



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](mailto: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](mailto:marie.brunet@usherbrooke.ca)

27

28 **ABSTRACT:**

29 Novel functional coding sequences (altORFs) are camouflaged within annotated ones (CDS) in a  
30 different reading frame. We discovered an altORF nested in the FUS CDS encoding a conserved  
31 169 amino acid protein, altFUS. AltFUS is endogenously expressed in human tissues, notably in  
32 the motor cortex and motor neurons. Overexpression of wild-type FUS and/or amyotrophic lateral  
33 sclerosis-linked FUS mutants is known to trigger toxic mechanisms in different models. These  
34 include an inhibition of autophagy, loss of mitochondrial potential, and accumulation of  
35 cytoplasmic aggregates. We show here that altFUS, not FUS, is responsible for the inhibition of  
36 autophagy. AltFUS is also pivotal in the mechanisms leading to the mitochondrial potential loss  
37 and accumulation of cytoplasmic aggregates. Suppression of altFUS expression in a *Drosophila*  
38 model of *FUS*-related toxicity protects against neurodegeneration. Some mutations found in ALS  
39 patients are overlooked because of their synonymous effect on the FUS protein. Yet we showed  
40 they exert a deleterious effect via their missense consequence on the overlapping altFUS protein.  
41 These findings demonstrate that *FUS* is a bicistronic gene and suggest that both proteins, FUS and  
42 altFUS, cooperate in toxic mechanisms.

43

44 **MAIN**

45 FUS is a nuclear RNA-binding protein, with a C-terminal nuclear localization signal (NLS)<sup>1,2</sup>. The  
46 protein is involved in RNA processing, DNA repair and cellular proliferation, although its functions  
47 are not precisely elucidated<sup>1</sup>. 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<sup>1</sup>. 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<sup>3-5</sup>. 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<sup>6-10</sup>. The mechanism of the wild-type or ALS-linked mutated FUS toxicity remains  
55 unclear<sup>1,11,12</sup>.

56 With currently non-annotated proteins being increasingly reported<sup>13-16</sup>, we hypothesized that the  
57 toxicity resulting from wild-type FUS over-expression may come from another, unseen, actor<sup>16</sup>.  
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<sup>15-17</sup>.  
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<sup>16</sup>. For example, mass spectrometry-based proteomics has become the gold standard  
64 for protein identification and has been extensively used in ALS studies<sup>18,19</sup>. 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<sup>16,20</sup>.

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<sup>21</sup> and the OpenProt database<sup>17</sup>, which offer a data-driven broader  
72 view of eukaryotic proteomes. The OpenProt database is based on a polycistronic model of ORF  
73 annotation<sup>17</sup> 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](http://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<sup>22,23</sup>.

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

80 We began by querying OpenProt<sup>17</sup> 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<sup>24</sup> 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**)<sup>25</sup>. 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<sup>26</sup>, 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<sup>( $\emptyset$ )</sup>.  
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<sup>(∅)</sup> 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)<sup>27,28</sup> 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  $\beta$ -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<sup>29–32</sup>. 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<sup>33–35</sup>, 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<sup>34,35</sup>. This mutant  
169 leads to a premature stop codon before FUS NLS and is linked to severe fALS and sALS cases<sup>1</sup>. In  
170 this construct, altFUS is still present and not affected by the mutation (**Fig EV4B**). Similar to FUS<sup>(∅)</sup>,  
171 we also generated the monocistronic construct FUS<sup>(∅)</sup>-R495x, which contains synonymous  
172 mutations for FUS-R495x, but prevents altFUS expression. V5-FUS<sup>(∅-FLAG)</sup> and V5-FUS<sup>(∅-FLAG)</sup>-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<sup>34</sup>,  
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<sup>(∅)</sup> or FUS<sup>(∅)</sup>-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<sup>36</sup> (**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<sup>37</sup>. 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<sup>10,11,35,38</sup>.

#### 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 alt*FUS*. 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 alt*FUS* displayed a significant accumulation of yellow foci (**Fig 4A**). Furthermore,  
208 our results were consistent with previously published data<sup>39</sup> 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 alt*FUS* expression  
211 (*FUS*<sup>(∅)</sup> or *FUS*<sup>(∅)</sup>-R495x). We used bafilomycin followed by LC3 probing to further validate the  
212 impact of alt*FUS* on autophagy (**Fig 4B, Fig EV6C**). Similarly, an inhibition of autophagy was  
213 observed only in cells over-expressing alt*FUS*. Furthermore, in cells over-expressing monocistronic  
214 *FUS*<sup>(∅)</sup>-R495x, the inhibition of autophagy could be restored by co-transfecting alt*FUS* (**Fig 4B, Fig**  
215 **EV6C**). These results establish alt*FUS*, rather than *FUS*, as the protein responsible of the inhibition  
216 of autophagy.

217 Furthermore, alt*FUS* 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<sup>40,41</sup>. In  
219 *FUS*-linked ALS and some sALS cases, *FUS* cytoplasmic aggregates or mislocalization are also

220 observed<sup>42-44</sup>. 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<sup>42,44-49</sup>. In cells over-  
224 expressing the monocistronic FUS<sup>(∅)</sup>-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<sup>(∅)</sup>-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<sup>40</sup> (**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<sup>(∅)</sup> and FUS<sup>(∅)</sup>-R495x, constructs. We used the  
237 Elav-GeneSwitch-GAL4 Driver strain, as previously described<sup>50</sup>, 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<sup>50,51</sup>. First, we generated flies containing the sequences for UAS-  
240 altFUS, UAS-FUS, UAS-FUS<sup>(∅)</sup>, UAS-FUS-R495x or UAS-FUS<sup>(∅)</sup>-R495x. These flies were then  
241 crossed with the Elav-GeneSwitch-GAL4 driver strain (**Fig 5A**). UAS-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<sup>52</sup>. 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<sup>53,54</sup> (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<sup>8</sup>, 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<sup>( $\emptyset$ )</sup> (**Fig 5D**)  
258 and FUS<sup>( $\emptyset$ )</sup>-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<sup>( $\emptyset$ )</sup>, 30% vs. 70% for FUS<sup>( $\emptyset$ )</sup>-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<sup>1</sup>. 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<sup>(P31L-FLAG)</sup>-S44=; GFP-FUS<sup>(A38V-FLAG)</sup>-G51=; GFP-FUS<sup>(A46V-FLAG)</sup>-G59=  
276 and GFP-FUS<sup>(R64P-FLAG)</sup>-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<sup>16</sup>. 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<sup>22,23</sup>. 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<sup>55</sup>. 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<sup>56</sup>. Our  
297 discovery thus suggests that *FUS* edited cells or models, notably targeting its last exons<sup>57-59</sup>, 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<sup>10,39</sup>, 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<sup>44</sup>, it has already been reported<sup>42,44-49</sup> 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<sup>38,60</sup>. 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<sup>38</sup>, 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<sup>61</sup>. 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<sup>51</sup>. 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*<sup>51</sup>. 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<sup>16</sup>. 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<sup>62</sup>. 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<sup>63</sup>.  
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<sup>41</sup>.

