The American Journal of Human Genetics, Volume *87*

Supplemental Data

Mutations in Centrosomal Protein CEP152

in Primary Microcephaly Families Linked to MCPH4

Duane L. Guernsey, Haiyan Jiang, Julie Hussin, Marc Arnold, Khalil Bouyakdan, Scott Perry, Tina Babineau-Sturk, Jill Beis, Nadine Dumas, Susan C. Evans, Meghan Ferguson, Makoto Matsuoka, Christine Macgillivray, Mathew Nightingale, Lysanne Patry, Andrea L. Rideout, Aidan Thomas, Andrew Orr, Ingrid Hoffmann, Jacques L. Michaud, Philip Awadalla, David C. Meek, Mark Ludman, and Mark E. Samuels

Figure S1. Multiple Sequence Alignment (MSA1) of *CEP152* Orthologs

Residues are shaded according to sequence conservation (only the region with the mutation is shown).

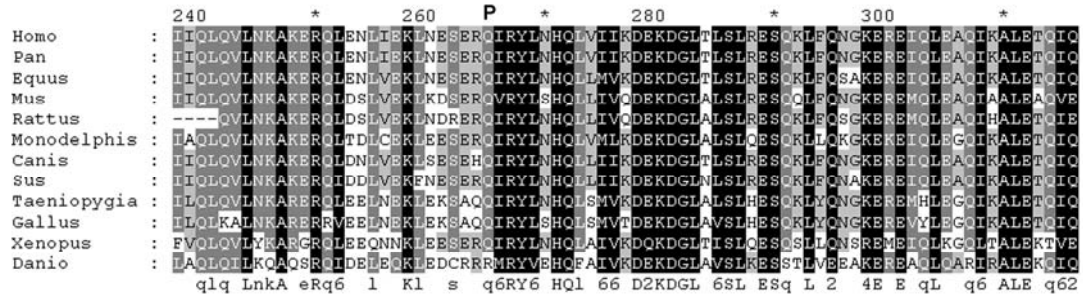


Table S1. Effects of the Mutations Predicted by SIFT, PolyPhen, and Align-GVGD

'+' and '-' indicate the mutation is predicted to have deleterious or benign effect on protein function, respectively.

Mutation	Method			
	SIFT	PolyPhen	PANTHER	Align-GVGD*
Q265P	+	+	NA	MSA1, – MSA2, +

*With the input alignment of MSA1, mutation p.Q265P is classified to C15 class, a mutation class which is less likely to affect protein function. With the input alignment of MSA2 (no Danio rerio), mutation Q265P is classified to C65 class, a class which is most likely to interfere with protein function.

Figure S2. Conserved Domain Search Results for CEP152

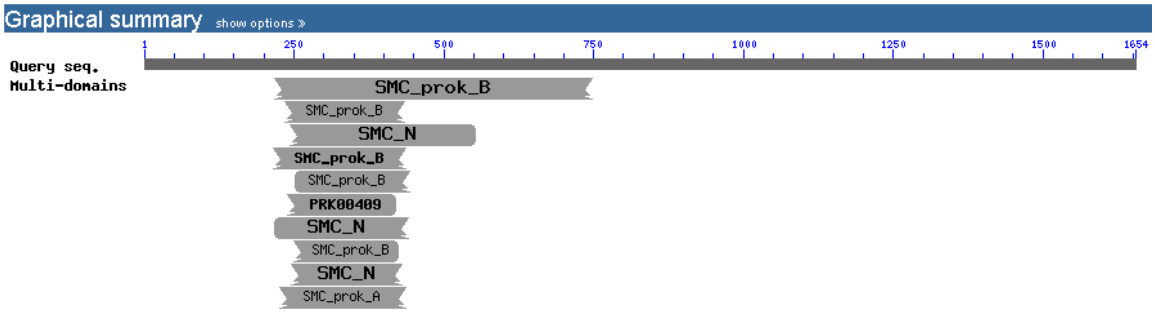


Table S2. Detailed Conserved domain Search Results for *CEP152*

Description	PssmId	E-value
TIGR02168, SMC_prok_B, chromosome segregation protein SMC, common bacterial type	131223	3e-08
TIGR02168, SMC_prok_B, chromosome segregation protein SMC, common bacterial type	131223	7e-07
pfam02463, SMC_N, RecF/RecN/SMC N terminal domain	111369	1e-06
TIGR02168, SMC_prok_B, chromosome segregation protein SMC, common bacterial type	131223	5e-06
TIGR02168, SMC_prok_B, chromosome segregation protein SMC, common bacterial type	131223	7e-06
PRK00409, PRK00409, recombination and DNA strand exchange inhibitor protein; Reviewed	134259	2e-05
pfam02463, SMC_N, RecF/RecN/SMC N terminal domain	111369	2e-05
TIGR02168, SMC_prok_B, chromosome segregation protein SMC, common bacterial type	131223	3e-05
pfam02463, SMC_N, RecF/RecN/SMC N terminal domain	111369	4e-05
TIGR02169, SMC_prok_A, chromosome segregation protein SMC, primarily archaeal type	131224	9e-05

Figure S3. COILS Predictions for Wild Type (Upper Panel) and Mutant (Lower Panel) *CEP152*

The y-axis represents the probability that the protein will form a coiled-coil at each position (x axis) based on sequence similarity with known coiled- coils. Comparison of the COILS predictions shows that position 265 is likely to fall within a coiled-coil, but for the mutant sequence, COILS predicts decreased support for a coiled-coil in the 35-residue window beginning at the position of the missense mutation (Table S3).

