	UniprotKB ID	Entry name	organism	full name	oglcnacscore	oglcnac sites	phosphorylation sites	PMIDS	sequence	intracellular	extracellular	cytosol	nucleus	mitochndrion	endoplasmic reticulum	golgi apparatus	plasma membrane	extracellular region
0	Q01149	CO1A2_MOUSE	Mus musculus	Collagen alpha-2(I) chain	25.629171	NaN	NaN	36288343	MLSFVDTRTLLLLAVTSCLATCQYLQ SGSVRKGPTGDRGPRGQRGPAGPRG RDGVDGPMGPPGPPGSPGPPGSPAP PGLTGNFAAQYSDKGVSSGPGPMGL MGPRGPPGAVGAPGPQGFQGPAGE PGEPGQTGPAGPRGPAGSPGKAGED GHPGKPGRPGERGVVGPQGARGFP GTPGLPGFKGVKGHSGMDGLKGQP GAQGVKGEPGAPGENGTPGQAGAR GLPGERGRVGAPGPAGARGSDGSVG PVGPAGPIGSAGPPGFPGAPGPKGEL GPVGNPGPAGPAGPRGEVGLPGLSG PVGPPGNPGTNGLTGAKGATGLPGV AGAPGLPGPRGIPGPAGAAGATGAR GLVGEPGPAGSKGESGNKGEPGSVG AQGPPGPSGEEGKRGSPGEAGSAGP AGPPGLRGSPGSRGLPGADGRAGVM GPPGNRGSTGPAGIRGPNGDAGRPG EPGLMGPRGLPGSPGNVGPSGKEGP VGLPGIDGRPGPIGPAGPRGEAGNIG FPGPKGPSGDPGKPGERGHPGLAGA RGAPGPDGNNGAQGPPGPQGVQGG KGEQGPAGPPGFQGLPGPSGTTGEV GKPGERGLPGEFGLPGPAGPRGERG TPGESGAAGPSGPIGSRGPSGAPGPD GNKGEAGAVGAPGSAGASGPGGLPG ERGAAGIPGGKGEKGETGLRGDTGN TGRDGARGIPGAVGAPGPAGASGDR GEAGAAGPSGPAGPRGSPGERGEVG PAGPNGFAGPAGAAGQPGAKGEKGT KGPKGENGIVGPTGSVGAAGPSGPN GPPGPVGSRGDGGPPGMTGFPGAA GRTGPPGPSGIAGPPGPPGAAGKEGI RGPRGDQGPVGRTGETGASGPPGFV GEKGPSGEPGTAGAPGTAGPQGLLG APGILGLPGSRGERGLPGIAGALGEP GPLGISGPPGARGPPGAVGSPGVNG APGEAGRDGNPGSDGPPGRDGQPG HKGERGYPGSIGPTGAAGAPGPHGS VGPAGKHGNRGEPGPAGSVGPVGA VGPRGPSGPQGIRGDKGEPGDKGHR GLPGLKGYSGLQGLPGLAGLHGDQG APGPVGPAGPRGPAGPSGPVGKDGR SGQPGPVGPAGVRGSQGSQGPAGPP GPPGPPGPPGVSGGGYDFGFEGDFY RADQPRSQPSLRPKDYEVDATLKSL NNQIETLLTPEGSRKNPARTCRDLR LSHPEWNSDYYWIDPNQGCTMDAI KVYCDFSTGETCIQAQPVNTPAKNS YSRAQANKHVWLGETINGGSQFEY NVEGVSSKEMATQLAFMRLLANRA SQNITYHCKNSIAYLDEETGSLNKAV LLQGSNDVELVAEGNSRFTYSVLVD GCSKKTNEWGKTIIEYKTNKPSRLPF LDIAPLDIGGADQEFRVEVGPVCFK	False	True	2.054	2.695	2.376	3.385	1.369	3.085	5.0