

Serial Analysis of Gene Expression: Applications in Human Studies

Renu Tuteja* and Narendra Tuteja

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

Received 7 August 2003; accepted 13 November 2003

Serial analysis of gene expression (SAGE) is a powerful tool, which provides quantitative and comprehensive expression profile of genes in a given cell population. It works by isolating short fragments of genetic information from the expressed genes that are present in the cell being studied. These short sequences, called SAGE tags, are linked together for efficient sequencing. 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. It does not require a hybridization probe for each transcript and allows new genes to be discovered. This technique has been applied widely in human studies and various SAGE tags/SAGE libraries have been generated from different cells/tissues such as dendritic cells, lung fibroblast cells, oocytes, thyroid tissue, B-cell lymphoma, cultured keratinocytes, muscles, brain tissues, sciatic nerve, cultured Schwann cells, cord blood-derived mast cells, retina, macula, retinal pigment epithelial cells, skin cells, and so forth. In this review we present the updated information on the applications of SAGE technology mainly to human studies.

INTRODUCTION

Almost all the biological events are associated with changes in expression of key genes. During the onset and progression of disease, extensive changes take place in gene expression. By comparing gene expression profiles under different conditions, individual genes or group of genes can be identified that play an important role in a particular signaling cascade or process or in disease etiology. Serial analysis of gene expression (SAGE) method is a highly efficient technology that can give a global gene expression profile of a particular type of cell or tissue [1, 2, 3, 4]. It also helps 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 [2]. SAGE technique works by isolating short fragments of genetic information from the expressed genes that are present in the cell under study. These unique sequence tags (9–10 base pairs in length) are concatenated serially into long DNA molecules for lump-sum sequencing [3]. This serial analysis of many thousands of gene-specific tags allows the simultaneous accumulation of information from genes expressed in the tissue of interest and gives rise to an expression profile of that tissue [3]. These sequencing data are then 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. This technology can be used for elucidation of quantitative gene expression pattern that does not depend on the prior availability of transcript information [1]. The SAGE technique can also be used in a wide variety of applications such as to analyze the effect of drugs on tissues, to identify disease-related genes, and to provide insights into disease pathways. Here we are focusing the applications of SAGE technology in human studies.

SAGE IN HUMAN STUDIES

SAGE technology has been widely used in a number of human studies. Some examples of these studies are described in the following sections.

Circulatory system

Dendritic cells (DCs) are professional antigen-presenting cells in the immune system and can be generated in vitro from hematopoietic progenitor cells in the bone marrow, CD34(+) cord blood cells, precursor cells in the peripheral blood, and blood monocytes by culturing with granulocyte-macrophage colony-stimulating factor (GM-CSF), interleukin-4, and tumor necrosis factor- α . SAGE was performed in DCs derived from human blood monocytes [5]. A total of 58 540 tag sequences from a DC cDNA library represented more than 17 000 different genes, and these data were compared with SAGE analysis of tags from monocytes and GM-CSF-induced macrophages. Many of the genes that were differentially

expressed in DCs were identified as genes encoding proteins related to cell structure and cell motility. The identification of specific genes expressed in human blood monocyte-derived DCs should provide candidate genes to define subsets of, the function of, and the maturation stage of DCs and possibly also to diagnose diseases in which DCs play a significant role, such as autoimmune diseases and neoplasms [5]. In continuation of this study, SAGE was conducted in lipopolysaccharide (LPS)-stimulated mature and activated DCs (MADCs) derived from human blood monocytes [6]. Many of the genes, such as germinal center kinase-related protein kinase, cystatin F, interferon (IFN)-alpha-inducible protein p27, EBI3, HEM45, actin-bundling protein, ELC, DC-LAMP, serine/threonine kinase 4, and several genes in expressed sequence tags, were differentially expressed in MADCs, and those encode proteins related to cell structure, antigen-processing enzymes, chemokines, and IFN-inducible proteins. The profile of MADCs was also compared with that of LPS-stimulated monocytes. The comprehensive identification of specific genes expressed in human IMDCs and MADCs should provide candidate genes to define heterogeneous subsets as well as the function and maturation stage of DCs [6].

