FUS gene is dual-coding with both proteins united in FUS-mediated toxicity

Marie A. Brunet[1, 2,*], Jean-Francois Jacques[1,2], Sonya Nassari[3], Giulia E. Tyzack[4,5], Philip McGoldrick[6], Lorne Zinman[7], Steve Jean[3], Janice Robertson[6], Rickie Patani[4,5], Xavier Roucou[1, 2,*]

[1] Department of Biochemistry and Functional Genomics, Université de Sherbrooke, Sherbrooke, Québec, Canada;

[2] PROTEO, Quebec Network for Research on Protein Function, Structure, and Engineering;

[3] Immunology and Cell Biology Department, Université de Sherbrooke, Sherbrooke, Québec, Canada;

[4] The Francis Crick Institute, 1 Midland Road, London, UK;

[5] Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, Queen Square, London, UK;

[6] Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Canada;

[7] Division of Neurology, Department of Medicine, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Canada.

[*]Corresponding authors:

21 Xavier Roucou, Department of Biochemistry and Functional Genomics, Université de Sherbrooke,
22 3201 Jean Mignault, Sherbrooke, Québec J1E 4K8, Canada, Tel. (819) 821-8000x72240; E-Mail:
23 xavier.roucou@usherbrooke.ca

24 Marie A. Brunet, Department of Biochemistry and Functional Genomics, Université de
25 Sherbrooke, 3201 Jean Mignault, Sherbrooke, Québec J1E 4K8, Canada, Tel. (819) 821-
26 8000x72189; E-Mail: marie.brunet@usherbrooke.ca

27

ABSTRACT:

Novel functional coding sequences (altORFs) are camouflaged within annotated ones (CDS) in a different reading frame. We discovered an altORF nested in the FUS CDS encoding a conserved 169 amino acid protein, altFUS. AltFUS is endogenously expressed in human tissues, notably in the motor cortex and motor neurons. Overexpression of wild-type FUS and/or amyotrophic lateral sclerosis-linked FUS mutants is known to trigger toxic mechanisms in different models. These include an inhibition of autophagy, loss of mitochondrial potential, and accumulation of cytoplasmic aggregates. We show here that altFUS, not FUS, is responsible for the inhibition of autophagy. AltFUS is also pivotal in the mechanisms leading to the mitochondrial potential loss and accumulation of cytoplasmic aggregates. Suppression of altFUS expression in a *Drosophila* model of *FUS*-related toxicity protects against neurodegeneration. Some mutations found in ALS patients are overlooked because of their synonymous effect on the FUS protein. Yet we showed they exert a deleterious effect via their missense consequence on the overlapping altFUS protein. These findings demonstrate that *FUS* is a bicistronic gene and suggest that both proteins, FUS and altFUS, cooperate in toxic mechanisms.

44 **MAIN**

45 FUS is a nuclear RNA-binding protein, with a C-terminal nuclear localization signal (NLS)[1,2]. The
46 protein is involved in RNA processing, DNA repair and cellular proliferation, although its functions
47 are not precisely elucidated[1]. Mutations in *FUS* gene associate with Amyotrophic Lateral Sclerosis
48 (ALS), frontotemporal lobar dementia (FUS-FTLD) and essential tremor, all characterized by FUS
49 cytoplasmic inclusions in neurons and glial cells[1]. Such cytoplasmic FUS aggregates are
50 pathological features in patients with *FUS* mutations or sporadic disease.

51 Recently, mutations in *FUS* 3'UTR were described in ALS patients and linked to an increased level
52 of *FUS* mRNA and protein[3–5]. Surprisingly, over-expression of wild-type FUS provokes an
53 aggressive ALS phenotype in mice and fruit flies, in accordance with findings in yeast and
54 mammalian cells[6–10]. The mechanism of the wild-type or ALS-linked mutated FUS toxicity remains
55 unclear[1,11,12].

56 With currently non-annotated proteins being increasingly reported[13–16], we hypothesized that the
57 toxicity resulting from wild-type FUS over-expression may come from another, unseen, actor[16].
58 These novel proteins are coded by alternative open reading frames (altORFs) that are located
59 within "non-coding" RNAs (ncRNA), within the 5' or 3' "untranslated" regions (UTR) of mRNAs, or
60 overlapping a known coding sequence (CDS) within a different frame of an mRNA[15–17].
61 Serendipitous discoveries and ribosome profiling have recently highlighted the distribution of
62 altORFs throughout the human genome, and the consequences of their absence from current
63 databases[16]. For example, mass spectrometry-based proteomics has become the gold standard
64 for protein identification and has been extensively used in ALS studies[18,19]. However, if a protein
65 is not annotated, it is not included in the protein database (e.g. UniProtKB), and thus cannot be

66 detected by mass spectrometry. An estimated 50 % of mass spectra from a proteomics
67 experiment are unmatched at the end of the analysis[16,20].

68 Genome annotations must avoid spurious ORF annotations. Thus, unless functional
69 characterization has been published, they rely upon 2 arbitrary criteria: a minimum length of 100
70 codons and a single ORF per transcript. Several groups developed tools to challenge such criteria,
71 such as the sORFs repository[21] and the OpenProt database[17], which offer a data-driven broader
72 view of eukaryotic proteomes. The OpenProt database is based on a polycistronic model of ORF
73 annotation[17] and reports any ORF longer than 30 codons within any frame of an mRNA or ncRNA.
74 It contains currently annotated proteins (RefProts), novel Isoforms, and novel alternative proteins
75 (altProts). Here, we used the OpenProt database (www.openprot.org) to ask whether *FUS* may
76 encode additional proteins that could explain the toxicity of the wild-type protein over-
77 expression. In support of this hypothesis, FUS displays an N-terminal prion-like domain. These
78 low-complexity domains are known to harbour overlapping ORFs[22,23].

79 **ALTFUS IS A NOVEL 170 AMINO ACID PROTEIN, ENDOGENOUSLY EXPRESSED IN CELL LINES AND TISSUES**

80 We began by querying OpenProt[17] predictions for *FUS* canonical mRNA (*ENST00000254108* or
81 *NM_004960*), which led to 8 predicted altORFs, either overlapping the coding sequence (CDS) or
82 within the 3'UTR (**Appendix Table 1**). Amongst these, IP_243680 or altFUS, a 170 codon altORF
83 overlapping FUS N-terminal prion-like domain, presents convincing experimental evidence of
84 expression (OpenProt v1.3). AltFUS overlaps the FUS CDS in an open reading frame shifted by one
85 nucleotide (**Fig 1A**, **Fig EV1A**). *FUS* is a complex gene with 13 annotated transcripts resulting from
86 alternative splicing. Based on the GTEx expression data in brain tissues and nerves, five transcripts
87 are more abundant and represent 85 % of all transcripts (**Fig EV1B**). Three of them (*FUS-206*, *FUS-*
88 *211*, *FUS-203*) are non-coding according to Ensembl, while only two (*FUS-211*, *FUS-203*) are non-
89 coding according to OpenProt (**Fig 1B, Fig EV1C**). Ensembl[24] annotates two transcripts as coding
90 (*FUS-201* and *FUS-202*), either for the 526 amino acid FUS protein or its 525 amino acid isoform
91 (**Fig EV1D**). From OpenProt prediction, these two transcripts also encode altFUS (IP_243680), or
92 its 169 amino acid isoform (IP_243691) respectively (**Fig EV1E**). Moreover, the second most
93 abundant transcript in brain tissues and nerves (*FUS-206*), representing about 20 % of all
94 transcripts, is non-coding according to Ensembl, but OpenProt predicts it contains the altFUS CDS
95 (**Fig 1B**, **Fig EV1C**). Thus, of the five most abundant transcripts in brain tissues and nerves, two
96 code for both FUS and altFUS proteins, one codes for altFUS alone, and the remaining two are
97 non-coding.

98 We retrieved nucleotide conservation scores (PhyloP) for *FUS* transcripts over 100 vertebrates.
99 PhyloP scores range from -10 (highly variable) to 10 (highly conserved). Scores over the FUS CDS
100 are under a constraint at the altFUS CDS locus (average score of 2.6 instead of 4 elsewhere on the
101 FUS CDS), which is consistent with a selection pressure across 2 overlapping frames (**Fig 1C**)[25]. We
102 then retrieved altFUS protein sequences over 84 species and observed a strong protein

103 conservation across mammals, and primates notably (75 to 99.4 % of sequence identity -
104 **Appendix Table 2**, **Fig EV1F and Appendix Source Data 1**). Thus, AltFUS is well conserved, with
105 domains showing little to no sequence variations (**Fig 1D**). No functional domain nor clear
106 secondary structures could be inferred from bioinformatics predictions (data not shown).

107 Published RIBO-seq data in Human, retrieved from the Gwips portal[26], revealed an accumulation
108 of initiating ribosomes around the altFUS initiating methionine, in association with an increase in
109 the density of elongating ribosomes over altFUS CDS (**Fig 1E**). These results suggest that altFUS is
110 translated. Similar results were observed in Mouse (**Fig EV2A**).

111 Based on the OpenProt database, AltFUS was identified in multiple proteomics datasets, with up
112 to 7 confident peptides (**Fig 1F, Appendix Table 1**). These peptides are unique to altFUS, or shared
113 with its isoform (IP_243691), and represent a 41 % sequence coverage (**Fig EV2B**). Furthermore,
114 we used a peptide-centric approach to query the TCGA datasets for altFUS expression with the
115 PepQuery algorithm. This approach allowed us to identify 28 peptides unique to altFUS, or shared
116 with its isoform (IP_243691), confidently mapped to mass spectra that could not be better
117 explained by any known protein (hg38_Ensembl database) (**Appendix Table 3**). These peptides
118 span through the entire altFUS sequence, representing a full sequence coverage (**Fig EV2C**). Out
119 of these, 20 peptides were confidently mapped to mass spectra that could not be better explained
120 by any known protein with any post-translational modification and/or chemical artefact (**Fig 1G**,
121 **Fig EV2C**).

122 To further validate altFUS protein expression, we developed a custom antibody targeting two
123 unique altFUS peptides (**Fig EV3A**) and tested it using three constructs: FUS, altFUS and FUS$^{(\emptyset)}$.
124 The latter is a monocistronic *FUS* version, where all altFUS methionines are mutated for
125 threonines in a manner synonymous for FUS (**Fig EV3B-D**). Thus, the FUS protein sequence is

126 unchanged, but the altFUS sequence does not contain any methionines. Transfection of HEK293
127 cells revealed expression of both proteins, FUS and altFUS, from the FUS nucleotide sequence (**Fig**
128 **2A**). As expected, altFUS expression was lost with the monocistronic FUS$^{(\emptyset)}$ construct. HEK293 cells
129 transfected with a siRNA targeting *FUS* mRNA showed a significant knockdown of both proteins,
130 FUS and altFUS; whereas altFUS endogenous expression was visible in scrambled control siRNA
131 and mock transfected cells (**Fig 2B**). These results validate the specificity of the custom antibody
132 for altFUS protein detection by western blot and demonstrate altFUS endogenous expression in
133 HEK293 cultured cells.

134 AltFUS endogenous expression was visible in control human tissues, HEK293 and HeLa cell lines
135 (**Fig 2C**). In order to test altFUS expression in pathological tissues, we retrieved motor cortex
136 lysates from 3 ALS patients with a *C9orf72* mutation (most common genetic cause) and 3 sporadic
137 ALS patients (most common aetiology). AltFUS endogenous expression was detected in all cases
138 (**Fig 2D**). Furthermore, as ALS is a motor neuron disease, we derived functional ventral spinal
139 motor neurons from induced pluripotent stem cells (iPSCs)[27,28] from healthy controls and ALS
140 patients carrying valosin-containing protein mutations (3 lines per group). AltFUS endogenous
141 expression was detected in all samples (**Fig 2E**). We noticed that brain (**Fig 2C**) and motor cortex
142 lysates (**Fig 2D**), as well as iPSCs-derived motor neurons (**Fig 2E**), from healthy controls and ALS
143 patients, presented a higher band detected with the custom altFUS antibody. This band is not
144 present in cultured cell lines or other tissues. It could come from a non-specific signal or a post-
145 translational modification on altFUS that is specific to the motor cortex and spinal cord motor
146 neurons (**Fig 2C-E**). Deep learning predictions of post-translational modifications on altFUS
147 revealed an extensive propensity to phosphorylation and O-GlcNAcylation (O-linked β-N-
148 acetylglucosamine glycosylation), with up to 19 and 18 sites respectively under a high stringency
149 model (**Appendix Table 4**). These two post-translational modifications are abundant in the

150 eukaryotic brain, and known to be dysregulated in neurodegenerative diseases or aging[29–32]. We
151 could also observe a lower band in the line 2 of controls motor neurons, which may correspond
152 to a degradation product or an initiation at a downstream methionine in altFUS sequence. This
153 band has never been observed in other samples so far. We demonstrate that the *FUS* gene
154 encodes two proteins, FUS and altFUS, both endogenously expressed in human tissues, iPSCs and
155 cell lines.

156 **ALTFUS IS A MITOCHONDRIAL PROTEIN**

157 FLAG-tagged altFUS (altFUS-FLAG) displayed a strong co-localization with a common mitochondrial
158 marker, TOMM20 (**Fig 2F**). Additionally, mitochondrial extracts showed an enrichment in altFUS-
159 FLAG (**Fig 2G**). Cellular fractionation of cells over-expressing untagged altFUS further validated
160 altFUS mitochondrial localization (**Fig 2H**). The endogenous altFUS protein was found in the
161 mitochondrial fraction, although it displayed a weak cytoplasmic signal as well (**Fig 2I**), consistent
162 with the immunofluorescence data (**Fig 2F**). Furthermore, cells over-expressing altFUS showed an
163 altered mitochondrial network, with a significant increase in fragmented mitochondria (globular)
164 compared to mock cells that displayed more tubular structures (**Fig 2J-K**, **Fig EV4A**).

165 Mitochondrial fragmentation is observed in models of overexpression of FUS mutants[33–35], and
166 we reproduced here a similar effect when over-expressing altFUS alone. Thus, we wondered
167 whether altFUS played a role in other mitochondrial dysfunctions observed in *FUS*-linked toxicity
168 models. To this end, we reproduced an ALS-associated FUS mutant: FUS-R495x[34,35]. This mutant
169 leads to a premature stop codon before FUS NLS and is linked to severe fALS and sALS cases[1]. In
170 this construct, altFUS is still present and not affected by the mutation (**Fig EV4B**). Similar to $FUS^{(Ø)}$,
171 we also generated the monocistronic construct $FUS^{(Ø)}$-R495x, which contains synonymous
172 mutations for FUS-R495x, but prevents altFUS expression. V5-$FUS^{(Ø\text{-}FLAG)}$ and V5-$FUS^{(Ø\text{-}FLAG)}$-R495x

173 did not express altFUS, but only the FUS protein, wild-type or ALS-linked mutant R495x
174 respectively (**Fig EV4C**). We first investigated the effect of altFUS on the mitochondrial membrane
175 potential using the potential sensitive dye TMRE (**Fig 3A-B, Fig EV4D**). As previously described[34],
176 over-expression of bicistronic FUS or FUS-R495x led to a decrease in mitochondrial membrane
177 potential. The mitochondrial membrane potential remained normal when over-expressing
178 monocistronic FUS$^{(\emptyset)}$ or FUS$^{(\emptyset)}$-R495x, underlining the role of altFUS. However, over-expression of
179 altFUS alone did not alter the mitochondrial membrane potential, which suggests both proteins
180 cooperate for this FUS-associated toxicity hallmark.

181 To further characterize altFUS, we investigated its protein interactors. Using stimulated emission
182 depletion microscopy (STED), we observed that altFUS localized in puncta following a cristae-like
183 pattern inside the mitochondria, delimited using an outer-membrane mitochondrial marker,
184 TOMM20 (**Fig 3C-D**). We then used size-exclusion chromatography on mitochondrial extracts to
185 isolate altFUS-FLAG macromolecular complexes (**Fig EV5A-B**). Following a FLAG affinity purification
186 and mass spectrometry (AP-MS) analysis, using a no-bait control for quantitative comparison, we
187 confidently identified 12 interacting proteins (**Fig 3E**, **Fig EV5C-D**, **Appendix Table 5**). These
188 proteins were identified with a minimum of 2 unique peptides and displayed over a 2-fold
189 enrichment to the control. A gene enrichment analysis to the Human proteome on subcellular
190 localization showed that the identified interactors are in majority proteins known to localize at
191 the mitochondria (**Fig 3F**). Amongst the identified interactors is the heat shock protein HSPA9, a
192 chaperone crucial in the mitochondrial iron-sulfur cluster biogenesis[36] (**Fig EV5D, Appendix Table**
193 **5**). Although HSPA9 was identified with 6 unique peptides and a fold-change of 6.53 to the control,
194 this protein is commonly detected in AP-MS datasets and thus likely non specific nor functionally
195 relevant[37]. Several altFUS interacting proteins are known to interact together, such as PHB,
196 ATAD3A, ERLIN2, RANBP2 and EMD (**Fig EV5D**). A refined gene enrichment analysis to Human

197 mitochondrial proteome identified three significantly enriched biological processes: autophagy-
198 related pathways, mitochondrial metabolism and cellular response to stress (**Fig 3G**). Disruptions
199 within these pathways are pathological hallmarks of *FUS*-linked toxicity and ALS[10,11,35,38].

200 **ALTFUS INHIBITS AUTOPHAGY AND DRIVES THE ACCUMULATION OF FUS- AND TDP43-POSITIVE CYTOPLASMIC**
201 **AGGREGATES**

202 Following on these results, we hypothesized that the inhibition of autophagy observed with ALS-
203 associated FUS mutants may instead be attributed to altFUS. We used the mCherry-GFP-LC3
204 reporter to track the autophagic flux by confocal microscopy (**Fig EV6A**). Under basal conditions,
205 cells displayed red and yellow foci as expected (**Fig 4A**). An accumulation of yellow foci was
206 observed when cells were treated with bafilomycin, an inhibitor of autophagy. Similarly, cells
207 over-expressing altFUS displayed a significant accumulation of yellow foci (**Fig 4A**). Furthermore,
208 our results were consistent with previously published data[39] as cells transfected with FUS or FUS-
209 R495x displayed a decreased autophagic flux (**Fig 4A, Fig EV6B**). This accumulation of yellow foci
210 was absent in cells that express monocistronic FUS constructs, thus lacking altFUS expression
211 ($FUS^{(\emptyset)}$ or $FUS^{(\emptyset)}$-R495x). We used bafilomycin followed by LC3 probing to further validate the
212 impact of altFUS on autophagy (**Fig 4B, Fig EV6C**). Similarly, an inhibition of autophagy was
213 observed only in cells over-expressing altFUS. Furthermore, in cells over-expressing monocistronic
214 $FUS^{(\emptyset)}$-R495x, the inhibition of autophagy could be restored by co-transfecting altFUS (**Fig 4B, Fig**
215 **EV6C**). These results establish altFUS, rather than FUS, as the protein responsible of the inhibition
216 of autophagy.

217 Furthermore, altFUS interactome analysis suggested a role in the cellular stress response, which
218 is known to be altered in ALS with a TDP-43 cytoplasmic accumulation in 98 % of patients[40,41]. In
219 *FUS*-linked ALS and some sALS cases, FUS cytoplasmic aggregates or mislocalization are also

220 observed[42–44]. In our hands, cells over-expressing FUS-R495x displayed cytoplasmic aggregates
221 that were positive for both FUS-R495x and TDP-43 (**Fig 4C**). Although TDP-43 aggregates are not
222 a common observation with FUS-associated mutants, co-aggregation has already been reported
223 in patients, animal models and cultured cell lines across multiple studies[42,44–49]. In cells over-
224 expressing the monocistronic FUS$^{(\emptyset)}$-R495x construct, thus lacking altFUS expression, FUS-R495x
225 displayed a more diffuse cytoplasmic localization, and TDP-43 remained in the nucleus (**Fig 4C**).
226 FUS cytoplasmic aggregates were significantly more numerous and larger when altFUS was co-
227 expressed (**Fig 4D-E**). Accumulation of FUS-R495x and TDP-43 in cytoplasmic aggregates could be
228 reconstituted by co-transfecting altFUS and the monocistronic FUS$^{(\emptyset)}$-R495x construct (**Fig 4C**).
229 These observations were repeated across all 7 ALS-associated FUS mutations tested (**Fig EV6D-F**).
230 The cytoplasmic aggregates were also TIAR-positive as observed with ALS-linked FUS mutants in
231 previous work[40] (**Fig EV7**). These results suggest that altFUS enhances the assembly of cytoplasmic
232 FUS mutants aggregates and is responsible for the recruitment of TDP-43 in these aggregates.

233 **ALTFUS PROTECTS AGAINST NEURODEGENERATION IN FUS-ASSOCIATED DROSOPHILA MODELS**

234 In order to investigate the role of altFUS in an already established *in vivo* model of FUS-related
235 neurodegeneration, we generated *Drosophila* models expressing either the bicistronic, FUS and
236 FUS-R495x constructs, or the monocistronic, FUS$^{(\emptyset)}$ and FUS$^{(\emptyset)}$-R495x, constructs. We used the
237 Elav-GeneSwitch-GAL4 Driver strain, as previously described[50], as it allows for an inducible over-
238 expression in motor neurons and avoids lethality at the larval stage from FUS over-expression in
239 the central nervous system[50,51]. First, we generated flies containing the sequences for UASt-
240 altFUS, UASt-FUS, UASt-FUS$^{(\emptyset)}$, UASt-FUS-R495x or UASt-FUS$^{(\emptyset)}$-R495x. These flies were then
241 crossed with the Elav-GeneSwitch-GAL4 driver strain (**Fig 5A**). UASt-mCherry flies were used as
242 controls. Selected F1 individuals were then divided into 2 groups with equal proportions of
243 males/females. The first group received standard food, while the other received RU-486 treated

244 food. The treatment induces a conformational change in the Elav-GeneSwitch driver, which allows
245 activation of the UAS promoter and thus expression of the target protein. We retrieved flies at
246 selected time points to validate protein expression in the RU-486 treated population through
247 time, while the controls showed no expression (**Fig 5B**).

248 The motor neuron degeneration linked to ALS provokes a progressive locomotion loss measurable
249 with a well-described climbing assay[52]. The control populations did not show any significant
250 locomotion loss at day 1, 10 nor 20 (**Fig 5C-E**). Similarly to the RU-486 treated control group did
251 not show a significant effect, although a decrease in climbing ability could be observed as
252 previously reported[53,54] (mCherry transgenic flies – **Fig 5C**). AltFUS flies did not show any
253 significant locomotion loss through time (**Fig 5C**). This result is consistent with the *in cellulo* data
254 showing altFUS alone is not sufficient to provoke pathological hallmarks. As previously shown with
255 this model[8], the bicistronic FUS flies, which express both FUS and altFUS proteins, displayed a
256 significant locomotion loss (**Fig 5D**). Bicistronic ALS-linked FUS-R495x flies showed an even greater
257 motor neuron degeneration through time compared to FUS (**Fig 5E**). Monocistronic FUS$^{(\emptyset)}$ (**Fig 5D**)
258 and FUS$^{(\emptyset)}$-R495x (**Fig 5E**) flies, which do not express altFUS, displayed both a delay in the onset
259 of motor neuron degeneration and a reduced drop in climbing success at 20 days post-induction
260 (40% vs. 60% for FUS$^{(\emptyset)}$, 30% vs. 70% for FUS$^{(\emptyset)}$-R495x). These results in *Drosophila* confirm a role
261 for altFUS in *FUS*-related neurodegeneration *in vivo*, are consistent with our *in cellulo*
262 observations and highlight the toxic cooperation between FUS and altFUS.

263 **ALS-ASSOCIATED MUTATIONS, SYNONYMOUS FOR FUS, ALTER ALTFUS AND LEAD TO TDP-43 CYTOPLASMIC**
264 **AGGREGATES**

265 As of today, over 50 mutations in the *FUS* gene have been associated with ALS[1]. However, most
266 of these locate at the carboxyl end of the protein and as such have no effect on altFUS (**Fig 1A**).

267 We wondered whether mutations altering altFUS might have been overlooked as non-
268 consequential in the FUS reading frame. We retrieved FUS synonymous mutations found in ALS
269 patients, with an allelic frequency below 0.01 %, from previous studies and the ALS Variant Server
270 (http://als.umassmed.edu/ - **Appendix Table 6**). The retrieved mutations clustered on the altFUS
271 locus (**Fig 6A**), with 60 % of FUS synonymous mutations found in sALS patients and 50 % of FUS
272 synonymous mutations found in fALS patients, which is significantly higher than expected by
273 chance (34 %) (**Appendix Table 6**). We selected 4 mutations for further analysis based on the
274 residue conservation: altFUS-P31L, altFUS-A38V, altFUS-A46V and altFUS-R64P. We generated
275 them in GFP-FUS constructs: GFP-FUS$^{(P31L\text{-}FLAG)}$-S44=; GFP-FUS$^{(A38V\text{-}FLAG)}$-G51=; GFP-FUS$^{(A46V\text{-}FLAG)}$-G59=
276 and GFP-FUS$^{(R64P\text{-}FLAG)}$-S77=. All altFUS mutants still localized to the mitochondria (**Fig EV8A**). To
277 investigate whether these mutations may provoke an ALS-like phenotype, we quantified the
278 number of cells presenting TDP-43 aggregates. All 4 altFUS mutants displayed clear TDP-43
279 aggregates, and showed a 1.8 to 2.4 fold increase compared to wild-type altFUS (**Fig 6B-C, Fig**
280 **EV8B**). This result indicates that altFUS mutations potentiate TDP-43 cytoplasmic aggregation, a
281 pathological hallmark in 98 % of ALS cases. Hence, some *FUS* mutations, synonymous for the FUS
282 protein, exert a deleterious effect through their missense consequence on the altFUS protein.

283 **CONCLUSIONS: FUS AND ALTFUS, ENCODED BY THE SAME GENE, COOPERATE IN FUS-RELATED TOXICITY**

284 Despite considerable advances in the field, current genome annotations still uphold arbitrary
285 assumptions, such as the monocistronic nature of eukaryotic genes[16]. Here, we demonstrate *FUS*
286 is a bicistronic gene. We discovered *FUS* CDS contains a second protein-coding sequence in a
287 shifted frame overlapping its prion-like intrinsically disordered domain, regions known to host
288 dual-coding events[22,23]. This novel protein, named altFUS, is not an isoform but an entirely new
289 sequence of 170 amino acids. AltFUS is endogenously expressed in human tissues and cultured
290 cell lines, as demonstrated by ribosome profiling, mass spectrometry and with a custom antibody.

291 AltFUS is notably expressed in the motor cortex and iPSCs-derived motor neurons of healthy
292 controls and ALS patients. Because altFUS is embedded within the *FUS* CDS, this discovery is of
293 crucial importance to the field. Indeed, over-expression studies on *FUS* actually implicate two
294 proteins: FUS and altFUS. Similarly FUS knockdown or knockout studies actually inhibit expression
295 of both proteins[55]. Moreover, previous work has shown that gene editing techniques targeting a
296 specific CDS do not necessarily result in knockout of the gene in case of dual-coding gene[56]. Our
297 discovery thus suggests that *FUS* edited cells or models, notably targeting its last exons[57–59], might
298 only impair FUS protein expression but not altFUS, thus not resulting in a true *FUS* knockout. Our
299 work provides a more accurate view of *FUS* coding potential to better understand its physiological
300 function and models of FUS-related neurodegeneration.

