In this review, the principle of the next generation sequencers, and their major research areas have been described. As shown above, the current applications are centered on continuation of works already started before appearance of the second-generation sequencers, and mainly restricted to experts in genomics. However, one of the most important aspects of this technical revolution should be the easy access to the large sequence data by scientists in other areas and doctors. For the wide spread use, sequence data will soon be available from outsourcing companies, but data analysis will still remain a difficult task. Development of software systems easily accessible to non-experts is essential for utilization of large sequence data.

Considering the steady increase of sequence capacity and decrease of cost, application to diagnosis will be realized in the near future. Routine neonatal diagnosis will be replaced with the routine sequence of the entire genome or the exome. This new type of diagnosis would reveal affected alleles of all known genetic diseases, including genes currently screened in postnatal diagnosis. From the data of a couple, risks of genetic diseases in their children can be accurately predicted.

Application to diagnostics for genetic diseases is easily predictable, but how to apply the next generation sequencers to medical science is rather difficult to predict. Large-scale data production such as the human genome project, the "1,000 genomes" project and "The Cancer Genome Atlas", are productive as far as construction of resources. However, utilization of large data sets to solve a specific problem is usually difficult as exemplified by GWAS. The obstacles are mainly statistical problems inherent to large data sets. One is multiplicity in statistical tests. In general, there is no method of choice to control multiplicity, and the method is chosen through practical applications. For example, Bonferroni correction is used with GWAS, and the q-value or FDR is used for gene expression analysis. Although the validity of each method has been confirmed with repeated use, it should be noted that some true positives must be excluded. In particular, GWAS detected loci representing only a fraction of the genetic background of common diseases. One possibility is that true positive loci may be excluded by stringent criteria set by Bonferroni correction.

Another problem is the "curse of dimensionality". The "curse of dimensionality" is the problem caused by the exponential increase in volume associated with adding extra dimensions to a space. This problem results in an increased number of samples needed for analysis. In the cancer classification problem, where cancer samples are classified into two classes by means of gene expression profiling, the "curse of dimensionality" is avoided by dimension reduction, i.e., reduction of the number of genes by gene selection. For example, when the initial data set contains 10,000 genes, the problem is to classify cancers in a 10,000 dimensional space. By selection of differentially expressed genes, classification is usually performed in the reduced dimensional space, requiring adequate number of samples. On the contrary, it is impractical to reduce the number of SNP markers in GWAS. Thus, the aim of GWAS is limited to discovery of individual loci associated with the disease. GWAS cannot identify association of a disease with a combination of more than two genes. This is due to the "curse of dimensionality", but an alternative explanation is as follows. When the number of tag SNP markers is 50,000, the number of the combination of two genes would be 1,249,975,000. It is impractical to perform this huge number of statistical tests, because of requirement of far larger cohorts and very low threshold p-value. The next-generation sequencers do not solve these statistical problems. Complex diseases are most likely to be mediated by numerous loci (both coding and non-coding) that interact with many environmental factors. Some argue that the whole genome sequencing would be useful for identification of such loci, but the real obstacle would be the statistical problem, which suggests requirement of a huge number of samples. This would also be the case in cancer genome projects.

Probably the most reasonable application would be identification of genes responsible for familial disorders. Familial disorders with small pedigrees, which cannot be subjected to linkage analysis, would be good targets. The above example of familial pancreatic cancer is a good example.

Address correspondence to: Kikuya Kato, MD, PhD, Research Institute, Osaka Medical Center for Cancer and Cardiovascular Diseases, 1-3-3 Nakamichi, Higashinari-ku, Osaka, 537-8511, Japan. E-mail address: katou-ki@mc.pref.osaka.jp

References

[1] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF and Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005; 437: 376-380.

[2] Bentley DR. Whole-genome re-sequencing. Curr Opin Genet Dev 2006; 16: 545-552.

[3] Pandey V, Nutter RC and Prediger E: Applied Biosystems SOLiD™ System: Ligation-Based Sequencing. Next Generation Genome Sequencing: Towards Personalized Medicine. Edited by Janitz M. Wiley, San Francisco, 2008, pp.29-41.

[4] Nutter R: New Frontiers in Plant Functional Genomics Using Next Generation Sequencing Technologies. The Handbook of Plant Functional Genomics, Concepts and Protocols. Edited by Kahl G and Meksem K. Wiley, San Francisco, 2008, pp.431-444.

[5] Adessi C, Matton G, Ayala G, Turcatti G, Mermod JJ, Mayer P and Kawashima E. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. Nucleic Acids Res 2000; 28: E87.

[6] Ronaghi M, Karamohamed S, Pettersson B, Uhlen M and Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem 1996; 242: 84-89.

[7] Ruparel H, Bi L, Li Z, Bai X, Kim DH, Turro NJ and Ju J. Design and synthesis of a 3'-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. Proc Natl Acad Sci U S A 2005; 102: 5932-5937.

[8] Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J and Corcoran K. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 2000; 18: 630-634.

[9] Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG and Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. Science 2003; 299: 682-686.

[10] Korlach J, Marks PJ, Cicero RL, Gray JJ, Murphy DL, Roitman DB, Pham TT, Otto GA, Foquet M and Turner SW. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. Proc Natl Acad Sci U S A 2008; 105: 1176-1181.

[11] Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J and Turner S. Real-time DNA sequencing from single polymerase molecules. Science 2009; 323: 133-138.

[12] A haplotype map of the human genome. Nature 2005; 437: 1299-1320.

[13] Goldstein DB. Common genetic variation and human traits. N Engl J Med 2009; 360: 1696-1698.

[14] Hirschhorn JN. Genomewide association studies--illuminating biologic pathways. N Engl J Med 2009; 360: 1699-1701.

[15] Kraft P and Hunter DJ. Genetic risk prediction--are we there yet? N Engl J Med 2009; 360: 1701-1703.

[16] Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL and Venter JC. The diploid genome sequence of an individual human. PLoS Biol 2007; 5: e254.

[17] Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA and Rothberg JM. The complete genome of an individual by massively parallel DNA sequencing. Nature 2008; 452: 872-876.

[18] Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM,

Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara ECM, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R and Smith AJ. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008; 456: 53-59.

[19] Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Zhang J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, Ma L, Li G, Yang Z, Zhang G, Yang B, Yu C, Liang F, Li W, Li S, Li D, Ni P, Ruan J, Li Q, Zhu H, Liu D, Lu Z, Li N, Guo G, Zhang J, Ye J, Fang L, Hao Q, Chen Q, Liang Y, Su Y, San A, Ping C, Yang S, Chen F, Li L, Zhou K, Zheng H, Ren Y, Yang L, Gao Y, Yang G, Li Z, Feng X, Kristiansen K, Wong GK, Nielsen R, Durbin R, Bolund L, Zhang X, Li S, Yang H and Wang J. The diploid genome sequence of an Asian individual. Nature 2008; 456: 60-65.

[20] Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S and Frazer KA. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol 2009; 10: R32.

[21] Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ and McCombie WR. Genome-wide in situ exon capture for selective resequencing. Nat Genet 2007; 39: 1522-1527.

[22] Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, Lin JC, Palmisano E, Brune K, Jaffee EM, Iacobuzio-Donahue CA, Maitra A, Parmigiani G, Kern SE, Velculescu VE, Kinzler KW, Vogelstein B, Eshleman JR, Goggins M and Klein AP. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science 2009; 324: 217.

[23] Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, Menzies A, Mironenko T, Perry J, Raine K, Richardson D, Shepherd R, Small A, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Cahill DP, Louis DN, Goldstraw P, Nicholson AG, Brasseur F, Looijenga L, Weber BL, Chiew YE, DeFazio A, Greaves MF, Green AR, Campbell P, Birney E, Easton DF, Chenevix-Trench G, Tan MH, Khoo SK, Teh BT, Yuen ST, Leung SY, Wooster R, Futreal PA and Stratton MR. Patterns of somatic mutation in human cancer genomes. Nature 2007; 446: 153-158.

[24] Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE and Vogelstein B. The genomic landscapes of human breast and colorectal cancers. Science 2007; 318: 1108-1113.

[25] Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008; 455: 1061-1068.

[26] Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA, Jr., Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE and Kinzler KW. An integrated genomic analysis of human glioblastoma multiforme. Science 2008; 321: 1807-1812.

[27] Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, Gordon D, Chinwalla A, Zhao Y, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson M, Baty J, Ivanovich J, Heath S, Shannon WD, Nagarajan R, Walter MJ, Link DC, Graubert TA, DiPersio JF and Wilson RK. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 2008; 456: 66-72.

[28] Hait WN and Hambley TW. Targeted cancer therapeutics. Cancer Res 2009; 69: 1263-1267; discussion 1267.

[29] Hambley TW and Hait WN. Is anticancer drug development heading in the right direction? Cancer Res 2009; 69: 1259-1262.

[30] Weber G, Shendure J, Tanenbaum DM, Church GM and Meyerson M. Identification of foreign gene sequences by transcript filtering against the human genome. Nat Genet 2002; 30: 141-142.

[31] Xu Y, Stange-Thomann N, Weber G, Bo R, Dodge S, David RG, Foley K, Beheshti J, Harris NL, Birren B, Lander ES and Meyerson M. Pathogen discovery from human tissue by sequence-based computational subtraction. Genomics 2003; 81: 329-335.

[32] Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan PL, Hui J, Marshall J, Simons JF, Egholm M, Paddock CD, Shieh WJ, Goldsmith CS, Zaki SR, Catton M and Lipkin WI. A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med 2008; 358: 991-998.

[33] Feng H, Shuda M, Chang Y and Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 2008; 319: 1096-1100.

[34] Okubo K, Hori N, Matoba R, Niiyama T, Fukushima A, Kojima Y and Matsubara K. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nat Genet 1992; 2: 173-179.

[35] Velculescu VE, Zhang L, Vogelstein B and Kinzler KW. Serial analysis of gene expression. Science 1995; 270: 484-487.

Breast Cancer (2009) 16:248–251
DOI 10.1007/s12282-009-0141-9

CONFERENCE PAPER

Presidential symposium: From standardization to personalization in breast cancer treatment

Algorithm for in vitro diagnostic multivariate index assay

Kikuya Kato

Received: 16 December 2008 / Accepted: 27 May 2009 / Published online: 8 July 2009

Abstract Recently emerging diagnostic tools such as MammaPrint and oncotype-DX are beginning to have impact on clinical practice of breast cancer. They are based on gene expression profiling, i.e., gene expression analysis of a large number of genes. Their unique characteristic is the use of a score calculated from expression values of a number of genes, for which the Food and Drug Administration (FDA) created a new diagnostic category entitled "in vitro diagnostic multivariate index assay (IVDMIA)." In contrast to conventional biomarkers, IVDMIA requires an algorithm to calculate the diagnostic score. The linear classifier is the preferred algorithm. When the number of diagnostic genes is n, each tumor is represented by a point in an n-dimensional space made from gene expression values. Diagnostic algorithms (linear classifier) make an $(n-1)$-dimensional plane in the n-dimensional space to separate two patient groups. Calculation of the diagnostic score is achieved by dimension reduction. Currently, IVDMIA is restricted to gene expression profiling, and will also be applied to malignancies other than breast cancer.

Keywords Gene expression profiling · Prognostic factor · Linear classifier

This article is based on a presentation delivered at Presidential Symposium, "From standardization to personalization in breast cancer treatment," held on 26 September 2008 at the 16th Annual Meeting of the Japanese Breast Cancer Society in Osaka.

