Biochemical and Biophysical Research Communications **273,** 364–369 (2000)
doi:10.1006/bbrc.2000.2925, available online at http://www.idealibrary.com on IDEAL®

Isolation of Two Novel Human RhoGEFs, *ARHGEF3* and *ARHGEF4,* in 3p13-21 and 2q22

Signe Thiesen,* S. Kübart,* H.-H. Ropers,* and H. G. Nothwang*,†
**Max Planck Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin-Dahlem, Germany; and*
†Abteilung Tierphysiologie, Universität Kaiserslautern, D-67653 Kaiserslautern, Germany

Received May 2, 2000

RhoGEFs play an important role in various signaling cascades and are implicated in human conditions like cancer and mental retardation. A database search combined with screening of a human neuronal teratocarcinoma library identified two novel RhoGEFs, *ARHGEF3* and *ARHGEF4* (HGMW-approved symbols). The widely expressed *ARHGEF3* transcript of 3561 nucleotides encodes a polypeptide of 526 amino acids with homology to NET1. The *ARHGEF4* gene generates two transcripts of 3665 and 4000 nucleotides that translate into 720 amino acid residues. Expression of *ARHGEF4* is restricted to brain and the encoded protein shows homology to collybistin. FISH analysis of genomic clones mapped *ARHGEF3* to 3p13-21 and *ARHGEF4* to 2q22.

Rho-like GTPases play an important role in many cellular processes like cytoskeletal rearrangements, transcriptional activation, regulation of cell morphology and cell aggregation, cytokinesis, endocytosis and secretion [1]. They are furthermore involved in oncogenic transformation and more recently in human disorders like mental retardation [2, 3]. Rho-GTPases are activated by binding GTP and inactivated by conversion of GTP to GDP by their intrinsic GTPase activity. Mainly three types of proteins regulate this cycling between active and inactive conformation. GTPase activating proteins (GAPs) stimulate the intrinsic GTPase activity, and guanine nucleotide dissociation inhibitors (GDIs) bind both to GTP-bound and GTP-bound Rho-GTPases and thereby stabilize the respective conformation. Finally, guanine nucleotide exchange factors (GEFs) catalyze the release of bound GDP that is replaced by GTP. Several studies indicate that the main form of regulation is exercised by RhoGEFs [4]. Their large size enables them to contain several functional domains that can integrate various signals.

Two domains are contained in all RhoGEF proteins, a *Dbl* homology (DH) and an adjacent pleckstrin homology (PH) domain that present in their tandem organization a hallmark for RhoGEFs [5]. Individual members feature in addition other domains or motifs like a *Src* homology 3 (SH3) or a PDZ domain. To identify novel members, an EST database search was combined with the screening of a neuronal teratocarcinoma library. This resulted in the identification of two novel genes, *ARHGEF3* and *ARHGEF4*, with homology to the known family members NET1 and collybistin [6, 7]. Further characterization revealed different expression patterns and alternative splicing for *ARHGEF4*. To determine their chromosome localization, genomic clones were identified and analyzed by FISH.

GenBank Accession No. for *ARHGEF3* is AF249744 and for *ARHGEF4* is AF249745.

MATERIALS AND METHODS

Library screening and sequence analysis. To obtain cDNA clones, a Uni Zap XR human neuronal teratocarcinoma cDNA library (Stratagene) was hybridized under low stringency conditions (50°C instead of 65°C in a solution containing 1 mM EDTA, 0.25 M NaCl, 0.125 mM Na_2HPO_4 (pH 7.0), 10% polyethylene glycol 6000, 7% SDS) and filters were washed at 50°C in phosphate buffer (40 mM Na_2HPO_4). Bluescript KS(+) plasmids were isolated by *in vivo* excision according to the manufacturer's instruction. To obtain genomic clones, a PAC and a BAC filter library (RPCIP704 and RPCIB753, German Resource Centre) containing genomic human DNA inserts were screened by using standard protocols [8]. Double-stranded sequencing of cDNA clones was performed with labeled universal M13 primers and the Thermo sequenase fluorescent labeled primer cycle sequencing kit (ABI) on a LICOR sequencer (MWG-Biotech). Sequences were determined on both strands. Sequence analysis was performed with BLAST [9]. Human brain total RNA was purchased (Clontech) and poly(A)$^+$-RNA was obtained by subsequent purification on magnetic beads using an mRNA isolation kit (Boehringer-Mannheim). For synthesis of cDNA, approximately 40–200 ng of purified mRNA was randomly primed by using SuperScript II RNase H-reverse transcriptase (Gibco-BRL). Reverse transcription was carried out for 40 min at 42°C, 10 min at 45°C, and 10 min at 50°C. Negative controls without reverse transcriptase were run under

