Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

Alternative pre-mRNA splicing switches modulate gene expression in late erythropoiesis

Permalink

https://escholarship.org/uc/item/3pn8366f

Author

Yamamoto, Miki L.

Publication Date

2009-10-15

eScholarship.org Powered by the California Digital Library
University of California

Alternative Pre-mRNA Splicing Switches Modulate Gene Expression in Late Erythropoiesis

Miki L. Yamamoto[1], Tyson A. Clark[2], Sherry L. Gee[1], Jeong-Ah Kang[3], Anthony C. Schweitzer[2], Amittha Wickrema[3] and John G. Conboy[1§]

[1]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA 94720

[2]Affymetrix, Inc., Santa Clara, CA, USA 95051

[3]Department of Medicine, University of Chicago, Chicago, IL, USA 60637

Running title: Erythroid Stage-Specific Alternative Splicing

Number of words in text (4884) and abstract (195)

Number of Figures (5) and Tables (1)

Number of Supplementary Files: 5

Number of references: 46

[§]Corresponding author: John G Conboy, Lawrence Berkeley National Lab, 1 Cyclotron Rd., Mail Stop 74R0157, Berkeley, CA 94720. Telephone: 510-486-6973; Fax: 510-486-6746, jgconboy@lbl.gov

Scientific Heading: Red Cells

ABSTRACT

Differentiating erythroid cells execute a unique gene expression program that insures synthesis of the appropriate proteome at each stage of maturation. Standard expression microarrays provide important insight into erythroid gene expression, but cannot detect qualitative changes in transcript structure, mediated by RNA processing, that alter structure and function of encoded proteins. We analyzed stage-specific changes in the late erythroid transcriptome via use of high resolution microarrays that detect altered expression of individual exons. Ten differentiation-associated changes in erythroblast splicing patterns were identified, including the previously known activation of protein 4.1R exon 16 splicing. Six new alternative splicing switches involving enhanced inclusion of internal cassette exons were discovered, as well as three changes in use of alternative first exons. All of these erythroid stage-specific splicing events represent activated inclusion of authentic annotated exons, suggesting they represent an active regulatory process rather than a general loss of splicing fidelity. The observation that three of the regulated transcripts encode RNA binding proteins (SNRP70, HNRPLL, MBNL2) may indicate significant changes in the RNA processing machinery of late erythroblasts. Together these results support the existence of a regulated alternative pre-mRNA splicing program that is critical for late erythroid differentiation.

INTRODUCTION

Differentiation of erythroid progenitors into mature red cells requires a carefully orchestrated gene expression program to insure synthesis of the appropriate stage-specific proteome as the cells become progressively more specialized. Previous studies of erythroid gene expression have focused predominantly on quantitative or semi-quantitative assays including Western and Northern blotting, proteomic analysis, cDNA cloning, and microarray analysis. Together these studies have provided considerable insights into the late erythroid gene expression program.

However, gene-level expression analysis cannot identify important qualitative changes in erythroid gene expression predicted to result from alternative pre-mRNA processing pathways. Exclusion or inclusion of individual exons can substantially alter the structure and function of the encoded protein isoforms independent of changes in transcript expression levels. Given that the majority of genes in the human genome exhibit alternative splicing (e.g., ref. 1,2), and that alternative splicing pathways may be regulated during differentiation and development, it is likely that stage-specific alternative splicing "switches" play an important role in modulating protein function during erythroid differentiation. Such changes in erythroid transcript structure may be regulated at the transcriptional level in the case of alternative first exons[3], or at the pre-mRNA splicing level in the case of internal exons[4]. To further understand the erythroid gene expression program, the erythroid transcriptome will need to be explored at different stages of erythropoiesis and at the resolution of individual exons.

The best-studied example of regulated pre-mRNA splicing in erythroid cells is the stage-specific splicing switch of protein 4.1R exon 16. Alternative exon 16 is tightly regulated such that it is excluded in early erythroid progenitor cells, but efficiently included in late erythroblasts[5,6]. This splicing switch

is functionally important: exon 16 inclusion leads to synthesis of 4.1R protein isoforms with high affinity for spectrin and actin, and increased ability to mechanically stabilize the erythroid membrane[7-9]. Mechanistically, the splicing switch is regulated at least in part by changes in expression of antagonistic splicing factors, in particular, a decrease in expression of the splicing inhibitory factor hnRNP A1 relative to that of stimulatory factors Fox-2 and SF2/ASF[10-13]. It seems reasonable to propose that these changes in splicing factor activity would regulate not only protein 4.1R pre-mRNA splicing, but also a subset of other alternative splicing events that together may constitute an erythroid alternative splicing program.

To explore the hypothesis that alternative splicing switches in other erythroid transcripts are executed during late erythropoiesis, we have undertaken a genome-wide expression analysis of erythroblast transcripts at the level of individual exons. We used Affymetrix exon microarrays[1,14,15] to identify erythroid stage-specific changes in exon expression, using RNA isolated from basophilic versus orthochromatic erythroblasts differentiated *in vitro* from human CD34+ erythroid progenitors[16-18]. Candidate alternative splicing events identified by microarray analysis were further validated by RT-PCR. Together these experiments revealed a number of internal alternative exons that exhibit significant changes in splicing efficiency, as well as differential expression of several alternative first exons that suggest changes in transcriptional promoter usage. These results suggest that an erythroid splicing program mediates stage-specific changes in transcript (and ultimately protein) structure and function that likely are critical for proper erythropoiesis.

MATERIALS AND METHODS

Cell Isolation and Culture

An *in vitro* primary culture system was employed to generate cells at different stages of human erythroid differentiation[16]. The culture system used was CD34-positive early hematopoietic progenitors isolated from growth factor mobilized peripheral blood (purchased from ALL Cells, Berkeley, CA) to promote the erythroid differentiation program. The culture media contained 15% FCS, 15% human AB serum, Iscove's modified Dulbecco's medium (IMDM), 10ng/ml IL-3, 2 units ml Epo, and 50 ng/ml SCF. SCF was gradually decreased, and IL-3 was not added after day 3. Cells were collected at day 7, 10, and 14, which respectively correspond to basophilic (early), polychromatic (intermediate), and orthochromatic (late) erythroblasts. 97-99% of these cells were erythroid as determined by flow cytometry for glycophorin A and CD71 (Figure 1A).

Total RNA extraction and Sample Preparation

Total cellular RNA was extracted from each culture at early (day 7), intermediate (day 10), and late (day 14) stages of differentiation. Alpha- and beta-globin mRNAs were depleted from the RNA samples using GLOBINclear^{TM} Kit (Ambion, Inc.). 1 μg of total globin-depleted RNA was used to subtract ribosomal RNA using the Ribominus^{TM} Human/Mouse Transcriptome Isolation kit (Invitrogen). Globin- and ribosomal RNA-depleted RNA was amplified and labeled for the exon array hybridization according to the Affymetrix WT Sense Target Labeling Assay Manual (http://affymetrix.com/products/arrays/specific/hugene_1_0_st/hugene_reagent_instr_solution.affx. The GeneChip^{TM} Hybridization Oven 320 and the GeneChip^{TM} Fluidics Station 450 were used for the hybridization and washing, respectively. The GeneChip^{TM} Scanner 3000 was used for scanning.

Exon Array Analysis

In order to identify candidate alternative splicing switches in differentiating erythroid cells, we assessed genome-wide changes in exon expression using two platforms: the Affymetrix Human Exon 1.0 ST Array[1] and a non-commercial exon junction array (described below). The commercial Human Exon 1.0 ST array was designed to be as inclusive as possible; it contains probes for virtually all known and predicted exons in the human genome and has the potential to discover novel splicing events[1]. In contrast, the junction array focuses on a smaller set of well-annotated exons and includes probes not only to each exon, but also to exon-exon junctions. This latter feature provides additional information that aids in predicting splicing differences between RNA samples, since a change in alternative splicing should yield reciprocal changes in intensity of junction probes for the exon skipping and exon inclusion event, respectively.

For the Human Exon 1.0 ST array, comparison of exon expression patterns in early versus late erythroblasts was performed using XRAY software, version 2.5 Excel Add-In (Biotique Systems, Inc.). The 8 input CEL files (3 replicates of day 7 samples, 2 replicates of day 10 samples, and 3 replicates of day 14 samples) were analyzed to identify genes that were significantly differentially expressed or displayed significant differential alternative splicing between the groups of interest. Replicate 1 of day 10 was removed due to poor data quality.

