

Supplementary Tables S1–S5

Human and mouse proteases are divided into five classes, which are subdivided into families according to the MEROPS database criteria (Tables S1–S5). We have provided the MEROPS code for all enzymes for which they are available. There are some conflicting cases in which different codes have been previously assigned to human and mouse protease genes that were shown in this work to be true orthologues. In these cases, the human code is proposed for both orthologues. The genes encoding protease-like proteins that show changes in crucial residues for proteolytic activity are indicated as ‘*np*’ (non-protease homologues) after the code.

The Locus link or nucleotide accession number is provided for each protease. The information for human enzymes is labelled in green and for mouse in yellow. Genes that are absent from human or mouse are labelled in red. Genes that have been inactivated by mutation in one species, but are functional in the other, are labelled in pink. Although these specific pseudogenes have been included in the Tables to emphasize the human–mouse difference, they have not been incorporated into the final counts of protease genes. Genes that have been verified experimentally, but the sequence of which is missing from the available genome sequences, are indicated in red and in parentheses. ‘*Y*’ indicates that the corresponding human and mouse genes are syntenic. The percentage of identities between orthologous proteases are also shown.

Table S1 | Aspartic proteases

Code	Peptidase	Human Gene	LocusLink	Locus	Mouse Gene	LocusLink	Locus	Syntenic	Identity
A01.001	pepsin A	PGA3/4/5	5219	11q12	Pepf	58803	19B	y	55
A01.003	pepsin C	PGC	5225	6p12	Pgc	109820	17C	y	73
A01.004	β-secretase 1	BACE	23621	11q23	Bace	23821	9B	y	96
A01.041	β-secretase 2	BACE2	25825	21q22	Bace2	56175	16C4	y	88
A01.006	chymosin	#CYMP	1542	1p13	Cymp	229697	3F3	y	
A01.007	renin	REN	5972	1q31	Ren1	19701	1E4	y	69
A01.009	cathepsin D	CTSD	1509	11p15	Ctsd	13033	7F5	y	81
A01.010	cathepsin E	CTSE	1510	(1q31)	Ctse	13034	1E4	y	82
A01.046	napsin A	NAP1	9476	19q13	Kdap	16541	7B2	y	71
A01.008	submandibular renin				Ren2	19702	1E4		
A02.059	DDI-related protease	DDI-RP	151516	2p13	Ddi-rp	67855	6D2	y	88
A02.xxx	DNA-damage inducible protein	DDI1	AK093336	11q22	Ddi1	71829	9A1	y	81
A02.xxx	DNA-damage inducible protein 2	DDI2	BN000122	1p36	Ddi2	BC021415	4E1	y	96
A02.xxx	Nuclear recept. interacting prot. 2	NRIP2	83714	12p13	Nrip2	60345	6F3	y	83
A02.xxx	Nuclear recept. interacting prot. 3	NRIP3	56675	11p15	Nrip3	78593	7E3	y	93

A22.001	presenilin 1	PSEN1	5663	14q24	Psen1	19164	12D3	y	92
A22.002	presenilin 2	PSEN2	5664	1q42	Psen2	19165	1H4	y	95
A22.005	presenilin homologue 1/SPPL3	PSH1	121665	12q24	Psh1	83678	5F	y	96
A22.006	presenilin homologue 2	PSH2	162540	17q21	Psh2	237958	11D	y	70
A22.003	presenilin homologue 3/SPP	PSH3	81502	20q11	Psh3	14950	2H2	y	96
A22.004	presenilin homologue 4/SPPL2B	PSH4	56928	19p13	Psh4	73218	10C1	y	83
A22.007	presenilin homologue 5	PSH5	84888	15q21	Psh5	66552	2F2	y	83
Ax1.xxx	GCDFP15	PIP	5304	7q34	Pip	18716	6B2	y	47
Ax1.xxxnp	seminal vesicle antigen				Sva	20939	6B2		
Ax1.xxxnp	seminal vesicle antigen-like 1				Sval1	71578	6B2		
Ax1.xxxnp	seminal vesicle antigen-like 2				Sval2	84543	6B2		
Ax1.xxxnp	seminal vesicle antigen-like 3				Sval3	232737	6B2		

These are divided into four families: A01, A02, A22 and Ax1. There are several pepsinogen A isozymogens encoded by highly related genes (>95% identities) that form part of a cluster located at 11q12. The individual pepsinogen A isozymogens result from haplotypes that contain different number of genes (ranging from 1 to 4)^{1,2}. In agreement with other databases, this region has been annotated as a single gene in human. According to the criteria discussed above, we have assigned mouse pepsinogen F as the orthologue of human pepsinogen A, despite notable divergence of their structure and regulation³. *Ren2* is absent in some strains of laboratory mice. The gene that encodes prochymosin has been inactivated by mutations and frameshifts in the human genome and is classified as a pseudogene, although in mouse and other species it is functional⁴.

The genes *DDI1*, *DDI2*, *DDI-RP*, *NRIP2* and *NRIP3* are included in the family A02 that contains predicted retroviral-like aspartic proteases⁵. All of these have mouse orthologues at syntenic regions, and are not embedded in endogenous retroviral elements. The human and mouse genomes also contain several aspartic protease-related sequences derived from endogenous retrovirus, but we have not annotated these as human or mouse proteases. In this regard, it is remarkable that most of the retroviruses embedded in both genomes have suffered inactivating mutations, also affecting the putative proteases that are encoded by these viral elements. However, HERV-K113, for example, which is located at 19p13 in ~30% of the human population, has intact open-reading frames for all viral proteins, including the corresponding aspartic protease, and remains capable of reinfecting human today⁶. The catalogue of aspartic proteases also includes a new family that is derived from the protein prolactin inducible protein/gross cystic disease fluid protein-15 (PIP/GCDFP15), which has recently been characterized as a protease belonging to this class of enzymes⁷. The four PIP-related proteins lack residues proposed to be essential for PIP proteolytic activity and have been classified as non-protease homologues.

Table S2 | Cysteine proteases

Code	Peptidase	Human Gene	LocusLink	Locus	Mouse Gene	LocusLink	Locus	Syntenic	Identity
C01.060	cathepsin B	CTSB	1508	8p23	Ctsb	13030	14C3	y	78
C01.070	cathepsin C	CTSC	1075	11q14	Ctsc	13032	7E1	y	77
C01.018	cathepsin F	CTSF	8722	11q13	Ctsf	56464	19A	y	78
C01.040	cathepsin H	CTSH	1512	15q24	Ctsh	13036	9E3	y	83
C01.036	cathepsin K	CTSK	1513	1q21	Ctsk	13038	3F2	y	86
C01.032	cathepsin L	CTSL	1514	9q21					
C01.009	cathepsin L2	CTSL2	1515	9q22	Ctsl	13039	13B3	y	75
C01.034	cathepsin S	CTSS	1520	1q21	Ctss	13040	3F2	y	73
C01.037	cathepsin W	CTSW	1521	11q13	Ctsw	13041	19A	y	68
C01.013	cathepsin Z	CTSZ	1522	20q13	Ctsz	64138	2H4	y	83
C01.038	cathepsin J				Ctsj	26898	13B3		
C01.023	cathepsin M				Ctsm	64139	13B3		
C01.051	cathepsin Q				Ctsq	104002	13B3		
C01.042	cathepsin R				Ctsr	56835	13B3		
C01.016	cathepsin-1				Cts1	116909	13B3		
C01.031	cathepsin-2				Cts2	56094	13B3		
C01.053	cathepsin-3				Cts3	117066	13B3		
C01.045	cathepsin-6				Cts6	58518	13B3		
C01.973np	tubulointerstitial nephritis antigen	TINAG	27283	6p12	Tinag	26944	9E1	y	85
C01.975np	TINAG related protein	LCN7	64129	1p35	Lcn7	94242	4D3	y	90
C01.972np	testin				Cmb22/23	214639	13B3		
C01.xxxnp	testin-2				Cmb24	70202	13B3		
C01.xxx	testin-3				Cmb25	BY736040	13B3		
C01.084	bleomycin hydrolase	BLMH	642	17q11	Blmh	104184	11B5	y	93
C02.001	calpain 1	CAPN1	823	11q13	Capn1	12333	19A	y	89

C02.002	calpain 2	CAPN2	824	1q42	Capn2	12334	1H4	y	93
C02.004	calpain 3	CAPN3	825	15q15	Capn3	12335	2F1	y	93
C02.011	calpain 5	CAPN5	726	11q13	Capn5	12337	7F1	y	92
C02.971np	calpain 6	CAPN6	827	Xq23	Capn6	12338	XF2	y	95
C02.008	calpain 7	CAPN7	23473	3p25	Capn7	12339	14B	y	95
C02.007	calpain 8	CAPN8	AA043093	(1q42)	Capn8	170725	1H4	y	72
C02.006	calpain 9	CAPN9	10753	1q42	Capn9	73647	8E2	y	85
C02.018	calpain 10	CAPN10	11132	2q37	Capn10	23830	1D	y	81
C02.013	calpain 11	CAPN11	11131	6p21	Capn11	103998	17C	y	83
C02.017	calpain 12	CAPN12	147968	19q13	Capn12	60594	7A3	y	87
C02.020	calpain 13	CAPN13	92291	2p23	Capn13	240159	17E2	y	62
C02.xxx	calpain 14	CAPN14	114773	2p23					
C02.010	calpain 15/Sol protein	SOLH	6650	16p13	Solh	50817	17B1	y	89
C12.001	ubiquitin C-terminal hydrolase 1	UCHL1	7345	4p14	Uchl1	22223	5D	y	94
C12.003	ubiquitin C-terminal hydrolase 3	UCHL3	7347	13q22	Uchl3	50933	14E2	y	98
C12.004	ubiquitin C-term. hydrolase BAP1	BAP1	8314	3p21	Bap1	104416	14B	y	93
C12.005	ubiquitin C-terminal hydrolase 5	UCHL5	51377	1q31	Uchl5	56207	1F	y	96
C12.007	ubiquitin C-terminal hydrolase 4				Uchl4	93841	9D		
C12.xxx	cylindromatosis protein	CYLD1	1540	16q12	Cyld1	74256	8C4	y	95
C13.004	legumain	LGMN	5641	14q32	Lgmn	19141	12F1	y	82
C13.xxx	legumain-2	LGMN2	122199	13q21					
C13.005	hGPI8	PIGK	10026	1p31	Pigk	66613	3H4	y	94
C14.001	caspase-1	CASP1	834	11q22	Casp1	12362	9A1	y	62
C14.006	caspase-2	CASP2	835	7q34	Casp2	12366	6B2	y	89
C14.003	caspase-3	CASP3	836	4q35	Casp3	12367	8B2	y	87
C14.007	caspase-4/11	CASP4	837	11q22	Casp11	12363	9A1	y	60
C14.008	caspase-5	CASP5	838	11q22					
C14.005	caspase-6	CASP6	839	4q25	Casp6	12368	3H1	y	90

