Differential domain evolution and complex RNA processing in a family of paralogous EPB41 (protein 4.1) genes facilitates expression of diverse tissue-specific isoforms

Marilyn Parra[1], Sherry Gee[1], Nadine Chan[1], Dmitriy Ryaboy[1], Inna Dubchak[1], Mohandas Narla[2], Philippe D. Gascard[1], John G. Conboy[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, California 94720; [2]New York Blood Center, New York, New York 10021,

corresponding author:

John G. Conboy
Lawrence Berkeley National Laboratory
Life Sciences Division
Mailstop 74-157
1 Cyclotron Road
Berkeley, CA 94720

telephone 510-486-6973
fax 510-486-6690
e-mail jgconboy@lbl.gov

Abstract

The EPB41 (protein 4.1) genes epitomize the resourcefulness of the mammalian genome to encode a complex proteome from a small number of genes. By utilizing alternative transcriptional promoters and tissue-specific alternative pre-mRNA splicing, *EPB41*, *EPB41L2*, *EPB41L3*, and *EPB41L1* encode a diverse array of structural adapter proteins. Comparative genomic and transcript analysis of these 140kb-240kb genes indicates several unusual features: differential evolution of highly conserved exons encoding known functional domains, interspersed with unique exons whose size and sequence variations contribute substantially to intergenic diversity: alternative first exons, most of which map far upstream of the coding regions; and complex tissue-specific alternative pre-mRNA splicing that facilitates synthesis of functionally different complements of 4.1 proteins in various cells. Understanding the splicing regulatory networks that control protein 4.1 expression will be critical to a full appreciation of the many roles of 4.1 proteins in normal cell biology and their proposed roles in human cancer.

Introduction

The human genome encodes a highly complex proteome via elegant regulatory machinery that facilitates expression of many polypeptides from a much smaller number of genes. This study focuses on the structural organization of the *EPB41*-related (protein 4.1) genes, a group of four paralogous genes [1, 2, 3, 4, 5, 6] that together encode an impressively complex array of cytoskeletal proteins that are expressed in unique tissue-specific patterns and play many diverse roles in cell biology. The prototypical *EPB41* gene is best known for encoding the protein familiarly known as protein 4.1R, which plays a structural role in the red cell membrane skeleton via its interactions with skeletal and integral proteins that are critical for normal erythroid mechanical and morphological properties [7, 8, 9]. More generally, protein 4.1 polypeptides organize membrane protein domains and/or to link membranes to internal cytoskeletal networks. Among the proposed sites of 4.1 function are tight junctions of epithelial cells, the contractile apparatus in muscle cells [10], neuronal membrane receptor complexes [11, 12, 13], paranodal axoglial junctions at nodes of Ranvier [14], centrosomes [15], mitotic spindles [16, 17], and nuclear matrix [18, 19].

Defects in the *EPB41*-related genes have been demonstrated or proposed to underlie human disease. One manifestation of EPB41 (protein 4.1R) deficiency is the production of red blood cells with abnormal morphology and unstable membranes leading to hereditary elliptocytosis and hemolytic anemia [20]. Studies with *Epb4.1* knockout mice [21] revealed phenotypes that extend beyond the familiar erythroid disorders caused by EPB41 deficiency, to include neuronal defects in fine motor coordination and spatial learning [22]. Moreover, recent studies suggest that EPB41 and its paralogs may function as tumor suppressors [23, 24, 25, 26]. Therefore, it will be important to understand the regulatory mechanisms whereby expression of these genes is controlled.

Characterization of *EPB41* gene expression has revealed a complex pattern of tissue- and developmental- specific alternative pre-mRNA splicing. Combinatorial expression of at least 10 alternative splicing events can potentially facilitate synthesis of hundreds of distinct protein isoforms, although only a subset of these is likely expressed in a given tissue (reviewed in [27]). Of these, a select few are subject to exquisite regulation. Particularly important in erythroid cells is a dramatic splicing switch in EPB41 structure that has important consequences for membrane mechanical stability: whereas early progenitors skip exon 16 and synthesize protein isoforms with low affinity for spectrin and actin, later erythroblasts efficiently include exon 16 and produce high affinity isoforms of EPB41 protein [28, 29]. Exon 16 splicing is influenced by multiple cis elements in the pre-mRNA [30, 31], and recent studies indicate that the splicing switch may be

mediated by downregulation in the expression of the splicing silencer protein, hnRNP A1 [31]. Additional alternative splicing events have been characterized in muscle, epithelial cells, kidney, and brain [2, 32, 33, 34, 35].

In this paper we present the first characterization of the large *EPB41L1*, *EPB41L2*, and *EPB41L3* genes and establish a consistent nomenclature to facilitate comparison with the prototypical *EPB41* gene. Moreover, we present new data to demonstrate that several notable features of the well-characterized *EPB41* gene are conserved among these three additional genes, including the presence of far upstream alternative first exons, extensive alternative pre-mRNA splicing of internal exons, and a unique pattern of differential domain evolution involving both highly conserved and highly variable regions within each of the four genes. Understanding the regulated expression of these genes will not only shed light on the tremendous resourcefulness of the human genome to generate a complicated proteome from many fewer genes, but will also be important for the cell biology of the 4.1 proteins.

Results

Protein conservation patterns indicate differential evolutionary rates of individual domains. We assembled composite cDNAs for the four *EPB41*-related genes, each of which encodes proteins with the general domain structure depicted in the top panel of Figure 1A. Other genes with similar names (*EPB41L4A*, *EPB41L4B*, *EPB41L5*, and *EPB41LO* contain only the FERM domain and represent a different subclass within the protein 4.1 superfamily. Comparison of the deduced amino acid sequences for the EPB41-related proteins revealed an alternating pattern of highly conserved domains interspersed with poorly conserved domains (Figure 1). Pairwise comparisons of the human and mouse orthologs showed that the most conserved region is the membrane binding FERM domain (average 98% identity), followed by the carboxy terminal domain (CTD; 96% identity) and the spectrin-actin binding domain (SABD; 93% identity). Much greater variability occurs in unique regions U1 (average 80% identity between mouse and human), U2 (74%), and U3 (70%). Importantly, a similar evolutionary pattern was observed in all four genes, indicating a consistent differential evolution of domains in the 75 million years since the divergence of human and mouse.

Differential domain evolution was explored further by comparing sequences among the human EPB41-related paralogs, since these genes clearly diverged from one another much earlier in vertebrate evolution. Human EPB41 (protein 4.1R), EPB41L2 (protein 4.1G), and EPB41L3 (protein 4.1B) proteins exhibited substantial homologies in the FERM (72-81% identity), SAB (53-66% identity), and CTD domains (72-74% identity) (Figure 1B); EPB41L1 (protein 4.1N) was also very similar except for partial domain deletions in the SAB (due to loss of E16) and in the CTD (due to loss of E18). In marked contrast, very little homology was observed in pairwise comparisons of the unique regions (not shown). These results confirm the overall finding of differential evolutionary rates for the functional FERM, SAB and CTD domains versus the unique domains.

Gene organization. The *EPB41* gene is ~240kb in length, including about 140kb that spans coding exons 2-21 and another 100kb of 5' region containing three known alternative promoters/ first exons [33, 36]. Mapping of the composite cDNAs for the three orthologous genes onto the corresponding genomic assembly (Figure 2)revealed that they are similarly complex genetic loci of approximately 140kb (*EPB41L1*) or 240kb (*EPB41L2*, and *EPB41L3*). A common organizational pattern among the members of this gene family includes the following unusual features (Figure 2 and Tables I and II): (1) a coding region consisting of three sets of highly conserved paralogous exons, interspersed with three sets of highly variable exons that are unique to each gene; (2) the finding of multiple first exon/promoter regions, most of which map very far upstream of the coding region(s); and (3) complex tissue-specific alternative splicing patterns.