K. Kato (✉)
Research Institute, Osaka Medical Center for Cancer and Cardiovascular Diseases, 1-3-3 Nakamichi, Higashinari-ku, Osaka 537-8511, Japan
e-mail: katou-ki@mc.pref.osaka.jp

Introduction

Recently emerging diagnostic tools such as MammaPrint [1, 2] and oncotype-DX [3] are beginning to have impact on clinical practice of breast cancer. Both should be considered in the context of personalized medicine for determination of chemotherapy through genetic information. They are based on gene expression profiling, i.e., gene expression analysis of a large number of genes. The unique characteristic of these tools is the use of a score calculated from expression values of a number of genes. This is a new feature that no previous diagnostic procedure has, for which the FDA created a new diagnostic category entitled "in vitro diagnostic multivariate index assay (IVDMIA)." Currently, MammaPrint and Pathwork of Origin Test (a microarray-based diagnostic system to determine tissue origin of cancer whose tissue origin is unknown) have been cleared by the FDA.

So far, when using a biomarker, for example, for estrogen receptor or blood cholesterol, the concentration or amount of the molecule is used as a score for diagnosis. However, in IVDMIA, the score is calculated from a number of measurement values, which are gene expression values in the cases of MammaPrint and oncotype-DX. Thus, in IVDMIA, the algorithm, i.e., the method of calculating the score, is critical. However, such an algorithm is a "black box" for clinicians. In this short review, I present a simplified explanation of the algorithm for IVDMIA.

Overview of the IVDMIA diagnostic system

There are two types of statistical analysis for gene expression profiling: unsupervised analysis and supervised

prediction [4, 5]. For diagnostic purposes, supervised prediction has usually been adopted. In supervised analysis, parameters of a diagnostic algorithm are determined with a learning set, and the performance of the algorithm is evaluated with a test set. With a small sample set, cross-validation procedures are usually applied. Leave-one-out (LOO) cross-validation is the most frequently used. In LOO, one sample is withdrawn, and the diagnostic classifier is built with the rest of the samples. The performance of the classifier is evaluated on the withdrawn sample. Repeating with all the samples, the overall performance of the classifier is determined. A schematic representation of LOO is shown in Fig. 1.

Although there have been many studies on supervised prediction, there have been very few comparing algorithms. The main obstacle is multiplicity of statistical test. Statistical significance should be adjusted when the test is repeated. For example, a prize could be obtained easily by increasing the number of lots drawn. The chance estimation of the prize should be adjusted with the number of trials. Similarly, when a number of classifiers are tested, some classifiers yield a good performance by chance. In addition, such studies do not guarantee consistency of results with other data sets. It should also be noted that many studies lacked proper evaluation of classifiers [4]. Thus, it is extremely difficult to determine the real performance of a classifier.

Although only a few comparative studies have been reported, a trend in choice of algorithm has been established. Relatively simple algorithms, categorized as linear classifiers, such as weighted voting [6] and nearest centroid [7], are now preferred. Complex algorithms such as artificial neural network [8] are the minority. One reason is that linear classifiers have sufficient performance. The other reason is that it is difficult to control overfitting with complex algorithms. Overfitting is a phenomenon inherent to supervised prediction: parameters of any algorithm are optimized with the learning data set, and its performance with other data sets is usually not as good as that with the learning set. In general, prevention of overfitting is easier

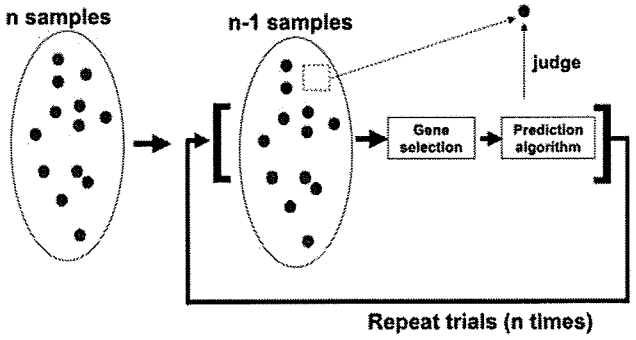

Fig. 1 Schematic representation of leave-one-out validation

with linear classifiers. The algorithm for MammaPrint belongs to the nearest-centroid type [1]. Oncotype-DX also employs a linear classifier, but it has not been described in detail.

Simplified explanation of diagnostic algorithm

When the number of diagnostic genes is n, each tumor is represented by a point in an n-dimensional space made from gene expression values. Diagnostic algorithms (linear classifier) make an $(n-1)$-dimensional plane in the n-dimensional space to separate two patient groups, e.g., high-risk and low-risk groups. As mentioned earlier, the main feature of IVDMIA is the use of a score calculated from expression values of all the diagnostic genes. I present a simplified explanation for calculation of the score using two diagnostic genes ($n = 2$).

When $n = 2$, each case is represented by a point in a two-dimensional plane made with expression of genes 1 and 2. Two coordinates correspond to expression values of genes 1 and 2. As shown in Fig. 2a, red cases (those belonging to the good prognosis group) and black cases (those belonging to the poor prognosis group) make clusters, respectively, and a border line can be drawn. It should be noted that this border line is determined by the learning set and the algorithm used.

With the coordinates in Fig. 2a, each case is represented by expression values of genes 1 and 2, e.g., (x_1, x_2). To convert these two values into a single diagnostic score (DS), the coordinates are rotated by the angle θ so that one axis (the score axis) is perpendicular to the border line (Fig. 2b). In the new coordinates, (x_1, x_2) in the old coordinates is converted to $(x_1\cos\theta - x_2\sin\theta, x_1\sin\theta + x_2\cos\theta)$. Assigning the value of the score axis at the border line is b, two groups are classified as follows.

$a_1x_1 + a_2x_2 \geq b$: good prognosis group
$a_1x_1 + a_2x_2 < b$: poor prognosis group
$a_1 = \sin\theta, a_2 = \cos\theta$

Thus, $a_1x_1 + a_2x_2$ acts as the score of the diagnosis.

The above example is for $n = 2$. For greater n values, DS can be simply extended as

$$\mathrm{DS} = \sum_{i=1}^{n} a_i x_i.$$

In this formula, x_i is the gene expression value for gene i, and a_i is a coefficient determined by the diagnostic algorithm with the learning data set. By defining the threshold, DS can be used to classify patients into two groups.

With oncotype-DX, a diagnostic score, named the recurrent score, is used in Paik et al. [3]. Their recurrent score is a sum of gene expression values multiplied by

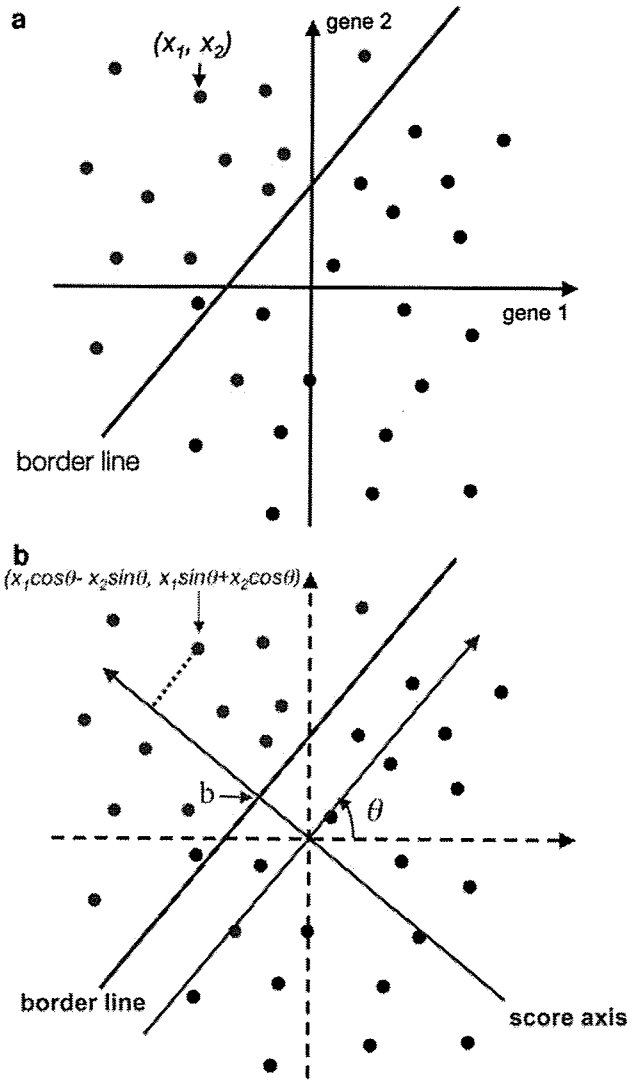

Fig. 2 Schematic representation of classification with a diagnostic algorithm. Two diagnostic genes are used. **a** Each *dot* represents a case (patient) in a two-dimensional space made from expression gene 1 and gene 2. Axes gene 1 and gene 2 indicate the expression of each gene. *Red dot*, good prognosis group; *black dot*, poor prognosis group. The borderline is constructed to separate the two groups optimally with a learning set. **b** New coordinates after axis rotation so that one axis (score axis) is perpendicular to the border line

coefficients, and conforms to the above formula. The algorithm for MammaPrint can be converted to a form of the above formula.

Additional comments

It should be noted that the above explanation is for two genes, and the process to yield a diagnostic score is somewhat different with a larger n, depending on the algorithm. However, for a linear classifier, the classification is done with an $(n-1)$-dimensional plane in the n-dimensional space made with n gene expression values. The difference is within the process of dimension reduction to yield diagnostic score.

From its definition, IVDMIA includes gene expression as well as protein expression. Large-scale identification of a protein can be achieved with mass spectrometry. This type of analysis is called proteome analysis. Proteome analysis has been mainly applied to blood samples for detection of early cancer. In spite of early studies reporting successes, e.g., a study on ovarian cancer [9], this approach is now viewed skeptically. There have been several technical advances, but more time is required to make proteome analysis plausible [10, 11]. Thus, under the current situation, IVDMIA is limited to gene expression.

MammaPrint and oncotype-DX are prognostic predictors, and will be used to provide information for decision on chemotherapy. On the other hand, there is another approach, i.e., direct prediction of effects of a particular anticancer drug. This approach has been taken with a single drug (docetaxel) [12, 13] or combined chemotherapy [14]. However, because the life of a particular regimen is short, it is not possible to recruit enough patients to establish diagnostic systems. Thus, direct prediction of anticancer drug efficacy is not a popular approach anymore.

When it first appeared, DNA microarray was expected to revolutionize medical science. The expectation was exaggerated, and now we know its limitation. However, as demonstrated by MammaPrint, it can be a powerful diagnostic tool. We are also developing IVDMIA for prognosis prediction of glioma [15, 16], which is expected to help clinical decision on temozolomide, a new alkylating agent. IVDMIA will be one of the main tools for personalized medicine.

References

1. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415:530–6.
2. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 2002;347:1999–2009.
3. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004;351:2817–26.
4. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst. 2007;99:147–57.
5. Kato K, Ishii S. Statistical analysis of gene expression profiles. Tanpakushitsu Kakusan Koso. 2003;48:2300–9.
6. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science. 1999;286:531–7.

7. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning; data mining, inference and prediction. New York: Springer; 2001.
8. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med. 2001;7:673–9.
9. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet. 2002;359:572–7.
10. Service RF. Proteomics. Will biomarkers take off at last? Science. 2008;321:1760.
11. Service RF. Proteomics. Proteomics ponders prime time. Science. 2008;321:1758–61.
12. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet. 2003;362:362–9.
13. Iwao-Koizumi K, Matoba R, Ueno N, Kim SJ, Ando A, Miyoshi Y, et al. Prediction of docetaxel response in human breast cancer by gene expression profiling. J Clin Oncol. 2005;23:422–31.
14. Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, Mejia JA, et al. Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. J Clin Oncol. 2006;24:4236–44.
15. Shirahata M, Iwao-Koizumi K, Saito S, Ueno N, Oda M, Hashimoto N, et al. Gene expression-based molecular diagnostic system for malignant gliomas is superior to histological diagnosis. Clin Cancer Res. 2007;13:7341–56.
16. Shirahata M, Oba S, Iwao-Koizumi K, Saito S, Ueno N, Oda M, Hashimoto N, et al. Using gene expression profiling to identify a prognostic molecular spectrum in gliomas. Cancer Sci. 2009; 100:165–72.

BRAIN RESEARCH 1287 (2009) 136–145

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Transcriptional induction and translational inhibition of Arc and Cugbp2 in mice hippocampus after transient global ischemia under normothermic condition

Noboru Otsuka[a,b,*], *Katsuki Tsuritani*[a,b], *Takanobu Sakurai*[b], *Kikuya Kato*[c], *Ryo Matoba*[d], *Jiro Itoh*[e], *Shigeru Okuyama*[b], *Kiyofumi Yamada*[f], *Yukio Yoneda*[a]

[a]*Laboratory of Molecular Pharmacology, Division of Pharmaceutical Sciences, Kanazawa University Graduate School of Natural Science and Technology, Kakuma-machi, Kanazawa, Ishikawa 920-1192, Japan*

[b]*Research Center, Taisho Pharmaceutical Co. Ltd., 403, Yoshino-cho 1-chome, Kita-ku, Saitama 331-9530, Japan*

[c]*Research Institute, Osaka Medical Center for Cancer and Cardiovascular Disease, 1-3-3 Nakamichi, Higashinari-ku, Osaka, 537-8511, Japan*

[d]*DNA Chip Research Inc., 1-1-43 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan*

[e]*Gifu Research Laboratories, JBS Inc., Kaizu-shi, Gifu 503-0628, Japan*

[f]*Department of Neuropsychopharmacology and Hospital Pharmacy, Nagoya University Graduate School of Medicine, 65, Tsurumai-cho, Showa-ku, Nagoya 466-8560, Japan*

ARTICLE INFO

Article history:
Accepted 16 June 2009
Available online 24 June 2009

Keywords:
Transient global ischemia
Hypothermia
ATAC-PCR
Arc
Cugbp2
Translational inhibition

ABSTRACT

Mild hypothermia protects against neuronal damage after transient global ischemia in experimental animals. The exact mechanism of this protective effect remains to be elucidated. The purpose of the present study was to investigate the molecular mechanisms relevant to different neurologic responses to hypothermia and normothermia. Transient global ischemia was induced in C57BL/6 mice by bilateral common carotid artery occlusion for 10 min. Hypothermia provided robust neuroprotection in the hippocampus region and dramatically reduced the mortality rate. Using adaptor-tagged competitive polymerase chain reaction, we obtained the relative transcription levels of 1210 genes in the hippocampal region and compared the expression patterns of these genes. Two genes, Activity-regulated cytoskeleton-associated protein (Arc) and CUG-binding protein-2 (Cugbp2), showed remarkable and persistent increases in their expression levels in normothermic mice, compared with in both sham and hypothermic mice. Despite the increased transcription of Arc and Cugbp2, an immunohistochemistry analysis did not show comparable increases in the translations of both genes. Only a transient increase in Arc protein was observed in the granule cells of the dentate gyrus at 6 h after reperfusion. A remarkable decrease in Cugbp2 protein was observed in the pyramidal cells of the hippocampal CA1–CA3, in accordance with the progress of neuronal degeneration. A decrease in Cugbp2 protein was not observed in

* *Corresponding author.* Prescription Pharmaceuticals, Taisho Pharmaceutical Co. Ltd., 24-1, Takada 3-chome, Toshima-ku, Tokyo 170-8633, Japan. Fax: +81 3 3985 0650.
E-mail address: noboru.otsuka@po.rd.taisho.co.jp (N. Otsuka).

Abbreviations: Arc, activity-regulated cytoskeleton-associated protein; ATAC-PCR, adaptor-tagged competitive polymerase chain reaction; BCCAO, bilateral common carotid artery occlusion; Cugbp2, CUG-binding protein-2; HE, hematoxylin and eosin; IHC, immunohistochemistry; MCAO, middle cerebral artery occlusion; PBS, phosphate buffered saline; PI, propidium iodide; RMSE, root mean square error

0006-8993/$ – see front matter
doi:10.1016/j.brainres.2009.06.050

hypothermic mice. These results suggest that transient global ischemia induces the translational inhibition of genes with increased expression not in hypothermic, but in normothermic mice. Thus, translational inhibition might play an important role in the progress of neuronal injury after transient global ischemia.

1. Introduction

Mild to moderate hypothermia has been shown to decrease neurologic damage and to improve the neurologic outcome in patients experiencing out-of-hospital cardiac arrest (The Hypothermia after Cardiac Arrest Study Group, 2002; Bernard et al., 2002). The use of therapeutic hypothermia is supported and recommended by organizations from many areas of the world (Nolan et al., 2003; Froehler and Geocadin, 2007).

Experimental models show that therapeutic hypothermia protects the brain from cerebral injury, reduces mortality, and improves neurologic outcome (Froehler and Geocadin, 2007). While therapeutic hypothermia has already entered clinical practice, further studies are required to define the detailed mechanisms of the neuroprotective effect. Although gene expression profiling is useful for identifying candidate genes from among large numbers of genes, there is only one report of gene expression profiling in the brains of rats after the ischemic insult under hypothermic or normothermic conditions (Sugahara-Kobayashi et al., 2008).

To screen candidate genes relevant to different neurologic responses to hypothermic and normothermic conditions, we compared the gene expression profiles in mice brains after transient global ischemia. We investigated the following four points prior to performing the gene expression profile analysis. First, we examined which animal model would be most suitable for such an analysis. Two animal models, the bilateral common carotid artery occlusion (BCCAO) model in mice and the middle cerebral artery occlusion (MCAO) model in rats, have been extensively used in the study of neuronal damage after ischemia and reperfusion (Carmichael, 2005). We selected the BCCAO model induced in C57BL/6 mice because of its reproducibility (Yang et al., 1997; Wellons et al., 2000; Yonekura et al., 2004) and because the neuroprotective effect of hypothermia has been confirmed in this model (Yang et al., 1997; Tsuchiya et al., 2002, Olsson et al., 2003). Second, we examined the availability of gene sequence information and its annotation. We previously established an in-house database of 3′-end sequences from cDNA libraries made from several regions of mouse brain, such as the cerebellum, hippocampus and striatum (Matoba et al., 2000). Using the information in this database, we were able to select genes and to construct the gene-specific oligonucleotide primers necessary for performing an adapter-tagged competitive polymerase chain reaction (ATAC-PCR) assay for gene expression. Third, we examined the amount of RNA required for the assay. This point is particularly important for experiments using animal models. ATAC-PCR requires a relatively small amount of RNA, and we confirmed that a sufficient amount of RNA could be obtained from the mouse hippocampus region (Kato, 1997; Ooe et al., 2007). And fourth, we examined the cell population in the targeted tissue. Marked changes in the cell population of a tissue or organ can make comparisons of gene expression profiles difficult. We studied the time course of neurologic injury, identified a time period during which minimal changes occurred in the hippocampus region, and defined a suitable time period for the gene expression profile of analysis.

2. Results

2.1. Control of body temperature in mice

The body temperatures of the mice were monitored during the 10 min of ischemia and 60 min of reperfusion by continuous measurement of the rectal temperature in each mouse. The mean rectal temperature of the surviving mice in the hypothermia group (HT-alive) before ischemia was 24 °C, while the mean rectal temperatures of the sham mice, the surviving mice in the normothermia group (NT-alive), and the mice that died during ischemia in the normothermia group (NT-dead) were each 38 °C (Fig. 1A). Thus, the mean body temperatures of the sham, NT-alive and NT-dead groups were not significantly different. The rectal temperature of the NT-alive group was maintained within 37–39 °C, while that of the HT-alive group was maintained within 20–27 °C, during the 10 min of ischemia and 60 min of reperfusion (Figs. 1B, C).

2.2. Mortality rate and behavioral analysis

Hypothermia enabled a dramatic reduction in the mortality rate during the 10 min of transient global ischemia (Table 1). The mortality rates in the hypothermia and normothermia groups were 6% and 51%, respectively. These rates were significantly different ($P<0.05$, Chi-square test).

The frequencies of neurologic signs and the neurologic scores were subsequently assessed in a series of behavioral studies (Fig. 2). No neurologic signs were observed in the sham mice. Comparable frequencies of the loss of righting reflex and coma were observed in the HT-alive, NT-alive and NT-dead groups. Seizure was not observed in the HT-alive group but was observed in both the NT-alive and the NT-dead groups (Figs. 2A, B). The frequencies of seizure in the NT-alive and NT-dead groups were 20% and 50%, respectively. A significant difference in the frequency of seizure was observed between the HT-alive and the NT-dead groups ($P<0.05$, Kruskal–Wallis test followed by Dunn's test) (Figs. 2A, B). The neurologic scores of the NT-alive and NT-dead groups were higher than that of the HT-alive group. The neurologic scores of the HT-alive and the NT-dead groups were significantly different ($P<0.05$, Kruskal–Wallis test followed by Dunn's test) (Fig. 2C).

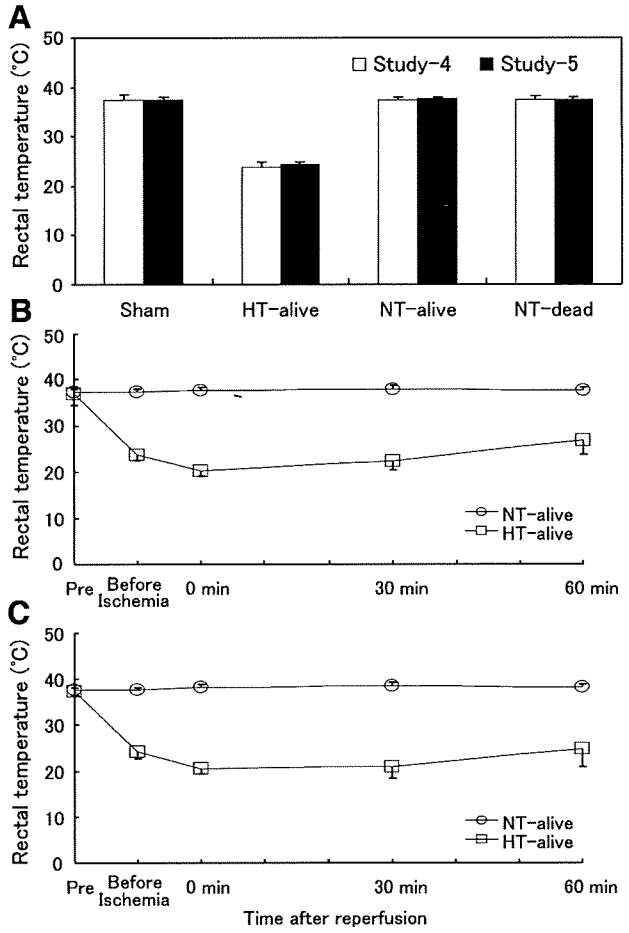

Fig. 1 – Body temperatures before ischemia (A) and over time (B, C). NT-alive means mice that survived global ischemia for 10 min under normothermic conditions. NT-dead means mice that died during the 10 min of ischemic insult under normothermic conditions. In the normothermic mice (NT-alive), the body temperature was maintained within 37–39 °C during the 10 min of ischemia and 60 min of reperfusion. In hypothermic mice (HT-alive), the rectal temperature was maintained within 20–27 °C during the 10 min of ischemia and 60 min of reperfusion. Comparable results were obtained for both studies (B, Study-4; C, Study-5). No significant difference in the rectal temperatures before ischemia was seen among the sham, NT-alive and NT-dead mice. The body temperatures are shown as mean±S.D.; N=20 except for N=27 (NT-dead, Study-4), N=25 (NT-dead, Study-5) and N=19 (sham, Study-5).

