Research

Duplications on Human Chromosome 22 Reveal a Novel Ret Finger Protein-Like Gene Family with Sense and Endogenous Antisense Transcripts

Eyal Seroussi,¹ Darek Kedra,¹ Hua-Qin Pan,² Myriam Peyrard,¹ Charles Schwartz,³ Peter Scambler,⁴ Dian Donnai,⁵ Bruce A. Roe,² and Jan P. Dumanski^{1,6}

¹Department of Molecular Medicine, Karolinska Hospital, 171 76 Stockholm, Sweden; ²Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019 USA; ³Center for Molecular Studies, J.C. Self Research Institute, Greenwood Genetic Center, Greenwood, South Carolina 29646 USA; ⁴Molecular Medicine Unit, Institute of Child Health, London, WC1N 1EH, UK; ⁵Regional Genetics Service, St. Mary's Hospital, Manchester M13 0JH, UK

Analysis of 600 kb of sequence encompassing the beta-prime adaptin (*BAM22*) gene on human chromosome 22 revealed intrachromosomal duplications within 22q12–13 resulting in three active *RFPL* genes, two *RFPL* pseudogenes, and two pseudogenes of *BAM22*. The genomic sequence of *BAM22* shows a remarkable similarity to that of *BAM22*. The cDNA sequence comparison of *RFPL1*, *RFPL2*, and *RFPL3* showed 95%–96% identity between the genes, which were most similar to the Ret Finger Protein gene from human chromosome 6. The sense *RFPL* transcripts encode proteins with the tripartite structure, composed of RING finger, coiled-coil, and B30-2 domains, which are characteristic of the RING–B30 family. Each of these domains are thought to mediate protein–protein interactions by promoting homo- or heterodimerization. The *MIDI* gene on Xp22 is also a member of the RING–B30 family and is mutated in Opitz syndrome (OS). The autosomal dominant form of OS shows linkage to 22q11–q12. We detected a polymorphic protein-truncating allele of *RFPL1* in 8% of the population, which was not associated with the OS phenotype. We identified 6-kb and 1.2-kb noncoding antisense mRNAs of *RFPL1S* and *RFPL3S* antisense genes, respectively. The *RFPL1S* and *RFPL3S* genes cover substantial portions of their sense counterparts, which suggests that the function of *RFPL1S* and *RFPL3S* is a post-transcriptional regulation of the sense *RFPL* genes. We illustrate the role of intrachromosomal duplications in the generation of *RFPL* genes, which were created by a series of duplications and share an ancestor with the RING–B30 domain containing genes from the major histocompatibility complex region on human chromosome 6.

[The sequence data described in this paper have been submitted to GenBank under the following accession nos: AJ010228–AJ010233, AC000025, AC000041, AC000045, and AC002059.]

The ability of artificially produced antisense oligonucleotide/RNA to suppress gene translation of sense transcripts is well documented and widely used in cell biology. Natural antisense RNAs (NARs) are well described in prokaryotic systems. The biological activities of NARs are diverse and affect phage development, transposition, chromosomal gene expression, as well as plasmid replication, compatibility, and conjugation (Wagner and Simons 1994). In all prokaryotic examples studied so far, antisense transcripts were found to down-regulate the expression of sense transcripts. In eukaryotes, natural, endogenous antisense transcription has been shown or suggested to regulate a limited number of diverse genes (for review, see Dolnick 1997; Vanhee-Brossollet and Vaquero 1998). One of the best understood examples is that of the basic fibroblast growth factor (*bFGF*) sense transcript and its antisense

counterpart (*gfg*). The *bFGF* transcripts, which are present in the unfertilized oocyte, disappear shortly after fertilization and are subsequently reexpressed in later stages of embryonic development. The *gfg* has been shown to regulate *bFGF* negatively in *Xenopus laevis* and human oocytes (Kimelman and Kirschner 1989; Knee et al. 1994). In *Xenopus*, the sense and antisense transcripts share 900 bp of sequence at their 3' ends and coexist as double-stranded (ds) RNA duplexes in the cytoplasm of the immature oocyte. The NARs present in 20-fold excess over the sense transcript suggesting that all of the sense transcripts in the unfertilized oocyte may exist as heteroduplexes. Antisense transcription has also been suggested to regulate negatively members of the *myc* family of proto-oncogenes by NARs, which are incapable of producing proteins (Krystal et al. 1990; Robertson et al. 1991). It is believed that NARs carry out post-transcriptional control on endogenous counterpart sense genes, or closely related sequences, by at least three mechanisms: (1) nuclear

⁶Corresponding author.
E-MAIL Jan.Dumanski@cmm.ki.se; FAX 46-8-517 73909.

dsRNA “unwindase” recognizes dsRNA and converts adenosine residues to inosine, which results in A → G conversions. This modification is temporally related to a rapid degradation of sense mRNA, suggesting a role for the RNA duplex in the regulation of mRNA stability; (2) areas of dsRNA prevent normal splicing by protecting the primary transcript from splicing enzyme complex; and (3) by controlling availability of selected forms of sense RNA for translational machinery (Kimelman and Kirschner 1989; Krystal et al. 1990; Lee et al. 1993; Wightman et al. 1993; Dolnick 1997; Vanhee-Brossollet and Vaquero 1998). In the majority of reported antisense regulated genes, however, the molecular effect of NARs remains unknown.

The present study was initiated with the large-scale genomic sequencing and detailed transcriptional analysis of the beta-prime adaptin [*BAM22*, Genome Database (GDB) symbol *ADTB1*] locus on chromosome 22. After our characterization of the *BAM22* gene (Peyrard et al. 1994), we obtained indications that an additional closely related gene may be present in the vicinity of *BAM22*. In the course of the study we detected several intrachromosomal duplications and uncovered a novel family composed of three closely related genes [Ret finger protein-like 1, 2, and 3 (*RFPL1*, *RFPL2*, and *RFPL3*)] that display both sense and antisense endogenous transcripts. We also characterized two *RFPL* pseudogenes and two pseudogenes of *BAM22*. One of the latter pseudogenes, which is located upstream of the active *BAM22* gene, shows remarkable conservation of its genomic sequence.

RESULTS

Sequencing of 22q12–q13 Reveals Two Pseudogenes of *BAM22* Within Two Duplicated Regions

Previously, we have determined the genomic organization and the promoter structure of the *BAM22* gene by sequencing a cosmid containing the 5′ end of the gene (GenBank accession no. L48038; Fig. 1A) (Peyrard et al. 1996). To characterize fully the *BAM22* locus, we sequenced additional genomic clones, for example, cosmid *E42H1* (GenBank accession no. AC000041), which was positive upon hybridization with the probe covering the 5′ end of the *BAM22* cDNA. Sequencing of *E42H1* revealed that it contains exons identical to those of the previously characterized *BAM22*. To resolve the issue whether the *BAM22* locus may contain an additional active β′ adaptin gene, several additional genomic clones, fully covering the *BAM22* region, were sequenced: PAC 704f1059q13 (GenBank accession no. AC002059), BAC 566c1 (AC000025), BAC 58b8 (AC000026), and cosmid N47G11 (AC000035) (Fig. 1A). A sequence contig of 287 kb was edited to 99.99% accuracy. Filtering of repetitive elements (Repeat Masker) and database searches (BLASTN) were carried

out on this sequence. The positions and transcriptional orientations of the seven known genes (*EWS*, *GAR22*, *RRP22*, *BAM22*, *NEFH*, *pK1.3*, and *NIPSNAP1*) are shown in Figure 1A. The exon–intron junctions of the previously cloned *pK1.3* gene (GenBank accession no. L18972) were deduced, and this gene is composed of 20 exons. All introns contain the conserved first and last two bases, gt and ag, for donor and acceptor splice sites, respectively (not shown).

