**Molecules
and
Cells**

©KSMCB 2007

Gateway RFP-Fusion Vectors for High Throughput Functional Analysis of Genes

Jae-Yong Park[†], Eun Mi Hwang[†], Nammi Park, Eunju Kim, Dong-Gyu Kim, Dawon Kang, Jaehee Han, Wan Sung Choi[1], Pan-Dong Ryu[2], and Seong-Geun Hong*

Department of Physiology, Institute of Health Science, Medical Research Center for Neural Dysfunction, College of Medicine, Gyeongsang National University, Jinju 660-751, Korea;

[1] *Department of Anatomy, Institute of Health Science, and Medical Research Center for Neural Dysfunction, College of Medicine, Gyeongsang National University College of Medicine, Jinju 660-751, Korea;*

[2] *Department of Veterinary Pharmacology, College of Veterinary Medicine and BK21 Program for Veterinary Science, Seoul National University, Seoul 151-742, Korea.*

There is an increasing demand for high throughput (HTP) methods for gene analysis on a genome-wide scale. However, the current repertoire of HTP detection methodologies allows only a limited range of cellular phenotypes to be studied. We have constructed two HTP-optimized expression vectors generated from the red fluorescent reporter protein (RFP) gene. These vectors produce RFP-tagged target proteins in a multiple expression system using gateway cloning technology (GCT). The RFP tag was fused with the cloned genes, thereby allowing us localize the expressed proteins in mammalian cells. The effectiveness of the vectors was evaluated using an HTP-screening system. Sixty representative human C2 domains were tagged with RFP and overexpressed in HiB5 neuronal progenitor cells, and we studied in detail two C2 domains that promoted the neuronal differentiation of HiB5 cells. Our results show that the two vectors developed in this study are useful for functional gene analysis using an HTP-screening system on a genome-wide scale.

Keywords: Gateway Cloning System; Red Fluorescent Protein; High Throughput Screening System.

[†] These authors contributed equally to this work.

* To whom correspondence should be addressed.
Tel: 82-55-751-8741; Fax: 82-55-759-0169
E-mail: hong149@gsnu.ac.kr

Introduction

Over the last decade rapid progress in genome sequencing has led to the identification of great numbers of genes, and has made it possible to create and apply repositories of defined functional elements to perform high throughput (HTP) genome-wide analyses (Gibbs *et al.*, 2004; Waterston *et al*., 2002). However, most gene products predicted from the currently available genome sequences remain functionally uncharacterized. One essential step in the development of genome-wide analyses is the systematic mapping of macromolecular interactions and biochemical reactions using reverse proteomics approaches (Walhout and Vidal, 2001). Reverse proteomics projects, in turn, require the cloning and manipulation of large numbers of protein-encoding sequences, or open reading frames (ORFs).

Recently, cloning systems employing *in vitro* recombination have been developed, and are amenable to HTP cloning (Hartley *et al.*, 2000; Siegel *et al.*, 2004). Gateway cloning technology (GCT) is one such system, which uses *in vitro* site-specific recombination to clone and subsequently transfer DNA segments between vector backbones (Hartley *et al.*, 2000). In this system, a fragment carrying *att*B1 and *att*B2 recombination sites at its ends is prepared by the polymerase chain reaction (PCR), and is introduced into a donor vector carrying *att*P1 and *att*P2 recombination sites by an *in vitro* recombination reaction driven by BP clonase. In turn, master vectors (so-called 'entry clones') are produced, which contain the fragment of interest carrying *att*L1 and *att*L2 recombination sites.

Abbreviations: GCT, gateway cloning technology; HTP, high throughput; ORF, open reading frames; RFP, red fluorescent protein.

This system has been used to generate several large clone collections, such as those for *Xenopus* (http://xgc.nci.hih.gov) and zebrafish (http://zgc.nci.nih.gov), as well as many human, mouse, and rat genes (http://mgc.nci.nih.gov). Some full-length clones have also been used to establish entry clones through GCT. Furthermore, an initial version of the human ORFeome, comprising 8,076 human ORFs, was recently generated using GCT (Rual *et al.*, 2004).

Although the GCT is a powerful tool for multi-gene cloning, few practical expression vectors for tagging with fluorescent proteins are commercially available. In this work we generated novel vectors that express red fluorescent protein (RFP)-tagged proteins using GCT. These vectors can be used with any entry clones and can produce any protein of interest fused to RFP in mammalian cell lines. In addition, we tested whether the RFP-tagging vectors could be used in HTP analysis to obtain cellular images of 60 representative human C2 domains.

