ELSEVIER

Experimental Hematology

Experimental Hematology 2009;37:285–293

Identification of four novel developmentally regulated gamma hemoglobin mRNA isoforms

Michelle Alvarez[a,c,*] and Jack Ballantyne[a,b,c]

[a]*Graduate Program in Biomolecular Science;* [b]*Department of Chemistry, University of Central Florida, Orlando, Fla., USA;* [c]*National Center for Forensic Science, Orlando, Fla., USA*

(Received 29 August 2008; revised 1 October 2008; accepted 9 October 2008)

Objective. **In the course of developing assays for the molecular prediction of biological age, we serendipitously discovered four novel isoforms of gamma hemoglobin mRNA, designated HBG1n1, HBG1n2, HBG2n2, and HBG2n3, collectively termed HBGn isoforms. Here we report the molecular characterization and tissue expression of these isoforms.**

Materials and Methods. **RNA obtained from human peripheral blood and various fetal and adult tissues was amplified with duplex reverse transcription polymerase chain reaction (RT-PCR) assays to determine the expression profiles of the HBGn isoforms. To determine if their molecular origin was either a genomic recombination or an undefined RNA rearrangement event, DNA and RNA samples were pretreated with RNaseI and/or DNaseI and then analyzed with the duplex RT-PCR assays.**

Results. **Alignment analysis indicated that the HBG1n(1/2) and HBG2n(2/3) isoforms were identical to the expected parental HBG1 or HBG2 sequences except for unique deletions that spanned the 3′ end of exon two, the entire second intron, and the 5′ end of exon three. RT-PCR duplex reactions revealed that the isoforms exhibit restricted expression to fetal hematopoietic tissue and newborn peripheral blood and appear to be temporally regulated during development. Finally, nucleic acid digestion experiments illustrate that the isoforms appear to be created by an RNA rearrangement event between penta- or octanucleotide direct repeats in adjacent exons.**

Conclusion. **The precise genesis and function of these novel isoforms is unknown at present. Interestingly, each of the four isoforms identified maintain an open reading frame as well as 5′ and 3′ regulatory regions invoking the possibility that they encode a family of related polypeptides.**

The life cycle of humans is composed of a number of developmentally recognized stages, some of which are manifested by differential gene expression. A good example of developmental age−related differential gene expression involves β-globin gene switching [1]. The human β-hemoglobin locus is located on the short arm of chromosome 11 (11p15.5), and encodes five functional β-like globin genes, ϵ, $^{G}\gamma$, $^{A}\gamma$, δ, and β, and a nonfunctional β-pseudogene ($^{\beta}\psi$) [2]. The expression of embryonic hemoglobin (ϵ-globin) commences in the yolk sac in the early stages of gestational development, approximately during week 2 and continues until 6 weeks (37 days) postconception [3]. During the next 6 weeks of gestation (days 37−79), the newly developed fetal liver and fetal spleen begin to produce the gamma globin chains ($^{A}\gamma$ and $^{G}\gamma$) of fetal hemoglobin [3]. This increased production of γ-globin is accompanied by a shutdown of ϵ-globin synthesis. Beginning at approximately 20 weeks' gestation and continuing throughout life, adult β-globin gene expression commences in the bone marrow and γ-globin expression is downregulated [3].

Previously, in the course of developing assays for the molecular prediction of biological age in a forensic context, we investigated the possibility that detection of γ-globin (HBG) mRNA in a bloodstain would be indicative of the stain originating from a newborn baby [4]. Although γ-globin is expressed in the fetal liver, we predicted (and confirmed) that circulating red blood cells from newborn infants would contain HBG mRNA produced during the terminal stages of

*Dr. Alvarez is currently at Science Applications International Corporation, SAIC, San Diego, CA, USA.

Offprint requests to: Jack Ballantyne, Ph.D., Department of Chemistry, University of Central Florida, Building 5, 4000 Central Boulevard, Orlando, FL 32816-2366; E-mail: jballant@mail.ucf.edu

0301-472X/09 $–see front matter.
doi: 10.1016/j.exphem.2008.10.003

differentiation. While investigating HBG expression, we serendipitously discovered four novel isoforms of gamma-hemoglobin mRNA, designated HBG1n1, HBG1n2, HBG2n2, and HBG2n3, collectively termed HBGn, which exhibited age-dependent gene expression in that they were confined to newborn individuals. These transcript isoforms represent a novel class of truncated HBG mRNAs that possess a partial deletion of exon 2 and exon 3, potentially arising from an unreported RNA rearrangement event between direct repeats in adjacent HBG exons. Here we report the molecular characterization and tissue expression of these four isoforms.

Materials and methods

Sample preparation and bloodstain RNA extraction

Human blood samples were collected from Florida Hospital (Orlando, FL) after approval by the Hospital's Institutional Review Board and in accordance with procedures approved by the University's Institutional Review Board. Bloodstains were made by dispensing 50-μL aliquots onto sterile cotton gauze, allowed to air-dry overnight at room temperature, and stored at −45°C until needed. Total RNA was extracted from the 50-μL bloodstains using a modified Chomczynski and Sacchi method, as described in previous publications [4–6]. Fetal tissue (i.e., brain, heart, kidney, liver, spleen, thymus) and adult tissue (i.e., adipose, bone marrow, brain, kidney, liver, lung, skeletal muscle, skin, salivary gland, spleen, thymus, testis, uterus) total RNA was obtained from BD Biosciences Clontech (Palo Alto, CA, USA). Fetal liver, fetal spleen, adult adipose, and adult bone marrow DNA was obtained from BioChain Institute, Inc. (Hayward, CA, USA).

DNA isolation

Genomic DNA was extracted from 50-μL bloodstains by an organic solvent extraction method [7]. Briefly, samples were incubated overnight at 56°C in stain extraction buffer (0.1 M NaCl, 10 mM Tris-HCl [pH 8.0], 25 mM ethylene diamine tetraacetic acid [EDTA; pH 8.0], 20 mM sodium dodecyl sulfate) supplemented with 0.5 mg/mL proteinase K. An equal volume of phenol/chloroform/isoamyl alcohol (25:24:1 [pH 6.6]) was added to the extract, mixed gently by inversion and centrifuged for 5 minutes at 16,000*g* to separate the phases. The DNA containing aqueous layer was transferred to a sterile microcentrifuge tube, precipitated with 1 mL ice-cold 100% ethanol overnight, centrifuged at 16,000*g* for 15 minutes, washed once with 1 mL 70% ethanol and solubilized in 100 μL TE^{-4} (10 mM Tris, 0.1 mM EDTA) at 56°C overnight.