In order to facilitate intergenic comparisons, we have adopted a consistent nomenclature in which the highly conserved coding regions have been assigned the same exon numbers originally designated in the *EPB41* gene (Figure 2, shaded regions). Thus, the FERM domain is encoded by the end of exon 4 plus exons 5-12; the SABD by exons 16-17; and the CTD by exons 18-21. Except for the lack of exons 16 and 18 in the *EPB41L1* gene, exon sizes and exon/intron boundaries are highly conserved among these domains, particularly in the central "core" coding regions of each domain. Most remarkable is the precise conservation of size of exons 5-12 in all four members of the gene family. These core regions correspond well to previously defined structural and functional domains of the protein, since all binding sites mapped thus far at the protein level reside within these core regions.

The unique regions, designated as U1, U2, and U3, exhibit substantially more variation in size, number, and sequence of exons (Figure 2, unshaded regions; Table I). The U1 domains, encoded by exon 2 and part of exon 4, comprise the N-terminal extension, or headpiece domain, of 4.1 proteins. Although the predicted amino acid sequence of headpiece domains exhibits limited homology among the four paralogs, a common origin is suggested by the presence of a conserved tetrapeptide (MTTE) at the N-terminus of each U1 domain. The U2 domain separates the FERM and SAB domains, and is encoded by a single poorly conserved, constitutively spliced exon 13, plus up to three additional alternatively spliced exons of unknown function. Finally, the most complex region among these genes is the U3 domain, a unique and highly variable domain encoded by up to six alternatively spliced exons designated 17A through 17F. It should be noted that U3 exons assigned the same number among the *EPB41* paralogs exhibit little or no sequence homology and are paralogs in position only.

Differential RNA processing events in the 4.1 gene family. A prominent hallmark of *EPB41* is the expression of multiple protein isoforms translated from differentially processed RNA transcripts. Both the use of alternative first exons, as well as alternative pre-mRNA splicing of internal exons, contribute to transcript diversity. With regard to *EPB41L1*, *EPB41L2*, and *EPB41L3*, published studies indicate there are differences in RNA processing between kidney and brain [2, 35, 37], but differential RNA processing in these genes has generally been less well characterized. The new data below demonstrate that (1) similar to *EPB41*, the three paralogous genes all possess candidate alternative first exons, many of which map far upstream of the coding regions, and (2) all three genes also exhibit tissue-specific alternative splicing of internal cassette exons in the U2 and U3 regions. Together this data suggests that regulated RNA processing events play a major role in facilitating expression of a distinct protein isoforms from the small *EPB41*-related gene family.

Alternative first exons. Multiple first exons, each with its own transcriptional promoter, were identified in *EPB41* through the finding of cDNAs that contain a common set of internal exons but possess unique 5' end sequences [36]. Similar analysis has now revealed multiple 5' end sequences in the *EPB41L1*, *EPB41L2*, and *EPB41L3* genes (Figure 2; Table II). All of these predicted first exons meet the expected criteria of proper chromosomal location upstream of, and proper splicing to, the authentic coding exons; all are conserved between mouse and human genomes. For *EPB41L1* and *EPB41L3* these alternative first exons map far upstream (20-130kb) of the coding exons in the human genome assembly. *EPB41L2* contains a single far upstream (~106kb) first exon, designated exon 1A, and a candidate internal first exon designated as exon 1B. Exon 1B is located just upstream of exon 16; it is conserved in the human, monkey, and mouse genomes; and it splices accurately to the downstream coding exons. These features strongly suggest a bona fide role for exon 1B in *EPB41L2* gene expression, and thus it is included in the gene maps presented here.

Alternative pre-mRNA splicing. Analysis of prototypical *EPB41* gene expression has demonstrated numerous alternative splicing events, several of which are regulated in tissue-specific fashion. This section combines the existing *EPB41* data with evidence for four novel tissue-specific alternative splicing events in *EPB41L1*, *EPB41L2*, and *EPB41L3* (Figure 4 below). Together these analyses suggest the following: (1) Alternative splicing is more common in the unique coding regions of the *EPB41*-related genes than in the conserved domains; (2) alternative exons in the unique domains are more highly regulated in tissue-specific patterns than those in the conserved domains (with the notable exception of alternative exon 16 in the spectrin-actin binding domain); (3) there are gene-specific differences in splicing regulation of paralogous exons; and (4) several alternative exons in the U3 domain are considerably larger (369-1932nt) than the average alternative exon in mammals (<100nt).

Among the conserved domains, only exon 16 in the SABD (*EPB41*, *EPB41L2*, and *EPB41L3*) and exon 21 in the CTD (*EPB41L3* only) exhibit tissue-specific alternative splicing. This data has been published in separate reports [2, 5, 6, 32, 35, 37, 38], and is brought together in Figure 3 to emphasize a couple of interesting features. First, exon 16 exhibits gene- and tissue-specific splicing patterns, such that *EPB41*, *EPB41L2*, and *EPB41L3* each directs synthesis of proteins capable of binding spectrin and actin in a limited but distinct spectrum of cell types. Second, the regulated splicing of exon 21 in only one of the four paralogs represents another example of evolutionary divergence in splicing patterns among highly conserved exons [2]. Alternative splicing events are otherwise rare in the conserved domains, although minor *EPB41* isoforms lacking exons 5, 8, 18, 19, and 20 are predicted from analysis of cloned cDNAs [32, 38].

Considerably more complex splicing is evident in the unique regions of the *EPB41*-family genes. In both the U2 (Figure 4A) and U3 (Figure 4B) regions, comparison of available cDNA and EST clones from the genetic databases suggests that most of the exons are alternatively spliced. Although a complete exploration of tissue-specificity is beyond the scope of this paper, our recent studies have revealed four new examples of tissue-specific splicing in these regions. Figure 4A shows that *EPB41L1* exon 15 is strongly included in brain (lane 1), but mainly excluded in all other tissues tested (lanes 2-7). Similarly, *EPB41L1* (exon 17D), *EPB41L2* (exon 17B), and *EPB41L3* (exon 17D) are all included in brain but predominantly or completely skipped in other tissues (Figure 4B bottom panel, compare lanes 2 with other lanes). Combined with previous reports of tissue-specific alternative splicing of *EPB41* exon 17A (muscle) and 17B (epithelial cells) [33, 34], this new data reveals an extraordinary abundance of regulated alternative splicing events among these unique domains.

Comparison of the human and mouse EPB41 family genes reveals a highly conserved organization. The *EPB41*-related genes of human and mouse were compared using the VISTA program developed for large-scale genome comparisons (Figure 5). These VISTA profiles revealed significant peaks of homology for all coding exons, indicating a conserved exon/intron organization. Interestingly, substantial homology is observed in selected intronic regions that may represent previously unrecognized exons or regulatory sequences important for gene expression. An example of the former may be the extensive peak of homology downstream of *EPB41L2* exon 13, which maintains an open reading frame of >170 amino acids that is conserved in the human, mouse, and rat genomes and exhibits significant homology to alternative exon 13A in the corresponding region of the *EPB41L3* gene (57% identity over 101 amino acids). Regarding the latter, we note that the most extensive noncoding homologies occur in the U3 region and speculate that these could represent important regulatory sequences for controlling the tissue-specific alternative splicing of U3 exons.

Discussion

This report presents a comparative genomics view of the complex *EPB41*-related (protein 4.1) family, and highlights a number of mechanisms by which these genes can encode a highly diverse set of tissue-specific polypeptides. Through the use of alternative first exons and alternative pre-mRNA splicing, the structure and function of these important proteins can be specifically adapted to the needs of each individual cell type. Differential domain evolution may play a particularly important role, perhaps allowing the unique regions to modify basic functions of the FERM, SABD, and CTD domains, resulting in altered subcellular targeting [39, 40, 41] or changes in affinity of these domains for interacting proteins. Alternatively, the unique domains may themselves encode novel localization signals or interacting domains that alter protein 4.1 function. It is important to note that an integral component of this evolutionary process must be a tissue-specific regulatory network of RNA processing events that orchestrates expression of a structurally and functionally unique complement of polypeptides in various cell types. Future studies will focus on understanding both the regulation of expression at the RNA level, and the cell biological functions at the protein level, of the complex EPB41-related family.