C14.004	caspase-7	CASP7	840	10q25	Casp7	12369	19D2	y	82
C14.009	caspase-8	CASP8	841	2q33	Casp8	12370	1C2	y	62
C14.010	caspase-9	CASP9	842	1p36	Casp9	12371	4E1	y	72
C14.011	caspase-10	CASP10	843	2q33					
C14.013	caspase-12	#CASP12	120329	11q22	Casp12	12364	9A1	y	
C14.018	caspase-14	CASP14	23581	19p13	Casp14	12365	10C1	y	74
C14.026	paracaspase	MALT1	10892	18q21	Malt1	240354	18E1	y	90
C14.020np	homologue ICEY	ICEYH	120332	11q22					
C14.971np	casper	CFLAR	8837	2q33	Cflar	12633	1C2	y	68
C14.975np	caspase-14-like	CASP14L	197350	16p13	Casp14L		17A3	y	78
C15.010	pyroglutamyl peptidase I	PGPEP1	65074	19p13	Pgpi	66522	8C1	y	95
C15.011	pyroglutamyl-peptidase II	PGPEP2	145814	15q26	Pgpep2	78444	7C	y	71
C19.019	USP1	USP1	7398	1p31	Usp1	230484	4C6	y	88
C19.013	USP2	USP2	9099	11q23	Usp2	53376	9B	y	95
C19.026	USP3	USP3	9960	15q22	Usp3	235441	9D	y	98
C19.010	USP4	USP4	7375	3p21	Usp4	22258	9F2	y	90
C19.001	USP5	USP5	8078	12p13	Usp5	22225	6F2	y	98
C19.009	USP6	USP6	9098	17p13					
C19.016	USP7	USP7	7874	16p13	Usp7	108732	16A3	y	99
C19.011	USP8	USP8	9101	15q21	Usp8	84092	2F2	y	82
C19.017	USP9X	USP9X	8239	Xp11	Usp9x	22284	XA1	y	98
C19.028	USP9Y	USP9Y	8287	Yq11	Usp9y	107868	(Y)		82
C19.018	USP10	USP10	9100	16q24	Uchrp	22224	8E1	y	83
C19.014	USP11	USP11	8237	Xp11	Usp11	236733	XA2	y	85
C19.020	USP12	USP12	9959	13q12	Ubh1	22217	5G2	y	98
C19.012	USP13	USP13	8975	3q26					
C19.015	USP14	USP14	9097	18p11	Usp14	59025	18A1	y	96
C19.022	USP15	USP15	9958	12q14	Usp15	14479	10D3	y	94

C19.021	USP16	USP16	10600	21q21	Usp16	74112	16C3	y	82
C19.023	USP17	USP17	23661	4p16					
C19.xxx	USP17-like	USP17L	BN000116	8p23					
C19.030	USP18	USP18	11274	22q11	Usp18	24110	6F2	y	70
C19.024	USP19	USP19	10869	3p11	Usp19	71472	9F2	y	79
C19.025	USP20	USP20	10868	9q34	Usp20	74270	2B	y	94
C19.034	USP21	USP21	27005	1q22	Usp21	30941	1H2	y	96
C19.035	USP22	USP22	23326	17p11	Usp22	216825	11B4	y	93
C19.047	USP24	USP24	23358	1p32	Usp24	72686	4C7	y	97
C19.041	USP25	USP25	29761	21q11	Usp25	30940	16C3	y	95
C19.046	USP26	USP26	83844	Xq26	Usp26	83563	XA3	y	36
C19.075	USP27	USP27	AW851065	Xp11	Usp27	54651	XA1	y	97
C19.054	USP28	USP28	57646	11q23	Usp28	235323	9B	y	98
C19.040	USP29	USP29	57663	19q13	Usp29	57775	7A1	y	45
C19.060	USP30	USP30	84749	12q23	Usp30	100756	5F	y	90
C19.071	USP31	USP31	57478	16p12	Usp31	209833	7F2	y	90
C19.044	NY-REN-60	USP32	84669	17q23	Usp32	237899	11B5	y	94
C19.037	VDU1	USP33	23032	1p31	Usp33	170822	3H4	y	92
C19.067	USP34	USP34	9736	2p15					
C19.059	USP35	USP35	57558	11q13	Usp35	244144	7E3	y	82
C19.042	USP36	USP36	57602	17q25	Usp36	72344	12F2	y	74
C19.053	USP37	USP37	57695	2q35					
C19.056	HP43.8KD	USP38	84640	4q31	Usp38	74841	8C3	y	72
C19.972np	SAD1	USP39	10713	2p11	Usp39	28035	6C3	y	87
C19.069	USP40	USP40	55230	2q37	Usp40	227334	1C5	y	81
C19.xxx	USP41	USP41	150200	22q11					
C19.048	USP42	USP42	84132	7p22	Usp42	76800	5G2	y	81
C19.xxx	USP43	USP43	124739	17p12	Usp43	216835	11B3	y	76
C19.057	USP44	USP44	84101	12q21	Usp44	214955	10C2	y	87
C19.975	USP45	USP45	85015	6q16	Usp45	77593	4A3	y	79
C19.052	USP46	USP46	64854	4q12	Usp46	100664	5C3	y	99

C19.055	USP47	USP47	55031	11p15	Usp47	320745	7F2	y	94
C19.068	USP48	USP48	84196	1p36	Usp48	170707	4D3	y	95
C19.073np	USP49	USP49	25862	6p21	Usp49	224836	17C	y	80
C19.058np	USP50	USP50	AI990110	15q21	Usp50	75083	2F2	y	75
C19.065	USP51	USP51	BF741256	Xp11					
C19.xxxnp	USP52	USP52	9924	12q13	Usp52	103135	10D3	y	97
C19.031	DUB-1				Dub1	13531	7F2		
C19.032	DUB-2				Dub2	13532	7F1		
C19.xxx	DUB2a				Dub3	AF393638	7F1		
C19.xxx	DUB2a-like				Dub4	AF393637	7F1		
C19.xxx	DUB2a-like2				Dub5	BAC40791	7F1		
C19.xxx	DUB6				Dub6	BN000117	7F1		
C26.001	γ -glutamyl hydrolase	GGH	8836	8q12	Ggh	14590	4A3	y	69
C44.001	Gln-PRPP amidotransferase	PPAT	5471	4q12	Ppat	231327	5E1	y	93
C44.971np	Gln-fructose-6-P transamidase 1	GFPT1	2673	2p13	Gfpt1	14583	6D2	y	99
C44.972np	Gln-fructose-6-P transamidase 2	GFPT2	9945	5q35	Gfpt2	14584	11B1	y	98
C44.973np	Gln-fructose-6-P transamidase 3	GFPT3	203431	Xq21	#Gfpt3		XC3	y	
C46.002	sonic hedgehog protein	SHH	6469	7q36	Shh	20423	5A3	y	92
C46.003	indian hedgehog protein	IHH	3549	2q35	Ihh	16147	1C3	y	95
C46.004	desert hedgehog protein	DHH	50846	12q13	Dhh	13363	15F2	y	97
C48.002	sentrin/SUMO protease 1	SENP1	29843	12q13	Senp1	223870	15F2	y	88
C48.007	sentrin/SUMO protease 2	SENP2	59343	3q27	Senp2	75826	16B1	y	71
C48.003	sentrin/SUMO protease 3	SENP3	26168	17p13	Senp3	80886	11B4	y	95
C48.008	sentrin/SUMO protease 5	SENP5	205564	3q29	Senp5	AK043171	16B2	y	71
C48.004	sentrin/SUMO protease 6	SENP6	26054	6q14	Senp6	215351	9E2	y	81
C48.009	sentrin/SUMO protease 7	SENP7	57337	3q12	Senp7	72869	16B1	y	87

C48.011	sentrin/SUMO protease 8	SENP8	123228	15q23	Senp8	71599	9C	y	92
C48.016	sentrin/SUMO protease 9				Senp9	236870	XA7		
C48.013	sentrin/SUMO protease 11				Senp11	216394	10D3		
C48.017	sentrin/SUMO protease 12				Senp12	208231	16B5		
C48.015	sentrin/SUMO protease 13				Senp13	114671	10A3		
C48.xxx	sentrin/SUMO protease 14				Senp14	278823	1B		
C48.xxx	sentrin/SUMO protease 15				Senp15	278824	1B		
C50.001	separase	ESPL1	9700	12q13	Espl1	105988	15F3	y	78
C54.003	autophagin-1	AUTL1	23192	2q37	Autl1	66615	1D	y	92
C54.002	autophagin-2	AUTL2	115201	Xq23	Autl2	102926	XF1	y	89
C54.004	autophagin-3	AUTL3	84938	1p31	Autl3	242557	4C6	y	86
C54.005	autophagin-4	AUTL4	84971	19p13	Autl4	235040	9A3	y	86
C56.002	DJ-1	DJ-1	11315	1p36	Dj-1	57320	4E1	y	90
Cx1.xxx	Hin-1	HSHIN1	54726	4q31	Hshin1	234484	8C3	y	91
Cx1.xxx	Hin-1-like	HSHIN1L	BN000160	12p13					
Cx1.xxx	Hin-2	HSHIN2	79868	Xq23	Hshin2	245656	XF2	y	65
Cx1.xxx	Hin-3	HSHIN3	254897	1p36	Hshin3	73162	4D3	y	73
Cx1.xxx	Hin-4	HSHIN4	220213	10p12	Hshin4	71198	2A2	y	90
Cx1.xxx	Hin-5	HSHIN5	55593	Xp11	Hshin5	54644	XA1	y	91
Cx1.xxx	Hin-6	HSHIN6	139562	Xq13	Hshin6	236924	XC2	y	64
Cx1.xxx	Hin-7	HSHIN7	BI829009	1q32	Hshin7	226418	1E4	y	92
Cx1.xxx	Otubain-1	OTUB1	55611	11q13	Otub1	107260	19A	y	99
Cx1.xxx	Otubain-2	OTUB2	78990	14q32	Otub2	68149	12F1	y	95
Cx1.xxx	TNFa-induced protein 3/A20	TNFAIP3	7128	6q23	Tnfaip3	21929	10A2	y	90
Cx1.xxxnp	TRAF-binding protein domain	TRABID	54764	10q26	Trabid	BN000126	7F4	y	99
Cx1.xxx	Cezanne	CEZANNE	56957	1q21	Cezanne	AAH37040	3F2	y	93
Cx1.xxx	Cezanne-2	LOC161725	161725	15q13	AJ430384	170711	7C	y	95

Cx1.xxx	CGI-77	CGI77	51633	8q21	Cgi77	72201	4A2	y	87
Cx1.xxxnp	CGI-77b				Cgi77b	236778	XA3		
Cx2.xxxnp	HetF-like	HETFL	23331	22q12	Hetfl	209683	5F	y	85

The cysteine proteases belong to 16 different families, and include proteins such as hedgehog family members, the protease function of which is only used for the autolytic processing of their respective precursors⁸. The C01 family is largely expanded in the mouse as a result of the presence of placental cathepsins and testins. We have annotated two further mouse testins, including testin-3, which was the first member of this subfamily predicted to be a functional protease. There are two functional human cathepsin L-like genes (*CTSL* and *CTSL2*) at 9q21, and a single gene in the mouse, which is more closely related to *CTSL2*. The cylindromatosis protein contains an ubiquitin C-terminal hydrolase domain and has been included in the C12 family. The genes for calpain 14, caspase 5 and caspase 10 are absent in mice, and the human gene for caspase 12 has been inactivated and is therefore classified as a pseudogene. We have annotated a second human legumain-like gene that is absent in mouse.

