Published in final edited form as:

*Gene*. 2018 March 30; 648: 42–53. doi:10.1016/j.gene.2018.01.029.

# The human GCOM1 complex gene interacts with the NMDA receptor and internexin-alpha☆

**Raymond S. Roginski**[a,b,*], **Chi W. Lau**[a,1], **Phillip P. Santoiemma**[b,2], **Sara J. Weaver**[c,3], **Peicheng Du**[d], **Patricia Soteropoulos**[e], and **Jay Yang**[c]

[a]Department of Anesthesiology, CMC VA Medical Center, Philadelphia, PA 19104, United States

[b]Department of Anesthesiology and Critical Care, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, United States

[c]Department of Anesthesia, University of Wisconsin at Madison, Madison, WI 53706, United States

[d]Bioinformatics, Rutgers-New Jersey Medical School, Newark, NJ 07103, United States

[e]Department of Genetics, Rutgers-New Jersey Medical School, Newark, NJ 07103, United States

## Abstract

The known functions of the human GCOM1 complex hub gene include transcription elongation and the intercalated disk of cardiac myocytes. However, in all likelihood, the gene's most interesting, and thus far least understood, roles will be found in the central nervous system. To investigate the functions of the GCOM1 gene in the CNS, we have cloned human and rat brain cDNAs encoding novel, 105 kDa GCOM1 combined (Gcom) proteins, designated Gcom15, and identified a new group of GCOM1 interacting genes, termed Gints, from yeast two-hybrid (Y2H) screens. We showed that Gcom15 interacts with the NR1 subunit of the NMDA receptor by co-expression in heterologous cells, in which we observed bi-directional co-immunoprecipitation of human Gcom15 and murine NR1. Our Y2H screens revealed 27 novel GCOM1 interacting genes, many of which are synaptic proteins and/or play roles in neurologic diseases. Finally, we showed, using rat brain protein preparations, that the Gint internexin-alpha (INA), a known interactor of the NMDAR, co-IPs with GCOM1 proteins, suggesting a GCOM1-GRIN1-INA interaction and a novel pathway that may be relevant to neuroprotection.

### Keywords

GRIN1 gene; INA gene; Yeast two-hybrid; Gdown1/POLR2M; Gup1/MYZAP; Gcom15 protein

☆DNA sequences deposited in GenBank: Human Gcom1, AY207007; human Gcom15, JF419331; rat Gcom15, JF440303.

*Corresponding author at: Room 5A140, 5th Floor Bldg 2, Anesthesia Department, CMC Philadelphia VA Medical Center, 3900 Woodland Avenue, Philadelphia, PA 19104, United States., raymond.roginski@va.gov (R.S. Roginski).
[1]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, United States.
[2]Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, United States.
[3]Department of Chemistry, California Institute of Technology, Pasadena, CA 91125, United States.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2018.01.029.

## 1. Introduction

The human GCOM1 gene (original gene symbol, GRINL1A) produces three sets of transcripts and proteins from two neighboring genes transcribed in the same direction on the long arm of chromosome 15 by a mechanism we termed the complex transcription unit (CTU) (Roginski et al., 2001; Roginski et al., 2004). Since our discovery and description of its genomic organization (Roginski et al., 2004), many ground-breaking findings have emerged from studies on the upstream (centromeric) gene, which generates the Gup transcripts and proteins, as well as from the downstream (telomeric) gene, which produces the Gdown transcripts and proteins. The individual components of the CTU were therefore assigned the gene symbols MYZAP for Gup1 and POLR2M for Gdown1.

Significant functions have been identified for the genes that produce the Gup1 and Gdown1 proteins. POLR2M encodes Gdown1, which was recognized as the 13th subunit of RNA polymerase II in 2006 (Hu et al., 2006), a seminal finding in the field of transcription which has subsequently been investigated in at least two major laboratories. Gdown1 performs essential functions in regulating transcription elongation (Cheng et al., 2012; Jishage et al., 2012). MYZAP codes for the Gup1 protein, aka Myozap (Seeger et al., 2010), which plays a vital role in the intercalated disk of cardiac myocytes, cell–cell adhesions in stable tissues and cell migration during tissue development and remodeling (Huo et al., 2011). Knock-out of Gup1 in zebrafish results in cardiomyopathy (Seeger et al., 2010).

We have focused our research on the GCOM1 combined (Gcom) proteins, whose predominant expression occurs in the CNS, because these proteins may play roles in neuroprotection, neuropsychiatric and neurodegenerative diseases. In our prior work on the GCOM1 gene, we cloned a 550 amino acid ORF from human adult brain mRNA (Roginski et al., 2004) and subsequently identified a corresponding 64 kDa protein, named Gcom1, for GCOM1 combined protein 1, in rat brain (Roginski et al., 2008). Gcom1 was detected with antibodies raised against a synthetic peptide corresponding to amino terminal residues S23-E38. As expected, these antibodies also revealed the predicted 54 kDa Gup1 protein (466 aa), which shares many amino terminal exons with Gcom1. Surprisingly, this antibody preparation also reacted with a novel, larger, putative GCOM1 protein migrating as a 105 kDa band, which was provisionally named Gcom15 (Roginski et al., 2008). Gcom15 was not predicted based upon the GCOM1 transcripts we cloned as of 2008. Such a protein could be highly significant, because our co-immunoprecipitation experiments with anti-GCOM1 antibodies showed co-IP with the rat NMDA receptor subunit 1 (NR1; human gene symbol GRIN1) in rat brain protein preparations. Furthermore, we demonstrated that anti-GCOM1 antibodies protected rat cortical neuronal cultures against NMDA toxicity, implying that a GCOM1 protein may be involved in a novel, neuroprotective cascade. However, these experiments did not reveal which GCOM1 protein(s) were responsible for the pull-down of the NMDAR subunit (Roginski et al., 2008). The most likely candidates would be Gcom1 and the novel Gcom15.

Therefore, we sought to isolate mRNAs/cDNAs capable of generating the larger protein with the goals of determining its size using heterologous cell expression and performing co-IP with co-expressed NR1 subunits to provide further evidence for the interaction with the

NMDAR observed in rat brain. At the same time, we carried out yeast two-hybrid screens to search for genes capable of interacting with Gcom1, because additional mediators would be needed to explain the neuroprotective effect. Here we describe our isolation and characterization of Gcom15 clones from human and rat sources, expression and co-immunoprecipitation of Gcom15 and NR1 subunits in heterologous cells, yeast two-hybrid screens using Gcom1 cDNA clones as bait, and immunologic experiments to confirm one of these Y2H interactions, with the neuronal intermediate filament gene, internexin-α (gene symbol INA), in a mammalian system. Our results suggest an interaction or interactions among three genes, GCOM1, GRIN1 and INA, that may activate a new pathway relevant to normal CNS function and neurologic disease.

## 2. Materials and methods

### 2.1. Cloning of Gcom1 and Gcom15 cDNAs

Human GRINL1A combined transcript 1 (Gcom1) cDNA was cloned using a single-tube reverse transcriptase PCR procedure, as previously described (Roginski et al., 2008). Human and rat Gcom15 cDNAs were cloned from adult brain mRNA as overlapping 5′ and 3′ segments using the Titan RT PCR system (Roche; Indianapolis, IN). The overlapping segments were combined by hybridization-extension PCR. The resulting large amplicons were ligated and transformed into competent *E. coli* using the TOPO-TA vector system (Invitrogen; Carlsbad, CA). Plasmid DNA aliquots from positive colonies were sequenced revealing intact open reading frames (ORFs) coding for 765 and 761 amino acids for human and rat Gcom15 respectively. The inserts were subcloned into the pCIneo expression vector (Promega; Madison, WI).

### 2.2. Bioinformatic analyses

Identities of the Gcom1 interacting cDNA clones (Gints) were determined by NCBI BLAST. Participation in biological networks was assessed by Ingenuity Pathways Analysis (Ingenuity Systems, Redwood City, CA) software and other bioinformatics tools available on the internet such as the European Bioinformatics Institute and GeneCards. Interactors of Gcom1 from the Y2H experiment were analyzed by placement of the molecules in networks. Interaction networks of Gcom1 and its interactors were developed according to the likelihood of occurrence in vivo. The networks were then given a numerical value score based on the hypergeometric distribution, to rank the networks to determine the degree of relevance. Networks with the highest scores were recorded and analyzed to determine their relevance to neurologic disease, neurologic and other functions.

### 2.3. Transfection, protein isolation, immunoprecipitation and Western blotting

Each Gcom15 cDNA (0.5 μg) was transfected into HEK293 cells with or without the mouse NR1-1a NMDAR subunit in pcDNA3 and incubated for 24 h at 37 °C. Membrane plus cytosolic proteins were isolated from transfected cells for both Western blots and immunoprecipitation (IP) experiments with anti-GCOM1 Ab against residues S23-E38 (GenScript order ID 416739, chickens 602 and 604) and T423–Q440 (GenScript ID 109531, rabbits 1585 and 1586) or with anti-NR1 Ab. After determining protein concentration in the sample with the BCA protein assay (BioRad), 100 μg of both membrane and cytosolic

proteins were transferred to a total of 6 microcentrifuge tubes (600 μg total protein). The lysate was precleared by adding 20 μL of Protein A-agarose (Invitrogen) and shaken for 30 min at 4 °C. The lysate was centrifuged to pellet the beads at 12,400 rpm for 15 min at 4 °C. After transferring the supernatant to new tubes, 2 μg of anti-Gcom1 T423-Q423 primary antibody were added to 2 tubes, 2 μg of anti-NR1 antibody (mouse monoclonal, BD; goat polyclonal, Santa Cruz) were added to 2 tubes and 2 tubes were given 2 μg of IgG instead of specific antibody. All tubes were incubated for 2 h on a rocker at 4 °C. After incubation, 40 μL of Protein-A agarose was added to the tubes followed by incubation at 4 °C on a rocker overnight (14–18 h). Each tube was then centrifuged at 12,400 rpm for 15 min at 4 °C. The supernatant was aspirated, and the beads were washed again two more times. After the final wash, tubes were centrifuged for 15 min at 12,400 rpm, the supernatant was aspirated, and 40 μL of 4× concentrated sample buffer was added to each tube. The beads were boiled for 10 min to elute proteins and finally centrifuged at 12,400 rpm for 15 min. The supernatant from each tube was then loaded into a 12% SDS-PAGE Bis-Tris gel in 1× SDS PAGE running buffer for 90 min at 120 V. The gel was placed in 1× transfer buffer (20% methanol in running buffer without SDS). A gel sized piece of polyvinyldifluoride (PVDF) was placed in 100% methanol and then on top of gel in between Brillo pads and Whatman leaves. Gel and PVDF complex was placed in transfer box and run at 100 V for 1 h in 4 °C. After PVDF was rinsed in 1× PBS (phosphate buffered saline, pH 7.4), PVDF was stained with Ponceau S to visualize protein, and then destained. After PBS rinses, PVDF was stored overnight in blocking buffer (1% BSA and 5% non-fat dry milk in 0.05% Tween-20 in 1× PBS). The membrane was then cut into 2 pieces; one for the anti-Gcom1 and the other for anti-NR1 Ab. Each membrane piece was then placed in primary antibody in blocking buffer (1:500 for anti-NR1, 1:1000 for anti-Gcom1) for 2 h at 20 °C. PVDF membranes were rinsed in 0.1% Tween in 1× PBS and then immersed in anti-rabbit secondary horseradish peroxidase (1:10,000; Amersham, Piscataway, NJ) for 1 h at 20 °C. After washes with 0.1% Tween in PBS, secondary antibodies were visualized with chemiluminescent detection (Figs. 3 and 6) using a Chemiluminescence kit (Amersham). ECL solution was placed on membranes in a dark room. Membranes were placed in developing cassette and X-ray photos were taken at varying exposure times.