Materials and Methods

Construction of gateway destination vectors pDsRed-Monomer-N1 (Clontech) was used as a backbone vector for pDEST-N-RFP (Fig. 1A). The suicide *ccd*B box, which contains the *ccdB* gene (Bernard and Couturier 1992), the chloramphenicol (CM) *cat* gene, and *att*R sites, was amplified with pDEST 15 (Invitrogen) as template, and with a *Bam*HI site in the upper primer and a *Bgl*II site in the lower primer (5′-GGG**GGATCC**CCACAAGT-TTGTACAAAAAAGC-3′ and 5′-GGG**TGATCA**ACCACTTT-GTACAAGAAAGC-3′, respectively). The PCR product was partially digested with the same enzymes, and cloned into the *Bam*HI site of pDsRed-Monomer N1. For pDEST-RFP-C, we used pEGFP-C1 (Clontech) as a backbone vector (Fig. 1B). The DsRed-Monomer gene was amplified from pDsRed-Monomer N1 vector, with the upper and lower primers containing *Age*I and *Bam*HI sites, respectively (5′-GGG**ACCGGT**CGCCACC-ATGGACAACACCGAGG-3′ and 5′-GGG**GGATCC**GGACT-GGGAGCCGGAGTGGCG-3′). The suicide *ccd*B box was also amplified from the above-mentioned pDEST15 vector, with upper and lower primers containing *Bgl*II and *Bam*HI sites, respectively (5′-GGG**AGATCT**AACAAGTTTGTACAAAAAAGC-3′ and 5′-GGG**GGATCCTCA**AACCACTTTGTACAAGAAA-GC-3′). A stop codon (underlined in bold) was added to the lower primer to prevent tagging unnecessary amino acids. The two products were partially digested with the same enzymes and were cloned into the *Age*I and *Bam*HI sites of pEGFP-C1 vector.

Construction of expression vectors for RFP-fused C2 domains Entry clones of human C2 domains were constructed as follows. We initially identified 175 published C2 domains with conserved regions from 250 proteins listed in the Pfam (www.sanger.ac.uk/Software/Pfam) and NCBI (www.ncbi.nlm.nih.gov) databases. In order to obtain human C2 domains, reverse transcription (RT)-PCR was performed using total RNAs isolated from human brain, placental tissues and HeLa cells. Human C2 domains were PCR amplified from the pooled complementary DNAs (cDNAs) using target sequence-specific *att*B1/*att*B2 primers. The upper primer (5′-GGGG<u>ACAAGTTTGTACAAA-AAAGCAGGCT</u>**CCACCATG**-[gene specific 20–25 sequences]-3′) and the *att*B1 sequences (underlined) followed by a Kozak sequence and start codon sequences (bold), were introduced. The lower primer was 5′-GGGG<u>ACCACTTTGTACAAGAAA-GCTGGGGTT</u>**TTA**-(gene specific 20–25 sequences)-3′. The *att*B2 sequence (underlined) was introduced immediately after the last stop codon (bold). The PCR products were cloned into the pDONR207 donor vector (Invitrogen), and their sequences confirmed (Macrogen Inc.). 150 human C2-domain entry clones were obtained by GCT. From these, we selected 60 entry clones at random and introduced them into the pDEST-RFP-C destination vector via the LR reaction of GCT.

Cell culture, HTP transfection, and adenovirus infection Immortalized HiB5 cells were cultured at 33°C and 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM) with 0.11 mg/ml sodium pyruvate, 3.7 mg/ml $NaHCO_3$, 50 U/ml penicillin, 50 μg/ml streptomycin, and 10% fetal bovine serum (FBS). For differentiation as previously reported (Kwon, 1997; Sung *et al.*, 2001), cells were shifted to the chemically defined N2 medium containing 5 μg/ml insulin, 100 μg/ml transferrin, 20 nM progesterone, 30 nM selenium, 60 μM putrescine, 0.11 mg/ml sodium pyruvate, and 2 mM glutamine at 39°C for 2 d.

