Supplemental Data for

SKELETAL MUSCLE-SPECIFIC CALPAIN IS AN INTRACELLULAR Na^+-DEPENDENT PROTEASE

Yasuko Ono[1], Koichi Ojima[1,2], Fukuyo Torii[3], Emi Takaya[1], Naoko Doi[1,2], Kazuhiro Nakagawa[3], Shoji Hata[1], Keiko Abe[3] and Hiroyuki Sorimachi[1,2]

From Calpain Project, The Tokyo Metropolitan Institute of Medical Science (Rinshoken), Tokyo 156-8506, Japan[1], CREST, JST, Saitama 332-0012, Japan[2], and Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan[3]

Table S1. Amino acid sequences of vertebrate p94s used in this study

Data are from NCBI (NP_xx...x and ACY78226) and ENSEMBL (ENSAxx...x and GSTENT10006672001) databases. See also Fig. S1.

Abbreviation	Species	General name	Ac. No.
H. sapiens	*Homo sapiens*	human	NP_000061
M. musculus	*Bos taurus*	bovine	NP_031627
B. taurus	*Mus musculus*	mouse	NP_776685
G. gallus	*Gallus gallus*	chicken	NP_001004405
T. guttata	*Taeniopygia guttata*	zebra finch	XP_002200345
A. carolinensis	*Anolis carolinensis*	green anole lizard	ENSACAP00000015362
X. laevis	*Xenopus laevis*	African clawed frog	NP_001098739[*1]
X. tropicalis	*Xenopus tropicalis*	Western clawed frog	ENSXETP00000026925[*2]
T. nigroviridis	*Tetraodon nigroviridis*	puffer fish	GSTENT10006672001[*3]
T. rubripes	*Takifugu rubripes*	fugu	ENSTRUP00000046251[*4]
H. hippoglossus	*Hippoglossus hippoglossus*	halibut	ACY78226
S. salar	*Salmo salar*	Atlantic salmon	NP_001158880

*1: Contains only domain III and the sequence C-terminal to it.

*2: The sequence was revised using the *Xenopus tropicalis* genomic (ENSXETG00000012304) and EST sequences.

*3: The sequence was revised by us for this study by comparison with the *Takifugu rubripes* genome sequence (ENSTRUG00000018064); no obvious IS1 sequence was found.

*4: The sequence was revised by us for this study by comparison with the *Tetraodon nigroviridis* genome sequence (GSTENG10006672001); no obvious IS1 sequence was found.

Table S2. Identity and similarity among 12 vertebrate p94 sequences

Amino acid sequence identities (upper right) and similarities (lower left) (%) were calculated by GENETYX Software Ver.10 (GENETYX Co.). Since only the C-terminal part of the *X. laevis* sequence was found in the database, the values for *X. laevis* are shown in parentheses. For species names, see Table S1.

% identity / % similarity	*mammals*			*birds*		*reptile*	*amphibians*		*fishes*			
	H. sapiens	*M. musculus*	*B. taurus*	*G. gallus*	*T. guttata*	*A. carolinensis*	*X. laevis*	*X. tropicalis*	*T. nigroviridis*	*T. rubripes*	*H. hippoglossus*	*S. salar*
H. sapiens		93.4	92.3	77.7	76.9	72.0	(76.8)	64.4	61.6	61.7	65.7	64.9
M. musculus	99.5		93.3	77.4	77.1	72.0	(76.5)	65.8	62.5	62.2	66.4	63.6
B. taurus	98.9	99.0		77.2	76.4	72.5	(76.1)	65.3	61.4	59.5	60.8	62.6
G. gallus	95.5	95.2	95.7		92.0	76.4	(74.2)	63.0	65.7	61.6	62.6	61.1
T. guttata	95.4	95.3	95.8	98.8		76.7	(75.2)	64.0	66.2	61.0	66.8	60.6
A. carolinensis	94.2	94.2	94.0	94.3	93.6		(70.3)	61.7	65.0	61.8	62.4	60.4
X. laevis	(94.5)	(95.8)	(95.5)	(97.1)	(97.7)	(93.5)		(86.7)	(64.7)	(65.4)	(66.6)	(66.0)
X. tropicalis	87.9	88.0	87.8	89.5	89.5	87.4	(96.4)		55.6	54.1	55.1	56.1
T. nigroviridis	86.9	87.0	86.9	93.7	93.3	91.8	(92.8)	81.9		91.5	83.7	75.5
T. rubripes	85.9	86.2	85.5	90.5	90.5	89.6	(92.2)	81.2	98.3		85.6	75.2
H. hippoglossus	91.5	91.7	86.0	90.8	93.5	86.9	(93.2)	82.4	97.5	98.3		78.0
S. salar	85.8	84.7	85.5	86.3	89.8	85.2	(91.9)	80.9	94.9	94.9	95.6	