The finding of alternative first exons in the *EPB41*-related family expands the rapidly growing list of mammalian genes that possess multiple first exons. We speculate that expression of multiple alternative first exons is a very common phenomenon among mammalian genes, perhaps as common as alternative pre-mRNA splicing, and that it may be particularly enriched in that exhibit complex splicing patterns. It will be interesting in the future to explore whether these 5' transcription events and downstream splicing events are functionally coupled [42, 43, 44, 45].

The comprehensive picture of 4.1 genomics also provides new insights into what could be referred to as "4.1 proteomics". It now seems evident that complex expression patterns detected in older immunoblot studies was due not only to the presence of multiple genes, but also to the ability of each gene to express multiple polypeptides. Theoretically, combinatorial expression of *EPB41* exons 5, 8, 14, 15, 16, 17A, 17B, 18, 19, and 20 could generate >1000 distinct isoforms; this number is expanded further via the expression of isoforms with the extended N-terminal domain (exon 2) and the use of alternative splice donor sites in exons 17A and 17B. A comparable or even greater number of isoforms may arise from the *EPB41L1*, *EPB41L2*, and *EPB41L3* genes, due to the greater complexity in the U3 region. Thus, the total number of 4.1 polypeptides could potentially approach 10,000. Recent surveys of *EPB41* cDNAs actually expressed in a single cell type have reported,

however, that only a few major cDNA isoforms are present in human erythroblasts [39], mouse skeletal muscle [10], and mouse kidney [35]. Many rare isoforms can also be expressed at low levels in these cells.

The versatility in 4.1 protein structure is in direct contrast to the related ERM family of proteins that includes ezrin, radixin, moesin and the tumor suppressor schwannomin/ merlin. Although the 4.1 and ERM proteins share a conserved membrane binding domain that is similar in primary sequence and crystal structure, the ERM proteins are encoded by simpler genes that lack the impressive plasticity of the true 4.1 genes. To our knowledge, no alternative first exons, relatively little alternative splicing, and little or no evidence of rapidly evolving unique domains have been reported in the ERM family.

Protein 4.1R was originally characterized as a major spectrin-actin binding protein in the membrane skeletal network of erythrocytes, and much of our current understanding of 4.1 function has focused on these interactions. Recent studies have shown that isoforms of EPB41L2 and EPB41L3 can also form ternary complexes with spectrin and actin [46]. However, it is important to recognize that much, perhaps most, of the nonerythroid protein 4.1 expressed in mammals lacks a high affinity SAB domain (no exon 16 paralog in EPB41L1; splicing out in many tissues of exon 16 from *EPB41*, *EPB41L2*, and *EPB41L3*). Interestingly, a more ancient homolog of protein 4.1 in *Drosophila melanogaster* [47] does not contain an obvious SAB domain, suggesting that this is a relatively recent adaptation to 4.1 function. The bulk of nonerythroid 4.1 proteins might therefore exhibit distinct functions. Based on recent reports demonstrating interaction of 4.1 proteins with various membrane proteins, it appears that 4.1 proteins play important roles as versatile membrane adapter proteins to localize and link functionally related integral membrane proteins.

In conclusion, the *EPB41*-related gene family serves as an important model system for understanding the complexities of human genome expression patterns, including the exquisite regulatory controls that facilitate differential expression of specific transcripts in various cell types. This knowledge will be critical for elucidating the functional correlates of the corresponding 4.1 polypeptides, in what is likely to be an equally exquisite array of adapter functions for 4.1 protein isoforms in various physiological settings in these cells.

Materials and Methods

Assembly of composite *EPB41*-related cDNAs. The composite transcripts used for mapping to the human genome assembly were derived from the relevant Refseq entries, annotated as needed to include additional exons not represented in Refseq. ***EPB41*** **(protein 4.1R)**. Refseq NM_004437 includes alternative first exon 1A plus most of the coding exons with the exception of exons 14, 15, 17A, 17B, and 19. The source of other exons is as follows: exons 1B and 1C, [36]; exons 14 and 15, [32]; exon 17A, [33]; exon 17B, [34]; exon 19 [9]. ***EPB41L2*** **(protein 4.1G)**. Refseq NM_001431 includes alternative first exon 1A plus most of the coding exons with the exception of 17A, 17B, and 17C. The source of other exons is as follows: exon 1B, AK127088, exons 17A and 17C, AJ251209; exon 17B, by homology to the corresponding mouse exon in AJ428901. ***EPB41L3*** **(protein 4.1B)**. Refseq NM_012307 includes alternative first exon 1B plus most of the coding exons with the exception of exons 13A, 15, 16, 17A, and 17B. The source of other exons is as follows: exon 1A, BC008377; exons 15 and 17D, AB023204; exon 17A, W28202, exon 17B, AB032828; exon 17C, predicted by Genescan plots and contained fully in the overlap of CB738972 and AK027921. Exon 16 was reported earlier [2]. ***EPB41L1*** **(protein 4.1N)**. Refseq NM_012156 includes alternative first exon 1B plus most of the coding exons with the exception of 17A, 17B, and 17C. The source of other exons is as follows: exon 1A, AK096848; exon 17A, AB002336; exons 17B and 17C, BX537978.

Comparison of human and mouse genes. VISTA is a program for visualizing global DNA sequence alignments [48] produced by AVID [49]. The results in this paper show the percent identity between mouse and human 4.1 genes using a sliding window of 50nt. Regions of high conservation were defined as sequences at least 70% identical over a span of at least 50nt.

Reverse transcriptase/PCR amplification of 4.1 mRNA sequences. Splicing patterns of 4.1N transcripts in the U2 region were characterized using RT-PCR techniques to amplify 4.1N mRNA from several different mouse tissue sources. RNA was prepared using RNeasy columns according to the manufacturer's instructions (Qiagen, Valencia, CA). One μg total RNA was transcribed into cDNA using a random hexamer primers in a total volume of 10 μL. Then, 2 μL cDNA was amplified using the following primers: forward, 5'-GCCGATCAGAGGCTGAAGAAGG-3'; reverse, 5'-CCTCTTGAGTTCATTGATGCTGG-3'. Thirty-five cycles of amplification were performed under the following conditions: denaturation for 30 s at 94 °C; annealing for 30 s at 60 °C; extension for 40 s at 72 °C. DNA fragments were analyzed by 5% polyacrylamide gel electrophoresis. The identity of PCR products was confirmed by DNA sequence analysis.

Acknowledgements

Supported by National Institutes of Health (NIH) grants HL45182 and DK32094; by the Director, Office of Biological and Environmental Research, US Department of Energy under contract DE-AC03-76SF00098.

References

1. Peters L. L., Weier H. U., Walensky L. D., Snyder S. H., Parra M., Mohandas N., and Conboy J. G., Four paralogous protein 4.1 genes map to distinct chromosomes in mouse and human. *Genomics* **54** (1998), pp. 348-50.

2. Parra M., Gascard P., Walensky L. D., Gimm J. A., Blackshaw S., Chan N., Takakuwa Y., Berger T., Lee G., Chasis J. A., Snyder S. H., Mohandas N., and Conboy J. G., Molecular and functional characterization of protein 4.1B, a novel member of the protein 4.1 family with high level, focal expression in brain. *J Biol Chem* **275** (2000), pp. 3247-55.

3. Parra M., Gascard P., Walensky L. D., Snyder S. H., Mohandas N., and Conboy J. G., Cloning and characterization of 4.1G (EPB41L2), a new member of the skeletal protein 4.1 (EPB41) gene family. *Genomics* **49** (1998), pp. 298-306.

