FEBS Letters 573 (2004) 214–218 FEBS 28729

Identification of region-specific transcription factor genes in the adult mouse brain by medium-scale real-time RT-PCR

Harukazu Suzuki[a,b,*], Rie Okunishi[a,b], Wataru Hashizume[a,b], Shintaro Katayama[a,b], Noriko Ninomiya[a,b], Naoki Osato[a,b], Kenjiro Sato[a,b], Mari Nakamura[a,b], Juri Iida[a,b], Mutsumi Kanamori[a,b], Yoshihide Hayashizaki[a,b]

[a]*Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan*
[b]*Genome Science Laboratory, RIKEN, 2-1 Hirosawa, Wako 351-0198, Japan*

Received 9 June 2004; revised 26 July 2004; accepted 27 July 2004

Available online 11 August 2004

Edited by Takashi Gojobori

Abstract We established a medium-scale real-time RT-PCR system focusing on transcription factors and applied it to their expression profiles in the adult mouse 11 brain regions (http://genome.gsc.riken.jp/qRT-PCR/). Almost 90% of the examined genes showed significant expression in at least one region. We successfully extracted 179 region-specific genes by clustering analysis. Interestingly, the transcription factors involved in the development of the pituitary were still expressed in the adult pituitary, suggesting that they also play important roles in maintenance of the pituitary. These results provide unique molecular markers that may account for the molecular basis of the unique functions of specific brain regions.

Keywords: Real-time RT-PCR; Region-specific; Transcription factor

1. Introduction

The mammalian brain is a highly complex organ, where many unique functions are performed in discrete regions. Exploration of the molecular basis of these unique functions is a great challenge in neurobiology. Recent advances in genomic science, in both large-scale genomic- and cDNA-sequencing projects, have provided information on complete sets of genes in mammals such as humans, mice, and rats [1–8]. Such structural information enables us to systematically analyze the functional properties of genes, even in the complex tissues of the mammalian brain.

Gene expression is one of the most important analyses performed in functional studies of the brain. DNA microarray analysis – both cDNA- and oligo-microarray – has been extensively applied to the exploration of global expression profiles in many organisms, and enables several tens of thousands of genes to be searched on one chip at one time. Many excellent studies using microarray technology have been reported in a wide variety of fields, including brain research [9–11]. However, application of the DNA microarray to the complex tissues seems to be limited in its sensitivity of detection; the sensitivity limit of the DNA microarray has been estimated to be 1–10 copies/cell [12], whereas many transcripts that are important for brain functions are expressed at lower levels or in a relatively small number of cells. An evaluation of DNA microarray sensitivity in rat hippocampal tissue showed that the sensitivity is not high enough for reliable detection of low-abundance transcripts [13]. Further, especially in cDNA microarrays, cross-hybridization may occur among homologous family genes, inhibiting detection of the expression profiles of specific genes in homologous gene families [14].

Real-time RT-PCR is a highly sensitive and specific method that is widely used to explore gene expression and utilizes several detection chemistries [15,16]. Direct detection of amplified double-stranded DNA is carried out by chemical methods with fluorescent dyes such as SYBR Green I. Although such fluorescent dyes bind to any amplified double-stranded DNA, including non-specific double-stranded DNA, the specificity of the test can be checked by the dissociation curve of PCR products after amplification, in which the amplified fragment is confirmed to be unique. Real-time RT-PCR is generally applied to several target genes of interest at one time although a high-throughput detection instrument that uses 384-well assay plates is now commercially available [15,16].

Here, we report the exploration of gene expression profiles of several hundred genes in different regions of the mouse brain using a medium-scale real-time RT-PCR system. We selected transcription factors as target genes of interest, because transcription factors play central roles in the complex regulation of gene expression.