The C19 family of ubiquitin specific proteases (USPs) is large and complex. We have annotated 21 human members (USP30, 31, 34–52) and assigned their corresponding mouse orthologues. We have not found mouse orthologues for human *USP6*, *-13*, *-34*, *-37*, *-42* and *-51*. *USP17* is located within the *RS447* human megasatellite at 4p15⁹. This region is highly polymorphic in the human genome, containing a variable number of *USP17*-related intronless tandemly-repeated sequences (>95% identical), which have probably been generated by retrotransposition. Forty-four distinct alleles in 74 unrelated chromosomes containing 20–103 copies of the *RS477* unit have been identified¹⁰. We have also identified several *USP17*-related sequences in a cluster located at 8p25. This cluster would contain at least seven *USP17*-like (*USP17L*) intronless genes (three of these are classified as non-protease homologues) and pseudogenes. The proteins encoded by these polymorphic and variable regions have been annotated as two single proteases (USP17 and USP17L) in this table. The closest relatives of *USP17* genes in the mouse genome are those that code for proteins called DUBs (deubiquitinating enzymes). DUB1, DUB2, and DUB2A have been extensively characterized as members of a novel group of cytokine-inducible deubiquitylating enzymes that are produced by lymphocytes^{11–13}. We have annotated three further members of this subfamily of haematopoietic proteases. The classification of mouse DUBs as orthologues of human *USP17* genes is doubtful because, despite sequence similarities, their syntenic relationship is unclear. Accordingly, we have tentatively classified them as paralogous genes.

We have annotated six members of the C48 family of SUMO-1 proteases in the mouse genome, which are absent in the human genome. We have also included a family of recently described cysteine proteases with deubiquitylating activity containing the OTU-protease domain and tentatively called otubains^{14,15}. This family should comprise 14 orthologues and one specific member in both human and mouse. All of them contain characteristic features of active proteases with the exception of TRABID and murine *Cgi77b*. The last protease included in our list of cysteine proteases is called HetF-like and forms part of the superfamily of caspase-haemoglobinase fold proteases¹⁶. Human and mouse HetF-like have a serine residue instead of the active-site cysteine present in cysteine proteases, and have been classified as non-protease homologues.

Table S3 | Metalloproteases

Code	Peptidase	Human Gene	LocusLink	Locus	Mouse Gene	LocusLink	Locus	Syntenic	Identity
M01.003	aminopeptidase A	ENPEP	2028	4q26	Enpep	13809	3H1	y	77
M01.014	aminopeptidase B	RNPEP	6051	1q32	Rnpep	215615	1F	y	86
M01.023	aminopeptidase MAMS	AMPEP	64167	5q15					
M01.001	aminopeptidase N	ANPEP	290	15q25	Anpep	16790	7D2	y	76
M01.018	aminopeptidase PILS	ARTS1	51752	5q21	Arts1	80898	13C1	y	85
M01.004	leukotriene A4 hydrolase	LTA4H	4048	12q23	Lta4h	16993	10C2	y	92
M01.008	pyroglutamyl-peptidase II	TRHDE	29953	12q21	Trhde	237553	10D1	y	94
M01.010	cytosol alanyl aminopeptidase	NPEPPS	9520	17q21	Psa	19155	11D	y	97
M01.011	leucyl-cystinyl aminopeptidase	LNPEP	4012	5q15	Lnpep	266720	13C1	y	88
M01.022	aminopeptidase B-like 1	RNPEPL1	57140	2q37	Rnpepl1	98480	1D	y	95
M01.028	aminopeptidase O	AOPEP	84909	9q22	Aopep	BAC31943	13B3	y	72
M01.027	aminopeptidase Q	AQPEP	BG623101	5q23	Aqpep	74574	18C	y	68
M01.972np	TBP-associated factor 2	TAF2	6873	8q24	Taf2	319944	15D	y	99
M02.001	angiotensin-converting enzyme 1	ACE	1636	17q23	Ace	11421	11E1	y	83
M02.006	angiotensin-converting enzyme 2	ACE2	59272	Xp21	Ace2	70008	XF5	y	82
M02.971np	angiotensin-converting enzyme 3	#ACE3		17q23	Ace3	217246	11E1	y	
M03.001	thimet oligopeptidase	THOP1	7064	19p13	Thop1	50492	10C1	y	89
M03.002	neurolysin	NLN	57486	5q13	Nln	75805	13D1	y	90
M03.006	mitochondrial intermediate peptidase	MIPEP	4285	13q12	Mipep	70478	14C3	y	84
M08.003	leishmanolysin-2	LMLN	89782	3q29	Lmln	239833	16B2	y	73
M10.034	collagenase-like B				Mcolb	83996	9A1		
M10.001	collagenase 1	MMP1	4312	11q22	Mcola	83995	9A1	y	59
M10.003	gelatinase A	MMP2	4313	16q22	Mmp2	17390	8C5	y	95
M10.005	stromelysin 1	MMP3	4314	11q22	Mmp3	17392	9A1	y	76
M10.008	matrilysin	MMP7	4316	11q22	Mmp7	17393	9A1	y	70

M10.002	collagenase 2	MMP8	4317	11q22	Mmp8	17394	9A1	y	72
M10.004	gelatinase B	MMP9	4318	20q13	Mmp9	17395	2H3	y	72
M10.006	stromelysin 2	MMP10	4319	11q22	Mmp10	17384	9A1	y	76
M10.007	stromelysin 3	MMP11	4320	22q11	Mmp11	17385	10B5	y	81
M10.009	macrophage elastase	MMP12	4321	11q22	Mmp12	17381	9A1	y	61
M10.013	collagenase 3	MMP13	4322	11q22	Mmp13	17386	9A1	y	86
M10.014	MT1-MMP	MMP14	4323	14q11	Mmp14	17387	14C1	y	96
M10.015	MT2-MMP	MMP15	4324	16q22	Mmp15	17388	8C5	y	87
M10.016	MT3-MMP	MMP16	4325	8q22	Mmp16	17389	4A3	y	98
M10.017	MT4-MMP	MMP17	4326	12q24	Mmp17	23948	5F	y	87
M10.021	MMP19	MMP19	4327	12q13	Mmp19	58223	10D3	y	78
M10.019	enamelysin	MMP20	9313	11q22	Mmp20	30800	9A1	y	89
M10.026	MMP21	MMP21	118856	10q26	Mmp21	214766	7F4	y	80
M10.022	MMP23A	MMP23A	8511	(1p36)	Mmp23	26561	4E2	y	83
M10.022	MMP23B	MMP23B	8510	1p36					
M10.023	MT5-MMP	MMP24	10893	20q11	Mmp24	17391	2H2	y	92
M10.024	MT6-MMP	MMP25	64386	16p12	Mmp25	240047	17A3	y	80
M10.029	matrilysin-2	MMP26	56547	11p15					
M10.027	MMP27	MMP27	64066	11q22	Mmp27	234911	9A1	y	57
M10.030	epilysin	MMP28	79148	17q12	Mmp28	118453	11B5	y	79
M12.002	mepirin α -subunit	MEP1A	4224	6p12	Mep1a	17287	17C	y	76
M12.004	mepirin β -subunit	MEP1B	4225	18q12	Mep1b	17288	18A2	y	77
M12.005	procollagen C-protease	BMP1	649	8p21	Bmp1	12153	14D1	y	92
M12.016	mammalian tolloid-like 1 protein	TLL1	7092	4q32	Tll1	21892	8B3	y	93
M12.018	mammalian tolloid-like 2 protein	TLL2	7093	10q24	Tll2	24087	19D1	y	91
M12.245	hatching-metalloprotease	HAMET	AJ537600	2q11	Hamet	215095	2F3	y	67
M12.219	DECYSIN	ADAMDEC1	27299	8p21	Adamdec1	58860	14D1	y	65
M12.201	ADAM1a	#ADAM1	8759	12q24	Adam1a	280668	5F	y	
M12.xxx	ADAM1b				Adam1b	280667	5F		

M12.950np	ADAM2/Fertilin-β	ADAM2	2515	8p11	Adam2	11495	14D1	n	59
M12.975np	ADAM3B	#ADAM3B		16q12	Adam3	11497	8A3	y	
M12.952np	ADAM4	#ADAM4		14q24	Adam4	11498	12D3	y	
M12.xxxnp	ADAM4B	#ADAM4B		14q24	Adam4b	AV274161	12D3	y	
M12.953np	ADAM5	#ADAM5	8757	8p11	Adam5	11499	8A3	y	
M12.xxxnp	ADAM6	#ADAM6		14q24	Adam6	238406	12F2	y	
M12.xxxnp	ADAM6B				Adam6b	238405	12F2		
M12.956np	ADAM7	ADAM7	8756	8p21	Adam7	11500	14D1	y	66
M12.208	ADAM8	ADAM8	101	10q26	Adam8	11501	7F5	y	65
M12.209	ADAM9	ADAM9	8754	8p11	Adam9	11502	8A3	y	86
M12.210	ADAM10	ADAM10	102	15q21	Adam10	11487	9D	y	96
M12.976np	ADAM11	ADAM11	4185	17q21	Adam11	11488	11D	y	91
M12.212	ADAM12	ADAM12	8038	10q26	Adam12	11489	7F4	y	81
M12.215	ADAM15	ADAM15	8751	1q21	Adam15	11490	3F1	y	80
M12.217	ADAM17	ADAM17	6868	2p25	Adam17	11491	(12)	y	91
M12.957np	ADAM18	ADAM18	8749	8p11	Adam18	13524	8A3	y	62
M12.214	ADAM19	ADAM19	8728	5q33	Adam19	11492	11B3	y	82
M12.218	ADAM20	ADAM20	8748	14q24					
M12.234	ADAM21	ADAM21	8747	14q24	Adam21	56622	12D3	y	68
M12.978np	ADAM22	ADAM22	53616	7q21	Adam22	11496	5A1	y	92
M12.979np	ADAM23	ADAM23	8745	2q33	Adam23	23792	1C2	y	94
M12.227	testase 1				Adam24	13526	8B1		
M12.228	testase 2	#ADAM25	137491	8p22	Adam25	23793	8B1	y	
M12.229	testase 3				Adam26	13525	8B1		
M12.224	ADAM28	ADAM28	10863	8p21	Adam28	13522	14D1	y	70
M12.981np	ADAM29	ADAM29	11086	4q34	Adam29	244486	8B3	y	58
M12.232	ADAM 30	ADAM30	11085	1p11	Adam30	71078	3F3	y	63
M12.960np	ADAM 32	ADAM32	203102	8p11	Adam32	209192	8A3	y	60
M12.244	ADAM 33	ADAM33	80332	20p13	Adam33	110751	2F3	y	71
M12.xxx	testase 4				Adam34	252866	8B1		
M12.xxx	testase 5				Adam35	XM_146316	8B1		