Quantification of nucleic acids

RNA was quantified using a sensitive fluorescence assay based upon the binding of the unsymmetrical cyanine dye RiboGreen (Molecular Probes, Eugene, OR, USA) [8]. The manufacturer's "high-range assay" instructions were followed. Briefly, 200 μL total volume assays comprised 2 μL TURBO DNase-treated RNA extract, 98 μL TE buffer (10 mM Tris-HCl, 1 mM EDTA [pH 7.5] in nuclease free water), and 100 μL 750 nM RiboGreen reagent in a 96 well plate format. After RiboGreen addition and a 3-minute incubation at room temperature protected from light, fluorescence emission at 535 nm (excited at 485 nm) was determined using a Wallac Victor2 microplate reader (Perkin Elmer Life Sciences, Boston, MA, USA). RNA concentration was calculated using an appropriate standard curve as described by the manufacturer [8]. All RNA samples were diluted to 5 ng/μL with nuclease-free water (Ambion Inc., Austin, TX, USA).

DNA was quantified by electrophoresis on 1% agarose gels stained with ethidium bromide and compared to known reference standards. All DNA samples were diluted to 1 ng/μL with TE^{-4} (10 mM Tris, 0.1 mM EDTA).

cDNA synthesis

For blood and tissue RNA samples 6 μL RNA (30 ng) was heated at 75°C for 3 minutes, snap-cooled, then added to the reverse transcription (RT) reaction tube. The (+RT) reaction tube contained 4 μL of a 10-mM dNTP mix (Applied Biosystems, Foster City, CA, USA), 2 μL 10× first strand buffer (500 mM Tris-HCl [pH 8.3], 750 mM KCl, 30 mM $MgCl_2$, 50 mM dithiothreitol), 2 μL Random Decamer primers (50 μM) (Ambion), 20 U SUPERase In RNase Inhibitor (20 U/μL) (Ambion), and 100 U Moloney Murine Leukemia Virus Reverse Transcriptase (100 U/μL) (Ambion). Nuclease free water (Ambion) was added to yield a final reaction volume of 20 μL. For the (−RT) reaction tubes, the Moloney Murine Leukemia Virus Reverse Transcriptase was replaced with nuclease free water. Reaction mixtures were incubated at 42°C for 1 hour, followed by a 10-minute incubation at 95°C to inactivate the RT enzyme [9].

PCR conditions

All amplification reactions were conducted with genomic DNA (2 ng) or cDNA (3 ng) (except for HBG1 and HBG2 singleplex reactions, which contained 5 ng cDNA). A standard reaction mix containing 1× PCR buffer (10 mM Tris-HCl [pH 8.3], 50 mM KCl), 1.5 mM $MgCl_2$, 0.125 mM each dNTP, 1.25-units AmpliTaq Gold DNA polymerase (5 U/μL) (Applied Biosystems) and nuclease free water (Ambion) were combined to yield a 25-μL final reaction volume. Primer sequences for all genes assayed with their corresponding amplimer sizes are listed in Table 1. All single-gene reactions contained 400 nM each primer. Primer concentrations for duplex reactions were as follows: HBG1n1-GNAS and HBG1n2-GNAS consisted of 400 nM HBG1n(1/2) primers and 100 nM GNAS primers, and HBG2n2-GNAS and HBG2n3-GNAS consisted of 200 nM HBG2n(2/3) primers and 400 nM GNAS primers.

Standard PCR conditions were used for all amplifications and consisted of an initial incubation step (95°C; 11:00) followed by repeating cycles of [denaturation (94°C; 20 seconds), annealing (60°C; 30 seconds), and extension (72°C; 40 seconds)] with a final incubation of (72°C; 10 minutes) [10,11]. Cycle number for amplifications were as follows: singleplex HBG1 and HBG2 (32 cycles; 55°C annealing), singleplex HBG1-GNAS and HBG2-GNAS (35 cycles), duplexes HBG1n1-GNAS and HBG1n2-GNAS (30 cycles) and duplexes HBG2n2-GNAS and HBG2n3-GNAS (28 cycles).

Molecular origin of novel HBG transcripts

Tissue DNA (43 ng) and blood DNA (64 ng) was treated with 5 U TURBO DNase (RNase Free) (2 U/μL) (Ambion) or tissue DNA (50 ng) and blood DNA (75 ng) was treated with 100 U RNaseI (100 U/μL) (Ambion) at 37°C for 30 minutes. The DNaseI enzyme was inactivated by incubation at 75°C for 10 minutes [12,13]. The RNaseI enzyme was inactivated by the addition of 50 μL phenol/chloroform/isoamyl alcohol (25:24:1 [pH 6.6]).

Table 1. Gene names, primer sequences, and expected amplicon sizes for DNA and reverse transcribed-mRNA amplifications.

Gene	Primer Sequences	DNA (bp)	mRNA (bp)
GNAS			
Forward	5' AAG-ATC-GAC-GTG-ATC-AAG-CA 3'	855	371
Reverse	5' CCA-TCC-TTG-GGA-GAT-GCC-AC 3'		
HBG1			
Forward	5' ACT-TCC-TTG-GGA-GAT-GCC-AC 3'	1157	277
Reverse	5' AAA-GCC-TAT-CCT-TGA-AAG-CTC-TGA 3'		
HBG2			
Forward	5' ACT-TCC-TTG-GAT-GCC-AT 3'	1060	274
Reverse	5' GCC-TAT-CCT-TGA-AAG-CTC-TGC 3'		
HBG1n1			
Forward	5' GAA-AGC-TCT-GAA-TCA-TCC-AGG-TG 3'*	0	207
Reverse	5' GGG-CAA-GGT-GAA-TGT-GGA-AG 3'		
HBG1n2			
Forward	5' AGT-GAG-CTC-AGT-GGC-ATC-TC 3'*	0	190
Reverse	5' GGG-CAA-GGT-GAA-TGT-GGA-AG 3'		
HBG2n2			
Forward	5' CTG-GAG-GAC-AGG-GCA-AAG-G 3'*	0	225
Reverse	5' GGG-CAA-GGT-GAA-TGT-GGA-AG 3'		
HBG2n3			
Forward	5' GGC-AGT-GAG-CTC-AGT-GCA-GTT-C 3'*	0	161
Reverse	5' CAG-CTT-TGG-CAA-GTC-CT 3'		

*Underlined sequence identifies the location of the newborn hemoglobin isoform breakpoint.

