InCeP: Intracellular Pathway Based on mKIAA Protein–Protein Interactions

Masatoshi Murakami,1 Kiyo Shimada,1 Makoto Kawai,1 and Hisashi Koga1,2,*

Chiba Industry Advancement Center, 2-6 Nakase, Mihama-ku, Chiba 261-7126, Japan1 and Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan2

(Received 19 May 2005; revised 2 September 2005; published online 11 January 2006)

Abstract

Since December 2001, we have been conducting a project to isolate and determine entire sequences of mouse KIAA cDNA clones, which encode polypeptides corresponding to human KIAA proteins. The ultimate goal of this project has been elucidation of the functions of KIAA proteins. To address this issue, we have been generating ‘libraries’ of antibodies against mKIAA proteins. We have, to date, already generated >800 antibodies. Using our ‘libraries’ of antibodies, we are now identifying endogenous mKIAA protein–protein interactions. In the present study, novel interactions were identified by MS/MS analysis following immunoprecipitation with anti-mKIAA antibodies. The interactions with biologically known molecules should enable us to predict the function of mKIAA/KIAA proteins, including hypothetical proteins identified in our cDNA project. These interactions are subsequently used for construction of an intracellular pathway related to the mKIAA protein, and the pathway is distributed through the InCeP (IntraCellular Pathway based on mKIAA protein–protein interactions) database. Users can freely access the InCeP through the internet and download the graphical display as well as the curated information.

Key words: InCeP; mKIAA; protein–protein interactions; intracellular pathway

1. Introduction

Accumulation of a large amount of biological evidence has led to a need for systematization according to the cellular function of each molecule. Recent progress in computational science provides the scope for a possible solution to this issue, with database archiving of cellular pathways already having begun. For example, the Biomolecular Interaction Network Database (BIND) is a quickly growing database designed to archive molecular interactions.1–3 The BIND has already recorded >170,000 interactions and the resultant cellular pathways. In addition, the Kyoto Encyclopedia of Genes and Genomes (KEGG) is a widely used bioinformatics resource for understanding higher-order cell function.4–6 The KEGG includes genomic and chemical information as well as pathway information. Furthermore, 166 pathway resources are now available online (Pathway Resource List, a catalog of pathway data resources, http://www.cbio.mskcc.org/prl/index.php). Although these databases have been based primarily on published experimental research, de novo molecular interactions are required not only to expand the pathways, but also to curate the details.

With this in mind, we have attempted to develop a pathway resource based on an accumulation of information regarding de novo molecular interactions. We have especially focused on the KIAA genes that were identified in our human cDNA sequencing project and that were functionally unknown at the time they were sequenced.7,8 Specific molecules that capture proteins such as antibodies have become strong tools in the identification of molecular interactions. We have therefore begun to generate ‘libraries’ of antibodies against mouse counterparts of human KIAA proteins in order to overcome the legal and ethical restrictions on the use of human materials.9,10 Using the ‘libraries’ of antibodies, we have begun to identify the endogenous mKIAA protein–protein interactions. Identified interactors are subsequently checked for their function and used to build the intracellular pathway...
involving the interaction. The inaugural version of the InCeP reveals only 18 intracellular pathways of mKIAA proteins, but further progress in our project promises to rapidly elucidate the function of many mKIAA proteins.

