Serial Analysis of Gene Expression: Applications in Malaria Parasite, Yeast, Plant, and Animal Studies

Renu Tuteja* and Narendra Tuteja

International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110067, India

The serial analysis of gene expression (SAGE) method is based on the isolation of unique sequence tags from individual transcripts and concatenation of tags serially into long DNA molecules. SAGE is an innovative technique that offers the potential of cataloging both the identity and relative frequencies of mRNA transcripts in a given RNA preparation. It can quantify low-abundance transcripts and reliably detect relatively small differences in transcript abundance between cell populations. SAGE data can be used to complement studies in cases where other gene expression methods may be more convenient or efficient. SAGE can be used in a wide variety of applications to identify disease-related genes, to analyze the effect of drugs on tissues, and to provide insights into the disease pathways. The most important application of SAGE is the identification of differentially expressed genes. In this review, we describe various applications of this powerful technology in malarial parasite, yeast, plant, and animal systems.

INTRODUCTION

The SAGE method is a highly competent technology that can give a global gene expression profile of a particular type of cell or tissue and also help in identifying a set of specific genes to the cellular conditions by comparing the profiles constructed for a pair of cells that are kept at different conditions [1, 2, 3, 4]. Since the discovery of SAGE, for several years now, it has been used to provide a comprehensive analysis of a variety of different tissue samples, each usually consisting of millions of cells. The approach has been extended recently to permit analysis of gene expression in substantially fewer cells, thereby allowing analysis of heterogeneous tissues or microanatomical structures. SAGE data can also be used to complement studies in cases where other gene expression methods may be more convenient or efficient. The SAGE technique is mainly based on two principles. (1) A short oligonucleotide sequence tag (10–11 base pairs) contains sufficient information to uniquely identify a transcript [3]. These tags are used to identify genes and relative abundance of their transcripts within mRNA. (2) Concatenation of short sequence tags allows the efficient analysis of transcripts in a serial manner since SAGE uses serial processing such that 25–50 SAGE tags are analyzed on each lane of DNA sequencer. The resulting sequence data are analyzed to identify each gene expressed in the cell and the levels at which each gene is expressed [4]. This information forms a library that can be used to analyze the differences in gene expression between cells. The frequency of each SAGE tag in the cloned multimers directly reflects the transcript abundance. Therefore, SAGE results in an accurate picture of gene expression at both the qualitative and the quantitative levels. SAGE technology has been used in a variety of cell lines and in many systems. The following sections describe the significant studies performed in malarial parasite, yeast, plant, and animal systems.

SAGE STUDIES IN MALARIA PARASITE

SAGE is particularly well suited for malarial systems, as the genomes of *Plasmodium* species remain to be fully annotated. By simultaneously and quantitatively analyzing mRNA transcript profiles from a given cell population, SAGE allows for the discovery of new genes. The successful application of SAGE in *Plasmodium falciparum*, 3D7 strain parasites, from which a preliminary library of 6880 tags corresponding to 4146 different genes was generated, has been reported recently [5]. It was demonstrated that *Plasmodium falciparum* is amenable to this technique, despite the remarkably high A-T content of its genome. SAGE tags as short as 10 nucleotides were sufficient to uniquely identify parasite transcripts from both nuclear and mitochondrial genomes. Moreover, the skewed A-T content of parasite sequence did not preclude the use of enzymes that are crucial for generating representative SAGE libraries. Finally, a few modifications

to DNA extraction and cloning steps of the SAGE protocol proved useful for circumventing specific problems presented by A-T rich genomes [5]. In a related study, SAGE was applied to the malarial parasite *Plasmodium falciparum* to characterize the comprehensive transcriptional profile of erythrocytic stages [6].

A SAGE library of approximately 8335 tags representing 4866 different genes was generated from 3D7 strain parasites. Basic local alignment search tool analysis of high abundance SAGE tags revealed that a majority (88%) corresponded to 3D7 sequence, and despite the low complexity of the genome, 70% of these highly abundant tags matched unique loci. Characterization of these suggested the major metabolic pathways that are used by the organism under normal culture conditions. Furthermore, several tags expressed at high abundance (30% of tags matching unique loci of the 3D7 genome) were derived from previously uncharacterized open reading frames, demonstrating the use of SAGE in genome annotation [6]. The open platform "profiling" nature of SAGE also leads to the important discovery of a novel transcriptional phenomenon in the malarial pathogen: a significant number of highly abundant tags that were derived from annotated genes (17%) corresponded to antisense transcripts. These SAGE data were validated by two independent means: strand specific RT-PCR and Northern analysis, where antisense messages were detected in both asexual and sexual stages [6].