For the tissue-specific exon expression analysis of the beta spectrin gene, we compared three replicates of erythroid day 14 to three replicates of the muscle tissue Exon Array data[1]. The splicing index[1,15] was calculated to create figures for the candidate genes showing alternative exon usage in the array data.

The exon junction array contains probesets that were designed to interrogate both exons and the

observed exon-exon junctions for approximately 35,000 genes in the human genome. The design uses transcript annotations from RefSeq, Ensembl, and ExonWalk. Approximately 250,000 exon clusters are represented. Additional details on the junction array design are included in the supplementary material.

Labeled target generated from five biological replicates of total RNA extracted from day 7 and day 14 was hybridized to the junction array. Target was prepared using the same method as the Human Exon 1.0 ST arrays. One replicate of the day 14 samples was deemed an outlier and removed from subsequent analyses. Data from the junction arrays was processed using the Affymetrix Power Tools. Candidate alternative splicing events were established using the splicing index approach[1]. Each exon or junction probeset was analyzed independently. We then selected splicing events for which than one informative probeset showed a significant change in utilization between the day 7 and day 14 samples. In particular, we selected cassette exons that exhibited reciprocal behavior of the probeset specific to the skip junction and one or more of the probesets monitoring inclusion of the exon. Additional details regarding analysis of the junction array data is included in the supplementary material.

Reverse transcription PCR

Validation of splicing changes predicted by microarray experiments was performed by RT-PCR analysis of the same RNA samples used for the array hybridizations. Total RNA was DNase-treated, and 1μg was used to synthesize first strand cDNA with SuperScript III (Invitrogen) and random hexamer primers in a total volume of 20 μl. 1μl of the cDNA was used for standard 20μl total volume PCR reactions. 25 to 35 cycles of amplification were performed under the following conditions: denaturation for 30 s at 94°C; annealing for 30 s at 55°C; extension for 30 s at 72°C. Primer sequences are shown in Supplementary Table 1. PCR products were analyzed on 5% polyacrylamide gels.

Densitometry was carried out to compare bands using Alpha Imager 2200 v. 5.5 (Alpha Innotech Corp.). At least two replicates were used to validate the array predictions by RT-PCR, and the identity of all PCR fragments was confirmed by DNA sequence analysis. It is worth noting that alternative splicing events supported by reciprocal changes in probesets for both exon inclusion versus exclusion events had a higher validation rate than those predicted by only one or the other.

RESULTS

Analysis of differentiation stage-specific alternative splicing switches in human erythroblasts

Human $CD34^+$ primary erythroid progenitors from growth factor-mobilized peripheral blood can differentiate in culture and undergo many of the programmed changes in gene expression that are characteristic of late erythropoiesis[16-18]. To investigate changes in alternative splicing that occur specifically in differentiating erythroblasts, we analyzed cultures highly enriched ($\geq$ 97%) for erythroid cells as determined by FACS analysis using CD71 and glycophorin A antibodies (Figure 1A). RNA from cells cultured for 7 days (basophilic erythroblasts) and 14 days (orthochromatic erythroblasts) was then examined by exon microarray analysis and by RT-PCR in order to identify changes in gene expression at the level of alternative pre-mRNA splicing.

A key alternative splicing event during late erythropoiesis involves protein 4.1R exon 16, which is excluded in early erythroblasts but included efficiently in late erythroblasts[5,8,10] (Figure 1A). Consistent with RNA studies, isoforms of 4.1R protein that include the exon 16-encoded peptide are increased in these cells[5,6]. To confirm that the exon 16 alternative splicing switch occurs during differentiation of these primary erythroid cultures, we used RT-PCR to assay the abundance of

transcripts that either include or exclude exon 16. Figure 1B shows that exon 16 was included much more efficiently in late erythroblasts (day 14) than in early erythroblasts (day 7). This result validates the use of these cultures for analysis of stage-specific splicing changes in late erythropoiesis.

Exon microarray analysis of changes in exon expression during late erythroid differentiation

To identify new cases of alternative splicing in differentiating erythroid cells, we assessed genome-wide changes in exon expression using the Affymetrix Human Exon 1.0 ST Array. This microarray was designed to be as inclusive as possible, containing probes designed to detect expression of virtually all known and predicted exons in the human genome[1]. We hybridized these arrays with probes prepared from erythroblast RNA at three stages of differentiation (basophilic, polychromatic, and orthochromatic erythroblasts) from three independent cultures. Array data from these three groups were then analyzed using XRAY software to look for differentiation-associated changes in expression, both at the level of whole transcripts and at the level of individual exons. For most comparisons, we focused on basophilic versus orthochromatic samples to maximize potential differences in gene expression.

First, gene-level changes in expression were deduced by combining the hybridization data for all informative probes from each annotated transcript. Previous studies have shown that differentiation in this erythroid culture system is characterized by increased expression of known red cell proteins, including band 3, ankyrin, and protein 4.1, whose mRNAs begins to accumulate around day 8[17]. As expected, exon expression analysis confirmed the up-regulation of these transcripts, as well as others encoding enzymes involved in heme biosynthesis (e.g., *ALAS2*) and proteins associated with the membrane skeletal network (e.g., *EPB49, EPB42, AQP1, GYPB*). Conversely, many other transcripts exhibit a significant decrease in expression in late erythroblasts. For example, the transcript encoding

HNRPA1, a negative regulator of 4.1 exon 16 splicing known to be down-regulated in late erythroblasts[10], was lower at late time points (e.g., day 14) compared to the early time point (day 7). Supplementary Tables 2 and 3 list the transcripts that show greater than two-fold increase or decrease in expression in early versus late erythroid progenitor cells.

Next, to determine whether the exon array data could detect known alternative splicing events, we considered the case of beta spectrin (*SPTB*). Beta spectrin mRNAs are known to possess distinct 3' terminal exons in eythroid cells and in muscle cells, due to alternative splicing and alternative polyadenylation (Figure 2, upper panel; ref. 19). Comparison of the new erythroid exon expression data to publicly available Exon Array data for human skeletal muscle[1] was performed using XRAY analysis software to identify predicted splicing differences. As expected, beta-spectrin ranked near the top of the list of genes predicted to splice differently in muscle versus erythroid cells. To visualize these differences, the array data was used to calculate the splicing index, a measure for detection of alternative splicing differences by identifying probesets that exhibit differential expression between RNA samples after normalizing for transcript levels (ref. 1 and supplementary information). The splicing index is shown in Figure 2 (lower panel). Probesets representing the 5' and central regions of the gene exhibited no significant differences in relative expression in erythroblasts versus muscle, as indicated by consistent values for the splicing index close to zero. In contrast, substantial differences were observed at the 3' end. The upward peak labeled "erythroid" indicates enriched expression of probesets corresponding to the known erythroid-specific 3' end, while the downward peak "muscle" represents probesets that map to the muscle-specific 3' exons and are under-expressed in erythroid RNA relative to their higher expression in muscle. This finding confirms exon array detection of a well-known erythroid-specific gene expression event.

Identification of stage-specific changes in utilization of alternative first exons.

Array data was next examined for evidence of differential exon expression in late erythropoiesis, as manifested by significant changes in probeset expression between basophilic (day 7) and orthochromatic (day 14) erythroblasts. After filtering out probesets with low expression and low variance, transcripts with highly significant differences in alternative splicing were manually inspected by mapping to the human genome to eliminate those with ambiguous transcript cluster assignments[1]. Transcripts that displayed consistent day 7 / day 14 expression ratios across most of their probesets, but differed markedly in one discrete region, were chosen for further characterization. This process eliminated many transcripts with poor annotation, inconsistent probe behavior, and/or overly complex splicing patterns.