RNaseI-treated DNA samples were then centrifuged at 16,000*g* (5 minutes) and the aqueous phase transferred to a new tube. Only 1 μL of these DNA samples (untreated, +DNaseI, +RNaseI) was added to the PCR reaction.

Blood and tissue total RNA (250 ng) was treated with 250 U RNaseI (100 U/μL) (Ambion) at 37°C for 30 minutes and inactivated with 60-μL acid phenol:chloroform 5:1 [pH 4.5] (Ambion). The samples were centrifuged (10 minutes at 16,000*g*) and the aqueous phase precipitated at −20°C overnight, with 15 μg GlycoBlue glycogen carrier (Ambion) and 50 μL isopropanol. Samples were then centrifuged at 16,000*g* (20 minutes), the supernatant removed and the pellet washed once with 200 μL 75% ethanol/25% diethylpyrocarbonate treated water, recentrifuged at 16,000*g* (10 minutes), the pellet was dried in a vacuum centrifuge, and finally resolubilized in 10 μL RNAsecure Resuspension Solution (Ambion) at 60°C for 10 minutes. The RNA samples were then reverse transcribed (see above) and 2 μL added to PCR.

Cloning and sequencing of PCR products

Low molecular weight amplimers from the HBG1 and HBG2 amplifications were excised from agarose gels and purified using MERmaid SPIN columns, which specifically isolate low molecular weight DNA products (10−200 bp) (Q-BIOgene, Carlsbad, CA, USA). Purified products were cloned into TOP10F' One Shot chemically competent cells using the TOPO TA Cloning Kit (pCR2.1-TOPO) (Invitrogen, Carlsbad, CA, USA). Positive colonies were isolated and plasmids purified using the RapidPURE Plasmid Mini Kit (Q-BIOgene). Plasmids that contained the inserted product were sequenced by standard Sanger sequencing (Lark Technologies, Inc., Houston, TX, USA).

Postamplification electrophoresis

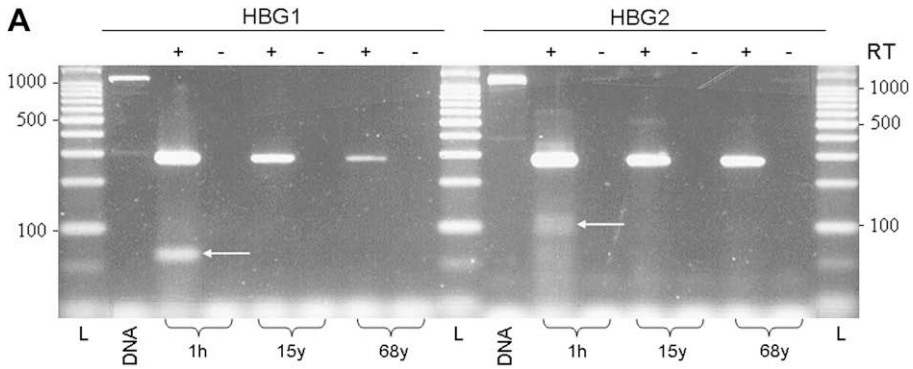
PCR and RT PCR amplified products were visualized on 4% NuSieve GTG Agarose gels (Cambrex Bio Science Rockland, Inc., Rockland, ME, USA). Electrophoresis was carried out at 100 V for 1.25 hours in TAE (0.04 M Trisacetate, 0.001 M EDTA) buffer. Gels were stained with SYBR Gold nucleic acid stain (Molecular Probes, Eugene, OR, USA), visualized on the Omega10 Chemiluminescence Imaging System (ULTRA LUM, Inc., Claremont, CA, USA) and analyzed with ONE Dscan 2.05, 1 D Gel Analysis Software for Windows (Scanalytics, Inc., Fairfax, VA, USA).

Results

Identification of newborn-specific gamma-hemoglobin mRNA isoforms

Expression of hemoglobin A- gamma (HBG1) and G- gamma (HBG2) in blood from individuals of different ages was evaluated by extracting RNA from venous blood samples donated from a 1-hour-old newborn, a 15-year-old juvenile, and a 68-year-old elderly individual. Messenger RNA was reverse-transcribed and the corresponding cDNA, along with a genomic DNA control, was amplified using primers designed to specifically recognize either the HBG1 or HBG2 gene transcripts (Fig. 1A). The forward and reverse primers were designed to bind to sequences in exons two and three of both genes in order to electrophoretically distinguish *bona fide* cDNA from genomic DNA amplified products. The expression of both HBG1 (277 bp) and HBG2 (274 bp) transcripts was detected in all three ages tested (Fig. 1A). Despite uniform HBG mRNA expression, we noted the presence of distinct lower molecular weight products in the 1-hour-old (+RT) newborn blood sample (illustrated with arrows in Fig. 1A) at approximately 65 bp and 100 bp, in the HBG1 and HBG2 assays, respectively. To verify the presence and specificity of these lower molecular weight products to newborn individuals, we repeated the HBG1 and HBG2 amplifications with multiple individuals aged 1 hour to 89 years. All newborns tested (n = 16) successfully produced a lower molecular weight product with both assays, whereas all other individuals (n = 20) did not produce the low-molecular weight amplimers (data not shown) [4]. These low molecular weight newborn-specific amplimers were then excised from the gel and sequenced.

Sequencing results from the HBG1 and HBG2 low molecular weight amplicons revealed that each band in the gel was actually comprised of two separate amplimers of indistinguishable size. We utilized the (NCBI) human genome BLAST alignment tool to determine the potential origin of the four amplified products. Alignment analysis indicated that the lower molecular weight products (designated HBG1n1, HBG1n2 and HBG2n2, HBG2n3) were identical



B HBG1

[1]acactcgcttctggaacgtctgaggttatcaataagctcctagtccagacgccatgggtcatttcacagagg
aggacaaggctactatcacaagcctgtggggcaaggtgaatgtggaagatgctggaggagaaaccctgg
gaag[145]gctcctggttgtctacccatggacccagaggttctttgacagctttggcaacctgtcctctgcctctg
ccatcatgggcaaccccaaagtcaaggcacatggcaagaaggtgctgacttccttgggagatgccacaaa
gcacctggatgatctcaagggcaccttgcccagctgagtgaactgcactgtgacaagctgcatgtggatcc
tgagaacttcaag[368]ctcctgggaaatgtgctggtgaccgttttggcaatccatttcggcaaagaattcaccc
ctgaggtgcaggcttcctggcagaagatggtgactgcagtggccagtgccctgtcctccagataccactga
gctcactgcccatgattcagagctttcaaggataggctttattctgcaagcaatacaaataataaatctattctg
ctgagagatcac[584]

