Erythropoiesis• Research Paper

Analysis of remnant reticulocyte mRNA reveals new genes and antisense transcripts expressed in the human erythroid lineage

[haematologica]
2004;89:1434-1440

Béatrice Bonafoux
Mireille Lejeune
David Piquemal
Ronan Quéré
Aurélie Baudet
Léa Assaf
Jacques Marti
Patricia Aguilar-Martinez
Thérèse Commes

A B S T R A C T

Background and Objectives. We studied the gene expression profile of human purified reticulocytes to provide a transcriptional basis for the study of erythroid biology, differentiation and hematologic disorders.

Design and Methods. We screened highly purified blood reticulocytes from ten healthy adult volunteers. We chose a modified protocol of serial analysis of gene expression (SAGE), the serial analysis of downsized extracts (SADE).

Results. Data analysis revealed that 64% of gene signatures (tags) matched with known genes; mainly hemoglobin. In addition to the abundant globin mRNA, SAGE analysis identified previously described genes and new transcripts. In reticulocytes, which are poor in mRNA, we also identified 9% of EST and 27% of tags that did not match with any known genes. Mining our data, 70% of the unknown tags and 39% of tags identifying EST were found to be specific to the reticulocyte. We demonstrated the presence of a mRNA that matched with the reverse sequence of the hemoglobin β (*HBB*) transcript.

Interpretation and Conclusions. This is the first description of an antisense transcript of the human *HBB* gene suggesting regulation by way of sense-antisense pairing. The well-characterized genes found in the SAGE library were genes specific to the blood cell lineage, housekeeping genes and, interestingly, genes not previously described in the reticulocyte. Furthermore the study provides markers of the erythroid lineage regulated during the differentiation process as observed in *in vitro* experiments.

Key words: SAGE, red blood cell, reticulocyte, blood, mRNA.

From the Institut de Génétique Humaine, CNRS UPR 1142, Montpellier, France (BB, ML, RQ, AB, LA, JM, PA-M, TC); Skuld-Tech, Faculty of Sciences, University Montpellier II, Montpellier, France (DP). BB and ML should be regarded as joint first authors.

*Correspondence:
Thérèse Commes,
Institut de Génétique Humaine
UPR CNRS 1142, 34396
Montpellier cedex 5, France.
E-mail: commes@univ-montp2.fr*

Reticulocytes, anucleated cells which circulate for 2 to 3 days in blood, are the last stage of erythropoiesis before formation of definitive red blood cells (RBC). Although most of the RBC proteins have been characterized, little is known about expression profiles of their mRNA during cell differentiation. As reticulocytes are abundant and easily accessible in blood and contain remnant erythroid transcripts, the gene expression analysis of these cells allows us to expand our knowledge on RBC structure and functions. Our aim was to initiate a genomic global approach to provide a transcriptional basis on which to follow erythroid differentiation and study RBC disorders. Gene expression in differentiating human erythroid nucleated cells was previously studied using suppressive subtractive hybridization (SSH).[1] We used a modified protocol of serial analysis of gene expression (SAGE),[2] the serial analysis of downsized extracts (SADE)[3] which allows identification of genes from small amounts of mRNA and is thus an appropriate method starting from reticulocytes. SAGE does not depend on previous knowledge of genes. The method produces a set of cellular transcribed gene signatures (tags), and provides a comprehensive view of cell phenotypes by computational analyses (*http://cgap.nci.nih.gov/SAGE; http://www.ncbi.nih.gov/SAGE/*).[4] Here, the SAGE analysis was performed on highly purified reticulocytes to allow identification of the specificity of expressed genes. Blood cells, and in particular reticulocytes, are difficult to eliminate from complex tissues and remain the major contaminant in global analysis approaches. Indeed, this gene profile could be used for global analysis of transcriptomes. These data represent the first gene profiling information for RBC

studies and were applied here to analyze gene regulation during erythroid differentiation.

Design and Methods

Sample preparation

Reticulocytes for the SADE library

Reticulocytes were purified from fresh blood samples obtained from 10 healthy adult volunteers (5 women and 5 men) after informed consent. The donors had normal blood cell counts, blood smears, hemoglobin electrophoresis, red cell membrane resistance tests and no biological evidence of hemolysis. Each adult donor gave 5 mL of whole blood from which reticulocytes were purified. Whole blood samples were centrifuged, the supernatant and the buffy coat containing white blood cells (WBC) and platelets were removed. The red cell pellet was then purified using the method described by Brun *et al*.[5] Briefly, fresh venous blood was filtered through a cellulose column, retaining WBC and platelets.[6] Purity, assessed by a Coulter® STKS™ hematology flow cytometer (Beckman-Coulter, Roissy, France), was 1 leukocyte per 15 million RBC and the efficiency of the purification process was 99.8% and 1 platelet per 10,000 RBC.