4. Walensky L. D., Gascard P., Fields M. E., Blackshaw S., Conboy J. G., Mohandas N., and Snyder S. H., The 13-kD FK506 binding protein, FKBP13, interacts with a novel homologue of the erythrocyte membrane cytoskeletal protein 4.1. *J Cell Biol* **141** (1998), pp. 143-53.

5. Ohara R., Yamakawa H., Nakayama M., and Ohara O., Type II brain 4.1 (4.1B/KIAA0987), a member of the protein 4.1 family, is localized to neuronal paranodes. *Brain Res Mol Brain Res* **85** (2000), pp. 41-52.

6. Ohara R., Yamakawa H., Nakayama M., Yuasa S., and Ohara O., Cellular and subcellular localization of a newly identified member of the protein 4.1 family, brain 4.1, in the cerebellum of adult and postnatally developing rats. *Brain Res Dev Brain Res* **117** (1999), pp. 127-38.

7. Bennett V., The spectrin-actin junction of erythrocyte membrane skeletons. *Biochim Biophys Acta* **988** (1989), pp. 107-21.

8. Takakuwa Y., Tchernia G., Rossi M., Benabadji M., and Mohandas N., Restoration of normal membrane stability to unstable protein 4.1-deficient membranes by incorporation of purified protein 4.1. *J. Clin. Invest.* **78** (1986), pp. 80-85.

9. Conboy J. G., Structure, function, and molecular genetics of erythroid membrane skeletal protein 4.1 in normal and abnormal red blood cells. *Sem. Hemat.* **30** (1993), pp. 58-73.

10. Kontrogianni-Konstantopoulos A., Huang S. C., and Benz E. J., Jr., A nonerythroid isoform of protein 4.1R interacts with components of the contractile apparatus in skeletal myofibers. *Mol Biol Cell* **11** (2000), pp. 3805-17.

11. Shen L., Liang F., Walensky L. D., and Huganir R. L., Regulation of AMPA receptor GluR1 subunit surface expression by a 4. 1N-linked actin cytoskeletal association. *J Neurosci* **20** (2000), pp. 7932-40.

12. Maximov A., Tang T. S., and Bezprozvanny I., Association of the type 1 inositol (1,4,5)-trisphosphate receptor with 4.1N protein in neurons. *Molecular and Cellular Neuroscience* **22** (2003), pp. 271-283.

13. Binda A. V., Kabbani N., Lin R. W., and Levenson R., D2 and D3 dopamine receptor cell surface localization mediated by interaction with protein 4.1N. *Molecular Pharmacology* **62** (2002), pp. 507-513.

14. Denisenko-Nehrbass N., Oguievetskaia K., Goutebroze L., Galvez T., Yamakawa H., Ohara O., Carnaud M., and Girault J. A., Protein 4.1B associates with both Caspr/paranodin and Caspr2 at paranodes and juxtaparanodes of myelinated fibres. *European Journal of Neuroscience* **17** (2003), pp. 411-416.

15. Krauss S. W., Chasis J. A., Rogers C., Mohandas N., Krockmalnic G., and Penman S., Structural protein 4.1 is located in mammalian centrosomes. *Proc. Natl. Acad. Sci. USA* **94** (1997), pp. 7297-7302.

16. Mattagajasingh S. N., Huang S. C., Hartenstein J. S., Snyder M., Marchesi V. T., and Benz E. J., A nonerythroid isoform of protein 4.1R interacts with the nuclear mitotic apparatus (NuMA) protein. *J Cell Biol* **145** (1999), pp. 29-43.

17. Delhommeau F., Vasseur-Godbillon C., Leclerc P., Schischmanoff P. O., Croisille L., Rince P., Moriniere M., Benz E. J., Jr., Tchernia G., Tamagnini G., Ribeiro L., Delaunay J., and Baklouti F., A splicing alteration of 4.1R pre-mRNA generates 2 protein isoforms with distinct assembly to spindle poles in mitotic cells. *Blood* **100** (2002), pp. 2629-36.

18. De Carcer G., Lallena M. J., and Correas I., Protein 4.1 is a component of the nuclear matrix of mammalian cells. *Biochem J* **312** (1995), pp. 871-7.

19. Krauss S. W., Larabell C. A., Lockett S., Gascard P., Penman S., Mohandas N., and Chasis J. A., Structural protein 4.1 in the nucleus of human cells: dynamic rearrangements during cell division. *J. Cell Biol.* **137** (1997), pp. 275-289.

20. Tchernia G., Mohandas N., and Shohet S. B., Deficiency of skeletal membrane protein band 4.1 in homozygous hereditary elliptocytosis. Implications for erythrocyte membrane stability. *J Clin Invest* **68** (1981), pp. 454-60.

21. Shi Z. T., Afzal V., Coller B., Patel D., Chasis J. A., Parra M., Lee G., Paszty C., Stevens M., Walensky L., Peters L. L., Mohandas N., Rubin E., and Conboy J. G., Protein 4.1R-deficient mice are viable but have erythroid membrane skeleton abnormalities. *J Clin Invest* **103** (1999), pp. 331-40.

22. Walensky L. D., Shi Z. T., Blackshaw S., DeVries A. C., Demas G. E., Gascard P., Nelson R. J., Conboy J. G., Rubin E. M., Snyder S. H., and Mohandas N., Neurobehavioral deficits in mice lacking the erythrocyte membrane cytoskeletal protein 4.1. *Curr Biol* **8** (1998), pp. 1269-72.

23. Tran Y. K., Bogler O., Gorse K. M., Wieland I., Green M. R., and Newsham I. F., A novel member of the NF2/ERM/4.1 superfamily with growth suppressing properties in lung cancer. *Cancer Res* **59** (1999), pp. 35-43.

24. Gutmann D. H., Donahoe J., Perry A., Lemke N., Gorse K., Kittiniyom K., Rempel S. A., Gutierrez J. A., and Newsham I. F., Loss of DAL-1, a protein 4.1-related tumor suppressor, is an important early event in the pathogenesis of meningiomas. *Hum Mol Genet* **9** (2000), pp. 1495-500.

25. Sun C. X., Robb V. A., and Gutmann D. H., Protein 4.1 tumor suppressors: getting a FERM grip on growth regulation. *J Cell Sci* **115** (2002), pp. 3991-4000.

26. Robb V. A., Li W., Gascard P., Perry A., Mohandas N., and Gutmann D. H., Identification of a third Protein 4.1 tumor suppressor, Protein 4.1R, in meningioma pathogenesis. *Neurobiol Dis* **13** (2003), pp. 191-202.

27. Conboy J., The role of alternative pre-mRNA splicing in regulating the structure and function of skeletal protein 4.1. *Proc Soc Exp Biol Med* **220** (1999), pp. 73-8.

28. Discher D., Parra M., Conboy J. G., and Mohandas N., Mechanochemistry of the alternatively spliced spectrin-actin binding domain in membrane skeletal protein 4.1. *J. Biol. Chem.* **268** (1993), pp. 7186-7195.

29. Horne W. C., Huang S. C., Becker P. S., Tang T. K., and Benz E. J., 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* **82** (1993), pp. 2558-63.

30. Deguillien M., Huang S. C., Moriniere M., Dreumont N., Benz E. J., Jr., and Baklouti F., Multiple cis elements regulate an alternative splicing event at 4.1R pre-mRNA during erythroid differentiation. *Blood* **98** (2001), pp. 3809-16.

31. Hou V. C., Lersch R., Gee S. L., Ponthier J. L., Lo A. J., Wu M., Turck C. W., Koury M., Krainer A. R., Mayeda A., and Conboy J. G., Decrease in hnRNP A/B expression during erythropoiesis mediates a pre-mRNA splicing switch. *EMBO J* **21** (2002), pp. 6195-204.

32. Huang J. P., Tang C. J., Kou G. H., Marchesi V. T., Benz E. J., Jr., and Tang T. K., Genomic structure of the locus encoding protein 4.1. Structural basis for complex combinational patterns of tissue-specific alternative RNA splicing. *J Biol Chem* **268** (1993), pp. 3758-66.