2. Materials and Methods

2.1. Identification of a novel protein–protein interaction

Approximately 0.4 g of each adult mouse tissue (ICR strain, 8 weeks) or cell lines derived from mouse tumor was homogenized with CelLytic M (Sigma, St Louis, MO) containing 0.5% Protease Inhibitor Cocktail (Sigma) and then subjected to immunoprecipitation with anti-mKIAA antibody. The resulting precipitates were recovered in 20 µl of 2x SDS sample buffer containing of 0.2 M DTT by boiling for 10 min. The supernatant was resolved by 8% 1D-SDS–PAGE and then subjected to imidazole-zinc reverse staining. All excised bands were digested with 10 µg/ml trypsin (Promega, Madison, WI) for 16 h. After dilution with 1% TFA, the resulting peptide mixture was subjected to a high-pressure liquid chromatography (HPLC) separation on a MAGIC 2002 (Michrom BioResources, Auburn, CA). The peptides were first loaded onto polymeric reverse-phase packing material (Peptide Captrap, Michrom BioResources) for desalting and concentration, and then separated onto a reverse-phase capillary HPLC column (C18, 200 A, 0.2 mm × 50 mm, Michrom BioResources) with a flow rate of 5 µl/min. As solvents, 2% v/v acetonitrile in 0.1% v/v formic acid (solvent A) and 90% v/v acetonitrile in 0.1% v/v formic acid (solvent B) were used with a linear gradient from 5 to 65% of solvent B over 30 min. The chromatography system was coupled via a HTS-PAL (CTC Analytics, Zwingen, Switzerland) to an ion trap mass spectrometer LCQ (Thermo Finnigan, CA). The resulting MS/MS data were analyzed using the Mascot search engine (Matrix Science, London, UK). Proteins identified with a combined peptide score of >80 were considered significant, and lower-scoring proteins were rejected.

2.2. Pathway analysis

Ingenuity Pathway Analysis software (Ingenuity Systems, Mountain View, CA, USA) is the world’s largest curated database consisting of millions of individually modeled relationships between proteins, genes, complexes, cells, tissues, drugs and diseases. If researcher inputs a set of proteins into the Ingenuity Pathway Analysis, the software presents relevant pathways and diseases. Therefore, we first imported accession numbers of mKIAA and their interacting proteins identified as described in Section 2.1 into the Ingenuity Pathway Analysis. The software then computed a score for each pathway according to the fit of the imported set of proteins. The score is derived from a P-value and indicates the likelihood of the focus proteins in a pathway being found together owing to random chance. A score of 2 indicates that there is a 1 in 100 chance that the focus proteins are together in same pathway owing to random chance. Therefore, scores of 2 or higher have at least a 99% confidence of not being generated by random chance alone. Biological functions were then calculated and assigned to each pathway. Subsequently, accession number of each component for the most relevant cellular pathway statistically selected was exported to another pathway analysis tool, PathwayAssist (Ariadne Genomics, Inc., Rockville, MD). The information of each node and edge was manually curated and updated through computerized searches in PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi). The curation of each record was performed by an MD or PhD-level scientist. Essential links to PubMed citations and our original data were performed on PathwayAssist, because the software is superior to Ingenuity Pathway Analysis in manual reconstruction of the graphical views.

2.3. Database construction

The InCeP database is written in Java. The Java Servlet and the JavaServer Pages (JSP) are implemented by the Apache Jakarta Tomcat Servlet/JSP container and communicate with the Oracle Database 10g (Relational Database Server) and the JDBC driver (the interface programs for the database). The recommended web client is Internet Explore 6.0 browser or higher.

3. Results and Discussion

3.1. Identification of a novel mKIAA protein–protein interaction

To identify novel mKIAA protein–protein interactions, we performed MS/MS analysis following immunoprecipitation with anti-mKIAA antibodies. Expecting efficient identification, we selected highly expressed tissues or cell lines based on the information obtained from western blotting. Some of these data are freely available through our InGaP database (Integrative Gene and Protein expression database; http://www.kazusa.or.jp/ingap2) in which information about mKIAA gene/proteins consisting of cDNA microarray analysis, subcellular localization of the ectopically expressed gene and experimental data generated from use of the anti-mKIAA antibody such as western blotting and immunohistochemical analyses are recorded. Functional annotation of mKIAA-interactors provides information important in conceiving a functional hypothesis of mKIAA proteins based on the molecular network. Although the identified molecules may interact with corresponding mKIAA in an indirect manner, it is
conceivable that at least these molecules function along the same cellular pathway. One example is the mKIAA1027 protein, also known as talin 1. Already distinct functional annotation had been obtained from numerous experimental studies. This protein plays a significant role in the cell–cell and cell–extracellular matrix adhesions through interactions with the intracellular domain of integrin $\beta$.\textsuperscript{14–16} Although these previous efforts have clarified the functional importance of mKIAA1027/talin 1 in the transduction of integrin signal, we have identified a novel interaction of mKIAA1027/talin 1 supposing a completely different functional aspect. Promyelocytic leukemia (Gene Symbol in Fig. 1 is PML) was found in immunoprecipitant with anti-mKIAA1027 antibody (Fig. 2). Pml was originally identified as fusion protein with retinoic acid receptor $\alpha$, which causes acute promyelocytic leukemia.\textsuperscript{17} Although the tumor-suppressive role of PML has been attributed to its ability to regulate the transcriptional function of nuclear tumor suppressors such as p53 and Rb,\textsuperscript{18} several cytoplasmic PML (cPML) isoforms of unknown function have also been described.\textsuperscript{19,20} Most recently, Lin \textit{et al.}\textsuperscript{21} have identified a novel function of cPML as an essential modulator of TGF-$\beta$ signaling. Likewise, our finding suggests that cPML might modulate integrin $\beta$-mKIAA1027/talin 1 signaling.