301 Following the discovery of altFUS, we developed tools in order to differentiate the specific roles
302 and phenotypes of FUS and altFUS. Our study demonstrates that altFUS is necessary for three
303 toxic molecular hallmarks previously attributed to FUS: mitochondrial fragmentation and loss of
304 mitochondrial membrane potential, inhibition of autophagy, and cytoplasmic aggregation of FUS.
305 The inhibition of autophagy was observed when over-expressing altFUS alone, absent when over-
306 expressing FUS alone (wild-type or ALS-associated R495x mutant) and reconstituted when co-
307 expressing FUS and altFUS. This demonstrates that the inhibition of autophagy, previously
308 described in *FUS*-ALS[10,39], has been incorrectly associated to the FUS protein. AltFUS inhibits
309 autophagy. Moreover, altFUS is necessary but not sufficient for the mitochondrial membrane
310 potential loss and cytoplasmic aggregation of FUS and TDP-43. Although TDP-43 aggregates are
311 not commonly seen with FUS mutations[44], it has already been reported[42,44–49] and this study lends
312 support for altFUS conspiring with FUS and TDP-43 to lead to this molecular hallmark. Our data
313 suggests the stoichiometry between FUS and altFUS may be important for the development of
314 cytoplasmic aggregates. Both proteins are required to observe the phenotype, highlighting a

315 functional alliance between FUS and altFUS. This pathological synergy was also observed in the
316 *Drosophila* model.

317 The physiological function of altFUS is still unclear, although our work provides evidence for its
318 role in mitochondrial dynamics and the cellular response to stress. One mechanism put forward
319 in ALS is that the disease originates from a sub-optimal resolution of cellular stresses, which can
320 come from environmental sources or mutated proteins[38,60]. We have shown that altFUS, not FUS,
321 inhibits autophagy, most likely via its interaction partners (**Appendix Table 5**). Since an inhibition
322 of autophagy inhibits the dissociation of stress granules[38], we suggest altFUS potentiates stress
323 granule accumulation under stress conditions, and that FUS phase separation properties then lead
324 to the formation of solid and toxic aggregates[61]. Further work is needed to fully understand the
325 role of altFUS in physiological and pathological conditions, yet our study shows this novel protein
326 plays a crucial role in *FUS*-linked gain-of-toxic dysfunctions in models of *FUS*-related
327 neurodegeneration.

328 Recent studies have addressed the toxicity resulting from over-expression of wild-type FUS.
329 Bogaert and colleagues used FUS domain truncation mutants to investigate wild-type FUS
330 toxicity[51]. A FUS mutant lacking its N-terminal intrinsically disordered domain, thus lacking altFUS,
331 displayed reduced toxicity. This study corroborates our findings that the absence of altFUS
332 reduces the toxicity. Furthermore, Bogaert and colleagues concluded that FUS N-terminal
333 synergizes with the C-terminal domain to mediate toxicity in *Drosophila*[51]. Here, we showed that
334 altFUS synergizes with FUS to mediate ALS-like toxic features in cultured cells and toxicity in
335 *Drosophila*. Despite different laboratories and different techniques, our data are in agreement
336 with theirs. Yet, we point to an alternative (not necessarily mutually exclusive) explanation
337 whereby altFUS, not the FUS N-terminal domain, synergizes for toxicity. Additionally, this

338 discussion shows that not being aware of overlapping CDSs, especially in deletion studies,
339 precludes alternative interpretations.

340 Current genome annotations guide the interpretation of data and the design of studies, however
341 they also affect the way we screen for pathological mutations[16]. Most of the ALS-linked *FUS*
342 mutations affect the carboxyl end of the protein, as shown with the example used in this study,
343 the R495x mutation[62]. These mutations do not alter the altFUS protein, which is embedded at the
344 beginning of the FUS CDS and span on exons 3 to 6. How could altFUS be important for the disease
345 if most pathological mutations do not alter it? To answer this question, one needs to grasp how
346 much genome annotations shape today's research. For example, when screening for pathological
347 mutations, those that are synonymous for FUS are discarded early in analyses as insignificant[63].
348 Yet, a synonymous mutation for FUS may not be for altFUS. Our work shows that FUS synonymous
349 mutations found in patients cluster on altFUS genomic locus. We tested 4 of these mutations,
350 synonymous for the FUS protein and missense for the altFUS protein, and we found that each of
351 them potentiated TDP-43 cytoplasmic aggregation, a pathological hallmark of ALS[41].

352 Overall, we have shown that *FUS* is a bicistronic gene and our results indicate that altFUS
353 interferes with mitochondrial homeostasis and autophagy in cell culture and induces motor
354 neuron toxicity in *Drosophila* models. It will be important to further characterize the function of
355 altFUS and determine if it has a role in ALS and/or FTLD.

356

357 **METHODS**

358 **FUS constructs**

359 FUS and altFUS sequences were obtained from Bio Basic Gene Synthesis service. All FUS constructs
360 were subcloned into pcDNA3.1- (Invitrogen) using Gibson assembly (New England Biolabs,
361 E26115). FUS and altFUS wild-type sequences correspond to that of the Human FUS canonical
362 transcript (*ENST00000254108* or *NM_004960*). FUS and altFUS proteins were tagged with V5
363 (GKPIPNPLLGLDST) and 2 FLAG (DYKDDDDKDYKDDDDK) respectively. FUS was tagged on the N-
364 terminal, altFUS was tagged on the C-terminal. For immunofluorescence assays, N-terminal GFP-
365 tagged FUS was also cloned into pcDNA3.1- by Gibson assembly. The necessary gBlocks were
366 purchased from IDT. The monocistronic constructs $FUS^{(\emptyset)}$ and $FUS^{(\emptyset)}$-R495x were generated by
367 mutating all altFUS methionines (ATG) to threonines (ACG). These mutations are synonymous in
368 the FUS CDS (TAT > TAC, both coding for tyrosine). The altFUS mutated sequence was obtained
369 from Bio Basic Gene Synthesis service, and then subcloned in FUS sequences in pcDNA3.1- using
370 Gibson assembly. The bicistronic constructs are named as follows throughout the article: FUS,
371 FUS-R495x, or $FUS^{(FLAG)}$ and $FUS^{(FLAG)}$-R495x when altFUS is FLAG-tagged in the +2 reading frame. The
372 monocistronic constructs are named as follows throughout the article: $FUS^{(\emptyset)}$ or $FUS^{(\emptyset)}$-R495x to
373 indicate altFUS absence.

374 **Cell culture, transfections, western blots and immunofluorescence**

375 HEK293 and HeLa cells cultures tested negative for mycoplasma contamination (ATCC 30–1012K).
376 Transfections, immunofluorescence, confocal analyses and western blots were carried out as
377 previously described[64]. For FUS knock-down, 150 000 HEK293 cells in a 6-well plate were
378 transfected with 25 nM FUS SMARTpool: siGENOME siRNA (Dharmacon, Canada, L-009497-00-
379 0005) or ON-TARGET plus Nontargeting pool siRNAs (Dharmacon, D-001810-10-05) with
380 DharmaFECT one transfection reagent (Dharmacon, T-2001–02) according to the manufacturer's

381 protocol. Cell media was changed every 24 hrs and cells were processed 72 hrs after transfection.
382 For immunofluorescence, primary antibodies were diluted as follows: anti-Flag (Sigma, F1804) 1/
383 1000, anti-TOMM20 (Abcam, ab186734) 1/500, anti-V5 (Cell Signalling Technologies, #13202)
384 1/1000, anti-TDP-43 (ProteinTech, 10782-2-AP) 1/500, and anti-TIAR (Cell Signalling Technologies,
385 #8611) 1/1600. For western blots, primary antibodies were diluted as follows: anti-Flag (Sigma,
386 F1804) 1/8000, anti-V5 (Sigma, V8012) 1/8000, anti-actin (Sigma, A5441) 1/40000, anti-FUS
387 (Abcam, ab84078) 1/500, anti-altFUS (Abcam, custom antibody) 1/3000, anti-LC3 (Cell Signalling
388 Technologies, #2775) 1/1000, anti-Hsp70 (Thermo Fisher Scientific, MA3-028) 1/1000, anti-
389 Tubulin (Thermo Fisher Scientific, a11126) 1/2000 and anti-VDAC (Abcam, ab15895) 1/10000. The
390 altFUS antibody was generated by injection two rabbits, each with 2 unique altFUS peptide. The
391 purified antibody from rabbit 2 was used in this study at a 1/2000 dilution. Mitochondrial
392 morphology was evaluated using the microP tool[65]. A minimum of 100 cells per replicate were
393 counted across 3 independent experiments (n = 3, i.e. minimum 300 cells for each experimental
394 condition). Colocalization analyses were performed using the JACoP plugin (Just Another
395 Colocalization Plugin) implemented in Image J software, as previously described[13]. When
396 specified, images obtained by confocal microscopy on the Leica TCS SP8 STED 3X were
397 deconvolved using the Huygens software (Scientific Volume Imaging B.V., Hilversum,
398 Netherlands). The software uses a signal reassignment algorithm for deconvolution, identical
399 deconvolution parameters were applied to all images. The default parameters were used,
400 including the Classic Maximum Likelihood Estimation (CMLE) algorithm, signal to noise ratio, and
401 background estimation radius. The maximum iteration number was set at 30. Human tissue
402 lysates for altFUS endogenous expression were purchased from Zyagen Laboratories (San Diego,
403 California, USA).

404 **RIBO-seq data and conservation analyses**

405 Global aggregate reads for initiating ribosomes and elongating ribosomes footprints across all
406 available studies were downloaded from the Gwips portal (https://gwips.ucc.ie/), for *Homo*
407 *sapiens* and for *Mus musculus*. For altFUS protein conservation analysis, all FUS mRNAs with at
408 least EST evidence were retrieved across all available species from NCBI RefSeq. We performed
409 an *in silico* 3-frame translation and retrieved the best matching protein sequence per species that
410 displayed a minimum of 20 % sequence identity with the Human altFUS sequence over 25 % of
411 Human altFUS length. AltFUS homologous sequences were found in 83 species, and we manually
412 added that of *Drosophila melanogaster* that displayed a 37.5 % sequence identity over 19 % of
413 the Human altFUS length. All retrieved altFUS sequences were then aligned using Clustalω with
414 default parameters.

415 **Peptide-centric analysis of proteomics datasets**

416 The stand alone PepQuery tool (v1.0) was downloaded from the PepQuery website
417 (http://www.pepquery.org/). The tool was run on the following datasets from the TCGA
418 consortium: colon cancer (COCA) proteome, ovarian cancer (OVCA) proteome, phosphoproteome
419 and glycoproteome, and the breast cancer (BRCA) proteome and phosphoproteome. The
420 reference database was set to the Ensembl datababase (hg38_Ensembl_20190910). The following
421 parameters were set for all runs unless specified: carbamidomethylation of cysteine as fixed
422 modification (as well as iTRAQ 4-plex of K, iTRAQ 4-plex of peptide N-term for BRCA and OVCA);
423 oxidation of methionine as variable modification (as well as iTRAQ 4-plex of Y for BRCA and OVCA);
424 a maximum of 3 modifications per peptides; trypsin digestion with maximum of 1 miscleavage;
425 precursor tolerance of 10 ppm (20 ppm for COCA); fragment mass tolerance of 0.05 Da (0.6 Da
426 for COCA); the hyperscore was used as a scoring metric, and 10 000 randoms. For
427 phosphoproteomes, phosphorylation of Y, T and S were added as variable modifications. For the
428 glycoproteome, deamidation of Q and N were added as variable modifications. PepQuery was run

429 in the protein mode, with altFUS (IP_243680) whole sequence as input. Spectra were visualised
430 and drawn using an in-house python script.

431 **Human induced pluripotent stem cell differentiation into motor neurons**

432 Directed differentiation to human iPSC-motor neurons was performed as previously reported[27].
433 Briefly, iPSCs were maintained on Geltrex (Life Technologies) with Essential 8 Medium media (Life
434 Technologies), and passaged using EDTA (Life Technologies, 0.5mM). All cell cultures were
435 maintained at 37°C and 5% carbon dioxide. For motor neuron differentiation, iPSCs were
436 differentiated to neuroepithelium by plating to 100% confluency in chemically defined medium
437 consisting of DMEM/F12 Glutamax, Neurobasal, L-Glutamine, N2 supplement, non-essential
438 amino acids, B27 supplement, β-mercaptoethanol (Life Technologies) and insulin (Sigma).
439 Treatment with the following small molecules from day 0-7: 1µM Dorsomorphin (Millipore), 2µM
440 SB431542 (Tocris Bioscience), and 3.3µM CHIR99021 (Miltenyi Biotec). At day 8, cells patterned
441 for 7 days with 0.5µM retinoic acid and 1µM Purmorphamine. At day 14 spinal cord motor neuron
442 precursors were treated with 0.1µM Purmorphamine for a further 4 days before being terminally
443 differentiated for >10 days in 0.1 µM Compound E (Enzo Life Sciences) to promote cell cycle exit.
444 Throughout the neural conversion and patterning phase (D0-18) the neuroepithelial layer was
445 enzymatically dissociated twice (at D4-5 and D10-12) using dispase (GIBCO, 1 mg ml-1).

446 **Preparation of tissue lysates of the motor cortex of ALS patients**

447 Approximately 100mg of motor cortex from 4 sporadic ALS and 4 C9orf72-ALS cases was lysed in
448 10x RIPA (50mM Tris HCl pH7.8, 150mM NaCl, 0.5% sodium deoxycholate, 1% NP40;
449 supplemented with protease inhibitors and EDTA) volume using TissueLyzer equipment (Qiagen).
450 Lysates were incubated on ice 20 minutes followed by centrifugation at 20,000xg for 20 minutes
451 at 4°C. Supernatant was taken as 'RIPA fraction' and pellets resuspended in RIPA and SDS (final

452 concentration of 2%). 3 sporadic ALS and 3 C9orf72-ALS samples were subsequently used as they
453 were sufficiently concentrated to load 100 ug of proteins onto SDS-page gels.

454 **Mitochondrial extracts and cellular fractionation**

455 Mitochondrial extracts were prepared as previously described[56]. Briefly, HEK293 cells grown up
456 to 80 % confluence, were rinsed twice with PBS and gathered using a cell scraper. Cells were
457 pelleted by centrifugation at 500 x g for 10 mins at 4 °C. Supernatant was discarded and cells
458 suspended in mitochondrial buffer (210 mM mannitol, 70 mM sucrose, 1 mM EDTA, 10 mM
459 HEPES-NaOH, pH 7.5, 0.5 mM PMSF and EDTA-free protease inhibitor (Thermo Fisher Scientific)).
460 Cells were disrupted by 15 consecutive passages through a 25G1 0.5 x 25 needle syringe on ice,
461 followed by a 3 min centrifugation at 2 000 x g at 4 °C. Supernatant was collected and the pellet
462 suspended in mitochondrial buffer. The cell disruption was repeated four times and all retrieved
463 supernatants containing mitochondria were again passed through syringe needle in mitochondrial
464 buffer and cleared by centrifugation for 3 mins at 2000 × g at 4 °C. Supernatants were pooled and
465 centrifuged for 10 mins at 13 000 x g at 4 °C to pellet mitochondria. The pellet was suspended in
466 200 µL of mitochondrial buffer until further processing. Cellular fractionation was performed
467 using the Cell Fractionation Kit (#9038S, Cell Signaling Technology). Briefly, HEK293 cells were
468 grown up to 80 % confluence, washed twice with PBS and gathered using a cell scraper. Cells were
469 spun at 350 x g for 5 mins at 4 °C and 2.5 x 10^6 cells were suspended in 500 µL of ice-cold PBS. An
470 aliquot of 100 µL was spun at 350 x g for 5 mins at 4 °C and resuspended in SDS buffer (4 % SDS,
471 Tris-HCl 100 mM pH 7.6) and kept as WCL (Whole Cell Lysate). The rest of the collected cells
472 (remaining 400 µL) were spun at 500 x g for 5 mins at 4 °C. Supernatant was discarded and pellet
473 resuspended in 500 µL of CIB (Cytoplasmic Isolation Buffer) from the kit, vortexed for 5 secs and
474 incubated on ice for 5 mins. After centrifugation at 500 x g for 5 mins at 4 °C, the supernatant was
475 collected as the cytosplasmic fraction. The pellet was resuspended in 500 µL of MIB buffer

476 (Membrane Isolation Buffer) from the kit, vortexed for 15 secs and incubated on ice for 5 mins.
477 After centrifugation at 8 000 x g for 5 mins at 4 °C, the supernatant was collected as the membrane
478 and organelles fraction. To each 100 µL of fraction was added 60 µL of loading buffer 1 x (from
479 ColdSpring Laemmli sample buffer: 50 mM Tris pH 6.8, 2 % SDS, 10 % Glycerol, 5 % β-
480 mercaptoethanol) before processing for western blot.

481 **Mitochondrial membrane potential measurements**

482 Mitochondrial membrane potential was measured by flow cytometry in HEK293 cells using TMRE
483 (Tetramethylrhodamine, Ethyl ester, Abcam, ab113852). FCCP was used as a positive control to
484 validate each independent experiment. Cells were grown up to 80 % confluence and washed twice
485 with PBS. The cells were then incubated for 5 mins at 37 °C, 5 % CO_2 with PBS/A (0.2 % BSA in PBS)
486 solution (experimental) or 3 µM FCCP in PBS/A solution (positive control). Then, 100 nM of TMRE
487 was added and cells were incubated 15 mins at 37 °C, 5 % CO_2. After incubation, cells were
488 trypsinized and centrifuged at 800 x g for 5 mins at 4 °C and resuspended in 500 µL of PBS and
489 kept on ice. Cells were immediately analysed by flow cytometry. A gate for living cells was set, as
490 well as a second gate to filter out cell doublets. TMRE fluorescence (PE-A) was recorded over a
491 minimum of 50 000 gated cells for each experimental condition. The mean TMRE fluorescence
492 intensity was measured over 3 independent experiments for each experimental condition.

493 **Stimulated Emission Depletion (STED) microscopy**

494 Samples were prepared as described above for confocal microscopy. A Leica TCS SP8 STED 3X was
495 used with a 100x objective lens and immersion oil for dual-color STED images. Images were
496 obtained by sequential scanning of a given area. The combination of Alexa Fluor 488 (Thermo
497 Fisher Scientific, A-11017) and Alexa Fluor 568 (Thermo Fisher Scientific, A-21069) dyes was
498 chosen for STED imaging. Alexa Fluor 488 dye was excited with a white light laser (WLL) at 488 nm
499 and was depleted using the 660 nm STED laser. Alexa Fluor 568 dye was excited with a WLL at

500 561 nm and was depleted using the 660 nm STED laser. The STED laser (660 nm) was applied at
501 80 % of maximum power.

502 **Fast Protein Liquid Chromatography (FPLC) and affinity-purification mass spectrometry (AP-MS)**

503 Mitochondrial extracts of HEK293 cells were centrifuged at 13 000 x g for 10 mins at 4 °C to
504 remove the supernatant and were resuspended in FPLC buffer (50 mM Tris-HCl, 1 mM EDTA, 150
505 mM NaCl, 1 % Triton X-100, pH 7.5, filtered with 0.2 µm filters) at 2 mg/mL for a total of 4 mg of
506 mitochondrial proteins. Samples were incubated on ice for 15 mins and then centrifuged at 10 000
507 x g for 5 mins at 4 °C and the supernatant was loaded in the injection syringe without disrupting
508 the pellet. The FPLC was performed on a HiLoad 16/60 Superdex 200 pg column (GE Healthcare,
509 Chicago, USA) at 4 °C. The column was pre-equilibrated with the FPLC buffer for up to 0.2 CV
510 (column volume) and the sample was applied at a flow rate of 0.5 mL/min with a pressure alarm
511 set at 0.5 MPa. The elution was performed over 72 fractions of 1.5 mL for a maximum of 1.1 CV.
512 For altFUS probing by western blot, proteins were precipitated from 150 µL of each 4 fractions in
513 technical duplicates. First, 600 µL of methanol was added to each tube and mixed gently, before
514 adding 150 µL of chloroform. Tubes were gently inverted 10 times before adding 450 µL of milliQ
515 H_2O and vortexing briefly. After centrifugation at 12 000 x g for 3 mins, the upper phase was
516 discarded, and 400 µL of methanol was added. Tubes are centrifuged at 16 000 x g for 4 mins and
517 the pellet was resuspended in loading buffer. For interactome analysis by mass spectrometry,
518 fractions of interest (8 to 14) were pooled together and incubated at 4 °C overnight with magnetic
519 FLAG beads (Sigma, M8823) pre-conditioned with FPLC buffer. The beads were then washed 3
520 times with 5 mL of FPLC buffer, and 5 times with 5 mL of 20 mM NH_4HCO_3 (ABC). Proteins were
521 eluted and reduced from the beads using 10 mM DTT (15mins at 55 °C), and then treated with 20
522 mM IAA (1 hour at room temperature in the dark). Proteins were digested overnight by adding 1
523 µg of trypsin (Promega, Madison, Wisconsin) in 100 µL ABC at 37 °C overnight. Digestion was

524 quenched using 1 % formic acid and supernatant was collected. Beads were washed once with
525 acetonitrile/water/formic acid (1/1/0.01 v/v) and pooled with supernatant. Peptides were dried
526 with a speedvac, desalted using a C18 Zip-Tip (Millipore Sigma, Etobicoke, Ontario, Canada) and
527 resuspended into 30 µl of 1% formic acid in water prior to MS analysis.

528 **Mass-spectrometry analysis**

529 Peptides were separated in a PepMap C18 nano column (75 µm × 50 cm, Thermo Fisher Scientific).
530 The setup used a 0–35% gradient (0–215 min) of 90% acetonitrile, 0.1% formic acid at a flow rate
531 of 200 nL/min followed by acetonitrile wash and column re-equilibration for a total gradient
532 duration of 4 h with a RSLC Ultimate 3000 (Thermo Fisher Scientific, Dionex). Peptides were
533 sprayed using an EASYSpray source (Thermo Fisher Scientific) at 2 kV coupled to a quadrupole-
534 Orbitrap (QExactive, Thermo Fisher Scientific) mass spectrometer. Full-MS spectra within a m/z
535 350–1600 mass range at 70,000 resolution were acquired with an automatic gain control (AGC)
536 target of 1e6 and a maximum accumulation time (maximum IT) of 20 ms. Fragmentation (MS/MS)
537 of the top ten ions detected in the Full-MS scan at 17,500 resolution, AGC target of 5e5, a
538 maximum IT of 60 ms with a fixed first mass of 50 within a 3 m/z isolation window at a normalized
539 collision energy (NCE) of 25. Dynamic exclusion was set to 40 s. Mass spectrometry RAW files were
540 searched with Andromeda search engine implemented in MaxQuant 1.5.5.1. The digestion mode
541 was set at Trypsin/P with a maximum of two missed cleavages per peptides. Oxidation of
542 methionine and acetylation of N-terminal were set as variable modifications, and
543 carbamidomethylation of cysteine was set as fixed modification. Precursor and fragment
544 tolerances were set at 4.5 and 20 ppm respectively. Files were searched using a target-decoy
545 approach against UniprotKB (*Homo sapiens* 03/2017 release) with the addition of altFUS
546 sequence for a total of 92,949 entries. The false discovery rate (FDR) was set at 1% for peptide-
547 spectrum-match, peptide and protein levels. Protein interactions were then scored using the

548 SAINT algorithm, with Mock cells as control and the magnetic FLAG beads in HEK293 cells
549 crapome[66]. Proteins with a SAINT score above 0.99 were considered, as well as those presenting
550 a SAINT score above 0.88 with a minimum of two unique peptides.

551 **Biological processes and cellular compartment enrichment analysis**

552 Proteins identified in altFUS interactome were screened for cellular compartment and biological
553 processes enrichment using Gene Ontology (GO) enrichment. Proteins were queried against the
554 whole Human Proteome for cellular compartment and against the Human mitochondrial
555 proteome (MitoCarta 2.0) for biological processes. The statistical analysis used a Fisher's Exact
556 test with a FDR set at 1 %.

557 **Autophagic flux measurements**

558 The mCherry-GFP-LC3 was used to evaluate the autophagic vesicles within HeLa cells by confocal
559 microscopy. Before fusion with the lysosome, the LC3 molecules on the autophagosome display
560 a yellow fluorescence (combined mCherry and GFP fluorescence). After fusion, the GFP
561 fluorescence is quenched by the lysosomal pH, and as such the LC3 molecules display a red signal
562 (mCherry alone). This allows a visual representation of the autophagic flux in a given cell. Cells
563 treated with 50 nM Bafilomycin for 4 hours were used as a positive control to validate each
564 independent experiment. Observations were made across 2 technical duplicates for each
565 biological condition, across 3 independent experiments (n=3). Alternatively, the autophagic flux
566 was also evaluated by LC3 probing before and after bafilomycin treatment (50 nM for 4 hours).
567 The quantification corresponds to the treated / untreated ratio of LC3-II abundance.

568 **Cytoplasmic aggregates measurements**

569 Images of HeLa cells were taken by confocal microscopy and then processed using the Image J 3D
570 Objects Counter plugin. FUS cytoplasmic aggregates were then quantified in number and size

571 (μm^2) for each cell. A total of 100 cells across two technical replicates were taken for each
572 independent experiment (n=3, i.e. a minimum of 300 cells per biological conditions).

573 **Transgenic *Drosophila* and climbing assay**

574 The bicistronic constructs, FUS and FUS-R495x, and the monocistronic constructs, altFUS, $FUS^{(\emptyset)}$
575 and $FUS^{(\emptyset)}$-R495x, were subcloned in the pUASTattB expression vector for site specific insertion
576 into attP2 on chromosome 3. Transgenic flies were generated by Best Gene (Best Gene Inc.,
577 California, USA). The Elav-GeneSwitch-GAL4 driver (stock number: 43642, genotype: y[1] w[*];
578 P{w[+mC]=elav-Switch.O}GSG301) and the UAS-mCherry flies (stock number: 35787, genotype:
579 y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8]=UAS-mCherry.VALIUM10}attP2) was purchased from
580 Bloomington (Bloomington Drosophila Stock Center, Indiana, USA). All stocks were in a w^{1118}
581 background and were cultured on standard medium at 25°C or room temperature. Transgenic
582 flies were crossed with the Elav-GeneSwitch-GAL4 driver strain. The F1 was equally divided in two
583 groups with equal proportion of males and females: one group will feed on standard food
584 supplemented with ethanol (0.2 % - control flies), the other on standard food supplemented with
585 RU-486 at 10 µM diluted in ethanol (induced flies). The climbing assay was performed as
586 previously described[52]. Briefly, flies were transferred into an empty vial and tapped to the bottom.
587 After 18 s, the number of flies at the top of the tube were considered successful. The assay was
588 done at day 1, 10 and 20 post-induction, across 4 independent F1. Five flies were taken at day 1,
589 10 and 20 post-induction to validate expression of the proteins of interest.

590 **Statistical analyses and representation**

591 Unless otherwise stated, the statistical analysis carried was a two-way ANOVA with Tukey's
592 multiple comparison correction. The box plots represent the mean with the 5 to 95 % percentile.
593 The bar graphs represent the mean, and error bars correspond to the standard deviation. When

using parametric tests, normality of data distribution was verified beforehand using the Shapiro-Wilk test.

DATA AVAILABILITY STATEMENT

The OpenProt database is available at www.openprot.org. The GTEx portal is available via www.gtexportal.org. The Gwips portal is available at www.gwips.ucc.ie. The proteomics data are available on the PRIDE repository with the accession PXD------. Any other relevant data are available from the corresponding authors upon reasonable request.

ACKNOWLEDGEMENTS

This research was supported by CIHR grants MOP-137056 and MOP-136962, by an ALS Canada Project Grant, and by a Canada Research Chair in Functional Proteomics and Discovery of Novel Proteins to X.R. X.R and S.J are members of the Fonds de Recherche du Québec Santé (FRQS)-supported Centre de Recherche du Centre Hospitalier Universitaire de Sherbrooke. P.M. is supported by the ALS Double Play Christopher Chiu Postdoctoral Fellowship. We thank J. Ule, M-J. Boucher and D. Hunting for helpful discussions. The autopsy program is supported by the James Hunter and Family ALS Initiative. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on April 2019.

