Tissue Antigens ISSN 0001-2815

BRIEF COMMUNICATION

Nomenclature of prominin-1 (CD133) splice variants – an update

C. A. Fargeas[1], W. B. Huttner[2] & D. Corbeil[1]

1 Tissue Engineering Laboratories, BIOTEC, Technical University of Dresden, Dresden, Germany
2 Max-Planck-Institute of Molecular Cell Biology and Genetics, Dresden, Germany

Key words
AC133 antigen; CD133; prominin-1; splice variant; stem cell marker

Correspondence
Denis Corbeil
Tissue Engineering Laboratories, BIOTEC
Technical University of Dresden
Tatzberg 47-51
D-01307 Dresden
Germany
Tel: +49 0351 4634 0118
Fax: +49 0351 4634 0244
e-mail: corbeil@biotec.tu-dresden.de

Received 14 January 2007; revised 30 January 2007; accepted 1 February 2007

doi: 10.1111/j.1399-0039.2007.00825.x

Abstract

Prominin-1 (CD133), a pentaspan membrane glycoprotein that constitutes an important cell surface marker of various, either normal or cancerous, stem cell populations is widely used to isolate or characterize such cells in different systems. Occurring throughout the metazoan evolution with a remarkably conserved genomic organization, it may be expressed as different splice variants with distinctive characteristics. A rational nomenclature has been proposed earlier for their consistent designation across species. Although generally accepted, it seems to be misunderstood in view of the recent report of novel prominin-1 complementary DNAs in rhesus monkey and humans with improper naming. As this may lead to confusion, we have reexamined the genomic organization of prominin-1 in various primates to provide an update that should further clarify the rationale of the nomenclature for prominin-1 gene products. This report comprises (i) the determination of the genomic organization of *prominin-1* gene in two non-human primates, i.e. *Macaca mulatta* and *Pan troglodytes*, commonly used in research, (ii) the mapping of a new exon that creates an alternative cytoplasmic C-terminal end of prominin-1, (iii) the identification of various potential PDZ-binding domains generated by alternative cytoplasmic C-terminal tails, suggesting that different prominin-1 splice variants might interact with distinct protein partners, and (iv) a summing up of the different prominin-1 splice variants.

In the article entitled 'Isolation, molecular cloning and *in vitro* expression of rhesus monkey (*Macaca mulatta*) prominin-1.s1 complementary DNA (cDNA) encoding a potential hematopoietic stem cell antigen', recently published in *Tissue Antigens* by Husain et al. (1), the authors refer to a previous article from our group proposing a unifying nomenclature for the designation of the prominin family gene products (2) based on their conserved genomic organization across species (3, 4). The aim of this nomenclature was to attribute the same designation to a given splice variant irrespective of the species as splice variants described in one organism could be predicted to exist in others, with similar characteristics at least among mammals. This nomenclature seems to be generally accepted in the field (5–9), but the rationale underlying it may remain unclear as the rhesus prominin-1 splice variant described by Husain et al. does not correspond to the s1 splice variant (1), and several unpublished human prominin-1 splice variants appear in the NCBI GenBank database (accession numbers AY449690 to AY449693) with the incorrect suffix according to this same nomenclature. Given the wide use of prominin-1 (also termed CD133) for the characterization of stem and progenitor cell populations in different normal tissues (10–13) as well as in cancers [(14–16); for review, see (17)], it is important to clarify the rationale of this nomenclature. We believe that it is essential to maintain a consistent designation with regard to potential cross-specificity toward particular epitopes, e.g. stem-cell-associated AC133 epitope (18). Moreover, when dissecting the differential tissue expression and function of these prominin-1 splice variants, it is important to refer to homologous molecules in the different model organisms.

Anarchic reference to the nomenclature can only bring in confusion as to the interpretation of data generated in different models. The chronology of the identification of prominin-1 splice variants in one species or the other is not necessarily related to the predominance of one variant in one species compared with the others and therefore cannot serve as a base for their designation. For instance, prominin-1.s1 was first described in mouse then in humans, while prominin-1.s2 was first described in human then in mouse, or in the same line, only prominin-1.s1 and s7 have been described in rat (19). Cross-species sequence comparison should be made between homologous variants in order to be significant. We therefore propose to sum up and update the nomenclature of the different prominin-1 splice variants and the underlying genomic features in humans and different animal models.

