GEDA: New Knowledge Base of Gene Expression in Drug Addiction

Young Ju Suh[1,2], Moon Hee Yang[2], Suk Joon Yoon[2] and Jong Hoon Park[2,*]

[1]The Research Institute of Natural Sciences, Sookmyung Women's University, Seoul 140-742, Korea
[2]Research Center for Women's Diseases, Department of Biological Sciences, Sookmyung Women's University, Seoul 140-742, Korea

Abuse of drugs can elicit compulsive drug seeking behaviors upon repeated administration, and ultimately leads to the phenomenon of addiction. We developed a procedure for the standardization of microarray gene expression data of rat brain in drug addiction and stored them in a single integrated database system, focusing on more effective data processing and interpretation. Another characteristic of the present database is that it has a systematic flexibility for statistical analysis and linking with other databases. Basically, we adopt an intelligent SQL querying system, as the foundation of our DB, in order to set up an interactive module which can automatically read the raw gene expression data in the standardized format. We maximize the usability of this DB, helping users study significant gene expression and identify biological function of the genes through integrated up-to-date gene information such as GO annotation and metabolic pathway. For collecting the latest information of selected gene from the database, we also set up the local BLAST search engine and non-redundant sequence database updated by NCBI server on a daily basis. We find that the present database is a useful query interface and data-mining tool, specifically for finding out the genes related to drug addiction. We apply this system to the identification and characterization of methamphetamine-induced genes' behavior in rat brain.

Keywords: Database, Drug addiction, Gene expression, Microarray, Rat brain

Introduction

Abuse of drugs elicits compulsive drug seeking behaviors upon repeated administration, and ultimately leads to the phenomenon of addiction (Everitt and Robbins, 2005).

*To whom correspondence should be addressed.
Tel: 82-2-710-9414; Fax: 82-2-2077-7322
E-mail: parkjh@sookmyung.ac.kr

Addiction (specially methamphetamine addiction) becomes a great problem not only for individual but also for the society (Klag *et al.*, 2005). However, the action mechanism has not been clearly elucidated yet, and thus no therapeutics is available for the treatment of drug addiction. For development of treatment, it is important to clarify addiction mechanism.

Recently, many researchers have studied addictive mechanism using high throughput screening tools. Microarray experiment is one of the most popular high throughput screening techniques for the comparative study of gene expression in disease-related cells and tissues (Rouse and Hardiman, 2003; Alberts *et al.*, 2005). This technique has been commonly used in the addiction problem of drug such as methamphetamine (Noailles *et al.*, 2003; Funada *et al.*, 2004; Thomas *et al.*, 2004; Yamada *et al.*, 2005; Yamamoto *et al.*, 2005), cocaine (McClung and Nestler, 2003; Albertson et al., 2004; Ahmed *et al.*, 2005; Bahi and Dreyer, 2005) and alcohol (Arlinde *et al.*, 2004; Iwamoto *et al.*, 2004; Sommer *et al.*, 2005). However, there was no systematic approach to gather these biological data in a large scale. We have only limited information on drug abuse and the significantly expressed genes without a systematic database (DB).

We develop a procedure for the standardization of microarray gene expression data of rat brain in drug addiction and stored them in a single integrated database system, named as Gene Expression in Drug Addiction (GEDA: http://compbio.sookmyung.ac.kr/~geda/search3_newwin.html), focusing on more effective data processing and interpretation. Repeated drug exposure leads to gene expression change and consequently behavior alteration (Nestler, 2001). Therefore, we divided the experimental samples into four types of addiction model in rats: Saline, MAP1, MAP2 and MAP3 that based on behavior alteration measurement. We used CPP (conditioned preference place) test for behavior alteration measurement (Zhao *et al.*, 2003). Saline is a group without any drug treatment. MAP1 is an addiction formation group (after treatment of methamphetamine, strong addictive drug). MAP2 is an addictive behavior disappearance group after addictive behavior pattern formation. MAP3 is a maintenance group of addictive behavior. We use oligonucleotide microarray to
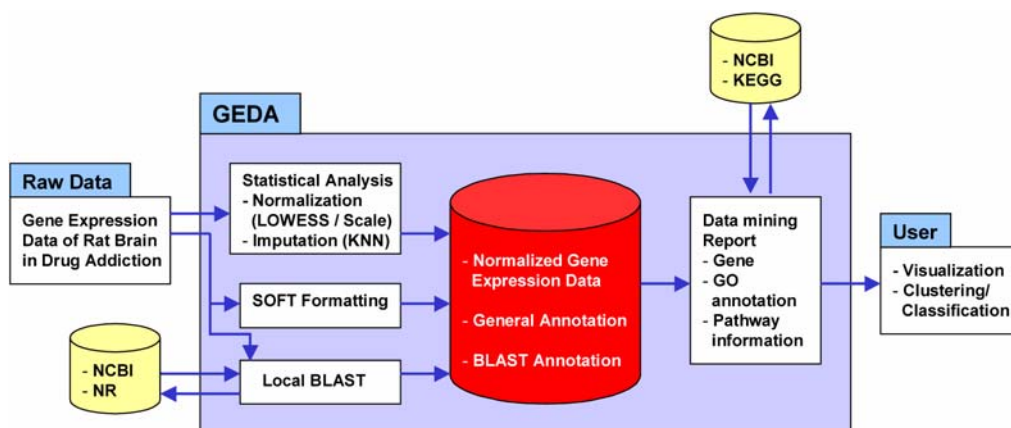
Fig. 1. Database structure for gene expression data related to drug addiction.