The analysis uncovered a regional duplication of 8.5 kb stretching over exons 2 and 3 of *BAM22* (denoted as exons 2′ and 3′; Fig. 1A). Sequences of both exons 2′ and 3′, which are 63 and 105 bp long, respectively, are identical to their counterparts in *BAM22*. The regions of nucleotide sequence identity stretch into the introns. For instance, exons 2′ and 3′ are embedded within the stretches of 172 and 337 bases, which are identical in both the 8.5-kb duplication and the *BAM22* gene. The overall nucleotide sequence similarity within the 8.5-kb duplicated region is 89.5%. This partial duplication of *BAM22* was named *BAM22* pseudogene 1 (*BAM22ψ1*).

In the course of characterization of the *BAM22* locus, another chromosome 22 cosmid (*E90G5*, GenBank accession no. AC000045; Fig. 1B) was sequenced. During the construction of the genomic contig covering the *BAM22* locus, this cosmid was assumed to be located immediately distal to the *BAM22* gene, as it displayed positive hybridization signal with a genomic probe from cosmid *E42H1* (Xie et al. 1993). However, sequence of *E90G5* was only partially in agreement with those from the genomic clones shown in Figure 1A, which is an indication that a further intrachromosomal duplication of the *BAM22* locus might exist. Recent sequencing results from the region 2–2.5 Mb telomeric to *BAM22*, generated at the Sanger Centre (Hinxton, UK) confirmed this hypothesis. Figure 1B displays a 316-kb sequence contig located telomeric to the human Na⁺/glucose cotransporter 1 gene (*SGLT1*, *SLC5A1*) (Turk et al. 1993), which fully incorporates the sequence from *E90G5* and contains the second pseudogene of *BAM22* (*BAM22ψ2*). This pseudogene is composed of three distinct segments. The first is 1.9 kb similar to the region surrounding exon 6 of *BAM22*, named *BAM22* pseudo-exon 6′. When compared with exon 6 of *BAM22*, exon 6′ contains 10-bp substitutions. The second segment of *BAM22ψ2* shows 3.1 kb of similarity to the region surrounding exon 3 of *BAM22* and exon 3′ of *BAM22ψ1*. This exon was named *BAM22* pseudo-exon 3′′, and it contains 5-bp substitutions when compared with exon 3 of *BAM22*. The position of these two pseudo-exons with regard to each other is also aberrant when compared with *BAM22*. The third segment of *BAM22ψ2* is 25.8 kb sequence with similarity to the *BAM22ψ1*, in the region immediately centromeric to pseudo-exon 3′ of *BAM22ψ1*.

Figure 1 Comparison of two sequence contigs from 22q reveals several intrachromosomal duplications distal to the *BAM22* gene. The sequenced genomic clones are labeled by their EMBL/GenBank accession numbers above the bars. Contig B is located 2–2.5 Mb to the telomere as compared with the other contig (www.sanger.ac.uk/HGP/Chr22) (Collins et al. 1995). (A) Position and transcriptional orientation of eight genes located in the 287 kb contig formed by seven completely sequenced genomic clones. The gene abbreviations are as follows: (*EWS*) Ewing's sarcoma gene (Delattre et al. 1992); (*GARG2*) a gene homologous to the mouse *Gas2* gene (Zucman-Rossi et al. 1996); (*RRP22*) a member of the Ras family of genes (Zucman-Rossi et al. 1996); (*BAM22*) a beta-prime adaptin gene (Peyrard et al. 1994); (*ADTB1*) GDB symbol; (*NEFH*) neurofilament heavy chain polypeptide gene (Lees et al. 1988); (*pK1.3*) a gene of unknown function (Xie et al. 1993); (*NIPSNAP1*) gene (Seroussi et al. 1998); (*RFPL1*) Ret Finger Protein-Like 1 gene; (*RFPLψ1*) Ret Finger Protein-Like pseudogene 1; (*BAM22ψ1*) *BAM22* pseudogene 1. *BAM22ψ1* is a regional duplication of a 8.5-kb genomic segment from *BAM22* that contains identical copies of exons 2 and 3. *RFPL1* gene is transcribed in both directions and the antisense transcript is denoted as *RFPL1S*. The arrows delineate the transcriptional orientation of genes. For the pseudogenes, the direction of arrows denote pseudogene orientations as if these were transcribed. (B) Delineation of the positions and transcriptional orientations of the genes located in the 316-kb contig telomeric to the human Na⁺/glucose cotransporter 1 gene (*SGLT1*, *SLC5A1*) (Turk et al. 1993). The second pseudogene of *BAM22* (*BAM22ψ2*) contains copies of exons 6 and 3 and the orientation of these exons is indicated above the bars. This pseudogene contains also 25.8 kb of sequence similar to *BAM22ψ1*. Contig B contains also *RFPL2*, *RFPL3*, and *RFPLψ2*, which is composed of two exons arranged in opposite orientations, which are indicated by two separate arrows. The transcription of *RFPL3* has been detected only as the antisense transcript, denoted as *RFPL3S*. The exons forming the putative sense and the antisense transcript are numbered and presented above (sense) and below (antisense) the genomic contig.

The *RFPL* Gene Family

Analysis of the genomic sequence (GenBank accession nos. AC002059, AC000025) using the BLASTX program revealed a putative *RFPL1* gene located in the region telomeric to the *BAM22ψ1* (Fig. 1A). We detected 2 exons strongly resembling the B30-2 domain from several proteins and the RING-like motif, which is also present at the amino terminus of B30-2 domain-containing proteins (Henry et al. 1997). On the basis of this genomic sequence, we designed PCR primers (1–8 and 16, Table 1) to characterize the cDNA of the gene. We tested 11 cDNA libraries and detected a 1.5-kb band only in testis (Table 1, primers 1 and 16). Sequencing of this PCR product (GenBank accession no. AJ010229;

Fig. 2) and comparison with the sequence of PAC 704f1059q13 (GenBank accession no. AC002059) revealed that the gene is composed of two exons with an ORF encoding 287 amino acids. Because this gene was similar to the previously characterized human *RFP* gene (BLASTP, 43% identity and 58% similarity with *RFPL1* between residues 94 and 272) and partially shared the protein domain structure (see below), we named it *RFPL1*. The first exon encodes a putative RING-like motif. Although the ring domain in the previously characterized proteins (e.g., Ro52, RFP, and MID1) contains the C3HC4 protein signature, the histidine residue is replaced by the cysteine at position 28 in the *RFPL1* protein (Figs. 2–4). The second exon con-

Table 1. PCR Primers

No.	Lab ref.	Primer sequence (5' → 3')
1	RFPL.97F	GCCGTGTCACTGGCTCAG
2	RFPL.283F	GCTGGGACGCAATACCTTT
3	RFPL.3676R	AAGTCTGCCCCACCAAGTA
4	RFPL.3811R	AGTTGGATGGTGGCAATAA
5	RFPL.ex1.f	CTTCCCTTGTGCACTTTTC
6	RFPL.ex1.r	TTTCCTGGTCTGGTCGTAG
7	RFPL.ex2.f	GGGTTGGCTTCTGTCCATT
8	RFPL.ex2.r	TTTTCAGTGGGGGCTTAT
9	RFPL.4F	GCTGCACTCTTCCAAG
10	RFPL.701R	TCAGTAGTGCCTGGGT
11	ex2u.Opitz	CACAGCCAACAACCTTCCTC
12	ex2L.mut.Opitz	GAAAAATCCCCACTGCTA
13	ex2L.WT.Opitz	GAAAAATCCCCACTGCTG
14	RFPL.anti3.F	CAGGTGAACATGCTTGCT
15	RFPL.anti3.R	TACTGTGTTCTTGGGGTCA
16	RFPL.1.3'R	GGAAATTGATCCCAGAATAC
17	RFPL.anti1.F	CTGGGGTGAGGTTGTTGGT
18	RFPL.anti1.R	GGCTTGGGGACTTTGCTGA

tains the putative B30-2 domain. The two above-mentioned domains are bridged by a coiled-coil domain (predicted residues 65–93, using the COILS program with weights and MTIDK matrix, maximum score 0.959 with a 14-scan window and maximum score 0.803 with a 28-scan window; Fig. 4). Coiled-coil motifs have been characterized in many proteins. These domains form stable, rodlike structures that mediate protein–protein interactions by formation of two or three α -helices coiled around each other. Pairwise protein comparisons were also performed using GAP program from the GCG package. We restricted this analysis to the two domains (RING and B30-2) shared between RFPL1 and RFP. Within the RING motif, the similarity and identity was 35% and 29%, respectively. Similarly, within the B30-2 domain the similarity and identity was 48% and 41%, respectively. Analysis of genomic sequences (GenBank accession nos. AC002059, AC000041, AC000025) revealed a frequent polymorphism in *RFPL1*. The two latter genomic clones revealed a variant with one extra amino acid (288), due to a 3-bp insertion, which we termed a long form (lf, GenBank accession no. AJ010228; Fig. 4). We tested 14 unrelated individuals and found that the lf allele occurs at a frequency of 50%.