For the experiments in Fig. 5 (and other experiments, not shown), rat brain samples were derived from humanely euthanized white adult male Wistar rats. Brains (hemispheres) were either removed and immersed in liquid nitrogen or were removed after perfusion with cold saline and stored at −80 °C. Rat brain tissue was homogenized in a buffer of 0.32 M sucrose, 20 mM Tris–HCl (pH 7.5), 1 mM EDTA, 1 mM PMSF, 1 μL/mL protease inhibitor cocktail (P-8304, Sigma) and then centrifuged at 1000 ×*g* for 15 min. The supernatant was collected and centrifuged at 100,000 ×*g* for 60 min; the supernatant was again collected. The pellet was washed and resuspended with lysis buffer (20 mM Tris–HCl (pH 7.5), 1 mM EDTA, 1 mM PMSF, 1 μL/mL protease inhibitor cocktail). The final precipitate and the supernatant, containing cellular and membrane proteins, were collected and stored separately at –80 °C. Membrane and cytosolic proteins were then isolated from the cell supernatant using a Plasma Membrane Protein Extraction Kit (MBL International; Woburn, MA). Also in Fig. 5, we used the Odyssey protocol (Li-Cor) for detection of fluorescently labeled secondary antibodies. For immunoprecipitations, 8 μg of anti-GCOM1 T423-Q440 (rabbit 1586)

antibody was added to 400 μg rat brain proteins (RBP) followed by precipitation overnight at 4 °C. On the next day, 1.5 mg Dynabeads Protein G (Novex, catalog # 10003D) was added to the IP mixture; 1.5 mg Dynabeads is an excess to bind immune complexes from these amounts of Ab/protein. The unbound fraction (flow-through) from the Dynabeads-Ab-Ag complexes (pellet) was retained for Western blot analysis. As a negative control, 400 μg RBP were immunoprecipitated with anti-synaptophysin Ab (Millipore, MAB329). Each sample well of the WB was loaded with 25% (by volume, equivalent to the IP complexes from 100 μg of RBP) of the material eluted from the Dynabeads. (Due to the small amounts of protein in the immune complexes, it is impractical to quantitate the amounts loaded into the gel.) Each IP was carried out at least three times with the same results as shown in Fig. 5.

### 2.4. Yeast two-hybrid screening

Amino (codons 22–414; plasmid G113D) and carboxy (332–550; plasmid G923C) cDNA segments of the human Gcom1 mRNA/protein were cloned into the Yeast Two Hybrid (Y2H) vector pCWX200 and screened using a TetR system against an adult human brain cDNA library (ProteinLinks; San Diego, CA). The N- and C-terminal screens yielded 11 and 48 verified clones, respectively, which were amplified using *E. coli* transformation, PCR or both and sequenced.

## 3. Results

### 3.1. Cloning, sequence and structural features of Gcom15

The most parsimonious hypothesis to explain Gcom15 is to splice exon 21a between exons 13 and 22 of the Gcom1 cDNA sequence (Table 1). In the human genome, the addition of exon 21a (645 bases) predicts a 765 aa ORF, while in rat, a 761 aa ORF would arise because rat has a shorter exon 21 (633 bases). To verify these ORFs, adult human and rat brain mRNA samples were amplified by RT PCR with two sets of primers designed to generate long, overlapping upstream and downstream products from the predicted Gcom15 mRNAs. The human products were mixed, heated to separate the strands, self-annealed and extended for several cycles, followed by amplification with a pair of nested primers located in the 5′ and 3′ untranslated regions of the human ORF. This strategy, which we call hybridization-extension PCR, is illustrated in Fig. 1A. The resulting 2.3 kb amplicon contained the entire human Gcom15 coding region (Fig. 1B, lane 1). Restriction endonuclease digestion and DNA sequencing of the cloned insert revealed intact reading frames with the splicing patterns shown in Table 1. The cloned human Gcom15 ORF was excised, inserted into the expression vector pCIneo and sequenced again to confirm the integrity of the ORF. For the rat Gcom15 ORF, cloning of upstream and downstream products resulted in plasmids having orientations amenable to direct (i.e., without PCR amplification) reconstitution of the entire rat Gcom15 reading frame by utilizing the unique HpaI restriction site. The exon structures of the human Gcom1 and Gcom15 cDNAs and proteins are shown in Fig. 2A. The cloned human and rat Gcom15 ORFs predict 765 and 761 amino acid proteins, respectively, with unmodified masses of 87.8 and 86.9 kDa (Table 1).

Homology searches revealed that the Gcom15 proteins are not orthologs or paralogs of any known gene. However, identity and similarity comparisons (e.g., Bestfit and similar

programs) using the human Gcom15 sequence revealed results similar to those we reported previously for Gcom1 (Roginski et al., 2004). Briefly, the Gcom1 and Gcom15 proteins display similarity to several proteins that interact with the NR1 subunit (human gene symbol GRIN1) of the NMDA receptor (NMDAR). Chief among these are the NR2A and NR2B NMDAR subunits (human gene symbols GRIN2A and GRIN2B); the rat NR3A subunit; and yotiao [gene symbol AKAP9 (Lin et al., 1998)]. In the interval since our 2004 publication, we have found additional GRIN1-interacting proteins that display similarity to the Gcom1/Gcom15 amino acid sequences (Supplementary Table 1): kalirin-7 (KALRN); α-actinin (ACTN1); neurofilament-L (NEFL); and internexin-α (INA). To assess whether the Gcom1/Gcom15 sequence similarities were restricted to genes that interact with the GRIN1 subunit, three genes known to interact with the NR2 subunit genes GRIN2A and B were compared with the large GCOM1 proteins by the same method: SAP102 (gene symbol DLG3), PSD95 (DLG4) and nNOS (NOS1) did not show similarity to Gcom1. Finally, NCBI homology searching using the stringent software blastp (run date 1 Sep 2015) revealed one gene that exceeded the significance threshold (E-value 4.47 e–03): lebercilin (gene symbol LCA5), a ciliary protein involved in Leber congenital amaurosis 5 and severe early-childhood-onset retinal dystrophy (Dharmaraj et al., 2000). Lebercilin displayed a 194 aa overlap of its amino terminal residues 11–193 with amino terminal residues 126–312 of human Gcom15. This alignment scored 24.2% identity and 44.8% similarity. Interestingly, one of the four interacting genes for LCA5 listed in GeneCards is the GRIN2B subunit, another suggestive link to the NMDAR. Taken together, identity and similarity comparisons suggest that Gcom15's amino acid sequence resembles extensive segments of proteins that interact with the GRIN1 subunit of the NMDAR. However, no significant similarity was detected between Gcom1 and several proteins that interact with GRIN2 subunits, suggesting that a possible GCOM1-GRIN1 interaction may be more likely than with GRIN2 subunits. Next, several of the genes showing similarity to Gcom1/Gcom15 have been implicated in neurologic diseases. Finally, GeneCards lists one Gene Ontology biological process in the GCOM1 entry: intracellular signal transduction.

The primary sequences of the human and rat Gcom15 proteins contain features characteristic of type 1 membrane proteins. First, visual inspection of Gcom15's amino terminal residues and the output files from the programs Signal P-4.1 (run dates 18 May 2014 and 27 June 2015 for the human and rat sequences respectively) and Phobius (run dates 20 and 27 June 2015) predict a preferred signal peptide cleavage site between residues 16 and 17, with a possible secondary site at 20/21 (only in Signal P). In both sequences, the N-region covers residues 1–3; the H-region corresponds to 4–11; and the C-region comprises residues 12–16. Second, the key features of the Gcom15 amino acid sequence are the same as we originally reported for the Gcom1 sequence (Roginski et al., 2004). However, due to the presence in human Gcom15 of exon 21a, there are two additional consensus N-linked glycosylation sites compared with human Gcom1, which has only one such site, in exon 22. (Note that the exon 22 site is not conserved in rat exon 22.) All 3 consensus N-glycation sites in human Gcom15 (at residues 609, 614 and 656) are located amino to the putative membrane spanning region (MSR; amino acids 685–703). The presence of glycosyl moieties at one or more of the consensus sites would strongly favor an extracellular destination for the long amino terminus. Rat Gcom15 conserves only the first consensus N-glycosylation site (at N605,

which corresponds to N609 of human Gcom15). Next, the residues bracketing the putative MSR, DD683–684 and EE704–705 in human Gcom15, are the same in both species. The amino acids of the MSR, I685–I703 in the human sequence, are highly conserved in both species. Finally, both Gcom15 sequences contain the identical, consensus endoplasmic reticulum retention-retrieval sequence in their carboxy termini, RYR, which are amino acids 749–751 in the human ortholog.

Secondary and tertiary structure predictions of the human Gcom15 sequence using the Phyre software also are consistent with a type 1 membrane protein. Residues 1–40 display coiled coil and α-helical segments; 41–489 are α-helix; 490–593 are coiled coil and α-helix; 594–670 (containing three N-glycosylation sites in human Gcom15) have no consensus; residues 671–730 (which include the putative MSR) display coiled coil and α-helix structure; and finally 731–764 (part of the putative intracellular segment) show coiled coil and beta sheet. According to the ProtParam tool (ExPASy software; run date 20 Nov 2014), human Gcom15 has instability index of 58.48, and rat Gcom15's index is 58.95, both of which classify these proteins as “unstable.” Phyre predicts structural overlap with smooth muscle heavy chain (MYH11) with an E-value of $4.2 \times e^{-35}$ and a 6% identity overlap. Amazingly, three interactors of MYH11 in GeneCards are GRIN1, GRIN2A and GRIN2B. Therefore, MYH11 is another instance of a protein with similarity to Gcom15 which interacts with NMDAR subunits. A space-filling model of human Gcom15, which incorporates the predicted secondary structure features, is depicted in Fig. 2B. The large amino terminus connects to the smaller, putative intracellular carboxyl end by several extended segments, which include the predicted MSR. Finally, Western blots of rat brain protein fractions stained with anti-amino and anti-T423-Q440 antibodies revealed significant amounts of rat Gcom15 in the plasma membrane fraction (results not shown).

