Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Global gene expression profile in myelodysplastic syndromes using SAGE

C.F. Mendiburu[1], W.A. Silva Jr.[2], O. Ricci Jr.[3], C.R. Bonini-Domingos[1] and A.C. Fett-Conte[4]

[1]Departamento de Biologia, Instituto de Biociência Letras e Ciências Exatas, Universidade Estadual de São Paulo, São José do Rio Preto, SP, Brasil
[2]Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brasil
[3]Departamento de Medicina, Faculdade de Medicina de São José do Rio Preto, São José do Rio Preto, SP, Brasil
[4]Departamento de Biologia Molecular, Faculdade de Medicina de São José do Rio Preto, São José do Rio Preto, SP, Brasil

Corresponding author: C.F. Mendiburu
E-mail: fmendiburu@yahoo.com.ar

Genet. Mol. Res. 7 (4): 1245-1250 (2008)
Received September 3, 2008
Accepted October 1, 2008
Published November 11, 2008

ABSTRACT. The molecular pathogenesis of myelodysplastic syndromes (MDS) is poorly understood. In order to expand our knowledge of genetic defects in MDS, we determined the overall profile of genes expressed in bone marrow from patients with refractory anemia with excess blasts (RAEB) by serial analysis of gene expression (SAGE). The present report describes a partial transcriptome of RAEB bone marrow derived from 56,694 sequenced tags that provides information about expressed gene products. This is the first attempt to determine an overall profile of gene expression specifically in RAEB at diagnosis

using SAGE, which should be useful in the understanding of the physiopathology of MDS and in identifying the genes involved.

Key words: Serial analysis of gene expression methodology; Myelodysplastic syndrome; Gene expression profile; Refractory anemia with excess blasts

INTRODUCTION

Myelodysplastic syndromes (MDS) are a group of hematological disorders characterized by peripheral cytopenias with a high risk of progression to acute myeloid leukemia (AML) (Hellström-Lindberg and Malcovati, 2008). MDS mainly affects the elderly. Recently, an annual incidence of 5.4-36.2/100,000 was reported in the United States in individuals between 60 and 84 years old (Ma et al., 2007).

The clinical course of MDS can be divided into several stages. After a variable interval, about 10-40% of the patients in the early indolent chronic stage may progress to an advanced stage of refractory anemia with excess blasts (RAEB) and subsequently to AML (Mano, 2006).

Although MDS has been recognized as an important disease for more than 50 years, its molecular pathogenesis and the basis for its leukemic progression remain undefined (Look, 2005). Several genes, including oncogenes, cell cycle regulatory genes, apoptotic genes and genes that regulate DNA methylation and histone deacetylation, have been implicated in the molecular pathogenesis of MDS. The application of cDNA microarray technology has facilitated the detection of previously unidentified genes, demonstrating the potential of genome-wide approaches in investigating the molecular pathogeneses of MDS (Nishino and Chang, 2005).

As serial analysis of gene expression (SAGE) is a powerful tool to obtain comprehensive and quantitative gene expression profiles from cell populations (Velculescu et al., 1995), we performed the first analyses using SAGE on bone marrow cells from patients with the RAEB subtype of MDS, in order to expand our knowledge of genetic defects in this disease. The present report describes a partial transcriptome of bone marrow from patients with RAEB.

MATERIAL AND METHODS

Bone marrow samples

The study was approved by the Research Ethics Committee of the Institute of Biosciences, Letters and Exact Sciences, State University of São Paulo - UNESP, and written informed consent was obtained from all patients.

Bone marrow samples were obtained from three male and three female adult patients (aged 26 to 68 years old; mean [SD] age, 52 [17.25] years) with primary RAEB with diagnosis based on the FAB criteria (Bennett et al., 1982). Only samples from patients with a normal karyotype in cytogenetic analysis were included. Patients were seen in the Hematology and Hemotherapy Service of the University Hospital of São José do Rio Preto, State of São Paulo, Brazil. After aspiration, fresh bone marrow samples were treated with red blood cell lysis solution, and the nucleated cell pellet was resuspended in PBS and Trizol (Invitrogen, USA) and stored at -80°C until nucleic acid extraction.

RNA extraction and pooling

Total RNA was extracted using Trizol following the protocol supplied by the manufacturer. To ensure that the gene expression measured by the SAGE assay was not affected by degradation of the RNA extracted, gel electrophoresis and spectrophotometer readings were used to evaluate the quality and quantity of the RNA, respectively. Equal amounts of RNA from each patient were mixed to generate a pool of total RNA.