M12.xxx	testase 6				Adam36	BN000114	8B1		
M12.xxx	testase 7				Adam37	BN000115	8B1		
M12.xxx	testase 8				Adam38	BN000119	8B1		
M12.247	testase 9				Adam39	BN000121	8B1		
M12.222	ADAMTS1	ADAMTS1	9510	21q21	Adamts1	11504	16C3	y	84
M12.301	ADAMTS2	ADAMTS2	9509	5q35	Adamts2	26550	11B1	y	88
M12.220	ADAMTS3	ADAMTS3	9508	4q21	Adamts3	BAC27597	5E2	y	65
M12.221	ADAMTS4	ADAMTS4	9507	1q23	Adamts4	11505	1H2	y	91
M12.225	ADAMTS5/11	ADAMTS5	11096	21q21	Adamts5	23794	16C3	y	91
M12.230	ADAMTS6	ADAMTS6	11174	5q12	Adamts6	238832	13D1	y	73
M12.231	ADAMTS7	ADAMTS7	11173	15q24	Adamts7	209798	9E3	y	67
M12.226	ADAMTS8	ADAMTS8	11095	11q24	Adamts8	30806	9A5	y	81
M12.021	ADAMTS9	ADAMTS9	56999	3p14	Adamts9	69070	6D3	y	90
M12.235	ADAMTS10	ADAMTS10	81794	19p13	Adamts10	224698	17B2	y	91
M12.237	ADAMTS12	ADAMTS12	81792	5p13	Adamts12	239227	15A2	y	88
M12.241	ADAMTS13	ADAMTS13	11093	9q34	Adamts13	279028	2A3	y	71
M12.024	ADAMTS14	ADAMTS14	140766	10q22	Adamts14	237360	10B4	y	81
M12.025	ADAMTS15	ADAMTS15	170689	11q24	Adamts15	235130	9A5	y	91
M12.026	ADAMTS16	ADAMTS16	170690	5p15	Adamts16	271127	13C1	y	83
M12.027	ADAMTS17	ADAMTS17	170691	15q26	Adamts17	244028	7C	y	78
M12.028	ADAMTS18	ADAMTS18	170692	16q23	Adamts18	208937	8E1	y	73
M12.029	ADAMTS19	ADAMTS19	171019	5q23	Adamts19	240324	18D2	y	82
M12.246	ADAMTS20	ADAMTS20	80070	12q12	Adamts20	223838	15F1	y	70
M13.001	neprilysin	MME	4311	3q26	Mme	17380	3E1	y	94
M13.008	neprilysin-2	MMEL2	79258	1p36	Mell1	27390	4E2	y	79
M13.002	endothelin-converting enzyme 1	ECE1	1889	1p36	Ece1	230857	4D3	y	93
M13.003	endothelin-converting enzyme 2	ECE2	9718	3q29	Ece2	107522	16B1	y	87
M13.007	DINE peptidase	ECEL1	9427	2q37	Ecel1	13599	1C5	y	94
M13.090	Kell blood-group protein	KEL	3792	7q35	Kel	23925	6B2	y	74

M13.091	PHEX endopeptidase	PHEX	5251	Xp22	Phex	18675	XF4	y	96
M14.001	carboxypeptidase A1	CPA1	1357	7q32	Cpa1	109697	6A3	y	74
M14.002	carboxypeptidase A2	CPA2	1358	7q32	Cpa2	232680	6A3	y	86
M14.010	carboxypeptidase A3	CPA3	1359	3q24	Cpa3	12873	3A3	y	81
M14.017	carboxypeptidase A4	CPA4	51200	7q32	Cpa4	215225	6A3	y	84
M14.020	carboxypeptidase A5	CPA5	93979	7q32	Cpa5	76649	1A3	y	84
M14.018	carboxypeptidase A6	CPA6	57094	8q13	Cpa6	329093	1A3	y	86
M14.003	carboxypeptidase B	CPB1	1360	3q25	Cpb1	76703	3A3	y	72
M14.009	carboxypeptidase U	CPB2	1361	13q14	Cpb2	56373	14D2	y	82
M14.021	carboxypeptidase O	CPO	130749	2q33	#Cpo	269201	1C2	y	
M14.005	carboxypeptidase E	CPE	1363	4q32	Cpe	12876	8B3	y	97
M14.004	carboxypeptidase N	CPN	1369	10q25	Cpn	93721	19D1	y	66
M14.006	carboxypeptidase M	CPM	1368	12q15	Cpm	70574	10D2	y	79
M14.011	carboxypeptidase D	CPD	1362	17q11	Cpd	12874	11B4	y	93
M14.012	carboxypeptidase Z	CPZ	8532	4p16	Cpz	242939	5B1	y	82
M14.015np	carboxypeptidase X1	CPX1	56265	20p13	Cpx1	56264	2F3	y	86
M14.019np	carboxypeptidase X2	CPX2	119587	10q26	Cpx2	55987	7F4	y	89
M14.951np	adipocyte-enhancer binding prot. 1	AEBP1	165	7p13	Aebp1	11568	11A1	y	90
M16.002	insulysin	IDE	3416	10q24	Ide	15925	19C3	y	97
M16.003	mitochondrial processing pept. β -sub	PMPCB	9512	7q22	Pmpcb	73078	5A3	y	90
M16.005	nardilysin	NRD1	4898	1p32	Nrd1	230598	4C7	y	93
M16.009	pitrilysin metalloprotease 1	PITRM1	10531	10p15	Pitrm1	69617	13A1	y	86
M16.971np	mitochondrial processing protease	INPP5E	23203	9q34	Inpp5e	66865	2A3	y	91
M16.973np	UCR1	UQCRC1	7384	3p21	Uqrc1	22273	9F2	y	88
M16.974np	UCR2	UQCRC2	7385	16p12	Uqrc2	67003	7F3	y	85
M16.976np	mitoch. processing protease-like	AMPP	133083	4q22					
M17.001	leucyl aminopeptidase	LAP3	51056	4p15	Lap3	66988	5B3	y	90

M17.006	aminopeptidase-like 1	NPEPL1	79716	20q13					
M18.002	aspartyl aminopeptidase	DNPEP	23549	2q36	Dnpep	13437	1C3	y	90
M19.001	membrane dipeptidase	DPEP1	1800	16q24	Dpep1	13479	8E2	y	73
M19.002	membrane dipeptidase 2	DPEP2	64174	16q22	Dpep2	244632	8D2	y	70
M19.004	membrane dipeptidase 3	DPEP3	64180	16q22	Dpep3	71854	8D2	y	73
M20.005	glu-carboxypeptidase-like 1	CPGL	55748	18q22	Cpgl	66054	18E3	y	91
M20.006	glu-carboxypeptidase-like 2	CPGL2	84735	18q22	Cpgl2	240478	18E3	y	73
M20.971np	HmrA-like protease	HMRALP	135293	6q15	Hmralp	242377	4A5	y	83
M20.973np	aminoacylase	ACY1	95	3p21	Acy1	109652	9F1	y	85
M22.003	O-sialoglycoprotein endopeptidase	OSGEP	55644	14q11	Osgep	66246	14C1	y	93
M22.004	O-sialoglycoprotein endopeptidase 2	OSGEP2	64172	2q32	Osgep2	72085	1C1	y	84
M24.001	methionyl aminopeptidase I	METAP1	23173	4q24	Metap1	75624	3H2	y	92
M24.002	methionyl aminopeptidase II	METAP2	10988	12q23	Metap2	56307	10C3	y	88
M24.028	methionyl aminopeptidase-like 1	METAPL1	254042	2q31	Metapl1	66559	2C3	y	95
M24.005	X-prolyl aminopeptidase 2	XPNPEP2	7512	Xq26	Xpnpep2	170745	XA3	y	81
M24.007	X-Pro dipeptidase	PEPD	5184	19q13	Pepd	18624	7B1	y	90
M24.009	aminopeptidase P1	XPNPEPL	7511	10q25	Xpnpep1	170750	19D2	y	81
M24.026	aminopeptidase P homologue	PEPP	63929	22q13	Pepp	321003	15E3	y	93
M24.973np	proliferation-association protein 1	PA2G4	5036	12q13	Pa2g4	18813	10D3	y	98
M24.974np	suppressor of Ty 16 homologue	SUPT16H	11198	14q11	Supt16h	114741	14C1	y	98
M28.010	glutamate carboxypeptidase II	FOLH1	2346	11p11	Folh1	53320	7E1	y	85
M28.011	NAALADASE L peptidase	NAALADL	10004	11q13	NAALADL	BN000129	19A	y	80
M28.012	NAALADASE II	NAALAD2	10003	11q14	Naalad2	72560	9A3	y	89
M28.975np	NAALADASE III	NAALAD3	254827	3q26	Naalad3	229149	3A3	y	63
M28.014	plasma Glu-carboxypeptidase	PGCP	10404	8q22	Pgcp	54381	15B3	y	93

