The biological significance of RBCC (N-terminal RING finger/B-box/coiled coil) proteins is increasingly being appreciated following demonstrated roles in disease pathogenesis, tumorogenesis, and retroviral protective activity. Found in all multicellular eukaryotes, RBCC proteins are involved in a vast array of intracellular functions; but as a general rule, they appear to function as part of large protein complexes and possess ubiquitin-protein isopeptide ligase activity. Those members characterized to date have diverse C-terminal domain compositions and equally diverse subcellular localizations and functions. Using a bioinformatics approach, we have identified some new RBCC proteins that help define a subfamily that shares an identical domain arrangement (MID1, MID2, TRIM9, TNL, TRIM36, and TRIFIC). Significantly, we show that all analyzed members of this subfamily associate with the microtubule cytoskeleton, suggesting that subcellular compartmentalization is determined by the unique domain architecture, which may in turn reflect basic functional similarities. We also report a new motif called the COS box, which is found within these proteins, the MURF family, and a distantly related non-RBCC microtubule-binding protein. Notably, we demonstrate that mutations in the COS box abolish microtubule binding ability, whereas its incorporation into a nonmicrotubule-binding RBCC protein redirects it to microtubule structures. Further bioinformatics investigation permitted subclassification of the entire human RBCC complement into nine subfamilies based on their varied C-terminal domain compositions. This classification schema may aid the understanding of the molecular function of members of each subgroup and their potential involvement in both basic cellular processes and human disease.

Members of the RBCC (N-terminal RING finger/B-box/coiled coil) or TRIM (tripartite motif) family of proteins perform a diverse array of cellular roles, yet are believed to share some functional properties: 1) act as a scaffold for the assembly of larger multiprotein complexes and 2) possess RING-dependent ubiquitin ligase activity (1, 2). The RBCC domain can be found in isolation or in combination with a variety of other C-terminal domains, including the NHL (NCL-1/HT2A/LIN-41 repeat), immunoglobulin, MATH (meprin and tumor necrosis factor receptor-associated factor homology), B30.2-like/RFP (Ret finger protein)/SPRY (SpiA and tyrosine receptor) (the largest subgroup in humans), ARF (ADP-ribosylation factor), PHD (plant homeodomain finger), and BROMO domains (1, 2). As a family of proteins, their biological significance is perhaps best highlighted by the growing number that have a demonstrated role in disease pathogenesis, including immunological and developmental disorders, tumorogenesis, and retroviral protective activity (3–7).

We have previously identified and characterized two RBCC proteins, MID1 and its closely related homolog, MID2. MID1 and MID2 contain a B30.2-like domain at their C terminus and a single fibronectin type III (FN3) motif between it and their N-terminal RBCC domain (8). Both proteins have been shown to associate with microtubules (8, 9) and do so through both homo- and heterodimerization (5). Consistent with their high level of sequence and structural similarity, MID1 and MID2 can even interact with some of the same proteins (5), and both likely possess ubiquitin-protein isopeptide ligase activity (10). Mutations in MID1 have been identified in patients with Opitz syndrome, a disorder recognized by a combination of congenital anomalies that includes cleft lip and palate as well as heart and anogenital defects. Although mutations have been found in most domains of MID1, mutations in the C-terminal domain predominate and result in significantly altered cellular localization of the protein (4, 11) as well as disrupted targeting of its ubiquitin ligase activity (10).

At the time of their identification, MID1 and MID2 represented the only known RBCC superfamily members with an FN3 motif. A third mammalian protein (Spring/TRIM9) that shares the same domain organization as MID1 and MID2 was subsequently discovered (12, 13). This protein is specifically expressed in neural tissue and appears to play a regulatory role in synaptic vesicle exocytosis (12). To identify additional proteins with a similar domain arrangement, we undertook an analysis of published expressed sequence tag (EST) and genomic data, which indicated that other related proteins may indeed be encoded by the human genome. Support for this notion came recently from the report of Haprin, a new testis-specific FN3 motif- and B30.2-like domain-containing RBCC protein that is clearly the murine ortholog of a sequence identified in our screen. Functional data suggest that Haprin plays a regulatory role in exocytosis of the sperm vesicle (14).

In this study, we report the identification of two novel proteins, TRIFIC and TNL, which share the same overall domain organization as MID1, MID2, TRIM9 (the human ortholog of rat Spring), and TRIM36 (the human ortholog of mouse Haprin), and show for the first time that all analyzed members of this RBCC subfamily associate with the microtubule...
cytoskeleton. We also describe a novel amino acid signature adjacent to the coiled coil (named the COS box) that, when mutated at key conserved residues, results in a complete loss of microtubule localization, but not dimerization ability. Additionally, a fusion protein generated from a C terminus containing a COS box from a microtubule-associated protein and an RBCC/TRIM domain from a non-cytoskeleton-associated protein redirects localization to microtubule structures. We therefore reanalyzed the entire human RBCC family, enabling classification into nine subgroups based on a conserved C-terminal domain organization provided by Pfam, SMART, PRINTS, PROFILE, and PATTERN analyses. The implications of this subclassification in understanding the broader function of RBCC proteins are discussed.

**EXPERIMENTAL PROCEDURES**

Alignment and Phylogenetic and Hidden Markov Model (HMM) Analysis—General amino acid and nucleotide sequence handling was aided by Vector NTI Suite (version 7.0.1, Invitrogen, Mulgrave, New South Wales, Australia), and multiple sequence alignments (MSAs) were post-processed for publication using CHROMA, removing regions of low homology across the family and producing an 85% consensus according to convention (15). MSAs were generated using a combination of manual, ClustalX (16) with Blosum62 (17) and Gonnet (18) matrices, and T-COFFEE (19) alignments using a best fit approach. Phylogenetic analysis was performed on MSA data using PHYLIP Version 3.62 (20). Human sequences were used together with *Drosophila melanogaster* TRIM9 as a representative of an ancestral RBCC/FN3/H11032 B30.2-like subfamily member. A maximum likelihood approach was used to infer phylogeny using the SEQBOOT, ProML, CONSENSE, and DRAWTREE programs, with 100 bootstrap resample data sets being used to infer phylogeny vigor. Bootstrap consensus figures were placed on the tree with post-processing using Adobe Illustrator Version 10 (Adobe Systems Inc., San Jose, CA). HMM analysis used pre- and post-processing Perl scripts3 for the handling of MSA data and HMMER execution (21). Two Sun Blade 100 workstations and an Apple G5 dual CPU workstation were used for data generation, and all HMMs were created using Sun Solaris Version 6.0 and G5 Altivec optimized versions of HMMER Version 2.3.2 (22). Subsequently, Microsoft Excel was used for converting post-processed HMMER data to a graphical output format using in-built pivot table functions. CHROMA-processed MSAs, the phylogenetic tree, and HMMER graphical output was presented using Adobe Illustrator Version 10.

