

Supplementary Data: Supplementary Tables and Figures

Comparative genomics of monotremes provides insights into the early evolution of mammalian epidermal differentiation genes

Julia Steinbinder, Attila Placido Sachslehner, Karin Brigit Holthaus, Leopold Eckhart

Content

Supplementary Tables S1-S5

Supplementary Figures S1-S9

Supplementary Table S1. Platypus (*Ornithorhynchus anatinus*) EDC genes

Gene name	Genomic DNA accession nr.	CDS start	CDS end	Sequence complete	Expression confirmed by RNA-seq	GenBank Gene ID*	Notes
S100A9	NC_041753.1	1462198	1460138	yes	yes	114808162	intron interrupts CDS: see GenBank
PGLYRP3	NC_041753.1	1471001	1479972	yes	no	114808147	introns interrupt CDS: see GenBank
LOR	NC_041753.1	1489992	1489390	yes	yes	114808056	
PRR9	NC_041753.1	1515784	1515245	yes	no	120638162	
SPRR9L	NC_041753.1	1521102	1520755	yes	yes	114808177	record withdrawn from GenBank
SPPR2EL	NC_041753.1	1533529	1534337	yes	no	120638142	
SPRR3L	NC_041753.1	1535878	1535081	yes	yes	n. a.	not predicted by GenBank
SPRR2_1	NC_041753.1	1549573	1549364	yes	yes	n. a.	not predicted by GenBank
SPRR2_2	NC_041753.1	1558052	1558297	yes	yes	n. a.	not predicted by GenBank
SPRR2_3	NC_041753.1	1570427	1570236	yes	yes	114808057	differently predicted by GenBank
SPRR1BL	NC_041753.1	1580196	1579939	yes	yes	114808057	differently predicted by GenBank
SPRR1B	NC_041753.1	1593099	1592749	yes	yes	114808058	differently predicted by GenBank
IVL	NC_041753.1	1604635	1604015	yes	no	n. a.	not predicted by GenBank
SPRRL	NC_041753.1	1611433	1611188	yes	yes	n. a.	not predicted by GenBank
LCE1L	NC_041753.1	1619089	1618679	yes	yes	n. a.	not predicted by GenBank
KPRP	NC_041753.1	1636928	1638505	yes	yes	114808060	
KPLCE	NC_041753.1	1647438	1646704	yes	yes	100681800	
LCE2L1	NC_041753.1	1655080	1654703	yes	yes	114807948	differently predicted by GenBank
LCE2L2	NC_041753.1	1661991	1662368	yes	yes	100089382	differently predicted by GenBank
LCE2AL	NC_041753.1	1668467	1668096	yes	yes	114807964	
CRNN	NC_041753.1	1695432	1701191	yes	no	n. a.	intron: 1695570-1697501 not predicted by GenBank
SFTP1	NC_041753.1	1716442	1725763	yes	yes	n. a.	intron: 1716580-1717441 not predicted by GenBank
RPTN	NC_041753.1	1748434	1748571	no	no	n. a.	pseudogene; premature stop codon in first coding exon; differently predicted by GenBank
TCHH	NC_041753.1	1773891	1773965	no	no	n. a.	pseudogene; premature stop codon in first coding exon; differently predicted by GenBank
TCHHL1	NC_041753.1	1786007	1786869	yes	no	n. a.	intron: 1786145-1786401 not predicted by GenBank
S100A11	NC_041753.1	1814721	1818318	yes	yes	100089848	intron interrupts CDS: see GenBank

Notes: "Expression confirmed by RNA-seq" refers to RNA-seq peaks shown in the section "Genomic regions, transcripts, and products" of the Gene view in NCBI GenBank. CDS, coding sequence (presence of introns between CDS start and end is indicated in the Notes column); n. a., not applicable. *, GenBank ID refers to the most similar gene prediction in GenBank. Differences between gene predictions in this study and in GenBank are indicated in the column "Notes".

Supplementary Table S2. Australian Echidna (*Tachyglossus aculeatus*) EDC genes

Gene name	Genomic DNA accession nr.	CDS start	CDS end	Sequence complete	Expression confirmed by RNA-seq	GenBank Gene ID*	Notes
S100A9	NC_052096.1	10922408	10925241	yes	yes	119947338	intron interrupts CDS: see GenBank
PGLYRP3	NC_052096.1	10913655	10902775	yes	no	119947327	introns interrupt CDS: see GenBank
Loricrin	NC_052096.1	10891662	10892564	yes	no	n. a.	not predicted by GenBank
PRR9	NC_052096.1	10858457	10859044	yes	no	n. a.	not predicted by GenBank
SPRR2HL	NC_052096.1	10852258	10852605	yes	no	119947002	differently predicted by GenBank
SPRR2BL1	NC_052096.1	10845371	10845147	yes	no	119947001	differently predicted by GenBank
SPRR2BL2	NC_052096.1	10838431	10838207	yes	no	n. a.	not predicted by GenBank
SPRR2BL3	NC_052096.1	10833395	10833589	yes	no	n. a.	not predicted by GenBank
SPRR2BL4	NC_052096.1	10830687	10830487	yes	no	n. a.	not predicted by GenBank
SPRR3L	NC_052096.1	10828172	10829452	yes	no	n. a.	not predicted by GenBank
SPRR2_1	NC_052096.1	10815743	10815988	yes	no	n. a.	not predicted by GenBank
SPRR2_2	NC_052096.1	10808129	10807854	yes	no	n. a.	not predicted by GenBank
SPRR2_3	NC_052096.1	10799930	10800124	yes	no	n. a.	not predicted by GenBank
SPRR2_4	NC_052096.1	10788457	10788666	yes	no	n. a.	not predicted by GenBank
IVL	NC_052096.1	10763197	10763961	yes	no	119947000	differently predicted by GenBank
SPRRL	NC_052096.1	10755731	10755976	yes	no	n. a.	not predicted by GenBank
LCE1L	NC_052096.1	10745514	10745840	yes	no	n. a.	not predicted by GenBank
KPRP	NC_052096.1	10726118	10724499	yes	no	n. a.	not predicted by GenBank
KPLCE	NC_052096.1	10712308	10713177	yes	no	119946852	
LCE2DL1	NC_052096.1	10703387	10703689	yes	no	119946862	
LCE2DL2	NC_052096.1	10690910	10690647	yes	no	119946821	
LCE2DL3	NC_052096.1	10682762	10682472	yes	no	119946817	
LCE2DL4	NC_052096.1	10675079	10674789	yes	no	119946818	
LCE2DL5	NC_052096.1	10667406	10667092	yes	no	119946815	
LCE2DL6	NC_052096.1	10662950	10662612	yes	no	119946819	
LCE2DL7	NC_052096.1	10658456	10658118	yes	no	119946816	
LCE2DL8	NC_052096.1	10651278	10650964	yes	no	119946857	
SFTP1	NC_052096.1	10598727	10581597	yes	no	119946998	intron: 10598589-10597725 differently predicted by GenBank
SFTP2	NC_052096.2	10568140	10556308	yes	no	n. a.	intron: 10568002-10567141 not predicted by GenBank
SFTP3	NC_052096.2	10542843	10533536	yes	no	n. a.	intron: 10542705-10541843 not predicted by GenBank
RPTN	NC_052096.1	10510100	10505629	yes	no	119946997	intron: 10509962-10508908
TCHH	NC_052096.1	10481535	10474690	yes	no	119946996	intron: 10481397-10479514 differently predicted by GenBank
S100A11	NC_052096.1	10439882	10435252	yes	yes	119947258	intron interrupts CDS: see GenBank

Notes: "Expression confirmed by RNA-seq" refers to RNA-seq peaks shown in the section "Genomic regions, transcripts, and products" of the Gene view in NCBI GenBank. CDS, coding sequence (presence of introns between CDS start and end is indicated in the Notes column); n. a., not applicable. *, GenBank ID refers to the most similar gene prediction in GenBank. Differences between gene predictions in this study and in GenBank are indicated in the column "Notes".

Supplementary Table S3. Opossum (*Monodelphis domestica*) EDC genes

Gene name	Genomic DNA accession nr.	CDS start	CDS end	Sequence complete	Expression, RNA-seq	GenBank Gene ID*	Notes
S100A9	NC_008802.1	187669498	187671576	yes	no	100019177	intron interrupts CDS: GenBank
PGLYRP3	NC_008802.1	187632813	187614636	no	no	100019140	introns interrupt CDS: GenBank
LOR	NC_008802.1	187574345	187575694	yes	no	100019077	differently predicted by GenBank
PRR9	NC_008802.1	187478785	187479243	yes	no	n. a.	not predicted by GenBank
LELP1	NC_008802.1	187457476	187457757	yes	no	103097902	record withdrawn from GenBank
SPRR2EL	NC_008802.1	187191344	187192111	yes	no	107651172	pseudogene
SPRR1AL1	NC_008802.1	187164920	187165447	yes	yes	103097554	differently predicted by GenBank
SPRR2L1	NC_008802.1	187147131	187147334	yes	no	n. a.	not predicted by GenBank
SPRR2L2	NC_008802.1	187133443	187133892	yes	no	n. a.	not predicted by GenBank
SPRR2L3	NC_008802.1	187124163	187124564	yes	no	n. a.	not predicted by GenBank
SPRR2L4	NC_008802.1	187110819	187111073	yes	no	n. a.	not predicted by GenBank
SPRR2L5	NC_008802.1	187102044	187101676	yes	no	n. a.	not predicted by GenBank
SPRR1AL2	NC_008802.1	187095883	187096371	yes	no	n. a.	not predicted by GenBank
SPRR1BL1	NC_008802.1	187000467	187000748	yes	yes	107651345	
SPRR1AL3	NC_008802.1	186984304	186984669	yes	yes	100018715	
SPRR1L	NC_008802.1	186942943	186942530	yes	yes	100018686	
SPRR1BL2	NC_008802.1	186906790	186907146	yes	yes	100018648	
SPRR4	NC_008802.1	186884536	186884775	yes	yes	100018614	
IVL	NC_008802.1	186796173	186797150	yes	yes	103096090	
LCE1F1	NC_008802.1	186746776	186747108	yes	yes	103095325	
LCE1F2	NC_008802.1	186684922	186685410	yes	no	130456908	
KPRP	NC_008802.1	186660356	186662080	yes	yes	107651344	
LCE3L1	NC_008802.1	186643224	186642907	yes	no	n. a.	not predicted by GenBank
KPLCE	NC_008802.1	186622089	186623045	yes	no	100618972	
LCE5AL1	NC_008802.1	498496862	498497263	yes	no	n. a.	record withdrawn from GenBank
LCE1EL1	NC_008802.1	498464042	498464464	yes	no	103098023	record withdrawn from GenBank
LCE1EL2	NC_008802.1	498453465	498453890	yes	no	n. a.	record withdrawn from GenBank
LCE5AL2	NC_008802.1	498423442	498423615	yes	no	130457020	pseudogene
LCE1EL3	NC_008802.1	498413017	498413439	yes	yes	130457019	
LCE2AL1	NC_008802.1	498402949	498403392	yes	yes	103097443	
LCE2DL	NC_008802.1	498393418	498393834	yes	yes	103097328	
LCE2CL	NC_008802.1	498385673	498386089	yes	yes	103097215	
LCE5AL3	NC_008802.1	498358345	498358728	yes	yes	103096878	
LCE1EL4	NC_008802.1	498349205	498349609	yes	yes	107651024	
LCE2AL2	NC_008802.1	498322733	498323047	yes	no	103096569	record withdrawn from GenBank
LCE3L2	NC_008802.1	498287133	498287444	yes	no	n. a.	not predicted by GenBank
LCE3L3	NC_008802.1	498273489	498273800	yes	no	n. a.	not predicted by GenBank
LCE3L4	NC_008802.1	498253326	498253024	yes	no	n. a.	not predicted by GenBank
LCE3L5	NC_008802.1	498244429	498244740	yes	no	n. a.	not predicted by GenBank
LCE3L6	NC_008802.1	498235136	498234834	yes	no	n. a.	not predicted by GenBank
LCE3L7	NC_008802.1	498230068	498230379	yes	no	n. a.	not predicted by GenBank
LCE3CL	NC_008802.1	498221020	498220718	yes	yes	103096160	
CRNN	NC_008802.1	498013410	498009667	yes	no	103095863	intron: 498013272-498012118
FLG	NC_008802.1	497947213	497939263	yes	yes	103106047	intron: 497947075-497945758
							differently predicted by GenBank
FLG2	NC_008802.1	497812400	497808648	yes	yes	103105843	intron: 497812262-497810556
HRNRL1	NC_008802.1	497763900	497756035	yes	no	100019988	intron: 497763762-497761933
HRNRL2	NC_008802.1	497672244	497660012	yes	no	103105680	intron: 497672106-497670182
							differently predicted by GenBank
HRNRL3	NC_008802.1	497629422	497621790	yes	no	103093607	intron: 497629284-497627574
HRNR	NC_008802.1	497589690	497585535	yes	no	103093297	intron: 497589552-497588469
RPTN	NC_008802.1	497488086	497484078	yes	no	100019364	intron: 497487948-497486631
TCHH	NC_008802.1	497447799	497443415	yes	yes	100019329	intron: 497447661-497445275
							differently predicted by GenBank
TCHHL1	NC_008802.1	497398205	497394602	yes	yes	100617572	intron: 497398067-497397050
TCHHL2	NC_008802.1	497363668	497360971	yes	yes	103105566	intron: 497363530-497361832
							record withdrawn in GenBank
S100A11	NC_008802.1	497330944	497322695	yes	yes	100016326	intron interrupts CDS: GenBank

Notes: "Expression confirmed by RNA-seq" refers to RNA-seq peaks shown in the section "Genomic regions, transcripts, and products" of the Gene view in NCBI GenBank. CDS, coding sequence (presence of introns between CDS start and end is indicated in the Notes column); n. a., not applicable. *, GenBank ID refers to the most similar gene prediction in GenBank. Differences are indicated in the column "Notes".

Supplementary Table S4. CRCT1, LELP1 and SMCP genes

Gene name	GenBank Gene ID	Genomic DNA accession nr.	Gene start	Gene end
Hs_CRCT1	54544	NC_000001.11	152514482	152516008
Cd_CRCT1	119517646	NC_051308.1	113824553	113822998
Sh_CRCT1	n. a.	NC_045429.1	427947250	427947615
Hs_LELP1	149018	NC_000001.11	153203430	153205120
Em_LELP1	126073628	NC_064821.1	189252129	189252434
Vu_LELP1	114049637	NW_020954576.1	2161697	2161996
Hs_SMCP	4184	NC_000001.11	152878322	152885047
Cd_SMCP	n. a.	NC_051308.1	113363067	113362813

Notes: n. a., not applicable; Hs, *Homo sapiens*; Cd, *Choloepus didactylus*; Sh, *Sarcophilus harrisii*; Em, *Elephas maximus*; Vu, *Vombatus ursinus*

Supplementary Table S5. CASP14-like genes of the Australian echidna (*Tachyglossus aculeatus*)

Gene name	GenBank Gene name	Genomic DNA accession nr.	Gene start	Gene end	GenBank Gene ID	Note
CASP14L1	LOC119923804	NW_024044983.1	34094	39458	119923804	
CASP14L2	LOC119922773	NW_024044852.1	891398	886058	119922773	
CASP14L3	LOC119923380	NW_024044901.1	298412	303799	119923380	
CASP14L4	LOC119923953	NW_024045091.1	102697	110405	119923953	
CASP14L5P	LOC119923317	NW_024044893.1	267752	259027	119923317	pseudogene
CASP14L6P	LOC119922765	NW_024044852.1	1065150	1073209	119922765	pseudogene

Notes: The numbers under Gene start and Gene end indicate nucleotide positions corresponding to the start of the first exon and the end of the last exon, as provided in GenBank (last accessed on 27 November 2023).