CONTRIBUTIONS

MA.B and X.R designed and wrote the study. MA.B and JF.J did the experiments, and MA.B did the analyses and figures. S.N and S.J assisted with the Drosophila experiments. P.M, L.Z and J.R

616 provided the motor cortex tissue lysates. GE.T and R.P provided the iPSCs-derived motor neurons.

617 All authors proofread the manuscript.

618 **REFERENCES**

619 1. Deng, H., Gao, K. & Jankovic, J. The role of FUS gene variants in neurodegenerative diseases.
620 *Nat. Rev. Neurol.* **10**, 337–348 (2014).
621 2. Vance, C. *et al.* Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic
622 Lateral Sclerosis Type 6. *Science* **323**, 1208–1211 (2009).
623 3. Zou, Z.-Y. *et al.* Screening of the FUS gene in familial and sporadic amyotrophic lateral
624 sclerosis patients of Chinese origin. *Eur. J. Neurol.* **19**, 977–983 (2012).
625 4. Dini Modigliani, S., Morlando, M., Errichelli, L., Sabatelli, M. & Bozzoni, I. An ALS-associated
626 mutation in the FUS 3'-UTR disrupts a microRNA–FUS regulatory circuitry. *Nat. Commun.* **5**,
627 4335 (2014).
628 5. Sabatelli, M. *et al.* Mutations in the 3' untranslated region of FUS causing FUS overexpression
629 are associated with amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **22**, 4748–4755 (2013).
630 6. Ajmone-Cat, M. A. *et al.* Increased FUS levels in astrocytes leads to astrocyte and microglia
631 activation and neuronal death. *Sci. Rep.* **9**, 4572 (2019).
632 7. Ju, S. *et al.* A Yeast Model of FUS/TLS-Dependent Cytotoxicity. *PLoS Biol.* **9**, (2011).
633 8. Chen, Y. *et al.* Expression of human FUS protein in Drosophila leads to progressive
634 neurodegeneration. *Protein Cell* **2**, 477–486 (2011).
635 9. Miguel, L. *et al.* Accumulation of insoluble forms of FUS protein correlates with toxicity in
636 Drosophila. *Neurobiol. Aging* **33**, 1008.e1–15 (2012).
637 10. Ling, S.-C. *et al.* Overriding FUS autoregulation in mice triggers gain-of-toxic dysfunctions
638 in RNA metabolism and autophagy-lysosome axis. *eLife* **8**, e40811 (2019).
639 11. Taylor, J. P., Brown Jr, R. H. & Cleveland, D. W. Decoding ALS: from genes to mechanism.
640 *Nature* **539**, 197–206 (2016).

641 12. Nolan, M., Talbot, K. & Ansorge, O. Pathogenesis of FUS-associated ALS and FTD: insights
642 from rodent models. *Acta Neuropathol. Commun.* **4**, (2016).
643 13. Samandi, S. *et al.* Deep transcriptome annotation enables the discovery and functional
644 characterization of cryptic small proteins. *eLife* **6**, e27860 (2017).
645 14. Saghatelian, A. & Couso, J. P. Discovery and characterization of smORF-encoded
646 bioactive polypeptides. *Nat. Chem. Biol.* **11**, 909–916 (2015).
647 15. Delcourt, V., Staskevicius, A., Salzet, M., Fournier, I. & Roucou, X. Small Proteins Encoded
648 by Unannotated ORFs are Rising Stars of the Proteome, Confirming Shortcomings in Genome
649 Annotations and Current Vision of an mRNA. *Proteomics* (2017)
650 doi:10.1002/pmic.201700058.
651 16. Brunet, M. A., Levesque, S. A., Hunting, D. J., Cohen, A. A. & Roucou, X. Recognition of
652 the polycistronic nature of human genes is critical to understanding the genotype-phenotype
653 relationship. *Genome Res.* (2018) doi:10.1101/gr.230938.117.
654 17. Brunet, M. A. *et al.* OpenProt: a more comprehensive guide to explore eukaryotic coding
655 potential and proteomes. *Nucleic Acids Res.* **47**, D403–D410 (2019).
656 18. Umoh, M. E. *et al.* A proteomic network approach across the ALS-FTD disease spectrum
657 resolves clinical phenotypes and genetic vulnerability in human brain. *EMBO Mol. Med.*
658 e201708202 (2017) doi:10.15252/emmm.201708202.
659 19. Collins, M. A., An, J., Hood, B. L., Conrads, T. P. & Bowser, R. P. Label-Free LC–MS/MS
660 Proteomic Analysis of Cerebrospinal Fluid Identifies Protein/Pathway Alterations and
661 Candidate Biomarkers for Amyotrophic Lateral Sclerosis. *J. Proteome Res.* **14**, 4486–4501
662 (2015).

663 20. Chick, J. M. *et al.* A mass-tolerant database search identifies a large proportion of
664 unassigned spectra in shotgun proteomics as modified peptides. *Nat. Biotechnol.* **33**, 743–
665 749 (2015).
666 21. Olexiouk, V., Van Criekinge, W. & Menschaert, G. An update on sORFs.org: a repository
667 of small ORFs identified by ribosome profiling. *Nucleic Acids Res.* **46**, D497–D502 (2018).
668 22. Kovacs, E., Tompa, P., Liliom, K. & Kalmar, L. Dual coding in alternative reading frames
669 correlates with intrinsic protein disorder. *Proc. Natl. Acad. Sci.* **107**, 5429–5434 (2010).
670 23. Pancsa, R. & Tompa, P. Coding Regions of Intrinsic Disorder Accommodate Parallel
671 Functions. *Trends Biochem. Sci.* **41**, 898–906 (2016).
672 24. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
673 25. Pavesi, A. Asymmetric evolution in viral overlapping genes is a source of selective
674 protein adaptation. *Virology* **532**, 39–47 (2019).
675 26. Michel, A. M., Kiniry, S. J., O'Connor, P. B. F., Mullan, J. P. & Baranov, P. V. GWIPS-viz:
676 2018 update. *Nucleic Acids Res.* doi:10.1093/nar/gkx790.
677 27. Hall, C. E. *et al.* Progressive Motor Neuron Pathology and the Role of Astrocytes in a
678 Human Stem Cell Model of VCP-Related ALS. *Cell Rep.* **19**, 1739–1749 (2017).
679 28. Luisier, R. *et al.* Intron retention and nuclear loss of SFPQ are molecular hallmarks of
680 ALS. *Nat. Commun.* **9**, 1–15 (2018).
681 29. Hart, G. W. & Akimoto, Y. The O-GlcNAc Modification. in *Essentials of Glycobiology* (eds.
682 Varki, A. et al.) (Cold Spring Harbor Laboratory Press, 2009).
683 30. Thompson, J. W., Sorum, A. W. & Hsieh-Wilson, L. C. Deciphering the Functions of O-
684 GlcNAc Glycosylation in the Brain: The Role of Site-Specific Quantitative O-GlcNAcomics.
685 *Biochemistry* **57**, 4010–4018 (2018).

686 31. Didonna, A. & Benetti, F. Post-translational modifications in neurodegeneration.
687 *Biophys. 2016 Vol 3 Pages 27-49* (2015) doi:10.3934/biophy.2016.1.27.
688 32. Santos, A. L. & Lindner, A. B. Protein Posttranslational Modifications: Roles in Aging and
689 Age-Related Disease. *Oxid. Med. Cell. Longev.* **2017**, (2017).
690 33. Carrì, M. T., D'Ambrosi, N. & Cozzolino, M. Pathways to mitochondrial dysfunction in ALS
691 pathogenesis. *Biochem. Biophys. Res. Commun.* **483**, 1187–1193 (2017).
692 34. Deng, J. *et al.* FUS Interacts with HSP60 to Promote Mitochondrial Damage. *PLOS Genet.*
693 **11**, e1005357 (2015).
694 35. Nakaya, T. & Maragkakis, M. Amyotrophic Lateral Sclerosis associated FUS mutation
695 shortens mitochondria and induces neurotoxicity. *Sci. Rep.* **8**, 15575 (2018).
696 36. Shan, Y. & Cortopassi, G. Mitochondrial Hspa9/Mortalin regulates erythroid
697 differentiation via iron-sulfur cluster assembly. *Mitochondrion* **26**, 94–103 (2016).
698 37. Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity purification-
699 mass spectrometry data. *Nat. Methods* **10**, 730–736 (2013).
700 38. Monahan, Z., Shewmaker, F. & Pandey, U. B. Stress granules at the intersection of
701 autophagy and ALS. *Brain Res.* **1649**, 189–200 (2016).
702 39. Marrone, L. *et al.* FUS pathology in ALS is linked to alterations in multiple ALS-associated
703 proteins and rescued by drugs stimulating autophagy. *Acta Neuropathol. (Berl.)* (2019)
704 doi:10.1007/s00401-019-01998-x.
705 40. Aulas, A. & Vande Velde, C. Alterations in stress granule dynamics driven by TDP-43 and
706 FUS: a link to pathological inclusions in ALS? *Front. Cell. Neurosci.* **9**, (2015).
707 41. Hergesheimer, R. C. *et al.* The debated toxic role of aggregated TDP-43 in amyotrophic
708 lateral sclerosis: a resolution in sight? *Brain* **142**, 1176–1194 (2019).

709 42. Deng, H.-X. *et al.* FUS-immunoreactive inclusions are a common feature in sporadic and
710 non-SOD1 familial amyotrophic lateral sclerosis. *Ann. Neurol.* **67**, 739–748 (2010).
711 43. Tyzack, G. E. *et al.* Widespread FUS mislocalization is a molecular hallmark of
712 amyotrophic lateral sclerosis. *Brain* doi:10.1093/brain/awz217.
713 44. Farrawell, N. E. *et al.* Distinct partitioning of ALS associated TDP-43, FUS and SOD1
714 mutants into cellular inclusions. *Sci. Rep.* **5**, 13416 (2015).
715 45. Wiesner, D. *et al.* Reversible induction of TDP-43 granules in cortical neurons after
716 traumatic injury. *Exp. Neurol.* **299**, 15–25 (2018).
717 46. Cohen, N. R., Hammans, S. R., Macpherson, J. & Nicoll, J. A. R. New neuropathological
718 findings in Unverricht–Lundborg disease: neuronal intranuclear and cytoplasmic inclusions.
719 *Acta Neuropathol. (Berl.)* **121**, 421–427 (2011).
720 47. Shan, X., Chiang, P.-M., Price, D. L. & Wong, P. C. Altered distributions of Gemini of
721 coiled bodies and mitochondria in motor neurons of TDP-43 transgenic mice. *Proc. Natl.*
722 *Acad. Sci. U. S. A.* **107**, 16325–16330 (2010).
723 48. Keller, B. A. *et al.* Co-aggregation of RNA binding proteins in ALS spinal motor neurons:
724 evidence of a common pathogenic mechanism. *Acta Neuropathol. (Berl.)* **124**, 733–747
725 (2012).
726 49. Kryndushkin, D., Wickner, R. B. & Shewmaker, F. FUS/TLS forms cytoplasmic aggregates,
727 inhibits cell growth and interacts with TDP-43 in a yeast model of amyotrophic lateral
728 sclerosis. *Protein Cell* **2**, 223–236 (2011).
729 50. Lanson, N. A. *et al.* A Drosophila model of FUS-related neurodegeneration reveals
730 genetic interaction between FUS and TDP-43. *Hum. Mol. Genet.* **20**, 2510–2523 (2011).
731 51. Bogaert, E. *et al.* Molecular Dissection of FUS Points at Synergistic Effect of Low-
732 Complexity Domains in Toxicity. *Cell Rep.* **24**, 529-537.e4 (2018).

733 52. Chambers, R. P. *et al.* Nicotine increases lifespan and rescues olfactory and motor
734 deficits in a Drosophila model of Parkinson's disease. *Behav. Brain Res.* **253**, 95–102 (2013).
735 53. Mawhinney, R. M. S. & Staveley, B. E. Expression of GFP can influence aging and climbing
736 ability in Drosophila. *Genet. Mol. Res. GMR* **10**, 494–505 (2011).
737 54. Robles-Murguia, M., Hunt, L. C., Finkelstein, D., Fan, Y. & Demontis, F. Tissue-specific
738 alteration of gene expression and function by RU486 and the GeneSwitch system. *Npj Aging*
739 *Mech. Dis.* **5**, 1–5 (2019).
740 55. Scekic-Zahirovic, J. *et al.* Toxic gain of function from mutant FUS protein is crucial to
741 trigger cell autonomous motor neuron loss. *EMBO J.* **35**, 1077–1097 (2016).
742 56. Delcourt, V. *et al.* The Protein Coded by a Short Open Reading Frame, Not by the
743 Annotated Coding Sequence, Is the Main Gene Product of the Dual-Coding Gene MIEF1. *Mol.*
744 *Cell. Proteomics MCP* **17**, 2402–2411 (2018).
745 57. Kino, Y. *et al.* FUS/TLS deficiency causes behavioral and pathological abnormalities
746 distinct from amyotrophic lateral sclerosis. *Acta Neuropathol. Commun.* **3**, (2015).
747 58. Hicks, G. G. *et al.* Fus deficiency in mice results in defective B-lymphocyte development
748 and activation, high levels of chromosomal instability and perinatal death. *Nat. Genet.* **24**,
749 175–179 (2000).
750 59. An, H. *et al.* ALS-linked FUS mutations confer loss and gain of function in the nucleus by
751 promoting excessive formation of dysfunctional paraspeckles. *Acta Neuropathol. Commun.* **7**,
752 7 (2019).
753 60. Al-Chalabi, A. & Hardiman, O. The epidemiology of ALS: a conspiracy of genes,
754 environment and time. *Nat. Rev. Neurol.* **9**, nrneurol.2013.203 (2013).
755 61. Qamar, S. *et al.* FUS Phase Separation Is Modulated by a Molecular Chaperone and
756 Methylation of Arginine Cation-π Interactions. *Cell* **173**, 720-734.e15 (2018).

757 62. Naumann, M. *et al.* Impaired DNA damage response signaling by FUS-NLS mutations
758 leads to neurodegeneration and FUS aggregate formation. *Nat. Commun.* **9**, 335 (2018).
759 63. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a
760 joint consensus recommendation of the American College of Medical Genetics and Genomics
761 and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **17**,
762 405–424 (2015).
763 64. Vanderperre, B. *et al.* An overlapping reading frame in the PRNP gene encodes a novel
764 polypeptide distinct from the prion protein. *FASEB J.* **25**, 2373–2386 (2011).
765 65. Peng, J.-Y. *et al.* Automatic Morphological Subtyping Reveals New Roles of Caspases in
766 Mitochondrial Dynamics. *PLOS Comput. Biol.* **7**, e1002212 (2011).
767 66. Choi, H. *et al.* SAINT: Probabilistic Scoring of Affinity Purification - Mass Spectrometry
768 Data. *Nat. Methods* **8**, 70–73 (2011).
769

770 **LEGENDS**

771

772 ***Main Figures***

773

774 **Figure 1: *FUS* is a bicistronic gene**

775 **A**, FUS gene bicistronic annotation, with the canonical FUS CDS (in blue, +1 frame) and altFUS CDS
776 (in green, +2 frame) represented on the FUS canonical transcript (*ENST00000254108* or
777 *NM_004960*). Sequence length proportions are respected, the scale bar corresponds to 300
778 nucleotides. **B**, Genome browser view of FUS gene. The five most abundant transcripts in the brain
779 are shown in the 'Transcript' track. Transcripts predicted coding by the OpenProt resource are
780 coloured in blue, and in grey if predicted non-coding. The 'Protein' track contains all predicted
781 protein products. The known FUS protein (ENSP00000254108) and its isoform
782 (ENSP00000369594) are coloured in green. The novel predicted altFUS protein (IP_243680) and
783 its isoform (IP_243691) are coloured in red. **C**, PhyloP nucleotidic conservation scores are
784 represented in grey across the *FUS* mRNA (ENST00000254108). The noise reduction after FFT (Fast
785 Fourrier Transformation) is outlined in blue. The average PhyloP score on the bicistronic and the
786 monocistronic region are represented as dotted red lines. The position of the FUS CDS is
787 represented by a blue rectangle, and that of altFUS CDS by a green rectangle. **D**, Alignment
788 (Clustalω) of altFUS protein sequences in Human (*Homo sapiens*), Chimpanzee (Chimp. - *Pan*
789 *troglodytes*), Rat (*Rattus norvegicus*), Mouse (*Mus musculus*) and Dog (*Canis lupus familiaris*).
790 Residues are coloured based on their identity across species, from white (not conserved) to red
791 (conserved in all species). **E**, RIBO-seq data over the *FUS* gene from the Gwips portal. Initiating
792 ribosome reads are indicated by blue bars, and elongating ribosomes footprints are indicated by
793 the blue curve. The graph capture the beginning of the FUS gene with FUS and altFUS methionines
794 indicated by blue arrows. The genomic positions are indicated relative to the start of exon 1. **F-G**,
795 Genome browser view of *FUS* gene, centered on altFUS. The 'Transcript' track contains the
796 beginning of the canonical FUS transcript (*ENST00000254108*) in blue. The 'Protein' track contains
797 the beginning of the FUS protein (green) and the whole altFUS protein (red). In **F**, the 'Peptide'
798 track contains all the peptides identified by the OpenProt resource using the classical spectrum-
799 centric approach. The peptides sequences are indicated and are unique to altFUS or its isoform
800 (see **Fig EV2B** for an example spectra). In **G**, the 'Peptide' track contains all the peptides identified
801 by our peptide-centric approach. The peptides indicated matched better to at least one spectra
802 than any known protein, and are coloured in yellow if they matched better than any known
803 protein with any PTM (see **Fig EV2C** for an example spectra). The peptides sequences are indicated
804 and are unique to altFUS or its isoform.

805

806 **Figure 2: altFUS, a novel endogenous mitochondrial protein**

807 **A**, Expression of both FUS and altFUS proteins from transfection of the *FUS* cDNA in HEK293 cells
808 by western blot, and expression of FUS with the monocistronic construct $FUS^{(\emptyset)}$ (representative
809 image from n=3). The western blot corresponds to a low exposition so that the endogenous FUS

810 and altFUS are non visible. **B**, AltFUS endogenous expression in HEK293 cells using a siRNA
811 targeting *FUS* mRNA as negative control and over-expression of altFUS CDS as positive control
812 (representative image from n=3). **C**, AltFUS (arrow) endogenous expression in Human tissues
813 (brain, muscles and kidney – 100 ug), in HEK293 and HeLa cultured cells (100 ug) and using the
814 over-expression of altFUS CDS in HEK293 cells (50 ug) as positive control (representative image
815 from n=3). The asterisk indicates a protein species detected with the anti-altFUS antibody
816 specifically in the brain. **D-E**, AltFUS (arrow) endogenous expression in the motor cortex of three
817 *C9orf72* and three sporadic ALS patients (**D**) or in iPSC-derived motor neurons of three lines from
818 controls and from ALS patients (**E**) (representative image from n=3). The asterisk indicates a
819 protein species detected with the anti-altFUS antibody specifically in the brain. **F**, Images by
820 confocal microscopy of altFUS-FLAG (green) in HeLa cells, using TOMM20 (red) as a mitochondrial
821 marker (representative image from n=3, Pearson's correlation r = 0.92). The white scale bar
822 corresponds to 10 µm. **G**, AltFUS-FLAG enrichment in mitochondrial extracts from transfected
823 HEK293 cells (representative image from n=3) with Hsp70 used as a mitochondrial marker (WCL
824 = Whole Cell Lysate, Mito = Mitochondrial extract). **H**, AltFUS mitochondrial expression in
825 transfected HEK293 cells following fractionation (representative image from n=3), with Tubulin as
826 a cytosolic fraction marker and VDAC as a mitochondrial fraction marker (WCL = Whole Cell Lysate,
827 Cyto = Cytosol fraction, Mito = mitochondrial fraction). **I**, Endogenous altFUS mitochondrial
828 expression in HEK293 cells following fractionation (representative image from n=3), with Tubulin
829 as a cytosolic fraction marker and VDAC as a mitochondrial fraction marker. We used si*FUS*
830 transfected cells as a negative control and altFUS transfected cells as a positive control for altFUS
831 expression (WCL = Whole Cell Lysate, Cyto = Cytosol fraction, Mito = mitochondrial fraction). **J**,
832 Representative images of the mitochondrial network (TOMM20 in red) in mock and altFUS-FLAG
833 (green) transfected HeLa cells (n=3). The white scale bar corresponds to 10µm. **K**, Proportion of
834 tubules and globules in the mitochondrial network of mock and altFUS-FLAG transfected HeLa cells
835 (see **Fig EV4A**). Quantification was done over a minimum of 100 cells across a technical duplicate
836 per independent experiments (n=3, i.e. a minimum of 300 cells per biological conditions, p-value
837 < 0.001, Mann-Whitney U test).

838

839 **Figure 3: altFUS is involved in mitochondrial dysfunction and autophagy processes**

840 **A**, Representative traces of TMRE fluorescence measured by flow cytometry in mock transfected
841 cells and cells overexpressing the bicistronic FUS-R495x or the monocistronic FUS$^{(\emptyset)}$-R495x
842 constructs (n=4, minimum of 50 000 live cells per independent replicates). Mean fluorescence
843 intensity of mock transfected cells treated with a decoupling agent, FCCP, is indicated by a grey
844 dotted line. **B**, Mean TMRE fluorescence intensity measures in mock transfected cells, cells over-
845 expressing altFUS, FUS, FUS$^{(\emptyset)}$, FUS-R495x or FUS$^{(\emptyset)}$-R495x, or mock transfected cells treated with
846 FCCP across 3 independent experiments (n=4, also see **Fig EV4D**). Statistical significance is relative
847 to the mock condition unless otherwise indicated (*** = p value < 0.001, **** = p value < 0.0001,
848 n.s. = non-significant). **C**, Representative image by stimulated emission depletion microscopy
849 (STED) of altFUS-FLAG (green) localization within mitochondria (TOMM20 marker in red). The white
850 bar across the mitochondria represents the region of interest quantified in panel **D**. The white
851 scale bar corresponds to 3µm. **D**, Relative fluorescence histogram for altFUS-FLAG and TOMM20
852 across the region of interest highlighted by a white line on panel **C**. **E**, Scatter plot of the proteins
853 identified by AP-MS (see **Fig EV5**) indicating their enrichment (Fold-change over control) and their
854 SAINT probability score. Proteins above the 0.8 threshold are indicated in black, others in grey.
855 AltFUS is indicated in red (bait) and preys known to regulate the autophagy or the cellular stress

856 response are indicated in green. **F**, Subcellular localizations of proteins identified by AP-MS from
857 (see **Fig EV5**). **G**, Enrichment of biological processes in altFUS interacting proteins compared to
858 the Human mitochondrial proteome (n=2, Fisher's Exact test with FDR < 0.1 %). The number of
859 proteins identified in each GO term is indicated next to the corresponding bar.

860

861 **Figure 4: AltFUS is necessary for FUS-associated inhibition of autophagy and accumulation of**
862 **FUS/TDP-43 cytoplasmic aggregates**

863 **A**, Images by confocal microscopy of mCherry-GFP-LC3 signal in HeLa cells across biological
864 conditions: untreated mock, bafilomycin treated mock, altFUS, FUS, FUS$^{(Ø)}$, FUS-R495x and FUS$^{(Ø)}$-
865 R495x (representative images of n=3). The white scale bar corresponds to 10 µm and the zoomed
866 in region (right panels) is delimited by a white box. **B**, LC3-II accumulation after bafilomycin
867 treatment from mock, altFUS, FUS, FUS$^{(Ø)}$, FUS-R495x, FUS$^{(Ø)}$-R495x transfected cells and FUS$^{(Ø)}$-
868 R495x and altFUS co-transfected cells across 3 independent experiments (n=3). The quantification
869 corresponds to the treated/untreated ratio of LC3-II abundance. Statistical significance is relative
870 to the mock condition unless otherwise indicated (**** = p value < 0.0001, *** = p value < 0.001,
871 ** = p value < 0.01, n.s. = non-significant). **C**, Images by confocal microscopy of altFUS (FLAG
872 tagged- white), FUS (GFP tagged - green) and TDP-43 (red) signals in HeLa cells transfected with
873 the bicistronic GFP-FUS$^{(FLAG)}$-R495x or the monocistronic GFP-FUS$^{(Ø\text{-}FLAG)}$-R4955x constructs, or co-
874 transfected with the monocistronic GFP-FUS$^{(Ø\text{-}FLAG)}$-R495x and altFUS-FLAG constructs
875 (representative images from n=3). The white scale bar corresponds to 10 µm. **D-E**, Quantification
876 of FUS cytoplasmic granules, number (**D**) and area (µm^2) (**E**) in cells over-expressing the bicistronic
877 (+) or monocistronic (-) construct for FUS, FUS-G156E, FUS-R495x, FUS-K510E, FUS-Q519x, FUS-
878 Q519I-fs527x, FUS-R521C, and FUS-P525L. Statistical comparisons are made between bicistronic
879 and monocistronic versions of each construct (**** = p value < 0.0001, *** = p value < 0.001, **
880 = p value < 0.01, * = p value < 0.05, n.s. = non-significant, one-way ANOVA test with Sidak's
881 multiple comparison).

882

883 **Figure 5: AltFUS expression is necessary for the full *FUS*-linked ALS phenotype in *Drosophila***

884 **A**, Cross-breeding strategy for *Drosophila* generation using the Elav-GeneSwitch-GAL driver as an
885 inducible expression system specific to the motor neurons. **B**, FUS and altFUS expression in
886 mCherry (control), altFUS, FUS, FUS$^{(Ø)}$, FUS-R495x or FUS$^{(Ø)}$-R495x expressing *Drosophila* from the
887 control F1 and the RU-486 treated F1 (see panel **A**) at 1, 10 or 20 days post-induction
888 (representative image from n=2). **C-E**, Locomotion assay using the percentage of climbing success
889 in control and RU-486-treated transgenic *Drosophila* expressing mCherry or altFUS (**C**), the
890 bicistronic FUS or the monocistronic FUS$^{(Ø)}$ (**D**), and the bicistronic FUS-R495x or the
891 monocistronic FUS$^{(Ø)}$-R495x (**E**) at day 1, 10 and 20 post-induction. Statistical comparison were
892 made between each population (n=4). Indicated significance are between the monocistronic and
893 the bicistronic transgenic flies of the RU-486-treated population (n.s. = non-significant, * = p value
894 < 0.05, *** = p value < 0.001).

895

896 **Figure 6: FUS mutations, synonymous for FUS but missense for altFUS, potentiate TDP-43**
897 **cytoplasmic aggregation**

898 **A**, Graphical representation of FUS synonymous mutations (yellow) found in ALS patients. The
899 canonical FUS mRNA is represented in dark blue (*ENST00000254108* or *NM_004960*). The FUS
900 protein coding sequence is indicated in light blue, and the altFUS protein coding sequence is
901 indicated in green. **B**, Images by confocal microscopy of TDP-43 (yellow) and altFUS (FLAG tagged
902 - red) in HeLa cells over-expressing GFP-FUS$^{(FLAG)}$ or GFP-FUS$^{(P31L\text{-}FLAG)}$-S44= (representative images
903 from n=3). The white scale bar corresponds to 10 µm. **C**, Quantification of cells with TDP-43
904 aggregates in HeLa cells (see panel **B** and **Fig EV8B**). The data are represented as the fold-change
905 compared to the GFP-FUS$^{(FLAG)}$ expressing cells. Statistical significance is indicated above the bars
906 (n=3, **** = p value < 0.0001).

