Available online at www.sciencedirect.com
ScienceDirect
Genomics 90 (2007) 249–260
GENOMICS
www.elsevier.com/locate/ygeno

Identification of three novel proteins (SGSM1, 2, 3) which modulate small G protein (RAP and RAB)-mediated signaling pathway

Hao Yang [a], Takashi Sasaki [a], Shinsei Minoshima [a,b], Nobuyoshi Shimizu [a,*]

[a] *Department of Molecular Biology, Keio University School of Medicine, Tokyo 160-8582, Japan*
[b] *Department of Medical Photobiology, Photon Medical Research Center, Hamamatsu University School of Medicine, Shizuoka, Japan*

Received 29 November 2006; accepted 26 March 2007
Available online 23 May 2007

Abstract

We report a novel protein family consisting of three members, each of which contains RUN and TBC motifs and appears to be associated with small G protein-mediated signal transduction pathway. We named these proteins as small G protein signaling modulators (*SGSM1/2/3*). Northern blot analysis revealed that human *SGSM2/3* are expressed ubiquitously in various tissues, whereas *SGSM1* is expressed mainly in brain, heart, and testis. Mouse possessed the same protein family genes, and the *in situ* hybridization and immunohistochemical staining of tissue sections revealed that mouse *Sgsm1/2/3* are expressed in the neurons of central nervous system, indicating the strong association of *Sgsm* family with neuronal function. Furthermore, endogenous Sgsm1 protein was localized in the *trans*-Golgi network of mouse Neuro2a cells by immunofluorescence microscopy. Expression of various cDNA constructs followed by immunoprecipitation assay revealed that human SGSM1/2/3 proteins are coprecipitated with RAP and RAB subfamily members of the small G protein superfamily. Based on these results, we postulated that the SGSM family members function as modulators of the small G protein RAP and RAB-mediated neuronal signal transduction and vesicular transportation pathways.

Keywords: Chromosome 22; RUN motif; TBC motif; SGSM; Small GTPase RAP; Small GTPase RAB; Intracellular signaling; Vesicle transportation

The G protein is a general term for guanine nucleotide binding proteins with GTPase activity comprising many similar but distinct proteins which are associated with a variety of cellular functions such as differentiation, proliferation, and intracellular molecular mechanisms. The small G proteins are low-molecular-weight GTPases including the RAS superfamily which serve as molecular switches in the pathways of signal transduction and vesicular transport by shuttling between GTP-bound active and GDP-bound inactive forms. Furthermore, these small G protein-mediated signaling pathways affect each other and such regulation is considered essential for maintenance of cellular homeostasis [1–3].

The RAP family is a subfamily of the RAS superfamily and consists of four members, RAP1A, RAP1B, RAP2A, and RAP2B. All members of RAP family have a highly conserved amino acid sequence (60–90% homology) and are regulated through exchange reactions controlled by guanine nucleotide exchange factors (GEFs) or GTPase-activating proteins (GAPs) [4]. The RAP proteins are shown to play key roles in various cellular phenomena, such as integrin-mediated cell adhesion [5–7], cadherin-mediated cell junction formation [8,9], actin dynamics [10–12], differentiation of neuronal cells [13,14], and synaptic plasticity and Wnt/β-catenin signaling pathway [15–17].

The RAB family is the largest subfamily of RAS superfamily and is composed of over 60 members [18]. Similar to RAP proteins, RAB proteins act as molecular switches by interchanging between active and inactive forms. RAB family proteins are localized on various intracellular membranes, where they play key roles in regulating the vesicle formation and its motility, docking, and fusion. Some RAB family members are correlated with particular cellular functions such as synaptic transmission [19,20], endosomal regulation [21–26], traffic between *trans*-Golgi network (TGN) and plasma membrane [27], peripheral distribution of melanosomes [28], and secretion in cytotoxic T lymphocytes [29].

* Corresponding author. Fax: +81 3 3351 2370.
E-mail address: shimizu@dmb.med.keio.ac.jp (N. Shimizu).

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2007.03.013

In 1999, we finished the DNA sequencing of the euchromatic region of chromosome 22 as the first significant achievement of the Human Genome Project with international collaboration [30]. Since then, we have continued the comprehensive analysis of all the probable/putative genes to confirm their existence by expression profiles, cDNA isolation, and protein characterization [31–33]. In this paper, we report a novel gene family, small G protein signaling modulator (*SGSM*), consisting of three members: *SGSM1* on chromosome 22q11.2, *SGSM2* on chromosome 17p13.3, and *SGSM3* on chromosome 22q13.1. All three *SGSM* proteins possess both RUN domain (a motif commonly found in RPIP8, UNC-14, and NESCA) and TBC domain (a motif commonly found in Tre-2, BUB2p, and Cdc16p), which are reported as domains for binding to RAP proteins and RAB proteins, respectively. The SGSMs are expressed in the neuronal cells of the central nervous system (CNS) and are found to be associated with all the members of RAP family and several members of RAB family. Furthermore, endogenous SGSM1 protein is localized at TGN of neuronal cells. Based on these findings, we postulated that the *SGSM* family members act as modulators of RAP and RAB-mediated cellular signaling.

Results

Cloning, genomic structure, and motif prediction of SGSM family

In our previous computer analysis of human chromosome 22 DNA sequence, we predicted a novel gene located between two genes *PIWIL3* and *CRYBB2* in 22q11.2 [30]. Applying RT-PCR and rapid amplification of cDNA ends (RACE) to human adult brain cDNA library, we isolated the full-length cDNA of this novel gene and determined its sequence (Accession No. **AB275761**). This novel gene was composed of 25 exons with 3.3-kb open reading frame (ORF) encoding a protein of 1093 amino acid residues (Fig. 1). To determine whether any related family genes exist in the human genome, we searched for the homologous genes by BLAST against the public databases [34]. Subsequently, we found two homologs, *KIAA0397* (Accession No. **BC039204**) on 17p13.3 and *DJ1042K10* (Accession No. **CR456441**) on 22q13.1. We determined the full-length cDNA sequences of these two genes and found that these genes encode proteins of 1051 and 749 amino acid residues, respectively. Motif analysis revealed that these three proteins possess RUN and TBC motifs in common. Interestingly, the order of two motifs was inverted in SGSM3 to which additional SH3 motif was inserted (Fig. 1). Based on the high sequence homology and the functional features demonstrated below (see text in details), we designated these three genes/proteins as the members of a novel family small G protein signaling modulator (SGSM).

Comparison of genomic structure

The exon/intron structure revealed that the length of most exons in *SGSM1* is the same as that of the corresponding exons in *SGSM2*. Therefore, *SGSM1/2* would have been generated from the same ancestral gene by gene duplication during the evolutionary process. Because of a peculiar difference in the genomic structure in *SGSM3* and *SGSM1/2*, we also compared

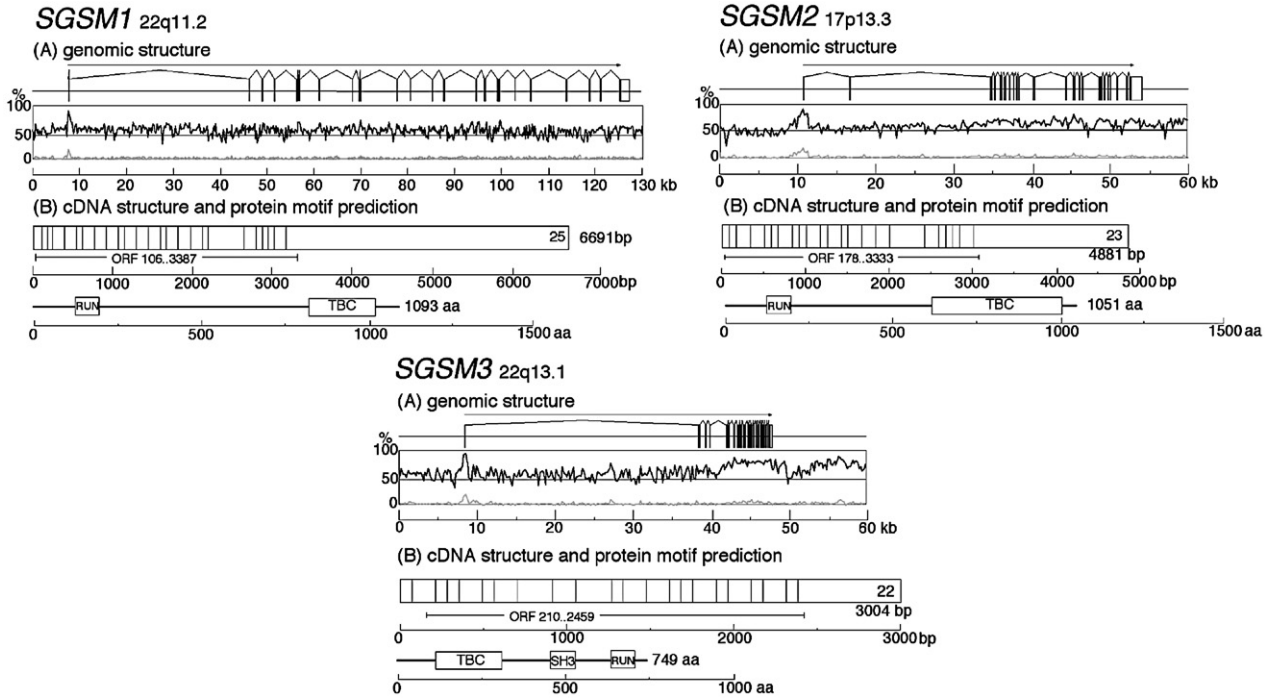


Fig. 1. Genomic structures, cDNA structures, and motif predictions of *SGSM* gene family. Three *SGSM* genes are shown with chromosomal localization and the following features. (A) Vertical bars indicate the positions of exons and arrows indicate the direction of transcription. Black and gray lines indicate GC and CpG contents, respectively. (B) Boxes indicate the size of exons and horizontal bars indicate the ORF in cDNA structure. Motif prediction was carried out using Pfam and is indicated by boxes. RUN and TBC motifs were predicted in all SGSM proteins. These two motifs are inverted in SGSM3 and SH3 motif was found.

their protein amino acid sequence and the length of each exon but there was no evidence for direct duplication.