SAGE IN YEAST STUDIES

SAGE analysis has been successfully applied for transcript profiling in yeast [7]. Of the genes identified in yeast, 1981 genes had known functions while other 2684 genes were previously uncharacterized. The integration of positional information with gene expression data allowed for the generation of chromosomal expression maps identifying physical regions of transcriptional activity and also identified genes that had not been predicted by sequence information alone [7]. A genome-wide characterization of mRNA transcript levels in yeast grown on the fatty acid oleate has been determined using SAGE [8]. Comparison of this SAGE library with that reported for glucose-grown cells revealed the dramatic adaptive response of yeast to a change in carbon source. In oleate-grown cells, this was exemplified by the huge increase of mRNAs encoding the peroxisomal beta-oxidation enzymes required for degradation of fatty acids. The data provide evidence for the existence of redox shuttles across organellar membranes that involve peroxisomal, cytoplasmic, and mitochondrial enzymes. Induction of genes under the immediate control of these factors was abolished; other genes were upregulated, indicating an adaptive response to the changed metabolism imposed by the genetic impairment [8]. Analysis of global gene expression in *Saccharomyces cerevisiae* by the SAGE technique has permitted the identification of at least 302 previously unidentified transcripts

from nonannotated open reading frames (NORFs) [9]. Transcription of one of these, *NORF5/HUG1*, is induced by DNA damage, and this induction requires *MEC1*, a homologue of the ataxia telangiectasia mutated (ATM) gene. *HUG1* is the first example of a NORF with important biological functional properties and defines a novel component of the MEC1 checkpoint pathway [9]. In a recent study, 10 genome expression data sets have been analyzed by large-scale cross-referencing against broad structural and functional categories [10]. This analysis enabled to determine features more prevalent in the transcriptome than the genome, that is, those that are common to highly expressed proteins. Starting with simplest categories, it has been found that, relative to the genome, the transcriptome is enriched in alanine and glycine and depleted in asparagine and very long proteins. In particular, some enzymatic folds, such as the TIM barrel and the G3P dehydrogenase folds, are much more prevalent in the transcriptome than the genome, whereas others, such as the protein-kinase and leucine-zipper folds, are depleted. Furthermore, for a given functional category, transcriptome enrichment varies quite substantially between the different expression data sets, with a variation an order of magnitude larger than for the other categories cross-referenced (eg, amino acids) [10].

SAGE IN PLANT STUDIES

SAGE was applied for profiling expressed genes in rice seedlings (Oryza sativa L.) [11]. Only 1367 genes (23.1%) matched the rice cDNA or EST sequences in the DNA database. SAGE showed that most of the highly expressed genes in rice seedlings belong to the category of housekeeping genes. Unexpectedly, the most highly expressed gene in rice seedlings was a metallothionein (MT) gene, and together with three other messages for MT, it accounts for 2.7% of total gene expression. SAGE was also applied to identify differentially expressed genes between anaerobically treated and untreated rice seedlings. In combination with microarray analysis, SAGE serves as a highly efficient tool for the identification and isolation of differentially expressed genes in plant [11]. The global gene expression patterns of *Arabidopsis* pollen using SAGE were characterized recently [12]. It was interesting to note that the number of unique tags in pollen was low compared with the SAGE library of the leaf constructed on a similar scale. Functional classification of the expressed genes reveals that those involved in cellular biogenesis such as polygalacturonase, pectate lyase, and pectin methylesterase make up more than 40% of the total transcripts. The expression level of the great majority of transcripts was unaffected by cold treatment at $0°C$ for 72 hours, whereas pollen tube growth and seed production were substantially reduced. These results strongly suggest that poor accumulation of proteins that play a role in stress tolerance may be why *Arabidopsis* pollen is cold-sensitive [12].