The *prominin-1* gene is located on chromosome 4 in humans and chromosome 5 in mice and spans more than 150 kb (3, 20, 21). Its genomic structure, i.e. exon/intron boundaries, is strikingly similar across species (3). Since the original description of prominin-1 (22), several splice variants affecting the open reading frame have been identified in mouse (2, 23, 24) and in humans (25, 26) and their expression was characterized (4). Factors regulating the *in vivo* expression of prominin-1 remain to be determined, but the messenger RNA profile suggests that the

Table 1 Genomic structure of primate *prominin-1* genes[a]

Exon	Human cDNA (AF027208), 3794 nt	Human cDNA (AY449689), 2493 nt	Human chromosome 4 (NC_000004)	Rhesus cDNA (AY903606), 3927 nt	Rhesus chromosome 5 (NC_007862)	Chimpanzee predicted (XM_517115)	Chimpanzee chromosome 4 (NC_006471)
1	1–257	1–220	15686664–15686408	5–440	10892907–10892472	308–739	16021345–16020914
2	258–313	221–276	15649722–15649667	441–496	10853557–10853502	740–795	15984217–15984162
3	314–340	—	15646482–15646456	497–523	10850408–10850382	796–822	15980970–15980944
4	341–546	277–482	15644230–15644025	524–729	10848142–10847937	823–1028	15978717–15978512
5	547–667	483–603	15636033–15635913	730–850	10839541–10839421	1029–1149	15970521–15970401
6	668–731	604–667	15635079–15635016	851–914	10838634–10838571	1150–1213	15969576–15969513
7	732–821	668–757	15634136–15634047	915–1004	10837679–10837590	1214–1303	15968613–15968524
8	822–1039	758–975	15629261–15629044	1005–1222	10832868–10832651	1304–1521	15962399–15962182
9	1040–1114	976–1050	15626961–15626887	1223–1297	10830557–10830483	1522–1596	15960098–15960024
10	1115–1178	1051–1114	15624059–15623996	1298–1361	10827613–10827550	1597–1660	15957207–15957144
11	1179–1338	1115–1274	15619829–15619670	1362–1521	10823685–10823526	1661–1820	15952979–15952820
12	1339–1491	1275–1427	15617411–15617259	1522–1674	10820925–10820773	1821–1973	15950562–15950410
13	1492–1615	1428–1551	15611340–15611217	1675–1798	10814875–10814752	1974–2097	15944639–15944516
14	1616–1719	1552–1655	15609209–15609106	1799–1902	10813072–10812969	2098–2201	15942516–15942413
15	1720–1804	1656–1740	15604792–15604708	1903–1987	10808780–10808696	2202–2286	15938123–15938039
16	1805–1948	1741–1884	15603112–15602969	1988–2131	10807105–10806962	2287–2430	15936442–15936299
17	1949–2020	1885–1956	15602014–15601943	2132–2203	10806037–10805966	2431–2502	15935353–15935282
18	2021–2113	1957–2049	15600545–15600453	2204–2296	10804562–10804470	2503–2595	15933920–15933828
19[b]			ND		ND		ND
20	2114–2167	2050–2103	15598437–15598384	2297–2350	10802763–10802710	2596–2649	15931721–15931668
21	2168–2248	2104–2184	15596758–15596678	2351–2431	10801104–10801024	2650–2730	15930033–15929953
22	2249–2317	2185–2253	15596549–15596481	2432–2500	10800895–10800827	2731–2799	15929824–15929756
23	2318–2410	2254–2346	15595076–15594984	2501–2593	10799403–10799311	2800–2892	15928351–15928259
24	2411–2526	2347–2462	15591258–15591143	2594–2709	10795923–10795808	2893–3008	15924544–15924429
25	—	2463–2487	15591022–15590998	—	10795690–10795666	—	15924302–15924278
26a[c]	—	—	15590886–15590846	—	10795553–10795513	—	15924165–15924125
26b[c]	2527–2550	2488–2493	15590625–15590602	2710–2733	10795292–10795269	3009–3032	15923905–15923882
27	2551–2619		15590184–15590116	—	10794849–10794781	3033–3101	15923464–15923396
28	2620–2659		15581803–15581764	2734–2770	10786491–10786455	3102–3141	15917934–15917895
29	2660–3794		15580089–15578955	2771–3904	10784777–10783647	3142–4284	15916215–15915073

cDNA, complementary DNA; ND, not determined; Nt, nucleotide; —, absence of the exon in the cDNA under consideration.