The BLASTN analysis of genomic sequences from Figure 1A also suggested the existence of two additional genes very similar to *RFPL1* (Figs. 1B and 4), which are the result of several intrachromosomal duplications. The two putative active genes were named *RFPL2* (GenBank accession no. AJ010231) and *RFPL3* (GenBank accession no. AJ010232), and both are localized in the contig distal to *SGLT1*. Comparison of cDNA sequences of the three genes (GAP program from the GCG package) revealed the following results: 95%

exon 1
gcccgtgtcactggctcaggctcagttgctgggtcaccagtagaattggacccttctccacaccca
gttcagagcacagtgatgattcgtgacttttcttcccaatagaacttcaaatctctgaagacgg
gggggtgggggatgtgcttgagtggtttgtactcatggtctgttcttagaagtgacaaagctggg
acgcaatacctttatgcatgaaaaggttgctactgtgcacaaactaacagggtttccacctcagg
aaattttcttcccttctgacttttcccttgggagtgag
ATG GCT GCA CTC TTC CAA GAA GCA AGC AGC TGT CCC GTC TGC TCA GAC
M A A L F Q E A S S C P V C S D
TAT CTA GAG AAA CCA ATG TCC CTA GAG TGT GGA TGC GCT GTC TGC TTC
Y L E K P M S L E C G A V C F
AAG TGC ATC AAT TCA CTA CAG AAG GAG CCC CAT GGG GAG GAT CTA CTT
K C I N S L Q K E P H G E D L L
TGC TGT TGC TGT TCC ATG GTC TCT CAG AAG AAC AAA ATC AGG CCC AGT
C C C C S M V S Q K N K I R P S
TGG CAG CTA GAG AGG CTG GCT TCC CAC ATC AAG GAA CTG GAG CCC AAG
W Q L E R L A S H I K E L E P K
CTG AAG AAG ATT CTG CAG ATG AAC CCA AGG ATG CGG AAG TTC CAA G
L K K I L Q M N P R M R K F Q
exon 1, 582 bp
gtgagggatctgtata **intron 1, 2448 bp** acttttttttcccttttccacag
exon 2
TG GAT ACG ACC TTG GAT GCC GAC ACA GCC AAC AAC TTC CTC ATT
V D T T L D A D T A N N F L I
TCT GAC GAC CTC AGG AGC GTC CGA AGT GGC TCT ACA CAG AAT CGG
S D D L R S V R S G C I T Q N R
CAA GAC CTT GCC GAG AGA TTT GAC TGT TCC ATT TGC ATC CTC GGC TCC
Q D L A E R F D V S I C I L G S
CCT CGC TTT ACC TGT GGC CGC CAC TAC TGG GAG GTG GAC GTG GGA ACA
P R F T C G R H Y W E V D V G T
AGC ACA GAA TGG GAC CTG GGA GTC TGC AGA GAA TCT GTT CAC CGC AAA
S T E W D L G V C R E S V H R K
GGG AGG ATC CAT CTG ACC ACA GAG CGT GGA TTT TGG ACT GTG AGT TTG
G R I H L T T E R G F W T V S L
AGG GAT GGA AGC CGC CTC TCT GCC AGC ACG GTG CCG CTG ACT TTC CTC
R D G S R L S A S T T V P L T F L
TTC GTA GAC CGC AAG TTA **CAG** CGA GTG GGG ATT TTT CTG GAT ATG GGC
F V D R K L **G** R V G I F L D M G
ATG CAG AAC GTT TCC TTT TTT GAT GCT GAA GGT GGT TCC CAT GTC TAT
M Q N V S F F D A E G G S H V Y
ACA TTC AGG AGT GTC TCT GCT GAG GAG CCA CTG CAC TTG TTT TTT GCT
T F R S V S A E E P L H L F F A
CCT CCA ACT CCA CCT AAT GGT GAT AAG AGT GTC TGG AGT ATC TGT CCT
P P S P P N G D K S V L S I C P
GTG ATA AAC CGC GGC ACT ACT GAT GCT CCA GTC CAT CCT GGG GAG GCC
V I N P G T T D A A P V H P G E A
AAA TAA gcccctcactgcaaaaaacaaaaacaggttaagaaaattacttgggtgggacagac
K stop
ttaggaaacgctcactcgtgtaaaagcgttatcaaaagtcataaggagaaaatattgggacaattc
atgattatacttaataatattgattagttataaaattgagtccttaagtaattattattggccac
catcccaactcattgagcttattggttcacattctgttccctatagaataattcgtattctggg
atcaattttccaaatgctttttttattttctgttaagttcaaatgaattattatagaagttatgagt
taataaagcattggaatatca **exon 2, 911 bp**

Figure 2 Structure of the *RFPL1* gene. The putative protein-coding sequence is capitalized and the amino acid sequence is shown below. Nucleotides (7 of 13), which are in agreement with the Kozak consensus sequence for ribosome binding (Kozak 1996) and the sequence of the polyadenylation signal, are double underlined. Exon–intron borders of intron 1 were deduced by comparing the cDNA and the genomic sequences and 20 bases extending into the intron 1 are shown. The first and last two bases of intron 1 (gt and ag, for donor and acceptor splice sites, respectively) are written in boldface type. The underlined sequence of exon 2 displays the part of the sense *RFPL1* gene, which is complementary to a part (exon 4) of the *RFPL3S* antisense transcript. Eight cysteines encoded by exon 1 and forming the signature of the RING domain are in boldface and italics. The cysteine residue that replaced the histidine in the RFPL1 protein is double underlined. The codon affected by the polymorphic C → T transition in exon 2 at the position 933 in *RFPL1* cDNA, introducing a TAG stop codon and truncating the RFPL1 protein by 75 amino acids, is written in boldface and italics and double underlined.

Figure 3 Partial multiple alignment of amino acid sequences from RFPL1 and three proteins that belong to the RING-B30 family: Ro52, RFP, and MID1. (A) Comparison between RING domain located at the amino terminus of the proteins. Arrows indicate the residues forming the RING signature C3HC4; the double arrow denotes histidine residue that is not conserved in RFPL1. (B) Comparison between B30-2 domain located at the carboxyl terminus of all proteins. Identity and similarity between the amino acid sequences is indicated by black and gray boxes, respectively. White boxes display nonconservative amino acid changes. Dashes stand for gaps introduced by the alignment program.

identity between *RFPL1-lf* and *RFPL2*; 94.7% identity between *RFPL1-lf* and *RFPL3*; 96% identity between *RFPL2* and *RFPL3*. At the protein level (Fig. 4) (GAP program), there was 91% identity and 91.3% similarity between *RFPL1-lf* and *RFPL2*, 91.3% identity and 92.4% similarity between *RFPL1-lf* and *RFPL3*, and 94%

identity and 94.8% similarity between *RFPL2* and *RFPL3*. It should be noted that these genes have identical exonic structure and the strong nucleotide similarity extends to the surrounding genomic sequence. The domain structure of the putative proteins is also similar, with a RING-like motif at the amino terminus,

followed by coil-coiled and B30-2-like domains. However, in *RFPL2* a serine residue substitutes the cysteine in the last position of the RING-like signature (Fig. 4). The *RFPL2* is represented in the dBEST by a single EST (GenBank accession no. AA659898) from prostate, spanning the 3' end of the transcript and partially covering exon 2 of the gene. This EST displays a poly(A) tail and a polyadenylation signal. The position of the putative polyadenylation signals is similar in all three *RFPL* genes (see Fig. 2).