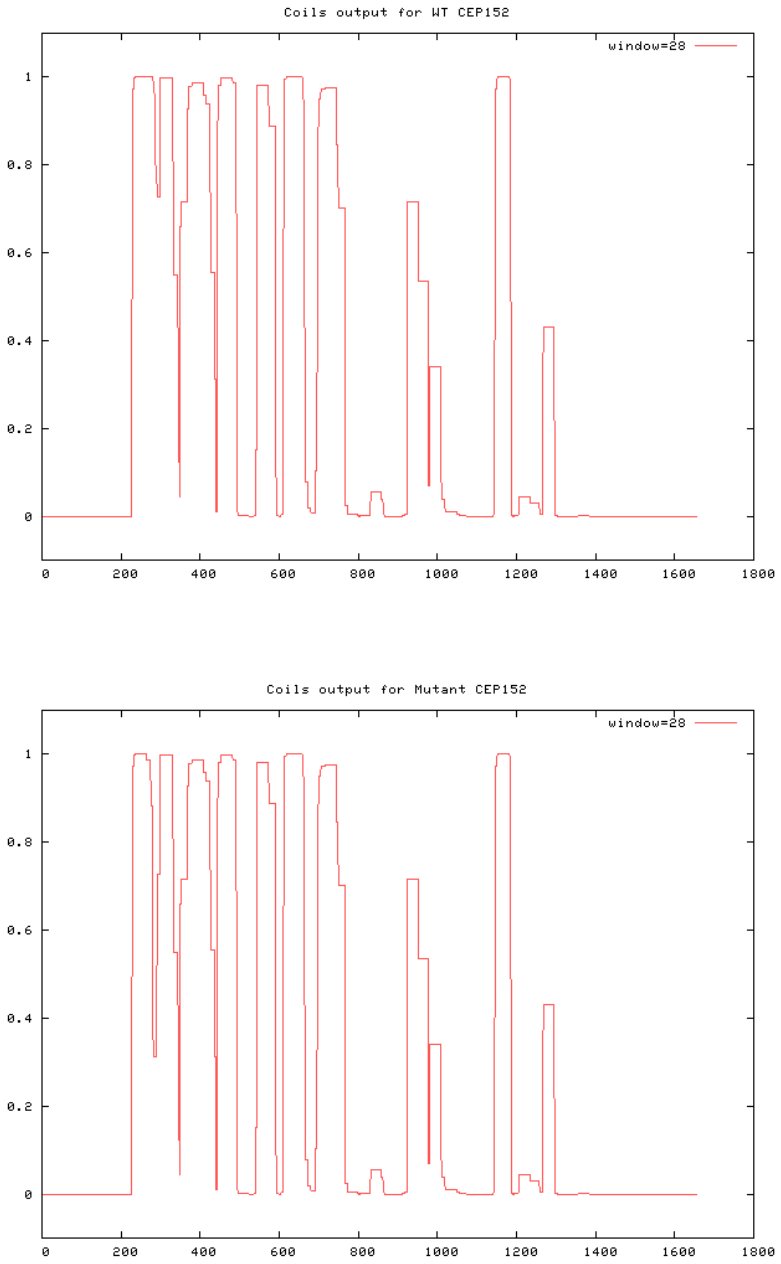


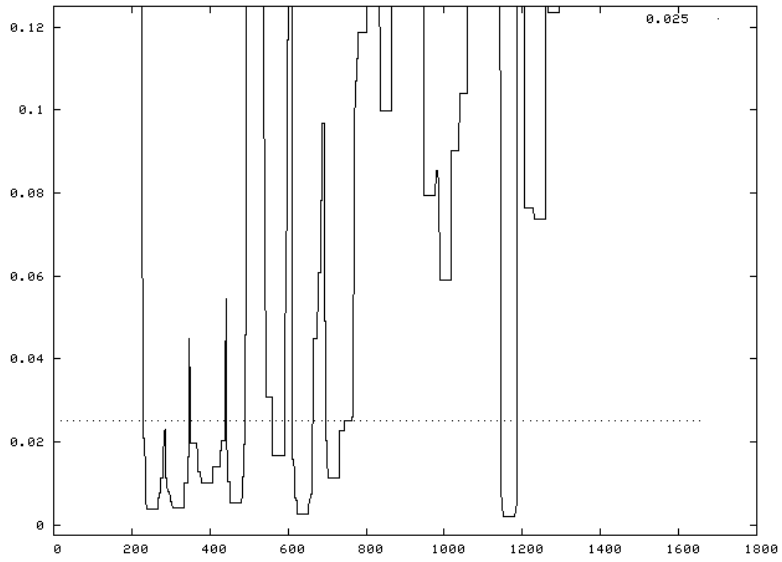
Table S3. Comparison of COILS Predictions for Wild Type and Mutant *CEP152*

Shaded rows indicate residues for which COILS predicts decreased support for a coil-coil in the mutant sequence.

	Wild type			Mutant		
Position	**AA**	**Register**	**P**	**AA**	**Register**	**P**
263	E	e	1.000	E	e	1.000
264	R	f	1.000	R	f	1.000
265	**Q**	g	1.000	**P**	g	0.987
266	I	a	1.000	I	a	0.987
267	R	b	1.000	R	b	0.987
268	Y	c	1.000	Y	c	0.987
269	L	d	1.000	L	d	0.987
270	N	e	1.000	N	e	0.987
271	H	f	1.000	H	f	0.987
272	Q	g	1.000	Q	g	0.987
273	L	a	1.000	L	a	0.987
274	V	b	1.000	V	b	0.941
275	I	c	1.000	I	c	0.936
276	I	d	1.000	I	d	0.936
277	K	e	1.000	K	e	0.882
278	D	f	1.000	D	f	0.882
279	E	g	1.000	E	g	0.882
280	K	a	0.997	K	a	0.488
281	D	b	0.995	D	b	0.387
282	G	c	0.988	G	c	0.312
283	L	d	0.988	L	d	0.312
284	T	e	0.988	T	e	0.312
285	L	f	0.940	L	f	0.312
286	S	g	0.808	S	g	0.312
287	L	a	0.808	L	a	0.312
288	R	b	0.788	R	b	0.312
289	E	c	0.788	E	c	0.312
290	S	d	0.788	S	d	0.312
291	Q	e	0.726	Q	e	0.726
292	K	f	0.726	K	f	0.726

Figure S4. Paircoil2 Predictions for Wild Type (Upper Panel) and Mutant (Lower Panel) *CEP152*

The y axis represents the p-score, with smaller p-scores signifying support for a coiled-coil conformation at each position (x axis). Paircoil2 predictions are highly consistent between the wild type and mutant sequences. Comparison of the results shows that positions 265-282 are predicted to be included in a coiled-coil in the wild type sequence, but are not included in the predictions for the mutant sequence using a default p-score cut-off of 0.025 (Table S4).



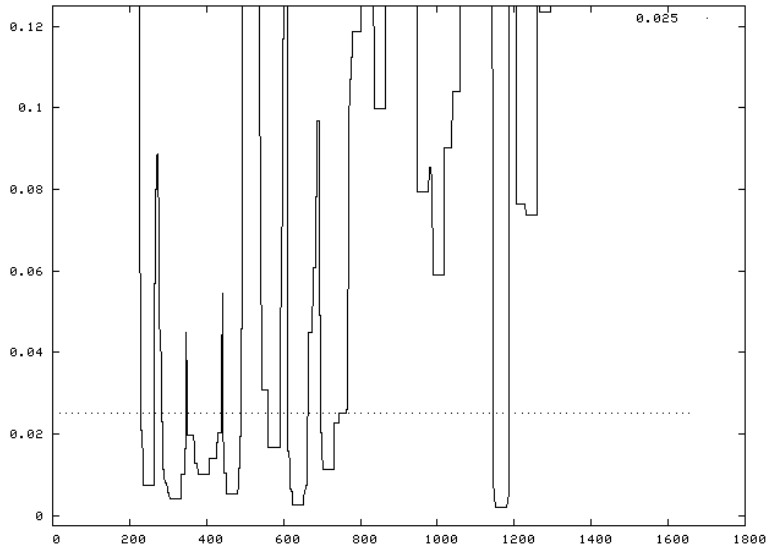


Table S4. Comparison of Paircoil2 Predictions for Wild Type and Mutant *CEP152*

Wild type CEP152		Mutant CEP152	
p-value	**Range**	**p-value**	**Range**
0.0039	229-346	0.0074	229-**264**
-	-	0.0042	**283**-346
0.0196	350-367	0.0196	350-367
0.0102	368-429	0.0102	368-429
0.0204	430-439	0.0204	430-439
0.0052	443-490	0.0052	443-490
0.0168	560-590	0.0168	560-590
0.0027	610-663	0.0027	610-663
0.0113	698-763	0.0113	698-763
0.0019	1146-1185	0.0019	1146-1185

Table S5. MARCOIL Predictions for Wild Type and Mutant *CEP152*, Using 9FAM Emission Matrix (Upper Panel), MTIDK Emission Matrix (Middle Panel), MTK Emission Matrix (Lower Panel)