identify the drug addiction mechanism in these four groups. Especially we investigated specific brain regions in rats: cingulate cortex (Cing), hippocampus (Hipp), nucleus accumbens (Na) and striatum (Str). These four regions are involved in mesolimbic dopamine system. The mesolimbic dopamine system is the main region of drug addiction. For identifying addictive mechanism, we screened meaningful genes using microarray in the specific brain region of each addiction group.

It is be worthwhile to set up a user-friendly, easy-accessible web interface to carry out genetic analysis for drug addiction. Thus we also tried to maximize the usability of this DB, helping users study significant gene expression and identify biological function of the genes accurately and easily through up-to-date gene information such as GO annotation, metabolic pathway, BLAST search, and etc.

Methods

Preliminary data analysis. GEDA database is composed of gene expression profiles of four rat brain regions (Cing, Hipp, Na, Str) in four drug addiction groups (Saline, MAP1, MAP2, MAP3 grouped by CPP test). We preliminarily analyzed the 10 k BioArray oligoneucleotide chip data using appropriate statistical method (Fig. 1). We filtered out several genes in accordance with a measured status of fluorescent intensity of gene expression. We considered the unqualified data as missing values. After the filtering process, 10,399 genes were included in each of the 16 slides. We analyzed the gene expression microarray data of Saline, MAP2, MAP3, comparing to MAP1 in each of four brain regions. We used intensity dependent (LOWESS) normalization, and scale adjustment normalization to adjust the scale differences in each pin. Finally, 6,300 informative genes were obtained for our analysis, after filtering out genes where more than one (8.4%) of the expression values was missing. For informative genes, the K-nearest neighbor (KNN) method with $k = 10$ was used for missing data imputation. Statistical analysis was performed using the R-Bioconductor.

Database format. At present NCBI website of America has a DB named GEO (Gene Expression Omnibus) (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=geo), which has been saving the results of various gene expression experiments having been carried out in the world, using a standardized format based on raw data. Accordingly, we standardize a number of gene expression profile data of rat brain in drug addiction in the mode of SOFT (Simple Omnibus Format in Text) used in the GEO to be compatible with other DBs in the future, setting up a systematic DB based on the standardized materials. Another characteristic of the present database is that it has a systematic flexibility for statistical analysis and linking with other databases (Fig. 1). Basically, we adopt an intelligent SQL querying system, as the foundation of our DB, in order to set up an interactive module which can automatically read the raw gene expression data in the SOFT format.

Application module for data classification. We develop Java application module which enables printing out the results of data classification with multiple-choice options on website. For example, Fig. 2 shows an interface to find genes related to the methamphetamine drug addiction in the microarray data. "Target group" includes main subgroups we focus on, and "Control group" contains subgroups to be compared with the target group. Users can specify the "Fold change" value for gene expression profile ratio (for example, 1 or −1), and select ">" (greater than specified cutoff value) or "<" (less than the cutoff value). Three different groups of "Saline", "MAP2", and "MAP3" are composed of the normalized ratios of gene expression profiles for Saline against MAP1, those for MAP2 against MAP1, and those for MAP3 against MAP1, respectively. Each of groups consists of four subgroups of different brain regions such as "Cing", "Hipp", "NA", "Str".

