Supplementary methods

Polymerase chain reaction (PCR) method

Sequence information for human primers was derived from Ensembl Genome Browser. The first gene candidate corresponds to accession number (ENST_00000333071), the second (ENST_00000332871) and the third (ENST_00000302973). The primers were following: for the first gene candidate forward primer (HF1) 5'-TCTATAGGTGCCTGGTGC-3' and reverse primer (HR1) 5'-GTGCATCTCCATAGCCTG-3' with the resulting PCR product of 372 bp; for the second gene candidate forward primer (HF1) and reverse primer (HR2) 5'-CTCTCTCCAGCATGGCAT-3' with the resulting product of 298 bp; for the third gene candidate forward primer (HF2) 5'-CTTCCGAGGCTATGACTC-3' and reverse primer (HR3) 5'-AGATTCACTGAGAGCCCTG-3' with the resulting product of 380 bp. All gene candidates should have been recognized by forward primer (HF2) and reverse primer (HR4) 5'-AGTAGCTCGAGGTGTTGG-3' with the resulting products of 444 bp for the first, 416 bp for the second and 425 bp for the third gene candidate.

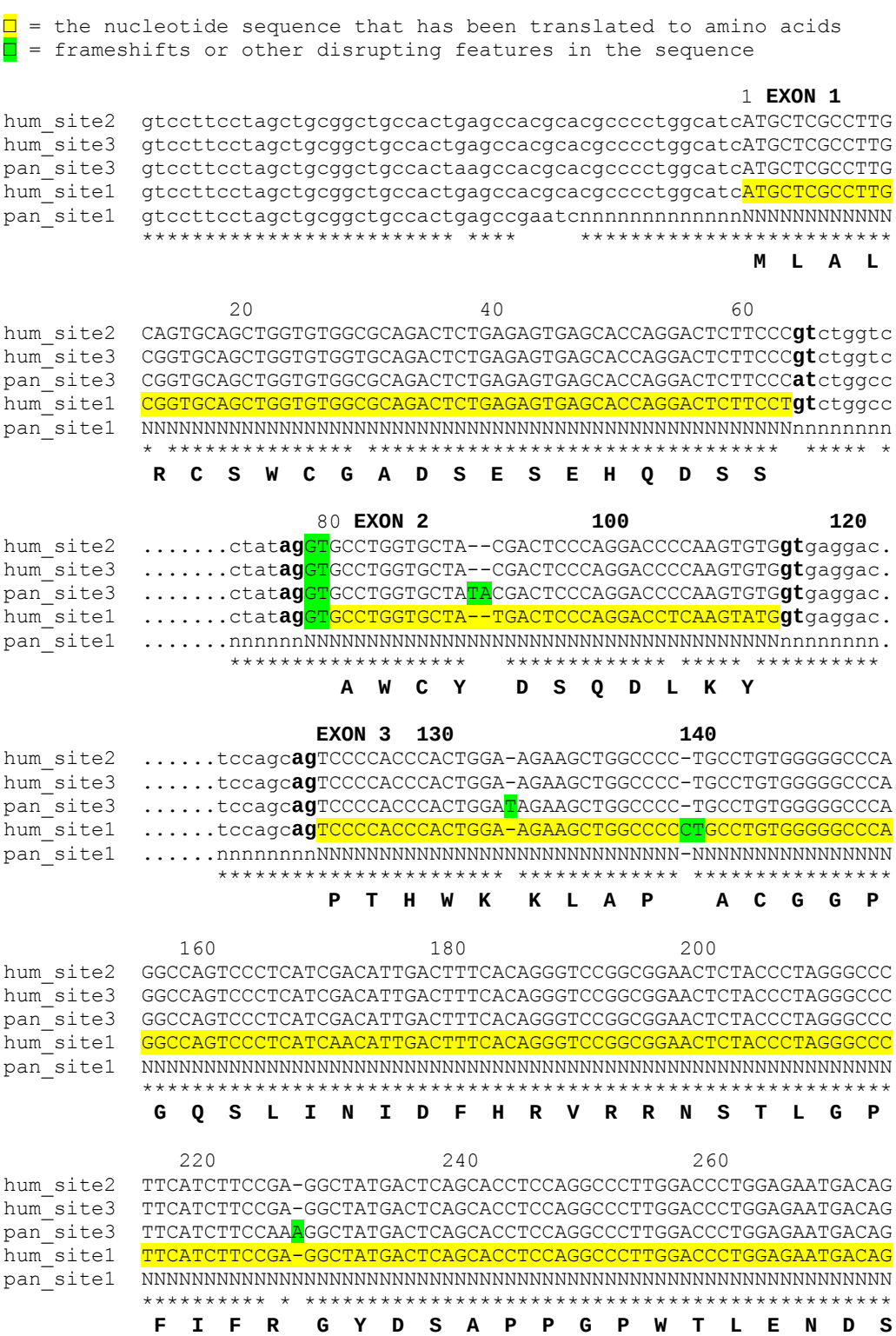