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 FTLT.

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 (GKPIPPLLGLDST) 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<sup>(∅)</sup> and FUS<sup>(∅)</sup>-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<sup>(FLAG)</sup> and FUS<sup>(FLAG)</sup>-R495x when altFUS is FLAG-tagged in the +2 reading frame. The  
372 monocistronic constructs are named as follows throughout the article: FUS<sup>(∅)</sup> or FUS<sup>(∅)</sup>-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<sup>64</sup>. 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<sup>65</sup>. 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<sup>13</sup>. 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 Clustalw 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<sup>27</sup>.  
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,  $\beta$ -mercaptoethanol (Life Technologies) and insulin (Sigma).  
439 Treatment with the following small molecules from day 0-7: 1 $\mu$ M Dorsomorphin (Millipore), 2 $\mu$ M  
440 SB431542 (Tocris Bioscience), and 3.3 $\mu$ M CHIR99021 (Miltenyi Biotec). At day 8, cells patterned  
441 for 7 days with 0.5 $\mu$ M retinoic acid and 1 $\mu$ M Purmorphamine. At day 14 spinal cord motor neuron  
442 precursors were treated with 0.1 $\mu$ M Purmorphamine for a further 4 days before being terminally  
443 differentiated for >10 days in 0.1  $\mu$ M 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<sup>-1</sup>).

#### 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<sup>56</sup>. 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<sup>6</sup> 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 cytoplasmic 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<sub>2</sub> 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<sub>2</sub>. 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<sub>2</sub>O 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<sub>4</sub>HCO<sub>3</sub> (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<sup>66</sup>. 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 ( $\mu\text{m}^2$ ) 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<sup>( $\emptyset$ )</sup>  
575 and FUS<sup>( $\emptyset$ )</sup>-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  $y[1] sc[*] v[1]; P\{y[+7.7] v[+t1.8]=UAS-mCherry.VALIUM10\}attP2$ ) 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  $\mu\text{M}$  diluted in ethanol (induced flies). The climbing assay was performed as  
586 previously described<sup>52</sup>. 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

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

596

#### 597 **DATA AVAILABILITY STATEMENT**

598 The OpenProt database is available at [www.openprot.org](http://www.openprot.org). The GTEx portal is available via  
599 [www.gtexportal.org](http://www.gtexportal.org). The Gwips portal is available at [www.gwips.ucc.ie](http://www.gwips.ucc.ie). The proteomics data are  
600 available on the PRIDE repository with the accession PXD----- . Any other relevant data are  
601 available from the corresponding authors upon reasonable request.

#### 602 **ACKNOWLEDGEMENTS**

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

#### 613 **CONTRIBUTIONS**

614 MA.B and X.R designed and wrote the study. MA.B and JF.J did the experiments, and MA.B did  
615 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 Crielinge, 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. Farrowell, 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-Murguía, 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. Scekcic-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- $\pi$  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 alt*FUS* 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 alt*FUS* 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 Fourier 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 alt*FUS* CDS by a green rectangle. **D**, Alignment  
788 (Clustalw) of alt*FUS* 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 alt*FUS* 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 alt*FUS*. 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 alt*FUS* 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 alt*FUS* 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 alt*FUS* or its isoform.

805

#### **806 Figure 2: alt*FUS*, a novel endogenous mitochondrial protein**

807 **A**, Expression of both *FUS* and alt*FUS* proteins from transfection of the *FUS* cDNA in HEK293 cells  
808 by western blot, and expression of *FUS* with the monocistronic construct *FUS*<sup>( $\emptyset$ )</sup> (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  $\mu\text{m}$ . **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 10 $\mu\text{m}$ . **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*<sup>( $\Delta$ )</sup>-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*<sup>( $\Delta$ )</sup>, *FUS*-R495x or *FUS*<sup>( $\Delta$ )</sup>-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 3 $\mu\text{m}$ . **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<sup>(∅)</sup>, FUS-R495x and FUS<sup>(∅)</sup>-  
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<sup>(∅)</sup>, FUS-R495x, FUS<sup>(∅)</sup>-R495x transfected cells and FUS<sup>(∅)</sup>-  
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<sup>(FLAG)</sup>-R495x or the monocistronic GFP-FUS<sup>(∅-FLAG)</sup>-R495x constructs, or co-  
874 transfected with the monocistronic GFP-FUS<sup>(∅-FLAG)</sup>-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<sup>2</sup>) (**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<sup>(∅)</sup>, FUS-R495x or FUS<sup>(∅)</sup>-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<sup>(∅)</sup> (**D**), and the bicistronic FUS-R495x or the  
891 monocistronic FUS<sup>(∅)</sup>-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<sup>(FLAG)</sup> or GFP-FUS<sup>(P31L-FLAG)</sup>-S44= (representative images  
903 from n=3). The white scale bar corresponds to 10  $\mu$ m. **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<sup>(FLAG)</sup> expressing cells. Statistical significance is indicated above the bars  
906 (n=3, \*\*\*\* = p value < 0.0001).

907

908

909

\*\*\*Extended View Figures\*\*\*

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 (Clustalw) 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<sub>∅</sub>). **C-D**, Protein alignment of altFUS (**C**) and FUS (**D**) proteins from FUS bicistronic  
953 construct and the monocistronic (FUS<sup>(∅)</sup>) 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<sup>(∅)</sup> and FUS<sup>(∅)</sup>-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<sup>(∅)</sup> or FUS<sup>(∅)</sup>-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<sup>(∅-FLAG)</sup>,  
984 FUS-R495x and FUS<sup>(∅-FLAG)</sup>-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<sup>(∅)</sup>, FUS-R495x, FUS<sup>(∅)</sup>-R495x, or co-transfected with FUS<sup>(∅)</sup>-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<sup>(FLAG)</sup>-R495x or the monocistronic GFP-FUS<sup>(∅-FLAG)</sup>-R495x  
997 constructs, or co-transfected with the monocistronic GFP-FUS<sup>(∅-FLAG)</sup>-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<sup>(A38V-FLAG)</sup>-G51=, GFP-FUS<sup>(A46V-FLAG)</sup>-G59= or GFP-FUS<sup>(R64P-FLAG)</sup>-S77=  
1013 constructs (representative images from n=3). The white scale bar corresponds to 10 μm.

