



Article scientifique

Article

2023

Supplemental data

Open Access

This file is a(n) Supplemental data of:

TP53BP1, a dual-coding gene, uses promoter switching and translational reinitiation to express a smORF protein

Inchingolo, Marta Angela; Diman, Aurélie; Adamczewski, Maxime; Humphreys, Tom; Jaquier-Gubler, Pascale; Curran, Joseph

This publication URL:

<https://archive-ouverte.unige.ch/unige:168662>

Publication DOI:

[10.1016/j.isci.2023.106757](https://doi.org/10.1016/j.isci.2023.106757)

Supplemental information

***TP53BP1*, a dual-coding gene, uses promoter switching and translational reinitiation to express a smORF protein**

Marta A. Inchingolo, Aurélie Diman, Maxime Adamczewski, Tom Humphreys, Pascale Jaquier-Gubler, and Joseph A. Curran

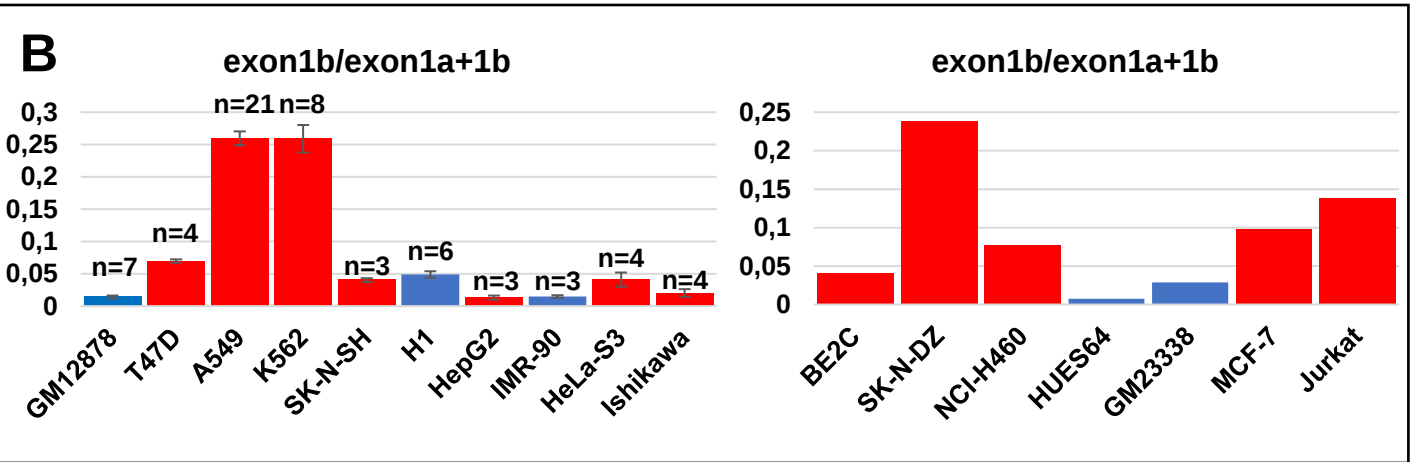
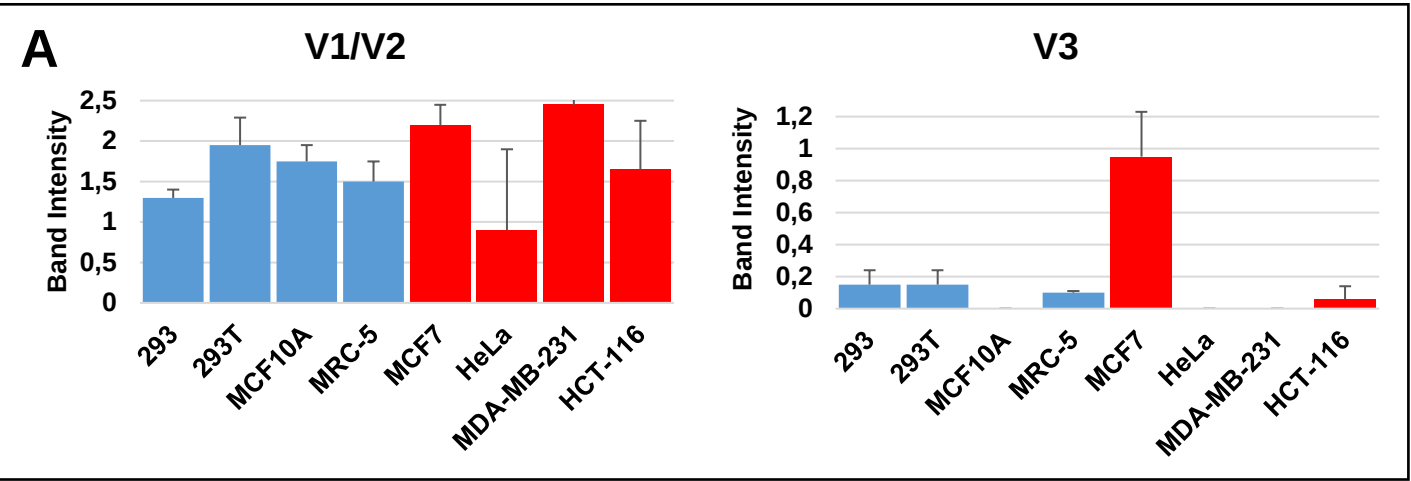