Serial analysis of gene expression

An RAEB bone marrow SAGE library was constructed using 27 µg of the total RNA pool using the I-SAGE™ Kit (Invitrogen, USA) and strictly following the manufacturer protocol. The key steps included the following: poly(A+) mRNAs in the sample were captured by oligo (dT) magnetic beads to synthesize cDNA; the cDNA was digested with *Nla*III (anchoring enzyme) and 3' cDNA isolated by the magnetic beads; the resultant 3' cDNA was split into two fractions and ligated to two SAGE adapters, A and B; SAGE tags were released by the tagging enzyme *Bsm*FI blunt ended with the Klenow polymerase fragment, and the tags from the two fractions were ligated to form ~100-bp ditags; a 1:120 dilution of the ligation product was amplified in 200 polymerase chain reactions (PCR); precipitated PCR products were run on a 12% polyacrylamide gel electrophoresis (PAGE), and only the 100-bp band containing ditags was isolated and digested with *Nla*III; the products of the digestion were run on a 12% PAGE and the 26-bp bands containing ditags purified and used for self-ligation to form concatamers; these were run on an 8% PAGE and a fraction from 500 to 1000 bp was isolated and cloned into a pZEro-1 vector digested with *Sph*I; cloned concatamers were used as templates for sequencing reactions. Sequencing reactions were carried out using DYEnamic ET Dye Terminator Sequencing Kit (Amersham Biosciences, USA) and a MegaBace 1000 DNA sequencer (Amersham Biosciences, USA).

Data analysis

Sequence files were analyzed with the SAGE2000 software, which removes the two tags flanked by *Nla*III sites. Gene identification and UniGene cluster assignment of each tag was obtained using the "reliable" tag-to-gene map from the SAGEmap NCBI site (http://www.ncbi.nlm.nih.gov/). The extracted tags were uploaded to the SAGEmap and corresponding accession numbers were retrieved using the *Homos sapiens* NCBI-GenBank database. The FatiGO+ web tool was used for functional analysis of expressed genes based on the Gene Ontology database (Al-Shahrour et al., 2007).

RESULTS

The RAEB bone marrow SAGE library denoted SAGE_BoneMarrow_MDS_RAEB was deposited in the Genome Data Mining Website (http://gdm.fmrp.usp.br/h2g/library/1153) of the Molecular Genetics and Bioinformatics Laboratory. This SAGE library yielded 56,694 sequenced tags, of which 16,244 corresponded to unique tags. Among the unique tags, 10,728 were from annotated genes and 5516 were no-matches with gene sequences. Table 1 summa-

Table 1. General serial analysis of gene expression (SAGE) data statistics of library obtained.

Origin	Frequency	Unique tags
All	56,694	16,244
Annotated	47,776	10,728
EST/cDNA	315	212
Ribosomal	8,693	269
No-matches	8,918	5,516
Frequency distribution[a]	No. of tags sequenced (%)	No. of unique tags (%)[b]
>200 tags	12,761 (22.5%)	29 (0.2%)
100-200 tags	5,722 (10.1%)	40 (0.3%)
20-99 tags	8,530 (15.0%)	216 (1.3%)
5-19 tags	9,810 (17.3%)	1,177 (7.2%)
2-4 tags	8,463 (15.0%)	3,374 (20.8%)
1 tag	11,408 (20.1%)	11,408 (70.2%)
Total	56,694 (100%)	16,244 (100%)

[a]Calculation of the frequency distribution of a given tag was based on total unique tags sequenced in the library.
[b]Some genes have more than one tag.

rizes the general statistical data of this library.

Among the unique tags, 12,761 (22.5%) tags corresponding to 29 genes showed frequencies >200 and 32,525 (57.4%) tags of the 4807 genes had frequencies between 2 and 200. Tags with a frequency of one represented 70.2% of the unique tags; however, these were not analyzed because they likely represent artifacts of sequencing or of the SAGE procedure.

The 20 most expressed genes in the RAEB bone marrow SAGE library are shown in Table 2. Eight (40%) of these 20 most expressed genes correspond to ribosomal genes, while the

Table 2. The 20 most expressed genes in the refractory anemia with excess blast (RAEB) bone marrow serial analysis of gene expression (SAGE) library.