33. Baklouti F., Huang S. C., Vulliamy T. J., Delaunay J., and Benz E. J., Jr., Organization of the human protein 4.1 genomic locus: new insights into the tissue-specific alternative splicing of the pre-mRNA. *Genomics* **39** (1997), pp. 289-302.

34. Schischmanoff P. O., Yaswen P., Parra M. K., Lee G., Chasis J. A., Mohandas N., and Conboy J. G., Cell shape-dependent regulation of protein 4.1 alternative pre-mRNA splicing in mammary epithelial cells. *J Biol Chem* **272** (1997), pp. 10254-9.

35. Ramez M., Blot-Chabaud M., Cluzeaud F., Chanan S., Patterson M., Walensky L. D., Marfatia S., Baines A. J., Chasis J. A., Conboy J. G., Mohandas N., and Gascard P., Distinct distribution of specific members of protein 4.1 gene family in the mouse nephron. *Kidney International* **63** (2003), pp. 1321-1337.

36. Parra M. K., Gee S. L., Koury M. J., Mohandas N., and Conboy J. G., Alternative 5' exons and differential splicing regulate expression of protein 4.1R isoforms with distinct N-termini. *Blood* **101** (2003), pp. 4164-71.

37. Yamakawa H., Ohara R., Nakajima D., Nakayama M., and Ohara O., Molecular characterization of a new member of the protein 4.1 family (brain 4.1) in rat brain. *Brain Res Mol Brain Res* **70** (1999), pp. 197-209.

38. Conboy J. G., Chan J. Y., Chasis J. A., Kan Y. W., and Mohandas N., Tissue- and development-specific alternative RNA splicing regulates expression of multiple isoforms of erythroid membrane protein 4.1. *J Biol Chem* **266** (1991), pp. 8273-80.

39. Gascard P., Lee G., Coulombel L., Auffray I., Lum M., Parra M., Conboy J. G., Mohandas N., and Chasis J. A., Characterization of multiple isoforms of protein 4.1R expressed during erythroid terminal differentiation. *Blood* **92** (1998), pp. 4404-14.

40. Gascard P., Nunomura W., Lee G., Walensky L. D., Krauss S. W., Takakuwa Y., Chasis J. A., Mohandas N., and Conboy J. G., Deciphering the nuclear import pathway for the cytoskeletal red cell protein 4.1R. *Mol Biol Cell* **10** (1999), pp. 1783-98.

41. Luque C. M., Lallena M. J., Perez-Ferreiro C. M., de Isidro Y., De Carcer G., Alonso M. A., and Correas I., The N-terminal 209-aa domain of high molecular-weight 4.1R isoforms abrogates 4.1R targeting to the nucleus. *Proc Natl Acad Sci U S A* **96** (1999), pp. 14925-30.

42. Cramer P., Caceres J. F., Cazalla D., Kadener S., Muro A. F., Baralle F. E., and Kornblihtt A. R., Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol Cell* **4** (1999), pp. 251-8.

43. Auboeuf D., Honig A., Berget S. M., and O'Malley B. W., Coordinate regulation of transcription and splicing by steroid receptor coregulators. *Science* **298** (2002), pp. 416-9.

44. Nogues G., Kadener S., Cramer P., Bentley D., and Kornblihtt A. R., Transcriptional activators differ in their abilities to control alternative splicing. *J Biol Chem* **277** (2002), pp. 43110-4.

45. Pagani F., Stuani C., Zuccato E., Kornblihtt A. R., and Baralle F. E., Promoter architecture modulates CFTR exon 9 skipping. *J Biol Chem* **278** (2003), pp. 1511-7.

46. Gimm J. A., An X., Nunomura W., and Mohandas N., Functional characterization of spectrin-actin-binding domains in 4.1 family of proteins. *Biochemistry* **41** (2002), pp. 7275-82.

47. Fehon R., Dawson I. A., and Artavanis-Tsakonas S., A Drosophila homologue of membrane-skeleton protein 4.1 is associated with septate junctions and is encoded by the coracle gene. *Devel.* **120** (1994), pp. 545-557.

48. Mayor C., Brudno M., Schwartz J. R., Poliakov A., Rubin E. M., Frazer K. A., Pachter L. S., and Dubchak I., VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16** (2000), pp. 1046-7.

49. Bray N., Dubchak I., and Pachter L., AVID: A global alignment program. *Genome Res* **13** (2003), pp. 97-102.

Figure legends

Figure 1. Differential domain conservation in the EPB41 protein family. A. Conservation between orthologous 4.1 proteins in human versus mouse. Top panel shows the domain structure of the 4.1 proteins including the conserved FERM, SAB, and CTD domains (shaded) interspersed with unique regions U1-U3. Shown below are the comparisons between each 4.1 domain in the human proteins versus their ortholog in the mouse. Numbers indicate the percent amino acid identity. **B.** Conservation between paralogous 4.1 proteins in human. FERM alignments measure homology in the region encoded by the 3' portion of exon 4 through the end of exon 12; SABD alignments measure homology in the 3' coding region of exon 16 plus exon 17 (EPB41, EPB41L2, EPB41L3) or exon 17 only (EPB41L1, which lacks a close homolog of exon 16); CTD alignments span the region encoded by exons 18-21 (EPB41, EPB41L2, EPB41L3) or exons 19-21 (EPB41L1, which lacks a close homolog of exon 18). Numbers indicate the percent amino acid identity.

Figure 2. Comparative structure of the human *EPB41*-related genes. Diagram (not to scale) is aligned so as to emphasize the organization features of the gene family. Arrows indicate predicted transcription start sites at alternative first exons. For ease of comparison, orthologous conserved exons in the FERM, SABD, and CTD are given the same exon numbers originally assigned in the *EPB41* gene. Approximate length of each 4.1 gene locus (140-240kb) is indicated at the right. Sizes of exons and introns are presented in Tables I and II.

Figure 3. Alternative splicing in the conserved domains of *EPB41*-related transcripts. A. Alternative splicing in the SAB domain. Shown is the exon 16-17 region of *EPB41*, *EPB41L2*, and *EPB41L3*, and the major regulated splicing event in this portion of the gene. Exon 16 is strictly regulated in the *EPB41* and *EPB41L3* genes, being selectively included in a few cell types but excluded in most others. Exon 16 and 17 are co-regulated in *EPB41L2*, being included or excluded together in a non-tissue specific manner. **B**. Alternative splicing in the CTD domain. Shown is the only regulated splicing event in the CTD, in which *EPB41L3* exon 21 is differentially excluded in kidney epithelial cells but included in brain and heart [2].

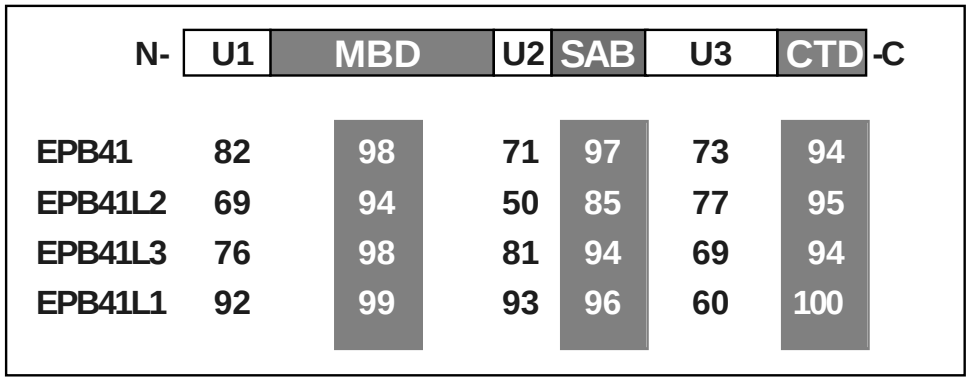
Figure 4. Alternative splicing in unique domains of the *EPB41*-related transcripts. **A**. Novel alternative splicing event in the U2 domain. Top panel shows a tissue-specific alternative splicing event involving *EPB41L1* exon 15. Bottom panel shows a polyacrylamide gel of the major amplification products of the *EPB41L1* U2 region in RNA from several human tissues. Lanes: 1, brain; 2, heart; 3, kidney; 4, liver; 5, lung; 6, ovary; 7, pancreas. **B**. Alternative splicing in the U3 domain. Top panel shows U3 region coding exons