2. Materials and methods

2.1. Preparation of gene-specific PCR primers

We constructed a list of genes for mouse transcription factors in accordance with the TRANSFAC public database (version 7.4; [17]). For the list, we collected genes for both transcription factors of DNA-binding properties and their regulation factors. We also added several related genes from the RIKEN mouse full-length cDNA collection, according to the annotation information of the FANTOM2 database [5] and the InterPro domain information [18].

Gene-specific primer pairs were designed using Primer3 software [19], with an optimal primer size of 20 bases, amplification size of 140 bp, and annealing temperature of 60 °C. The primer pairs were

* Corresponding author. Fax: +81-45-503-9216.
E-mail address: rgscerg@gsc.riken.jp (H. Suzuki).

0014-5793/$22.00
doi:10.1016/j.febslet.2004.07.068

designed to be located within 3 kb of the 3′-end of each gene. The uniqueness of the designed primer pairs was checked by a BLAST homology search (http://www.ncbi.nlm.nih.gov/BLAST/) so that homologous genes were not cross-amplified by the same primer pair.

2.2. RNA preparation and first-stranded cDNA synthesis

Eleven-week-old male C57BL/6J mice were killed according to the Institute's guidelines and the whole brain was removed. The brain, cooled in ice-cold saline, was dissected within 20 min into 11 regions: anterior and posterior cerebral cortices, cerebellum, hippocampus, hypothalamus, medulla oblongata, midbrain, olfactory bulb, pituitary gland, striatum, and thalamus. Total RNA was extracted by the acid phenol–guanidinium thiocyanate–chloroform method [20] and was checked by agarose gel electrophoresis. Total RNA (50 μg) was suspended in 100 μl of RNase-free H_2O with 1 mM $MgCl_2$ solution at final concentration. The RNA solution was centrifuged at $12\,000 \times g$ for 3 min and the supernatant was moved to another micro tube; this was an essential step in the complete removal of contaminated genomic DNA by the following DNase I treatment. RNA was treated with 25 U of DNase I (Stratagene) at 37 °C for 30 min, followed by heat inactivation of the enzyme at 70 °C for 10 min. Complete removal of the contaminated genomic DNA was confirmed by real-time PCR (described below), using the PCR primer pairs for three ubiquitously expressed genes in every sample (glyceraldehyde-3-phosphate dehydrogenase (*GADPH*), β2-macro globulin (β2m), and phosphoglycerate kinase (*Pgk1*)). First-stranded cDNA synthesis (5 μg total RNA per 20-μl reaction) was carried out with the ThermoScript RT-PCR System (Invitrogen), in accordance with the manufacturer's protocol.

2.3. Real-time PCR

PCR amplification, triplicate in each sample, was carried out with an ABI Prism 7900HT instrument (Applied Biosystems). The tailor-made reaction (20 μl) on the 384-well plates was as follows: 0.5 U of HotStar Taq DNA polymerase (Qiagen) and its attached ×1 amplification buffer, additional 1 mM $MgCl_2$, 160 μM dNTPs, 1/38 000 SYBR Green I (Molecular Probes), 7% DMSO, 0.4% ROX Reference Dye (Invitrogen), 300 nM of each primer (forward and reverse), and 2 μl of 40-fold diluted first-stranded cDNA synthesis reaction mixture. The polymerase activation step at 95 °C for 15 min was followed by 40 cycles of 15 s at 94 °C, 30 s at 60 °C, and 30 s at 72 °C. The dissociation curve analysis, which evaluates each PCR product to be amplified from single cDNA, was carried out in accordance with the manufacturer's protocol. Several PCR products were also checked by agarose gel electrophoresis (data not shown).

2.4. Data analysis

Data were analyzed for each 384-well plate. We first made 384 amplification plots, in which the fluorescence signal detected was plotted against the PCR cycle. The amplification curves with the top 10–50% of fluorescence signals in the plateau phase were used to calculate the mean fluorescence signal in this phase. The threshold was empirically set as 4% of the value of the mean fluorescence signal. We then calculated each threshold cycle (C_t) value from the 384 plots.