SGSM gene family in other species

To identify the orthologs of human *SGSM* genes from other species, we carried out homology search using BLAST in the Ensembl database. The *SGSM1/2/3* orthologs were found in *Takifugu rubripes*, *Drosophila melanogaster*, and *Mus musculus*. Interestingly only two genes *SGSM1* and *SGSM3* were identified in *Caenorhabditis elegans*. Because the motif architecture is different in *SGSM3* and *SGSM1/2*, the multiple alignment by CLUSTAL W was carried out separately and the phylogenetic trees were drawn using TreeView (Fig. 2) [35]. These phylogenetic trees indicate that a primordial gene of vertebrate *SGSM* had been duplicated into *SGSM1/2* before the divergence of pisces and tetrapod. *SGSM1/2* were also identified in *Drosophila*, but these two genes were clustered. Therefore, *Drosophila SGSM1/2* may have been generated by gene duplication in the fly lineage.

Tissue expression profiles of SGSM genes

To confirm the expression profiles in human and mouse tissues, we performed Northern blot hybridization using the partial sequence of ORFs of human *SGSM1/2/3* and mouse *Sgsm1/2/3*. In human, *SGSM2/3* were expressed in a variety of organs and tissues, whereas *SGSM1* was expressed mainly in brain, heart, and testis (Fig. 3A). In mouse, the expression patterns of *Sgsm2/3* are basically the same as these of human *SGSM2/3*. However, *Sgsm1* was expressed only in adult mouse brain. Furthermore, the expression analysis at different developmental stages revealed that *Sgsm2/3* are expressed in all the developmental stages, but *Sgsm1* is expressed only in later stages of development (Fig. 3B).

The transcripts of human *SGSM2/3* were detected as 9.4 kb for *SGSM2* and 4.0 kb for *SGSM3*. The transcripts of mouse *Sgsm1/2/3* were detected in different sizes, with particular size differences notable between adult and embryo: 7.0 and 5.0 kb for *Sgsm1*, 6.0 and 4.4 kb for *Sgsm2*, and 3.4 and 3.0 kb for *Sgsm3*. These transcript variants must be generated by the alternative splicing in different tissues or at different developmental stages.

We also carried out RT-PCR using human cDNAs derived from different tissues and obtained similar expression profiles with Northern blot analysis (data not shown). These results suggested that *SGSM1* is mainly expressed in brain, and hence this gene might be associated with the CNS function. To confirm localization of *SGSM1* in the CNS, we carried out Northern blot analysis using human mRNAs prepared from different parts of human brain and spinal cord. In fact, *SGSM1* transcript of 7.0 kb was detected in most parts of human CNS (Fig. 3C).

Expression profiles in mouse brain

The transcripts of *Sgsm* family, especially *Sgsm1* in mouse, were expressed predominately in brain. To obtain a more detailed expression profile of *SGSM* family in brain, we performed *in situ* hybridization with adult mouse brain tissue

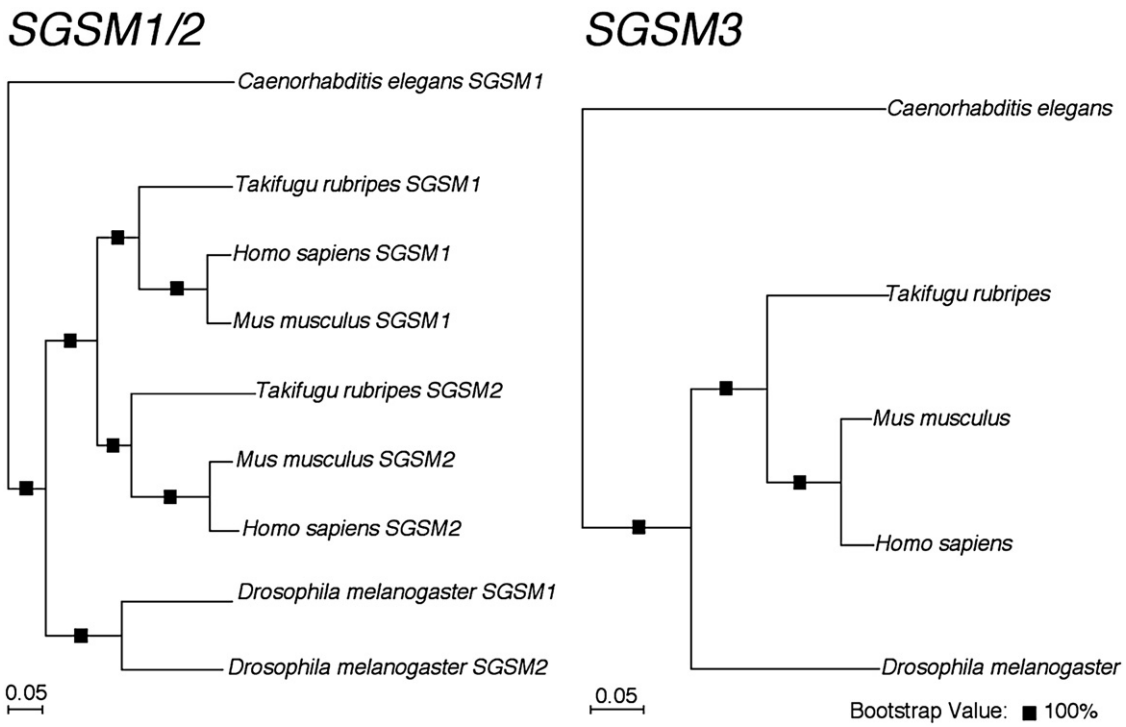


Fig. 2. Phylogenetic trees of SGSM1/2 and SGSM3 family proteins. The deduced amino acid sequences of *SGSM* orthologs were analyzed for *Homo sapiens*, *Mus musculus*, *Takifugu rubripes*, *Caenorhabditis elegans*, and *Drosophila melanogaster*. According to the relative order of RUN and TBC motifs, these three proteins were divided into two groups, SGSM1/2 and SGSM3. Phylogenetic trees were calculated and drawn using CLUSTAL W and TreeView. Horizontal lengths of branches are proportional to the extent of sequence divergence from the common ancestor.