The expression-ready subset of human C2 domains was transfected into cultured HiB5 cells at 33°C in 96-well plates. In a semi-automated process, each expression vector (200 ng) was incubated with a liposomal transfection reagent (lipofectamin 2000; Invitrogen) and introduced into a well containing ~10,000 HiB5 hippocampal progenitor cells. After further incubating for 48 h, cell images were acquired using an In Cell Analyzer 1000 automated high-content imaging system (Amersham Biosciences). Finally, we screened for human C2 domains that induced neurite outgrowth of HiB5 cells. We also constructed two replication-defective adenoviruses (rAd5) containing GFP-tagged CPNE9-C2 and PKCδ-C2, respectively (Neurogenex Co., Korea). HiB5 cells were infected at the 90% level at a multiplicity of infection (MOI) of 50.

Imaging analysis and immunocytochemistry The screened C2 domains were re-analyzed by the RFP fluorescence-detection method with a confocal microscope (Olympus Fluoview FV 1000) mounted on an inverted microscope (Olympus BX81) and fitted with a 60× objective. The excitation light source was the 543-nm line of a green HeNe laser. The emitted fluorescence was passed through a 580-nm (40-nm bandwidth) primary barrier filter before it reached the photo-multiplier tube. The laser intensity was minimized to prevent dye bleaching. The digital image output was 512 × 512 pixels with a 32-bit resolution. For the imaging analysis the cells were plated onto glass coverslips at 4×10^4 cells/ml. After 2 days they were washed twice with
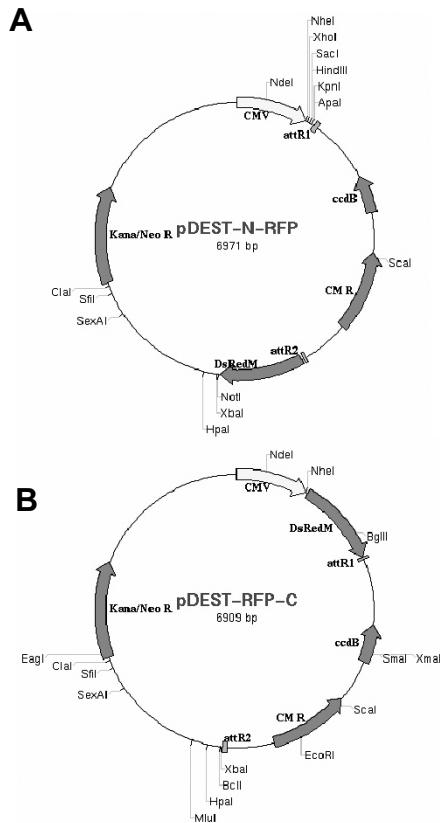
A

B

Fig. 1. Schematic of the new RFP-tagging expression vectors. The two destination vectors, pDEST-N-RFP (A) and pDEST-RFP-C (B) are depicted. These vectors contain the suicide *ccdB* box (*att*B1 site, *ccd*B gene, chloramphenicol cat gene (CM R), and *att*R2 site). The cyan boxes indicate the *att*R1/ *att*R2 recombination sites. The red and yellow arrows show the locations and orientation of several ORFs and the cytomegalovirus immediate-early promoter (CMV), respectively. Restriction enzyme sites are also shown. Kana/Neo R, Kanamycin/Neomycin resistance gene.

DMEM and monitored with the confocal microscope. The immunocytochemisty protocol was modified from Choi *et al.* (2005). Briefly, HiB5 cells grown on coverslips (4×10^4 cells) were fixed in 4% paraformaldehyde and permeabilized in cold methanol for 5 min. They were then incubated with primary pan-neuronal neurofilament marker monoclonal antibody (SMI 311, Covance Research Products) against neurofilament (NF; 1:500) overnight at 4°C. The following day the cells were washed and treated with fluorescein isothiocyanate (FITC)-conjugated secondary antibody (HTP transfection) or Cy3-conjugated secondary antibody (rAd infection) for 1 h at the room temperature. All images were captured on the same confocal microscope.

Results and Discussion

A new set of RFP-tagging expression vectors for HTP
We decided to base our new RFP-tagging expression vec-

tors on recombination-based cloning using the GCT series of plasmids (Invitrogen). GCT uses a site-specific recombination reaction whereby a gene of interest is cloned into a so-called entry vector (pENTR) that is then recombined with a so-called destination vector (pDEST) to produce the desired expression vector. This cloning system permits the easy assembly of a variety of expression vectors in one step. In our case we designed the destination vectors to contain RFP flanked by a suicide *ccdB* box.