### 3.2. Expression of Gcom15 proteins in heterologous cells

Each Gcom15 cDNA was transfected into HEK293 cells with or without the mouse NR1-1a NMDAR subunit cDNA. Western blots of membrane proteins isolated from transfected cells revealed 105 kDa bands when stained with anti-GCOM1 Ab (Fig. 3A, B, top panels). Co-transfection of human or rat Gcom15 cDNA with mouse NR1-1a cDNA revealed co-IP of NR1 protein by anti-GCOM1 Ab (Fig. 3A, lower panels). The “reverse” co-IP, using anti-NR1 antibodies, revealed co-IP of the human (but not rat) Gcom15 (Fig. 3B, lower panels). The likelihood that GCOM1-GRIN1 PPIs play a role in neuroprotection, as shown in our previous work (Roginski et al., 2008), implies that GCOM1 proteins interact with other proteins besides the NR1 subunit. Therefore, we searched for additional GCOM1-interacting genes.

### 3.3. Yeast two-hybrid (Y2H) studies

We screened an adult human brain cDNA library with the 550 aa human Gcom1 cDNA (AY207007) to find additional candidate interacting genes. The Gcom1 sequence was chosen primarily because it was the largest GCOM1 cDNA available prior to the cloning of the Gcom15 cDNAs and should detect more interacting genes than the smaller Gdown1 and Gup1 cDNAs. Second, the combined proteins are a distinguishing feature of the GCOM1 gene, and none had been screened for interacting partners prior to our study.

### 3.4. Rationale for design of Gcom1 bait clones

Because Gcom1 cDNA has a larger ORF than typical Y2H clones, it was engineered into 2 smaller, overlapping amino- and carboxy-terminal clones. A second issue was that the Gcom1 protein begins with a candidate signal peptide (residues 1–16), a sequence that might cause the hybrid protein to bind with yeast membranes and prevent its transit to the nucleus. If these residues were to be included in the amino-terminal bait clone, they might limit the detection of interacting genes, and could also lead to a false negative result when tested in the bait self-activation control. Therefore, we chose cDNA segments that spanned codons 22–414 and 332–550. Finally, the overlapping design also allowed us to test the hypothesis that one or more interacting clones would occur in both screens, thereby potentially localizing the interaction of such clones to residues 332–414.

### 3.5. Identities of Y2H Gcom1 interacting genes (Gints)

The COOH and NH2 screens identified and confirmed 48 and 11 clones, respectively (Supplementary Data, Supplementary Table 2A and B). Plasmid DNA minipreps of prey clones were sequenced and identified by BLAST searches. Several prey clones occurred more than once, such that the 59 clones represented 27 unique gene symbols. Three genes, internexin-α (INA), MIF and PDE4DIP, were revealed in both screens, suggesting that these proteins may interact preferentially with Gcom1 amino acids 332–414, which are derived from exons 9/10/12/13 and have the same residue numbers in human Gcom15. INA was the most frequently observed gene symbol, occurring twice in the amino screen and 7 times in the carboxy screen (9 total hits).

### 3.6. Chromosome sites of GCOM1 interacting genes

The human chromosomal locations of our Y2H Gints, shown in Table 2, reveal no obvious, genome-wide pattern. However, the distribution of Gints does not appear to be completely random, in that several chromosomes contain multiple Gints, especially chromosomes 1 and 11, which have 5 and 3 Gints, respectively, and may hint at functional significance. Furthermore, it is intriguing that several pairs of Gints reside on the same chromosome arm, relatively close to each other. On chromosome 1, DLGAP3 and KCNAB2 lie at bands 1p34 and 1p36, respectively; PDE4DIP and RIIAD1 are less than 300 kb apart at 1q21.2 and 1q21.3, respectively. On chromosome 11, FTH1 and FAM89B reside at 11q12.3 and 11q13.1, respectively.

An even more interesting cluster of 4 genes, which may provide insights into the functions of GCOM1, occurs at chromosome 9q34.3: the Y2H Gint olfactomedin 1 (OLFM1), the neurodevelopmental hub gene NOTCH1, the immunologically determined Gint GRIN1 and the IPA Gint NSMF (see Fig. 4) all reside in a 2 Mb segment between bases 135,075,243 (start of OLFM1; GeneCards accessed 4 September 2016) and 137,459,357 (end of NSMF). Several genes proximal to human GCOM1 on chromosome 15q21.3/q22.1 may also shed light on the roles of GCOM1 as well as the complex gene model (see Discussion, Section 4.4).

### 3.7. Functions of Y2H Gints

The data available in GeneCards entries suggested that the Y2H Gints could be allocated to four main groups based on functions: (1) synapses, channels and neurodevelopment: ATP6V1D, DLGAP3, GPM6B, INA, KCNAB2, OLFM1, PQLC1, SCOC, SNAP25, SRI and TPM4; (2) transcription, RNA processing, gene regulation, cell cycle and chromatin: CCNB1IP1, CUL1, ID2, PDE4DIP, PRKRA, PRPF38B, RPGRIP1L and ZNF418; (3) kinases and signaling: CSNK2B, MIF and RIIAD1; and (4) hematopoiesis, leukemia, cancer, blood plasma and platelets: ANKRD26, CT45A, FAM89B, FTH1 and KIAA1549L (formerly known as C11orf41). FAM89B and KIAA1549L showed their highest levels of protein expression in blood plasma and platelets and were therefore assigned to group 4. Several Gints have multiple functions, such that they could be assigned to multiple groups. GeneCards data indicate that all 27 Gints are expressed at detectable levels in human brain. Furthermore, eight Gints displayed their highest levels of expression in brain: ATP6V1D, DLGAP3, FTH1, GPM6B, INA, PRKRA, SNAP25 and SRI.

### 3.8. Bioinformatic analyses suggest GCOM1's involvement in CNS functions and other physiologic processes

To determine more convincingly that Gcom1 interactors, especially the novel Y2H genes, could be assigned to biologically relevant functional groups, we analyzed them further using Ingenuity Systems software (IPA) and other methods, in two ways, beginning in 2007. First, we analyzed our novel 27 Gcom1 interactors (narrowly defined Gints) from our Y2H studies as a group. Second, we combined our Y2H Gints with our immunologic GRIN1 interaction and with the broad GCOM1 interactome which includes: all GCOM1 proteins for which there are interactants in GeneCards and IPA. Specifically, GeneCards Gints (as of 1 January 2016) included 16 for the gene symbol GCOM1, 32 for the Gup proteins (aka MYZAP) and 24 for the Gdown proteins (aka POLR2M). As of 14 February 2017, these numbers have increased to 17, 68 and 27, respectively. Note that Gup and Gdown proteins may or may not also interact with Gcom proteins.

Fig. 4 depicts a network diagram (IPA Path Designer) using the broad GCOM1 interactome made by the Grow 3 algorithm with the command “Grow 3 molecules at a time to connect Gints with NSMF.” NSMF was chosen based upon its function in NMDA receptor mediated synaptonuclear signaling and neuronal migration (see Section 4.4). The network shows a moderate level of connectivity, suggesting likely biologic relevance. The orange lines (edges) indicate new connections predicted by Grow. Key signaling pathways shown involve ERK, CREB and NSMF. For example, the Gint CUL1 connects to ERK, which networks to the Gint INA, which links to NSMF. Another path to NSMF runs from the Gint RPGRIP1L to RPSA to RPS21 to FMR1, which links to NSMF. CUL1 also connects through RPS21 to RNA polymerase II (the 13th subunit of which is the GCOM1 protein Gdown1, aka POLR2M) to NSMF. Overall, the known and putative novel edges in Fig. 4 suggest the involvement of GCOM1 proteins in the glutamate system, in particular with the NMDA receptor via the GRIN1 subunit. It should be noted that the IPA software does not depict all known connections. For example, the GRIN1-INA interaction, which has been known since the year 2000 (Husi et al., 2000), does not have an edge in this network, but should. Another aspect which this IPA algorithm does not address is that many Gints can be shown to

network with other Gints through one or two intervening interacting genes from GeneCards. For example, the Gint OLFM1 interacts with the hub gene disrupted in schizophrenia 1 (DISC1), two of whose interactors are AKAP9 (yotiao) and KALRN, both of which interact with the Gint GRIN1. Another example includes the following Y2H Gints (bold print): **OLFM1** ↔ PAX3 ↔ **ID2** ↔ ERBB2 ↔ UL1.

In a different Path Designer format called “Connect,” IPA produced a larger, more complicated network diagram (Supplementary Fig. 1), which also includes canonical pathways. This network reveals links to several neurologically relevant pathways, especially glutamate and calcium signaling, long-term potentiation, amyloid processing, SLE and Huntington's disease signaling, dopaminergic transmission and mTOR and ERK signaling. As with the Grow algorithm, the GCOM1 hub was connected to the ERK-NSMF hub, with GRIN1 interactions key to both pathways. Furthermore, in contrast to the Grow diagram, Supplementary Fig. 1 does show the GRIN1-INA interaction, suggesting a possible triple interaction, **GCOM1** ↔ **GRIN1** ↔ **INA**.

### 3.9. Demonstration of GCOM1-INA interaction by co-immunoprecipitation in a mammalian system

To determine whether the GCOM1-Gint interactions revealed in the yeast two-hybrid system might reflect physiologically relevant interactions in mammalian tissues, we performed co-immunoprecipitation studies using adult rat brain protein preparations (RBP), anti-GCOM1 primary antibodies and anti-Gint secondary antibodies. We hypothesized that internexin-α would be a prime candidate for a significant interaction with GCOM1 proteins in the mammalian CNS for reasons described above. In addition, INA is involved in NIFID, a form of fronto-temporal dementia (Cairns et al., 2004), and is therefore likely to be a clinically relevant interacting gene. Finally, INA is known to interact with the NMDA receptor (Husi et al., 2000).