Figure S1 The human 53BP1 variants and their expression levels in tumoural and non-tumoural cell lines, related to Figure 1.

(A) Levels of each transcript variant in a range of cell lines available in the lab. RT-PCR was performed with primer sets specific for the V1/2 and V3 variants. Amplicons were analysed on agarose gels and band intensities (arbitrary units) were normalised to an actin amplicon control. The results are from biological triplicates. Bars indicate the SD.

(B) The abundance of the exon 1b (derived from the P2 promoter and present in V3: exon 1a arises from the P1 promoter and is present on V1/2) relative to total exon 1 expression (P1 + P2 or exon 1a + 1b) from a series of cell lines. Transcriptome data was extracted from the ENCODE site (<https://www.encodeproject.org>) selecting only data sets that had undergone validation and initial filtering. The total number of reads corresponding to each exon 1 was determined and used to estimate the fraction arising from the P2 promoter (reads exon 1b/reads 1a+1b). The left hand panel shows cell lines in which multiple independent data sets were available (indicated by n above the columns). SD values are shown. The right hand panel shows cell lines for which only a single study was available. In both panels red columns indicate tumoural and blue columns non-tumoural cells.

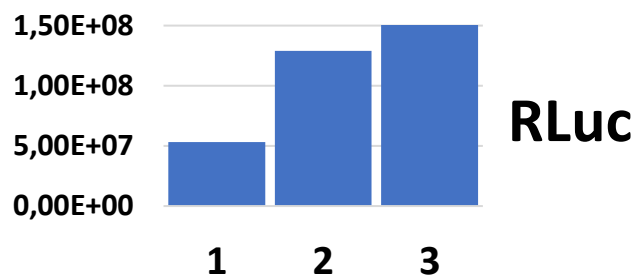
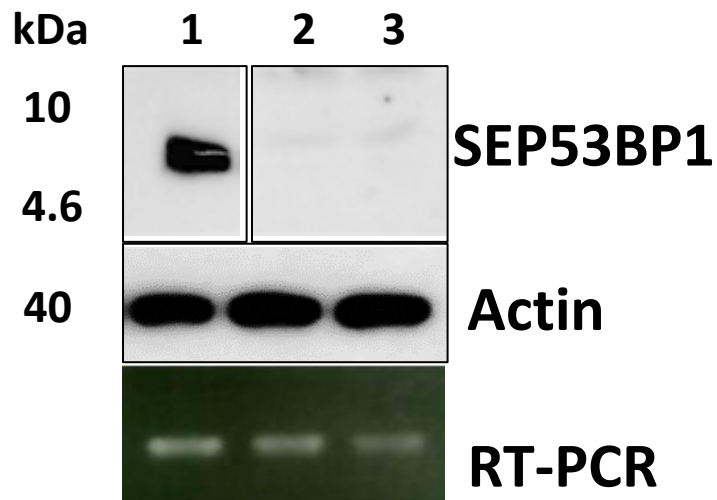
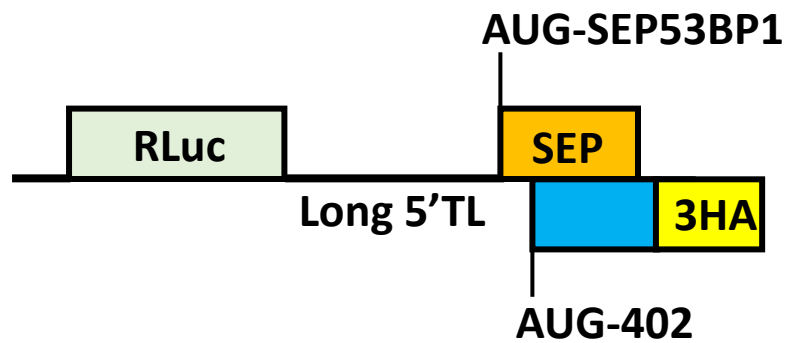
53BP1 (N-term)