The destination vectors we developed are essentially modifications of the existing pDsRed-Monomer-N1 and pDsRed-Monomer-C1, which encode the DsRed-Monomer protein (Fig. 1A). This reporter protein is a monomeric mutant derived from the tetrameric *Discosoma sp.* RFP DsRed (Matz *et al.*, 1999). By performing the gateway recombination reaction with pENTR containing a gene of interest, new destination vectors are produced that express RFP-tagging proteins under the control of the cytomegalovirus immediate-early promoter (CMV).

HTP screening of the putative roles of C2 domains in neuronal differentiation of HiB5 cells Our destination vectors were designed for the HTP screening of gene functions on a genome-wide scale and the effectiveness of these destination vectors for HTP screening was evaluated using C2 domains. The C2 domain, a conserved protein module of ~120 amino acids, was originally defined as homologous to the C2 regulatory region of protein kinase C (reviewed in Newton and Johnson, 1998). It is now known to be present in numerous eukaryotic signaling proteins, including kinases, GTPase-activating proteins, ubiquitination enzymes and proteins involved in vesicular trafficking. In addition, Benes *et al.* (2005) recently showed that the C2 domain of PKCδ can directly bind phosphotyrosine.

To screen for C2 domains involved in neuronal differentiation we selected 60 entry clones of human C2 domains at random and converted them into pDEST-RFP-C expression vectors by the LR reaction of GCT (Fig. 2A and Table 1). The RFP-tagging expression vectors were transiently transfected into HiB5 hippocampal progenitor cells in 96-well plates. After 48h incubation, cellular images were captured, and the effects of the C2 domains on neurite outgrowth were assessed (Fig. 2B).

HiB5 cells comprise a multipotent hippocampal stem cell line isolated from the embryonic day 16 rat hippocampus, in which precursors of pyramidal cells initiate proliferation (Renfranz *et al.*, 1991). They are immortalized by the temperature-sensitive SV40 large T antigen, grow at a permissive temperature of 33°C, and express the stem-cell marker nestin. When incubated in N2 medium at the non-permissive temperature of 39°C, they stop growing and many die. However, 30–40% of the cells survive, and less than 30% of the surviving cells differentiate into cells with neurite-like structures. We

Table 1. Human genes containing C2 domains used in this study. The listed genes are identified by arrows and numbers matched with the images shown in Fig. 2B.

Gene name	Accession NO.	Nucleotide region	Gene name	Accession NO.	Nucleotide region	Gene name	Accession NO.	Nucleotide region
PIK3C2A	NM_002645	4720-4986	ABR	U01147	1516-1785	RGS3	NM_017790	118-381
UNC13B	NM_006377	4318-4590	doc2	D31897	316-585	PIK3C2B	NM_002646	4555-4824
RIMS1	NM_014989	2317-2592	doc2 β	NM_003585	427-696	UNC13C	XM_930037	1462-1734
SYT7	NM_004200	451-714	RASA3	NM_007368	40-288	MCTP2	NM_018349	295-537
dysferlin	NM_003494	4012-4263	PLCL1	NM_006226	1873-2145	DKFZP586P0123	NM_015531	1186-1461
GAP1 like	BC014420	403-648	RASA2	NM_006506	517-819	PLCH2	NM_014638	2493-2890
PLA2G4B	NM_005090	727-981	PLCB1	NM_015192	1-240	SYTL2	NM_032943	259-528
PKCγ	NM_002739	517-780	SYT11	NM_152280	520-786	SYTL2	NM_206930	706-969
PKCβ	NM_002738	517-780	SYT11	NM_152280	922-1191	RPGRIP1	NM_020366	2339-2730
BCR	NM_004327	2737-3006	CPNE9	NM_153635	229-492	NEDD4L	NM_015277	64-327
PKCδ	NM_006254.	1-375	PLCB3	NM_000932	2053-2301	B/K protein	NM_016524	601-882
PKCα	NM_002737	517-780	SYT3	NM_032298	946-1206	WWP1	NM_007013	508-774
SYT1	NM_005639	868-1134	SYT2	NM_177402	466-726	FLJ12548	AK022610	556-834
PLCD1	NM_006225	1891-2160	SYTL4	NM_080737	1582-1851	PLCE1	NM_016341	556-834
cadps2	AF401638	1096-1338	MTAC2D1	NM_152332	1102-1368	RIMS1	NM_014989	4550-4980
PLCG2	NM_002661	3187-3456	PLC ε	AF170071	5611-5835	NEDD4LUL1	AB048365	556-843
myoferlin	NM_013451	4-255	myoferlin	NM_013451	1078-1374	myoferlin	NM_013451	3907-4179
DAB2IP	NM_032552	64-306	C21orf25	NM_199050	793-1038	FAM62B	NM_020728	718-987
SYT13	NM_020826	949-1221	KIAA1301	AB037722	586-873	kiaa1228	NM_020728	1378-1617
PKCε	NM_005400	22-297	KIAA0941	AB023158	43-306	rabphilin-3a	NM_014954	1225-1494