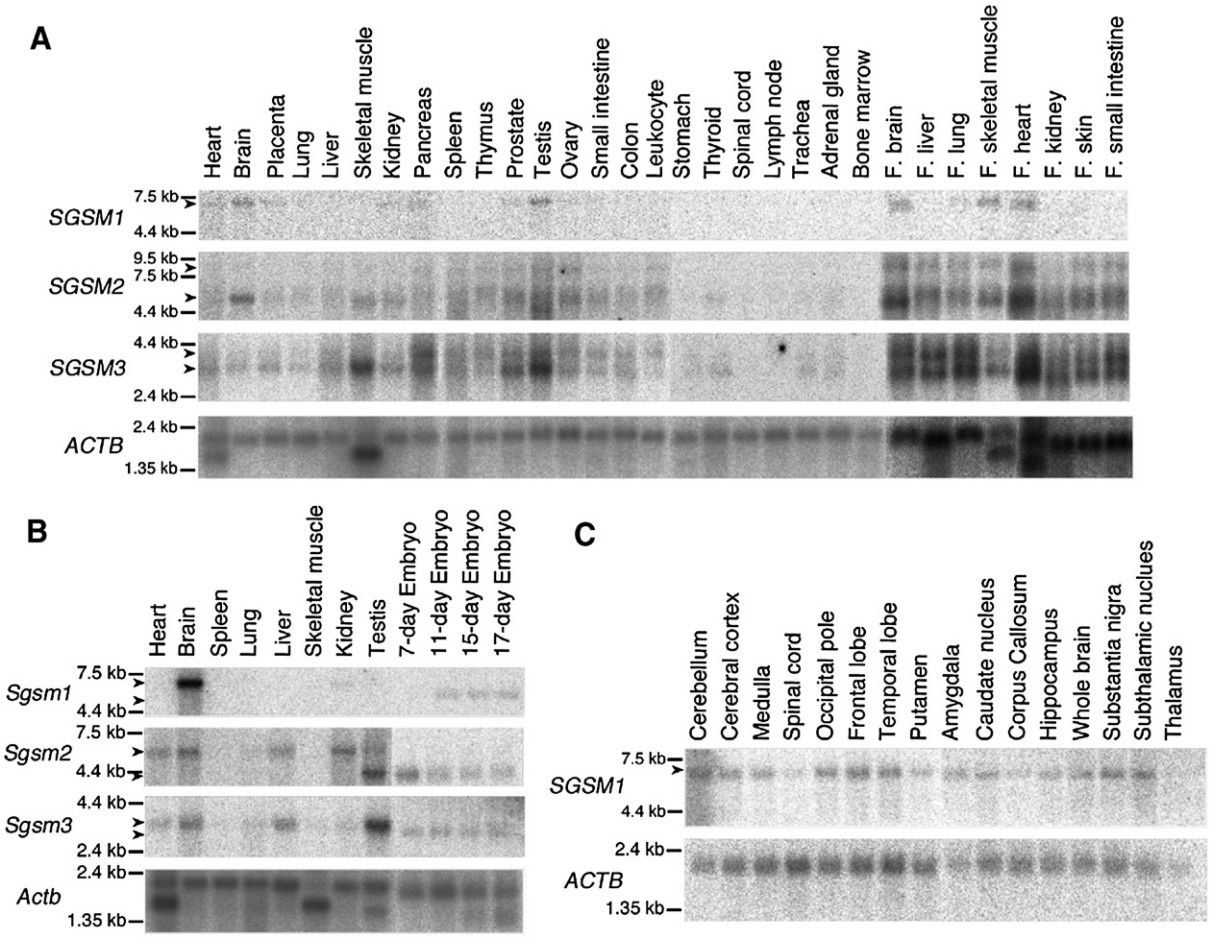


Fig. 3. Expression profiles of the members of *SGSM* family in human and mouse. (A) Northern blot analyses using human multiple tissues/organs (*SGSM1/2/3*). The lengths of transcripts of human *SGSM* genes are 7.0 kb for *SGSM1*, 6.0 and 9.4 kb for *SGSM2*, and 3.4 and 4.0 kb for *SGSM3*. (B) Northern blot analyses using mouse multiple tissues/embryos (*Sgsm1/2/3*). The lengths of transcripts of mouse *Sgsm* genes are 5.0 and 7.0 kb for *Sgsm1*, and 4.4 and 6.0 kb for *Sgsm2*, and 3.0 and 3.4 kb for *Sgsm3*. (C) Northern blot analyses using human brain multiple tissues (*SGSM1*). A single transcript of human *SGSM1* was detected at 7.0 kb in most tissues. The hybridization with human β-*ACTIN* (*ACTB*) or mouse β-*Actin* (*Actb*) probe was used as the control.

sections. The transcripts of *Sgsm1* gene were detected by antisense RNA probes in the layer of Ammon's horn (CA1, CA2, and CA3) and dentate gyrus (DG) of hippocampus (Hip), granular layer (GL) and Purkinje cell layer (PL) of cerebellum (Cb), and neuronal cells of cerebral cortex (Cx) (Figs. 4A–4F). The sense RNA probes exhibited no specific signals (Figs. 4D′–4F′). We obtained similar expression profiles with the RNA probes for *Sgsm2/3* (data not shown). To confirm the expression of Sgsm1 protein in CNS, immunohistochemistry was carried out using anti-SGSM1/Sgsm1 antibody (Figs. 4G–4I). The expression profile of Sgsm1 protein is similar to that of *Sgsm1* transcript in mouse brain sections. These results indicate that *Sgsm* family genes are expressed in brain and associated with the neurons of CNS.

SGSM family proteins interact with RAP family proteins via novel RAP interacting domain (RAPID) motif

All members of SGSM family proteins possess one RUN motif. The exact function of RUN motif is still unknown, however, several proteins containing RUN motif are known to interact with some small GTPase, namely RAP family proteins [36,37]. For example, Rap2 interacting protein 8 (RPIP8) contains a RUN motif in the region 44–188 amino acids (aa) to which RAP2 interacts as a putative effector for G protein signaling. The N-terminal region (1–298 aa) of RPIP8 is necessary for the interaction with GTP-bound form of RAP2 [37]. Our study with Dot matrix analysis of amino acid sequence revealed that a RUN motif-containing region (16–330 aa) of SGSM1 has a significant homology to RPIP8 (data not shown). We also obtained similar results for SGSM2/3. These results suggested that SGSM proteins would also interact with RAP2. We performed immunoprecipitation assay on the lysate extracted from Neuro2a cells using anti-SGSM1 antibody and identified endogenous RAP1 and RAP2 as proteins interacting with SGSM1 protein (Supplementary Fig. 1). However, anti-RAP1 antibody could not distinguish between RAP1A and RAP1B, and anti-RAP2 antibody could not distinguish between RAP2A and RAP2B. Therefore, we analyzed the possible interaction between SGSM family and RAP family proteins using 293FT cells in which FLAG-tagged SGSM1/2/3 and HA-tagged expression constructs of RAP1A/1B/2A/2B were

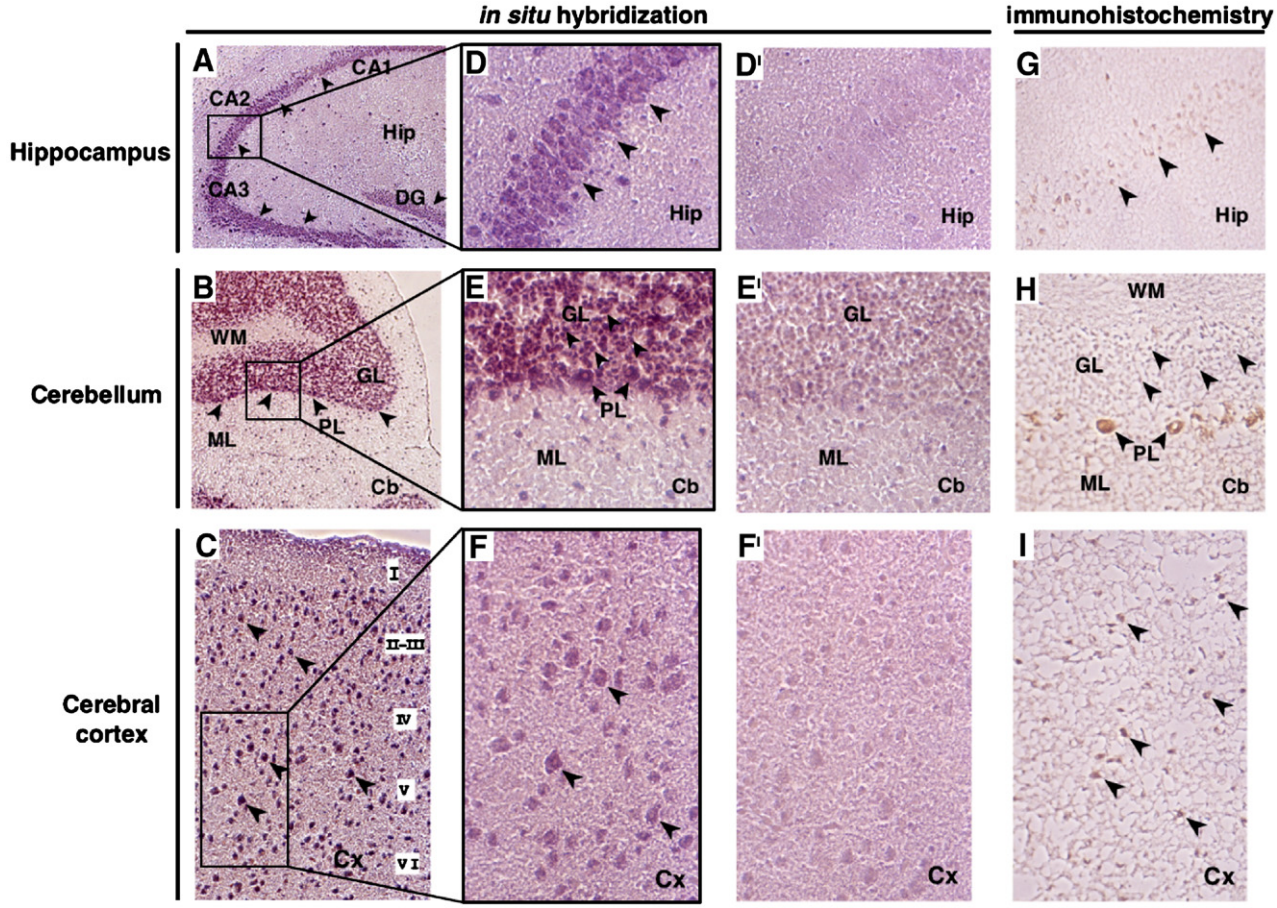


Fig. 4. *In situ* hybridization and immunohistochemistry analyses of *Sgsm1* in mouse brain. Horizontal sections of adult mouse brain were used. For the *in situ* hybridization, the magnifications were 50× (A–C) and 200× (D–F, D′–F′). Arrows indicate positive signals using antisense DIG-labeled RNA probes (A–F). For immunohistochemistry analyses, the magnification was 100× (G–I). Arrows indicate positive signals using anti-Sgsm1 antibodies. DG, dentate gyrus; Hip, hippocampus; GL, granular layer; PL, Purkinje cells layer; Cb, cerebellum; Cx, cerebral cortex; ML, molecular layer; WM, white matter.

transfected. Then, the immunoprecipitation was performed with anti-HA tag antibody for the lysates of these transfectants, and the precipitates were analyzed by Western blotting with anti-FLAG tag antibody (Fig. 5). All members of FLAG-tagged SGSM proteins were coimmunoprecipitated with all members of HA-tagged RAP1A/1B/2A/2B. Essentially the same results were obtained when anti-FLAG tag antibody was used for immunoprecipitation and anti-HA tag antibody was used for Western blotting (data not shown). These results indicate that all members of SGSM family bind to all members of RAP family.