Figure 1. Schematic representation of the pathway including mKIAA1027. The proteins are essentially represented by ellipses. A modification of the shape refers to a protein’s functions (e.g. an eclipsed shape represents protein kinase). The proteins are also displayed with several different colors. Dark blue indicates the targeted mKIAA protein. Light blue indicates identified mKIAA-interactors. Red indicates other components of the cellular pathway selected by Ingenuity Pathway Analysis. The interactions and regulations among the proteins are illustrated by the different lines of connection. The detail explanation of these differences is recorded in Search Results of InCeP (Intracellular pathway based on mKIAA protein–protein interactions) database (http://www.kazusa.or.jp/create/). These data will be freely available through our InCeP database. Each component of the pathway is represented by the gene symbol (e.g. Promyelocytic leukemia, PML).
Non-specifically binding proteins to sepharose column and antibodies reduce the reliability of identified interactors. Considering this point, we have also determined all non-specifically binding proteins in each tissue or cell line by MS/MS analysis (for instance, Ogdh protein and transitional endoplasmic reticulum ATPase are observed in most of the tissues). If we detect these proteins in the excised band which is assumed to be mKIAA-interactor, these proteins are excluded from further construction of corresponding mKIAA-pathway. The reproducibility of these interactions should also be assessed. To strengthen the reliability, we performed at least two independent experiments and confirmed the interactions (mKIAA0035, mKIAA0675, mKIAA0994 and mKIAA1465) at initial stage of our study.

### 3.2. Data representation and usage of the InCeP database

Figure 2. Identification of endogenous mKIAA protein and its interactor. (A) Approximately 60 mg of the kidney lysate was subjected to immunoprecipitation with antibody against mKIAA1027, and the resulting precipitates were resolved in 8% SDS–PAGE. The arrowheads indicate the positions of the bands that were excised for digestion by trypsin (gray and black indicate endogenous mKIAA protein and its interactor, respectively). All picked bands were identified by subsequent LC-MS/MS analysis and the resulting MS/MS dataset was analyzed using the Mascot search engine. IgG indicates the position of the rabbit immunoglobulin heavy chain. (B) The sequences of each protein are depicted in the single-letter code. The sequence stretches that are covered by peptide ion signals are shown in bold. Percent coverages of mKIAA1027 and PML are 5% and 7%, and the Mowse scores are 308 and 268, respectively.

The inaugural version of the InCeP database provides users 18 intracellular pathways, including endogenous mKIAA protein interactions, in HTML format (Table 1). This information has been manually curated from selected publications. The InCeP database is accessible through the CREATE portal (http://www.kazusa.or.jp/create/) (Fig. 3A). Clicking on InCeP, the InCeP home page will open. Similar to this operation, users
Table 1. List for the 18 mKIAA pathways publicly released on the inaugural version of InCeP database.