References for Supplemental Data

1. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) *Bioinformatics* **23**, 2947-2948
2. Knowles, D. G., and McLysaght, A. (2009) *Genome Res* **19**, 1752-1759
3. Strobl, S., Fernandez-Catalan, C., Braun, M., Huber, R., Masumoto, H., Nakagawa, K., Irie, A., Sorimachi, H., Bourenkow, G., Bartunik, H., Suzuki, K., and Bode, W. (2000) *Proc Natl Acad Sci U S A* **97**, 588-592
4. Hanna, R. A., Campbell, R. L., and Davies, P. L. (2008) *Nature* **456**, 409-412
5. Hosfield, C. M., Moldoveanu, T., Davies, P. L., Elce, J. S., and Jia, Z. (2001) *J Biol Chem* **276**, 7404-7407
6. Moldoveanu, T., Gehring, K., and Green, D. R. (2008) *Nature* **456**, 404-408
7. Moldoveanu, T., Hosfield, C. M., Lim, D., Elce, J. S., Jia, Z., and Davies, P. L. (2002) *Cell* **108**, 649-660
8. Moldoveanu, T., Hosfield, C. M., Lim, D., Jia, Z., and Davies, P. L. (2003) *Nat Struct Biol* **10**, 371-378
9. Hata, S., Sorimachi, H., Nakagawa, K., Maeda, T., Abe, K., and Suzuki, K. (2001) *FEBS Lett* **501**, 111-114
10. Tompa, P., Emori, Y., Sorimachi, H., Suzuki, K., and Friedrich, P. (2001) *Biochem Biophys Res Commun* **280**, 1333-1339
11. Ono, Y., Kakinuma, K., Torii, F., Irie, A., Nakagawa, K., Labeit, S., Abe, K., Suzuki, K., and Sorimachi, H. (2004) *J. Biol. Chem.* **279**, 2761-2771

Supplemental Figure legends

Fig. S1. Amino acid sequence alignment of CAPN3 products from representative vertebrates. Amino acid sequences of homologues of human p94/calpain 3 were retrieved from cDNA, genome, and/or EST databases (NCBI and ENSEMBL). These included 25 sequences from various vertebrates (12, 2, 1, 2, and 8 from mammals, birds, reptile, amphibians, and fishes, respectively), which are more similar to human p94 than the other 14 human calpains. These sequences were aligned using Clustal W Ver.2 (1), which showed that *T. nigroviridis* and *T. rubripes* have two independent genes that encode proteins significantly and equally similar to mammalian *CAPN3*. One of the two genes (ENSTNIP00000013441 for *T. nigroviridis*, and ENSTRUP00000016935 for *T. rubripes*), however, does not contain a genomic sequence that encodes the IS2 consensus (PIxFVSDRxxxNK), and was eliminated from alignment. Finally, we selected p94 sequences from 12 representative vertebrates (3 mammals, 2 birds, 1 reptile, 2 amphibians, and 4 fishes) and show their alignment here. Since the IS1 sequences are rather divergent, we cannot conclude at present whether IS1 is missing from fish *CAPN3* or if the computer programs failed to identify the exon. For the abbreviations used for species names, see Table S1. The white-on-black residues indicate 100% conservation among the species compared here. Dots, commas, and colons over the sequences indicate every 10, 50, and 100 residues in the human sequence. Green areas indicate three p94 characteristic sequences, NS, IS1, IS2, which are encoded by exons 1, 6, and 15+16, respectively. Pink stars denote the positions of residues involved in Ca^{2+} binding, and yellow stars denote the residues in the proteolytic active site. Yellow ovals indicate where calpainopathy-associated mutations have been identified (from Leiden Muscular Dystrophy pages: http://www.dmd.nl/; see also Fig. 3A). Note that four of these mutations are in the highly conserved CBS-IIa residues (D120N, S194C, E199Q, W201R), and three of them are in the N-terminal half of IS2 (S606L, R608K, and A609E). Most of the sequences shown here, except for those from mammals and chicken, are deduced from the genomic sequences. Therefore, they may contain sequencing errors and/or the exon organization may have been deduced incorrectly (2). For some of these sequences, we performed revisions by comparing them with other species and/or by comparing the genomic and EST sequences (see Table S1).