The primers were produced by Sigma Genosys (Cambridge, UK) and all the other reagents were from BD Biosciences except for the dNTP mix that was from Finnzymes (Espoo, Finland). 5 ng of cDNA was used as template. The PCR reactions were carried out on a thermal cycler (Gene Amp PCR system 9700, Applied Biosystems, Foster City, CA). For mouse cDNA studies a touchdown PCR protocol was carried out: it consisted of 94°C denaturation step for 1 min followed by 3 cycles of 94°C for 30 s, annealing at 60°C for 30 s, extension at 68°C for 1 min 30 s, followed by 4 cycles where annealing temperature was 58°C, followed by 26 cycles where annealing temperature was 56°C and finally extension at 68°C for 3 min. For human cDNA studies a single step protocol was used that consisted of 94°C for 1 min followed by 33 cycles of 94°C for 30 s, 54°C for 30 s and 72°C for 1 min 30 s and finally 72°C for 3 min. Primers HF2 and HR4 were used to study all the 15 human tissues and in addition, primer pairs HF1/HR1, HF1/HR2 and HF2/HR3 were used to study kidney, brain and heart.

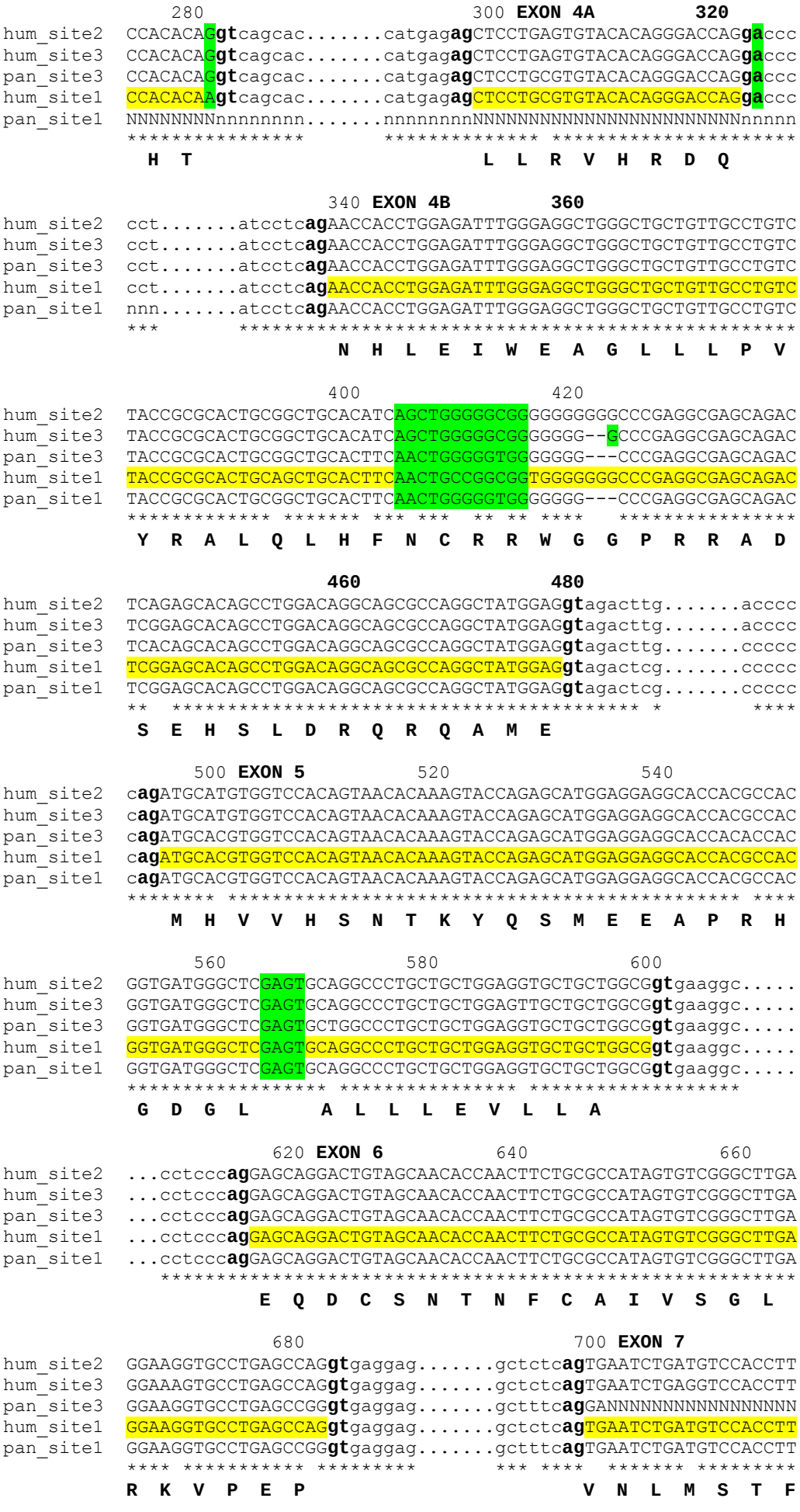
Sequencing of the PCR products

