The human reticulocyte transcriptome

Sung-Ho Goh,1 Matthew Josleyn,1 Y. Terry Lee,1 Robert L. Danner,3 Robert B. Gherman,4 Maggie C. Cam,2 and Jeffery L. Miller1

1Molecular Medicine Branch, 2Microarray Core Facility, National Institute of Diabetes, Digestive and Kidney Diseases; 3Critical Care Medicine Department, Clinical Center, National Institutes of Health; and 4National Naval Medical Center, Bethesda, Maryland

Submitted 9 November 2006; accepted in final form 26 March 2007

Goh S-H, Josleyn M, Lee YT, Danner RL, Gherman RB, Cam MC, Miller JL. The human reticulocyte transcriptome. Physiol Genomics 30: 172–178, 2007. First published April 3, 2007; doi:10.1152/physiolgenomics.00247.2006.—RNA from circulating blood reticulocytes was utilized to provide a robust description of genes transcribed at the final stages of erythroblast maturation. After depletion of leukocytes and platelets, Affymetrix HG-U133 arrays were hybridized with probe generated from the reticulocyte total RNA (blood obtained from 14 umbilical cords and 14 healthy adult humans). Among the cord and adult reticulocyte profiles, 698 probe sets (488 named genes) were detected in each of the 28 samples. Among the highly expressed genes, promoter analyses revealed a subset of transcription factor binding motifs encoded at higher than expected frequencies including the hypoxia-related arylhydrocarbon receptor repressor family. Over 100 probe sets demonstrated differential expression between the cord and adult reticulocyte samples. For verification, the array expression patterns for 21 genes were confirmed by real-time PCR (correlation coefficient 0.98). Only four transcripts (MAP17, FLJ32009, ARRB2, and FLJ27365) were identified as being upregulated in the adult blood transcriptome. Further analysis revealed that the lipid-regulating protein MAP17 was present in the membrane fraction of adult erythrocytes, but not detected in cord blood erythrocytes. Combined with other clinical and experimental data, these reticulocyte transcriptome profiles should be useful to better understand the molecular bases of terminal erythroid differentiation, hemoglobin switching, iron metabolism and malarial pathogenesis.

microarray; erythropoiesis; hemoglobin; malaria; iron

ERYTHROPOIESIS IS DEFINED by lineage commitment and proliferation of hematopoietic stem cells followed by terminal differentiation of erythroblasts into mature erythrocytes. After the final cell division, enucleation marks the complete cessation of erythroid transcription. For several days after enucleation, the cells retain some RNA in their cytoplasm and enter the circulation. Until that residual RNA is fully degraded, the cells are referred to as reticulocytes.

Reticulocytes have been utilized for a variety of research assays and other studies pertaining to erythroid biology (16). As a laboratory reagent, reticulocyte lysate continues to be useful for studies of in vitro translation (10). Reticulocytes also provide an important reagent to study the effects of erythroid gene manipulation in mice (7, 17). In the clinic, reticulocyte enumerations are regularly performed in patients with anemia or erythrocytosis. In some patients, erythroid diseases are caused by dysregulated or abnormal production of proteins during terminal differentiation. Terminal erythroid differentiation is manifest by many specialized processes including protection from oxidant damage, accumulation of hemoglobin, cytoskeleton formation, and autophagocytosis of organelles. After the final cell division and enucleation, reticulocytes complete the phenotypic specialization required for transporting oxygen and survival in the absence of new gene activity. Due to their unique position at the final stage of the erythroid developmental pathway, reticulocytes are one of the most functionally specialized cellular populations in humans.