0006-291X/00 $35.00

identical conditions. 5′-RACE was performed by using essentially the protocol of Frohman [10] with some minor modifications. After reverse transcription, the cDNA was purified by using QIAquick spin columns (QIAGEN). For second strand synthesis, the primer CTAATACGACTCACTATAGGGCTCGAGCGGCTTTTTTTTTTTTTTTTTTV was used. PCR was performed with gene specific primers and AP1 CTAATACGACTCACTATAGGGC and AP2 ACTCACTATAGGGCTCGAGCGGC at an annealing temperature of 60°C (1 min) and extension at 72°C (2 min). PCR amplification was performed on 1/50 of the original RT reaction in a final volume of 50 μl containing 40 pmol of each primer and 1 U of *Taq* polymerase. One-fifth of the PCR product was analyzed on 1.5% agarose gels with 0.5 μg/mL ethidium bromide in 1× TAE.

Northern blot analysis. An adult multiple tissue total RNA (Invitrogen) was purchased. The blot was hybridized in Ultrahyb solution (Ambion) with indicated ^{32}P-labeled cDNA probes for 15 h at 65°C. Hybridized filters were washed for 2× 30 min in 2× SSC, 0.1% SDS and 30 min in 0.1× SSC, 0.1% SDS at 65°C. Autoradiography took 16–40 h at −70°C using two intensifying screens.

Fluorescence in situ hybridization (FISH). Standard FISH protocols were followed [11]. The PAC clones were labeled with biotin-14-dATP by using a standard nick translation procedure [12]. Metaphase spreads were prepared using standard procedures [12]. Hybridization was done for 45 h in a moist chamber at 37°C. Immunocytochemical detection of the hybridization probes was achieved as described elsewhere [12]. Images were taken with a Zeiss epifluorescence microscope equipped with a thermoelectronically cooled charge coupled device (CCD) camera (Photometrics CH250), which was controlled by an Apple Macintosh computer. Merging and pseudocoloring were accomplished using the ONCOR Image and ADOBE Photoshop software.

RESULTS

Identification of Two Novel RhoGEF Genes

In order to identify novel RhoGEF gene family members, an EST database search was initiated by using the amino acid sequence of the DH domain as query in the tblastn program. This algorithm compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames. About 70 human ESTs were identified encoding this hallmark of RhoGEFs. Most of them were derived from known RhGEFs. 7 EST clones, EST26703, EST56928, EST28207, EST00232, 784813, 166961, 1009218 (GenBank Accession Nos. AA323861, AA349948, AA325176, AA448233, M62170, R89026, AA225792) showed only similarity but no identity to known RhoGEFs. Further sequence analysis demonstrated that clones EST26703 and EST56928 were derived from the same transcript (ARHGEF3 cluster).

Several RhoGEFs contain large 3′ untranslated regions. It was therefore possible that some DH domains escaped detection in the database search, because the available EST sequence did not reach the DH domain. We therefore performed a screening of 2.5×10^6 plaques of a neuronal teratocarcinoma cDNA library with nucleotide probes encoding the DH domain of *ECT2*, *P1115-RhoGEF*, *KIAA0006*, and of clone EST56928 (GenBank Accession Nos. AA206473, U64105, D13631, AA349948 [13, 14]) under low stringency conditions. 11 clones were identified and sequenced. 6 were identical to known genes. Among them, only one clone represented a RhoGEF, *NET1A* [6], whereas the other clones were derived from genes without a DH domain: *TRA1* (3 clones), *EF1α*, and *GAP43* (GenBank Accession Nos. X15187, X16869, M25667). One clone had no homology to any database entry, and 4 clones (9i, 19i, 40i, 42i,) showed similarity to RhoGEFs. Clone 42i with *ECT2*, 19i with *KIAA0942* (GenBank Accession No. AB023159), and clones 9i and 40i revealed homology with *NET1*. The latter two clones in addition overlapped with the cluster ARHGEF3, identified by the database search.