For the sequencing, PCR products were first purified with a GFX PCR DNA and Gel Band Purification Kit (Amersham Biosciences, Poole, UK) from the gel, then cloned into pGEM-T Easy Vector System I (Promega, Madison, WI) and finally the vectors were transformed into TOP10 cells (Invitrogen) according to manufacturers' instructions. The plasmids were purified using the Qiagen Spin Miniprep Kit (Hilden, Germany) following the protocol given by the manufacturer. The sequencing was carried out using ABI PRISM Big Dye Terminator Cycle Sequencing Ready

Reactions Kit version 3.1 (Applied Biosystems). The sequencing was performed in both directions and the primers were designed for vector's SP6 and T7 promoter regions and ordered from Sigma Genosys. 5 µl of purified plasmid was mixed with 4 µl of Big Dye mix and 1.6 pmol of primers were added. The reactions were amplified by cycle sequencing on a thermal cycler (Gene Amp PCR system 9700) according to the manufacturer's protocol. The products were purified by ethanol precipitation, resuspended in HiDi formamide (Applied Biosystems) and denatured according to the manufacturer's instructions. The sequencing was performed with an ABI PRISM Genetic Analyser instrument 9100 (Applied Biosystems).

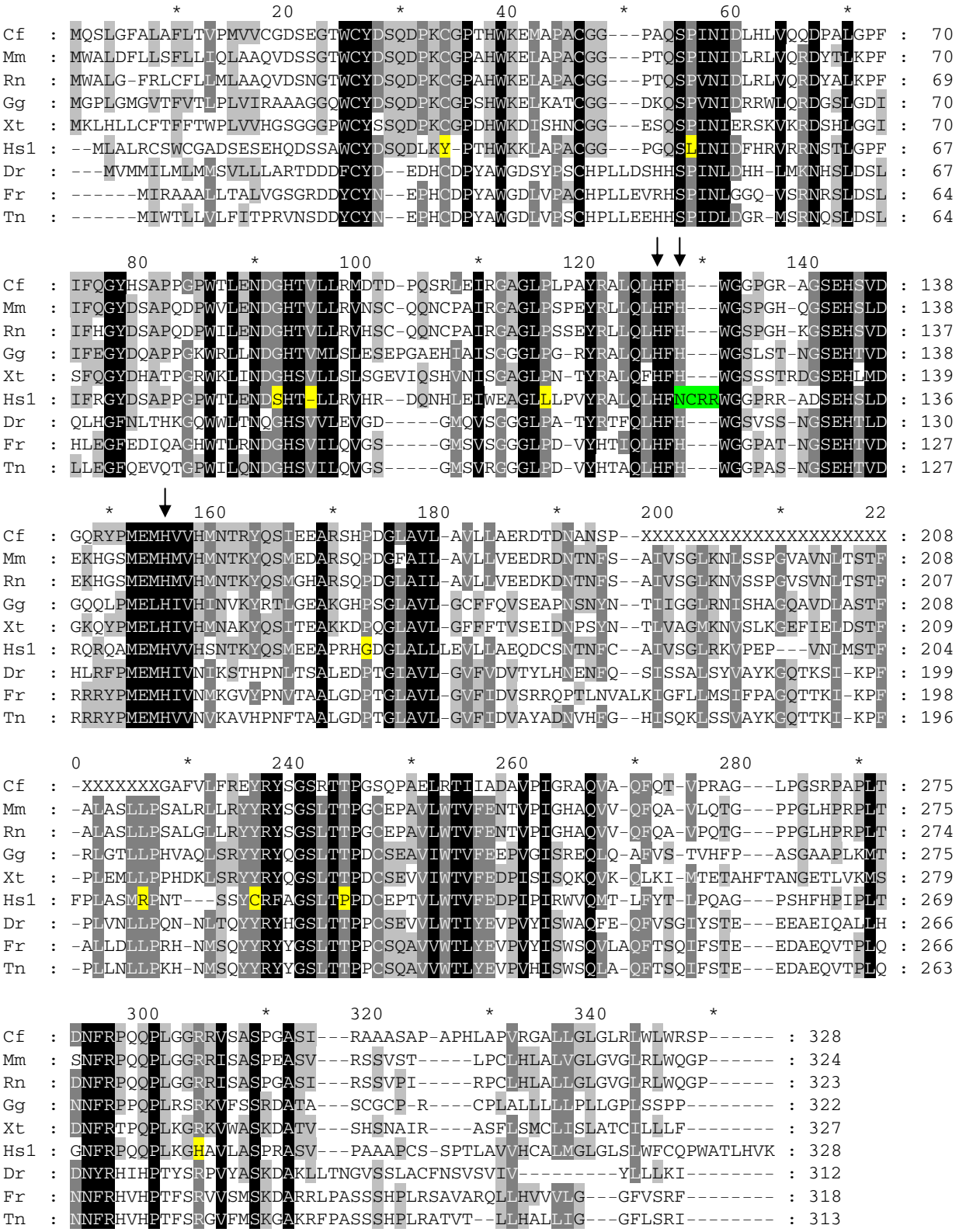
Supplementary Figure 1. The aligment contains all the gene candidates for *CA15* that were found in the human and chimpanzee genomes. A protein sequence was constructed by using the functional mouse CA XV and the alignment in Figure 1. For clarity, only the construction of one protein sequence is shown; it corresponds to human gene candidate 1 (hum_site1). The yellow color shows the nucleotide sequence used to make the construct. The errors in the sequences are highlighted with green color. All the gene candidates contain frameshifts that would destroy the protein sequence for CA XV: the frameshifts are found at positions 79-80, 283, 566-569. The exon 8 in each gene is disrupted by an AluY repeat sequence. Exon 4 is also split into two exons (named A and B), and the intron between them does not conform to the GT-AG rule (nucleotide 325). In addition, every gene candidate has its own frameshift(s), shown in green. These errors confirm that humans and chimpanzees are not able to produce a functional CA XV protein.