	Tag sequence	Tags	TPM	UniGene ID	Gene	Description
1	TACCTGCAGA	2050	36080	Hs.416073	S100A8	S100 calcium-binding protein A8
2	GTGGCCACGG	1336	23514	Hs.112405	S100A9	S100 calcium-binding protein A9
3	GCCTGCTATT	1089	19166	Hs.380781	DEFA1	Defensin, alpha 1
4	CTTCTTGCCC	656	11546	Hs.654744	HBA2	Hemoglobin, alpha 2
5	GAGGGAGTTT	480	8448	Hs.523463	RPL27A	Ribosomal protein L27a
6	GCAAGAAAGT	478	8413	Hs.523443	HBB	Hemoglobin, beta
7	GTTGTGGTTA	443	7797	Hs.534255	B2M	Beta-2-microglobulin
8	GGATTTGGCC	406	7146	Hs.437594	RPLP2	Ribosomal protein, large, P2
9	GGGCTGGGGT	363	6389	Hs.425125	RPL29	Ribosomal protein L29
10	ATGTAAAAAA	344	6054	Hs.706744	LYZ	Lysozyme (renal amyloidosis)
11	TAGGTTGTCT	328	5773	Hs.374596	TPT1	Tumor protein, translationally controlled 1
12	CTGGGTTAAT	308	5421	Hs.438429	RPS19	Ribosomal protein S19
13	CCCAACGCGC	302	5315	Hs.654744	HBA2	Hemoglobin, alpha 2
14	GAAATAAAGC	300	5280	Hs.510635	IGHG1	Immunoglobulin heavy constant gamma 1 (G1m marker)
15	TGCACGTTTT	293	5157	Hs.265174	RPL32	Ribosomal protein L32
16	CCCATCGTCC	289	5086	Hs.559716	-	Transcribed locus, moderately similar to NP_536846.1 cytochrome c oxidase subunit II [*Homo sapiens*]
17	ATAATTCTTT	286	5034	Hs.156367	RPS29	Ribosomal protein S29
18	TTGGTCCTCT	282	4963	Hs.632703	RPL41	Ribosomal protein L41
19	AGGGCTTCCA	279	4910	Hs.534404	RPL10	Ribosomal protein L10
20	GAGCCCAGCC	276	4858	Hs.654826	SPATS2	Spermatogenesis associated, serine-rich 2

[a]TPM = tags per million; [(Tag frequency)(1,000,000)/Total No. of sequenced tags].

most expressed tags correspond to two calcium-binding protein genes (*S100A8* and *S100A9*).

To obtain better knowledge of the functional categories of the overall gene expression profile, 1673 genes with tag frequencies ≥3 were analyzed in the FatiGO+ web tool. Figure 1 illustrates the distribution of expressed genes in 21 principal functional categories (with number of genes >1%) as defined by the Gene Ontology (GO) Consortium. The most common transcripts correspond to genes involved in cellular metabolism. On the other hand, genes involved in other processes such as cytokine production, cell division, cellular homeostasis, and cell activation were not so frequently expressed.

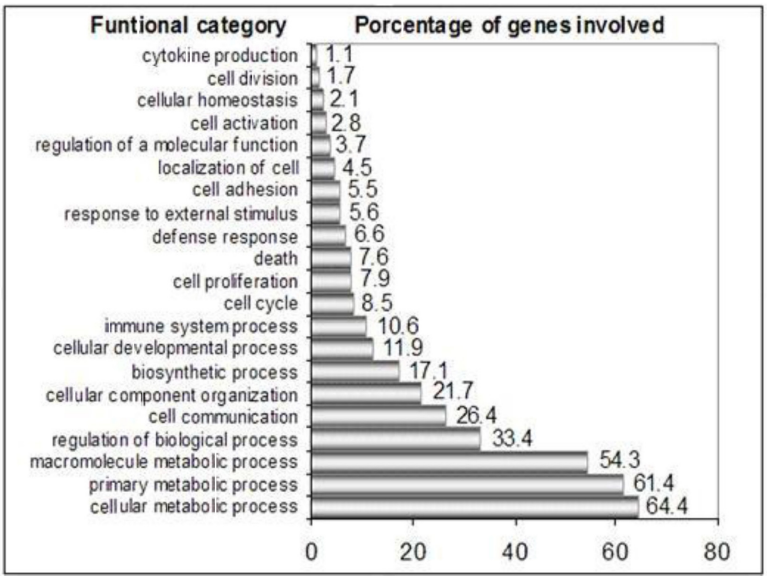

Figure 1. Functional categories assigned to individual genes identified in the refractory anemia with excess blast (RAEB) bone marrow serial analysis of gene expression (SAGE) library.