907

908

909 ***<u>Extended View Figures</u>***

910

911 **Figure EV1: *FUS* gene protein diversity**

912 **A**, AltFUS is encoded within an alternative open reading frame (ORF) overlapping the FUS
913 canonical CDS. When read in the +1 frame, the *FUS* mRNA (here *ENST00000254108*) codes for a
914 526 amino acid protein (highlighted in blue), named FUS. In the +2 frame, the *FUS* mRNA contains
915 a second ORF (highlighted in green) that codes for a novel 170 amino acid protein, named altFUS.
916 The two proteins are not isoforms. **B**, GTEx portal data on *FUS* mRNA expression in the brain and
917 nerves are shown in blue colored scale (TPM = Transcripts Per Million). Transcripts are identified
918 with Ensembl accessions (the number after the dot corresponds to the version used in the
919 analysis) and are quantified across 14 tissues. The five transcripts framed in red share 85 % of all
920 mRNA expression level. **C**, Table compiling the protein information relayed by Ensembl and
921 OpenProt resources for the five transcripts highlighted in panel **B**. The FUS protein is highlighted
922 in blue. The altFUS protein is highlighted in green. D-E, Alignment (Clustalω) of protein sequences
923 for FUS (ENSP00000254108) and its isoform (ENSP00000369594) in panel **D** (blue) and altFUS
924 (IP_243680) and its isoform (IP_243691) in panel **E** (green). The residues are coloured based on
925 their degree of identity. **F**, Heatmap of altFUS protein sequences identity across 84 species (see
926 **Appendix Table 2** and **Source Data 1**). Primates, Rodents and Mammals in general display strong
927 protein conservation. The sequence identity is colored from blue (0 %) to red (100 %).

928

929 **Figure EV2: AltFUS is endogenously expressed**

930 **A**, RIBO-seq data over the mouse *FUS* gene from the Gwips portal (*Mus musculus*). Initiating
931 ribosome reads are indicated by green bars, and elongating ribosomes footprints are indicated by
932 the green curve. The graph capture the beginning of the *FUS* gene with FUS and altFUS
933 methionines indicated by green arrows. The genomic positions are indicated relative to the start
934 of exon 1. **B**, MS/MS spectra confidently mapped to a peptide unique to altFUS by OpenProt
935 resource. Peaks are represented by their mass over charge ratios (m/z) and their intensity relative
936 to the highest (relative intensity). The y ions are coloured in blue, the b ions in red and the
937 unannotated peaks appear in grey. The charge state and m/z error of the matched peptide are

938 indicated on the top right corner of the spectra. The original study, file and spectra number are
939 indicated below the graph. **C**, Comparison of one MS/MS spectra confidently mapped to a peptide
940 unique to altFUS (left spectra) and its best possible annotation with any known protein with any
941 post-translational modification (PTM - right spectra). Peaks are represented by their mass over
942 charge ratios (m/z) and their intensity relative to the highest (relative intensity). The y ions are
943 coloured in blue, the b ions in red and the unannotated peaks appear in grey. The charge state
944 and m/z error for each of the matched peptides are indicated on the top right corner of each
945 panel. The original study, file and spectra number are indicated below the graph. On the top of
946 each panel is indicated the matched spectra with its corresponding protein and PTMs.

947

948 **Figure EV3: Tools to explore altFUS expression and roles**

949 **A**, AltFUS custom antibody strategy, with epitopes highlighted in bold on altFUS sequence (green),
950 and ELISA test results on both bleeds from each rabbits. B, Nucleotidic mutations on *FUS* mRNA
951 (*ENST00000254108*) are highlighted in red and mutate all altFUS methionines to threonines
952 (altFUS_Ø). **C-D**, Protein alignment of altFUS (**C**) and FUS (**D**) proteins from FUS bicistronic
953 construct and the monocistronic ($FUS^{(Ø)}$) construct. The residues are coloured based on sequence
954 identity (dark blue for identical residues).

955

956 **Figure EV4: AltFUS is involved in mitochondrial functions**

957 **A**, MicroP processing of confocal images of mock and altFUS transfected HeLa cells. The nucleus
958 is coloured in blue, altFUS in green and the mitochondrial network in red (TOMM20). Tubules are
959 coloured in purple, globules are coloured in green. **B**, Graphical representation of bicistronic
960 constructs FUS and FUS-R495x (early stop codon highlighted in red), and their monocistronic
961 equivalent $FUS^{(Ø)}$ and $FUS^{(Ø)}$-R495x. AltFUS CDS is highlighted in green when present, FUS CDS is
962 highlighted in blue. **C**, V5-FUS, V5-FUS-R495x and altFUS-FLAG expression from bicistronic or
963 monocistronic constructs in HEK293 cells (representative image from n=3). **D**, Representative
964 traces of TMRE fluorescence measured by flow cytometry in mock cells and cells overexpressing
965 the bicistronic constructs FUS or FUS-R495x (in red) in the top panels, and the monocistronic
966 constructs, $FUS^{(Ø)}$ or $FUS^{(Ø)}$-R495x (in blue) in the bottom panels (n=3, minimum of 50 000 live
967 cells per independent replicates).

968

969 **Figure EV5: AltFUS mitochondrial interactome**

970 **A**, Workflow of mitochondrial extraction and size exclusion chromatography (FPLC) on mock or
971 altFUS-FLAG transfected HEK293 cells. **B**, AltFUS expression in FPLC fractions on transfected
972 HEK293 cells mitochondrial extracts (representative image from n=2). Every 4 fractions were
973 loaded, the size scale through the fractions is indicated in red. **C**, Workflow of magnetic beads
974 FLAG trap on FPLC fractions 8 to 14, from mock or altFUS-FLAG transfected HEK293 cells. **D**, Protein-
975 protein interaction network of altFUS (bait, in green). Identified proteins (preys) are indicated in
976 blue. The network was expanded using String database, with a maximum of 2 additional nodes

977 (grey nodes and grey dotted edges). Interactions (edges) identified by our AP-MS are in black.
978 Interactions between preys were added using the String database (blue edges).

979

980 **Figure EV6: AltFUS inhibits the autophagy and potentiate FUS/TDP-43 cytoplasmic aggregates**

981 **A**, Graphical representation of the mCherry-GFP-LC3 construct to study the autophagic flux. **B**,
982 Images by confocal microscopy of mCherry-GFP-LC3 and altFUS (FLAG, in blue) signal in HeLa cells
983 across biological conditions: untreated mock, bafilomycin treated mock, altFUS, FUS, FUS$^{(Ø\text{-}FLAG)}$,
984 FUS-R495x and FUS$^{(Ø\text{-}FLAG)}$-R495x (representative images of n=3). The white scale bar corresponds
985 to 10 µm and the zoomed in region (right panel) is delimited as a white box (related to **Fig4A**). **C**,
986 LC3-II expression from untreated (-) or bafilomycin treated (+) cells transfected with mock, altFUS,
987 FUS, FUS$^{(Ø)}$, FUS-R495x, FUS$^{(Ø)}$-R495x, or co-transfected with FUS$^{(Ø)}$-R495x and altFUS. **D-F**, Images
988 by confocal microscopy of FLAG (white), GFP (green) and TDP-43 (red) signals in HeLa cells
989 transfected with the bicistronic constructs (**D**), monocistronic constructs (**E**) or co-transfected
990 with altFUS-FLAG and the monocistronic constructs (**F**) of 6 ALS-associated mutants: FUS-G156E,
991 FUS-K510E, FUS-Q519x, FUS-Q519I-fs527x, FUS-R521C, and FUS-P525L (representative images
992 from n=3). The white scale bar corresponds to 10 µm.

993

994 **Figure EV7: AltFUS potentiate FUS recruitment to stress granules**

995 **A**, Images by confocal microscopy of FLAG (white), GFP (green) and TIA-1 (red) signals in HeLa cells
996 transfected with the bicistronic GFP-FUS$^{(FLAG)}$-R495x or the monocistronic GFP-FUS$^{(Ø\text{-}FLAG)}$-R495x
997 constructs, or co-transfected with the monocistronic GFP-FUS$^{(Ø\text{-}FLAG)}$-R495x and altFUS-FLAG
998 constructs (representative images from n=3). The white scale bar corresponds to 10 µm. **B-D**,
999 Images by confocal microscopy of FLAG (white), GFP (green) and TIA-1 (red) signals in HeLa cells
1000 transfected with the bicistronic constructs (**B**), monocistronic constructs (**C**) or co-transfected
1001 with altFUS-FLAG and the monocistronic constructs (**D**) of 6 ALS-associated mutants: FUS-G156E,
1002 FUS-K510E, FUS-Q519x, FUS-Q519I-fs527x, FUS-R521C, and FUS-P525L (representative images
1003 from n=3). The white scale bar corresponds to 10 µm.

1004

1005 **Figure EV8: Mitochondrial altFUS mutants potentiates TDP-43 cytoplasmic aggregates**

1006 **A**, Images by confocal microscopy of altFUS (FLAG tagged - green) and mitochondria (TOMM20
1007 marker, red) in HeLa cells over-expressing altFUS-FLAG, altFUS-P31L-FLAG, altFUS-A38V-FLAG,
1008 altFUS-A46V-FLAG or altFUS-R64P-FLAG constructs (representative images from n=3). The white
1009 scale bar corresponds to 10 µm. Deconvolution over a maximum of 30 iterations on the green and
1010 red channels was performed for the zoomed in pictures. The white scale bar corresponds to 3 µm.
1011 **B**, Images by confocal microscopy of TDP-43 (yellow) and altFUS (FLAG tagged - red) in HeLa cells
1012 over-expressing GFP-FUS$^{(A38V\text{-}FLAG)}$-G51=, GFP-FUS$^{(A46V\text{-}FLAG)}$-G59= or GFP-FUS$^{(R64P\text{-}FLAG)}$-S77=
1013 constructs (representative images from n=3). The white scale bar corresponds to 10 µm.

1014

1015 ***<u>Appendix Tables</u>***

1016

1017 **Appendix Table 1: *FUS* gene annotation on OpenProt**

1018 OpenProt search results for FUS transcripts *ENST00000254108* and *NM_004960.3*. The OpenProt
1019 version 1.3 was used. Each protein prediction is identified with a unique protein accession
1020 (starting with II_ for novel predicted isoform, and IP_ for predicted alternative proteins). All
1021 predicted alternative proteins are coloured in light blue. The protein localization is relative to the
1022 canonical FUS coding sequence (CDS). The furthered studied protein, named altFUS (IP_243680)
1023 is highlighted in dark blue. The workbook also contains the list of peptides identified in re-analysis
1024 of mass spectrometry experiments using the OpenProt database (version 1.3).

1025

1026 **Appendix Table 2: altFUS protein sequences across species**

1027 AltFUS conservation across vertebrates. The workbook contains the list of *FUS* transcripts
1028 retrieved from NCBI RefSeq database (transcript accession and nucleotidic sequences); the list of
1029 transcripts containing an altFUS sequence (transcript accession, species, and altFUS protein
1030 sequence); and the matrix of altFUS protein sequence identity between each species.

1031

1032 **Appendix Table 3: Peptide-Centric proteomic results for altFUS endogenous expression**

1033 Peptide-spectrum matches confidently identified using PepQuery (v1.0) on the TCGA datasets.
1034 The first sheet of the workbook contains the legends for the summary and results tables. The
1035 summary table lists the number of peptides and PSMs in each dataset. The details of results for
1036 each dataset queried is also listed.

1037

1038 **Appendix Table 4: Post-translational modification predictions on altFUS sequence**

1039 Predictions for post-translational modifications (PTMs) on altFUS sequence. The summary sheets
1040 contains the name of the modification predicted, the target residues, the tool used and the
1041 number of predicted sites with their confidence score. Results for the 2 more abundant PTMs
1042 predicted (phosphorylation and O-Glc-NAcylation) are detailed.

1043

1044 **Appendix Table 5: AltFUS interacting proteins**

1045 Proteins identified by mass spectrometry following size exclusion chromatography (FPLC) and
1046 FLAG affinity purification on mitochondrial extracts from HEK293 cells over-expressing altFUS-
1047 FLAG. The proteins are listed by their UniProtKB accession alongside their annotated function, the
1048 number of unique peptides supporting the detection and the number of peptide spectrum
1049 matches (PSMs) within each experimental condition. Confidence of the interaction was scored
1050 using the SAINT algorithm. The SAINT score is also reported. Proteins involved in autophagy
1051 related processes are indicated in red. In light greys are proteins removed from the analysis
1052 because lacking protein coverage (unique peptides < 2).

1053

1054 **Appendix Table 6: FUS synonymous mutations in ALS patients and their consequence on altFUS**

1055 *FUS* mutations, synonymous for FUS but missense for altFUS. The first sheet contains mutations
1056 identified from a manually curated literature review. The second sheet retrieves variants in sALS
1057 cases from the ALS Variant Server (http://als.umassmed.edu/). The third sheet retrieves variants
1058 in fALS cases from the ALS Variant Server.

1059

1060 ***<u>Source Data</u>***

1061

1062 **Source Data 1: altFUS protein alignment across species**

1063 Alignment of altFUS protein sequences across all 82 species. The alignment was done using
1064 Clustalω, and coloured based on conservation from white (highly variable residue) to dark blue
1065 (highly conserved residue).

1066

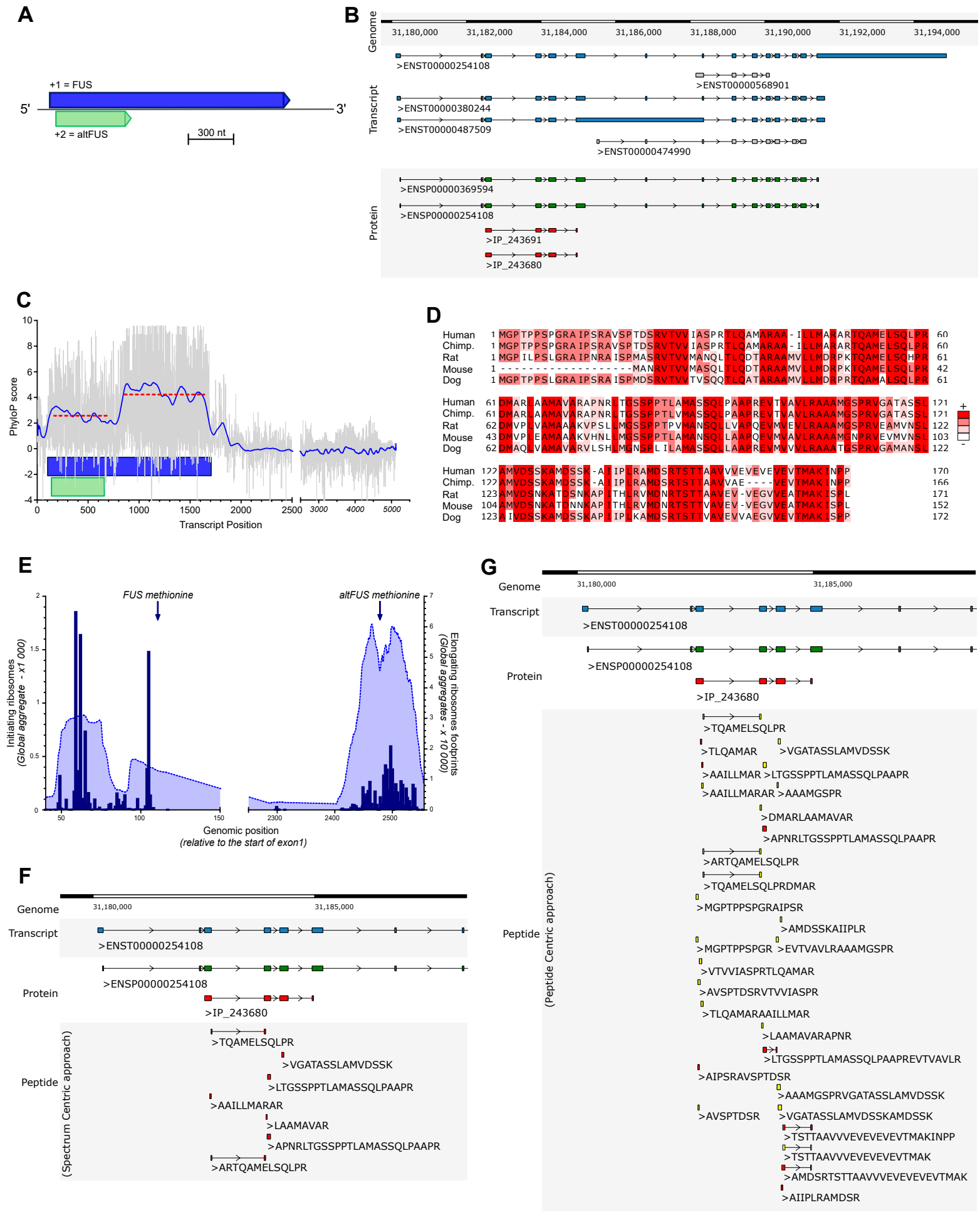

A
+1 = FUS
5'
3'
+2 = altFUS
300 nt
B
Genome
31,180,000
31,182,000
31,184,000
31,186,000
31,188,000
31,190,000
31,192,000
31,194,000
Transcript
>ENST00000254108
>ENST00000568901
>ENST00000380244
>ENST00000487509
>ENST00000474990
Protein
>ENSP00000369594
>ENSP00000254108
>IP_243691
>IP_243680
C
PhyloP score
Transcript Position
D
Human
Chimp.
Rat
Mouse
Dog
E
FUS methionine
altFUS methionine
Initiating ribosomes
(Global aggregate - x1 000)
Elongating ribosomes footprints
(Global aggregates - x 10 000)
Genomic position
(relative to the start of exon1)
F
Genome
31,180,000
31,185,000
Transcript
>ENST00000254108
Protein
>ENSP00000254108
>IP_243680
Peptide
(Spectrum Centric approach)
>TQAMELSQLPR
>VGATASSLAMVDSSK
>LTGSSPPTLAMASSQLPAAPR
>AAILLMARAR
>LAAMAVAR
>APNRLTGSSPPTLAMASSQLPAAPR
>ARTQAMELSQLPR
G
Genome
31,180,000
31,185,000
Transcript
>ENST00000254108
Protein
>ENSP00000254108
>IP_243680
Peptide
(Peptide Centric approach)
>TQAMELSQLPR
>TLQAMAR
>VGATASSLAMVDSSK
>AAILLMAR
>LTGSSPPTLAMASSQLPAAPR
>AAILLMARAR
>AAAMGSPR
>DMARLAAMAVAR
>APNRLTGSSPPTLAMASSQLPAAPR
>ARTQAMELSQLPR
>TQAMELSQLPRDMAR
>MGPTPPSPGRAIPSR
>AMDSSKAIIPLR
>MGPTPPSPGR
>EVTVAVLRAAAMGSPR
>VTVVIASPRTLQAMAR
>AVSPTDSRVTVVIASPR
>TLQAMARAAILLMAR
>LAAMAVARAPNR
>LTGSSPPTLAMASSQLPAAPREVTVAVLR
>AIPSRAVSPTDSR
>AAAMGSPRVGATASSLAMVDSSK
>AVSPTDSR
>VGATASSLAMVDSSKAMDSSK
>TSTTAAVVVEVEVEVEVTMAKINPP
>TSTTAAVVVEVEVEVEVTMAK
>AMDSRTSTTAAVVVEVEVEVEVTMAK
>AIIPLRAMDSR

Figure 1

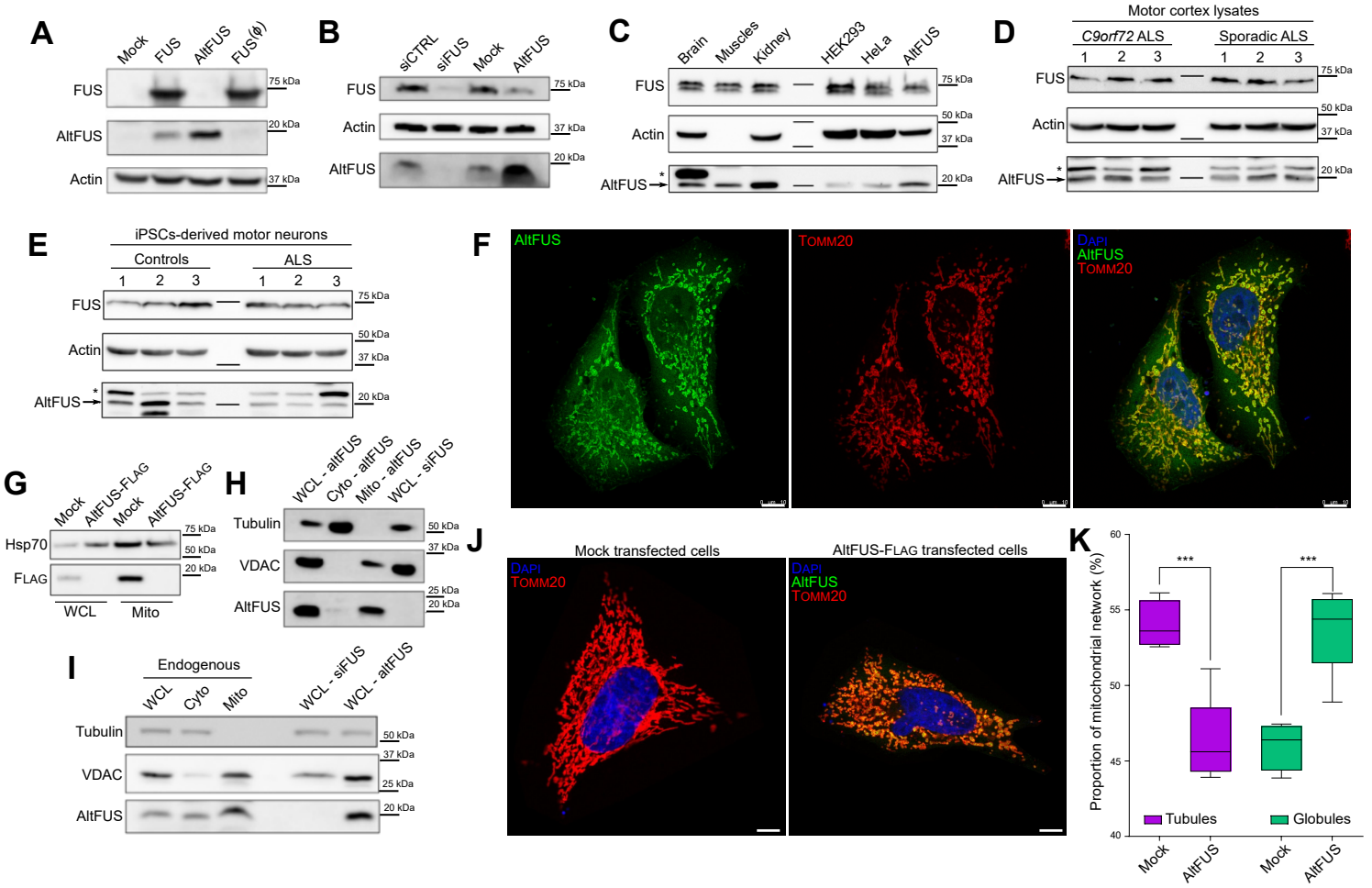

A
Mock
FUS
AltFUS
FUS(Φ)
FUS
AltFUS
Actin
75 kDa
20 kDa
37 kDa
B
siCTRL
siFUS
Mock
AltFUS
FUS
Actin
AltFUS
75 kDa
37 kDa
20 kDa
C
Brain
Muscles
Kidney
HEK293
HeLa
AltFUS
FUS
Actin
AltFUS
75 kDa
50 kDa
37 kDa
20 kDa
D
Motor cortex lysates
C9orf72 ALS
Sporadic ALS
1 2 3
1 2 3
FUS
Actin
AltFUS
75 kDa
50 kDa
37 kDa
20 kDa
E
iPSCs-derived motor neurons
Controls
ALS
1 2 3
1 2 3
FUS
Actin
AltFUS
75 kDa
50 kDa
37 kDa
20 kDa
F
AltFUS
TOMM20
DAPI
AltFUS
TOMM20
G
Mock
AltFUS-FLAG
Mock
AltFUS-FLAG
Hsp70
FLAG
WCL
Mito
75 kDa
50 kDa
20 kDa
H
WCL - altFUS
Cyto - altFUS
Mito - altFUS
WCL - siFUS
Tubulin
VDAC
AltFUS
50 kDa
37 kDa
25 kDa
20 kDa
I
Endogenous
WCL
Cyto
Mito
WCL - siFUS
WCL - altFUS
Tubulin
VDAC
AltFUS
50 kDa
37 kDa
25 kDa
20 kDa
J
Mock transfected cells
AltFUS-FLAG transfected cells
DAPI
TOMM20
DAPI
AltFUS
TOMM20
K
Proportion of mitochondrial network (%)
60
55
50
45
40
***
***
Tubules
Globules
Mock
AltFUS
Mock
AltFUS

Figure 2

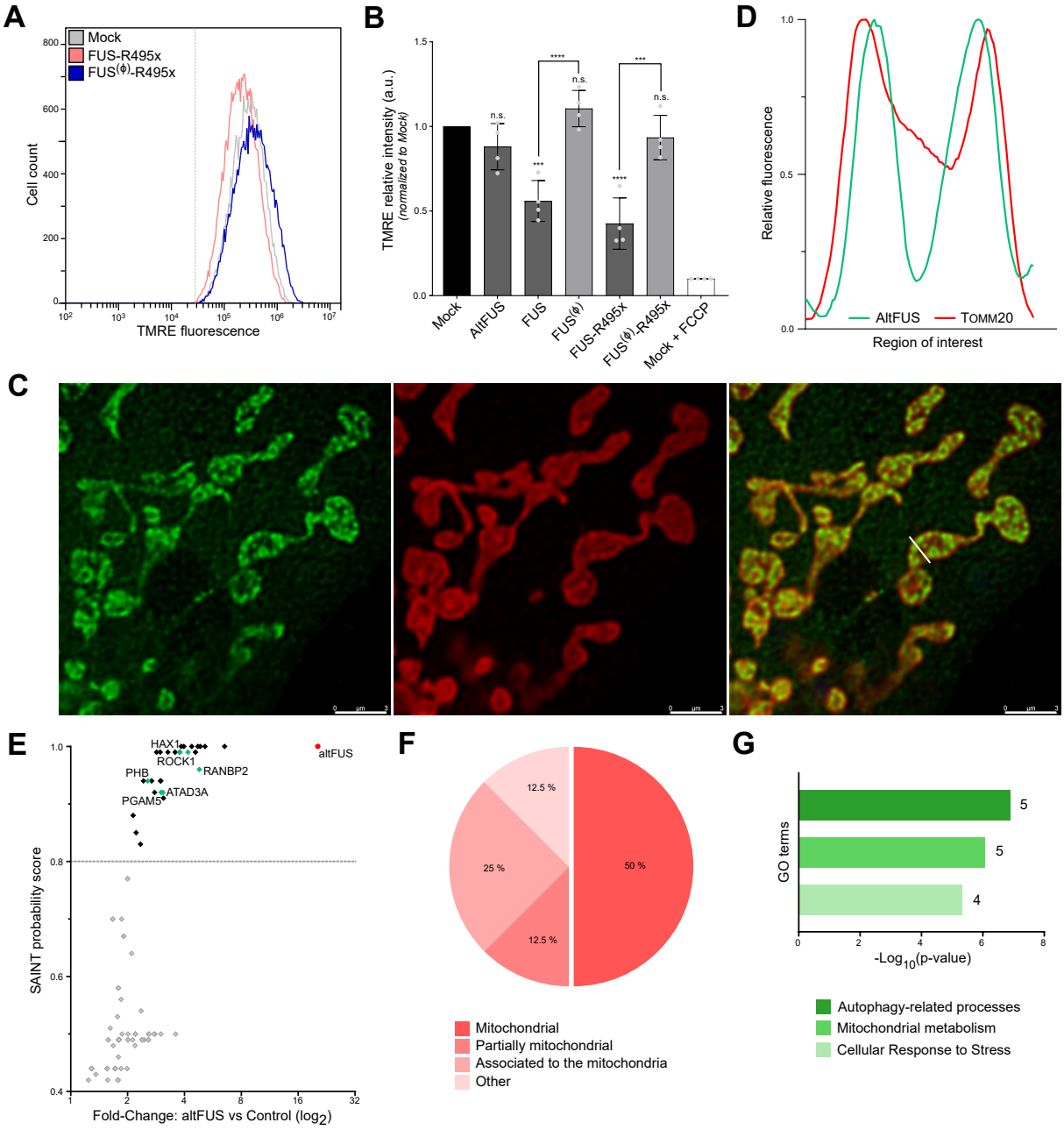

A
Mock
FUS-R495x
FUS(ϕ)-R495x
Cell count
TMRE fluorescence
B
TMRE relative intensity (a.u.)
(normalized to Mock)
n.s.
****
***
Mock
AltFUS
FUS
FUS(ϕ)
FUS-R495x
FUS(ϕ)-R495x
Mock + FCCP
D
Relative fluorescence
AltFUS
TOMM20
Region of interest
C
E
HAX1
ROCK1
PHB
RANBP2
ATAD3A
PGAM5
altFUS
SAINT probability score
Fold-Change: altFUS vs Control (log2)
F
12.5 %
25 %
50 %
12.5 %
Mitochondrial
Partially mitochondrial
Associated to the mitochondria
Other
G
GO terms
5
5
4
-Log10(p-value)
Autophagy-related processes
Mitochondrial metabolism
Cellular Response to Stress