280
300 EXON 4A
320
hum_site2 CCACACAGgtcagcac.......catgagagCTCCTGAGTGTACACAGGGACCAGgaccc
hum_site3 CCACACAGgtcagcac.......catgagagCTCCTGAGTGTACACAGGGACCAGgaccc
pan_site3 CCACACAGgtcagcac.......catgagagCTCCTGCGTGTACACAGGGACCAGgaccc
hum_site1 CCACACAAgtcagcac.......catgagagCTCCTGCGTGTACACAGGGACCAGgaccc
pan_site1 NNNNNNNNnnnnnnnn.......nnnnnnnnNNNNNNNNNNNNNNNNNNNNNNnnnnn
H T
L L R V H R D Q
340 EXON 4B
360
hum_site2 cct.......atcctcagAACCACCTGGAGATTTGGGAGGCTGGGCTGCTGTTGCCTGTC
hum_site3 cct.......atcctcagAACCACCTGGAGATTTGGGAGGCTGGGCTGCTGTTGCCTGTC
pan_site3 cct.......atcctcagAACCACCTGGAGATTTGGGAGGCTGGGCTGCTGTTGCCTGTC
hum_site1 cct.......atcctcagAACCACCTGGAGATTTGGGAGGCTGGGCTGCTGTTGCCTGTC
pan_site1 nnn.......atcctcagAACCACCTGGAGATTTGGGAGGCTGGGCTGCTGTTGCCTGTC
N H L E I W E A G L L L P V
400
420
hum_site2 TACCGCGCACTGCGGCTGCACATCAGCTGGGGGCGGGGGGGGGCCCGAGGCGAGCAGAC
hum_site3 TACCGCGCACTGCGGCTGCACATCAGCTGGGGGCGGGGGG--GCCCGAGGCGAGCAGAC
pan_site3 TACCGCGCACTGCGGCTGCACTTCAACTGGGGGTGGGGGG---CCCGAGGCGAGCAGAC
hum_site1 TACCGCGCACTGCAGCTGCACTTCAACTGCCGGCGGTGGGGGGGCCCGAGGCGAGCAGAC
pan_site1 TACCGCGCACTGCGGCTGCACTTCAACTGGGGGTGGGGGG---CCCGAGGCGAGCAGAC
Y R A L Q L H F N C R R W G G P R R A D
460
480
hum_site2 TCAGAGCACAGCCTGGACAGGCAGCGCCAGGCTATGGAGgtagacttg.......acccc
hum_site3 TCGGAGCACAGCCTGGACAGGCAGCGCCAGGCTATGGAGgtagacttg.......acccc
pan_site3 TCACAGCACAGCCTGGACAGGCAGCGCCAGGCTATGGAGgtagacttg.......ccccc
hum_site1 TCGGAGCACAGCCTGGACAGGCAGCGCCAGGCTATGGAGgtagactcg.......ccccc
pan_site1 TCGGAGCACAGCCTGGACAGGCAGCGCCAGGCTATGGAGgtagactcg.......ccccc
S E H S L D R Q R Q A M E
500 EXON 5
520
540
hum_site2 cagATGCATGTGGTCCACAGTAACACAAAGTACCAGAGCATGGAGGAGGCACCACGCCAC
hum_site3 cagATGCATGTGGTCCACAGTAACACAAAGTACCAGAGCATGGAGGAGGCACCACGCCAC
pan_site3 cagATGCACGTGGTCCACAGTAACACAAAGTACCAGAGCATGGAGGAGGCACCACACCAC
hum_site1 cagATGCACGTGGTCCACAGTAACACAAAGTACCAGAGCATGGAGGAGGCACCACGCCAC
pan_site1 cagATGCACGTGGTCCACAGTAACACAAAGTACCAGAGCATGGAGGAGGCACCACGCCAC
M H V V H S N T K Y Q S M E E A P R H
560
580
600
hum_site2 GGTGATGGGCTCGAGTGCAGGCCCTGCTGCTGGAGGTGCTGCTGGCGgtgaaggc.....
hum_site3 GGTGATGGGCTCGAGTGCAGGCCCTGCTGCTGGAGTTGCTGCTGGCGgtgaaggc.....
pan_site3 GGTGATGGGCTCGAGTGCTGGCCCTGCTGCTGGAGGTGCTGCTGGCGgtgaaggc.....
hum_site1 GGTGATGGGCTCGAGTGCAGGCCCTGCTGCTGGAGGTGCTGCTGGCGgtgaaggc.....
pan_site1 GGTGATGGGCTCGAGTGCAGGCCCTGCTGCTGGAGGTGCTGCTGGCGgtgaaggc.....