To comprehensively analyze the genes involved in B-cell antigen receptor (BCR)-mediated apoptosis, the SAGE has been applied to B-cell lymphoma WEHI231 [7]. Comparison of expression patterns revealed that BCR cross-linking caused coordinate changes in the expression of genes involved in polyamine metabolism. The coordinate expression of the polyamine-related genes was confirmed by semiquantitative RT-PCR analysis. During apoptosis, the genes involved in polyamine biosynthesis were downregulated, whereas those involved in polyamine catabolism were upregulated, suggesting that intracellular polyamines play a role in BCR-mediated apoptosis. The levels of intracellular putrescine, spermidine, and spermine were reduced after BCR cross-linking. These effects were prevented by concurrent CD40 stimulation, which blocked BCR-mediated apoptosis. Furthermore, addition of spermine could repress the BCR-mediated apoptosis by attenuating the mitochondrial membrane potential (Deltapsim) loss and activation of caspase-7 induced by BCR signaling. These findings strongly suggest that polyamine regulation is involved in apoptosis during B-cell clonal deletion [7].

The gene expression profile of human cord blood-derived mast cells (MCs) was investigated using SAGE [8]. By selecting tags that were detected more frequently in MCs than in other tissues, genes characteristic of MCs were enriched. This inventory will be useful to identify novel genes with important functions in MCs [8, 9]. A genome-wide analysis of gene expression in primary human CD15(+) myeloid progenitor cells was performed [10]. By using the SAGE technique, the quantitative information was performed for the expression of 37 519 unique SAGE-tag sequences. Of these unique tags, (i) 25%

were detected at high and intermediate levels, whereas 75% were present as single copies, (ii) 53% of the tags matched known expressed sequences, 34% of which were matched to more than one known expressed sequence, and (iii) 47% of the tags had no matches and represent potentially novel genes. The correct genes were confirmed by application of the generation of longer cDNA fragments from SAGE tags for gene identification (GLGI) technique for high-copy tags with multiple matches. A set of genes known to be important in myeloid differentiation were expressed at various levels and used different spliced forms [10].

Studies in keratinocytes

Keratinocyte gene expression was surveyed by means of SAGE [11]. A total of 25 694 tags derived from expressed mRNA were analyzed in a model for normal differentiation and in a model where cultured keratinocytes were stimulated for a prolonged period of time with tumor necrosis factor-alpha, thus mimicking aberrant differentiation in the context of cutaneous inflammation. SAGE revealed many transcripts derived from unknown genes and a large number of genes that are not expressed in keratinocytes; furthermore, these data provide quantitative information about the relative abundance of transcripts, allowing the identification of differentially expressed genes. A major part of the identified transcripts accounted for genes involved in energy metabolism and protein synthesis. A large proportion of all transcripts corresponded to genes associated with terminal differentiation and barrier formation. Another highly expressed functional group of genes corresponded to proteins involved in host protection such as antimicrobial proteins and proteinase inhibitors. Three of these genes were not expressed in keratinocytes, and some were up regulated after prolonged tumor necrosis factor-alpha exposure. The data on expressed genes in keratinocytes are consistent with the known function of human epidermis and provide a first step to generate a transcriptome of human keratinocytes [11]. In a recent study, cluster analysis was applied on multiple SAGE libraries derived from premalignant epidermal tissue (actinic keratosis), normal human epidermis, and cultured keratinocytes [12]. Two clusters of genes, strongly upregulated in the tumor tissue compared with normal epidermis, were investigated in more detail. The differential expression of genes could be confirmed in actinic keratosis from four patients. Several of these genes have been previously associated with carcinogenesis or are likely to be important on the basis of their presumed function. Automated literature search tools show that a subgroup of these genes is coexpressed in other tissues and is part of an epidermal differentiation gene cluster on chromosome 1q21. It has been predicted that these partitions will lead to biological interpretations that can be relevant for understanding the processes of carcinogenesis and tumor progression [12]. Using SAGE on cultured human keratinocytes, high expression levels

of genes putatively involved in host protection and defense, such as proteinase inhibitors and anti-microbial proteins, have been found [13]. It has been reported that cystatin M/E has a tissue-specific expression pattern in which high expression levels are restricted to the stratum granulosum of normal human skin, the stratum granulosum/spinosum of psoriatic skin, and the secretory coils of eccrine sweat glands [13]. Low expression levels were found in the nasal cavity. It was found that *in vitro*, cystatin M/E expression in cultured keratinocytes is up regulated at the mRNA and protein levels, upon induction of differentiation.