Among the probesets exhibiting stage-specific changes in expression, several mapped to the 5' ends of transcripts and were considered to be candidate alternative first exons. Many erythroid genes are known to possess alternative first exons[3], but few have been analyzed for changes in expression during late erythropoiesis. RT-PCR analysis confirmed three examples of switches in alternative first exon usage that occur in late erythropoiesis (Figure 3, left). Gene models at the right depict 5' exon/intron structure of these genes, based on data from the Refseq track in the UCSC genome browser, and illustrate the predominant expression pattern at day 7 (black lines) and day 14 (red lines) that was deduced from array and PCR experiments. Two alternative 5' ends are shown for SLC12A6 (KCC3, a KCl co-transporter) and three each for GCNT2 (an acetylglucosaminyltransferase important in the formation of human blood group I antigen) and TNPO2 (transportin-2). For each gene, PCR reactions were performed using multiple forward primers (one in each alternative 5' exon) and a single reverse primer in the downstream constitutive exon ("exon 2"). Competitive PCR reactions of this nature have been used previously to detect variations in expression of terminal exons (e.g., ref. 20). The

top panel shows the SLC12A6 transcript, for which exon microarray data indicated a stage-specific decrease in expression of exon 1B relative to exon 1A, as cells differentiate from day 7 to 14. To validate this prediction, RT-PCR analysis was performed with two forward primers (in exons 1A and 1B) and a single reverse primer (in exon 2). As shown in Figure 3B, the relative amounts of amplified exon 1B product declined significantly relative to that of exon 1A, confirming the microarray prediction.

Two additional examples of alternative first exon usage were similarly predicted by the exon array data and verified by RT-PCR, in the GCNT2 and TNPO2 genes (Figure 3, middle and lower panels). PCR clearly showed a substantial difference in relative expression of alternative first exons between basophilic and orthochromatic erythroblasts for GCNT2, with exon 1A predominating at day 7 and exon 1C relatively much more abundant at day 14. Exon 1B was barely expressed in these erythroblast cultures; however, it could be amplified efficiently from brain cDNA (lane br), providing further evidence for tissue-specific regulation of first exon usage in the GCNT2 gene. For TNPO2, there was a substantial switch in relative use of exons 1A and 1B between day 7 and day 14 erythroblasts (bottom panel). Relatively little expression exon 1C was observed, with only small amounts detected in day 14 erythroblasts and in brain. For both of these genes, the RT-PCR data was consistent with the exon microarray predictions and thus two independent assays confirm the switch in expression of alternative first exons.

Stage-specific switches in alternative pre-mRNA splicing of internal cassette exons

The Human Exon 1.0 ST Array can also detect alternative splicing changes within internal regions of transcripts[1,14,15]. We used this array in conjunction with a prototype exon junction array, which focuses

on a smaller set of well-annotated exons, in order to identify additional splicing changes in differentiating erythroid cells. Junction probes provide additional information that can aid in predicting splicing differences between RNA samples, since a change in alternative splicing should yield reciprocal changes in intensity of junction probes for the exon skipping and exon inclusion event, respectively.

Figure 4 illustrates detection of the protein 4.1R exon 16 splicing switch by the junction array. The structures of two Refseq transcripts that include or exclude exon 16 are shown in the lower part of the Figure, with the position of alternative exon 16 indicated by the arrow. The position of junction probesets exhibiting a significant change in expression, normalized to transcript levels, is shown above. The two probesets representing exon 16 inclusion products exhibited strong up-regulation in day 14 orthochromatic erythroblasts (probeset ID 504900 for the exon 13/16 junction, p=7.5E-07, and probeset 224956 for the exon 16/17 junction, p=1.6E-09). Importantly, the probeset interrogating the exon 16 skipping event demonstrated a significant *decrease* in expression in these more mature cells (probeset ID 366937, spanning the exon 13/17 junction, p=1.2E-04). Thus, the reciprocal increase in inclusion probes and decrease in the exclusion probe validates the ability of the junction array to detect a known splicing switch, the activation of exon 16 splicing in late erythroblasts.

Further analysis of junction array data predicted several novel erythroid stage-specific alternative splicing switches. All of these cases exhibited a significant decrease at day 14 in the junction probeset representing the exon skipping event, and a reciprocal increase for at least one probeset representing the alternative exon inclusion event. The diagram in Figure 5A illustrates schematically the locations of these diagnostic probesets for a model alternative splicing event. Table 1 summarizes hybridization data of specific probesets for six candidate stage-specific exons, and also provides some information regarding the identity of the genes / exons undergoing the putative splicing switches. In order to evaluate the performance of the microarray, we sought independent verification of these splicing predictions

using PCR methods to amplify across the relevant region of each mRNA using primers in the flanking exons. Alternative splicing switches should be manifested by changes in the relative amounts of exon inclusion/exon exclusion products when comparing PCR products derived from basophilic and orthochromatic erythroblast samples. As shown in Figure 5B, this was indeed the case, as six new exons were validated by RT-PCR analysis to have substantial erythroid stage-specific changes in alternative splicing in two independent biological replicates. In each case, consistent with the microarray predictions, the relative efficiency of alternative exon inclusion was higher in the more mature day 14 (orthochromatic) erythroblasts. Control experiments confirmed the expected sequence of these PCR products, and demonstrated that exon inclusion:exclusion ratios were independent of PCR cycle number within the range examined (Supplementary Figure 1). The latter is an important criterion typically used in splicing studies to validate cell-specific differences in alternative splicing[22-25]. Together, the combined analysis of erythroblast mRNA by exon microarrays and RT-PCR strongly supports the hypothesis that these splicing switches are *bona fide* features of the gene expression program in late erythropoiesis.

Effects of stage-specific splicing switches on protein structure

In cases where alternative first exons contain coding information or influence expression of downstream coding sequences, switches in alternative first exon expression can directly impact N-terminal protein structure. The late erythroid switches in first exon use for SLC12A6 and GCNT2 both fall into this category, since these exons contain translation initiation sites. The GCNT2 gene has an unusual structure in which three long alternative first exons encode alternative but structurally similar N-terminal domains of ~308 amino acids each[26]. Our finding that exon 1C represents the major isoform in

late erythroblasts is consistent with previous reports that exon 1C is strongly expressed in terminally differentiating K562 cells[21]. In contrast, first exons in the TNPO2 gene do not possess start codons and thus switching does not affect protein structure.

All of erythroid stage-specific splicing switches involve internal cassette exons that exhibit enhanced inclusion at day 14, relative to less efficient or undetectable inclusion at day 7. The exons switched on in late erythroblasts range in size from 66-114nt, significantly smaller than the average human exon but consistent with previous observations that alternative exons are generally shorter than average[27,28]. In each case the splicing switch is within the coding portion of the transcript and thus it will alter the protein product. For PLD1 (phospholipase D1) and ARFIP1 (arfaptin1, or ADP-ribosylation factor interacting protein 1), the splicing switches are predicted to insert in-frame peptides of ~3-4kDa, without alteration of C-terminal sequences. There are no obvious functional motifs encoded in these alternative peptides. Interestingly, ARFIP1 has been reported to inhibit PLD1 activity[29], suggesting the possibility that splicing changes in both genes might coordinately affect some aspect of signaling or membrane trafficking. The four remaining splicing switches differ in that they should induce C-terminal alterations or truncations in the encoded proteins. In MBNL2 (muscleblind2, an RNA binding protein), the switch introduces a new penultimate coding exon that inserts a new peptide and also changes the reading frame. The resulting mRNA should terminate translation at an alternative site in the final exon. In three other genes, the inserted exon contains an in-frame stop codon so that the splicing switch will result in C-terminal truncations of the predicted protein product. This is the predicted outcome of splicing switches in HNRPLL (heterogeneous nuclear ribonucleoprotein L-like), SNRP70 (U1 small nuclear ribonucleoprotein 70K, a component of the U1 snRNP), and ATP11C (ATPase class VI type 11C). Whether the truncated proteins have any physiological role is unknown. However, since the truncations would delete one of the two RNA binding motifs in HNRPLL and half of the single RNA

binding domain in SNRP70, any function would be quite different from, and potentially antagonistic to, the full length proteins. Because MBNL2, HNRPLL, and SNRP70 have known or predicted affects on pre-mRNA splicing, these results suggest that alternative splicing of selected factors may potentially affect a wider range of RNA processing events in late erythropoiesis.