HUMAN	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFSMLSRHLPNLQTHKENPVL--DVVSNPEQTAGEERGDSNGSGFNEHLKENKVA-DPVDSSNLDTCGSISQVIEQLPQPN
Chimp	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFSMLSRHLPNLQTHKENPVL--DVVSNPEQTAGEERGDSNGSGINEHLKENKVA-EPVDSSNLDTCGSISQVIEQLPQPN
Rattus	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQALEEDAGSHFSVLSRHLPNLQMHKENPVL--DIMSNEQTVA-EQGDN-SSFSEHLKEKKAS-DPVESSHLGTSGPISQVIERLPQPN
Green	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFSMLSRHLPNLQTHKENPVL--DVVSNPEQTAGEERGDSNGSGFNEHLKENKVA-DPVDSSNLDTCGSISQVIEQLPQPN
Orangutan	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFSMLSRHLPNLQTHKENPVL--DVVSNPEQTAGEERGDSNGSGFNEHLKENKVA-DPMDSSNLDTCGSISQVIEQLPQPN
Squirrel	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFSMLSRHLPNLQTHKENPVL--DVVSNPEQTAGEERGDSNGSGFNEHLKENKVA-DPVDSSNLDTCGSISQVIEQLPQPN
Snub	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFSMLSRHLPNLQTHKENTVL--DVVSNPEQTAGEERGDSNGSGFNEHLKENKVA-DPVDSSNLDTCGSISQVIEQLPQPN
Macaque	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFSMLSRHLPNLQTHKENPVL--DVVSNPEQTAGEERGDSNGSGFNEHLKENKVA-DPVDSSNLDTCGSISQVIEQLPQPN
Sooty	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFSMLSRHLPNLQTHKENPVLTDVFKPKPTTGEERGDSNGSGFNEHLKENKVA-DPVDSSNLDTCGSISQVIEQLPQPN
Tarsier	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFGVLSRHLLNLQTHKENPVL--DVVSNPERTAGEEQGDSNGSGFSEHLKENKAA-DAIDSSHMDACGSISQVIEQLPQPN
Tarsius	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFGVLSRHLLNLQTHKENPVL--DVVSNPERTAGEEQGDSNGSGFSEHLKENKAA-DAIDSSHMDACGSISQVIEQLPQPN
Mouse	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFGVLSRHLPNLQTHKENPVL--DVVSNPEQTAGEEQGDSNGSGFSEHLKENKAA-DSMDSSHLDTCGSISQVIEQLPQPN
Squirrel	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFSMLSRHLPNLQTHKENPVL--DVVSDPEQAAGEEQGDSNGSGFTEHLKD-KAA-VPVDSSHLDTCGSISQVIEQLPQPN
Hamster	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEEDAGSHFSVLSRHLPNLQTHKENPVL--DIVSNPDQAAEEQGDNNSSF-KHLKENKASADPVDSSHLGTCDISQVIEQLPQPN
Rat	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQALEEDAGSHFSVLSRHLPNLQMHKENPVL--DIMSNEQTVA-EQGDSSNSFSEHLKEKKAS-DPVESSHLGTSGPISQVIERLPQPN
Mole	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFSMLSRHLPNLQTHKENPVL--DAVSNPEQTT-EQGDRNSGFSEHLKENKAA-DPMDSSHLDTCCSSISQVIDRLPQPN
Pig	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSHFSMLSRHLPNLQTHKENPVL--DVVSNTEQTGGEEQGDNNSSFNECLKENKAA-DSMDSSHLDTCGSISQVIEQLPQPN
Elephant	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLDD-SGSHFSMLSRHLPNLQTHKENPVL--DVVSNPEQAAGEEQGDNNSGSNEHLKENKPS-EPMDSSHLDTCCSVSQVIEQLPQPN
Whale	MDPTGSQLESDFSQQDTPCLIIEDSQPESQVLDDSGSRFSMLSRHLPNLQTHKENPLL--DVVSNSEQTAGEEQGDSNGSGFNEQLKENKAA-ESMDSSHLDTCGSISQVIEQLPQPN
Cat	MDPTGSQQLSDFSQQDTPCLIIEDSQPESQVLEDDSGSRFSLSRHLPNLQTHKENPVL--DVVSNPEQTAGE-GGDSNGSFNEDLKENKAA-DPMESSHLDTGGTISQVIEQLPQPN