1014

1015

\*\*\***Appendix Tables**\*\*\*

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 alt*FUS* (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: alt*FUS* protein sequences across species**

1027 Alt*FUS* 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 alt*FUS* sequence (transcript accession, species, and alt*FUS* protein  
1030 sequence); and the matrix of alt*FUS* protein sequence identity between each species.

1031

1032 **Appendix Table 3: Peptide-Centric proteomic results for alt*FUS* 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 alt*FUS* sequence**

1039 Predictions for post-translational modifications (PTMs) on alt*FUS* 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: Alt*FUS* 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 alt*FUS*-  
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

**\*\*\*Source Data\*\*\***

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 Clustal $\omega$ , and coloured based on conservation from white (highly variable residue) to dark blue  
1065 (highly conserved residue).

1066



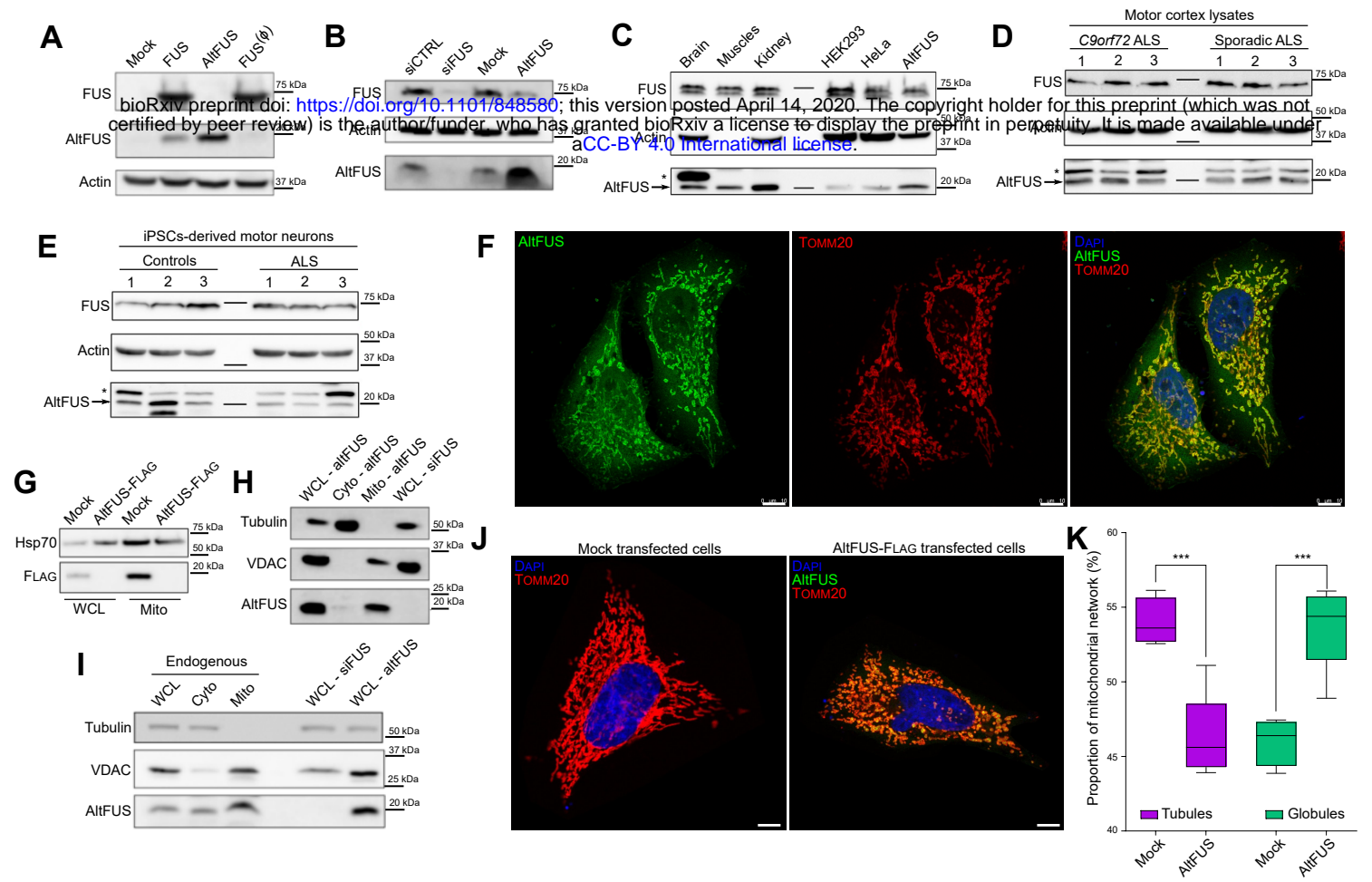


Figure 2

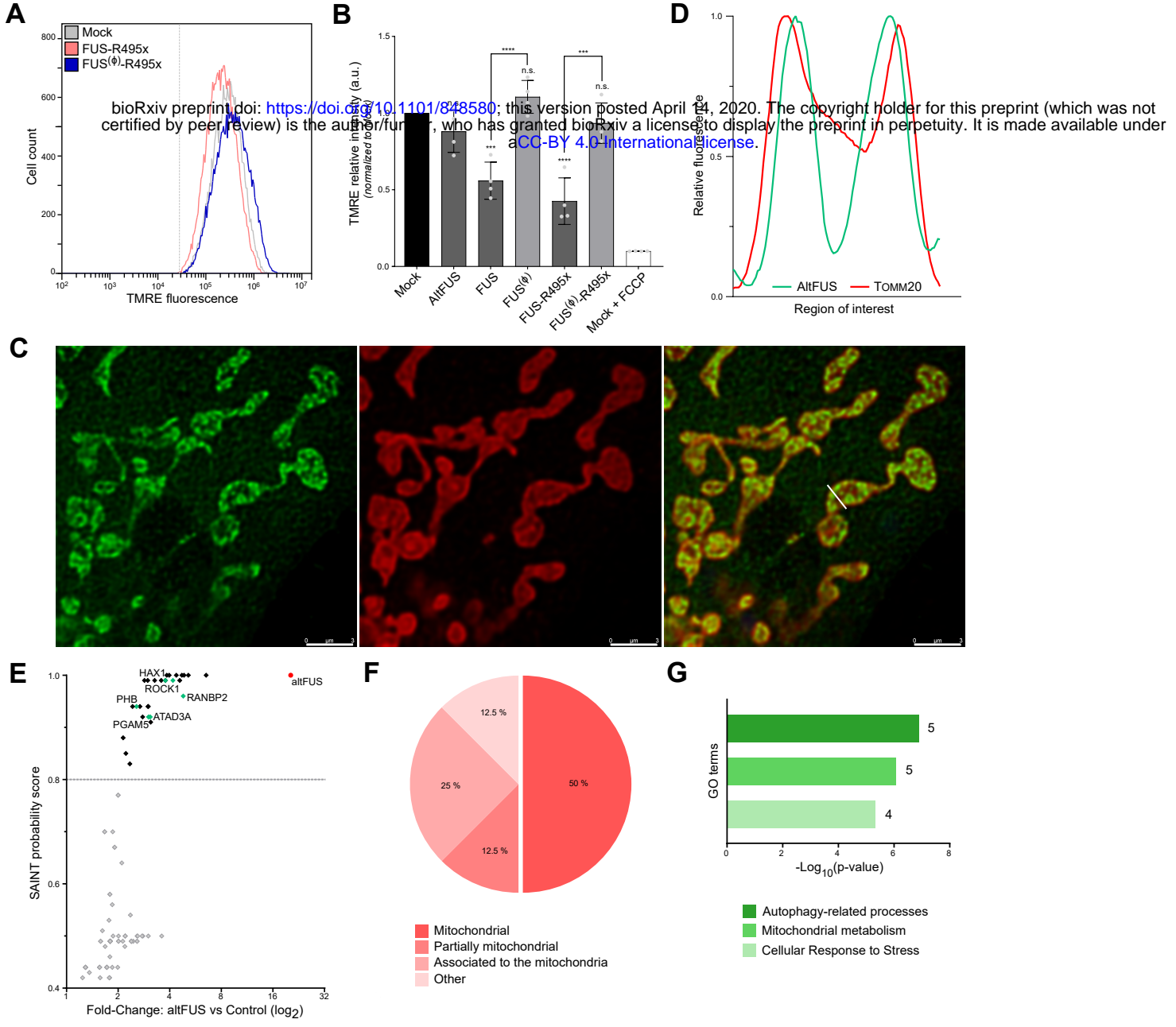


Figure 3

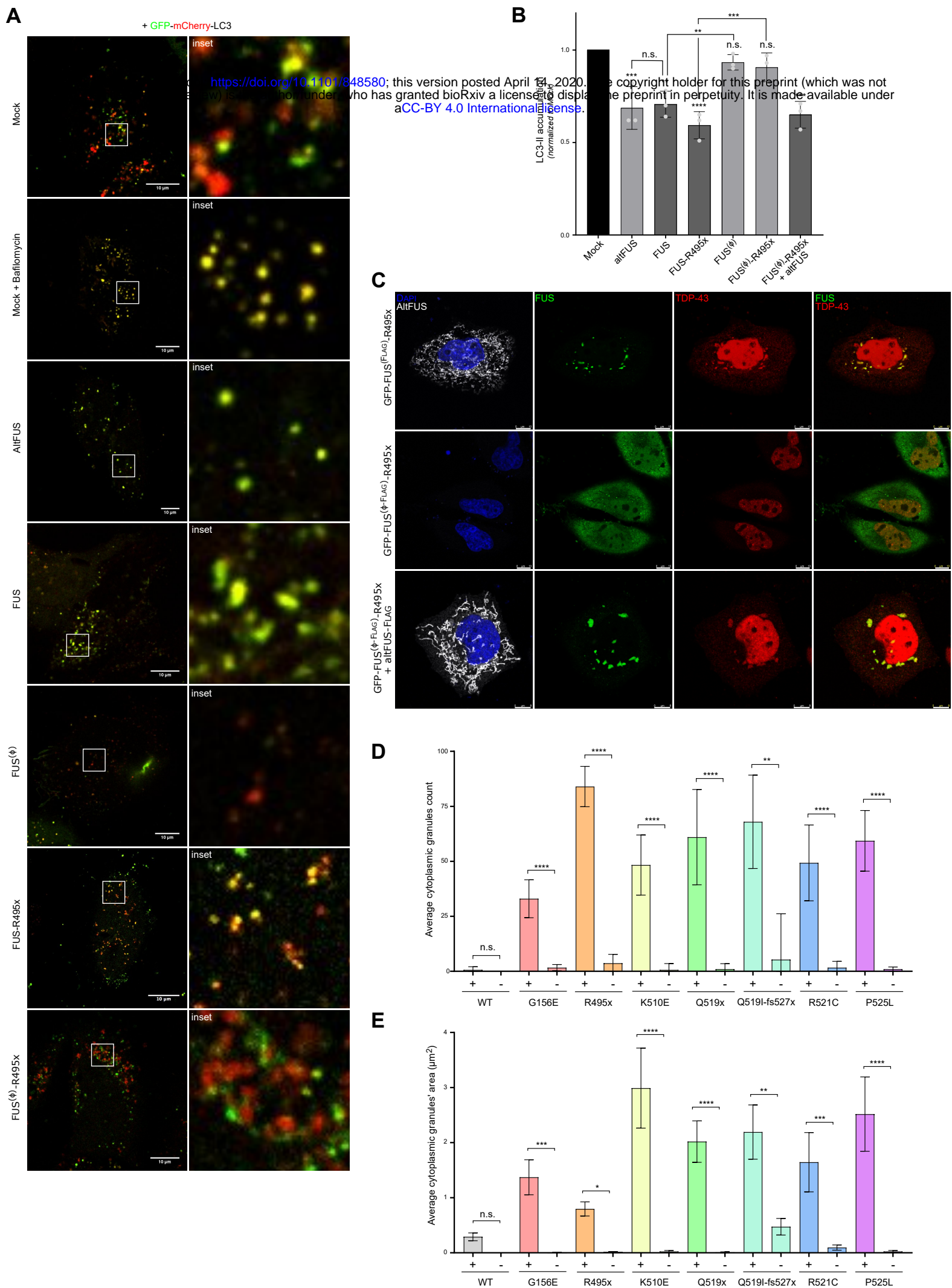


Figure 4



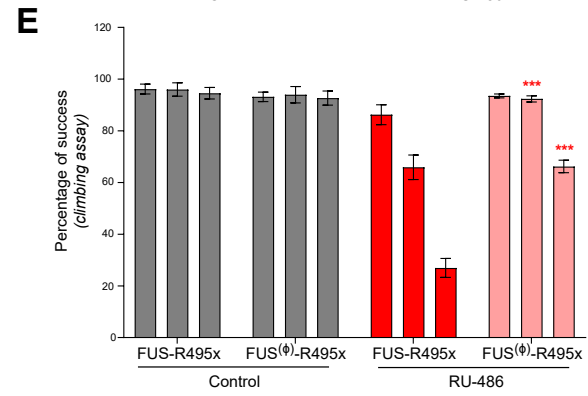
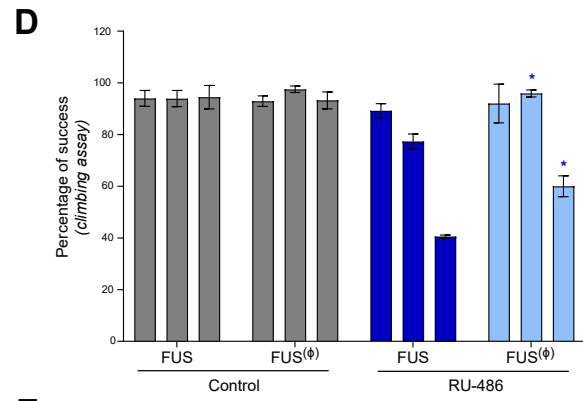
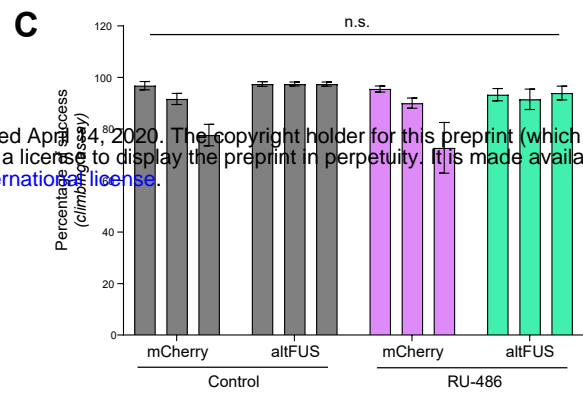
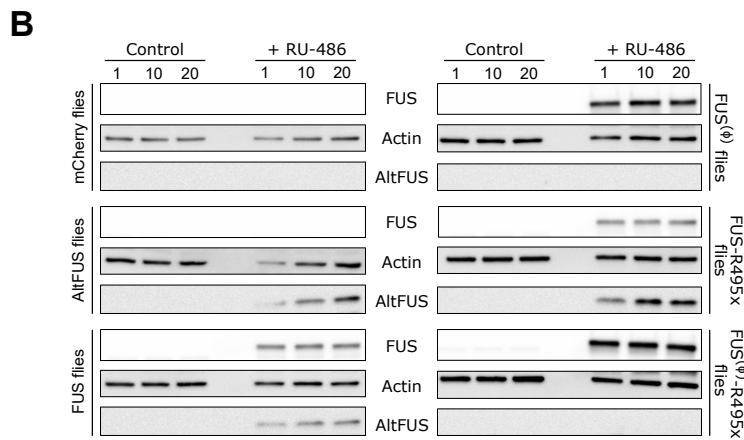
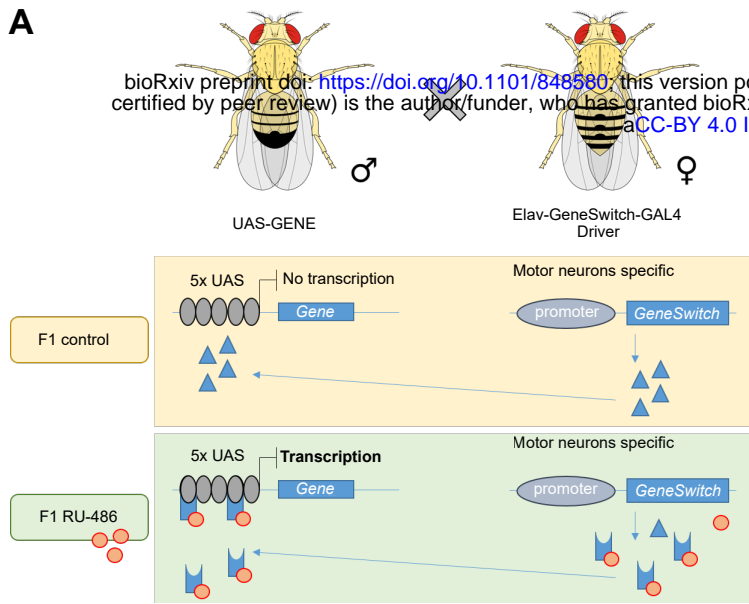


Figure 5

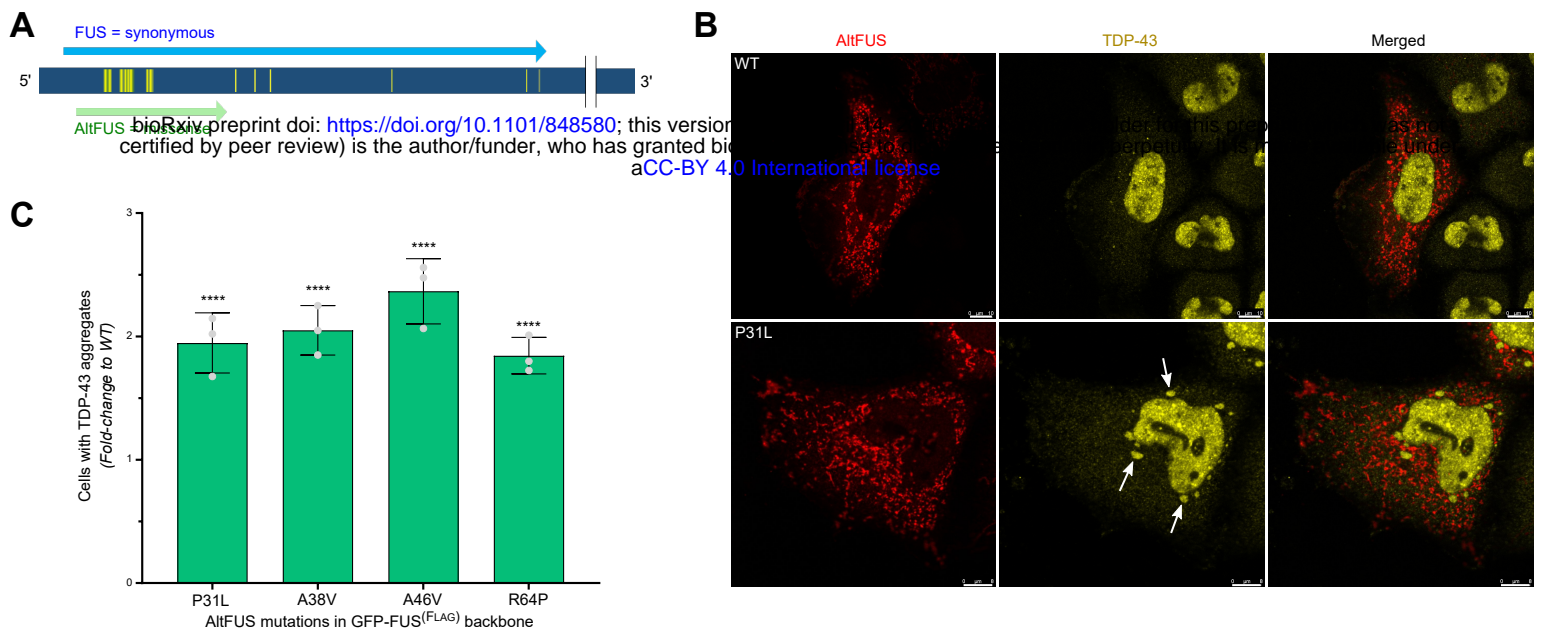
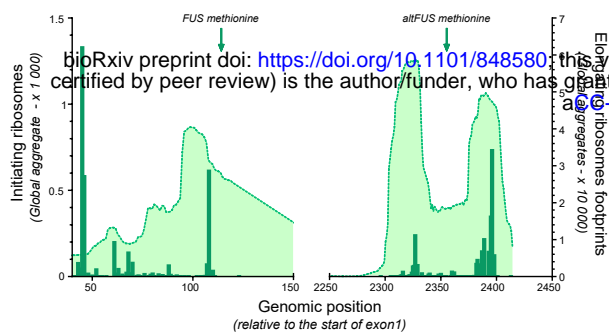


