGGAAGAGGTGTTGAGCCTGGAC
Hsa6    GCCTATATTTCTGTGTGTGGATAAGCCACCTTAGATTAGCAAGAAGTGATGAGCCTGGAC
        ****  *******  * * ******** **** ** *** ** * *** ***********

                70        80        90        100       110       120
                |         |         |         |         |         |
Hsa7    TGTGGGAGGAAAGAGCTGCGTGGATAGATTCAAACTTCCTGCAAAAGTGCTCCCAGTCTG
Hsa15   TGTGGGAGGAAAGAGCTGCGTGGATAGATTCAAACTTCCTGTGGTAGTGCTCCCAGTCTG
Hsa6    TGTGGGAGGAAAAAGCTGTTTTGGGACATTCAAACTTCCTGTGTCAGGGCTCCCAGTCTG
        ************ *****  * *  * **************    ** ************

                130       140       150       160       170       180
                |         |         |         |         |         |
Hsa7    AGCTCTGTAGATCTTCAGCACTCACTTCTTCTTGCATGGGCTCTCTGTATGTTGAAAACC
Hsa15   ACCTCTGTAGACCTTCAGTACTCACT-CTTCTTGCTTAGGCTCTCTGTGTGTTGAAAGCC
Hsa6    ACCTCTGTAGATCTTCAGTACTCACC-CTTCATGCTTGGGCTCACTATGTGTTGAAAACC
```

Figure 2 Similarities of sequences from region 15q25.3–q26.1 with other chromosomes. (*A*) PFGE hybridization of probe *SHGC-34665* against high molecular weight DNA from unrelated subjects from the general population. Four bands were detected with probe *SHGC-34665* in a *Pme*I digestion. (*B*) CLUSTAL W alignment of sequences corresponding to marker *SHGC-34665*. Asterisks show matched nucleotides. The identity between *SHGC-34665* sequences was 79.6%.

mya, chromosomes 6 and 15 sequences 38 mya, and chromosomes 6 and 7 sequences 42 mya. These data and the fact that the *SHGC-34665* sequence from chromosome 7 has an ORF allow us to postulate that an ancestral *SHGC-34665* sequence from chromosome 7 jumped to chromosome 15 and then to chromosome 6.

BLASTN analysis of other genomic sequences of this contig, distal to marker *SHGC-34665*, also uncovered other similarities. NIX analysis of sequence from pDJ443n8 clone on 15q26.1 (from McDermott Center contig that overlap with pDJ68d5 clone and contains the *IQGAP1* gene; GenBank accession no. AC004587) showed similarities with sequences from clones RPCI-11–13c13 and RPCI-11–656e20 on 12p13. This similarity spans 1.85 kb, with an identity of 95%. RPCI-11–13c13 clone contains markers *D12S1916*, *D12S1696*, and *D12S1690*, located at ~2.7 Mb of the *ETV6* gene on 12p13 (from 579 cR to 693 cR on the G3 radiation hybrid panel; Gyapay et al. 1996). It is interesting that the physical distances of this region of sequence similarity on 12p13 and 15q26.1 from *ETV6* and *NTRK3* genes, are very similar (2.7 Mb and 2.0 Mb, respectively), both genes being the target of recurrent rearrangements in CFS.

Similarities Between Sequences of Chromosome Regions 15q11–q13, 15q24, and 15q26 Identify a New Set of Chromosome 15 Duplicons

In addition to similarities with sequences on chromosome 7p14–p15 (see previous section), BLASTN analysis of REP471 (forward arm RG-471g13) revealed an identity of 90% with the RPCI-11–2m12 sequence mapping on 15q24. This BAC clone contains, by NIX analysis, the *lysyl oxidase-like* (*LOXL1*) and *promyelocytic leukemia* (*PML*) genes, and markers *WI-6717* and *D15S1326*. We are currently building a BAC/PAC contig map of the 15q24 region, of which a partial map containing the *LOXL1* gene is shown in Figure 1B. FISH data indicate that genomic clones containing *LOXL1* map to 15q24, in agreement with Szabo et al. (1997). Seven BAC/PAC/YAC clones from 15q24 were positive for amplimers derived from REP471 (Table 1). These seven clones could be grouped in three blocks that do not overlap, defining at least three copies of this sequence at 15q24: (1) RG-38a7, RZPD-161c1, and RG-10d9; (2) RZPD-34o9 and HGMP-126h9; and (3) CEPH-875a3 (see Fig. 1B).

To determine the extension of the similarity between regions 15q24 and 15q26.1 revealed by REP471 common amplification, we performed a Southern blot analysis. The analysis of clones from the 15q26.1 region with probes belonging to 15q24 (RG-38a7 or RZPD-161c1) allowed us to estimate the extension of the region of similarity at 15q26.1 in a maximum of 22 kb. In the same way, the analysis of clones from 15q24 with a probe belonging to region 15q26.1 (RZPD-1070i12) showed that the extension of this region of similarity at 15q24 extends a maximum of 26 kb (Fig. 3). These regions of similarity were named chromosome 15 low copy repeat 1 (LCR15–1) on 15q26.1, and LCR15–2 on 15q24. LCR15–1 maps within REP471 sequence and extends toward HGMP-

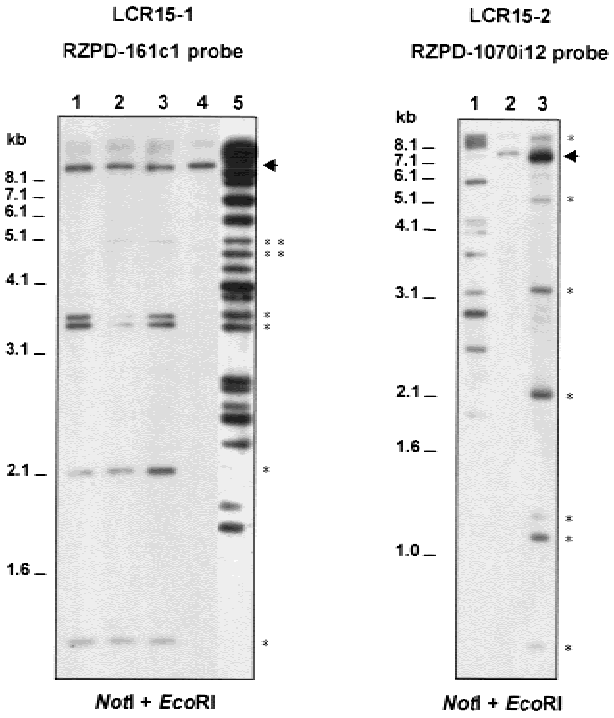


Figure 3 Hybridization of RZPD-161c1 and RZPD-1070i12 full clone probes against *Not*I–*Eco*RI digests of clones mapping on regions 15q26.1 (*left*) and 15q24 (*right*), respectively. For RZPD-161c1 probe: lane *1*, BAC RZPD-1017c8 (15q26.1); lane *2*, BAC RZPD-415o6 (15q26.1); lane *3*, BAC RZPD-1070i12 (15q26.1); lane *4*, BAC RZPD-416g15 (15q26.1, negative control); and lane *5*, BAC RZPD-161c1 (15q24, positive control). Clone RZPD-416g15 constitutes the negative control as it overlaps at the opposite end of the REP471 sequence (see Fig. 1). Clone RZPD-1017c8 does not hybridize with the two larger low copy repeat fragments (**), further confirming the distal orientation of this repeated sequence from REP471 sequence. For probe RZPD-1070i12: lane *1*, BAC RG-471g13 (15q26.1, positive control); lane *2*, BAC RZPD-215f12 (BAC clone mapped on 15q24 that does not show amplification with REP471 sequence; negative control); lane *3*, BAC RZPD-161c1 (15q24). Hybridization fragments that constitute the two LCR15s are marked by asterisks. The positive fragments add up to 20 kb for RZPD-161c1 and 26 kb for RZPD-1070i12, defining the estimated sizes of the LCR15 at 15q26.1 and 15q24, respectively (to the 20 kb we have to add 2 kb detected with RG-471g13; not shown). An arrow marks the fragment that corresponds to the vector hybridization.