[a] The numbering of the exons begins with the exon bearing the initial start codon. The 5′ untranslated regions (UTRs) were ignored for the sake of simplicity in comparing the different gene sequences. Exon 29 contains the 3′ UTR with the polyadenylation sequence. The exon–intron structure of primate *prominin-1* genes were completed by comparing the sequences of the humans, rhesus, and chimpanzee genomic DNA (accession numbers: NC_000004, NC_007862 and NC_006471, respectively) with that of human prominin-1.s2 and −s9 cDNAs (AF027208 and AY449689, respectively) and that of mouse prominin-1.s3 and −s8 cDNAs (AF305215 and BC028286, respectively). Nucleotide and protein databases were searched at the National Center for Biotechnology Information (NCBI) using the BLAST network services.

[b] Exon 19 has been only identified in murine prominin-1.s8 so far (see Table 2).

[c] Exons 26a and 26b are created by alternative splice acceptor sites, and they are initially referred as exons 25a and b in a previous publication (4).

mouse prominin-1 splice variants are tissue specific and developmentally regulated (4).

To date, alternative splicing was found to affect the N-terminal domain, the first and second extracellular loops or mostly the cytoplasmic C-terminal domain, which might implicate distinct cytoplasmic protein-interacting partners. The splice variants s1 and s3–s8 were first isolated in mouse and defined four alternative C-termini for prominin-1, which occur through (i) intron retention, (ii) exon skipping, or (iii) usage of a cryptic acceptor site (4). This splicing cassette, which locates within a cluster of short exons and introns, is conserved across species allowing the prediction of similar splice variants (4). Additional splice variants that define two more alternative C-termini have been isolated in humans (AY449689, AY449691, and AY449693). By comparing the prominin-1 cDNA sequence (nucleotides 2463–2487) isolated from KG-1a cell line (see AY449689) with the human prominin-1 genomic sequence (NC_000004), we have identified a novel facultative exon (SSWVTSVQ), i.e. exon 25, within this splicing cassette (positions 15591022 to 15590998 of the genomic clone) that conforms to the consensus GT-AG rule and encodes residue 822–829 of this particular prominin-1 variant (from here on referred to as s9) (Tables 1 and 2). Importantly, exon 25 being 25 nucleotides long, its insertion induces a frameshift on the following exon 26b, and hence, a premature ending following a C-terminal cysteine (residue 830) compared with the s1 and s2 splice variants (Table 3).

Thus, the prominin-1 coding region spans 28 exons among which 7 are facultative with exon 26 bearing two mutually exclusive acceptor sites a and b (Table 2). Remarkably, we have identified a nearly identical genomic structure for two others primates, i.e. rhesus and chimpanzee (on chromosomes 5 and 4, respectively) by similar sequence comparison (Tables 1 and 3), lending further weight to the rationale underlying the nomenclature for prominin-1 gene products. Hence, prominin-1 from chimpanzee is predicted to display the characteristic five transmembrane domains (see updated sequence XM_517115) like in the other species, and the s1 variant would be predicted to contain 865 residues, contrarily to what appears in table 1 of Husain et al. (1). Together, 12 different prominin-1 splice variants affecting the protein sequence have been described to date in rodents and primates (Table 2). Some of them, e.g. s1, s2, s7, s12, have been detected in several species. Completing the chart for a given species might thus only be a question of time.