Wild type and mutant CEP152
PREDICTED DOMAINS AT THRESHOLD 90% :
1. from 238 to 343 (length = 106) with max = 100.0%
2. from 371 to 423 (length = 53) with max = 100.0%
3. from 446 to 482 (length = 37) with max = 100.0%
4. from 547 to 582 (length = 36) with max = 99.3%
5. from 624 to 762 (length = 139) with max = 100.0%
6. from 839 to 856 (length = 18) with max = 91.4%
7. from 927 to 939 (length = 13) with max = 95.6%
8. from 1152 to 1179 (length = 28) with max = 100.0%
PREDICTED DOMAINS AT THRESHOLD 99% :
1. from 246 to 340 (length = 95) with max = 100.0%
2. from 376 to 413 (length = 38) with max = 100.0%
3. from 452 to 479 (length = 28) with max = 100.0%
4. from 566 to 573 (length = 8) with max = 99.3%
5. from 627 to 662 (length = 36) with max = 100.0%
6. from 701 to 759 (length = 59) with max = 100.0%
7. from 1155 to 1177 (length = 23) with max = 100.0%

Wild type and mutant CEP152
PREDICTED DOMAINS AT THRESHOLD 90% :
1. from 239 to 343 (length = 105) with max = 100.0%
2. from 371 to 423 (length = 53) with max = 100.0%
3. from 446 to 481 (length = 36) with max = 100.0%
4. from 564 to 579 (length = 16) with max = 93.2%
5. from 624 to 660 (length = 37) with max = 100.0%
6. from 703 to 762 (length = 60) with max = 100.0%
7. from 927 to 939 (length = 13) with max = 94.2%
8. from 972 to 974 (length = 3) with max = 90.5%
9. from 1151 to 1179 (length = 29) with max = 100.0%
PREDICTED DOMAINS AT THRESHOLD 99% :
1. from 247 to 340 (length = 94) with max = 100.0%
2. from 377 to 413 (length = 37) with max = 100.0%
3. from 453 to 479 (length = 27) with max = 100.0%
4. from 627 to 654 (length = 28) with max = 100.0%
5. from 713 to 758 (length = 46) with max = 100.0%
6. from 1155 to 1177 (length = 23) with max = 100.0%

Wild type	Mutant
PREDICTED DOMAINS AT THRESHOLD 90% :	PREDICTED DOMAINS AT THRESHOLD 90% :
1. from 246 to 343 (length = 98) with max = 100.0%	1. from 246 to 343 (length = 98) with max = 100.0%
2. from 370 to 423 (length = 54) with max = 100.0%	2. from 370 to 423 (length = 54) with max = 100.0%
3. from 447 to 482 (length = 36) with max = 100.0%	3. from 447 to 482 (length = 36) with max = 100.0%
4. from 624 to 656 (length = 33) with max = 99.9%	4. from 624 to 656 (length = 33) with max = 99.9%
5. from 709 to 762 (length = 54) with max = 100.0%	5. from 709 to 762 (length = 54) with max = 100.0%
6. from 1152 to 1179 (length = 28) with max = 100.0%	6. from 1152 to 1179 (length = 28) with max = 100.0%
PREDICTED DOMAINS AT THRESHOLD 99% :	PREDICTED DOMAINS AT THRESHOLD 99% :
1. from **250** to **340** (length = 91) with max = 100.0%	1. from 250 to **263** (length = 14) with max = 99.8%

- 2. from 376 to 413 (length = 38) with max = 100.0% 3. from 454 to 479 (length = 26) with max = 100.0% 4. from 627 to 649 (length = 23) with max = 99.9% 5. from 717 to 757 (length = 41) with max = 100.0% 6. from 1161 to 1177 (length = 17) with max = 100.0%	2. from **271** to 340 (length = 70) with max = 100.0% 3. from 376 to 413 (length = 38) with max = 100.0% 4. from 454 to 479 (length = 26) with max = 100.0% 5. from 627 to 649 (length = 23) with max = 99.9% 6. from 717 to 757 (length = 41) with max = 100.0% 7. from 1161 to 1177 (length = 17) with max = 100.0%

Figure S5. SOSUIcoil Predictions for Wild Type (Upper Panel) and Mutant (Lower Panel) *CEP152*

Wild type

Predicted coil-coil regions:

7-44, 226-499, 520-1039, 1141-1186, 1218-1260, 1283-1310, 1343-1372

Predicted coiled-coil fragile points:

```
226-499              LGANENSAENMQIIQLQVLNKAKERQLENLIEKLNESERQIRYLNHQLVIIKDEKDGLTLSL
register             efgabcdefgabcdefabcdefgabcdefgabcdefgabcabcdefgabcdefgabcdefga
point types          ---------------D----------------------B----------------------
H-philic  core(I)    ---II-----------IIIIII-----------------------------------------
H-phobic  outF(O)    -----------OOOOOOOOO-----------------------------------------
```

Legend

Fragile point by Heptad breaks:	A:Stutter B:Stammer C:Skip(+) D:Skip(-)
Fragile point by Hydrophilic core:	I
Fragile point by Hydrophobic outfield:	O

Mutant

Predicted coil-coil regions:

7-44, 226-499, 520-1039, 1141-1186, 1218-1260, 1283-1310, 1343-1372

Predicted coiled-coil fragile points:

```
226-499              LGANENSAENMQIIQLQVLNKAKERQLENLIEKLNESERPIRYLNHQLVIIKDEKDGLTLSL
register             efgabcdefgabcdefabcdefgabcdefgabcdefgabcabcdefgabcdefgabcdefga
point types          ---------------D----------------------B----------------------
H-philic  core(I)    ---II-----------IIIIII-----------------------------------------
H-phobic  outF(O)    -----------OOOOOOOOO-----------------------------------------
```

Figure S6. Conservation of *CEP152* across Coding Region of Gene

The plot was obtained for protein sequences using 100-amino-acid-long overlapping windows, sliding using a step of 10 amino acids. For DNA sequences, the window was 300 pb and the step was 30 pb. Nucleotide (in red) and protein (in black); vertebrates (solid lines) and primates (dashed lines).

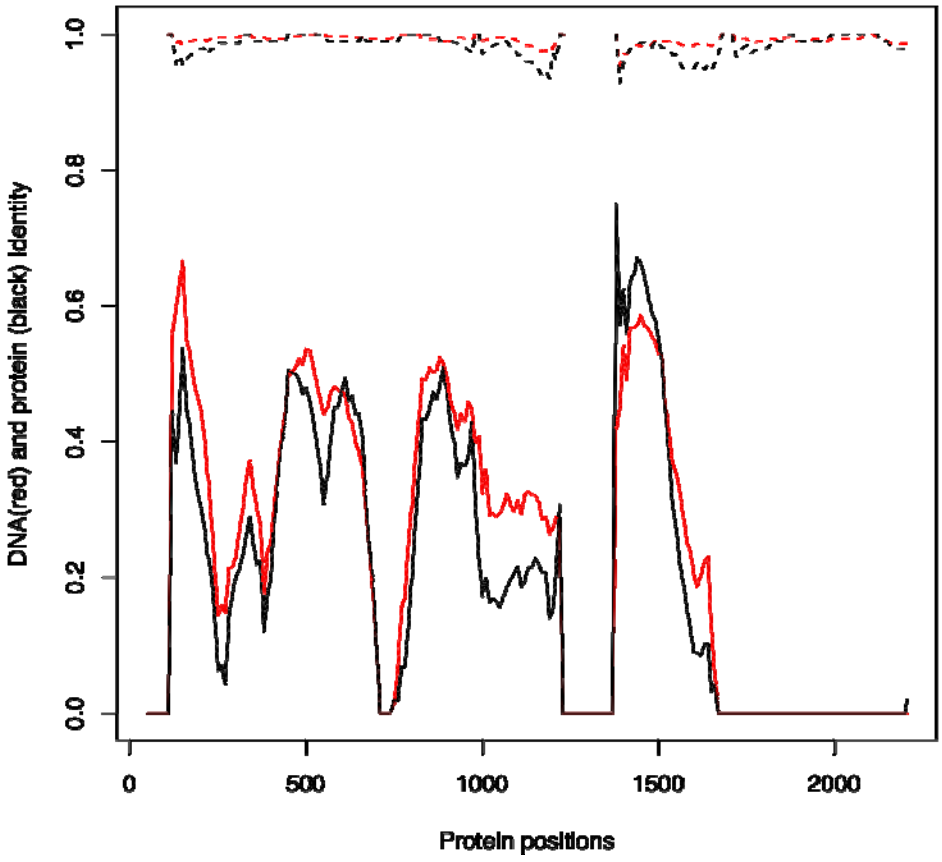


Figure S7. Sequences of CEP152 Orthologs Used for Evolutionary Selection Analysis