68e5 clone (LCR15–1.a on Fig. 1A). LCR15–2 maps close to the *LOXL1* gene on 15q24. Two other copies of the LCR15–2 could be present on 15q24 as REP471 sequences are present in three groups of nonoverlapping clones at this region. BAC/PAC clones containing LCR15–1 and LCR15–2 were used as probes in FISH analysis. Positive hybridization for the two bands 15q26.1 and 15q24 was detected using either probe, further confirming the common identity between these two regions. Interestingly, BAC RZPD-161c1, containing LCR15–2, also showed a FISH signal on 15q11–q13 (Fig. 4A,B).

Southern blot analysis of human genomic DNA revealed several fragments of hybridization with probe REP471 (Fig. 5A). The analysis of a complete chromosome panel (Dubois and Naylor 1993) revealed two copies of REP471 on chromosome Y, one on chromosomes 2 and 7, and at least eight copies on chromosome 15 (Fig. 5B). BLASTN search of REP471 sequence on public databases showed identities between 85% and 100% with genomic sequences located at human chromosomes 3, 7, 10, 15, 18, and Y. However, the majority of the entries correspond to clones that have been mapped to chromosome 15. Thus, at least 14 different RPCI-11 BAC clones in the HTGS database division, known to map on chromosome 15 (one being clone RPCI-11–2m12), have similarities with the REP471 sequence. Moreover, three of these RPCI-11 clones (152f13, 156n7, and 624n5) have two copies of REP471. The 17 sequences from chromosome 15 showing REP471 similarities were aligned with the CLUSTAL W program. This analysis revealed a common identity of 30% and showed that these 17 REP471 similarities could be grouped into 10 different sequences on human chromosome 15 (Fig. 5C). The original REP471 sequence was identical to a sequence belonging to clone RPCI-11–697e2. By NIX analysis we confirmed that this clone maps on 15q26.1, as it contains the same markers as RG-471g13 (*WI-20237*, *D15S909*, and *SHGC-34665*) from which REP471 was derived. All the other sequences showed several nucleotide changes, with identities between 88% and 95%, with the exception of clones 483e23, 291o21, and 546i14, lacking relatively large stretches of nucleotides.

To evaluate whether clones containing REP471 sequences were included in LCR15 elements, in a similar way as LCR15–1 and LCR15–2, we performed a NIX analysis for all the identified RPCI-11 clones from the HTGS database. These analyses showed that at least 13 kb of a common genomic sequence was present within all of REP471 clones mapping on chromosome 15. Within this 13-kb sequence we found similarities for the *golgin-like protein* (*GLP*) gene (its 3′ end and upstream genomic sequence; GenBank accession nos. AF263742 and AF266285) and for the *SH3 domain-containing protein* (*SH3P18*) gene (except for the first 800 bp; GenBank accession no. U61167). The original sequence corresponding to the ORF of the *GLP* gene was found in clone RPCI-11–44J20 (GenBank accession no. AC012527). This indicates that LCR15–1 and LCR15–2 have the characteristics of duplicons (Eichler 1998). The RPCI-11–44J20 clone contains markers *D15S838* and *D15S1270* from the 15q24 region and probably overlaps with clone RPCI-11–361m10 (group 6 on Fig. 5C). These results confirmed the presence of at least three different copies of the LCR15 on 15q24 and suggest that other LCR15 are also present in the chromosome. The corresponding ORF for the other gene similarity identified within the LCR15 (*SH3P18*)

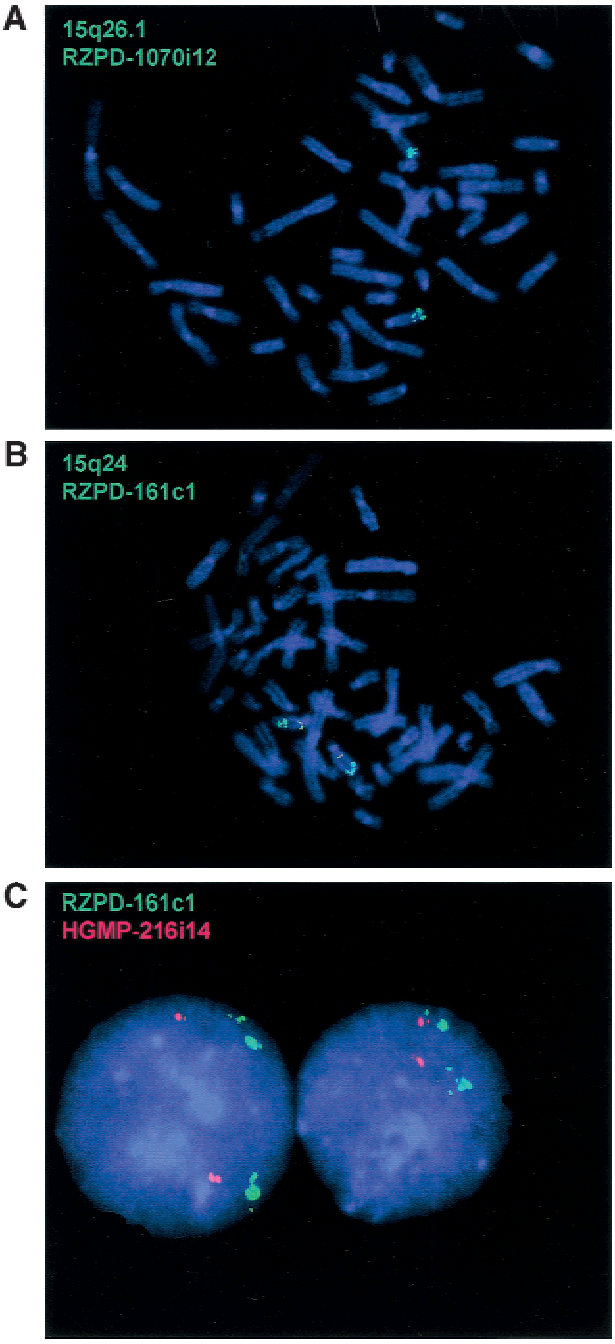


Figure 4 Localization of LCR15–1, LCR15–2, and LCR15–3 on 15q by FISH analysis. Probes used were RZPD-1070i12 (15q26.1) (*A*) and RZPD-161c1 (15q24) (*B*). RZPD-1070i12 clone hybridizes to two regions on 15q, the signal on 15q24 being weaker than on 15q26.1. RZPD-161c1 clone hybridizes on 15q24 and on 15q26.1, but also on 15q11-q13 (weaker signal). (*C*) Interphase nuclei analysis with probe RZPD-161c1 and control probe HGMP-216i14 (15q25.3; see Fig. 1A). The hybridization conditions for all three experiments were identical. At lower stringency conditions other FISH signals on chromosomes 2 and Y could be seen (not shown).