Clustering of expression profiles was carried out with the HOPACH program (http://stat-www.berkeley.edu/~laan/Research/Research_subpages/Papers/hopach.pdf) with Euclidean dissimilarity metric. Genes specifically expressed in brain regions were selected as follows. We first removed genes that had C_t values larger than 30 in all 11 brain regions, because C_t values more than 30 are less-reliable due to large variations in the reproducibility test (data not shown). Then, we selected genes whose C_t values in some brain regions were 2.5 cycles smaller than the average C_t values of all 11 brain regions. This meant that the level of expression in some regions was ideally more than 5.6-fold higher than the average expression. Finally, we calculated the deviation values of each C_t value for the 11 brain regions and selected genes that had deviation scores of more than 60 in at least one region.

3. Results

3.1. Expression profiles of transcription factor genes in adult mouse brain regions

To detect the levels of mRNA expression for several hundred transcription factor genes, we have established a medium-scale real-time RT-PCR system that used a tailor-made reaction with SYBR Green I dye as a detector of PCR products. Using the system, we constructed the expression profiles of 756 transcription factor genes (http://genome.gsc.riken.jp/qRT-PCR/). A high percentage of transcription regulation factors seems to be expressed in adult mouse brain: 543–612 genes (72–81%) showed C_t values less than 30 (corresponding to 69.7 copies/reaction, as estimated from the standard curves) in each brain region. Further, 677 genes (89.6%) showed C_t values less than 30 in at least one brain region, reflecting the high complexity of mouse brain tissues. Next, we clustered the expression profiles,

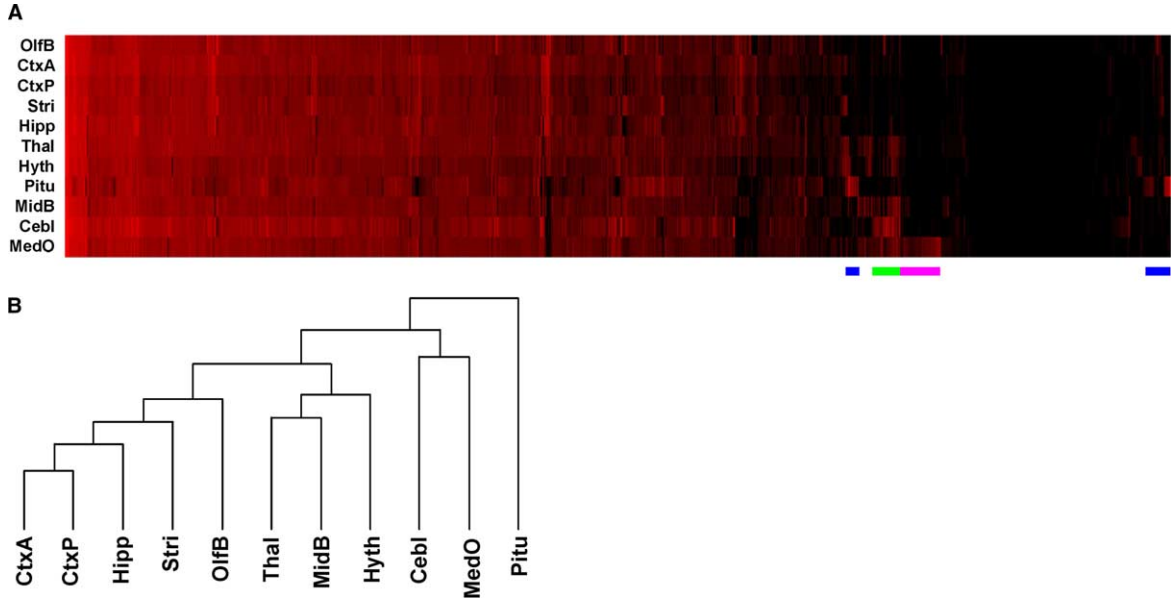

Fig. 1. Global insights into the expression profiles of transcription factor genes in adult mouse brain regions. (A) Clustering analysis using the HOPACH program. Gene expressions with Ct values from 30 to 17 appear as light red to red. The blue, light green, and pink bars show genes enriched in the pituitary, cerebellum, and medulla. (B) Phylogenic tree of the 11 brain regions, based on the expression of transcription factor genes.