In vitro experiments

Reticulocytes were sampled from cord blood of healthy neonates, less than one hour after delivery and were purified as described above. The human erythroleukemia cell line TF1 was grown as reported by Murate *et al*. in the presence of 500 IU/mL of recombinant granulocyte-monocyte colony-stimulating factor (GM-GSF).[7] Differentiation with recombinant human erythropoietin (rHuEPO, 5 IU/mL, Neorecormon®) was controlled by flow cytometry. The expression of HLA-DR and CD34 declined and expression of glycophorin A (*GYPA*) increased with maturation.

Generation of SADE libraries and sequence data analysis

Total RNA from all ten donors was extracted individually and then pooled for mRNA purification. Total RNA was extracted with Trizol™ (Invitrogen, Cergy Pontoise, France) from a total of 5×10^9 reticulocytes. Poly(A) mRNA were prepared and selected by hybridization to oligo (dT) 25-coated magnetic beads according to the manufacturer's instructions (Dynal, Compiegne, France). A total of 200 ng mRNA was obtained. The library was constructed using a modified SAGE procedure as previously described.[3,4] *Sau3*AI was chosen as anchoring enzyme because of the absence of its restriction site (GATC) on α globin, *HBA1* and *HBA2* mRNA. C+tag software (Skuld-Tech, France, *http://www.skuldtech.com*) was used for automatic tag detection and counting. This program provides criteria for assessing the quality of the SAGE library, such as ditags [ditags are the combination of two tags into a unit], frequency of repeated ditags, and detection of linkers. For the definition of these terms see[4] and *http://www.embl-heidelberg.de/info/sage*. The rate of contamination by linker sequences was 0.5%. For this particular reticulocyte library, repeated ditags (mainly from the *HBB* gene) represented 40% of the total ditag population. A high frequency reveals an artifactual loss of complexity during the library construction or, as expected in this cell, a low complexity of the original mRNA population observed in highly specialized tissues. Our goal was to determine the most abundantly expressed genes; for this reason and due to the specificity of this library, we sequenced only 3000 tags. Preditag® software (Skuld-tech, France) was used for tag prediction and *tag-to-gene* mapping. Furthermore, the Preditag® software allows identification of canonical tags [first tag obtained following Sau3A digestion from the 3' end of the mRNA sequence] as well as more unusual tags, second and third tags [corresponding to the second and third Sau3A sites from the 3' end of the mRNA] and also tags in reverse orientation. Full data sets are available on Internet at URL *http://www.igh.cnrs.fr/equip/transcriptome*. No-match tags observed once were excluded. In order to compare our data with other published SAGE data, we used the SAGE Genie web site (*http://www.ncbi.nih.gov/SAGE*). This database gathers the expression level of every transcript from more than 200 SAGE libraries constructed with the *Nla*III anchoring enzyme and obtained from a large range of human cell types.[8] SAGE Genie provides an automatic link between gene name, Genbank number, SAGE tag occurrence and tissue expression.

Polymerase chain reaction (PCR) analysis and quantification by real-time PCR

To confirm tag identification, we used real-time PCR analysis to check the presence of mRNA from purified normal human reticulocytes and different cell types. The cDNA were synthesized using mRNA and oligo (dT)-primers. For the analysis of hemoglobin β (*HBB*) antisense mRNA, primer-specific reverse transcription was performed with Thermoscript (Invitrogen, France) at 70°C to avoid non-specific interaction. Amplicons were characterized by their size and, in case of ambiguities, by sequence analysis. Quantification of gene expression was carried out on cDNA by real-time PCR using a Light Cycler apparatus (Roche, France) as already described.[9] The primers used were as follows: hemoglobin β (*HBB*, NM_000518): forward primer 5'-GCAACCTCAAACAGACACCA-3'; reverse primer 5'-