>HUMAN

MSLDFGSVALPVQNEDEEYDEEDYEREKELQQLLTDLPHDMLDDDLSSPELQYSDCSEDGTDGQPHHPEQL
EMSWNEQMLPKSQSVNGYNEIQSLYAGEKCGNVWEENRSKTEDRHPVYHPEEGGDEGGS----------
----------GYSPPSKCEQTDLYHLPENFRPYTNGQKQEFNNQATNVIKFSDPQWNHF-Q------
GPSCQGLEPYNKVTYKPYQSSAQNN-GSPAQEIT-GSDTFEGLQQQFLG-
ANENSAENMQIIQLQVLNKAKERQLENLIEKLNESERQIRYLNHQLVIIKDEKDGLTLSLRESQKLFQNGK
EREIQLEAQIKALETQIQALKVNEEQMIKKSRTTEMALESLKQQLVDLHHSESLQRAREQHESIVMGLTKK
YEEQVLSLQKNLDATVTALKEQEDICSRLKDHVKQLERNQEAIKLEKTEIINKLTRSLEESQKQCAHLLQS
GSVQEVAQLQFQLQQAQKAHAMSANMNKALQEELTELKDEISLYESAAKLGIHPSDSEGELNIELTESYVD
LGIKKVNWKKSKVTSIVQEEDPNEELSKDEFILKLKAEVQRLLGSNSMKRHLVSQLQNDLKDCHKKIEDLH
QVKKDEKSIEVETKTDTSEKPKNQLWPESSTSD-VVRDDILLLKNEIQVLQQQN-------
QELKETEGKLRNTNQDLCNQMRQMVQDFDHDKQEAVDRCERTYQQHHEAMKTQIRESLLAKHALEKQQLFE
AYERTHLQLRSELDKLNKEVTAVQECYLEVCREKDNLELTLRKTTEKEQQTQE----------
----
KIKEKLIQQLEKEWQSKLDQTIKAMKKKTLDCGSQTDQVTTSDVISKKE-
MAIMIEEQKCTIQQNLEQEKDIAIKGAMKKLEIELELKHCENITKQVEIAVQNAHQRWLGELPELAEYQAL
VKAEQKKWEEQHEVSVNKRISFAVSEAKEKWKSELENMRKNILPGKELEEKIHSLQKELELKNEEVPVVIR
AELAKARSEWNKEKQEEIHRIQEQNEQDYRQFLDDHRNKINEVLAAAKEDFMKQKT----------
----

ELLLQKETELQTCLDQSRREWTMQEAKRIQLEIYQYEEDILTVLGVLLSDTQKEHISDSEDKQLLEIMSTC
SSKWMSVQYFEKLKGCIQKAFQDTLPLLVENADPEWKKRNMAELSKDSASQGTG-
QGDPGPAAGHHAQPLALQATEAEA----------

EENNKVVEELIEENNDMKNKLEELQTLCKTPPRSLSAGAIENACLPCSGGALEELRGQYIKAVKKIKCDML
RYIQESKERAAEMVKAEVLRERQETARKMRKYYLICLQQILQDDGK-E-
GAEKKIMNAASKLATMAKLLETPISSKSQSKTTQS----
ALPLTSEMLIAVKKSKRNDVNQKIPCCIESKSNSVNTITRTLCEQAPKRRAACNLQRLLENSEHQSIKHVG
SKETHLEFQFGDGSCKHLNSLPRNVSPEFVPCEGEGGFGLHKKKDLLSDNGSESLPHSAAYPFL----
GTLGNKPSPRCTPGPSESGCMHITFRDSNERLGLKVYKCNPLMES-ENAASEKSQGLDVQEPPVKDGG-
DLSDCLGWPSSSA-TLSFDSREASFVHGRPQGTLEIPSESVKSKQFSPSGYLSDT-
EESNMICQTMKCQRYQTPYLSEETTYLEPGKISVNCGHPSRHKADRLKSDFKKLSSTLPSSVCQQPSRKLI
V-PLSSQQDSGFDSPFVNLD