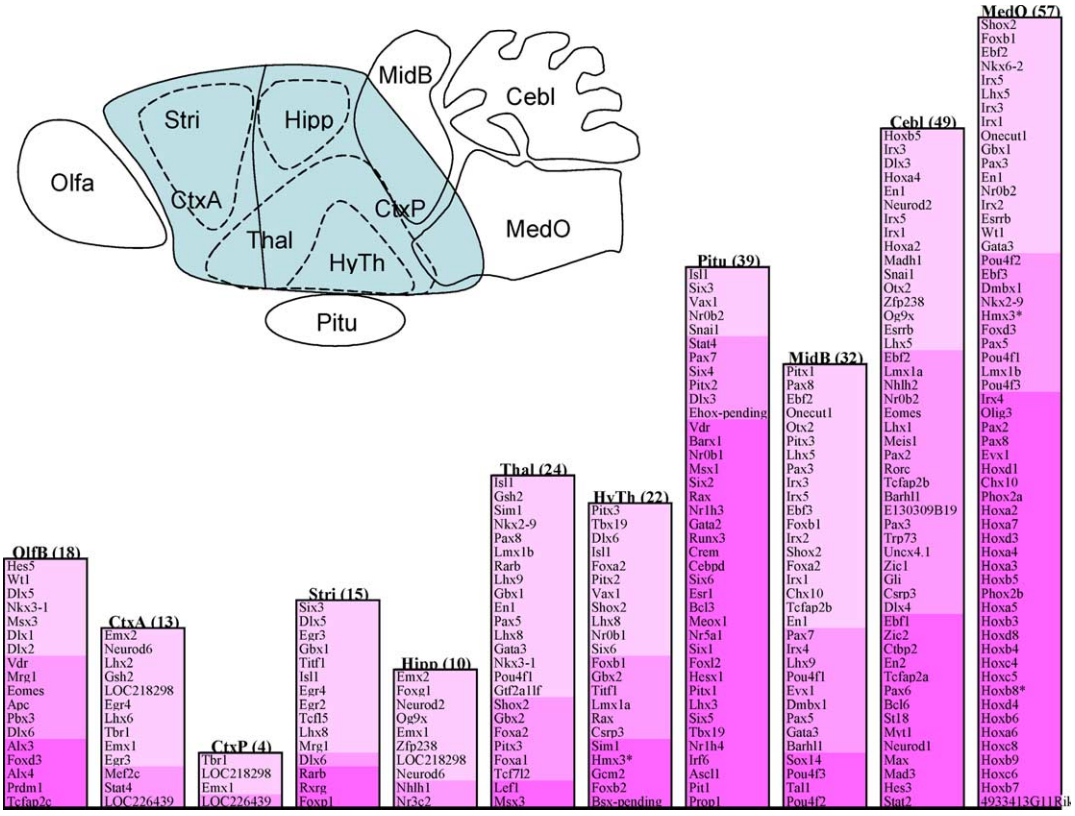

Fig. 2. Region-specific genes identified by the medium-scale real-time RT-PCR system. The 11 dissected brain regions and their abbreviated names are schematically represented in the line diagram. The brain regions and numbers of region-specific genes are shown above each column. In the columns, region-specific genes are shown in the order of smaller to larger deviation score. Light pink, pink, and dark pink represent deviation scores of 60.0–64.9, 65.0–69.9, and 70.0 and above, respectively.

using the HOPACH program, to gain an overview of transcription factor gene expression (Fig. 1A). All 11 brain regions examined had similar expression profiles, although several genes seemed to be expressed in a region-specific manner. The phylogenic tree of the brain regions, with the exception of the pituitary gland, which includes cells derived from early ectoderm (see below), seemed to be consistent with the development of each brain region (Fig. 1B); the regions derived from the