C HBG2

[1]acactcgcttctggaacgtctgaggttatcaataagctcctagtccagacgccatgggtcatttcacagagg
aggacaaggctactatcacaagcctgtggggcaaggtgaatgtggaagatgctggaggagaaaccctgg
gaag[145]gctcctggttgtctacccatggacccagaggttctttgacagctttggcaacctgtcctctgcctctg
ccatcatgggcaaccccaaagtcaaggcacatggcaagaaggtgctgacttccttgggagatgccataaa
gcacctggatgatctcaagggcaccttgcccagctgagtgaactgcactgtgacaagctgcatgtggatcc
tgagaacttcaag[368]ctcctgggaaatgtgctggtgaccgttttggcaatccatttcggcaaagaattcaccc
ctgaggtgcaggcttcctggcagaagatggtgactggagtggccagtgccctgtcctccagatacactgga
gctcactgcccatgatgcagagctttcaaggataggctttattctgcaagcaatcaaataataaatctattctgc
taagagatcac[583]

Figure 1. Identification of novel newborn-specific HBG mRNA isoforms. (**A**) Total RNA was isolated from three individuals of differing ages [newborn (1 hour), juvenile (15 years), and elderly (68 years)] reverse transcribed and amplified using HBG1- and HBG2-specific primers. Arrows indicate the newborn-specific low molecular weight amplimers. Gene sequences of (**B**) HBG1 and (**C**) HBG2 with the forward and reverse primer binding sites indicated (^^^^). RNA rearrangement breakpoints of the HBGn isoforms occur within short direct repeats (labeled by circles or boxes) that result in deletion of the intervening sequences. DNA = genomic DNA control; L = 100-bp DNA ladder; h = hour; y = year; RT = reverse transcriptase; superscript numerals refer to base positions within the three contiguous exons, and define the exon-intron boundaries.

to the parental HBG1 and HBG2 sequences and did not exhibit significant sequence homology with any other region of the human genome. Although the isoforms aligned with their respective parent transcripts, unique deletion segments were identified within each isoform which spanned the 3′ end of exon two, the entire second intron and the 5′ end of exon three. Significantly, the deletion breakpoints of the isoforms were associated with, and appeared to occur within, penta- or octanucleotide direct repeat sequences that were present in both exons 2 and 3 (Fig. 1B and C). The HBG1n1 and HBG1n2 isoforms of HBG1 possessed deletion breakpoints within the ATGAT and AGATG/ACCA direct repeats, respectively, resulting in 217 and 214 base deletions (Fig. 1B). The HBG2n2 and HBG2n3 isoforms of HBG2 possessed deletion breakpoints within TGCCC and CACTG direct repeats that, in both instances, coincidentally resulted in a deletion of 162 bases (Fig. 1C).

Tissue expression of the HBG1n(1/2) and HBG2n(2/3) isoforms

Based upon the newly discovered HBG1n(1/2) and HBG2n(2/3) sequences, four separate gel-based duplex RT-PCR assays were designed to detect the presence of these transcripts in various tissue samples. Isoform-specific forward primers were designed to span each of the breakpoints to prevent coamplification of the parental HBG1 and HBG2 genes (Table 1). Each assay coamplified a specific HBGn transcript, either HBG1n1, HBG1n2, HBG2n2, or

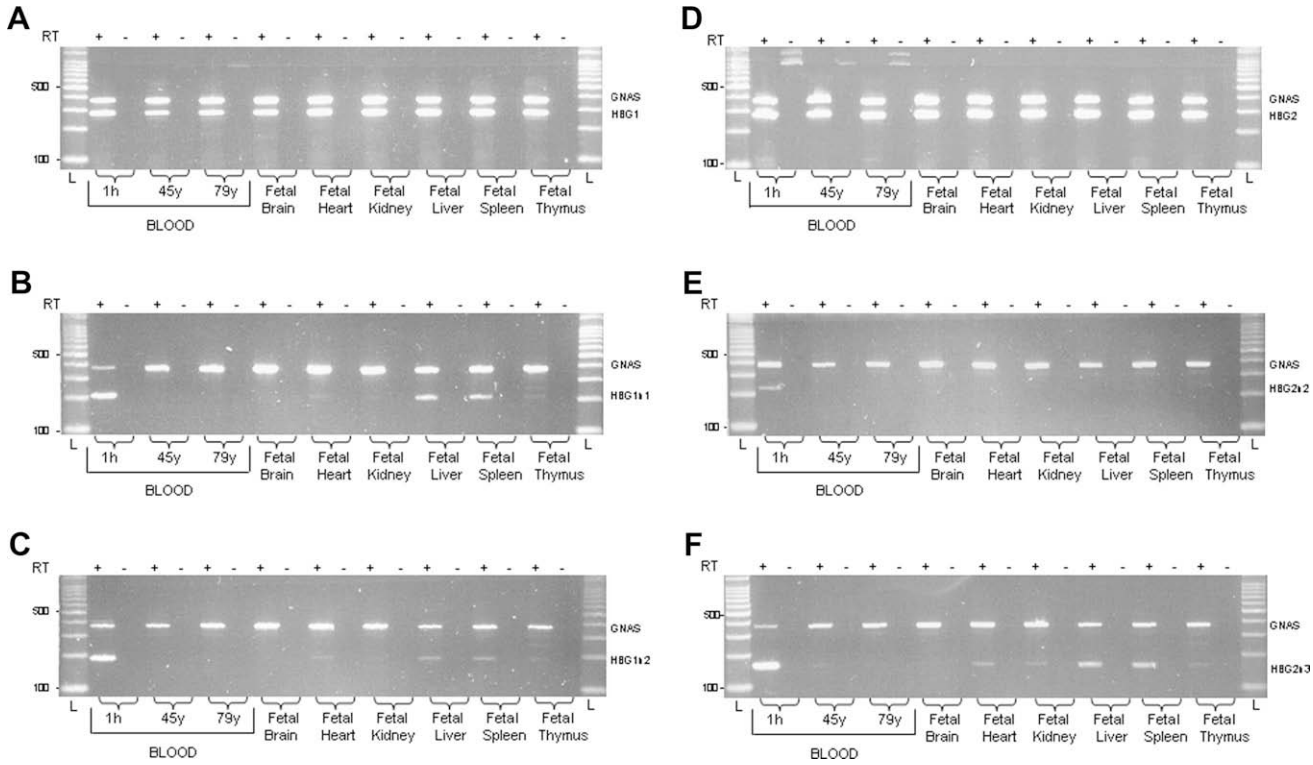


Figure 2. Expression of the HBGn isoforms in fetal tissue. Tissue expression of HBG1 and HBG2 and their isoforms in fetal brain, heart, kidney, liver, spleen, and thymus are illustrated. Each assay amplified the GNAS housekeeping gene to control for RNA quality and quantity, with the (**A**) HBG1, (**B**) HBG1n1, (**C**) HBG1n2, (**D**) HBG2, (**E**) HBG2n2, and (**F**) HBG2n3 genes. As controls, blood extracted RNA from a newborn (1 hour), a middle-aged (45 years), and an elderly (79 years) individual was also amplified. L = 100 bp DNA ladder; RT = reverse transcriptase; h = hour; y = year.