DISCUSSION

High-resolution analysis of the erythroid transcriptome provides new insights into the mechanisms by which gene regulatory mechanisms orchestrate expression of a highly dynamic proteome in a developmentally appropriate, tissue-specific manner. The current study complements and extends previous analyses of the erythroid transcriptome[30,31] by exon-level analysis of switches in alternative splicing during late erythropoiesis. Based on the finding of several novel stage-specific splicing switches in erythroid gene expression, we propose the existence of an alternative splicing program that is essential for proper differentiation of late erythroid cells. All of the switches involve *bona fide* splicing events that occur precisely at exon-intron boundaries of annotated exons, supporting the hypothesis that they represent programmed splicing changes rather than a general de-regulation of the splicing machinery in late erythropoiesis. Our results indicate that this program is comprised of distinct regulatory mechanisms involving transcription events at alternative promoters, and splicing events at internal alternative exons, as both processes contribute to modulation of gene expression in important ways that are generally not detectable by standard expression analysis.

Alternative first exons are common among erythroid genes[3,33,36-40], and indeed among the whole spectrum of genes in the human genome[34,35]. In this report we show that switches in expression of alternative first exons can occur even late in erythroblast differentiation. Such changes can affect

protein levels quantitatively and/or qualitatively, depending on the relative expression levels of the transcripts and the location of translation start codons. Mechanistically, it is likely that these switches are regulated at the transcriptional level, since alternative first exons generally possess their own promoters (e.g., ref. 33), but changes in mRNA stability of alternative isoforms could also play a role in some transcripts. Our finding that GCNT2 exon 1C represents the major isoform in late erythroblasts is consistent with previous reports that exon 1C is up-regulated in terminally differentiating K562 cells by activity of transcription factor CCAAT/enhancer binding protein α (C/EBPα)[21]. The switches in SLC12A6 and TPNO2 first exons could potentially be determined by a related mechanism, perhaps operating so as to allow the genes to escape downregulation of expression in late erythroblasts.

A second class of gene expression changes involves internal cassette exons that are regulated at the level of alternative pre-mRNA splicing. Interestingly, in each of the six new cases described here, alternative exon inclusion was increased in late erythroblasts. This observation suggests that a common pathway or shared splicing program may be coordinating these splicing events. Preliminary analysis of potential splicing regulatory motifs has revealed that five of the six stage-specific exons possess in the flanking proximal introns a highly specific UGCAUG binding site for the alternative splicing regulator Fox-2 (data not shown). This frequency is much higher than expected by chance and is consistent with a coordinating role of Fox-2 in mediating some of these late erythroid alternative splicing switches. Indeed, Fox-2 has already been reported to play an important role in activating the protein 4.1R exon 16 splicing switch[12,13]. Moreover, Fox-1/Fox-2 binding sites are highly over-represented in the introns near many muscle-enriched exons[14] , and also near brain-enriched exons[41,42], but not near nontissue-specific alternative exons or constitutive exons. However, the erythroid-, muscle- and brain-enriched exon datasets identified in these studies are largely unique. We speculate that Fox proteins are necessary to activate splicing switches of many tissue-specific exons, but not sufficient to determine the

spatial or temporal patterns of these switches. Tissue-specific differences in Fox protein structure (mediated by regulated alternative splicing of the Fox pre-mRNAs themselves[43]) or in the complement of putative Fox interacting proteins (e.g., ref. 44), are presumably required to execute the appropriate alternative splicing programs through development. Future studies will be required to test this hypothesis experimentally.

It is noteworthy that three of the alternative splicing switches in late erythropoiesis involve exons containing premature termination codons (PTC), such that the resulting mRNAs would encode truncated proteins. While it is possible that the truncated proteins have unique functions, recent studies suggest an alternative purpose for this type of regulated splicing. Premature stop codons often induce rapid mRNA degradation via nonsense mediated decay (NMD), and alternative splicing coupled to nonsense mediated decay can represent an evolutionarily conserved strategy for regulating gene expression[45,46]. This process, termed "regulated unproductive splicing and translation" (RUST), provides a control mechanism for modulating gene expression by regulating the balance between functional mRNAs that encode full length proteins and PTC-containing mRNAs that do not. RUST has been demonstrated most convincingly in the case of RNA binding proteins with roles in pre-mRNA splicing[45,46] and it may therefore be very relevant that two of the PTC-containing erythroid splicing events involve RNA binding proteins (HNRPLL and SNRP70).

The modest number of alternative splicing switches identified in this study is comparable to other recent reports[15,25], but likely represents only a fraction of the total changes in splicing that occur in differentiating erythroblasts. Here we examined only splicing switches that occur in late stages of erythropoiesis; only annotated alternative splicing events are represented on the microarray; and we can detect only exons represented by "good" oligonucleotide probes that hybridize well under standard conditions. Moreover, given the observation that some splicing switches introduced premature stop

codons, NMD may have masked splicing changes that result in rapidly degraded mRNAs. We anticipate that further advances in array technology, together with application of high throughput sequencing[32] and analyses of earlier stages of erythroid differentiation, will lead to discovery of many additional switches in promoter use and alternative splicing that have important implications for erythroid physiology.

AUTHORSHIP

Contribution: M.L.Y. designed research, performed research, and wrote the paper; T.A.C. designed research, performed research, and wrote the paper; S.L.G. performed research; J.-H.K. performed research, A.C.S. designed research, A.W. designed research, performed research, and wrote the paper; J.G.C. designed research and wrote the paper.

Conflict of interest disclosure: T.A.C. and A.C.S. are employees of Affymetrix, Inc.

Correspondence: John G. Conboy, Lawrence Berkeley National Laboratory, Life Sciences Division, Building 84-171, One Cyclotron Road, Berkeley, CA 94720; USA; e-mail: jgconboy@lbl.gov.

ACKNOWLEDGMENTS

The authors thank Thomas Gallagher for careful review of the manuscript. This work was supported by the National Institutes of Health NIH grant DK75021 and HL45182 and by the Director, Office of Biological and Environmental Research, US Department of Energy under contract DE-AC03-76SF00098.

This work was supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

REFERENCES

1. Clark TA, Schweitzer AC, Chen TX, et al. Discovery of tissue-specific exons using comprehensive human exon microarrays. Genome Biology. 2007;8:R64.

2. Johnson JM, Castle J, Garrett-Engele P, et al. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science. 2003;302:2141-2144.

3. Tan JS, Mohandas N, Conboy JG. High frequency of alternative first exons in erythroid genes suggests a critical role in regulating gene function. Blood. 2006;107:2557-2561.

4. Hou VC, Conboy JG. Regulation of alternative pre-mRNA splicing during erythroid differentiation. Curr Opin Hematol. 2001;8:74-79.

5. Chasis JA, Coulombel L, Conboy J, et al. Differentiation-associated switches in protein 4.1 expression. Synthesis of multiple structural isoforms during normal human erythropoiesis. J Clin Invest. 1993;91:329-338.

6. Baklouti F, Huang SC, Tang TK, Delaunay J, Marchesi VT, Benz EJ, Jr. Asynchronous regulation of splicing events within protein 4.1 pre-mRNA during erythroid differentiation. Blood. 1996;87:3934-3941.

7. Discher D, Parra M, Conboy JG, Mohandas N. Mechanochemistry of the alternatively spliced spectrin-actin binding domain in membrane skeletal protein 4.1. J Biol Chem. 1993;268:7186-7195.

8. Horne WC, Huang SC, Becker PS, Tang TK, Benz EJ, Jr. Tissue-specific alternative splicing of protein 4.1 inserts an exon necessary for formation of the ternary complex with erythrocyte spectrin and F-actin. Blood. 1993;82:2558-2563.

9. Discher DE, Winardi R, Schischmanoff PO, Parra M, Conboy JG, Mohandas N.

Mechanochemistry of protein 4.1's spectrin-actin-binding domain: ternary complex interactions, membrane binding, network integration, structural strengthening. J Cell Biol. 1995;130:897-907.

10. Hou VC, Lersch R, Gee SL, et al. Decrease in hnRNP A/B expression during erythropoiesis mediates a pre-mRNA splicing switch. EMBO J. 2002;21:6195-6204.

11. Yang G, Huang SC, Wu JY, Benz EJ, Jr. An erythroid differentiation-specific splicing switch in protein 4.1R mediated by the interaction of SF2/ASF with an exonic splicing enhancer. Blood. 2005;105:2146-2153.

12. Ponthier JL, Schluepen C, Chen W, et al. Fox-2 Splicing Factor Binds to a Conserved Intron Motif to Promote Inclusion of Protein 4.1R Alternative Exon 16. J Biol Chem. 2006;281:12468-12474.