2.3. *Histologic injury after transient global ischemia*

Specimens of the hippocampus and striatum obtained from mice after 10 min of global ischemia and 7 days of reperfusion were stained using hematoxylin and eosin (HE). Neuronal damage in the CA1–CA3 region of the hippocampus was observed in the normothermic mice. No histologic injury was observed in the sham (N=10) or hypothermic mice (N=9). The histologic injury score in the normothermic mice was 0.20 ±0.28 (mean±S.D., N=11) (Fig. 3A). Relatively higher histologic injury scores and higher numbers of injury sites were seen in the normothermic mice with seizure than in the mice without seizure (Fig. 3B).

Table 1 – Mortality rates during 10 min of bilateral common carotid artery occlusion in mice.

Group	Number of dead mice/surviving mice (mortality rate: %)				
	Study-1	Study-3	Study-4	Study-5	Total
Sham	0/10 (0)	0/12 (0)	0/20 (0)	0/19 (0)	0/61 (0)
Hypothermia	2/9 (18)	1/12 (8)	1/20 (5)	0/20 (0)	4/61 (6)*
Normothermia	7/11 (39)	9/14 (39)	27/20 (57)	25/20 (55)	68/65 (51)

Mortality rate in hypothermic mice was significantly reduced compared to normothermic mice (*P<0.05, Chi-square test).

The time course and extent of neuronal cell death in the normothermic mice were examined by observing propidium iodide (PI)-stained cells in mice after transient global ischemia. The prevalence of positively PI-stained cells at 18, 24, 30 and 48 h after reperfusion were 25%, 18%, 45% and 75%, respectively. The PI staining intensity at 18 and 24 h was mild compared with that at 30 and 48 h. Both the prevalence and

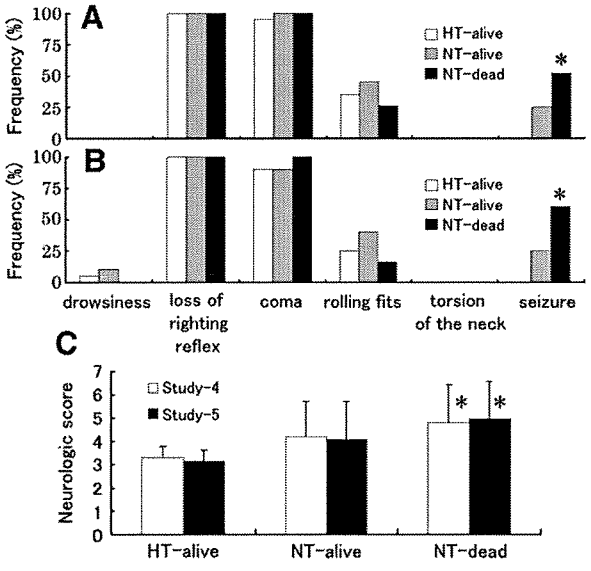

Fig. 2 – Frequency of neurologic signs and neurologic deficit scores under hypothermia and normothermia. C57BL/6 mice were subjected to 10 min of bilateral common carotid artery occlusion under conditions of hypothermia (HT) or normothermia (NT). During the ischemic insult, neurologic signs were observed and the frequencies of the signs (A: Study-4, B: Study-5) and the neurologic deficit scores (C) were obtained. Comparable and reproducible results were obtained from two experiments. Data from the sham groups are not shown because no neurologic signs were observed. No seizures were observed in the HT-alive mice. The frequencies of seizure and the neurologic deficit scores were significantly higher in the NT-dead mice than in the HT-alive mice (*P<0.05, Kruskal–Wallis followed by Dunn's test). The values of the neurologic deficit scores are mean±S.D., N=20 except for N=27 (NT-dead, Study-4), N=25 (NT-dead, Study-5) and N=19 (sham, Study-5).

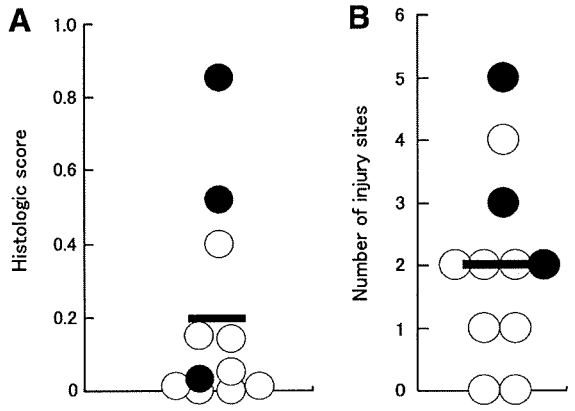

Fig. 3 – Histologic injury scores and numbers of injury sites in normothermic mice. Animals that survived the ischemic insult were sacrificed 7 days after reperfusion (N=11). An incremental 4-point scoring system was used: 0, normal; 1, <10%; 2, 10–49%; and 3, >50% for the C1 to C3 region in the hippocampus. The extent of the CA1 to CA3 sector with each grading was measured and the mean histologic score was obtained for each mouse by dividing the integration of each grading and its length by the total length of the CA1 to CA3 sector (A). The number of sites with neuronal injury in the CA1 to CA3 sector was counted on both sides. Closed circles are the data from mice with seizures, open circles are the data from mice without seizures, and the bars are the mean values. A relatively high histologic score and number of injured sites were observed in mice with seizures. No histologic injuries were observed in the sham (N=10) or HT mice (N=9).

the degree of staining (data not presented) increased with the extension of the reperfusion period.

The Nissl-stained hippocampal region was also assessed at 6 and 24 h after reperfusion. No changes were observed in the hippocampal neuronal regions of the sham and hypothermic mice (Figs. 5A, B, M, N). No change or primary nuclear pyknosis was observed in the granule cells of the dentate gyrus, and minor degenerative changes were observed in the pyramidal cells of the CA1 and CA3 hippocampal regions in normothermic mice at 6 h (Figs. 5C, D). Advanced changes were observed in the pyramidal cells of the CA1–CA3 region and the granule cells of the dentate gyrus in normothermic mice at 24 h after reperfusion (Fig. 5O, P).

2.4. *Gene expression analysis using adaptor-tagged competitive-PCR*

The expression levels of 1210 genes in the hippocampus region at 18 h after reperfusion were assayed using ATAC-PCR. After evaluating the regression line according to a previously described algorithm (Kita-Matsuo et al., 2005), 244 genes with a root mean square error (RMSE) of over 0.6 were discarded and 966 successfully assayed genes were classified into 9 clusters according to their expression patterns (Table 2). We prioritized Cluster-2 (Sham=HT<NT; 78 genes) and Cluster-8 (Sham<HT<NT; 29 genes), and 62 genes were selected for further assay.

The quantitative performances of these 62 genes at 6 h after reperfusion in Study-4 and at 24 h after reperfusion in Study-5 were evaluated using a coefficient of variance (CV) and absent data. Four genes with a CV of over 0.35 and/or at least one absent data were discarded. Remaining 58 genes that had a mean CV of 0.11 were classified into 9 clusters (Table 2). The relative transcription levels of 14 genes that had been classified into Cluster-2 at either 6 or 24 h are shown (Fig. 4). Only two genes, Activity-regulated cytoskeleton-associated protein (Arc) and CUG-binding protein-2 (Cugbp2), were classified into Cluster-2 at both 6 and 24 h after reperfusion.

2.5. *Protein expressions of Arc and Cugbp2 in the hippocampus region*

We examined the protein expressions of Arc and Cugbp2 (N=3 for each group) in the hippocampus regions of normothermic, hypothermic, and sham mice at 6 and 24 h after reperfusion using immunohistochemistry (IHC). Two antibodies were used for both Arc (C-7; sc-17839 and H-300; sc-15325) and Cugbp2 (AP; 12921-AP and PG; LS-C31676). Similar results for Arc were obtained using both C-7 and H-300. Similar results for Cugbp2 were also obtained using AP and PG. As representative results, photographs of Arc using C-7 antibody and of Cugbp2 using PG antibody are shown (Fig. 5). The specificity of the antibodies was confirmed by the disappearance of a positive reaction using normal mouse IgG or normal rabbit IgG instead of the primary antibodies for Arc or Cugbp2.

An increase in Arc protein expression was transiently observed in the nuclei of neuronal granule cells in the dentate gyrus of normothermic mice at 6 h after reperfusion (Figs. 5G, H). Then, a decrease in Arc expression was observed in the dentate gyrus of normothermic mice at 24 h after reperfusion

Table 2 – Gene expression profiles classified into nine clusters.

Cluster	Study-3:18 h	Study-4:6 h	Study-5:24 h
0: S=HT=NT	210 (17)	33 (53)	39 (63)
1: S=HT>NT	100 (8)	3 (5)	6 (10)
2: S=HT<NT	78 (6)	8 (13)	7 (11)
3: S>HT=NT	89 (7)	12 (19)	2 (3)
4: S>HT>NT	32 (3)	0 (0)	1 (2)
5: S>HT<NT	137 (11)	2 (3)	0 (0)
6: S<HT=NT	96 (8)	0 (0)	22 (3)
7: S<HT>NT	195 (6)	0 (0)	1 (2)
8: S<HT<NT	29 (2)	0 (0)	0 (0)
Sub-total	966 (80)	58 (94)	58 (94)
Discarded	244 (20)	4 (6)	4 (6)
Total	1210 (100)	62 (100)	62 (100)

S: sham; HT: hypothermic; NT: normothermic. In Study-3, Study-4, and Study-5, the genes were assayed at 18 h, 6 h and 24 h of reperfusion following 10 min of global ischemia. In Study-3, genes that had a root mean square error (RMSE) of over 0.6 were discarded. For Study-4 and Study-5, 62 genes were selected from among 107 genes that were classified into Cluster-2 or Cluster-8 in Study-3. In Study-4 and Study-5, genes that had a CV of more than 0.35 and/or at least one absent data were discarded. Only two genes were classified into Cluster-2 in both Study-4 and Study-5. A value of log (NT/HT)>0.2 was regarded as NT>HT. ( ):%.