| mKIAA No. | LocusLink ID | Alias name | Tissue or cell Lines | Interactor | LocusLinkID | Pathway assigned by ingenuity pathways analysis | Score | Related diseases |
|-----------|--------------|------------|----------------------|------------|------------|-----------------------------------------------|-------|-----------------|
| 1         | mKIAA0035    | 70769      | Nucleolar and coiled-body | P388D1     | mKIAA1434  | Innate immune response processing of RNA       | 7     | Lupus erythematosus |
|           |              | 70769      | Phosphoprotein 1       |            |            |                                               |       | Neurodegenerative disease |
| 2         | mKIAA0202    | 20362      | Septin 8               | B16 melanoma | Septin 9   | Cytokinesis                                    | 3     | Acute myeloid leukemia |
|           |              | 20362      | Septin 9               |            |            |                                               |       |                 |
| 3         | mKIAA0336    | 70297      | GRIP and coiled-coil domain-containing 2 adult brain | Valosin-containing protein | 209523 | Fusion of Golgi membranes                       | 7     | Frontotemporal dementia |
|           |              | 70297      |                     |            |            |                                               |       | Familial amyotrophic lateral sclerosis |
|           |              | 70297      |                     |            |            |                                               |       | Carcinoma |
|           |              | 70297      |                     |            |            |                                               |       | IBMPFD |
|           |              | 70297      |                     |            |            |                                               |       | Hunting |
|           |              | 70297      |                     |            |            |                                               |       | Frontotemporal |
|           |              | 70297      |                     |            |            |                                               |       | dementia |
| 4         | mKIAA0531    | 16574      | Kinesin family member 5C | adult brain | Amphiphysin | Endocytosis of synaptic vesicles               | 10    | Encephalomyelitis |
|           |              | 16574      |                     |            |            |                                               |       |                 |
| 5         | mKIAA0675    | 9666       | Zinc finger DAZ interacting protein 3 Neuro 2a | Heterogenous nuclear ribonucleoprotein U Nucleolin | 51810 | Transcription | 11 | Male infertility |
|           |              | 9666       |                     |            |            | Survival of germ cells | Sertoli cell-only syndrome | |
|           |              | 9666       |                     |            |            |                                               |       |                 |
| 6         | mKIAA0769    | 207278     | Hypothetical adult testis | DEAH (Asp-Glu-Ala-His) box polypeptide 9 Gem (nuclear organelle) associated protein 4 | 13211 | Processing of RNA | 7 | Neurodegenerative disease |
|           |              | 207278     |                     |            |            |                                               |       | Alzheimer’s disease |
|           |              | 207278     |                     |            |            |                                               |       |                 |
| mKIAA No. | LocusLink ID | Alias name | Tissue or^a cell Lines | Interactor | LocusLinkID | Pathway assigned by ingenuity pathways analysis | Score^b | Related diseases^c |
|-----------|--------------|------------|------------------------|------------|------------|-----------------------------------------------|--------|---------------------|
| 7 mKIAA0988 108903 | Tubulin-specific chaperone d | adult heart | Dynactin 1 | 216766 | Vesicle transport | 7 | Spinal muscular atrophy |
| 8 mKIAA0994 18600 | Peptidyl arginine deiminase, type II | adult muscle | Dihydrolipoamide S-acetyltransferase | 235339 | Lipid metabolism | 9 | Cerebral ischemia |
| 9 mKIAA1027 21894 | Talin 1 | adult kidney | Promyelocytic leukemia | 18854 | Development | 16 | Acute promyelocytic leukemia |
| 10 mKIAA1101 108737 | Oxidative-stress responsive 1 | adult heart | Heat shock 70 kDa protein 1B, Heat shock 70 kDa protein 1A | 15511, 193740 | Phosphorylation of protein (oxidative stress) | 43 | Inflammation (oxidative stress) |
| 11 mKIAA1102 77569 | Hypothetical | adult lung | C6.1a protein | 210766 | Repair of DNA Cell cycle progression | 8 | Breast cancer |
| 12 mKIAA1115 243819 | Hypothetical | adult testis | Small glutamine-rich tetra-riptide repeat-containing, α | 64667 | Formation of synaptic vesicles | 7 | Huntington disease |
| No. | Gene ID | Score | Description                                                                 | Disease | Tissue/Cell Line        |
|-----|---------|-------|-----------------------------------------------------------------------------|---------|-------------------------|
| 13  | mKIAA1250 | 77480 | Kinase D-interacting substance of 220 kDa                                  | Cellular development | adult brain              |
|     |         |       | Heat shock 70 kDa protein 2                                                |         |                         |
|     |         |       | Adaptor-related protein complex 2, beta 1 subunit                          | Transport of protein |                         |
|     |         |       | Adaptor-related protein complex 2, alpha 1 subunit                         |         |                         |
| 14  | mKIAA1301 | 329152| HECT, C2 and WW domain containing E3 ubiquitin protein ligase 2            | Repair of DNA Apoptosis | adult testis             |
|     |         |       | Ubiquitin specific protease 25                                             |         |                         |
| 15  | mKIAA1406 | 99152 | Anaphase-promoting complex subunit 2                                       | Mitosis | adult kidney             |
|     |         |       | Anaphase-promoting complex subunit 4                                       |         |                         |
|     |         |       | mKIAA0723 (matrin 3)                                                       |         |                         |
| 16  | mKIAA1465 | 320563| Hypothetical                                                                | Apoptosis | Neuro 2a                |
|     |         |       | mKIAA0871(rap2 interacting protein x)                                      |         |                         |
| 17  | mKIAA1684 | 56013 | SNAP-25-interacting protein                                                 | Exocytosis | adult brain             |
|     |         |       | Glutamate receptor, ionotropic, delta 2                                    |         |                         |
| 18  | mKIAA4025 | 15525 | Heat shock protein 4                                                        | Autophagy | adult heart             |
|     |         |       | Heat shock protein, A                                                       |         |                         |
|     |         |       | Heat shock protein 8                                                        | Cell cycle progression |                     |
|     |         |       | Glutamate oxaloacetate transaminase 2                                      |         |                         |
|     |         |       | Branched chain ketoacid dehydrogenase E1, α polypeptide                |         |                         |