Plasmid Constructions—*MID1* and *MID2* cDNAs and fusions were already cloned and available (4, 5, 8). Human full-length cDNA clones from the IMAGE collection were identified for TRIM9 (IMAGE clone 5243288) and TRIM36 (IMAGE clone 4826462), but not for TRIFlc. Instead, a murine full-length *Trifl* cDNA (IMAGE clone 3813814) that encodes a protein with 98% amino acid identity to human TRIFlc was identified. A full-length cDNA clone was not available for either human *TnI* or mouse *TnI* at the time of this study. Full-length cDNAs were amplified for the purpose of recloning using the following sequence-specific primer pairs: hTRIM9-dTop (5′-GACCCGGGATCCCTCTCGTGCCACTGACGCGTGTCGAC-3′) and hTRIM9-rEco (5′-GAATATCATCATCATCATGACGCGTGTCGAC-3′) for the fusion region from the MID1 C-terminal fragments. The MID1 fragments were amplified with forward primer COS-F (5′-AAATCCGCGGATCCCTCTCGTGCCACTGACGCGTGTCGAC-3′) and reverse primer M1SPRY-R (5′-TCACGGCAGCTGCTCTGTGAGATATGCTGACGCGTGTCGAC-3′) for the COS box fusion, or with wild-type MID1 cDNA template (containing an introduced TGA codon at position 380) for the COS box fusion and were generated by site-directed mutagenesis (Stratagene) using primers MIDCosFLQ-AAA (5′-GGGAAGATGATTCCCTCGGCGCGCTACTGCTAAGATATCCAGGCGCCAGCAGTCTGACGCGTGTCGAC-3′), rMIDcosFLQ-AAA (5′-CTGGTGATATTCTTAGCATGACGCGCGCCACGCGGTATGCTCTCGC-3′), rMIDcosFLQ-AAA (5′-GAAGAAACCTGCTGAAATGTCGCGCGCGCTTACAGCTCCACAACCTG-3′), and rMIDcosFLQ-AAA (5′-GAGGGTGGAGAGGCTGTAAGGGCGCCGCCACACATTAGCTGCTTCTG-3′). All introduced mutations were confirmed by automated sequencing.

Fusions of the TRIM37 tripartite domain with the COS box or C-terminal domains (including the COS box) were generated by PCR amplification of TRIM37D cDNA (kindly supplied by Anna-Elena Lehesjoki, Neuroscience Center and Folkhålsan Institute of Genetics, Biomedicum Helsinki, University of Helsinki, Finland) using primers TRIMD-F (5′-CACCATGGATGACAACAGGCCGTCGAGACGCGCGCGACGCGAGACGCGCGCGACGCGGTAGTCTCTCTGAAC-3′) and TRIMD-R (5′-CACCATGCGGAGATGGTGATGATGCTAGCAATCTGCTGACGCGTGTCGAC-3′) and rMIDcosLDY-AAA (5′-GAAGAAACCTGCTGAAATGTCGCGCGCGCTTACAGCTCCACAACCTG-3′), and rMIDcosLDY-AAA (5′-GAAGAAACCTGCTGAAATGTCGCGCGCGCTTACAGCTCCACAACCTG-3′). All introduced mutations were confirmed by automated sequencing.

Cell Culture, Transfection, and Immunofluorescence—COS-1 cells were grown in Dulbecco’s modified Eagle’s medium (Invitrogen) supplemented with heat-inactivated 10% fetal calf serum (Invitrogen) at 37 °C in 5% CO2. Transfection of plasmid DNA was performed using FuGENE 6 transfection reagent (Roche Applied Science, Castle Hill, New South Wales). Transfected cells were grown on coverslips for 18 h in Dulbecco’s modified Eagle’s medium plus 10% fetal calf serum and then fixed as described previously (4). Microtubule staining was performed post-fixation using an anti-α-

3 Modified Perl scripts used for large-scale GenBank™ HMM scanning (which conserve disk space) can be requested from the authors, as can a Cygwin-packaged version to run the analysis on a Windows PC running Cygwin (Redhat) with Perl Version 5.8.
tubulin primary antibody and Texas Red-conjugated anti-mouse secondary antibody (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA). Fused Myc tag was detected using mouse anti-Myc antibody 9E10, and the same mouse secondary antibody was used as per tubulin. Nuclei were visualized using the DNA-specific Hoechst stain (Sigma, Sydney, New South Wales). GFP and Texas Red fluorescence was detected at appropriate wavelength light using an Olympus AX70 microscope. Images were captured using a Photomicrotics CE200A camera electronics unit and processed using Photoshop Version 7 software (Adobe Systems Inc.).

**Yeast Two-hybrid Analysis**—The ProQuest yeast two-hybrid system (Invitrogen) was employed to assess interactions between MID1, MID2, and other subfamily members. All yeast two-hybrid fusion constructs were transfected into the Saccharomyces cerevisiae MaV203 yeast strain (MATa, leu2-3,112, trp1-901, his3Δ200, ade2-101, gal4Δ, gal80Δ, SPAL10::ura3, GAL1::lacZ, LYS2::P_His3JASGAL::his3, can1Δ, cyh2Δ) by polyethylene glycol/LiAc/Tris-EDTA transfection (23). The HIS3 reporter was used to determine the level of self-activation of the fusions based on the level of 3-aminotriazole resistance of the transformed yeast. 50 mM 3-aminotriazole was found to be an adequate level for the assay of all generated fusions. Transfections were plated on selective synthetic complete medium lacking Leu and Trp, from which colonies were picked, and cultures were established and grown for 24 h at 30 °C in selective medium. Normalized diluted cultures (10 μl at 0.1 A600) were then spotted onto interaction-selective synthetic complete medium lacking Leu, Trp, and His and containing 50 mM 3-aminotriazole and incubated at 30 °C for 48 h. For X-gal assays, spots of selectively cultured transfected yeast were placed on nitrocellulose Hybond-N™ membrane (Amersham Biosciences, Castle Hill) overlaid on complete medium and allowed to grow at 30 °C for 24 h. X-gal assays were performed as reported previously (24), and all post-assay plates and filters were scanned using a Hewlett-Packard Scanjet scanner with Adobe Photoshop.