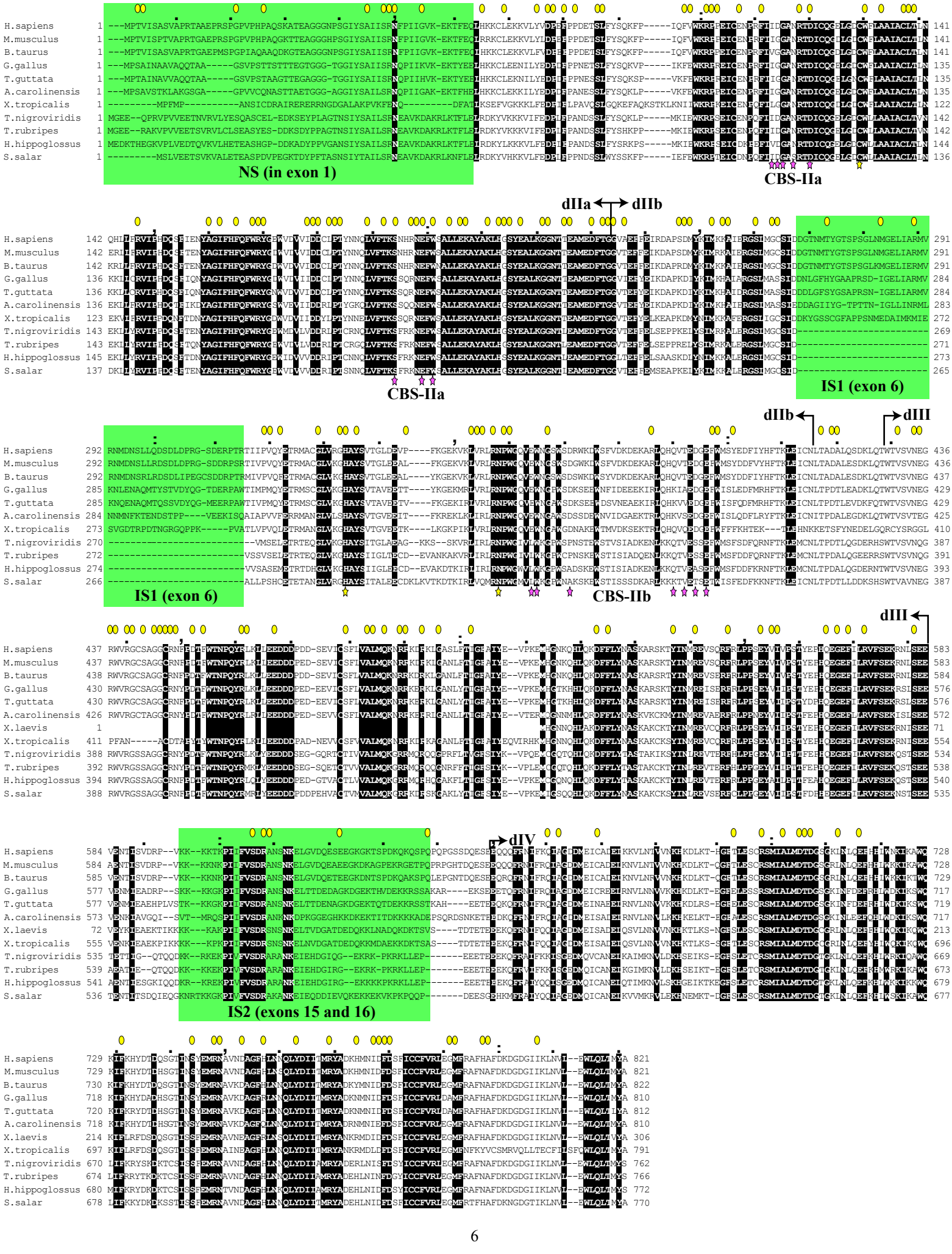
Fig. S2. Partial purification of rat wild-type p94. For procedures, see Materials and Methods. The upper panel shows an elution profile for DEAE column chromatography of rat skeletal muscle extract. The curve and the dotted line represent the A280 absorbance and [CsCl], respectively. The middle panel shows the SDS-PAGE analysis of each fraction, stained by CBB. T, P1, P2, S, and FT indicate total extract, first and second precipitate factions, supernatant fraction, and flow-through fraction. The fraction numbers are given above the gel. The lower panel shows a western blot of the same samples probed with the anti-pIS2 antibody. Closed and open arrowheads indicate the full-length and autolyzed p94 protein, respectively. Fraction 6 (circled) was used for further studies, which are described in Fig. 2B.