G D G L A L L L E V L L A
620 EXON 6
640
660
hum_site2 ...cctcccagGAGCAGGACTGTAGCAACACCAACTTCTGCGCCATAGTGTCGGGCTTGA
hum_site3 ...cctcccagGAGCAGGACTGTAGCAACACCAACTTCTGCGCCATAGTGTCGGGCTTGA
pan_site3 ...cctcccagGAGCAGGACTGTAGCAACACCAACTTCTGCGCCATAGTGTCGGGCTTGA
hum_site1 ...cctcccagGAGCAGGACTGTAGCAACACCAACTTCTGCGCCATAGTGTCGGGCTTGA
pan_site1 ...cctcccagGAGCAGGACTGTAGCAACACCAACTTCTGCGCCATAGTGTCGGGCTTGA
E Q D C S N T N F C A I V S G L
680
700 EXON 7
hum_site2 GGAAGGTGCCTGAGCCAGgtgaggag.......gctctcagTGAATCTGATGTCCACCTT
hum_site3 GGAAAGTGCCTGAGCCAGgtgaggag.......gctctcagTGAATCTGAGGTCCACCTT
pan_site3 GGAAGGTGCCTGAGCCGGgtgaggag.......gctttcagGANNNNNNNNNNNNNNNN
hum_site1 GGAAGGTGCCTGAGCCAGgtgaggag.......gctctcagTGAATCTGATGTCCACCTT
pan_site1 GGAAGGTGCCTGAGCCGGgtgaggag.......gctttcagTGAATCTGATGTCCACCTT
R K V P E P
V N L M S T F

```
           720                 740                 760
hum_site2  CTTCCTGCTGGCGTCGATGCGGCCCAACACCTCGAGCTACTGTCGCTTCGCTGGGTCACT
hum_site3  CTTCCTGCTGGCGTCGATGCGGCCCAACACCTCGAGCTACTGTCGCTTCGCTGGGTCACT
pan_site3  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
hum_site1  CTTCCCGCTGGCGTCGATGCGGCCCAACACCTCGAGCTACTGTCGCTTCGCTGGGTCGCT
pan_site1  CTTCCCGCTGGCGTCGATGTGGCCCAACATCTCGAGCTACTGTCGCTTTGCTGGGTCACT
           ***** ************* ********* ****************** ******** **
             F  P  L  A  S  M  R  P  N  T  S  S  Y  C  R  F  A  G  S  L

           780                 800                 820
hum_site2  GACCCCGCCTGACTGCGAGCCCACGGTGCTCTGGACCGTCTTCGAGGACCCCATACCCAT
hum_site3  GACCCCGCCTGACTGCGAGCCCACGGTGCTCTGGACCGTCTTCGAGGACCCCATACCCAT
pan_site3  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
hum_site1  GACCCCGCCTGACTGCGAGCCCACGGTGCTCTGGACCGTCTTCGAGGACCCCATACCCAT
pan_site1  GACCCCGCCTGACTGTGAGCCCACGGTGCTCTGGACCGTCTTTGAGGACCCCATACGCCT
           ************** ************************ ************ * *
             T  P  P  D  C  E  P  T  V  L  W  T  V  F  E  D  P  I  P  I

           840                         860     EXON 8A     880
hum_site2  CCGGTGGGTGCAGgtggagc.......accctcagATGACCCTGTTCCACACC-GTGCCC
hum_site3  CCGGTGGGTGCAGgtggagc.......accctcagATGACCCTGTTCCACAGC-GTGCCC
pan_site3  NNNNNNNNNNNNNnnnnnnn.......-----------------------------------
hum_site1  CCGGTGGGTGCAGgtggagc.......accctcagATGACCCTGTTCTACACC-CTGCCC
pan_site1  CCGGTGGGTGCAGgtggagc.......accctcagATGACCCTGTTCCACACCCGTGCCC
           ******************       ******************** *****  *****
             R  W  V  Q                      M  T  L  F  Y  T   L  P

                  900                 1000                1020
hum_site2  CAGGCTGGACCCTCCCACTTTCACCCCATAC-CGCTCACGGGTAACTTCCGCCCACAGCA
hum_site3  CAGGCTGGACCCTCCCACTTTCACCCCATAC-CGCTCACGGGTAACTTCCGCCCACAGCA
pan_site3  ------------------------------------------------------------
hum_site1  CAGGCTGGACCCTCCCACTTTCACCCCATAC-CGCTCACGGGTAACTTCCGCCCCCAGCA
pan_site1  CAGGCTGGACCCTCCCACTTTCACCCCATACGCGCTCACGGGTAATTTCCGCCCACAGCA
           ****************************** ************* ******** *****
            Q  A  G  P  S  H  F  H  P  I   P  L  T  G  N  F  R  P  Q  Q

                    1040                1060                1080
hum_site2  GCCTC--------ttttttttttttgagacgg-ggtcttgctgtgtcgcccag.......
hum_site3  GCCTC--------ttttttttttttgagacgg-ggtcttgctgtgtcgcccag.......
pan_site3  --------------ctttttctttttgagacgg-ggacttgctgtgtcgcccag.......
hum_site1  GCCTCTTttttttttttttttttttttgagacgg-aggctcactctgtcgcccag.......
pan_site1  GCCTCTTtt---ttttttttttttttgagacggagggctcactctgtcgcccag.......