>Oa LOR (XP 028911301.1)

MSYQQHQTTQPTHTIPFVCGGGCSGGGGGGGGSSGGSCGGCCGGGSSGGSSGGGGYFPSQQQTQPTQGNLPSGGGGGGGGGGCGG
SSGGGKASGGCSWSGGGVVIGGGGGKSSGGGCSSGGCGGSSSGSSGGCGGGGCSYTQQTYYVPTVQTSSGGCSCGGSSSGGSS
SGGCSSGGSSSGGCSSGGSSAFQTQKQPTSWPTK

>Oa_PRR9 (XP_039766785.1)

MFHAQQRKQFCLPPPCLOQQQERLKSQEACPPPAKKEQEPAKQDLCPPLCQEQGPAKCPACQEQGFPVKCQDLCPFVCQEQ
GPAKQDLCPFCQEQGPAKQDLCPFCQEQGPAKCPDLCPPVCLDQGPAKQDLCPFVCLEQGPAKQDLCPFECKEQGSV
KCQDLCPFKCQCK

>Oa SPRR9L

MAEDKDHS**SS**CPFKFTCPATPK**CS**EF**SK**V**SN**CLRK**LL**GG**CS**AF**S**PPPKCPPRD**PP**K**CF**HK**CP**PK**E**PPK**CP**PK**E**PPK**CL**PK**Q**PK**E**
PK**CL**PK**Q**PAE**PP**K**CP**PK**E**PK**CP**PPDKAEN**F**

>Oa SPPR2EL (XP_039766675.1)

MSYQDQQQV**KQ**PCPPPP**I**PQPCPPPVQCPQPCPPPVQCPQPCPPPVQCPQPCPPPAQN**PQ**PC**S**PTCPEPPVPWQQQE

>0a SPRR3L

MSFQHYQCKQVCQPPPCQPKCPTSSPCPRHPEPYQPNFESCHPGYPEFSQPRYPEPSYERYPEPTYPKYEQPYQPRYF
EQNYPKHPEPCQERYPEPCPRSENRYPEPCHERHPQCPCPRSEPCPRHPEFCHPRHPELQCPERSPEPCPSRHPPECHPR
RFELCPERSPEPCPRHQEPCPERSFVSGQPHFPEPCPERSPESSPSGRHRHREPKYPAQCSFECFLQCPERPSQFQQWKCTP
VQFSPPCCQGKFFQEYK

>0a SPRR2 1

M S Q Q Y Q Q Q T K Q P C P P P V I K E P C P S P P V I K D P C H P K T P E P C H P K T S T P C P H P S Q Q K A P P T Q Q V P Q T K Q K

>0a SPRR2 2

MSQQYQQQT**KQ**PCPPPPVIKEPCPPPPVVKEPC**L**PPPVVKEPCPPAPVVKEPCPPKNPTPCPP**T**SQQKVPPVQQVPQT**KQ**K

>0a SPRR2 3

M S Q Q A K Q P T Q V P P K E P C H P K T P A P T P C H P K T P E P C H P K T P T P S P A P T Q Q K A P P A Q Q V P K T K Q K

>0a SPRR1BL

MSQQHHQCPPETQQT**KQPCQPPEKCEPCQA**K**CPPPAHQCCEFCQPKCPPFAAQKCCEPCQPKCPPFAAQKCCEFCQPKCPP**
PK

>Oa SPRR1B

MSQQQQQPCPFLQQTKQPCQPPPKCPEFCHPKCPPPAQQKCFQPCQPKCPPEAQQKCFEPCQPKCPPPAQQKCFEFCQPKCPP
FAQQKCFEPCQPKCPPPAQQKCPFFAAGCQKSK

>0a IVL

MSQHVQETCQQQRKQPTVLPTTFRFPCPPQMESQLKVPQKQSSQLPGKTQVKSPNPVQVSDFHLDQKQKKVNLKDDQH
LEKERQQLKELEEKKKEQLKQLEEQLKELEQKEPKDDHLEKKQQLLAKELEDKKGGKQQLLEEQLKEKELSLKDRKDAQHL
EKKQQLLEKELEDKKERLKQQLLEEQLAKELEQKDAKDAQHLEKETQLLEKEMKKKEQLKKQLEEQLAKELEQKEFKYDQLE
KEKOLLEKEMKKKEQLKKQLEEQLKELEQKGAQDAQQLKKELLGKGPGEKKQLE

>0a SPRRL

MADHQQRKQQNLPPPVSCPEPQFPQPKCFEPQPKCFEPQPKCFEPQPKCLEPQPKCFEPQPKCFEPQPNCPSPAGGQSKRPQ

>0a LCE1L

MSQQQSQQQCPPPPKCLPSPKVTETSPAPCPAPAPAPAPSCCGSGSGGGCCGDSNSSGGCCGSGGGCCGNSGGCCGSGGGCCGNSSG
CGCGSGGGCCSSSSGGCCLTSPFHRRRRRRRQRSSCGGGQRDQFCEESGSGCCC

>Oa KPRP (XP 028911306.1)

MCDHQQLQPCPFVTVCCVKGLGLGFSADQGRKIFCSAPGSAQAQAFFTLCSSSQVSSTSQSRVSLTPGQDSSVGYQCPQAG
 YVMCQSSSQVSSTSQGDSSAGCQPFCEARTGYETCQSSSSAATFDIQDSSLIAQFQGGFALASVYVHCSFVYVKETYVYVCAI
 PCFVHTYYVECFVQTSYVQCEALCQSQTSYVRCFAFCQFQASARSAAALCQFQAFAGSHFALNQFQAPARSRALDQFQVPPER
 RCPGLRPAQSCQWCPPLRFRSASTVLPAFAPGASLGLCGPPRCALQGPWRRCCFKYRAEICSENYFQQVPPERPFVRIPIRR
 SVSCSDQQFVLDLGFSCRVRFPFPPGRSRSWDDLAGDTSDLSEFKARGSGFGRSSGSGCAREEFQSGNSGCCQRRGECRG
 SGNSSGGNGGEGCGSSNGCGRRGVEQEAENS SGGSSGECQSGSNSGGSSGGEYQCSENSGCCGEEARGLGNSSGSSCRGED
 RLIPENSECLFEDCKKPANNNCEKQRF

>Oa KPLCE (XP 028910439.1)

MCDQQKQSQTLPFHLVKGVGFSVYQVTKGEAENIEALVKCPFERETQTYVKCPFFCPTQTILKCPFFCPIQTYVKCPFFCVTQ
 TYMKCAPPGFQNIIVKCPFFCQTQTCTYVQCFAFCFPFKCPPFCFIQKCYVQCFAFCQAPTYYIQRRQTYVVPQREAWRPAPQC
 CPPFCPRPAALSCSSSSSSCCNLAFRSFGIRLRRWVRGPECCDNSGGYDECEDESGCCGLGIIPMRASGFACCEDEC

[illegible]

MSCQQNQKQC**PPPH**QCPQQC**PPEKCPPKC**PK**CPKCP**PK**CPAFCFAPASCCSSSGGCCSSSGGCCSSSGGCCSSSGG**
CCGSSSGGGGCCLSHHHHRRPLRFHRRRHQSPECCDDNSGCC

MSQQNQNGPPPPQCPQQCQPPEKCPPKCPPKCPKCKPCKCAPFCFAPASCCSSSGGGCCSSSGGCCSSSGGGGCCLSHH
RERRSHRLRHRSROECEDSSGSCCHGSDSGCCHCSGSGCC

B

M^HQ^LL^GLNIV^GI^IQAFN^SYARTE^GD^CTTL^SR^GEL^{KK}LIEREF^AE^VI^VK^HYD^PETV^DTV^LHL^LDD^DAD^GK^VG^FTE^FL^AL^VFR^VQ^A
 A^CYN^LQ^I^S^G^G^G^AC^GM^GARA^QA^PT^GQ^OT^GG^AV^RT^RQ^T^G^SAAT^G^S^S^S^G^TT^V^S^G^RT^G^S^S^VT^Q^VT^QA^Q^Q^G^S^G^RT^V^G^T^G^S^T^T^V^T^S^T
^P^S^L^T^T^T^A^A^T^S^T^Q^S^G^T^C^Q^T^G^T^S^G^VA^A^R^T^S^S^G^Q^T^S^G^T^V^R^D^Q^R^S^S^N^Q^Q^T^G^T^S^T^T^V^T^G^G^Q^T^T^T^S^T^T^T^V^T^G^S^RA^A^T^S^G^S^Q^V^T^Q^G^S^G^S
^T^G^G^T^T^T^V^T^A^T^G^S^P^T^Q^K^T^T^T^T^T^T^T^G^G^Q^T^S^T^G^A^T^R^D^P^T^T^T^T^V^T^T^G^Q^T^A^S^T^A^S^R^N^P^T^T^T^M^V^T^G^G^Q^T^T^T^T^V^T^R^G^P^T^T^T^T^A^T^T^G^P^T^A
^P^T^A^T^T^G^G^Q^T^T^T^T^T^V^T^R^G^R^S^G^T^C^Q^A^G^T^P^V^T^Q^A^R^Q^A^S^S^S^Q^T^G^G^P^A^A^A^T^V^T^V^T^R^T^G^T^P^S^Q^A^Q^T^T^T^T^T^G^G^P^T^T^T^A^T^G^G^Q^T^S^T^T^T^T^R
^Q^S^S^S^T^T^T^G^S^P^T^T^A^T^V^S^P^A^A^T^G^T^Q^T^S^T^T^A^T^T^R^P^A^T^T^A^T^T^G^G^Q^T^T^T^A^T^A^G^P^T^T^T^V^T^R^G^S^Q^T^T^T^T^G^S^T^T^G^T^T^D^Q^T^S^T^T^A^T^T^R^A
^T^T^A^T^T^G^G^S^G^S^G^T^C^Q^T^G^T^S^G^S^Q^V^T^A^Q^Q^S^S^S^Q^T^G^G^VA^A^T^V^T^V^T^R^G^T^P^S^Q^T^Q^T^T^G^G^T^S^T^A^V^T^G^R^T^S^T^A^A^T^G^G^Q^T^T^A^T^G
^P^T^S^T^A^T^V^S^Q^T^S^T^A^T^T^G^T^S^T^A^T^T^R^P^A^T^T^A^T^G^G^Q^S^S^T^A^T^A^P^A^T^T^T^A^T^G^S^Q^T^S^T^A^T^R^G^T^T^T^A^T^G^G^Q^G^T^C^Q^T^G^S^E
^Q^V^T^Q^T^Q^Q^S^S^S^Q^T^G^G^P^A^A^T^V^T^V^T^R^T^G^T^P^S^Q^T^Q^S^T^A^T^S^T^G^G^Q^T^S^T^A^V^T^G^G^R^T^T^T^A^A^T^G^G^Q^T^T^T^A^T^G^G^Q^T^A^T^T^A^T^G^G^Q^T^T^T^A
^T^G^T^P^T^T^T^A^T^G^S^Q^T^S^T^T^G^T^T^Q^T^S^T^T^T^T^R^P^A^T^T^A^T^G^G^Q^A^S^T^S^G^S^Q^V^T^Q^T^Q^S^S^S^Q^T^G^G^P^A^A^T^V^T^V^T^R^T^G^T^P^S^Q^T^Q^S^T^A
^S^T^G^G^Q^T^S^T^A^V^T^R^G^R^T^T^A^A^T^G^G^Q^T^T^T^A^T^E^G^Q^T^A^T^A^T^G^G^Q^T^T^T^A^T^A^G^P^T^T^T^V^T^G^S^Q^T^S^T^T^G^T^T^G^Q^T^S^T^T^A^T^T^R^P^A

MSQLLKSIIVTVIDIFYHYTQGDGCELTSLKELKELLEKEFRPILKHNDEPDTVDVIMQILDRDHRRVDFTEFLLMVFKLTQ
AONKVVSKDFCFASGSKQKSGRHTWQGHENELKQEDDEEGEEEEEEDGEEEAQKDGQSSWSAGEEWSEFSGGTVKHGHK
GGFEHTARHGHCSSGKQKHKSQGGKHVGEKQGSKKRGSRRSRRTCSLTRQDSEHSECGGGQGVSWGFEGGEECPSSSSQSEACG
KQGGQSSKQGSTSCGGHSSSGGCGSGGCGGQHSGSAGRRSGGCGGQGHSGSAGRRSGGCGGQGHSGCAGGQSSGACGCGQHQS
CSGGHSEYEEQSSSSGRQASGHGRGFSSGQCSSEGVKSGSRGCASTRGGGCGQSTYRGHGHFSGRCQSGYEEHWSSSGQS
TCYGGSSSSSQSSYGGHSGSGSEFTRHGGYTSSSGQSTSYGCGSGSGGQSSYGGHSGSGCKFTSSGGHGTSSSGQS
SYGGKCSGSGGQSSHGGHSGSGGEFTSQGGYGTSSGQSTSHGCGSGSGGQSSHGGQASGSGGCKFTSQGGYGTSSGQSTS
HGGQCSGSGGQSSHGGQASGSGGCKFTSQGGHGTSSGQSTSYGCKCSGSGGQSSYGGHRSSSGGEFTSYGGHGTSSGQATSH
GGQSSSSGGHSGCGGQESTSGGQSGGHWTTRSWHSWSDGSDGGRHNNQSWTSESEGTQSSGSESDFEQACKSWTQTQSG
RGHSSESSTTERRCGHHYQTWDSRRCQSSQSAESTHGQTGSTHGQTTDSYQQGTSYEQGTYSHEQTGSSHGQTGSYQQTSG
NYGQTGISHGQSSFSTGERARHCSGQSESERSWYSTTGRDSSPQYEGEETHYASSYQGTGTHGQTADSYQGTCSRHRQT
CSHHRQSSSHGQSRDSYQGTGTHGQTEYTHQTSSGQQTDSYRQTCSHHRQTCSHGGTSSYQQTDSYQGTGTHGQ
SCYSHGQTSSHGQTSSHGQTSQYQAAGTSQGTSTHRTQDTSYQQGTRQGTSSHGQASSHGQTRDTYRQTSSHG
QTSSHGQTSSYQQTDSYQQGTSHRQSGYSQGTSSHGQTCSHHEQTGGRYCQRGTSSGQTGSTHGQTTDSYQQGSSQ
QGTSSHRQAGSSHGQTRDSYRQTSSHGQICSSHGQTSSYQQTDSYQQGTSHRQSGYSQGTSSHGQTSHGQTTDS
YQAAGTSQGTGSTHGQTTDSYQQGTRQGTSSHGQAGSSHGQTRDTYRQTSSHGQTSSHGQTSSYQQTDSYQQGT
SHRQSGYSQGTSSHGQTSSHEQTGGRYCQRGTSSGQTSHGQTTDSYQQGSSGQGTSSHGQAGSSHGQTRDSYRQT
SSHGQTSSHGQTSSYQQTDSYQQGTSHRQSGYSQGTSSHGQTSHGQTTDSYQQGTRQGTSSHGQAGSSHGQTRDTYRQTSSHG
CQTRQGTSSHGQAGSSHGQSRDSYRQTSSHGQTSSHGQTSSYQQTDSYQQGTSHRQSGYSQGTSSHGQTSSHEQ
ARCSGGRGTSSGQGTSTHGQTTDSYQQGTRQGTSSGQGTSSHGQTRDSYRQTSSHGQTESHGQTGSTHGQTTDSY
QQGTSSSSGYSHGQTSSHGQICSSHGQTRDSYQAAGTSQGTGSTHGQTTDSYQQGSSGQGTSSHEQAGSRHGQTRDSY
RQTSSRGQTESHGQTGSTHGQTTDSYQQGTSSHSQSGYSHQAASSHGQTRSGYQAAGTSHGQSSSTHGQTTDSYRQGT
QCGTSSHGQAASSHGQSRDSYRQTSSHGQTSSHGQTSSYQQTGSYQGTGISHGERTGRHSGCQSSSGRSGYSTTRSD
SSGQYEGEETHYASSYQGTGTHGQTTDSYQQGTSSSSGSSHGQNRDSYQGTGTHGQTTSSHGQTRGTYQQGTSSHG
CSHGQNRDSYQGTGTHGQTTSSHGQNRDSYQGTGTHGQTTSSHGQNRDSYQGTGTHGQTTSSHGQTRGTYQQGTSSHG
SSHGQTRDSYQGTGTHGQTTSSHGQNRDSYQGTGTHGQTTSSHGQNRDSYQGTGTHGQTTSSHGQNRDSYQGTGTHGQ
QTSSHGQTRGTYQQGTSSHGQSSSSHGQTRDSYQGTGTHGQSSSSHGQTRDSYQGTGTHGQSSSSHGQTRDSYQGTGTH
GQTSSHGQTTDSYQGTGTHGQSSSSHGQTRDSYQGTGTHGQSSSSHGQTRDSYQGTGTHGQSSSSTGERARHSGCQSSSGT
TTGSRDSSSGQYEGEETHYASSYQGTGTHGQTTADSYQGTSSRNRQTSSHRQSSSSHGQSSDSYQGTGTSQGSSSHEQAR
DSYQTRTSHGQTSSYQGTDSHGQSSSGSQTGSTHGQTSSHRQTSSGYSQSSSYQGTSYTSSQSSDWSGSGTSYVQT
GGRGRQRQRSSWRHSGYNLAYDSESGFGSGHRTGRNSSSGSGSRTGRQSSHGQSSDTYGQSSSSTNGQSSSY
CGSSDTYGQSSSSTNGQASYSQSSDTYGQSSSNTKQSSSYQSSDTYGQSSSNTKQSSHGQSSDTYGQSSSNTKQSS

QGS^QGH^QSTD^TNG^QSG^SASE^SSK^QSS^HQSS^DTS^QSW^SGH^WQSS^QSW^SQSS^FFG^QTE^SSS^SGWT^GTR^SQTE^HQAE^FSIL^SRG
SR^SSTR^TPT^SQSG^GRRR^QE^SSS^SLL^SASE^GST^SPLLLI^AGP^QRR^GSES^SEN^FYIRVRTV^SCS^FY^SSS^TPLYEYI^QEQ^RGLY

>Oa_TCHHL1

MP^RLLK^SSVTDVMEVFH^KYA^QEDGG^QAVLT^KDGLR^QLLR^SELGDIL^QRS^SDGRYV^GSEVDLLDANHD^GVIDFNEFILLVF^GLLN^G
AC^YLDIR^SLVRS^KSAH^QRD^EEE^PK^GGL^EGNRRR^ELYREE^GEEEEDEE^GRGWYR^QQSDAA^SSRNWLTE^KE^GKAY^SASLD^GQGM^S
AQ^KSN^FVAH^EEP^QKMM^GKA^QSE^FIC^GQR^ENN^SSE^RHR^S

C

>Oa_PGLYRP3

MLRLVVFLAV^GVRF^SASV^PPAMT^SHA^QQ^FSTA^HENATDTLARFETLL^GCFRDVF^QDP^TIT^FRAEW^GAQA^RCTA^PLK^TTF^S
YLLVHHIAG^TDC^GAQ^GSS^WCLRQLD^HHTLTNG^WCDIAYNFLIME^SGEVFEG^TGWTV^QGHHTAGYNEVAL^GFAFFT^NMTDRAP^S
SQ^AALAS^AQH^LIS^FAV^QKR^HLS^PNYI^QEF^LFRGED^CLQA^FGS^TES^GVL^CPT^IIPRADW^GAK^GSMAN^CRK^LDRPA^KYV^IIHT^S
AG^QFC^TELDS^KCELVR^GIQ^DFHVNG^RK^FCDVGYNFLV^GEDGNVYE^GVGW^DTEGAHTY^GYNDIALG^VAF^LGLFED^KPPNAAALM^S
AAQRLIR^CSVDRDYLD^ENYLLVAH^SSDVIN^SISE^GGRATHDIIKTW^EHF^KKG^S

>Oa_S100A9 (XP_028911801.1)

ME^KSL^ENIINV^FHQ^YSVRV^GNE^DTILT^KRELROLINK^ELNFLK^DQ^QCFADVA^KILEDL^DSN^QDS^QLS^FEFEFVVLITRLTVAS^H
NKMHENAG^SSG^EGH^SHC^FGL^GES^GHGH^SHC^FGH^SHK^H

>Oa_S100A11 (XP_028911439.1)

MA^KVILN^ETETER^CIES^LIAVFORYA^GQ^EGNN^TLS^KTEFLK^FFMNTELA^AFS^KKNQ^DEGVLD^RMMKK^LDLN^CDG^QLD^FQ^EFLN^S
LI^GGLAQ^ACH^TS^FTAA^PPTAHHQ^HKK^I

Supplementary Figure S1. Amino acid sequences of proteins encoded by EDC genes of platypus. (A) Amino acid sequences of proteins encoded by SEDC genes of platypus. **(B)** Amino acid sequences of platypus SFTPs. **(C)** Amino acid sequences of proteins encoded by other EDC genes of the platypus. To show the peculiar amino acid compositions of SEDCs and SFTPs and the importance for protein cross-linking the following amino acid residues are highlighted: lysine (K) and glutamine (Q) as potential sites of transglutamination; cysteine residues (C) as potential sites of disulfide bonds; glycine (G), proline (P) and serine (S) are highly abundant residues not directly involved in cross-linking. When available, the GenBank accession number is shown behind the protein name. Only the S100A proteins whose genes are flanking *PGLYRP3* and *TCHHL1* are included here. SEDC, single coding exon epidermal differentiation complex; SFTP, S100 fused-type protein; Oa, *Ornithorhynchus anatinus*.