To identify the binding site of SGSM1 protein for RAP family protein, we produced the deletion series of FLAG-tagged SGSM1 and transfected these together with HA-tagged RAP2A to 293FT cells (Fig. 6A). The coimmuoprecipitation results indicated that intact SGSM1 and the N-terminal deletion mutants of SGSM1, namely Del1 (201–1093 aa) and Del2 (301–1093 aa), bind to HA-tagged RAP2A, whereas all the N-terminal deletion mutants truncated over 350 aa, namely Del3-6, lost their binding ability to HA-tagged RAP2A (Fig. 6B). We obtained the same results when these SGSM1 mutants were transfected with the other RAP family members (data not shown). These results suggested that the 301–350 aa region of SGSM1 is necessary for the binding to RAP family members. To confirm the presence of this binding site, we also carried out coimmunoprecipitation using the C-terminus-deleted FLAG-tagged SGSM1 with HA-tagged RAP2A (Fig. 6C). This experiment supported the idea that the 301–350 aa region of SGSM1 is essential for the binding to RAP family members. These results strongly indicated that the binding site is not RUN motif but the adjacent region (301–350 aa). To examine whether this binding site sequence is present in other related proteins, we compared the amino acid sequences of RUN-containing proteins of human and other species using the CLUSTAL W program. We found five additional blocks of conserved sequence (1–5) in the neighboring region of RUN motif (Fig. 7). Therefore, we defined these conserved blocks as a novel motif RAPID which acts as an interacting site between RUN-containing proteins and RAP proteins. Especially, the 301–350 aa region is essential for the binding between SGSM1 and RAP proteins, so that the 301–350 aa region is considered the “core” of RAPID motif (Fig. 7).

SGSM family proteins interact with several members of RAB family

The small GTPase RAB family in human is composed of over 60 members and they are believed to function as key

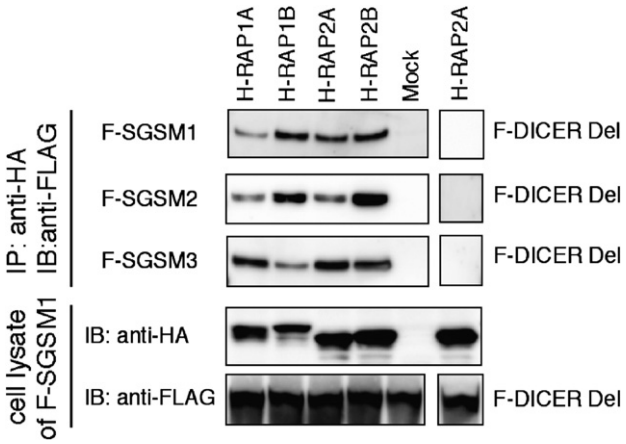


Fig. 5. SGSM family proteins interact with small GTPase RAP proteins. The expression constructs of FLAG-tagged SGSM and HA-tagged RAP cDNAs were cotransfected in 293FT cells and the cell lysates were processed for immunoprecipitation using mouse anti-HA antibody. Western blot analyses were performed using rabbit anti-FLAG antibody. Cotransfection of p3XHA-N vector with FLAG-tagged SGSM and HA-tagged RAP2A with FLAG-tagged DICER deletion mutant 1–1075 aa were used as negative control. The expression of tagged proteins in cell lysates was confirmed by antibodies for each tag. The results for SGSM1 are shown as representative. F, FLAG-tagged; H, HA-tagged; DICER del, DICER deletion mutant of 1–1075 aa; IP, immunoprecipitation; IB, immunoblotting.

regulators in intracellular vesicle transportation [18]. In this system, a series of proteins containing TBC motif has been identified as RAB-associated proteins and it was demonstrated that TBC motif acts as a catalytic site for GDP/GTP exchange reaction and interaction site for the GTP-bound RAB [40,41,46]. Based on these findings, we assumed that SGSM family proteins interact with RAB proteins through TBC motif. We then carried out coimmunoprecipitation analysis to examine the possible interaction between SGSM family and RAB family members using Neuro2a cells which are derived from mouse neuroblastoma of spinal cord and have been shown to have significant expression of various RAB proteins (data not shown). FLAG-tagged SGSM constructs were transfected into Neuro2a cells and the cell lysates of transfectants were processed for immunoprecipitation using anti-FLAG tag antibody. To detect the endogenous RAB proteins in the immunoprecipitates, Western blotting analysis was performed using anti-RAB antibodies against each of human RAB3, 4, 5, 8, 11, and 27. As a result, the interaction of SGSM proteins with several RAB proteins was confirmed, but the pattern of interaction is apparently different depending on the combination of SGSM and RAB members (Fig. 8). RAB 4/11 interacted with all the members of SGSM and RAB 3/5/8 interacted with SGSM1 and SGSM3; however, RAB27 did not interact with SGSM family proteins. This indicates a certain specificity in the interaction between SGSM members and RAB members.

SGSM1 is localized in trans-Golgi network in neuronal cells

To determine the intracellular localization of endogenous SGSM1 protein, Neuro2a cells were analyzed by immunofluorescence microscopy. The endogenous SGSM1 protein was detected mainly at a Golgi-like apparatus in the cytosol (Figs. 9F–9J). To further confirm this observation, we carried out dual immunofluorescence assay using antibodies against two Golgi markers, GM130 as a marker for *cis*-Golgi and p230 as a marker for *trans*-Golgi network (Figs. 9A and 9B) [42]. Fluorescent images showed that SGSM1 is colocalized with p230 and not with GM130, indicating the TGN location of SGSM1 (Figs. 9K′–9L′).

Subcellular localizations of other RAP and RAB proteins such as RAP1, RAP2, RAB3, RAB4, RAB5, RAB8, RAB11, and RAB27 were also analyzed using various antibodies. The endogenous RAP1 was detected strongly in the nucleus and weakly in the cytoplasm (Fig. 9C), whereas the endogenous RAP2 was localized only in the cytoplasm (Fig. 9D). Endogenous RAB3, RAB4, RAB5, RAB11, and RAB27 were also localized in the cytoplasm (data not shown). The endogenous RAB8 was detected in a region near the nucleus (Fig. 9E). Thus, the majority of RAP and RAB proteins are localized in the cytoplasm, so that it is reasonable to assume that SGSM1 was occasionally detected together with RAP and RAB family proteins at multiple locations in the cytoplasm (Figs. 9M′–9O′). Therefore, we conclude that SGSM1 interacts with RAP or RAB family proteins in the cytoplasm.

Discussion

Recent studies have provided abundant evidence that the modulation of various signaling pathways plays critical roles for growth, proliferation, and differentiation of cells. It is shown that the small G proteins such as RAP and RAB proteins are the key molecules in intracellular signal transduction and vesicle transportation. In this study, we identified a novel protein family SGSM consisting of three members and demonstrated that they bind to RAP and RAB family proteins. Furthermore, it was found that *SGSM1* gene is mainly expressed in the CNS, and at the subcellular level *SGSM1* protein is localized at the TGN. These results suggested that SGSM family members would play key roles as novel mediators in the RAP-signaling pathway and RAB-associated intracellular vesicle transportation.

SGSM genes exist in animal species with multicellular systems

Through the homology search against public databases, we found the *SGSM*-like genes (orthologs) in many animals such as nematode (*C. elegans*), fly (*D. melanogaster*), fish (*T. rubripes*), and mouse (*M. musculus*); however, we could not find any orthologs in plant, fungi, and protozoan. All the animal SGSM proteins possessed both RUN and TBC motifs. In addition, we identified a novel motif RAPID near RUN motif in most of the animal RUN motif-containing proteins.