SAGE IN ANIMAL STUDIES

To characterize gene expression in activated mast cells more comprehensively, the changes in genetic transcripts were surveyed by the method of SAGE in the RBL-2H3 line of rat mast cells before and after they were stimulated through their receptors with high affinity for immunoglobulin E (FcepsilonRI) [13]. Among the diverse genes that had not been previously associated with mast cells and that were constitutively expressed were those for the cytokine macrophage migration inhibitory factor neurohormone receptors such as growth hormone-releasing factor and melatonin and components of the exocytotic machinery. In addition, several dozen transcripts were differentially expressed in response to antigen-induced clustering of the FcepsilonRI. Included among these were the genes for preprorelaxin, mitogen-activated protein kinase kinase 3, and the dual specificity protein phosphatase, rVH6 [13].

SAGE method was used to systematically analyze transcripts present in a microglial cell line [14]. Among the diverse transcripts that had not been previously detected in microglia were those for cytokines, such as endothelial monocyte-activating polypeptide I (EMAP I), and for cell surface antigens, including adhesion molecules such as CD9, CD53, CD107a, CD147, CD162 and mast cell high affinity IgE receptor. In addition, transcripts that were characteristic of hematopoietic cells or mesodermal structures, such as E3 protein, A1, EN-7, B94, and ufo were also detected. Furthermore, the profile contained a transcript, Hn1, that is important in hematopoietic cells and neurological development suggesting the probable neural differentiation of microglia from the hematopoietic system in development. mRNA expression of these genes was confirmed by RT-PCR in primary cultures of microglia [14]. The mouse *Otx2* gene is a homeobox transcription factor required as early as gastrulation for the proper development of the head. The gene expression profiles were compared in wild-type and *Otx2(-/-)* 6.5-day post-coitum embryos by using a SAGE assay adapted to microdissected structures [15]. Chronic renal disease initiation and progression remain incompletely understood. Using SAGE, a tag expression library from ROP-+/+ mouse kidney has been constructed [16]. Tag sequences were sorted by abundance, and identity was determined by sequence homology searching. Previously characterized transcripts were clustered into functional groups, and those encoding metabolic enzymes, plasma membrane proteins (transporters/receptors), and ribosomal proteins were most abundant. The most common, kidney-specific transcripts were kidney androgen-regulated protein, sodium-phosphate cotransporter, renal cytochrome P-450, parathyroid hormone receptor, and kidney-specific cadherin [16]. In a recent study, the transcriptome of a highly differentiated mouse clonal cortical collecting duct (CCD) principal cell line (mp-kCCD(cl4)) and the changes in the transcriptome in-

duced by aldosterone and vasopressin have been analyzed [17]. SAGE was performed on untreated cells and on cells treated with either aldosterone or vasopressin for 4 hours. Statistical comparison of the three SAGE libraries revealed 34 AITs (aldosterone-induced transcripts), 29 ARTs (aldosterone-repressed transcripts), 48 VITs (vasopressin-induced transcripts), and 11 VRTs (vasopressin-repressed transcripts). A selection of the differentially expressed, hormone-specific transcripts (5 VITs, 2 AITs, and 1 ART) has been validated in the mp-kCCD(cl4) cell line either by northern blot hybridization or reverse transcription-PCR. The hepatocyte nuclear transcription factor HNF-3-alpha (VIT39), the receptor activity modifying protein RAMP3 (VIT48), and the glucocorticoid-induced leucine zipper protein (GILZ) (AIT28) are candidate proteins playing a role in physiological responses of this cell line to vasopressin and aldosterone [17].

The development of cardiovascular diseases such as heart failure involves functional changes that are beneficial short-term, but may be fatal long-term. In a recent study, the current state of genomic research for determination of the transcriptome by the first limited SAGE analysis of rodent heart gene expression has been described [18]. It has also been discussed that how these results generated with this approach can be applied to the study and treatment of cardiovascular diseases [18]. Molecular inventories of the developing mouse neocortex before and after birth were generated using the global gene expression profiling tool SAGE [19]. The libraries were generated from embryonic day 15 and postnatal day 1 mouse neocortices. The differentially expressed transcripts included genes known to be important in neocortical development (eg, brain factor 1, neuroD2, and Id2), genes not previously associated with neocortical development (such as brahma-related gene 1, receptor for activated C-kinase I, hypermethylated in cancer 2, and Evi9), and genes of unknown identity or function [19]. SAGE was applied to study differentially expressed genes in mouse brain 14 hours after the induction of focal cerebral ischemia [20]. Metallothionein-II (MT-II) was the most significantly upregulated transcript in the ischemic hemisphere. MT-I and MT-II are induced by metals, glucocorticoids, and inflammatory signals in a coordinated manner, yet their function remains elusive. MT-I- and MT-II-deficient mice developed approximately threefold larger infarcts than wild-type mice and a significantly worse neurological outcome [20]. FTL-1, -3, and -10 are three murine day 14 fetal thymocyte cell lines produced in order to model developmental stages within early (CD3-CD4-CD8) thymocyte differentiation. In a recent study, the SAGE method was used to perform a systematic analysis of transcripts present in these cell lines [21]. Differentially expressed mRNA transcripts representing different gene classes were identified, including T cell functional genes, cytokine receptors, adhesion molecules, and transcription factors. Expression of the