A

>Ta_LOR

MSYQQHQTTQPTFLPPVVCGGGSSVGGGGGGVGVCGGGSSGGGVYYPSQQPSQPQGISSGGGSSGCGGGVWSGGSSGSGGGGKSIG
GGVSSGGCGGGGWSSGSSGEGKSTGGGGVSSGGCGGGGWSSGSSGGGGGGKSIIGGGSSGGCGGGSSWSGGSSGGGGGGKSIIGGG
SSGGCGGGGWFGSSSSGGGKSIGGGSSGGCGGGGWSSGSSSSGGGKSIGGGSSGGCGTGGSSSGGGGFYFQQPTYVFWQPSG
GGCLTTGGGSSSGCGFGGGSISGGGSSSGGCFGGGSAQTQQTQPPSWPTK

>Ta_PRR9

MEFHSQQHKQFCLPPFCLOQQQERLKSQKACFQLACQEQGFPAKCQDLCPPVCQEQGFPAKCFPFACFEQGFVKCQDLCPPVCQKQ
GPAKRQDLCPPFCQEQHFAKCQDLCPPFCQAQGFPTKCQDLCPPFCQEQGFPAKCQDLCPPFCQEQGFPAKCQDLCPPVCLEQGFPA
KCQDLCPPFEFKEQGFVKCQDLCPPKQCK

>Ta_SPRR2HL (XP_038624405.1)

MAEDKDHSSCPPKICFATPKCSFESKVSNCRLKLLCCSAPSPPKCPKDPKPCPPKCPKEPPKCPKPEPPKCLPKCPEKE
PKCLPKCQPKPPKCPKDPPICLPFDAENP

>Ta_SPRR2BL1

MSYQNQQQVKQPCPPPPHIPPCCPPVVKCPQPCPPVVKCPQPCPPVVKCPQPCPPVVKYLKCKQPCPPPPCQQ

>Ta_SPRR2BL2

MSYQDQQQVKQPCPPPPHIPPCCPPVVKCPQPCPPVVKCPQPCPPVVKCPQPCPPVVKYLKCKQPCPPPPCQQ

>Ta_SPRR2BL3

MSYQDQQQVKQPCPPPPHIPPCCPPVVKCPQPCPPVVKCPQPCPPVVKYLKCKQPCPPPPCQQ

>Ta_SPRR2BL4

MSYQNQQQVKQPCPPPPHIPPCCPPAKCPQIPESVQCPQFPFPVQCPQPCPFACQEPFLPCQQ

>Ta_SPRR3L

MSFQONHFQVQCPPPHNQPKCPRRQEEFYHPQYEEFSYKLEPCQERYEFPERSNLRYPEPCHRRHGPCPPRSRCP
PPRHFEFCHPQYEQYPERSLPCHRRHFEPCPPRSLEFYPPRHFEPCHEQHFEPCPPGSGPCTPRHFEFCHPQYFQLYPFR
SPKPCCHRRHFEACPPRSFPCPPRHFEFCHPQYFELYPERSPFCHRRHFEPCPPRSRCPSPRRHFKPCHTQYFELHFPERS
PCHRRHFEPCPPRSRCPPPRHFEFCHPQYFEPCCPPRSFEPCCPPRQAEFCHLQYFELYPERSTGCPPPRHFEFCYFQYFELY
PRSPFEPWPPRHFEFYPERSSVPCQPHFEFYPPRSFEFCSPGRNRRHFKNPAFCPEPCPPQCPEPRSPFQQWDAPLIVQFSP
CCRCKFFREYK

>Ta_SPRR2_1

MSQQYQQQTQKQPCQPPHIVKEPCPPPVVKEPCLPPFVVKEPCLPPFVVKEPFLPKTSTFCPPPSQQKLPFAQQVFKTKQK

>Ta_SPRR2_2

MSQQYQEQTQKQPCQPPHIVKEPCPPPVVKEPCPPPVVKEPCLPPFVVKEPCPPFVVKEPFPKTKSTFCPPPSQQKVPAQ
KVEQTQKQK

>Ta_SPRR2_3

MSQIQAKQTTQVPPKEFCCLKIPTTLCHFKTPKLSQPKFCLLCPAFQQQKAPFAQQVFKSKQK

>Ta_SPRR2_4

MSQQHQQSCTPPQQTQKQCQRPPKCFEFCQPKCLPFAQQKCFPCQPKCFFPAQQKCFPCQPKCPLPK

>Ta_IVL

MSQFVQFPCQQQCKQFTVLPPALRFEFCPPQMEFQKVFQKPFTELECKTFQVKSSIFVQVSSDFQFDQKEGKEDQTDDE
FQFDQEEGKKEDQSDSSSQFEQTECKKEDQMDQHLKEKQLEKELEEKKEHLKKQLEEQLEEKLLKDEDDQNLKEKQ
LLEKELEEKKEHLEKELEQEEFKDDQNLKEKQLEKELEEDKKEQLKKQLEEGLEKELEQKEFKDDQHLKEKQLEKELEEK
KQOLE

>Ta_SPRR1

MADQQQRKQNNLPPFVSCPEFQFQKCFEFQPKYEFQPKYEFQPKCFEFQPKYEFQFNYEFQPKCFSAFGGHSKRRQ

>Ta_LCE1L

MSYQNNQQCCPPLPKLPPKVPETCFAPCPAPVPAFTSSCCSGSGGCGGDSNSRGCCSGGCGGSSSGCCCLTSFPHRRRRR
RQRRSGCGGQRDQEESSGSCC

>Ta_KPRP

MC DHQQLQPCPPFVCCVKGGLVGSVDQGIKIFCFAPCFAPAQAAGFTLCQSSSQVSSAQSQQVSLTSCQDSSVGYYCFAQAG
SVICQSSSQVSSSQDSSAWCQPPCHTQASYGTQSSSSAATFDIQDSALIQTFOGSGFALASYAQCSFVYYKETYYVCFPA
PCFVHTYYVECFVQTSYVQREVLCSQTSYVRCFAFCQFQVSVGSHAALCQFQAFAGSRFALYQTEAFARSRFALDQEQVPER
RCPLGLRFQATCQWCPRFPPRSASAALEAFAPGTSLGLCFFPRCALQGWRRCGPKYRAEICSFNPYQQVALERRPPVRIPIRR

SVSSDRQFVLDTGFSNCRVPSPPRGRSPACWDLGAGDTSDLCKARGFGFCQESGGSGCGREEFQCSGYSGCGQRGEFRG
SGNSGWGSGRECGSWNFGCSRREEVQGSKNSSGGCGGEGQSGNSSGSRGKLGSSNSGGGSGGEFGCSENFSCKCKEAFQ
RLGNSGGGISRGEDCLIFGNSECLFEDCKEPAKTNCEEQR

>Ta_KPLCE (XP_038624254.1)

MCDOQKQFQTFPAVAQVKGVGPFVQVTKCPAPNIEALVKCPPPCFAQTYVKCPPPCFTQTFVKCPPPCFTQTYVKCPPPC
FTETLVKCPPPCFTQTYVKCPPPCFTQTYVKCPPPCVTQNIIVKCPPPCQTQTCYVQCFAFCPPKCPPPCFIQKYVQCPPPCQ
VPTCYIQRPSTQTYVQRPFAWRPAPQYCPPPCPRPCPRPCPRPCPRPCPRPAPSSCFSSSNCCNLACSFQIRPLRRWVR
GPECDDNSCGGYGDEGEDSGCCLGITPMRASGPACCGDEC

>Ta_LCE2DL1 (XP_038624268.1)

MSWQONQKQFPPTEQFFQQCQPPPKCPPKCPAPCPAFCPASSSCCGFSSGGCCGFSSGDCCGFSSGGCCLSHHHHRRPRLFHR
RQHQSFCDDNGSSGCC

>Ta_LCE2DL2 (XP_038624210.1)

MSWQONQKQCPPPFQFQKQCPPPKCPPKCPAPCPAFSSCCGFSSGGCCGSSSSGGCCLSHHHHRRPHLFHRRRHQSFECCDDN
SDCC

>Ta_LCE2DL3 (XP_038624207.1)

MSWQONQKQCPPPFQFQKQCPPPKCPPKCPPKCPPKCPAPCPFLSSCCGFSSGGCCGFSSGGCCCLSHHHHRRPHLFHRRRH
QSFECCDDNSDCC

>Ta_LCE2DL4 (XP_038624208.1)

MSWQONQKQCPPPFQFQKQCPPPKCPPKCPPKCPPKCPAPCPFLSSCCGFSSGGCCGFSSGGCCCLSHHHHRRPHLFHRRRH
QSFECCDDNSDCC

>Ta_LCE2DL5 (XP_038624205.1)

MSWQONQKQFPPTEQFFQQCQPPPKCPPKCPPKCPPKCPAPCPFLSSCCGFSSGGCCGFSSGGCCGFSSGGCCCLSHHHHRR
HLFHRHRHQSFECCDDNSDCC

>Ta_LCE2DL6 (XP_038624209.1)

MSWQONQKQFPPTEQFFQQCQPPPKCPPKCPPKCPPKCPAPCPFLSSCCGFSSGGCCGFSSGGCCGFSSGGCCGFSSGGCCCL
SHHHHRRPHLFHRRRHQSFECCDDNSDCC

>Ta_LCE2DL7 (XP_038624206.1)

MSWQONQKQFPPTEQFFQQCQPPPKCPPKCPPKCPPKCPAPCPFLSSCCGFSSGGCCGFSSGGCCGFSSGGCCGFSSGGCCCL
SHHHHRRPHLFHRRRHQSFECCDDNSDCC

>Ta_LCE2DL8 (XP_038624262.1)

MSWQONQKQFPPTEQFFQQCQPPPKCPPKCPPKCPPKCPAPCPFLSSCCGFSSGGCCGFSSGGCCGFSSGGCCCLSHHHHRR
HLFHRHRHQSPESSCNDNSDCC

B

>Ta_SFTP1

MSQILLESIATIIDVFYHYTQGDGHCE TLSKRELKELKKEFRFILKNNDKVDAIMQMLDQDHNRRVNFTEFLLLVFKLTQ
ACNMVSKDFCSASCSKQKSGRHPWCQHPSQEEGEDEDEEGEEEEEEDGEAEAGRCASQSSWSAGEEWGSEFSGETVRH
GHKARHEHQGGSGKQKHLGGQYVGEKQSSKKRSGSRFRTESLTKQSGSQSGGCGGQGVSWFGGEECSGSNQSGACGRQGG
QSGKCSSTSYIGRSSSTGGQSGRACQGHSGAGDHSYEQQWSNSGSSQTTGHGGHSGISGGQSSSSGNRSGSRGGATKDGQ
TLD SYGQOETSHGQSGSSHGQTTDSYGQVTSHSQTGSAHQTTDSYGQOETSYGQSGSSHGQTRDSYGQHGTSHGHSGSSHG
QTRGTYGQTGTRQGQTGYSHGQTSSSHGQTSSSHGQTD SYRQTGSNHGQTSSSHGQTSSSHGQATDSHGQGTSHGQTYSH
GQTGYSYGQNGSGHGQTRDSYRQGSSSHGQTSSSHGQNTGSYGQOETSHGQKGYSGQTSSSHGQTSSSHGQTTDNYGQTS
HGQSSYSTGERAGRHSQSSSRSQYNTTGSRDSSQYEGEETHYSSSYGQTSNHCQTTDSYGQOETSHGQSGSRHGQTR
DSYGQOETSHGQTSSSHGQTRGTYGQTGTRQGQTSSSHGQNTDSYGQOETGTRQGQTSSTHGQNTESYRQOETSHGQSSAHQON
TDSYGQGTSHGQTSSSHGQTRDSYGQOETSHGQSGSRHGQNTDSYGQOETSHGQTSSSHGQTRGNYGQOETSHGQSGYSHGQ
NTDSYGQOETSHGQTSSSHGQITDSYGQOETSHGQSGSSHGQTRDSYGQOETSHGQTYTGQSTD SYGQOETSHGQSGYSHG
QNTDSYGQOETSHGQTSSSHGQITDSHGQOETSHGLSGSSHGQTRDSYGQOETSHGQTSRHGQTRGTYGQOETSHGQSGYSH
GQNTDSYGHQGTSHGQSGYSHGQNTDSYGHQGTSHGQIQSSHGQVTD SYGQOETSHGQSGYSHGQNTDSYGHQGTSHGQIES
HGQVTD SYGQOETSHGQSGYSHGQNTDSYGHQGTSHGQIESSHGQVTD SYGQOETSHGQSGYSHGQNTDSYGQOETSHGQ
SHGQTRGNYGQOETSHGQSGSNRQTTDSYGQOETSHGQTYTGQTTDSYGQOETSHGQSGYSHGQNTDSYGQOETSHGQTS
SSHGQTRGNYGQOETSHGQSGYSHGQTRDSYGQOETSHGQTSSSHGQTRGNYGQOETSHGQSGSSHGQTRDSYGQOETSHGQ
GSSHGQTRGNYGQOETSHGQSGYSHGQNTDIYGQOETSHGQTSRHGQTRGNYGQOETSHGQSGYSHGQNTDSYGHQGTSHGQ
LGYSHGQNTDSYGQOETSHGQSGYSHGQNTDSYGHQGTSHGQSGYSHGQNTDSYGHQGTSHGQIESSHGQVTD SYGQOETSHG
QSGYSHGQNTDSYGQOETSHGQIESSHGQVTD SYGQOETSHGQSGYSHGQNTDSYGHQGTSHGQIESSHGQVTD SYGQOETSH
GQSGSSHGQTRDSYGQOETSHGQTSSSHGQTRGNYGQOETSHGQSGSNRQTTDSYGQOETSHGQTYTGQTTDSYGQOETSH
HGQSGYSHGQTRDSYGQOETSHGQTSSSHGQTRGNYGQOETSHGQSGYSHGQTRDSYGQOETSHGQTSSSHGQTRGNYGQOET
SHGQSGSSHGQTRDSYGQOETSHGQTSSSHGQTRGNYGQOETSHGQSGSIHGQTRDSYGQOETSHGQTSRHGQTRGNYGQOET