was found in a clone mapped on chromosome 2 (RPCI-11–507m3; GenBank accession no. AC008073). This result is in agreement with the hybridization of REP471 on DNA from somatic cell hybrid containing human chromosome 2. Within the 13-kb minimal LCR15 extension is also included identities higher than 90% for several markers: *SHGC-14665*, *SHGC-82310*, *SHGC-100268*, *SHGC-103176*, *SHGC-104376*, *WI-6362*, and *WI-30306*. None of these markers have been assigned to specific regions of human chromosome 15. Moreover, the LCR15 contained additional duplicon features. LCR15–1 on RPCI-11–697e2 present a G-rich sequence of ~360 bp, a repeated sequence resembling a VNTR $[TAAC(A)(T)_{3-7}(A)_{1-4}TC(T/C)C]_7$ and a TATG repeat. Similar features are also included within the LCR15–2 and LCR15–3 (not shown).

The NIX analysis of positive REP471 RPCI-11 sequences allowed to localize these clones according to their marker content (Fig. 5C). This analysis demonstrate the existence of at least three different copies of the LCR15–1 on 15q26 (groups 1 [LCR15–1.a], 7 [LCR15–1.b], and 8 [LCR15–1.c]); three copies of the LCR15–2 on 15q24 (groups 2, 5, and 6); and an undetected LCR15 element on 15q11–q13 (group 9, LCR15–3). These results are in agreement with previous presented data and revealed the origin of the FISH cross-hybridization signals on 15q11–q13 when probe RZPD-161c1, containing LCR15–2, was used. The RPCI-11 clones 483e23 and 291o21 (group 9, LCR15–3) overlap with clone pDJ778a2 (GenBank accession no. AC004583) located at the distal end of the 15q11–q13 PWS/AS region, contain *D15S1274* and *D15S1276* markers (291o21 clone only) and have a copy of the *HERC2* sequence (*HEct domain* and *RCc1 domain protein 2* gene; Ji et al. 1999; also known as *ERY-1*). The *HERC2* sequence is localized within the duplicons responsible for PWS/AS rearrangements (also known as *END* repeats; Amos-Landgraf et al. 1999; Christian et al. 1999). When we used the 15q26.1 RZPD-1070i12 clone containing the LCR15–1 in FISH experiments, we only observed signals on 15q11–q13 at low stringency conditions (data not shown). This apparent discordance with the 15q24 RZPD-161c1 probe could be explained by the lesser similarity between LCR15–1 and LCR15–3 or by the fact that LCR15–3 includes additional sequences that are also present on 15q24. FISH analysis in interphase nuclei with the RZPD-161c1 probe showed multiple but clustered signals, proving the existence of multiple LCR15 copies on chromosome 15 (see Fig. 4C).

In addition to the LCR15–3 duplicon detected on 15q11–q13, two PAC clones at the distal end of the contig, HGMP-142g11 and HGMP-143g18, showed positive FISH hybridization signals on two regions of 15q (q11–13 and q26.1; green rectangles in Fig. 1A). These clones do not contain the LCR15–1 element. The

hybridization signals of these clones at 15q11–q13 were weaker than at 15q26.1 (data not shown), indicating that they map on 15q26.1 and suggesting the existence of similar sequences within the 15q11–q13 region. Southern blot analysis of these two clones with probes from the 15q11–q13 region (*SNRPN* pseudoexon u1C [Färber et al. 1999]; *GABRA5* exon 1 [Ritchie et al. 1998]; *MN7*/*D15F37*, kindly supplied by Dr. Horsthemke [Buiting et al. 1998]; and *D15S114* [Korenberg et al. 1999]) failed to detect hybridization signals (not shown). Because most pDJ443n8 and pDJ68d5 clones from the McDermott Center 15q26.1 contig, containing the *IQGAP1* gene, are sequenced and correspond to the HGMP-142g11/143g18 covered region, a NIX analysis was performed. A sequence similarity was detected within clone pDJ276c12 that maps on 15q11–q13 (GenBank accession no. AC004737). This similarity spans ~5 kb, distributed in different fragments, the longest one being of 563 bp with an identity of 100%. The pDJ276c12 clone contains markers *D15S63* and *D15S128* and upstream 1C and 1D exons of the small nuclear ribonucleoprotein N (SNRPN), and therefore, it is included in the PWS/AS region, at a maximum of 130 kb from the imprinting center (IC) (Färber et al. 1999).

DISCUSSION

We present here the first BAC/PAC contig of the 15q25.3–q26.1 region and the identification of two regions within the contig that contain similarities with other chromosomes and other regions of chromosome 15. The contig has a minimal overlap set of 17 PAC/BAC clones covering ~2 Mb, 16 genes or UniGene clusters, and a total of 26 new STSs. This contig should be a valuable resource for the study of the molecular basis of 15q rearrangements involving this region, including CFS, and for the isolation of potential genes involved in different human disorders mapping to this region. In this regard, two disorders have recently been mapped to the 15q25–q26 region (a new locus for autosomal recessive hypercholesterolemia [MIM 603813]; Ciccarese et al. 2000; and a locus for autosomal dominant pyogenic arthritis, pyoderma gangrenosum, and acne syndrome, [MIM 604416]; Yeon et al. 2000). At present, we are not able to assign candidate sequences from mapped genes/UniGene/ESTs for these disorders.