Figure 3

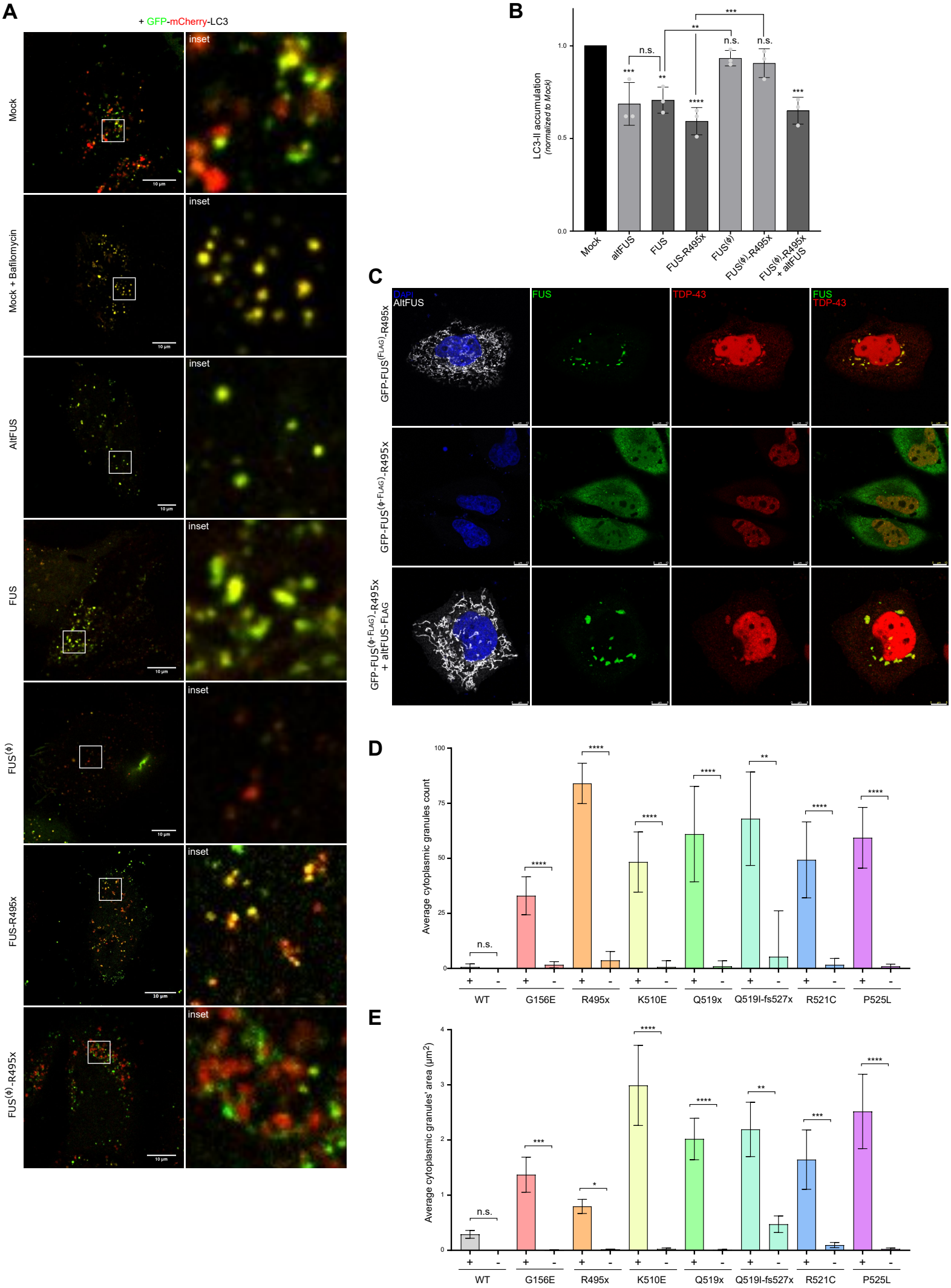

A
+ GFP-mCherry-LC3
Mock
Mock + Bafilomycin
AltFUS
FUS
FUS(ϕ)
FUS-R495x
FUS(ϕ)-R495x
inset
10 µm
B
LC3-II accumulation
(normalized to Mock)
1.0
0.5
0.0
n.s.
***
**
****
Mock
altFUS
FUS
FUS-R495x
FUS(ϕ)
FUS(ϕ)-R495x
FUS(ϕ)-R495x + altFUS
C
DAPI
AltFUS
FUS
TDP-43
GFP-FUS(FLAG)-R495x
GFP-FUS(ϕ-FLAG)-R495x
GFP-FUS(ϕ-FLAG)-R495x + altFUS-FLAG
D
Average cytoplasmic granules count
100
75
50
25
0
WT
G156E
R495x
K510E
Q519x
Q519I-fs527x
R521C
P525L
E
Average cytoplasmic granules' area (µm²)
4
3
2
1
0

Figure 4

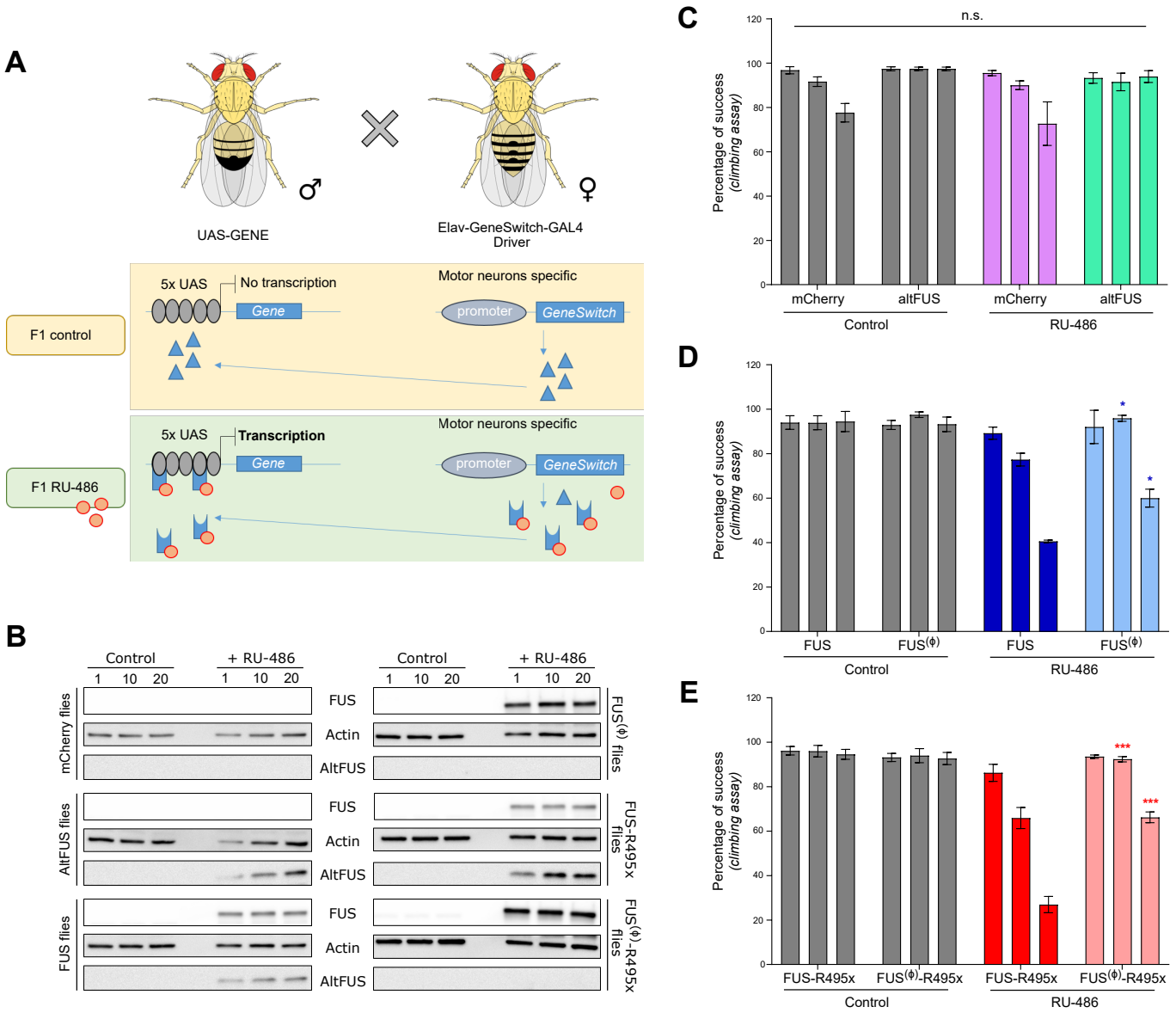

A
UAS-GENE
Elav-GeneSwitch-GAL4 Driver
F1 control
5x UAS
No transcription
Gene
Motor neurons specific
promoter
GeneSwitch
F1 RU-486
5x UAS
Transcription
Gene
Motor neurons specific
promoter
GeneSwitch
B
Control
+ RU-486
1 10 20
1 10 20
mCherry flies
AltFUS flies
FUS flies
FUS
Actin
AltFUS
FUS(ϕ) flies
FUS-R495x flies
FUS(ϕ)-R495x flies
C
n.s.
Percentage of success (climbing assay)
mCherry
altFUS
mCherry
altFUS
Control
RU-486
D
Percentage of success (climbing assay)
FUS
FUS(ϕ)
FUS
FUS(ϕ)
Control
RU-486
E
Percentage of success (climbing assay)
FUS-R495x
FUS(ϕ)-R495x
FUS-R495x
FUS(ϕ)-R495x
Control
RU-486

Figure 5

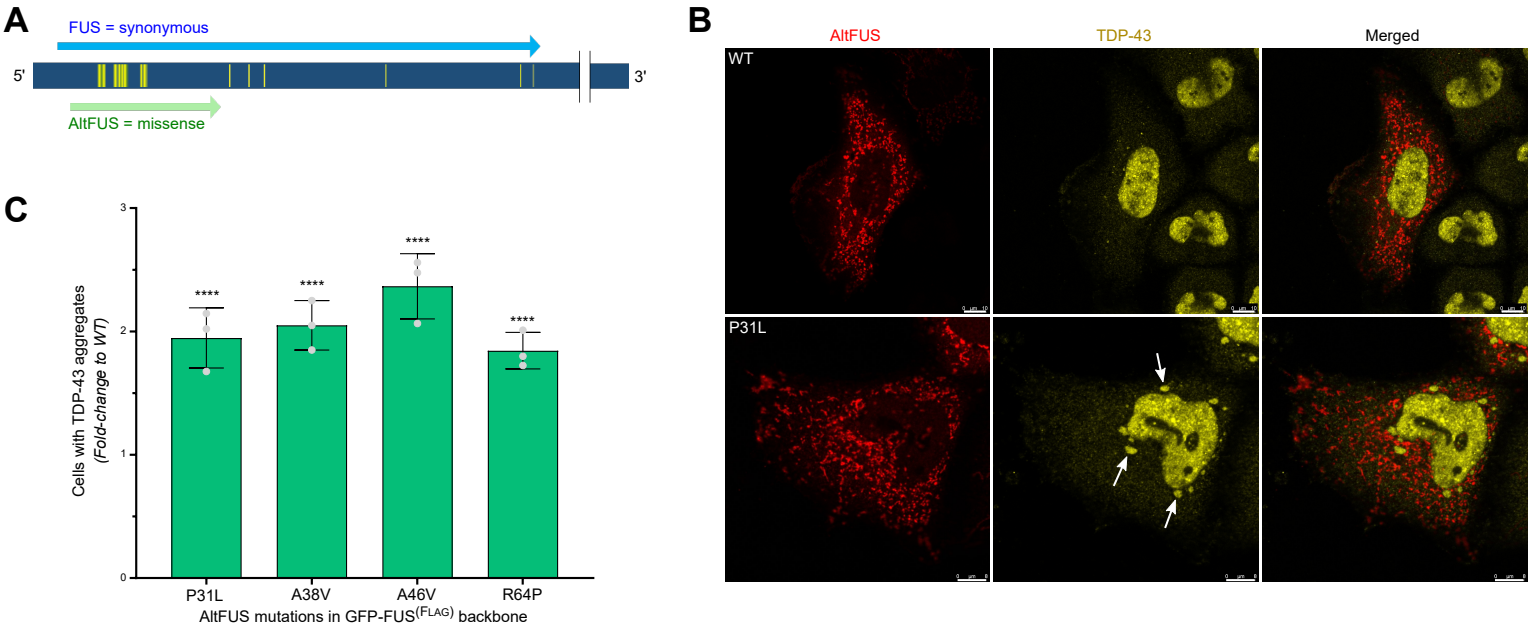

A
FUS = synonymous
5'
3'
AltFUS = missense
B
AltFUS
TDP-43
Merged
WT
P31L
C
Cells with TDP-43 aggregates
(Fold-change to WT)
****
P31L
A38V
A46V
R64P
AltFUS mutations in GFP-FUS(FLAG) backbone

Figure 6

A

>FUS – NM_004960.3

ACTTAAGCTTCGACGCAGGAGGCGGGGCTGCTCAGTCCTCCAGGCGTCGGTACTCAGCGGTGTTGGAACTTCGTTGCTTGCTTGCCTGTGCGCGCGTGCGCGGACATGGCCTCA
AACGATTATACCCAACAAGCAACCCAAAGCTATGGGGCCTACCCCACCCAGCCCGGGCAGGGCTATTCCCAGCAGAGCAGTCAGCCCTACGGACAGCAGAGTTACAGTGGTTAT
AGCCAGTCCACGGACACTTCAGGCTATGGCCAGAGCAGCTATTCTTCTTATGGCCAGAGCCAGAACACAGGCTATGGAACTCAGTCAACTCCCCAGGGATATGGCTCGACTGGC
GGCTATGGCAGTAGCCAGAGCTCCCAATCGTCTTACGGGCAGCAGTCCTCCTACCCTGGCTATGGCCAGCAGCCAGCTCCCAGCAGCACCTCGGGAAGTTACGGTAGCAGTTCT
CAGAGCAGCAGCTATGGGCAGCCCCAGAGTGGGAGCTACAGCCAGCAGCCTAGCTATGGTGGACAGCAGCAAAGCTATGGACAGCAGCAAAGCTATAATCCCCCTCAGGGCTAT
GGACAGCAGAACCAGTACAACAGCAGCAGTGGTGGTGGAGGTGGAGGTGGAGGTGGAGGTAACTATGGCCAAGATCAATCCTCCATGAGTAGTGGTGGTGGCAGTGGTGGCGGT
TATGGCAATCAAGACCAGAGTGGTGGAGGTGGCAGCGGTGGCTATGGACAGCAGGACCGTGGAGGCCGCGGCAGGGGTGGCAGTGGTGGCGGCGGCGGCGGCGGTGGTGGT
TACAACCGCAGCAGTGGTGGCTATGAACCCAGAGGTCGTGGAGGTGGCCGTGGAGGCAGAGGTGGCATGGGCGGAAGTGACCGTGGTGGCTTCAATAAATTTGGTGGCCCTCGG
GACCAAGGATCACGTCATGACTCCGAACAGGATAATTCAGACAACAACACCATCTTTGTGCAAGGCCTGGGTGAGAATGTTACAATTGAGTCTGTGGCTGATTACTTCAAGCAG
ATTGGTATTATTAAGACAAACAAGAAAACGGGACAGCCCATGATTAATTTGTACACAGACAGGGAAACTGGCAAGCTGAAGGGAGAGGCAACGGTCTCTTTTGATGACCCACCT
TCAGCTAAAGCAGCTATTGACTGGTTTGATGGTAAAGAATTCTCCGGAAATCCTATCAAGGTCTCATTTGCTACTCGCCGGGCAGACTTTAATCGGGGTGGTGGCAATGGTCGT
GGAGGCCGAGGGCGAGGAGGACCCATGGGCCGTGGAGGCTATGGAGGTGGTGGCAGTGGTGGTGGTGGCCGAGGAGGATTTCCCAGTGGAGGTGGTGGCGGTGGAGGACAGCAG
CGAGCTGGTGACTGGAAGTGTCCTAATCCCACCTGTGAGAATATGAACTTCTCTTGGAGGAATGAATGCAACCAGTGTAAGGCCCCTAAACCAGATGGCCCAGGAGGGGGACCA
GGTGGCTCTCACATGGGGGGTAACTACGGGGATGATCGTCGTGGTGGCAGAGGAGGCTATGATCGAGGCGGCTACCGGGGCCGCGGCGGGGACCGTGGAGGCTTCCGAGGGGGC
CGGGGTGGTGGGGACAGAGGTGGCTTTGGCCCTGGCAAGATGGATTCCAGGGGTGAGCACAGACAGGATCGCAGGGAGAGGCCGTATTAATTAGCCTGGCTCCCCAGGTTCTGG
AACAGCTTTTT[…]GTAAGAA

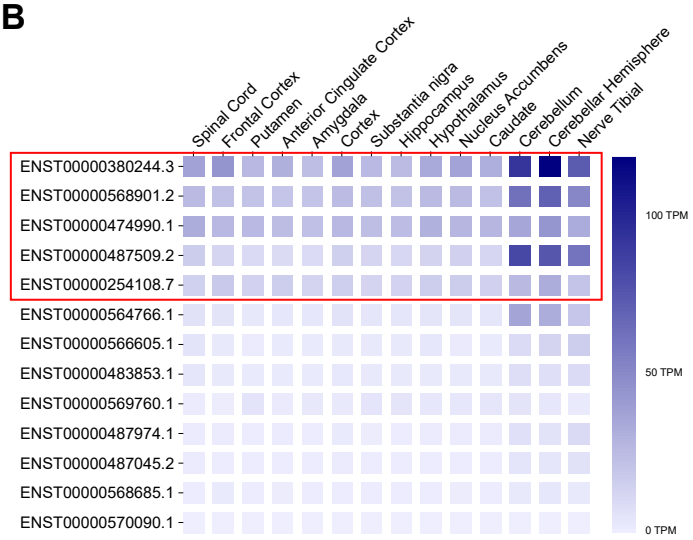
>Frame +1 = FUS protein

MASNDYTQQATQSYGAYPTQPGQGYSQQSSQPYGQQSYSGYSQSTDTSGYGQSSYSSYGQSQNTGYGTQSTPQGYGSTGGYGSSQSSQSSYGQQSSYPGYGQQPAPSSTSGSYG
SSSQSSSYGQPQSGSYSQQPSYGGQQQSYGQQQSYNPPQGYGQQNQYNSSSGGGGGGGGGGNYGQDQSSMSSGGGSGGGYGNQDQSGGGGSGGYGQQDRGGRGRGGSGGGGGGG
GGGYNRSSGGYEPRGRGGGRGGGMGGSDRGGFNKFGGPRDQGSRHDSEQDNSDNNTIFVQGLGENVTIESVADYFKQIGIIKTNKKTGQPMINLYTDRETGKLKGEATVSFD
DPPSAKAAIDWFDGKEFSGNPIKVSFATRRADFNRGGGNGRGGRGRGGPMGRGGYGGGGSGGGGRGGFPSGGGGGGGQQRAGDWKCPNPTCENMNFSWRNECNQCKAPKPDGPG
GGPGGSHMGGNYGDDRRGGRGGYDRGGYRGRGGDRGGFRGGRGGDRGGFGPGKMDSRGEHRQDRRERPY

>Frame +2 = altFUS protein

MGPTPPSPGRAIPSRAVSPTDSRVTVVIASPRTLQAMARAAILLMARARTQAMELSQLPRDMARLAAMAVARAPNRLTGSSPPTLAMASSQLPAAPREVTVAVLRAAAMGSPRVG
ATASSLAMVDSSKAMDSSKAIIPLRAMDSRTSTTAAVVVEVEVEVEVTMAKINPP

B

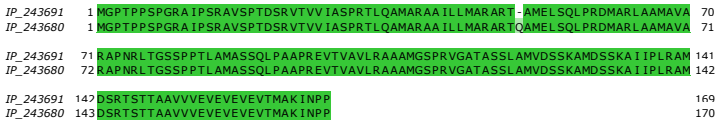


C

Transcript	Ensembl Proteins	OpenProt Proteins
ENST00000380244 (FUS-202)	ENSP00000369594 525 amino acids - FUS isoform	ENSP00000369594 525 amino acids - FUS isoform IP_243691 169 amino acids - altFUS isoform
ENST00000487509 (FUS-206)	No protein	IP_243680 170 amino acids - altFUS
ENST00000568901 (FUS-211)	No protein	No protein
ENST00000474990 (FUS-203)	No protein	No protein
ENST00000254108 (FUS-201)	ENSP00000254108 526 amino acids - FUS	ENSP00000254108 526 amino acids - FUS IP_243680 170 amino acids - altFUS

D

E



F

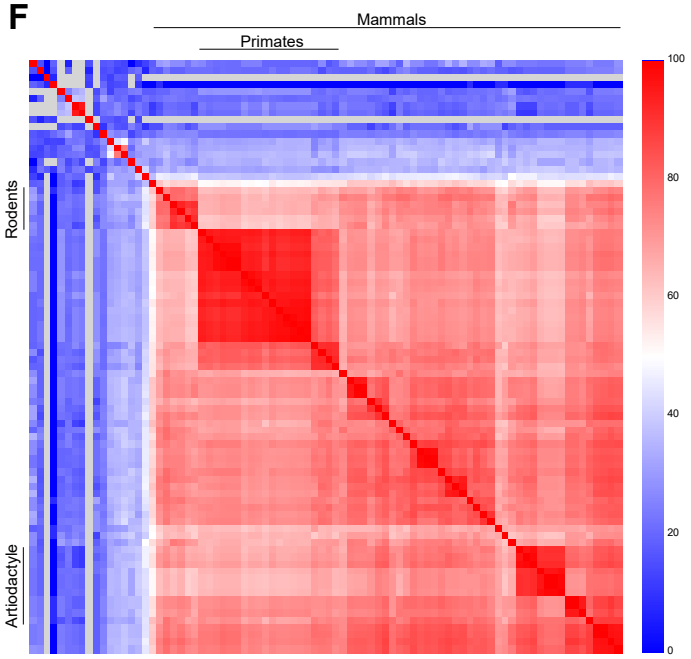


Figure EV1

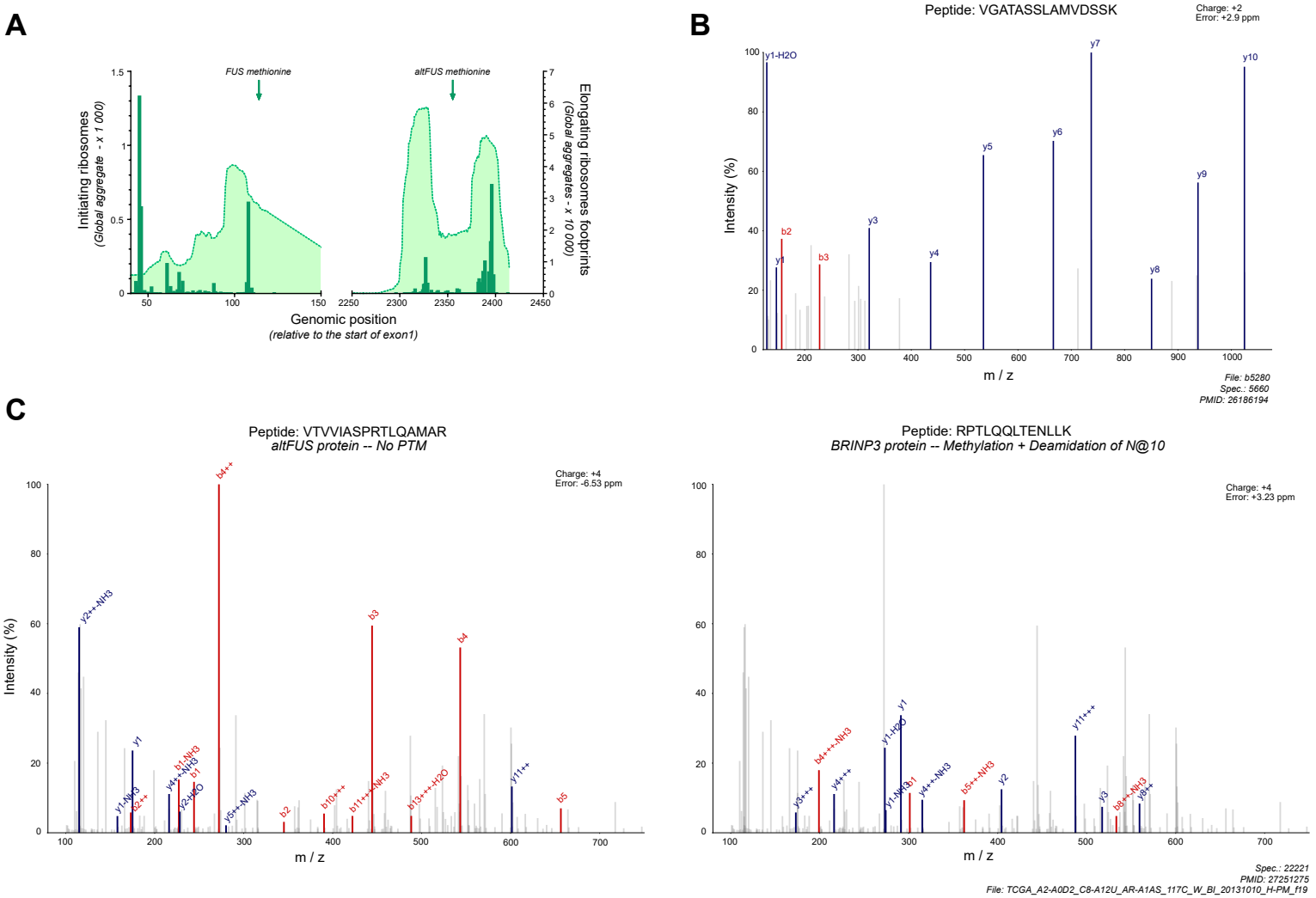
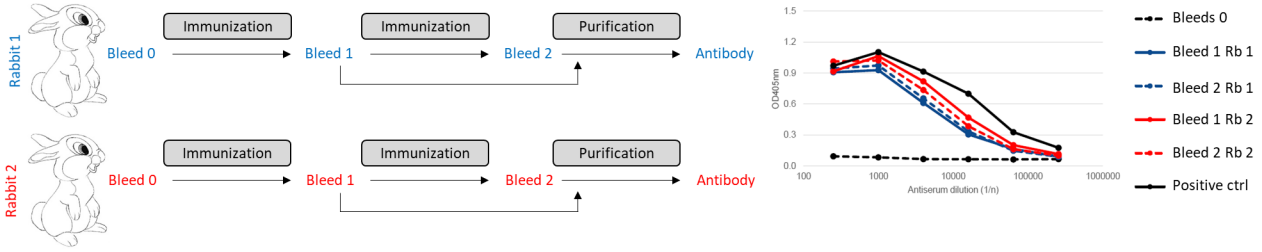