TSHCQSGYSHGQNTDSYQQGTSHGQSGYSHGQNTDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQNRGNYQQ
GTSHCQSGYSHGQNTDSYGHQGTSHCQSGYSHGQNTDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQNRGNYQQ
QGTSHCQSGYSHGQNTDIYGHQGTSHCQTESSHGRITDSYQQGTSHCQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGYG
QGTSTNGQSGYSHGQNTDSYQQGTSHGQSGYSHGQNTDSYQQGTSHGQIESSHGQVTDYQQGTSHGQSGYSHGQNTDSY
GHQGTSHGQIESSHGQVTDYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGSNRQTTDS
YQQGTSHGQTYTYGTTDSYQQGTSHGQSGYSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGYSHGQTRD
SYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGYSHGQNT
DIYGHQGTSHGQTESSHGRITDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGYGQGNSHGQSGYSHGQNT
TDSYQQGTSHGQSGYSHGQNTDSYGHQGTSHGQIESSHGQVTDYQQGTSHGQSGSSHGQTTDSYQQGTSHGQTSSSHGQ
TRGNYGQGNSHGQSGSNRQTTDSYQQGTSHGQTYTYGTTDSYQQGTSHGQSGYSHGQTRDSYQQGTSHGQTSSSHGQ
QTTDSYQQGTSHGQSGYSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGYSHGQNTDSYGHQGTSHGQSGY
GQTRGNYGQGNSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGYSHGQNTDSYGHQGTSHGQSGY
HGQNTDSYQQGTSHCQSGSSHGQTRDSYQQGTSHCQTSSSHGQNRGNYGQGNSHGQSGYSHGQNTDSYQQGTSHGQSGY
SHGQNTDSYQQGTSHGQTESSQTTDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSG
YSHGQNTDSYGHQGTSHGQSGYSHGQNTDSYQQGTSHGQTESSLGQVTDYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTS
SSHGQTRGNYGQGNSHGQSGSNRQTTDSYQQGTSHGQTYTYGTTDSYQQGTSHGQSGYSHGQNTDSYQQGTSHGQ
TESSSHGQITDSYQQGTSHGQSGASHGQTRDSYQQGTSLGQTSRHRQTRGNYGQGNSHGQSGYSHGQNTDSYGHQGTSHG
QSGYSHGQNTDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQNRGNYGQGNSHGQSGYSHGQNTDSYGHQGTSH
GQSGYSHGQNTDSYGHQGTSHGQTESSHGQITDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGN
HGQSGYSHGQNTENYGHQGTSHGQSGYSHGQNTDSYGHQGTSHGQTESSHGQVADSYQQGTSHGQSGSSHGQTRDSYQQGT
SRGQTSSSHGQTRGNYGQGNSHGQSGSNRQTTDSYQQGTSHGQTYTYGTTDSYQQGTSHGQTSSSHGQITDSYQQGT
TSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQNRGNYGQGNSHGQSGYSHGQNTDSYQQGTSHGQTESSHGQITDSYGHQ
GTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGYSHGQNTDSYGHQGTSHGQSGYSHGQNTDSYGH
LGISHGQIESSHGQVTDYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGYSHGQNTDSYGH
QTGTSHGQTSSSHGQSRGTYGTTGTGQGTSSHGQTTDSYQQGTSHGQSGSSHGQTRDSYQQGTGHQGTSSHGQNRGNYG
GQGTSHGQSGYSHGQNTDSYGHQGTSHGQTESSHGQITDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGN
YQQGTSHGQSGYSHGQNTDSYGHQGTSHGQSGYSHGQNTDSYGHQGTSHGQIESSHGQVTDYQQGTSHGQSGSSHGQTTD
SYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGYSHGQNTENYGHQGTSHGQSGYSHGQNTDSYGHQGTSHGQTESSHGQVT
DSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGSNRQTTDSYQQGTSHGQTYTYGTT
TDSYQQGTSHGQTSSSHGQITDSYQQGTSHGQSRSSHGQTRDSYQQGTSHGQTSSSHGQNRGNYGQGNSHGQSGYSHGQ
NTDSYQQGTSHGQSGYSHGQNTDSYGHQGTSHGQIESSHGQVTDYQQGTSHGQSGSSHGQTTDSYQQGTSHGQTSSSHG
QTRGNYGQGNSHGQSGSRHGQTRDSYQTGTSHGQTSSSHGQSRGTYGTTGTGQGTSSHGQTTDSYQQGTSHGQSGSSH
GQTRDSYQQGTSHGQTSSSHGQNRGNYGQGNSHGQSGYSHGQNTDSYGHQGTSHGQTESSHGQITDSYQQGTSHGQSGSS
HGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGYSHGQNTDSYGHQGTSHGQSGYSHGQNTDSYGHQGTSHGQIE
SHGQVTDYQQGTSHGQSGSSHGQTTDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGSRHGQTRDSYQTGTSHGQTS
SSHGQSRGTYGTTGTGQGTSSHGQTTDSYQQGTSHGQSGYSHGQNTDSYGHQGTSHGQSGSSHGQTRDSYQQGTSHGQ
GSSHGQTRGDYGHQGTSHGQIESSHGQVTDYQQGTSHGQSGYSHGQNTDSYGHQGTSHGQSGSRHGQTRDSYQQGTSHDQ
SGSSHGQITDSYQQGTSHGQTSSSHGQTRGTYGTTGTGQGTSSHGQNTDSYQTGTSHGQSSYSTERTGRHSGQSSSG
RSYSATSSRDSSQYEGEETHYSSYGTGTSHGQTTDRYGTSSHRQTSSSVGQSGYSHGQTRDSYQTGASHGQTS
DGTGTSHGQTHSGQGTSSSHGQTRDIYGTSSHRQTETSHGQTSNQGTSSQGTSSYGTSSYGTSSYQRTSSYQRTSS
SWAGTSSVQTGGRGRQRQRSSGNWRHSSYGNLAYDYGEFFGQSSYGVTVSRNSSQTSSTETGSSYGTSSDYGQ
SSSTFKQSSSYGQSSDYGQGYSYGQSSDYEQSSSTFGNQSSSYGQSSDYEQSSSTFGNQSSRYGQSSDYEQSS
YGQSSDTHRQSSSTTEGQSSIHGQSSDTYEQSSSETFKQSSSHGQSSDYEQSSWHQSSQSSSSQSSSSQSSSSQSS
GTVSSSFYSSSTFLYEYIQEQRGLY

>Ta_SFTP2

MSQLLKSIATIIDVFYHYTQRDGDYETLSKRELKELLKKEFFHPIIKHNDPDTVDIIMQMLDQDHDRRVNFAEFLLMVFKLTQ
ACNMVSKDFRSASGSKQKSGRHFQCGHSSQPEEGEGEEDEEGEGEGEEEDGEEAERGAGQSSWSAGEERGSESGET
VKHCHQGDGHKACHGHGSGSKOKHKLGGKSYVGEKQSSKKRSGSSRTESLTKQSGQSGGCGGVSWFGGEGEGSSGN
SGAGCRQGGGSGKQSTSYVGRSSTGGQSGRSGSSSGAGDHSYEGQWNSGSGQTTGHGGHSTSGEGGSSSGSGNSG
SRGATKDGQTLDSYQQGTSHGQTSSSHGQTRGTYGTTGTGQGTSSHGQTTDSYQQGTSHGQSGYSHGQNTDSYGHQGT
RHGQIESSHGQITDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQSGSSHGQTRGNYGQGNSHGQSGSNHQTDDSYQQGT
TSHGQTYTYGTTDSYQQGATHGQSGYSHGQNTDSYGHQGTSHGQTESSHGQTRDSYQQGTSHGQSGSSHGQTRDSYQQ
GTSHGQSGSSHGQTRGNYGQGNSHGQSGSNHQTDDSYQQGTSHGQTYTYGTTDSYQQGAIHGQSGYSHGQNTDSYGH
QGTSHGQTESSHGQTRDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGSNHQTDDSY
HQGTSHGQIESSHGQITDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGSNHQTDDSY
QQGTSHGQTYTYGTTDSYQQGATHGQSGYSHGQNTDSYGHQGTSHGQTESSHGQTRDSYQQGTSHGQSGSSHGQTRDS
YQQGTSHGQSGSSHGQTRGNYGQGNSHGQSGSNHQTDDSYQQGTSHGQTYTYGTTDSYQQGAIHGQSGYSHGQNTD
SYGHQGTSHGQTESSHGQTRDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGYSHGQNT
DSYGHQGTSHGQIESSHGQITDSYQQGTSHGQSGSSHGQTRDSYDQGTSHGQTSSSHGQTRGNYGQGNSHGQTS
TRGTYGTTGTGQGTSSHGQNTDSYQQGTSHGQSGYSHGQNTDSYGHQGTSHGQSGYSHGQNTDSYGHQGTSHGQTESSHG
QVTDYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGYSHGQNTDSYQQGTSHGQSGYSH
GQNTDSYGHQGTSHGQIESSHGQVTDYQQGTSHGQSGSSHGQTTDSYQQGTSHGQTSSSHGQTRGNYGQGNSHGQSGSR
HGQTRDSSEGTGTSHGQTSSSHGQSRGTYGTTGTGQGTSSHGQTTDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTS
SHGQNRGNYGQGNSHGQSGYSHGQNTDSYGHQGTSHGQTESSHGQITDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTS
SSHGQTRGNYGQGNSHGQSGYSHGQNTDSYGHQGTSHGQIESSHGQITDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQTS
GSSHGQTRGNYGQGNSHGQSGSNHQTDDSYQQGTSHGQTYTYGTTDSYQQGATHGQSGYSHGQNTDSYGHQGTSHGQ
TESSSHGQTRDSYQQGTSHGQSGSSHGQTRDSYQQGTSHGQSGSSHGQTRGNYGQGNSHGQSGSNHQTDDSYQQGTSHG
QTGYTYGTTDSYQQGAIHGQSGYSHGQNTDSYGHQGTSHGQTESSHGQTRDSYQQGTSHGQSGSSHGQTRDSYQQGTSH
GQTSSSHGQTRGNYGQGNSHGQSGYSHGQNTDSYGHQGTSHGQIESSHGQITDSYQQGTSHGQSGSSHGQTRDSYQQGTSH

HGQTGSSHGQTRNGYQQGNSHGQSSGNHGQTTDSYQQGTSHGQTYTYGQTTDSYQQGATHGQSGYSHGQNTDSYGHQGT
SHGQTESSHGQTRDSYQQGTSHGQSSSSHGQTRDSYQQGTSHGQSSSSHGQTRNGYQQGNSSRGQSSGNHGQTTDSYQQG
TSHGQTYTYGQTTDSYQQGAIHGQSGYSHGQNTDSYGHQGTSHGQTESSHGQTRDSYQQGTSHGQSSSSHGQTRDSYQQ
TSHGQTESSHGQTRNGYQQGTSHGQSGYSHGQNTDSYGHQGTSHGQIESSHGQITDSYQQGTSHGQSSSSHGQTRDSYD
QQGTSHGQTESSHGQTRNGYQQGNSHGQTESSHGQTRTYGQGTTRQGTGSSHGQNTDSYQQGTSHGQSGYSHGQNTDSY
GHQGTSHGQSSGYSHGQNTDSYGHQGTSHGQTESSHGQVTDYQQGTSHGQSSSSHGQTRDSYQQGTSHGQTESSHGQTRNG
YQQGTSHGQSSGYSHGQNTDSYQQGTSHGQSSGYSHGQNTDSYGHQGTSHGQIESSHGQVTDYQQGTSHGQSSSSHGQTT
SYQQGTSHGQTESSHGQTRNGYQQGNSHGQSSSRHGQTRDSSGQGTSHGQTESSHGQSRGTGQGTTRQGTGSSHGQTT
DSYQQGTSHGQSSSSHGQTRDSYQQGTSHGQTESSHGQNRNGYQQGTSHGQSSGYSHGQNTDSYGHQGTSHGQTESSHGQ
TDSYQQGTSHGQSSSSHGQTRDSYQQGTSHGQTESSHGQTRNGYQQGTSHGQSSGYSHGQNTDSYGHQGTSHGQSSGYSHG
NTDSYGHGHSIHHGQIESSHGQVTDYQQGTSHGQSSSSHGQSSSSHGQSSSSHGQSSSSHGQSSSSHGQSSSSHGQSSSSHG
QTRDSYQQGTSHGQTESSHGQSRGTGQGTTRQGTGSSHGQTTDSYQQGTSHGQSSGYSHGQNTDSYGHQGTSHGQSSSSHG
GQTRDSYQQGTSHGQTESSHGQTRDYGQGNSHGQTESSHGQNTADSYQQGTSHGQSSGYSHGQNTDSYGHQGTSHGQSSSR
HGQTRDSYQQGTSHGQSSSSHGQITDSYQQGTSHGQTESSHGQTRTYGQGTTRQGTGSSHGQNTDSYQQGTSHGQSSSY
STGERTGRHSGGSESESSGYSATSRDSSFGQYEGEETHYSSSYGQGTSHGQTTDRYQGTGSSSHRQTGSSVQSSGYSHGQ
RDYQGTGASHGQTESSHGQGTSTHGQTHSQQIESSHGQTRDIYQGTGSSSHRQTETSHGQTESSNQSSGTSQGTGSSGYQ
TSSYGTSTSYQRQISDSWQSSGTSSVQAGGRGRQRSSQNWRHSSYGNLAYDYEFGFGQSSYGVTVSRNSSFGQSSGRT
GTQSSSYGQSSDITYQQSSRTFGKQSSSYGQSSDITYQQSSSYGQSSDITYQQSSSTFGNQSSSYGQSSDITYQQSSSTFGNQ
SSSYGQSSDITYQQSSSTFGNQSSSYGQSSDITYQQSSSTFGNQSSRYGQSSDITYQQSSSYGQSSDTHRQSSGTTGQCS
HGQSTDTYQGFSSSETCKQSSSHGQSSDITYQSSWFGHWQSSSSSGSAGSASFGQRQSSASGTVSSGSFYSSSTFLYEYIQE
RGLY

>Ta_SFTP3

MSQLLKSSISTIIDVFYHYTQRDGDYETLSKRELKELLKKEFRPILKHNDPDTVDVIMQMLDQDHDRRVNFAEFLLMVFKLTQ
ACNMVSKDFRSASGSKQKSGRHPWCQCHPSQPEEGEGEEGEEDEEGEGEEEEEEDGEEEAERGAGQSSWSAGEERGSESGET
VKHCHQGDGCHKACHGHQGSSEKQKHLGGKYVGEKQSSKKRSGSSSRTESLTKQSSQSGGCGGQVSWFGGEECSGSN
QSGAQGRQGGQSGKHSTSYVGRSSSTGGQSSRSAGQGHGSGAGDHSYAAQQWNSNGSQTTHGGHGSTSGGQSSSSGNGRSG
SRGGATKDGQTLDSYQQGTSHGQSSSSHGQTRTYGQGTTRQGTGSSHGQNTDSYQQGTSHGQSSSAHGQNTDSYGHQGT
SHGQTESSHGQITDSYQQGTSHGQSSSSHGQTRDSYQQGTSHGQTESSHGQTRNGYQQGTSHGQSSGYSHGQNTDSYGHQ
TSHGQTVSSHGQITDSYQQGTSHGQSSSRHGQTRDSYQGTSHGQTESSHGQTRTYGQGTTRQGTGSSHGQNTDSYQQ
GTSHGQSSGYSHGQNTDSYQQGTSHGQSSGYSHGQNTDSYGHQGTSHGQTESSHGQITDSYQQGTSHGQSSSSHGQTRDSYQ
QGTSHGQTESSHGQTRDYGQGNSHGQSSGNHGQTTDSYQQGTSHGQTYTYGQTTDSYQQGTSHGQSSGYSHGQNTDSY
GHQGTSHGQTESSHGQTRDSYGHQGVSHGQSSSRHGQTRDSYQQGTSHGQTESSHGQTRNGYQQGNSHGQSSGYSHGQNTDSY
GHQGTSHGQTESSHGQITDSYQQGTSHGQSSGYSHGQNTDSYGHQGTSHGQTESSHGQITDSYQQGTSHGQSSSSHGQTRDS
YQQGTSHGQTESSHGQTRDYGQGNSHGQSSGNHGQTTDSYQQGTSHGQTYTYGQTTDSYQQGTSHGQSSGYSHGQNTDSY
SYGHQGTSHGQTESSHGQTRDSYGHQGVSHGQSSSRHGQTRDSYQQGTSHGQTESSHGQTRNGYQQGNSHGQSSGYSHGQNT
DSYGHQGTSHGQTESSHGQITDSYQQGTSHGQSSSGHGQTRDSYQGTSHGQTESSHGQTRNGYQQGTSHGQSSGYSHGQNT
TDSYGHQGTSHGQTESSHGQVTDYQQGTSHGQSSSSHGQTRDSYQQGTSHGQTESSHGQTRDYGQGNSHGQSSGNHGQ
NTDSYQQGTSHGQTYTYGQTTDSYQQGTSHGQSSGYSHGQNTDSYQQGTSHGQSSSRHGQTRDSYQGTSHGQTESSHG
QTRTYGQGTTRQGTGSSHGQTTDSYQQGTTRHGQSSGYSHGQNTDSYQQGTSHGQSSSSHGQTRDSYQQGTSHGQTESSH
GQTRDYGQGNSHGQSSGNHGQTTDSYQQGTSHGQTYTYGQTTDSYGHQGTSHGQSSSRHGQTRDSYQGTSHGQTESS
HGQTRTYGQGTTRQGTGSSHGQNTDSYQQGTSHGQSSGYSHGQNTDSYQQGTSHGQSSSSHGQTRDSYQQGTSHGQTES
SHGQTRDYGQGNSHGQSSGNHGQTTDSYQQGTSHGQTYTYGQTTDSYGHQGTSHGQSSSRHGQTRDSYQGTSHGQTES
SSHGQTRTYGQGTTRQGTGSSHGQNTDSYQQGTSHGQSSGYSHGQNTDSYQQGTSHGQSSSSHGQTRDSYQQGTSHGQ
SSHGQTRNGYQQGTSHGQSSGYSHGQNTDSYQQGTSHGQSSGYSHGQNTDSYGHQGTSHGQTESSHGQITDSYQQGTSHGQ
SSSSHGQTRDSYQQGTSHGQTESSHGQTRDKYQQGNSHGQSSGNHGQTTDSYQQGTSHGQTYTSGQTTDSYQQGTSHS
QSSGYSHGQNTDSYQQGTSHGQTESSHGQVTDYQQGTSHGQSSSSHGQTRDYGQGNSHGQSSGNHGQNTDSYQQGT
SHGQTESSHGQNTDSYQQGTSHGQTYTYGQTTDSYQQGTSHGQSSGYSHGQNTDSYQQGTSHGQSSSRHGQTRDSYQGT
HGQTESSHGQTRTYGQGTTRQGTGSSHGQTTDSYQQGTSHGQSSGYSHGQNTDSYGHQGTSHGQTESSHGQTRDSYQQGT
SHGQTESSHGQTRDYGQGNSHGQTESSHGQNTADSYQQGTSHGQSSGYSHGQNTDSYGHQGTSHGQTESSHGQTRDSYQ
TSHGQSSSSHGQITDSYQQGTSHGQTESSRKGQTRTYGQGTTRQGTGSSHGQNTDSYQGTSHGQSSSYSTGERTGRHSG
QSESSRSGYSATSRDSSFGQYEGEETHYSSSYGQGTSHGQTTDRYQGTGSSSHRQTGSSVQSSGYSHGQTRDSYQGTGASH
QTSSDGQGTSTHGQTHSQQIESSHGQTRDIYQGTGSSSHRQTETSHGQTESSNQSSGTSQGTGSSYGTSTSYSQ
RQISDSWQSSGTSSVQAGGRGRQRSSQNWRHSSYGNLAYDYEFGFGQSSYGVTVSRNSSFGQSSGRTGTQSSSYGQSSD
TYQQGTSSSTFGKQSSSYGQSSDITYQQSSSYGQSSDITYQQSSSTFGNQSSSYGQSSHTYQQSSSTFGNQSSSYGQSSDITY
QQSSSTFGNQSSSYGQSSDITYQQSSSTFGNQSSSYGQSSDITYQQSSSYGQSSDITYQQSSGTTGQCSIHGQSTDIYQGF
SSSETCKQSSSHGQSSDITYQSSWFGHWQSSSSSGSAGSASFGQRQSSASGTVSSGSFYSSSTFLYEYIQEORGLY