between the SAB and CTD domains in all four *EPB41*-related genes. Boxes filled in the checkerboard pattern represent alternatively spliced exons; those known to be specifically regulated are indicated with asterisks (*EPB41* exon 17A and 17B, *EPB41L1* exon 17D, *EPB41L2* exon 17B, *EPB41L3* exon 17B). Bottom panel shows polyacrylamide gel analysis of an RT-PCR experiment to examine alternative splicing in the U3 region. Location of primers is indicated by arrows. Lanes 1, negative control; lanes 2, brain; lanes 3, skeletal muscle; lanes 4, heart; lanes 5, kidney: lanes 6, liver: lanes 7, lung; lanes 8, testis.

Figure 5. Comparative genomics of human and mouse *EPB41*-related genes. The sequences of *EPB41*, *EPB41L2*, *EPB41L3* and *EPB41L1* genes of human and mouse were compared using VISTA [51]. Peaks corresponding to exons are indicated above the plot and labeled with the exon numbers. The X-axis of the generated plot represents the base sequence and the Y-axis represents the percent identity. Asterisk indicates a candidate new exon 13A in the *EPB41L2* gene.

Figure 1

A.



B.

	MBD				SABD				CTD			
	EPB 41	EPB 41L2	EPB 41L3	EPB 41L1	EPB 41	EPB 41L2	EPB 41L3	EPB 41L1	EPB 41	EPB 41L2	EPB 41L3	EPB 41L1
EPB41	100	75	75	72	100	66	53	32	100	73	74	64
EPB41L2		100	79	78		100	59	37		100	72	68
EPB41L3			100	81			100	28			100	71
EPB41L1				100				100				100

Figure 2

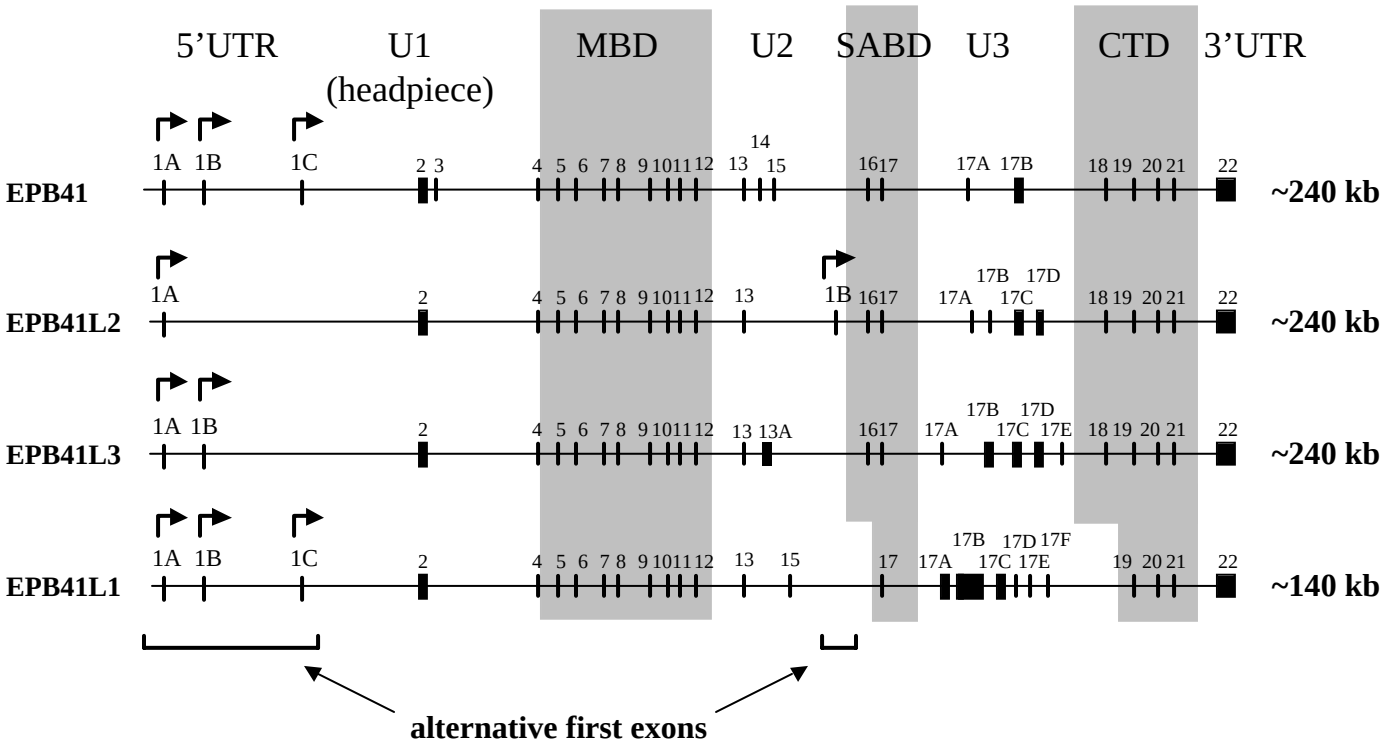

5'UTR
U1
(headpiece)
MBD
U2
SABD
U3
CTD
3'UTR
EPB41
EPB41L2
EPB41L3
EPB41L1
~240 kb
~240 kb
~240 kb
~140 kb
alternative first exons