A
FUS methionine
altFUS methionine
Initiating ribosomes
(Global aggregate - x 1 000)
Elongating ribosomes footprints
(Global aggregates - x 10 000)
Genomic position
(relative to the start of exon1)
B
Peptide: VGATASSLAMVDSSK
Charge: +2
Error: +2.9 ppm
Intensity (%)
m / z
File: b5280
Spec.: 5660
PMID: 26186194
C
Peptide: VTVVIASPRTLQAMAR
altFUS protein -- No PTM
Charge: +4
Error: -6.53 ppm
Intensity (%)
m / z
Peptide: RPTLQQLTENLLK
BRINP3 protein -- Methylation + Deamidation of N@10
Charge: +4
Error: +3.23 ppm
m / z
Spec.: 22221
PMID: 27251275
File: TCGA_A2-A0D2_C8-A12U_AR-A1AS_117C_W_BI_20131010_H-PM_f19

Figure EV2

a

MGPT**PPSPGRAIPSRAVSPTDSR**VTVVIASPRTLQAMARAAILLMARARTQAMELSQLPRDMARLAAMAVA**RAPNRLTGSSPPTLA**MASSQLPAAPREVTVAVLRA
AAMGSPRVGATASSLAMVDSSKAMDSSKAIIPLRAMDSRTSTTAAVVVEVEVEVEVTMAKINPP



b

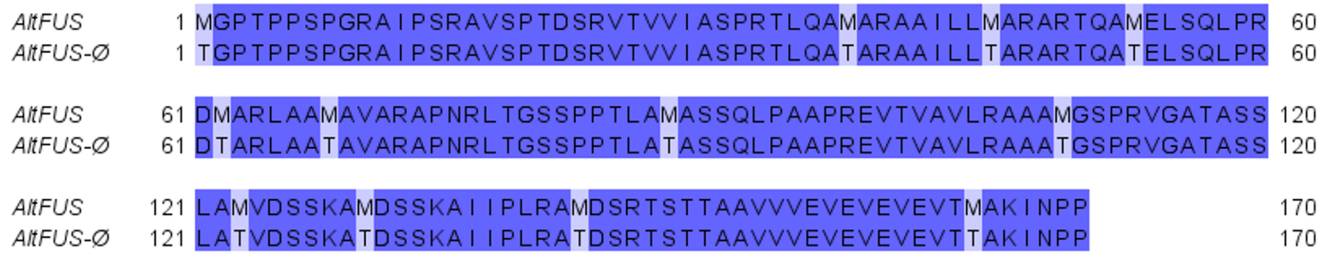
>Mutated_sequence

ACGGGGCCTACCCCACCCAGCCCGGGCAGGGCTATTCCCAGCAGAGCAGTCAGCCCTACGGACAGCAGAGTTACAGTGGTTATAGC
CAGTCCACGGACACTTCAGGCT**ACG**GCCAGAGCAGCTATTCTTCTT**ACG**GCCAGAGCCAGAACACAGGCT**ACG**GAACTCAGTCAAC
TCCCCAGGGAT**ACG**GCTCGACTGGCGGCT**ACG**GCAGTAGCCAGAGCTCCCAATCGTCTTACGGGCAGCAGTCCTCCTACCCTGGCT**A**
CGGCCAGCAGCCAGCTCCCAGCAGCACCTCGGGAAGTTACGGTAGCAGTTCTCAGAGCAGCAGCT**ACG**GGCAGCCCCAGAGTGGG
AGCTACAGCCAGCAGCCTAGCT**ACG**GTGGACAGCAGCAAAGCT**ACG**GACAGCAGCAAAGCTATAATCCCCCTCAGGGCT**ACG**GACA
GCAGAACCAGTACAACAGCAGCAGTGGTGGTGGAGGTGGAGGTGGAGGTGGAGGTAACT**ACG**GCCAAGATCAATCCTCCATGA

>altFUS_Ø

TGPTPPSPGRAIPSRAVSPTDSRVTVVIASPRTLQA**T**ARAAILL**T**ARARTQA**T**ELSQLPRD**T**ARLAA**T**AVARAPNRLTGSSPPTLA**T**ASSQLP
AAPREVTVAVLRAAA**T**GSPRVGATASSLA**T**VDSSKA**T**DSSKAIIPLRA**T**DSRTSTTAAVVVEVEVEVEVT**T**AKINPP

c



d

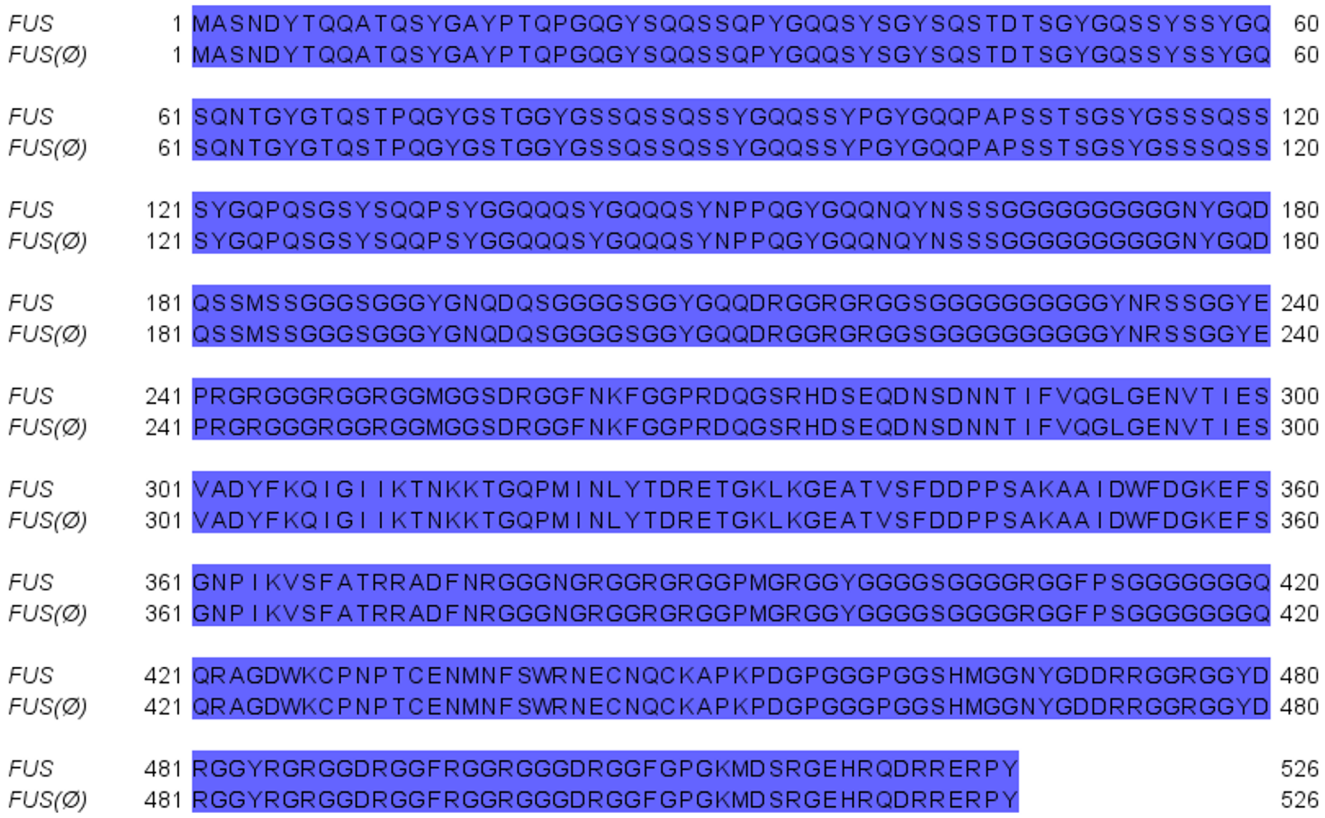


Figure EV3

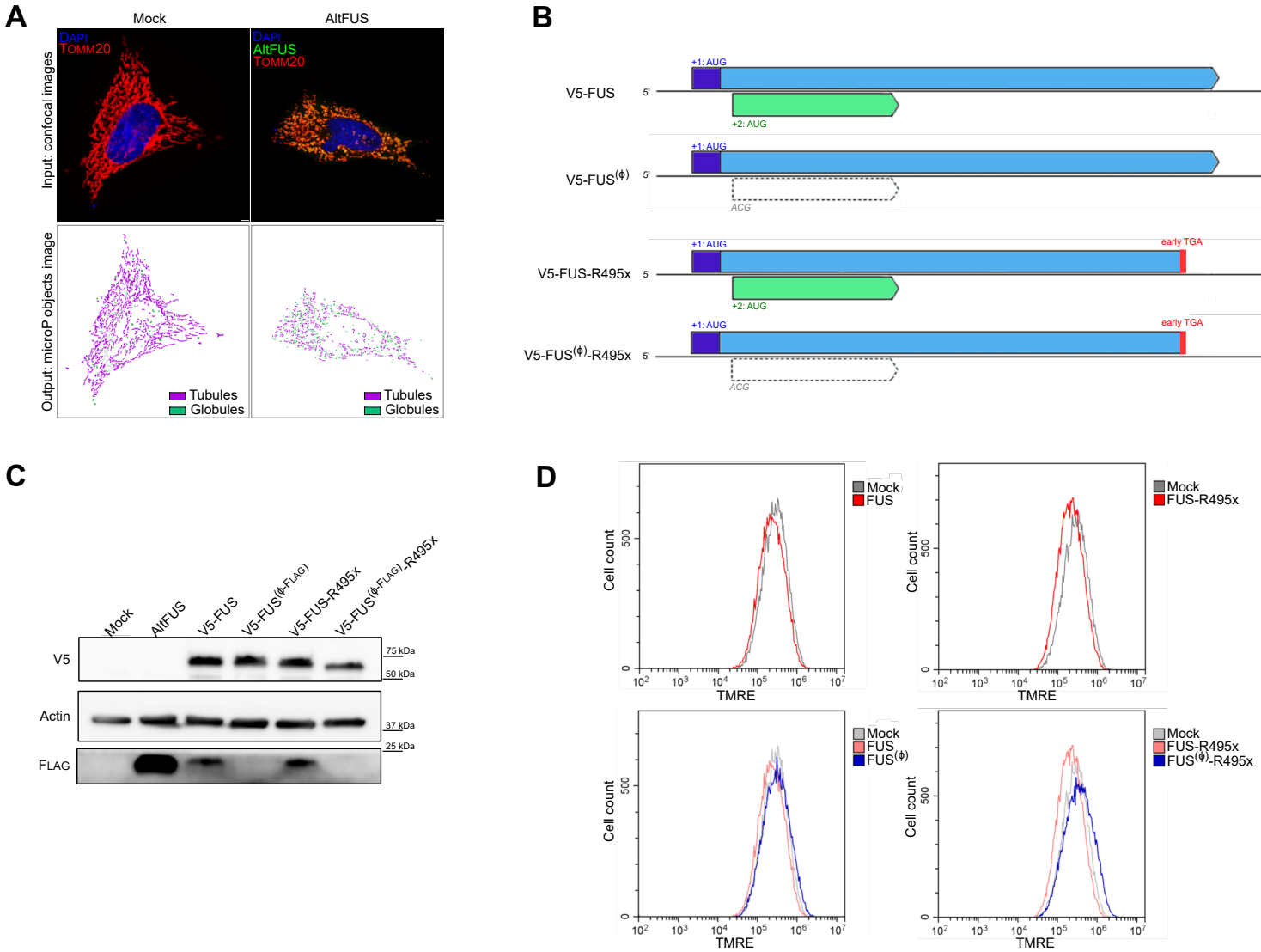

A
Mock
AltFUS
DAPI
TOMM20
DAPI
AltFUS
TOMM20
Input: confocal images
Output: microP objects image
Tubules
Globules
Tubules
Globules
B
V5-FUS
V5-FUS(ϕ)
V5-FUS-R495x
V5-FUS(ϕ)-R495x
+1: AUG
+2: AUG
ACG
early TGA
5'
3'
C
Mock
AltFUS
V5-FUS
V5-FUS(ϕFLAG)
V5-FUS-R495x
V5-FUS(ϕFLAG)-R495x
V5
Actin
FLAG
75 kDa
50 kDa
37 kDa
25 kDa
D
Cell count
TMRE
Mock
FUS
FUS-R495x
FUS(ϕ)
FUS(ϕ)-R495x

Figure EV4

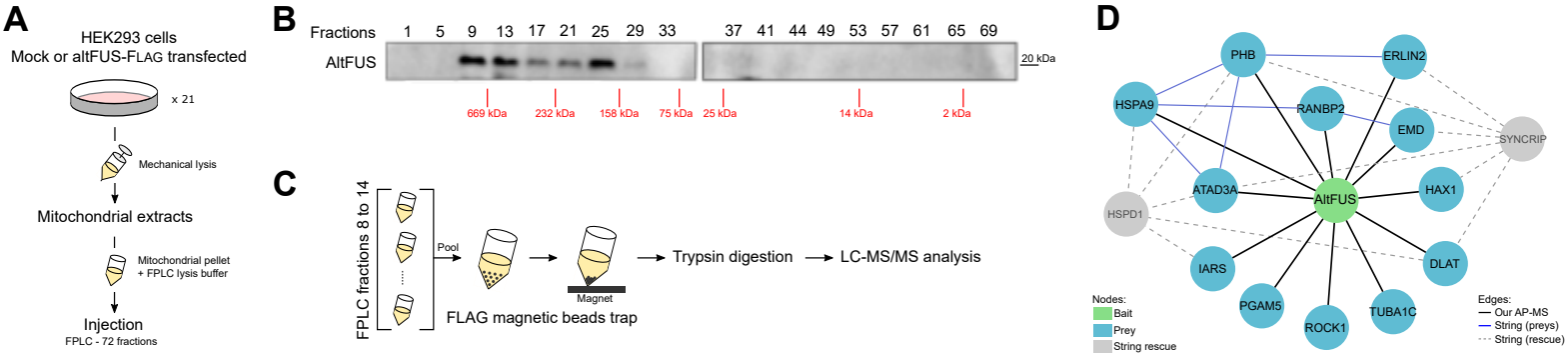

A
HEK293 cells
Mock or altFUS-FLAG transfected
x 21
Mechanical lysis
Mitochondrial extracts
Mitochondrial pellet
+ FPLC lysis buffer
Injection
FPLC - 72 fractions
B
Fractions
1 5 9 13 17 21 25 29 33 37 41 44 49 53 57 61 65 69
AltFUS
20 kDa
669 kDa
232 kDa
158 kDa
75 kDa
25 kDa
14 kDa
2 kDa
C
FPLC fractions 8 to 14
Pool
Magnet
FLAG magnetic beads trap
Trypsin digestion
LC-MS/MS analysis
D
PHB
ERLIN2
HSPA9
RANBP2
EMD
SYNCRIP
ATAD3A
AltFUS
HAX1
HSPD1
IARS
DLAT
PGAM5
ROCK1
TUBA1C
Nodes:
Bait
Prey
String rescue
Edges:
Our AP-MS
String (preys)
String (rescue)

Figure EV5

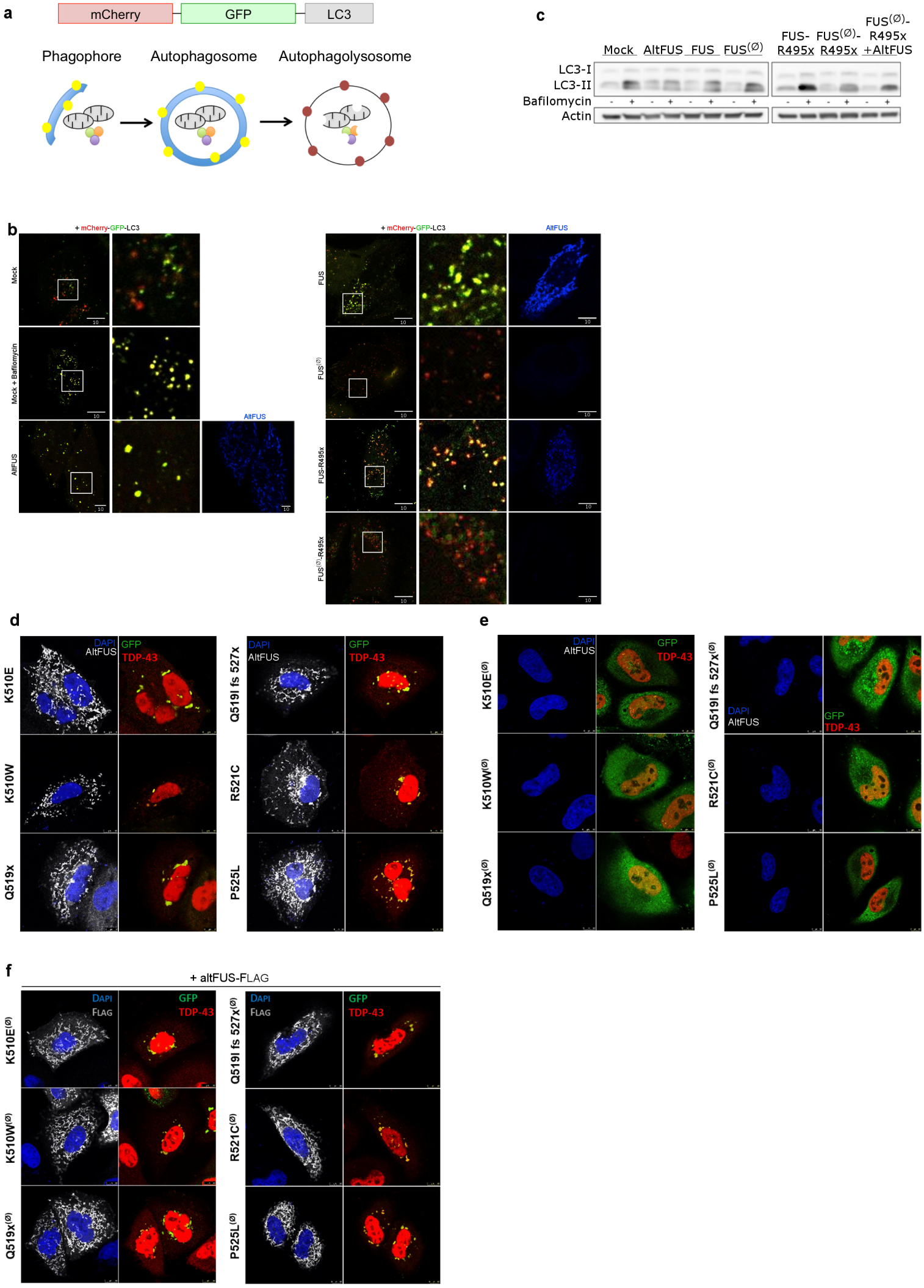

a
mCherry
GFP
LC3
Phagophore
Autophagosome
Autophagolysosome
c
Mock
AltFUS
FUS
FUS(Ø)
FUS-R495x
FUS(Ø)-R495x
FUS(Ø)-R495x +AltFUS
LC3-I
LC3-II
Bafilomycin
Actin
b
+ mCherry-GFP-LC3
AltFUS
Mock
Mock + Bafilomycin
FUS
FUS(Ø)
FUS-R495x
FUS(Ø)-R495x
d
K510E
K510W
Q519x
Q519I fs 527x
R521C
P525L
DAPI
AltFUS
GFP
TDP-43
e
K510E(Ø)
K510W(Ø)
Q519x(Ø)
Q519I fs 527x(Ø)
R521C(Ø)
P525L(Ø)
f
+ altFUS-FLAG
FLAG

Figure EV6

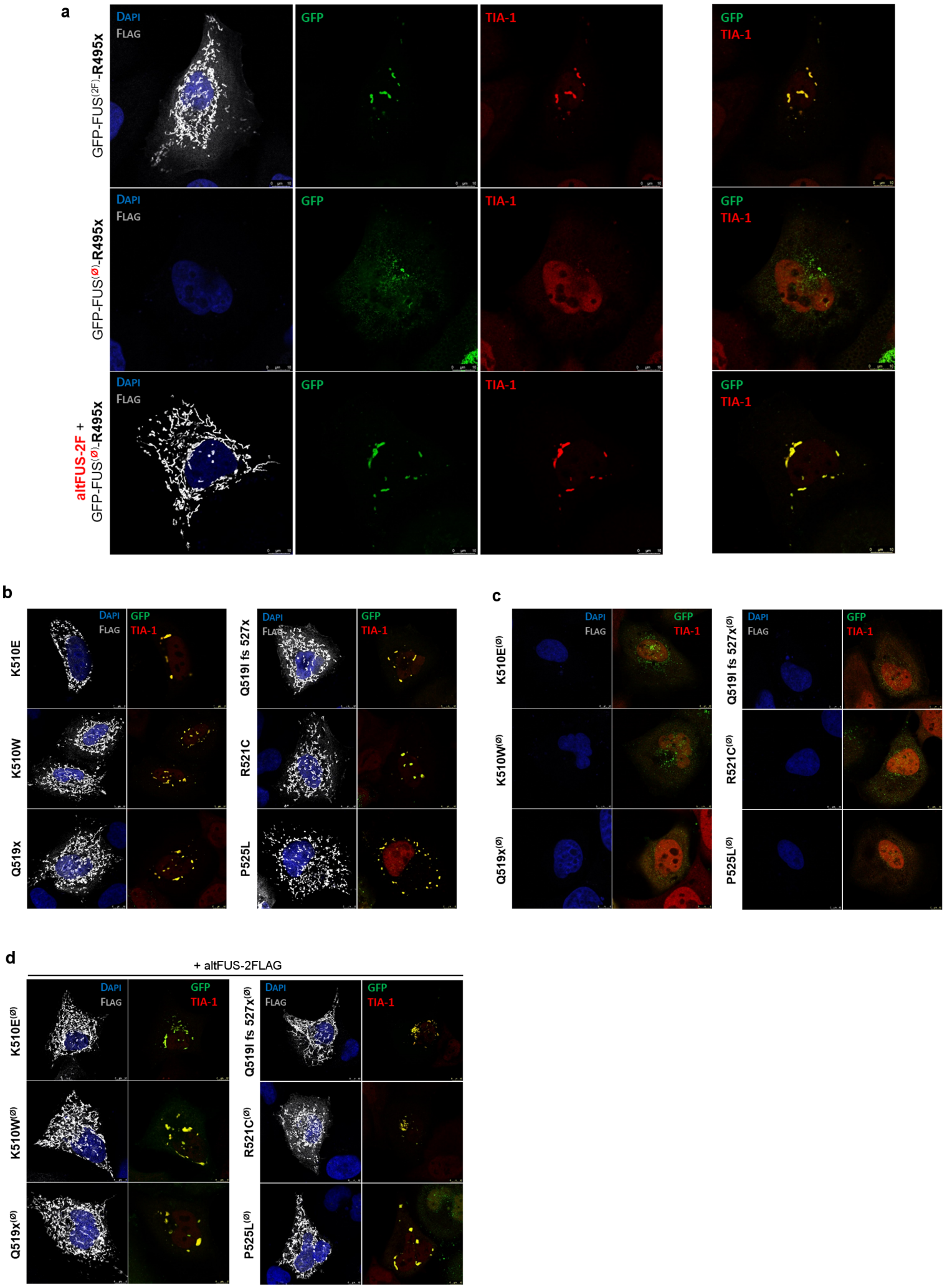

a
GFP-FUS(2F)-R495x
GFP-FUS(Ø)-R495x
altFUS-2F +
GFP-FUS(Ø)-R495x
DAPI
FLAG
GFP
TIA-1
b
K510E
K510W
Q519x
Q519I fs 527x
R521C
P525L
c
K510E(Ø)
K510W(Ø)
Q519x(Ø)
Q519I fs 527x(Ø)
R521C(Ø)
P525L(Ø)
d
+ altFUS-2FLAG