As a result, the gene expression profile ratios can be compared within or between groups through the interface. Eventually, the established DB enables to verify significantly expressed genes that meet the conditions suggested by users. For example in Fig. 2, focusing on genes relatively down regulated in Hipp region (target group), we explore genes that are not down regulated in the other three brain regions (control groups) within Saline group (against MAP1 group). In this search, we found 148 genes differently
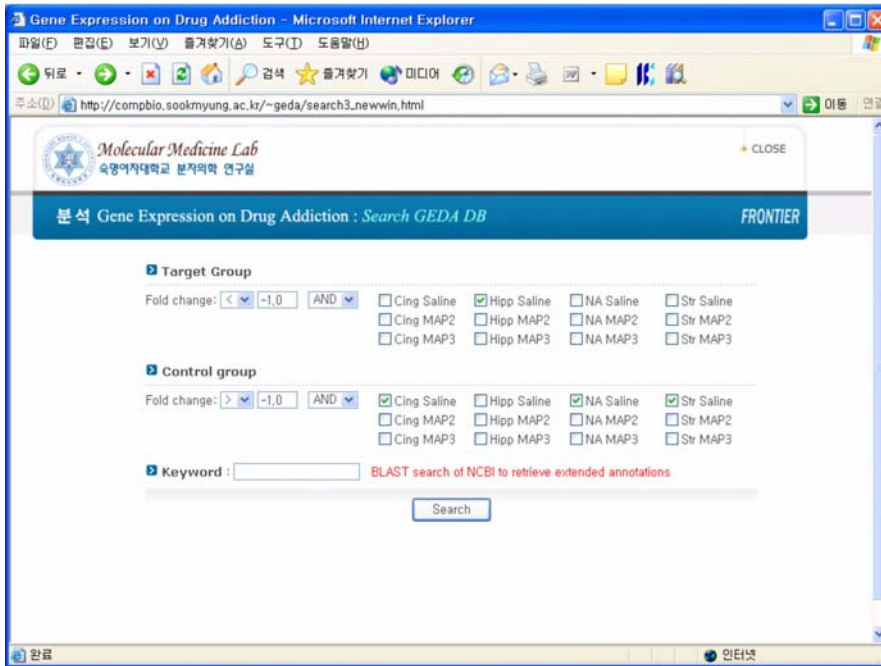
Fig. 2. Main web interface.

expressed at the Hipp region, comparing to other three brain regions (Fig. 3).

Linking to other databases. The established DB enables to confirm their biological functions by using their DNA chip annotation information. We try to link the expression profile genes with the existing DB, referring gene annotation ID of oligoneucleotide chip

used in this study (Fig. 1). We maximize the usability of this DB, helping users study significant gene expression and identify biological function of the genes through integrating up-to-date gene information such as GO annotation and KEGG metabolic pathway. We can link to other databases such as NCBI and KEGG using RefSeq ID (mRNA sequence ID provided in NCBI), gene name and pathway information for genes hit by GEDA (Fig. 3). In this
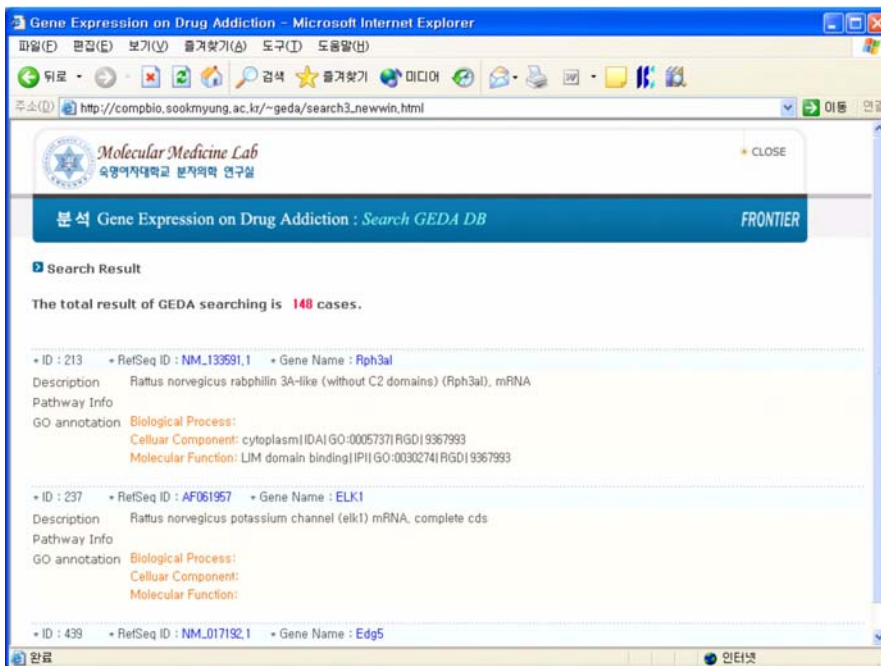