**Immunoprecipitation and Western Blot Analysis**—Preparations of the various GFP- and Myc-tagged expression constructs were made using a DNA plasmid midi kit (Promega Corp., Annandale, New South Wales). 6 pmol (~3 μg) of each construct was transfected into 1 × 10⁶ COS-1 cells using FuGENE 6 transfection reagent. After a 24-h incubation, cells were scraped from the culture dish and lysed on ice for 30 min in 1 ml of lysis buffer (50 mM Tris-HCl (pH 7.4), 300 mM NaCl, 5 mM EDTA, and 1.0% Triton X-100). Cell lysates were cleared by centrifugation at 4 °C for 15 min at 16,000 × g, and protein extract was recovered as the supernatant. After preclowing 200 μl of protein extract with 10 μl of 50% protein G-Sepharose bead slurry (Sigma), extracts were incubated with 1 μg of antibody for 2 h at 4 °C and then for another 2 h with 20 μl of fresh 50% protein G-Sepharose bead slurry. The beads were washed four times with wash buffer (50 mM Tris-HCl (pH 7.4), 300 mM NaCl, 5 mM EDTA, and 0.1% Triton X-100), and protein was eluted from the beads by boiling in 2× SDS loading buffer. Proteins were separated by 10% SDS-PAGE and blotted onto Hybond-C membranes (Amersham Biosciences) using a semidry transfer apparatus (Owl Separation Systems, Portsmouth, NH). Membranes were blocked, incubated with the appropriate primary antibody, washed, incubated with the appropriate horseradish peroxidase-conjugated secondary antibody, and washed again according to established methods (43). Detection was carried out using an enhanced chemiluminescence kit (ECL, Amersham Biosciences) following the manufacturer’s instructions. The antibodies used in immunoprecipitation and Western blot analysis included rabbit anti-GFP polyclonal antibody (gift from Pam Silver, Dana-Farber Cancer Institute, Boston), mouse anti-Myc monoclonal anti-body 9E10 (gift from Stephen Dalton, University of Adelaide), and horseradish peroxidase-conjugated anti-rabbit and anti-mouse secondary antibodies (Amersham Biosciences).

**RESULTS**

**Identification of New RBCC/FN3/B30.2 Proteins**—To identify additional RBCC/FN3/B30.2-like proteins, we performed separate BlastX and PSI-BLAST searches on both the publicly available human genome and EST data bases using the full-length sequences of the three reported human FN3 motif-containing RBCC proteins (MID1, MID2, and TRIM9) as well as their individual RBCC, FN3, and B30.2-like sequences. Included in this analysis were all previously identified human full-length RBCC sequences (Fig. 1A), excluding those TRIM proteins that were not of human origin or that were missing part or all of the RBCC domain components (see Fig. 1 legend). All RBCC sequences to be analyzed were retrieved and further scrutinized using InterPro (25), SMART (26), Pfam (27), PROSITE (28), PRINTS (29), and MultiCoil (30). These programs were chosen because they use a range of analyses to determine sequence domain similarity by comparison with domain data bases from HMM to PROFILE and pattern matching methods as well as primary sequence analysis. The data provided by these searches allowed the thorough identification of particular subgroups of RBCC proteins based on the presence or absence of currently identified domain families in the Pfam Database (Version 16), PROSITE PROFILE Database (Version 18.37), and MultiCoil when no coil was detected using domain search tools (Fig. 1, A and B).

These analyses identified two human RBCC proteins with significant matches to the FN3 consensus: TRIM36 (14, 31) and FLJ23229 (TRIM46). Although Pfam and SMART Database scores were not significant for a SPRY domain (a submotif of the B30.2-like domain) match over the C terminus of each of these proteins, both were shown to have significant scores for B30.2-like domains using a PROFILE search (available at PROSITE), which is the only tool available to detect complete B30.2-like domains. To provide a similar comparison, the B30.2-like profile matrix (PS00188) was converted to a HMMER data base using ptoh (32). Results of scans performed against the identified RBCC/FN3/ B30.2-like proteins were also significant matches, similar to the PROFILE scan. Using the domains from homologous proteins as a guide, available EST data and genomic sequence representing FLJ23229/TRIM46 were assembled to build a predicted cDNA, which was assisted with predictions available from the GenBank™ Data Bank. From this, we identified that the putative protein contained the same domain architecture as the other identified proteins, and we hereafter refer to this novel protein as TRIFIC (tripartite, fibronectin type III, and C-terminal B30.2-like motif; human GenBank™ accession numberAY251386 (nucleotide) and NCBI accession number AAPP1206 (protein)) and mouse GenBank™ accession number AY251388 (nucleotide) and NCBI accession number AAPP1208 (protein).

**Alignment and Phylogenetic Analysis of the RBCC/FN3/B30.2 Proteins**—Protein MSA indicated a similar motif arrangement in each of the human FN3 motif-containing RBCC sequences (supplemental Fig. 1), with the most notable differences being within the core of the C3HC4 RING finger domain, which is the most variable part of all RING fingers (refer to Pfam accession number zf-C3HC4 and SMART accession number SM00184). The alignment also reiterated the similarity across the B30.2-like region for the entire subfamily and significantly identified regions of high homology not previously associated with domains or protein function.

MSA was used as the basis to perform a maximum likelihood phylogenetic analysis to determine the relative interrelationships between
MID1, MID2, TRIM9, and a novel TRIM9 homolog (TNL; discussed below). All human sequences and the D. melanogaster TRIM9 sequence were used as the basis for a maximum likelihood analysis, and an unrooted phylogenetic tree was plotted (Fig. 2). The tree shows the relationship between the RBCC/FN3/B30.2-like human sequences and MID1 and MID2, TRIFIC and TRIM36, and TRIM9 and TNL forming paired groups, indicating they are the result of progressive gene duplication events. When murine and chick sequences were added to the analysis (data not shown), the orthologous sequences formed monophyletic groups to the exclusion of paralogs, indicating that the gene duplication events predated the speciation of humans, mice, and chicks.

HMMER-based Analysis of the RBCC Superfamily—Verification that we had identified all members containing the FN3 motif came with the reanalysis of each of the 51 human full-length RBCC proteins using HMMER (22). A concatenated HMM data base was generated (as per Ref. 21) using 698 50-amino acid “window” alignments derived from an MSA of four of the five human FN3 motif-containing RBCC proteins (and their orthologs, based on the MSA depicted in supplemental Fig. 1). The calibrated data base containing 698 HMMs representing significant sequence identity was then used for scanning all human RBCC proteins in an attempt to identify sequence window homology. We chose to omit the TRIFIC sequence from the HMM data base so it could be used as a “positive” reference sequence in these analyses, such that each recognized motif would be identifiable within the histogram output to aid interpretation and visual presentation of the data for the entire protein family (Fig. 3A). The sensitivity and visual presentation were aided using 50-amino acid wide “sliding windows” of the alignment so that a score for an HMM search (22) identified.