Since clones 9i and 40i added sequence information to the ARHGEF3 cluster and since M62170, identified by the database search, was expressed in brain (see below), these two putative RhoGEF transcripts were assembled to full-length transcripts. Several rounds of 5′-RACE experiments on RNA from adult human brain resulted in two composite full-length cDNAs, *ARHGEF3* and *ARHGEF4* (M62170 based).

Characterization of ARHGEF3

The composite *ARHGEF3* transcript of 3561 nucleotides (nts) contains an open reading frame (ORF) from nucleotide 128 to 1708 and a large 3′ untranslated region (UTR) of 1853 nts (Fig. 1A). The initiation codon ATG is imbedded in a strong Kozak consensus sequence (CCGCCATGG) [15]. To test the expression of *ARHGEF3*, a multiple tissue Northern blot analysis was performed. In order to avoid cross hybridization of the nucleotide sequences encoding the DH and PH domains with other RhoGEFs, a probe representing the 3′ UTR was used. The probe detected widespread expression in heart, brain, kidney, lung, pancreas, spleen and skeletal muscle with a transcript size of about 3800 nts (Fig. 2A). This conforms to a poly(A) tail of about 200 bases and confirms that *ARHGEF3* represents the full-length transcript. The ubiquitous expression was confirmed by an EST database analysis that identified 65 *ARHGEF3* derived ESTs from various tissues: spleen, brain, colon, connective tissue, ear, germ cells, kidney, lymph node, muscle, pancreas, parathyroid, pineal gland, prostate, retina, testis, uterus, and whole blood. No apparent alternative splicing products were observed. During the preparation of this manuscript, a patented sequence (GenBank Accession No. AX002254.1) was entered into the database that showed identity to *ARHGEF3*, but no further information was provided.

The ORF encodes a 526 amino acid (aa) protein with a calculated molecular weight of 59,782 and a calculated isoelectric point (p*I* value) of 6.03. The only domains in the polypeptide are the DH and the PH do-