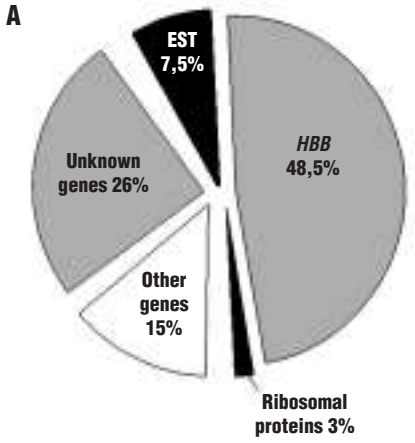
AGCTCACTCAGTGTGGCAAA-3'; ferritin, light polypeptide (FTL, NM_000146): forward primer 5' TGTACCTGCAGGCCTCCTAC-3', reverse primer 5'-AGAACCCAGGGCATGAAGAT-3'; CGI-69 (NM_016016) forward primer 5'-CCCCAAGTTCAAGACCAAAT-3', reverse primer 5'-GGGTCAGGGAGACTACACCA-3'; ornithine decarboxylase antizyme 1 (OAZ1, NM_004152) forward primer 5'-GGATTCTCAACGTCCAGTCC-3', reverse primer 5'-CGGTTCTTTGTGGAAGCAAA-3'. For the global calibration of samples, RPS19 (NM_001022) was quantified. The following primers were designed: *RPS19* forward 5'-GGCTGAAAATGGTGGAAAAG-3' and reverse 5'-AACCCAGCATGGTTTGTTC-3'. In order to normalize the amount of RNA of each experiment and to compare various samples, the results are presented as a ratio between the number of copies of each gene and *RPS19*.

For the quantification of *HBB* antisense mRNA *versus* sense *HBB* mRNA, specific reverse transcription was performed with, respectively, the A and B primers described in Figure 2. Real-time PCR was performed with the *HBB* primers mentioned below.

Results

Global analysis of the library

In order to inventory the most abundant remnant mRNA expressed in blood reticulocytes, samples from 10 healthy adults were used to build a SADE library using Sau3A1 as anchoring enzyme.[3,4] Sequence data analysis revealed that 64% (1920 tags) of the obtained tags matched with known genes (HUGO gene nomenclature), 9% (270 tags) with EST, while 27% (810 tags)



B	Number of different tags	Sum of occurences
EST	33	226
No match	135	775
Known genes	88	1992
HBB		1458
Ribosomal proteins	24	86
Other known gens	63	448

Figure 1. Tags distribution in the reticulocyte. A) Percentage of each group of genes calculated with the sum of occurrences. *HBB*, ribosomal protein, other genes correspond to known genes with a Refseq accession number (*http://www.ncbi.nlm.nih.gov/RefSeq/*). B) For each group of genes, number of different tags (diversity) and sum of occurrences are reported. Details are available on line at *http://igh.cnrs.fr/equip/transcriptome/*.

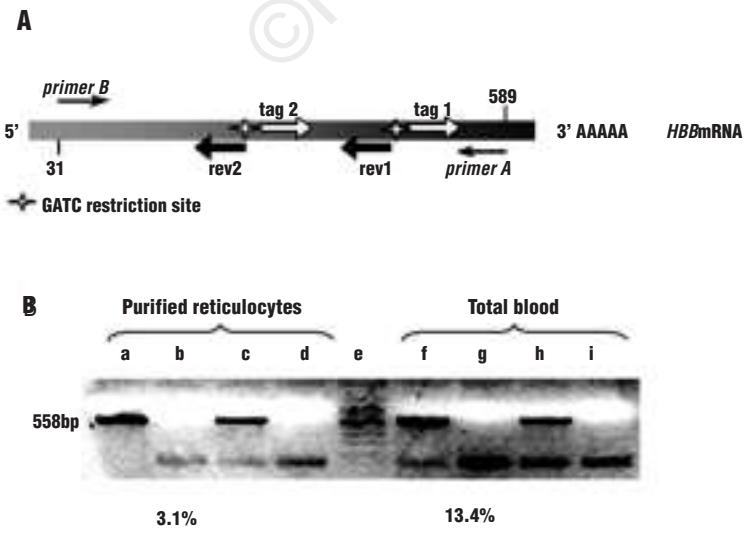


Figure 2. *HBB* SAGE-tags analysis. (A) Localization of tags and reverse tags on *HBB* transcript. (B) Validation of *HBB* antisense transcript. Reverse transcription was performed using primers which specifically hybridized to either HBB sense (lane a and f) or *HBB* antisense transcript (lane c and h), respectively primer A 5'-GATGCTCAAGGCCTTTCATA-3' and primer B 5'-GCAACCTCAAACAGACACCA-3'. Controls without reverse-transcriptase were performed to check for genomic DNA contamination (lanes b, d, g and i). From lane a to d PCR were performed on RNA extracted from a pool of purified reticulocytes, from lane f to i PCR were performed on RNA extracted from a unique sample of total blood. PCR were performed with primers A and B, analyzed by electrophoresis on 1.5% agarose gel and PCR products were sequenced. Product lengths were assessed by the used of a ladder (lane e). Percentages represent the antisense vs sense counterpart.