>RHESUS

MSLDFGSVALPAQNEDEEYDKEDYEREKELQQLLTDLPHDMLDDDLSSPEPQYSDCSEDGTDRQPHHPEQL
EMSWNEQMLPKSQSVNDYNEIQSLYAGEKCGNVWEENRSKTEDRHPVYHPEEGGDEGGS----------
----------GYSPPSKCEQTDLYHLPENFRPYTNGQKQEFNNQPTNVIKFSDPQWNHF-Q------
GPSCQGLEPYNKVTYKPYQSSAQNN-GSPAQAIT-GSDTFEGLQQQFLG-
ANENSAENMQIIQLQVLNKAKERQLENLVEKLNESECQIRYLNHQLVIIKDEKDGLTLSLRESQKLFQNGK
EREIQLEAQIKALETQIQALKVNEEQMIKKSRTTEMALESLKQQLVDLHHSESLQRAREQHESIVMGLTKK
YEEQVLSLQKNLDATVTALKEQEDICSRLKDHVKQLERNQEAIKLEKTEIINRLTRSLEESQKQCAHLLQS
GSVQEVAQLQFQLQQAQKAHTMSENMNKALQEELTELKDEISLYESAAKLGIHPSDSEGELNIELTESYVD

```
LGIKKVNWKKSKVNSIVQQEDPNEELSKDEFILKLKAEVQRLLGSNSMKRRLVSQLQNDLKDCHKKIEDLH
QVKRDEKSIEVETKTDTSEKPKNQLWLESSTSD-IVRDDILLLKNEIQVLQQQN-------
QELKETEEKLRNTNQDLCNQMRQMVQDFDHDKREAVDRCERTYQQHHEAMKTQIRESLLAKHALEKQQLFE
AYERTHLQLRSELDKVNKEMTAVQECYLEVCREKDHLESTLRKTTEKEQQTQE------------------
---------------------------------------------
KIKEKLIQQLEKEWQSKLDQTIKAMKTKTSDRGSQTDQVTTSDVISKKE-
MAIMIEEQKCIIQQHLEQEKDIAIKGAMKKLEIELELKHCENIAKQVEIAVQNARQRWLGELPELAEYQAL
VRAEQKKWEEQHEVSVNKRISFAVSEAKEKWKSELENMKKNILPGTELEEKIHSLQRELELKNEEVPVVIR
AELAKARSEWNKEKQEEIRRIQEQNEQDYRQFLDDHRNKINEVLAAAKEDFMKQKT---------------
---------------------------------------------------------------------
---------------------------------------------------------------------
---------------------------------------------------------------------
------------
ELLLQKETELQTCLDQSRREWTMQEAKRIQLEIYQYEEDILTVLGVLLRDTQKEHVSDSEDKQLLEIMSTC
SSKWMSVQYFEKLKGCIQKAFQDTLPLLVENADPQWEKRNMAELSKDSASRSTG-
QGVPGPAAGHHAQPLALQETEAEA-----------------------------------------------
----------------------------------
EENNKVIEELIEENNDMKNKLEELRTLCKTPPRSLSAGAIENTCLPCSGGALEELRGQYIKAVKKIKRDML
RYIQESKERAAEMVKAEVLRERQETARKMRKYYLICLQQILQDDGK-E-
GAEKKIMNAASKLATMAKLLETPISSKSESKTTQS------
PLTSEMLIGVEKSKRNDVNQKISCCIESKSNSVNTITRSVCEQAPKRRAACNLQRLLENSDHRSIKHVGSK
ETHLEFQFGDGSCKHLHSLPRNVSPEFVPCEGEGGFGLHKKKDLLSDNGSESLPHSAAYPFL----
GTLGNKPSPRCTSSPSESGCMHITFRDSNERLGLKVYKCNPLMES-ENAASKKSQGLDVQEPPVKDGG-
DLSDCLGWPSSSA-TLSFESREASFVHGRPQGTLEIPNESVNSKQFSPSGYLSDT-
ETSNMICQTMKCQHDQTPYLSEETMYLEPGKISVTHGYPSHHKADRLKPDFKKLSSTLPSSVCQQPSRKLI
V-PLSSQQDSGFDSPFVNLD

>CHIMP

MSLDFGSVALPVQNEDEEYDKEDYEREKELQQLLTDLPHEMLDDDLSSPELQYSDCSEDGTDGQPHHPEQL
EMSWNEQMLPKSQSVNGYNEIQSLYAGEKCGNVWEKNRSKTEDRHPVYHPE-GGDEGGS-----------
----------GYSPPSKCEQTDLYHLPENFRPYTNGQKQEFNNQATNVIKFSDPQWNHF-Q------
GPSCQGLEPYNKVTYKPYQSSAQNNDGSPAQEIT-GSDTFEGLQQQFLG-
ANENSAENMQIIQLQVLNKAKERQLENLIEKLNESERQIRYLNHQLVIIKDEKDGLTLSLRESQKLFQNGK
EREIQLEAQIKALETQIQALKVNEEQMIKKSRTTEMALESLKQQLVDLHHSESLQRAREQHESIVMGLTKK
YEEQVLSLQKNLDATVTALKEQEDICSRLKDHVKQLERNQEAIKLEKTEIINRLTRSLEESQKQCAHLLQS
GSVQEVAQLQFQLQQAQKAHAMSANMNKALQEELTELKDEISLYESAAKLGIHPSDSEGELNIELTESYVD
LGIKKVNWKKSKVTSSVQEEDPNEELSKDEFILKLKAEVQRLLGSNSMKRHLVSQLQNDLKDCHKKIEDLH
QVKKDEKSIEVETKTDTSEKPKNQLWPESSTSD-VVRDDILLLKNEIQVLQQQN-------
QELKETEGKLRNTNQDLCNQMRQMVQDFDHDKQEAVDRCERTYQQHHEAMKTQIRESLLAKHALEKQQLFE
AYERTHLQLRSELDKLNKEMTAVQECYLEVCREKDNLELTLRKTTEKEQQTQE------------------
---------------------------------------------
KIKEKLIQQLEKEWQSKLDQTIKAMKKKTSDCGSQTDQVTTSDVISKKE-
MAIMIEEQKCTIQQNLEQEKDIAIKGAMKKLEIELELKHCENITKQVEIAVQNARQRWLGELPELAEYQAL
VRAEQKKWKEQHEVSVNKRVSFAVSEAKEKWKSELENMRKNILPGKELEEKIHSLQRELELKNEEVPVVIR
AELAKARSEWNKEKQEEIHRIQEQNEQDYRQFLDDHRNKINEVLAAAKEDFMKQKT---------------
---------------------------------------------------------------------
---------------------------------------------------------------------
---------------------------------------------------------------------
------------
```

```
ELLLQKETELQTCLDQSLREWTMQEAKRIQLEIYQYEEDILTVLGVLLSDIQKEHISDSEDKQLLEIMSTC
SSKWMSLQYFEKLKGCIQKAFQDTLPLLIENTDPEWKKRNMAELSKDSASRGTG-
QGDPGPAAGHHAQPLALQATEAEADK------------------------
KKVLEIKDLCCGHCFQELEKAKQECQDLKGKLEKCCRHLQHLERKHKAVVEKIGEENNKVVEELIEENNDM
KNKLEELRTLCKTPPRSLSAGAIENACLPCSGGALEELRGQYIKAVKKIKHDMLRYIQESKERAAEMVKAE
VLRERQETARKMRKYYLICLQQILQDDGK-E-GAEKKIMNAASKLATMAKLLETPISSKSQSKTTQS----
ALPLTSEMLIAVKKSKRNDVNQKIPCCIESKSNSVNTITRSLCEQAPKRRAACNLQRLLENSEHQSIKHVG
SKETHLEFQFGDGSCKHLNSLPRNVSPEFVPCEGEGGFGLHKKKDLLSDNGSESLPHSAAYPFL----
GTLGNKPSPRCTPGPSESGCMHITFRDSNERLGLKVYKCNPLMES-ENAASEKSQGLDVQEPPVKDGG-
DLSDCLGWPSSSA-TLSFDSREASFVHGRPQGTLEIPSESVKSKQFSPSGYLSDT-
EESNMICQTMKCQRDQTPYLSEETTYLEPGKISVNCGHPSHHKADRLKSDFKKLSSTLPSSVCQQPSRKLI
V-PLSSQQDSGFDSPFVNLD

>MARMOSET

MSLDFGSVALPAQNEDEEYDEEDYEREKELQQLLTDLPHDMLDDDLSSPELHDSDYSEDGTDREPHHPEQL
EMNWNEQVLPKSQRINDYNEIQNLYAGEKCGNVWEENRNKTEDRHPGYHPEEGGDEGGS-----------
----------GYSPPSKCEQTDLYHLPENFRPYTNGQKQEFNNQPTNIIKFSDPQWNHF-Q------
GPSCQGLEPYNKVIYKPYQSSAPNN-GSPAPEIT-GSDTFEGLQQQFLG-
ANEKSAENMQIIQLQVLNKAKERQLENLVEKLNESERQIRYLNHQLLIIKDEKDGLTLSLRESQKLFQDGK
EREIQLEAQIKALETQIQALKVNEEQMIKKSRTTEMALESVKQQLVDLHHSESLQRAREQHESIVMGLTKK
YEEQVLSLQKNLDATVAALKEQEDICSHLKDHVKQLERNQEAIKLEKTEIINRLTRSLEESQKQCAHLLQS