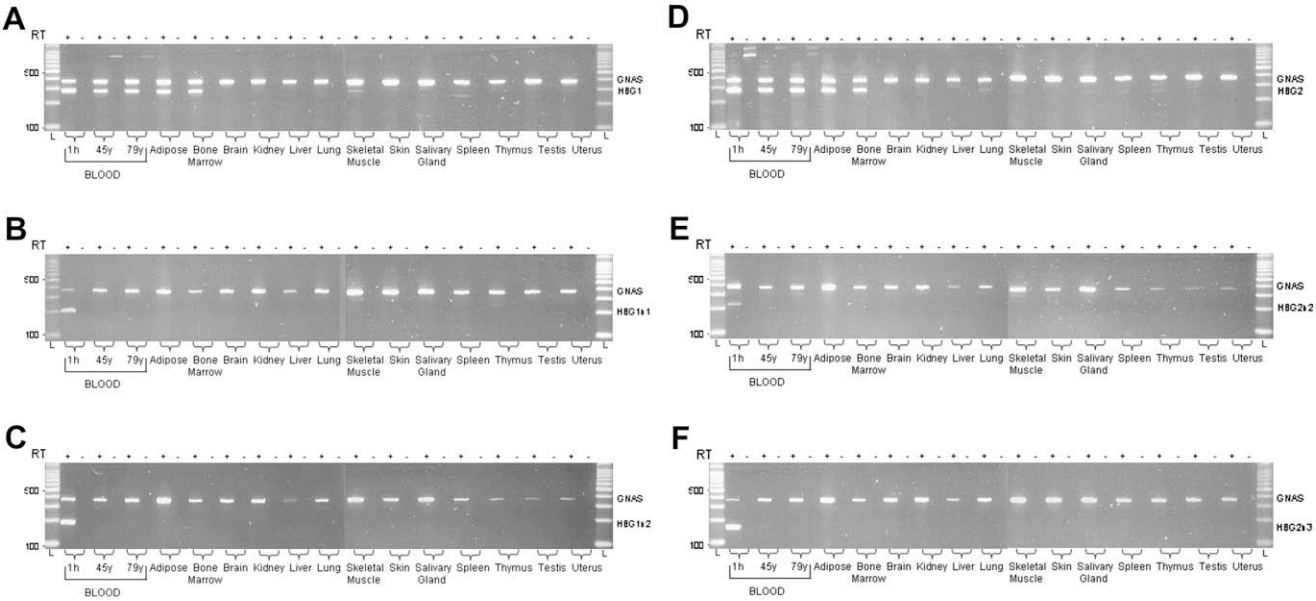


Figure 3. Expression of the HBGn isoforms in adult tissue. The tissue expression of HBG1 and HBG2 and their isoforms in various adult tissues are indicated. Each assay amplified the GNAS housekeeping gene to control for RNA quality and quantity, with the (**A**) HBG1, (**B**) HBG1n1, (**C**) HBG1n2, (**D**) HBG2, (**E**) HBG2n2, and (**F**) HBG2n3 genes. As controls, blood extracted RNA from a newborn (1 hour), a middle-aged (45 years), and an elderly (79 years) individual was also amplified. L = 100-bp DNA ladder; RT = reverse transcriptase; h = hour; y = year.