It has been demonstrated that cystatin M/E, which has a putative signal peptide, is indeed a secreted protein and is found *in vitro* in culture supernatant and *in vivo* in human sweat by enzyme-linked immunosorbent assay or western blotting. Cystatin M/E showed moderate inhibition of cathepsin B but was not active against cathepsin C [13]. This analysis was done to evaluate the extent to which the age-related differences in gene expression in murine muscle are also evident in human muscle [14]. RNA extracted from muscle of young (21–24 year) and old men (66–77 year) was studied both by SAGE and by oligonucleotide microarrays. SAGE tags were detected for 61 genes homologous to genes reported to be differentially expressed in young and old murine muscle. For 17 genes, there was evidence for a similar age-related change in expression in muscles of mice and men. For 32 other genes, there was evidence that the effect of age on the level of expression is not the same in mice and men. There was no evidence that older human muscle has increased expression of the stress response genes that are increased in old murine muscle [14]. To study the phenotypic changes in human skin associated with repeated sun exposure at the transcription level, a comparative SAGE of sun-damaged preauricular skin and sun-protected postauricular skin as well as sun-protected epidermis was undertaken [15]. SAGE libraries, containing multiple mRNA-derived tag recombinants, were made to poly(A+)RNA isolated from human postauricular skin and preauricular skin, as well as epidermal nick biopsy samples. 5330 mRNA-derived cDNA tags from the postauricular SAGE library were sequenced and these tag sequences were compared to cDNA sequences identified from 5105 tags analyzed from a preauricular SAGE library. Of the total of 4742 different tags represented in both libraries, 34 tags were found with at least a 4-fold difference of tag abundance between the libraries. Among the mRNAs with altered steady-state levels in sun-damaged skin, those encoding keratin 1, macrophage inhibitory factor, and calmodulin-like skin protein were detected. In addition, a comparison of cDNA sequences identified in the SAGE libraries obtained from the epidermal biopsy samples (5257 cDNA tags) and from both full-thickness skin samples indicated that many genes with altered steady-state transcript levels upon sun exposure were expressed in epidermal keratinocytes [15].

Nervous system

Approximately 6% of human beings harbor an unruptured intracranial aneurysm. Each year in the United States, more than 30000 people suffer a ruptured intracranial aneurysm, resulting in subarachnoid hemorrhage. To characterize the molecular pathology of intra-cranial aneurysm, a global gene expression analysis approach (SAGE-lite) in combination with a novel data-mining approach has been used to perform a high-resolution transcript analysis of a single intracranial aneurysm, obtained from a 3-year-old girl [16]. SAGE-lite provides a detailed molecular snapshot of a single intracranial aneurysm. These data suggest that, at least in this specific case, aneurysmal dilation results in a highly dynamic cellular environment in which extensive wound healing and tissue/extracellular matrix remodeling are taking place. Specifically significant overexpression of genes encoding extracellular matrix components (eg, COL3A1, COL1A1, COL1A2, COL6A1, COL6A2, elastin) and genes involved in extracellular matrix turnover (TIMP-3, OSF-2), cell adhesion and anti-adhesion (SPARC, hevin), cytokinesis (PNUTL2), and cell migration (tetraspanin-5) was observed [16]. Bipolar disorder is a serious brain disease affecting more than a million individuals living in the USA. Epidemiological studies indicate a role for both genetic and environmental factors in the pathogenesis of this disorder [17]. SAGE and RT-PCR were used to identify RNA transcripts, which are differentially expressed in the frontal cortex of brains obtained postmortem from individuals with bipolar disorder compared with other psychiatric and control conditions. Levels of RNA transcripts encoding the serotonin transporter protein and components of the NF-kappaB transcription factor complex are significantly increased in individuals with bipolar disorder compared with unaffected controls [17]. Increased levels of expression of these RNA transcripts were also detected in the brains of some individuals with schizophrenia and unipolar depression [17]. Tay-Sachs and Sandhoff diseases are lysosomal storage disorders characterized by the absence of beta-hexosaminidase activity and the accumulation of GM2 ganglioside in neurons. In each disorder, a virtually identical course of neurodegeneration begins in infancy and leads to demise generally by 4–6 years of age. Through SAGE, gene expression profiles in cerebral cortex from a Tay-Sachs patient, a Sandhoff disease patient, and a pediatric control have been determined [18]. Examination of genes that showed altered expression in both patients revealed molecular details of the pathophysiology of the disorders related to neuronal dysfunction and loss. A large fraction of the elevated genes in the patients could be attributed to activated macrophages/microglia and astrocytes, and included class II histocompatibility antigens, the pro-inflammatory cytokine osteopontin, complement components, proteinases and inhibitors, galectins, osteonectin/SPARC, and prostaglandin D2 synthase [18].