The remnant mRNA contained in reticulocytes is hypothesized to encode a reservoir of information regarding in vivo erythropoiesis as well as more general information regarding postmitotic maturation of human cells in the absence of apoptosis. With the development of functional genomics, it has become plausible to analyze gene expression among cellular populations as entire transcriptomes rather than individual transcripts. Hematology is ideally suited for this type of genome-based research based upon the relative ease with which purified populations of hematopoietic cells may be isolated (5). Unfortunately, the amount of information gained from the reticulocyte transcriptome has been largely limited by the abundance of globin gene transcripts. According to serial analysis of gene expression among adult human reticulocytes, >70% of sequenced transcript tags encoded the adult betaglobin gene even in the absence of alpha-globin transcripts (3). Separate attempts to perform high-throughput sequencing of reticulocyte mRNA were thwarted by the high percentage of globin transcripts contained within reticulocyte expression libraries (J. L. Miller; unpublished observations). In this study, oligonucleotide arrays were utilized to overcome this problem to generate more robust profiles of the genes encoded by reticulocyte mRNA. Reticulocyte transcriptomes at the newborn and adult stages of human development are presented and compared.

MATERIALS AND METHODS

Preparation of reticulocyte RNA. All cells were collected according to approved human subjects’ guidelines. The research studies were performed at the National Institutes of Health under an exemption from 45 CFR 46, which is in accord with the principles of Helsinki. Acid citrate dextrose anticoagulated blood was collected from 14 healthy adult donors by venipuncture and 14 placental umbilical cords at term delivery. All the samples were stored at 4°C for <48 h prior to RNA extraction. After the centrifugation at 200 g for 10 min, the plasma and the buffy-coat layers were removed by aspiration. Next, the packed red blood cells were diluted with 4 volumes of 1× PBS and filtered through two RCX12 high-efficiency leukocyte reduction filters connected in series (Pall). Platelets were removed by duplicate low-speed centrifugation at 200 g at 4°C for 10 min. Special attention

Address for reprint requests and other correspondence: J. L. Miller, Molecular Medicine Branch, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, 10 Center Dr., Bldg. 10, Rm. 9B17, Bethesda, MD 20892 (e-mail: jm7f@nih.gov).

Submitted 9 November 2006; accepted in final form 26 March 2007

172
was made to process all samples in an equivalent manner. To check the purification quality, complete blood counts with reticulocyte before and after purification process (Supplemental Table S1) were performed. Contaminating white blood cells were not seen on Giemsa-stained blood smears. For RNA extraction, the purified samples were thoroughly mixed with three volumes of TRIzol-LS (Invitrogen) and stored at −80°C prior to processing. Batched samples were treated with DNaseI to degrade residual genomic DNA, and the extracted RNA was purified using RNasey Mini columns (Qiagen). Prior to array analysis, a consistent pattern of RNA molecular weight distribution with distinct 28S and 18S rRNA peaks was detected using Agilent Technologies 2100 Bioanalyzer (Agilent Technologies) in each of the 28 samples used for this study.

**Microarray data analysis.** Affymetrix HG-U133A and HG-U133B microarray hybridizations were performed using 5 μg of total RNA from each samples with one cycle of complementary RNA amplification according to the manufacturer’s protocol. After hybridization and washing, the microarrays were scanned using MicroArray Suite 5.0 software. The data were analyzed using Partek Pro 6.0 software (Partek) for statistical evaluation. Mean signal intensities and the P value measurement of significance were calculated for each probe set according to the software default settings. The probe sets encoding highly expressed genes were compiled according to the geometric mean values of signal intensities (Table 1) and Supplemental Table S2. In general, the genes identified in this manuscript are defined according to Refseq reference terminology (http://www.ncbi.nlm.nih.gov/projects/Refseq).

The raw data for 28 samples were also imported into GeneSpring 7.2 (Agilent Technologies) according to the default settings. Genes were scored as “present” or “absent” according to the MicroArray Suite 5.0 software. For this study, differentially expressed genes were identified according to the following strict criteria: 1) raw signal intensity of the mean adult blood or cord blood ≥1,000, 2) normalized intensity difference between adult blood and cord blood of more than fivefold difference, 3) P value <0.0001, 4) standard deviation of normalized intensities for each probe set within sample group <0.0625, 5) gene distinguished as present in at least 7 of 14 adult blood or cord blood samples according to the MicroArray Suite 5.0 software.