It appears therefore that a prominin-1 splice variant with a cytoplasmic C-terminal tail that is 24 amino acid shorter than the human prominin-1.s2 (AF027208) like the rhesus prominin-1.s1 splice variant described by Husain and colleagues does exist in humans. Consequently, this rhesus prominin-1 splice variant aligns better with the 842 amino acid long human splice variant encoded by an unpublished cDNA sequence deposited in GenBank (accession number AY449693; here referred to as s12; Table 2) than with human prominin-1.s2, yielding 95.7% amino acid sequence identity rather than 93.2% (using the stretcher program from the EMBOSS package with a EBLOSUM 62 matrix and gap and extend penalty of 12 and 2, respectively). The prominin-1 sequence from *Macaca mulatta* (AY903606) has therefore been renamed prominin-1.s12 (S. M. Husain, personal communication, St. Jude Children's Research Hospital, Memphis, TN).

The importance of using the same suffix to design a given splice variant irrespective of the species is best illustrated when considering the cytoplasmic C-terminal tail of

Table 2 Presence or absence of facultative exons in various prominin-1 splice variants[a]

	Structure (facultative exons included in the coding sequences)												
Splice variant	3	9	19	25	26a	26b	27	28	*Mus musculus*	*Rattus norvegicus*	*Homo sapiens*	*Macaca mulatta*	*Pan troglodytes*
s1	−	+	−	−	−	+	+	+	AF026269 (858)	AF386758 (857)	AF507034 (856)		
s2	+	+	−	−	−	+	+	+	AF039663 (867)		AF027208 (865)	XM_001100223 (864)	XM_517115 (865)
s3	−	+	−	−	+	−	−	−	AF305215 (834)				
s4	−	−	−	−	+	−	−	−	AY223521 (804)				
s5	−	−	−	−	+	−	−	−	AY223522 (809)				
s6	−	+	−	−	−	−	−	−	AY099088 (823)				
s7	−	+	−	−	−	−	−	+	AK029921 (827)	AY262731 (826)	AY449690 (825)		
s8	+	+	+	−	−	−	−	+	BC028286 (842)				
s9	−	+	−	+	−	+	−	−			AY449689 (830)		
s10	−	+	−	−	−	+	−	+			AY449691 (833)		
s11	+	+	−	−	−	−	−	+			AY449692 (834)		
s12	+	+	−	−	−	+	−	+			AY449693 (842)	AY903606 (841)	

[a] The structure of the different prominin-1 splice variants is indicated by the presence (+) or absence (−) of the facultative exons numbered according to Table 1. Prominin-1 splice variants identified in each species are indicated by their respective GenBank accession number and protein sequence length (number of amino acid residue).

Table 3 Alternative prominin-1 C-termini[a]

Splice variant	Exon	Species	C-terminal sequence
s1, s2	24, 26b, 27, 28	H	RRMDSEDVYDDVETIPMKNMENGNNGYHKDHVYGIHNPVMTSPSQH
		C	*RRMDSEDVYDDVETIPMKNMENGNNGYHKDHVYGIHNPVMTSPSQH*
		R	*RRMDSEDVYDDVETIPMKNMENGNNGYHKDH**L**YGIHNPVMTSPS-H*
s3, s4, s5	24, 26a	H	*RRMDSEDVYDDSSLSGTWHFTL*
		C	*RRMDSEDVYDDSSLSGTWHFTL*
		R	*RRMDSEDVYDDSSLSGTWHFTL*
s6	24	H	*RRMDSEDVYDE*
		C	*RRMDSEDVYDE*
		R	*RRMDSEDVYDE*
s7, s8, s11	24, 28	H	RRMDSEDVYDDPSQH
		C	*RRMDSEDVYDDPSQH*
		R	*RRMDSEDVYDDPS-H*
s9	24, 25, 26b*	H	RRMDSEDVYDDSSWVTSVQC*
		C	*RRMDSEDVYDDSSWVTSVQC*
		R	*RRMDSEDVYDDSSWVISIQC*
s10, s12	24, 26b, 28	H	RRMDSEDVYDDVETIPMKNPSQH
		C	*RRMDSEDVYDDVETIPMKNPSQH*
		R	RRMDSEDVYDDVETIPMKNPS-H

cDNA, complementary DNA; C, chimpanzee; H, human; R, rhesus.