Western blots of unprecipitated brain lysate (Fig. 5A–C, lanes 1) reveal the expected bands for GCOM1 and INA proteins, respectively: a 54 kDa Gup1 species and a 66 kDa INA species. (Three faster migrating INA species, especially the intense 40 kDa band, may represent proteolytic cleavage products derived from the major 499 aa isoform, UniProtKB Q16352.) Although these blots are not rigorously quantitative, it is evident that INA proteins show much higher levels of immunofluorescence than GCOM1 proteins in the unprecipitated samples and are therefore likely to be much more abundant. The Gcom1 and Gcom15 proteins were not detectable because only 20 μg of RBP were loaded in Fig. 5A, B, lanes 1. To demonstrate that our rabbit polyclonal, anti-peptide, anti-GCOM1 antibodies raised against Gcom1 amino acids T423-Q440 precipitate Gcom1 and other GCOM1 proteins containing exons with epitopes present in the peptide, we precipitated RBP with these Ab and stained Western blots with the same antibody. Fig. 5A, lane 2, containing proteins precipitated from 400 μg RBP, shows intense immunofluorescence at the 54 kDa Gup1 position and less intense bands at 64 kDa and 105 kDa corresponding to Gcom1 and Gcom15, respectively. These GCOM1 protein bands and relative amounts are similar to what we have observed previously with rabbit anti-S23-E38 antibodies (Roginski et al., 2008). Proteins not precipitated by the anti-GCOM1 Ab (Fig. 5A, lane 3) and IP proteins

pulled down by an anti-synaptophysin Ab (Fig. 5A; lane 4) show neither Gcom1 nor Gcom15. Trace amounts of Gup1 (in lane 3) may be due to sub-quantitative immunoprecipitation of the abundant Gup1 protein and/or “spillover” from loading lane 2. Therefore, Fig. 5A demonstrates that the anti-Gcom1 T423-Q440 antibodies precipitate several expected GCOM1 protein species which are present in the RBP preparation in a quantitative fashion and do not bind to proteins pulled down by anti-synaptophysin Ab.

A replicate Western blot (Fig. 5BCD) was used to determine whether anti-Gcom1 Ab can pull down INA bound to GCOM1 proteins. As in Fig. 5A, anti-T423-Q440 antibodies were employed for IP (lanes 1–3). However, due to the fact that the anti-INA Ab preparation is a rabbit polyclonal IgG, we could not use our rabbit anti-T423-Q440 to detect GCOM1 protein bands in this experiment. Therefore, we designed a chicken anti-amino terminal Ab against residues S23-E38, which are common to Gcom1, Gcom15 and Gup1. For reasons that may be related to the epitopes present in the S23-E38 peptide or the accessibility of the cognate sequences in the mature GCOM1 proteins, this antibody preparation does not react as strongly with the IP GCOM1 proteins as the anti-T423-Q440 antibodies (Fig. 5B, lane 2, versus Fig. 5A, lane2). In fact, there appears to be more immunoreactive signal migrating at discrete molecular sizes in the *unbound* fraction (Fig. 5B, lane 3) than in the IP fraction (lane 2) with this antibody. A possible explanation for this phenomenon is that (a) the *chicken* anti-S23-E38 Ab binds more weakly to its target sequence and (b) the target sequence may be masked to different degrees in the various GCOM1 proteins. Also, during the mixing and binding step, some of the more abundant INA proteins may have bound to and formed complexes with GCOM1 proteins, which, due to conformational changes, were not brought down by the anti-T423-Q440 antibodies and therefore ended up in the unbound fraction. Interestingly, the band sizes in Fig 5B and D (yellow in merged image), lane 3, are about 40 kDa larger than the GCOM1 species bands in Fig. 5A lane 2. Therefore, it is possible that only some of the GCOM1-INA complexes may be precipitable by the anti-T423-Q440 antibodies.

Fig. 5C (lane 2) reveals a robust signal from the ~50 kDa putative doublet INA band, strongly supporting an interaction of GCOM1 and INA proteins. We also observed a small signal from the 66 kDa INA species, although it is clearly not the predominant band. INA is not precipitated by anti-SYP Ab (negative control, lane 4), indicating that the observed co-IP is most likely due to a specific interaction with one or more of the GCOM1 proteins. The unbound fraction (Fig. 5C, lane 3) contains much more total INA immunofluorescence than the IP lane, most probably because INA proteins are much more abundant than GCOM1 proteins. Therefore, only a fraction of the total INA proteome would be able to associate with the less abundant GCOM1 proteins. Moreover, INA had 36 interactants in GeneCards as of 2 November 2016; an updated value of 211 was found on 14 February 2017. Such an abundance of interacting genes would be expected to reduce the pool of INA proteins available to associate with GCOM1 proteins.

Fig. 6 addresses the same question as Fig. 5 using both preparations of anti-Gcom1 T423-Q440 antibodies (GenScript rabbits 1585 and 1586), with chemiluminescent detection instead of fluorescent secondary antibodies. Another difference from Fig. 5 is that lower amounts of rat brain proteins were used for these immunoprecipitations. As a result, Fig. 6A

shows that the predominant GCOM1 protein in the un-precipitated lysate (lane 1) is Gup1, migrating at 54 kDa. Fig. 6A again demonstrates that Gcom15 is a relatively low abundance protein, because 20 μg of RBP does not reveal a band at the 105 kDa position (lane 2). However, IP from 100 μg RBP clearly demonstrates efficient staining and IP of the Gcom15 protein (bold arrow). Note that IP is more efficient by the antibodies from rabbit 1585 than from rabbit 1586. (This result may reflect differential loss of binding ability of the 1586 antibodies as a result of transport or storage conditions.) Fig. 6B demonstrates co-IP of the 66 kDa INA band (lane 3) by 1585 antibodies. No INA immunostaining is seen in lane 2 (no antibody control). As in Fig. 5, we observe a 66 kDa band in the unbound fraction (lane 4), indicating that co-IP is not quantitative, most likely because internexin-α protein levels greatly exceed those of GCOM1 proteins. We also performed the reverse co-IP using anti-INA antibodies in an attempt to pull down GCOM1 proteins: Fig. 6C reveals a weak Gcom15 band (boxed, lane 3), which is not present in the flow-through and no antibody controls (lanes 4 and 2, respectively). Our finding of co-IP between rat Gcom15 and INA is strong evidence that these proteins can interact in the mammalian CNS.

## 4. Discussion

### 4.1. Our results encompass five major findings

(1) isolation of the ORFs of the largest GCOM1 cDNAs and proteins detected to date, Gcom15, from human and rat sources; (2) the demonstration in heterologous cells of co-IP of Gcom15 and the Grin1 subunit of the NMDA receptor; (3) the discovery of 27 genes (Gints) that interact with the Gcom1 protein in yeast two-hybrid screens; (4) roles for the GCOM1 combined proteins (especially Gcom15) in normal CNS function, neurologic disease and neuroprotection based upon the identities, known functions and human chromosomal locations and clustering of Gints; and (5) initial evidence for verification of the Gcom15-internexin-α protein–protein interaction in a mammalian system, rat brain proteins. In the sections to follow, we shall discuss the implications and ramifications of these findings, with particular emphasis on the functions they suggest for the GCOM1 complex hub gene.

### 4.2. Significance and broader relevance of the Gcom15 transcripts and proteins

In our most recent publication (Roginski et al., 2008), we detected a GCOM1 protein in rat brain larger than any of the ORFs that we had cloned previously (Roginski et al., 2004). To confirm and extend this result to human brain, it was clearly necessary to isolate and express cDNAs capable of encoding this 105 kDa protein. Next, the demonstration of another combined GCOM1 protein displaying a novel splice between the upstream and downstream segments of the complex transcription unit, which continues the reading frame in both species tested thus far, would further support this model. Finally, we hypothesized that the large Gcom15 proteins would be biologically relevant, especially because our 2008 immunoprecipitation studies suggested only a weak co-IP of transfected mouse NR1-1a protein by human Gcom1-HA. Therefore, the question as to which GCOM1 protein(s) was (were) responsible for co-IP of the GRIN1 subunit of the NMDA receptor awaited resolution.

To shed light on this issue, and to generate cDNAs that would facilitate study of the respective proteins in heterologous cells, we cloned the hypothesized ORFs from both human and rat brain sources. This task proved difficult because of the length and high GC content of these mRNAs, as well as their lower abundance compared to other GCOM1 mRNAs in the complex brain mRNA population. In fact, attempts to amplify the ORFs directly from rat brain polyadenylated RNA, i.e., by employing primer combinations from the 5′ and 3′ untranslated regions, were unsuccessful. Therefore, we cloned overlapping upstream and downstream products, which were joined as described (Results Section 3.1 and Fig. 1). The rat PCR products were able to be cloned using standard techniques because the upstream and downstream products cloned in orientations favorable for the reconstitution of the rat Gcom15 reading frame. However, this method did not work for the human ORF, because each of the human products cloned in only one orientation. Therefore we denatured and re-annealed a mixture of the overlapping upstream and downstream amplicons, followed by self-extension of the resulting upstream-downstream hybrid duplexes for several cycles. We successfully amplified the 765-codon Gcom15 ORF in toto from the reconstituted amplicon, using a second set of primers nested relative to the upstream (sense) primer of the 5′ product and the downstream (antisense) primer of the 3′ product. Thus, we used atypical methods to clone the Gcom15 ORFs, especially human Gcom15. One might contend, therefore, that we created entities which do not occur naturally in their respective species. However, at least 3 lines of evidence argue in favor of the biological reality of Gcom15: (1) RNA data: mRNA species large enough to encode the human ORF have been detected by Northern blot analysis; at least 2 of these large mRNAs were revealed with both upstream (exons 1–6) and downstream (exons 21–23) probes (Roginski et al., 2004), suggesting that the exon 13/21a/22 splicing pattern occurs in vivo; (2) when expressed in heterologous cells, the cloned human and rat Gcom15 cDNAs generate stable proteins which have the same mobility in Western blots (105 kDa) as the Gcom15 species detected in rat brain (Roginski et al., 2008); and (3) both human and rat Gcom15 display co-IP with the mouse NMDA receptor subunit NR1-1a in heterologous cells (Fig. 3), consistent with our results observed in rat brain (Roginski et al., 2008). Such results would be unlikely if these proteins did not represent true biologic entities. Finally, Gcom15 may not be the only large GCOM1 protein larger than Gcom1. For example, the exon 13–exon 21a splice suggests the possibility of splicing to the shorter exon 21b, which would lead to an ORF comprised of exons 1–10/12/13/21b/22/23 (608 aa). Another potential variant could incorporate the exon 22–exon 25 splice observed in the Gdown2 transcript (Roginski et al., 2004), resulting in an ORF with exons 1–10/12/13/21a/22/25–28 (760 aa). Because of their distinct amino acid sequences, these hypothetical proteins might have different biological functions from Gcom15, though still be potentially capable of interacting with the NMDAR.