* We selected appropriate tissues or cell lines for MS/MS analysis based on the information of InGaP database.
* The score was derived from a p-value and indicates the likelihood of the focus proteins in a pathway. Scores ≥ 2 have at least a 99% confidence of not being generated by random chance.
* The participation of each mKIAA protein in listed diseases was expected from corresponding pathway.
can easily access the InGaP database. To locate pathways, a user can simply search using the mKIAA number, the accession number, or the clone name. Alternatively, a user can find a pathway related to a particular disease or tissue (Fig. 3B). On the summary page, an overview of each pathway is represented by edges (protein, protein complex, small molecule and phenomena) and nodes (several kinds of interactions) (Fig. 3C). Clicking on each edge or node, a user can directly obtain a variety of supplementary data such as detailed information for the molecules and a PubMed citation for the interaction (Fig. 3D). mKIAA protein interactions developed from ‘in-house’ assays are also displayed by simple clicking of the corresponding node. At this time, the node links to the mKIAA knowledge page, which contains all information about the mKIAA gene and protein on InGaP database.

4. Concluding Remarks

We have previously placed a strong emphasis on integration of several kinds of biological knowledge.13 With respect to this issue, we have launched the InGaP
database, a comprehensive database of gene/protein expression profiles using cDNA microarray and anti-mKIAA antibodies. The InCeP is defined as a sister database of the InGaP; concomitant use of these databases may therefore accelerate our appreciation of the function of mKIAA/KIAA proteins. Furthermore, a bidirectional accumulation of information is the expected next step of InCeP development. We are currently developing a graphical analysis tool using the SVG format that will enable us to supplement user’s data on the internet through our resources. The user’s data as well as the records of our novel interactions would facilitate consolidation of the intracellular pathway involving mKIAA/KIAA proteins. We expect that our strategy will be an ‘InCePitive’ approach to discovering molecular mechanisms related to other hypothetical proteins.

Acknowledgements: We are grateful to all the staff who have worked on our project. This study was supported by the CREATE Program (Collaboration of Regional Entities for the Advancement of Technological Excellence) from the JST (Japan Science and Technology Corporation).

References

1. Alfarano, C., Andrade, C. E., Anthony, K., et al. 2005, The Biomolecular interaction network database and related tools 2005 update, Nucleic Acids Res., 33, D418–D424.