FIGURE 1. Summary of the human RBCC superfamily and their domain-based subfamily classification. A, classification relied on the identification of domains within the amino acid sequences of the proteins listed. The identity, alternative names, human chromosome location, and subfamily allocation for each gene are indicated. B, shown is a schematic representation of each subfamily, with classification of subfamilies based on the differences in C-terminal domain composition: C-I (COS/FN3/B30.2-like domains), C-II (COS/acid-rich (AR) region), C-III (COS/FN3 domains), C-IV (B30.2-like domains), C-V (as yet no C-terminal motifs identified), C-VI (PHD/BROMO domains), C-VII (immunoglobulin/NHL repeats), C-VIII (MATH domain), and C-IX (ADP-ribosylation factor (ARF) domain). Domains in dashed boxes may be variably presented. The TRIM proteins excluded from the study do not follow the TRIM nomenclature and lack the characteristic tripartite (RING (R), B-box, and coiled coil (CC) RBCC motif: TRIM14, TRIM16 (EBBP), TRIM20 (PYRIN/MEFV), TRIM29 (ATDC), TRIM44 (DIPB), and TRIM66, all of which do not have RING fingers; TRIM51 and TRIM58, which have no RING or B-boxes; TRIM48, TRIM49, TRIM52, and TRIM61, which have no coiled coil; and TRIM65, which has no RING, B-boxes, or coiled coil. TRIM12 and TRIM30 were also excluded because they represent murine proteins with no obvious human ortholog.
FIGURE 3. HMMER analysis of the RBCC superfamily using a C-I subfamily window data base. A, shown are the results from HMMER analysis of human full-length RBCC proteins, which assisted delineation into nine RBCC subgroups. The HMM data base utilized 50-amino acid windows from the established alignment (supplemental Fig. 1). The scores (in color) indicating a significantly matching region (E-values in black) are shifted relative to the sequence due to the window size and similarity of surrounding residues (21). The position relating to the COS box in the C-I subfamily (highlighted on the window axis) is shown; the COS box region scored higher than almost all previously established motifs for the C-I subfamily (represented here by TRIFIC and TNL) and is seen as a large peak between windows 310 and 360 at the C-terminal end of the coiled coil (CC). Likewise, C-terminal peaks at the end of the coiled coil of the MURF proteins and TRIM42 also produced significant, albeit weaker scores for COS box sequences. It should be noted that, due to the progressive windows of the C-I subfamily MSA used to build the HMM data base, scores for non-C-I subfamily members where sequence identity is lower are weighted toward the origin of higher identity.

RF, RING finger.

B, Ensembl Version 11 was used to annotate the human gene structure homologous to a partial rat clone encoding a TRIM9-like (TNL) sequence that was identified by a COS box HMM search of the GenBank™ non-redundant protein sequence data base. We predict that the human TNL gene (GenBank™ accession number AY253917) consists of 12 exons, spanning ~53 kb of chromosome 1q42, and encodes a previously undiscovered RBCC/COS/FN3/B30.2 protein.
A New Subfamily of Microtubule-associated RBCC Proteins

| C-I | BB1 | BB2 | CC | CO | FN3 | B30.2 |
|-----|-----|-----|----|----|-----|-------|
|     | RING |     |     | GLFNĐ |     |      |
| C-II | BB2 | CC | CO | AR |     |

FIGURE 4. The COS box motif. The domains of the known microtubule-associated proteins with COS box homology are represented (A), and an alignment of the peptide sequences of representative proteins was performed using the following sequences: B: MID1 (human), MID2 (human and mouse), Murf1 (human and rat), Mid1 (human and mouse); C: MID1 (human and mouse), GLFNĐ (human and mouse), Murf1 (human and rat), Murf2 (human and predicted mouse), and Murf3 (human and mouse). Asterisks represent regions that are mutated as shown in Fig. 6. Cylinders above the sequences represent the approximate consensus positions of JPred/PSIPRED-predicted α-helical coils across the COS boxes of the proteins. a, amino acids; h, hydrophobic; p, polar; a, aromatic; l, aliphatic; s, small; b, big; c, charged. The accession numbers for other sequences used here are available: human MID1, AF269101 and AAG31330; mouse Mid1, NM_010797 and NP_034927; chicken Mid1, AF269102 and AAK58598; rat Mid1, NM_020927 and NP_075216; human MID2, AF106481 and AA07341; mouse Mid2, NM_011845 and NP_035975; human TRIM9, AF220037 and AAG53491; mouse TRIM9, NM_051367 and NP_444397; rat TRIM9, AF350422 and AA27988; tri, NM_226563 and XP_226563; human TRIM36, NM_018700 and NP_061170; human MID1, NM_032588 and NP_115977; rat Murf1, NM_080903 and NP_543179; human Murf2, AA31704 and CAD24432; mouse Murf2, XM_355438 and XP_355438; human Murf3, AF294790 and AAG03076; mouse Glfnđ, NM_181378 and NP_899001. The fugu MID1 sequence was obtained from A. Ashworth (personal communication).

fying the level of similarity between every progressive window against each sequence from the RBCC superfamily could be plotted in a fashion similar to that used by Truong and Ikura (21). This approach also revealed a greater level of sequence similarity (and therefore likely evolutionary relatedness) between various RBCC members that would not have been discernible using phylogenetic algorithms on similarly large numbers of highly variant sequences. Together with domain data base and coil prediction programs, these analyses have permitted delineation of nine subfamilies based on their C-terminal domain composition (called the C-I through C-IX subfamilies) (Figs. 1 and 3). Notably, this analysis categorizes the Murf (muscle-specific RING finger) proteins (MURF1–3) as a distinct subfamily of RBCC/TRIM proteins, a grouping suggested previously (33), but not identified in the original TRIM classification (2).