Fig. S3. Molecular basis for the activation of p94/calpain 3 by physiological concentrations of Na^+ and Ca^{2+}. Schematic structures of p94/calpain 3 and conventional calpains, μ- and m-calpains, are shown. p94/calpain 3 and the catalytic subunit of conventional calpains consist of the N-terminal domain (NS or domain I (dI)), the protease domain (dII) composed of sub-domains IIa and IIb (3) (dIIa and dIIb; also called sub-domains I and II (4-8)), the C2-domain-like Ca^{2+}-binding domain (dIII), and five EF-hand domains (PEF domain, dIV). Conventional calpains are composed of a large catalytic subunit and a small regulatory subunit, which contains a Gly-rich, hydrophobic domain (dV) and another PEF domain (dVI). Each of dIIa and dIIb contains one Ca^{2+}-binding site (CBS-IIa and CBS-IIb, respectively) (7-9), and dIII and dIV also bind multiple Ca^{2+} ions (4,6,10). Although conventional calpains are activated by Ca^{2+}, but not Na^+, p94's activity is dependent on both Ca^{2+} and Na^+, the effects of which are additive (see Figs. 1, 2). Biochemical analyses using p94 mutants revealed that CBS-IIa is the best candidate for the Na^+-binding site, and it is probably also used for Ca^{2+} binding (see Figs. 3-4). The insertion sequences, IS1 and IS2, and dIII are also involved in p94's Na^+-dependency, whereas dIV is dispensable. Na^+ and Ca^{2+} both activate p94 but differentially influence its substrate specificities. (See Figs. 5-7)