### 4.3. New insights into the complex gene model based on Gcom15

Other complex genes meeting the criteria that we established previously based on GCOM1 (Roginski et al., 2004) have been reported, as have studies of more traditional genes with very large numbers of splice variants. The term “transcription induced chimeras” (TICs) was coined in 2006 to describe complex genes; these authors cited over a dozen examples (including GCOM1) for which they provided computational and experimental support (Parra et al., 2006). Another group (Hernandez-Sanchez et al., 2006) published detailed evidence

for chimeric tyrosine hydroxylase-insulin transcripts during early avian development. However, the hybrid transcripts were not detected in mouse embryos, suggesting that some complex genes may show species-specific regulation. Classic examples of “mega-splicing” are the neurexin genes (Rowen et al., 2002), which generate thousands of isoforms using multiple promoters and alternative splicing at up to 5 different positions in their pre-mRNAs. The GCOM1 gene employs both mechanisms to augment its proteome and could be part of a larger transcript including exons from two neighboring upstream genes (see Section 4.4). Our previous conservative estimate for the total number of Gup, Gdown and Gcom mRNA transcripts prior to the discovery of the Gcom15 splicing pattern (Roginski et al., 2004) was 284. The recognition of the novel exon 13–exon 21a splice would increase this number at least by a factor of 2.

### 4.4. Functional implications of chromosomal sites of GCOM1 and Y2H Gints

Intriguingly, the genomic neighborhood surrounding GCOM1 on human chromosome 15 from NEDD4 to ADAM10 (~2.9 Mb) includes several genes that encode functions that may be related to and possibly shed light on those of GCOM1. First and foremost, two genes located immediately upstream from and centromeric to GCOM1, TCF12 and CGNL1, are transcribed in the same direction (on the “plus” strand, towards the telomere) as GCOM1. The segment harboring these genes spans ~864 kb and could therefore constitute a ***megagene*** composed of 4 genes (recall that GCOM1 is formed from the upstream and downstream genes). If the same combinatorial mechanism that occurs in the GCOM1 complex transcription unit were applied to the megagene, 15 groups of transcripts and proteins could be produced. Amazingly, TCF12, which is involved in ERK signaling and other pathways, is a transcription factor (like Gdown1). Among TCF12's interacting genes are the Y2H Gints ID2 and SRI. CGNL1 is a coiled-coil protein (as are Gcom1 and Gcom15) involved in tight cell–cell junctions (as is Gup1). NEDD4, a GeneCards Gint of both Gdown1 and Gup1, which is expressed in neural precursor cells and is involved in NOTCH signaling, is located ~1.6 Mb centromeric to GCOM1. NEDD4 lists the Y2H Gint CUL1 among its interacting genes. Finally, ADAM10, an α-secretase involved in Alzheimer's disease, lies ~1 Mb telomeric to GCOM1 and encodes a 748 aa type 1 membrane protein, about the same size as Gcom15 (765 aa), which is also a putative type 1 membrane protein.

Intriguingly, the cluster of 4 genes at human 9q34.3, OLFM1-NOTCH1-GRIN1-NSMF, three of which are Gints, strongly suggests nervous system roles for GCOM1, based on information in GeneCards. First, the Gints OLFM1, GRIN1 and NSMF all interact with INA. Second, OLFM1 also interacts with the Gints SNAP25 and GRIN1 and with KALRN, which interacts with the OLFM1-proximal genes GRIN1 and NOTCH1. KALRN also interacts with DISC1, which is an interactor of OLFM1. Finally, among the interactors of NOTCH1 are the Y2H Gints CUL1 and ATP6V1D and the GCOM1-proximal genes TCF12 and ADAM10 on 15q. The finding of multiple interactions between GCOM1-interacting genes with each other and with neurologically related genes, especially in clusters on human chromosomes 15 and 9, appears to be nonrandom and may facilitate coordinate regulation.

### 4.5. Possible roles of GCOM1-NMDAR interaction and implications for neuroprotection

Our demonstration of bidirectional co-IP between human Gcom15 and mouse NR1-1a proteins strongly suggests that this interaction may also occur between the orthologous proteins in their respective species, e.g., human GCOM1 and human GRIN1, due to the high degree of sequence conservation of the GCOM1 and GRIN1 genes in mammalian species. The finding of a GCOM1-GRIN1 interaction by co-IP is also consistent with our previous results in rat CNS neuronal cultures which suggested an interaction between these genes (Roginski et al., 2008), as well as our finding of enrichment of rat Gcom15 in the plasma membrane fraction, a location necessary for a role in synaptic function. Although further studies will be necessary to elucidate the details of the GCOM1-GRIN1 interaction in vivo, the likelihood of its existence led us to ask two broad questions about its possible role in the CNS: (1) Do GCOM1 proteins participate in a subset of NMDA receptor complexes, and, if so, might they modulate the functions of NMDA receptors? (2) Could stimulation of NMDARs lead to changes in neuronal gene expression by a novel pathway that transduces the glutamate signal from the NMDAR via the GRIN1 subunit through GCOM1 protein(s) to downstream effector molecules?

Regarding the first question, GCOM1 proteins may alter the calcium permeability of NMDARs. In a preliminary study, one of us (RSR) has shown that pre-treatment of rat neuronal cultures with anti-GCOM1 antibodies (S23-E38 of the Gcom1/Gcom15 amino acid sequence) reduced both the rate of rise and peak values of $Ca^{2+}$ influx in response to applied NMDA (Roginski, 2007). Based on our data in this paper, it appears probable that the effect on calcium entry may result primarily from binding of the anti-GCOM1 antibodies to Gcom15 proteins, which could disrupt Gcom15's interaction with the GRIN1 subunit. Decreased calcium entry into neurons is also a plausible explanation for the protection against NMDA toxicity afforded by anti-GCOM1 antibodies (Roginski et al., 2008). Tying these several threads together, we speculate that GCOM1 combined proteins may participate in a subset of NMDA receptor complexes (NRCs) and thereby modulate NMDAR-mediated $Ca^{2+}$ influx into neurons. It remains to be determined whether Gcom proteins are associated primarily with synaptic or extra-synaptic NRCs. By avoiding strategies that *directly* inhibit NMDARs (which have failed in clinical trials), we suggest that GCOM1 proteins may provide attractive targets to investigate as one component of a combination approach for achieving clinically useful neuroprotection.

The answer to the second question may also have implications for neuroprotective strategies in at least two ways. First, the Gcom15 protein contains the same downstream carboxy-terminal exons (21a/22/23) as in the Gdown1 protein, which is the 13th subunit of RNA polymerase II (Hu et al., 2006). This might effect changes in gene expression as follows. Stimulation of NMDARs complexed with Gcom15 might result in the proteolytic cleavage of fragments derived from Gcom15 and/or other Gints (such as INA). Such fragments could be transported to the nucleus wherein they might affect transcription by competition with Gdown1 already bound to RNAP2, thus interfering with the functionality Gdown1 (Espinosa, 2012). A second mode would invoke recruitment of Gints that interact with Gcom15 (and possibly Gcom1) after activation of NMDAR-GCOM1 synapses (see Section 4.7).

### 4.6. Yeast two-hybrid GCOM1 interacting genes: consistency with hypotheses and implications for GCOM1 functions

At least two thirds of the protein–protein interactions revealed in Y2H screens can be verified in mammalian or other eukaryotic organisms. For example, Zoghbi and colleagues verified 83% of their Y2H interactions in mammalian cells (Lim et al., 2006). Therefore, it is likely that most of our Y2H Gints are capable of physiologically important interactions with GCOM1 proteins. Nevertheless, such likely PPIs need to be verified by other means. Pending such verification, however, we may draw some tentative inferences about the roles of GCOM1 proteins based upon the classes of genes revealed in our Y2H screens. First of all, we screened an adult brain cDNA library because the brain expresses the largest fraction of the proteome. Furthermore, we knew from EST databases and our prior work (Roginski et al., 2004) that GCOM1 expresses exons from the entire complex gene in the brain, i.e., the combined, upstream and downstream genes, such that the CNS has a larger number of GCOM1 proteins with which interactions might occur. Therefore, it is not surprising that all of the Gints are expressed in brain and that 8 Gints show their highest levels of tissue-specific expression in the brain (GeneCards). Second, the Y2H Gints could easily be assigned to groups which are consistent with our hypotheses, especially the role in synaptic function of NMDA receptors and possible regulation of neuronal gene expression. Third, bioinformatic analyses revealed interaction networks indicative of several important neurophysiologic pathways. Fourth, the observation that several pairs of our Y2H Gints and the larger group of GCOM1 interacting genes (which includes the Gcom1/Y2H + Gup1/MYZAP + Gdown1/POLR2M Gints) are located close together on specific human chromosomes may also suggest functional significance, consistent with insights gleaned from IPA networks. Also of note, we did not expect to detect the GRIN1 gene in our screens because the yeast two-hybrid system inherently favors the identification of soluble (i.e., cytosolic and nuclear) rather than membrane-associated proteins. Because of this constraint, Gcom1's 27 Y2H Gints represent a lower limit on the number of interactors for this protein. Tallying the GCOM1 interacting genes (for Gcom, Gup and Gdown proteins) from all sources (Section 3.8), we estimate a lower limit of 140 for the number of genes in the GCOM1 interactome.

### 4.7. GCOM1-INA protein–protein interaction in mammalian CNS

One of the likely functions of the GCOM1 proteins is a role in glutamatergic neurotransmission, mediated by an interaction with the GRIN1 subunit of the NMDAR. This hypothesis is bolstered by the Gene Ontology section of the GCOM1 entry in GeneCards, which lists intracellular signal transduction as the sole biological process. Additional roles in normal neurologic function and diseases are suggested by sequence similarities between Gcom1/Gcom15 and INA and LCA5 proteins, which are involved in neurologic diseases. How might the GCOM1 combined proteins play a role in signal transduction? We propose that activation of NRCs results in cleavage of fragments from GCOM1 proteins (most probably Gcom15) associated therewith. Such fragments may bind to internexin-α to facilitate transfer of these complexes to the nucleus (Fig. 7). There, the fragments of GCOM1 proteins, which contain partial sequences of Gdown1/POLR2M, may alter transcriptional programs, e.g., by displacing Gdown1 already bound to RNA polymerase II or by interacting with other transcription-relevant Gints expressed in the nucleus, such as the

Y2H Gints PRPF38B, RIIAD1 and ID2 and/or the GeneCards Gints TANK, NFKB1A and NFKB1/2. Precedent exists for such a role of internexin-α: INA fragments associate with NSMF to facilitate the nuclear transit of signals initiated at synaptic (but not extrasynaptic) NMDARs. Once inside the nucleus, NSMF induces the dephosphorylation of CREB proteins (Karpova et al., 2013). Therefore, the Gcom15-internexin-α interaction might be the entry point to a previously unrecognized, biologically important pathway, which may play a role in neuroprotection (Roginski et al., 2008).