M28.018	Ojeda peptidase	OJP	79956	9p24	Ojp	BAC38286	19C2	y	87
M28.972np	transferrin receptor protein	TFRC	7037	3q29	Trfr	22042	16B3	y	77
M28.973np	transferrin receptor 2 protein	TFR2	7036	7q22	Trfr2	50765	5G1	y	84
M28.974np	glutaminyl cyclase	QPCT	25797	2p22	Qpct	70536	17E3	y	81
M28.016	glutaminyl cyclase 2	QPCT2	54814	19q13	Qpct2	67369	7A2	y	84
M38.972np	dihydroorotase	CAD	790	2p23	Cad	69719	5B1	y	94
M38.973np	dihydropyrimidinase	DPYS	1807	8q22	Dpys	64705	15C	y	88
M38.xxxnp	dihydropyrimidinase-related prot. 1	CRMP1	1400	4p16	Crmp1	12933	5B2	y	96
M38.xxxnp	dihydropyrimidinase-related prot. 2	DPYSL2	1808	8p21	Dpysl2	12934	14D1	y	98
M38.xxxnp	dihydropyrimidinase-related prot. 3	DPYSL3	1809	5q32	Dpysl3	22240	18B3	y	98
M38.xxxnp	dihydropyrimidinase-related prot. 4	DPYSL4	10570	10q26	Dpysl4	26757	7F5	y	93
M38.xxxnp	dihydropyrimidinase-related prot. 5	DPYSL5	56896	2p23	Dpysl5	65254	5B1	y	98
M41.004	i-AAA protease	YME1L1	10730	10p12	Yme1l1	27377	2A3	y	95
M41.006	paraplegin	SPG7	6687	16q24	Spg7	234847	8E2	y	89
M41.010	Afg3-like protein 1	#AFG3L1	172	16q24	Afg3l1	114896	8E2	y	
M41.007	Afg3-like protein 2	AFG3L2	10939	18p11	Afg3l2	69597	18E1	y	94
M43.004	pappalysin-1	PAPPA	5069	9q32	Pappa	18491	4C1	y	93
M43.005	pappalysin-2	PLAC3	60676	1q25	Plac3	240848	1H1	y	78
M47.001	procol. III N-endopeptidase	PCOLN3	5119	16q24	#Pcoln3	BI690732	8E2	y	
M48.003	FACE-1/ZMPSTE24	FACE1	10269	1p34	Face1	230709	4D1	y	91
M48.017	VVML	VVML	115209	1p32	Vvml	67013	4C6	y	71
M49.001	dipeptidyl-peptidase III	DPP3	10072	11q13	Dpp3	75221	19A	y	92
M50.001	S2P protease	MBTPS2	51360	Xp22	Mbtps2	270669	XF4	y	97

M67.001	Pad1-homologue	POH1	10213	2q24	Poh1	59029	2C3	y	99
M67.002	JAB1	COPS5	10987	8q13	Cops5	26754	1A2	y	98
M67.xxxnp	COPS6	COPS6	10980	7q22	Cops6	26893	5G1	y	100
M67.xxx	AMSH	AMSH	10617	2p13	Stamp	70527	6D1	y	83
M67.003	AMSH 2	AMSH2	57559	10q23	Amsh2	76630	19C3	y	89
M67.004	C6.1A	C6.1A	79184	Xq28	C6.1a	210766	XA6	y	97
M67.xxx	C6.1A-like				C6.1al	BN000130	10D1		
M67.005	jammin-like protease 1	JAMML1	114803	1p32	Jamml1	230448	4C5	y	91
M67.xxx	jammin-like protease 2	JAMML2	84954	19p13	Jamml2	68047	17D	y	87
M67.xxxnp	PSMD7	PSMD7	5713	16q23	Psm7	17463	8D2	y	97
M67.xxxnp	PRPF8	PRPF8	10594	17p13	Prpf8	192159	11B4	y	99
M67.xxxnp	eukar. translation initiation F3S3	EIF3S3	8667	8q24	Eif3s3	68135	15D1	y	97
M67.xxxnp	eukar. translation initiation F3S5	EIF3S5	8665	11p15	Eif3s5	66085	7F2	y	93
M67.xxxnp	eukar. translation initiation F3S5B	EIF3S5B	120963	12p13					
M67.xxxnp	IFP38	IFP38	83880	(2p11)					
Mx1.xxx	FACE-2/RCE1	FACE2	9986	11q13	Face2	19671	19A	y	95
Mx2.xxxnp	aspartoacylase-2	ASPA/ACY-2	443	17p13	Aspa/Acy-2	11484	11B4	y	86
Mx2.xxxnp	aspartoacylase-3	ACY-3	91703	11q13	Acy-3	71670	19A	y	68

These belong to 26 distinct families. The M01 family contains 13 members in human and 12 in mouse, which lacks aminopeptidase MAMS. We propose the names aminopeptidases O and Q for the M01 proteases previously annotated as human hypothetical proteins FLJ14675 and BG623101. We have also identified orthologues for these genes located at mouse chromosomes 13B3 and 18C. In the M02 family, we have tentatively annotated a mouse gene for a third angiotensin-converting enzyme-like (*Ace3*), which is located at chromosome 11E1. We have classified *Ace3* as a non-protease homologue because it contains the HQMGH sequence instead of the consensus Zn-binding HExxH motif. No expressed sequence tags (ESTs) have been found for mouse *Ace3*, which could be an inactive pseudogene, although the locus is apparently complete and conserved in the rat. The corresponding human gene is a pseudogene as a result of the accumulation of stop codons and frameshifts.

There are some differences between human and mouse members of the M10 family of matrix metalloproteases (MMPs). Mouse *McolB*, a diverging counterpart of human MMP1 is absent in human, whereas human matrilysin-2 (MMP26) is absent from mouse, although there are some gaps in the mouse genome region which could contain this missing gene. *MMP23* has been recently duplicated in the human genome¹⁷, generating two closely related genes *MMP23A* and *MMP23B*. This region is artefactually collapsed in the available public and private genome sequences owing to the high sequence identity between both genes, and is erroneously considered as containing a single gene. Apparently, there is a single mouse *MMP23* gene, although the possibility

that this region is duplicated in the mouse genome and has also been computer-collapsed can not be ruled out. In the family M12, we have annotated a new member within the meprin/tolloid subfamily¹⁸.

The ADAM (a disintegrin and metalloprotease) subfamily of M12 metalloproteases¹⁹ shows important differences between both organisms. The genes for ADAM-1, -3, -4, -5, -6 and -25 are pseudogenes in the human but active genes in the mouse. ADAM-1 and -6 are duplicated in mouse, whereas ADAM-20 is duplicated in human (ADAM-20 and ADAM-21). Also, testases — a subgroup of ADAMs located at 8B1 — are mouse specific. We have annotated five further members of this family (testases 5–9), although they are intronless and their functional relevance remains to be shown. The group of ADAMTSs (ADAMs with thrombospondin domains) is completed with the inclusion of human and mouse ADAMTS-20. In the M14 family of carboxypeptidases, we have found that mouse carboxypeptidase O has been specifically inactivated by mutation and is annotated as a pseudogene²⁰. Dihydroorotase and several dihydropyrimidinases have been included as non-protease homologues of bacterial isoaspartyl dipeptidases. The gene that encodes procollagen III N-endopeptidase is inactivated in mouse, thereby representing an interesting difference between both human and mouse degradomes, as there are no other functional members in the M47 family that could compensate this specific loss in mouse. We have annotated 14 human and 13 mouse proteins in the recently described M67 family of metalloisopeptidases^{21,22}. All of them contain the JAMM motif, although some lack conserved residues that are predicted to be essential for proteolytic activity, and have therefore been classified as non-protease homologues.

There are doubts about the ascription of the FACE-2/RCE1 prenyl endopeptidase to the cysteine or metalloprotease classes of enzymes²³; however, in agreement with recent structural comparisons²⁴, we have included it as the only human and mouse representative of a new family of membrane-bound metalloproteases. Finally, we have included three aminoacylases in our catalogue of metalloproteases. These enzymes are not, strictly speaking, proteases because they cleave peptide bonds that connect an acyl derivative with an amino acid²⁵. However, the structure of ACY1 clearly allows its inclusion in the M20 family of metalloproteases, whereas those of ACY2 and ACY3 have also been proposed to be part of a superfamily of metalloproteases that contains members of the M14 family of carboxypeptidases²⁶.

Table S4 | Serine proteases

Code	Peptidase	Human gene	LocusLink	Locus	Human gene	LocusLink	Locus	Syntenic	Identity
S01.160	kallikrein hK1	KLK1	3816	19q13	mGk6	16612	7B2	y	65
S01.161	kallikrein hK2	KLK2	3817	19q13	#mGk25		7B2	y	
S01.162	kallikrein hK3	KLK3	354	19q13					
S01.251	kallikrein hK4	KLK4	9622	19q13	mKlk4	56640	7B2	y	69
S01.017	kallikrein hK5	KLK5	25818	19q13	mKlk5	68668	7B2	y	70
S01.236	kallikrein hK6	KLK6	5653	19q13	mKlk6	19144	7B2	y	68
S01.300	kallikrein hK7	KLK7	5650	19q13	mKlk7	23993	7B2	y	75
S01.244	kallikrein hK8	KLK8	11202	19q13	mKlk8	259277	7B2	y	72
S01.307	kallikrein hK9	KLK9	23579	19q13	mKlk9	73832	7B2	y	76
S01.246	kallikrein hK10	KLK10	5655	19q13	mKlk10	69540	7B2	y	68
S01.257	kallikrein hK11	KLK11	11012	19q13	mKlk11	56538	7B2	y	80
S01.020	kallikrein hK12	KLK12	43849	19q13	mKlk12	69511	7B2	y	71
S01.306	kallikrein hK13	KLK13	26085	19q13	mKlk13	13647	7B2	y	79
S01.029	kallikrein hK14	KLK14	43847	19q13	mKlk14	233190	7B2	y	73
S01.081	kallikrein hK15	KLK15	55554	19q13	mKlk15	XM_145570	7B2	y	75
S01.164	glandular kallikrein mK1				mGk1	16623	7B2		
S01.170	glandular kallikrein mK3				mGk3	18050	7B2		
S01.066np	glandular kallikrein mK4				mGk4	18048	7B2		
S01.037	glandular kallikrein mK5				mGk5	16622	7B2		
S01.067	glandular kallikrein mK8				mGk8	16624	7B2		
S01.071	glandular kallikrein mK9				mGk9	13648	7B2		
S01.041	glandular kallikrein mK11				mGk11	16613	7B2		
S01.068	glandular kallikrein mK14				mGk14	16614	7B2		
S01.163	glandular kallikrein mK16				mGk16	16615	7B2		
S01.038	glandular kallikrein mK21				mGk21	16616	7B2		
S01.039	glandular kallikrein mK22				mGk22	13646	(7B2)		
S01.069	glandular kallikrein mK24				mGk24	16617	7B2		
S01.070	glandular kallikrein mK26				mGk26	16618	7B2		
S01.073	glandular kallikrein mK27				mGk27	16619	7B2		