A TBLASTN analysis using predicted RFPL1 protein sequence revealed two *RFPL* pseudogenes (*RFPL ψ 1* and *RFPL ψ 2*), which are located distal to the *BAM22* and *SGLT1* loci, respectively (Fig. 1). *RFPL ψ 1* is interrupted in exon 1 by an AluSg element. Moreover, exons 1 and 2 each contain one truncating stop codon (Fig. 5A). Exons 1 and 2 of *RFPL ψ 2* are rearranged, and their orientation is "tail to tail" (see Fig. 1B). Exon 2 is truncated by a LINE1 element and is missing its 5' end (Fig. 5B).

Figure 4 Multiple alignment of amino acid sequences of the putative proteins encoded by three *RFPL* genes. The first 120-amino-acid residues of the 287-amino-acid and the 288-amino-acid variants of *RFPL1*, termed short form (sf) and long form (lf), respectively, are displayed. The dash denotes the missing amino acid at position 110 in *RFPL1sf*, which is highlighted by an asterisk (*). White and black boxes indicate amino acid substitutions and identities, respectively. Vertical arrows point to the cysteine residues of the RING-like domain; the double arrow displays the position corresponding to that of the histidine residue in the C3HC4 signature, which is typical of other RING domain-containing proteins. The predicted coiled-coil (residues 65-93) and B30-2 (residues 90-288) domains are indicated by solid lines above and below the sequences, respectively.

Figure 5 Alignment of predicted amino acid sequences of *RFPL1* and the two pseudogenes *RFPLψ1* (A) and *RFPLψ2* (B). TBLASTN analysis was carried out using *RFPL1* protein as the query. The alignments are shown, with the query on top and the database match labeled as *RFPLψ1* or *RFPLψ2*. A plus sign (+) indicates a similarity between the two amino acid residues. One or more dashes denote insertions or deletions. Stop codons are marked by asterisks (*). Numbers flanking the query and the database match indicate the amino acid and the nucleotide positions, respectively. The nucleotide positions are those of genomic sequences AC000041 (in A) and AL008723 (in B). Vertical arrows point to the cysteine residues of the RING-like domain. Double arrow displays the position parallel to this of the histidine residue in the C3HC4 signature, which is typical to other RING domain containing proteins. AluSg repetitive element is inserted into the pseudo-exon 1 of *RFPLψ1* and LINE1 element is truncating the 5' end of pseudo-exon 2 in *RFPLψ2*.

Opitz G/BBB syndrome (OS), (MIM [Mendelian inheritance in man]nos. 300000 and 145410) is a genetically heterogeneous disease, with X-linked and autosomal dominant inheritance linking to genes on chromosome 22. The main manifestations of OS include facial abnormalities and hypospadias. The critical region on 22q encompasses 32 cM, which is bordered distally by D22S685 (Robin et al. 1995). *RFPL1* is located within the critical region and ~5.7 cM telomeric to marker D22S345, which was linked to OS (maximum lod score 4.06, $\Theta = 0.0$). The OS gene from chromosome X (*MID1*) has been characterized recently (Quaderi et al. 1997) and displays striking similarities to *RFPL1* (see Fig. 3). In view of the above findings, the *RFPL1* gene was considered a candidate for the OS gene from chromosome 22. Therefore, we tested whether *RFPL1* gene is mutated in a previously reported OS family with male-to-male transmission, which would exclude X-linked inheritance (Farndon and Donnai 1983), and the results are summarized in Figure 6. Exons 1 and 2 of *RFPL1* were amplified and the products were sequenced using primers 5–8 (Table 1). The sequence was also confirmed after cloning the PCR products into a pCR2.1-TOPO vector (Invitrogen). Both sequences indicated that the analyzed subjects were homozygous for *RFPL1sf*. The affected father, who had

both long and short alleles (Fig. 6A), displayed a C → T transition at position 933 in the *RFPL1* cDNA sequence (GenBank accession no. AJ010228), introducing a TAG stop codon instead of the glutamine codon (CAG) and truncating the *RFPL1* protein by 75 amino acids (Fig. 2). Because this allele was not inherited by the affected son, this mutation was eliminated as a cause of the son's phenotype. To confirm the pattern of inheritance in this OS family, we resampled the DNA from affected father and son. We performed PCR analysis using allele-specific primers (Table 1, primers 11–13; Fig. 6B), which confirmed that the son did not inherit the truncated allele. Samples of 50 unrelated, normal individuals also were PCR tested, and truncated alleles were detected in four cases (not shown), indicating that it is a polymorphic form of the gene.

RFPL1 and *RFPL3* Genes Reveal Antisense Transcription

To verify the expression of *RFPL1*, we performed Northern blot analyses using probes for exon 1 (Table 1, primers 5 and 6) and exon 2 (primers 7 and 8). The expression pattern for both probes was similar, confirming the existence of a 1.5-kb transcript that was dominant in prostate and less abundant in adult brain, fetal liver, and fetal kidney (Fig. 7A–C). Two other bands were also detected: One was exclusively ob-

Figure 6 Allele-specific PCR of OS family was performed with primers designed to distinguish between the wild-type, nontruncated (WT) and truncated (TR) *RFPL1* forms. (A) Schematic delineation of OS family with male-to-male transmission of the disease phenotype (Farndon and Donnai 1983). Affected individuals are indicated by black boxes. *RFPL1* alleles (the short form, sf, and the long truncated form, lf*) and sample numbers of the template DNA taken for the PCR analysis (B) are shown under each individual. (B) Allele-specific PCR (2% agarose gel) was performed on four members of the OS family delineated at left. The sample numbers of the template DNA used and the specificity of the used PCR primer pairs are indicated in boxes above the lanes. A size marker (123-bp ladder) was loaded in the left lane. It should be stressed that the affected individuals were sampled twice (affected father 96-309 and 98-953; affected son 96-009 and 98-954) to confirm that the truncated allele was not transmitted from the father to the son.

served in testis (1.2 kb) and was specific for the exon 2 probe of *RFPL1*; the other (6 kb) was detected by probes for both exons as strong bands in adult and fetal brain, and weak bands in testis, ovary, and fetal kidney. Similarly, we verified the expression of *RFPL2*, using primers 9 and 10 (Table 1), which allowed us to amplify a PCR product only from the Marathon brain cDNA library. This cDNA contained exons 1 and 2 of *RFPL2* spliced together, at splicing sites that are equivalent to those of *RFPL1* (see Fig. 2). Then we performed a Northern blot analysis using this *RFPL2* cDNA probe, which revealed the same pattern of expression as for the combined exon 1/2 probe of the *RFPL1* gene (Fig.

7). These cross-hybridization results, even under very stringent hybridization and washing conditions, illustrate the strong sequence similarity between the genes.

We were puzzled by the fact that the above-described *RFPL1* and *RFPL2* probes produced intense bands on Northern blots, but we were unable to detect correctly spliced ESTs containing both exons 1 and 2 for the *RFPL* genes. Analysis of one crucial EST clone from testis, corresponding to the *RFPL3* locus (forward sequence accession no. AA398586; reverse sequence accession no. AA393375), indicated that the *RFPL3* gene has an antisense transcript, which is composed of four exons (Fig. 1B; Table 2). Two of these exons (2 and 3) were identified previously and correctly spliced by exon-trapping procedure of a cosmid from chromosome 22 (GenBank accession no. H55552). We named the antisense transcript of the gene as *RFPL3S*. The structure and position of the splicing sites for *RFPL3S* indicate that it is formed by trans-

cription, which proceeds in the opposite direction to that of the putative sense *RFPL3* transcript and covers the entire exon 2 of *RFPL3*. No apparent ORF and no repetitive elements could be detected in *RFPL3S*. We hypothesize that it may have a role in the antisense regulation of the *RFPL* genes. Other *RFPL3S* ESTs (GenBank accession nos. AA868889, AI002159, AI015976) are all from testis, suggesting that this antisense transcript is expressed there and may correspond to the 1.2 transcript detected by Northern blot analysis. To verify this, we designed PCR primers 14 and 15 (Table 1), which allow amplification of a 167-bp segment containing exons 1–3 of *RFPL3S*. This fragment does not

Figure 7 Northern blot hybridization using cDNA probes specific for the human *RFPL* genes. (A) Human (Clontech accession no. 7760-1); (B) human II (Clontech accession no. 7759-1); (C) human fetal II (Clontech accession no. 7756-1). The size of transcripts of the *RFPL* genes are indicated at right of each autoradiograph. The sense forms of the transcripts of the *RFPL* genes migrate ~1.5 kb, and the two antisense transcripts migrate at 6 and 1.2 kb (*RFPL1S*, *RFPL3S*). Size markers are shown at left. The autoradiographs presented in A–C were obtained using the probe for both exons of *RFPL2* gene.