Table 1. Well annotated genes and specific reticulocyte markers in reticulocyte and leukocyte SAGE libraries. [1]For an easier reading of the results, tag frequencies were calculated per 100 HBB tags. [2]Absolute number of tags in reticulocyte library. [3]Frequencies of *NlaIII* tags observed among all the CATG-constructed libraries available on *http://cgap.nci.nih.gov/SAGE*). *Expression observed only in reticulocytes by PCR verification in hematopoietic cell lines (U937, NB4, Jurkatt, TF1). °Genes newly described in the reticulocyte.

Tag NlaIII	*Tag Sau3A1*	*HUGO symbol*	*Leukocyte library[1]*	*Reticulocyte library* [1]	[2]	*GenBank acc. no.*	*SAGE genie[3]*
CATGGCAAGAAAGT	GATCCTGAGAACTT	HBB (tag 1)*	100	100	647	NM_000518	2227
CATGGTGCACCTGA	GATCTGTCCACTCC	HBB (tag 2)*	2	83	539	NM_000518	30
CATGAGCCTTCACC	GATCCACGTGCAGC	HBB (rev 1)*	4	10	67	NM_000518	94
CATGGTGTCTGTTT	GATCCCCAAAGGAC	HBB (rev 2)*°	3	4	29	NM_000518	101
CATGGTGGGGATTC	GATCGCACCGCCCC	OLIG2*°	0	3	16	NM_005806	55
CATGTACCGGCACA	GATCCAGCTCAGAA	GYPC*°	1	3	16	NM_002101	7
CATGAATAAGGGTT	GATCTGTCCACTTC	NDP52*°	1	3	13	NM_005831	20
CATGTCTCCAGGAA	GATCATCACTTACC	CGI-69°	1	3	13	NM_016016	697
CATGCCCTGGGTTC	GATCTTCATGCCCT	FTL	8	2	8	NM_000146	4911
CATGGCATAATAGG	GATCAGAAAAAGAA	RPL21	11	2	8	NM_000982	15447
CATGCAGATCTTTG	GATCTGCCGCAAGT	UBA52	1	2	8	NM_003333	1415
CATGGGATTTGGCC	GATCTTGGACAGCG	RPLP2	10	2	5	NM_001004	12232
CATGTCTGTACACC	GATCGGTGACATCG	RPS11	1	2	5	NM_001015	607
CATGTAGGTTGTCT	GATCTATCACCTGT	TPT1 b	21	2	5	NM_003295	13463
CATGGAGGGAGTTT	GATCGGGCTACTAC	RPL27A	22	2	5	AY044167	14440
CATGACATCATCGA	GATCATTTGACAAC	RPL12	3	2	5	AK026491	6323
CATGATTAACAAAG	GATCAGCACTGCCA	GNAS°	1	2	5	NM_080425	1862
CATGGCCGTGTCCG	GATCAGACTCTGAA	RPS6	2	2	5	AK027187	5769
None	GATCACAGCCGAAG	RPS24	0	2	5	NM_033022	0
CATGTTGTAATCGT	GATCCCGGGCGGCG	OAZ1°	2	2	5	NM_004152	4373
CATGGGCTTTGGTC	GATCAATGCCCTCA	RPLP1	1	1	5	NM_001003	518

Table 2. EST identification and undetermined tags (no-match). *Sau3A1 tag occurrences observed in GATC-constructed libraries (*http://www.igh.cnrs.fr/equip/transcriptome and personal data*). The reticulocyte full-database is available on the web site *http://www.igh.cnrs.fr/equip/transcriptome*.