S01.107	glandular kallikrein mKx				mGkx	76999	7B2		
S01.217	thrombin	F2	2147	11p11	F2	14061	2E1	y	82
S01.215	coagulation factor VIIa	F7	2155	13q34	F7	14068	8A2	y	70
S01.214	coagulation factor IXa	F9	2158	Xq27	F9	14071	XA5	y	82
S01.216	coagulation factor Xa	F10	2159	13q34	F10	14058	8A2	y	76
S01.213	coagulation factor XIa	F11	2160	4q35	F11	109821	8B2	y	78
S01.211	coagulation factor XIIa	F12	2161	5q35	F12	58992	13B2	y	72
S01.218	protein C	PROC	5624	2q21	Proc	19123	18B3	y	68
S01.979np	protein Z	PROZ	8858	13q34	Proz	66901	8A2	y	67
S01.212	plasma kallikrein	KLKB1	3818	4q35	Klkb1	16621	8B2	y	76
S01.228	hepatocyte growth factor activator	HGFAC	3083	4p16	Hgfac	54426	5B1	y	81
S01.033	hyaluronan-binding ser-protease	HABP2	3026	10q25	Habp2	226243	19D2	y	80
S01.998np	protein C-like	PROCL	25891	11p12	Procl	210622	2E3	y	90
S01.303	mastin	#MASTIN	257157	16p13	Mastin	207224	17A3	y	
S01.242	tryptase β -1	TPSB1	7177	16p13	Mcpt7	17230	17A3	y	75
S01.242	tryptase β -2	TPSB2	64499	16p13	Mcpt6	17229	17A3	y	77
S01.028	tryptase γ -1	TPSG1	25823	16p13	Tpsg1	26945	17A3	y	73
S01.074	marapsin	MPN	83886	16p13	Mpn	213171	17A3	y	80
S01.075	tryptase homologue 2	EOS	260429	16p13	Eos	BE646687	17A3	y	81
S01.076	tryptase homologue 3	TESSP1	BN000124	16p13	Tessp1	71003	17A3	y	62
S01.011	testisin	PRSS21	10942	16p13	Prss21	57256	17A3	y	67
S01.252	brain serine protease 2	PRSS22	64063	16p13	Prss22	70835	17A3	y	75
S01.314	implantation serine protease 1				Isp1	114661	17A3	y	
S01.315	implantation serine protease 2	#ISP2	123787	16p13	Isp2	114662	17A3	y	
S01.098	intestinal serine protease 1	#DISP	124221	16p12	Disp	30943	17A3	y	
S01.295	intestinal serine protease 2				Disp2	69814	17A3		
S01.054	tryptase δ -1	TPSD1	23430	16p13					
S01.143	tryptase α	TPS1	7176	(16p13)					
S01.159	prostasin	PRSS8	5652	16p11	Prss8	76560	7F4	y	77

S01.414	prostasin-like 1	PSTL1	146547	16p11	PSTL1	77613	7F4	y	62
S01.xxx	prostasin-like 2	PSTL2	79001	16p11	PSTL2	27973	7F4	y	84
S01.xxx	epidermis-specific SP-like	ESSPL	BN000134	4q31	Esspl	BN000135	3F1	y	68
S01.318	marapsin 2	MPN2	BN000131	1q42	Mpn2	216797	11B2	y	57
S01.993np	testis-specific protein tsp50	TSP50	29122	3p21	Tsp50	235631	9F2	y	61
S01.317	testis serine protease 2	TESSP2	AJ544583	3p21	Tessp2	235628	9F2	y	64
S01.106	testis serine protease 3	#TESSP3		3p21	Tessp3	73336	9F2	y	62
S01.xxx	testis serine protease 4	#TESSP4		3p21	Tessp4	272643	9F2	y	
S01.968np	testis serine protease 5	TESSP5	BN000137	3p21	Tessp5	260408	9F2	y	68
S01.xxx	testis serine protease 6	#TESSP6		3p21	Tessp6	74306	9F2	y	66
S01.985np	TESP1				Tesp1	21755	1B		
S01.045	TESP2	#TESP2		2q21	Tesp2	21756	1B	y	
S01.088	TESP3	#TESP3		9q22	Tesp3	218304	13B3	y	
S01.140	chymase	CMA1	1215	14q11	Mcpt5	17228	14C2	y	74
S01.010	granzyme B	GZMB	3002	14q11	Gzmb	14939	14C1	y	67
S01.133	cathepsin G	CTSG	1511	14q11	Ctsg	13035	14C1	y	69
S01.147	granzyme H	GZMH	113155	14q11	Gzmc	14940	14C1	y	60
S01.141	mast cell protease 1				Mcpt1	17224	14C2		
S01.003	mast cell protease 2				Mcpt2	17225	14C1		
S01.149	mast cell protease 4				Mcpt4	17227	14C2		
S01.254	mast cell protease 8				Mcpt8	17231	14C1		
S01.304	mast cell protease 9				Mcpt9	17232	14C1		
S01.004	mast cell protease 10				Mcpt10	AF361939	14C1		
S01.xxx	mast cell protease L				Mcptl	17233	14C1		
S01.398	granzyme D				Gzmd	14941	14C1		
S01.399	granzyme E				Gzme	14942	14C1		
S01.401	granzyme F				Gzmf	14943	14C1		
S01.402	granzyme G				Gzmg	14944	14C1		
S01.xxx	granzyme N				Gzmn	245839	14C1		
S01.xxx	granzyme O				Gzmo	239106	14C1		

S01.135	granzyme A	GZMA	3001	5q11	Gzma	14938	13D2	y	68
S01.146	granzyme K	GZMK	3003	5q11	Gzmk	14945	13D2	y	70
S01.139	granzyme M	GZMM	3004	19p13	Gzmm	16904	10C1	y	70
S01.134	protease 3	PRTN3	5657	19p13	Prtn3	19152	10C1	y	63
S01.131	neutrophil elastase	ELA2	1991	19p13	Ela2	50701	10C1	y	72
S01.971np	azurocidin	AZU1	566	19p13					
S01.156	enteropeptidase	PRSS7	5651	21q21	Prss7	19146	16C3	y	75
S01.xxx	enteropeptidase-like	PRSS7L	BQ638967	2q37	Prss7l	332474	1C5	y	89
S01.224	hepsin	HPN	3249	19q13	Hpn	15451	7B1	y	88
S01.291	HAT-related protease	HATRP	283471	12q13	Hatrp	75002	15F1		59
S01.301	airway-trypsin-like protease	HAT	9407	4q13	Hat	231382	5E1	y	66
S01.292	HAT-like 1	HATL1	BN000133	4q13	Hatl1	194597	5E1	y	76
S01.xxx	HAT-like 2	#HATL2		4q13	Hatl2	320454	5E1	y	
S01.294	HAT-like 3	#HATL3		4q13	Hatl3	231381	5E1	y	
S01.321	HAT-like 4	#HATL4	132722	4q13	Hatl4	243083	5E1	y	
S01.xxx	HAT-like 5	HATL5	132724	4q13	Hatl5	BAC29606	5E1	y	52
S01.021	DESC1 protease	DESC1	28983	4q13	Desc1	243084	5E1	y	75
S01.019	corin	PRSC	10699	4p12	Lpr4	53419	5D	y	82
S01.302	matriptase	MTSP1	6768	11q24	Mtsp1	19143	9A5	y	80
S01.247	epitheliasin	TMPRSS2	7113	21q22	Tmprss2	50528	16C4	y	77
S01.079	transmembrane Ser-protease 3	TMPRSS3	64699	21q22	Tmprss3	140765	17B2	y	82
S01.034	transmembrane Ser-protease 4	TMPRSS4	56649	11q23	Tmprss4	214523	9B	y	76
S01.313	spinesin	TMPRSS5	80975	11q22	Tmprss5	80893	9B	y	78
S01.308	matriptase-2	TMPRSS6	164656	22q12	Tmprss6	71753	15E1	y	84
S01.xxx	matriptase-3	TMPRSS7	BN000125	3q13	Tmprss7	208171	16B5	y	91
S01.298	polyserase	TMPRSS8	AJ488946	19p13	Tmprss8	270749	10C1	y	80
S01.087	membrane-type mosaic Ser-prot.	MSPL	84000	11q23	Mspl	AAH42878	9B	y	90
S01.320	oviductin-like	OVTN	BN000130	11p15	Ovtn	BN000123	7F2	y	71
S01.322	ovochoymase-like	OVCH	BN000128	12p11					

S01.152	chymotrypsin B	CTRB1	1504	16q23	Ptrb	66473	8D3	y	85
S01.256	chymopasin	CTRL	1506	16q22	Ctrl	109660	8D2	y	86
S01.157	chymotrypsin C	CTRC	11330	1p36	Ctrc	76701	4E1	y	77
S01.127	cationic trypsin	PRSS1	5644	7q34	Try4	22074	6B2	y	77
S01.060	trypsin 3	#TRY3		7q34	Try3	22073	6B2	y	
S01.059	trypsin 10	#TRY10		7q34	Try10	AAB69058	6B2	y	
S01.258	anionic trypsin (II)	PRSS2	5645	7q34	Try2	22072	6B2	y	77
S01.063	trypsin C	#TRY6	154754	7q34	Try10l	BN000136	6B2	y	
S01.062	trypsin 15	#TRY15		7q34	Try15	AAB69087	6B2	y	
S01.xxxnp	trypsin X1	#TRYX1		7q34	Tryx1	272341	6B2	y	
S01.989np	trypsin X2	TRYX2	136242	7q34	Tryx2	67690	6B2	y	78
S01.174	mesotrypsin	PRSS3	5646	9p13					
S01.151	trypsin 1				Try1	67373	6B2		
S01.058	trypsin 9				Try9	BAB25300	6B2		
S01.061	trypsin 12				Try12	AAB69086	(6B2)		
S01.063	trypsin 16				Try16	114228	6B2		
S01.984np	trypsin X3				Tryx3	194359	6B2		
S01.xxx	trypsin X4				Tryx4	194360	6B2		
S01.129	trypsin 4				Try4bis	73626	6B2		
S01.092	trypsin V				Tryv	232718	6B2		
S01.105	trypsin X5				Tryx5	73481	6B2		
S01.153	pancreatic elastase	#ELA1	1990	12q13	Ela1	109901	15F3	y	
S01.155	pancreatic elastase II (IIA)	ELA2A	63036	1p36	Ela2a	13706	4E1	y	75
S01.154	pancreatic endopeptidase E (A)	ELA3A	10136	1p36	Ela3a	242711	4D3	y	76
S01.205	pancreatic endopeptidase E (B)	ELA3B	23436	1p36	Ela3b	67868	4D3	y	84
S01.206	pancreatic elastase II form B	ELA2B	51032	1p36					
S01.194	complement component 2	C2	717	6p21	C2	12263	17B2	y	76
S01.196	complement factor B	BF	629	6p21	Bf	14962	17B2	y	84
S01.995np	complement C1r-homologue	C1RL	51279	12p13	C1rl	232371	6F2	y	73