Table 2. Genomic Organization of the *RFPL3S* Gene

Intron	Exon		Intron	Intron size
	no.	size		
gtcatgttggtcctttc ag TTCCCC	1	87	TTAAAG gt atgttcacattggcctgt	2332
tttcttctctaattgc ag GTGACC	2	71	ACCAAG gt aaagtagcgtgtcttgc	782
gaccctgtttttttgc ag TGGCGG	3	25	TAAAT gt atgtattctggaatcatt	6850
	4	934		

span the sequence of *RFPL* sense forms. We tested the panel of 11 cDNA libraries by PCR, and detected and sequenced the correct size product, which was predominant in the testis cDNA library. Similar, less intense bands were detected in Marathon cDNA libraries (brain, placenta, and pancreas), suggesting a weak expression in other tissues. Furthermore, we used this PCR fragment as a probe on Northern blot analysis and detected exclusively a 1.2-kb transcript in testis. This demonstrates that it corresponds to *RFPL3S* and that the original ESTs from testis represent the full-length *RFPL3S* transcript (1117 bp, excluding the poly(A) tail; accession no. AJ010233).

The dBEST database contains EST clones covering the genomic sequence of *RFPL1* locus, which suggests that a similar antisense transcription mechanism may function here as well. The majority of the ESTs originate from brain/neuroepithelium (GenBank accession nos. D61008, D61208, D81014, D81153, D80620, H51938, N64407, N68987, N76400, R61476, R61477, AA708002, AA127191) and one from colon (accession no. AA948403). When assembled (*RFPL1S*, 5112 bp, accession no. AJ010230), this EST contig corresponds to >5 kb of genomic sequence including exon 2, intron 1, and exon 1 of *RFPL1*, up to and including the second *Alu* repeat located on the centromeric side of exon 1. The position of putative polyadenylation signal (AATAAA at position 5093) and the orientation of the poly(A) tail indicate that this gene is transcribed in the opposite direction than that of the sense form of *RFPL1*. Because it was likely that this large transcript corresponds to the 6-kb band detected predominantly in brain on Northern blots (Fig. 7), PCR primers 17 and 18 (Table 1) within intron 1 of *RFPL1* were designed to amplify a 155-bp repeat-free fragment. We tested by PCR the expression of this fragment in the panel of cDNA libraries as described above and obtained an appropriate PCR product in brain and testis. When used as a probe in Northern blot analysis it detects exclusively a 6-kb band, which confirms that it corresponds to the *RFPL1S* antisense transcript.

We confirmed the identity of sense and antisense transcripts on Northern blots, as summarized in Figure 7. The same panel of Northern blots was hybridized with five different probes covering (1) *RFPL1* exon 1;

(2) *RFPL1* exon 2; (3) *RFPL1S*-specific probe, which is a 155-bp fragment of *RFPL1S* transcript that did not contain repetitive elements, and is located in intron 1 of *RFPL1*; (4) probe specific to *RFPL3S*, exons 1–3; and (5) both exons of *RFPL2*. Using *RFPL1* exon 1 probe all bands were detected, except for *RFPL3S* band. The probe for *RFPL1* exon 2 and the probe for

both exons of *RFPL2* detect all the above bands. The *RFPL1S*-specific probe detected only the 6-kb band. Finally, the *RFPL3S* probe detected only the 1.2-kb band.

DISCUSSION

We report a novel family of three very similar *RFPL* genes. Comparisons between nucleotide sequences of exons in sense orientation for *RFPL1*, *RFPL2*, and *RFPL3* revealed a 95%–96% identity. This explains a strong cross-hybridization of probes derived from sense exons of these genes on Northern blots, despite stringent filter washing conditions. The *RFPL1* and *RFPL3* genes express NARs (*RFPL1S* and *RFPL3S*), which can be detected as abundant transcripts on Northern blots in testis, adult brain, and fetal brain, as well as less intense bands in prostate, ovary, and fetal kidney. We confirmed the existence of antisense mRNAs using RT-PCR followed by sequencing and we identified unequivocally the *RFPL1S* and *RFPL3S* transcripts as 6- and 1.2-kb bands on Northern blots, respectively. The hypothesized role of these mRNAs is post-transcriptional regulation of *RFPL* genes at different spatial and temporal windows. Considering the high degree of similarity between sense exons 1 and 2 of *RFPL1*, *RFPL2*, and *RFPL3*, it is plausible that an antisense transcript of one of the genes could exert a regulatory effect on other family members. Both *RFPL1S* and *RFPL3S* genes cover substantial portions of their sense counterparts. *RFPL3S* covers the entire coding part of sense exon 2 of *RFPL3* (591 bp), whereas *RFPL1S* covers the entire coding region of exons 1 and 2 of the sense *RFPL1* gene. Furthermore, the *RFPL1S* and *RFPL3S* antisense transcripts have no apparent protein product. Their predicted ORFs are short and putative peptides that could be encoded by these ORFs do not display significant similarities to any known proteins. In addition, *RFPL1S* contains *Alu* elements, the first report of a NAR that has repetitive DNA elements. In summary, it is most likely that the normal function of *RFPL1S* and *RFPL3S* is to regulate the expression of the sense *RFPL* genes post-transcriptionally.

Although Northern blot analysis indicates that antisense *RFPL1S* and *RFPL3S* transcripts as well as sense *RFPL* mRNAs are abundantly expressed, very few ESTs are present in dBEST, especially for the *RFPL3S* and the

sense *RFPL* genes. This may suggest that duplex formation of sense and antisense transcripts promotes rapid degradation of these mRNAs (Kimelman and Kirschner 1989). Alternatively, the duplex formation may prevent cDNA synthesis during the preparation of the cDNA libraries. As a consequence, ESTs for other, as yet uncharacterized, genes with antisense transcripts may be underrepresented in the current cDNA libraries used for generation of ESTs.

Domain Structure, Presumed Function, and Origin of *RFPL* Genes

The putative *RFPL* proteins are members of a large protein family with zinc finger motifs. The sense *RFPL* transcripts encode proteins with the tripartite structure, composed of RING-finger, coiled-coil, and B30-2 domains, which are characteristic of the RING-B30 family (Henry et al. 1997; Quaderi et al. 1997). One distinct difference between the RING-B30 subfamily and the *RFPL* proteins is that the histidine residue in the C3HC4 motif of their RING-finger domains is replaced with a cysteine. Another difference is lack of the B-box domain, which is usually located between RING-finger and coiled-coil domains (Henry et al. 1997; Quaderi et al. 1997). Several of the proteins containing this tripartite structure were found in multiprotein complexes within cells. Each of the domains that form *RFPL* has been suggested to mediate protein-protein interactions by promoting homo- or heterodimerization (Lupas et al. 1991; Borden and Freemont 1996; Borden et al. 1996; Quaderi et al. 1997).