SEP53BP1

HUMAN	MILVLTSVCYLDTFILISRRTKKILCWMLCPILNKQLEKNEETVIVGSMNI
Chimp	MILVLTSVCYLDTFILISRRSKKILCWMLCPILNKQLEKNEETVIVGSMNI
Green_Monkey	MILVLTSVCYLDTFILIFRRTKKILYWMLCPILNKQLEKNEETVIVGSMNI
Orangutan	MILVLTSVCYLDTFILISRRTKKILCWMLCPTILNKQLEKNEETVIVGSMNI
Squirrel_Monkey	MILVLTSVCCLDTFILIFRRTKKILCWMLCPILNRQLKNEETVIGSSMNI
Snub-nosed_monkey	MILVLTSVCYLDTFILIFRRTKKIPYWMMLCPILNKQLEKNEETVIVGSMNI
Macaque	MILVLTSVCYLDTFILIFRRTKKILYWMLCPILNKQLEKNEETVIVGSMNI
Sooty_mangabey	MILVLTSVCYLDTFILIFRRTKKILYWMLCPILNKQLEKNEETVIVGSMNI
Tarsier	MILGLTSVCSLDTFILYRRTKRTLWCMLCPILNEQLEKNKETVIVGSVNI
Tarsius_syrichtha	MILGLTSVCSLDTFILYRRTKRTLWCMLCPILNEQLEKNKETVIVGSVNI
Mouse_lemur	MILVLTLVCCLDTFILYRRTKRILCWMLCPILNKQLEKNREIVIVGSVNI
Squirrel	MILVLTSVCCLDTFIRISRRTKRTLWCMLCPILNKLEKNKKTIVIVGSVNI
Hamster	-MRALTSVCCLDTFPICRYTKRTLWCILPILNKQLOKSKETI IAPSMNI

Figure S2 ClustalW2 Alignments used to generate data related to Figure 2A.



- 1: Long 5'TL 3HA monocistronic+ β -actin R-Luc
- 2: Long 5'TL 3HA bicistronic clone 3
- 3: Long 5'TL 3HA bicistronic clone 6

Figure S3 The human Long 5'TL does not drive second cistron expression in a plasmid-based bicistronic assay, related to Figure 4.

We generated a bicistronic construct with a first cistron renilla (Rluc) fused upstream of the Long 5'TL-3HA sequence, as depicted in the upper panel. Independent clones of this construct were transiently expressed in HEK293T. In parallel cells were transfected with the monocistronic Long 5'TL-3HA plus a monocistronic β -actin R-Luc. SEP53BP1 and actin levels were monitored by immunoblotting and transcript levels by RT-PCR using a primer set specific for the SEP53BP1 region. Renilla activity in each extract was also recorded. All transcripts carried 5'TL sequences derived from β -actin. Lanes 1, and 2/3 are separates lanes from the same gel.