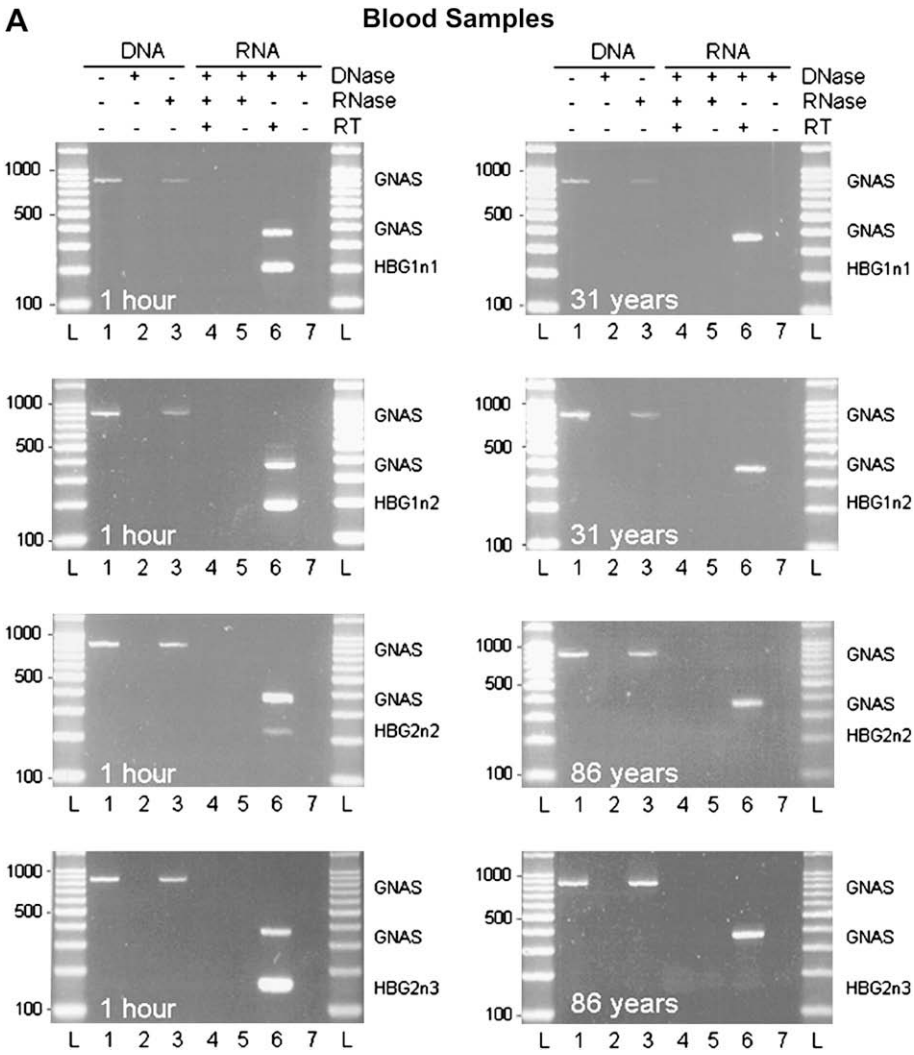


Figure 4. Novel HBGn transcripts originate from an RNA-based rearrangement event. DNA and RNA from multiple blood donors along with fetal liver and spleen and adult bone marrow (BM) and adipose tissues were tested for the presence of the GNAS housekeeping gene (internal control) and one of the newborn hemoglobin isoforms using reverse transcription polymerase chain reaction duplex assays. (**A**) Expression of the HBG1n1, HBG1n2, HBG2n2, and HBG2n3 isoforms was detected specifically in reverse transcribed newborn (1 hour) blood RNA samples predigested with DNaseI but were not detected in RNA from adult (31 years) and elderly (86 years) blood or in any of the DNA samples from the same individuals. (**B**) The HBG1n1, HBG1n2, and HBG2n3 isoforms were amplified in DNaseI-treated reverse-transcribed RNA isolated from fetal liver and spleen, but not in adult adipose or bone marrow tissues. Lane 1, untreated DNA; lane 2, DNaseI-treated DNA; lane 3, RNaseI-treated DNA; lanes 4−5, DNaseI and RNaseI-treated RNA (+ and - RT); lanes 6−7, DNaseI only treated RNA (+ and − RT). L = 100-bp DNA ladder; RT = reverse transcriptase enzyme.

HBG2n3, as well as a housekeeping gene. The housekeeping gene GNAS (guanine nucleotide binding protein, α stimulating) [14] was used as an internal positive control that serves to both normalize RNA quantity and ensure its quality. Additional positive and negative controls included RNA isolated from blood from a 1-hour-old newborn infant and two adult individuals (45 and 79 years old): blood from the newborn is expected to be positive for both the parental HBG genes and all four HBGn isoforms, HBG1n(1/2) and HBG2n(2/3), whereas the adult blood samples should be negative for the HBGn isoforms and positive for the parental HBG genes.

The expression of the novel HBGn transcripts together with their parental genes was investigated in a variety of fetal tissues, including brain, heart, kidney, liver, spleen, and thymus. HBG1 and HBG2 were expressed in all fetal tissues examined, as well as in blood from the newborn child and two adults (Figs. 2A and 2D). Thus, not only did HBG1 and HGB2 appear to be expressed in blood in a non−age-specific manner, but they were also ubiquitously expressed in fetal tissues. In contrast, the four novel hemoglobin transcripts exhibited variable expression patterns in fetal tissue. The HBG1n1, HBG1n2, and HBG2n3 isoforms

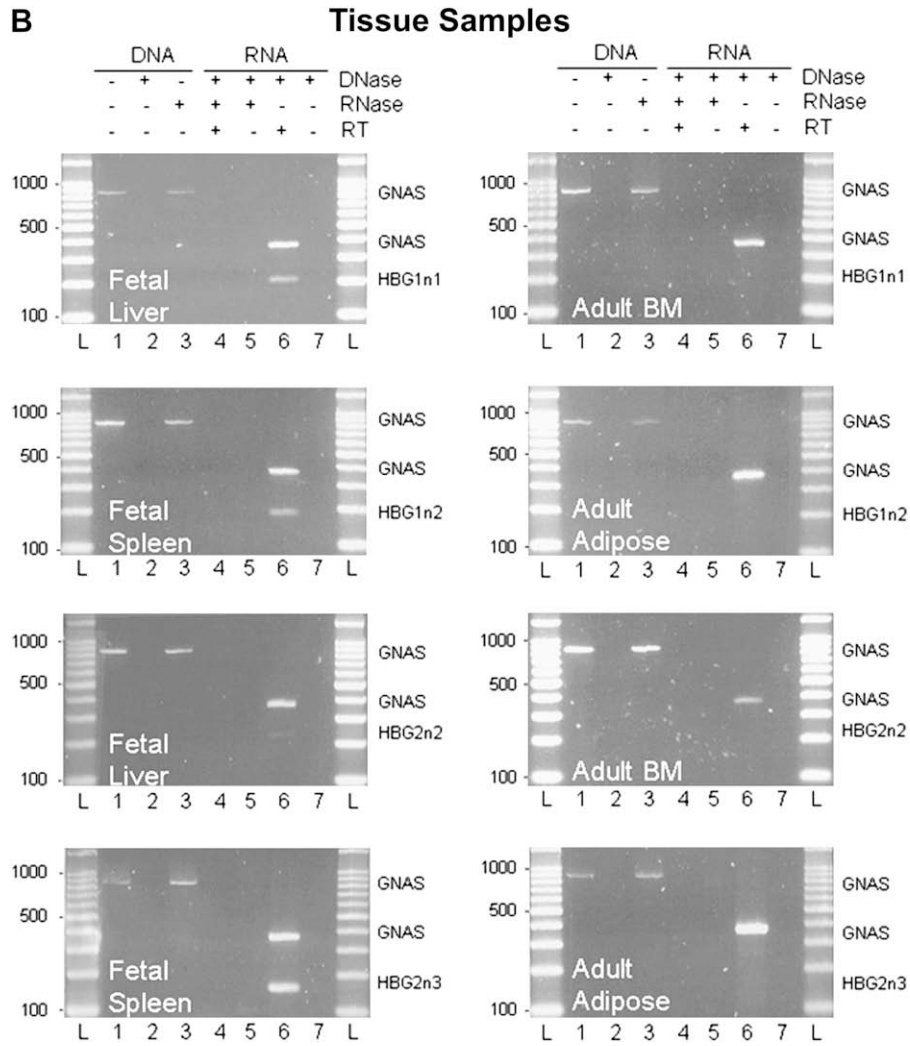


Figure 4. (*continued*).

were expressed at high levels in fetal liver and spleen. The liver and, to a lesser extent, the spleen are the major sites of fetal hematopoiesis, after the transition from vitelline to placental circulation and disappearance of the embryonic yolk sac [3]. Although significantly weaker expression of these isoforms was observed in other fetal tissues such as the heart (HBG1n1, HBG1n2), kidney (HBG2n3), and thymus (HBG1n1), it is not possible to preclude the possibility that this may be due to the presence of contaminating circulatory fetal blood within the tissue isolates.