Further database search revealed that the TBC-containing proteins exist ubiquitously in eukaryotes, whereas the RUN-and-RAPID-containing proteins are present only in animals with multicellular systems. Therefore, we presume that the TBC motif is involved in the fundamental functions for both plant and animal cells, whereas the RUN and RAPID motifs are involved in the intracellular signaling essential for

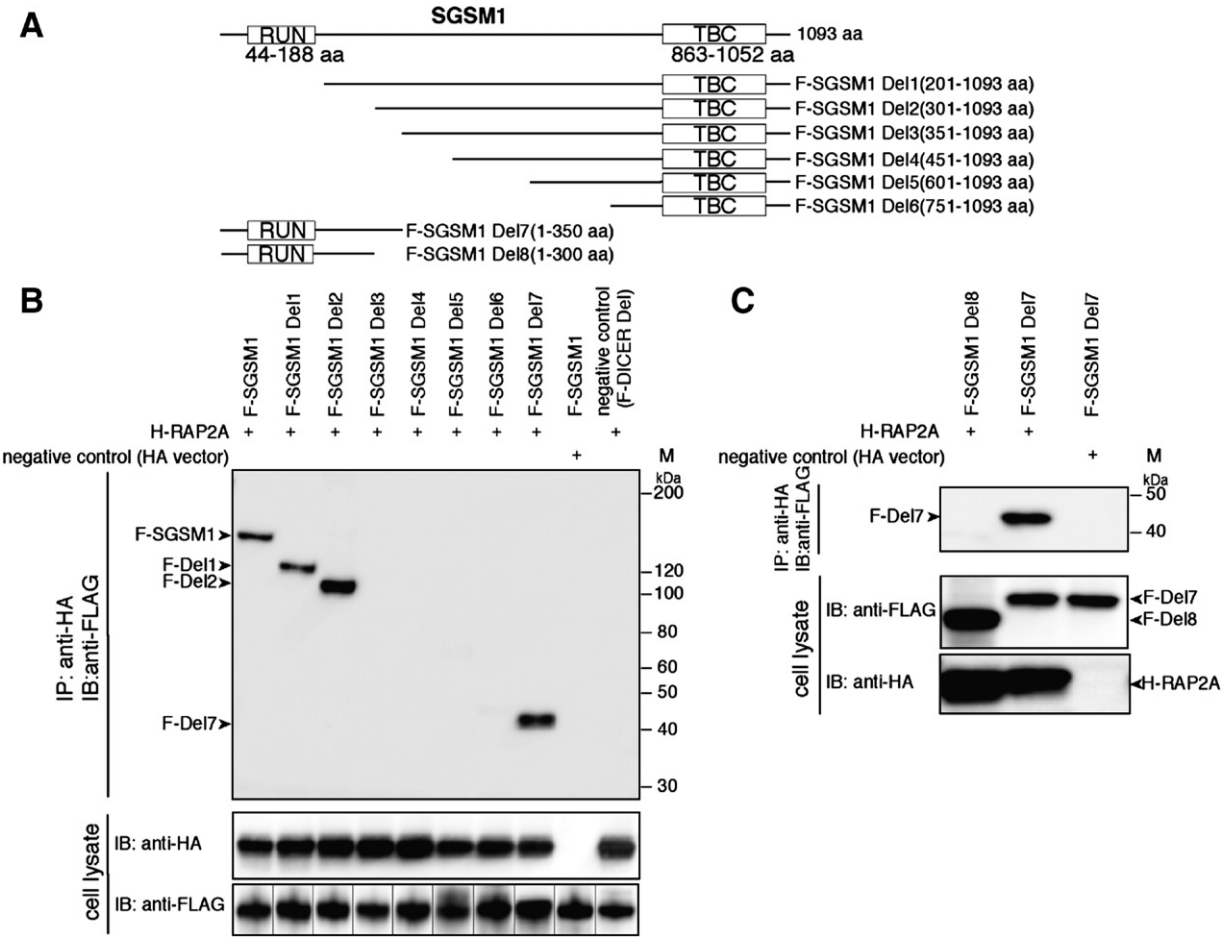


Fig. 6. Interaction analysis between RAP2A and SGSM1 deletion mutants. (A) A series of SGSM1 deletion mutants was constructed and tagged with 3×FLAG at N-terminal (F-SGSM1 Del 1–8). (B) and (C) The HA-tagged RAP2A protein was coexpressed in 293FT cells with a series of SGSM1 deletion mutants and analyzed by immunoprecipitation and Western blotting. p3XHA-N vector with FLAG-tagged SGSM1 and HA-tagged RAP2A with FLAG-tagged DICER Del were used as negative controls. F, FLAG-tagged; H, HA-tagged; DICER del, DICER deletion mutant of 1–1075 aa; IP, immunoprecipitation; IB, immunoblotting.

multicellular animal systems. Since SGSM proteins possess all these TBC, RUN, and RAPID motifs, it is possible that they function as modulators between RUN/RAPID-mediated signaling and TBC-mediated signaling to maintain vesicle transportation regulated by various intracellular and extracellular signals.

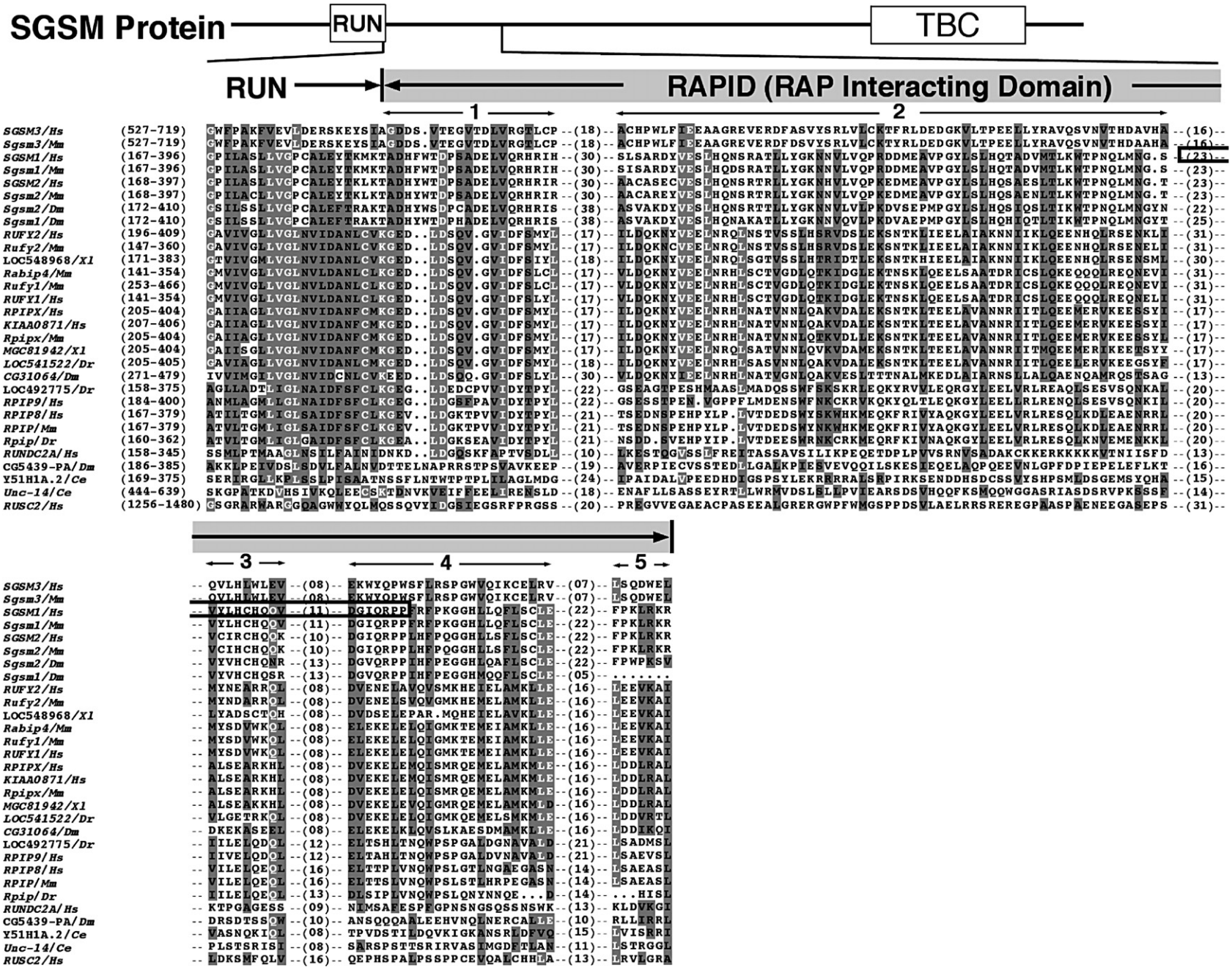
SGSM proteins are involved in intracellular transportation in neuronal cells

Human *SGSM2/3* and mouse *Sgsm2/3* are expressed ubiquitously in both human and mouse tissues, whereas human *SGSM1* and mouse *Sgsm1* are expressed in limited tissues, especially in brain. Human *SGSM2/3* and mouse *Sgsm2/3* are expressed in brain, but human *SGSM1* and mouse *Sgsm1* appear more specifically related to the function of CNS. Because mouse *Sgsm1* was expressed only at the later stages of embryonic development, we suspected that it may be related to the formation of ectoblast from which brain develops. We confirmed the human *SGSM1* and mouse *Sgsm1* expression in the neurons of CNS. Furthermore, the Sgsm1 protein was localized at TGN in Neuro2a cells. Based on these findings, we speculated that SGSM1 may play a role in modulating two distinct G protein-associated pathways, namely signal transduction by RAP family and vesicle transportation by RAB family in brain. Because human *SGSM2/3* and mouse *Sgsm2/3* are expressed ubiquitously, it is possible that they are involved in similar functions in other tissues and organs.

It is known that GAPs generally exhibit different specificity to effector proteins. In this study, we determined that SGSM1/3 interact with RAB3/5/8, which function in sorting between endosome, Golgi apparatus, and plasma membrane [20,23,24,27], whereas SGSM2 interact with RAB4 and RAB11, which are known to function in sorting and recycling of vacuoles between early endosome and plasma membrane [21,22,25]. This suggests the specificity in the interaction between SGSM family proteins and RAB family proteins. Therefore we deduced that SGSM proteins would affect several processes of intracellular transportation by interacting with distinct RAB members.