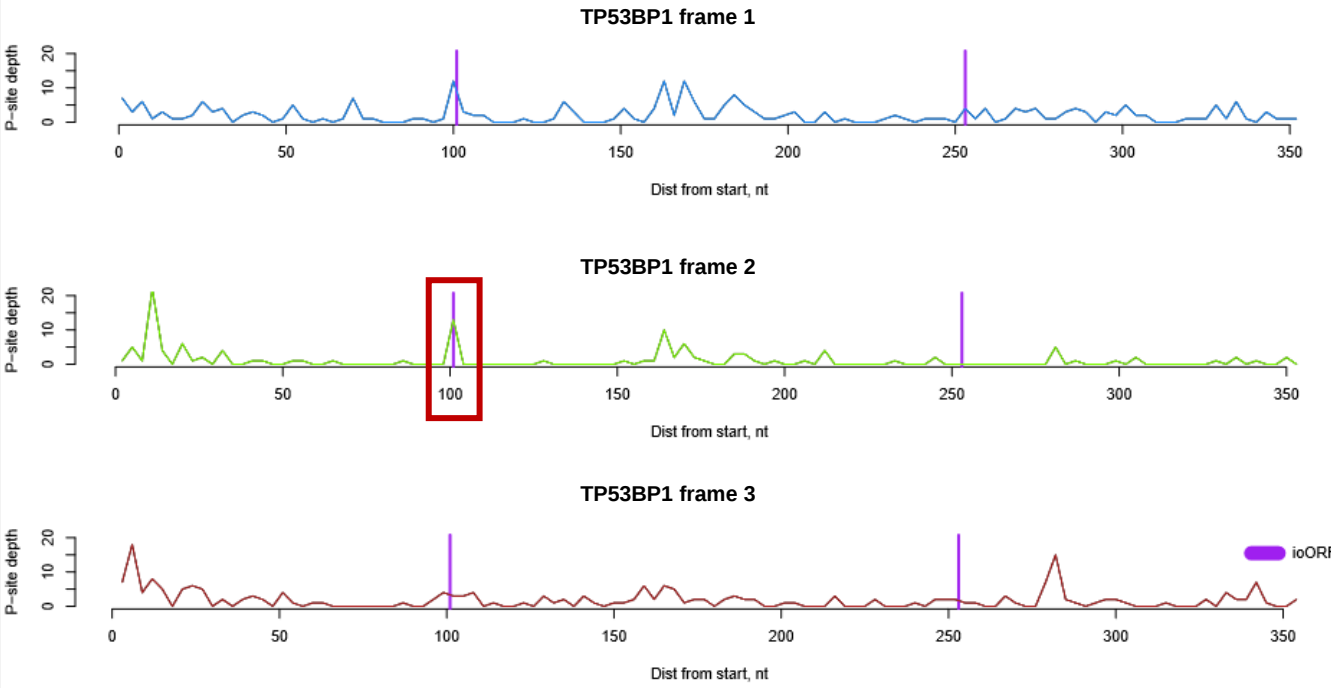


Figure S4 Ribosome Profiling across the SEP53BP1 ioORF region in THP-1 cells. (the violet vertical lines indicate the AUG start and UGA stop sites of the ioORF: the +1 position starts after the AUG^{53BP1(b)}), related to Figure 5.

The data was extracted from the RPF database <http://sysbio.sysu.edu.cn/rpfd/index.html>. The read scoring follows P-site occupancy across the three possible ORFs, with frame 1 corresponding to that of 53BP1 and frame 2 to that of SEP^{53BP1}. The red rectangle highlights the increase in P-site read count that places the AUGSEP53BP1 start codon in the P site.