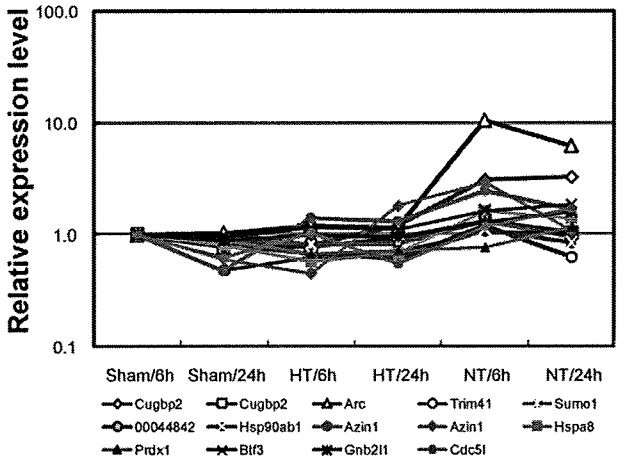

Fig. 4 – Relative transcription levels of 14 genes at 6 and 24 h after reperfusion. Fourteen out of 58 genes were classified into Cluster-2 at either 6 or 24 h after reperfusion. The hippocampal gene expression profiles of the 58 genes were studied after 6 and 24 h of reperfusion following 10 min of global ischemia. HT and NT mean hypothermic and normothermic mice, respectively. The gene expression levels were assayed using ATAC-PCR, and the transcription levels relative to those in the sham control at 6 h after reperfusion are shown. Arc and Cugbp2 were up-regulated in normothermic mice at 6 and 24 h after reperfusion.

(Figs. 5S, T). Slight or no expression of Arc protein was observed in the hypothermic or sham mice (Figs. 5E, F, Q, R).

Cugbp2 was localized and constitutively expressed in the nuclei of granule cells in the dentate gyrus and of pyramidal cells in the hippocampal CA1–CA3 region in sham and hypothermic mice at 6 and 24 h after reperfusion (Figs. 5I, J, U, V). A decrease in Cugbp2 expression was observed in the CA1–CA3 pyramidal cells of the hippocampus in normothermic mice at 6 h after reperfusion (Fig. 5K, L). The apparent decrease coincided with the degeneration of neuronal cells in the CA1–CA3 region of the hippocampus in normothermic mice at 24 h after reperfusion (Figs. 5W, X). No changes in the histology or expression level of Cugbp2 protein were observed in the hypothermic mice (Figs. 5B, N, J, V).

3. Discussion

Before analyzing gene expression, we confirmed the adequacy of our mouse BCCAO model and the effect of hypothermic conditions. Our findings showed that the neurologic signs and neurologic deficit scores as well as the histologic injury scores in BCCAO mice were consistent with those described in a previous report (Yang et al., 1997). HE staining and Nissl staining showed neuronal loss in the hippocampal CA1 and CA3 and no neuronal loss in the dentate gyrus. The neuronal loss in the hippocampal CA1 and CA3 were protected by hypothermia. A reduction in the mortality rate was obtained in the hypothermic mice. These findings were also consistent with previous reports (Yang et al., 1997; Tsuchiya et al., 2002; Olsson et al., 2003). Thus, we concluded that a gene expression analysis using this animal model would be meaningful.

We screened for candidate genes relevant to different neurologic responses to hypothermic and normothermic conditions from among a large number of genes by comparing gene expression profiles. The genomic responses following the ischemic insult are complex and involve increases in the expressions of both protective and detrimental genes. We expected that a few key genes with detrimental roles might exist among genes that were specifically up-regulated under normothermic conditions, compared with hypothermic conditions.

We used the ATAC-PCR method, which is an advanced version of quantitative PCR, because this method has several advantages. Using ATAC-PCR, we can perform a quantitative assay for a large number of genes with a small amount of RNA (Kato, 1997; Matoba et al., 2000). A persistent and specific increase under normothermic conditions was only observed for two out of the 1210 genes that were screened. We found that the expressions of Arc and Cugbp2 were increased in normothermic mice, compared with in hypothermic mice.

Arc, also known as Arg3.1, was initially discovered as a novel member of the immediate-early genes induced by neuronal activity (Lyford et al., 1995; Link et al., 1995). Arc mRNA and/or protein is strongly induced by synaptic activation caused by high-frequency stimulation (Messaoudi et al., 2007; Huang et al., 2007), the administration of kainate (Palop et al., 2005; Li et al., 2005) and behavioral experience (Guzowski et al., 2006; Fletcher et al., 2006). In rat studies, the gene expression of Arc was induced by MCAO in the peri-infarct cortex and the dentate gyrus (Kunizuka et al., 1999; Rickhag et al., 2007). Persistent up-regulation was observed in the dentate gyrus and the hippocampus CA1–CA3 from the beginning until 12 h after reperfusion (Rickhag et al., 2007). Considering those reports, up-regulated gene expression seems to be triggered by the ischemic insult and maintained by reperfusion.

However, the protein expression of Arc in mouse or rat brain after the ischemic insult and reperfusion has not been previously investigated. We investigated the protein expression of Arc using IHC with two antibodies (C-7 and H-300) to confirm antibody performance and reproducibility. One antibody (H-300) was used for IHC (Li et al., 2005; Messaoudi et al., 2007) and the other antibody was used to stain cultured cells (Bloomer et al., 2007). The expression level in the hippocampus region was low in the hypothermic and sham mice at 6 and 24 h after reperfusion. Meanwhile, the transient expression of Arc was observed in normothermic mice at 6 h after reperfusion. Arc protein was localized to the nuclei of the neuronal granule cells of the dentate gyrus. The localization and subcellular localization of Arc expression were consistent with the results of previous reports (Palop et al., 2005; Li et al., 2005; Bloomer et al., 2007). Unexpectedly, we found that the expression of Arc protein was reduced in normothermic mice at 24 h after reperfusion, in spite of the increased expression of the Arc gene. One previous report has suggested that Arc may have an anti-apoptotic role (Irie et al., 2000). Neuronal degeneration was observed not in the dentate gyrus but in the CA1–CA3 of the hippocampus region. Thus, the transient increase in Arc protein might be related to the invulnerability of the granule cells of the dentate gyrus.

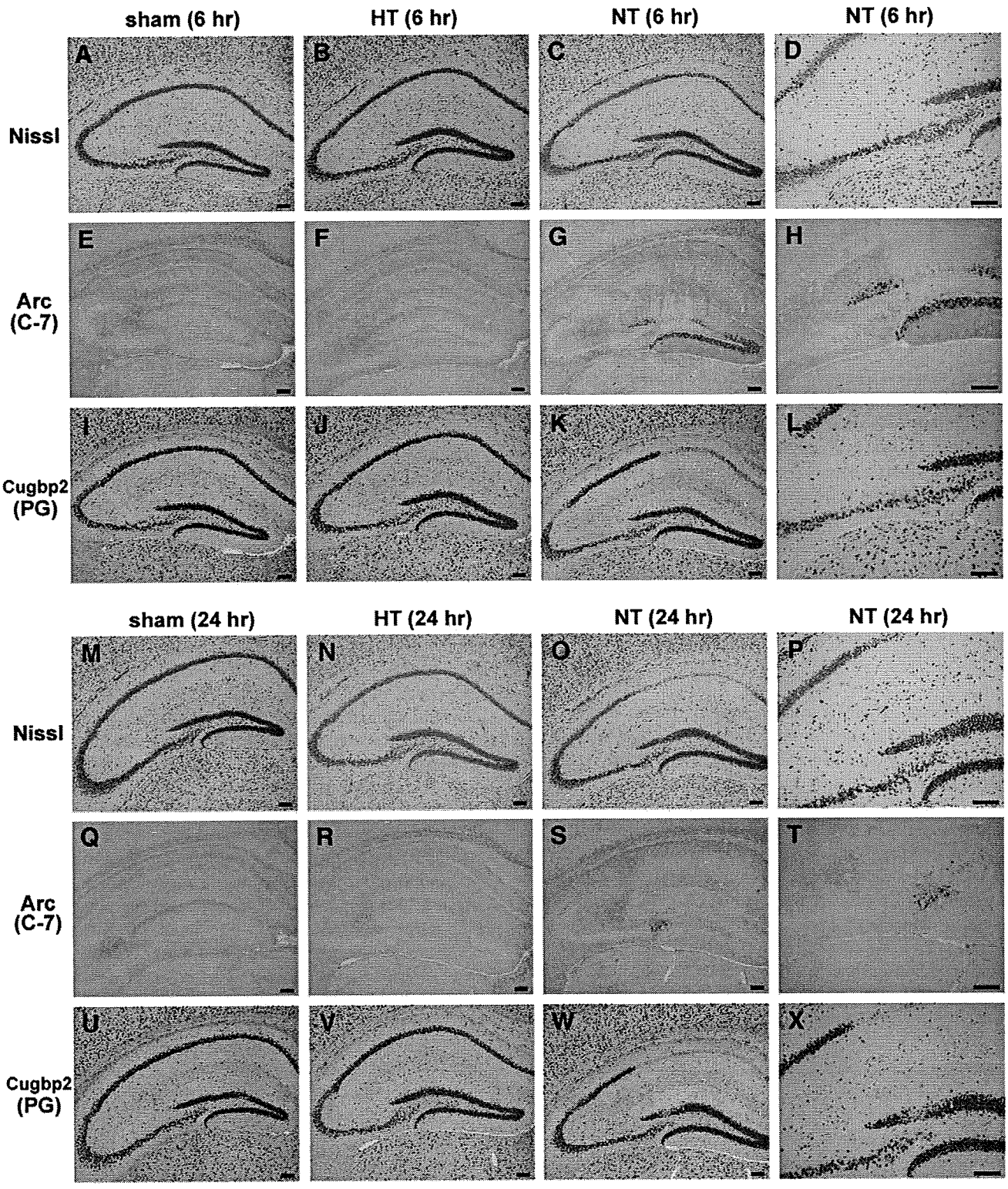

Fig. 5 – Immunohistochemistry findings of Arc and Cugbp2 protein expressions in mice brain after the ischemic insult. The protein expressions of Arc (C-7; sc-17839) and Cugbp2 (PG; LS-C31676) were observed in the brain sections at the level of the hippocampus in C57BL/6N mice at 6 and 24 h of reperfusion following 10 min of global ischemia. A representative photograph is shown for each group (N=3 for each treatment group), and high magnification photographs of the normothermic mice are shown (D, H, L, P, T, X). A reduction in the Nissl-stained sections in the CA1 to CA3 sector of the hippocampus region and dentate gyrus was only observed in the normothermic mice (O, P). Hypothermia enabled dramatic neuroprotection (B and N). Arc was up-regulated in the neuronal granule cells of the dentate gyrus only at 6 h (G and H). Cugbp2 was spontaneously expressed in the neuronal granule cells of the hippocampus and dentate gyrus of sham and hypothermic mice (I, J, U, V). Decreased expression was observed in the hippocampus of normothermic mice in accordance with the progressive degeneration of the neuronal granule cells (K, L, W, X). Scale bars are 100 μm.