**Other bioinformatic analyses.** To analyze transcription factor binding site motif frequencies, a web-based Gene2Promoter software application was used (http://www.genomatix.de). For this purpose, the Affymetrix probe sets listed in Table 1 were queried to predict 67 human promoter sequences. Among those 67 sequences, 26 sequences were identified from RefSeq gene entries or full-length open reading frames (ORFs). The remaining 41 sequences (derived from expressed sequence tags or comparative genomics predictions) were discarded. Those 26 promoter sequences were searched for common transcription factor (TF) motifs (see Supplemental Table S5). The frequencies of shared TF motifs among the 26 promoter sequences were calculated (%frequency = number of promoters containing the TF motif/26×100). For comparison, the frequencies of TF motifs among all the vertebrate promoters contained in the database were utilized. For investigators wishing to perform analytical approaches that are not described here, the original array data were deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) (accession numbers: GSM143572–143599, GSM143671–143682, GSM143703, GSM143706–143716, GSM143718–143721).

**Quantitative real-time PCR.** For confirmation of the microarray results, quantitative PCR (QPCR) was performed on selected genes. Pooled RNA samples from five adult vs. five cord blood donors were used for the synthesis of cDNA. We amplified 50 ng of cDNA from adult and cord blood reticulocytes with the 2× TaqMan master mix and 20× Assay-on-Demand kit (Applied Biosystems), and the signal was detected by ABI 7700 real-time PCR machine. Copy numbers were derived from comparisons with the standard curve that was obtained by the simultaneous amplification of a positive control plasmid harboring a targeted amplicon using Sequence Detection Systems 1.7a (Applied Biosystems). Each reaction was performed in triplicate.

**MAP17 protein expression.** Protein from the membrane fraction of erythrocytes was prepared according to published methods (21). For the Western analysis, 15 μg of the erythrocyte membrane fraction was electrophoresed, transferred to a membrane, and hybridized with polyclonal MAP17 antibody (Novus Biologicals). The protein band

---

Table 1. **Highly represented genes in cord and adult blood reticulocytes**

<table>
<thead>
<tr>
<th>Rank CB</th>
<th>Rank AB</th>
<th>Probe Set</th>
<th>Gene Title</th>
<th>Gene Symbol</th>
<th>CB Geo Mean</th>
<th>AB Geo Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>21174.5_x_at</td>
<td>hemoglobin, alpha</td>
<td>HBA</td>
<td>440,019</td>
<td>574,210</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>209116_x_at</td>
<td>hemoglobin, beta</td>
<td>HBB</td>
<td>352,016</td>
<td>459,172</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>204848_x_at</td>
<td>hemoglobin, gamma A/gamma G</td>
<td>HBG1/ HBG2</td>
<td>339,713</td>
<td>232,207</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>231078_at</td>
<td>mitochondrial solute carrier protein, Mitoferrin</td>
<td>MSCP (MFRN)</td>
<td>292,261</td>
<td>124,047</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>200633_at</td>
<td>ubiquitin B</td>
<td>UBB</td>
<td>267,557</td>
<td>356,773</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>223649_s_at</td>
<td>CGI-69 protein</td>
<td>CGI-69</td>
<td>263,117</td>
<td>359,287</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>200077_s_at</td>
<td>ornithine decarboxylase antizyme 1</td>
<td>OAZ1</td>
<td>231,649</td>
<td>253,334</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>204466_s_at</td>
<td>synuclein, alpha</td>
<td>SNCA</td>
<td>214,052</td>
<td>130,634</td>
</tr>
<tr>
<td>9</td>
<td>19</td>
<td>240336_at</td>
<td>hemoglobin, mu</td>
<td>HBM</td>
<td>190,356</td>
<td>126,459</td>
</tr>
<tr>
<td>10</td>
<td>27</td>
<td>211560_s_at</td>
<td>aminolevulinate, delta-, synthase 2</td>
<td>ALAS2</td>
<td>171,838</td>
<td>89,844</td>
</tr>
<tr>
<td>11</td>
<td>18</td>
<td>223266_at</td>
<td>amyotrophic lateral sclerosis 2 (juvenile)</td>
<td>ALS2CR2</td>
<td>171,426</td>
<td>127,352</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>55705_s_at</td>
<td>hypothetical protein MGC16353</td>
<td>MGC16353</td>
<td>171,352</td>
<td>191,717</td>
</tr>
<tr>
<td>13</td>
<td>22</td>
<td>221479_s_at</td>
<td>BCL2 adenovirus E1B 19 kDa interacting protein 3-like</td>
<td>BNIPL3</td>
<td>164,496</td>
<td>115,566</td>
</tr>
<tr>
<td>14</td>
<td>6</td>
<td>212869_x_at</td>
<td>tumor protein, translationally-controlled 1</td>
<td>TPT1</td>
<td>147,376</td>
<td>250,092</td>
</tr>
<tr>
<td>15</td>
<td>22</td>
<td>226811_at</td>
<td>hypothetical protein FLJ20202</td>
<td>FLJ20202</td>
<td>145,293</td>
<td>20,949</td>
</tr>
<tr>
<td>16</td>
<td>42</td>
<td>224752_at</td>
<td>Homo sapiens, clone IMAGE:5285814</td>
<td>IMAGE:5285814</td>
<td>136,496</td>
<td>58,932</td>
</tr>
<tr>
<td>17</td>
<td>12</td>
<td>221700_s_at</td>
<td>ubiquitin A-52 residue ribosomal protein fusion product 1</td>
<td>UBA52</td>
<td>135,863</td>
<td>180,789</td>
</tr>
<tr>
<td>18</td>
<td>25</td>
<td>213187_x_at</td>
<td>ferritin, light polypeptide</td>
<td>FTL</td>
<td>130,600</td>
<td>108,991</td>
</tr>
<tr>
<td>19</td>
<td>14</td>
<td>201178_at</td>
<td>F-box only protein 7</td>
<td>FBXO7</td>
<td>125,467</td>
<td>156,995</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>210774_s_at</td>
<td>nuclear receptor coactivator 4</td>
<td>NCOA4</td>
<td>124,879</td>
<td>13,373</td>
</tr>
</tbody>
</table>