Figure 6



**A**

bioRxiv preprint doi: <https://doi.org/10.1101/848580>; this version posted April 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

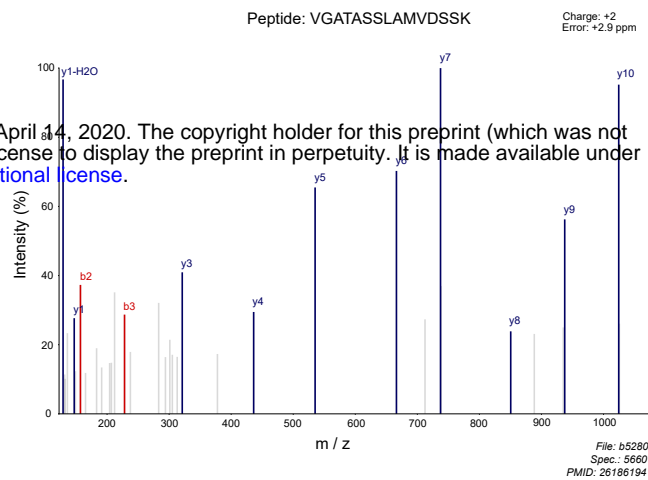
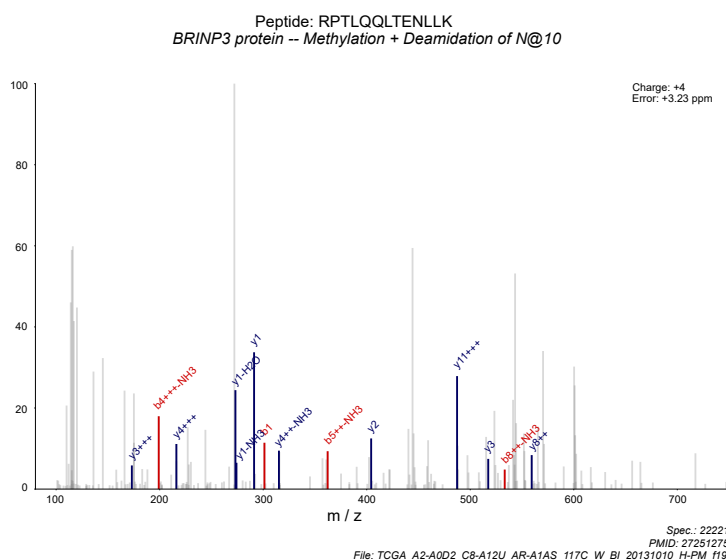
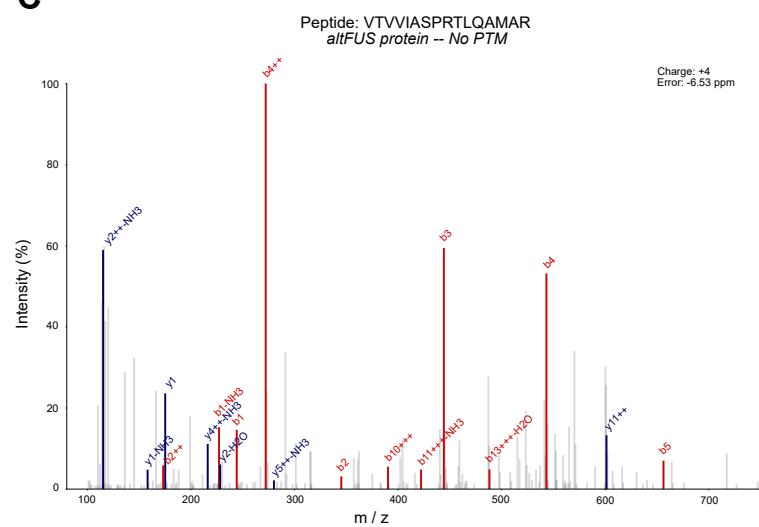
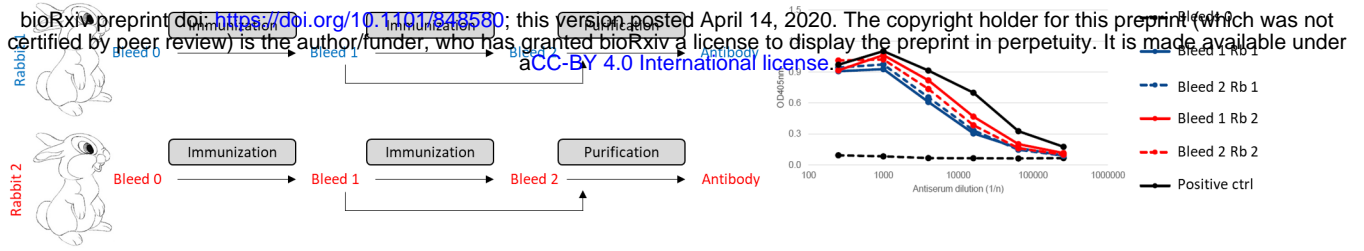
**B****C**