As expected for a member of the first RBCC subfamily (C-I), the TRIFIC sequence scored significantly across each known motif (Fig. 3A). However, it was unexpected to find the second highest score (79.3) and E-value (9.4 × 10^-22) for a region of 67 amino acids extending from the end of the TRIFIC coiled coil bridging through to the beginning of the FN3 domain (Fig. 3A, large peak at the end of the coiled coil). Alignment of this region from the C-I subfamily revealed a high level of sequence conservation, comparable with the level among known motifs discovered by domain search engines (Fig. 4B). Members of the C-II subfamily, consisting of Murf1–3, revealed scores above cutoff levels (Fig. 3A) over the same region (scores of 15.9–18.8 and E-values of 0.015–0.0031, corresponding to amino acids 255–311, 254–310, and 256–312 within the proteins, respectively), as did TRIM42 from the C-III subfamily (score of 20.4 and E-value of 0.00075, corresponding to amino acids 435–491). The weighted nature of the HMM analysis resulted in peaks with high scores and significance being positioned at the end of the coiled coil in the MURF proteins and TRIM42 and even a higher peak in TRIFIC, corresponding to sequence identity C-terminal to the coiled coil. Like MID1 and MID2 from the C-I subfamily, the Murf proteins have an ability to form both homo- and heterodimers with each other and to associate with the microtubule cytoskeleton (33, 34). Wider heterodimerization between these subfamilies has not been investigated but is unlikely in light of the findings presented in this study. Although loss of mutation of the MID1/MID2 B30.2-like domain is associated with loss of cytoskeletal binding (4, 8, 12, 35), the structural and localization similarities with the Murf proteins (C-II subfamily), which do not possess the B30.2-like domain, suggest that another region of the protein is directly responsible for microtubule binding and that the impact of mutation in the B30.2-like domain is indirect. In this regard, partitioning of the coiled coil from the rest of the RBCC domain in members of these subfamilies has shown that it is an essential domain required for both dimerization and cytoskeletal association (5, 34). Reanalysis of all C-I subfamily members using the MultiCoil program (30) aided delineation of the coiled coil from the conserved 67-amino acid region adjacent to the coiled coil. We have called this novel signature the COS (C-terminal subgroup one signature) box because we first identified it in the C-I subfamily. Notably, the Murf proteins and TRIM42 also show some domain similarity to the C-I subfamily across the N-terminal RBCC domain (Fig. 1B), but have little to no similarity C-terminal to the 67-amino acid signature. TRIM42 is the only other RBCC protein that we also identified as possessing an FN3 domain.
A New Subfamily of Microtubule-associated RBCC Proteins

However, TRIM42 has no C-terminal B30.2-like domain, a novel cysteine-rich motif N-terminal to the RBCC domain, an expanded region between the coiled-coil and FN3 domains (Fig. 1B), and also has low overall sequence similarity to the C-I subfamily. We have detected mouse (GenBank™ accession number NP_084495), rat (accession number AAH87151), and predicted chicken (accession number XP_422632) orthologs of human TRIM42, indicating that this indeed represents a bona fide unique and separate member of the RBCC superfamily.

A GenBank™-wide Scan for Human COS Box-containing Proteins—To further investigate the specificity of the 67-amino acid COS box, we generated an HMM data base of these sequences from the five human C-I subfamily members and available orthologs (based on the MSA depicted in Fig. 4B), and this was scanned against the GenBank™ non-redundant protein sequence data base containing all entries from GenPept, Swiss-Prot, Protein Information Resource, PDF, Protein Data Bank, and RefSeq (ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz) using HMMpfam (22). As expected, all previously identified members of the C-I, C-II, and C-III subfamilies produced significant high scores. In addition, this analysis identified an incomplete rat sequence highly related to, but distinct from, TRIM9 (score of 108.5 and E-value of 1.3 × 10⁻²⁸) as well as a recently deposited Trim9 sequence from the mosquito Anopheles gambiae (score of 121.6 and E-value of 1.5 × 10⁻³⁵). Although no human sequence orthologous to the novel rat sequence was found in GenBank™, a search of Ensemble Version 11 identified a predicted human transcript with high identity. Based on subsequent alignments with and homology to the C-I subfamily, we have re-annotated this sequence and called this gene TNL (TRIM nine-like; GenBank™ accession number AY253917) (Fig. 3B). Strikingly, the predicted human TNL protein shares high identity with human TRIM9 across its entire length (76% similarity and 65% identity) and as such represents an additional member of the C-I subfamily. Of interest, an additional sequence gave a lower but above cutoff score in the HMM analysis: GLFND (score of 14.9 and E-value of 0.0035, corresponding to amino acids 105–162). The microtubule-associated protein GLFND shares C-terminal similarity with the C-I subfamily (including coiled-coil, COS box, FN3, and B30.2-like domains), but lacks RING and B-box motifs, so is not an RBCC/TRIM protein (36–38). An HMM data base generated from the alignment of all available COS box sequences was used to perform an additional search for COS box-containing proteins in the most current human GenBank™ peptide sequence data base using the current NCBI 35 assembly available from Ensembl. This search detected all known human COS box-containing proteins, although no additional high scoring matches were revealed.

Subcellular Localization of the C-I Subfamily Members—Early analysis of MID1 and MID2 showed that both proteins associate with the microtubule network and have the ability to form homo- and hetero-multimers (5, 8, 9, 35). To investigate the subcellular distribution of the remaining C-I subfamily members, we transiently transfected either GFP- or Myc-tagged expression constructs of each available full-length protein (see “Experimental Procedures”). The localization patterns of GFP-TRIM36, GFP-Trific, and Myc-TRIM9 in COS-1 cells resembled those of GFP-MID1 and GFP-MID2 observed in simultaneous experiments, and this was confirmed by immunohistochemical detection of α-tubulin. These data indicate that all C-I subfamily proteins associate with the microtubule cytoskeleton (Fig. 5). Myc-TRIM9 was used because N-terminally GFP-tagged TRIM9 exhibited cellular localization similar to that of GFP alone when compared with immunohistochemical detection (12), indicating a disruption of normal localization (data not shown). Similar disruption of MID1 and MID2 has been seen with some tags such as FLAG.4 Although a full-length TNL sequence was not available, given the extensive sequence similarity to TRIM9, it too would be expected to have a microtubule distribution in line with that reported here for other members of the C-I subfamily. Of note, the original TRIM9 EST harboring the L653F missense change (see “Experimental Procedures”) in the B30.2 domain localized in cytoplasmic clumps (Fig. 5F). This localization is reminiscent of MID1 B30.2 domain mutants (Fig. 5G), which are associated with Opitz syndrome, and suggests that other C-I subfamily members associate with microtubules in a manner that is similarly disrupted by B30.2 domain mutation.