CB, cord blood; AB, adult blood; GeoMean: geographic mean.

---

1 The online version of this article contains supplemental material.
was visualized using anti-rabbit-horseradish peroxidase secondary antibody (Amersham, Piscataway, NJ). Human kidney lysate (ProSci) was used as the positive control.

RESULTS

Purification of reticulocytes. A negative selection strategy was utilized to obtain purified populations of erythroid elements circulating in peripheral blood. Leukocytes were removed by filtration, and platelets were depleted by repeated low-speed centrifugation. Depletion of leukocytes and platelets was confirmed to levels near the sensitivity limits of automated counting as shown in Supplemental Table S1. While erythroid cells were the only population visualized on the blood smears, contaminating white blood cells were estimated as 0.001% of the total cellular populations. Density gradient purification was attempted separately, but contamination by white blood cells was more pronounced (data not shown). Despite the depletion of leukocytes and platelets, the percentage of reticulocytes in the purified populations remained stable. As predicted (2), a higher percentage of reticulocytes were detected in the cord blood samples (3.3 ± 0.9% for cord blood after purification vs. 2.2 ± 1.2% for adult blood after purification, t-test P value = 0.01) (Supplemental Table S1). Additional purification or separation of reticulocytes into subpopulations was not performed.

Transcripts detected at high levels. To generate a subset of genes with transcripts detected at the highest levels among remnant mRNA from cord and adult blood reticulocytes, the mean signals were sorted for the 44,854 probe sets contained on the microarrays. The 27 gene transcripts detected at the highest levels (ranked among the top 20 according to mean signal intensity from either cord or adult profiles) were compiled as shown in Table 1. A broader list containing the 100 transcripts with the highest mean signal intensities in reticulocyte mRNA from cord vs. adult blood is provided in Supplemental Table S2. For gene transcripts hybridized with multiple probe sets, the probe set with the highest mean intensity is shown. A complete list of probe sets encoding the globin genes is provided in Supplemental Table S3. The alpha1 globin gene was detected at the highest levels in both of the cord and adult blood groups. Other globin gene transcripts identified as being very highly represented in the transcriptome profiles included beta-globin, gamma-globins and mu-globin. The high level of beta-globin mRNA in cord blood reflects the relatively advanced stage of globin gene switching in reticulocytes at the time of birth (11). Even though the gamma-globin genes were detected at lower levels in adult blood, their signals were still ranked 9th among the highly expressed genes. Interestingly, the newly discovered alpha-like globin gene, mu-globin was ranked 9th in cord blood and 19th in adult blood. Delta-globin was ranked 24th and 48th in adult and cord blood, respectively (Supplemental Table S2). Eight other genes (CGI-69, OAZ1, ALAS2, BNIP3L, FTL, MSCP, TPT1, and GYPC) known to be expressed during terminal erythroid differentiation or play key roles in erythroid specialization were also identified here. Several ribosomal proteins were also noted on the broader list. FBXO7, UBB, and UBA52 encode genes that function in protein ubiquitination. The polyamine regulator OAZ1, the Parkinson’s disease-associated gene SNCA, the transcription related factor NSEPI, and three uncharacterized transcripts were also identified.