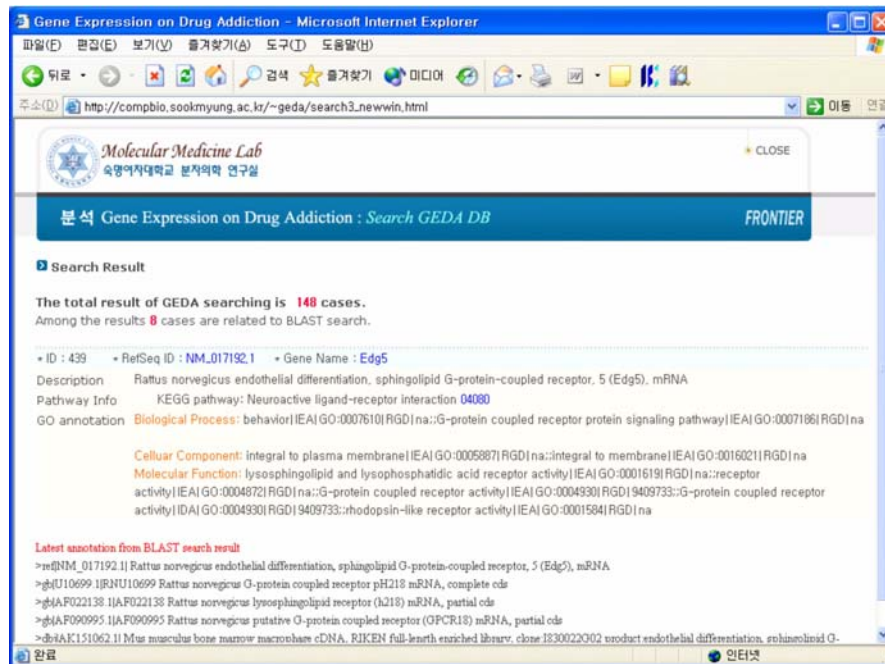**Fig. 3.** Results of gene expression database search.

Fig. 4. Results of database search with keyword.

process we try to make GEDA DB do an important role to find genes susceptible to drug abuse.

Local BLAST search engine. For collecting the latest information of selected gene from the database, we also set up the local BLAST search engine and non-redundant (NR) sequence database updated by NCBI server on a daily basis (Fig. 1). In Fig. 2, doing BLASTn search by denoting functional terminology in the "Keyword", we can identify genes including the functional terminology related to BLAST updated annotated information. Fig. 4 shows that 8 genes (among 148 genes) are detected by GEDA using keyword as receptor. Among the selected 8 genes, we can investigate meaningful genes involved in receptors associated with drug addiction in the next step.

Results

Through GEDA, we could search and classify the drug addiction data in several aspects. First of all, we focused on genes with different expression ratio among three CPP groups of Saline, MAP2 and MAP3 in each of four brain regions. Here we considered gene expression log (base 2) ratio greater than 0.5 as U (up-regulated) and gene expression log (base 2) ratio less than –0.5 as D (down-regulated). Classifying the data by possible distribution of U in combination with D in three CPP groups, we obtained several genes hit by the GEDA in each of four brain regions (Table 1). In particular, we focused on hits (genes) in four distribution such as "UUD", "UDU", "DUD" and "DDU" for Saline, MAP2 and Map3 group, since it is important to see genes having remarkably different distribution between MAP2 (a group that

addictive behaviors are disappeared for a month after addiction formation) and MAP3 (a group that addictive behaviors are maintained for a month after addiction formation). Among the selected genes in these four groups, we found several hits related to drug addiction such as Hear shock protein 4 (Hspa4, HSP70) and Galanin (Gal). Hspa4, detected in "UDU" group in the Str brain region, is considered as an oxidative injury related marker in methamphetamine addiction (Ishigami *et al.*, 2003). Found in "DUD" group in the Na brain region, Gal is known as an important neuropeptide in opiate addiction. Gal and its receptor are expressed in locus coeruleus and ventral tegmental region (Hawes *et al.*, 2004). Especially Gal acts attenuator of opiate reinforcement and withdrawal (Picciotto *et al.*, 2005). Gal is also related to alcoholism. In vivo model, Gal is increased by ethanol intake and decreased by withdrawal in hypothalamus (Leibowitz *et al.*, 2003).