Figure EV2

**a** MGPTPPSPGRAIPSRVSPDTSRVTVVIASPRTLQAMARAAILLMARARTQAMELSQLPRDMARLAAMAVARAPNRLTGSSPPTLAMASSQLPAAPREVTVAVLRA  
AAMGSPRVGATASSLAMVDSSKAMDSSKAIPLRAMDSRTSTTAAVVVEVEVEVETMAKINPP



**b** >Mutated\_sequence  
**ACG**GGGCCTACCCACCCAGCCGGGGCAGGGCTATTCCAGCAGAGCAGTCAGCCCTACGGACAGCAGAGTTACAGTGGTTATAGC  
 CAGTCCACGGACACTTCAGGCT**ACG**GCCAGAGCAGCTATTCTTCT**ACG**GCCAGAGCCAGAACACAGGCT**ACG**GAACTCAGTCAAC  
 TCCCAGGGAT**ACG**GCTCGACTGGCGGCT**ACG**GCAGTAGCCAGAGCTCCAATCGTCTTACGGGCAGCAGTCTCTACCTGGCT**A**  
**CG**GCCAGCAGCCAGTCCCAGCAGCACCTCGGGAAGTTACGGTAGCAGTTCTCAGAGCAGAGCT**ACG**GGCAGCCAGAGTGGG  
 AGCTACAGCCAGCAGCCTAGCT**ACG**GTGGACAGCAGCAAAGCT**ACG**GACAGCAGCAAAGCTATAATCCCCTCAGGGCT**ACG**GACA  
 GCAGAACCAGTACAACAGCAGCAGTGGTGGTGGAGGTGGAGGTGGAGGTGGAGGTA**ACTACG**GCCAAGATCAATCTCCATGA

>altFUS\_Ø  
 TGPTPPSPGRAIPSRVSPDTSRVTVVIASPRTLQATARAAILLTARARTQATELSQLPRDTARLAATAVARAPNRLTGSSPPTLATASSQLP  
 AAPREVTVAVLRAAATGSPRVGATASSLATVDSSKATDSSKAIPLRATDSRTSTTAAVVVEVEVEVETAKINPP

**c**

|                 |     |  |     |
|-----------------|-----|--|-----|
| <i>AltFUS</i>   | 1   | MGPTPPSPGRAIPSRVSPDTSRVTVVIASPRTLQAMARAAILLMARARTQAMELSQLPR  | 60  |
| <i>AltFUS-Ø</i> | 1   | TGPTPPSPGRAIPSRVSPDTSRVTVVIASPRTLQATARAAILLTARARTQATELSQLPR  | 60  |
| <i>AltFUS</i>   | 61  | DMARLAAMAVARAPNRLTGSSPPTLAMASSQLPAAPREVTVAVLRAAAMGSPRVGATASS | 120 |
| <i>AltFUS-Ø</i> | 61  | DTARLAATAVARAPNRLTGSSPPTLATASSQLPAAPREVTVAVLRAAATGSPRVGATASS | 120 |
| <i>AltFUS</i>   | 121 | LAMVDSSKAMDSSKAIPLRAMDSRTSTTAAVVVEVEVEVETMAKINPP             | 170 |
| <i>AltFUS-Ø</i> | 121 | LATVDSSKATDSSKAIPLRATDSRTSTTAAVVVEVEVEVETAKINPP              | 170 |

**d**

|               |     |  |     |
|---------------|-----|--|-----|
| <i>FUS</i>    | 1   | MASNDYTQQATQSYGAYPTQPGQYSQQSSQPYGQQSYSGYSQSDTSGYGQSSYSSYGQ     | 60  |
| <i>FUS(Ø)</i> | 1   | MASNDYTQQATQSYGAYPTQPGQYSQQSSQPYGQQSYSGYSQSDTSGYGQSSYSSYGQ     | 60  |
| <i>FUS</i>    | 61  | SQNTGYGTQSTPQGYGSTGGYGSQSSQSSYQGQSSYPGYGQQPAPSSSTSGSYGSSSQSS   | 120 |
| <i>FUS(Ø)</i> | 61  | SQNTGYGTQSTPQGYGSTGGYGSQSSQSSYQGQSSYPGYGQQPAPSSSTSGSYGSSSQSS   | 120 |
| <i>FUS</i>    | 121 | SYGQPQSGSYSQQPSYGGQQQS YGQQQSYNPPQGYGQQNQYNSSSGGGGGGGGGNYGQD   | 180 |
| <i>FUS(Ø)</i> | 121 | SYGQPQSGSYSQQPSYGGQQQS YGQQQSYNPPQGYGQQNQYNSSSGGGGGGGGGNYGQD   | 180 |
| <i>FUS</i>    | 181 | QSSMS SGGGSGGYGNQDQSGGGGSGGYGQDRGGRGRGGS SGGGGGGGGYNRSSGGYE    | 240 |
| <i>FUS(Ø)</i> | 181 | QSSMS SGGGSGGYGNQDQSGGGGSGGYGQDRGGRGRGGS SGGGGGGGGYNRSSGGYE    | 240 |
| <i>FUS</i>    | 241 | PRGRGGGRGGRGGMGSDRGGFNKFGGPRDQGSRHDSEQDNDNNTIFVQGLGENVTIES     | 300 |
| <i>FUS(Ø)</i> | 241 | PRGRGGGRGGRGGMGSDRGGFNKFGGPRDQGSRHDSEQDNDNNTIFVQGLGENVTIES     | 300 |
| <i>FUS</i>    | 301 | VADYFKQIGI IKTNKKTGQPMINLYTDRETGKLGKGEATVSFDDPPSAKAAIDWFDGKEFS | 360 |
| <i>FUS(Ø)</i> | 301 | VADYFKQIGI IKTNKKTGQPMINLYTDRETGKLGKGEATVSFDDPPSAKAAIDWFDGKEFS | 360 |
| <i>FUS</i>    | 361 | GNP I KVSFATRRADFNRRGGNGRGRGRGGPMGRGGYGGGGSGGGRRGGFP SGGGGGGGQ | 420 |
| <i>FUS(Ø)</i> | 361 | GNP I KVSFATRRADFNRRGGNGRGRGRGGPMGRGGYGGGGSGGGRRGGFP SGGGGGGGQ | 420 |
| <i>FUS</i>    | 421 | QRAGDWKCPNPTCENMNF SWRNECNQCKAPKPDGPGGGPGGSHMGGNYGDDRRGGRGGYD  | 480 |
| <i>FUS(Ø)</i> | 421 | QRAGDWKCPNPTCENMNF SWRNECNQCKAPKPDGPGGGPGGSHMGGNYGDDRRGGRGGYD  | 480 |
| <i>FUS</i>    | 481 | RGGYRGRGGDRGGFRGGRRGGDRGGFGPGKMDSRGEHRQDRRERPY                 | 526 |
| <i>FUS(Ø)</i> | 481 | RGGYRGRGGDRGGFRGGRRGGDRGGFGPGKMDSRGEHRQDRRERPY                 | 526 |