GSVQEAAQLQFQLQQAQKAHAMSENMNKALQEELTELKDEISLYESAAKLGIHPSDSEGVLNIELTESYVN
LGIKKANWKKSKVKSIVQHEYPNEELSKDEFILKLKAEVQRLLGSNSMKRRLVTQLQNDLKDCHKKIEDLQ
QVKEDEKSIEVETKTDTSEKQTNQLWPESSTPN-TVRDDILLLKNEIQVLQQQN-------
QELKETEEKLRTTNQDLCNQMRQMVQDFDHDKQEAVDRCERTYQQHHEAMKTQIRESLLAKHALEKQQLFE
VYERTHLQLRSELDKVNKEMSAVQECYLEVCREKDNLESILRKTAEKEQQAQE-----------------
-----------------------------------------------
KIKEKLTQQLEKEWQSKLDQTIKAVKKKTSDCGSQTDQVTTSDVISKKE-
MAIMIEEQKHTIQQNLEQEKDIAIKEAVKKLEMELELKYCENITKQVEIAVQNARQRWLGELPELAEYQAL
VRAEHKKWEEQHEISVNKRILFAVSEAKEKWKSELENRKKNILPGKELEEKFHSLQRELKLKNEEVPVVIR
AELAKARSEWNKEKQEEIHRIQERNEQDYRQFLDDHRNKINEVLAAAKEDFMKQKT---------------
----------------------------------------------------------------------
----------------------------------------------------------------------
----------------------------------------------------------------------
------------ELLLQKETELQTCLDQSRREWTMQEAKRIQLEIRQYEEDILTVLEFLLR-
TQKEHVNDSENTQLLEILSTCSSKWMSVQYFEKLKACIQKAFQDTLLQLTENIDPEWEKRNMSELAKDSAR
RGTG-QGDPGPTAGHHAQPLALQQTEAEV----------------------------------------
---------------------------------------
EENNKVVEELIEENNDMKNKLEELRTLCKTPPRSLSAGAIENGCLPCSGGALEELRGQYIKAVKKIKRDML
RYIQESKERAAEMVKAEVLRERQETARKMRKYYLICLQQILQDDGK-E-
GAEKKIMNAASKLATMAKLLETPISSKSHSKTTQS----
ALPLTSERLIGVEKSKRNDVNQKIPCYDESKSNSLNTITRGVCEQVPKRRAACNLQRLLENSEHGSRKHVG
SKETHLEFQFGDSSCKHLNSLPRKASHEFVPCEGEGGFGLHKKKHLLTDNGSESFLHSDAYF-L----
GTLGSKPSLRCTPSPSESGCMNIIFHDPNERLGLKVHKCSLPMES-ENVASEKSQGLGVQEPPVRDGG-
DLSNSLGWPSSSA-ALPVDSHQASFVCSRPQGTLEIPTESVNSKQFSPSSCLSDT-
ENSNMICQTMKCKRDQTPYLSEETTYLQPGEISVTHGHPSHHKANILKSDFEKLSSTLPSSVCQQPSRKLI
V-PLSSQQDSGFDSPFVNLD

>OPOSSUM
```

```
MSQDFGSV------------------------------------------------------------
------------------------------------------------------------------------
------------------
FDNVALQSQNDEEEYAQEDYAREKELHQLLTDLPHDMLDDDLSSPDLNLSDCSEDATVGQTHKIQDPEDSW
NEHHVLP-KPQSGYDDVQNLYPGEKHVNGWEEHHRKNEDLYTGCNIEEHHDGGGS---------------
------GYSSTNEYEQNDLYQLPENFRPYTNGQKKEFNNQPTKIINFPDGQREHF-
QQFGTSHVPNSQALESY-KVTYKPYQTSSQNN-VSSNQERTRRNDEFEEMQQKFLE-
PAENSLENMQIAQLQVLNKAKERQLTDLCEKLEESERQIRYLNHQLVMLKDEKDGLALSLQESQKLLQKGK
EKEIQLEGQIKALETQIQTLKANEEQMIKKARVSEVAMENLRQQLLELRHSDSLQRAKEQHENIVLCLRKK
YEDQVTTLQKNLEVTNAVLQEQKDICCRLRDHVKQVERSQEENKMEKTEIINRLTKSLEESQKQCANLLET
GSVQEMNQLKLQLQQAQSAHTLSDSMNKALQEELAELKEEITLYESAAKLGVFSNSPDEELNLDLAESYVD
LGIKKTNWKKSNILSTMQKSEPNEELPKDEIIMRLKSELQRLLTSNTVKRRLVSQLQNDLKDCHRRLEDLS
QQKKSQEKCELETKSINVEKSPNRLWHDSSVYDKTVKEDILRLKNEIQSLQQEK-------
EDLKATEEKLKSTNQDLCSQMRQMVQDFDRDKQETVDRCERTFQQHHEAVKNQIRESLLAKHTLEKQQLFE
VYESNRLQLRADLDKVNKEMATVQECYLAVCKEKDNLETTIRKAVEKEQQTKEEMLKNKLLQEKEEDLNKL
KVELEEKLKSTITATKSQWLKEKESDVKQQIENEVILAKTQWEKEQKEIKDKVIQQLEREWRFKLDQTIKE
MKKKTFDCGIQTDQIINPDVISKKE-
MEAVIEEQQLKIQQDVLQEKEIAIKEALRKQEKEMETHYRENIIKQVEIAVLNARSRWVQELPELAEYRAN
IKAEQKKWETEHELSVAKKVSLALLEAKEKWKIELGNTEESV---
KELEAKIFSLQRELELKEEEIPVIVRAKLAKSRSEWNKEKQEEIHSIQEQNEKDYRKFLDDHRNKINEVLA
AAKEDFMKQKT---------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-------------------------------------
ELILQKEAELQACLEQKQREWTIQEGKQLQMEIHQYEEDVLTEIELLVEEIYEELSGI-
ENKHLSKTKSPCFPRRLNAQYLEKLRACIQMAFRGTLYRLLENSNPEWKMCRKILFESQEDS---------
------------------------------------------------------------
CCEQCCQKLEKAKQECQDLKKMLEKSCRHLQHAVREHKATVEKIGEENSRVVEELMEENNEMKMKLEELRA
LSSTPPRSFSEGAVDHSCLICSGKGLEELRGHYIRAVKKIKSDMIRYIQASKERAAELLKAEIFKERQETA
QRMQKYYLVCLHQLLQDDGKEE-GAEKKIMNAASKLAAMAKMLETPVVNRPQIKNSPSVSR-----
SSELLPGVEQSKRNYMNQSRASHMENKSSSRNKAGG---
EQALQKRLPCNLRQRLEDTEHTETKHVAFREAPSGIPISVTACKQLDDLPRNVLLELAPCDGGRSKTGPHV
--------------------------
QFANTHTSTRAPGLSDSGSSHVTFQGLKEKRILKERKCNSLTECCQQIVAEKSQDFDIQETPVKDEESGPS
DSLDWSSKNVTSQALSSSDTSFLHPLQQGILDKQVESVSFMQFPANTCLSETEEKANKLCRQFQCQNDHIP
AVKKEDKRTQTEQVNKSPGHASHHNSDVLKSDFKKLNSAVQSSSYHQHSRKLLMAPQASQQDSGFDSPFAN
LD

>RATTUS
MTTNAQDFPNGGSGKLPAATGALFEQPRSRNCGRLLNRSCGRRGFRLGVPPSGQAHRAAAAIGGNRCSSSG
GVRGLGCTGKVLAGWKSRQRVAGGDCSLTAGRAAAGSGPAFESPAPGRCVFAPCSVFSRGVSPVDLSAPVI
PGRRTQLFGKGAARTMSLEFGSVALQTQNEDEEFDKEDYEREKELQQLLTDLPHDMLDDELSSPERRDSDC
SVDGIAGELYPSEHLERKWIERDILP-KPH----------GVNCGNGWEENRSEPEDQRLGYHPG-
EGDEGGS---------------------GYSPPGKQEQADLYRLPEDFRPYTGRQKQ-----
AASVITFSDPQRDSF-Q-----------------------------------------------------
-----------------
VLNKAKERQLDSLVEKLNDRERQIRYLNHQLLIVQDEKDGLALSLRESQKLFQSGKEREMQLEAQIHALET
QIEAFKVSEEKLTKKLRTTEITLESLKQQLVELHHSESLQRAREQHESVVAGLTQKY-------------
----
EEQEDVCTRLKDHVQQLERNQEAVRLEKTELINRLTRSLEDSQKQCAHLLQSGSVHEVTQLQLQLQQAQKA
```

```
HILSESMNKALQEELTGLKEEISLYESAAELGVLPGDVEGDLSVELTESCVGLGIKNVSWKQSKANSVAQQ
EPPNEKLSKDEVILKLKTQVQRLLSSNSMKRHLVSELQRDLRDCRETMEALQQSKDSDRG--
METKTETSEKTTNQLWLDSPEAIN--REDILRLKNEVEVLQQQN-------
QELKEAEEKLRSTNQDLCNQMRQMVQEFDHDKQEAVDRCERTYQQHHEAMKAQIRESLLAKHAAEKQHLSE
VYEGTQSQLRSDLDKLNKEVAAVQECYLEVCREKDELESTLRKTMEKAQEQTRQLLE-----