In summary, we have cloned, expressed and performed immunologic experiments with Gcom15 cDNAs and proteins from two mammalian species, demonstrating that Gcom15 is the most likely candidate responsible for the GCOM1-GRIN1 interaction we observed previously. These studies lend further support to the complex gene model of the GCOM1 gene. We have expanded our knowledge of the interactome of the GCOM1 gene by yeast two-hybrid screens using cDNAs corresponding to the Gcom1 protein, revealing 27 novel interacting genes, termed Gints, which strengthens our conclusion that GCOM1 is a hub gene. Furthermore, GCOM1's status as a hub gene suggests many areas for future study with likely clinical relevance to neurodegenerative and neuropsychiatric disorders. Finally, evidence supporting the Y2H interaction of Gcom15 with internexin-α was obtained in a mammalian system. Owing to the fact that INA also interacts with the key GRIN1 subunit of the NMDA receptor, our findings suggest the existence of a novel pathway, the details of which should be investigated.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

These studies were supported by: Education and Development funding from the Department of Anesthesiology and Critical Care of the University of Pennsylvania Perelman School of Medicine (2006-2009); a University of Pennsylvania Research Foundation award to RSR (2005-2006); a Department of Veterans Affairs Competitive Pilot Project Funding award to RSR (2013-2015); and NIH RO1 GM107054 to JY. Lee Vandevier received a stipend from the University of Pennsylvania's Provost Undergraduate Research Mentoring (PURM) Program in the summer of 2008 under the mentorship of RSR. The PURM Program also provided the PI with a small amount of funding for supplies. The mouse NR1-1a cDNA was a kind gift from R. S. Zukin, PhD. VA Disclaimer: The contents of this paper do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

## Abbreviations

| | |
|---|---|
| **aa** | amino acid |
| **Ab** | antibody(ies) |
| **BSA** | bovine serum albumin |
| **cDNA** | DNA complementary to RNA |
| **co-IP** | IP by a second Ab of a complex precipitated by a primary Ab |
| **CTU** | complex transcription unit |

| | |
|---|---|
| **GCOM1** | GRINL1A combined gene 1 aka and originally assigned as gene symbol GRINL1A |
| **Gcom1** | GCOM1 combined protein 1 |
| **Gcom15** | GCOM1 combined protein 15 |
| **Gdown1** | GCOM1 downstream protein 1/POLR2M/13th subunit of RNA polymerase II |
| **Gup1** | GCOM1 upstream protein 1/MYZAP/myocardial intercalated disc protein |
| **Gint(s)** | GCOM1-interacting gene(s) |
| **IP** | immunoprecipitation |
| **kDa** | kiloDalton(s) |
| **NMDA(R)** | N-methyl D-aspartate (receptor) |
| **NR1** | NMDA receptor subunit 1/GRIN1/Grin1 |
| **NRC(s)** | NMDA receptor complex(es) |
| **ORF** | open reading frame |
| **PAGE** | polyacrylamide gel electrophoresis |
| **PMSF** | phenylmethylsulfonyl fluoride |
| **PPI** | protein–protein interaction |
| **RBP** | rat brain protein preparation |
| **(RT)PCR** | (reverse transcriptase) polymerase chain reaction |
| **Y2H** | yeast two-hybrid |

## References


Cairns NJ, Zhukareva V, Uryu K, Zhang B, Bigio E, et al. Alpha-internexin is present in the pathological inclusions of neuronal intermediate filament inclusion disease. Am J Pathol. 2004; 164:2153–2161. [PubMed: 15161649]

Cheng B, Li T, Rahl PB, Adamson TE, Loudas NB, et al. Functional association of Gdown1 with RNA polymerase II poised on human genes. Mol Cell. 2012; 45:38–50. [PubMed: 22244331]

Dharmaraj S, Li Y, Robitaille JM, Silva E, Zhu D, et al. A novel locus for Leber congenital amaurosis maps to chromosome 6q. Am J Hum Genet. 2000; 66:319–326. [PubMed: 10631161]

Espinosa JM. Get back TFIIF, don't let me Gdown1. Mol Cell. 2012; 45:3–5. [PubMed: 22244325]

Hernandez-Sanchez C, Bartulos O, Valenciano AI, Mansilla A, de Pablo F. The regulated expression of chimeric tyrosine hydroxylase-insulin transcripts during early development. Nucleic Acids Res. 2006; 34:3455–3464. [PubMed: 16840532]

Hu X, Malik S, Negroiu CC, Hubbard K, Velalar CN, et al. A mediator-responsive form of metazoan RNA polymerase II. Proc Natl Acad Sci U S A. 2006; 103:9506–9511. [PubMed: 16769904]

Huo L, Wen W, Wang R, Kam C, Xia J, et al. Cdc42-dependent formation of the ZO-1/MRCKbeta complex at the leading edge controls cell migration. EMBO J. 2011; 30:665–678. [PubMed: 21240187]

Husi H, Ward MA, Choudhary JS, Blackstock WP, Grant SG. Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. Nat Neurosci. 2000; 3:661–669. [PubMed: 10862698]

Jishage M, Malik S, Wagner U, Uberheide B, Ishihama Y, et al. Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II. Mol Cell. 2012; 45:51–63. [PubMed: 22244332]

Karpova A, Mikhaylova M, Bera S, Bar J, Reddy PP, et al. Encoding and transducing the synaptic or extrasynaptic origin of NMDA receptor signals to the nucleus. Cell. 2013; 152:1119–1133. [PubMed: 23452857]

Lim J, Hao T, Shaw C, Patel AJ, Szabo G, et al. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell. 2006; 125:801–814. [PubMed: 16713569]

Lin JW, Wyszynski M, Madhavan R, Sealock R, Kim JU, et al. Yotiao, a novel protein of neuromuscular junction and brain that interacts with specific splice variants of NMDA receptor subunit NR1. J Neurosci. 1998; 18:2017–2027. [PubMed: 9482789]

Parra G, Reymond A, Dabbouseh N, Dermitzakis ET, Castelo R, et al. Tandem chimerism as a means to increase protein complexity in the human genome. Genome Res. 2006; 16:37–44. [PubMed: 16344564]

Roginski RS. Decrease in rate of NMDA-induced $Ca^{2+}$ entry may explain neuroprotection by anti-GRINL1A antibodies. J neurosurgical Anesthesiology. 2007; 19:325.

Roginski RS, Mohan Raj BK, Finkernagel SW, Sciorra LJ. Assignment of an ionotropic glutamate receptor-like gene (GRINL1A) to human chromosome 15q22.1 by in situ hybridization. Cytogenet Cell Genet. 2001; 93:143–144. [PubMed: 11474202]

Roginski RS, Mohan Raj BK, Birditt B, Rowen L. The human GRINL1A gene defines a complex transcription unit, an unusual form of gene organization in eukaryotes. Genomics. 2004; 84:265–276. [PubMed: 15233991]

Roginski RS, Goubaeva F, Mikami M, Fried-Cassorla E, Nair MR, et al. GRINL1A colocalizes with N-methyl D-aspartate receptor NR1 subunit and reduces N-methyl D-aspartate toxicity. Neuroreport. 2008; 19:1721–1726. [PubMed: 18849881]

Rowen L, Young J, Birditt B, Kaur A, Madan A, et al. Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. Genomics. 2002; 79:587–597. [PubMed: 11944992]

Seeger TS, Frank D, Rohr C, Will R, Just S, et al. Myozap, a novel intercalated disc protein, activates serum response factor-dependent signaling and is required to maintain cardiac function in vivo. Circ Res. 2010; 106:880–890. [PubMed: 20093627]

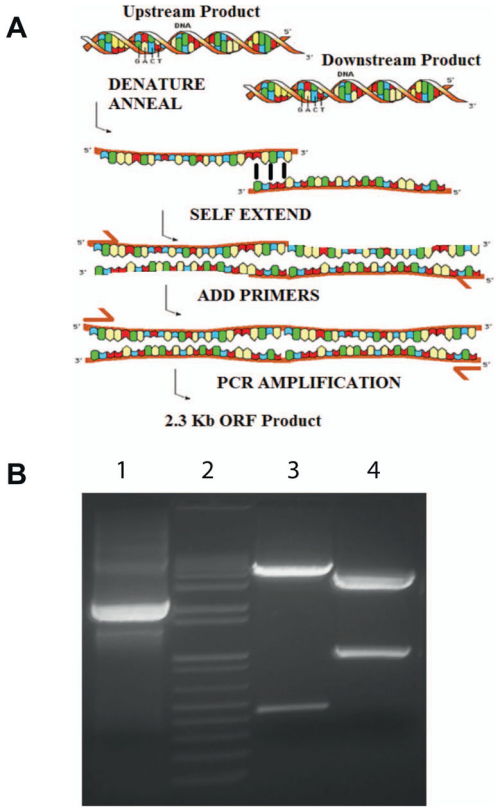


**Fig. 1.**

Hybridization-extension amplification of overlapping human Gcom15 PCR products to reconstitute the complete ORF. (A) Extension of hybrid template formed by annealing of overlapping upstream (from 5′ product) and downstream (from 3′ product) strands followed by self-primed extension (filled in bases). Of the four possible hybridization products, only the 5′ sense strand/3′ anti-sense strand hybrid can be converted to double-stranded DNA in the self-extension phase. After addition of nested primers targeting the 5′ and 3′ untranslated regions, only this template, containing the entire, reconstructed ORF plus segments of the 5′ and 3′ untranslated regions, can be amplified. (B) Agarose gel electrophoresis of human and rat Gcom15 cDNAs. 1, Human Gcom15 amplicon containing the complete ORF plus segments of the 5′ and 3′ non-coding regions, 2.3 kb. 2, 1 kb DNA ladder. 3 and 4, rat Gcom15 cDNA plasmids, both digested with BglII and SpeI, containing overlapping 5′ and 3′ segments, respectively. Rat Gcom15 PCR products were clonable in the proper orientations to permit reconstitution of the ORF by combining the A fragment from the plasmid in lane 3 with the B fragment from the plasmid in lane 4.

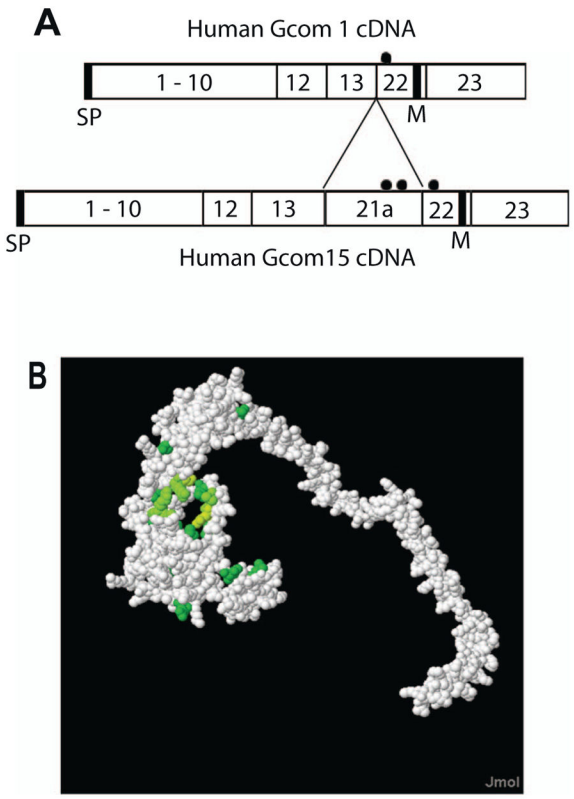


**Fig. 2.**

Human GCOM1 combined proteins. (A) Exon diagrams of human Gcom1 (AY207007) and Gcom15 (JF419331) cDNAs, showing conceptually how Gcom15 can be derived from Gcom1 by splicing exon 21a between exons 13 and 22. SP, signal peptide. M, putative membrane spanning region. Dots, consensus asparagine-linked glycosylation sites. (B) Space-filling model of human Gcom15 protein from Phyre software (Jmol).