The mammalian brain is estimated to contain about a hundred billion neurons, making it the most complex biological structure on earth [19]. The information to build a brain is encoded by no more than a subset of 80 000 genes present in the genome, a more manageable number. The use of SAGE technology has been described to decode the genetic repertoire of genes that are differentially expressed in time and in space during development of the neocortex, the part of the mammalian brain responsible for complex traits [19]. In a recent study, a multifactorial, multi-step approach called genomic convergence that combines gene expression with genomic linkage analysis to identify and prioritize candidate susceptibility genes for Parkinson's disease has been used [20]. SAGE was used to identify genes expressed in two normal substantia nigras (SN) and adjacent midbrain tissue. This identified over 3700 transcripts, including the three most abundant SAGE tags, which did not correspond to any known genes or ESTs [20]. The peripheral nerve contains both nonmyelinating and myelinating Schwann cells. The interactions between axons, surrounding myelin, and Schwann cells are thought to be important for the correct functioning of the nervous system.

A SAGE analysis of human sciatic nerve and cultured Schwann cells was performed to get insight into the genes involved in human myelination and maintenance of the myelin sheath and nerve [21]. In the sciatic nerve library, high expression of genes encoding proteins related to lipid metabolism, the complement system, and the cell cycle was found, while cultured Schwann cells showed mainly high expression of genes encoding extracellular matrix proteins [21]. Trisomy 21, or Down syndrome (DS), is the most common genetic cause of mental retardation. Changes in the neuropathology, neurochemistry, neurophysiology, and neuropharmacology of DS patients' brains indicate that there is probably abnormal development and maintenance of central nervous system structure and function. The segmental trisomy mouse (Ts65Dn) is a model of DS that shows analogous neurobehavioral defects. The global gene expression profiles of normal and Ts65Dn male and normal female mice brains (P30) have been studied using the SAGE technique [22]. There are 14 ribosomal protein genes (nine under-expressed) among the 330 statistically significant differences between normal male and Ts65Dn male brains, which possibly implies abnormal ribosomal biogenesis in the development and maintenance of DS phenotypes. This study contributes to the establishment of a mouse brain transcriptome and provides the first overall analysis of the differences in gene expression in aneuploid versus normal mammalian brain cells [22]. Neuropsychiatric diseases such as schizophrenia and bipolar disorder are major causes of morbidity throughout the world. Despite extensive searches, no single gene, RNA transcript, or protein has been found which can, on its own, account for these disorders. Recently, the availability of genomic tools such as cDNA mi-

cro arrays, SAGE, and large-scale sequencing of cDNA libraries has allowed researchers to assay biological samples for a large number of RNA transcripts. Similarly, proteomic tools allow for the quantitation of a large number of peptides and proteins. These methods include two-dimensional electrophoresis and surface-enhanced laser desorption/ionization (SELDI). Recently these techniques have been applied to study the comparison of RNAs and proteins expressed in clinical samples obtained from individuals with psychiatric diseases and controls [23]. These methods have the potential to identify pathways that are involved in the pathogenesis of complex psychiatric disorders. The characterization of these pathways may allow for the development of new methods for the diagnosis and treatment of schizophrenia, bipolar disorder, and other human psychiatric diseases [23].

Studies in liver and viral infection

To investigate the gene expression profile of a normal human liver, polyadenylated RNA was obtained from a bulk normal human liver sample and SAGE was performed [24]. RT-PCR was also performed in each of the 3 different normal liver samples to evaluate the validity of the profile in each individual. The genes highly expressed in the normal liver were those encoding plasma proteins, cytoplasmic proteins, enzymes, protease inhibitors, complements, and coagulation factors. This study identifies candidate genes to be examined in relation to various human liver diseases, including viral hepatitis, liver cirrhosis, and hepatocellular carcinoma [24].