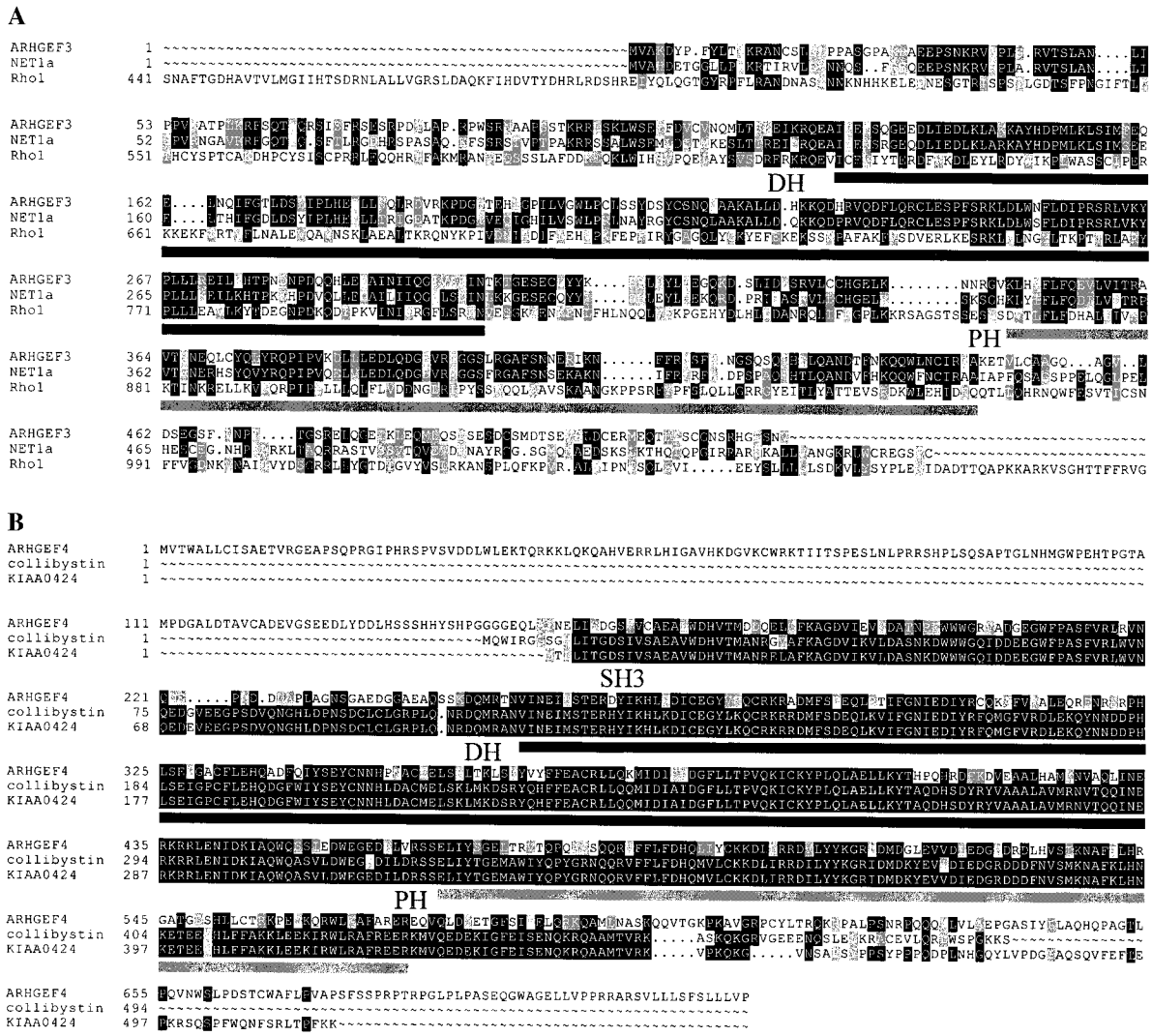

FIG. 1. Sequence alignment of the novel RhoGEFs with similar family members. Amino acid sequences were compared using the pileup program. Conserved amino acid residues are shaded. Numbers at left indicate amino acid position and DH, PH, and SH3 domains are underlined. (A) Sequence similarity of ARHGEF3 to human NET1 and yeast rho1 (*Saccharomyces pombe*) (GenBank Accession Nos. U02081 and AL049498). (B) Sequence similarity of ARHGEF4 to human KIAA0424 and its rat ortholog collybistin (GenBank Accession Nos. AB007884 and AJ250425).

mains, characteristic of RhoGEFs. A homology search revealed identity of 64% and similarity of 82% on the amino acid level to human NET1A and 23% identity and 42% similarity to rho1 of *Schizosaccharomyces pombe* (Fig. 1A).

To localize the gene within the human genome, a human genomic BAC clone library was screened with the 3′ UTR of *ARHGEF3* and 4 positive BAC clones identified: A214, B234, D0819, and B06174. Restriction and southern blot experiments confirmed that these clones contained *ARHGEF3* (data not shown). FISH analysis of these clones mapped *ARHGEF3* to 3p13-p21 (Fig. 3A). This confirms radiation hybrid mapping data that localized the gene between markers D3S1289 and D3S1547 on 3p [16].

Characterization of ARHGEF4

ARHGEF4 was assembled to a full length transcript of 3665 nts. Furthermore, two alternative splice variants were observed during cloning and sequence analysis. 5′-RACE and RT-PCR analysis resulted in an alternative 5′ UTR sequence with a 335 nucleotide insertion at position 412 (splice variant I, data not shown). This insertion likely represents an additional exon, since a PCR using a forward primer upstream of position 412 and a reverse primer downstream of this position did amplify this insertion from reverse transcribed brain RNA, but not from genomic DNA (data not shown).

Comparison with the partial cDNA clone KIAA11112 revealed that the latter contains an additional 591-