>Ta_RPTN (XP_038624401.1)

MSQLINSILRIIEVFQNYANADDNVSLTKTELSSELLQAEFGNLLRRERDEKTVDTIILQLLDRDRDGVVGFNEFLLLVLVFKMAQ
ACHQALGNDACGRGRDQGEPEEEEGEKREGEAEAAQRRRTGQTEVTGRRREDTPRGQAGQQRRESSESQSERQRSDFDQDNHSES
ERWTSSYRQSERRDSRSGQSERQRSDFRDQSSERWTSSQGSERRDSHSGQSERQRSDFRDQGYESRRTSSYQSSKRDS
RSQSERQRSDSRDDQSSSRRTSSYDHSERRDSRSGQSERWGGQSSNGQYESRRTTSSRSEKQWSDSQSSQSSRRAS
SYGQSERRDSCSQSESRGERERQKRVVDGGREQSGHGLQWDPETEAQSRKKLTSQATSLCHQDQSWQSSYESLRAHRAG
RTGQGRDEGQSHQPDGYTDDEEQSCQAHGQRRQSRGRQVEEEDQTTAGGGQRCQSLGRQQSPSQDWQTDDEESSRQTRDQ
RRQSRANAMEEEDQSSQTRDQWRQSRGGYDEDEDQSRQARDQRRQARGGYSEEEHSHREKKRTSQSATSLCHQGTWQSC
FLRARRGGRTGQRDEQSRQFQRYTDDEEQSRQARDQRCQSQGGQAEENQSVQTRQSRGRQGLSQDRQTDDEEIRQT
LDRQRQARGEYTEEEEQSHQTDQGWQSRRAANTADEQSRQAWDQRRQSRGGYTEEEEQSREKKRSSQFVTSLCHQGTW
QSCFLRGRRTGRTGQGRDEQSRQFQDRYTEDEEQSRQAGGQRRQSRGRQFESQDWQTDDEESSRRTQDQRCQSRANM
EDEDQSHQAQDQRRQSRGGYTEDEDQSRQARDQRCQSRRAADMEDEDQSQOARGGYTEEEEQSREKKRTYQATSLCHQGT

WQSCFLRARREGRTGCKRDEGQSRQPDRTYDDEDDRRQAWGGQQSQSRGGQAEKENQSVHAGGEQRRQFRDRQONFSQDWQ
 TDDEEQSRQTRDGQRRQSRGGYPKDEEKSRQAQDGQRRQARGGQAEDEDHVSQAGGGQRRRSSELARSQAQCSQCPDPRATST
 TRSTVGNSAWKETQCRGGARGRREPKASESDSNADSHLSVFPQFLYEYVHEQSVHHYQ

>Ta_TCHH

MSFLKSIIDISEIFNYATCDGDDVKLNKGELRTLLQREFRDVLRRLQDFTQVDLVIQLLDRDRDGSVDFNEFLLLVFKVAQ
 ACYSALSQVAGSTKREKGRDFSQERQSRQEDDQRCGRDGERHQQQEQDQELQRRERCKEEEEEEEEQROKQOEEERREQLC
 NEEAERERQRLCEEEERRQRQHERQREQRSEEEERERRREQLCDEEEQRRERFSEQRFEEEEQFLGQRKQRCHEEEERLRQ
 RLRRREEKRRLKQOELCEDRRQRQSRQDEALEEAKHREVSQRVQVEREGQARQNKLYSQFRVPERVRRRTARDQRLQGAEESS
 RRREKERRHTEEEEFRRLLDQDRLREEEKDEEERSTRDLEQNQNLNWERQVEEEKAOHRNKLYSKERSQOETREKQERELR
 EELEELRQOERDRQSRGEDHVRREQERKLREELEELRQOESDRQSRGEDQVRRQERERNLREELEELRQOERVRQCGEDQV
 RRQERERKLREELEELRQOERDRQSWGEDQVRRQERERKLREELEELRQOGRDRQSRGVNQFRQERERKLREELEELRQERD
 RQSRGEDQVRRQERERKLREELEELRQOERNRQSRGEDQVRRQERERKLREELEELRQOGRDSSQSRGEDQVRRQERERKLREE
 LEELRQOERDRQSRGEDQVRRQERERKLREELEELRQOERDRQSQGEDQVRRQERERKLREELEELRQOERDRQSRGEDQVRRQ
 RERKLREELEELRQOERDRQSRGEDQVRRQERERKLREELEELRQOERDRQSRGEDQVRRQERERKLREELEELRQOARDSSQ
 RGEDQVRRQERERKLREELEELRQOERDRQSRDEDQVRRQERERKLTEELEELRQOGRDLQSRGEDQFRQERERKLREELEE
 RQOERDRQSRGENQVRRQERERKLREELEELRQOERDRQSRGEDQVRRQERERKLREELEELRQOERDRQCGEDQVRRQER
 ERKLREELEELRQOGRDRQSQGEDQVRRQERERKLREELEELRQOERDRQSRGEDQVRRQERERKLREELEELRQOERDRQSR
 GEDQVRRQERETQREELEELRQOERDRQSRGEDQVRRQERERKLREELDELROOGRDRQSRGEDQVRRQERERKLREELDEL
 RQOERDRQSRDEDQVRRQERERKLREELEELRQOERDRQCGEDQVRRQERERKLREELEELRQOGRDRQLGEDRVRRQEEERE
 LREELEELQORETCKRYEEDQNRLLKSRQPERENVVRNKEVFSKAYEVDAEQLSQDLSTEEGYCCQEVLSRETLEEEER
 RRQARDQQFYRDAQNHRESQEEERLLWERERKFREEERRCEEEEEVEQRRRQERERELQEEELLEEVEQLQRLRRDRDKL
 REEERLRREEEVEEELEELRRERNRKLREEEQLRREEVEQELEELCWERNRKLREEEQLLQEEEREEKRRQERERKLRE
 EEEQRFRERDRKFREEEHLGEEGEEQRFRERERKLREEEQLRKEEEREEKKRRQGRERKQEEEQDLKRLRRERDRKL
 REEQQLRKEEEVEQELKKLRRGRDRRVREEKREQNRFOIQDROTDFRSPLSNPTARSSPLYEYIQEQRSQYCP

C

>Ta_PGLYRP3

MLRRVLLLLAVGMRSSDSSPPAVTSHAPQPPAYSNATEALARFKILLGCFRDIFQEPKKIIPRAEWGAQAFHCTVPLKTTTF
 YLLFHHIAGMDCAQGSWCVRQLQDHHHTNGWCDAYNFLIMESGEVFEGTGWTVQGHHTAGYNEVALGFALLTNMTDQAF
 SQAALASQHLISFAVQKRHLSYNIQFLVHGEDCLQVSGSTPSGLGCTIIPRADWGAMGSMANCRKLDRPTKYVIIH
 TTGQCTESDSCKELVRGIQDFHMGKFCFQVGYNFLVGEDGNVYEGVGEWTEGAHTYGYNDIALGVAFGLFEEKPPNAVAL
 LAAQLRLIRCSVDQYLDENYLLVAHSDIINSISPERATYDIKTPHFKG

>Ta_S100A9 (XP_038624912.1)

MEKALGDVIDVFHQYVVRVGNEDTLTKRELRLQINKELNFKLDQCCAEVAKILEDLDSNQDFELSFEEFVVLITRLTVASH
 NKMHENAGSGPGHSHGFLGESGHGCHSHGPGHGHSHKH

>Ta_S100A11 (XP_038624799.1)

MAKIVNPTETERCTESLIADFQRYAGQEGNNTLSKTEFLKFMNTELAASFSSQKDEGVLDMMKKLDLNCQGQLDFQEFIN
 LIGGLAQACHASFTAAPEATHLQHKKI

Supplementary Figure S2. Amino acid sequences of proteins encoded by EDC genes of echidna. (A) Amino acid sequences of proteins encoded by SEDC genes of echidna. **(B)** Amino acid sequences of echidna SFTPs. **(C)** Amino acid sequences of proteins encoded by other EDC genes of the echidna. To show the peculiar amino acid compositions of SEDCs and SFTPs and the importance for protein cross-linking the following amino acid residues are highlighted: lysine (K) and glutamine (Q) as potential sites of transglutamination; cysteine residues (C) as potential sites of disulfide bonds; glycine (G), proline (P) and serine (S) are highly abundant residues not directly involved in cross-linking. When available, the GenBank accession number is shown behind the protein name. Only the S100A proteins whose genes are flanking *PGLYRP3* and *TCHH* are included here. SEDC, single coding exon epidermal differentiation complex; SFTP, S100 fused-type protein; Ta, *Tachygllossus aculeatus*.

[illegible]

MSSQQQQQQQEQCPSSQPPQQQVVKQPCQPPPVKQDPCAPSTKDCPPPKGTVVPAQKCPPAQKCPPAQKCPPAQKPKQK

MSQHGYKQFVSLPPFVSQEQQCKQLPQIPDFVPEQVQKQPTVPFPFCFVPEQVLDVQEETTVKIVFTLTPEQLEQKGGQELE
EQLEQQLEHKQEQQLDHKQVQQLEEQLEHKQEQQLEGEQREHKHEQQLDHKQVQQLEEQLEHKQEQLHGEQLDQQLEEQLEHKQ
EQQLEHKQEQQLEEQLEHKQEQLHGEKLDQQLEGEQLEHKQEQQLEHKQVQQLEGEQLDPLGEQLEHKHEQQLEHKQEQLHGEQ
LDQQLEGEQLEHKQEQQLEGHKQVQQLEEQLEKQEEDLKQOLEQQLEEKKEEMKQOLEQQLEEKKEELKQOLEEKKEEQQ

MSQQSQQQCQPPPKKSTFKCPKKPPPVSSCGSSSGGCCSSCGGGGGCCLCSHRRRRRQRQSSCGSSGHSSGGG
CCSGHSSGGSGCCSSGHSSGSGCCSSGHSSGSGCCSSGHSSGGGCCSSHGSSGGGCCSSSQSGGC

MSQQSQQQCQAPFKKCNFKCPKQVSAPTLAPCPKPFESSCEPSSGCCCVTSGGACCASSTGGCCNSGAGTGGGFSLFHPRW
RSYRFWNLRSSSDCCGSGGNQPGSGGCC

[illegible]

MS^{SS}QQNQ^QQ^QPY^QG^QPP^QPK^QQT^QPK^QDL^QPK^QCP^QP^QQT^QPA^QAF^QSC^QSSNT^QGG^QG^QY^QSS^QSG^QPY^QS^QFS^QQWR^QRS^QIRRRV^QHS^QDYR^QGR^QNG^QSS^QQK^Q
ST^QEC^QGG^QY^QGS^QNG^QGF^QSW^QLS^QPQ^QEQ^QH

MCDQQKQFPQFLPCCVKGSALGSIETAQGPASSQVDSCEYGEPCQSKSSFIKSTSSFFQKQSKCAFVKCPPPCQTTYVKCF
 PFCQTTYVKCAPPCQTTFFVKCAFQTTFFVKCPFCQTTYVTCPFCQTTFFVKCSPPPCQTTCKCFAPCQTTSVKCFAPCQT
 TSVKCFAPCQTTCKVCFAPCQTTSSVQCQAPCQVQTFQQTQTQYVQSSAPCQTQTYVQTFPSVQYLAASSYYTSSSYGTTCCPA
 AFMSFGVRLRLRWIRGPQCRENTGCCEDSGCCSSCCSSGCCGCGCCCLGILPMRSRGPVCCANDDDCCC

MSCQQSQQQQ**CQA**E**KC**P**PK**C**PPK**CQT**FKC**P**PK**C**PEK**C**PPK**C**PPK**C**PP**Q**A**P**CP**PPP**S**CC**GSSSGG**CC**SSGG**CC**GSSSGG**CC**SSGS**
GGCC**LFS**HRRRSHRRRHQRSDCD**SS**GRSQQ**SGGC**CG**GSSGG**CC**SSGG**CC

MS C Q Q S Q Q Q C Q A E K C P P K C P P K C Q T F K C P P K C P P K C P P K C P P K Q A P C P P F S C C G S S S G G C C S S G G C C G S S G G C C S S S G G
G C C L F S H H H R R S H R H R R Q K S D C D S G S G H S Q Q S G G C C G S S G G C C S S G G C C S S G G C C

MS C Q Q S Q Q Q Q C Q A E K C P P K C P P K C Q T E K C P P K C Q T E K C P P K C P P K C P P Q A P C P P P S C C G S S S G G C C S S G G C C G S S G G C C S S G S G
G C C C L F S H H H R R S H R H R R E K S D C C D S S G S R S Q Q S G G C C G S S G G C C G S S G G C C G S S G G C C

MS C Q Q S Q Q Q C Q A F K C P P K C P P K C Q T F K C P P K C P P K C P P K C P P K C P P Q A P C P P F S C C G S S S G G C C S S G G C C G S S G G C C S S S G G
G C C L F S H H H R R S H R H R R P K S D C D S G S G R S Q Q S G G C C G S S G G C C S S G G C C G S S G D C C

MSCCQQSQQQCCQA**EK**CP**P**K**C**PP**K**CQT**EK**CP**P**K**C**PE**K**CP**P**K**C**PP**K**C**PP**QA**PC**PPP**S**CC**GSSSGG**CC**SSGG**CC**GSSSGG**CC**SSSG**
GGCC**L**F**SHHHRR****SHRHR**R**EK****S**DCCD**SSSGRS**QQ**SGG**CC**GSSGG**CC**SSGG**CC**SSGG**CC**GSSGG**CC

MSCCQQSQQQCCQA**EK**FCPPKC**PPK**CQT**FKC**PPKC**PEPK**CPPKC**PP**QA**EC**PPPVSSCCGSSSGGGCCSSGGCCGSSSGGGCCSSGSGGG
CCLF**S**HRRRS**S**HRRRHQRSDCCDS**G**SR**S**QQSGGC**G**SSGGCC**G**SSGGCCGGSGGCC

MSCCQSSQQQCQAFCCKPPKCPKPCQTFFCKPCKCPEKCPPKCQPPQAFCPPPVS^{SC}^CSSSGGCCSSGGCCSSSGGCCSSSGGD
CCLFSSHRRRS^HSHRRRHQRSDCCDSG^SGRSQQS^GGGCCGGSSGGCCGSSGGCCGGSSGGCC

MSCQQSSQQQCQA**K**CPKPPK**P**PKQT**F**KCPKCPK**C**PP**A**PCPPV**S**CCGSSSGGCCSSGGCCGSSSGGCCSSSGGGCCLF
SHHRRSSHRRRHQRSDCCDSGSRSQSOGGCCSSGGCCSSGGCC

MS C Q Q S Q Q Q C Q A P K C P P K C P P K C Q T P K C P P K C P P K C P T Q A P C P P P V S S C C G S S S G G C C S S G G C C G S S S G G C C S S G S G G G C C L F

SHHRRSSHRRRHQRSDCCDSGSGRSQQSGGCCSSGGSCGSSGGCC

>Md_LCE2AL2 (XP_007485553.1)

MSCQQNQQQCQPPPKCQAFKCPKPPQAFCEPAFASSCCGSSSGGCCGSSSGGCHSSSSGGGCCLFSSHHRSSHRRRHQRSDNC
CDSSGSHSQQSGDCCGNSGGCC

>Md_LCE3L2

MSCQQNQQQCQPPPKCQTFKCPPKSQAFCTPPVSSCGASSSVGCHSSGSGAGCGSGSEGGCCLFSSHHRSSHRRNRQHSDCC
GSGSGHSQQSGDSCGNSGGCC

>Md_LCE3L3

MSCQQNQQQCQPPPKCQTFKCPPKSQAFCTPPVSSCGASSSVGCHSSGSGAGCGSGSEGGCCLFSSHHRSSHRRNRQHSDCC
GSGSGHSQQSGDSCGNSGGCC

>Md_LCE3L4

MSCQQNQQQCLPPPKCQAFKCPQAFCSFFISSCSGSSFGGCGSSGSEICGSSGSEGGCSFFPHHNRSSHRRGSGSDCFDSSG
SGHSQQFEDSCSSGNCC

>Md_LCE3L5

MSCQQNQQQCQPPPKCQTFKCPPKSQAFCTPPVSSCGASSSVGCHSSGSGAGCGSGSESGCCLFSSHHRSSHRRNRQHSDCC
GSGSGHSQQSGDSCGNSGGCC

>Md_LCE3L6

MSCQQNQQQCLPPPKCQAFKCPQAFCSFFISSCSGSSFGGCGSSGSEICGSSGSEGGCRLFPHHNRSSHRRGHQGSDFDNG
SGHSQQFEDSCSSGNCC

>Md_LCE3L7

MSCQQNQQQCQPPPKCQTFKCPPKSQAFCTPPVSSCGASSSVGCHSSGSGAGCGSGSEGGCCLFSSHHRSSHRRNRQHSDCC
GSGSGHSQQSGDSCGNSGGCC

>Md_LCE3CL (XP_007485552.2)