Next expression of the four HBGn transcripts together with their parental genes was investigated in a variety of adult tissues, including adipose, bone marrow, brain, kidney, liver, lung, skeletal muscle, skin, salivary gland, spleen, thymus, testis, and uterus. The novel isoforms were not expressed in any of the adult tissues examined nor the adult blood controls, but were clearly identifiable in the 1-hour-old newborn blood sample (Fig. 3B, C, E, and F). The parental HBG genes were expressed in the blood of newborns and adults, as well as in adult bone marrow and adipose, but were not detected in other adult tissues (Figs. 3A and D).

Molecular origin of newborn hemoglobin isoforms

Sequencing of the four mRNA isoform transcripts revealed inter-exon deletion breakpoints within octanucleotide (HBG1n2) and varying pentanucleotide (HBG1n1, HBG2n2, HBG2n3) direct repeats (Fig. 1B and C). At the molecular level, we proposed that the developmentally regulated newborn transcripts could be formed by either a DNA-based genomic recombination event or, conceivably, by a heretofore undefined RNA rearrangement event [4]. Both possibilities were investigated by ascertaining the effect of nuclease pretreatment of DNA and RNA isolates from samples (fetal liver, fetal spleen, and newborn blood) known to express the HBGn isoforms, followed by duplex PCR analysis. Samples that do not express the isoforms (blood from 31- and 86-year-olds, adult adipose and bone marrow) were used as negative controls. Each DNA

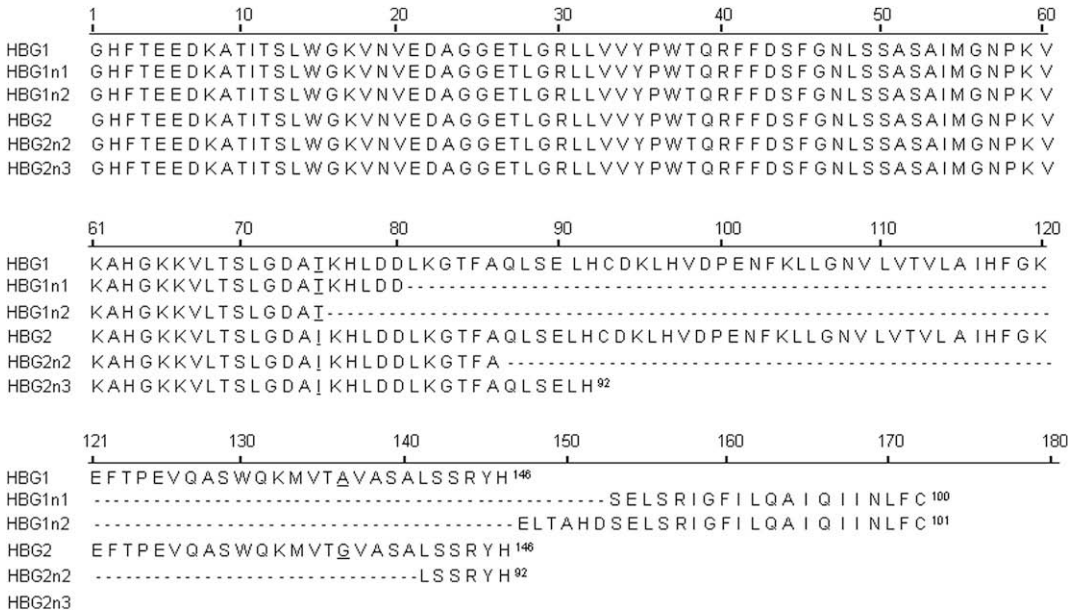


Figure 5. HBGn isoforms maintain putative open reading frames. Putative protein sequences of the HBGn mRNA isoforms. HBG1n1 and HBG1n2 align with the HBG1 sequence from AA1-80 and AA1-75, respectively, while HBG2n2 and HBG2n3 align with the HBG2 sequence from AA1-86 and AA1-92, respectively. At the C-terminal region the two A-gamma–derived isoforms regain their open reading frames in the 3′ UTR of the HBG1 gene, while the G-gamma isoform (HBG2n2) contains the same six terminal amino acids as the parental HBG2 gene. In contrast to the other isoforms, the deletion event creating HBG2n3 introduces an immediate stop codon at AA92. The underlined amino acids at positions 75 and 136 represent the two common amino acid substitutions found in the HBG1 and HBG2 protein sequences.

and RNA sample was pretreated with RNaseI and/or DNaseI and then analyzed with the individual duplex HBGn/GNAS assays.

No evidence to support the presence of the four HBGn isoforms was obtained from the DNA samples amplified (Fig. 4A and B, lanes 1−3). The DNA was of adequate quality for PCR analysis as judged by the presence of GNAS amplicons of the appropriate size (855 bp) in untreated aliquots (Fig. 4, lane 1) that were subsequently abolished when pretreated with DNaseI (Fig. 4, lane 2) but unaffected by RNaseI (Fig. 4, lane 3) treatment. For the RNA extracts, all samples were DNaseI digested to eliminate any contaminating DNA before analysis. Subsequently, the digested total RNA was either treated (Fig. 4, lanes 4, 5) or untreated (Fig. 4, lanes 6 and 7) with the RNaseI enzyme. Following ±RNaseI treatment, the RNA samples were reverse transcribed with (lanes 4 and 6) or without (lanes 5 and 7) the reverse transcriptase enzyme (±RT) and amplified with the duplex PCR reactions. The results illustrate that all RNA samples failed to amplify GNAS and HBGn products after treatment with both DNaseI and RNaseI enzymes (lanes 4 and 5). In contrast, when total RNA was treated with both the DNaseI and RT enzymes (no RNaseI digestion) we were able to detect an amplicon corresponding to GNAS mRNA (371 bp, lane 6) in all samples tested, while only the newborn blood samples (1 hour), fetal liver and fetal spleen amplified the newborn hemoglobin isoforms, HBG1n(1/2) and HBG2n(2/3). The HBG2n2 gene did not amplify in the fetal liver sample, which was expected due to it only being expressed in newborn blood (Fig. 2E). Collectively, these results are consistent with the newborn HBG1n(1/2) and HBG2n(2/3) transcripts being created by an RNA rearrangement event.