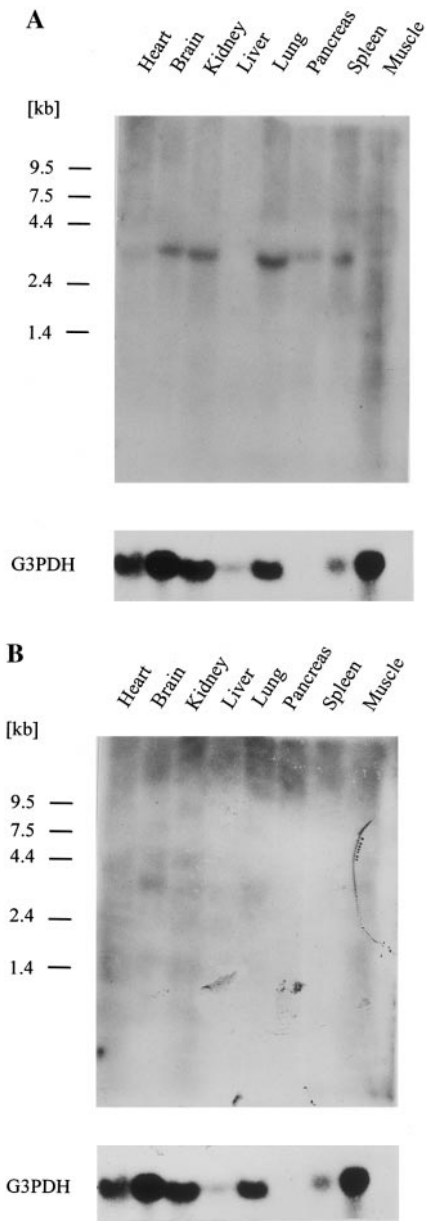

FIG. 2. Northern blot analysis of *ARHGEF3* and *ARHGEF4*. Hybridization of the 3′ UTR of *ARHGEF3* (A) or *ARHGEF4* (B) to an adult multiple tissue Northern blot containing total RNA (Invitrogen). As a control, glyceraldehyde 3-phosphate dehydrogenase (G3PDH) cDNA was hybridized to the Northern blot.

nucleotide insertion at position 2441 of *ARHGEF4* (splice variant II). This insertion corresponds to the last intron of *ARHGEF4*, as revealed by PCR and sequence analysis (data not shown). None of 6 EST clones (IMAGE clones 47898, 51711, ai60154, and DKFZ clones p578O01144 and p434G2016) analyzed by us contained this insertion. RT-PCR analysis did also not result in amplification of the insertion present in clone KIAA11112. It is therefore most likely that the clone variant KIAA11112 corresponds to an incomplete spliced mRNA or is expressed at a very low amount. Northern blot analysis indicated weak expression of a transcript of 3800 nucleotides in brain (Fig. 2B). This size corresponds well with the length of *ARHGEF4* (3665 nts). Faint bands were also observed in kidney, lung, and muscle (Fig. 2B). The predominant expression in brain was supported by RT-PCR data, available for the partial cDNA clone KIAA11112, that showed weak expression in adult and fetal brain, and very weak expression in testis [17]. Furthermore, 27 out of 36 *ARHGEF4* derived human ESTs, identified in the dbEST database, originated from brain tissues (data no shown).

ARHGEF4 contains three AUG at positions 521, 584, and 620, preceded by a stop codon at position 347. None of the ATGs is embedded in a Kozak consensus sequence, but the ATG at position 620 fulfills the minimum criteria of G^{+4} and a purine in position -3 [15]. This initiation codon would results in an ORF encoding a 720 amino acid polypeptide with a calculated molecular weight of 81,780 and a p*I* of 6.36. The 591 nts insertion at position 2441 would led to an altered

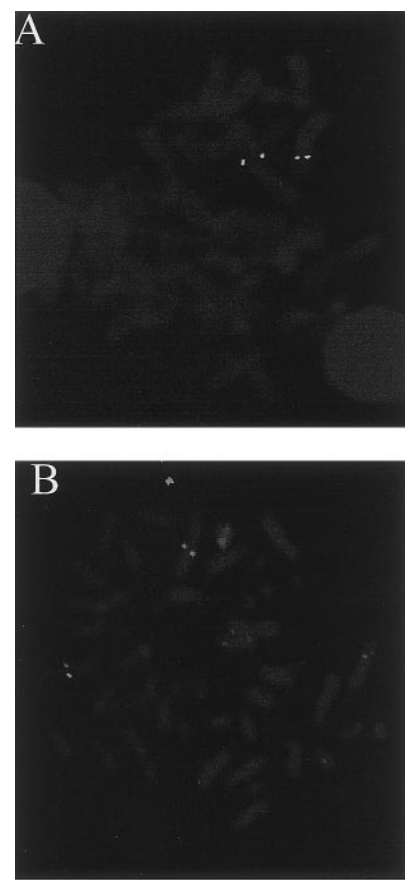

FIG. 3. FISH mapping of *ARHGEF3* and *ARHGEF4*. Fluorescence *in situ* hybridization of BAC clone D0819 containing *ARHGEF3* (A) to 2q22 and PAC clone L05585 containing *ARHGEF4* (B) to 3p13-21.

C-terminus of ARHGEF4 with a shorter polypeptide containing an alternative carboxyl terminus of 29 novel amino acid residues instead of the last 112 aa of ARHGEF4.