Genomic overview of the highly expressed genes. It is well known that the globin transcripts are expressed from two developmentally regulated groups of genes on chromosomes 11 and 16. The very high level of gene activity within those genes may be related to insulated or protected chromosomal domains (12, 13). To determine whether genes in the vicinity of the globin loci possess similarly high-level gene expression, the signal intensities for probes that hybridize to adjacent genes (defined here as genomic regions defined by RefSeq ID) were examined (Fig. 1). The developmentally predicted patterns of high-level expression were observed for the alpha- and beta-globin gene clusters in the cord and adult blood samples. However, comparably high levels of transcripts encoding the surrounding genes were not detected. Analyses of other highly abundant nonglobin transcripts from Table 1 demonstrated similarly low probe intensity levels from adjacent genes (Supplemental Table S4).

Promoter motif comparison among highly expressed genes. Based upon the shared pattern of very high level transcripts among those genes shown in Table 1, it was hypothesized that
these genes may possess shared mechanism(s) of transcriptional regulation. As an initial study in this regard, transcription factor binding motifs were determined for the promoter regions of the genes shown in Table 1. As described in MATERIALS AND METHODS, 26 promoter sequences were identified using promoter prediction software from the RefSeq or full-length ORFs corresponding to the Table 1 gene list. The frequencies of particular transcription factor binding motifs within these promoters (number of promoters containing the TF motif/26 × 100) were compared with the frequency of those motifs within all the vertebrate promoters contained in the database (Genomatix MatBase vertebrate TF motif frequency). 143 TF families were determined to possess motifs within the highly expressed gene promoters. Among these, the 15 TF families with the highest frequencies within the 26 promoters are shown in Fig. 2 (for a complete list, see Supplemental Table S5). With the exception of the TATA binding motif (TBPF), the most commonly identified motifs were overrepresented within this group of 26 promoters compared with the frequencies predicted by the software for all vertebrate promoters. Among the TF group, both EKLF and Tal1 (EBOX) binding motifs were included. The most overrepresented TF motif was the AHRR consensus binding motif (73.1% of the promoters from highly expressed genes versus vs. 36.6% of vertebrate promoters). The AHRR (arylhydrocarbon receptor repressor) family members include AHR, a transcription factor that may be involved in hypoxia-related signaling through binding to arylhydrocarbon nuclear translocator (8). GATA binding motif was identified in 58% of the promoters of the genes described in Fig. 1 compared with 73% of vertebrate promoters identified by the database (Supplemental Table S6).

Comparison of the cord and adult blood transcriptomes. In addition to those transcripts detected at very high levels, analyses of the entire data set were also performed according to the MAS 5.0 software-defined absent or present algorithm; 1,285 and 806 probe sets were annotated as present in all cord blood or in adult blood samples, respectively. The probe set IDs are provided in Supplemental Table S7 and summarized by the Venn diagram shown in Fig. 3. In total, 698 probe sets (488 named genes) were identified as present in all 28 samples. An additional 587 probe sets (454 named genes) were identified in each of the cord blood samples but not among the entire group of adult blood samples. In contrast, 108 probe sets (98 named genes) were identified in each of the adult blood samples but not consistently among the cord blood samples.