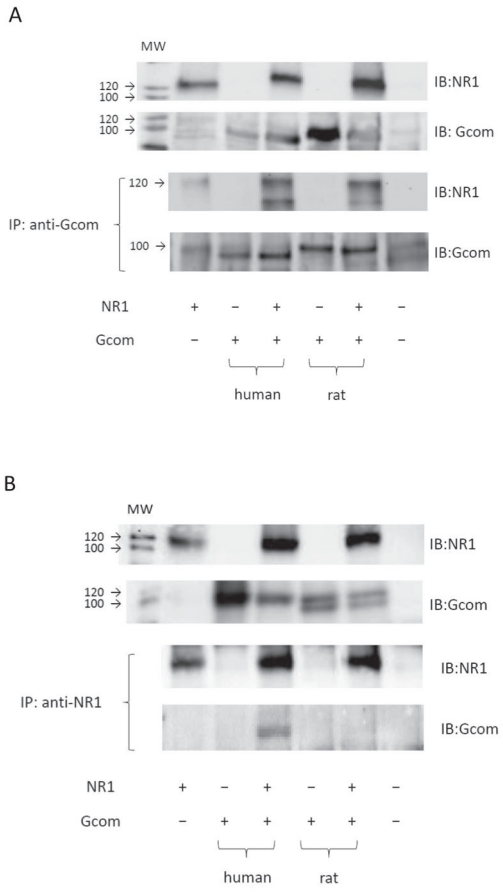


**Fig. 3.**

Expression, immunoprecipitation and co-immunoprecipitation of GCOM1 and GRIN1 proteins in HEK293 cells. Human and rat Gcom15 and mouse NR1-1a cDNAs were transfected as described in Methods. Anti-Gcom1 S23-E38 (rabbit polyclonal antibody) and anti-NR1 SC-1467 (goat polyclonal antibody) were used for detection (IB) and immunoprecipitation (IP).

(A) IP with anti-GCOM; (B) IP with anti-NR1.

Upper panels. Detection of Gcom15 and NR1 immunoreactivity. The human and rat Gcom15 proteins both migrate as 105 kDa bands. The NR1 subunit migrates at 120 kDa. Note that intensity levels of NR1 immunoreactive bands are greater in the presence of co-transfected Gcom15 proteins. GAPDH loading controls (not shown) indicated that equal amounts of transfected cell proteins were electrophoresed and transferred.

Lower panels. Co-immunoprecipitation. (A) NR1 immunoreactivity is detected in the presence of co-transfected human and rat Gcom15 cDNAs/proteins. (B) Gcom15 immunoreactivity is detected in the presence of co-transfected human (but not rat) cDNA/protein.

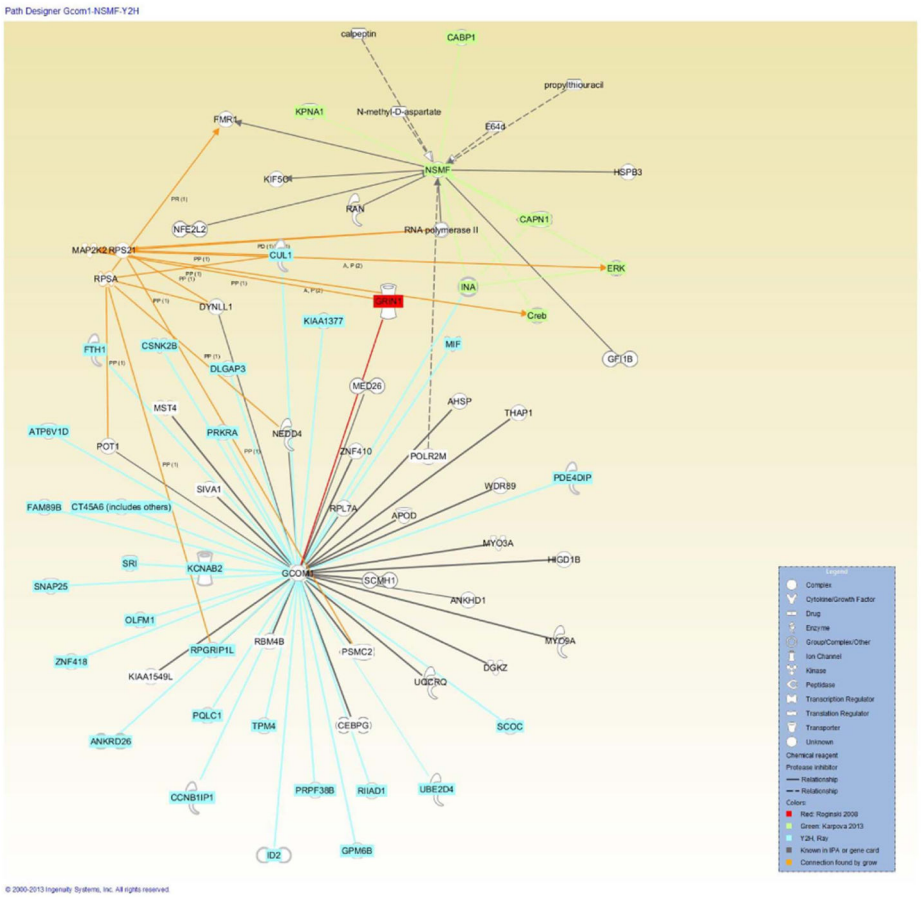


**Fig. 4.**

Network analysis (IPA Path Designer, Grow 3) incorporating gene symbols of the broad GCOM1 interactome. The program predicts novel possible connections (orange arrows) and a putative interaction between the GCOM1 and NSMF hubs. As discussed in the text, not all known interactions are included in the diagram. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

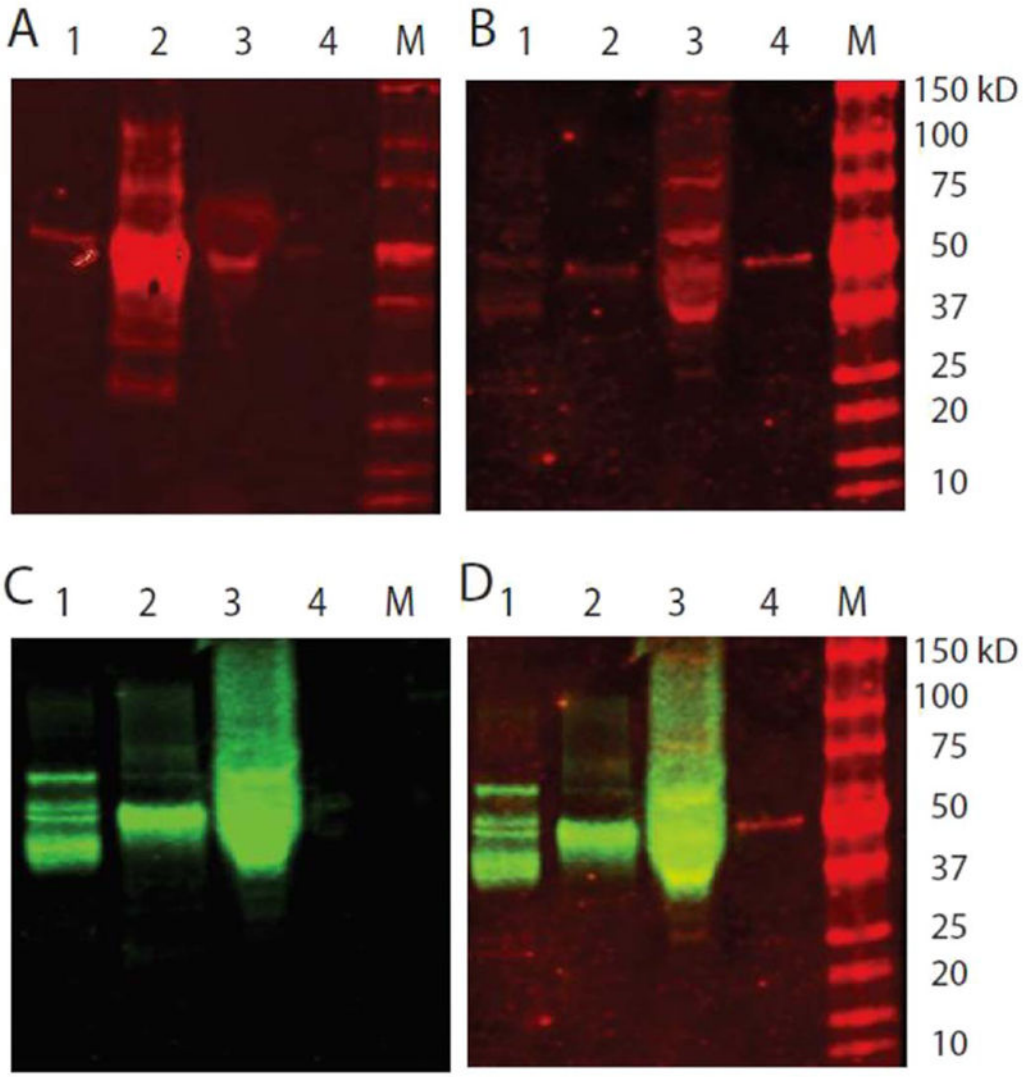


**Fig. 5.**

Immunoprecipitation with anti-GCOM1 antibody reveals co-IP of internexin-α. Replicate blots (A and BCD) were loaded, immunoprecipitated (IP), stained with primary and secondary antibodies and detected at appropriate wavelengths with the Odyssey system (Li-Cor) as follows. Lane identities: 1, Total rat brain protein lysate (RBP), 20 μg. 2, Proteins IP from 400 μg RBP with anti-Gcom1 T423-Q440 Ab (GenScript rabbit 1586). 3, Flow through from 400 μg RBP IP with anti-Gcom1 T423-Q440. 4, Proteins IP from 400 μg RBP with anti-synaptophysin Ab (mouse monoclonal antibody; Millipore MAB329). 5, Size standards (Li-Cor fluorescent and Bio-Rad pre-stained). Staining: Primary antibodies: (A) rabbit anti-Gcom1 T423-Q440; (B) chicken anti-Gcom1 N-terminus S23-E38 (GenScript SC1492); (C) anti-INA (rabbit polyclonal antibody; Millipore AB5354). Secondary antibodies (all Li-Cor): (A) goat anti-rabbit IgG IRDye 680RD; (B) donkey anti-chicken IgG IRDye 680RD; (C) goat anti-rabbit IgG IRDye 800CW. (D) merged images of B and C. Note that the ~50 kDa band in Fig. 5B, lane 4 (anti-SYP IP proteins stained with chicken

anti-Gcom1 S23-E38 Ab) is probably an artifact because no immunoreactivity is present in Fig. 5A, lane 4.