Hepatitis C virus (HCV) causes chronic hepatitis C (CH-C) and is epidemiologically linked with the occurrence of hepatocellular carcinoma (HCC). To elucidate the comprehensive gene expression profiles of CH-C and HCC, SAGE libraries were made from CH-C and HCC tissues of a patient, and compared with a reported SAGE library of a normal liver (NL) [25]. Upregulation of IFN-gamma-inducible genes and oxidative stress-inducible genes were identified in both the CH-C and HCC libraries, and some unpublished new genes were specifically up- or down-regulated in the HCC library. This genome-wide scanning study discloses the molecular portraits of CH-C and HCC, and provides novel candidate genes that should help clarify the mechanism of hepatocarcinogenesis in the chronically HCV-infected liver [25]. Human cytomegalovirus (HCMV) has been shown to have the potential to alter cellular gene expression early after infection. SAGE has been used to investigate the transcriptional program of human fibroblasts in response to HCMV in the immediate-early phase of infection [9]. Differential expression of various cellular genes was monitored. Transcriptional expression changes of genes coding for ribosomal proteins reflected a general cellular response to starvation and stress. But differential regulation of genes coding for transcription factors and proteins associated with cellular metabolism, homeostasis, and cell structure may represent transcriptional

alterations in response to HCMV infection. Expression kinetics by 5' nuclease fluorogenic RT-PCR of selected genes revealed partial protection of infected cells against initial stress-associated alterations of gene expression and indicated fluctuations of transcriptional levels over time. Additionally, agreement with the quantitative results obtained by SAGE was observed only for genes upregulated in HCMV-infected cells [9].

Ocular studies

In an interesting study recently, SAGE technique has been used to catalogue and measure the relative levels of expression of the genes expressed in the human peripheral retina, macula, and retinal pigment epithelium (RPE) from one or two humans, aged 88 and 44 years [26]. The cone photoreceptor contribution to all transcription in the retina was found to be similar in the macula versus the retinal periphery, whereas the rod contribution was greater in the periphery versus the macula. Genes encoding structural proteins for axons were found to be expressed at higher levels in the macula versus the retinal periphery, probably reflecting the large proportion of ganglion cells in the central retina. In comparison with the younger eye, the peripheral retina of the older eye had a substantially higher proportion of mRNAs from genes encoding proteins involved in iron metabolism or protection against oxidative damage and a substantially lower proportion of mRNAs from genes encoding proteins involved in rod phototransduction. The RPE library had numerous previously unencountered tags, suggesting that this cell type has a large, idiosyncratic repertoire of expressed genes [26].

Miscellaneous studies

The profile of genes expressed in human lung fibroblast cell line WI-38 was analyzed using SAGE [27]. The expression pattern of the genes coding for cytoskeletal proteins, extracellular matrix components (ECM), and cytokines has been obtained. The results demonstrate the unique gene expression pattern in human WI-38 fibroblast cell line and support a possibility of useful application of SAGE method for differential expression analysis [27].

Consecutive application of PCR and SAGE was used to generate a catalog of approximately 50 000 SAGE tags from nine human oocytes [28]. Matches for known genes were identified using the National Institutes of Health SAGE tag database. Matches in the oocyte SAGE catalog were found for surface receptor, second-messenger systems, and cytoskeletal, apoptotic, and secreted proteins. The relative abundance of transcripts for cytoskeletal proteins and proteins known to be in oocytes are consistent with their documented expression, suggesting an absence of representational distortion by the PCR step. The expression profile of the human oocyte may help identify factors that reprogram somatic cell nuclei to totipotency [28]. A gene expression profile of the human germinal vesicle (GV) oocyte has recently been established by SAGE

[29]. A significant number of the genes identified in this profile have not previously been associated with mammalian oocytes. Two receptors found in the human catalogue, CCR6 and PAR3, were not found in mouse eggs, whereas myosin light chain, LLGL, beta-actin, 5HT receptor, bad, bak, DFF45, and Caspase homologue (cash) were present. Individual SAGE tags can match more than one gene and, in some cases, more than ten. Examination of transcript sequences that generate multiple gene assignments identified a common denominator of short interspersed elements or Alu sequences. For reasons, which are, as yet, unclear, the human GV oocyte SAGE catalogue contains relatively high abundances of SAGE tags in Alu sequences. This may reflect normal expression of Alu-containing genes in eggs or upregulated expression of Alu elements following stress. The degeneracy of gene matches in SAGE generated by Alu sequences makes independent confirmation of candidate genes essential [29]. The great frontier in reproductive medicine is implantation biology. Recent investigations utilizing molecular biology approaches have scratched the surface of the interaction between the developing trophoblast and the receptive endometrium [30].