MSCQQNQQQCQPPPKCQTFKILPQAFCSFFISSCSGSSSGTCSGSSSGRCSSSSDCCCLFSSHHRSSHRRQRSSDHDGCGG
SGHSQQSGGSCGSSGGCC

B

>Md_CRNN (XP_007485551.1)

MEQLLGNIDGIIQAFSRYAKTEGDCITLTKEGELKKLLEQELADVIVKEHDEATVDQVFHLLDEDSKGTVDKKEFLVLVFKVAQ
ACYKTLNESSSQAGCTQKPGSQHSGSTQQRNDGDQRSHTEIGRVEREQSHTRHGGEETSTGQNRNSNLTTGTQTQGGDVTSTQ
ICDQDRWFKTQGERGTQOVRDKIEGQRQGETGTTQTREQDGRTHQSETVTGGQSQIHTGTTQSSREQGWTQKTDKVIQSQQ
VHTGTTQTREQDSSHQTRDRVTTGGQSQIHTGTTQTREQDSSHQTRDRVTTGGQSQIHTGTTQTVEQDRSRHQTRDRVTTGGQS
QIHTGTTQTREQDGRTHQSETVTGGQSQIHTGTTQSSREQGWTQKTDKVIQSQQVHTGTTQTREQDSSHQTRDRVTTGGQS
IHTGTTQTREQDRSHQTRDRVTTGGQSQIHTGTTQTVEQDRSHHQTRDRVTTGGQSQIHTGTTQTVEQDRSHHQTRDRVTTGGQS
QIHTGTTQTVEQDRSHHQTRDRVTTGGQSQIHTGTTQTVEQDRSHHQTRDRVTTGGQSQIHTGTTQTVEQDSSHQTRDRVTTGGQ
SQIHTGTTQTVEQDSSHQTRDRVTTGGQSQIHTDTTQTVEQDSSHQTRDRVTTGGQSQIHTDTTQTVEQDSSHQTRDRVTTGGQ
QTQIHTGTTQVRGQGRTHQTSVTIVTRGQSQITGTTQTNQTQQLQTYGQTRRPETQGGQTHQIIGGTQTQTGCSQGWGQE
RSQQTREDMVEGHGHTQVGSSTLEGGQDWSDQRSCSVTGGLOENDYCQTOEBSFGGQEWSRHQTRETVAQQQDHSRESEIHCC
GSTQLEAGQIQGKRLTARGLYSYFKCNKQ

>Md_FLG

MSHLLSSILSIIIEVYKYTSQDDQDNTLCKRELKKLLENEFRPILKNDDDEDTVEIFMQMLDRDHDKKVDFIEFLLMIFKLTM
ACNQSIGKEYCQASGSGQKHPPHHRRQEQSKTKEEEEEETDSSNSWSAGEEP RSHTGRSKKIRHRSKSNRSGHGKLDRSSSS
GYKDIFGRKQHESKPRHRSKRSENMENRSSSSDLEKMRKSSISFSRIYGGKHEFSDDWSGSSGRKIRVYSSEGLSYGKSMG
YGSNSTQSKECGEQQRCKSEQTRNCRDHHEHACAQSDKCGKQRKRSSSCQSDESCREQAYSKMCSENYGENVTNSCCQSQN
SRGKGHWSCSSQSASYGQSDSGECISGQSSRDQCQCFSSGQQRSASSRGRDVPSRRQQQDSSRQSGSSQRQSSDRRSTSR
RSSRSQQDDISRQSGSGQRESDDSRSGRGESSRRQQQDAFRGSGTSDRQARDASSRGRDITSRQQQDSSRRSGSGQRQSSD
SSRSRTRGSSSRQCFESSHSTSGQRQSSDSSRSRGGSHRRQONDSSRQSGSGQRETRDASSRGRGVPSRRQQQDSSRQSGS
SQRQSSDRRSTSRRRSSRSQQDDISRQSGSGQRESDDSRSGRGESSRRQQQDAFRGSGTSDRQARDASSRGRDITSRQQQD
SSRRSGSGQRQSSDRSSIQSSSRQCFESSHSTSGQRQSSDSSRSRGGSSRRRQHDSSRQSGSGQRETSDDSRDRDAES
GRIQDQSSGSGYQQRQSSGDRSRGRGEADRRQCFSSSRQSVSSQRQTTDASSRGRDVPSRRQQQDSSRQSGSSQRQSSDR
STSRRRSSRSQQDDISRQSGSGQRESDDSRSGRGESSRRQQQDAFRGSGTSDRQARDASSRGRDITSRQQQDSSRRSGSGQ
RQSSDSSRSRTQSSSRQCFESSHSTSGQRQSSDSSRSRGGSSRRRQONDSSRQSGSDQRETRDASSRGRGVPSRRQQQDSS
RQSGSSQRQSSDRRSTSRRRSSRSQQDDISRQSGSGQRESDDSRSGRGESSRRQQQDAFRGSGTSDRQARDASSRGRDITSR
REQQDSSRWSGSGQRQSSDSSRSIQSSSRQCFESSHSTSGQRQSSDSSRSRGGSSRRRQHDSSRQSGSGQRETSDDSRSR
DRAESGRIQEQDSGSGYQQRQSSDSSRSRGRGEADRRQCFSSSRQSGSSQRQTRDASSRGRDVPSRRQQQDSSRQSGSSQRQ
SSDRRSTSRRRSSRSQQDDISRQSGSGQRESDDSRSGRGESSRRQQQDAFRGSGTSDRQARDASSRGRDITSRQQQDSSRR
SGSGQRQSSDRRSTSRRTQSSSRQCFESSHSTSGQRQSSDSSRSRGGSSRRRQONDSSRQSGSGQRETRDASSRGRGVPSRRQ
QQDSSRQSGSSQRQSSDRRSTSRRRSSRSQQDDISRQSGSGQRESDDSRSGRGESSRRQQQDAFRGSGTSDRQARDASSRGR

DITSRRREQDSSRWSSGSGQRQSSDSRRSSIQSSSSRQQLSSSTHSTFCQRQSSDFRSRSRGGSSRRRQHDSSSRQSSGSGQRETS
DSRSRDRAESSGRIQEQDSSQSGYQGRQSSDSSSRGRGEAGRRQQLDSSSRQSSSSQQRTRDASSRGRDVSSRRQQDSSSRQSSG
SSQQRSSDRRSRRRRSSSQDDISRQSSSGQRESSDSSRRGRGESSRQQQDAFRCSSTSDRQARDASSRRDITSSRQQQ
DSSRRSSGSGQRQSSRRSSSIQSSSSRQQLSSSTHSTFCQRQSSDSSSRGGSSRRRQHDSSSRQSSGSGQRETDSSSRDRAE
SGRIQEQDSSQSGYQGRQSSDSSLRQKKEAGRRQYEYSFRQSDCDRESIGFISLGEASGHGENQYSSGHSYELIQSNHL
IRSSQASDSPGAFVAFHRHFGVRAQQQKQESGRHWVHCYSYSAEYNYQTFRHSFGSRTSIHNSSSMSGSEHVSLSHLSSLLCHS
CLRNDYSSDDMSTFGQEVFDESIEQSGEISFYSGLSSTEMSRDRFSRILGSGQESSHQAEVFSQVKNESSKNNKQSSKKRNT
PTKSTCSQSSSRKNGIHEPICDQSSDRNSQSSSSHQITYEHTYNKNSNNQSFDSFDQCTREPRHNNSSSTGRKRSVSSSS
VPGDSNGCKQDIVGSRQLEVFHIRERTNSSSYYPSSITFLYEYMQEQRCNNY

>Md_FLG2 (XP_016285131.1)

MSVLLQSIIVNIIDIFYQYAEDDGECETLSKKEMKELLEKELSSIMSNKKDEQIIEAIFHILDQDHDHKVNFAEFLLMVFKLAM
SFNKSVSKEYCEASSGSKHKSQDHHHKEKSVTEEEEDETGSDTDAEEQCCSFHGSQKKCKYRNQSRNSFWKRRRSSSSNSEH
KKDSRKRRHHKSRKNEKEKNGLSSKKNGRRYKSSSTHEKHGSGTGRRELSDQOTDSEGGQTRDQSRHKVSEQQSSFTHSSSSSR
GRKSSSCQSSSGSERQSKHRESSSTKRNSDHTRGRQKSNQNSQESDSCQIIDHSRESVSVHEKKNSSDSCAISSHRERCSQCC
PSNSESQTEDSSGESRLGEQKTSSKEKKKGNQRSRQQAEDTSAHLGSGKCKSKDRSKTRRQTSSHHSSSENKQGSTERSHHS
GSGHQSSSSQGHRCSSQRGRQOSTSSQVSDSEGSTREHTRESVSGHRQSSSSSSSSTSRRQSSSHGSSSHRQRQQSDRSSH
SSSGHQSSSSQGRHGSQRGRQOSTSSQVSDSEGSTREHTRESVSGHRQSSSTSSSGSRSSRRQSSSHGSSSHRQRQQSDRSSH
HSGSGHQSSSSQGRHGSQRGRQOSTSSQVSDREGSTREHTRESVSGHRQSSSPSSGSSSSRRQSSSHGSSSHRQRQTTRQI
QLRIWSSSTILIISETT

>Md_HRNRL1 (XP_016284680.1)

MSQLLRISIVTVIDVFYNYGQDEECDTMKKELKELLYKEMGELLKNNNNTVDVIFMQILDRDHDRRVDFTEYLLMVFKLTM
ACNKSVCKEYCHASGSKRKQGRHHQEKQSETEDQELKCHESSSSSSWSSGEEYSSRSRSRQSSKHGHKYSSTGWRKKD
YSGNGEGSKNYHRSSSSSSRSSQEKHGFKDNEQWRRRNSSISSRKSGEEYESGTDQSGRWRRQRHGSNSQTGFGVQ
EEGSKRGRGREQHWRSRSHQSSSYERESGANQSSTYEYQHGSSSEQYSNYGQNESGSGQSSSHGKCECGSGQSSNQSSSFG
SGSGQSSRRRRHGSSSSGQSGSWGQCYSSGSNQSRHRQHSGTGQSTRSGQQSGSGQSSSYGQHSGSGSGHSGSCGQGH
YSSSHQSSSYGNHESGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSHGQYSGSG
QSSRYGQHSGSGQSSSHGQYSGSGQSSRYGQHSGSGQSSSYGQHSGSGQSSSYGHHGSSSGHSEFCGQEQYSSSQS
SSHGYHGSSSGQSSNYGQHSGSGQSSSYGQRGSGSSSGYQHGSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQ
HSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSG
SGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSG
HSECCGSEHYSSSKSSSHGHGSGSGQSSYGQRSGSSQSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSS
SYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGSGHSECCGSEHYSSSQSSSH
GHHGSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSHSQHSGSGQSSSYGQHSG
SGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSG
SYGQHSGSGQSSSHSQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQH
GSGSGQSSSYGQHSGSGQSSSHSQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSG
HSGSGQSSAYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSHSQHSGSGQSSSYGQHSGSG
QHSGSGQSSSHSQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSHQHSGSG
SGQSSSHSQHSGSGQSSSYGQHSGSGQSSSHQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSS
SHSQHSGSGQSSSHSQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQH
SSQRRQRSSSGSRSESQGGRRYSSTRSSSYGRCSGCRSSSYGQRRSYGQSSSCSGSSSSSQRRYRSSGQCGRSS
SSHRSSSYGRHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSGQSSSYGQHSGSG
GQSSSCSGSGQSSQDQYSSSGGQSMNCGQQGHSSSYGQSSSYGQCCSGCDQSYQYCCSSSSGQSGAAQYSSSGCE
SCVTEHTFDWGRHKFSSNSICGEFGKQTVGSTFIQYEDINNESQVDRNHENKEVCLSQGVDDSLKVRGNYGRVRTHSGSY
FCSTITFLYEYCQEQRFYHK

>Md_HRNRL2

MSQLLTRIATIIDVFNQYCGQDKCDMISQOELKEFLENELQFIVQNSEDEAVDVIMLNMDDLQSGRGLHTDLVFIKIL
LMKKCNKKSSQNSYHMEIEKEEKRWKDLMRSNWNQKEEYERRSKYSLETEKVRYRLHGEERKDSVSSSDEEKDSEKK
NYTSNSSKAGNNKDEKTVSRKQWRKKESERPSKKSGGEKRETDYERSSQGERKKQDMKSTVGNNESSNRGQTVRFSE
EKORTCQEKTNIIYRRNYAMGNENCDIETNDSNKFSESQIGSGSKSYNSKRHGTTSCKESETCETKKSSSGHQSSKSSQY
SFFYINGTYSDSGSDSECCTCSDYRNDTGTSSSSSSNQYGSQSGHSSSTKCHKSSSKKASSMGGRYNSSSSSAQTHSSCG
SGSGSDSGSGTDSGSGSESCSGSGSES SGHSSSKRQGSRSYSSRSRHHGSSSSQSSHRGHRSSSGEESVSSGRKSSSS
GSSHRSSSNTOYGSQSGRSSSKHSGSSGKQSGSSGAKGSSSSSHQSSSSGQYSGSKSSSSSGSGSGSGSGSGSRSS
SSDTSSSSSYGSRSGRSSSKHSGSSGKHSSSGAKGSSSSSHQSSSSGQYSGSKSSSGSGSGSGSGSGSGSNKSSSSG
QYSGSGHSSSKHSGSSGKSSSSGGRHSSSSSSSRHRSGSASGSGSESCSGSGSES SGHSSSKKQGRSGHSSR
SGRHGSSSSQSSHRGHRSSSGEESVSSGRKSSSSSSHQSSSSSYGSRSGRSSSKHSGSSGKQSGSGAKGSSSSSHQSS
SSGQYSGSKSSSG
QYSGSGSKSSSG
SGSGSES CS SGSES SGHSSSKQGRSGHSSSRGRHSSSSQSSHRGHRSSSGEESVSSGRKSSSSSSHQSSSSSYGSR
SGRSSSKHSGSSGKQSGSGAKGSSSSSHQSSSSQYSGSKSSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG
GSKHSGSSSGKSSSSGGRHSSSSSRHRSGSASGSGSES CS SGSES SGHSSSKQGRSGHSSSRGRHSSSS
SQSSHRGHRSSSGEESVSSGRKSSSSSSHQSSSSSYGSRSGRSSSKHSGSSGKHS SGSGAKGSSSSSHQSSSSQYSG
SKSSSG
SG
CSG
HSSGSKQGRSGHSSSRGRHSSSSQSSQRGHRSSSGEESVSSGRKSSSSSSHQSSSSSYGSRSGRSSSKHSGSSGKQ

[illegible]

GFQDSSYGQTHRGGLDTGYGYSEVQGNSSQFQNSGRRGQAEQDLNRNDHVSGEDRGHFHQGLWQPERESESRQHKVYSQS
 QQERTLCHKGREWQKDTGTETGGQAQDRRGCEEKKSRYRYDQOQTHKEEQSHQTRDRQGHEEQGTSSQQRNRQSCGNEQSHWTQ
 DNQTQNRQTRDNEKYHQTRDRQDHQCSNEQSHQDRQTRGEEESSQQTQSTREAHEGQRQOKSQNRQYQGNQONHQPOTRGS
 THESHHVRHQIQGSRRRSGDQSSYTQNHGNRGGASNGRHTTNSAVASNLFDYDVEQKAQGHNN

>Md_TCHH

MEQLLKSIIVNIVEVYNQYAKSDCDGTALSKKALKNLQREFGDI LRHDEQTVDLVLQLLDRDCNGVVDNFNEFLLLLFKVAQ
 ACYYALSQATGNEGKRRVKLEEKRERSQDFLREENRRRDTRERQLQEEELQQAEEQERERQRSDARSLEFERQRQERVQSA
 SEEEQLQRQEQERRLQERERQSLDEQLQRPKRERQERQIREEEQRQROELEQERRRREQRQESAEQYKROEQERQREQR
 ERQEESSLOERADQSLSEEEQLLRQERKVRKESRSLRLQWLQEREVERGRQNKVYSKRRQERQFLEEESSLRQOREEERRR
 QQERERQLLEEEALQRDRFGRLRDEDFEDSFQEEERARQLREEEQIRRRQIRDRQLQEEERRRDLKIOWLFEETERRRRNRLY
 AKSEQQQERERRRRVQRERQLREEERELQQLERRRQEQLYREEERLEQEEELRRRQEREEKRQRQERERQLREEEALQQQ
 DRFRRLRDGDLDDSFQEEERARQLLEEEQIRQOQIRGRQLQEEELGRDLKIOWLFEETERRRRNRLYAKSEQQQERKRREQL
 RQRKEEEELQEEELQRLQLEQELEKKREELEKRRRQERERQYREEELQEEELQRLQLELENKLEELEKRRRQERQREKTLF
 W

>Md_TCHHL1 (XP_007485516.1)