Figure 3

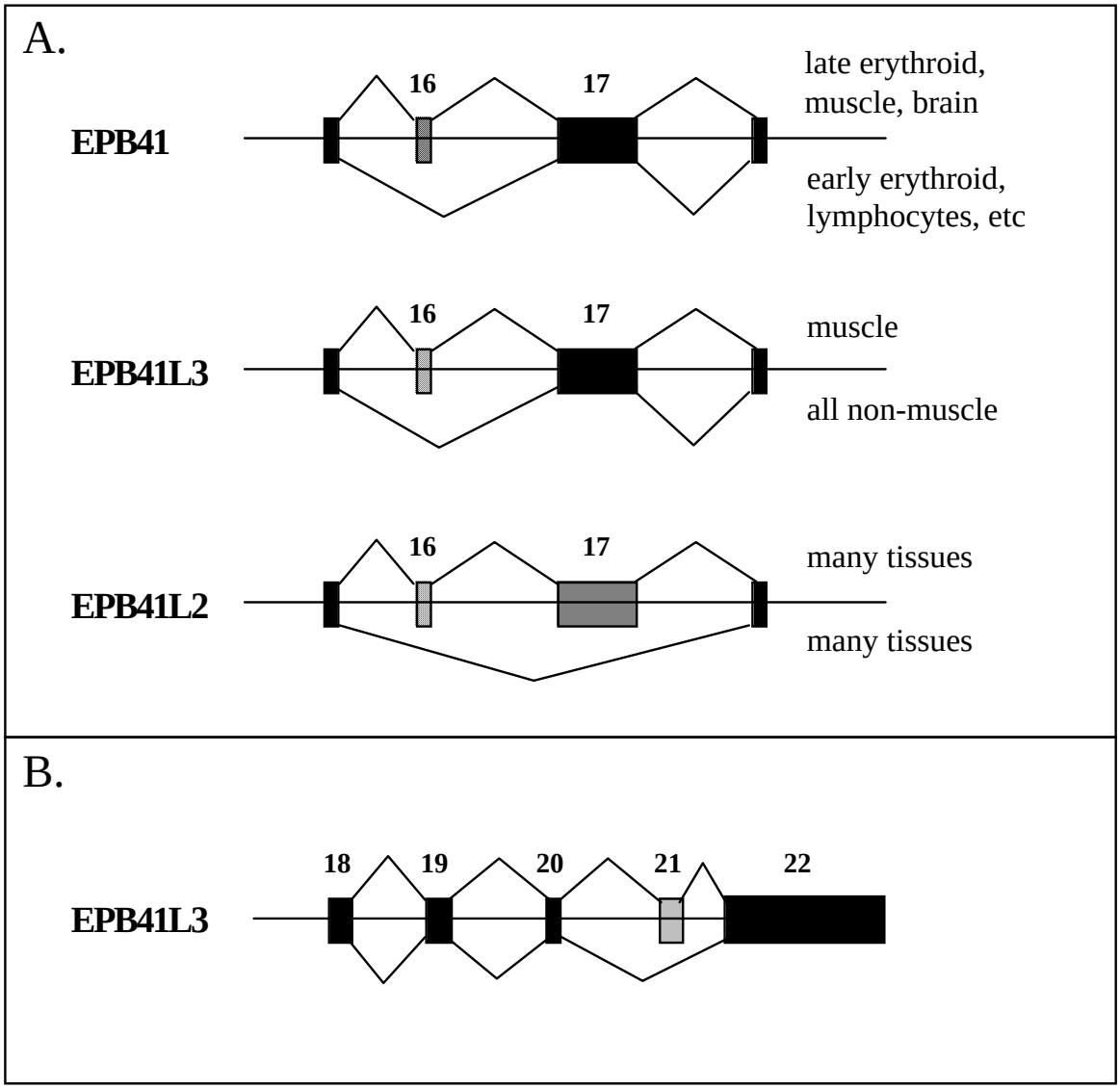
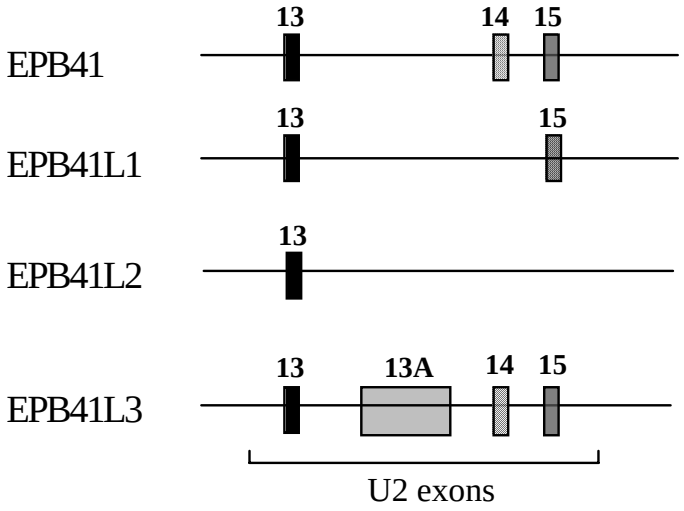

A.
EPB41
16
17
late erythroid, muscle, brain
early erythroid, lymphocytes, etc
EPB41L3
16
17
muscle
all non-muscle
EPB41L2
16
17
many tissues
many tissues
B.
EPB41L3
18
19
20
21
22