Discussion

In this report, we describe the molecular structure and tissue expression of a novel class of truncated HBG mRNAs. These four newly discovered isoforms are characterized by the presence of deleted gene segments of varying sizes that span the 3′ region of exon 2 and the 5′ end of exon 3. Uniquely these deletion breakpoints occur between direct repeats found in adjacent HBG exons, specifically in exons 2 and 3. The novel isoforms, designated HBG1n1, HBG1n2, HBG2n2, and HBG2n3, or collectively termed HBGn, exhibit restricted patterns of developmental- and tissue-specific expression, with levels only being significantly detected in newborn blood, as well as in fetal liver or spleen, the primary sites of fetal hematopoiesis. In all adult tissues tested, including blood, expression of the HBGn isoforms was not detected. In contrast, the HBG1 and HBG2 transcripts were found to be expressed in peripheral blood of all age groups, as well as in adult bone marrow and adipose tissue. Expression of the parental hemoglobin genes in adult bone marrow was not unexpected, due to it being the primary site of hematopoiesis taking over for the fetal liver at approximately 20 weeks gestation [3]. Alternatively,

the presence of the same transcripts in human adult adipose tissue was somewhat unexpected due to this tissue's lack of direct association with erythropoiesis. Interestingly, a human fetal HBG2 promoter linked to the viral simian virus 40 T antigen produced brown adipose tumors in the neck and shoulder areas of transgenic mice indicating the ability of the gamma hemoglobin promoter to be transcribed in adult mouse adipose tissue [15].

The mechanism of HBGn isoform formation is still not conclusively resolved, although evidence points to their genesis via a novel RNA rearrangement event between direct repeats in adjacent exons. One plausible hypothesis posits a looping out and deletion of sequences from the (presumably) parental HBG transcripts that is developmentally regulated. Conversely, their generation may represent an unreported alternative splicing mechanism because no known consensus alternative splicing sequences are associated with the particular RNA rearrangements described. It is still formally possible that the isoforms are produced by a somatic DNA rearrangement event occurring within a small restricted cellular compartment of the peripheral blood. In such a scenario, DNA recombination might be undetectable due to the low copy number of recombined genomic DNA. However, such cells may be programmed to express the novel isoforms at an extremely high level, resulting in their subsequent detection in the RNA isolates.

The locations of the apparent recombination events that create the HBGn isoforms raise the question as to whether they might encode proteins. All four isoforms maintained an open reading frame, despite the deletion events that resulted in their formation (Fig. 5). All isoforms maintain the initiation start codon present in exon one whereas three alternate stop codons are formed by the HBG1n1, HBG1n2, and HBG2n3 sequences; HBG2n2 maintains its original stop codon. Preliminary attempts to detect the predicted truncated protein products of the HBGn isoforms using immunochromatographic and electrophoretic methods proved unsuccessful, suggesting that such proteins are not expressed or expressed at low levels (data not shown). It should be noted that a number of other hemoglobin loci transcripts have been reported that do not apparently encode proteins, including the developmentally regulated θ [16,17] and μ [18] genes within the α-globin gene cluster on chromosome 16. In addition the β-globin locus also encodes an alternative δ-globin gene, which is transcribed in adult erythroid cells [19].

Acknowledgments

This work was supported under award number 2005-MU-BX-K075 from the Office of Justice Programs, National Institute of Justice, Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Justice. The authors would like to thank Sarah Province from Florida Hospital Laboratory for providing all human blood samples used in this study.

References

1. Stamatoyannopoulos G. Control of globin gene expression during development and erythroid differentiation. Exp Hematol. 2005;33:259–271.
2. Bernards R, Little PF, Annison G, Williamson R, Flavell RA. Structure of the human G gamma-A gamma-delta-beta-globin gene locus. Proc Natl Acad Sci U S A. 1979;76:4827–4831.
3. Brittain T. Molecular aspects of embryonic hemoglobin function. Mol Aspects Med. 2002;23:293–342.
4. Alvarez M, Ballantyne J. The identification of newborns using messenger RNA profiling analysis. Anal Biochem. 2006;357:21–34.
5. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem. 1987;162:156–159.
6. Alvarez M, Juusola J, Ballantyne J. An mRNA and DNA co-isolation method for forensic casework samples. Anal Biochem. 2004;335:289–298.
7. Comey CT, Koons BW, Presley KW, et al. DNA extraction strategies for amplified fragment length polymorphism analysis. J Forensic Sci. 1994;39:1254–1269.
8. Jones LJ, Yue ST, Cheung CY, Singer VL. RNA quantitation by fluorescence-based solution assay: RiboGreen reagent characterization. Anal Biochem. 1998;265:368–374.
9. Ambion. RETROscript Kit. 2008;1–26.
10. Shiga K, Yamamoto H, Okamoto H. Isolation and characterization of the human homologue of rig and its pseudogenes: the functional gene has features characteristic of housekeeping genes. Proc Natl Acad Sci U S A. 1990;87:3594–3598.
11. Kitagawa M, Takasawa S, Kikuchi N, et al. rig encodes ribosomal protein S15. The primary structure of mammalian ribosomal protein S15. FEBS Lett. 1991;283:210–214.
12. Huang Z, Fasco MJ, Kaminsky LS. Optimization of Dnase I removal of contaminating DNA from RNA for use in quantitative RNA-PCR. Biotechniques. 1996;20:1012–1014. 1016, 1018−1020.
13. Wiame I, Remy S, Swennen R, Sagi L. Irreversible heat inactivation of DNase I without RNA degradation. Biotechniques. 2000;29:252–254. 256.
14. Weinstein LS, Liu J, Sakamoto A, Xie T, Chen M. Minireview: GNAS: normal and abnormal functions. Endocrinology. 2004;145:5459–5464.
15. Perez-Stable C, Altman NH, Brown J, Harbison M, Cray C, Roos BA. Prostate, adrenocortical, and brown adipose tumors in fetal globin/T antigen transgenic mice. Lab Invest. 1996;74:363–373.
16. Clegg JB. Can the product of the theta gene be a real globin? Nature. 1987;329:465–466.
17. Albitar M, Peschle C, Liebhaber SA. Theta, zeta, and epsilon globin messenger RNAs are expressed in adults. Blood. 1989;74:629–637.
18. Goh SH, Lee YT, Bhanu NV, et al. A newly discovered human alpha-globin gene. Blood. 2005;106:1466–1472.
19. Goh SH, Jackson KA, Terry Lee Y, Miller JL. Identification of an alternate delta-globin mRNA in adult human erythroid cells. Genomics. 2004;84:431–434.