tag Sau3A1	*HUGO symbol*	*Other Sau3A1 libraries**	*Reticulocyte library*	*GenBank acc. no.*
GATCCTGAGAACTG	EST	5	26	AI024715
GATCTCCTGAGAAC	EST	0	21	AI417219
GATCGCCCCAGCTG	EST	2	13	AI093704
GATCATCCTGAGAA	EST	0	10	AI678637
GATCATCTGTCCAC	EST	1	10	W22959
GATCTGTCCAATTT	EST	0	10	BF184321
GATCGAGGCCCAGC	EST	7	8	BF032734
GATCGATAAGCTTG	EST	8	8	AW366294
GATCGCTTGAGCAT	EST	0	5	BF870697
GATCCTGTCCACTC	no match	0	42	
GATCGCGCCTGTGA	no match	1	29	
GATCCTGAGAACTC	no match	3	18	
GATCCCTGAGAACT	no match	0	16	
GATCTGAGAACTTA	no match	0	16	
GATCAGAGTAGTGG	no match	3	16	
GATCGCACTGCCCC	no match	1	16	
GATCGCACAGCCCC	no match	1	13	
GATCTCTGTCCACT	no match	0	13	
GATCTGTCCACCCC	no match	1	10	
GATCCCCACAGACT	no match	0	8	
GATCCTGAGAACCT	no match	0	8	
GATCTGAGAACTTC	no match	0	8	
GATCTGTCTACTCC	no match	1	8	

remained unidentified (no-match) (Figure 1A). The absence of matching with Unigene clusters does not mean that tags are artifactual, but only that they are currently *unknown*. In our strategy of analysis of the tags we voluntarily limited our identification to those genes which are well-known and defined, as is the case in the Unigene database. The comparison with other hematopoietic SAGE data from our laboratory (available at URL: *http://www.igh.cnrs.fr/equip/transcriptome* and *unpublished data*) revealed that 70% of the no-match tags and 39% of the tags matching with EST seemed to be specific to the erythroid lineage. The most abundant tags observed in the reticulocyte are reported in Tables 1 (known genes) and 2 (non-characterized genes).

The β, α1 and α2 globin gene transcripts (*HBB*, *HBA1* and *HBA2*, respectively) are known to be the most expressed mRNA in the reticulocyte. Interestingly, *HBA1* and *HBA2* were eliminated from the gene expression profile by the use of the Sau3AI anchoring enzyme, avoiding unnecessary sequencing. The *HBA1* and *HBA2* tags are present in the Leukocyte library constructed with the *NlaIII* enzyme from bulk peripheral blood cells [Gene Expression Omnibus accession (GEO GSM709); available at URL *http://www.ncbi.nih.gov/geo,* or URL *http://www.ncbi.nih.gov/SAGE*]. These data allow the ratio between *HBA1+HBA2* and *HBB* transcripts in these bulk blood cells to be estimated at 1.4.

No other globin gene was found in the library. This is not surprising as other globin chains account for less than 2% of the globin chains in normal adults.

Analysis of the HBB tags

Canonical SAGE tags are localized in the most 3'anchoring enzyme site of the mRNA.[2] In addition to the canonical tag of *HBB* (*HBB*-tag1), the most abundant of the tags (21.5%), we also found a second tag (*HBB*-tag2) at a high rate (18% of all tags) (Table 1 and Figure 2). This is quite unusual in the absence of polymorphism in the first tag, of an internal stretch of polyA or of alternative splicing, which were manually checked in databases. Second tags resulting from incomplete digestion on the first Sau3A site are usually less common (2% in previously reported SAGE libraries).[4] This second tag matched with two Genbank sequences (U90143 and U97502) corresponding to *BTN3A3* transcripts, members of the butyrophilin family.[10] These 2 sequences contain an additional Sau3A site (GATC) compared with the reference sequence of the UniGene cluster. We confirmed the expression of this member of the butyrophilin family by sequencing a fragment amplified by RT-PCR from reticulocyte RNA.

Two others *HBB* tags (*HBB*-rev1 and *HBB*-rev2) matched with reverse sequences of the *HBB* transcript (Figure 2). Screening of the already cited *Nla*III Leukocyte library revealed that the corresponding *HBB*-rev1 and *HBB*-rev2 tags were also present in this library. Performing reverse transcription with a primer that specifically hybridizes to either antisense or sense transcripts,[11] we confirmed that these tags corresponded with an *HBB* antisense mRNA. Sequencing analysis revealed that the sense and antisense *HBB* transcripts are perfectly aligned from position nt 31 to nt 589, a segment which includes the 3 *HBB* exons (Figure 2B).