The molecular basis of recurrent t(12;15)(p13;q25) rearrangements fusing the *ETV6* and *NTRK3* genes in CFS is unknown. We have found a sequence of 1.85 kb in length with a high degree of identity (95%) between 12p13 and 15q26.1, at similar distances from the *ETV6* and *NTRK3* genes, respectively. Although it seems unlikely that these two relatively short regions of similarity, located Mb apart from the region of the CFS translocation, would play a major role, this cannot be ruled out. It is possible that additional similarities between these chromosome regions are necessary to promote misalignments that lead to rearrangements between these genes.

The contig presented here harbors regions of similarity with other chromosomes in addition to 12p13. Two regions of this contig (one on 15q25.3 and the other on 15q26.1) showed sequence similarities (165 bp and 180 bp, with identities of 44% and 85%, respectively) with chromosome 6q12–q13, and 80% sequence identity with chromosome 7. These similarities could be the result of one or several independent chromosomal exchanges, with or without considerable sequence reorganization between them. It is unknown at this stage whether these sequence similarities belong to transchromosomal duplicons, but, interestingly, the two regions of sequence similarity on chromosome 6 are located at the pericentromeric region, which is usually enriched in duplicated sequences (Eichler et al. 1999).

The most striking similarities detected here for the 15q26.1 region correspond to the 15q11–q13 and 15q24 bands. A 15q11–q13 sequence similarity was localized at the PWS/AS region, relatively close to markers *D15S63* and *D15S128*, and upstream from exons 1C and 1D of the *SNRPN* gene. This region contains the 15q11–q13 maternal/paternal imprinting center (IC) (upstream of exon 5 of *SNRPN* in AS and within exon 1 of *SNRPN* in PWS; Färber et al. 1999 and for review, see Mann and Bartolomei 1999). The 15q26.1 sequence of similarity with the PWS/AS region maps close (2.7 kb) to the 3′ end of the *IQGAP1* gene on 15q26.1. This fact prompted us to analyze the methylation status of the *IQGAP1* gene. This analysis showed complete methylation of this region of DNA from peripheral blood lymphocytes with no maternal/paternal differences (not shown).

The 15q26.1 region also contains similarities with regions 15q11–q13 and 15q24 on the basis of a low copy repeat sequence (LCR15). This repeat is within a 250-kb genomic region on 15q26.1, also containing similarities with chromosomes 6q, 7p, and 12p, and the 15q11–q13 similarity described above. The size of LCR15 was estimated to be a maximum of 22 kb (15q24) to 26 kb (15q26.1) by Southern blot analysis. Several copies of this sequence were found in databases of human sequences, mostly on chromosome 15. Thus, we have found at least 10 different copies of this sequence on chromosome 15 and have localized one copy close to the distal 15q11–q13 *HERC2* sequence, three copies around the *LOXL1* gene on 15q24, and three copies on 15q26, one of them close to the *IQGAP1* gene on 15q26.1. The analysis of these sequences has defined the minimal extension of the region of similarity in 13 kb, with identities >90%. Because sequences of genomic regions containing LCR15

are still not complete, the final length of the LCR15 is estimated between 13 and 22 kb.

A large polymorphic fragment linked to *PML* and *LOXL1* genes has been described in the general population (Goy et al. 1995, 2000). Size differences between alleles are due to the putative insertion/deletion of a sequence of ~30 kb. This size is not far from the estimated size of the LCR15. These polymorphic fragments contain one or several LCR15 sequences as detected by PFGE with a REP471 probe (not shown).

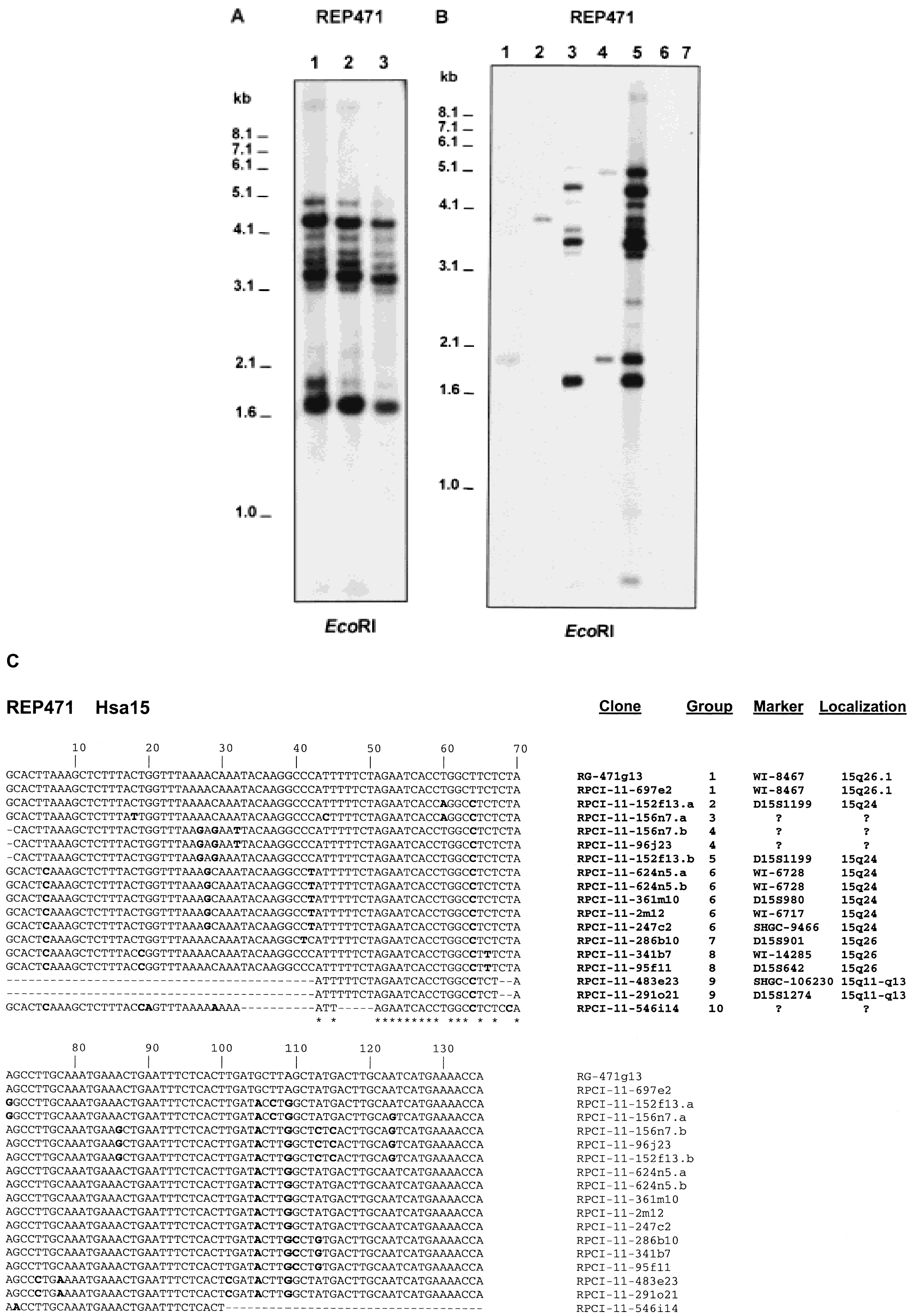


Figure 5 Legend on following page.