CUGBP2 has several synonymous names including ETR-3 (elav-type RNA-binding protein-3), NAPOR (neuroblastoma apoptosis-related RNA-binding protein), CELF2 (CUG-BP- and ETR-3-like factor) and BRUNAL3 (Bruno-like protein 3). Cugbp2 was initially identified as a gene induced during the apoptosis of human neuronal cells and named NAPOR (Choi et al., 1998). Mouse Cugbp2 was identified as a second member of the CUG-binding protein known as elav-type ribonucleoprotein-3 (ETR-3) (Lu et al., 1999). In mouse and rat brains, the localized expression of Cugbp2 mRNA has been observed in the hippocampus (Choi et al., 1999; Zhang et al., 2002). Here, we report for the first time that the expression of the Cugbp2 gene was increased in the hippocampus region after global ischemia and reperfusion. This is also the first report to describe the protein expression of Cugbp2 in mouse brain after transient global ischemia. We investigated the expression of Cugbp2 protein using IHC with two antibodies. In sham and hypothermic mice, Cugbp2 protein was localized to the nuclei of the pyramidal cells in the CA1 to CA3 region of the hippocampus and in the granule cells of the dentate gyrus and was constitutively expressed at 6 and 24 h (Figs. 5I, J, U, V). The expression of Cugbp2 protein was similar in the sham and hypothermic mice. However, the expression of Cugbp2 protein decreased in neuronal cells of the hippocampus at both 6 and 24 h (Figs. 5K, L, W, X), in spite of the increase in gene expression at both of these time points. An apparent decrease in Cugbp2 protein was demonstrated at 24 h after reperfusion, and this degeneration coincided with the neuronal cells in the hippocampus and dentate gyrus. Although Cugbp2 has been identified as an apoptosis-related gene, little is known concerning its function in neuronal cell death. There is a report of overexpressed mRNA and protein of Cugbp2 in an in vitro model of cerebral ischemia (Pacini et al., 2005). It was also reported that the down regulation of Cugbp2 gene expression protected hippocampal cultures against oxygen–glucose depletion induced apoptosis (Pacini et al., 2005). In contrast, Cugbp2 is rapidly induced in mouse intestinal epithelial cells after treatment with ionizing radiation, and the induction of Cugbp2 leads to an apoptotic response (Murmu et al., 2004). Before the IHC assessment, we speculated that Cugbp2 might be involved in apoptotic cell death via an increase in the expressions of its gene and protein.

Protein synthesis is known to be inhibited in neuronal cells immediately after ischemia and reperfusion (DeGracia et al., 2008). A persistent translational inhibition is correlated with the delayed neuronal cell death of vulnerable neurons following ischemia and reperfusion (Hossmann, 1993). The pyramidal neurons of the CA1 in the hippocampus are selectively vulnerable to transient global ischemia. Thus, the decrease in the expression of Cugbp2 protein was consistent with the degeneration of vulnerable pyramidal neurons.

We observed a transient increase in Arc protein and a decrease in Cugbp2 protein expression, in spite of persistent increases in the transcriptions of both genes. These results may indicate that ischemia and reperfusion under normothermic conditions may induce translational inhibition. In hypothermic mice, neither post-ischemic neuronal injury nor the decreased expression of Cugbp2 protein was observed in the hippocampus region. The control of body temperature might be related to the induction of translational inhibition, and the neuroprotective effects of hypothermia might be achieved, in part, by regulating the systems that control protein synthesis in vulnerable cells.

4. Experimental procedures

We screened candidate genes relevant to different neurologic responses to transient global ischemia in the presence of hypothermia or normothermia using a series of six studies. In Study-1, the neuroprotective effect of hypothermia on ischemic brain injury was confirmed. In Study-2, we examined the time course of neuronal damage after reperfusion. Minimal damage to the hippocampal neurons was observed after 18 h of reperfusion. In Study-3, the expression profiles of 1210 genes in the hippocampal region were assessed after 18 h of reperfusion and 62 genes that were up-regulated in the presence of normothermia, compared with in the sham and hypothermia groups, were selected. In Study-4 and Study-5, the expression profiles of these 62 genes were assessed after 6 and 24 h of reperfusion and two genes were selected. In Study-6, the expression levels of Arc and Cugbp2 in the hippocampal region at 6 and 24 h after reperfusion were examined using immunohistochemistry.

4.1. *Surgical preparation for transient global ischemia*

This study was approved by the Animal Care and Use Committee at Taisho Pharmaceutical Co. Ltd. Male adult C57BL/6N mice were obtained from Charles River Inc. (Atusugi, Kanagawa, Japan). All mice were 8 weeks old, weighed 18.6–23.6 g, and were given free access to food and water prior to surgery and after surgery until occlusion. The surgical procedures used to induce transient global ischemia were the same as previously reported (Himori et al., 1990). Briefly, under anesthesia with sodium pentobarbital (50 mg/kg, i.p.), both carotid arteries were carefully exposed and a ligature for occlusion was placed around them. Two snares for release were placed inside the occlusion loop. The other ends of the loose ligature and snares were exteriorized through the skin at the back of the neck. Twenty hours after the surgical operation, the mice were subjected to 10 min of ischemia by pulling the ends of the exteriorized ligatures and snare. Global ischemia was induced by bilateral common carotid artery occlusion under conscious conditions. Animals that underwent the same surgical procedures, except for occlusion, were used as sham controls. In each study, the number of animals subjected to surgery was adjusted to obtain the target number of surviving animals. For Study-1, Study-2 and Study-3, the target number was ten animals per group and the numbers of mice tested for these studies were 50, 100 and 52 respectively. For Study-4 and Study-5, the target number was 20 animals per group and the numbers of mice tested were 91 and 86. For Study-6, the target number was 3 animals per group and the number of mice tested was 42.

4.2. *Control of body temperature in mice*

The body temperatures of the mice were controlled by monitoring the rectal temperature using a digital thermometer (Omron, Japan). The body temperatures of the normothermic mice were maintained at around 38 °C, while those of the hypothermic mice were maintained at within 20 to 27 °C. The

normothermic mice were placed in an incubator in which the temperature was maintained at 35 °C. The hypothermic mice were kept in cold water at a temperature of 10 °C for 3 to 5 min and their rectal temperatures were then confirmed to be around 24 °C. The mice were then placed in a cooling box, the temperature of which was controlled at 10 °C, and the mice were kept in this box throughout the 10 min of occlusion and for 60 min of reperfusion.

4.3. Behavioral studies

During the ischemic insult, the mortality rate was evaluated and a behavioral assessment was performed to confirm that the occlusion had been successfully performed. For the behavioral assessment, neurologic signs including disturbance of consciousness, circling, torsion of the neck and seizure were assessed. Disturbance of consciousness was classified into three signs: drowsiness (paucity of movement), loss of righting reflex, and coma (Yang et al., 1997). A minor modification was made to the original method (Yang et al., 1997). A score of 3 was given for coma and seizure. Higher mortality rate and severer histologic damage were observed in mice with seizure. The score for coma was reduced from 7 to 3, because it seemed rather high compared with the score for seizure. A score of 2 was given for torsion of the neck and loss of righting reflex, and a score of 1 was given for drowsiness and circling. For disturbance of consciousness, the maximum score among the three scores was adopted. The total score was calculated as the sum of neurologic sign scores. The maximum score for each observation period was thus 9.

4.4. Histologic evaluation of neuronal cell death

Mice that survived the 10 min of global ischemia were sacrificed by cervical dislocation at 7 days after reperfusion in Study-1 and 18, 24, 30, or 48 h after reperfusion in Study-2.

In Study-1, the brains were removed and stored in 10% formalin in phosphate buffered saline (PBS). After dehydration with ethanol, the hippocampus and striatum regions were embedded in paraffin, sectioned (4–5 μm thick), and stained with HE. Neuronal injury in the CA1 to CA3 sector of the hippocampus was assessed according to the grading system reported by Kitagawa et al. (1998). Briefly, an incremental 4-point scoring system was used: 0, normal; 1, <10%; 2, 10–49%; and 3, >50%. The extent of the CA1 to CA3 sector with each grading was measured and the histologic damage score was obtained for each mouse by dividing the integration of each grading and its length by the total length of the CA1 to CA3 sector (Kitagawa et al., 1998). The number of sites with neuronal injury in the CA1 to CA3 sector was counted on both sides.

In Study-2, 0.1 ml/body of 0.1% propidium iodide (PI) in 0.9% saline was administered via the tail vein, and this PI injection was repeated 4 to 6 times at 20-min intervals. Twenty minutes after the final injection, the mice were anesthetized with 50 mg/kg of sodium pentobarbital (Nembtal®) and were subjected to a thoracotomy. Before removal from the skull, the brains were fixed by transcardiac perfusion with heparinised saline (10 IU/ml) followed by 10% buffered formalin. The hippocampus and striatum regions were removed and embedded in OCT compound (Miles Inc.), frozen in 100% acetone (−80 °C), and sectioned (10 μm thick). The sections were immersed for 1 h in propidium iodide solution (4.6 mg/ml) diluted with serum-free medium (75% MEM, 25% HBSS, 2 mM L-glutamine, and 6.5 g/l D-glucose). After rinsing in PBS, the sections were assessed fluoroscopically (excitation, 450 nm; emission, 520 nm). The time course and extent of the neuronal cell death were determined by observing the PI-stained cells.

4.5. Adaptor-tagged competitive-PCR

The ATAC-PCR protocol, including the primer sequences and the 6 sets of adapters, was the same as previously described (Kato, 1997; Matoba et al., 2000). Genes for the ATAC-PCR assay were selected from an in-house brain-expressed sequence tags database (BED, accessible from http://genome.mc.pref.osaka.jp/data_download.html). BED is based on a collection of 3′-end sequences from cDNA libraries made from mouse cerebellum, hippocampus and other brain regions (Matoba et al., unpublished). A total of 1210 genes were selected in order of abundance, with priority given to known genes.

In Study-3, total RNA from the cerebrum of a 6-week-old mouse was used as a control. The relative expression levels of the three treatment groups versus the control sample were calculated. RNA samples were obtained from the hippocampus region of ischemic mice with normothermia (NT) or hypothermia (HT) after 18 h of reperfusion and from a sham control (sham). The root mean square error (RMSE) was calculated according to a previously described algorithm and was used to evaluate the regression lines (Kita-Matsuo et al., 2005). Data with an RMSE of over 0.6 were discarded. The expression patterns of the genes were classified into 9 clusters. When log (NT/HT) was greater than 0.2, the expression level was recognized as NT > HT. The pattern for each cluster is shown in Table 2. Among the up-regulated genes in the NT group, compared with the sham and HT groups, 62 genes were selected and the expression profiles of these genes were further investigated in Study-4 and Study-5.

In Study-4 and Study-5, the procedure for the ATAC-PCR was almost the same as in Study-3 except that the three control cDNA samples were replaced with three treatment samples. Three samples at 6 h after reperfusion (Study-4) and three samples at 24 h after reperfusion (Study-5) were used for the assay. This modification enabled the direct comparison of six samples. ATAC-PCR was repeated twice. The quantitative performance of each gene was assessed as follows. The relative expression level of each sample to the mean of the six samples was obtained separately for the first and second assays. The difference in the relative expression levels between the two assays was calculated as the coefficient of variance (CV) for each gene, and the mean of the CV was calculated for each gene. Data, for which the mean of the CV was over 0.35 or for which a value was missing were discarded. The resulting genes, classified into Cluster-2 and Cluster-8, were recognized as specifically up-regulated in normothermic mice after global ischemia and reperfusion.

4.6. Immunohistochemical analysis of Arc/Arg3.1 and Cugbp2 protein

In Study-6, mice were anesthetized with 50 mg/kg of sodium pentobarbital (Nembtal®) and subjected to a thoracotomy. The

brains were removed from the skulls after transcardiac perfusion with PBS (0.1 M, pH 7.2) and fixed with 10% neutral buffered formalin at room temperature. The brains were then immersed in the same fixative for 2–3 days and embedded in paraffin. The hippocampal regions were sectioned at a thickness of 3 μm and a portion of the sections were stained with 0.1% cresyl violet (Nissl) for conventional microscopy.