To our knowledge, this is the first description of a human *HBB* antisense transcript. An antisense transcript was previously described in normal mouse erythroid tissue and accounted for 3% of its sense counterpart.[12] Using real-time PCR, we quantified the antisense transcript in human blood samples, purified reticulocytes and bulk blood cells. The antisense transcript represented 3.08% and 13.4% of its sense counterpart in these experiments (Figure 2B).

Specific reticulocyte genes and newly described genes

To evaluate the specificity of the tags identified in the reticulocyte SAGE library we searched for the presence of these signatures in other lineages. We selected the most abundant tags corresponding to known genes in the reticulocyte library (Table 1). This selection provided a well-characterized group of genes for further *in silico* and *in vitro* investigations.

We first compared the reticulocyte transcriptome to the above-cited *Nla*III Leukocyte library. As expected, 95% of the reticulocyte-selected genes were also found in this blood library. Secondly we carried out analyses using the SAGE Genie database (available at *http://cgap.nci.nih.gov/SAGE*) corresponding to a large range of human cell types. The tags frequencies are reported for the well-characterized group of reticulocyte genes (Table 1).

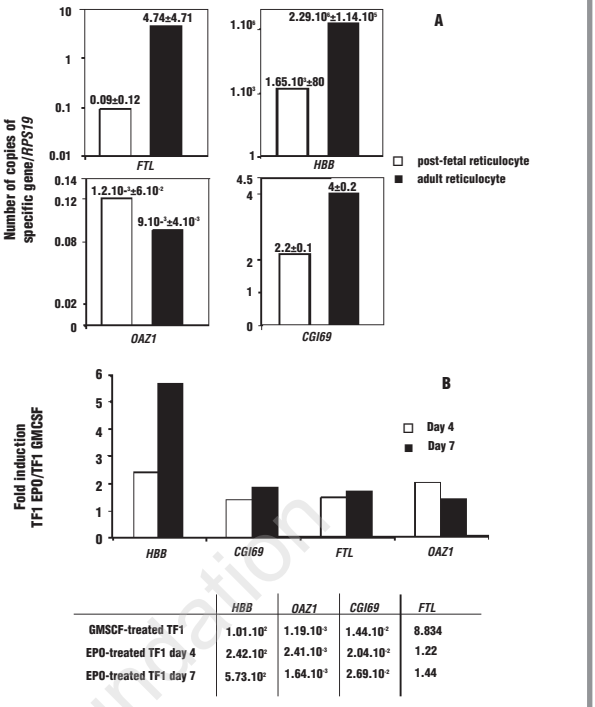
We then performed PCR analysis on hematopoietic cell lines (U937, NB4, Jurkatt, TF1) to further characterize erythroid-specific transcripts. *HBB*, *GYPC*, *NDP52* and OLIG2 showed restricted lineage expression. Moreover these transcripts were observed with a low frequency in the SAGE Genie database in non-hematopoietic tissues (Table 1). Newly described genes in reticulocytes are reported in Table 1. It is noteworthy that several tissue libraries contain a significant number of *HBB* tags; consequently, the most abundant and specific reticulocyte tags could be used as markers of tissue contamination by blood.

The most abundant and well-characterized genes were either specific to the blood cell lineage or housekeeping genes. Other transcripts such as *OLIG2* and *NDP52* were not expected in reticulocytes. It is interesting to note that the identified genes correspond to

genes involved in metabolic pathways for oxygen transport, protein biosynthesis, cell surface proteins, mitochondrial and cytoplasmic enzymes. Furthermore, only 2% of tags identified ribosomal proteins. This was lower than the percentages found in SAGE data established from cell lines or primary cells, which were 15.6% and 10%, respectively[4] (*and personal data*).

Aspects of in vitro and in vivo differentiation

As a preliminary *in vitro* approach, we studied the regulation of abundant transcripts during erythroid differentiation. For this purpose we analyzed *OAZ1*, *CGI-69*, *FTL* and *HBB* expression using real-time PCR. We tested these markers on a TF1 cell line treated with recombinant human erythropoietin, which provides a model for erythroid differentiation. Considering the difference of RBC maturation in post-fetal and adult life,[13] we also tested these markers in cells from cord blood and from adults. The data indicated that mRNA distribution was different in post-fetal and adult reticulocytes (Figure 3A). An increased expression of these genes was observed during TF1 differentiation after 4 and 7 days of treatment (Figure 3B).