S01.192	complement component C1ra	C1R	715	12p13	C1ra	50909	6F2	y	81
S01.xxx	complement component C1rb				C1rb	AF459018	(6F2)		
S01.193	complement component C1sa	C1S	716	12p13	C1sa	50908	6F2	y	74
S01.xxx	complement component C1sb				C1sb	317677	6F2		
S01.191	complement factor D	DF	1675	19p13	Df	11537	10C1	y	67
S01.xxx	complement factor D-like	DF2	199783	19p13	Df2	270746	10C1	y	79
S01.199	complement factor I	IF	3426	4q25	If	12630	3H1	y	69
S01.198	MASP1/3	MASP1/3	5648	3q29	Masp1/3	17174	16B1	y	86
S01.229	MASP2	MASP2	10747	1p36	Masp2	17175	4E1	y	81
S01.237	neurotrypsin	PRSS12	8492	4q28	Prss12	19142	3G3	y	82
S01.231	u-plasminogen activator	PLAU	5328	10q22	Plau	18792	14B	y	69
S01.232	t-plasminogen activator	PLAT	5327	8p11	Plat	18791	8A3	y	80
S01.233	plasminogen	PLG	5340	6q26	Plg	18815	17A2	y	79
S01.976np	hepatocyte growth factor	HGF	3082	7q21	Hgf	15234	5A3	y	91
S01.975np	macrophage-stimulating protein	MSP	4485	3p21	Msp	15235	9F2	y	80
S01.999np	apolipoprotein	LPA	4018	6q26					
S01.223	acrosin	ACR	49	22q13	Acr	11434	15F1	y	68
S01.972np	haptoglobin-1	HP	3240	16q22	Hp	15439	8D3	y	79
S01.974np	haptoglobin-related protein	HPR	3250	16q22					
S01.277	osteoblast serine protease	HTRA1	5654	10q26	Htra1	56213	7F4	y	91
S01.278	HTRA2	HTRA2	27429	2p12	Htra2	64704	6D1	y	84
S01.284	HTRA3	HTRA3	94031	4p16	Htra3	78558	5B1	y	86
S01.285	HTRA4	HTRA4	203100	8p11	Htra4	66943	8A3	y	66
S01.309	umbilical vein protease	SPUVE	11098	11q14	Spuve	76453	7E1	y	90
S01.994np	similar to SPUVE	SPUVE2	167681	6q14	Spuve2	244954	9E3	y	77
S01.104	plasma-kallikrein-like 1	KLKBL1	XP_116753	8p23	Klkb11	74215	(14C3)		66
S01.415	plasma-kallikrein-like 2	KLKBL2	203074	8p23	Klkb12	71037	14C3	y	71
S01.419	plasma-kallikrein-like 3	#KLKBL3		8p23	Klkb13	73382	14C3	y	

S01.992np	plasma-kallikrein-like 4	KLKBL4	221191	16q21	Klklbl4	BN000132	8C5	y	62
S01.286	similar to <i>Arabidopsis</i> Ser-prot.	SASP	219743	10q22	Sasp	71767	10B4	y	80
S01.991np	chymase-like serine protease				Clsp	75106	XC3		
S08.063	site-1 protease	MBTPS1	8720	16q23	Mbtps1	56453	8E1	y	96
S08.039	proprotein convertase 9	PCSK9	255738	1p32	Pcsk9	100102	4C7	y	73
S08.090	tripeptidyl-peptidase II	TPP2	7174	13q33	Tpp2	22019	1C1	y	95
S08.072	proprotein convertase 1	PCSK1	5122	5q15	Pcsk1	18548	13C1	y	93
S08.073	proprotein convertase 2	PCSK2	5126	20p12	Pcsk2	18549	2H1	y	97
S08.071	furin	PCSK3	5045	15q26	Pcsk3	18550	7D2	y	94
S08.074	proprotein convertase 4	PCSK4	5124	19p13	Pcsk4	18551	10C1	y	82
S08.076	proprotein convertase 5	PCSK5	5125	9q21	Pcsk5	18552	19B	y	92
S08.075	PACE4 proprotein convertase	PCSK6	5046	15q26	Pcsk6	18553	7C	y	93
S08.077	proprotein convertase 7	PCSK7	9159	11q23	Pcsk7	18554	9B	y	88
S09.001	prolyl oligopeptidase	PREP	5550	6q22	Prep	19072	10B2	y	96
S09.015	prolyl-oligopeptidase 2	PREP2	9581	2p21	Prep2	213760	17E4	y	94
S09.003	dipeptidyl-peptidase 4	DPP4	1803	2q24	CD26	13482	2C3	y	85
S09.973np	dipeptidyl-peptidase 6	DPP6	1804	7q36	Dpp6	13483	5A3	y	91
S09.018	dipeptidyl-peptidase 8	DPP8	54878	15q23	Dpp8	74388	9D	y	95
S09.019	dipeptidyl-peptidase 9	DPP9	91039	19p13	Dpp9	224897	17D	y	89
S09.974np	dipeptidyl-peptidase 10	DPP10	57628	2q14	Dpp10	269109	1E2	y	88
S09.007	Seprase	FAP	2191	2q24	Fap	14089	2C3	y	90
S09.004	acylaminoacyl-peptidase	APEH	327	3p21	Apeh	235606	9F2	y	91
S09.055	CGI-67 protein	CGI-67	51104	9q21	Cgi-67	BN000127	19C1	y	98
S09.052	CGI-67-like protease-1	CGI-67L1	81926	19p13	Cgi-67l1	216169	10C1	y	93
S09.053	CGI-67-like protease-2	CGI-67L2	58489	15q25	Cgi-67l2	70178	7D3	y	97
S09.051	BEM46-like 1	BEM46L1	84945	13q33	Bem46l1	68904	8A2	y	97

S09.054	BEM46-like 2	BEM46L2	26090	20p11	Bem46l2	76192	2H1	y	90
S09.xxx	BEM46-like 3	BEM46L3	BG74273	14q22	Bem46l3	278594	12C3	y	78
S10.002	lysosomal carboxypeptidase A	PPGB	5476	20q13	Ppgb	19025	2H3	y	87
S10.003	vitellogenic carboxypeptidase-L	CPVL	54504	7p15	Cpvl	71287	6B3	y	76
S10.013	serine carboxypeptidase 1	RISC	59342	17q23	Risc	74617	11C	y	82
S12.004	β -lactamase	LACTB	114294	15q22	Lactb	80907	9D	y	85
S14.003	endopeptidase Clp	CLPP	8192	19p13	Clpp	53895	17E1	y	87
S16.002	PIM1 endopeptidase	PRSS15	9361	19p13	Prss15	74142	17E1	y	88
S16.006	PIM2 endopeptidase	PIM2	83752	16q21	Pim2	66887	8C4	y	95
S26.009	signalase 18 kDa component	SPC18	23478	15q25	Spc18	56529	7D2	y	98
S26.010	signalase 21 kDa component	SPC21	90701	18q21	Spc21	66286	18E1	y	98
S26.xxx	signalase-like 1	SPCL1	158326	9p22	Spcl1	230344	4C3	y	76
S26.012	mitoc. inner membrane protease 2	IMMP2L	83943	7q31	Imp2l	93757	12B3	y	90
S26.013	mitochondrial signal peptidase	IMMP1	196294	11p13	Imp1	66541	2E3	y	95
S26.xxx	lactotransferrin	LTF	4057	3p21	Ltf	17002	9F2	y	70
S28.001	lysosomal Pro-X carboxypeptidase	PRCP	5547	11q14	Prcp	72461	7E2	y	77
S28.002	dipeptidyl-peptidase II	DPP7	29952	9q24	Dpp7	83768	2A3	y	80
S28.003	thymus-specific serine peptidase	PRSS16	10279	6p21	Prss16	54373	13A3	y	79
S33.009	$\alpha\beta$ -hydrolase dom. containing 4	ABHD4	63874	14q11	Abhd4	105501	14C1	y	96
S33.971np	epoxyde hydrolase	EPHX1	2052	1q42	Ephx1	13849	1H4	y	83
S33.972np	Mesoderm specific transcript hom.	MEST	4232	7q32	Mest	17294	6A3	y	97
S33.974np	epoxyde hydrolase related protein	EPHXRP	253152	1p22	Ephxrp	243192	5E	y	87
S33.xxxnp	CGI-58	CGI-58	51099	3p21	Cgi-58	67469	9F4	y	94
S53.003	tripeptidyl-peptidase I	CLN2	1200	11p15	Cln2	12751	7F1	y	88

S54.005	rhomboid-like protein 1	RHBDL	9028	16p13	Rhbd1	214951	17B1	y	97
S54.002	rhomboid-like protein 2	RHBDL2	54933	1p34	Rhbd12	230727	4D1	y	89
S54.006	rhomboid-like protein 4	RHBDL4	162494	17q11	Rhbd14	246104	11B5	y	95
S54.xxx	rhomboid-like protein 5	RHBDL5	84236	2q36	Rhbd15	76867	1C5	y	80
S54.953np	rhomboid-like protein 6	RHBDL6	79651	17q25	Rhbd16	276799	11E2	y	93
S54.xxxnp	rhomboid-like protein 7	RHBDL7	AC005067	7q11	Rhbd17	215160	5G1	y	88
S54.xxx	Presenilins associated rhomboid like	PARL	55486	3q27	Parl	208159	16B1	y	80
S54.952np	EGF Receptor Related Sequence	EGFR-RS	64285	16p13	Egfr-rs	13650	11A5	y	95
Sx1.xxx	Reelin	RELN	5649	7q22	Reln	19699	5A3	y	95
Sx2.xxx	tumor rejection antigen (gp96)	TRA1	7184	12q23	Tra1	22027	10C2	y	97
Sx2.xxxnp	heat shock 90kDa protein 1, α	HSPCA	3320	14q32	Hspca	15519	12F2	y	99
Sx2.xxxnp	heat shock 90kDa protein 1, β	HSPCB	3326	6p21	Hsp84-1	15516	17C	y	98
Sx2.xxxnp	heat shock protein 75	TRAP1	10131	16p13	Trap1	68015	16A1	y	88

Most of these belong to the S01 family, but there are representatives of 13 further serine protease families in the human and mouse degradomes. All differences between human and mouse serine proteases correspond to changes in members of this densely populated family. The kallikreins are duplicated in mouse almost entirely — there are 28 members in mouse and 15 in human. The genes for mastin, implantation serine protease-2 (ISP-2), intestinal serine protease (DISP-1), and testis serine proteases TESP-2 and -3, are inactivated in human hence their classification as pseudogenes. The absence of genes for human DISP-2, ISP-1 and TESP-1, together with the finding that human *DISP-1*, *ISP-2*, *TESP-2* and *TESP-3* are pseudogenes, indicates that the functions performed by ISP, DISP and TESP proteases might be mouse-specific. We have also annotated several new members of the testis-specific serine protease (TESSP) subfamily, with *TESSP-3*, *-4* and *-6* being pseudogenes in human and active genes in mouse. Mast-cell proteases (Mcpt), granzymes (Gzm), trypsins and human-airway trypsin-like (HAT-like) proteases are expanded in mouse; two tryptases, an ovochymase-like protease and a form of pancreatic elastase, are only present in human. Two well-known non-protease homologues, apolipoprotein (a) (LPA) and haptoglobin-related protein, are absent in mouse. Further characteristic features of the mouse degradome include the duplication of complement factors C1r and C1s, and the presence of an extra functional member of the plasma-kallikrein like subfamily (Kikbl3), and of a non-protease homologue called Clsp (chymase-like serine protease).

