								_	
Α	Species		mRNA and protein NCBI or Ensembl ID						
				CBP20L		CBP20S		_	
		Homo sapier	ns	NM_007362.3		NM_0010425			
				NP_031388.2		NP_00103600			
		Pan troglody	vtes	XM_516977.2		XM_0011654			
				XP_516977.2		XP_00116548			
		Macaca muli	ata	ENSMMUT00000		XM_0010967			
		C		ENSMMUP00000		XP_00109671 XM_0019287			
		Sus scrofa		XM_001928691.1 XP_001928726.1		XP 00192873			
		Equus cabal	Tue	XM 001501008.1		XM 0015010			
		Equus caban	ius	XP 001501058.1		XP 00150106			
		Monodelphis	e domastica	XM 001367619.1		XM 0013676			
		Monoueiphis	s uomesticu	XP 001367656.1		XP 00136770			
		Mus musculi	us	NM 026554.3		ENSMUST00			
		musemi	4.5	NP 080830.1		ENSMUSP00			
<u> </u>				111_000030.1		Di tollio di co	000110032	_	
B									
Human NM_001042540		10	20 TC 1 1 C C C	30	40	ACCTCCACA	50 C.T.C.A.C.C.C.A	60 C.T.A.C.C.C.C.C.C.	70 A C C A G C A C T T C C G G G
Pan XM_001165488.1									ACCAGCACTTCCGGG
Macaca XM_001096719	ATGTCGGGT	resected	TGAAGGC	GCTGCGCAGCG	GACTCCT	ACGTGGAGG	TGAGCCA	GTACCGGG	ACCAGCACTTCCGGG
Sus XM_001928701	ATGTCGGGI	GGCCTGC	TGAAGGC	GCTGCGCAGC	GACTCCT	A C G TCGAA	CTGAGCCA	GTACCGGG	ACCAGCACTTCCGGG
Equus XM_001501017.1									A C C A G C A C T T C C G G G A C C A G C A C T T T C G G G
Mus ENSMUST00000115178 Monodelphis XM_001367664.1									A C C A G C A G T T C C G A G
monodelpino Am_oo roon oo m									CCAGCACTTCCGGG
Human NM_001042540	80	90	100	110	12		130	140	150 A A A T G G G A C G C G T C T
Pan XM_001165488.1									AAATGGAACGCGTCT
Macaca XM_001096719									AAATGGAACACGTCT
Sus XM_001928701	GTGACAATG	GAAGAACA	GGAGAAG	TTACTGAAGAA	AAGCTA	TGCTGAAAA	TGCCATG	CGGTACATA	AAA <u>T</u> GGAACTCGTCT
Equus XM_001501017.1	GTGACAATG	AAGAACA	GGAAAAA	TTACTGAAGAA	AAGCTA	TGCAGAAA	TGCTATG	CGGTACATA	AAACGGAACTCGTCT
Mus ENSMUST00000115178 Monodelphis XM_001367664.1									AAACGGAACTCGTCT
Monodelphis XM_001367664.									AATGGACTCACCT
	or one and o	AAGAACA	AUAAAAA	TIACIONNONN	innoc i n	CCSOAAAA	reocearo	COOLACALA	inar ouarer corer
	160	170	180	190		200	210	220	230
Human NM_001042540									GGGCGATCTGGGGGC
Pan XM_001165488.1									GGGCGATCTGGGGGC
Macaca XM_001096719 Sus XM_001928701									GGGCGATCTGGGGGC GGACGATCTGGGGGC
Equus XM_001501017.1									GGGCGATCTGGGGGC
Mus ENSMUST00000115178	GGATGACCG	GATCATA	CGGACAG	ACTGGGATGCA	GGCTTT	AAGGAGGG	CAGGCAGT	ATGGACGT	G GAC GG T C T G G G G G T
Monodelphis XM_001367664.1									
	GGATGACCG	RATCATT	CGCACAG	ACTGGGACGCA	GGCTTT	AAGGAGGG	AGGCAGT	ACGGCCGT	GGGCGATCTGGGGGC
	240	250	260	270		280	290	300	310
Human NM_001042540					GCTGGGA				AGAACCAGTGAGTGG
	0.0000000								

C|G|C|G|C|G|G|A|T|G

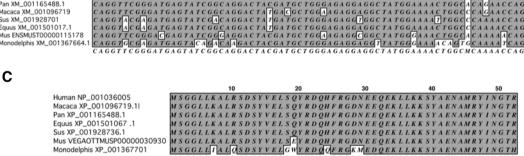
AGAGCAGATG

A G A G C A G A T G

AGAGCAGATG

C|G|T|G|C|G|A|C|G

 $MGMGCAG_{A}ATG$



MSGGLLKALRSDSYVELSQYRDQHFRGDNEEQEKLLKKSYAENAMRYINGTR Human NP_001036005 LDDRIIRTDWDAGFKEGRQYGRGRSGGQVRDEYRQDYDAGRGGYGKLAQNQ Macaca XP_001096719.11 LDDRIIRTDWDAGFKEGRQYGRGRSGGQVRDEYRQDYDAGRGGYGKLAQNQ LDDRIIRTDWDAGFKEGRQYGRGRSGGQVRDEYRQDYDAGRGGYGKLAQNQ Pan XP 001165488.1 Equus XP_001501067 .1 LDDRIIRTDWDAGFKEGRQYGRGRSGGQVRDEYRQDYDAGRGGYGKLAQNQ LDDRIIRTDWDAGFKEGRQYGRGRSGGQVRDEYRQDYDAGRGGYGKLAQNQ Sus XP_001928736.1 Mus VEGAOTTMUSP00000030930 LDDRIIRTDWDAGFKEGRQYGRGRSGGQVRDEYREDYDAGRGGYGKLAQKQ Monodelphis XP_001367701 LDDRIIRTDWDAGFREGRQYGRGRSGGQVRDEYRQDYDAGRGGYGKTVQNQ LDD R I I R T D WD A G F K E G R Q Y G R G R S G G Q V R D E Y R Q D Y D A G R G G Y G K L A Q N Q 5' splice site 3' splice site

GCTGTAC|G|

GCTGCACTC

 $\overline{GCTGTTACRTT}$

D Homo sapiens GCTGTACGPan troglodytes GCTGTACGSus scrofa GCTGTACAT Equus caballus GCTGTACAT

Mus musculus

Figure S1

Monodelphis domestica

Human NM_001042 Pan XM_00116548 Macaca XM_001096 Sus XM_00192870 Equus XM 0015010 Mus ENSMUSTO000 Monodelphis XM_00

Supplementary Figure 1. Sequence alignment of CBP20S RNA and protein found in mammalian species.

(A) A Blast search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was conducted to find CBP20S in other species. A tabular summary of the results is presented, showing the species where CBP20S ESTs were found. (B and C) CBP20S mRNA (A) and protein (B) sequences from all seven species were aligned, using MacVector. (D) Alignment of 5' and 3' splice sites used in the alternative splicing event leading to CBP20S formation. Yellow arrowheads indicate the conserved cleavage sites.