SGSM members mediate interaction between intracellular signaling pathway and vesicle transportation

All three members of SGSM family possess RUN and TBC motifs as the common conserved domains. Several proteins


SGSM Protein
RUN
TBC
RUN
RAPID (RAP Interacting Domain)
1
2
3
4
5
SGSM3/Hs
Sgsm3/Mm
SGSM1/Hs
Sgsm1/Mm
SGSM2/Hs
Sgsm2/Mm
Sgsm2/Dm
Sgsm1/Dm
RUFY2/Hs
Rufy2/Mm
LOC548968/Xl
Rabip4/Mm
Rufy1/Mm
RUFY1/Hs
RPIPX/Hs
KIAA0871/Hs
Rpipx/Mm
MGC81942/Xl
LOC541522/Dr
CG31064/Dm
LOC492775/Dr
RPIP9/Hs
RPIP8/Hs
RPIP/Mm
Rpip/Dr
RUNDC2A/Hs
CG5439-PA/Dm
Y51H1A.2/Ce
Unc-14/Ce
RUSC2/Hs

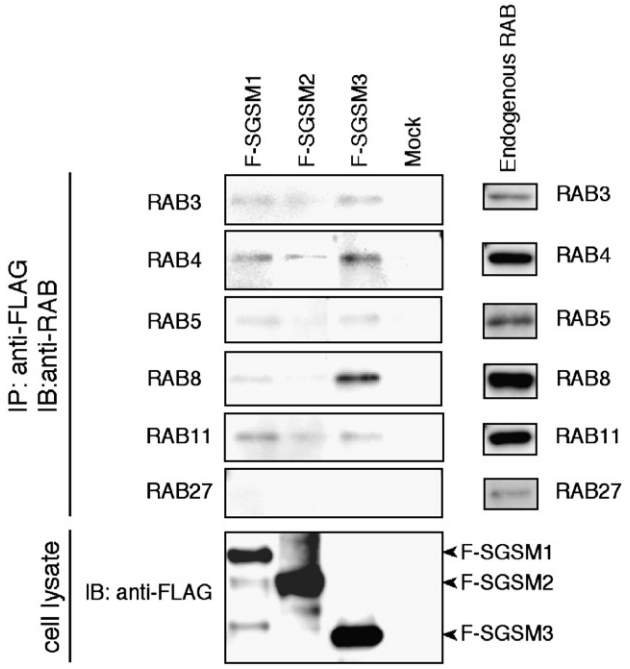


Fig. 8. SGSM family proteins interact with several members of RAB family proteins. FLAG-tagged SGSM members were expressed in Neuro2a cells. Immunoprecipitation was performed using anti-FLAG antibody and eluted with FLAG peptides. Endogenous RAB proteins were detected by anti-RAB antibodies specific for RAB3, 4, 5, 8, 11, and 27. p3XFLAG-7.1 vector was used as negative control. F, FLAG–tagged; IP, immunoprecipitation; IB, immunoblotting.

with RUN motif, such as RABIP4 (Rab4-interacting protein), RAB6IP1 (Rab6 interacting protein 1), and RPIP8, are shown to interact with RAP and RAB proteins, however, RUN motif was not the binding site to RAB proteins [36–39,43]. Moreover, the region (1–298 aa) containing RUN motif (60–189 aa) in the RPIP8 protein was proven to be necessary for the interaction with RAP2 [37]. Our coimmunoprecipitation assay demonstrated that RUN motif itself is not necessary for the association with RAP proteins; instead, a novel motif RAPID was identified as a true interaction site between SGSM and RAP proteins. In most of the animal RUN-containing proteins, the amino acid sequences of RAPID motif were conserved in the RUN-motif-neighboring region. Therefore, the interaction between SGSM family and RAP proteins would be mediated through RAPID domain. Furthermore, it is known that RAP1 and RAP2 share most effectors, such as GEFs and GAPs [4]. Our results confirmed that the SGSM family proteins show no selectivity for RAP1 and RAP2. As described above, RUN motif was found in the proteins which interact with the small GTPase RAP and RAB proteins, but it is not always required for the binding itself. For example, the RUN motif of RABIP4 is not involved in the binding to RAB4/5/11, but it is essential for targeting RABIP4 to its intracellular localization [39]. The RUN motif in new molecule containing SH3 at the carboxy terminus (NESCA), which is expressed in brain and is involved in the neurotrophin-dependent neurite outgrowth, is essential for the nuclear translocation of NESCA [44]. On the other hand, the RUN motif of UNCoordinated family member 14 (UNC-14) is necessary for its binding to UNC-16 to regulate the transportation or localization of synaptic vesicles [45]. Therefore, we postulate that RUN motif plays a role in spatial regulation for transferring target protein to its proper intracellular location, especially in the neuronal cells. Since most RUN motifs are located near RAPID motif, this regulation may be mainly associated with RAP-mediated signaling pathway.

In addition to RUN and RAPID motifs, SGSM proteins possess TBC motif. The TBC motif is found in most GAP proteins, and it has been shown to act as a catalytic site for GDP/GTP exchange reaction and interaction site for the GTP-bound RAB [40,41,46]. From the genome-wide search, at least 60 RABs and more than 50 GAPs were found in the human genome [18]. By immunoprecipitation of endogenous RAB proteins, we determined that all SGSM family members interact with several selected members of RAB family with different binding abilities. RAB proteins play key roles in intracellular vesicle transportation. Thus, it is suggested that SGSM proteins are involved in the regulation of vesicle transportation pathways. It was recently reported that SGSM3 has GAP activity specific for RAB5 [46] although our results showed that SGSM3 binds strongly to RAB8 instead of RAB5.

In this study, we used only six kinds of antibodies to detect the endogenous RAB proteins and therefore further comprehensive study is necessary to explore the detailed association between the SGSM family proteins and the over 60 members of RAB family proteins.

Taking all the data together, we predict that each SGSM protein forms a complex with distinct RAP and/or RAB proteins to perform its physiological function. It has been shown that several RAP family members play important roles in the regulation of synaptic transmission through protein kinase c-Jun-N-terminal-kinase- (JNK) or 38-kDa-mitogen-activated protein kinase (p38 MAPK)-mediated signaling cascades [4,16,17]. In our study, the SGSM family members showed strong association with neuronal functions and it is possible that they function downstream of the RAP signaling pathway and thereby regulate the vesicle trafficking through interaction with different RAB proteins. Thus, we postulate that the SGSM family proteins play significant roles as novel

Fig. 7. Identification of a novel RAPID motif in *SGSM* family proteins of various species. Five conserved blocks (1–5) were identified in the neighboring region of predicted RUN motif and constitute a novel motif RAP interacting domain (RAPID). The amino acid region of 301–350 aa in human SGSM1 protein is indicated by rectangle. The numbers in the parentheses indicate the numbers of amino acid residues between blocks. White letters on black background, black letters on light gray, and black letters on dark gray show identical, similar, and conserved residues, respectively. *Hs*, *Homo sapiens*; *Mm*, *Mus musculus*; *Dm*, *Drosophila melanogaster*; *Xl*, *Xenopus laevis*; *Dr*, *Danio rerio*; *Ce*, *Caenorhabditis elegans*.