However, it is very difficult to prove at this stage whether the repeats themselves are directly responsible for the polymorphism or whether there is another molecular mechanism involved.

Several reports have demonstrated the relationship between the existence of low copy repeat sequences and chromosomal rearrangements in different human genomic disorders (for review, see Ji et al. 2000). These rearrangements are caused by homologous recombination events mediated by the high sequence identity between the low copy repeats. The vast majority of these repeated sequences contain genes or pseudogenes and they have been named duplicons (Eichler 1998). The LCR15 presented in this report is a newly recognized duplicon on chromosome 15. The size of LCR15 is between 13 and 22 kb. The sizes of the duplicons vary for the different regions and genomic disorders, but range from 1.6 kb in Hunter syndrome (Lagerstedt et al. 1997) to ~300 kb in WBS (Peoples et al. 2000) or PWS/AS (Amos-Landgraf et al. 1999; Christian et al. 1999). The presence of genes or pseudogenes within duplicon sequences could promote recombination. It has been proposed that the presence of putative-expressed sequences results in an open chromatin structure that may further stimulate recombination (Chen et al. 1997). The LCR15 described here contains sequences that are highly similar with the *GLP* and *SH3P18* genes.

Because BAC clones containing LCR15–1/2 only show FISH signals on two bands of chromosome 15q (q24 and q26) (LCR15–2 containing clone also show 15q11–13 signals: LCR15–3), it is likely that the different copies of the LCR15 are mainly clustered within these regions (three LCR15 are shown to be around *LOXL1* on 15q24 and three on 15q26). The existence of more than two copies of low copy repeat sequences flanking rearranged regions has been documented for several genomic disorders, DiGeorge/VCFS (Edelmann et al. 1999), WBS (Peoples et al. 2000), and PWS/AS (Amos-Landgraf et al. 1999; Christian et al. 1999). The presence of several copies of low copy repeat sequences within a genomic region provides further complexity to rearrangement configurations. Although most low copy repeat sequences are chromosome specific, the existence of different copies of LCR15 elements in other human chromosomes is not a new feature for a duplicon. As has been detected here for LCR15, the Y chromosome contains duplicon copies of sequences involved in genomic disorders (X-linked ichthyosis and SMS; Li et al. 1992; Chen et al. 1997), suggesting that the Y chromosome is prone to accumulate this type of sequence. The phylogenetic analysis of complete sequences and FISH experiments in chromosomes of nonhuman primates would help in determining the order in evolution of the repeat sequences described here.

The LCR15 elements reported here have the molecular characteristics of duplicons (large size and high sequence identity, recombination-promoting features, and presence of several copies), which could cause genomic disorders involving chromosome 15. Because several genomic disorders could arise from duplicons from the same chromosome region and as there is a large number of LCR15 within chromosome 15, it is possible that some of these LCR15s are involved in genomic mutations. Several cases of 15q reorganizations affecting the most distal portion of this chromosome arm have been associated with different disease traits, including autism (Blennow et al. 1994; Verma et al. 1996; Cook et al. 1997; Bettelheim et al. 1998; Chen et al. 1998; Jewett et al. 1998; Han et al. 1999; Browne et al. 2000). It is tempting to speculate that the LCR15–1/2/3 described here or other LCR15s located within 15q could be involved in some of these rearrangements. In summary, we present here evidence of additional complexity on human chromosome 15q. The bacterial clone-based contig constructed and the identification of duplicons within 15q24 and 15q26 should be a valuable resource for the elucidation of the molecular basis of chromosome reorganizations affecting the distal part of human chromosome 15q.

METHODS

Construction of a Bacterial Clone-Based Map and Low Copy Repeat Analysis

BAC and PAC clones were isolated from several centers: Research Genetics (RG abbreviation; Cat. no. 96055; California Institute of Technology, CITB human BAC library; Kim et al.

Figure 5 Detection of several copies of REP471 on chromosome 15 and other human chromosomes. (*A*) Southern blot analysis of REP471 against human DNA shows a complex pattern with at least 11 fragments. (*B*) Southern blot analysis of REP471 against DNA from somatic cell hybrids containing single human chromosomes in a rodent background. Only those human chromosome lanes with positive hybridization are shown. Lane *1*: DNA from somatic cell hybrid containing human chromosome 2; lane *2*: human chromosome 7; lane *3*: human chromosome 15; lane *4*: human chromosome Y; lane *5*: human genomic DNA from line IMR91; lane *6*: mouse genomic DNA from line 3T6; lane *7*: Chinese hamster genomic DNA from line RJK88. (*C*) CLUSTAL W alignment of 18 REP471 sequences belonging to human chromosome 15 present in the HTGS database division. The 18 REP471 sequences were grouped in 10 groups according to the observed differences (clones 152f13, 156n7, and 624n5 have each two copies of REP471, designated as .1 and .2). The localization of the clones containing the REP471 sequences was based on the position of the respective markers on the G3 and Genebridge 4 ratiation hybrid maps. Bold type indicates nucleotides that are different from the original REP471 sequence of RG-471a13. Asterisks indicate matched nucleotides for all sequences.