Three RING-B30 proteins (*RFP*, *Ro52*, and *MID1*), which display highest similarity with *RFPL*s, are presented in Figure 3. The *RFP* protein acquires transforming activity when fused with the *RET* proto-oncogene and plays a role in regulation of cell differentiation (Takahashi et al. 1988; Cao et al. 1998). *Ro52* is associated with cytoplasmic ribonucleic particles (Deutscher et al. 1988; Pruijn et al. 1997). Both *RFP* and *Ro52* genes map to the human chromosome 6p21 region, in the vicinity of the major histocompatibility complex (MHC) class I genes (Vernet et al. 1993). Interestingly, the exon encoding the B30-2 domain was cloned originally from the MHC region. This exon was copied to several genes and made it a noted feature of the MHC region (Vernet et al. 1993). The B30-2 domain encoding exon was found in several genes of diverse functions, namely the myelin oligodendrocyte glycoprotein (*MOG*) and *RFP* genes, which are located ~0.2 and ~1 Mb telomeric to *HLA-A*, respectively. This exon was further duplicated to the hemochromatosis locus (*HLA-H* or *HFE*), 4.5-Mb telomeric to *HLA-A* (Ruddy et al. 1997). It was also detected in the butyrophilin gene family (*BTF 1*, *BTF 2*, *BTF 3*, *BTF 5*) and the *RoRet* gene, which contain a RING finger in the amino terminus (Ruddy et al. 1997). Thus, the protein sequence and

domain structure of *RFPL*s suggest they share a common ancestral gene with the RING-B30 genes of the MHC region.

The *MID1* gene is responsible for the development of the X-linked form of the OS (Quaderi et al. 1997). Because of the similarity between the *RFPL1* and *MID1* genes and the report of OS linkage to a region of 32 cM on 22q, which includes the *RFPL1* region (Robin et al. 1995), we tested the possibility that *RFPL1* is mutated in OS. However, we were unable to find disease-associated inactivating mutations. In the course of this study we detected a truncating mutation of the *RFPL1* gene in an affected father from the OS family. However, this truncating allele was not transmitted to the affected son of this patient, which excludes it as being the direct cause of OS in this family. Moreover, we showed that this truncated allele is present in the normal Swedish population at a frequency of ~4%. To our knowledge, polymorphic stop codons with no obvious phenotypic effects have been observed previously in two other genes: the *BRCA2* gene from chromosome 13 (Mazoyer et al. 1996) and the *MICB* (MHC class I chain-related B) gene from chromosome 6 (Ando et al. 1997). Considering the chromosome 22 linkage data from families with OS, the *RFPL2* and *RFPL3* genes are less likely candidates for the OS-causing genes. *RFPL1* is located on 22q at the position 14197 on the chromosome 22 map from the Sanger Centre (CHR22 map in which 1 unit $\equiv$ 1 kb; <http://webace.sanger.ac.uk/cgi-bin/>), within the OS critical region and ~5.7 cM telomeric to *D22S345* (position 8532.61), which was shown previously to be linked to OS (maximum lod score 4.06, $\Theta = 0.0$). The *RFPL2* and *RFPL3* genes are located in a much more telomeric position on 22q (*RFPL2*, position ~16774; *RFPL3*, position ~16938). The OS critical region is flanked distally by marker *D22S685* (map position 19336.2). Two additional, independent lines of evidence suggest that the OS gene on 22q is located toward the centromere, as compared with the location of *RFPL* genes. First, patients having stigmata of OS and displaying constitutional deletions in 22q11.2 have been reported (McDonald-McGinn et al. 1995; Lacassie and Arriaza 1996), suggesting that the OS gene is located centromeric to heparin cofactor II gene (*HCF2*, map position 4863). Second, several reports showed constitutional deletions (in the range of 0.6–7 Mbp), encompassing the neurofibromatosis type 2 (*NF2*) and the *RFPL* genes, which, in these patients, are associated with the *NF2* disease phenotype (Sanson et al. 1993; Watson et al. 1993; Bruder et al. 1999). These *NF2*-affected subjects did not reveal a phenotype related to OS.

Chromosome 22 Is a Puzzle of Intrachromosomal Duplications

The *BAM22* gene was cloned from a homozygous tu-

mor deletion and displayed a lack of transcript in a subset of human meningiomas (Peyrard et al. 1994). The starting point for this study was the investigation of a putative second *BAM22* gene. This search resulted in description of two *BAM22* pseudogenes. One of these (*BAM22 ψ 1*), located upstream of the functional *BAM22* gene, displays a remarkable degree sequence similarity with the functional *BAM22* gene. It is likely that this conservation of *BAM22 ψ 1* reflects a functional importance. One conceivable function of *BAM22 ψ 1* would be its involvement in the post-transcriptional antisense regulation of the *BAM22* gene. Another possibility would be its role in *trans*-splicing between primary transcripts of *BAM22* and *BAM22 ψ 1*. There is increasing evidence suggesting that *trans*-splicing occurs naturally in mammalian cells (Konarska et al. 1985; Dandekar and Sibbald 1990). A recent report on the rat carnitine actanoyltransferase (*COT*) gene showed that repetition of exons 2 and 3 in the *COT* gene transcript occurs secondary to *trans*-splicing mechanism, leading to production of two forms of the *COT* protein (Caudevilla et al. 1998).

We uncovered several intrachromosomal duplications on 22q and these genetic events were the underlying mechanism behind creation of three active *RFPL* genes, as well as four pseudogenes, *RFPL ψ 1*, *RFPL ψ 2*, *BAM22 ψ 1*, and *BAM22 ψ 2*. Although it is likely that several consecutive duplications/inversions were necessary to produce the complex picture shown in Figure 1, the exact number and order of these events is currently difficult to delineate. However, comparison of sequences from the *RFPL2* and *RFPL3* loci suggests that the duplication creating these two distinct genes occurred more recently. Many intrachromosomal duplications, or low copy repeats, on 22q have been described previously (Halford et al. 1993; Collins et al. 1995, 1997). As the full sequence of this chromosome will soon emerge, the number of characterized duplications on 22q is likely to increase significantly. The link between the presence of low copy repeats and genetic disease seems well established. On chromosome 22, a majority of dispersed, low copy repeats were so far reported in 22q11 region, which has been shown to be unstable, as it is often affected by deletions and other rearrangements leading to, for example, CATCH22 phenotype (Scambler 1993; Puech et al. 1997). It is assumed that genetic instability is caused by recombination between dispersed repeats, as seen for instance, on the X chromosome in cases of steroid sulfatase deficiency and hemophilia A (Mazzarella and Schlessinger 1997).

METHODS

Sequencing and Informatics

Large-scale genomic sequencing was performed as described

previously (Chisoe et al. 1991; Bodenteich et al. 1994; Kedra et al. 1997). Repetitive sequences were filtered out from genomic sequence using REPEAT MASTER server (ftp.genome.washington.edu). EST clones were obtained from Genome Systems, Inc. and resequenced using vector-specific primers and Prism-DyeTerminator (Perkin-Elmer) sequencing chemistry. The BLAST family of programs were used for database searches on the National Center for Biotechnology Information/National Institutes of Health (NCBI/NIH) server (www.ncbi.nlm.nih.gov/BLAST/). Trace files for the ESTs were imported via ftp from genome.wustl.edu and assembled using the GAP4 program from the Staden package (Staden 1994). Pairwise nucleotide and protein comparisons were calculated using the GAP program from the GCG package. Predicted amino acid sequences of the RFPL proteins were aligned using the CLUSTALX (Thompson et al. 1997) and the output was processed by the BOXSHADE program. Coiled-coil domains were predicted using the COILS (Lupas et al. 1991) (www.isrec.isb-sib.ch/software/COILSform.html) and the MULTICOIL programs (Wolf et al. 1997) (nightingale.lcs.mit.edu/cgi-bin/multicoil).