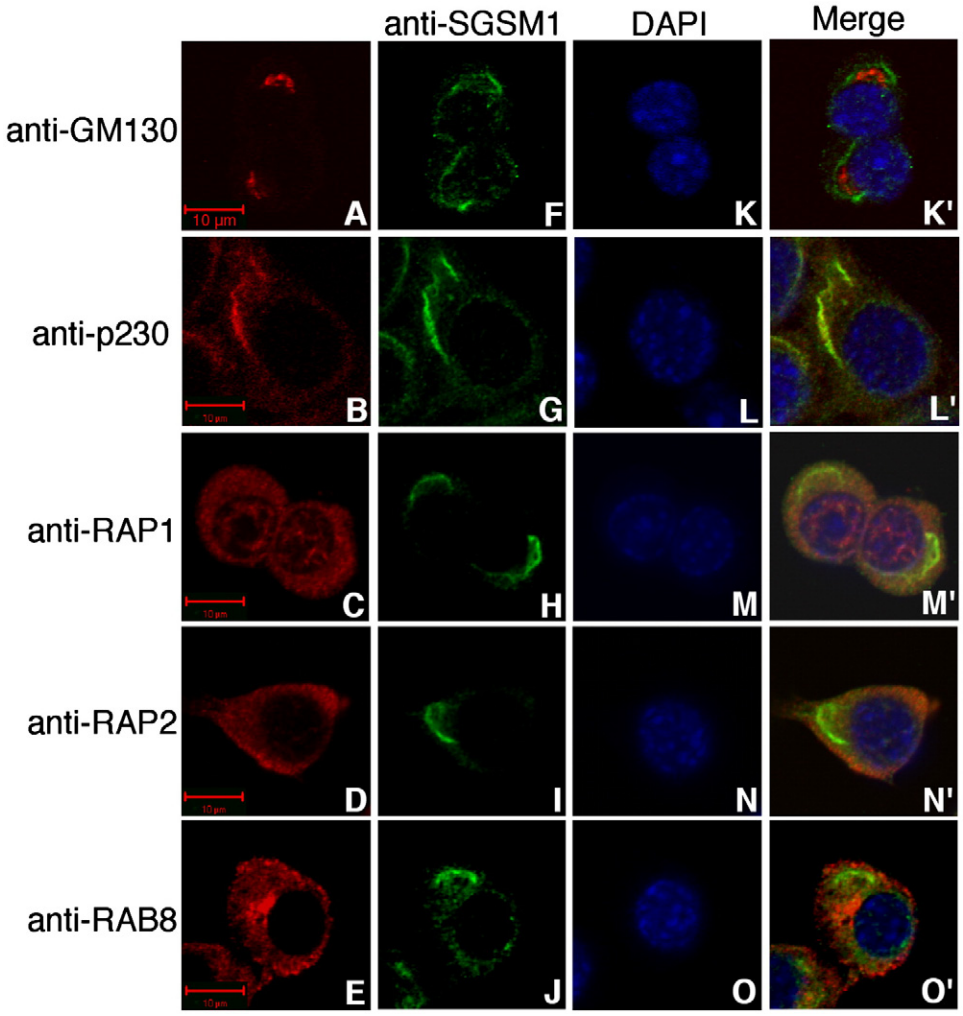


Fig. 9. Intracellular localizations of SGSM1, RAP members, and RAB members. Neuro2a cells were fixed and immunofluorescence microscopy was performed. (A) Anti–GM130 antibody was used as the marker of *cis*-Golgi. (B) Anti-p230 antibody was used as the marker of *trans*-Golgi network. (K–O) Cell nuclei were stained using DAPI. Digitally merged images are shown on the right (K′–O′).

modulators in the small G protein-mediated neuronal signal transduction and vesicular transportation pathways.

Materials and methods

Computer analysis of DNA and protein sequences

A genomic clone RP5-930L11 (Accession No. **AL049759**) on chromosome 22 was analyzed to identify novel genes. The nucleotide sequences and amino acid sequences were analyzed using several kinds of programs and databases. The details about the computational analysis and the accession numbers of the nucleotide acid and amino acid sequences are provided in Supplementary information [47–49].

cDNA cloning and construction of expression vectors

All PCR primers were designed using the software package OLIGO (National Biosciences Inc., Plymouth, MN) based on the results of computer-aided analyses (Supplementary Table 1). The cDNA cloning was performed using PCR-amplified products from the human or mouse adult brain cDNA library of Multiple Tissue cDNA Panels (Takara, Palo Alto, CA). The RACE reaction was performed to determine 5′ and 3′ ends of the novel genes using Marathon Ready human brain cDNA Amplification Kit (Takara, Japan). PCRs were performed using KOD plus PCR system (Toyobo, Japan) according to the manufacturer's protocol under the conditions of 94 °C for 20 s, 60 °C for 30 s, 68 °C for 2 min for 35 cycles. The amplified products were inserted in the *Hin*cII site of pUC118 plasmid vector (Takara, Japan) and used as the templates for nucleotide sequencing.

HA-tagged or FLAG-tagged *RAP1A/1B/2A/2B*, SGSM1/2/3, and various cDNA fragments encoding 1–300, 1–350, 201–1093, 301–1093, 351–1093, 451–1093, 601–1093, and 751–1093 aa of SGSM1 were constructed by inserting each cDNA into the expression vectors p3XHA-C, p3XHA-N, p3XFLAG-C, and p3XFLAG-7.1 (Sigma–Aldrich, St. Louis, MO). The details of the construction of expression vectors are described in Supplementary information.

Northern blot hybridization

The probes for human or mouse transcripts were PCR products amplified using primers described in Supplementary Table 1. Human Mutiple Tissue Northern (MTN) blot, Mouse MTN Blot, and Mouse Embryo Northern Blot (Takara, Japan) and Human Fetal Normal mRNA Northern Blot (BioChain, Hayward, CA) prepared from human adult and fetal issues as well as mouse adult and embryo tissues were hybridized under the following conditions: probe was labeled with [α-^{32}P] dCTP using Random Primer Extension Labeling System (NEN life Science). The hybridization was performed at

42 °C overnight. The filters were washed once at 65 °C in 2.0×standard saline citrate (SSC) 0.5% SDS and then once at 65 °C in 0.1×SSC/0.5% SDS. Autoradiogram was carried out using IP plate (Fujifilm, Japan) for 48 h and then hybridization signals were detected by FLA-3000G system (Fujifilm).

Generation of antiserum to SGSM1

Antiserum was raised against synthetic peptide of *SGSM1*. The amino acid sequence selected for peptide synthesis was CIPNGNLVNGT and showed 80% identity between human and mouse. The peptide was conjugated with keyhole limpet hemocyanin and processed for rabbit immunization. The antiserum was purified using HiTrap Protein G HP and NHS-activated HP (Amersham Bioscience, Uppsala, Sweden). Specificity was examined finally by Western blotting and immunoprecipitation of the lysates from 293FT cells transfected with FLAG-tagged *SGSM1* or FLAG-tagged *Sgsm1* (data not shown). The purified anti-SGSM1 antibody was used for intracellular immunolocalization and immunohistochemistry.

In situ hybridization, immunohistochemistry, and intracellular immunolocalization

Normal adult mouse brain was frozen and sliced into 6-μm horizonal sections to use for *in situ* hybridization or immunohistochemistry. For *in situ* hybridization, about 500-bp fragments of mouse *Sgsm1/2/3* were amplified and cloned with the vector of Zero Blunt TOPO PCR cloning kit (Invitrogen, Carlsbad, CA) to synthesize riboprobes using T3, T7 RNA polymerase and digoxigenin-11-uridine-5′-triphosphate (Roche Diagnostics, Mannheim, Germany). The hybridization signals were detected with anti-DIG antibody conjugated with alkaline phosphatase (Roche Diagnostics) and revealed by the reaction with NBT/BCIP (PerkinElmer Life Science, Boston, MA). Immunohistochemistry staining was performed using affinity-purified anti-SGSM1 antibody described above. Normal rabbit IgG was used for negative control. The signals were dected by horseradish-peroxidase (HRP)-conjugated anti-rabbit IgG antibody (Dako, Glostrup, Denmark). Sigma FAST 3,3′-diaminobenzidine tablet sets (Sigma–Aldrich) were used to reveal immunoreactivity. For the intracellular localization, Neuro2a cells were immunostained with affinity-purified anti-SGSM1 antibody. For the detections of RAP family proteins and RAB family proteins, a series of mouse monoclonal antibodies (BD Transduction, Lexington, KY) was used at 1:50–200 dilution. Alexa 488-conjugated goat anti-rabbit IgG and Alexa 594-conjugated goat anti-mouse IgG (Molecular Probes, Carlsbad, CA) were used as secondary antibodies. Fluroescent images were obtained by LSM 510 META (Zeiss, Germany).

The detailed protocols of *in situ* hybridization, immunohistochemistry, and intracellular immunolocalization are in Supplementary information.

Cell culture, transient transfection, immunoprecipitation, and Western blotting analysis

293FT cells were cultured in Dulbecco's modified Eagle's medium (Sigma–Aldrich) supplemented with 10% fetal bovine serum (Intergen, Japan) and 10% penicillin streptomycin solution stabilized (Sigma–Aldrich) at 37 °C with 5% CO_2. The composition of the medium for Neuro2a cells was basically the same as that for 293FT except for the addition of 1% G418 disulfate salt solution (Sigma–Aldrich). The cells were transfected with expression vectors using lipofectamine 2000 (Invitrogen). After being cultured for 30 h, transfected cells were lysed with buffer (50 mM Tris–HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, and 1% Triton X-100) containing protease inhibitor cocktail tablet (Roche Diagnostics) and 100 U/ml Benzonase (Merck, Whitehouse Station, NJ). The procedures of Western blotting analysis and immunoprecipitation using EZview Red Anti-FLAG M2 Affinity Gel or EZview Red Anti-HA Affinity Gel (Sigma–Aldrich) were described previously [50]. DICER del tagged by 3× FLAG tag at N terminus has a molecular weight similar to that of SGSM1 protein and was used as the control protein. The function predictions and previous experimental results indicate that DICER Del associates with neither SGSM family proteins nor RAP family proteins. p3XHA-N vector was used as another negative control. The detailed protocol for endogenous immunoprecipitation is described in Supplementary information.

Acknowledgments

This work was supported by a Fund for "Research for the Future" Program from the Japan Society for the Promotion of Science and a Grant-in-Aid for Scientific Research on Priority Areas "Medical Genome Science" from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2007.03.013.