Table 1
Confirmation of cerebellum-specific transcription factor genes identified by real-time RT-PCR

Locus ID	Gene Name	Dev.	Ct	Methods	Refs.
15207	*Hes3*	77.8	20.9	Northern	[38]
17121	*Mad3*	76.4	28.3		n.r.
18012	*Neurod1*	75.2	18.7	SAGE	[39]
21418	*Tcfap2a*	73.4	24.2	ISH	[40]
13799	*En2*	72.1	21.0		n.r.
22772	*Zic2*	70.3	20.9	Northern, SAGE	[27,39]
22771	*Zic1*	69.2	19.5	Northern	[28]
22255	*Uncx4.1*	68.6	22.8		n.r.
18505	*Pax3*	68.2	25.1	RT-PCR*	[41]
243187	*E130309B19*	68.1	21.8		n.r.
54422	*Barhl1*	67.9	25.6	ISH**	[29]
21419	*Tcfap2b*	67.7	22.3		n.r.
16869	*Lhx1*	66.6	21.2	ISH, IH	[30]
23957	*Nr0b2*	66.2	27.0		n.r.
13592	*Ebf2*	65.2	25.0	RNAse protection	[31]
16873	*Lhx5*	64.4	28.9	RT-PCR	[42]
18292	*Og9x*	64.0	24.3	ISH	[32]
18013	*Neurod2*	61.7	20.7	Northern, ISH	[43]

We selected 18 prominent cerebellum-specific genes whose Ct values were more than 4.0 cycles smaller than the average Ct values of all 11 brain regions (see http://genome.gsc.riken.jp/qRT-PCR/). We used PubMed (http://www.ncbi.nlm.nih.gov/) to search the literature using the keywords 'cerebellum', 'expression', and the gene name (e.g. '*Hes3*'), and then manually collated the results. The Section 2 and the literature references (**Refs.**) used to elicit cerebellum-specific expression of the target genes are shown in the Table. '**Dev.**', 'deviation score'; '**ISH**', 'in situ hybridization'; '**IH**', 'immunohistochemistry'; '**n.r.**', 'not reported'. * The alternatively spliced form of *Pax3A* is expressed in a cerebellum-specific manner. ** Strongly suggested from the embryonic expression pattern.

embryonic telencephalon (olfactory bulb, anterior and posterior cerebral cortices, striatum, and hippocampus), the diencephalons and mesencephalon (thalamus, hypothalamus, and midbrain), and metencephalon and myelencephalon (cerebellum and medulla oblongata) were clustered well.

3.2. Analysis of region-specific genes

We selected genes whose expression was enriched in a region-specific manner; this resulted in the identification of 179 genes (Fig. 2, and http://genome.gsc.riken.jp/qRT-PCR/). Such region-specific genes were observed frequently in the medulla oblongata, cerebellum, pituitary gland, and midbrain, but less often in the posterior cerebral cortex and hippocampus. In particular, transcription factor genes with high deviation scores (>70; shown in dark pink in Fig. 2), which suggest highly region-specific expression, were found frequently in the medulla oblongata and the pituitary gland. To evaluate our findings, we selected 18 cerebellum-specific genes that exhibited prominent cerebellum-specific expression in our experiments and searched the literature to find out whether they had already been identified as cerebellum-specific genes. In the literature, 12 of the 18 genes had been confirmed, or were strongly suspected, as being cerebellum-specific (Table 1).