1996), UK Human Genome Mapping Project Resource Centre (HGMP abbreviation; RPCI-1, human PAC library originated at Roswell Park Cancer Institute, RPCI, http://bacpac.med.buffalo.edu; Ioannou et al. 1994), and the Resource Center within the German Human Genome Project (RZPD abbreviation; RPCI-11, human BAC library, constructed by Osoegawa and Tateno; Osoegawa et al. 1998) by radioactive hybridization screening. YAC clones were obtained from the CEPH human library (Albertsen et al. 1990) and from the ICI human library (Anand et al. 1990) according to their STS content. Putative positive clones were colony isolated and verified by PCR and hybridization analysis. *NTRK3* 5′ end was analyzed by hybridization screening with a specific primer 5′ GCC GAGCGATCAGATGCAAAATCCTTCAGCGT-3′. Clones were assembled based on their STS content and new STSs were developed from end sequences. Assembly of clones was also ascertained by direct hybridization between clones (protocol based on Wapenaar et al. 1994 and Kern and Hampton 1997): briefly, BAC/PAC DNA minipreparations were labeled with the Megaprime DNA Labeling System (Amersham), denatured and blocked with 0.5 µg/µL of COT DNA (GIBCO BRL) and 0.5µg/µL of $(CA)_{20}$ and $(GT)_{20}$ oligonucleotides in a solution of 6× SSC and 0.5% SDS at 65°C for 5 h. Hybridization conditions were standard in Church's buffer (Church and Gilbert 1984) followed by stringency washes to 0.5–0.25× SSC and 0.1% SDS at 50°–65°C for 30 min. For end-sequencing, BAC/PAC DNA maxipreparations obtained by Qiagen Plasmid Maxi Kit (QIAGEN) were resuspended in TE buffer and sequenced according to the following protocol: a total of 0.5–2.0 µg of purified DNA was fluorescent sequenced with 10 µL of Big-Dye Terminator RR Mix (Applied Biosystems, Inc.), using 10 pmol of BAC or PAC vector primers (5′-GATTACGCCAAGCTATTTAGGTGACACTATAGAATAC-3′ and 5′-CCAGTCACGACGTTGTAAAACGACGGCCAGT GAAT-3′, forward and reverse RG BAC primers respectively; 5′-CACCGGAAGGAGCTGACTGGGTTG-3′ and 5′-GATG TTCATGTTCATGTCTCCTTCTGTATGTACTGT-3′, T7 and SP6 HGMP PAC and RZPD BAC primers, respectively) in a final volume of 25 µL. The thermal cycling parameters were as follows: 96°C for 2 min followed by 70 cycles of 96°C for 10 sec, 50°C for 5 sec, and 60°C for 4 min. The sequence reactions were analyzed on an ABI 377 automated sequencer (Applied Biosystems, Inc.). End-sequences from YAC clones were isolated using vectorette PCR amplification (Riley et al. 1990). Primers for PCR analysis and STS sequences suitable for hybridization screening were analyzed with the RepeatMasker program to avoid human repetitive sequences (A.F.A. Smit and P. Green, unpublished results at http://ftp.genome.washington.edu/RM/RepeatMasker.html). All new sequences generated are available from GenBank (http://www.ncbi.nlm.nih.gov; see above).

FISH Analysis

Metaphase chromosomes were prepared from human peripheral blood lymphocytes. Before hybridization, slides were baked at 55°C for 30 min. Probes, BAC, or PAC DNA minipreparations or Alu–PCR products from YAC clones were labeled with either biotin-16dUTP or digoxigenin-11dUTP (Boehringer Mannheim) and FISH protocol was performed as described elsewhere (Nadal et al. 1997). Slides were studied under a fluorescence microscope (AH3, Olympus) equipped with the appropriate filter set. Images were analyzed with the Cytovision system (Applied Imaging Ltd.).

Transcription Map

Genes and ESTs identified from public maps of the region (GDB and GeneMap '99 URLs: http://gdbwww.dkfz-heidelberg.de/; http://www.NCBI.nlm.nih.gov/Sitemap/index.html#GeneMap) were assayed by PCR with all the clones that form the contig. Other EST were also identified by BLASTN (Altschul et al. 1990) search against the dbEST division with unfinished genomic sequences from the HTGS division of GenBank that matched to our contig (pDJ10k5, GenBank accession no. AC005316; pDJ68d5, GenBank accession no. AC006411; and pDJ105i19, GenBank accession no. AC005318). These genomic sequences belong to the contig draft map of the Bloom's disease region constructed at the McDermott Center (http://gestec.swmed.edu/chromoso5.htm). Clone HGMP-68e5 in our contig and clone pDJ68d5 in McDermott Center contig differ only in the name, they have the same *Not*I insert size and they share the same markers, such as *IQGAP1*. Moreover, clone pDJ68d5 overlaps with two other clones that map on 15q25.3–q26.1 (pDJ250e21, also present in our contig, and pDJ443n8). Probably these two names represent the same clone.

PFGE Analysis

BAC/PAC DNA minipreparations from single colonies were obtained by the alkaline lysis method and digested with *Not*I restriction enzyme (New England Biolabs). Yeast high molecular weight DNA preparations was obtained as described elsewhere (Strughen et al. 1996). Human high molecular weight DNA was prepared from human peripheral blood lymphoblastoid cells lines EBV immortalized (Neitzel 1986). PFGE DNA MW markers were I Lambda-Ladder (Boehringer Mannheim) and DNA Size Standard Yeast Chromosomal (Bio-Rad 170–3605), and a CHEF-DRII Mapper apparatus (Bio-Rad) was used. PFGE agarose gels were depurinated in 0.25 M HCl for 15 min, washed, equilibrated, and blotted to a Hybond-N^+ (Amersham) membrane for 48 h in a solution of 0.4 N NaOH and 1.5 N NaCl. Hybridization conditions were standard in Church's buffer (Church and Gilbert 1984) followed by stringency washes of 0.5–0.2× SSC and 0.1% SDS at 50°–65°C for 30 min.

Similarity Analysis

The DNA source from somatic cell hybrids was a NIMGS 2 Panel (Dubois and Naylor 1993). PCR products were ligated by T4 DNA Ligase (Boehringer Mannheim) into hand-made T-vector (Marchuk et al. 1991) and transformed in to XL1-Blue *E. coli* F′. Plasmid minipreparations were carried out with QIAprep Spin Miniprep Kit (QIAGEN) and fluorescently sequenced using Big-Dye Terminator RR Mix (Applied Biosystems). Sequences obtained were aligned by CLUSTAL W Multiple Alignment (gap opening penalty, 10.0; gap extension penalty, 0.2) (Thompson et al. 1994; http://pbil.ibcp.fr/cgi-bin/align_clustalw.pl) and time since divergence of sequences was calculated based on the Kimura two-parameter model (Kimura 1980) with an estimated rate substitution for pseudogenes of 2.2×10^{-9} substitutions/bp per year (Eichler et al. 1999). Duplicated 15q11–q13 sequences analyzed were: pseudoexon u1C of *SNRPN* (Färber et al. 1999), exon 1 of *GABRA5* (Ritchie et al. 1998), *MN7* (*D15F37*, kindly supplied by Dr. Horsthemke, Institut für Humangenetik) (Buiting et al. 1998), and *D15S114* (Korenberg et al. 1999). The presence of mariner elements was analyzed with a specific *Hsmar2* primer (Reiter et al. 1999). Analysis of relatively extensive genomic se-

quences was performed through the NIX program (G.W. Williams, P.M. Woollard, and P. Hingamp: "NIX: A nucleotide identification system at the HGMP-RC"; http://www.hgmp.mrc.ac.uk/NIX/).