The assessment of the expression profile of normal human thyroid tissue using SAGE generated a collection of transcripts (tags) [31]. The presence and abundance of thyroid-specific transcripts showed the overall expression profile to be from a normal thyroid cell. Seventy percent of tags could not be attributed to a known human gene and, therefore, possibly correspond to novel genes putatively involved in thyroid function [31]. In a related study, the analysis of a human thyroid SAGE library shows the presence of an abundant SAGE tag corresponding to the mRNA of thyroglobulin (TG) [32]. The results show that the three putative TG SAGE tags can be attributed to TG transcripts and reflect the use of alternative polyadenylation cleavage sites downstream of a single polyadenylation signal *in vivo*. By screening more than 300 000 sequences corresponding to human, mouse, and rat transcripts for this phenomenon, it has been shown that a considerable percentage of mRNA transcripts (44% human, 22% mouse, and 22% rat) show cleavage site heterogeneity. Both experimental and *in silico* data show that the selection of the specific cleavage site for poly(A) addition using a given polyadenylation signal is more variable than was previously thought [32].

To identify target genes of the oncogenic transcription factor c-MYC, SAGE was performed after adenoviral expression of c-MYC in primary human umbilical vein endothelial cells; 216 different SAGE tags, corresponding to unique mRNAs, were induced, whereas 260 tags were repressed after c-MYC expression [33]. The induction of 53 genes was confirmed by using microarray analysis and quantitative RT-PCR; among these genes was MetAP2/p67, which encodes an activator of translational initiation and represents a validated target for inhibition of neovascularization. Furthermore, c-MYC induced

the cell cycle regulatory genes CDC2-L1, Cyclin E binding protein 1, and Cyclin B1. The DNA repair genes BRCA1, MSH2, and APEX were induced by c-MYC, suggesting that c-MYC couples DNA replication to processes preserving the integrity of the genome. MNT, a MAX-binding antagonist of c-MYC function, was upregulated, implying a negative feedback loop. In vivo promoter occupancy by c-MYC was detected by chromatin immunoprecipitation for CDK4, Prohibitin, MNT, Cyclin B1, and Cyclin E binding protein 1, showing that these genes are direct c-MYC targets [33].

The number of genes in the human genome is still a controversial issue. Whereas most of the genes in the human genome are said to be physically or computationally identified, many short cDNA sequences identified as tags by use of SAGE do not match these genes. By performing experimental verification of more than 1000 SAGE tags and analyzing 4 285 923 SAGE tags of human origin in the current SAGE database, the nature of the unmatched SAGE tags was examined [34]. This study shows that most of the unmatched SAGE tags are truly novel SAGE tags that originated from novel transcripts not yet identified in the human genome, including alternatively spliced transcripts from known genes and potential novel genes [34]. A sensitive molecular quantitation methodology that allows analysis of in vivo immune response to vaccination has been developed [35]. Metastatic melanoma patients were immunized with a synthetically modified peptide epitope (209-M) from the melanoma self-Ag gp100. Using SAGE analysis, the functional evidence of vaccine-induced CTL reactivity in fresh cells obtained directly from the peripheral blood of post immunized patients has been reported. The in vivo localization of vaccine-induced immune response within the tumor microenvironment has also been reported. The results of these molecular assays provide direct evidence that peptide immunization in humans can result in tumor-specific CTL that localize to metastatic sites [35].

PGP9.5 is a neurospecific peptide that functions to remove ubiquitin from ubiquitinated cellular proteins, thereby preventing them from targeted degradation by the proteasome-dependent pathway or regulating their localization, activity, or structure. Using the SAGE, it was initially found that the PGP9.5 transcript and protein were highly expressed in more than 50% of primary lung cancers and nearly all lung cancer cell lines but was not detectable in the normal lung [36]. This increased expression could be the result of transcriptional regulation accompanied by methylation changes at the CpG island of the promoter region. The methylation status of the cytosines at the promoter region of human PGP9.5 was studied using sodium bisulfite genomic sequencing in normal and neoplastic cells. Although no methylation of PGP9.5 promoter was observed in the normal lung, normal cervical tissue, and lung cancer cell lines, this region was densely methylated in the HeLa cell line. Exposure of HeLa cells to the demethylating agent, 5-aza-

2'-deoxycytidine, led to re-expression of PGP9.5. This data suggested that while other mechanisms may be involved in the frequent overexpression of PGP9.5 gene in lung tumors and lung cancer cell lines, promoter methylation may play a role in the transcriptional suppression of PGP9.5 gene expression in the cervical tissue-derived HeLa cell line [36].