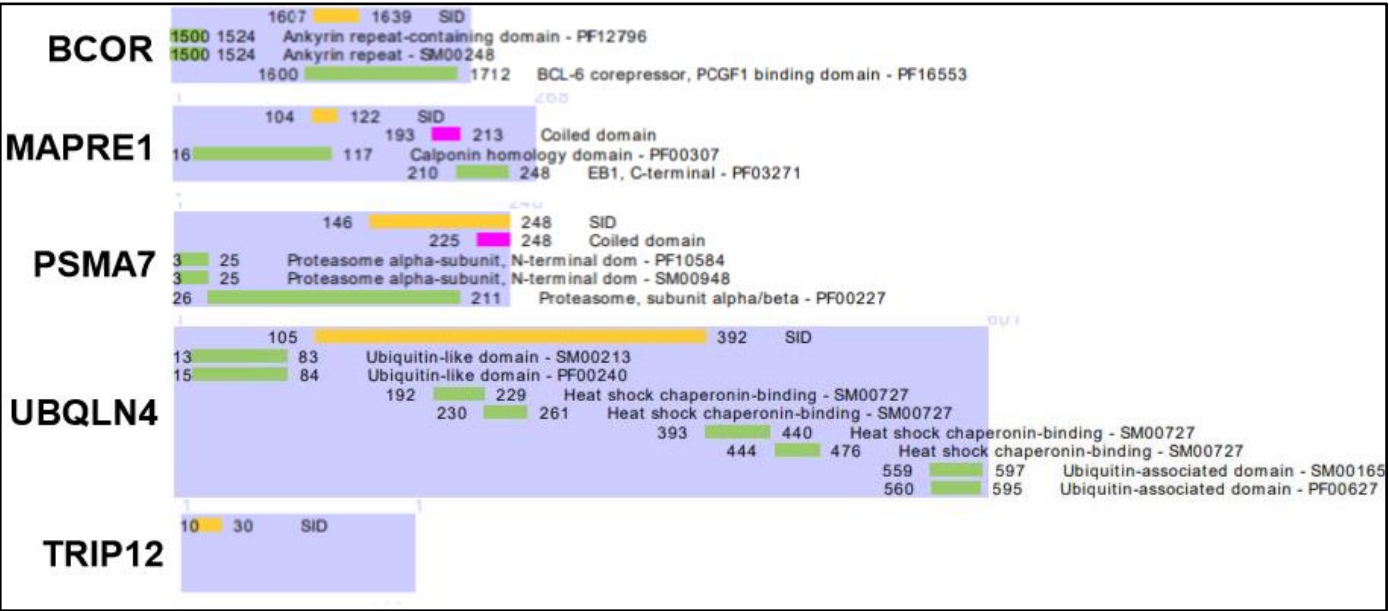


Figure S5 The SEP53BP1 partners as determined by Y2H are represented according to their protein chain length, related to Figure 7.

The selected interaction domains (SID) and the known functional domains of each protein are all indicated.