[a] Alternative splicing might generate six distinct cytoplasmic C-terminal tails of prominin-1. Sequences from isolated cDNAs (see Table 2) appear in regular characters, while predicted sequences are given in italics. Exons 25, 26a, 26b, and 27 are given in cyan, red, green, and blue, respectively. Asterisks indicate exon 25 being 25 nucleotides long introduces a frameshift on the following exon 26b. Nonconserved amino acid residues within the rhesus sequence are given in bold underlined.

prominin-1 (Table 3). Although no information is currently available as to potential cytoplasmic proteins interacting with prominin-1, the presence of distinct C-terminal domains suggests that different prominin-1 splice variants might interact with alternative protein partners. Interestingly, we found that the last four C-terminal amino acids of the s3, s4 and s5 variants, i.e. HFTL, exhibit the characteristics of a class II PDZ-binding domain with a hydrophobic residue in position 0 and −2 (X-ϕ-X-ϕ; X, unspecified amino acid and ϕ, hydrophobic residue), while the C-terminus of the s9 splice variant, i.e. SVQC, would be highly related to a class III PDZ-binding domain (X-X-C) and rodent s1, s2, s7, s8 and s11 (PSQR) to a class I PDZ-binding domain (X-S/T-X-ϕ) (27, 28). In keeping with a potential interaction with PDZ-domain-containing protein, we have recently found that prominin-2, the prominin-1 paralogue (3), binds to a novel splice variant of the glutamate receptor-interacting protein, a PDZ-domain-containing protein in a yeast two-hybrid screening (Kathrin Opherk and DC, unpublished data; see GenBank accession number AY255674). Further studies are needed to determine whether prominin-1 molecules interact with various PDZ-domain-containing proteins as well as to study the physiological relevance of such interaction.

In conclusion, we have shown (i) that the genomic organization of *prominin-1* gene in two non-human primates is highly similar to that in the human counterpart and (ii) the existence of a novel conserved facultative exon that creates an alternative cytoplasmic C-terminal end of prominin-1. Therefore, we recommend that the designation of the prominin-1 splice variants from *Homo sapiens* (AY449689 to AY449693) be corrected according to the present Table 2 in order to conform to the nomenclature for prominin-1 gene products they appear to refer to.

Acknowledgments

WBH was supported by the Deutsche Forschungsgemeinschaft (SPP 1109, Hu275/7-3; SPP 1111, Hu275/8-3; SFB/TR13-04 B1; SFB 655 A2) and DC by the Deutsche Forschungsgemeinschaft (SPP 1109, CO 298/2-2; SFB/TR13-04 B1; SFB 655 A13) and Sächsisches Ministerium für Wissenschaft und Kunst–Europäischer Fond für Regionale Entwicklung (4212/05-16).

References

1. Husain SM, Shou Y, Sorrentino BP, Handgretinger R. Isolation, molecular cloning and in vitro expression of rhesus monkey (Macaca mulatta) prominin-1.s1 complementary DNA encoding a potential hematopoietic stem cell antigen. *Tissue Antigens* 2006: **68**: 317–24.
2. Fargeas CA, Corbeil D, Huttner WB. AC133 antigen, CD133, prominin-1, prominin-2, etc.: prominin family gene products in need of a rational nomenclature. *Stem Cells* 2003: **21**: 506–8.
3. Fargeas CA, Florek M, Huttner WB, Corbeil D. Characterization of prominin-2, a new member of the