measured the effects of the C2 domains on the formation of these neurite-like structures at 33°C. Based on the results, we screened six C2 domains that appeared to promote the formation of neurite-like structures and settled on the C2 domains of PKCδ (PKCδ-C2) and of copine9 (CPNE9-C2) for further study.

The selected C2 domains promote the neuronal differentiation of HiB5 cells In order to confirm that neurite outgrowth was induced by PKCδ-C2 and CPNE9-C2, each RFP-tagged C2 domain was transfected into HiB5 cells and their ability to induce morphological changes was monitored with a confocal microscope. Since PKCδ-C2 and CPNE9-C2 were fused to RFP it was very easy to detect the transfected cells, which were examined for the presence of neurites, defined as processes longer than the length of two cell bodies. As shown in Fig. 3A, the cells expressing PKCδ-C2 and CPNE9-C2 had well-defined neurite-like structures. In addition, neuronal differentiation was checked by immunocytochemistry for neurofilament (NF), a differentiation marker for neuronal cells; overexpression of PKCδ-C2 or CPNE9-C2 led to an increase in the number of transfected cells bearing neurites at the permissive temperature (Fig. 3B).

Since only 30–40% of the HiB5 cells survive in N2 medium at the non-permissive temperature of 39°C, and transfection efficiency is very low in this condition, it was difficult to assess the activity of the transfected genes. Therefore we constructed recombinant adenoviruses containing GFP-tagged CPNE9-C2 and PKCδ-C2, to overcome the low efficiency of transfection and measure the effect of the two selected C2 domains on neuronal differentiation at 39°C. As shown in Fig. 3C, PKCδ-C2 and CPNE9-C2 clearly increased the number of cells bearing neurites at 39°C.

In accord with our result it has been reported that PKCδ is involved in the nerve growth factor signaling that elicits neuronal differentiation of PC12 cells (O'Driscoll *et al*., 1995) and in neurogenic fibroblast growth factor signaling in both PC12 cells and hippocampal cells (Corbit *et al*., 1999). Moreover the regulatory domain of PKCδ (RD-PKCδ) enhances neurite outgrowth of HiB5 cells independent of the catalytic activity of the protein (Ling *et al*., 2004; Trollér and Larsson, 2006). Since the regulatory domain contains C1 and C2 domains (reviewed in Newton and Johnson, 1998), the PKCδ-C2 -induced neurite outgrowth described above implies that the C2 domain of PKCδ plays an important role in RD- PKCδ-induced neu-
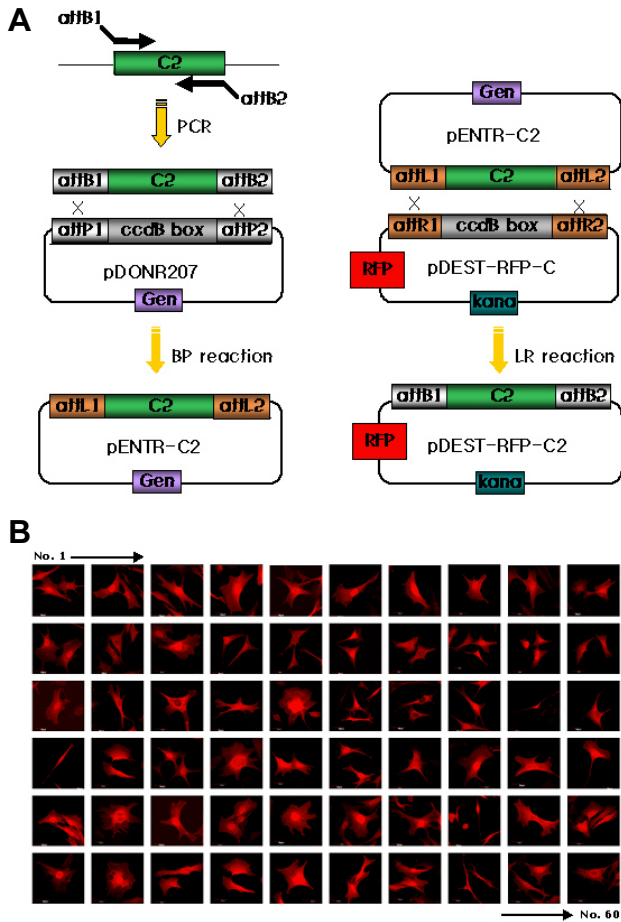
Fig. 2. Gene mining of C2 domains using GCT, and their expression images. **A.** C2 domains were amplified by RT-PCR with primers containing *att*B1/*att*B2 sites, and cloned into pDONR207 by the BP reaction. The C2 domains in pENTR-C2 were converted to pDEST-RFP-C destination vectors by the LR reaction. **B.** pDEST-RFP-C2 expression vectors were transfected into HiB5 cells in 96-well plates. Images were captured using a HTP imaging system (In Cell Analyzer 1000). The captured images are identified by arrows and numbers. Gen, Gentamycin resistance gene; Kana, Kanamycin resistance gene.