PCR Primers and Probes

Using PCR primers (Table 1) the following human cDNA libraries were tested for the transcript forms of the *RFPL1* gene: fetal brain (Stratagene, no. 936206), fetal muscle (Stratagene, no. 836201), adult skeletal muscle (Stratagene, no. 937209), fetal spleen (Stratagene, no. 937205), pancreatic adenocarcinoma (Stratagene, no. 937208), testis (Stratagene, no. 939202), fetal brain (Clontech, no. HL3003a) and thyroid (Clontech, no. HL3019a), brain Marathon-ready (Clontech, no. 7400-1), placenta Marathon-ready (Clontech, no. 7411-1), pancreas Marathon-ready (a generous gift of Dr. P. Zaphiropoulos, Karolinska Institute). PCR-amplified cDNA fragments were isolated in low melting point agarose gels and sequenced as described previously (Seroussi et al. 1998). Products of sequencing reactions were separated using LongRanger (FMC Bioproducts, Rockland, ME) acrylamide gels on ABI 377 sequencer (Perkin Elmer) using Big-DyeTerminator sequencing kit. Radioactive labeling of probes was performed according to standard methods (Feinberg and Vogelstein 1984; Sambrook et al. 1989). Southern and Northern blots were hybridized and washed using stringent conditions ($0.1\times$ SSC, 0.1% SDS, 65°C) (Sambrook et al. 1989). Allele specific PCR was performed using primers 11-13 (Table 1) and AmpliTaq Gold (Perkin-Elmer) polymerase in 33 cycles (92°C, 1 min; 63°C, 1 min; 72°C, 2 min).

ACKNOWLEDGMENTS

We thank Dr. Peter G. Zaphiropoulos for the pancreas Marathon-ready cDNA library and Kevin O'Brien for critical review of the manuscript. The mapping and sequence data for genomic clones with accession numbers Z83839, AL022321, AL008723, and AL021937 was produced by the human chromosome 22 mapping and sequencing groups at the Sanger Centre. This work was supported by grants from the Swedish Cancer Foundation, the Swedish Medical Research Council, the Cancer Society in Stockholm, the Berth von Kantzow Fond, the Ake Wiberg's Foundation, the Karolinska Hospital, and the Karolinska Institutet to JPD, grants from the National Human Genome Research Institute to B.A.R. and grants from the British Heart Foundation to P.S.

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

- Ando, H., N. Mizuki, M. Ota, M. Yamazaki, S. Ohno, K. Goto, Y. Miyata, K. Wakisaka, S. Bahram, and H. Inoko. 1997. Allelic variants of the human MHC class I chain-related B gene (MICB). *Immunogenetics* **46**: 499–508.
- Bodenteich, A., S. Chissoe, Y.F. Wang, and B.A. Roe. 1994. Shotgun cloning as the strategy of choice to generate templates for high-throughput dideoxynucleotide sequencing. In *Automated DNA sequencing and analysis techniques* (ed. C. Venter), pp. 42–50. Academic Press, London, UK.
- Borden, K.L. and P.S. Freemont. 1996. The RING finger domain: A recent example of a sequence-structure family. *Curr. Opin. Struct. Biol.* **6**: 395–401.
- Borden, K.L., J.M. Lally, S.R. Martin, N.J. O'Reilly, E. Solomon, and P.S. Freemont. 1996. In vivo and in vitro characterization of the B1 and B2 zinc-binding domains from the acute promyelocytic leukemia protooncoprotein PML. *Proc. Natl. Acad. Sci.* **93**: 1601–1606.
- Bruder, C.E.G., K. Ichimura, E. Blennow, T. Ikeuchi, T. Yamaguchi, Y. Yuasa, V.P. Collins, and J.P. Dumanski. 1999. Severe phenotype of the neurofibromatosis type 2 in a patient with a 7.4 Mbp constitutional deletion on chromosome 22; possible localization of a neurofibromatosis type 2 modifier gene? *Genes, Chromosomes, Cancer* **25**: 184–1900.
- Cao, T., E. Duprez, K.L. Borden, P.S. Freemont, and L.D. Etkin. 1998. Ret finger protein is a normal component of PML nuclear bodies and interacts directly with PML. *J. Cell Sci.* **111**: 1319–1329.
- Caudevilla, C., D. Serra, A. Miliar, C. Codony, G. Asins, M. Bach, and F.G. Hegardt. 1998. Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. *Proc. Natl. Acad. Sci.* **95**: 12185–12190.
- Chissoe, S.L., Y.F. Wang, S.W. Clifton, N.T. Ma, H.J. Sun, J.S. Lobsinger, S.M. Kenton, J.D. White, and B.R. Roe. 1991. Strategies for rapid and accurate DNA sequencing. *Methods Companion Methods Enzymol.* **3**: 55–65.
- Collins, J.E., C.G. Cole, L.J. Smink, C.L. Garrett, M.A. Leversha, C.A. Soderlund, G.L. Maslen, L.A. Everett, K.M. Rice, A.J. Coffey et al. 1995. A high resolution integrated yeast artificial chromosome clone map of human chromosome 22. *Nature* **377**: 367–379.
- Collins, J.E., A.J. Mungall, K.L. Badcock, J.M. Fay, and I. Dunham. 1997. The organization of the gamma-glutamyl transferase genes and other low copy repeats in human chromosome 22q11. *Genome Res.* **7**: 522–531.
- Dandekar, T. and P.R. Sibbald. 1990. Trans-splicing of pre-mRNA is predicted to occur in a wide range of organisms including vertebrates. *Nucleic Acids Res.* **18**: 4719–4725.
- Delattre, O., J. Zuckman, B. Plougastel, C. Desmaze, T. Melot, M. Peter, H. Kovar, I. Joubert, P. de Jong, G. Rouleau, A. Aurias, and G. Thomas. 1992. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **359**: 162–165.
- Deutscher, S.L., J.B. Harley, and J.D. Keene. 1988. Molecular analysis of the 60-kDa human Ro ribonucleoprotein. *Proc. Natl. Acad. Sci.* **85**: 9479–9483.
- Dolnick, B.J. 1997. Naturally occurring antisense RNA. *Pharmacol. Therap.* **75**: 179–184.
- Farndon, P.A. and D. Donnai. 1983. Male to male transmission of the G syndrome. *Clin. Genet.* **24**: 446–448.
- Feinberg, A.P. and B. Vogelstein. 1984. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Addendum. *Anal. Biochem.* **137**: 266–267.
- Halford, S., E. Lindsay, M. Nayudu, A.H. Carey, A. Baldini, and P.J. Scambler. 1993. Low-copy-number repeat sequences flank the DiGeorge/velo-cardio-facial syndrome loci at 22q11. *Hum. Mol. Genet.* **2**: 191–196.
- Henry, J., M. Ribouchon, D. Depetris, M. Mattei, C. Offer, R. Tazi-Ahnini, and P. Pontarotti. 1997. Cloning, structural analysis, and mapping of the B30 and B7 multigenic families to the major histocompatibility complex (MHC) and other chromosomal regions. *Immunogenetics* **46**: 383–395.
- Kedra, D., E. Seroussi, I. Fransson, J. Trifunovic, M. Clark, J. Lagercrantz, E. Blennow, H. Mehlin, and J.P. Dumanski. 1997. The germinal center kinase gene and a novel CDC25-like gene are located in the vicinity of PYGM gene on 11q13. *Hum. Genet.* **100**: 611–619.
- Kimelman, D. and M.W. Kirschner. 1989. An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in *Xenopus* oocytes. *Cell* **59**: 687–696.
- Knee, R.S., S.E. Pitcher, and P.R. Murphy. 1994. Basic fibroblast growth factor sense (FGF) and antisense (fgf) RNA transcripts are expressed in unfertilized human oocytes and in differentiated adult tissues. *Biochem. Biophys. Res. Comm.* **205**: 577–583.
- Konarska, M.M., R.A. Padgett, and P.A. Sharp. 1985. Trans splicing of mRNA precursors in vitro. *Cell* **42**: 165–171.
- Kozak, M. 1996. Interpreting cDNA sequences: Some insights from studies on translation. *Mamm. Genome* **7**: 563–574.
- Krystal, G.W., B.C. Armstrong, and J.F. Battey. 1990. N-myc mRNA forms an RNA-RNA duplex with endogenous antisense transcripts. *Mol. Cell. Biol.* **10**: 4180–4191.
- Lacassie, Y. and M.I. Arriaza. 1996. Opitz GBBB syndrome and the 22q11.2 deletion. *Am. J. Med. Genet.* **62**: 3180.
- Lee, R.C., R.L. Feinbaum, and V. Ambros. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**: 843–854.
- Lees, J.F., P.S. Shneiderman, S.F. Skuntz, M.J. Carden, and R.A. Lazzarini. 1988. The structure and organization of the human heavy neurofilament subunit (NF-H) and the gene encoding it. *EMBO J.* **7**: 1947–1955.
- Lupas, A., D.M. Van, and J. Stock. 1991. Predicting coiled coils from protein sequences. *Science* **252**: 1162–1164.
- Mazoyer, S., A.M. Dunning, O. Serova, J. Dearden, N. Puget, C.S. Healey, S.A. Gayther, J. Mangion, M.R. Stratton, H.T. Lynch et al. 1996. A polymorphic stop codon in BRCA2. *Nat. Genet.* **14**: 253–254.
- Mazzarella, R. and D. Schlessinger. 1997. Duplication and distribution of repetitive elements and non-unique regions in the human genome. *Gene* **205**: 29–38.
- McDonald-McGinn, D.M., D.A. Driscoll, L. Bason, K. Christensen, D. Lynch, K. Sullivan, D. Canning, W. Zavod, N. Quinn, J. Rome et al. 1995. Autosomal dominant "Opitz" GBBB syndrome due to a 22q11.2 deletion. *Am. J. Med. Genet.* **59**: 103–113.
- Peyrard, M., I. Fransson, Y.-G. Xie, F.-Y. Han, M.H. Rutledge, S. Swahn, J.E. Collins, I. Dunham, V.P. Collins, and J.P. Dumanski. 1994. Characterization of a new member of the human β -adapin gene family from chromosome 22q12, a candidate meningioma gene. *Hum. Mol. Genet.* **3**: 1393–1399.
- Peyrard, M., H.Q. Pan, D. Kedra, I. Fransson, S. Swahn, K. Hartman, S.W. Clifton, B.A. Roe, and J.P. Dumanski. 1996. Structure of the promoter and genomic organization of the human β '-adapin gene (BAM22) from chromosome 22q12. *Genomics* **36**: 112–117.
- Prujij, G.J., F.H. Simons, and W.J. van Venrooij. 1997. Intracellular localization and nucleocytoplasmic transport of Ro RNP components. *Eur. J. Cell Biol.* **74**: 123–132.
- Puech, A., B. Saint-Jore, B. Funke, D.J. Gilbert, H. Sirotkin, N.G. Copeland, N.A. Jenkins, R. Kucherlapati, B. Morrow, and A.I. Skoultschi. 1997. Comparative mapping of the human 22q11 chromosomal region and the orthologous region in mice reveals complex changes in gene organization. *Proc. Natl. Acad. Sci.* **94**: 14608–14613.
- Quaderi, N.A., S. Schweiger, K. Gaudenz, B. Franco, E.I. Rugarli, W. Berger, G.J. Feldman, M. Volta, G. Andolfi, S. Gilgenkrantz et al. 1997. Opitz g/bbb syndrome, a defect of midline development, is due to mutations in a new ring finger gene on Xp22. *Nat. Genet.* **17**: 285–291.
- Robertson, N.G., R.J. Pomponio, G.L. Mutter, and C.C. Morton. 1991. Testis-specific expression of the human MYCL2 gene. *Nucleic Acids Res.* **19**: 3129–3137.