We further analyzed genes enriched in the pituitary, because several transcription regulation factors are well known to be involved in pituitary development [21,22]. The mammalian pituitary is an endocrine organ consisting of several types of secretory cells; each type of cell secretes a unique peptide hormone. These cell types are derived from a common origin in the early ectoderm. Our results showed that most of the transcription factors involved in the development and differentiation of each type of secretory cell are still expressed specifically in the adult pituitary gland (pink columns in Fig. 3). The results suggest that such transcription factors play important roles in both the development and maintenance of secretory cells in the pituitary. Interestingly, we observed that expression of the *Nr1h3* and *Nr1h4* genes, known as activators of retinoic acid receptor (RAR) [23–25], was greatly enriched in the pituitary, although *RAR* mRNA, which is involved in somatotrope differentiation, is expressed in both the pituitary and other brain regions (http://genome.gsc.riken.jp/qRT-PCR/). The result suggests that Nr1h3 and Nr1h4 play critical roles as determinants of somatotrope differentiation (and also maintenance of somatotrophs) by regulating RAR activity.

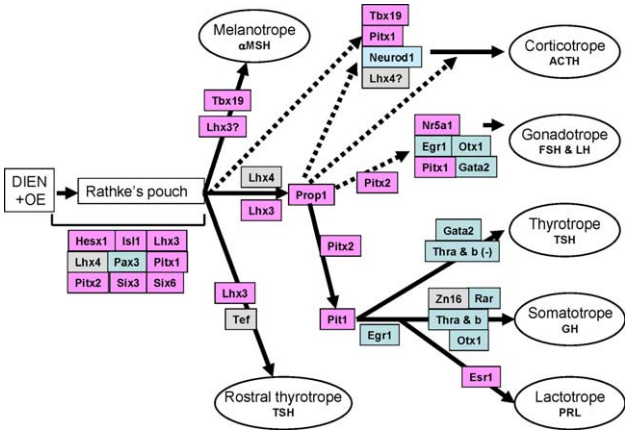

Fig. 3. Detection of the transcription factors regulating anterior pituitary gland development. This figure was made by referring the figure from [22]. After interaction with the ventral diencephalons (DIEN), oral ectoderm (OE) forms Rathke's pouch and then differentiates into several hormone-secreting cell types (shown in ellipses). Many transcription factors (shown in the colored columns) have been known to act throughout development of the pituitary. Transcription factors that showed and did not show pituitary-specific expression in our real-time RT-PCR studies are marked with pink and light blue columns, respectively. *Gata2* showed pituitary-enriched expression, but was not considered in our criteria as a region-specific gene (see Section 2 and http://genome.gsc.riken.jp/qRT-PCR/). Transcription factors that could not be explored in our studies are shown in gray columns.

4. Discussion

We successfully identified 179 region-specific genes and confirmed that the expression of several of them was consistent with data in previous reports, thus showing that our results were reliable. Several regions of the adult brain, such as the posterior cerebral cortex and hippocampus, had relatively few region-specific transcription factor genes. However, this may be biased by the effect of normalization, whereby genes highly expressed in a small area and/or a small number of cells in a particular dissected region may have been buried and thus not detected as region-specific genes. Thus, real-time RT-PCR analysis of more restricted brain regions and/or groups of cells may be valuable for further identification of region-specific genes. In fact, both here (data not shown) and in a previous report [26], we successfully detected Nr2f2 as an amygdala-specific transcription factor by real-time RT-PCR, although *Nr2f2* was not identified in our experiments as a region-specific gene of the anterior cerebral cortex, with which the amygdala is involved. We also stress that the identified region-specific genes should be further explored by other methods such as in situ hybridization. Together with the need to confirm our results, it is very important to explore whether the expression of these genes is spatially localized in specific nuclei and/or cells in the dissected region.