Secondly, we did BLAST search by GEDA using three keywords such as receptor, EST and factor (Table 1). The reason why we used these keywords is to find receptors, new genes, or factors in signaling pathway, related to drug addiction, respectively. We discovered several genes associated with drug addiction through GEDA using the keywords. For example, glutamate receptor (Grin1, NR1) in "UUU" group at the Na brain region plays an important role in excitatory neurotransmission, synaptic plasticity and brain development. In vivo model Grin1 is involved in methamphetamine induced abnormal behavior (Miyamoto *et al.*, 2004).

Third, we searched genes with various distribution of expression ratio among four brain regions of Cing, Hipp, NA, Str in each of three CPP groups. Through GEDA database analysis, we found genes in each class formed by combination

Table 1. Classified distribution of three different groups in each of brain regions

Distribution of CPP groups[a]			No. of hits (genes) in each of brain regions[b]														
			Without keyword				With keyword-receptor				With keyword-EST				With keyword-factor			
Sal	M2	M3	Cin	Hip	Na	Str	Cin	Hip	Na	Str	Cin	Hip	Na	Str	Cin	Hip	Na	Str
U[c]	U	U	28	82	37	40	2	5	5	1	2	4	2	2	0	1	2	0
U	U	D	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
U	D	U	1	0	0	3	0	0	0	0	0	0	0	1	0	0	0	0
D	U	U	17	27	1	5	0	0	0	1	0	0	0	0	0	0	0	0
U	D	D	18	3	5	18	1	0	0	1	3	0	0	3	2	0	0	0
D	U	D	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
D	D	U	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
D	D	D	53	66	88	47	3	5	2	2	3	1	3	2	0	3	3	3

[a]Three different groups determined by the CPP test. Sal, M2, M3 indicate the groups with normalized log ratio (base 2) of gene expression profile for Saline, MAP2, MAP3 group against MAP1 group, respectively.
[b]Number of genes searched by GEDA in each of four different brain regions. Cin, Hip, Na, Str indicate cingulate cortex, hippocampus, nucleus accumbens, striatum, respectively.
[c]U is up-regulated and D is down regulated genes.

Table 2. Classified distribution of four different brain regions in each of CPP groups

Distribution of brain regions[a]				No. of hits (genes) in each of CPP groups[b]								
				Without keyword			With keyword-receptor			With keyword-EST		
Cin	Hip	Na	Str	Sal	M2	M3	Sal	M2	M3	Sal	M2	M3
U[c]	U	U	U	4	2	5	0	0	0	0	0	0
U	U	U	D	0	0	0	0	0	0	0	0	0
U	U	D	U	2	3	0	0	0	0	0	0	0
U	D	U	U	0	1	0	0	0	0	0	0	0
D	U	U	U	2	0	2	0	0	0	0	0	0
U	U	D	D	1	0	0	0	0	0	0	0	0
U	D	U	D	0	0	0	0	0	0	0	0	0
U	D	D	U	0	0	0	0	0	0	0	0	0
D	U	U	D	0	1	0	0	0	0	0	0	0
D	U	D	U	0	0	1	0	0	0	0	0	0
D	D	U	U	0	0	0	0	0	0	0	0	0
U	D	D	D	1	0	0	0	0	0	1	0	0
D	U	D	D	0	1	1	0	0	0	0	0	0
D	D	U	D	0	1	0	0	0	0	0	0	0
D	D	D	U	0	1	0	0	0	0	0	0	0
D	D	D	D	6	23	4	0	1	1	1	1	0

[a]Distribution of four different brain regions. Cin, Hip, Na, Str indicate cingulate cortex, hippocampus, nucleus accumbens, striatum, respectively.
[b]Number of genes searched by GEDA in each of three different CPP groups. Sal, M2, M3 stand for Saline, MAP2, MAP3 group, respectively.
[c]U is up-regulated and D is down regulated genes.

of U and D for each of three CPP groups (Table 2). In this search, we could find several genes differently expressed among the four brain regions. Current literature searched through PubMed did not contain any evidence that the selected genes are related to drug abuse or dependency. Further experimental study remains to confirm the association of the identified genes on drug addition.