rite outgrowth.

Copine 9 (CPNE9) belongs to the copine family, a novel family of ubiquitous Ca^{2+}-dependent phospholipid-binding proteins (reviewed in Tomsig and Creutz, 2002) and was recently cloned (Xie *et al*., 2004). Its role in neural differentiation has not yet been elucidated. Although further studies will be required, our results strongly suggest that the two genes containing these C2 domains play pivotal roles in neuronal differentiation of HiB5 cells.

In conclusion, our novel destination vectors can be successfully applied to HTP screening of cellular image data. In addition, the two vectors allow simultaneous examination of two or more proteins of interest by com- bining them with the EGFP-tagging system, since the fluores-
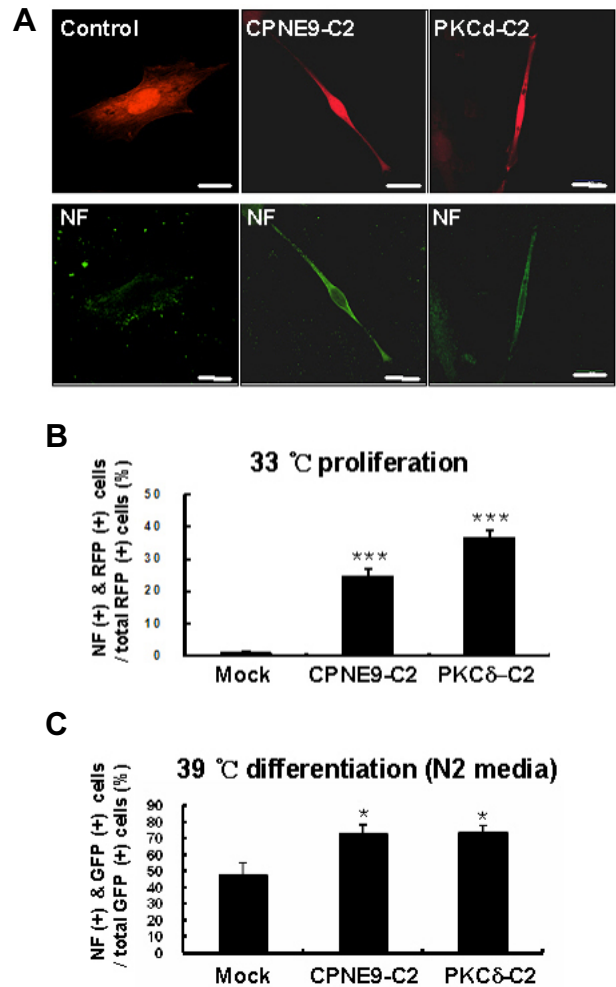

Fig. 3. Images of the screened C2 domains, and C2 domain-induced neuronal differentiation. **A.** Two C2 domains selected from the initial screening were reexamined in HiB5 cells with a confocal microscope, and their ability to induce neurite outgrowth was confirmed by immunoreactivity for neurofilament (NF). **B–C.** Quantification of the number of RFP or GFP-positive HiB5 cells with neuritis (NF-positive) longer than two cell bodies expressing RFP- or GFP-tagged CPNE9-C2 and PKCδ-C2. Data are means ± SEM of three separate experiments (*** $P < 0.001$ or * $P < 0.05$, by one-way ANOVA).

cence of RFP can be readily differentiated from that of EGFP.