	HBB	OAZ1	CGI69	FTL
GMSCF-treated TF1	1.01.10²	1.19.10⁻³	1.44.10⁻²	8.834
EPO-treated TF1 day 4	2.42.10²	2.41.10⁻³	2.04.10⁻²	1.22
EPO-treated TF1 day 7	5.73.10²	1.64.10⁻³	2.69.10⁻²	1.44

Figure 3. Analysis of erythroid differentiation markers. *HBB*, *CGI69*, *FTL* and *OAZ1* gene expression was studied using real-time PCR following different aspects of *in vivo* and *in vitro* differentiation. (A) Ratio between the number of copies of a specific gene and *RPS19* in post-fetal and adult stages. Standard deviation values are indicated in front of each experiment done with five donors. (B) Gene induction in TF1 cell line treated four and seven days with rHuEPO, *versus* GM-CSF. Ratio between the number of copies of a specific gene and *RPS19* for all experiments are indicated below.

Discussion

During the course of our study to create an inventory of remnant mRNA in blood reticulocytes, we confirmed already known data and found new and unexpected information on late RBC mRNA expression.

As expected the SAGE analysis showed that reticulocytes abundantly express hemoglobin transcripts and low levels of genes involved in protein biosynthesis or other metabolic pathways. In addition to the abundant globin mRNA, the SAGE analysis identified previously described genes as well as new transcripts. Among these, the discovery of antisense β globin transcript mRNA is of utmost interest. First of all, tags matching with the *HBB* mRNA in reverse orientation were sequenced and strand-specific RT-PCR allowed amplification of antisense transcripts. Reticulocyte-restricted expression was observed when mining other non-hematopoietic SAGE libraries, suggesting constitutive antisense transcription of the β-globin gene in this cell.

By mining the collection of 260 publicly available SAGE libraries, we also found reverse tags of *HBA1* and *HBA2* transcripts in the Leukocyte library. An *HBB* antisense transcript was previously described in normal mouse erythroid tissue[12] and, more recently, in a case of α-thalassemia, an antisense transcript of *HBA2* was described as a novel cause of human genetic disease.[11] Reverse tags for globin genes are also present, at a low rate, in the SAGE Genie database and seem to be specific to the erythroid lineage. Thus, the existence of globin antisense transcripts has been observed in the erythroid lineage under both, pathological[11] and physiological[12] conditions, suggesting a regulatory role in globin gene expression.

The sequence analysis revealed that the sense and antisense *HBB* transcripts are perfectly aligned in a region including the three *HBB* exons. The observation of reverse tags in the SAGE library constructed from oligo-dT cDNA may suggest that the antisense transcript detected is a polyadenylated product. We could therefore hypothesize that the antisense transcript is synthesized from the opposite strand of the sense counterpart followed by a splicing process. The biological synthesis process of the *HBB* antisense transcript remains to be clearly demonstrated.

The systematic analysis of reticulocyte gene expression allowed identification of well-known genes, such as FTL, GYPC and OAZ1. Some components of RBC structure or functions such as *ABO*, *RHD*, *GYPA*, *KEL*,

SLC4A1, *SPTB* and *SPTA1* were expected to be found in the profile. The corresponding tags were not detectable in our library. It is noteworthy that the corresponding *NlaIII* tags were similarly not detected in the Leukocyte library or in SAGE Genie databases among the over 5 millions tags reported. This demonstrates the low correlation between the usually described proteins expressed in reticulocytes and the expressed mRNA. Moreover, in a recent study, Hosoi *et al.*[14] showed that the transcription of both the FUT1(H) and *ABO* genes starts early in immature red cells but gradually decreases during cellular maturation while the antigen expression is maintained thereafter.

Among the genes newly identified in the reticulocyte SAGE library, we demonstrated the presence of a member of the butyrophilin (BT) family, the *BTN3A3* variant.[10] Several members of the BT gene family are widely expressed in hematopoietic cells (see UniGene cDNA source) and erythroid membrane associated protein (*ERMAP)*, also sharing the B30.2 domain common to all members of the BT family, has been recently reported in human erythroid tissues.[15,16] We also found other tags, corresponding to unexpected genes (*OLIG2*, *NDP52* and *GNAS*) that will require further investigation.