We found that most of the transcription factors involved in the development and differentiation of each type of secretory cell in the pituitary were still expressed in a region-specific manner in the adult pituitary. This observation seems to be applicable to other brain regions such as cerebellum [27–32]. Region-specific transcription factor genes may be good clues to our understanding of tissue development and functions. Because current chromatin immunoprecipitation (ChIP) technology allows us to identify the interaction loci of each transcription factor on the genomic DNA, it is possible to draw transcription regulation motifs and pathways by integrating the expression profiles with data from ChIP analysis, and thus show how transcription factors are regulated by one another [33,34]. Further, as we showed that Nr1h3 and Nr1h4 may be involved in somatotrope differentiation through RAR, it will also be useful to integrate the expression profiles with protein–protein interaction data, because most of the transcription factors act as homo- and hetero-dimers and are regulated by direct interaction with a variety of modulating factors. Finally, ChIP analysis may also disclose the target genes of region-specific transcription factors, which may also lead to identification of novel region-specific genes that play important roles in the functioning of brain regions.

In addition to the analysis of region-specific genes, our expression profiles can also be analyzed to find novel transcription factor genes involved in regionalization of the brain, where they are expressed in relatively large regions such as forebrain, midbrain or hindbrain rather than one of the 11 dissected regions. So far, more than 10 transcription factor genes have been identified as genes controlling regionalization of the brain such as *En1* and *2*, *Pax2*, *5* and *6*, *Emx1* and *2*, *Dlx1* and *2*, *Otx1* and *2*, *Gsh2* and *Gbx2* [35–37]. We found in our data that all the genes, except for *Pax6*, are still expressed distinctively in adult brain. Remarkably, we discovered that many genes, for example *Lhx* gene family, also exhibit similar expression patterns to those of genes described above. Thus, it is a very attractive way to explore the function of such genes with regard to the regionalization of the brain.

This medium-scale real-time RT-PCR system seems valuable for expression profile analysis of highly complex tissues such as the brain and/or relatively sparsely distributed transcript families such as transcription factors, although the system is basically applicable to all tissues. In addition, the system may be suitable for exploring the dynamic expression of genes during different developmental stages and/or conditions, such as during drug treatment. The system is almost ready for general use, although the primer sets for mouse transcription factors are still not complete. The major reason is that there is no comprehensive non-redundant transcription factor list in the mouse. Recently, we have started to construct a genome-wide, non-redundant mouse transcription factor list, which will provide us with additional transcription factor genes to be explored.

Acknowledgements: We thank Dr. K. Polland (UCSC) for providing the HOPACH program and A. Hasegawa for technical assistance. We also thank the members of the Laboratory for Genome Exploration Research Group. This study was supported by MEXT Research Grants for the RIKEN Genome Exploration Research Project, for Advanced and Innovational Research Program in Life Science and for National Project on Genome Network Analysis to Y.H.