transcription factors RUNX2 and PHD finger protein 2 and of the IGF type 1 receptor was shown to have differentially regulated expression patterns in sorted DN1-4 cells. These genes, and others identified by this analysis, are likely to play important roles in the development of T cells [21]. In order to identify genes developmentally regulated in the somatic cells of the testis, SAGE has been used to generate gene expression profiles from these cells in the fetal and adult mouse testes [22]. To avoid germ cell transcripts, a fetal SAGE library was generated from germ cell-free fetal W(v)/W(v) mice and an adult SAGE library from adult testes depleted of germ cells with busulfan. The differentially regulated genes are likely to provide insight into mechanisms regulating testis function both during development and in the adult animal [22].

SAGE technology has been utilized to contrast the differential gene expression profile in rat embryo fibroblast cells producing temperature-sensitive p53 tumor suppressor protein at permissive or nonpermissive temperatures [23]. Analysis of approximately 15 000 genes revealed that the expression of 14 genes was dependent on functional p53 protein, whereas the expression of 3 genes was significantly higher in cells producing nonfunctional p53 protein. Those genes whose expression was increased by functional p53 include *RAS, U6 snRNA*, cyclin *G, EGR-1*, and several novel genes. The expression of actin, tubulin, and HSP70 genes was elevated at the nonpermissive temperature for p53 function. Interestingly, the expression of several genes was dependent on a non-temperature-sensitive mutant p53 suggesting altered transcription profiles dependent on specific p53 mutant proteins. These results demonstrate the utility of SAGE for rapidly and reproducibly evaluating global transcriptional responses within different cell populations [23].

Kringle domain, a triple-disulfide-linked domain, is conserved in diverse proteins which play important roles in various biological processes. Kremen, a novel member of kringle-containing proteins, has been cloned using a newly developed unique strategy, Kringle-SAGE, which enables comprehensive analysis of kringle-containing proteins [24]. Kremen is likely to be a type-I transmembrane protein composed of 473 amino acid residues. Kremen has a kringle domain, a WSC domain, and CUB domains in the extracellular region, while the intracellular region has no conserved motif involved in signal transduction. In the mouse embryo, the Kremen mRNA level, which was increased during embryonic development, was localized in the apical ectodermal ridge of limb buds, myotome, and sensory organs (eg, optic vesicle, otic vesicle, and nasal pit). In the adult mouse, Kremen mRNA was expressed in a variety of tissues with a relatively strong expression in the lung, heart, and skeletal muscle. Kremen mRNA expression in C2C12 and NIE-115 cells increased during respective differentiation into muscular and neural cells. These results suggest a potential role for Kremen in the regulation of cellular responses upon extracel-

lular stimulus or cell-cell interaction in neuronal and/or muscle cells. Kringle-SAGE is expected to facilitate further elucidation of structure and functions of kringle proteins [24].

To elucidate the molecular basis of muscle atrophy, the SAGE method has been performed with control and immobilized muscles of 10 rats [25]. The genes that expressed greater than 0.5% in muscle are involved in the following three functions: (1) contraction (troponin I, C, and T; myosin light chain 1–3; actin; tropomyosin; and parvalbumin), (2) energy metabolism (cytochrome c oxidase I and III, creatine kinase, glyceraldehyde-3-phosphate-dehydrogenase, phosphoglycerate mutase, ATPase 6, and aldolase A), and (3) housekeeping (lens epithelial protein). Muscle atrophy appears to be caused by changes in mRNA levels of specific regulators of proteolysis, protein synthesis, and contractile apparatus assembling, such as polyubiquitin, elongation factor 2, and nebulin. Immobilization has produced a decrease more than threefold in gene expression of enzymes involved in energy metabolism, especially ATPase, cytochrome c oxidase, NADH dehydrogenase, and protein phosphatase 1. Differential gene expressions of selenoprotein W and uroporphyrinogen decarboxylase, which can be involved in oxidative stress, were also observed. Other genes with various functions, such as cholesterol metabolism and growth factors, were also differentially expressed. Moreover, novel genes regulated by immobilization were discovered. Thus, this study allows a better understanding of global muscle characteristics and the molecular mechanisms of sedentarity and sarcopenia [25].