13. Yang G, Huang SC, Wu JY, Benz EJJ. Regulated Fox-2 isoform expression mediates protein 4.1R splicing during erythroid differentiation. Blood. 2008;111:392-401.

14. Das D, Clark TA, Schweitzer A, et al. A correlation with exon expression approach to identify cis-regulatory elements for tissue-specific alternative splicing. Nucleic Acids Res. 2007;35:4845-4857.

15. Gardina PJ, Clark TA, Shimada B, et al. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. BMC Genomics. 2006;7:325.

16. Uddin S, Ah-Kang J, Ulaszek J, Mahmud D, Wickrema A. Differentiation stage-specific activation of p38 mitogen-activated protein kinase isoforms in primary human erythroid cells. Proc Natl Acad Sci U S A. 2004;101:147-152.

17. Wickrema A, Koury ST, Dai CH, Krantz SB. Changes in cytoskeletal proteins and their mRNAs during maturation of human erythroid progenitor cells. J Cell Physiol. 1994;160:417-426.

18. Wickrema A, Krantz SB, Winkelmann JC, Bondurant MC. Differentiation and erythropoietin receptor gene expression in human erythroid progenitor cells. Blood. 1992;80:1940-1949.

19. Chu ZL, Wickrema A, Krantz SB, Winkelmann JC. Erythroid-specific processing of human beta spectrin I pre-mRNA. Blood. 1994;84:1992-1999.

20. Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat Struct Mol Biol. 2007;14:185-193.

21. Twu YC, Chen CP, Hsieh CY, et al. I branching formation in erythroid differentiation is regulated by transcription factor C/EBPalpha. Blood. 2007;110:4526-4534.

22. Ule J, Stefani G, Mele A, et al. An RNA map predicting Nova-dependent splicing regulation. Nature. 2006;444:580-586.

23. Fagnani M, Barash Y, Ip JY, et al. Functional coordination of alternative splicing in the mammalian central nervous system. Genome Biol. 2007;8:R108.

24. Ip JY, Tong A, Pan Q, Topp JD, Blencowe BJ, Lynch KW. Global analysis of alternative splicing during T-cell activation. RNA. 2007;13:563-572.

25. Hung LH, Heiner M, Hui J, Schreiner S, Benes V, Bindereif A. Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. RNA. 2008;14:284-296.

26. Yu H, Yoon SH, Eber S, Prchal JT. Molecular defect of truncated β-spectrin associated with hereditary elliptocytosis. Blood. 1990;76, Suppl. 1:51a.

27. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860-921.

28. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. Science. 2001;291:1304-1351.

29. Williger BT, Provost JJ, Ho WT, Milstine J, Exton JH. Arfaptin 1 forms a complex with ADP-ribosylation factor and inhibits phospholipase D. FEBS Lett. 1999;454:85-89.

30. Keller MA, Addya S, Vadigepalli R, et al. Transcriptional regulatory network analysis of developing human erythroid progenitors reveals patterns of coregulation and potential transcriptional regulators. Physiol Genomics. 2006;28:114-128.

31. Welch JJ, Watts JA, Vakoc CR, et al. Global regulation of erythroid gene expression by transcription factor GATA-1. Blood. 2004;104:3136-3147.

32. Wold B, Myers RM. Sequence census methods for functional genomics. Nat Methods. 2008;5:19-21.

33. Parra MK, Gee SL, Koury MJ, Mohandas N, Conboy JG. Alternative 5' exons and differential splicing regulate expression of protein 4.1R isoforms with distinct N-termini. Blood. 2003;101:4164-4171.

34. Carninci P, Sandelin A, Lenhard B, et al. Genome-wide analysis of mammalian promoter architecture and evolution. Genome Res. 2006;38:626-635.

35. Kimura K, Wakamatsu A, Suzuki Y, et al. Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes. Genome Res. 2006;16:55-65.

36. Chretien S, Dubart A, Beaupain D, et al. Alternative transcription and splicing of the human porphobilinogen deaminase gene resulst either in tissue-specific or in housekeeping expression. Proc Natl Acad Sci U S A. 1988;85:6-10.

37. Aizencang GI, Bishop DF, Forrest D, Astrin KH, Desnick RJ. Uroporphyrinogen III synthase. An alternative promoter controls erythroid-specific expression in the murine gene. J Biol Chem. 2000;275:2295-2304.

38. Sahr KE, Taylor WM, Daniels BP, Rubin HL, Jarolim P. The structure and organization of the

human erythroid anion exchanger (AE1) gene. Genomics. 1994;24:491-501.

39. Gallagher PG, Birkenmeier C, Wang Y, Barker J, Forget BG. An alternate promoter of the erythrocyte ankyrin (ANK-1) gene directs expression of a muscle-specific transcript. Blood. 1995;86 Suppl. 1:415a.

40. Vannucchi AM, Linari S, Lin CS, Koury MJ, Bondurant MC, Migliaccio AR. Increased expression of the distal, but not of the proximal, Gata1 transcripts during differentiation of primary erythroid cells. J Cell Physiol. 1999;180:390-401.

41. Brudno M, Gelfand MS, Spengler S, Zorn M, Dubchak I, Conboy JG. Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. Nucleic Acids Res. 2001;29:2338-2348.

42. Sugnet CW, Srinivasan K, Clark TA, et al. Unusual Intron Conservation near Tissue-Regulated Exons Found by Splicing Microarrays. PLoS Comput Biol. 2006;2:e4.

43. Nakahata S, Kawamoto S. Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. Nucleic Acids Res. 2005;33:2078-2089.

44. Mauger DM, Lin C, Garcia-Blanco MA. hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. Mol Cell Biol. 2008;28:5403-5419.

45. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature. 2007;446:926-929.

46. Saltzman AL, Kim YK, Pan Q, Fagnani MM, Maquat LE, Blencowe BJ. Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. Mol Cell Biol. 2008;28:4320-4330.

FIGURE LEGENDS

Figure 1. Stage-specific switch of protein 4.1R exon 16 splicing in highly purified erythroblast cultures. A. FACS analysis of day 7 erythroblasts from three different preparations indicates purity is ≥97%, based on expression of erythroid markers for Glycophorin A and CD71. Quantitative analysis demonstrated erythroblast purities as follows: prep. #1, 97%; prep. #2, 97%; and prep. #3, 99%. B. RT-PCR scheme employed to analyze 4.1R pre-mRNA splicing in early (day 7) and late erythroblasts (day 14), using primers in the nearest constitutive exons 13 and 17. Gel image shows primarily exclusion of exon 16 in early erythoblasts (lower band), while substantial inclusion of exon 16 was observed in late erythroblasts (upper band). Alternative exons 14 and 15 are not expressed in erythroid cells.

Figure 2. Exon array detection of erythroid-specific beta spectrin mRNA 3' end. Upper panel depicts the exon structure of human beta spectrin transcripts in muscle versus erythroid cells, which express distinct 3' terminal exons. Lower panel shows the splicing index of probesets across the full length of the beta spectrin gene. Numbers along the horizontal axis represent probeset IDs. Positive values for the splicing index represent higher relative probeset expression in erythroblasts, while negative values indicate higher relative expression in muscle. The significant upward peak maps to the known erythroid-specific 3' end, while the downward peak represents the muscle-specific 3' end.

Figure 3. Stage-specific change in expression of alternative first exons. A. Splicing index shows normalized changes in probeset expression along the entire SLC12A6 transcript for day 10 (blue curve) and day 14 erythroblasts (red curve), relative to expression at day 7 (black line). Results show a stage-dependent increase in expression of probesets representing the 1A region (boxed), and a decrease in expression of probesets for exon 1B. B. RT-PCR validation of exon array predictions for SLC12A6 and two additional genes. Shown are gels of PCR products validating alternative splicing

switches in first exon expression (left) and diagrams of the relevant pre-mRNA regions (right). Black arrow / black lines indicate predominant pattern in day 7 erythroblasts, and red arrow / red lines indicate predominant pattern in day 14 erythroblasts. Alternative first exons are indicated by 1A, 1B, 1C and a shared constitutive exon indicated as exon 2. Common names of the alternatively spliced transcripts are as follows: SLC12A6, KCl co-transporter 3 (KCC3); TNPO2, transportin 2 (a nuclear import protein); GCNT2, glucosaminyl (N-acetyl) transferase 2 (generates the branched chain carbohydrate structure that constitutes the I antigen).