Variations in the level of expression of these genes during *in vitro* cell analysis (TF1 cell line, reticulocytes from adults and neonates) indicate that they could serve as markers of erythroid differentiation in further studies. These preliminary results provide perspectives for research on RBC physiology and pathologies.

BB: conception, design, carrying out of the human study, carrying out the SADE, analysis and data interpretation, drafting and article revision and final approval; ML: conception, design, carrying out of the SADE, the real time-PCR, analysis and data interpretation, tables and figures design, drafting and article revision and final approval; DP: conception, design, carrying out of the SADE, conception of bioinformatic analysis, supplementary material and web data, analysis and data interpretation, drafting and article revision and final approval; RQ: handling, design, data acquisition and interpretation, article revision and final approval. AB, LA: handling, design, data acquisition and interpretation, article revision and final approval; JM: design, data acquisition and interpretation, article revision and final approval; PAM: design, carrying out of the human study, data acquisition and interpretation, drafting the article, article revision and final approval; TC: design, analysis and data acquisition and interpretation, drafting the article, article revision and final approval. The authors reported no potential conflicts of interest.

Professor P. Boulot is gratefully acknowledged for the gift of cord blood samples.

The CNRS and the Ligue Nationale contre Le Cancer supported this work.

Manuscript received February 9, 2004. Accepted September 14, 2004.

References

1. Gubin AN, Njoroge JM, Bouffard GG, Miller JL. Gene expression in proliferating human erythroid cells. Genomics 1999; 59:168-77.
2. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science 1995;270:484-7.
3. Virlon B, Cheval L, Buhler JM, Billon E, Doucet A, Elalouf JM. Serial microanalysis of renal transcriptomes. Proc Natl Acad Sci USA 1999;96:15286-91.
4. Piquemal D, Commes T, Manchon L, Lejeune M, Ferraz C, Pugnere D, et al. Transcriptome analysis of monocytic leukemia cell differentiation. Genomics 2002; 80: 361-71.
5. Brun A, Skadberg O, Hervig TA, Sandberg S. Phenotyping autologous red cells within 1 day after allogeneic blood transfusion by using immunomagnetic isolation of reticulocytes. Transfusion 1994; 34: 162-6.
6. Beutler E, Gelbart T. The mechanism of removal of leukocytes by cellulose columns. Blood Cells 1986;12:57-64.
7. Murate T, Saga S, Hotta T, Asano H, Ito T, Kato K, et al. The close relationship between DNA replication and the selection of differentiation lineages of human erythroleukemia cell lines K562, HEL, and TF1 into either erythroid or megakaryocytic lineages. Exp Cell Res 1993; 208: 35-43.
8. Boon K, Osorio EC, Greenhut SF, Schaefer CF, Shoemaker J, Polyak K, et al. An anatomy of normal and malignant gene expression. Proc Natl Acad Sci USA 2002; 99:11287-92.
9. Varsamos S, Wendelaar Bonga SE, Flik G, Quere R, Commes T. Cloning of a proopiomelanocortin cDNA from the pituitary gland of the sea bass (Dicentrarchus labrax) and assessment of mRNA expression in different tissues by means of real-time PCR. J Endocrinol 2003;176:405-14.
10. Tazi-Ahnini R, Henry J, Offer C, Bouissou-Bouchouata C, Mather IH, Pontarotti P. Cloning, localization, and structure of new members of the butyrophilin gene family in the juxta-telomeric region of the major histocompatibility complex. Immunogenetics 1997;47:55-63.
11. Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, et al. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet 2003;34:157-65.
12. Volloch V, Schweitzer B, Rits S. Antisense globin RNA in mouse erythroid tissues: structure, origin, and possible function. Proc Natl Acad Sci USA 1996; 93:2476-81.
13. Stamatoyannopoulos G, Nienhuis AW, Majerus PW, Varmus. H. The Molecular Basis of Blood Diseases, 2nd ed. Philadelphia, 1994.
14. Hosoi E, Hirose M, Hamano S. Expression levels of H-type α(1,2)-fucosyltransferase gene and histo-blood group ABO gene corresponding to hematopoietic cell differentiation. Transfusion 2003;43:65-71.
15. Xu H, Foltz L, Sha Y, Madlansacay MR, Cain C, Lindemann G, et al. Cloning and characterization of human erythroid membrane-associated protein, human ERMAP. Genomics 2001;76:2-4.
16. Wagner FF, Poole J, Flegel WA. Scianna antigens including Rd are expressed by ERMAP. Blood 2003;101:752-7.