           *****          ***** ***********  * **  ** **********
             P  L

                                  EXON 8B 1100                1120
hum_site2  AluY repeat.......agcctcttAAGGGGCACACAGTCTTGGCCTCCCCCAGAGCCT
hum_site3  AluY repeat.......agcctcttAAGGGGCACGCAGTCTTGGCCTCCCCCAGAGCCT
pan_site3  AluY repeat.......agcctcttAAGGGGCACACAGTCTTGGCCTCCCCCAGAGCCT
hum_site1  AluY repeat.......agcctcttAAGGGGCACGCAGTCTTGGCCTCTCCGAGAGCCT
pan_site1  AluY repeat.......agcctcttAAGGGGCACGCAGTCTTGGCCTCCCCGAGAGCCT
                           ***************** ************* ** *******
                                    K  G  H  A  V  L  A  S  P  R  A

                        1140                1160                1180
hum_site2  CGGTCCCCGCAGCAGCCCCCC-GCTCTTCCCCCACCCTAGCAGGTGTGCACTGCGCTCTG
hum_site3  CGGTCCCCGCAGCAGCCCCCC-GCTCTTCCCCCACCCTAGCAGGTGTGCACTGCGCTCTG
pan_site3  CGGTCCCCGCAGCAGTCCCCCTGCTCTTCCCCCACCCTAGCAGGTGTGCACTGCGCTCTG
hum_site1  CGGTCCCCGCAGCAGCCCCCT-GCTCTTCCCCCACCCTAGCAGTTGTGCACTGCGCTCTG
pan_site1  CGGTCCCCGCAGCAGCCCCCT-GCTCTTCCCCCACCCTAGCAGTTGCGTACTGCGCTCTG
           ************** ****  ********************** ** * ***********
           S  V  P  A  A  A  P   C  S  S  P  T  L  A  V  V  H  C  A  L

                        1200                1220                1240
hum_site2  ATGCGCCTGGGGCTCAGCTTGTGGTTCTGTCAACCGTGGGCGACCCTCCGTGTGAAAtaa
hum_site3  ATGCGCCTGGGGCTCAGCTTGTGGTTCTGTCAACCGTGGGCGACCCTCCATGTGAAAtaa
pan_site3  ATGGGCCTGGGGCTCAGCTTGTGGTTCTGTCAACCGTGGGCGACCCTCCATGTGAAAtaa
hum_site1  ATGGGCCTGGGGCTCAGCCTGTGGTTCTGTCAACCGTGGGCGACCCTCCATGTGAAAtaa
pan_site1  ATGGGCCTGGGGCTCAGCCTGTGGTTCTGTCCACCGTGGGCGACCCTCCATGTGAAAtaa
           *** ************** ************ ***************** **********
            M  G  L  G  L  S  L  W  F  C  Q  P  W  A  T  L  H  V  K
```

Supplementary Figure 2. The alignment includes CA XV in eight species and additionally the hypothetical protein produced from human *CA15* pseudogene copy 1 (Hs1), in which frameshifts, splicing problems and the AluY insert have been ignored to reconstruct the maximal amount of the protein sequence. The yellow color indicates those residues that are conserved in all other species but have changed in our protein reconstruction. In addition, green color highlights the region that would disrupt the active site of CA XV. The abbreviations are the same as used in Figure 1.



Supplementary Figure 3. Alignment of hypothetical transcripts to human chromosome 22. Three transcript predictions from Ensembl (corresponding to three copies of CA XV in the human genome, ENSTnnnn), one RefSeq prediction (XM_377696) and our final reconstruction (hum1_site1_mRNA) were aligned to the human genome using BLAT. Results from chromosome 22 at 17.4 MB are shown. Similar hits for all five sequences can be found in chromosome 22 at 18.9 MB and 20.0 MB (data not shown). The sequences in RefSeq and Ensembl were found to be conflicting and to contain various errors due to frameshifts in the genomic sequences, which has misled automatic gene prediction programs. The direction of the sequence is shown by arrows in the intron sequences (exon 1 is on the right and 8 on the left).