We have included in the catalogue of serine proteases, a series of proteins such as lactoferrin, reelin and tumour rejection antigen (gp96), which have been recently reported to have this kind of proteolytic activity²⁷⁻²⁹. On the basis of structural analysis, lactoferrin has been tentatively classified as a member of the S26 family of serine proteases, whereas reelin, gp96 and their close relatives have been preliminarily ascribed to two Sx families of presently unclassified serine proteases. Gene Ontology annotation of the human proteome also predicts a series of serine proteases with minimal relationship to other members of this class of enzymes. They include torsin, NSP (novel serine protease) and Ufd1L (ubiquitin fusion degradation protein 1 homologue), but

owing to the absence of enough evidence to support its ascription as serine proteases, they have not been included in the present version of the human and mouse degradomes.

Table S5 | Threonine proteases

Code	Peptidase	Human Gene	LocusLink	Locus	Mouse Gene	LocusLink	Locus	Syntenic	Identity
T01.010	proteasome catalytic subunit 1	PSMB6	5694	17p13	Psmb6	19175	11B4	y	97
T01.011	proteasome catalytic subunit 2	PSMB7	5695	9q33	Psmb7	19177	2B	y	96
T01.012	proteasome catalytic subunit 3	PSMB5	5693	14q11	Psmb5	19173	14C1	y	93
T01.013	proteasome catalytic subunit 1i	PSMB9	5698	6p21	Psmb9	16912	17B2	y	88
T01.014	proteasome catalytic subunit 2i	PSMB10	5699	16q23	Psmb10	19171	8D2	y	88
T01.015	proteasome catalytic subunit 3i	PSMB8	5696	6p21	Psmb8	16913	17B2	y	79
T01.016	proteasome β -subunit LMP7-like	LMP7L	122706	14q11	Lmp7l	73902	14C1	y	84
T01.986np	proteasome β -1 subunit	PSMB1	5689	6q27	Psmb1	19170	17A2	y	93
T01.984np	proteasome β -2 subunit	PSMB2	5690	1p34	Psmb2	26445	4D2	y	96
T01.983np	proteasome β -3 subunit	PSMB3	5691	17q12	Psmb3	26446	11D	y	98
T01.987np	proteasome β -4 subunit	PSMB4	5692	1q21	Psmb4	19172	3F2	y	93
T01.976np	proteasome α -1 subunit	PSMA1	5682	11p15	Psma1	26440	7F2	y	98
T01.972np	proteasome α -2 subunit	PSMA2	5683	7p14	Psma2	19166	13A2	y	99
T01.977np	proteasome α -3 subunit	PSMA3	5684	14q23	Psma3	19167	12C3	y	99
T01.973np	proteasome α -4 subunit	PSMA4	5685	15q24	Psma4	26441	9C	y	99
T01.975np	proteasome α -5 subunit	PSMA5	5686	1p13	Psma5	26442	3F3	y	99
T01.971np	proteasome α -6 subunit	PSMA6	5687	14q13	Psma6	26443	12C1	y	99
T01.974np	proteasome α -7 subunit	PSMA7	5688	20q13	Psma7	26444	2H4	y	99
T01.978np	proteasome α -8 subunit	PSMA8	143471	18q11	Psma8	73677	18A2	y	95
T02.001	glycosylasparaginase	AGA	175	4q34	Aga	11593	8B3	y	82
T02.003	glycosylasparaginase-2	ASRGL1	80150	11q12	Asrgl1	66514	19A	y	77
T02.004	glycosylasparaginase-3	AGA3	55617	20p12	Aga3	75812	2G2	y	94
T03.006	γ -glutamyltransferase 1	GGT1	2678	22q11	Ggtp	14598	10B5	y	79
T03.017	γ -glutamyltransferase-like 3	GGTL3	2686	20q11	Ggtl3	207182	2H2	y	95
T03.015	γ -glutamyltransferase 2	GGT2	2679	22q11					

T03.016	γ -glutamyltransferase m-3	GGTL4	91227	22q11					
T03.002	γ -glutamyltransferase 5	GGTLA1	220522	22q11					

The most recently identified catalytic class of proteases, the threonine proteases³⁰, are classified into three families: T01, containing the proteasome components; T02, composed of three distinct glycosylasparaginases; and T03, including diverse γ -glutamyltransferases (GGTs). All members of the T01 and T02 families are conserved between human and mouse. There are, however, some differences in the number of GGT genes clustered in a region of chromosome 22, which has undergone successive duplications³¹. As a consequence of this dynamic evolution, there are four GGT genes in this region of the human genome but only one in the corresponding region of the mouse genome (10B5). An additional GGT gene located at 20q11 is conserved in the mouse genome at an equivalent position (2H2).

1. Evers, M. P. *et al.* Nucleotide sequence comparison of five human pepsinogen A (PGA) genes: evolution of the PGA multigene family. *Genomics* **4**, 232–239 (1989).
2. Taggart, R. T., Mohandas, T. K., Shows, T. B. & Bell, G. I. Variable numbers of pepsinogen genes are located in the centromeric region of human chromosome 11 and determine the high-frequency electrophoretic polymorphism. *Proc. Natl Acad. Sci. USA* **82**, 6240–6244 (1985).
3. Chen, X., Rosenfeld, C. S., Roberts, R. M. & Green, J. A. An aspartic proteinase expressed in the yolk sac and neonatal stomach of the mouse. *Biol. Reprod.* **65**, 1092–1101 (2001).
4. Ord, T., Kolmer, M., Villems, R. & Saarma, M. Structure of the human genomic region homologous to the bovine prochymosin-encoding gene. *Gene* **91**, 241–246 (1990).
5. Krylov, D. M. & Koonin, E. V. A novel family of predicted retroviral-like aspartyl proteases with a possible key role in eukaryotic cell cycle control. *Curr. Biol.* **11**, 584 (2001).
6. Turner, G. *et al.* Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr. Biol.* **11**, 1531–1535 (2001).
7. Caputo, E., Manco, G., Mandrich, L. & Guardioli, J. A novel aspartyl proteinase from apocrine epithelia and breast tumors. *J. Biol. Chem.* **275**, 7935–7941 (2000).
8. Lee, J. J. *et al.* Autoproteolysis in hedgehog protein biogenesis. *Science* **266**, 1528–1537 (1994).
9. Gondo, Y. *et al.* Human megasatellite DNA RS447: copy-number polymorphisms and interspecies conservation. *Genomics* **54**, 39–49 (1998).
10. Okada, T. *et al.* Unstable transmission of the RS447 human megasatellite tandem repetitive sequence that contains the USP17 deubiquitinating enzyme gene. *Hum. Genet.* **110**, 302–313 (2002).
11. Zhu, Y., Carroll, M., Papa, F. R., Hochstrasser, M. & D'Andrea, A. D. DUB-1, a deubiquitinating enzyme with growth-suppressing activity. *Proc. Natl Acad. Sci. USA* **93**, 3275–3279 (1996).
12. Zhu, Y. *et al.* DUB-2 is a member of a novel family of cytokine-inducible deubiquitinating enzymes. *J. Biol. Chem.* **272**, 51–57 (1997).
13. Baek, K. H., Mondoux, M. A., Jaster, R., Fire-Levin, E. & D'Andrea, A. D. DUB-2A, a new member of the DUB subfamily of hematopoietic deubiquitinating enzymes. *Blood* **98**, 636–642 (2001).

14. Evans, P. C. *et al.* A novel type of deubiquitinating enzyme. *J. Biol. Chem.* (in the press).
15. Balakirev, M. Y., Tcherniuk, S. O., Jaquinod, M. & Chroboczek, J. Otubains: a new family of cysteine proteases in the ubiquitin pathway. *EMBO Rep.* **4**, 517–522 (2003).
16. Aravind, L. & Koonin, E. V. Classification of the caspase-hemoglobinase fold: detection of new families and implications for the origin of the eukaryotic separins. *Proteins* **46**, 355–367 (2002).
17. Gururajan, R. *et al.* Duplication of a genomic region containing the *Cdc2L1-2* and *MMP21-22* genes on human chromosome 1p36.3 and their linkage to *D1Z2*. *Genome Res.* **8**, 929–939 (1998).
18. Bertenshaw, G. P., Norcum, M. T. & Bond, J. S. Structure of homo- and hetero-oligomeric meprin metalloproteases: dimers, tetramers, and high molecular mass multimers. *J. Biol. Chem.* **278**, 2522–2532 (2003).
19. Seals, D. F. & Courtneidge, S. A. The ADAMs family of metalloproteases: multidomain proteins with multiple functions. *Genes Dev.* **17**, 7–30 (2003).
20. Wei, S. *et al.* Identification and characterization of three members of the human metallocarboxypeptidase gene family. *J. Biol. Chem.* **277**, 14954–14964 (2002).
21. Verma, R. *et al.* Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298**, 611–615 (2002).
22. Yao, T. & Cohen, R. E. A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* **419**, 403–407 (2002).
23. Cadiñanos, J. *et al.* Identification, functional expression and enzymatic analysis of two distinct CaaX proteases from *Caenorhabditis elegans*. *Biochem. J.* **370**, 1047–1054 (2003).
24. Pei, J. & Grishin, N. V. Type II CAAX prenyl endopeptidases belong to a novel superfamily of putative membrane-bound metalloproteases. *Trends Biochem. Sci.* **26**, 275–277 (2001).
25. Biagini, A. & Puigserver, A. Sequence analysis of the aminoacylase-1 family: a new proposed signature for metalloexopeptidases. *Comp. Biochem. Physiol. B* **128**, 469–481 (2001).
26. Makarova, K. S. & Grishin, N. V. The Zn-peptidase superfamily: functional convergence after evolutionary divergence. *J. Mol. Biol.* **292**, 11–17 (1999).
27. Hendrixson, D. R. *et al.* Human milk lactoferrin is a serine protease that cleaves Haemophilus surface proteins at arginine-rich sites. *Mol. Microbiol.* **47**, 607–617 (2003).
28. Quattrocchi, C. C. *et al.* Reelin is a serine protease of the extracellular matrix. *J. Biol. Chem.* **277**, 303–309 (2002).
29. Menoret, A., Li, Z., Niswonger, M. L., Altmeyer, A. & Srivastava, P. K. An endoplasmic reticulum protein implicated in chaperoning peptides to major histocompatibility of class I is an aminopeptidase. *J. Biol. Chem.* **276**, 33313–33318 (2001).
30. Seemuller, E. *et al.* Proteasome from *Thermoplasma acidophilum*: a threonine protease. *Science* **268**, 579–582 (1995).
31. Courtay, C., Heisterkamp, N., Siest, G. & Groffen, J. Expression of multiple γ -glutamyltransferase genes in man. *Biochem. J.* **297**, 503–508 (1994).