Figure 4. Exon junction array detection of the 4.1R exon 16 splicing switch. Probesets in the exon 16 region that exhibit significant changes in relative expression between day 7 and day 14 were mapped to the human genome using the UCSC BLAT alignment tool. Bottom panel shows the exon structure of the two mRNA isoforms expressed from the 4.1R gene; arrow indicates alternative exon 16. Upper panel shows that probesets interrogating the exon 16 inclusion event were up-regulated at day 14, while a reciprocal decrease in expression of the exon 13-17 skipping event was observed. RT-PCR validation of this splicing switch is shown in Figure 1. Alternative exons 14 and 15 are not expressed in these cultures.

Figure 5. Novel stage-specific alternative splicing switches in erythroid genes. A. General scheme for detection of alternative splicing of a pre-mRNA (left) into the mRNA isoform including the alternative exon (upper right) or the mRNA skipping this exon (lower right). Diagnostic isoform-specific microarray probes are indicated above the spliced mRNAs, while PCR primers used for validation are shown below the mRNAs (arrows). In addition, there are exon probes for the first and third exons that hybridize equally to both isoforms and are useful for determining overall transcript levels. B. Shown are gels of PCR products validating alternative splicing switches in late erythropoiesis (left) and diagrams of the relevant pre-mRNA regions (right). Gels demonstrate substantial increases in

exon inclusion products (upper bands, indicated by arrows), relative to exon skipping products (lower bands), at day 14. The deduced splicing patterns are indicated at the right, with black lines indicating major splice patterns at day 7 and red lines indicating predominant splice pattern at day 14. Asterisks indicate positions of stop codons (not shown for ARFIP1 and PLD1 because they are located farther downstream). Common names of the alternatively spliced transcripts are as follows: HNRPLL, heterogeneous nuclear ribonucleoprotein L-like (an hnRNP protein); SNRP70, U1 small nuclear ribonucleoprotein 70K (a component of the U1 snRNP); MBNL2, muscleblind 2 (RNA binding proteins with known splicing regulatory activity); ATP11C, ATPase class VI type 11C; ARFIP1, ADP-ribosylation factor interacting protein 1; PLD1, phospholipase D1.

Table 1.

gene (exon length)	probeset location	probeset ID	Splicing Index Magnitude	P value	direction	comments
EPB41 63nt	alt exon	217437	2.16	9.0E-03*	up	In-frame insertion of 21aa
	5' junction	504900	2.20	7.5E-07	up	
	3' junction	224956	2.33	1.6E-09	up	
	skip junction	366937	-0.93	1.2E-04	down	
ARFIP1 96nt	alt exon	497178	0.92	1.6E-05	up	In-frame insertion of 32aa
	5' junction	425212	0.82	1.8E-04	up	
	3' junction	79253	0.97	3.8E-03*	up	
	skip junction	524645	-0.62	2.4E-05	down	
ATP11C 66nt	alt exon	462451	0.67	7.2E-02*	up	Insertion of stop codon creating alternative C-term. sequence and possible NMD
	alt exon	20030	0.63	8.3E-03*	up	
	5' junction	521875	0.86	5.8E-04	up	
	3' junction	645995	1.01	1.8E-05	up	
	skip junction	603987	-1.03	3E-04	down	
HNRPLL 83nt	alt exon	376564	1.11	3.2E-04	up	Insertion of stop codon creating alternative C-term. sequence and possible NMD
	alt exon	418887	1.32	7.0E-04	up	
	5' junction	291050	1.39	3.6E-05	up	
	3' junction	679040	1.24	1.0E-02*	up	
	skip junction	376125	-0.50	3.6E-04	down	
MBNL2 95nt	alt exon	253023	0.69	5.6E-05	up	Creates alternative C-term. sequence; NMD unlikely due to 3' location of insertion
	5' junction	473479	1.01	1.3E-06	up	
	3' junction	636764	0.94	3.8E-04	up	
	skip junction	276844	-2.55	1.7E-05	down	
PLD1 114nt	alt exon	644361	0.52	7.7E-05	up	In-frame insertion of 38aa
	5' junction	657499	0.57	6.3E-05	up	
	3' junction	652587	0.56	7.3E-02*	up	
	skip junction	421462	-1.87	1.2E-04	down	
SNRP70 72nt	alt exon	178662	1.14	7.0E-03*	up	Insertion of stop codon creates alternative C-term. sequence and possible NMD
	5' junction	38132	1.74	2.7E-06	up	
	3' junction	475296	0.68	1.6E-03	up	
	skip junction	585925	-1.08	1.2E-04	down	

* Probeset did not pass significance filters (p-value < 0.001)

Figure 1

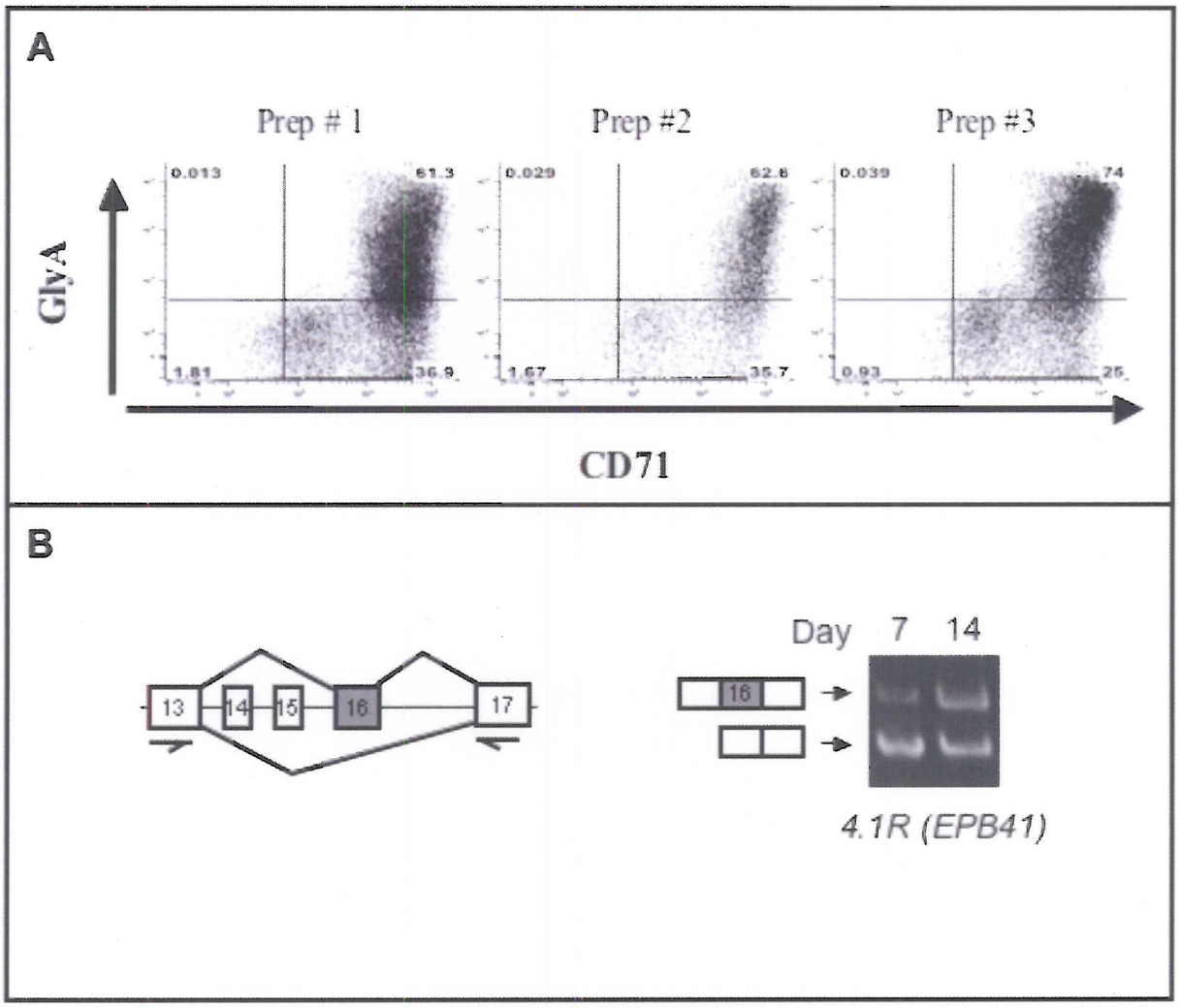

A
Prep # 1
Prep #2
Prep #3
0.013
61.3
1.81
36.9
0.029
62.8
1.67
35.7
0.039
74
0.93
25
GlyA
CD71
B
13
14
15
16
17
Day 7 14
16
4.1R (EPB41)