MEQLLKSVICVMEAFQKYAKEEGDCWTLNSSLKRLLLGEIGEFKIFDILTAGTSLHFLDRDGDGSSISFDEFILLIFDLLNI
 CYQDIHSFLNQEPKPKSNTTEKFSGDTEACEASEDYQKVAGPDQYEQRLTGTESPSLVNIEKVPDTSKEDPQDDLESFKLL
 GKEVEHNHFKSQYYGEEAKQSQERSQYVQATGADGIPPEEKPKPKFVKRTCSQKDERIVSEEHEVPEQVDDKDRDHSSEQE
 KEQNVIEIQSVHMEETAQRTSESSEGTIIKDVGEHTDTQGLASQEDHERKLGTTDLPAEKAEKLSETERLEPRDDDGTSIDIQ
 EPLQISEKQEQKEYETTKIPAKNGSRVPETEVFIDEKKEKKPETLDTVGKKDGEKTKQMETLLEQAEQKEKQLEGLEE
 KGEIRKSGGIDFKSVDDKNHYEHKFAAEIEIKVSEISESEYQGDTEETDRVQVSSQDDQSKGNRRRIQILDQIEQY
 DRHQVENQDDGTLSETYSNPEDEKSGSETRDLPAQKKSQGVDAQEQDQGDCKNYLGTQESLVEGDKSRTFEKEVLIDKDNE
 NITEEQEKLAKKKSDTAEESSQSESQESMKQKATETSPKIEIKNTDDIDEQLSIVSAEKEDHRKDFEAHGEAEKQGE
 ESESQEALEGGQEPETQGLALDARDTLNFIEREDKSAQKLAGEGSELQYAVNKEDGSSGLSELIKKARDTKLSSSGDEI
 QSNFIYENIQKTTEQDILLDTEHLDQTDTSQTSSELVEREIQKSECSDFEALLDLYDDTQELRNQTDGSESDEEYGNQQ
 EILAYQSQEDIHGGQEQEDQQLRDBSSRKY

>Md_TCHHL2 (XP_007485517.1)

MEQLLKGIITVIDVFYKNAWTDGGCQRLSKQELKQLLQGFGEALQKKGHSETTDKILQLLDKGDGTVDSEFVFLVFSVVK
 ACYACIQPLLCPLLEGGRRYETQEPQKGRTEIDPLGSNELPATTDNHRDDENNOPTQNAACRVTAGQKESSEETRIHIR
 DENTATRTSEGHHIGGETSTFRSNERKDFVNEDGGKYNVRDRFSRDDRDEVKQTEQRESGKRHFETITPMSDNRRNVI
 EEVETQMNGRMYHDREQFPAQGDDERLQLRDEISERDEERYTVREKSEKVRYNRRYQMSQIEPQALEKRNQIKVQISSAER

C

>Md_PGLYRP3

MKMFVGFLLFSSILCLGCCDGVFDVLSFDGNETESLFFNFQDLINNITLLIEKVKTPEYGGSFQMVSRSEWKKGPGGNIQLRTP
 VEYLIHHILQEQCHEKATCRQVKGLOELHIKINGWQDVAYNFLIGEDGNVYEGLGWTLGTHMGYNRKSLGFAFVGSAG
 SSLSAAALTAENLISFAVNGYLSKRYIQELFVQSESCLACYQKMSKKCEPDIVRSSWGAQDTCCKLSEPAKYVVIHT
 GGRNENETEECQIALRYIQSYHIEKMKFCDIAYNFLVGEDKAYEGVWDTEGAHTYGYNDIGLGIAFMGLFTDNPPNDAALK
 AAQDLIQCSVDKGYLDLDYLLVGHSDVVNTLSFAQALYDQIKTCFHFKH

>Md_S100A9 (XP_001372115.3)

MENCTMEKALDIIVNTFHHYSTRVGNEDTLVKGEMKQLITKELENFIKNAKDLQDVKHLMQELDTNNGQVDFKEFSMMMARL
 TMATHEKMHENAFDKDHHSHGEGLECKGSSCGSGHGHSH

>Md_S100A11 (XP_001366952.1)

MAKKVYETEIERCIESLIQVQRYACQEKTNSSTTLSKTEFLRFMNTELASFTQNDNGVLDMMKKLDLNCQGLDFQEFLL
 NLIQGLAQACHKSFQDSQNIKKK

Supplementary Figure S3. Amino acid sequences of proteins encoded by EDC genes of opossum. (A) Amino acid sequences of proteins encoded by SEDC genes of opossum. **(B)** Amino acid sequences of opossum SFTPs. **(C)** Amino acid sequences of proteins encoded by other EDC genes of the opossum. To show the peculiar amino acid compositions of SEDCs and SFTPs and the importance for protein cross-linking the following amino acid residues are highlighted: lysine (K) and glutamine (Q) as potential sites of transglutamination; cysteine residues (C) as potential sites of disulfide bonds; glycine (G), proline (P) and serine (S) are highly abundant residues not directly involved in cross-linking. When available, the GenBank accession number is shown. Only the S100A proteins whose genes are flanking *PGLYRP3* and *TCHH* are included here. SEDC, single coding exon epidermal differentiation complex; SFTP, S100 fused-type protein; Md, *Monodelphis domestica*.

1 87

Hs_CRNN M-PQLLQNINGIIIEAFRRYARTEGNCALTTRGELKRLLEQEFADVIKPHDPATVDEVLRLLDEEDHTGTVEFKEFLVLVFKVQAQCF

Bt_CRNN M-PQLLRNINIGIIIEAFRRYARMEGDCAVLERGELKRLLEKEFADVIKPHDPATVDEVLRLLDEEDTGTVEFKEFLVLVFKVQAQCF

Em_CRNN M-PQLLRNINHIIIEAFGRYAKTEGNCMVLTTRGELKRLLEHEFADVIKPHDPATVDEVLRLLDEEDTGTVEFKEFLVLVFKVQAQCF

Md_CRNN M-PQLLGNIIDGIIQAFSRYAKTEGDCITLTGELKRLLEQELADVIKPHDPATVDQVFHLDEDSKGTVDKEFLVLVFKVQAQCF

Oa_CRNN M-PQLLGNIIVGIIQAFNSYARTEGDCITLSRGEKRLLEKEFADVIKPHDPATVDVTLHLDDADGKVGFTFLALVFRVAQACF

Hs_FLG2 M-TDLRLRSVVTIDVIFYKYTKQDGECSLKGELKELKELEHLPVLKNPDDPTVDVIMHMLDRDHDRRLDFTEFLLMIFKLTMAACN

Em_FLG2 M-TDLRLRSVVTIIDIFYKYTRQDGECSLKEELKELKEFHPILKNPDDPTVDVIMHMLDRDHDRRLDFTEFLLMVFKLAMACN

Bt_FLG2 M-TNLLRSVVTIDIFYKYTKQDGECSLKGELKELKEFHPILKNPDDPTVDVIMHMLDRDHDRRLDFTEFLLMVFKLAMACN

Md_FLG2 M-SVLLQSVINIIDIFYQYAEEDGECSTLSKKEMKELKELELSSIMSNPKDPQIIIEAIFHILDDQDHDKVNFAEFLLMVFKLAMSFN

Hs_FLG M-STLLENIFAIIINLFQYSSKKDKNTDTLSKKELKELKEFHPILKNPDDPTVDVIMHMLDRDHDRRLDFTEFLLMVFKLAQAYY

Bt_FLG M-STLLENINDIKIFHKYSKTDKETDTLSEKELKELEVEFHPILKNPDDPTAEVFMYNLDRDHNNKIDFTEFLMVFKVAQVYVY

Em_FLG M-STLLENITAIIDLQYSSNDKENDTLSKKELKELELEFHPILKNPDDPTADDFMHILDLHNNKVDFTTEFLMVFKLAQAYY

Md_FLG M-SHLLSSILSIIEVYKYTSQDQDCNTLCKRELKELKEFHPILKNPDDPTVEIFMQMLDRDHDKKVDFTIEFLLMIFKLTMAACN

Oa_SFTP1 M-SQLLKSIVTVIDIFYHYTQGDGCETLSKGELKELKEFHPILKNPDDPTVDVIMQILDRDHDRRVDFTEFLLMVFKLTQACN

Ta_SFTP1 M-SQLLESIATIIDIFYHYTQGDGCETLSKRELKELKEFHPILKNPDDPKVDAIMQMLDQDHNRVNFTEFLLMVFKLTQACN

Ta_SFTP2 M-SQLLKSITATIIDIFYHYTQRDGYETLSKRELKELKEFHPILKNPDDPTVDIIMQMLDQDHDRRVNFTEFLLMVFKLTQACN

Ta_SFTP3 M-SQLLKSISTIIDIFYHYTQRDGYETLSKRELKELKEFHPILKNPDDPTVDVIMQMLDQDHDRRVNFTEFLLMVFKLTQACN

Hs_HRNR M-PKLLQGIVTIDIFYQYATQHGEYDTLNKAELKELKEFHPILKNPDDPTVDIILQSLDRDHNNKVDFTTEFLLMIFKLVQARN

Bt_HRNR M-PKLLQSVTVIDIFYQYANQAGCDMLNKAELKELKEFHPILKNPDDPTVDIIMQMLDQDHDRRVNFTEFLLMIFKLVQARN

Em_HRNR M-PKLLPSIVSVIEIYYQYATEEGECNSLNRAELKELKEFHPILKNPDDPTVDIIMQSLDRDHNNKVDFTTEFLLMIFKLVQARN

Md_HRNR M-SQLLTSMATIMDIFYYCGKDEECDTMSQSELKELKELELQFHPILKNPDDPTIDIMMLNMDLHNNKIDFTEFLLMVFKLIMILN

Md_HRNR1 M-SQLLRSIVTVIDIFYNYCGQDEECDTMCCKELKELKEFHPILKNPDDPTVDIIMQMLDQDHDRRVNFTEFLLMVFKLTMAACN

Md_HRNR2 M-SQLLTRIATIIDIFYNYCGQDCECDMSQSELKELKELELQFVQNSSEDSEAVDVIIMLMDLQSGRGHLDLVFQIIFKLTLMK

Md_HRNR3 M-SQLLTSITITLDIFYQYCGEDEECDTMSQSELKELKELELRFHPILKNPDDPTVDVIMLMDQDRDKRVDFTEFLLMIFKITMSFN

Md_TCHHL1 M-PQLLQSVICVMEAFQYAKTEEGDCWLTNSGQLKRLLEGEIGEFKLPF-DILTAGTSLHFLDRDGDGSSIFDEFLILIFDLNICY

Hs_RPTN M-AQLLNSILSVIDVFHYAKNGDCALLCKEELKQLLAEFGDILQRPNDPETVETILNLDRDQDGHIDFHEYLVLVFLVQACF

Bt_RPTN M-TELLNSILTVIRVFQYAKENGSTSLCKEELKQLLAEFGDILRRPNDPETVETILSLDRNRNEHVDFHEYLVLVFLVQACF

Em_RPTN M-AQLLSSILTIVIKVFQKHAENGDCSTLCKEELKQLLAEFGDILWRPNDPETVETILTIDRDSNGHIDFHEYLVLVFLVQACF

Ta_RPTN M-SQLINSILRIIEVFQYANADNCSVLTTELSELLQAEFGNLRRRPDPKTVDTILQLDRDRDGVVGFNEFLLMVFKMAQACH

Md_RPTN M-APLLNSILTVIEVFHEYAKENDCTSLGKELKQLLLEFQELRRPNDPETVDTILQLDRDHDEHIDFNEFLLMVFKLAQACF

Hs_TCHH M-SPLLRSICDITEIFNQYASHDCDGTLSKKDLKNLLEFEGDILRRPNDPETVDLLEQLDRDSCNGVDFNEFLLMIFKLVQACF

Bt_TCHH M-SPLLRSIFNITIKFNQYASHDCDGTLSKKDLKNLLEFEGDILRRPNDPETVDLLEQLDRDRGLIDFHEFLAIVFKVAAACF

Em_TCHH M-PALLRSICDITEIFNQYASNDCDRAALCKKDLKELLEFEGDILRRPNDPETVDLLEQLDRDHNLIDFHEYLVLVFLVQACF

Md_TCHH M-PQLLKSIVNIEVYNYQYAKSDCDGTALSKKALKNLQREFGDILRRPNDPETVDLVLQLDRDCNGVDFNEFLLMVFKVQACF

Ta_TCHH M-SPLLKSIIIDISEIFNYATCDGDDVKLNKGLRLLQREFRDVLRPPQDPQTVDLVLQLDRDRDGSVDFNEFLLMVFKVQACF

Hs_TCHHL1 M-PQLLRNVLVCIETFHKYASEDSNGATLTGRELKQLIQEFEGDFQPC-VLHAVEKNSNLLNIDSGAISDFEFLAIFNLLNLCY

Bt_TCHHL1 M-PQLLRDILCVIETFHKYAREDA--ATLTCTELKQLIQSEFEDIFQPC-AIHAVEKNLNLNIDSGAISDFEFLAIFNLLNLCY

Em_TCHHL1 M-PRLLRRVLVCIETFHKYAREDGVTLTHRELKQLLQEGFDILQPH-VMAHEKNVNLIDIGSDGTIRFDQFVLATCNLLNHCY

Oa_TCHHL1 M-PRLLKSIVTDVMEVFHYQAEQDGGQAVLTGDGLRQLLSELDGILQRSPPDGRYVGEVDLDDANHDGVDFNEFLILVFLGLNACY

Bt_TCHHL2 MTHRLRSIISIIDAFLPNAKSDGDCQSLNKTTELKQLLEFEGNALAESNNSETTGKILQQLDQDQDTIDFSELILLMFVAVTTAYY

Md_TCHHL2 M-PQLLKGIITVIDIFYKNAWTDGGCQRLSKQLKQLQEFGEALQKPGHSETTKILQLDKDGDGTVDSEFVFLVFSVVKACY

Supplementary Figure S4. Amino acid sequence alignment of S100 domains of SFTPs. Amino acid sequence alignment of SFTP proteins of human, cattle, elephant, opossum, platypus and echidna. Amino acid residues conserved in all species are colored in red. Amino acid residues and highly similar residues (D/N, D/E/Q, L/M, I/V, F/Y) conserved in >50% of all investigated species are colored in blue. Species: Human (*Homo sapiens*), cattle (*Bos taurus*), elephant (*Elephas maximus indicus*), opossum (*Monodelphis domestica*), platypus (*Ornithorhynchus anatinus*) and echidna (*Tachyglossus aculeatus*). Accession numbers: human CRNN, NP_057274.1; human FLG2, NP_001014364.1; human FLG, NP_002007.1; human HRNR, NP_001009931.1; human RPTN, NP_001116437.1; human TCHH, NP_009044.2; human TCHHL1, NP_001008536.1; cattle CRNN, NP_001178232.1; cattle RPTN, XP_010801377.1; cattle TCHH, XP_002686080.3; cattle TCHHL1, NP_001193524.1; cattle TCHHL2, XP_024845739.1.

A

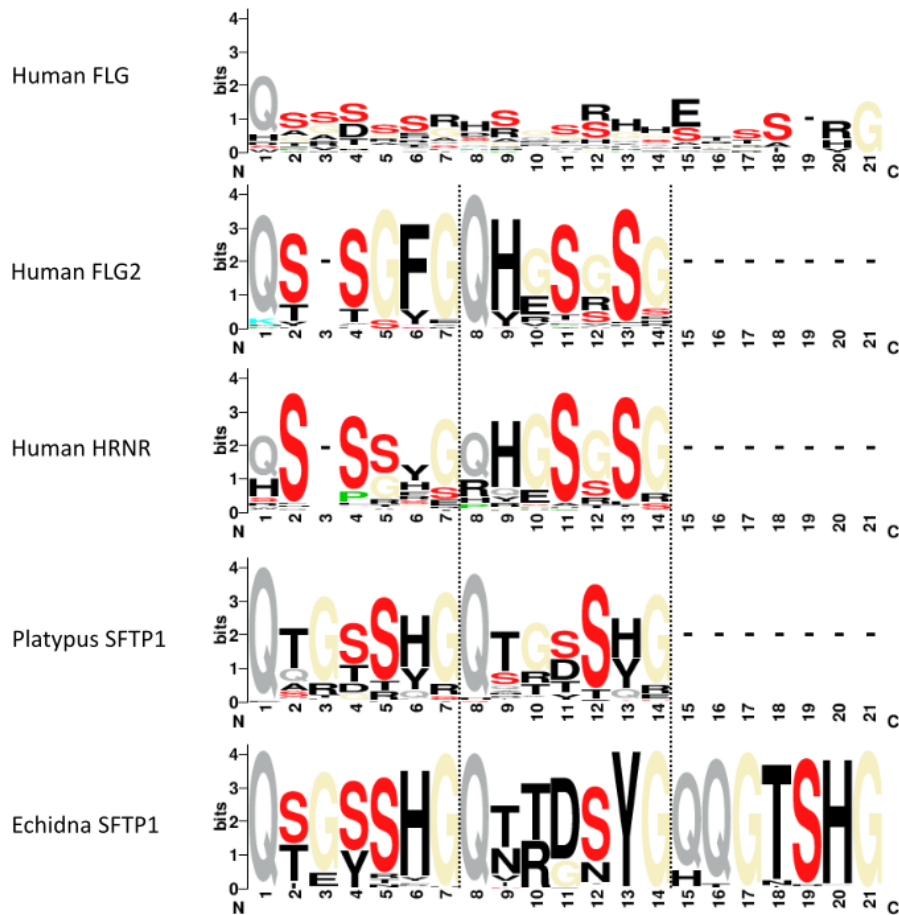
	M A Q L L N S I L S V I D V F H K Y A K G N G D C A L L C K E E L
Human <i>RPTN</i>	ATGGCTCAACTCCTGAATAGCATACTCAGTGTGATTGACGTATTCCACAAATATGCCAAAGG-GAATGGGACTGTGCCTTACTATGCAAGGAA-GAGTT
Platypus <i>RPTN</i>	ATGTCTCAGCTCGTCAGAGCATCCTCAGGGTCATCGAGGTTTCCAAAGCTATGCCAATGC■GGACGACA■ACTGCGTCTCGCTGGACCAAGAC■GAGCT
	M S Q L V R S I L R V I E V F Q S Y A N A G R Q L R L A G P R P S
	frameshift frameshift

	K Q L L L A E F G D I L
Human <i>RPTN</i>	GAAACA■ACTGCTCTTGCTGAGTTTGGAGACATCCTCC
Platypus <i>RPTN</i>	CAAGGA■ACTTCTGCAAGCAGAGTT■CGCAACCTTCTGA
	S R N F C K Q S S A T F -
	stop