2. Bader, G. D. and Hogue, C. W. 2000, BIND—a data specification for storing and describing biomolecular interactions, molecular complexes and pathways, Bioinformatics, 16, 465–477.

3. Bader, G. D., Betel, D., and Hogue, C. W. 2003, BIND: the biomolecular interaction network database, Nucleic Acids Res., 31, 248–250.

4. Kanehisa, M. and Goto, S. 2000, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res., 28, 27–30.

5. Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. 2002, The KEGG databases at GenomeNet, Nucleic Acids Res., 30, 42–46.

6. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. 2004, The KEGG resource for deciphering the genome, Nucleic Acids Res., 32, D277–D280.

7. Nakajima, D., Okazaki, N., Yamakawa, H., Kikuno, R., Ohara, O., and Nagase, T. 2002, Construction of expression-ready cDNA clones for KIAA genes: manual curation of 330 KIAA cDNA clones, DNA Res., 9, 99–106.

8. Nomura, N., Miyajima, N., Sazuka, T., et al. 1994, Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1, DNA Res., 1, 27–35.

9. Hara, Y., Shimada, K., Kohga, H., Ohara, O., and Koga, H. 2003, High-throughput production of recombinant antigens for mouse KIAA proteins in Escherichia coli: computational allocation of possible antigenic regions, and construction of expression plasmids of glutathione-S-transferase-fused antigens by an in vitro recombination-assisted method, DNA Res., 10, 129–136.

10. Koga, H., Shimada, K., Hara, Y., et al. 2004, A comprehensive approach for establishment of the platform to analyze functions of KIAA proteins: generation and evaluation of anti-mKIAA antibodies, Proteomics, 4, 1412–1416.

11. Fernandez-Patron, C., Madrazo, J., Hardy, E., Mendez, E., Frank, R., and Castellanos-Serra, L. 1995, Single-step electrotransfer of reverse-stained proteins from sodium dodecyl sulfate-polyacrylamide gel onto reversed-phase minicartridge and subsequent desalting and elution with a conventional high-performance liquid chromatography gradient system for analysis, Electrophoresis, 16, 911–920.

12. Fernandez-Patron, C., Hardy, E., Sosa, A., Secoane, J., and Castellanos, L. 1995, Double staining of coomassie blue-stained polyacrylamide gels by imidazolium-sodium dodecyl sulfate-zinc reverse staining: sensitive detection of coomassie blue-undetected proteins, Anal. Biochem., 224, 263–269.

13. Koga, H., Yuasa, S., Nagase, T., et al. 2004, A comprehensive approach for establishment of the platform to analyze functions of KIAA proteins II: public release of inaugural version of InGaP database containing gene/protein expression profiles for 127 mouse KIAA genes/proteins, DNA Res., 11, 293–304.

14. Campbell, I. D. and Ginsberg, M. H. 2004, The talin-tail interaction places integrin activation on FERM ground, Trends Biochem. Sci., 29, 429–435.

15. Calderwood, D. A. 2004, Integrin activation, J. Cell Sci., 117, 657–666.

16. Naylor, A., Webb, D. J., and Horwitz, A. F. 2004, Talin: an emerging focal point of adhesion dynamics, Curr. Opin. Cell Biol., 16, 94–98.

17. Goddard, A. D., Yuan, J. Q., Fairbairn, L., et al. 1995, Cloning of the murine homolog of the leukemia-associated PML gene, Mamm. Genome, 6, 732–737.

18. Salomoni, P. and Pandolfi, P. P. 2002, The role of PML in tumor suppression, Cell, 108, 165–170.

19. Fagioli, M., Alcalay, M., Pandolfi, P. P., et al. 1992, Alternative splicing of PML transcripts predicts coexpression of several carboxy-terminally different protein isoforms, Oncogene, 7, 1083–1091.

20. Jensen, K., Shiel, C., and Freemont, P. S. 2001, PML protein isoforms and the RBCC/TRIM motif, Oncogene, 20, 7223–7233.

21. Lin, H. K., Bergmann, S., and Pandolfi, P. P. 2004, Cytoplasmic PML function in TGF-beta signalling, Nature, 431, 205–211.