ARHGEF4 contains the DH and the PH domains, and in addition a SH3 domain at the N-terminal. A similarity search revealed high homology to human KIAA0424 (58% identity, 73% similarity), and its rat ortholog collybistin (62% identity, 79% similarity), a brain specific RhoGEF involved in receptor localization in neurons [7] (Fig. 1B).

Hybridization of the 3′ UTR of *ARHGEF4* to a human genomic PAC library identified 3 clones, L05585, H24814 and P2189 that contained *ARHGEF4*. The clones were verified by restriction and Southern blot experiments to contain *ARHGEF4* (data not shown). All three clones mapped to human 2q22 (Fig. 3B). This localization is in good agreement with radiation hybrid data that mapped an *ARHGEF4* derived EST to the interval D2S2215-D2S132 on 2q22-q23 [16].

DISCUSSION

We report on the identification and isolation of two novel RhoGEFs, *ARHGEF3* and *ARHGEF4*. ARHGEF4 shows strong similarity of 82% with NET1, another ubiquitously expressed RhoGEF. NET1 exhibits focus forming activity, and NIH3T3 cells transfected with a *NET1* expression construct showed altered growth properties *in vitro*, and were tumorigenic when injected into nude mice. *NET1* thus reveals features of an oncogene. Interestingly, *ARHGEF3* maps to a chromosomal region, 3p21-p13 that harbors 6 translocations and approximately 70 deletions, that are recurrently associated with tumors [18]. The identification of several genomic clones containing *ARHGEF3* that are suitable for FISH analysis will help to elucidate whether *ARHGEF3* is affected by any of these rearrangements and plays a role in tumorigenesis. However, an EST database survey did not indicate an increased expression of *ARHGEF3* in cancer cell lines (data not shown).

ARHGEF4 revealed significant similarity of 79% to collybistin, a recently isolated RhoGEF that plays a crucial role in gephyrin-mediated postsynaptic submembrane clustering of $GABA_A$ and glycine receptors [7]. Collybistin, that is predominantly expressed in the brain, might therefore play an important role in the formation of inhibitory postsynaptic membrane formation and plasticity [7]. Interestingly, *ARHGEF4* shows also brain specific expression and it will therefore be interestingly to analyze whether ARHGEF4 displays similar function as collybistin. Furthermore, recent studies on mental retardation revealed the involvement of several RhoGTPase interacting proteins in the etiology of this disorder. *Oligophrenin-1* encodes a RhoGAP, *αPIX* a RhoGEF, and *PAK3* (p21 activating kinase) encodes a downstream effector of RhoGTPases [19–21]. An analysis of the Mendelian Cytogenetic Network database (dbMCN, http://mcndb.imbg.ku.dk/index.php) indicated that more than 10 breakpoints associated with MR map to 2q22. FISH analysis of the PAC clones containing *ARHGEF3* will reveal whether any of these translocations affects this gene.

ACKNOWLEDGMENTS

We thank C. Menzel, S. Schneider, D. Vogt, P. Viertel, and S. Zacharias for expert technical assistance. We are grateful to the German Resource Center for providing genomic PAC and BAC clones. This work was supported by the Bundesministerium für Bildung und Forschung (Grant 4763 to H. H. Ropers).