B

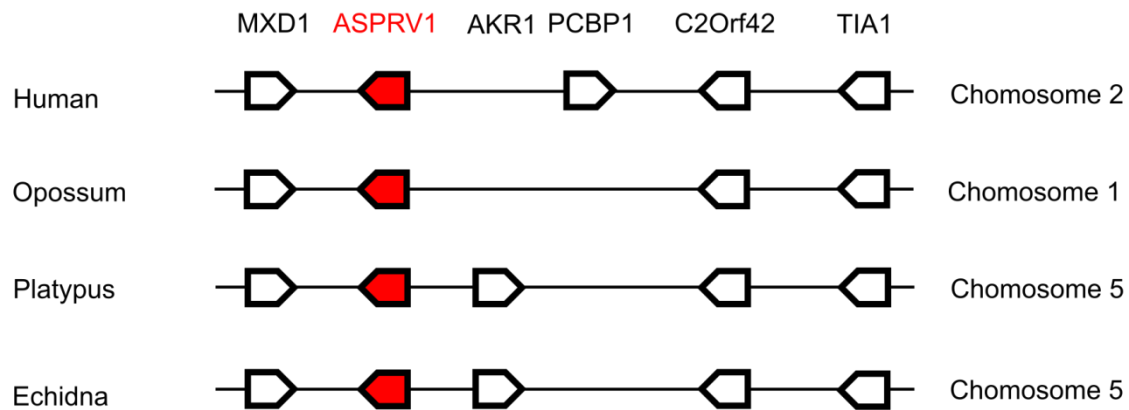
	M S P L L R S I C D I T E I F N Q Y V S H D C D G
Human <i>TCHH</i>	ATGTCCTCCACTTCTGAGAAGCATC-TGTGACATCACTGAAATTTCAATCAGTATGTCCTCATGATTGTGATGG
Platypus <i>TCHH</i>	ATGTCCTCCCTCCTGAAGAGCATC■TCCGACATTTCCGAGATCTTCAACCACTACGCCACGTGCGACGGCGATGA
	M S P L L K S I I R H F R D L Q P L R H V R R R -
	frameshift stop

Supplementary Figure S5. *RPTN* and *TCHH* are pseudogenes in the platypus. (A) Nucleotide sequence alignment of human and platypus (*Ornithorhynchus anatinus*) *RPTN* until the first in-frame stop codon of the platypus *RPTN* pseudogene. Nucleotide insertions leading to frameshifts are marked by red shading. The corresponding amino acid sequences are shown above and below the nucleotide sequences. **(B)** Nucleotide sequence alignment of human and platypus *TCHH* until the first in-frame stop codon of the platypus *TCHH* pseudogene. A nucleotide insertion leading to a frameshift is marked by red shading. The corresponding amino acid sequences are shown above and below the nucleotide sequences. Note that the premature stop codons in platypus *RPTN* and *TCHH* are located in coding exon 1. Supplementary Table S1 provides information about accession numbers and the positions of the platypus *RPTN* and *TCHH* pseudogenes from the start codon to the first in-frame stop codon.



Supplementary Figure S6. The sequence repeats in the carboxy-terminal domain of platypus and echidna SFTP1 proteins are similar to those of human FLG2 and HRNR. Comparison of sequence repeat motifs of human, platypus and echidna SFTPs. Vertical dotted lines mark the borders of segments of similar sequence within longer repeats. Amino acid residues are present in the one-letter code. Dashes mark gaps that were introduced to optimize the alignment. Note that human FLG does not contain sequence repeats of a length similar to those of the other SFTPs shown here, resulting in a sequence motif with low conservation (small height of letters) at all positions. Species: Human (*Homo sapiens*), platypus (*Ornithorhynchus anatinus*), echidna (*Tachyglossus aculeatus*). GenBank accession numbers: human *FLG*, NP_002007.1; human *FLG2*, NP_001014364.1; human *HRNR*, NP_001009931.1. SFTP, S100 fused-type protein.

A



B

Human	ASPRV1	MAGSGAR---	SEEGRRQHAFVPEPFDGANVVPNLWLHSFVINDLNHWDHITKLRFLKESLRGEALGVYNRLSPQDQDGYGTVKEALLKAFGVPGA---	100
Opossum	ASPRV1	MAGREVR---	SQSGRT-HAFTEHFDGHNLDVYRWLHQFNVISDLNNWDDSTQMKFLTQSLKGDALDVYTGSLPEAKGNCQSVKDALFEAFRSESDFEHK	
Platypus	ASPRV1	MAEAGDREAGSRGAREHAFLEPPFDGSRGDAGRWLHQLVFIADLNHWDEATKLRFLSSALRGGPLDLFRGLGPEARRSYPAVRDALLKAFG-----	G	
Echidna	ASPRV1	MAEAG-----	SRELPREHAFLEPPFDGSRGDAGRWLHQLVFIADLNHWDEATKLRFLGSAALRGGPLDLYGGLGPEARRHNYPAVRAALLGAFG-----	G
Human	ASPRV1	PSHLPKEIVFANSMGKGYLLKGGIKGVVPRFLVDSGAQVSVVHPNLWEVTDGDLDTLPFENVVKVANGAEMKILGVWDTAVSLGKLLKAQFLVANAS	200	
Opossum	ASPRV1	QSHGPREIVFANSMGKGYLLKGGIKGVVPRFLVDSGAQVSVVSPDLWEQATDGDVDTLRPFENVVKVANGAEMKILGIWDTKVTLGKFEMEAEFLVANAS		
Platypus	ASPRV1	PVPGPREILLARSMGKGYLLKGGIEGVVPRFLVDSGAQVSAHPDLWEQATDGDPSLTPFENVVKVANGAELKILGIWDTVITLGKLEMNAQFLVADEA		
Echidna	ASPRV1	P--GPREILLARSMGKGYLLKGGIEGVVPRFLVDSGAQVSAHPDLWEQATDGDPSLTPFENVVRVANGAELKILGVWDTVITLGKLEMNAQFLVAGEA		
Human	ASPRV1	AEAAIIGTDVLQDHNAILDFEHRTCTLKGGKFRLLPVGGSLEDEFDLELIEEDPSSEEGRQELSH	265	
Opossum	ASPRV1	TEAAIIGTDVLQDHGAAILDFKHRTCTLRGKKFRLLPVGGSIEDEFDLELIEEETSDH-----		
Platypus	ASPRV1	TEAAIIGTDVLQDHGAAILDFKHRTCTLKGGKFRLLPVGGSLEEEFDLELIEERP-----		
Echidna	ASPRV1	AEAAIIGTDVLQDHGAAILDFKHRTCTLKGGKFRLLPVGGSLEEEFDLELIEERP-----		

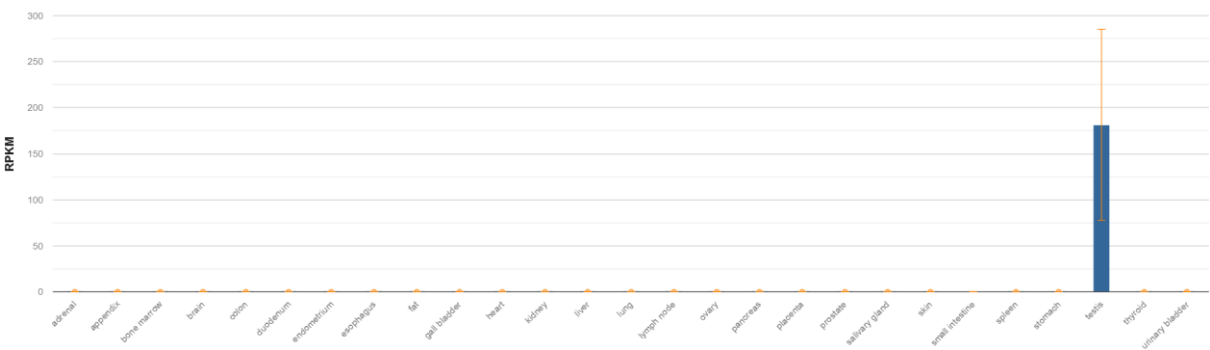
Supplementary Figure S7. ASPRV1 is conserved in monotremes and therian mammals. (A) Gene locus of *ASPRV1* in human, opossum, platypus and echidna. **(B)** Amino acid sequence alignment of *ASPRV1* proteins of human, opossum, platypus and echidna. Species: Human (*Homo sapiens*), opossum (*Monodelphis domestica*), platypus (*Ornithorhynchus anatinus*), echidna (*Tachyglossus aculeatus*). Accession numbers: Human *ASPRV1*, NP_690005.3; opossum *ASPRV1*, XP_007476366.1; platypus *ASPRV1*, XP_001519377.3; echidna *ASPRV1*, XP_038603102.1.

		Q F Q E E L E K F Q Q A I D S R E D P V S C A F V V L M A H G R E G F L K G
Human <i>CASP14</i>	CTCC	<u>AG</u> CAATTCAGGAAGAGCTGGAAAAATCCAGCAGGCCATCGATTCCGGGAAGATCCCGTCAGTTGTGCCTTCGTGGTACTCATGGCTCACGGGAGGGAAGGCTTCCTCAAGGGA
Echidna <i>CASP14L1</i>	TTTT	<u>AG</u> GGCTTTTCGGGACGAAGTGGTGCAATTCAGAGAGGAGATGGAAAAGCGGACAGACCCCATAAAGCTGCTGCTTCGTGGTGCTCATGGCTCACGGGAAGGAGGACGCCTGCTAGGG
Echidna <i>CASP14L5</i>	TTTT	<u>AG</u> GGCTTTTGGGATGAAGTGGTGCTTTCAGAGAGGAGATGGAAAAGCGGACAGACCCATAAACTGCTGCTTCGTGGTGCTCATGGCTCACGGGAAGGAGGAGTTCTGCTAGGG
Echidna <i>CASP14L6</i>		-----
		E D G E M V K L E N L F E A L N N K N C Q A L R A K P K V Y I I Q A C R G
Human <i>CASP14</i>	GAAGATGGGGAGATGGTCAAGCTGGAGAATCTCTTCGAGGCCCTGAACAACAAGAACTGCCAGGCCCTGCGAGCTAAGCCCAAGGTGTACATCATACAGGCCTGTCGAGGAGGTTGGGG	
Echidna <i>CASP14L1</i>	GCAGATGGGCAGGTGGTGGAACTGGAGGAGCTTTATGATGTCTTGACTAACAGACCTGCCGGGCACTCCTGGGCAAATCCAAAGTGTTCATCCTGCAGGCCTGCCGTGGGGGTCAGT	
Echidna <i>CASP14L5</i>	GCAGATGGGCAGGTGGTGGAACTGGAGGAGCTGTATG-----AAACCCAAAGTGTTCATCCTGCAGGCCTGCCCTGGGGGTCAGT	
Echidna <i>CASP14L6</i>	-----	

Supplementary Figure S8. *Caspase-14 (CASP14)* pseudogenes of the echidna contain partial or complete deletions of the exon encoding the catalytic site. Nucleotide sequence alignment of the third coding exon of the human (Hs, *Homo sapiens*) *CASP14* gene and the homologous sequences of echidna (Ta, *Tachyglossus aculeatus*) *caspase-14-like 1 (CASP14L1)*, *CASP14L5*, *CASP14L6* (Suppl. Table S5). Note that a deletion (indicated by dashes) of 38 nucleotides in Ta *CASP14L5* leads to a frame shift. This exon is entirely missing in the Ta *CASP14L6* gene, in which additional mutations have destroyed splice sites of other exons. The amino acid sequence encoded the human exon is shown above the nucleotide sequences. The cysteine within the motif QACRG (highlighted by yellow shading) at the end of the amino acid sequence is the active site of the caspase. Six nucleotides of the introns flanking the exonic sequences are included in the alignment. Splicing signals at the ends of introns are underlined.

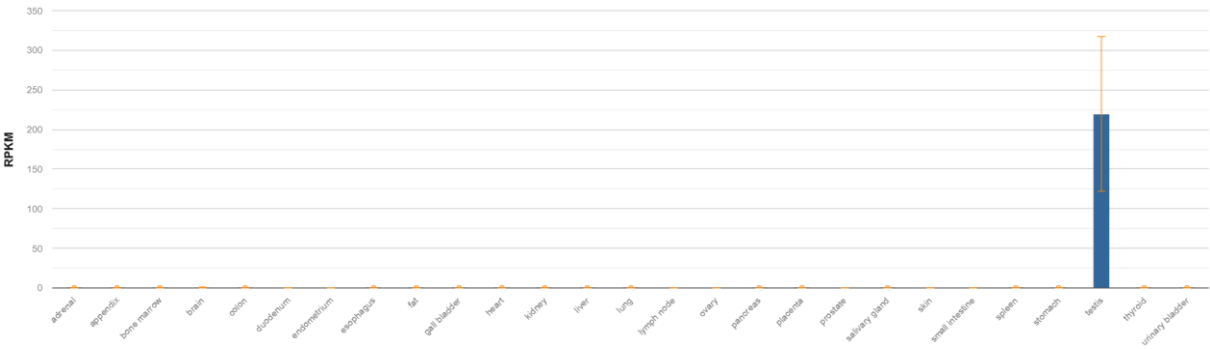
A

LELP1 mRNA



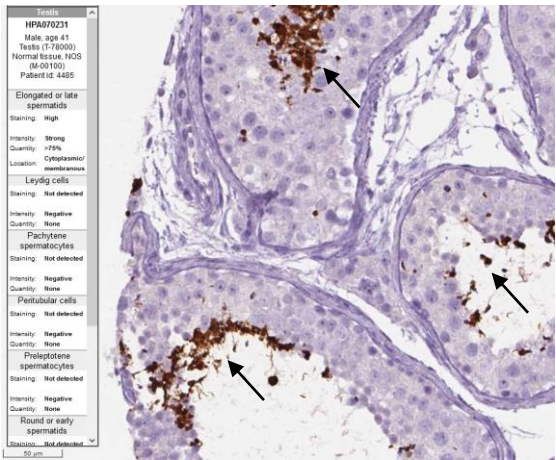
B

SMCP mRNA



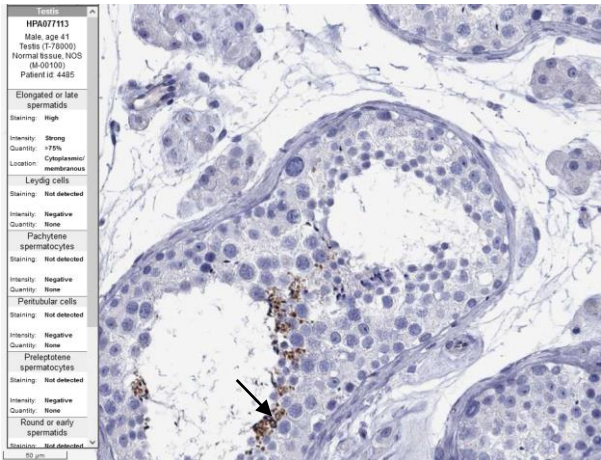
C

LELP1 protein



D

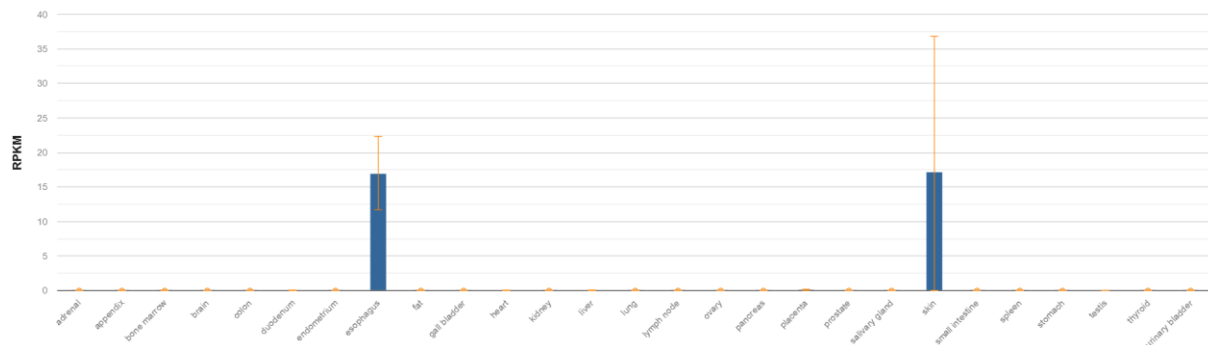
SMCP protein



Supplementary Fig. S9 (continued on the next page)

E

CRCT1 mRNA



Supplementary Figure S9. Expression patterns of *LELP1*, *SMCP* and *CRCT1* in human tissues. mRNA (A, B, E) data were downloaded from GenBank at <https://www.ncbi.nlm.nih.gov/gene/149018/?report=expression> (last accessed on 4 August 2023) (A), <https://www.ncbi.nlm.nih.gov/gene/4184/?report=expression> (last accessed on 4 August 2023) (B) and <https://www.ncbi.nlm.nih.gov/gene/54544/?report=expression> (last accessed on 4 August 2023) (E). Protein immunolocalization data were downloaded from Protein Atlas (version 23.0, release date: 2023.06.19) at <https://www.proteinatlas.org/ENSG00000203784-LELP1/tissue/testis#img> (last accessed on 4 August 2023) (C) and <https://www.proteinatlas.org/ENSG00000163206-SMCP/tissue/testis#img> (last accessed on 4 August 2023) (D) (Uhlén M et al. 2015). Arrows point to sites of immunostaining (brown). mRNA data are publicly available without restrictions at the NCBI website. They were obtained from BioProject: PRJEB4337 (Fagerberg et al. 2014). Protein immunolocalization data were not available for CRCT1. RPKM, reads per kilobase of transcript per million mapped reads.

References:

Fagerberg L, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell Proteomics* **13**, 397-406 (2014).

Uhlén M, et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).