Figure 2

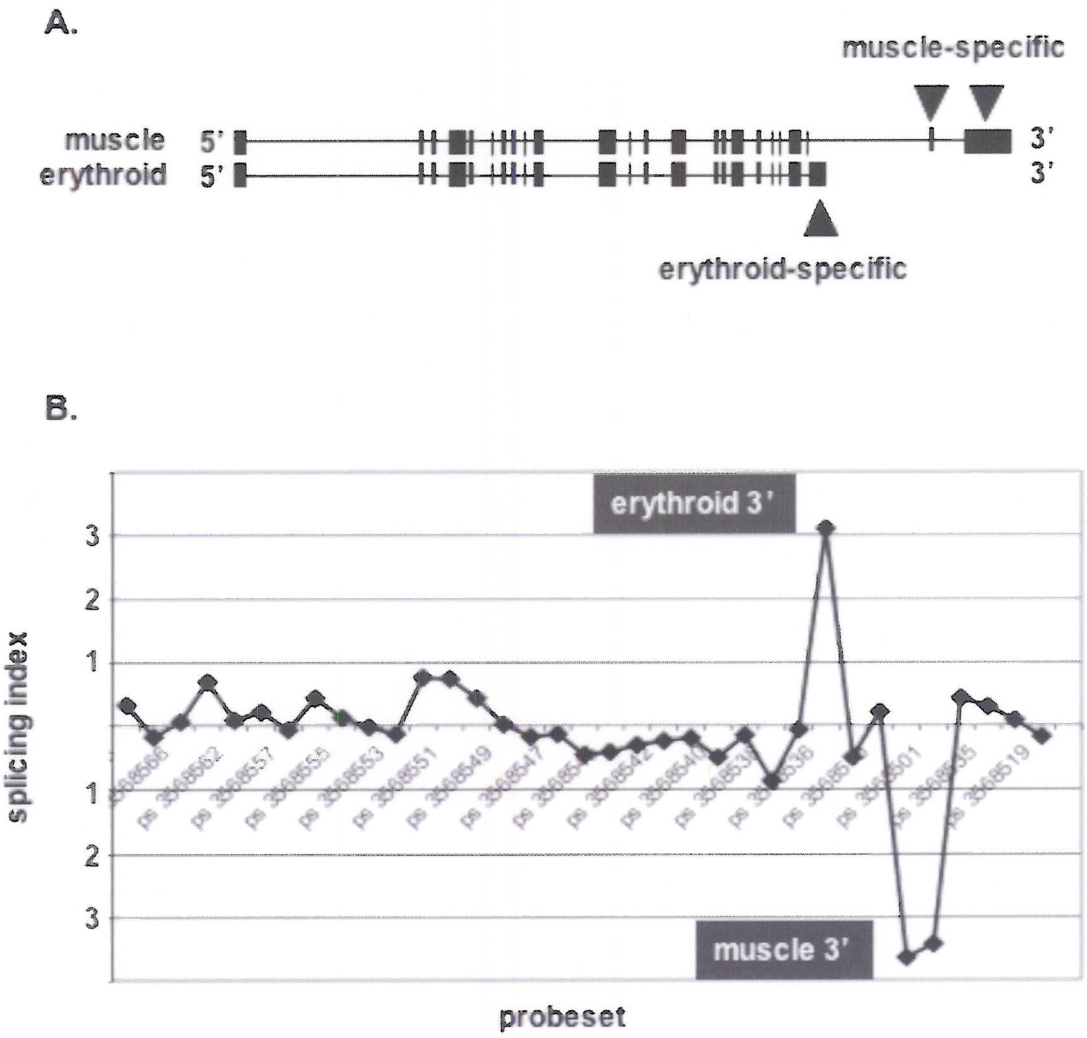

A.
muscle-specific
muscle 5' 3'
erythroid 5' 3'
erythroid-specific
B.
erythroid 3'
muscle 3'
splicing index
3
2
1
1
2
3
probeset

Figure 3

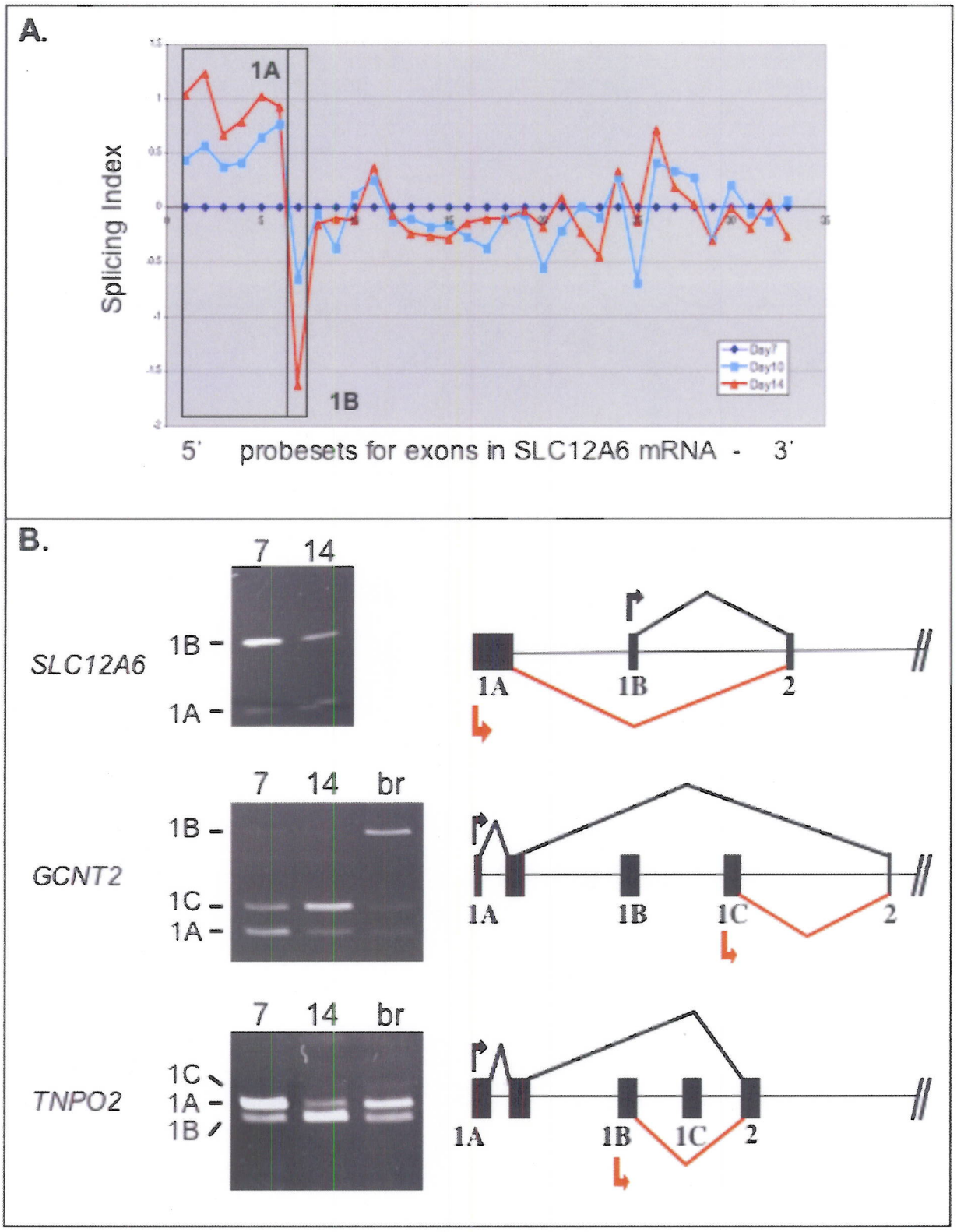

A.
1A
1B
Splicing Index
Day7
Day10
Day14
5' probesets for exons in SLC12A6 mRNA - 3'
B.
7 14
SLC12A6
1B
1A
1A 1B 2
7 14 br
GCNT2
1B
1C
1A
1A 1B 1C 2
7 14 br
TNPO2
1C
1A
1B
1A 1B 1C 2