CONCLUSIONS

SAGE technique allows the rapid and detailed analysis including the relative abundance of all the transcripts in a given sample (cell or tissue). SAGE has been widely used in biological, medical, and pharmaceutical areas of research. It has been used successfully to compare expression profiles between normal and diseased cells or any two differentially treated samples. In brief, this technique uses two samples that are ligated and tagged with separate primers and then amplified. Subsequently, the primers are removed, revealing sticky ends that form concatemers, which are cloned into a vector and sequenced; then that is followed by extensive computational analysis. In case of yeast and cancer transcriptome, databases based on SAGE are already accessible via the internet.

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 development 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 the 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. The high level of accuracy and sensitivity as well as the depth of coverage achieved by SAGE accounts for its comparative advantage over other methods of transcript profiling. The continuing evolution and compilation of various tissue-specific SAGE data will enable further refinements in the design of the human tissue-specific microarrays.

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;389:300–305.

- [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] Hashimoto S, Suzuki T, Dong HY, Nagai S, Yamazaki N, Matsushima K. Serial analysis of gene expression in human monocyte-derived dendritic cells. *Blood*. 1999;94(3):845-852.
- [6] Hashimoto SI, Suzuki T, Nagai S, Yamashita T, Toyoda N, Matsushima K. Identification of genes specifically expressed in human activated and mature dendritic cells through serial analysis of gene expression. *Blood*. 2000;96(6):2206-2214.
- [7] Nitta T, Igarashi K, Yamashita A, Yamamoto M, Yamamoto N. Involvement of polyamines in B cell receptor-mediated apoptosis: spermine functions as a negative modulator. *Exp Cell Res*. 2001;265:174-183.
- [8] Kuramasu A, Kubota Y, Matsumoto K, et al. Identification of novel mast cell genes by serial analysis of gene expression in cord blood-derived mast cells. *FEBS Lett*. 2001;498(1):37-41.
- [9] Kenzelmann M, Muhlemann K. Transcriptome analysis of fibroblast cells immediate-early after human cytomegalovirus infection. *J Mol Biol*. 2000;304(5):741-751.
- [10] Lee S, Zhou G, Clark T, Chen J, Rowley JD, Wang SM. The pattern of gene expression in human CD15+ myeloid progenitor cells. *Proc Natl Acad Sci USA*. 2001;98(6):3340-3345.
- [11] Jansen BJ, van Ruissen F, de Jongh G, Zeeuwen PL, Schalkwijk J. Serial analysis of gene expression in differentiated cultures of human epidermal keratinocytes. *J Invest Dermatol*. 2001;116(1):12-22.
- [12] van Ruissen F, Jansen BJ, de Jongh GJ, van Vlijmen-Willems IM, Schalkwijk J. Differential gene expression in premalignant human epidermis revealed by cluster analysis of serial analysis of gene expression (SAGE) libraries. *FASEB J*. 2002;16:246-248.
- [13] Zeeuwen PL, Van Vlijmen-Willems IM, Jansen BJ, et al. Cystatin M/E expression is restricted to differentiated epidermal keratinocytes and sweat glands: a new skin-specific proteinase inhibitor that is a target for cross-linking by transglutaminase. *J Invest Dermatol*. 2001;116(5):693-701.
- [14] Welle S, Brooks A, Thornton CA. Senescence-related changes in gene expression in muscle: similarities and differences between mice and men. *Physiol Genomics*. 2001;5(2):67-73.
- [15] Urschitz J, Iobst S, Urban Z, et al. A serial analysis of gene expression in sun-damaged human skin. *J Invest Dermatol*. 2002;119(1):3-13.
- [16] Peters DG, Kassam AB, Feingold E, et al. Molecular anatomy of an intracranial aneurysm: coordinated expression of genes involved in wound healing and tissue remodeling. *Stroke*. 2001;32(4):1036-1042.
- [17] Sun Y, Zhang L, Johnston NL, Torrey EF, Yolken RH. Serial analysis of gene expression in the frontal cortex of patients with bipolar disorder. *Br J Psychiatry*. 2001;178:S137-S141.
- [18] Myerowitz R, Lawson D, Mizukami H, Mi Y, Tifft CJ, Proia RL. Molecular pathophysiology in Tay-Sachs and Sandhoff diseases as revealed by gene expression profiling. *Hum Mol Genet*. 2002;11(11):1343-1350.
- [19] Tan SS, Gunnarsen J, Job C. Global gene expression analysis of developing neocortex using SAGE. *Int J Dev Biol*. 2002;46(4):653-660.
- [20] Hauser MA, Li YJ, Takeuchi S, et al. Genomic convergence: identifying candidate genes for Parkinson's disease by combining serial analysis of gene expression and genetic linkage. *Hum Mol Genet*. 2003;12(6):671-677.
- [21] de Jonge RR, Vreijling JP, Meintjes A, et al. Transcriptional profile of the human peripheral nervous system by serial analysis of gene expression. *Genomics*. 2003;82(2):97-108.
- [22] Chrast R, Scott HS, Pappasavvas MP, et al. The mouse brain transcriptome by SAGE: differences in gene expression between P30 brains of the partial trisomy 16 mouse model of Down syndrome (Ts65Dn) and normals. *Genome Res*. 2000;10(12):2006-2021.
- [23] Johnston-Wilson NL, Bouton CM, Pevsner J, Breen JJ, Torrey EF, Yolken RH. Emerging technologies for large-scale screening of human tissues and fluids in the study of severe psychiatric disease. *Int J Neuropsychopharmacol*. 2001;4(1):83-92.
- [24] Yamashita T, Hashimoto S, Kaneko S, et al. Comprehensive gene expression profile of a normal human liver. *Biochem Biophys Res Commun*. 2000;269(1):110-116.
- [25] Yamashita T, Kaneko S, Hashimoto S, et al. Serial analysis of gene expression in chronic hepatitis C and hepatocellular carcinoma. *Biochem Biophys Res Commun*. 2001;282(2):647-654.
- [26] Sharon D, Blackshaw S, Cepko CL, Dryja TP. Profile of the genes expressed in the human peripheral retina, macula, and retinal pigment epithelium determined through serial analysis of gene expression (SAGE). *Proc Natl Acad Sci USA*. 2002;99(1):315-320.
- [27] Shklyayev S, Namba H, Hara T, Ohtsuru A, Yamashita S. Sage transcript profiles in cultured human fetal fibroblasts, WI-38. *DNA Seq*. 2000;11(3-4):281-286.
- [28] Neilson L, Andalibi A, Kang D, et al. Molecular phenotype of the human oocyte by PCR-SAGE. *Genomics*. 2000;63(1):13-24.
- [29] Stanton JL, Bascand M, Fisher L, Quinn M, Macgregor A, Green DP. Gene expression profiling of human GV oocytes: an analysis of a profile obtained by Serial Analysis of Gene Expression (SAGE). *J Reprod Immunol*. 2002;53(1-2):193-201.
- [30] Hines RS. Molecular analysis of implantation. *Semin Reprod Med*. 2000;18(1):91-96.
- [31] Pauws E, Moreno JC, Tijssen M, Baas F, de Vijlder JJ, Ris-Stalpers C. Serial analysis of gene expression