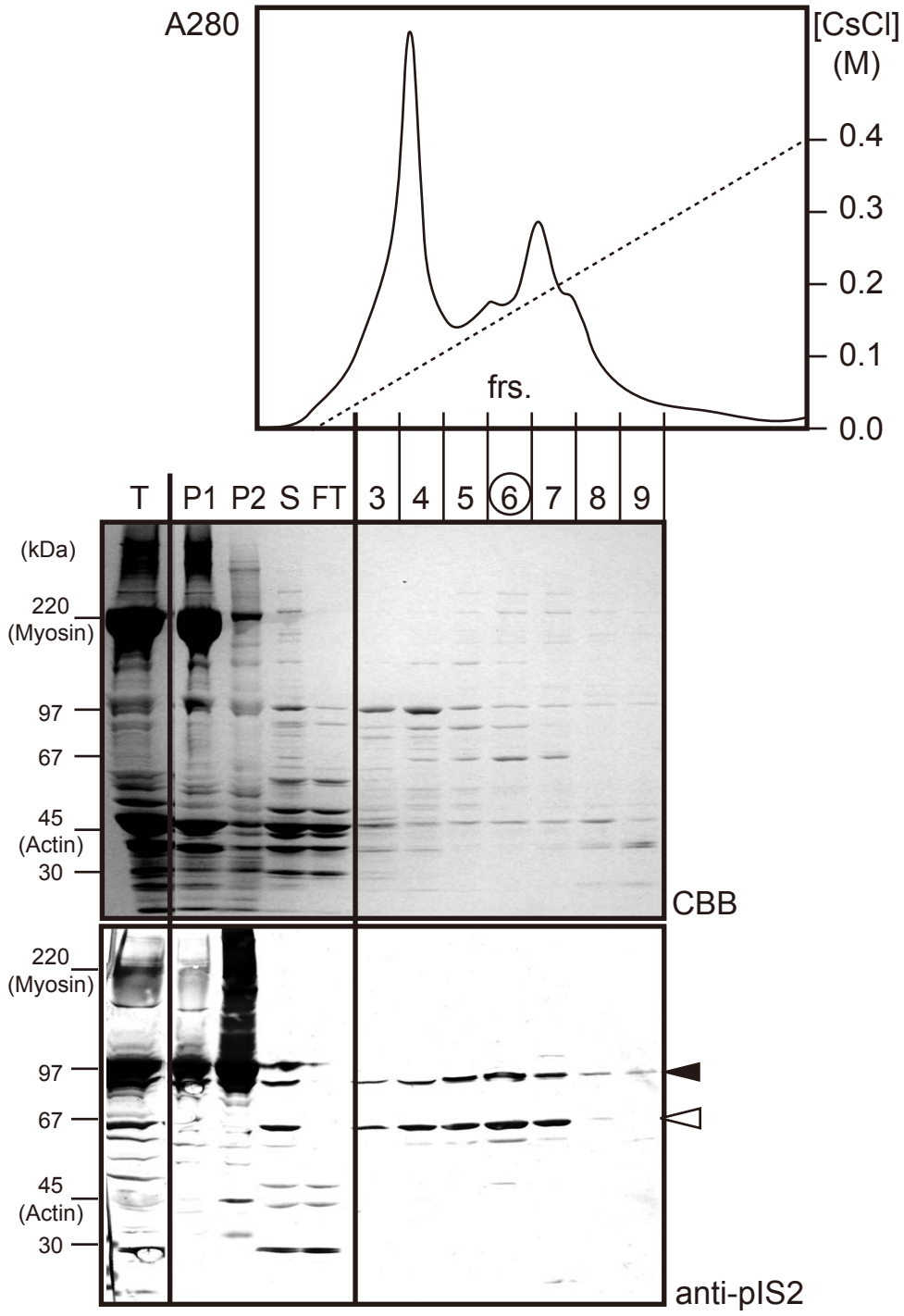
Fig. S4. Summary of the autolytic and fodrinolytic activities of various p94 missense mutants. Wild-type p94 and 37 missense mutants were expressed in COS7 cells, and the lysates were subjected to western analysis (see the "no incubation" condition under "autolysis", and the "fodrinolysis" column) using the anti-pIS2 (11) and anti-GMMPR antibodies (see Materials and Methods), or were incubated in the presence of Ca^{2+} or Na^+ followed by western analysis with the anti-pIS2 antibody ("incubation" condition, under "autolysis"). The activities are qualitatively indicated by -, ±, +, and ++, and compared with those for wild-type. The "domain" column indicates the domain/region where the mutation resides. Under "LAGMD2A mutation," + indicates the mutation was reported as pathogenic in LGMD2A patients. Under "major product," FL and 55K stand for the full-length and 55-kDa autolyzed fragments of p94, respectively. If both were comparably expressed, they are shown in the order of abundance.