Figure 4

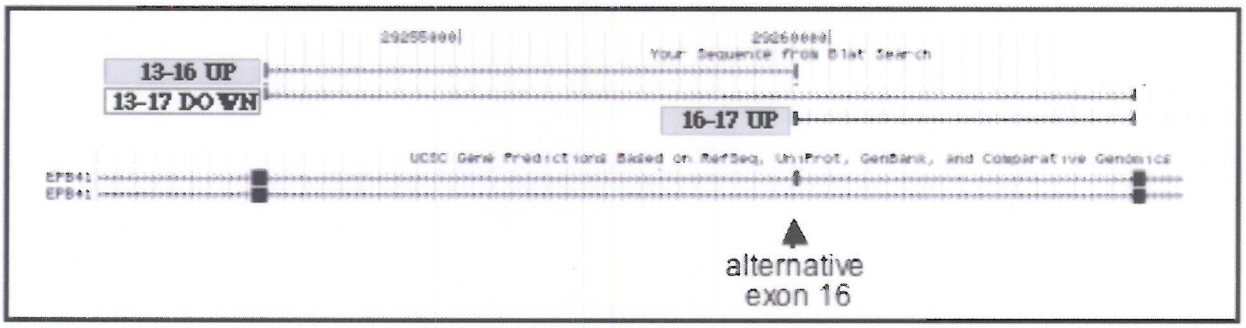


Figure 5A

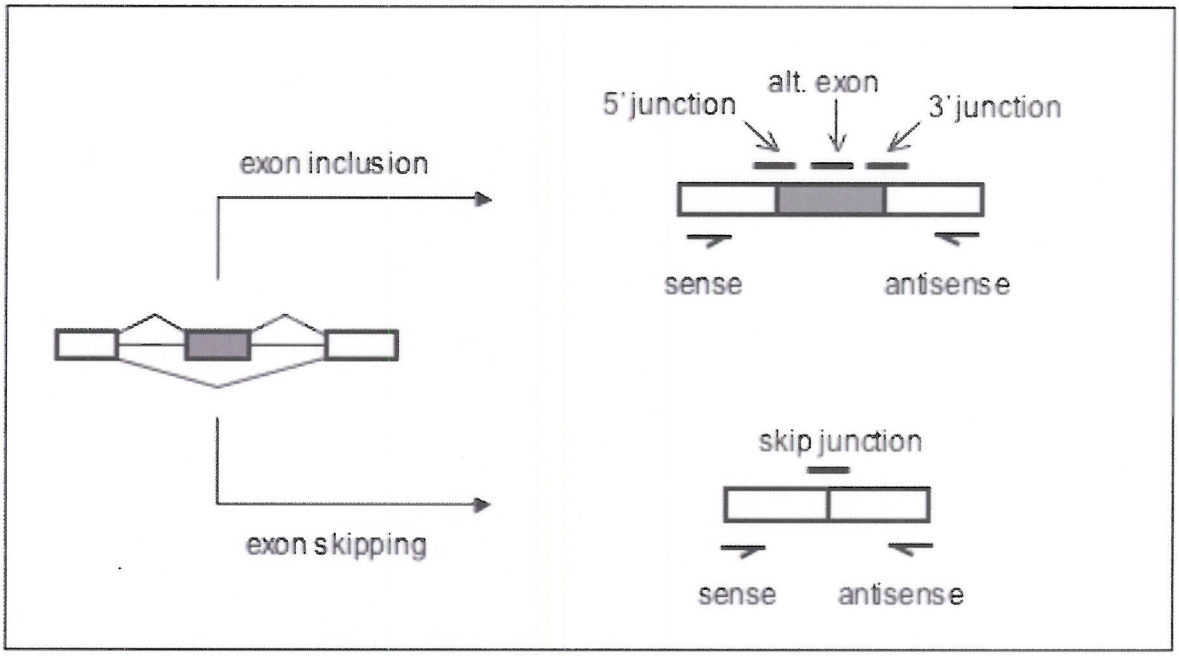
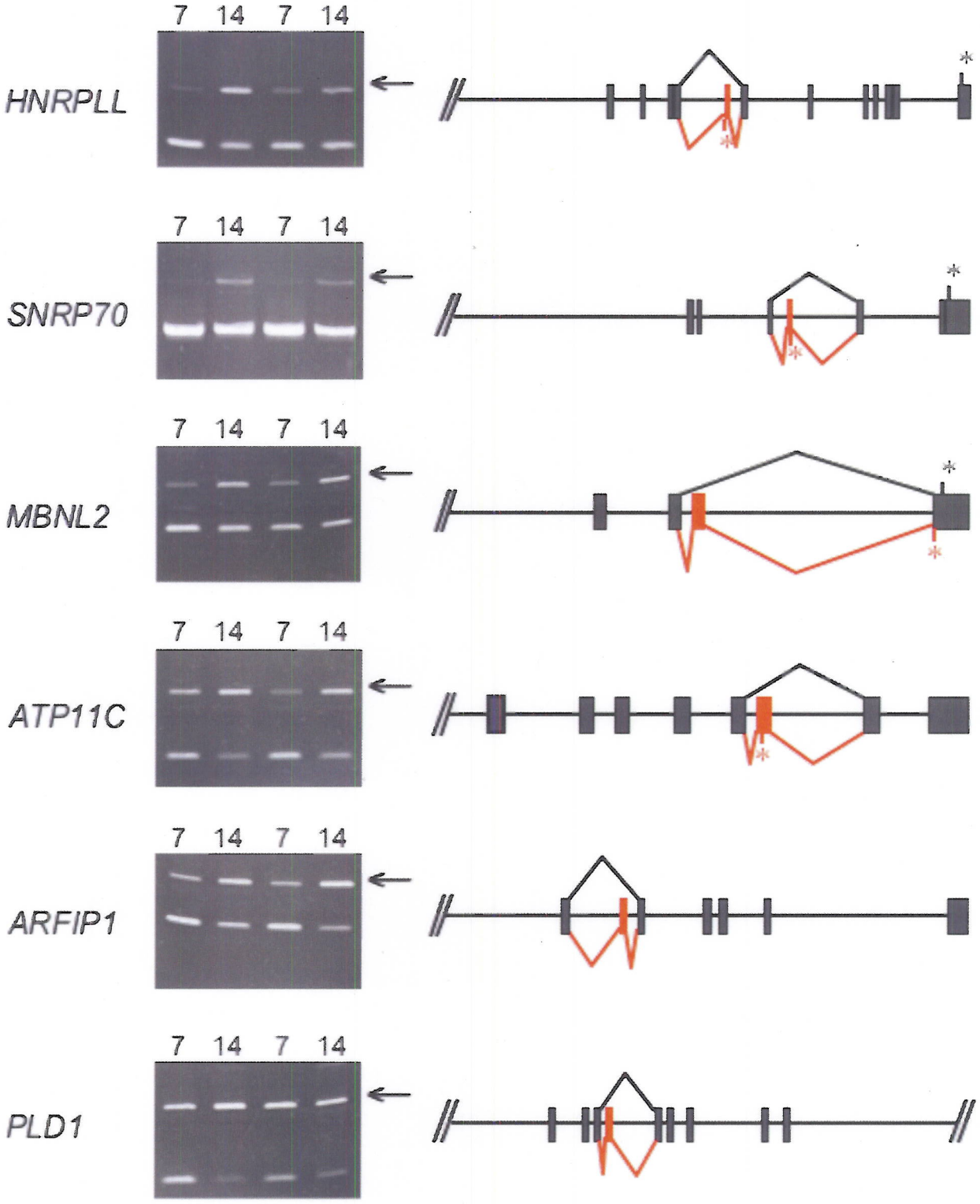

exon inclusion
exon skipping
5' junction
alt. exon
3' junction
sense
antisense
skip junction
sense
antisense

Figure 5B


7 14 7 14
HNRPLL
7 14 7 14
SNRP70
7 14 7 14
MBNL2
7 14 7 14
ATP11C
7 14 7 14
ARFIP1
7 14 7 14
PLD1