prominin family of pentaspan membrane glycoproteins. *J Biol Chem* 2003: **278**: 8586–96.
4. Fargeas CA, Joester A, Missol-Kolka E, Hellwig A, Huttner WB, Corbeil D. Identification of novel prominin-1/CD133 splice variants with alternative C-termini and their expression in epididymis and testis. *J Cell Sci* 2004: **117**: 4301–11.
5. Shmelkov SV, St Clair R, Lyden D, Rafii S. AC133/CD133/prominin-1. *Int J Biochem Cell Biol* 2005: **37**: 715–9.
6. Zhang CC, Lodish HF. Murine hematopoietic stem cells change their surface phenotype during ex vivo expansion. *Blood* 2005: **105**: 4314–20.
7. Hess DA, Meyerrose TE, Wirthlin L et al. Functional characterization of highly purified human hematopoietic repopulating cells isolated according to aldehyde dehydrogenase activity. *Blood* 2004: **104**:1648–55.
8. Wu X, Rabkin-Aikawa E, Guleserian KJ et al. Tissue-engineered microvessels on three-dimensional biodegradable scaffolds using human endothelial progenitor cells. *Am J Physiol Heart Circ Physiol* 2004: **287**: H480–7.
9. Yu Y, Flint AF, Mulliken JB, Wu JK, Bischoff J. Endothelial progenitor cells in infantile hemangioma. *Blood* 2004: **103**: 1373–5.
10. Yin AH, Miraglia S, Zanjani ED et al. AC133, a novel marker for human hematopoietic stem and progenitor cells. *Blood* 1997: **90**: 5002–12.
11. Uchida N, Buck DW, He D et al. Direct isolation of human central nervous system stem cells. *Proc Natl Acad Sci U S A* 2000: **97**: 14720–5.
12. Richardson GD, Robson CN, Lang SH, Neal DE, Maitland NJ, Collins AT. CD133, a novel marker for human prostatic epithelial stem cells. *J Cell Sci* 2004: **117**: 3539–45.
13. Lee A, Kessler JD, Read T-A et al. Isolation of neural stem cells from postnatal cerebellum. *Nat Neurosci* 2005: **8**: 723–9.
14. Singh SK, Hawkins C, Clarke ID et al. Identification of human brain tumour initiating cells. *Nature* 2004: **432**: 396–401.
15. O'brien CA, Pollett A, Gallinger S, Dick JE. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* 2007: **445**: 106–10.
16. Ricci-Vitiani L, Lombardi DG, Pilozzi E et al. Identification and expansion of human colon-cancer-initiating cells. *Nature* 2007: **445**: 111–5.
17. Fargeas CA, Fonseca AV, Huttner WB, Corbeil D. Prominin-1 (CD133): from progenitor cells to human diseases. *Future Lipidol* 2006: **1**: 213–25.
18. Florek M, Haase M, Marzesco A-M et al. Prominin-1/CD133, a neural and hematopoietic stem cell marker, is expressed in adult human differentiated cells and certain types of kidney cancer. *Cell Tissue Res* 2005: **319**: 15–26.
19. Corbeil D, Fargeas CA, Huttner WB. Rat prominin, like its mouse and human orthologues, is a pentaspan membrane glycoprotein. *Biochem Biophys Res Commun* 2001: **285**: 939–44.
20. Maw MA, Corbeil D, Koch J et al. A frameshift mutation in prominin (mouse)-like 1 causes human retinal degeneration. *Hum Mol Genet* 2000: **9**: 27–34.
21. Shmelkov SV, Jun L, St Clair R et al. Alternative promoters regulate transcription of the gene that encodes stem cell surface protein AC133. *Blood* 2004: **103**: 2055–61.
22. Weigmann A, Corbeil D, Hellwig A, Huttner WB. Prominin, a novel microvilli-specific polytopic membrane protein of the apical surface of epithelial cells, is targeted to plasmalemmal protrusions of non-epithelial cells. *Proc Natl Acad Sci U S A* 1997: **94**: 12425–30.
23. Miraglia S, Godfrey W, Buck D. A response to AC133 hematopoietic stem cell antigen: human homologue of mouse kidney prominin or distinct member of a novel protein family? *Blood* 1998: **91**: 4390–1.
24. Corbeil D, Röper K, Fargeas CA, Joester A, Huttner WB. Prominin: a story of cholesterol, plasma membrane protrusions and human pathology. *Traffic* 2001: **2**: 82–91.
25. Miraglia S, Godfrey W, Yin AH et al. A novel five-transmembrane hematopoietic stem cell antigen: isolation, characterization, and molecular cloning. *Blood* 1997: **90**: 5013–21.
26. Yu Y, Flint A, Dvorin EL, Bischoff J. AC133-2, a novel isoform of human AC133 stem cell antigen. *J Biol Chem* 2002: **277**: 20711–6.
27. Harris BZ, Lim WA. Mechanism and role of PDZ domains in signaling complex assembly. *J Cell Sci* 2001: **114**: 3219–31.
28. Sheng M, Sala C. PDZ domains and the organization of supramolecular complexes. *Annu Rev Neurosci* 2001: **24**: 1–29.