ACKNOWLEDGMENTS

We thank Rafa de Cid for his expertise in establishing immortalized human peripheral blood lymphoblastoid cells. We thank Helena Kruyer for the assistance in preparing the manuscript. We thank the UK Human Genome Mapping Project Resource Centre and the Resource Center within the German Human Genome Project for kindly supplying the clones. This work was supported by La Marató de TV3 (98/1810), European Union (BMH4-CT97–2284), and Spanish Government (CICYT; SAF99–0092-CO2–01). MAP is supported by La Marató de TV3.

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

REFERENCES

Albertsen, H., Abderrahim, H., Cann, H., Dausset, J., and Le Paslier, D. 1990. Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl. Acad. Sci.* **87:** 4256–4260.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Amos-Landgraf, J.M., Ji, Y. Gottlieb, W., Depinet, T., Wandstrat, A.E., Cassidy, S.B., Driscoll, D.J., Rogan, P.K., Schwartz, S., and Nicholls, R.D. 1999. Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. *Am. J. Hum. Genet.* **65:** 370–386.

Anand, R., Riley, J.H., Butler, R., Smith, J.C., and Markham, A.F. 1990. A 3.5 genome equivalent multi access YAC library: Construction, characterisation, screening and storage. *Nucleic Acids Res.* **18:** 1951–1956.

Bettelheim, D., Hengstschläger, M., Drahonsky, R., Eppel, W., and Bernaschek, G. 1998. Two cases of prenatally diagnosed diaphragmatic hernia accompanied by the same undescribed chromosomal deletion (15q24 *de novo*). *Clin. Genet.* **53:** 319–320.

Blennow, E., Telenius, H., de Vos, D., Larsson, C., Henriksson, P.,Johansson, O., Carter, N.P., and Nordenskjold, M. 1994. Tetrasomy 15q: Two marker chromosomes with no detectable alpha-satellite DNA. *Am. J. Hum. Genet.* **54:** 877–883.

Browne, C.E., Hatchwell, E., Protopapos, A., and Ramos, J. 2000. Duplication of medial 15q confirmed by FISH. *J. Med. Genet.* **37:** (online).

Buiting, K., Grob, S., Ji, Y., Senger, G., Nicholls, R.D., and Horsthemke, B. 1998. Expressed copies of the MN7 (D15F37) gene family map close to the common deletion breakpoints in the Prader-Willi/Angelman syndromes. *Cytogenet. Cell Genet.* **81:** 247–253.

Chen, C.P., Lee, C.C., Pan, C.W., Kir, T.Y., and Chen, B.F. 1998. Partial monosomy 8q and partial monosomy 15q associated with congenital hydrocephalus diaphragmatic hernia, urinary tract anomalies, congenital heart defect and kyphoscoliosis. *Prenat. Diagn.* **18:** 1289–1293.

Chen, K.S., Manian, P., Koeuth, T., Potocki, L., Zhao, Q., Chinault, A.C., Lee, C.C., and Lupski, J.R. 1997. Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nat. Genet.* **17:** 154–163.

Christian, S.L., Fantes, J.A., Mewborn, S.K., Huang, B., and Ledbetter, D.H. 1999. Large genomic duplicons map to sites of instability in the Prader-Willi/Angelman syndrome chromosome region (15q11–q13). *Hum. Mol. Genet.* **8:** 1025–1037.

Church, G.M. and Gilbert, W. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci.* **81:** 1991–1995.

Ciccarese, M., Pacifico, A., Tonolo, A., Pintus, P., Nikoshkov, A., Zuliani, G., Fellin, R., Luthman, H., and Maioli, M. 2000. A new locus for autosomal recessive hypercholesterolemia maps to human chromosome 15q25–q26. *Am. J. Hum. Genet.* **66:** 453–460.

Cook, E.H., Jr., Lindgren, V., Leventhal, B.L., Courchesne, R., Lincoln, A., Shulman, C., Lord, C., and Courchesne, E. 1997. Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *Am. J. Hum. Genet.* **60:** 928–934.

Dubois, B.L. and Naylor, S.L. 1993. Characterization of NIGMS human/rodent somatic cell hybrid mapping panel 2 by PCR. *Genomics* **16:** 315–319.

Edelmann, L., Pandita, R.K., Spiteri, E., Funke, B., Goldberg, R., Palanisamy, N., Chaganti, R.S., Magenis, E., Shprintzen, R.J., and Morrow, B.E. 1999. A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum. Mol. Genet.* **8:** 1157–1167.

Eichler, E.E. 1998. Masquerading repeats: Paralogous pitfalls of the human genome. *Genome Res.* **8:** 758–762.

Eichler, E.E., Archidiacono, N., and Rocchi, M. 1999. CAGGG repeats and the pericentromeric duplication of the hominoid genome. *Genome Res.* **9:** 1048–1058.

Färber, C., Dittrich, B., Buiting, K., and Horsthemke, B. 1999. The chromosome 15 imprinting centre (IC) region has undergone multiple duplication events and contains an upstream exon of *SNRPN* that is deleted in all Angelman syndrome patients with an IC microdeletion. *Hum. Mol. Genet.* **8:** 337–343.

Goy, A., Passalaris, T., Xiao, Y.-H., Miller, W.H., Siegel, D.S., and Zelenetz, A.D. 1995. The PML gene is linked to a megabase-scale insertion/deletion restriction fragment length polymorphism. *Genomics* **26:** 327–333.

Goy, A., Gilles, F., Remache, Y., and Zelenetz, A.D. 2000. Physical linkage of the lysyl oxidase-like (LOXL1) gene to the PML gene on human chromosome 15q22. *Cytogenet. Cell Genet.* **88:** 22–24.

Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillett, D., Muselet, D., Prud'Homme, J.F., Dib, C., Auffray, C., et al. 1996. A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5:** 339–346.

Han, J.Y., Kim, K.H., Lee, H.D., Moon, S.Y., and Shaffer, L.G. 1999. De novo direct duplication of 15q15→q24 in a newborn boy with mild manifestations. *Am. J. Med. Genet.* **87:** 395–398.

Ioannou, P.A., Amemiya, C.T., Garnes, J., Kroisel, P.M., Shizuya, H., Chen, C., Batzer, M.A., and de Jong, P.J. 1994. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* **6:** 84–89.