Discussion

We developed a useful interface and data-mining tool, GEDA for gene expression data analysis in drug addiction study. The GEDA DB is for the standardization of oligoneucleotide microarray data of rat brain in drug addiction and stored them in a single integrated database system, focusing on more

effective data processing and interpretation. We maximized the usability of this DB, helping users study significant gene expression and identify biological function of the genes through integrating the latest gene information such as GO annotation and metabolic pathway. For collecting the latest information of selected gene from GEDA DB, we also set up the local BLAST search engine and NR sequence database updated by NCBI server on a daily basis. It is an effective and user-friendly DB for analyzing concerned data on website, specifically for finding out the genes related to drug addiction. We applied this system to the identification and characterization of methamphetamine-induced genes' behavior in rat brain. It might be said that GEDA DB have a similar format with the search by expression ratio of Fei *et al.* (2006). However, GEDA has a unique feature of local BLAST search through keyword inputting. Actually we found GEDA DB to be a very useful tool to find genes related to keywords in our drug addiction experiment.

GEDA is a flexible platform to serve as a foundation of our on-going high throughput system studies using DNA microarrays. It aims for accelerating the procedure of selecting candidate genes for experimental characterization. Our work will integrate various biological informations, being useful to understand and study the biological materials systematically. Through developing an effective DB system, we are planning to maximize the application of our DB by helping our users confirm correctly and easily the formation of significantly expressed genes and their biological functions.

The GEDA DB available now is very useful instrument in a way, but it requires correction and supplementary measures to enhance structural flexibility and format standardization. We are aiming to get confidence from our users by providing them with a system of fast run-time for processing and analyzing complex microarray gene expression data. We plan to do more replicates of microarray experiments in each of groups. For the repeated data we are going to add various statistical analysis methods on web-module, such as pairwise T test and ANOVA test. At the same time, we plan to do cluster analysis on website, visualizing the results. It is expected that specific gene expression can be checked by means of various statistical methods, and biological functions are more correctly and easily confirmed. Besides that, we will present MIAME (minimum information about a microarray experiment) formatted information of microarray data in our system, enabling more clearly interpretation of the data. This DB will play an important role to find out genes related drug abuse/dependency by the public standardization.

References

Ahmed, S. H., Lutjens, R., van der Stap, L. D., Lekic, D., Romano-Spica, V., Morales, M., Koob, G. F., Repunte-Canonigo, V. and Sanna, P. P. (2005) Gene expression evidence for remodeling of lateral hypothalamic circuitry in cocaine addiction. *Proc. Natl. Acad. Sci. USA* **102**, 11533-11538.

Alberts, R., Fu, J., Swertz, M. A., Lubbers, L. A., Albers, C. J. and Jansen, R. C. (2005) Combining microarrays and genetic analysis. *Brief Bioinform.* **6**, 135-145.

Albertson, D. N., Pruetz, B., Schmidt, C. J., Kuhn, D. M., Kapatos, G. and Bannon, M. J. (2004) Gene expression profile of the nucleus accumbens of human cocaine abusers: evidence for dysregulation of myelin. *J. Neurochem.* **88**, 1211-1219.

Arlinde, C., Sommer, W., Bjork, K., Reimers, M., Hyytia, P., Kiianmaa, K. and Heilig, M. (2004) A cluster of differentially expressed signal transduction genes identified by microarray analysis in a rat genetic model of alcoholism. *Pharmacogenomics J.* **4**, 208-218.

Bahi, A. and Dreyer, J. L. (2005) Cocaine-induced expression changes of axon guidance molecules in the adult rat brain. *Mol. Cell. Neurosci.* **28**, 275-291.

Burgoon, L. D., Boutros, P. C., Dere, E. and Zacharewski, T. R. (2006) dbZach: A MIAME-compliant toxicogenomic supportive relational database. *Toxicol. Sci.* **90**, 558-568.

Everitt, B. J. and Robbins, T. W. (2005) Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* **8**, 1481-1489.