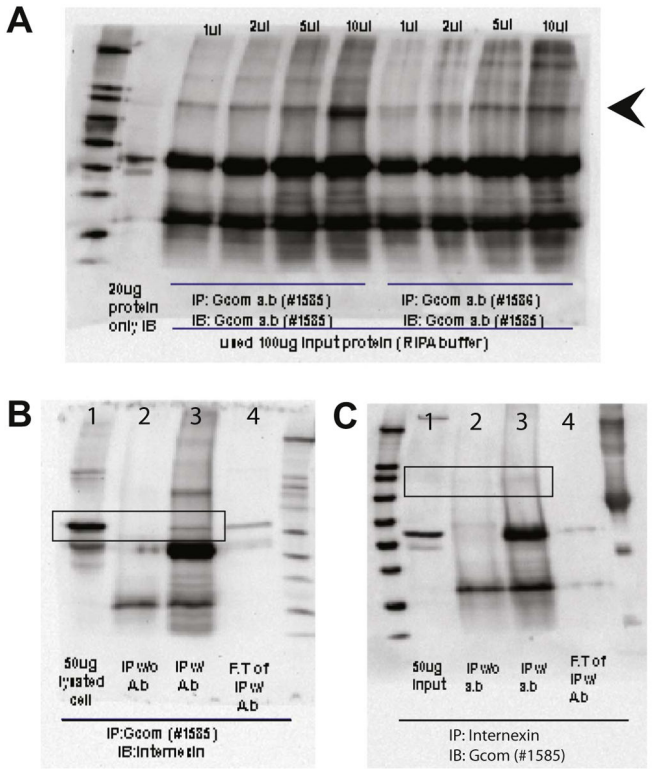


**Fig. 6.**

Co-immunoprecipitation of GCOM1 and INA proteins in heterologous cells. (A) Control IP of rat brain lysate (100 μg; lanes 3–6 and 7–10) with increasing amounts (1 μL to 10 μL) of two different rabbit polyclonal anti-Gcom1 antibody preparations (from rabbits 1585 and 1586, respectively). Lanes: 1, Molecular size standards. 2, rat brain lysate 20 μg. The predominant GCOM1 protein detected under these conditions is Gup1 (54 kDa). The intense bands migrating at ~50 and 25 kDa in lanes 3–10 are from IgG heavy and light chains, respectively. (B) IP of rat brain lysate (100 μg) with anti-Gcom1 antibody (rabbit 1585; 10 μL) stained with anti-INA antibody (1:500 dilution). Lanes: 1, input brain lysate. 2, negative control IP without anti-Gcom1 Ab. 3, IP with anti-Gcom1 Ab. 4, flow-through of unbound wash from lane 3 sample. The expected INA band is 66 kDa (boxed). (C) Same lane assignments as in panel B, except IP of brain lysate with anti-INA Ab and probed with anti-Gcom1 antibody. The rectangular box includes the 105 kDa position, which is the size of the faint Gcom15 protein band in lane 3 (see text). The predominant GCOM1 protein visible in lane 1 is Gup1 migrating at its expected size of 54 kDa. As in panel A, the low abundance Gcom1 and Gcom15 proteins are not seen in lane 1. Owing to the fact that this MW overlaps the IgG heavy chain position (~50 kDa) in lane 3, it is not possible to assess co-IP of INA and Gup1 proteins.

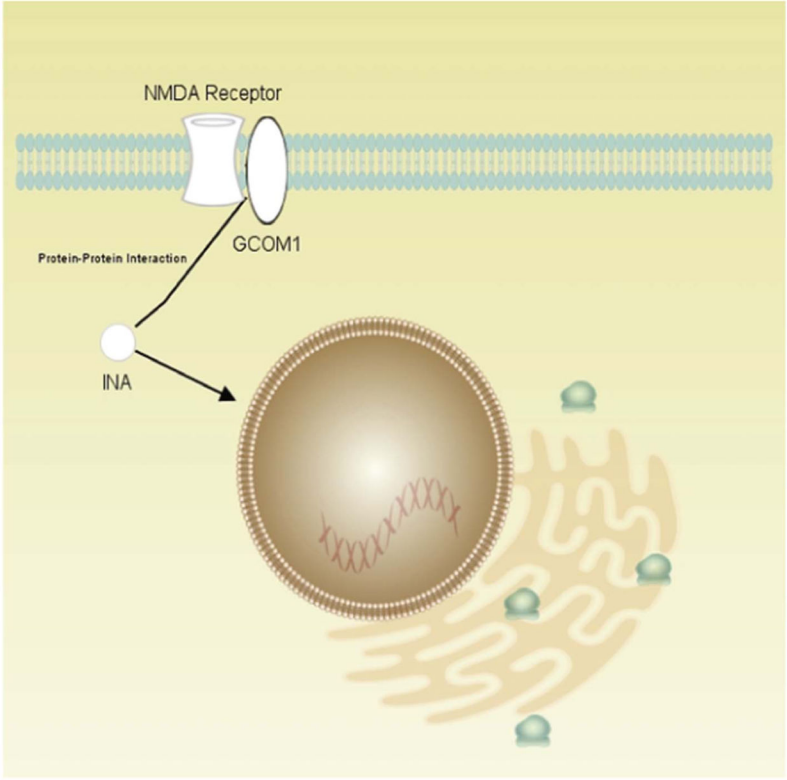


**Fig. 7.**

Proposed linear pathway based on interactions of GRIN1, GCOM1 and INA. The diagram illustrates one of the two hypothetical ways (see Sections 4.5 and 4.7) in which a synaptic signal that targets the NMDA receptor-GCOM1 complex may be transduced to the neuronal nucleus: the signal would travel from the NMDAR (via the GRIN1 subunit) through GCOM1 (most likely via the Gcom15 protein) to the neuronal intermediate filament protein internexin-α (INA). Because INA also is known to interact with GRIN1 and three of the GRIN2 subunits, the second hypothetical pathway (not shown) postulates a 3-way interaction of the NMDAR, GCOM1 proteins and INA which may transduce the signal using the same or different downstream effectors as the linear pathway.

**Table 1**

Comparison of Gcom1 with novel Gcom15 transcripts and proteins.

| GCOM1 protein | Length (aa) | MW (kDa) | Exon composition | cDNA (bp) | Accession number |
|---|---|---|---|---|---|
| Gcom1 human | 550 | 64.0 | 1–10/12/13/22/23 | 1721 | AY207007 |
| Gcom15 human | 765 | 87.8 | 1–10/12/13/21a/22/23 | 2329 | JF419331 |
| Gcom15 rat | 761 | 86.9 | 1–10/12/13/21/22/23 | 2328 | JF440303 |

*Gene*. Author manuscript; available in PMC 2018 June 20.

## Table 2

Human chromosome sites of GCOM1 Y2H Gints.

| | Gene symbol | Description | Chromosome band | Start base |
|---|---|---|---|---|
| 1 | ANKRD26 | Ankyrin repeat domain 26 | 10p12.1 | 26,991,914 |
| 2 | ATP6V1D | ATPase, H+ transporting, lysosomal 34 kDa, V1 subunit D | 14q23.3 | 67,294,371 |
| 3 | CCNB1IP1 | Cyclin B1 interacting protein 1, E3 ubiquitin protein ligase | 14q11.2 | 20,311,368 |
| 4 | CSNK2B | Casein kinase 2, beta polypeptide | 6p21.33 | 31,655,236 |
| 5 | CT45A3 | Cancer/testis antigen family 45, member A3 | Xq26.3 | 135,759,846 |
| 6 | CUL1 | Cullin 1 | 7q36.1 | 148,697,914 |
| 7 | DLGAP3 | Discs large homolog-associated protein 3 | 1p34.3 | 34,865,436 |
| 8 | FAM89B | Family with sequence similarity 89, member B | 11q13.1 | 65,572,349 |
| 9 | FTH1 | Ferritin, heavy polypeptide 1 | 11q12.3 | 61,959,718 |
| 10 | GPM6B | Glycoprotein M6B | Xp22.2 | 13,770,943 |
| 11 | ID2 | Inhibitor of DNA binding 2, dominant negative HLH protein | 2p25.1 | 8,678,845 |
| 12 | INA | Internexin neuronal intermediate filament protein, alpha | 10q24.33 | 103,277,163 |
| 13 | KCNAB2 | Potassium channel, voltage gated subfamily A regulatory beta subunit 2 | 1p36.31 | 5,991,466 |
| 14 | KIAA1549L | Chromosome 11 open reading frame 41; KIAA1549-like | 11p13 | 33,542,072 |
| 15 | MIF | Macrophage migration inhibitory factor (glycosylation-inhibiting factor) | 22q11.23 | 23,894,004 |
| 16 | OLFM1 | Olfactomedin 1 | 9q34.3 | 135,075,243 |
| 17 | PDE4DIP | Phosphodiesterase 4D interacting protein | 1q21.2 | 148,808,181 |
| 18 | PQLC1 | PQ loop repeat containing 1 | 18q23 | 79,902,420 |
| 19 | PRKRA | Protein kinase, interferon-inducible double stranded RNA dependent | 2q31.2 | 178,431,414 |
| 20 | PRPF38B | Pre-mRNA processing factor 38B | 1p13.3 | 108,692,310 |
| 21 | RIIAD1 | Regulatory subunit of type II PKA R-subunit (RIIa) domain containing 1 | 1q21.3 | 151,710,433 |
| 22 | RPGRIP1L | RPGRIP1-like | 16q12.2 | 53,597,683 |
| 23 | SCOC | Short coiled-coil protein | 4q31.1 | 140,257,286 |
| 24 | SNAP25 | Synaptosomal-associated protein, 25 kDa | 20p12.2 | 10,218,694 |
| 25 | SRI | Sorcin | 7q21.12 | 88,205,115 |
| 26 | TPM4 | Tropomyosin 4 | 19p13.12 | 16,067,021 |
| 27 | ZNF418 | Zinc finger protein 418 | 19q13.43 | 57,921,884 |