Jewett, T., Marnane, D., Stewart, W., Hayworth-Hodge, R., Finklea, L., Klinepeter, K., Rao, P.N., and Pettenati, M.J. 1998. Jumping translocation with partial duplications and triplications of chromosomes 7 and 15. *Clin. Genet.* **53:** 415–420.

Ji, Y., Walkowicz, M.J., Buiting, K., Johnson, D.K., Tarvin, R.E., Rinchik, E.M., Horsthemke, B., Stubbs, L., and Nicholls, R.D. 1999. The ancestral gene for transcribed, low-copy repeats in the Prader-Willi/Angelman region encodes a large protein implicated in protein trafficking, which is deficient in mice with neuromuscular and spermiogenic abnormalities. *Hum. Mol. Genet.* **8:** 533–542.

Ji, Y., Eichler, E.E., Schwartz, S., and Nicholls, R.D. 2000. Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res.* **10:** 597–610.

Kern, S. and Hampton, G.M. 1997. Direct hybridization of large-insert genomic clones on high-density gridded cDNA filter arrays. *BioTechniques* **23:** 120–124.

Kim, U.J., Birren, B.W., Slepak, T., Mancino, V., Boysen, C., Kang, H.L., Simon, M.I., and Shizuya, H. 1996. Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34:** 213–218.

Kimura, M. 1980. A simple method for estimating evolutionary rates

of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16:** 111–120.

Khan, N.L. and Wood, N.W. 1999. Prader-Willi and Angelman syndromes: Update on genetic mechanisms and diagnostic complexities. *Curr. Opin. Neurol.* **12:** 149–154.

Knezevich, S.R., McFadden, D.E., Tao, W., Lim, J.F., and Sorensen, P.H.B. 1998. A novel *ETV6-NTRK3* gene fusion in congenital fibrosarcoma. *Nature Genet.* **18:** 184–187.

Korenberg, J.R., Chen, X.-N., Sun, Z., Shi, Z.-Y., Ma, S., Vataru, E., Yimlamai, D., Weissenbach, J.S., Shizuya, H., Simon, M.I., et al. 1999. Human genome anatomy: BACs integrating the genetic and cytogenetic maps for bridging genome and biomedicine. *Genome Res.* **9:** 994–1001.

Lagerstedt, K., Karsten, S.L., Carlberg, B.M., Kleijer, W.J., Tonnesen, T., Pettersson, U., and Bondeson, M.L. 1997. Double-strand breaks may initiate the inversion mutation causing the Hunter syndrome. *Hum. Mol. Genet.* **6:** 627–633.

Li, X.-M., Yen, P., and Shapiro, L. 1992. 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.

Lupski, J.R. 1998. Genomic disorders: Structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* **14:** 417–422.

Mann, M.R.W. and Bartolomei, M.S. 1999. Towards a molecular understanding of Prader-Willi and Angelman syndromes. *Hum. Mol. Genet.* **8:** 1867–1873.

Marchuk, D., Drumm, M., Saulino, A., and Collins, F.S. 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* **19:** 1154.

Nadal, M., Moreno, S., Pritchard, M., Preciado, M.A., Estivill, X., and Ramos-Arroyo, M.A. 1997. Down syndrome: Characterisation of a case with partial trisomy of chromosome 21 owing to a parental balanced translocation (15;21)(q26;q22.1) by FISH. *J. Med. Genet.* **34:** 50–54.

Neitzel, H. 1986. A routine method for the establishment of permanent growing lymphoblastoid cell lines. *Hum. Genet.* **73:** 320–326.

Osoegawa, K.P., Woon, Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J., and de Jong, P.J. 1998. An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **52:** 1–8.

Peoples, R., Franke, Y., Wang, Y.K., Pérez-Jurado, L., Paperna, T., Cisco, M., and Franke, U. 2000. A physical map, including a BAC/PAC clone contig of the Williams-Beuren syndrome-deletion region at 7q11.23. *Am. J. Hum. Genet.* **66:** 47–68.

Potocki, L., Chen, K.S., Park, S.S., Osterholm, D.E., Withers, M.A., Kimonis, V., Summers, A.M., Meschino, W.S., Anyane-Yeboa, K., Kashork, C.D., et al. 2000. Molecular mechanism for duplication 17p11.2—the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nat. Genet.* **24:** 84–87.

Reiter, L.T., Liehr, T., Rautenstrauss, B., Robertson, H.M., and Lupski, J.R. 1999. Localization of *mariner* DNA transposons in the human genome by PRINS. *Genome Res.* **9:** 839–843.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J.C., and Markham, A.F. 1990. A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18:** 2887–2890.

Ritchie, R.J., Mattei, M-G., and Lalande, M. 1998. A large polymorphic repeat in the pericentromeric region of human chromosome 15q contains three partial gene duplications. *Hum. Mol. Genet.* **7:** 1253–1260.

Schinzel, A.A., Brecevic, L., Bernasconi, F., Binkert, F., Berthet, F., Wuilloud, A., and Robinson, W.P. 1994. Intrachromosomal triplication of 15q11→q13. *J. Med. Genet.* **31:** 798–803.

Straughen, J., Ciocci, S., Ye, T-Z., Lennon, D.N., Proytcheva, M., Alhadeff, B., Goodfellow, P., German, J., Ellis, N.A., and Groden, J. 1996. Physical mapping of the Bloom syndrome region by the identification of YAC and P1 clones from human chromosome 15 band q26.1. *Genomics* **35:** 118–128.

Szabo, Z., Light, E.,Boyd, C.D., and Csiszar, K. 1997. The human lysyl oxidase-like gene maps between STS markers D15S215 and GHLC.GCT7C09 on chromosome 15. *Hum. Genet.* **101:** 198–200.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Verma, R.S., Kleyman, S.M., Giridharan, R., and Ramesh, K.H. 1996. A de novo interstitial deletion of chromosome 15 band q25 as revealed by FISH-technique. *Clin. Genet.* **49:** 303–305.

Wapennaar, M.C., Schiaffino, M.V., Bassi, M.T., Schaefer, L., Chinault, A.G., Zoghbi, H.Y., and Ballabio, A. 1994. A YAC-based binning strategy facilitating the rapid assembly of cosmid contigs: 1.6 Mb of overlapping cosmids in Xp22. *Hum. Mol. Genet.* **7:** 1155–1161.

Webb, T. 1995. Inv dup(15) supernumerary marker chromosomes. *J. Med. Genet.* **31:** 585–594.

Yeon, H.B., Lindor, N.M., Seidman, J.G., and Seidman, C.E. 2000. Pyogenic arthritis, pyoderma gangrenosum, and acne syndrome maps to chromosome 15q. *Am. J. Hum. Genet.* **66:** 1443–1448.

Received July 13, 2000; accepted in revised form September 18, 2000.