Fei, Z., Tang, X., Alba, R. and Giovannoni, J. (2006) Tomato Expression Database (TED): a suite of data presentation and analysis tools. *Nucl. Acids Res.* **34**, 766-770.

Funada, M., Zhou, X., Satoh, M. and Wada, K. (2004) Profiling of methamphetamine-induced modifications of gene expression patterns in the mouse brain. *Ann. N. Y. Acad. Sci.* **1025**, 76-83.

Hawes, J. and Piccotto, M. R. (2004) Characterizatoin of GalR1, GalR2, and GalR3 immunoreactivity in catecholaminergic nuclei of the mouse brain. *J. Comp. Neurol.* **479**, 410-423.

Ishigami, A., Tokunaga, I., Gotohda, T. and Kubo, S. (2003) Immunohistochemical study of myoglobin and oxidative injury related markers in the kidney of methamphetamine abusers. *Leg. Med. (Tokyo)* **5**, 42-48.

Iwamoto, K., Bundo, M., Yamamoto, M., Ozawa, H., Saito, T. and Kato, T. (2004) Decreased expression of NEFH and PCP4/PEP19 in the prefrontal cortex of alcoholics. *Neurosci. Res.* **49**, 379-385.

Klag, S., O'Callaghan, F. and Creed, P. (2005) The use of legal coercion in the treatment of substance abusers: an overview and critical analysis of thirty years of research. *Subst. Use Misuse.* **40**, 1777-1795.

Leibowitz, S. F., Avena, N. M., Chang, G. Q., Karatayev, O., Chau, D. T. and Hoebel, B. G. (2003) Ethanol intake increases galanin mRNA in the hypothalamus and withdrawal decreases it. *Physiol. Behav.* **79**, 103-111.

McClung, C. A. and Nestler, E. J. (2003) Regulation of gene expression and cocaine reward by CREB and DeltaFosB. *Nat. Neurosci.* **6**, 1208-1215.

Miyamoto, Y., Yamada, K., Nagai, T., Mori, H., Mishina, M., Furukawa, H., Noda, Y. and Nabeshima, T. (2004) Behavioural adaptations to addictive drugs in mice lacking the NMDA

receptor epsilon1 subunit. *Eur. J. Neurosci.* **19**, 151-158.

Nestler, E. J. (2001) Molecular basis of long-term plasticity underlying addiction. *Nat. Rev. Neurosci.* **2**, 119-128.

Noailles, P. A., Becker, K. G., Wood, W. H. 3rd, Teichberg, D. and Cadet, J. L. (2003) Methamphetamine-induced gene expression profiles in the striatum of male rat pups exposed to the drug in utero. *Brain Res. Dev. Brain Res.* **147**, 153-162.

Picciotto, M. R., Hawes, J. J., Brunzell, D. H. and Zachariou, V. (2005) Galanin can attenuate opiate reinforcement and withdrawal. *Neuropeptides* **39**, 313-315.

Rouse, R. and Hardiman, G. (2003) Microarray technology-- intellectual property retrospective. *Pharmacogenomics* **4**, 623-632.

Sommer, W., Arlinde, C. and Heilig, M. (2005) The search for candidate genes of alcoholism: evidence from expression profiling studies. *Addict. Biol.* **10**, 71-79.

Thomas, D. M., Francescutti-Verbeem, D. M., Liu, X. and Kuhn, D. M. (2004) Identification of differentially regulated transcripts in mouse striatum following methamphetamine treatment--an oligonucleotide microarray approach. *J. Neurochem.* **88**, 380-393.

Yamada, K., Nagai, T. and Nabeshima, T. (2005) Drug dependence, synaptic plasticity, and tissue plasminogen activator. *J. Pharmacol. Sci.* **97**, 157-161.

Yamamoto, H., Imai, K., Takamatsu, Y., Kamegaya, E., Kishida, M., Hagino, Y., Hara, Y., Shimada, K., Yamamoto, T., Sora, I., Koga, H. and Ikeda, K. (2005) Methamphetamine modulation of gene expression in the brain: analysis using customized cDNA microarray system with the mouse homologues of KIAA genes. *Brain Res. Mol. Brain Res.* **137**, 40-46.

Zhao, R. J., Woo, R. S., Jeong, M. S., Shin, B. S., Kim, D. G. and Kim, K. W. (2003) Orphanin FQ/nociceptin blocks methamphetamine place preference in rats. *Neuroreport.* **14**, 2383-2385.