Functional Analysis of the C-I Subfamily: Protein-Protein Interaction—In addition to their ability to homo- and hetero-multimerize, microtubule-bound MID1 and MID2 also bind Alpha 4, a negative regulatory subunit of protein phosphatase 2A (5). Given the similar domain architecture and degree of sequence identity over specific domains such as the B-boxes, we chose to use the ProQuest yeast two-hybrid system to determine whether other members of this subfamily also show an ability to homo- and hetero-multimerize and to bind Alpha 4. The C-I subfamily members TRIM9 and TRIM36 both showed strong yeast two-hybrid interaction for homodimerization (Table 1). However, no evidence of heterodimerization or interaction with Alpha 4 could be found for any of the C-I subfamily members except MID1 and MID2 (Table 1). Interestingly, Trific was the only C-I subfamily member that did not show any self-interaction, although we cannot exclude the possibility that the particular N-terminal fusion interferes specifically with Trific protein folding and or dimerization capacity in this system, similar to that mentioned previously with the N-terminal GFP fusion to TRIM9. Although not tested, TNL is predicted to be able to both homo- and heterodimerize with TRIM9 because their level of amino acid similarity, especially across their coiled-coil domains (91% similarity and 78% identity), is comparable with that of MID1 and MID2 (83% similarity and 72% identity). The results indicate that het-

---

4 K. M. Short and T. C. Cox, unpublished data.
A New Subfamily of Microtubule-associated RBCC Proteins

TABLE 1
Protein-protein interaction between the C-I subfamily and associated proteins

|        | Gal4 DBD | MID1 | MID2 | TRIM9 | TRIM36 | Trific | Alpha 4 |
|--------|----------|------|------|-------|--------|--------|---------|
| Gal4 AD| -        | -    | -    | -     | -      | -      | -       |
| MID1   | -        | ++   | ++   | --    | -      | -      | ++      |
| MID2   | -        | ++   | ++   | --    | -      | -      | ++      |
| TRIM9  | -        | ++   | ++   | --    | -      | -      | ++      |
| TRIM36 | -        | ++   | ++   | +     | +      | -      | -       |
| Trific | -        | ++   | ++   | ++    | ++     | ++     | -       |
| Alpha 4| -        | ++   | ++   | ++    | ++     | ++     | -       |

Yeast two-hybrid analysis showed that heterodimerization was restricted to MID1 and MID2. Homodimerization was shown for all tested C-I members in both yeast two-hybrid orientations, except Trific (see “Results”). Two independent reporters were used to confirm interactions: HIS3 on 3-amino triazole (first marker) and lacZ analyzed by an X-gal assay (second marker). DBD, DNA-binding domain; AD, activation domain; +++, positive interaction; −, no interaction.

erodimerization within the C-I subfamily is restricted to those members that, through more recent gene duplication events, have enough similarity within the respective coiled-coil regions.

Analysis of a Potential Microtubule-associated Function for the COS Box—Homology to the COS box is present in the microtubule-associated MID1 and MID2 proteins (C-I subfamily), MURF1–3 (C-II), and GLFND (10, 39). Although these proteins show regions of similarity to each other, the coiled coil and COS box are the only motifs shared by all (Fig. 4A). We thus hypothesized that the new 67-amino acid COS box may, at least in part, play a role in the cytoskeletal association or, at a minimum, be a signature for microtubule-associated RBCC proteins. In this respect, microtubule association of several associated C-I and C-II proteins has been shown to be dependent on dimerization mediated by the coiled coil (2, 5, 33). In nearly all studies of homodimerization, the coiled coil has been expressed to capture full-length proteins in vitro or in yeast two-hybrid assays; yet, in all instances, we have revealed that part or all of the COS box has also been removed. This indicates that dimerization in the absence of an intact COS box is still strong (5, 34) and suggests that the COS box is not necessary for dimerization.

However, a number of studies have not disrupted COS boxes. Spencer et al. (34) generated a C-terminal truncation that removed 31 amino acids from the C terminus of MURF3, including the acid-rich domain. The mutation (amino acids 354–384 deleted) did not significantly impact on the predicted COS box (amino acids 298–354) or coiled coil (positioned N-terminal to the COS box), and it was noted that both microtubule association and homodimerization were not significantly affected by this deletion. However, removal of both the coiled coil and predicted COS box with a 154-amino acid C-terminal deletion of MURF3 abolished microtubule association and homodimerization (34). Similar studies of microtubule association in GLFND have indicated that removal of N-terminal portions of the protein (including the coiled coil, but with COS boxes intact) permits microtubule association, albeit with a lower affinity and of a whorl-like appearance (36, 37). Furthermore, GLFND without the coiled coil loses the ability to stabilize microtubules after treatment with nocodazole, which is thought to be the result of a loss of dimerization ability (36). Additionally, in these studies, fortuitous removal of regions containing the COS box motif severely inhibited the ability of the protein to associate with microtubules (36, 37). Therefore, to test our hypothesis, COS box mutations were introduced into the well characterized MID1 protein, and cellular localization patterns and homodimerization were tested in the presence of these mutations. To minimize effects on tertiary structure, two individual point mutations were generated in the COS box that replaced conserved amino acid triplets (FLQ and LDY) (highlighted in Fig. 4B and supplemental Fig. 1) with alanines. Replacement with triple alanine residues was anticipated to mildly disturb the local folding of the motif by shortening the predicted α-helical motif (data not shown). Mutation of FLQ to AAA had little effect on cellular localization patterns, with prominent microtubule association seen (Fig. 6A). In contrast, the LDY → AAA mutation resulted in a punctate speckled cytoplasmic appearance, with some short “tracks” of microtubule staining observed in a small percentage of cells (Fig. 6B). Both mutations were then introduced into MID1, and the resultant protein completely lost its ability to interact with microtubules (Fig. 6C). Consistent with predictions, the double mutant was still able to efficiently dimerize and also to bind Alpha 4 (Fig. 6D), indicating that the tertiary structure was not grossly disrupted.

FIGURE 6. Mutations within the COS box of MID1 affect microtubule localization, but not dimerization. Replacement of residues FLQ (A), LDY (B), and both FLQ and LDY (C) with AAA residues in the COS box affected the localization of GFP-MID1 (A–i–C–i) from a normal microtubule distribution (A–ii–C–ii) to one that was progressively disrupted (A–iii–C–iii) show the overlay, with Hoechst nuclear staining in blue). Co-immunoprecipitation (IP; D) revealed that the same mutations (FLQ → AAA (lanes 2 and 4), LDY → AAA (lanes 5 and 6), and both FLQ → AAA and LDY → AAA (lanes 7 and 8)) did not affect self-association (lanes 1, 3, 5, and 7) or interaction with Alpha 4 (lanes 4, 2, 4, 6, and 8).

MARCH 31, 2006 • VOLUME 281 • NUMBER 13
JOURNAL OF BIOLOGICAL CHEMISTRY 8977
A New Subfamily of Microtubule-associated RBCC Proteins

These data support a specific role for the COS box in directing microtubule association.