Figure EV7

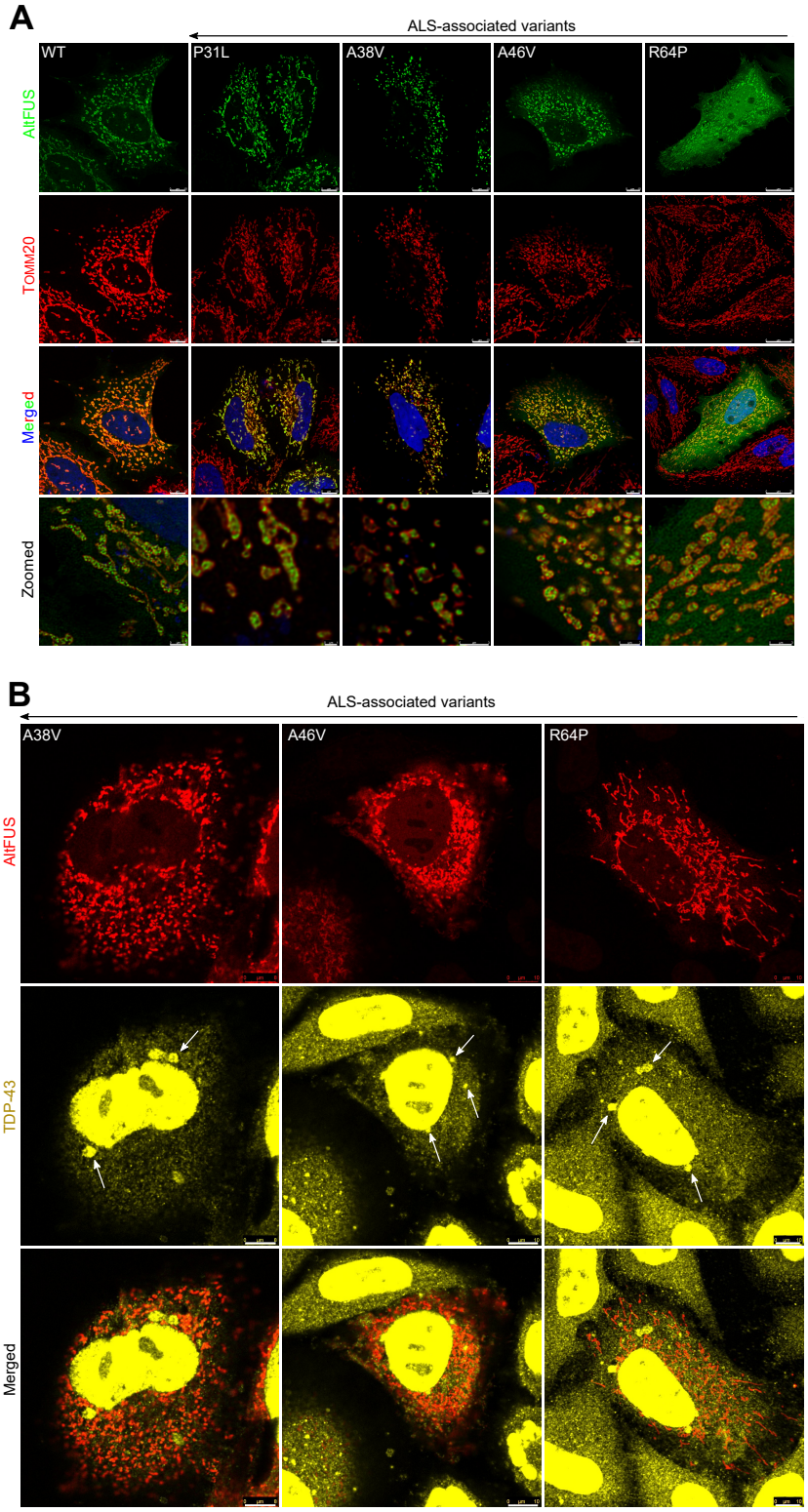
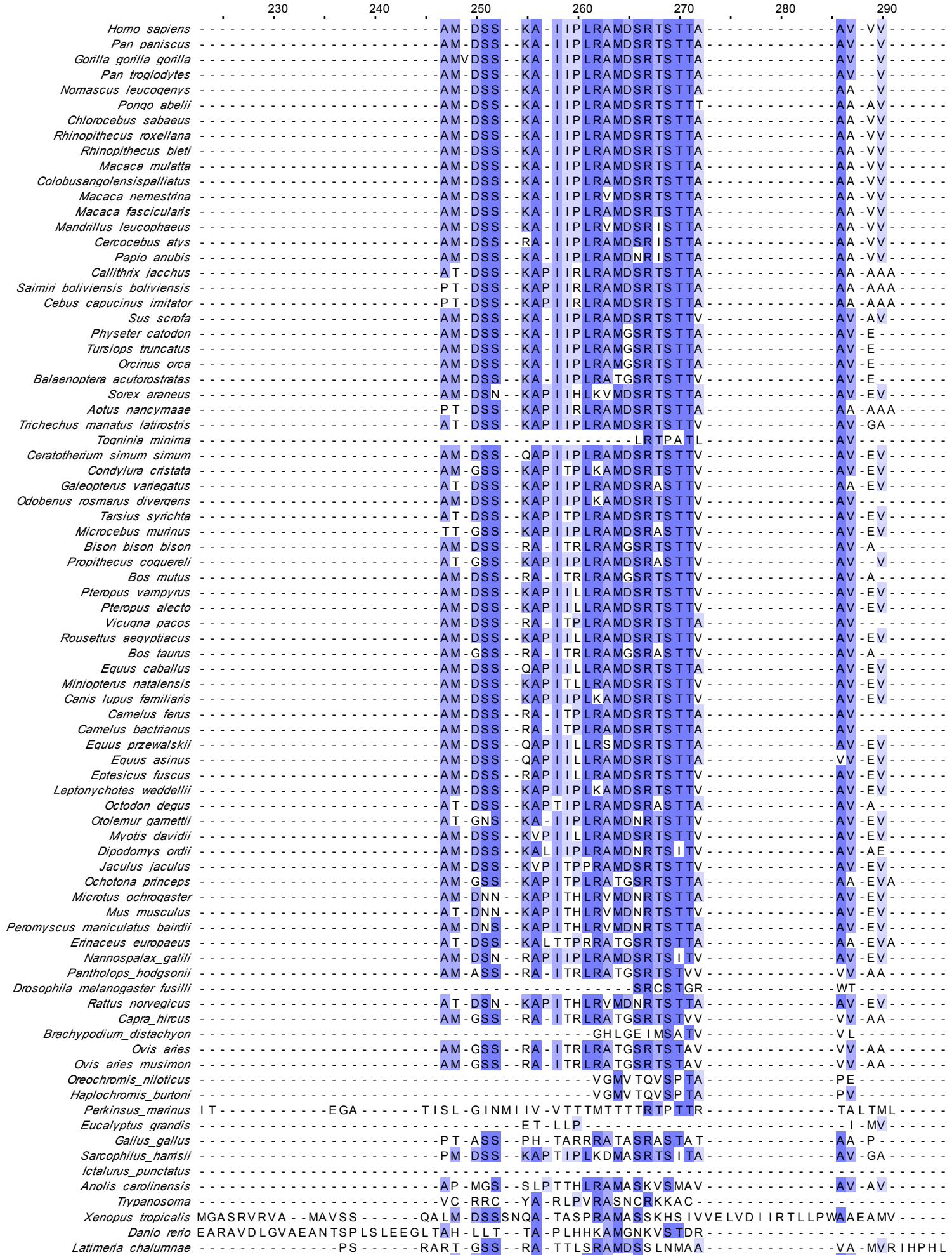

A
ALS-associated variants
WT
P31L
A38V
A46V
R64P
AltFUS
TOMM20
Merged
Zoomed
B
ALS-associated variants
A38V
A46V
R64P
AltFUS
TDP-43
Merged

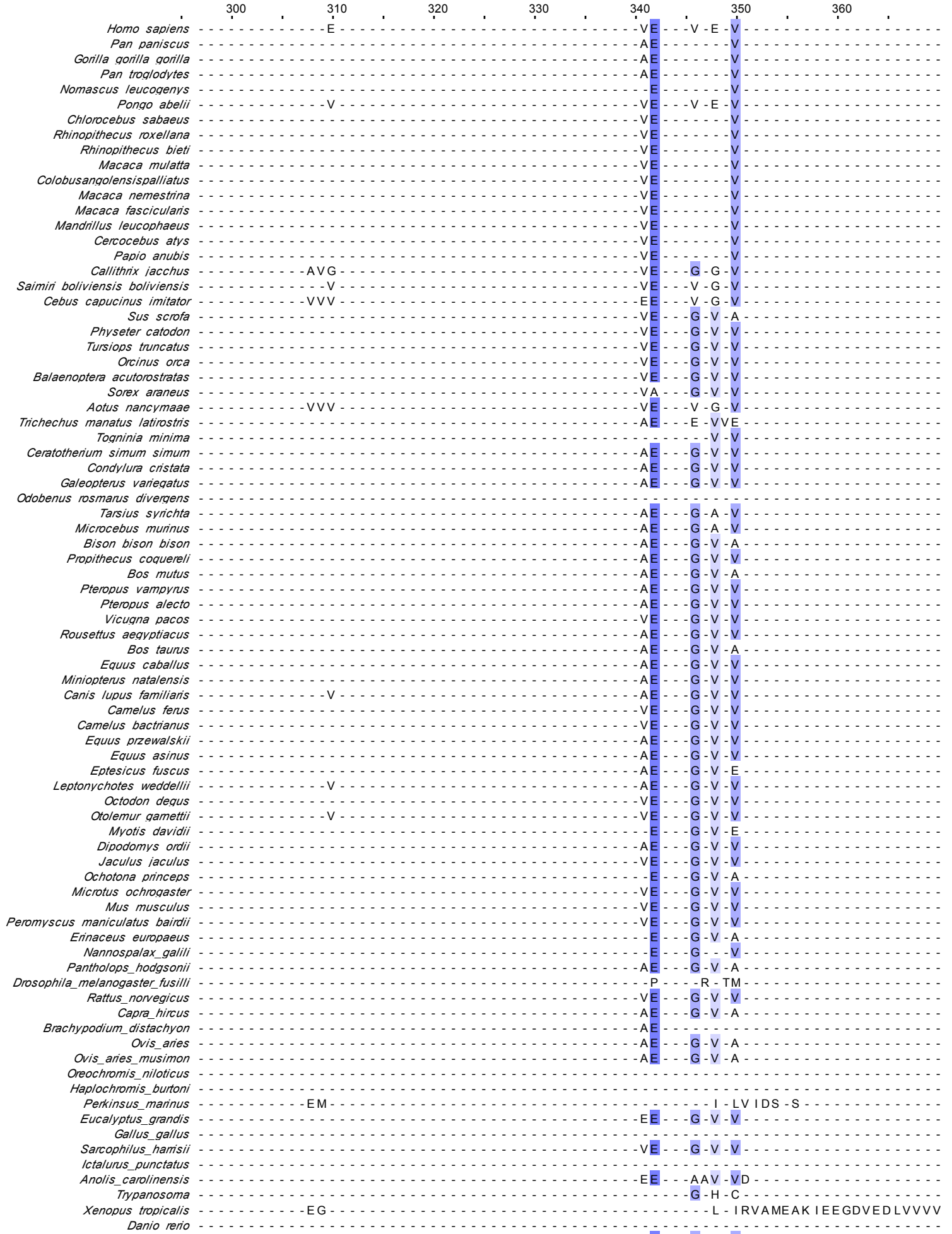
Figure EV8

```
                                        10        20        30        40        50        60        70
                                 ·         ·         ·         ·         ·         ·         ·         ·
Homo sapiens                     MGP----------T---PPS---PG--RA-IPSRAVSPTDSRVTVVIASPRTL-----QAM-ARAAI-L-LMAR
Pan paniscus                     MGP----------T---PPS---PG--RA-IPSRAVSPTDSRVTVVIASPRTL-----QAM-ARAAI-L-LMAR
Gorilla gorilla gorilla          MGP----------T---PPS---PG--RA-IPSRAVSPTDSRVTVVIASPRTL-----QAM-ARAAI-L-LMAR
Pan troglodytes                  MGP----------T---PPS---PG--RA-IPSRAVSPTDSRVTVVIASPRTL-----QAM-ARAAI-L-LMAR
Nomascus leucogenys              MGP----------T---PPS---PG--RA-IPSRAVSPTDSRVTAVIASPRTL-----QAM-ARAAI---LMAR
Pongo abelii                     MGP----------T---PPS---PG--RA-IPSRAVSPTDSRVTALIASPRTL-----QAM-ARAAI-L-LTAR
Chlorocebus sabaeus              MGP----------T---PPS---PG--RA-IPSRAVSPTDSRVIVVTASPRT------QAM-ARAAI-L-LTAR
Rhinopithecus roxellana          MGP----------T---PPS---PG--RA-IPSRAVSPTDSRVTAVTASPRTL-----QAM-ARAAI-L-LTAR
Rhinopithecus bieti              MGP----------T---PPS---PG--RA-IPSRAVSPTDSRVTAVTASPRTL-----QAM-ARAAI-L-LTAR
Macaca mulatta                   MGP----------T---PPS---LG--RA-IPSRAVSPTDSRVTAVTASPRTL-----QAM-ARAAI-L-LTAR
Colobusangolensispalliatus       MGP----------T---PPS---PV--RA-IPSRAVSPTDSRVTAVTASPRTL-----QAM-ARAAI-L-LTAR
Macaca nemestrina                MGP----------T---PPS---LG--RA-IPSRAVSPTDSRVTAVTASPRTL-----QAM-ARAAI-L-LTAR
Macaca fascicularis              MGP----------T---PPS---LG--RA-FPSRAVSPTDSRVTAVTASPRTL-----QAM-ARAAI-L-LTAR
Mandrillus leucophaeus           MGP----------T---PPS---LG--RA-IPSRAVSPTDSRVTAVTASPRTL-----QAM-ARAAI-L-LTAR
Cercocebus atys                  MGP----------T---PPS---LG--RA-IPSRAVSPTDSRVTAVTASPRTL-----QAM-ARAAI-L-LTAR
Papio anubis                     MGP----------T---PPS---LG--RA-IPSRAVSPTDSRVTAVTASPRTL-----QAM-ARAAI-L-LTAR
Callithrix jacchus               ----------------------------------------------------------------------------
Saimiri boliviensis boliviensis  MGP----------T---PPS---PG--RA-IPSRAVSPTDSRVTAVTASPQTL-----QAM-ARAAMVL-LTAR
Cebus capucinus imitator         MGP----------T---PPS---PG--RA-IPSRAVSPTDSRVTVVMGSLQIL-----QAM-ARAAMVL-LTAR
Sus scrofa                       ----------------------------------------------------------M-ARAAMAV-LMDR
Physeter catodon                 MGP----------T---QPS---LG--RA-IPSRAISPMDSRVTVAMASQRTL-----QAM-ARAAMVV-LMDR
Tursiops truncatus               MGP----------T---RPS---LG--RA-IPSRAISPMDSRVTVATASQRML-----QAM-ARAAMVV-LTDR
Orcinus orca                     MGP----------T---RPS---LG--RA-IPSRAISPMDSRVTVATASQRML-----QAM-ARAAMVV-LTDR
Balaenoptera acutorostratas      MGP----------T---QPS---LG--RA-IPSRAISPMDSRVTVATASQRTL-----QAM-ARAAMVV-LMDR
Sorex araneus                    ----------------------------------------------------------------------------
Aotus nancymaae                  MGP----------T---PRS---PG--RA-IPSRAVSPMDSRVTAVMASPQIL-----QAM-AKAAMVL-LTAR
Trichechus manatus latirostris   ----------------------------------------------------------M-ARAAMVL-LMDK
Togninia minima                  MTL----------L---QAR---H-----------------------------------------------
Ceratotherium simum simum        MGP----------T---PPS---LG--RA-TPSRAVSPMDSRVTVVTVSQQTL-----QAM-ARAAMVL-LMDR
Condylura cristata               MGP----------T---PPS---LG--RA-IPSRAISPMDSRVTVVMASQQTL-----QAM-ARAAMVV-LMDR
Galeopterus variegatus           MGP----------T---PPS---LG--RA-TPSRAISPTDSRVTVVTASQQTL-----QAM-ARAAIVL-LMDR
Odobenus rosmarus divergens      MGP----------T---PPS---LG--RA-IPSRAVSPMDSRVTVVMASQQTL-----QAM-ARAAMVL-LTDR
Tarsius syrichta                 MGP----------T---LPS---LG--RA-TPSRAVSPMGSRVTVVMASQRTL-----QAM-ARAAIVL-LMDR
Microcebus murinus               MGP----------T---PPS---RG--RA-TRSRAVSPMDSRVTVVTAS-QTL-----QAM-VRAATVV-LMDR
Bison bison bison                ----------------------------------------------------------M-ARAATVV-LMDR
Propithecus coquereli            MGP----------T---PPS---RG--RA-THSRAVSPMDSRVTVVTAS-QTL-----QAM-ARAAMVV-LMDR
Bos mutus                        ----------------------------------------------------------M-ARAATVV-LMDR
Pteropus vampyrus                MGL----------T---PPS---LD--RA-IPSRAVSPMDSRVTVVTASQRTL-----LAM-ARAAM-V-LMDR
Pteropus alecto                  MGL----------T---PPS---LD--RA-IPSRAVSPMDSRVTVVTASQRTL-----LAM-ARAAM-V-LMDR
Vicugna pacos                    MGP----------T---LPS---LG--RA-IPSRAISPTDSRVTVVMASHRTL-----QAM-ARAAIVV-LMDR
Rousettus aegyptiacus            MGL----------T---PPS---LD--RA-IPSRAVSPMDSRVTAVTASHRTP-----QAM-ARAAM-V-LMDR
Bos taurus                       ----------------------------------------------------------M-ARAATVV-LMDR
Equus caballus                   MGP----------T---PPS---LG--RA-IPSRAINPMDSRVTAVTVSQQTL-----QAM-ARAAMVL-LMDR
Miniopterus natalensis           MGP----------T---PPS---LG--RV-IPSRAISPMDNRVMVATASQGTL-----QAM-ARAPM-V-LMDR
Canis lupus familiaris           MGP----------T---PPS---LG--RA-IPSRAISPMDSRVTVVTVSQQTL-----QAT-ARAAMVL-LMDR
Camelus ferus                    MGP----------T---LPS---LG--RA-IPSRAISPMDSRVTVVMANHRTL-----QAM-ARAAMVV-LMDR
Camelus bactrianus               MGP----------T---LPS---LG--RA-IPSRAISPMDSRVTVVMANHRTL-----QAM-ARAAMVV-LMDR
Equus przewalskii                MGP----------T---PPS---LG--RA-IPSRAINPMDSRVTAVTVSQQTL-----QAM-ARAAMVL-LMDR
Equus asinus                     MGP----------T---PPS---LG--RA-IPSRAINPMDSRVTAVTVSQQTL-----QAM-ARAAMVL-LMDR
Eptesicus fuscus                 MGP----------T---PPS---LG--RV-IPSRATSPMDSRVTVATASLGTL-----QAM-ARAPT-V-LTGR
Leptonychotes weddellii          MGP----------T---PPS---LG--RA-IPSRAISPMDSRVTVVTASQQTL-----QAM-ARAAMVL-LTDR
Octodon degus                    MGP----------T---PPS---LG--RA-TPSRAISPMDSRVTVAMASQQTL-----QAM-ARAAMVL-LMDR
Otolemur garnettii               MGP----------T---PPS---LG--RA-TPNRAISPMDSRVTVVMASQ-IL-----QAM-ARAAMV--LMDR
Myotis davidii                   MGP----------T---PPS---LG--RA-IPSRAASPMDSRVTVATASLGTP-----QAM-ARAPM-V-PTDR
Dipodomys ordii                  MGH----------T---LLS---LG--RA-TPSRAASPMDSRATVVMASQQTL-----QAM-VRAAMVL-LMDR
Jaculus jaculus                  MGP----------I---PHS---LG--RA-TLSRAISLMDSRVTVVMASQQIL-----QDM-ARAAMVL-LMDR
Ochotona princeps                ----------------------------------MASRAMAATGSPRTL-----PAM-ARAAMAP-RTDR
Microtus ochrogaster             MGP----------I---LPS---LG--RA-TPNRAINPMASRVTVVMANQLTL-----QDT-ARAAMAL-LMDR
Mus musculus                     ----------------------------------MANRVTVVMASQLTL-----QDT-ARAAMVL-LMDR
Peromyscus maniculatus bairdii   MGL----------I---LPS---LG--RD-TPSRAINPMASRVTVVMASQLTL-----RDM-ARAAMGL-LMDR
Erinaceus europaeus              MGR----------T---PPS---LG--RA-IPSRATSPTDSRVTEVTASQRTR-----QAM-ARAAMAL-PMDK
Nannospalax_galili               MGP----------T---LPS---LG--RV-TPNRAISPMDSRVTVDMASQLTL-----QGM-ARAAMVL-LMDR
Pantholops_hodgsonii             MGP----------T---QLS---LG--KA-TPSRAVSPTGSRVTVATASLRTL-----RAM-ARAATVV-LTDR
Drosophila_melanogaster_fusilli  A---------------------------------------------MFL-----TGMRACCSLRS-RMDG
Rattus_norvegicus                MGP----------I---LPS---LG--RA-IPNRAISPMASRVTVVMANQLTL-----QDT-ARAAMVL-LMDR
Capra_hircus                     MGP----------T---QLS---LG--KA-THSRAVNPTGSRVTVATASLRTL-----RAM-ARAATVV-LTDR
Brachypodium_distachyon          MAT----------R---EGT-----------------------------------------------------
Ovis_aries                       MGP----------T---QLS---LG--KA-THSRAVNPTGSRVTVATASLQTL-----RAM-ARAATVV-LTDR
Ovis_aries_musimon               MGP----------T---QLS---LG--KA-THSRAVNPTGSRVTVATASLQTL-----RAM-ARAATVV-LTDR
Oreochromis_niloticus            MVE----------V---A-V---LA--RA-TVSRLARVTASRATEATTRART-----------------------
Haplochromis_burtoni             MVE----------V---A-V---LA--RA-TVSRLARVTASRATEATTRART-----------------------
Perkinsus_marinus                MEV----------VGRVPGS-----------------NSRVVAV-----------------------------
Eucalyptus_grandis               MASPAAVAPPGTTT---SPA---TT--AA-IPSRRTARIGTPTTPDLELVQVLYLLLVIGM-QNAAIV------
Gallus_gallus                    ----------------------------------------------------------------------------
Sarcophilus_harrisii             MGL----------T---LLS---LD--KV-IPNRAVSPMANRVIVDMASQLIP-----QVM-VRAAMAP-PTDK
Ictalurus_punctatus              MVD----------M---EVS---PV--RA-TASRVTEDTARVRTAVPGLTVTD-----TAP-ATARVDT-VPSR
Anolis_carolinensis              MGP----------T---QLH---PH--RV-ILSRLVSRMGSRAMVDMDSQQIL-----QDT-AKVVTAL-LMDR
Trypanosoma                      MQI----------F---VKT---L-----------------------TGKTI-----ALE-VESSDTI---EN
Xenopus tropicalis               MGH----------T---PPN---QA--RV-MGSRAASPTASKVTVVTASLQVA-----PRM-DRVATAL-HMVK
Danio rerio                      MIM----------A---KPQAMAMGVMED-SPVKATANLLPRTTVSRAMEDTT-----KAQ-RAAP-LLIIRED
Latimeria chalumnae              MV-----------------LQ---LG--RV-TLSKPLRRTASRDMVTTASQQTL-----RHM-DRVATLH-PMDK
Xenopus laevis                   MGH----------I---PPS---QA--RVMVSSRAARPTASKVTVATANLQVA-----PRM-DRVATAL-RMVK
```

```
                                        80        90        100       110       120       130       140
                                 Homo sapiens ART-----QAMELSQLPRD-MARLAAMAVAR----------------APNRL--------------TGSSPP-T
                                  Pan paniscus ART-----QAMELSQLPRD-MARLAAMAVAR----------------APNRL--------------TGSSPP-T
                         Gorilla gorilla gorilla ART-----QAMELSQLPRD-MARLAAMAVAR----------------APNRL--------------TGSSPP-T
                              Pan troglodytes ART-----QAMELSQLPRD-MARLAAMAVAR----------------APNRL--------------TGSSPP-T
                             Nomascus leucogenys ART-----QAMELSQLPRD-MARLAAIAVAR----------------APNRL--------------TGSSPP-T
                                   Pongo abelii ART-----QAMELSQLPRD-MARLAAMAVAR----------------APNRL--------------TGSSPP-T
                              Chlorocebus sabaeus ART-----QAMELSQLPRD-MARLAAMAVAR----------------APNLL--------------MGSSPP-T
                           Rhinopithecus roxellana ART-----QAMELSQLPRD-MAQLAAMAVAR----------------APNLL--------------TGSSPP-T
                              Rhinopithecus bieti ARTVSLSQQAMELSQLPRD-MAQLAAMAVAR----------------APNLL--------------TGSSPP-T
                                  Macaca mulatta ARA-----QAMELSQLPRD-MARLAAMAVAR----------------APNLL--------------MGSSPP-T
                       Colobusangolensispalliatus ART-----QAMELSQLPRD-MARLAAMAVAR----------------APNLL--------------TGSSPP-T
                                 Macaca nemestrina ARA-----QAMELSQLPRD-MARLAAMAVAR----------------APNLL--------------MGSSPP-T
                                Macaca fascicularis ARA-----QAMELSQLPRD-MARLAAMAVAR----------------APNLL--------------MGSSPP-T
                              Mandrillus leucophaeus ARA-----QAMELSQLPRD-MARLAAMAVAR----------------APNLL--------------MGSSPP-T
                                   Cercocebus atys ARA-----QAMELSQLPRG-MARLAAMAVAR----------------APNLL--------------MGSSPP-T
                                     Papio anubis VRAVSLSQQAMELSQLPRD-MARLAAMAVAR----------------APSLL--------------MGSSPP-T
                                  Callithrix jacchus ----------MALSQLPRD-MAQLVAMAVAR----------------APNHL--------------TGSSPP-T
                     Saimiri boliviensis boliviensis PRT-----QAMALSQLPRD-MAQLVAMAVAR----------------APNHL--------------TGSSPL-T
                           Cebus capucinus imitator PRT-----QAMALSQLPRD-MAQLVAMAVAR----------------APNHL--------------TGSSPL-T
                                        Sus scrofa PRT-----QAMAHSQLPRD-MAQLVAMVVAR----------------VLNRL--------------TGSSLH-T
                                     Physeter catodon PRT-----QAMAHSQLPRD-MAQLVAMAVAK----------------VLSRL--------------MGSSPH-T
                                    Tursiops truncatus PRT------AMAHSQLPRD-MAQLVAMAVAK----------------VLSRL--------------TGSSPH-T
                                        Orcinus orca PRT-----QAMAHSQLPRD-MAQLVAMAVAK----------------VLSRL--------------TGSSPH-T
                                Balaenoptera acutorostratas PRT-----QAMAHSQLPRD-MAQLVAMAVAK----------------VLSRL--------------TGSSPH-T
                                        Sorex araneus ----------MALSPLPRD-TAQLVAMAAVR----------------VLSRP--------------MDSSPP-T
                                        Aotus nancymaae PRT-----QAMALSQLPRD-MAQLVAIVVAR----------------APNHL--------------TGSSPP-T
                                Trichechus manatus latirostris PRT-----QVMAPSQLPRD-MAQLVAMAVAR----------------VPNHL--------------TGSSLR-I
                                        Togninia minima PRT-----A-PEMA-PPRR-TAR--------------------------------------------------------
                                Ceratotherium simum simum PRT-----QAMELSQLPKD-MAQLVAMAVAR----------------VLSRL--------------TGSSPP-T
                                      Condylura cristata PRT-----QVMALSQLPRD-MAQLAAMAVAR----------------VLNRL--------------TGSSPP-T
                                     Galeopterus variegatus PRT-----QAMALNQLPRD-MAQPVAMAVVR----------------VLNRL--------------TGSSPP-T
                                Odobenus rosmarus divergens PRT-----QAMALSQVPRD-MAQPVAMAVAR----------------VLSHL--------------MGNSPL-I
                                        Tarsius syrichta PRT-----QVIALSQLHRD-MAQ-VAMAVAR----------------VPSRL--------------MGSSPP-T
                                        Microcebus murinus LRT-----QAIVLSQLPRD-MAQLVAMAVAR----------------VPNRL--------------MGSSPL-T
                                        Bison bison bison PRT-----QATAHSQLPRD-MAQLVAMAVAR----------------VLSHL--------------MGNSPR-T
                                     Propithecus coquereli PRT-----QAIALRQLPRD-MAQLVAMAVAR----------------VPSRL--------------MGSSPP-T
                                        Bos mutus PRT-----QATAHSQLPRD-IAQLVAMAVAR----------------VLSHL--------------MGNSPR-I
                                        Pteropus vampyrus PRT-----QAITLSQLPRD-MAQQVAMA-AR----------------VLNLL--------------MGSSPP-T
                                        Pteropus alecto PRT-----QAITLSQLPRD-MAQRVAMA-AR----------------VLNLL--------------MGSSPP-T
                                        Vicugna pacos LRT-----QAMAHSQLPRV-MAQLVAMAAAK----------------VRSRL--------------MGSSPH-T
                                     Rousettus aegyptiacus PRT-----QAIALSQLPRD-MAQRVAMA-AR----------------VLNLL--------------MGSSPP-T
                                        Bos taurus PRT------ATAHSQLPRD-MAQLVAMAVAR----------------VLSHL--------------MGNSPR-T
                                        Equus caballus PRT-----QAMALSQLPRD-MAQLVDMAVAR----------------VLSRL--------------TGSSPP-T
                                     Miniopterus natalensis PRT-----QAMALSQLPRD-MAQLVAMAVDR----------------VLSRL--------------MGSSP--T
                                     Canis lupus familiaris PRT-----QAMALSQLPRD-MAQLVAMAVAR----------------VLSHL--------------MGNSPL-I
                                        Camelus ferus PRT-----QAMAHSQLPRV-MAQLVATAAAK----------------VLSRL--------------MGSSPH-T
                                        Camelus bactrianus PRT-----QAMAHSQLPRV-MAQLVATAAAK----------------VLSRL--------------MGSSPH-T
                                        Equus przewalskii PRT-----QAMALSQLPRD-MAQLVDMAVAR----------------VLSRL--------------TGSSPP-T
                                        Equus asinus LRT-----QAMALSQLPRD-MAQLVDMAVAR----------------VLSRL--------------TGSSPP-T
                                        Eptesicus fuscus PRT-----QAMAPSQLPRD-MAQPVAMVVAR----------------VLSHL--------------MGSSP--T
                                     Leptonychotes weddellii PRT-----QAMALSQVPRD-MAQPVAMAVAR----------------VLSHL--------------MGNSPL-I
                                        Octodon degus LRT-----QVMAHSQLPRV-MAPLVAMAVAR----------------VPSHL--------------MVSSPL-I
                                        Otolemur garnettii PRT-----QVIALSQLPRD-MAQLVAMAVAR----------------VLSHL--------------MGSSPL-I
                                        Myotis davidii PRT-----QAMAPSQLPRV-MAQPVAMVVAR----------------VLSHL--------------MGSSP--T
                                        Dipodomys ordii LRT-----QVMELSQLPRG-TAPLVGMAAAR----------------VPSRL--------------TGSSLP-T
                                        Jaculus jaculus PRT-----QVMALSQPPRD-MVPLVDMAVAR----------------VPN-L--------------MGSSPH-T
                                        Ochotona princeps HRT-----QAMALSQLPRV-MARLAAMVAAR----------------VPSRL--------------TDNSPP-T
                                        Microtus ochrogaster PKT-----QAMALSQLPRD-MVPLEAMAAAR----------------VPNLL--------------MGSSPP-T
                                        Mus musculus PKT-----QAMELSQLPRD-MVPLEAMAAAK----------------VHNLL--------------MGSSPP-T
                                     Peromyscus maniculatus bairdii PKT-----QAMALSQLPRD-MVPLEAMAAAK----------------VPSLL--------------MGSSPP-T
                                        Erinaceus europaeus PRT-----QAMVLSQPPRA-TAPPAAMAAVR----------------VLS-L--------------TGSSRP-T
                                        Nannospalax_galili PRT-----QAMALSRLPRD-MVPLEAMAVAK----------------VPNLL--------------MDNSPP-I
                                        Pantholops_hodgsonii PRI-----QATAHSQLPRD-MAQLVAMAVAR----------------VLSRL--------------TGNSLR-T
                                     Drosophila_melanogaster_fusilli PRV-----MPSCYSQTRRM-R----------------PRRW--------------AGTVNPLG
                                        Rattus_norvegicus PKT-----QAMELSQHPRD-MVPLVAMAAAK----------------VPSLL--------------MGSSPP-T
                                        Capra_hircus PRI-----QATARSQLPRD-MAQLVAMAVAR----------------VLSRL--------------TGNSLR-T
                                     Brachypodium_distachyon ----------VVQVLEMALYL----------------VVE--------------------------
                                        Ovis_aries PRI-----QATARSQLPRD-MAQLVAMAVAR----------------VLSRL--------------TGNSLR-T
                                        Ovis_aries_musimon PRI-----QATARSQLPRD-MAQLVAMAVAR----------------VLSRL--------------TGNSLR-T
                                        Oreochromis_niloticus ----------AAPAP---TIREAIAATAI----------------LSQVD--------------TGLHPP-T
                                        Haplochromis_burtoni ----------AAPAP---TVREAIAATAI----------------LSQVD--------------TGLHPP-T
                                        Perkinsus_marinus ----------------------------------------------------------------------
                                        Eucalyptus_grandis ----------------------------------------------------------------------
                                        Gallus_gallus ----------MAAAEGTAVGR----------------AHSHP--------------MGRPPP-T
                                        Sarcophilus_harrisii PRT-----QAMVPNRLPKD-MVQPVAMEVAR----------------VPNLL--------------MVSSHP-I
                                        Ictalurus_punctatus PHS-----PT------ARVVLTA-AAAVSLL----------------RAE--------------QRTISRPP-T
                                        Anolis_carolinensis LKT-----LAMALSPHHRP-MAHQQDMAAAK----------------LLRHP--------------MDSSHH-I
                                        Trypanosoma VKA-----KIQDKEGIPPD-QQRLIFAGKQL----------------E--------------------DGR-T
                                        Xenopus tropicalis TRV-----LDIVLSQLHKD-MD-LVVMETPN----------------HLSPLMAAGSHSSSPSPLTPATPS-S
                                        Danio rerio TAP-----TMDSPSQEDMD-LKPLLKATVSPVSRIALVATVTPASLHQLKA--------------EATASSPL-I
                                        Latimeria chalumnae HKL-----LGMEHSPRPRV-MAP-VAMEALK----------------HHR--------------TPMDNNLL-T
                                        Xenopus_laevis IRA-----LGIVLSQLHNK-GMDLEVMETPN----------------HLSPLMAA-----GS---HSSRPL-I
```

```
                                        150       160       170       180       190       200       210       220
                                         |         |         |         |         |         |         |         |
Homo sapiens                    --------LAMASSQLPAAPR-EV-T-VAVLRA---------AA-MGS-PRVGATAS-SLAMV--DSSK-----
Pan paniscus                    --------LVMASSQLPAAPR-EV-T-VAVLRA---------AA-MGS-PRVGATAS-SLAMV--DSSK-----
Gorilla gorilla gorilla         --------LAMASSQLPAAPR-EV-T-VAVLRA---------AA-MGS-PRVGATAS-SLAMV--DSSL-----
Pan troglodytes                 --------LVMASSQLPAAPR-EV-T-VAVLRA---------AA-MGS-PRVGATAS-SLAMV--DSSK-----
Nomascus leucogenys             --------LAMASSHLPAAPR-EV-T-VAVLRP---------AA-MGS-PRVGAMAS-SLAMV--DSSQ-----
Pongo abelii                    --------LAMASSQLPAAPR-EV-T-VAVLRP---------AA-MGS-PRVGATAS-SLAMV--DSSK-----
Chlorocebus sabaeus             --------LAMASSQLPAAPR-EV-T-VAVLRP---------AA-MGS-PRVGATAS-SLAMV--DSSK-----
Rhinopithecus roxellana         --------LVMASSQLPAAPR-EV-T-VAVLRP---------AA-MGS-PQVGATAS-SLAMV--DSSK-----
Rhinopithecus bieti             --------LVMASSQLPAAPR-EV-T-VAVLRP---------AA-MGS-PQVGATAS-SLAMV--DSSK-----
Macaca mulatta                  --------LAMASSQLPAAPR-EV-T-VAVLRP---------AA-MGS-PRVGATAS-SLAMV--DSSK-----
Colobusangolensispalliatus      --------LVMASSQLPAAPR-EV-T-VAVLRP---------AA-MGS-PQVGATAS-SLAMV--DSSK-----
Macaca nemestrina               --------LAMASSQLPAAPR-EV-T-VAVLRP---------AA-MGS-PRVGATAS-SLAMV--DSSK-----
Macaca fascicularis             --------LAMASSQLPAAPR-EV-T-VAVLRP---------AA-MGS-PRVGATAS-SLAMV--DSSK-----
Mandrillus leucophaeus          --------LAMASSQLPAAPR-EV-T-VAVLRP---------AA-MGS-RRVGATAS-SLAMV--DSSK-----
Cercocebus atys                 --------LAMASSQLPAAPR-EV-T-VAVLRP---------AA-MGS-PRVGATAS-SLAMV--DSSK-----
Papio anubis                    --------LAMASSQLPAAPR-EV-T-VAVLRP---------AA-MGS-PQVGATAS-SLAMV--DSSK-----
Callithrix jacchus              --------LAMVSSQLPAAPR-EV-T-GAVLRA---------AA-MGS-PRVAATAS-NLAMV--DSSK-----
Saimiri boliviensis boliviensis --------LAMASSQLPAAPR-EV-M-GAVLRA---------AA-TGSPPRVAATAS-SLAMV--DSSK-----
Cebus capucinus imitator        --------LAMASSQLPAAPR-EV-T-GAVLRA---------AA-MGSPPRVAAMAS-SLAMV--DSRK-----
Sus scrofa                      --------LAMASSQLLAAPR-EV-M-VLVRRA---------VA-MGS-PRVGAMAS-NLAMV--DSSK-----
Physeter catodon                --------LAMASSQLLAAPR-EV-T-VAVLRA---------AA-MGS-PRVGAMAS-SLAMV--DSSK-----
Tursiops truncatus              --------LAMASSQLLAAPR-EV-T-VAVLRA---------AA-MGS-PRVGAMAS-SLAMV--DSSK-----
Orcinus orca                    --------LAMASSQLLAAPR-EV-T-VAVLRA---------AA-MGS-PRVGAMAS-SLAMV--DSSK-----
Balaenoptera acutorostratas     --------LAMASSQLLAAPR-EV-T-VAVLRA---------AA-MGS-PRVGAMAS-SLAMV--DSLK-----
Sorex araneus                   --------LATASSQPRAVPR-EV-M-AAVLRA---------AA-MGS-PRVAAMAS-SLAMV--GSSK-----
Aotus nancymaae                 --------LAMASSQLRAAPR-EV-M-GAVLRA---------AA-MGSPPRVAATAS-SLAMV--YSSK-----
Trichechus manatus latirostris  --------LAMASSQLPAAPR-EA-M-EVVLRA---------AA-MGS-PRVEAMAN-SLDMV--DSSQ-----
Togninia minima                 ----------PTPSL-----------------------------------------------------------
Ceratotherium simum simum       --------LAMASNRLLAAPR-EV-T-AAVPRA---------AA-MGS-PRVVAMAS-SLAMV--DSSQ-----
Condylura cristata              --------LAMASSQLLAAPR-EV-M-GAVLRA---------AA-TDS-PRVEAMAS-SLATV--DSSK-----
Galeopterus variegatus          --------LAMVNSQLPAVPR-EV-M-VAVLRA---------AA-MGS-PQVGAMAN-SLAMV--DSSK-----
Odobenus rosmarus divergens     --------LAMASSQLLAALR-EV-T-VVVLRA---------AA-TGS-PRVGAMAS-SLPIV--DSSQ-----
Tarsius syrichta                --------LAMVSNQLPAAPR-EV-T-VAVLRA---------AA-MGS-PRVGAMAN-SLAMV--DSRK-----
Microcebus murinus              --------LAMASSRLPAAPR-EV-T-VAVLRA---------AA-TGS-PRVGAMAN-SLAMV--DSNK-----
Bison bison bison               --------LAMASSQLLAAPQ-EV-T-VAVLKA---------AA-MGS-PRVEATAS-SLAMV--GSSK-----
Propithecus coquereli           --------LAMANSRLPAAPR-EV-T-VAVPRA---------AA-MDS-PQAGAMAS-SLAMV--DSNK-----
Bos mutus                       --------LAMASSQLLAAPQ-EV-T-VAVLKA---------AA-MGS-PRVEATAS-SLAMV--GSSK-----
Pteropus vampyrus               --------LAMASSQPLAAPR-EV-M-VAVLRA---------AATMGS-PRVGVMVS-SLAMV--DSSK-----
Pteropus alecto                 --------LAMASSQPLAAPR-EV-M-VAVLRA---------AATMGS-PRVGVMVS-SLAMV--DSSK-----
Vicugna pacos                   --------LAMASSQLLVAPR-EA-T-VVVLRA---------AA-MAS-PRAGATDS-SLAMV--DSSR-----
Rousettus aegyptiacus           --------LAMASSQPLAAPR-EV-M-VAVLRA---------AATMGS-PRVGVMVS-SLAMV--DNSK-----
Bos taurus                      --------LAMASSQLLAAPQ-EV-T-VAVLKA---------AA-MGS-PRVEATAS-SLAMV--GSSK-----
Equus caballus                  --------LGMVSSQLLAAPR-EV-M-VAVLRA---------VA-MDS-PKVEAMAS-SLAMV--DSSQ-----
Miniopterus natalensis          --------LAMASSQLLAAPL-EV-T-VAVLRA---------AT-MGS-PRVGAMVS-SLAMV--DSSL-----
Canis lupus familiaris          --------LAMASSQLLAAPR-EV-M-VVVLRA---------AA-TGS-PRVGAMAN-SLAIV--DSSK-----
Camelus ferus                   --------LAMASSQLLVAPR-EA-T-VAVLRA---------AA-MAS-PRAGATAS-SLAMV--DSSR-----
Camelus bactrianus              --------LAMASSQLLVAPR-EA-T-VAVLRA---------AA-MAS-PRAGATAS-SLAMV--DSSR-----
Equus przewalskii               --------LGMVSSQLLAAPR-EV-M-VAVLRA---------VA-MDS-PKVEAMAS-SLAMV--DSSQ-----
Equus asinus                    --------LGMVSSQLLAAPR-EV-M-VAVLRA---------VA-MDS-PKVEAMAS-SLAMV--DSSQ-----
Eptesicus fuscus                --------LAMASSQLLAAPR-EV-T-VAVLRA---------AT-MGS-PRVGAMVS-SLAMV--DSSL-----
Leptonychotes weddellii         --------LAMASSQLLAAPR-EV-M-VVLLRA---------AA-TGS-PRVGAMAS-SRPIV--DSSK-----
Octodon degus                   --------LAMVSSQLPAAPR-EV-M-VVVLRA---------AA-MGS-PRVGAMAN-SLAMV--DSNR-----
Otolemur gamettii               --------LAMASSQLLAAPR-EV-T-AAVLRA---------AA-MDS-PQVGAMAN-SPAMA--DSSR-----
Myotis davidii                  --------PAMASSPLLAARR-EV-T-VAVLRA---------AT-MGS-PRVGAMVS-SLAM---DSSL-----
Dipodomys ordii                 --------LAMASSQLPAAPL-EV-M-VAVPRA---------AA-MGS-LKVGATAN-SLAME--DSNR-----
Jaculus jaculus                 --------LAMASSQLLAAPQ-EV-M-VAVLRA---------AA-MGS-LRVGAMAN-SLGMA--DSNK-----
Ochotona princeps               --------LAMGSSQLPAAPQ-EV-T-AAVLRA---------AA-TGS-PRAGAMAS-NLAMA--DSNK-----
Microtus ochrogaster            --------LAMANSQLLAAPR-EV-M-VAVLRA---------AA-MGS-PRVEAMAN-SPAMV--DSNK-----
Mus musculus                    --------LAMANSQLLAAPQ-EV-M-VAVLRA---------AA-MGN-PRVEVMVN-SLAMV--DSNK-----
Peromyscus maniculatus bairdii  --------LAMANSQLLAAPR-EV-T-VAVLRA---------AA-MGS-PRVEAMAN-SQAMV--DSNK-----
Erinaceus europaeus             --------PAMASSRLLATPL-EV-T-VAALRA---------AA-MGS-PRAGATGS-SRATA--DSSK-----
Nannospalax_galili              --------LAMANNQLLAALQ-EV-T-VGALRA---------AA-MGS-PRVEAMAN-SQAMV--DSNR-----
Pantholops_hodgsonii            --------LAMVSSQLLAAPQ-EV-T-VAVLKA---------AA-MGS-PRVEAMAS-SLAMV--GSSK-----
Drosophila_melanogaster_fusilli SGTLSYSV--------PRPPR-C-------------------------------------------------
Rattus_norvegicus               --------PVMANSQLLVAPQ-EV-M-VEVLRA---------AA-MGS-PRVEAMVN-SLAMV--DSNK-----
Capra_hircus                    --------LAMASSQLLAAPQ-EV-T-VAVLKA---------AA-MGS-PRVEAMAS-SLAMV--GSSK-----
Brachypodium_distachyon         ----------MGTGREAPAPH-SV-E-VEV-----------------------GAVATETVSMV--MTQR-----
Ovis_aries                      --------LAMASSQLLAAPQ-EV-T-VAVLKA---------AA-MGS-PRVEAMAS-SLAMV--GSSK-----
Ovis_aries_musimon              --------LAMASSQLLAAPQ-EV-T-VAVLKA---------AA-MGS-PRVEAMAS-SLAMV--GSSK-----
Oreochromis_niloticus           --------RAV-----------------A-V-------------------------------------------
Haplochromis_burtoni            --------RVV-----------------VVV-------------------------------------------
Perkinsus_marinus               ---------------------R----G-IAIVRA---------AI-MGV-PRTTTTRV-GVALGA--VTG----V
Eucalyptus_grandis              ------------------------Q-ILIVRR---------V-AVNF-LAEGSLVG-S----------------
Gallus_gallus                   --------PATTSRPPPAPPP-EA-M-GAAPRA---------PA-TGS-RPAPATAS-SPHTA--ASSP-----
Sarcophilus_harrisii            --------LATANSQLLAALQ-EA-M-VAVPRA---------VA-MDS-PRVGAMVR-SLVMV--DNKV-----
Ictalurus_punctatus             --------RPTASSSRRLLPQLQA-S-PAALG---------HQA-TGS-REEAG-------------------
Anolis_carolinensis             --------PATPSSQ-PVPLL-EV-M-APALKP---------PA-MGN-PRVEVMAS-NQAMA--ASSR-----
Trypanosoma                     --------LADYNIQ-----K-ES-TLHLVLRLR-------GG-V-ME-PTLEALAK-KYNW----EKK-----
Xenopus tropicalis              --------LQPVATQEVAAPS-HP-A-TSSSRV---------VD-MGS-NQVGVMEV-SRAAMEVNSNP----V
Danio rerio                     --------LAITS-LLLHLPL-VAIA-AALSR---------LA-TVSSSSRVEV--G-MEAVV--VSLEDMGAV
Latimeria chalumnae             --------LVTASRPPLRPPQ-PA-T-EEVAPS--------LPA-TGS-PRVEPMDR-ILLTA--SSSR-----
Xenopus_laevis                  --------LATRSSLQPVATQ-EV-M-VGAPSP-LPTSSNKVVD-MDS-RQVGVMEA-SRAAMEVNNNP----V
```