Based upon the large number of differences in the profiles determined by absent or present annotations, a more quantitative approach was explored to determine genes that may be differentially represented by the cord and adult blood transcriptomes. For this purpose, GeneSpring 7.2 software was utilized to compare the signal intensities for each probe set. With the application of filtering to identify probe sets with distinct signals between the cord and adult blood samples (see MATERIALS AND METHODS), a list of 107 probe sets encoding 94 genes was generated from 44,854 probe sets (Supplemental Table S8). Those probe sets were clustered into a conditional tree by unsupervised hierarchical clustering as shown in Fig. 4. Only four probe sets demonstrated higher level expression...
in the adult reticulocytes (219630_at-MAP17, 242957_at-FLJ32009, 203388_at-ARRB2, and 238058_at-FLJ27365). The vast majority of differentially expressed transcripts were detected at higher levels in the cord blood samples. However, the distinctly higher levels noted for those genes in Fig. 4 were not detected by pangenomic comparisons. When all 44,854 probe sets were compared, the geometric means of the raw signal intensities were $1,078_{\text{H11006}}^{38}$ (mean $SE$) for cord blood and $1,037_{\text{H11006}}^{43}$ (mean $SE$) for adult blood. Functional classifications of 107 differentially expressed probe sets were performed using GoMiner (22) software to demonstrate that the encoded genes are associated with a wide range of cellular and molecular functions (data not shown).

Validation of microarray data. For validation, 21 probe sets were chosen according to the ratio of signals from cord blood and adult blood arrays. The ratios were calculated using array signal intensities, and these ratios were compared with the copy number/ng cDNA ratio from real-time PCR (Fig. 5). Three genes (ADIPOR1, CFTR, and ATP6V0C) with adult blood/cord blood ratios between 1 and 2 demonstrated disparate results between the two measuring methods. The remaining 17 genes with array signal intensities ratios of $>0.5$ or $<2$
Differences between the cord and adult reticulocyte transcriptomes were clearly seen. Interestingly, increased representation of differentially expressed genes was highly biased toward the cord blood transcriptome. Since the methods used for collection, storage, and processing of the 28 samples were consistent, no obvious experimental or technical artifacts were identified for these differences. Instead, biological differences between cord and adult blood erythropoiesis and reticulocyte release into the circulation may be involved. It is well recognized that cord blood contains less-mature reticulocytes than adult blood (15). Therefore, the increased representation of genes in cord blood reticulocytes may be related to the presence of those less-mature reticulocyte populations. Beyond differences in the reticulocyte populations, the differential representation transcripts may also reflect developmental changes in erythroid gene expression (11) or mRNA stability. Curiously, two poly(A) binding proteins (215823_x_at, PABPC1; 222984_at, PAIP2) were identified as being more abundant in cord blood.

Fig. 5. Validation of microarray data using quantitative real-time PCR (Q-PCR). We selected 21 genes for real-time PCR for comparison with the array-based data. The y-axis shows Log (AB/CB ratio) values for array signals (array, black bars) vs. Q-PCR copy numbers (Q-PCR, gray bars). The gene annotations are shown on the x-axis.