**Modular Role for COS Box-assisted Microtubule Association**—To test whether the COS box is sufficient to confer microtubule association properties to RBCC/TRIM proteins normally not associated with microtubules, chimeric proteins were generated with the RBCC domain (amino acids 1–252) from TRIM37/MUL and a COS box (Fig. 7A) or a COS box with the C-terminal domain from MID1 (Fig. 7, B and C). Endogenous TRIM37 has been observed to natively localize to peroxisomes (40), with overexpressed full-length TRIM37 or the TRIM domain found in cytosolic bodies and aggresomes (41, 42). Cellular localization by direct fluorescence revealed that the fusion of the COS box alone to the TRIM domain of TRIM37 redirected the localization to a small number of large punctate speckles within the cytoplasm (Fig. 7A). The disruption of cytoskeletal organization, which is similar to the effect seen with overexpressed full-length TRIM37 (42), and the low viability of transfected cells indicated that the C-terminal fusion was harmful to cells (data not shown). However, partial microtubule overlap could be seen in fixed cells (Fig. 7B, upper arrowhead), with more definitive microtubule localization observed in living cells (where microtubules could not be counterstained) (Fig. 7C). In addition, some perinuclear aggresomal localization could also be noted (Fig. 7B, lower arrowhead), which is similar to the overexpressed full-length protein (data not shown) (42). Assessment of dimerization by co-immunoprecipitation revealed that dimerization did not occur between both the fusions and either full-length MID1 or the RBCC domain of TRIM37 (data not shown), consistent with this property being conferred by the coiled-coil domain. The inability of the chimeric TRIM-COS fusion to associate with microtubules may be explained by the observation that, in all COS box-containing proteins, domains C-terminal to the COS box are present, and these may assist in folding, placement, or protection of a microtubule-interfacing structure.

**DISCUSSION**

In a recent large study, Reymond et al. (2) identified numerous new RBCC proteins encoded in the human genome. Based on initial findings using N-terminal GFP tags of some of these factors, these researchers suggested that this protein family identifies a unique array of subcellular compartments, although few structures were identified. Here, we sought to determine the relationship of all human RBCC proteins using a bioinformatics approach with a view to better understand the functions of this important family. Underpinning this approach was our previous observation of an FN3 domain in the Opitz syndrome protein, MID1, and its homolog, MID2 (8), and the subsequent finding of the same domain architecture in an additional RBCC protein, TRIM9 (12). Like the microtubule-associated MID1 and MID2 proteins, TRIM9 had been reported to associate with the cytoskeleton, yet the actual cytoskeletal elements with which it associated had not been determined (12). We therefore surmised that the domain architecture might reflect or determine a common subcellular localization with perhaps even similarities in function. As the presence of FN3 domains was not noted in the structural representations of the various RBCC/TRIM proteins described by Reymond et al. (2), we first sought to determine whether any additional RBCC family members possess an FN3 domain. Using both available and newly developed bioinformatics resources, including PSI-BLAST-, PROFILE-, and HMMER-based analyses, we undertook a sensitive screen of both the known RBCC family members and the translated current releases of the human genomic sequence. This approach uncovered three additional RBCC proteins that share the same domain organization as MID1, MID2, and TRIM9. Two of these represent products of previously unidentified genes that we have called TRIFIC and TNL. The third, TRIM36, had been reported during the course of this study and also determined its homolog, MID2 (8), and the subsequent finding of the same domain architecture in an additional RBCC protein, TRIM9 (12). Like the microtubule-associated MID1 and MID2 proteins, TRIM9 had been reported to associate with the cytoskeleton, yet the actual cytoskeletal elements with which it associated had not been determined (12). We therefore surmised that the domain architecture might reflect or determine a common subcellular localization with perhaps even similarities in function. As the presence of FN3 domains was not noted in the structural representations of the various RBCC/TRIM proteins described by Reymond et al. (2), we first sought to determine whether any additional RBCC family members possess an FN3 domain. Using both available and newly developed bioinformatics resources, including PSI-BLAST-, PROFILE-, and HMMER-based analyses, we undertook a sensitive screen of both the known RBCC family members and the translated current releases of the human genomic sequence. This approach uncovered three additional RBCC proteins that share the same domain organization as MID1, MID2, and TRIM9. Two of these represent products of previously unidentified genes that we have called TRIFIC and TNL. The third, TRIM36, had been previously reported by Reymond et al. (2), but was represented as only possessing an N-terminal RBCC/TRIM domain. Validating our findings, two other groups identified and reported TRIM36 during the course of this study and also determined that it possesses a MID1/MID2-like domain architecture that includes both C-terminal FN3 and B30.2-like domains (14, 31). Significantly, we have also shown in this study that the five tested RBCC proteins that share the same domain architecture all localize to the microtubule cytoskeleton, representing the first report of a family of structurally related B30.2-like proteins functioning in the same subcellular compartment. In addition, we have shown that most of the members form dimers with the exception of murine Trific, although this may be due to issues with the system used (as described). Based on the findings of the structural and functional similarities to these homologous proteins, we re-evaluated the domain architecture of all remaining RBCC proteins and performed similar groupings on the basis of their
C-terminal domain composition. This classification revealed nine subgroups with distinct C-terminal architectures: COS/FN3/B30.2-like domains (C-I subfamily), COS/acid-rich region (C-II), COS/FN3 domains (C-III), B30.2-like or SPRY-containing domains (C-IV), no identifiable C-terminal domain homology (C-V), PHD/BROMO domains (C-VI), immunoglobulin/NHL repeats (C-VII), MATH domain (C-VIII), and ADP-ribosylation factor domain (C-IX). Representative proteins of six of the nine subgroups have been identified in *D. melanogaster* and *Caenorhabditis elegans* (data not shown), with exceptions being the C-II, C-III, and C-VIII subfamilies. In the latter, which is defined by a C-terminal MATH domain, the domain is nevertheless present and found C-terminal to RING and coiled-coil motifs (e.g. C. elegans TRF1). Interestingly, there is no evidence from EST and genomic sequences to date to support the existence of the RBCC tripartite domain in fungi and plants despite the presence of RING, B-box, and coiled-coil motifs individually in many proteins. This also indicates that rearrangement of the separate domains into a tripartite cluster may have occurred only during or after metazoan evolution in animals. Consistent with this, TRIM9 is the only invertebrate protein with the C-I (RBCC/FN3/B30.2) motif architecture, and as such, the likely role of this protein was perhaps not necessary until multicellularization in animals. It is therefore likely that these model organisms will be useful for investigating basic functions of the different RBCC subgroups that, in turn, might aid our understanding of various human disease processes in which RBCC proteins have been implicated.