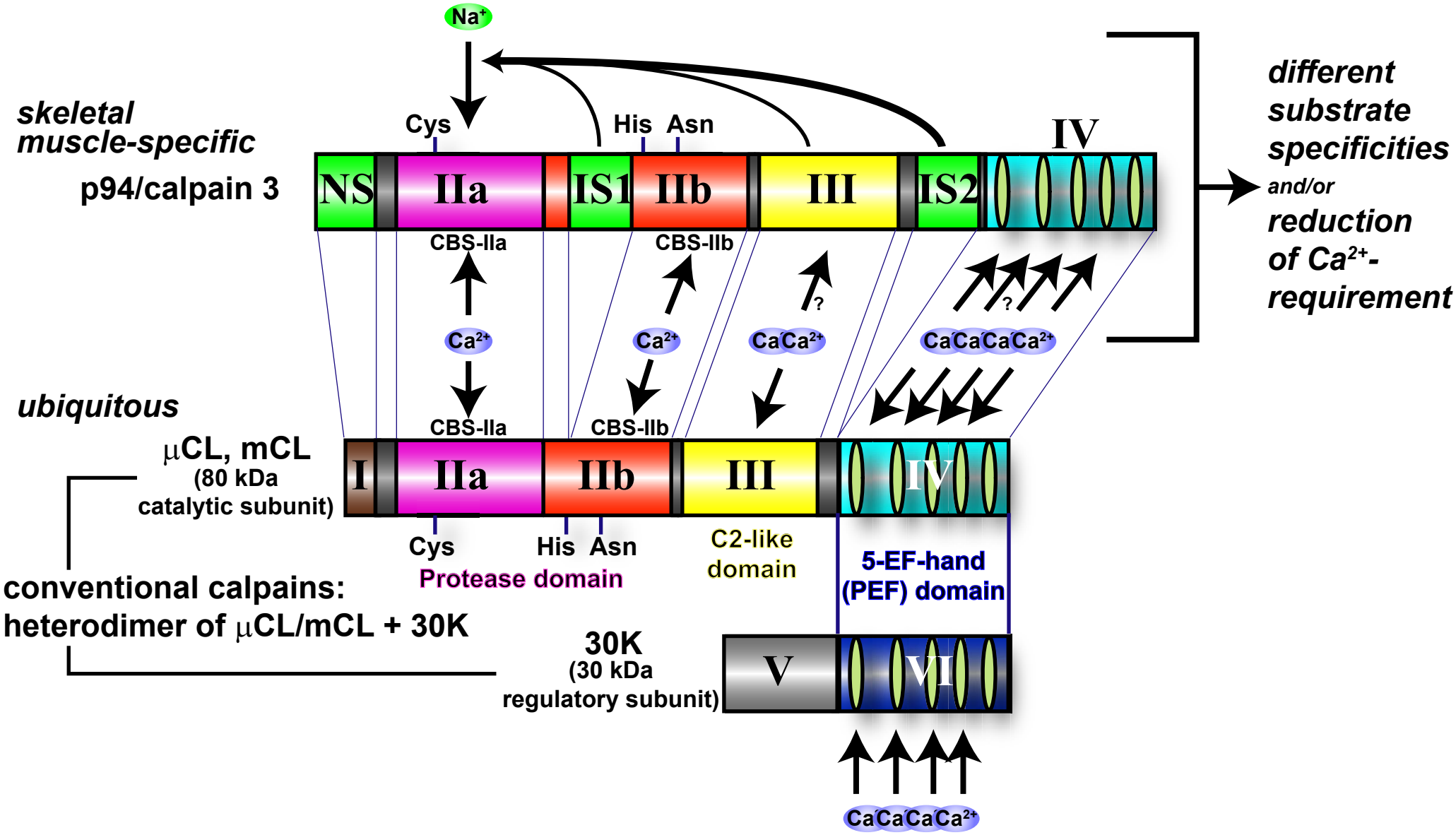
Fig. S5 for Videos. Time-lapse fluorescence imaging of GFP-fused p94WT (A) or p94:C129S (B) in skeletal muscle cell primary cultures prepared from WT B6 mice. In the cultured muscle cells, exogenously expressed GFP-p94WT and GFP-p94:C129S were predominantly localized to the M-lines, and no significant difference between the two was observed in the absence of ouabain. Ouabain (final concentrations of 1 mM) was added at the start of each recording. Images were scanned and recorded using an LSM510 confocal microscope (Zeiss) with frame rates between 6 and 300 frame/min (in order to better focus on the cultured muscle cells, which sometimes move during observation). After the recording, all the frame rates were adjusted to 6 frame/min using LSM510 software. GFP signals at the M-lines in WT cells disappeared within 30 min (15 sec in the movie) after the addition of ouabain. In contrast, those from GFP-p94:C129S were stably detected at the M-lines throughout the recording session (110 min). One second of each movie corresponds to two minutes in real time (12 frame/sec). Recording times: A (WT), 35 min (17.5 sec); B (C129S), 110 min (55 sec). Photo frames from 0 and 35 min (A, B), and 110min (B) are shown.