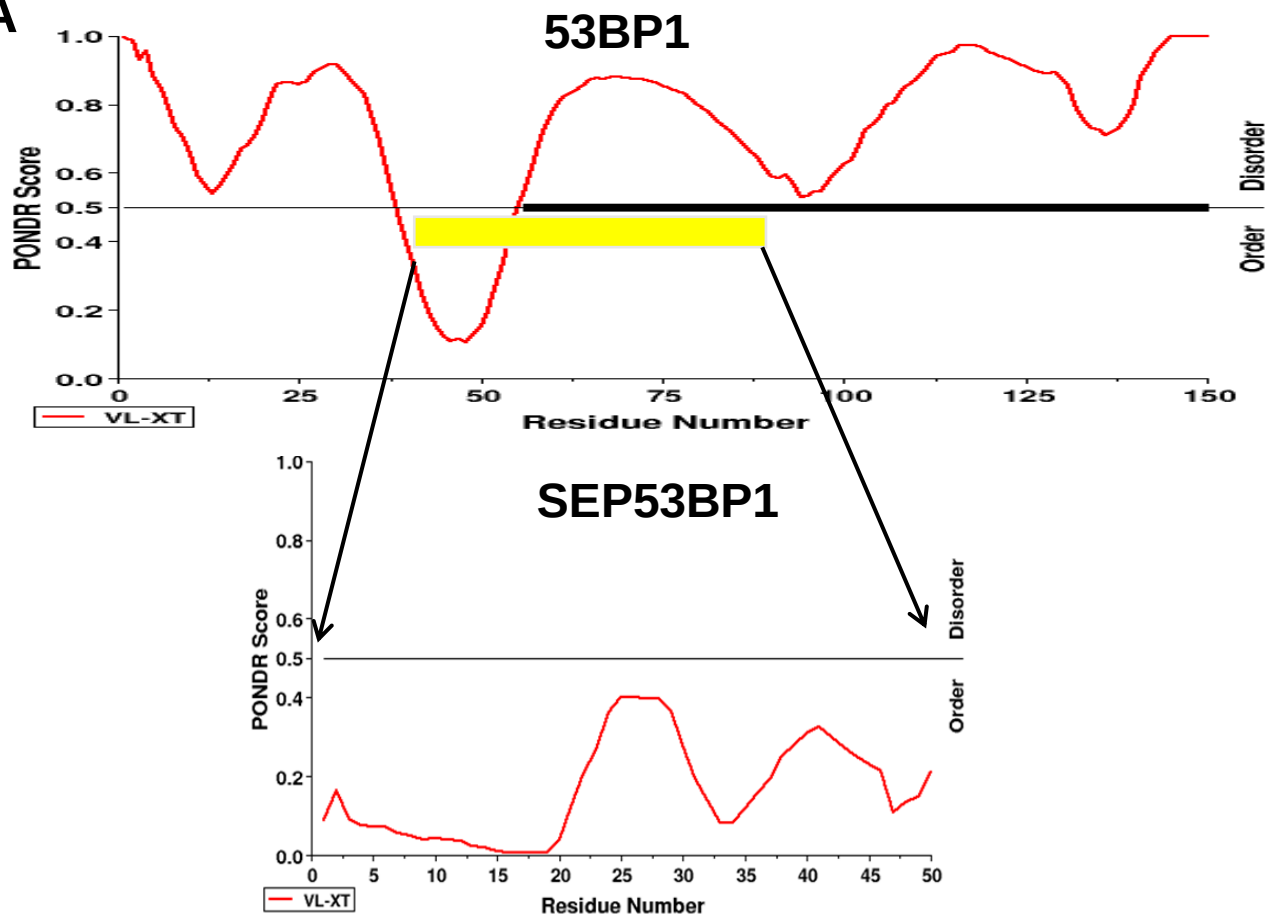
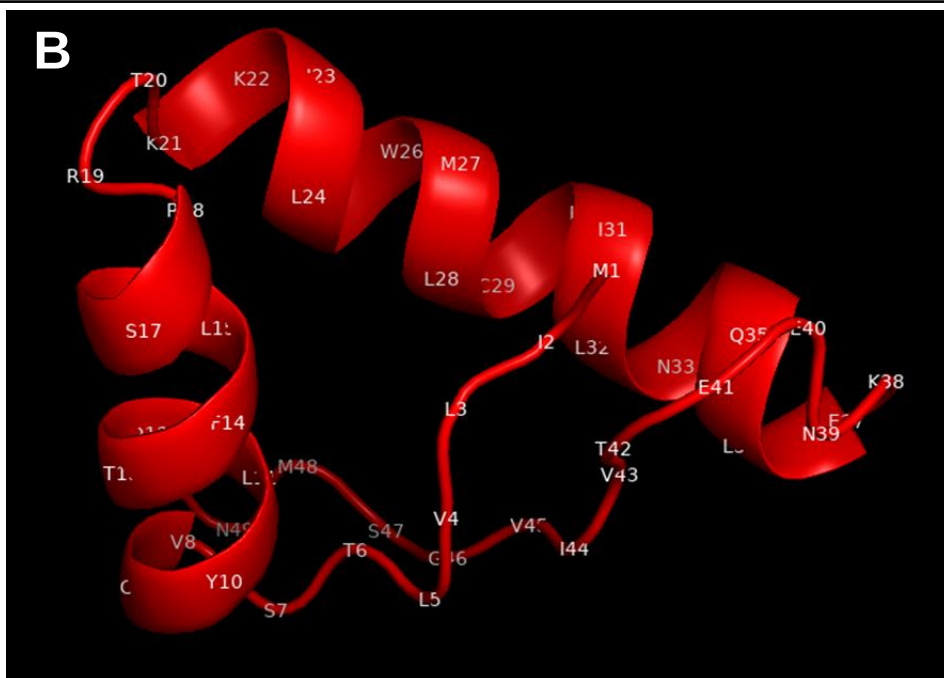
A**B**

Figure S6 Structural predictions for the human SEP53BP1 protein, related to Figure 6.

(A) Structural disorder as measured by PONDR (Predictor of Natural Disordered Region) score (<http://www.pondr.com>). The upper panel covers the N-terminal 150 aa of 53BP1 (starting from the AUG^b start codon) with the positioning of the overlapping SEP53BP1 ORF indicated in yellow. The disorder prediction for the SEP ORF is indicated below.

(B) A three dimensional tertiary structural model generated using I-TASSER (Iterative Threading ASSEmbly Refinement: <https://zhanggroup.org>).