Using the SAGE method, a gene expression profile of the rat hippocampus was generated [26]. A total of 76 790 SAGE tags were analyzed, allowing identification of 28 748 different tag species, each representing a unique mRNA transcript. The tags were divided into different abundancy classes, ranging from tags that were detected over 500 times to tags encountered only once in the 76 790 tags analyzed. The mRNA species detected more than 50 times represented 0.3% of the total number of unique tags while accounting for 22% of the total hippocampal mRNA mass. The majority of tags was encountered 5 times or less. The genes expressed at the highest levels were of mitochondrial origin, consistent with a high requirement for energy in neuronal tissue. At a lower level of expression, several neuron-specific transcripts were encountered, encoding various neurotransmitter receptors, transporters, and enzymes involved in neurotransmitter synthesis and turnover, ion channels and pumps, and synaptic components. Comparison of relative expression levels demonstrated that glutamate receptors are the most frequent neurotransmitter receptors expressed in the hippocampus, consistent with the important role of glutamatergic neurotransmission in the hippocampus, while GABA receptors were present at approximately ten-fold lower levels. Several kinases were present including CaMKII, which was expressed at high

levels, consistent with being the most abundant protein in the spines of hippocampal pyramidal cells [26]. Adrenal corticosteroids (CORT) have a profound effect on the function of the hippocampus. This is mediated in a coordinated manner by mineralocorticoid (MR) and glucocorticoid (GR) receptors via activation or repression of target genes. Using SAGE, CORT-responsive hippocampal genes regulated via MR and/or GR have been identified in a recent study [27]. SAGE profiles were compared under different conditions of CORT exposure, resulting in identification of 203 CORT-responsive genes that are involved in many different cellular processes like energy expenditure and cellular metabolism; protein synthesis and turnover; signal transduction, neuronal connectivity, and neurotransmission. In situ, hybridization revealed that six randomly chosen CORT-responsive genes had distinct expression patterns in neurons of the hippocampus. In addition, using in situ hybridization, it was confirmed that these six genes were indeed regulated by CORT, underscoring the validity of the SAGE data. Comparison of MR- and GR-dependent expression profiles revealed that the majority of the CORT-responsive genes was regulated either by activated MR or by activated GR, while only a few genes were responsive to both activated MR and GR. This indicates that the molecular basis for the differential effects of activated MR and GR is activation or repression of distinct, yet partially overlapping sets of genes. The putative CORT-responsive genes identified in this study will provide insight into the molecular mechanisms underlying the differential and sometimes opposing effects of MR and GR on neuronal excitability, memory formation and behaviour as well as their role in neuronal protection and damage [27].

Intraepithelial lymphocytes (IELs) are abundant, evolutionarily conserved T cells, commonly enriched in T cell receptor (TCR) gamma delta expression. However, their primary functional potential and constitutive activation state are incompletely understood. To address this, SAGE was applied to murine TCR gamma delta+ and TCR alpha beta+ intestinal IELs directly *ex vivo*, identifying 15, 574 unique transcripts that collectively portray an "activated yet resting," Th1-skewed, cytolytic, and immunoregulatory phenotype applicable to multiple subsets of gut IELs [28]. Expression of granzymes, Fas ligand, RANTES, prothymosin beta4, junB, RGS1, Btg1, and related molecules is high, whereas expression of conventional cytokines and high-affinity cytokine receptors is low. Differentially expressed genes readily identify heterogeneity among TCR alpha beta+ IELs, whereas differences between resident TCR gamma delta+ IELs and TCR alpha beta+ IELs are less obvious [28].