Ono Y. et al., Figure S1



Ono et al., Figure S2





Ono et al., Figure S3

Ono et al., Figure S4

No.	mutation	domain	LGMD2A mutation	major product	autolysis			fodorinolysis*1 (Proteolysis of fodrin)
					no incubation*1	incubation*2		
						+Na^+	+Ca^{2+}	
0	(WT)	-		55K	+	⊘	⊘	+
1	D120A	IIa		55K	+	⊘	⊘	+
3	D120A/E199A	IIa		FL	-	-	-	n.d.
4	D120A/N358A	IIa/IIb		FL	+	++	++	n.d.
2	E199A	IIa		55K	+	⊘	⊘	+
5	E199A/E364A	IIa/IIb		55K	+	⊘	⊘	n.d.
6	E199A/D371A	IIa/IIb		55K	+	⊘	⊘	n.d.
7	E199A/N358D	IIa/IIb		FL	-	-	-	n.d.
8	G234A	IIb		FL	+	-	-	+
9	G234C	IIb		FL	+	-	-	+
10	G234E	IIb	+	FL	-	-	-	-
11	G234F	IIb		FL	±	-	-	-
12	G234L	IIb		FL	±	-	-	-
13	G234R	IIb		FL	±	-	-	-
14	G234S	IIb		FL	+	-	-	+
15	G234T	IIb		FL	±	-	-	-
16	G234V	IIb		FL	±	-	-	-
17	G234Y	IIb		FL	±	-	-	-
18	N358D	IIb		FL	+	++	++	-
19	N358D/D371A	IIb		FL	+	++	++	n.d.
20	W360Y	IIb		FL	+	++	++	-
21	E364A	IIb		FL, 55K	+	⊘	⊘	++
22	E364A/D371A	IIb		FL, 55K	+	+	+	n.d.
23	D371A	IIb		55K	+	⊘	⊘	+
24	E393K	IIb		FL, 55K	+	+	++	n.d.
25	N358D/R490W	IIb/III		FL	-	±	±	n.d.
26	N358D/R572Q	IIb/III		FL	-	±	±	n.d.
27	N358D/S744G	IIb/IV		FL	+	++	++	n.d.
28	N358D/R769Q	IIb/IV		FL	+	-	++	n.d.
29	S423P	IIb~III		FL, 55K	+	±	++	n.d.
30	R490W	III	+	FL	+	+	+	-
31	R572Q	III	+	FL	+	±	+	-
32	V590M	III		55K, FL	+	⊘	⊘	n.d.
33	F604L	IS2		FL, 55K	+	+	++	n.d.
34	D607A	IS2		FL, 55K	+	+	++	++
35	D607N	IS2		FL, 55K	+	+	++	n.d.
36	E614A	IS2		55K, FL	+	++	++	++
37	Y777H	IV		55K	+	⊘	⊘	n.d.

⊘: could not be determined due to the low signal of the band; n.d.: not done

*1: activity is qualitatively shown by ++, +(=WT), ±, or -

*2: activity is qualitatively shown by ++(=N358D), +, ±, or -

Ono et al., Figure S5