Despite the validation of differential protein expression of MAP17 among cord and adult human erythrocytes, the transcriptome profiles are not expected to completely and quantitatively reflect the protein content of reticulocytes. Differential levels of mRNA and protein stability make the degree of overlap between reticulocyte transcriptome and proteome profiles difficult to predict. For instance, the differential representation of glycophorin A among cord and adult blood profiles is not borne out at the protein level (4). Furthermore, there are transcripts present at very high levels in all the samples for which the encoded proteins are difficult to detect. Perhaps the best example of this phenomenon is portrayed by the newly discovered mu-globin gene. Mu-globin mRNA is clearly detected in erythroid cells by array, Northern, and RT-PCR assays. However, mu-globin protein chains or assembled hemoglobin molecules are not expressed at sufficient levels for identification by standard chromatography techniques (5). Interestingly, mass spectroscopy techniques used for mature erythrocyte proteome profiling recently provided the first evidence that the mu-globin protein exists in vivo (14). In these and other examples, the data suggest that transcriptome and proteome profiling provide complementary, rather than duplicate, descriptions of genome expression. Comparisons of transcriptome and proteome data gathered concurrently from the same cellular samples may be necessary to most accurately determine the level of overlap between these two approaches. Ultimately, these transcriptome data should be integrated with proteome and other high-throughput studies of human reticulocytes.

Fig. 6. Detection of MAP17 protein in adult erythrocytes. Erythrocyte membrane proteins from three cord blood (lanes 1–3) and three adult blood (lanes 4–6) samples were analyzed. Kidney lysate was used as a positive control (C). Each lane was loaded with 15 µg of erythrocyte membrane protein lysate.
www.ncbi.nlm.nih.gov/geo/) that might be used by the global community to better understand the molecular bases of terminal erythroid differentiation, hemoglobin switching, iron metabolism, and malarial pathogenesis. It is predicted that the gene expression patterns that underlie chromatin condensation, enucleation, autophagocytosis, mitochondrial integrity, and membrane specialization are contained within the transcriptome profiles. Several nonglobin transcripts were detected at very high levels in reticulocytes from cord and adult blood including ALAS2, ferritin, and glycophin C. ALAS2CR2 (also known as ILPPIP) is an apoptosis-related gene identified within this select group of highly expressed genes (18). High-level expression of another apoptosis related gene Bnip3L (Nix) was also detected. Nix expression in erythroblast mitochondria was reported previously (1). Another mitochondrial gene identified among the highly represented transcripts is MSCP (mitochondrial solute carrier protein, also known as mitoferrin). Very recently, mitoferrin was determined to be essential for erythropoiesis in the zebrafish model system (19). The presence of mitoferrin transcripts present at such high levels in human reticulocytes suggests the importance of this gene for erythropoiesis may be conserved in humans. The promoter allocation of particular TF motifs including that of EKLF (Fig. 2) also suggests that these genes may also share a common mode(s) of transcriptional regulation. The relative overrepresentation of AHR binding motifs in the promoter regions of these genes is of particular interest. Since hypoxia is of fundamental importance for the regulation of erythropoiesis, further characterization of the AHR transcription factor in erythroid cells is warranted.

Preliminary studies of MAP17 RNA and protein expression were provided as an example of how this large body of novel information may be useful for future studies of erythroid biology. Despite decades of research regarding the composition of erythrocyte membranes, MAP17 expression in erythrocytes was not reported. Remarkably, MAP17 is the first erythrocyte membrane protein with a fetal-to-adult pattern of increased expression in humans. Originally, MAP17 was identified as a small protein associated with PDZK1, and it is thought to be involved in the regulation of plasma high-density lipoprotein levels (20). The developmentally increased MAP17 expression in adult erythroid cells is an especially interesting finding, since the human fetal-to-adult transition is also associated with a major shift in plasma lipid levels in humans (9). As such, this discovery provides an intriguing clue that erythrocytes may be involved in plasma lipid regulation. Erythrocytes were also very recently shown to regulate levels of sphingosine 1-phosphates in blood (6). Thus, our discovery of regulated MAP17 expression during human ontogeny provides an excellent example of how this massive dataset may facilitate the progress of new avenues of research. It is predicted that many other novel and unexpected discoveries will be forthcoming as this clinical dataset is integrated with experimental model systems and other sources of information regarding erythroid biology and disease.

ACKNOWLEDGMENTS
The Department of Transfusion Medicine at National Institutes of Health (NIH) is acknowledged for assistance in the collection of blood samples, as well as members of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Microarray Core Facility for assistance.

GRANTS
This research was supported by the Intramural Research Program of the NIH, NIDDK.

REFERENCES