DISCUSSION

The molecular pathogenesis of MDS remains uncertain. The lack of suitable experimental models for MDS hampers progress in understanding the biology of this disease. New approaches to elucidate the underlying abnormalities are required (Hofmann et al., 2002). Hence, we proposed the first attempt to determine an overall profile of gene expression specifically in RAEB at diagnosis using the SAGE methodology.

Since its establishment, SAGE technology has been successfully and widely applied in transcript profile studies in biology and oncology to characterize the genes that are responsible for observed biological phenotypic changes (Velculescu et al., 1995; Harbers and Carninci, 2005). Unlike microarray technology, SAGE can be performed on very few cells, and does not require *a priori* knowledge of the transcriptome of a given tissue (Velculescu et al., 1995). This opens the opportunity to discover novel genes involved in the pathogenesis of MDS.

All transcripts expressed with an *Nla*III site can be "tagged" and counted efficiently in

large numbers (typically >50,000 tags) using automated sequencing (Boon et al., 2002). With this in mind, we considered the 52,309 tags from RAEB bone marrow to be a reasonable number.

GO provides a tool for functional interpretation of expressed genes in RAEB. We classified the functional category of 1673 RAEB expressed genes by GO. These genes were classified into 21 major functional categories. The comparison of this functional profile with that of normal bone marrow may elucidate the biological processes affected in RAEB.

CONCLUSION

Because SAGE provides a qualitative and quantitative evaluation of messenger RNA abundance, our data provide an estimation of the genes expressed in RAEB bone marrow cells. Data generated by the SAGE library will help to identify genes and molecular processes involved in the etiology and/or progression of RAEB. This knowledge will be useful in understanding the physiopathology of MDS as well as in identifying possible diagnostic and prognostic biomarkers and potential therapeutic targets.

ACKNOWLEDGMENTS

We thank all patients who collaborated with this study. We also thank Dr. Carlos Maluf for the collection of bone marrow samples, Daniel G. Pinheiro for his help with bioinformatics analysis, Anemarie R. Dinarte for technical support, and David A. Hewitt for English correction and idiomatic adaptations. C.F. Mendiburu is grateful to the Brazilian agency Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for a Ph.D. scholarship.

REFERENCES

Al-Shahrour F, Minguez P, Tarraga J, Medina I, et al. (2007). FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Res.* 35: W91-W96.

Bennett JM, Catovsky D, Daniel MT, Flandrin G, et al. (1982). Proposals for the classification of the myelodysplastic syndromes. *Br. J. Haematol.* 51: 189-199.

Boon K, Osorio EC, Greenhut SF, Schaefer CF, et al. (2002). An anatomy of normal and malignant gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 99: 11287-11292.

Harbers M and Carninci P (2005). Tag-based approaches for transcriptome research and genome annotation. *Nat. Methods* 2: 495-502.

Hellström-Lindberg E and Malcovati L (2008). Supportive care, growth factors, and new therapies in myelodysplastic syndromes. *Blood Rev.* 22: 75-91.

Hofmann WK, de Vos S , Komor M, Hoelzer D, et al. (2002). Characterization of gene expression of CD34+ cells from normal and myelodysplastic bone marrow. *Blood* 100: 3553-3560.

Look AT (2005). Molecular Pathogenesis of MDS. *Hematology Am. Soc. Hematol. Educ. Program*, 156-160.

Ma X, Does M, Raza A and Mayne ST (2007). Myelodysplastic syndromes: incidence and survival in the United States. *Cancer* 109: 1536-1542.

Mano H (2006). DNA micro-array analysis of myelodysplastic syndrome. *Leuk. Lymphoma* 47: 9-14.

Nishino HT and Chang CC (2005). Myelodysplastic syndromes: clinicopathologic features, pathobiology, and molecular pathogenesis. *Arch. Pathol. Lab. Med.* 129: 1299-1310.

Velculescu VE, Zhang L, Vogelstein B and Kinzler KW (1995). Serial analysis of gene expression. *Science* 270: 484-487.