Figure 4A


EPB41
13
14
15
EPB41L1
13
15
EPB41L2
13
EPB41L3
13
13A
14
15
U2 exons

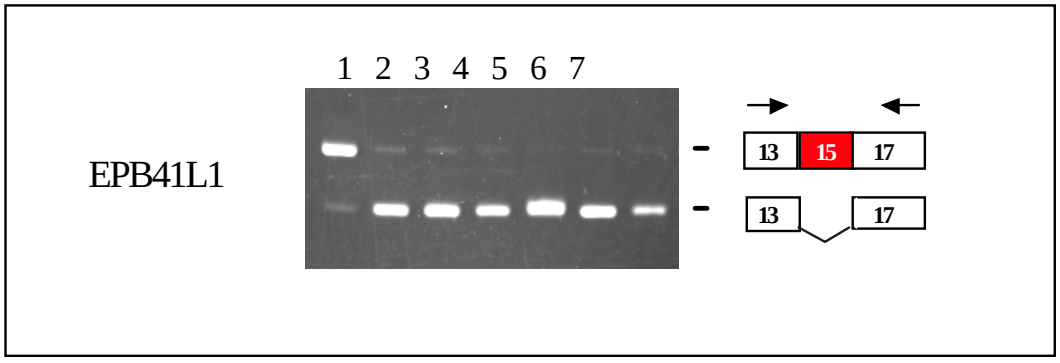

EPB41L1
1 2 3 4 5 6 7
13
15
17
13
17

Figure 4B

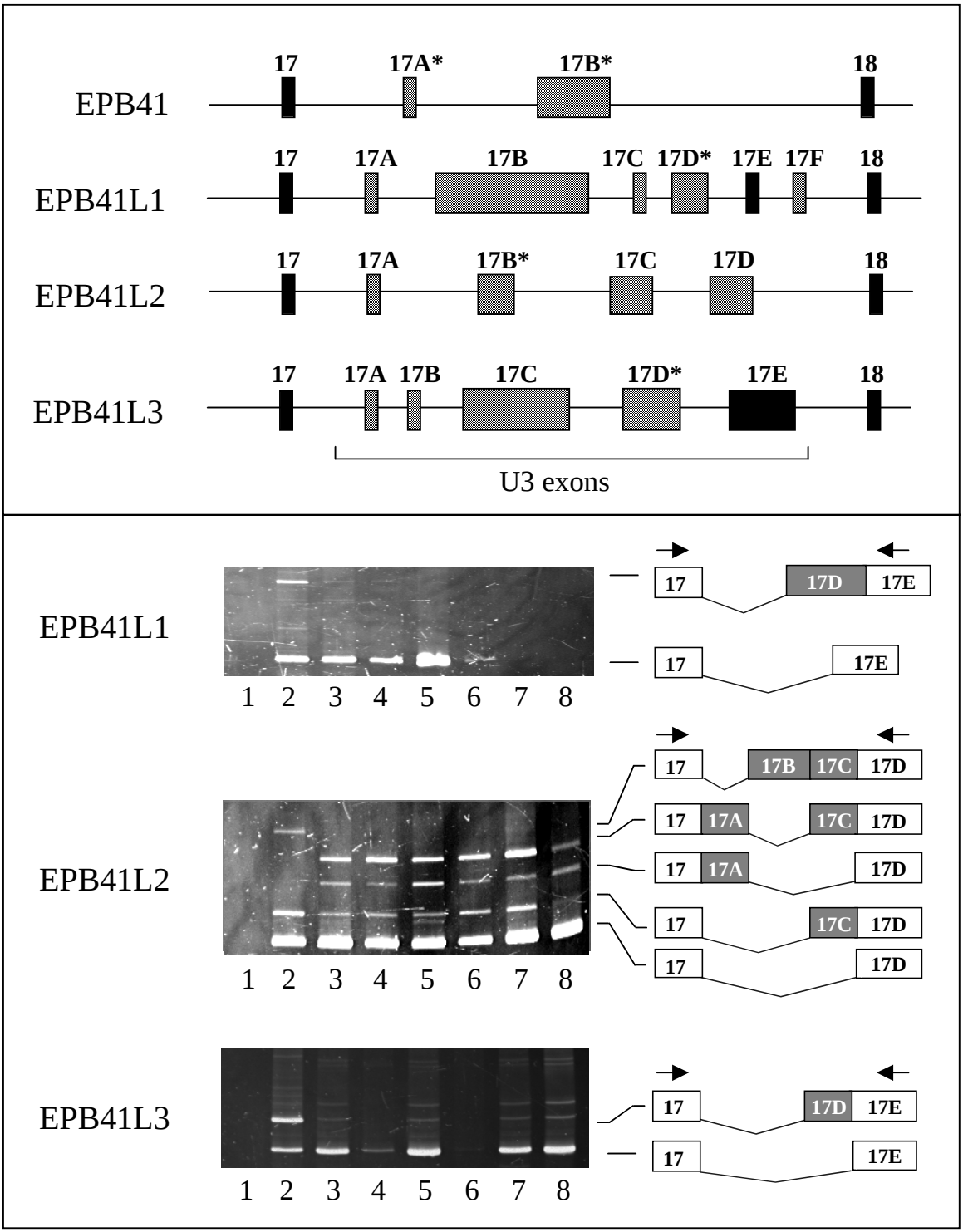

EPB41
17
17A*
17B*
18
EPB41L1
17
17A
17B
17C
17D*
17E
17F
18
EPB41L2
17
17A
17B*
17C
17D
18
EPB41L3
17
17A
17B
17C
17D*
17E
18
U3 exons
EPB41L1
1 2 3 4 5 6 7 8
17
17D
17E
17
17E
EPB41L2
1 2 3 4 5 6 7 8
17
17B
17C
17D
17
17A
17C
17D
17
17A
17D
17
17C
17D
17
17D
EPB41L3
1 2 3 4 5 6 7 8
17
17D
17E
17
17E

Figure 5

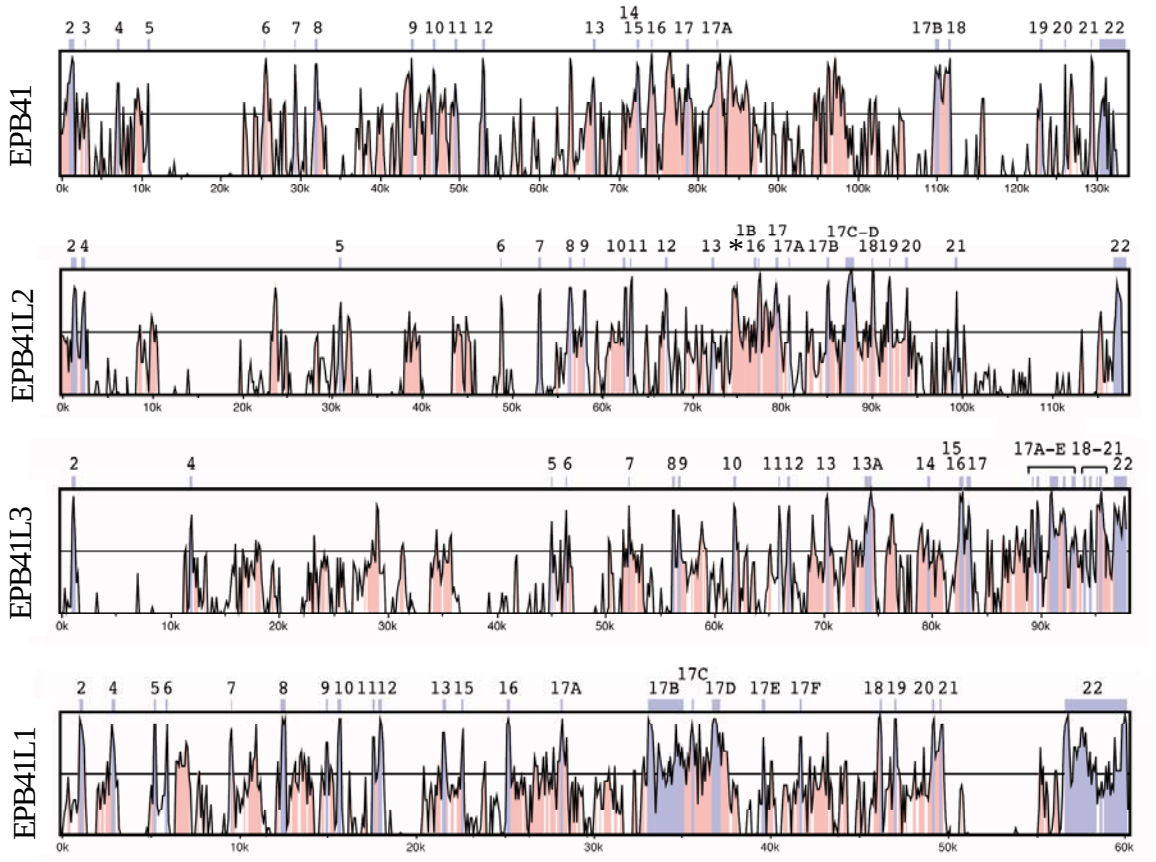

EPB41
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 17A 17B 18 19 20 21 22
0k 10k 20k 30k 40k 50k 60k 70k 80k 90k 100k 110k 120k 130k
EPB41L2
2 4 5 6 7 8 9 10 11 12 13 1B 17 * 16 17A 17B 17C-D 18 19 20 21 22
0k 10k 20k 30k 40k 50k 60k 70k 80k 90k 100k 110k
EPB41L3
2 4 5 6 7 8 9 10 11 12 13 13A 14 15 16 17 17A-E 18-21 22
0k 10k 20k 30k 40k 50k 60k 70k 80k 90k
EPB41L1
2 4 5 6 7 8 9 10 11 12 13 15 16 17A 17B 17C 17D 17E 17F 18 19 20 21 22
0k 10k 20k 30k 40k 50k 60k

TABLE I. Exon sizes for EPB41 genes in human and mouse

Region	Exon	EPB41	EPB41L2	EPB41L3	EPB41L1
U1	2	475 (481)	506 (485)	194 (218)	191
MBD	4	213 (210)	213	198	165
	5	105	105	105	105
	6	43	43	43	43
	7	76	76	76	76
	8	219	219	219	219
	9	88	88	88	88
	10	153	153	153	153
	11	98	98	98	98
	12	173	173	176	176
U2	13	209 (188)	173	167 (221)	149
	13A			561 (528)	
	14	57		54	
	15	42		36	36
SAB	16	63	63	66	-
	17	177	147	192	183
U3	17A	51	102 (101)*	75	105 (99)
	17B	450	168	99	1932**(1923)
	17C		54	735 (720)	108 (111)
	17D		564	123	411 (408)
	17E			369 (378)	105
	17F				84 (81)
CTD	18	129	123	132	87
	19	102	99	99	84
	20	81	81	81	81
	21	105	113 (109)	117	117
	22	2642	1228	1068	3441

Exon sizes are given in nucleotides. Numbers in parentheses represent mouse exons, given only where they differ in size from the human (only coding exons were examined for mouse).

* EPB41L2 exon 17A may be pseudoexon in the mouse since its expression would cause a frameshift.

** predicted by homology to rat cDNA seq

TABLE II. Intron sizes in human EPB41 genes

intron	hEPB41	hEPB41L2	hEPB41L3	hEPB41L1
1A	6.5 kb	106.6 kb	87.0 kb	61.7 kb
1B	20.7 kb		54.1 kb	18.9 kb
1C	72.7 kb			
2	5424	676	10.6 kb	1596
4	3672	28.4 kb	33.0 kb	2236
5	14.5 kb	17.7 kb	1259	573
6	3784	4280	5727	3599
7	2456	3284	3913	2769
8	12.0 kb	1383	346	2340
9	2605	4371	5003	583
10	2580	525	3953	1777
11	3330	3877	798	247
12	13.7 kb	5025	3500	3414
13	5276	4889	3332	968
13A			5198	
14	165		2829	
15	1569		198	2494
16	4497	1893	468	
17	3577	1328	5733	2831
17A	27.4 kb	4135	373	4871
17B	1202	1989	1110	520
17C		201	507	1079
17D		1981	594	2373
17E			729	1980
17F				4435
18	11.4 kb	1824	503	798
19	2930	1817	461	2019
20	3250	5394	273	333
21	1015	17.5kb	1198	6937

Intron length is given in nucleotides (for smaller introns) or kb (for introns >10 kb).