For IHC, deparaffinized sections were pre-incubated with 1% bovine serum albumin-PBS for 1 h at room temperature (R.T.). The sections were then rinsed with PBS and boiled in 0.01 M citrate buffer solution at pH 6.0 (00-5000; Zymed Laboratories, Inc., San Francisco, CA) for 30 min for antigen retrieval. The sections were allowed to cool to R.T. and were then incubated with mouse anti-Arc (C-7) antibody (sc-17839, ×1280, 0.15 μg/ml; Santa Cruz Biotechnology, Inc., Santa Cruz, CA), rabbit anti-Arc (H-300) antibody (sc-15325, ×320, 0.625 μg/ml; Santa Cruz Biotechnology), rabbit anti-CUGBP2 (ETR-3) antibody (LS-C31676, ×200, 5 μg/ml; Life-Span BioSciences, Inc., Seattle, WA), or rabbit anti-CUGBP2 antibody (12921-1-AP, ×200, 5 μg/ml; ProteinTech Group, Inc., Chicago, IL) overnight at 4 °C. Next, the sections were incubated with biotinylated horse anti-mouse IgG (BA-2001, 10 μg/ml; Vector Laboratories, Burlingame, CA) or biotinylated goat anti-rabbit IgG (BA-1000, 10 μg/ml; Vector Laboratories) for 1 h at R.T. and immersed in 0.3% H_2O_2-methanol solution for 30 min to quench endogenous peroxidase activity. The sections were washed with PBS and then incubated with Avidin Biotin Peroxidase Complex (PK-6100, Vector Laboratories) for 1 h at R.T., then washed again with PBS. Finally, peroxidase activity was visualized using an AEC staining kit (AEC-101; Sigma, St. Louis, MO) and was observed using an Axioplan 2 microscope (Carl Zeiss, Jena, Germany). Controlled sections were incubated with normal mouse IgG (17314, IBL, Gunma Japan) or normal rabbit IgG (20304, IMGENEX, San Diego, CA) instead of the primary antibodies.

4.7. *Statistical analysis*

All data are presented as mean±S.D. Comparisons between two groups were performed using the Chi-square test or the Fisher exact test. Comparisons among three groups were performed using Kruskal–Wallis test (one-way ANOVA by ranks) followed by Dunn's test for pairwise comparisons. The significance level was set at $P<0.05$.

Acknowledgments

We would like to thank Drs. K. Matsubara and K. Kitamura for their advice and discussions. We also wish to thank Ms. M. Ono, Ms. A. Ikano and Dr. T. Omura for their technical assistance.

REFERENCES

Bernard, S.A., Gray, T.W., Buist, M.D., Jones, B.M., Silvester, W., Gutteridge, G., Smith, K., 2002. Treatment of comatose survivors of out-of-hospital cardiac arrest with induced hypothermia. N. Engl. J. Med. 346, 557–563.

Bloomer, W.A, VanDongen, H.M., VanDongen, A.M., 2007. Activity-regulated cytoskeleton-associated protein Arc/Arg3.1 binds to spectrin and associates with nuclear promyelocytic leukemia (PML) bodies. Brain Res. 1153, 20–33.

Carmichael, S.T., 2005. Rodent models of focal stroke: size, mechanism, and purpose. NeuroRx 2, 396–409.

Choi, D.K., Ito, T., Mitsui, Y., Sakaki, Y., 1998. Fluorescent differential display analysis of gene expression in apoptotic neuroblastoma cells. Gene 223, 21–31.

Choi, D.K., Ito, T., Tsukahara, F., Hirai, M., Sakaki, Y., 1999. Developmentally-regulated expression of mNapor encoding an apoptosis-induced ELAV-type RNA binding protein. Gene 237, 135–142.

DeGracia, D.J., Jamison, J.T., Szymanski, J.J., Lewis, M.K., 2008. Translation arrest and ribonomics in post-ischemic brain: layers and layers of players. J. Neurochem. 106, 2288–2301.

Fletcher, B.R., Calhoun, M.E., Rapp, P.R., Shapiro, M.L., 2006. Fornix lesions decouple the induction of hippocampal Arc transcription from behavior but not plasticity. J. Neurosci. 26, 1507–1515.

Froehler, M.T., Geocadin, R.G., 2007. Hypothermia for neuroprotection after cardiac arrest: mechanisms, clinical trials and patient care. J. Neurol. Sci. 261, 118–126.

Guzowski, J.F., Miyashita, T., Chawlam, M.K., Sanderson, J., Maes, L.I., Houston, F.P., Lipa, P., McNaughton, B.L., Worley, P.F., Barnes, C.A., 2006. Recent behavioral history modifies coupling between cell activity and Arc gene transcription in hippocampal CA1 neurons. Proc. Natl. Acad. Sci. U. S. A. 103, 1077–1082.

Himori, N., Watanabe, H., Akaike, N., Kurasawa, M., Itoh, J., Tanaka, Y., 1990. Cerebral ischemia model with conscious mice involvement of NMDA receptor activation and derangement of learning and memory ability. J. Pharmacol. Methods 23, 311–327.

Hossmann, K.A., 1993. Disturbances of cerebral protein synthesis and ischemic cell death. Prog. Brain Res. 96, 161–177.

Huang, F., Chotiner, J.K., Steward, O., 2007. Actin polymerization and ERK phosphorylation are required for Arc/Arg3.1 mRNA targeting to activated synaptic sites on dendrites. J. Neurosci. 27, 9054–9067.

Irie, Y., Yamagata, K., Gan, Y., Miyamoto, K., Do, E., Kuo, C.H., Taira, E., Miki, N., 2000. Molecular cloning and characterization of Amida, a novel protein which interacts with a neuron-specific immediate early gene product Arc, contains novel nuclear localization signals, and causes cell death in cultured cells. J. Biol. Chem. 275, 2647–2653.

Kato, K., 1997. Adaptor-tagged competitive PCR: a novel method for measuring relative gene expression. Nucleic Acids Res. 25, 4694–4696.

Kitagawa, K., Matsumoto, M., Mabuchi, T., Yagita, Y., Ohtsuki, T., Hori, M., Yanagihara, T., 1998. Deficiency of intercellular adhesion molecule 1 attenuates microcirculatory disturbance and infarction size in focal cerebral ischemia. J. Cereb. Blood Flow Metab. 18, 1336–1345.

Kita-Matsuo, H., Yukinawa, N., Matoba, R., Saito, S., Oba, S., Ishii, S., Kato, K., 2005. Adaptor-tagged competitive polymerase chain reaction: amplification bias and quantified gene expression levels. Anal. Biochem. 339, 15–28.

Kunizuka, H., Kinouchi, H., Arai, S., Izaki, K., Mikawa, S., Kamii, H., Sugawara, T., Suzuki, A., Mizoi, K., Yoshimoto, T., 1999. Activation of Arc gene, a dendritic immediate early gene, by middle cerebral artery occlusion in rat brain. NeuroReport 10, 1717–1722.

Li, L., Carter, J., Gao, X., Whitehead, J., Tourtellotte, W.G., 2005. The neuroplasticity-associated Arc gene is a direct transcriptional target of early growth response (Egr) transcription factors. Mol. Cell Biol. 25, 10286–10300.

Link, W., Konietzko, U., Kauselmann, G., Krug, M., Schwanke, B., Frey, U., Kuhl, D., 1995. Somatodendritic expression of an

immediate early gene is regulated by synaptic activity. Proc. Natl. Acad. Sci. U. S. A. 92, 5734–5738.
Lu, X., Timchenko, N.A., Timchenko, L.T., 1999. Cardiac elav-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy. Hum. Mol. Genet. 8, 53–60.
Lyford, G.L., Yamagata, K., Kaufmann, W.E., Barnes, C.A., Sanders, L.K., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Lanahan, A.A., Worley, P.F., 1995. Arc, a growth factor and activity-regulated gene, encodes a novel cytoskeleton-associated protein that is enriched in neuronal dendrites. Neuron 14, 433–445.
Matoba, R., Saito, S., Ueno, N., Maruyama, C., Matsubara, K., Kato, K., 2000. Gene expression profiling of mouse postnatal cerebellar development. Physiol. Genomics 4, 155–164.
Messaoudi, E., Kanhema, T., Soulé, J., Tiron, A., Dagyte, G., da Silva, B., Bramham, C.R., 2007. Sustained Arc/Arg3.1 synthesis controls long-term potentiation consolidation through regulation of local actin polymerization in the dentate gyrus in vivo. J. Neurosci. 27, 10445–10455.
Murmu, N., Jung, J., Mukhopadhyay, D., Houchen, C.W., Riehl, T.E., Stenson, W.F., Morrison, A.R., Arumugam, T., Dieckgraefe, B.K., Anant, S., 2004. Dynamic antagonism between RNA-binding protein CUGBP2 and cyclooxygenase-2-mediated prostaglandin E2 in radiation damage. Proc. Natl. Acad. Sci. U. S. A. 101, 13873–13878.
Nolan, J.P., Morley, P.T., Vanden Hoek, T.L., Hickey, R.W., Kloeck, W.G., Billi, J., Böttiger, B.W., Okada, K., Reyes, C., Shuster, M., Steen, P.A., Weil, M.H., Wenzel, V., Carli, P., Atkins, D., International Liaison Committee on Resuscitation, 2003. Therapeutic hypothermia after cardiac arrest, an advisory statement by the advanced life support task force of the international liaison committee on resuscitation. Circulation 108, 118–121.
Olsson, T., Wieloch, T., Smith, M.L., 2003. Brain damage in a mouse model of global cerebral ischemia. Effect of NMDA receptor blockade. Brain Res. 982, 260–269.
Ooe, A., Kato, K., Noguchi, S., 2007. Possible involvement of CCT5, RGS3, and YKT6 genes up-regulated in p53-mutated tumors in resistance to docetaxel in human breast cancers. Breast Cancer Res. Treat. 101, 305–315.
Pacini, A., Toscano, A., Cesati, V., Cozzi, A., Meli, E., Di Cesare Mannelli, L., Paternostro, F., Pacini, P., Pellegrini-Giampietro, D.E., 2005. NAPOR-3 RNA binding protein is required for apoptosis in hippocampus. Mol. Brain Res. 141, 34–44.
Palop, J.J., Chin, J., Bien-Ly, N., Massaro, C., Yeung, B.Z., Yu, G.Q., Mucke, L., 2005. Vulnerability of dentate granule cells to disruption of arc expression in human amyloid precursor protein transgenic mice. J. Neurosci. 25, 9686–9693.
Rickhag, M., Teilum, M., Wieloch, T., 2007. Rapid and long-term induction of effector immediate early genes (BDNF, Neuritin and Arc) in peri-infarct cortex and dentate gyrus after ischemic injury in rat brain. Brain Res. 1151, 203–210.
Sugahara-Kobayashi, M., Asai, S., Ishikawa, K., Nishida, Y., Nagata, T., Takahashi, Y., 2008. Global profiling of influence of intra-ischemic brain temperature on gene expression in rat brain. Brain Res. Rev. 58, 171–191.
The Hypothermia after Cardiac Arrest Study Group, 2002. Mild therapeutic hypothermia to improve the neurologic outcome after cardiac arrest. N. Engl. J. Med. 346, 549–556.
Tsuchiya, D., Hong, S., Suh, S.W., Kayama, T., Panter, S.S., Weinstein, P.R., 2002. Mild hypothermia reduces zinc translocation, neuronal cell death, and mortality after transient global ischemia in mice. J. Cereb. Blood Flow Metab. 22, 1231–1238.
Wellons III, J.C., Sheng, H., Laskowitz, D.T., Burkhard Mackensen, G., Pearlstein, R.D., Warner, D.S., 2000. A comparison of strain-related susceptibility in two murine recovery models of global cerebral ischemia. Brain Res. 868, 14–21.
Yang, G., Kitagawa, K., Matsushita, K., Mabuchi, T., Yagita, Y., Yanagihara, T., Matsumoto, M., 1997. C57BL/6 strain is most susceptible to cerebral ischemia following bilateral common carotid occlusion among seven mouse strains: selective neuronal death in the murine transient forebrain ischemia. Brain Res. 752, 209–218.
Yonekura, I., Kawahara, N., Nakatomi, H., Furuya, K., Kirino, T., 2004. A model of global cerebral ischemia in C57 BL/6 mice. J. Cereb. Blood Flow Metab. 24, 151–158.
Zhang, W., Liu, H., Han, K., Grabowski, P.J., 2002. Region-specific alternative splicing in the ervous system: implications for regulation by the RNA-binding protein NAPOR. RNA 8, 671–685.