Although extraocular muscle (EOM) is a skeletal muscle, aspects of its biology are unlike other striated muscles. In a recent study, the broad molecular genetics profile underlying the novel EOM phenotype was examined [29]. SAGE was used to quantify adult rat EOM gene transcripts. SAGE isolates and sequences 10-bp tags from defined locations in mRNA-derived cDNA. Tag sequence-location was used to extract transcript identity from a curated SAGE database, and detection frequencies reflected abundance of corresponding mRNAs. Of the unique tags, 7.8% were detected at high to intermediate levels, 19.3% at lower levels, and 72.9% as single copies; 40% of the tags matched known expressed sequence tags (ESTs), most of which (85.7%) represented a unique EST. Tags without matches in the SAGE database and those expressed as single copies only were not considered further. SAGE tags expressed at more than 0.1% of total transcripts reflected several aspects of muscle biology, including sarcomeric structure, energy metabolism, and ribosomal protein expression. Genes highly expressed in EOM were compared with other existing muscle expression databases to identify conserved and novel patterns in EOM. These data provide a normative gene expression database and a novel molecular signature that will facilitate the study of EOM development and function and of the mechanisms behind its preferential targeting or sparing in neuromuscular disease [29].

CONCLUSIONS

Progress in large-scale cDNA analysis (EST analysis) in many organisms is a prerequisite for the useful application of SAGE, as the annotation of SAGE tags is based on preexisting EST databases. The uniqueness of SAGE is that it allows transcript profiles to be given as digital data. Accordingly, they become suitable for the construction of gene expression databases on computer networks. Yeast and cancer transcriptome databases based on SAGE are already accessible via the internet. In the organisms where transcript data are limited, SAGE may initially be the most efficient method of identifying new or differentially expressed genes. The SAGE data analysis could also be used as reference data for the relative expression data obtained by hybridization experiments on cDNA arrays, and may ultimately allow comparison of array data between different experiments in different laboratories.

SAGE is also used as a primary discovery engine that can characterize human diseases at the molecular level while illuminating potential targets and markers for therapeutic and diagnostic developments respectively. The ability of SAGE to define specific transcriptomes will aid in the development of gene therapies whereby cell- or tissue- specific promoters and genes can be utilized to appropriately express and deliver a given therapy. In general, SAGE alone or in combination with proteomic approaches can accelerate the identification of high-quality drug targets which could be a next generation of therapeutic products. SAGE, along with other methods, should yield valuable information about the fundamental biology and virulence mechanism of an important plant or human pathogen. In combination with microarray analysis, SAGE should serve as a highly efficient tool for the identification and isolation of differentially expressed genes in plants and animals.

REFERENCES

[1] Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science*. 1995;270(5235):484–487.

[2] Bartlett J. Technology evaluation: SAGE, Genzyme molecular oncology. *Curr Opin Mol Ther*. 2001;3(1):85–96.

[3] Polyak K, Xia Y, Zweier JL, Kinzler KW, Vogelstein B. A model for p53-induced apoptosis. *Nature*. 1997;389(6648):300–305.

[4] Yamamoto M, Wakatsuki T, Hada A, Ryo A. Use of serial analysis of gene expression (SAGE) technology. *J Immunol Methods*. 2001;250(1-2):45–66.

[5] Munasinghe A, Patankar S, Cook BP, et al. Serial analysis of gene expression (SAGE) in *Plasmodium falciparum*: application of the technique to A-T rich genomes. *Mol Biochem Parasitol*. 2001;113(1): 23–34.

[6] Patankar S, Munasinghe A, Shoaibi A, Cummings LM, Wirth DF. Serial analysis of gene expression in *Plasmodium falciparum* reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. *Mol Biol Cell*. 2001;12(10):3114–3125.

[7] Velculescu VE, Zhang L, Zhou W, et al. Characterization of the yeast transcriptome. *Cell*. 1997;88(2): 243–251.

[8] Kal AJ, van Zonneveld AJ, Benes V, et al. Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. *Mol Biol Cell*. 1999;10(6):1859–1872.

[9] Basrai MA, Velculescu VE, Kinzler KW, Hieter P. NORF5/HUG1 is a component of the MEC1-mediated checkpoint response to DNA damage and replication arrest in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 1999;19(10):7041–7049.

[10] Jansen R, Gerstein M. Analysis of the yeast transcriptome with structural and functional categories: characterizing highly expressed proteins. *Nucleic Acids Res*. 2000;28(6):1481–1488.