References

Benes, C. H., Wu, N., Elia A. E., Dharia, T., Cantley, L. C., *et al.* (2005) The C2 domain of PKCdelta is a phosphotyrosine binding domain. *Cell* **121**, 271–280.

Bernard, P. and Couturier, M. (1992) Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J. Mol. Biol.* **226**, 735–745.

Choi, S. H., Kim, D. W., Kim, S. Y., An, J. J., Lee, S. H., *et al.* (2005) Transduced human copper chaperone for Cu,Zn-SOD (PEP-1-CCS) protects against neuronal cell death. *Mol. Cells* **20**, 401–408.

Corbit, K. C., Foster, D. A., and Rosner, M. R. (1999) Protein kinase Cδ mediates neurogenic but not mitogenic activation of mitogen-activated protein kinase in neuronal cells. *Mol. Cell. Biol.* **19**, 4209–4218.

Gibbs, R. A., Weinstock, G. M., Metzker, M. L., Muzny, D. M., Sodergren, E. J., *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521.

Hartley, J. L., Temple, G. F., and Brasch, M. A. (2000) DNA cloning using in vitro site-specific recombination. *Genome Res.* **10**, 1788–1795.

Kwon, Y. K. (1997) Expression of brain-derived neurotrophic factor mRNA stimulated by basic fibroblast growth factor and platelet-derived growth factor in rat hippocampal cell line. *Mol. Cells* **7**, 320–325.

Ling, M., Trollér, U., Zeidman R., Lundberg, C., and Larsson, C. (2004) Induction of neurites by the regulatory domains of PKCδ and ε is counteracted by PKC catalytic activity and by the RhoA pathway. *Exp. Cell Res.* **292**, 135–150

Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zaraisky, A. G., *et al.* (1999) Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* **17**, 969–973.

Newton, A. C. and Johnson, J. E. (1998) Protein kinase C: a paradigm for regulation of protein function by two membrane-targeting modules. *Biochim. Biophysys. Acta* **1376**, 155–172.

O'Driscoll, K. R., Teng, K. K., Fabbro, D., Greene, L. A., and Weistein, I. B. (1995) Selective translocation of protein kinase C-δ in PC12 cells during nerve growth factor-induced neuritogenesis. *Mol. Cell. Biol.* **6**, 449–458.

Renfranz, P. J., Cunningham, M. G., and McKay, R. D. G. (1991) Region-specific differentiation of the hippocampal stem cell line HiB5 upon implantation into the developing mammalian brain. *Cell* **66**, 713–729.

Rual, J. F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., *et al.* (2004) Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res.* **14**, 2128–2135.

Siegel, R. W., Velappan, N., Pavlik, P., Chasteen, L., and Bradbury, A. (2004) Recombinatorial cloning using heterologous lox sites. *Genome Res.* **14**, 1119–1129.

Sung, J. Y., Lee, S. Y., Min, D. S., Eom, T. Y., Chung, K. C., *et al.* (2001) Differential activation of phospholipases by mitogenic EGF and neurogenic PDGF in immortalized hippocampal stem cell lines. *J. Neurochem.* **78**, 1044–1053.

Tomsig, J. L. and Creutz, C. E. (2002) Copine: a ubiquitous family of Ca^{2+}-dependent phospholipids-binding proteins. *Cell. Mol. Life Sci.* **59**, 1467–1477.

Trollér, U. and Larsson, C. (2006) Cdc42 is involved in PKCε- and δ- induced neurite outgrowth and stress fibre sidmantling. *Biochem. Biophysic. Res. Commun.* **349**, 91–98.

Walhout, A. J. and Vidal, M. (2001) Protein interaction maps for model organisms. *Nat. Rev. Mol. Cell. Biol.* **2**, 55–62.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562.

Xie, H., Diber, A., Pollock, S., Safer, H., Meloon, B., *et al.* (2004) Bridging expressed sequence alignments through targeted cDNA sequencing. *Genomics* **83**, 572–576.