DREEHVRKLKLELEERYQETLKAEKQSWLKAQAAGATQV--------------
EHESRQKLIQQLEEEWQSKLNHSLPARRKATSDCGSQTDQAAYPAAMPNAE-ALVLAEEQA----
HQVQQEKELTVKEALRKSEVELELKYCESIAQKVETAVQNARSRWILELPMLAEYKALLRAQQQEWAKQQE
VAVAHRLSLALSEAKEKWKSELQNMKPNVASVQELEEKIHSLQKELELKDEEVPVAVRAELAKARNEWNKE
KQEEIHKIQEQNEEDYRQFLEDHRNKINEVLTAAKEDFVKQKT--------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
ELLLQKETEFQACLDQSRKEWTLQEAQRTQLEVHRYEEDILTVLDFLLRDTQLEYDGDSQGKQLLEVMSVC
SSKWTSVQYFEKVKACIHKALQDTLSLLIDNIASEREKSHHQHLTSGIHWLCQCDASTTLRYEHKQHHWKD
PEQSGGLRGMSLDLTPKMTHTPSPRSPASVPPPAHHKKIFEMKDLCCGHCFQELEKEKQECQDLRRKLEKS
RRHLQHLERTHKATVEKLGG--------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------

>ORANGUTAN

MSLDFGSVALPVQNEDEEYDKEDYEREKEFQLLT--LPQDMLEDLYSEVQY---
SCSAVSTDGLPHHPEQLEMSWNEQGTSKSQSVKGYKEIKNLYVGEKCRKVWEENTSKTEDRHPVYHPEEGG
DEGGS----------------------
GYSPPSKCEQTDLYHLPENFRPYTNGQKQEFNNQATNVIKFSDPQWNHF-Q------
GPSCQGLEPYNKVTYKPYQCSAQNN-GSPAQEIA-GSDTFEGLQQQFLG-
ANENSVENMKIIQLQVLNKAKERQLENLIEKLNESERQIRYLNHQLVIIKDEKDGLTLSLRESQKLFQNGK
EREIQLEAQIKALETQIQALKVNEEQMIKKSRTTEMALESLKQQLVDLHHSESLQRAREQHESIVMGLTKK
YEEQVLSLQKNLDATVTALKEQEDICSRLKDHVKQLERNQEAIKLEKTEIINRLTRSLEESQKQCAHLLQS
GSVQEVAQLQFQLQQAQKAHAMSENMNKALQEELTELKDEISLYESAAKLGIHPSDSEGELNIELTESYVD
LGIKKVNWKKSKVPSIVQQEDPNEELSKDEFILKLKAEVQRLLGSNSMKRRLVSQLQNDLKDCHKKIEDLH
QVKKDEKSIEVETKTDTSEKPKNQLWPESSTSD-IVRDDILLLKNEIQVLQQQN-------
QELKETEEKLRNTNQDLCNQMRQMVQDFDHDRQEAVDRCERTYQQHHEAMKTQIRESLLAKHALEKQQLFE
AYERTHLQLRSELDKLNKEMTAVQECYLEVCREKDNLESTLRKTIEKEQQTQE-----------------
-----------------------------------------------
KIKEKLTQQLEKEWQSKLDQTIKAMKKKTSDCGSQTDQVTTSDVISKKE-
MAIMIEEQKCTIQQNLEQEKDIAIKGAMKKLEIELELKHCENITKQVEIAVQNARQRWLGELPELAEYQAL
VRAEQKKWEEQHEVSVNKRISFAVSEAKEKWKSELENMKKNTLPGKELEEKIHSLQRELELKNEEVPVVIR
AELAKARSEWNKEKQEEIHRIQEQNEQDYRQFLDDHRNKINEVLAAAKEDFMKQKT--------------
------------------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
------------
ELLLQKETELQTCLDQSRREWTMQEARRIQLEIYQYEEDILTVLGVLLSDTQKEHISDSEDKQLLEIMSTC
```

```
SSKWMSVQYFEKLKGCIQKAFQDTLPLLIENTDPEWKKRNMAELSKDSACRGTG-
RGDPGPAAGHHAQPLALQETEAEA-----------------------------------------------
---------------------------------
EENNKVVEELIEENNDMKNKLEELRTLCKTPPRSLSAGAIENACLPCSGGALEELRGQYIKAVKKIKRDML
RYIQESKERAAEMVKAEVLRERQETARKMRKYYLICLQQILQDDGK-E-
GAEKKIMNAASKLATMAKLLETPISSKSQSKTTQS----
ALPLTSEMLITVEKSKRNDVNQKIPCCIESKSNSVNTITRSLCEQAPKRRAASNLKRLLENSEHQSIKHVG
SKETHLEFQFGDGSCKHLNSLPRNVSPEFVPCEGEGGFGLHKKKDLLSDNGSESLLHSAAYPFL----
GTLGNKPSPRCTPGPSESGCMHITFRDSNERLGLKVYKCNPLMES-ENAASEKSQGLDVQEPPVKDGG-
DLSDCLGWPSSSA-TLSFDSREASFVHGRPQGTLEIPSESVNSKQFSPSGYLSDT-
EKSNMICQTMKCQRDQTPYLSEETMYLEPGKISVNCGHPSHRKADRLKSDFKKLSSTLPSSVCQQPSRKLI
V-PLSSQQDSGFDSPFVNLD

>CHICK

MSLDFDSGALQTQHEDEDYDQEDYAREQELQQLLTDLPHDMLEDSGDQLSN-
YSDSSIHETEEQSQEARKPDGRWNDHLLIN-
DAQNGYEQGQNLYPDQFLCDQQNDHVEKHAKNWNGLHSSEQMKHLYDVKEDYSGQNSQEDPDDIYLDRDGF
NAPGCYQQSNLYHLPENFRPYTNGHNPEFNSQQNQIINFPDAPKEHLKQPFVTSEVVSGQSAESY-
KVMYKPYQNDVHKKIPVTQEGMR-RNEVFEDLHHEFSG-
NDENSSENMQILQLKALNKARERRVEELNEKLEKSAQQIRYLSHQLSMVTDEKDGLAVSLHESQKLYQNGK
EREVYLEGQIKALETQIQTLTTNEEQILKQSRVAEVAMESMQKQLLELQRSDTLQRAREQHDAIVTALKQK
HEKQVVSLEQKLDATKSALQEQKELCQNLGERVKELEKMLEETKCEKTEIINRLTRSLEESQKQCANLLQT
GSIQEANQLRFQLQQAQSAHTISNDMNLALQPVTNQNDALVA----------------------------
----------------------------------------------------------------------
---------------------------
SPLVSDNLKEEVLRLRKANEALLREVQNHALTIEELKENEKKLRSSNQDLCCQMRKMVQEFDEDKQEAIDR
CERTYQQHHEDTKAQFEKDMMQRYAAEKQQLIQSYEETVLQLKATIEELNREMTTVKECYIAVCGEKDTLE
TTLRQTFAHEQQMKED-----------------------------------------------------
----------KKEDVIAKIEREWQNRLEEIKKTMFECK-DCSSQTDRVTRVSDASTEE-
LARIVEDQKLQIQKALKEKK--
ASEEALKELEIELESKYRKRLASQVEAALTQAHARSLQELTDLKEYEVNLGTEQEKQEREQTETAAKLLAL
AFSSDELKWKKERECTEQSGSIIRELEEKVVSLKKELELKEEEIPVTIKAAVAKARAQWNKEKQEEIAQIQ
EQNERDYRSFLDDHRNKIKEVLATTKEDLAKQVNDLSAQKEAEMKMILDQKQREWEAQETKRLQDEINRYE
EKTLIELECLLREVHEELVKSTYSQHSWEEKDPDAPVELNHQRKDKLKACLQKAYRRTVHKILEKHEQEWK
EVEAALTQAHARSLQELTDLKEYELALAFSSDELKWKKERECTEQSGSIIRELEEKVVSLKKELELKEEEI
PVTIKAAVAKARAQWNKEKQEEIAQIQEQNERDYRSFLDDHRNKIKEVLATTKEDLAKQVNDLSAQKEAEM
KMILDQKQREWEAQETKRLQDEINRYEEKTLIELECLLREVHEELVKSTYSQHSWEERDPDAPVELNHQRK
DKLKACLQKAYRRTVHKILEKHEQEWKEKNEELLNNANKEVCCCMQGGDGETGDMARSPLYNTGQQAETQR
-------------------------
RLRRRHRSGTDGVCMQASTEPGEQGADQMVSWSRLQSKGSVSKACSSRGSVK------------------
------------------------------
GATLEEMRAQYIKSVSKIKSDMLCYIRESKERAAAMIKAEVLRERQETARKMRKYYLTCLQQLLTDNGK-
QEGAEKKIMNAASKLVTMAKGLETPLRHTLQCRTTRS----
ALLLNSDLPTGVEYAKRDCLSQTRPNHMENKSCGKNTTEEANDKVV--
RKYVPRDLRQQFGARQTETQHMLHETDTSDIQNSNI-
CKHLVSAGIDGLSKILPNVDGRKEKYSPLVNLDKELLCATSNFAHPNVSPFACQGQLENIDAARLKNGPAI
SGSTHHMLKSKNERTTLKVDKCILSHGT-WQTTYERIQDFDFHRTPER----DEGSCSEWSSASG-
SLHLDGSDKSAVYPEEKPSTNEQRQSVYCEESFSAGDENFGTSWRELTNGNVKNFNVKNS----------
---------PIVYNRDKLRSHSDHISSLGPDKTNSANTRTNTL-PI---------------
```