Supplemental Material can be found at:
http://www.mcponline.org/cgi/content/full/M900179-MCP200/DC2

An Integrated Approach of Differential Mass Spectrometry and Gene Ontology Analysis Identified Novel Proteins Regulating Neuronal Differentiation and Survival*[S]

Daiki Kobayashi‡, Jiro Kumagai§, Takashi Morikawa‡, Masayo Wilson-Morifuji‡, Anthony Wilson‡, Atsushi Irie¶, and Norie Araki‡‖

MS-based quantitative proteomics is widely used for large scale identification of proteins. However, an integrated approach that offers comprehensive proteome coverage, a tool for the quick categorization of the identified proteins, and a standardized biological study method is needed for helping the researcher focus on investigating the proteins with biologically important functions. In this study, we utilized isobaric tagging for relative and absolute quantification (iTRAQ)-based quantitative differential LC/MS/MS, functional annotation with a proprietary gene ontology tool (Molecular Annotation by Gene Ontology (MANGO)), and standard biochemical methods to identify proteins related to neuronal differentiation in nerve growth factor-treated rat pheochromocytoma (PC12) cells, which serve as a representative model system for studying neuronal biological processes. We performed MS analysis by using both nano-LC-MALDI-MS/MS and nano-LC-ESI-MS/MS for maximal proteome coverage. Of 1,482 non-redundant proteins semiquantitatively identified, 72 were differentially expressed with 39 up- and 33 down-regulated, including 64 novel nerve growth factor-responsive PC12 proteins. Gene ontology analysis of the differentially expressed proteins by MANGO indicated with statistical significance that the up-regulated proteins were mostly related to the biological processes of cell morphogenesis, apoptosis/survival, and cell differentiation. Some of the up-regulated proteins of unknown function, such as PAIRBP1, translationally controlled tumor protein, prothymosin α, and MAGED1, were further analyzed to validate their significant functions in neuronal differentiation by immunoblotting and immunocytochemistry using each antibody combined with a specific short interfering RNA technique. Knockdown of these proteins caused abnormal cell morphological changes, inhibition of neurite formation, and cell death during each course of the differentiation, confirming their important roles in neurite formation and survival of PC12 cells. These results show that our iTRAQ-MANGO-biological analysis framework, which integrates a number of standard proteomics strategies, is effective for targeting and elucidating the functions of proteins involved in the cellular biological process being studied. ***Molecular & Cellular Proteomics 8:2350–2367, 2009.***

MS-based quantitative proteomics strategies such as iTRAQ[1] (1) and stable isotope labeling with amino acids in cell culture (2) are powerfully effective for the comprehensive characterization of biological phenomena (1–5). Although these methods have been applied for cancer biomarker (6, 7) and drug target (8) discovery, their use in the elucidation of biological and functional processes has been limited because of certain technical problems that arise when attempting to meaningfully process the immense amount of data obtained from such experiments. The following four main issues are typically the sources of such difficulties. 1) Quantitative identification by one type of MS system may fail to cover the total proteome because of ionization efficiency differences, such as those between ESI and MALDI, for certain peptides, leading to theoretical limitations in proteome coverage. 2) The public protein databases are often insufficient for searching non-human species because of the limited available genomic information. 3) The identification of the functions and biological processes of thousands of proteins is a formidable task because of the lack of simple and user-friendly software to automate gene ontology (GO) annotation. Furthermore it is

From the Departments of ‡Tumor Genetics and Biology and ¶Immunogenetics, Graduate School of Medical Sciences, Kumamoto University and §General Research Core Laboratory, Kumamoto University Medical School, 1-1-1 Honjo, Kumamoto 860-8556, Japan

Received, April 7, 2009, and in revised form, June 12, 2009
Published, MCP Papers in Press, June 13, 2009, DOI 10.1074/mcp.M900179-MCP200

[1] The abbreviations used are: iTRAQ, isobaric tagging for relative and absolute quantitation; NGF, nerve growth factor; GO, gene ontology; PC12PRS, PC12 proteome reference set; ICC, immunocytochemistry; siRNA, short interfering RNA; TCTP, translationally controlled tumor protein; ProTα, prothymosin α; TrkA, tropomyosin-related kinase A; 2-D, two-dimensional; RP, reverse phase; GOA, gene ontology annotation; PCNA, proliferating cell nuclear antigen; PI, propidium iodide; PAIRBP1, plasminogen activator inhibitor 1 RNA-binding protein; PAI, plasminogen activator inhibitor; mESC, mouse embryonic stem cell; MAGE, melanoma antigen; p75NTR, p75 neurotrophin receptor; QqTOF, quadrupole/quadrupole/time-of-flight mass spectrometers.

difficult to convert large lists of taxonomically diverse proteins into their human orthologs to obtain the richest GO information available. 4) Lastly biological validation strategies for identified proteins have not been standardized. Therefore, we believe an analysis framework that provides (*a*) comprehensive proteome data; (*b*) a simple and quick tool for organizing, enriching, and sorting those data to reveal candidate molecules for relation to certain processes; and (*c*) a standardized biological validation technique would greatly benefit this field. We therefore designed a concise, three-step, sequential proteomics strategy that addresses the above concerns and utilized it successfully in studying the mechanism of neuronal differentiation in PC12 cells.

PC12 cells (9) have been widely used as a model of neurons because of their unique advantages, such as stability, homogeneity, strong nerve growth factor (NGF) responsiveness, high differentiation potential, and a wealth of accessible background material, which help to facilitate their manipulation (10). This cell line has also been used for studying the mechanisms of neuronal disorders such as Alzheimer (11), Huntington (12), and Parkinson diseases (13) and neurofibromatosis type 1 (14–16). Here we used PC12 cells as a model for characterizing the mechanisms of neuronal differentiation and neurodegenerative disorders by means of MS-based quantitative proteomics.

NGF is one member of a family of structurally and functionally related dimeric polypeptides, neurotrophins, that are essential for the development and maintenance of distinct neuronal populations in the central and peripheral nervous systems (17). The initial signaling cascades in the neuronal cells right after NGF stimulation have been subjected to thorough investigation and characterization by using PC12 cells. After binding of extracellular NGF to the cell membrane-localized tropomyosin-related kinase A (TrkA) receptor, TrkA receptors dimerize and subsequently autophosphorylate each other. Then the phosphorylated receptors recruit a complex of signaling molecules and induce a number of intracellular signaling cascades involving the signaling molecules, such as phosphoinositide 3-kinase, phospholipase C-γ, and Ras (18). The posttranslational modifications, such as phosphorylation cascades, triggered by NGF stimulation play important roles in PC12 cell differentiation. However, knowledge of the precise dynamic molecular events of protein expression in response to NGF signaling in PC12 cells after an interval that allows the stimulation to take full effect and produce morphological changes remains far from complete.

Several reported studies have applied such methods as expressed sequence tag (19), restriction landmark cDNA scanning (20), targeted display (21), serial analysis of gene expression (22), and cDNA microarray (23) to survey the global change of differentially expressed genes in PC12 cells before and after NGF treatment (19–23). However, the genes and underlying mechanisms associated with the acquisition of a neuronal phenotype in these cells have not been clarified. Also a few proteomics approaches have been used for identifying the proteins related to NGF-inducible neurite formation in PC12 cells. For example, 2-D electrophoresis was applied in whole-cell extract separation to study the NGF modulation of protein synthesis (24); however, only two peptides were identified (25). Even currently available PC12 cell 2-D databases include merely a few proteins related to NGF stimulation (26–29). There is thus a paucity of functional proteomic information related to PC12 cell biological processes that may be attributed to technical limitations such as those listed above.

In this study, we performed the first proteomics survey of proteins differentially expressed in PC12 cells during NGF treatment by using a semiquantitative differential LC shotgun method, namely isobaric tagging for relative and absolute quantitation (iTRAQ) coupled with concurrent use of two tandem MS/MS systems, namely nano-LC-MALDI-TOF-TOF and nano-LC-ESI-Quadrupole/quadrupole/time-of-flight mass spectrometers. The total list of proteins identified was converted into a new file linked to the GO database by our proprietary GO analysis tool for proteomes (MANGO) and categorized by biological process and function using specific classification methods. Thereafter we classified the subset of proteins that were up- or down-regulated during neurite formation into specific molecular categories by combining the differential data obtained by iTRAQ with the proteomic GO analysis results. We then attempted to characterize the functional mechanism of NGF-induced PC12 cell neuronal differentiation. Interestingly the specific up-regulated groups classified in this study were related to apoptosis/cell survival in addition to cell motility, differentiation, stress stimulation, and morphogenesis. To investigate the molecular functions of the up-regulated proteins in relation to both PC12 cell differentiation and apoptosis/survival during neurite formation, some of them were further analyzed with a biochemical and cellular biological strategy using a combined antibody and siRNA technique. Lastly we demonstrated the advantages that our concise, sequential proteomics strategy offers for studying the molecular mechanisms of cellular biological events such as cell differentiation and survival/apoptosis.

EXPERIMENTAL PROCEDURES

Cell Culture, NGF Treatment, and Preparation of Cell Lysate—PC12 cells were cultured under 5% CO_2 at 37 °C in Dulbecco's modified Eagle's medium supplemented with 10% horse serum and 5% fetal bovine serum. We performed four independent cell cultures for a fourplex iTRAQ analysis. Two of them were used as duplicated samples for controls, and the other two samples were used as NGF-treated cells. For NGF stimulation, the cells were cultured onto collagen-coated culture dishes (Iwaki) in the same medium and stimulated with 50 ng/ml 2.5 S NGF (Wako) at 48 h. For preparation of cell lysate, cells were solubilized with the lysis buffer containing 8 M urea, 2% CHAPS, 2 mM Na_2VO_4, 10 mM NaF, 1 μM okadaic acid, and 1% (v/v) protease inhibitor mixture (Sigma) and passed through a 25-gauge syringe 15 times. Lysates were centrifuged at 13,000 × *g* for 20 min at 4 °C, and the protein concentration of the supernatants was determined using the Bio-Rad protein assay.