- as a tool to assess the human thyroid expression profile and to identify novel thyroidal genes. *J Clin Endocrinol Metab.* 2000;85(5):1923–1927.
- [32] Pauws E, van Kampen AH, van de Graaf SA, de Vijlder JJ, Ris-Stalpers C. Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis. *Nucleic Acids Res.* 2001;29(8):1690–1694.
- [33] Menssen A, Hermeking H. Characterization of the c-MYC-regulated transcriptome by SAGE: identification and analysis of c-MYC target genes. *Proc Natl Acad Sci USA.* 2002;99(9):6274–6279.
- [34] Chen J, Sun M, Lee S, Zhou G, Rowley JD, Wang SM. Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. *Proc Natl Acad Sci USA.* 2002;99(19):12257–12262.
- [35] Kammula US, Lee KH, Riker AI, et al. Functional analysis of antigen-specific T lymphocytes by serial measurement of gene expression in peripheral blood mononuclear cells and tumor specimens. *J Immunol.* 1999;163(12):6867–6875.
- [36] Bittencourt Rosas SL, Caballero OL, Dong SM, da Costa Carvalho Mda G, Sidransky D, Jen J. Methylation status in the promoter region of the human PGP9.5 gene in cancer and normal tissues. *Cancer Lett.* 2001;170(1):73–79.

* Corresponding author.

E-mail: renu@icgeb.res.in

Fax: +91 11 26162316; Tel: +91 11 26195007