>MOUSE

MSLEFGSVALQTQNEDEEFDKEDFEREKELQQLLTDLPHDMLDDELSSPERHDSDCSMDGRAAEPHPSEHL
ERKWIERDILP-KPH----------SMNCGNGWEENRSKTEDQHLGYHPGEGGDEGGS----------
----------GYSPPGKREQADLYRLPEDFRPYTGGSKQ-----AASVITFSDPQRDNF-
QQFGLSRGPSCGALEPY-KAVYKPYRNSSVQKNSSPAQEVA-
ASDMFEGLQQQFLGANETDSAENIHIIQLQVLNKAKERQLDSLVEKLKDSERQVRYLSHQLLIVQDEKDGL
ALSLRESQQLFQNGKEREMQLEAQIAALEAQVEAFRVSEEKLTKKLRTTEITLESLKQQLVELHHSESLQR
AREHHESIVASLTQKHEEQVSSLQKNLDATITALQEQESICTRLKDHVQQLERNQEAVRLEKTELINRLTR
SLEDSQKQCAHLLQSGSVQEVAQLQLQLQQAQKAHVLSESMNKALQEELTELKDEISLYESAAELGVLPGD
SEGDLSIELTESCVDLGIKKVNWKQSKANRVTQQESPDEDPSKDELILKLKTQVQRLLTSNSVKRHLVSQL
QSDLRECRETMEAFQQSKDGDSG--METKTDTSEKTTKQLWLESSEAIN--REDILQLKNEVQVLQKQN--
-----
QELKEAEEKLRSTNQDLCNQMRQMVQEFDHDKQEAVARCERTYQQHHEAMKAQIRESLLAKHAVEKQHLLE
VYEGTQSQLRSDLDKMNKEMAAVQECYLEVCREKDGLESTLRKTMEKAQEQKRQLLE-----
AREEYVRKLKLELEEKYQETLKTERQSWLQEQAAGATQQA--------------
EKESRQKLIQQLEKEWQSKLDDSLAAWRKTTSDRGSQTEQVACPAAVSKAEAAAVLAEEQA----
RQVQQEKELATKEALRKPEVELELKYCEIIAQKVETAVQNARSRWIQELPMLAEYKALLRAQQQEWAKQQE
LAVAHRLSLALSEAKEKWKSELENMKPNVMSVKELEEKVHSLQKELELKDEEVPVIVRAEVAKARTEWNKE
KQEEIHKIQEQNEEDYRQFLEDHRNKINEVLAAAKEDFVKQKA----------------------------
----

ELLLQKETEFQACLDQSRKEWTLQEAQQTQVEIRQYEEDTLTVLAYLLKDTQLEYGGDSQDKQLLEAMSAC
SSKWISVQYFEKVKACIQKALHDMLSLLTDSVASEQEKRKVVKSSADTVSWTSS-
EGDSAVPVPLPDSTSVRCAQSSAWLKAEAETD--------------------
KKICEIKGLRCGHCFQELEKEKQECQDLRRKLEKCRRHLQHLERTHRAAVEKLGEENSRVVEELIEENHDM
KNKLEALRALCRTPPRSLSAGAAESAGPSCSRQALEELRGQYIKAVRKIKRDMLRYIQESKERAAEMVKAE
VLRERQETARKMRNYYLSCLQQILQDNGK-
EEGAEKKIMSAASKLATMAELLGTIAESDCRVRCAQAGRSVALPLASEMLTGTERSERSGVNHNIPHYVES
KPNSGKTLPRSVCEQLPGRKAAPRSQRRLEESKHREMRPMASTALPSDCRCGDASCRHSGVLAKDVAPEFV
PCQGEGGFDLHEKRDALGA-GSEPLLYSAAHSFL----GGAEKNSSPRCISE-----
SRHTTLRSPSEMPRLKALMCGSPTET-DSIASEKSQGVGSQDSPVKDGV-GPSSSPAWPSDSTLPC--
GSPAVLFLGDGSQRTQEMLGDSVQWKQFSATSCHPDA-QKSNMVCRSSHTLD-----
LPKETLHSQQGKMGATLGHPSPQSTDMLKTDFKRLSGTGPSSLCQKPLIKLTA-PMPSQQDSGFDSPLE--
-