[11] Matsumura H, Nirasawa S, Terauchi R. Technical advance: transcript profiling in rice (*Oryza sativa L.*) seedlings using serial analysis of gene expression (SAGE). *Plant J*. 1999;20(6):719–726.

[12] Lee JY, Lee DH. Use of serial analysis of gene expression technology to reveal changes in gene expression in *Arabidopsis* pollen undergoing cold stress. *Plant Physiol*. 2003;132(2):517–529.

[13] Chen H, Centola M, Altschul SF, Metzger H. Characterization of gene expression in resting and activated mast cells. *J Exp Med*. 1998;188(9):1657–1668.

[14] Inoue H, Sawada M, Ryo A, et al. Serial analysis of gene expression in a microglial cell line. *Glia*. 1999;28(3):265–271.

[15] Zakin L, Reversade B, Virlon B, et al. Gene expression profiles in normal and Otx2-/- early gas-

trulating mouse embryos. *Proc Natl Acad Sci USA*. 2000;97(26):14388–14393.

[16] El-Meanawy MA, Schelling JR, Pozuelo F, et al. Use of serial analysis of gene expression to generate kidney expression libraries. *Am J Physiol Renal Physiol*. 2000;279(2):F383–F392.

[17] Robert-Nicoud M, Flahaut M, Elalouf JM, et al. Transcriptome of a mouse kidney cortical collecting duct cell line: effects of aldosterone and vasopressin. *Proc Natl Acad Sci USA*. 2001;98(5):2712–2716.

[18] Anisimov SV, Lakatta EG, Boheler KR. Discovering altered genomic expression patterns in heart: transcriptome determination by serial analysis of gene expression. *Eur J Heart Fail*. 2001;3(3):271–281.

[19] Gunnersen JM, Augustine C, Spirkoska V, Kim M, Brown M, Tan SS. Global analysis of gene expression patterns in developing mouse neocortex using serial analysis of gene expression. *Mol Cell Neurosci*. 2002;19(4):560–573.

[20] Trendelenburg G, Prass K, Priller J, et al. Serial analysis of gene expression identifies metallothionein-II as major neuroprotective gene in mouse focal cerebral ischemia. *J Neurosci*. 2002;22(14):5879–5888.

[21] Zhao FQ, Sheng ZM, Tsai MM, et al. Serial analysis of gene expression in murine fetal thymocyte cell lines. *Int Immunol*. 2002;14(12):1383–1395.

[22] O'Shaughnessy PJ, Fleming L, Baker PJ, Jackson G, Johnston H. Identification of developmentally regulated genes in the somatic cells of the mouse testis using serial analysis of gene expression. *Biol Reprod*. 2003;69(3):797–808.

[23] Madden SL, Galella EA, Zhu J, Bertelsen AH, Beaudry GA. SAGE transcript profiles for p53-dependent growth regulation. *Oncogene*. 1997;15(9): 1079–1085.

[24] Nakamura T, Aoki S, Kitajima K, Takahashi T, Matsumoto K, Nakamura T. Molecular cloning and characterization of Kremen, a novel kringle-containing transmembrane protein. *Biochim Biophys Acta*. 2001;1518(1-2):63–72.

[25] St-Amand J, Okamura K, Matsumoto K, Shimizu S, Sogawa Y. Characterization of control and immobilized skeletal muscle: an overview from genetic engineering. *FASEB J*. 2001;15(3):684–692.

[26] Datson NA, van der Perk J, de Kloet ER, Vreugdenhil E. Expression profile of 30,000 genes in rat hippocampus using SAGE. *Hippocampus*. 2001;11(4):430–444.

[27] Datson NA, van der Perk J, de Kloet ER, Vreugdenhil E. Identification of corticosteroid-responsive genes in rat hippocampus using serial analysis of gene expression. *Eur J Neurosci*. 2001;14(4):675–689.

[28] Shires J, Theodoridis E, Hayday AC. Biological insights into TCRgammadelta+ and TCRalphabeta+ intraepithelial lymphocytes provided by serial analysis of gene expression (SAGE). *Immunity*. 2001;15(3):419–434.

[29] Cheng G, Porter JD. Transcriptional profile of rat extraocular muscle by serial analysis of gene expression. *Invest Ophthalmol Vis Sci*. 2002;43(4):1048–1058.

* Corresponding author.
E-mail: renu@icgeb.res.in
Fax: +91 11 26162316; Tel: +91 11 26195007