230
240
250
260
270
280
290
Homo sapiens
Pan paniscus
Gorilla gorilla gorilla
Pan troglodytes
Nomascus leucogenys
Pongo abelii
Chlorocebus sabaeus
Rhinopithecus roxellana
Rhinopithecus bieti
Macaca mulatta
Colobusangolensispalliatus
Macaca nemestrina
Macaca fascicularis
Mandrillus leucophaeus
Cercocebus atys
Papio anubis
Callithrix jacchus
Saimiri boliviensis boliviensis
Cebus capucinus imitator
Sus scrofa
Physeter catodon
Tursiops truncatus
Orcinus orca
Balaenoptera acutorostratas
Sorex araneus
Aotus nancymaae
Trichechus manatus latirostris
Togninia minima
Ceratotherium simum simum
Condylura cristata
Galeopterus variegatus
Odobenus rosmarus divergens
Tarsius syrichta
Microcebus murinus
Bison bison bison
Propithecus coquereli
Bos mutus
Pteropus vampyrus
Pteropus alecto
Vicugna pacos
Rousettus aegyptiacus
Bos taurus
Equus caballus
Miniopterus natalensis
Canis lupus familiaris
Camelus ferus
Camelus bactrianus
Equus przewalskii
Equus asinus
Eptesicus fuscus
Leptonychotes weddellii
Octodon degus
Otolemur garnettii
Myotis davidii
Dipodomys ordii
Jaculus jaculus
Ochotona princeps
Microtus ochrogaster
Mus musculus
Peromyscus maniculatus bairdii
Erinaceus europaeus
Nannospalax galili
Pantholops hodgsonii
Drosophila melanogaster fusilli
Rattus norvegicus
Capra hircus
Brachypodium distachyon
Ovis aries
Ovis aries musimon
Oreochromis niloticus
Haplochromis burtoni
Perkinsus marinus
Eucalyptus grandis
Gallus gallus
Sarcophilus harrisii
Ictalurus punctatus
Anolis carolinensis
Trypanosoma
Xenopus tropicalis
Danio rerio
Latimeria chalumnae
Xenopus laevis


300
310
320
330
340
350
360
Homo sapiens
Pan paniscus
Gorilla gorilla gorilla
Pan troglodytes
Nomascus leucogenys
Pongo abelii
Chlorocebus sabaeus
Rhinopithecus roxellana
Rhinopithecus bieti
Macaca mulatta
Colobusangolensispalliatus
Macaca nemestrina
Macaca fascicularis
Mandrillus leucophaeus
Cercocebus atys
Papio anubis
Callithrix jacchus
Saimiri boliviensis boliviensis
Cebus capucinus imitator
Sus scrofa
Physeter catodon
Tursiops truncatus
Orcinus orca
Balaenoptera acutorostratas
Sorex araneus
Aotus nancymaae
Trichechus manatus latirostris
Togninia minima
Ceratotherium simum simum
Condylura cristata
Galeopterus variegatus
Odobenus rosmarus divergens
Tarsius syrichta
Microcebus murinus
Bison bison bison
Propithecus coquereli
Bos mutus
Pteropus vampyrus
Pteropus alecto
Vicugna pacos
Rousettus aegyptiacus
Bos taurus
Equus caballus
Miniopterus natalensis
Canis lupus familiaris
Camelus ferus
Camelus bactrianus
Equus przewalskii
Equus asinus
Eptesicus fuscus
Leptonychotes weddellii
Octodon degus
Otolemur garnettii
Myotis davidii
Dipodomys ordii
Jaculus jaculus
Ochotona princeps
Microtus ochrogaster
Mus musculus
Peromyscus maniculatus bairdii
Erinaceus europaeus
Nannospalax galili
Pantholops hodgsonii
Drosophila melanogaster fusilli
Rattus norvegicus
Capra hircus
Brachypodium distachyon
Ovis aries
Ovis aries musimon
Oreochromis niloticus
Haplochromis burtoni
Perkinsus marinus
Eucalyptus grandis
Gallus gallus
Sarcophilus harrisii
Ictalurus punctatus
Anolis carolinensis
Trypanosoma
Xenopus tropicalis
Danio rerio
Latimeria chalumnae
Xenopus laevis

380
390
400
410
420
430
440
Homo sapiens
Pan paniscus
Gorilla gorilla gorilla
Pan troglodytes
Nomascus leucogenys
Pongo abelii
Chlorocebus sabaeus
Rhinopithecus roxellana
Rhinopithecus bieti
Macaca mulatta
Colobusangolensispalliatus
Macaca nemestrina
Macaca fascicularis
Mandrillus leucophaeus
Cercocebus atys
Papio anubis
Callithrix jacchus
Saimiri boliviensis boliviensis
Cebus capucinus imitator
Sus scrofa
Physeter catodon
Tursiops truncatus
Orcinus orca
Balaenoptera acutorostratas
Sorex araneus
Aotus nancymaae
Trichechus manatus latirostris
Togninia minima
Ceratotherium simum simum
Condylura cristata
Galeopterus variegatus
Odobenus rosmarus divergens
Tarsius syrichta
Microcebus murinus
Bison bison bison
Propithecus coquereli
Bos mutus
Pteropus vampyrus
Pteropus alecto
Vicugna pacos
Rousettus aegyptiacus
Bos taurus
Equus caballus
Miniopterus natalensis
Canis lupus familiaris
Camelus ferus
Camelus bactrianus
Equus przewalskii
Equus asinus
Eptesicus fuscus
Leptonychotes weddellii
Octodon degus
Otolemur garnettii
Myotis davidii
Dipodomys ordii
Jaculus jaculus
Ochotona princeps
Microtus ochrogaster
Mus musculus
Peromyscus maniculatus bairdii
Erinaceus europaeus
Nannospalax_galili
Pantholops_hodgsonii
Drosophila_melanogaster_fusilli
Rattus_norvegicus
Capra_hircus
Brachypodium_distachyon
Ovis_aries
Ovis_aries_musimon
Oreochromis_niloticus
Haplochromis_burtoni
Perkinsus_marinus
Eucalyptus_grandis
Gallus_gallus
Sarcophilus_harrisii
Ictalurus_punctatus
Anolis_carolinensis
Trypanosoma
Xenopus tropicalis
Danio rerio
Latimeria chalumnae
Xenopus_laevis

```
                                    450       460       470
Homo sapiens                        ---------------AKI----------------NPP
Pan paniscus                        ---------------AKI----------------NPP
Gorilla gorilla gorilla             ---------------AKI----------------NPP
Pan troglodytes                     ---------------AKI----------------NPP
Nomascus leucogenys                 ---------------AKI----------------NPP
Pongo abelii                        ---------------AKI----------------NPP
Chlorocebus sabaeus                 ---------------AKI----------------NPP
Rhinopithecus roxellana             ---------------AKI----------------NPP
Rhinopithecus bieti                 ---------------AKI----------------NPP
Macaca mulatta                      ---------------AKI----------------NPP
Colobusangolensispalliatus          ---------------AKI----------------SPP
Macaca nemestrina                   ---------------AKI----------------NPP
Macaca fascicularis                 ---------------AKI----------------NPP
Mandrillus leucophaeus              ---------------AKI----------------NPP
Cercocebus atys                     ---------------AKI----------------NPP
Papio anubis                        ---------------AKI----------------NPP
Callithrix jacchus                  ---------------AKI----------------SPP
Saimiri boliviensis boliviensis     ---------------AKI----------------SPP
Cebus capucinus imitator            ---------------AKI----------------SPP
Sus scrofa                          ---------------AKI----------------SPP
Physeter catodon                    ---------------AKI----------------SPP
Tursiops truncatus                  ---------------AKI----------------SPP
Orcinus orca                        ---------------AKI----------------SPP
Balaenoptera acutorostratas         ---------------AKI----------------SPP
Sorex araneus                       ---------------AKI----------------SPP
Aotus nancymaae                     ---------------AKI----------------SPP
Trichechus manatus latirostris      ---------------AKI----------------SPP
Togninia minima                     ---------------TMM----------------MPP
Ceratotherium simum simum           ---------------AKI----------------SPP
Condylura cristata                  ---------------AKI----------------NPP
Galeopterus variegatus              ---------------AKI----------------SPP
Odobenus rosmarus divergens         ---------------AKI----------------SPP
Tarsius syrichta                    ---------------AKI----------------SLP
Microcebus murinus                  ---------------AKI----------------SPP
Bison bison bison                   ---------------AKI----------------SPP
Propithecus coquereli               ---------------AKI----------------SPP
Bos mutus                           ---------------AKI----------------SPP
Pteropus vampyrus                   ---------------VKI----------------NPQ
Pteropus alecto                     ---------------VKI----------------NPQ
Vicugna pacos                       ---------------AKT----------------SPP
Rousettus aegyptiacus               ---------------VKI----------------NPP
Bos taurus                          ---------------AKI----------------SPP
Equus caballus                      ---------------AKI----------------SPP
Miniopterus natalensis              ---------------AKI----------------SPP
Canis lupus familiaris              ---------------AKI----------------SPP
Camelus ferus                       ---------------AKT----------------SPP
Camelus bactrianus                  ---------------AKT----------------SPP
Equus przewalskii                   ---------------AKI----------------SPP
Equus asinus                        ---------------AKI----------------SPP
Eptesicus fuscus                    ---------------AKI----------------SPP
Leptonychotes weddellii             ---------------AKI----------------SPP
Octodon degus                       ---------------AKI----------------SPP
Otolemur garnettii                  ---------------AKI----------------SPP
Myotis davidii                      ---------------AKI----------------SPP
Dipodomys ordii                     ---------------AKI----------------SPP
Jaculus jaculus                     ---------------VKI----------------SPP
Ochotona princeps                   ---------------AKT----------------SPL
Microtus ochrogaster                ---------------AKI----------------SPP
Mus musculus                        ---------------AKI----------------SPL
Peromyscus maniculatus bairdii      ---------------AKI----------------SPP
Erinaceus europaeus                 ---------------AKT----------------SPP
Nannospalax_galili                  ---------------AKI----------------SPP
Pantholops_hodgsonii                ---------------AKI----------------SPP
Drosophila_melanogaster_fusilli     ---------------GTA----------------SRR
Rattus_norvegicus                   ---------------AKI----------------SPL
Capra_hircus                        ---------------AKI----------------SPP
Brachypodium_distachyon             ---------------------------------ALG
Ovis_aries                          ---------------AKI----------------SPP
Ovis_aries_musimon                  ---------------AKI----------------SPP
Oreochromis_niloticus               ---------------ATS----------------RPA
Haplochromis_burtoni                ---------------ATS----------------RPA
Perkinsus_marinus                   ---------------RTT----------------ITP
Eucalyptus_grandis                  ---------------V--AGVLMVLDWVLGHSEVNP
Gallus_gallus                       ---------------GRS----------------RPR
Sarcophilus_harrisii                ---------------AKI----------------NHL
Ictalurus_punctatus                 ---------------------------------TAE
Anolis_carolinensis                 ---------------ARI----------------SPL
Trypanosoma                         ---------------MKK----------------KLR
Xenopus tropicalis                  LTTSNKLESL-R-LTRRL----------------GSL
Danio rerio                         ---------------TKI----------------AHQ
Latimeria chalumnae                 -------------IKEDQG----------------MMQ
Xenopus_laevis                      ---------LRTTSNRLA----------------SSR
```