References

[1] Lander, E.S. et al. (2001) Nature 409, 860–921.
[2] Venter, J.C. et al. (2001) Science 291, 1304–1351.
[3] Kawai, J. et al. (2001) Nature 409, 685–690.
[4] Waterston, R.H. et al. (2002) Nature 420, 520–562.
[5] Okazaki, Y. et al. (2002) Nature 420, 563–573.
[6] Ota, T. et al. (2004) Nat. Genet. 36, 40–45.
[7] Imanishi, T. et al. (2004) PLoS Biol. 2, E162.
[8] Gibbs, R.A. et al. (2004) Nature 428, 493–521.
[9] Lockhart, D.J. and Barlow, C. (2001) Nat. Rev. Neurosci. 2, 63–68.
[10] Russo, G., Zegar, C. and Giordano, A. (2003) Oncogene 22, 6497–6507.
[11] Henry, G.L., Zito, K. and Dubnau, J. (2003) Curr. Opin. Neurobiol. 13, 570–576.
[12] Kane, M.D., Jatkoe, T.A., Stumpf, C.R., Lu, J., Thomas, J.D. and Madore, S.J. (2000) Nucleic Acids Res. 28, 4552–4557.
[13] Evans, S.J., Datson, N.A., Kabbaj, M., Thompson, R.C., Vreugdenhil, E., De Kloet, E.R., Watson, S.J. and Akil, H. (2002) Eur. J. Neurosci. 16, 409–413.
[14] Miller, N.A., Gong, Q., Bryan, R., Ruvolo, M., Turner, L.A. and LaBrie, S.T. (2002) Biotechniques 32, 620–625.
[15] Giulietti, A., Overbergh, L., Valckx, D., Decallonne, B., Bouillon, R. and Mathieu, C. (2001) Methods 25, 386–401.
[16] Bustin, S.A. (2002) J. Mol. Endocrinol. 29, 23–39.
[17] Matys, V. et al. (2003) Nucleic Acids Res. 31, 374–378.
[18] Mulder, N.J. et al. (2003) Nucleic Acids Res. 31, 315–318.
[19] Rozen, S. and Skaletsky, H. (2000) Methods Mol. Biol. 132, 365–386.
[20] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[21] Scully, K.M. and Rosenfeld, M.G. (2002) Science 295, 2231–2235.
[22] Savage, J.J., Yaden, B.C., Kiratipranon, P. and Rhodes, S.J. (2003) Gene 319, 1–19.
[23] Lu, T.T., Makishima, M., Repa, J.J., Schoonjans, K., Kerr, T.A., Auwerx, J. and Mangelsdorf, D.J. (2000) Mol. Cell. 6, 507–515.
[24] Yoshikawa, T. et al. (2001) Mol. Cell Biol. 21, 2991–3000.
[25] Urizar, N.L., Dowhan, D.H. and Moore, D.D. (2000) J. Biol. Chem. 275, 39313–39317.
[26] Zirlinger, M., Kreiman, G. and Anderson, D.J. (2001) Proc. Natl. Acad. Sci. USA 98, 5270–5275.
[27] Aruga, J., Nagai, T., Tokuyama, T., Hayashizaki, Y., Okazaki, Y., Chapman, V.M. and Mikoshiba, K. (1996) J. Biol. Chem. 271, 1043–1047.
[28] Aruga, J., Yokota, N., Hashimoto, M., Furuichi, T., Fukuda, M. and Mikoshiba, K. (1994) J. Neurochem. 63, 1880–1890.
[29] Bulfone, A. et al. (2000) Hum. Mol. Genet. 9, 1443–1452.
[30] Hayes, W.P., Yangco, N., Chin, H., Mill, J.F., Pu, L.P., Taira, M., Dawid, I.B. and Gallo, V. (2001) J. Neurosci. Res. 63, 237–251.
[31] Malgaretti, N. et al. (1997) J. Biol. Chem. 272, 17632–17639.
[32] Cinquanta, M., Rovescalli, A.C., Kozak, C.A. and Nirenberg, M. (2000) Proc. Natl. Acad. Sci. USA 97, 8904–8909.
[33] Lee, T.I. et al. (2002) Science 298, 799–804.
[34] Odom, D.T. et al. (2004) Science 303, 1378–1381.
[35] Nakamura, H. (2001) Trends Neurosci. 24, 32–39.
[36] Simeone, A., Puelles, E. and Acampora, D. (2002) Curr. Opin. Genet. Dev. 12, 409–415.
[37] Simeone, A. (2002) Trends Neurosci. 25, 119–121.
[38] Sakagami, T., Sakurada, K., Sakai, Y., Watanabe, T., Nakanishi, S. and Kageyama, R. (1994) Biochem. Biophys. Res. Commun. 203, 594–601.
[39] Siu, I.M., Lal, A. and Riggins, G.J. (2001) Gene Expr. Patterns 1, 33–38.
[40] Shimada, M., Konishi, Y., Ohkawa, N., Ohtaka-Maruyama, C., Hanaoka, F., Makino, Y. and Tamura, T. (1999) Neurosci. Res. 33, 275–280.
[41] Tsukamoto, K., Nakamura, Y. and Niikawa, N. (1994) Hum. Genet. 93, 270–274.
[42] Zhao, Y., Hermesz, E., Yarolin, M.C. and Westphal, H. (2000) Gene 260, 95–101.
[43] Olson, J.M. et al. (2001) Dev. Biol. 234, 174–187.