REFERENCES

1. Stam, J. C., and Collard, J. G. (1999) The DH protein family, exchange factors for Rho-like GTPases. *In* Cytoskeleton and Small G Proteins (Jeanteur, P., Kostovic, I., and Kuchino, Y., Eds.), pp. 51–83, Springer-Verlag, Berlin/Heidelberg/New York.
2. Van Aelst, L., and D'Souza-Schorey, C. (1997) Rho GTPases and signaling networks. *Genes Dev.* **11,** 2295–2322.
3. Chelly, J. (1999) Breakthroughs in molecular and cellular mechanism underlying X-linked mental retardation. *Hum. Mol. Genet.* **8,** 1833–1838.
4. Feig, L. A. (1994) Guanine-nucleotide exchange factors: A family of positive regulators of ras and related GTPases. *Curr. Opin. Cell. Biol.* **6,** 204–211.
5. Whitehead, I. P., Campbell, S., Rossman, K. L., and De, C. J. (1997) Dbl family proteins. *Biochim. Biophys. Acta* **1332,** F1–23.
6. Chan, A. M. L., Takai, S., Yamada, K., and Miki, T. (1996) Isolation of a novel oncogene, NET1, from neuroepithelioma cells by expression cDNA cloning. *Oncogene* **12,** 1259–1266.
7. Kins, S., Betz, H., and Kirsch, J. (2000) Collybistin, a newly identified brain-specific GEF, induces submembrane clustering of gephyrin. *Nat. Neurosci.* **3,** 22–29.
8. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A laboratory Manual, Cold Spring Harbor Laboratory Press, New York.
9. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.
10. Frohman, M. A. (1993) Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: Thermal RACE. *Methods Enzymol.* **218,** 340–356.
11. Ward, D. C., Boyle, A., and Haaf, T. (1995) Fluorescence *in situ* hybridization techniques. Metaphase chromosomes, interphase nuclei, and extended chromatin fibers. *In* Human Chromosomes. Principles and Techniques in Human Chromosomes: Principles and Techniques (Verma, R. S., and Babu, A., Eds.), pp. 184–192, MacGraw–Hill, New York.
12. Wirth, J., Nothwang, H. G., van der Maarel, S., Menzel, C., Borck, G., Lopez-Pajares, I., Brondum-Nielsen, K., Tommerup, N., Bugge, M., Ropers, H. H., and Haaf, T. (1999) Systematic characterization of disease associated balanced chromosome rearrangements by FISH: Cytogenetically and genetically an-

chored YACs identify microdeletions and candidate regions for mental retardation genes. *J. Med. Genet.* **36,** 271–278.

13. Miki, T., Smith, C. L., Long, J. E., Eva, A., and Fleming, T. P (1993).) Oncogene ect2 is related to regulators of small GTP-binding proteins. *Nature* **362,** 462–465.
14. Hart, M. J., Sharma, S., e Masry, N., Qiu, R. G., McCabe, P., Polakis, P., and Bollag, G. (1996) Identification of a novel guanine nucleotide exchange factor for the Rho GTPase. *J. Biol. Chem.* **271,** 25452–25458.
15. Kozak, M. (1987) At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* **196,** 947–950.
16. Deloukas, P., Schuler, G. D., Gyapay, G., Beasley, E. M., Soderlund, C., Rodrigueztome, P., Hui, L., Matise, T. C., Mckusick, K. B., Beckmann, J. S., Bentolila, S., Bihoreau, M. T., Birren, B. B., Browne, J., Butler, A., Castle, A. B., Chiannilkulchai, N., Clee, C., Day, P. J. R., Dehejia, A., Dibling, T., Drouot, N., Duprat, S., Fizames, C., Fox, S., *et al.* (1998) Physical map of 30,000 human genes. *Science* **282,** 744–746.
17. Kikuno, R., Nagase, T., Suyama, M., Waki, M., Hirosawa, M, and Ohara, O. (2000) HUGE: A database for human large proteins identified in the Kazusa cDNA sequencing project. *Nucleic Acids Res.* **28,** 331–332.
18. Mitelman, F., Mertens, F., and Johansson, B. (1997) A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nat. Genet.* **15,** 417–474.
19. Billuart. P., Bienvenu, T., Ronce, N., des Portes, V., Vinet, M. C., Zemmi, R., Crollius, H. R., Carrié, A., Fauchereau, F., Cherry, M., Briault, S., Hamel, B., Fryns, J. P., Beldjord, C., Kahn, A., Moraine, C., and Chelly, J. (1998) *Oligophrenin-1* encodes a rhoGAP protein involved in X-linked mental retardation. *Nature* **392,** 823–826.
20. Kutsche; K., Yntema, H., Brandt, A., Jantke, I., Nothwang, H. G., Orth, U., Boavida, M. G., David, D., Chelly, J., Fryns, J. P., Moraine, C., Ropers, H. H., van Bokhoven, H., and Gal, A. (2000) Mutations in *αPIX*, encoding a guanine nucleotide exchange factor for RhoGTPases, in patients with X-linked mental retardation. *Med. Genetik* **12.** [Abstract W7–6]
21. Allen, K. M., Gleeson, J. G., Bagrodia, S., Partington, M. W., MacMillan, J. C., Cerione, R. A., Mulley, J. C., and Walsh, C. A. (1998) PAK3 mutation in nonsyndromic X-linked mental retardation. *Nat. Genet.* **20,** 25–30.