References

[1] G.W. Reuther, C.J. Der, The Ras branch of small GTPases: Ras family members don't fall far from the tree, Curr. Opin. Cell Biol. 12 (2000) 157–165.
[2] D. Bar-Sagi, A. Hall, Ras and Rho GTPases: a family reunion, Cell 103 (2000) 227–238.
[3] V. Cavalli, M. Corti, J. Gruenberg, Endocytosis and signaling cascades: a close encounter, FEBS Lett. 498 (2001) 190–196.
[4] Y. Ohba, et al., Rap2 as a slowly responding molecular switch in the Rap1 signaling cascade, Mol. Cell. Biol. 20 (2000) 6074–6083.
[5] K.A. Reedquist, et al., The small GTPase Rap1, mediates CD31-induced integrin adhesion, J. Cell Biol. 148 (2000) 1151–1158.
[6] E. Caron, A.J. Self, A. Hall, The GTPase Rap1 controls functional activation of macrophage integrin aMb2 by LPS and other inflammatory mediators, Curr. Biol. 10 (2000) 974–978.
[7] J.M. Enserink, et al., The cAMP-Epac-Rap1 pathway regulates cell spreading and cell adhesion to laminin-5 through the a3b1 integrin but not the a6b4 integrin, J. Biol. Chem. 279 (2004) 44889–44896.
[8] L.S. Price, et al., Rap1 regulates E-cadherin-mediated cell-cell adhesion, J. Biol. Chem. 279 (2004) 35127–35132.
[9] C. Hogan, et al., Rap1 regulates the formation of E-cadherin-based cell-cell contacts, Mol. Cell. Biol. 24 (2004) 6690–6700.
[10] P.J. Rebstein, J. Cardelli, G. Weeks, G.B. Spiegelman, Mutational analysis of the role of Rap1 in regulating cytoskeletal function in dictyostelium, Exp. Cell Res. 231 (1997) 276–283.
[11] M. Torti, et al., Interaction of the low-molecular-weight GTP-binding protein rap2 with the platelet cytoskeleton is mediated by direct binding to the actin filaments, J. Cell Biochem. 75 (1999) 675–685.
[12] K. Taira, et al., The Traf2- andNck-interacting kinase as a putative effector of Rap2 to regulate actin cytoskeleton, J. Biol. Chem. 279 (2004) 15711–15714.
[13] J.C. Schwamborn, A.W. Puschel, The sequential activity of the GTPases Rap1B and Cdc42 determines neuronal polarity, Nat. Neurosci. 7 (2004) 923–929.
[14] J.D. Jordan, et al., Cannabinoid receptor-induced neurite outgrowth is mediated by Rap1 activation through $G\alpha_{o/i}$-triggered proteasomal degradation of Rap1GAPII, J. Biol. Chem. 280 (2005) 11413–11421.
[15] J.J. Zhu, Y. Qin, M. Zhao, L.V. Aelst, R. Malinow, Ras and Rap control AMPA receptor trafficking during synaptic plasticity, Cell 110 (2002) 443–455.
[16] Y. Zhu, et al., Rap2-JNK removes synaptic AMPA receptors during depotentiation, Neuron 46 (2005) 905–916.
[17] S.-C. Choi, J.-K. Han, Rap2 is required for Wnt/β-catenin signaling pathway in *Xenopus* early development, EMBO J. 24 (2005) 985–996.
[18] J. Colicelli, Human ras superfamily proteins and related GTPases, Sci. STKE 250 (2004) RE13.
[19] M. Geppert, Y. Goda, C.F. Stevens, T.C. Sudhof, The small GTP-binding

protein Rab3A regulates a late step in synapic vesicle fusion, Nature 387 (1997) 810–814.
[20] S. Giovedi, F. Darchen, F. Valtorta, P. Greengard, F. Benfenati, Synapsin is a novel Rab3 effector protein on small synaptic vesicles: II. Functional effects of the Rab3A-synapsin I interaction, J. Biol. Chem. 279 (2004) 43769–43779.
[21] P.V.D. Sluijs, et al., The small GTP-binding protein rab4 controls an early sorting event on the endocytic pathway, Cell 70 (1992) 729–740.
[22] B. Sonnichasen, R.S. De, E. Nielsen, J. Rietdorf, M. Zerial, Distinct membrane domains on endosomes in the recycling pathway visualized by multicolor imaging of Rab4, Rab5 and Rab11, J. Cell Biol. 149 (2000) 901–914.
[23] C. Bucci, et al., The small GTPase rab5 functions as a regulatory factor in the early endocytic pathway, Cell 70 (1992) 715–728.
[24] S. Chistoforidis, H.M. McBride, R.D. Burgoyne, M. Zerial, The Rab5 effector EEA1 is a core component of endosome docking, Nature 397 (1999) 621–625.
[25] O. Ullrich, S. Reinsch, S. Urbe, M. Zerial, R.G. Parton, Rab11 regulates recycling through the pericentriolar recycling endosome, J. Cell Sci. 135 (1996) 913–924.
[26] M. Trischler, W. Stoorvogel, O. Ullrich, Biochemical analysis of distinct Rab5- and Rab11-positive endosomes along the transferrin pathway, J. Cell Sci. 112 (1999) 4773–4783.
[27] L.A. Huber, et al., Rab8, a small GTPase involved in vesicular traffic between the TGN and the basolateral plasma membrane, J. Cell Biol. 123 (1993) 35–45.
[28] A.N. Hume, et al., Rab27a regulates the peripheral distribution of melanosomes in melanocytes, J. Cell Biol. 152 (2001) 795–808.
[29] J.C. Stinchcombe, et al., Rab27a is required for regulating secretion in cytotoxic T lymphocytes, J. Cell Biol. 152 (2001) 825–833.
[30] I. Dumhan, et al., The DNA sequence of human chromosome 22, Nature 402 (1999) 489–495.
[31] A. Shiohama, T. Sasaki, S. Noda, S. Minoshima, N. Shimizu, Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region, Biochem. Biophys. Res. Commun. 304 (2003) 184–190.
[32] T. Sasaki, A. Shiohama, S. Minoshima, N. Shimizu, Identification of eight members of the Argonaute family in the human genome, Genomics 82 (2003) 323–330.
[33] K. Hosono, T. Sasaki, S. Minoshima, N. Shimizu, Identification and characterization of a novel gene family YPEL in a wide spectrum of eukaryotic species, Gene 340 (2004) 31–43.
[34] S.F. Altschul, W. Gish, W. miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[35] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[36] I. Callebaut, J. De Gunzburg, B. Goud, J.-P. Mornon, RUN domains: a new family of domains involved in Ras-like GTPase signaling, Trends Biochem. Sci. 26 (2001) 79–83.
[37] I. Janoueix-Lerosey, E. Pasheva, M. De Tand, A. Tavitian, J. De Gunzburg, Identification of a specific effector of the small GTP-bingding protein RAP2, Eur. J. Biochem. 252 (1998) 290–298.
[38] I. Janoueix-Lerosey, F. Jollivet, J. Camonis, P.N. Marche, B. Goud, Two-hybrid system screen with the small GTP-binding protein Rab6, J. Biol. Chem. 270 (1995) 14801–14808.
[39] M. Mari, E. Macia, Y.L. Marchand-Brustel, M. Cormont, Role of the FYVE finger and the RUN Domain for the Subcellular Localization of Rabip4, J. Biol. Chem. 276 (2001) 42501–42508.
[40] A.F. Neuwald, A shared domain between a spindle assembly checkpoint protein and Ypt/Rab-specfic GTPase-activators, Trends Biochem. Sci. 22 (1997) 243–244.
[41] S. Albert, E. Will, D. Gallwitz, Identification of the catalytic domains and their functionally critical arginine residues of two yeast GTPase-activating proteins specfic for Ypt/Rab transport GTPases, EMBO J. 18 (1999) 5216–5225.
[42] P.A. Gleeson, et al., p230 is associated with vesicles budding from the trans-Golgi network, J. Cell Sci. 109 (1996) 2811–2821.
[43] E. Levivier, B. Goud, M. Souchet, T.P. Calmels, J.P. Mornon, I. Callebaut, uDENN, DENN, and dDENN: indissociable domains in Rab and MAP kinases signaling pathways, Biochem. Biophys. Res. Commun. 287 (2001) 688–695.
[44] J.I.S. MacDonald, C.J. Kubu, S.O. Meakin, Nesca, a novel adapter, translocates to the nuclear envelope and regulates neurotrophin-induced neurite outgrowth, J. Cell Biol. 164 (2004) 851–862.
[45] R. Sakamoto, D.T. Byrd, H.M. Brown, N. Hisamoto, K. Matsumoto, Y. Jin, The *Caenorhabditis elegans* UNC-14 RUN domain protein binds to the Kinesin-1 and UNC-16 complex and regulates synaptic vesicle localization, Mol. Biol. Cell 16 (2005) 483–496.
[46] A.K. Haas, E. Fuchs, R. Kopajtich, F.A. Barr, A GTPase-activating protein control Rab5 function in endocytic trafficking, Nat. Cell Biol. 7 (2005) 887–893.
[47] C. Burge, S. Karlin, Prediction of complete gene structures in human genomic DNA, J. Mol. Biol. 268 (1997) 78–94.
[48] M.Q. Zhang, Identification of protein coding regions in the human genome by quadratic discriminant analysis, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 565–568.
[49] E.C. Uberbacher, Y. Xu, R.J. Mural, Discovering and understanding genes in human DNA sequence using GRAIL, Methods Enzymol. 266 (1996) 259–281.
[50] J. Wang, J. Kudoh, A. Takayanagi, N. Shimizu, Novel human BTB/POZ domain-containing zinc finger protein ZNF295 is directly associated with ZFP161, Biochem. Biophys. Res. Commun. 327 (2005) 615–627.