We used a HMMER-based approach in part of the human RBCC superfamily analysis. Graphical depiction of the data generated from 50-amino acid wide sliding windows of the partial C-I subfamily MSA clearly highlights the similarities between the members of the C-I subfamily and the other subfamilies. Truong and Ikura (21) have previously highlighted the significance of high scoring HMM matches from a sliding window data base to regions of structural identity within the cadherin superfamily of proteins, linking statistical analysis findings to actual structural identities. A similar unbiased approach was used in this study, highlighting areas of significant conservation within the subfamily as well as throughout the RBCC superfamily. The analysis resolved highly significant matches for TRIFIC and TNL throughout the RING finger, B-boxes, and coiled-coil regions, but also aided the definition of a high scoring region immediately adjacent to the C-terminal end of the coiled coil in a number of the RBCC subfamilies. We have called this region the COS box, and it is a region consisting of two α-helical coils (Fig. 4). To determine whether this motif supports more than just conserved positioning of the neighboring coiled-coil and FN3 domains in the C-I proteins, an HMM data base representing the COS box was used to scan the GenBank™ FASTA Protein Database. This scan identified COS boxes in the non-FN3 domain-containing MURF proteins and in the non-TRIM domain-containing protein GLFND, although all these proteins also have in common a coiled-coil domain. Previous analyses suggested that the coiled-coil domains of the RBCC proteins (MID1, MID2 and MURF3) are required for dimerization and microtubule association (5, 34). Inadvertently, in these studies, the regions encoding the COS boxes were also retained when attempts were made to isolate coiled-coil function. The other COS box-containing protein, GLFND, has also been found to associate with microtubules and has a C terminus similar to that of the C-I subfamily with coiled-coil, FN3, and B30.2 domains (36), yet it does not possess a RING or B-box. We therefore considered whether the COS motif simply provides an adjunct to the coiled-coil domain or whether it has a more direct functional activity. A role for the coiled-coil/COS box region in dimerization and microtubule association is supported by the observation that mutations in MURF3 that remove amino acids either N-terminal to the coiled coil or C-terminal to the COS box maintain their microtubule association and dimerization properties (34). However, the presence of coiled-coil domains throughout the generally non-microtubule RBCC superfamily would suggest that a coiled coil alone is not enough for microtubule association and gives support to the notion that the COS box confers microtubule binding capacity assisted by the coiled coil. To investigate this, we introduced mutations (FLQ → AAA and LDY → AAA) into two conserved regions of the COS box in MID1 (or consensus sites FIQ and lca in Fig. 4). Whereas the individual mutations alone had relatively moderate effects on microtubule localization of the protein (Fig. 6, A and B), when introduced together, both mutations completely abolished microtubule association (Fig. 6C). Notably, the mutations did not affect coiled-coil function because dimerization with wild-type MID1 was not affected (Fig. 6D). Furthermore, these mutants were still capable of interacting with Alpha 4 (which binds B-box 1), indicating that the overall RBCC tertiary structure was likely to be intact. We therefore wanted to investigate whether the addition of a COS box to a non-microtubule-localized RBCC protein would direct the chimeric protein to the cytoskeleton. Fusion of the COS box alone to the RBCC/TRIM domain from the Mulibrey nanism protein, TRIM37, altered the localization of overexpressed TRIM37 to cytoplasmic clumps. Notably, however, fusion of a larger COS box C-terminal fragment permitted the chimeric protein to associate with cytoskeletal structures (Fig. 7C). These data indicate that the COS box, which is found only in the C-I, C-II, and C-III RBCC/TRIM proteins and GLFND, acts independently of the other general dimerization properties of the coiled coil to mediate direction of these proteins to the microtubule cytoskeleton. However, the association of the coiled coil with the COS box in all these proteins (including the chimeric TRIM37-COS protein) indicates a potential need of these two domains together for normal microtubule association and that these proteins all bind microtubules through a similar mechanism. Our bioinformatics-based approach used to subclassify the entire human RBCC complement (based on C-terminal domain composition) may thus assist in understanding the compartmentalization of the various RBCC subfamilies and in turn may better reflect the evolutionary and functional relatedness of the diverse members of this large family of proteins.
A New Subfamily of Microtubule-associated RBCC Proteins

9. Caimarca, S., Messali, S., Ballabio, A., and Meroni, G. (1999) *Hum. Mol. Genet.* 8, 1387–1396.
10. Trockenbacher, A., Suckow, V., Foerster, J., Winter, J., Krauss, S., Ropers, H. H., Schneider, R., and Schweiger, S. (2001) *Nat. Genet.* 29, 287–294.
11. Quaderi, N. A., Schweiger, S., Gauldenz, K., Franco, B., Ragaller, E. I., Berger, W., Feldman, G. J., Volta, M., Andolfi, G., Gilgenkrantz, S., Marion, R. W., Hennekam, R. C., Opitz, J. M., Muenke, M., Ropers, H. H., and Ballabio, A. (1997) *Nat. Genet.* 17, 285–291.
12. Li, Y., Chinn, L. S., Weigel, C., and Li, L. (2001) *Science* 299–310.
13. Berti, C., Messali, S., Ballabio, A., Reymond, A., and Meroni, G. (2002) *Trends Biochem. Sci.* 27, 403–405.
14. Kitamura, K., Tanaka, H., and Nishimune, Y. (2003) *J. Cell Sci.* 116, 159–162.
15. Henikoff, S., and Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 10915–10919.
16. Goodstadt, L., and Ponting, C. P. (2001) *Bioinformatics* 17, 845–846.
17. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998) *Nucleic Acids Res.* 26, 293–294.
18. Benner, S. A., Cohen, M. A., and Gonnet, G. H. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 6404–6408.
19. Notredame, C., Higgins, D. G., and Heringa, J. (2000) *Bioinformatics* 16, 897–908.
20. Felsenstein, J. (2004) *Trends Biochem. Sci.* 159–162.
21. Truong, K., and Ikura, M. (2002) *Exp. Cell Res.* 276, 24242–24252.
22. Eddy, S. R. (1998) *Nucleic Acids Res.* 26, 293–294.
23. Elble, R. (1992) *BioTechniques* 13, 18–20.
24. Myers, A. M., Tzagoloff, A., Kinney, D. M., and Lusty, C. J. (1986) *Gene* (Amst.) 45, 299–310.
25. Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Barrett, D., Bateman, A., Birney, E., Biswas, M., Bradley, P., Bork, P., Buheir, E., Copley, R. R., Courcelle, E., Das, U., Durbin, R., Falquet, L., Fleischmann, W., Griffiths-Jones, S., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krzywaninova, M., Lopez, R., Letunic, I., Lonsdale, D., Silventoinen, V., Orchard, S. E., Pagni, M., Peyruc, D., Ponting, C. P., Selengut, J. D., Servant, F., Sigrist, C. J., Vaughton, R., and Zdobnov, E. M. (2003) *Nucleic Acids Res.* 31, 315–318.
26. Letunic, I., Copley, R. R., Schmidt, S., Ciccarelli, F. D., Deeks, T., Schultz, J., Ponting, C. P., and Bork, P. (2004) *Nucleic Acids Res.* 32, D142–D144.