- Robin, N.H., G.J. Feldman, A.L. Aronson, H.F. Mitchell, R. Weksberg, C.O. Leonard, B.K. Burton, K.D. Josephson, R. Laxova, K.A. Aleck et al. 1995. Oritz syndrome is genetically heterogeneous, with one locus on Xp22, and a second locus on 22q11.2. *Nat. Genet.* **11**: 459–461.
- Ruddy, D.A., G.S. Kronmal, V.K. Lee, G.A. Mintier, L. Quintana, R.J. Domingo, N.C. Meyer, A. Irrinki, E.E. McClelland, A. Fullan et al. 1997. A 1.1-Mb transcript map of the hereditary hemochromatosis locus. *Genome Res.* **7**: 441–456.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanson, M., C. Marineau, C. Desmaze, M. Luthman, M. Ruttledge, C. Baron, S. Narod, O. Delattre, G. Lenoir, G. Thomas et al. 1993. Germline deletion in a neurofibromatosis type 2 kindred inactivates the NF2 gene and a candidate meningioma locus. *Hum. Mol. Genet.* **2**: 1215–1220.
- Scambler, P.J. 1993. Deletions of human chromosome 22 and associated birth defects. *Curr. Opin. Genet. Dev.* **3**: 432–437.
- Seroussi, E., H.-Q. Pan, D. Kedra, B. Roe, and J.P. Dumanski. 1998. Characterization of the human NIPSNAP1 gene from 22q12: A member of a novel gene family. *Gene* **212**: 13–20.
- Staden, R., Ed. 1994. *The Staden package. Methods in molecular biology*. Humana Press, Totawa, NJ.
- Takahashi, M., Y. Inaguma, H. Hiai, and F. Hirose. 1988. Developmentally regulated expression of a human “finger”-containing gene encoded by the 5′ half of the ret transforming gene. *Mol. Cell. Biol.* **8**: 1853–1856.
- Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The CLUSTALX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876–4882.
- Turk, E., I. Klisak, R. Bacallao, R.S. Sparkes, and E.M. Wright. 1993. Assignment of the human Na⁺/glucose cotransporter gene SGLT1 to chromosome 22q13.1. *Genomics* **17**: 752–754.
- Vanhee-Brossollet, C. and C. Vaquero. 1998. Do natural antisense transcripts make sense in eukaryotes? *Gene* **211**: 1–9.
- Vernet, C., J. Boretto, M.G. Mattei, M. Takahashi, L.J. Jack, I.H. Mather, S. Rouquier, and P. Pontarotti. 1993. Evolutionary study of multigenic families mapping close to the human MHC class I region. *J. Mol. Evol.* **37**: 600–612.
- Wagner, E.G. and R.W. Simons. 1994. Antisense RNA control in bacteria, phages, and plasmids. *Ann. Rev. Microbiol.* **48**: 713–742.
- Watson, C.J., L. Gaunt, G. Evans, K. Patel, R. Harris, and T. Strachan. 1993. A disease associated germline deletion maps the type 2 neurofibromatosis (NF2) gene between the Ewing sarcoma region and the leukemia inhibitory factor locus. *Hum. Mol. Genet.* **2**: 701–704.
- Wightman, B., I. Ha, and G. Ruvkun. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* **75**: 855–862.
- Wolf, E., P.S. Kim, and B. Berger. 1997. MultiCoil: A program for predicting two- and three-stranded coiled coils. *Protein Sci.* **6**: 1179–1189.
- Xie, Y.-G., F.-Y. Han, M. Peyrard, M.H. Ruttledge, I. Fransson, P. DeJong, J. Collins, I. Dunham, M. Nordenskjöld, and J.P. Dumanski. 1993. Cloning of a novel, anonymous gene from a megabase-range YAC and cosmid contig in the neurofibromatosis type 2/meningioma region on human chromosome 22q12. *Hum. Mol. Genet.* **2**: 1361–1368.
- Zucman-Rossi, J., P. Legoix, and G. Thomas. 1996. Identification of new members of the gas2 and ras families in the 22q12 chromosome region. *Genomics* **38**: 247–254.

Received March 18, 1999; accepted in revised form July 21, 1999.

Duplications on Human Chromosome 22 Reveal a Novel Ret Finger Protein-Like Gene Family with Sense and Endogenous Antisense Transcripts

Eyal Seroussi, Darek Kedra, Hua-Qin Pan, et al.

Genome Res. 1999 9: 803-814

Access the most recent version at doi:[10.1101/gr.9.9.803](https://doi.org/10.1101/gr.9.9.803)

References

This article cites 51 articles, 10 of which can be accessed free at:
<http://genome.cshlp.org/content/9/9/803.full.html#ref-list-1>

License

Email Alerting Service

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

Affordable, Accurate
Sequencing.

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