Figure EV3

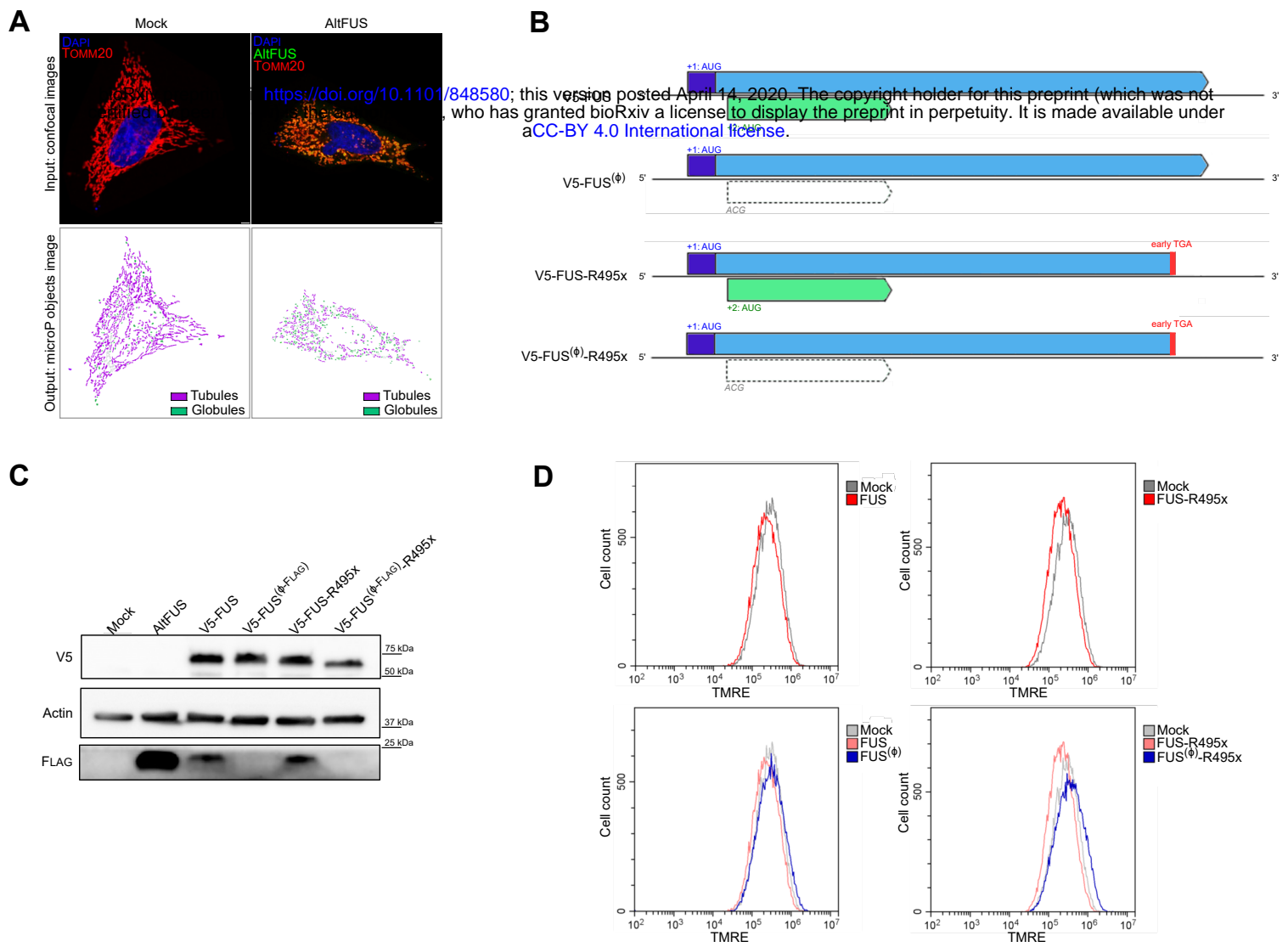


Figure EV4

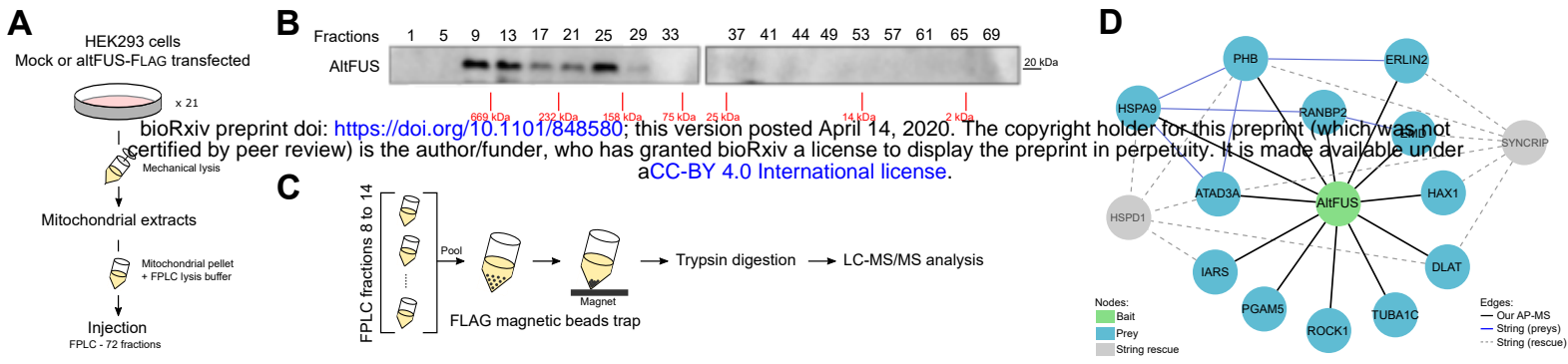


Figure EV5

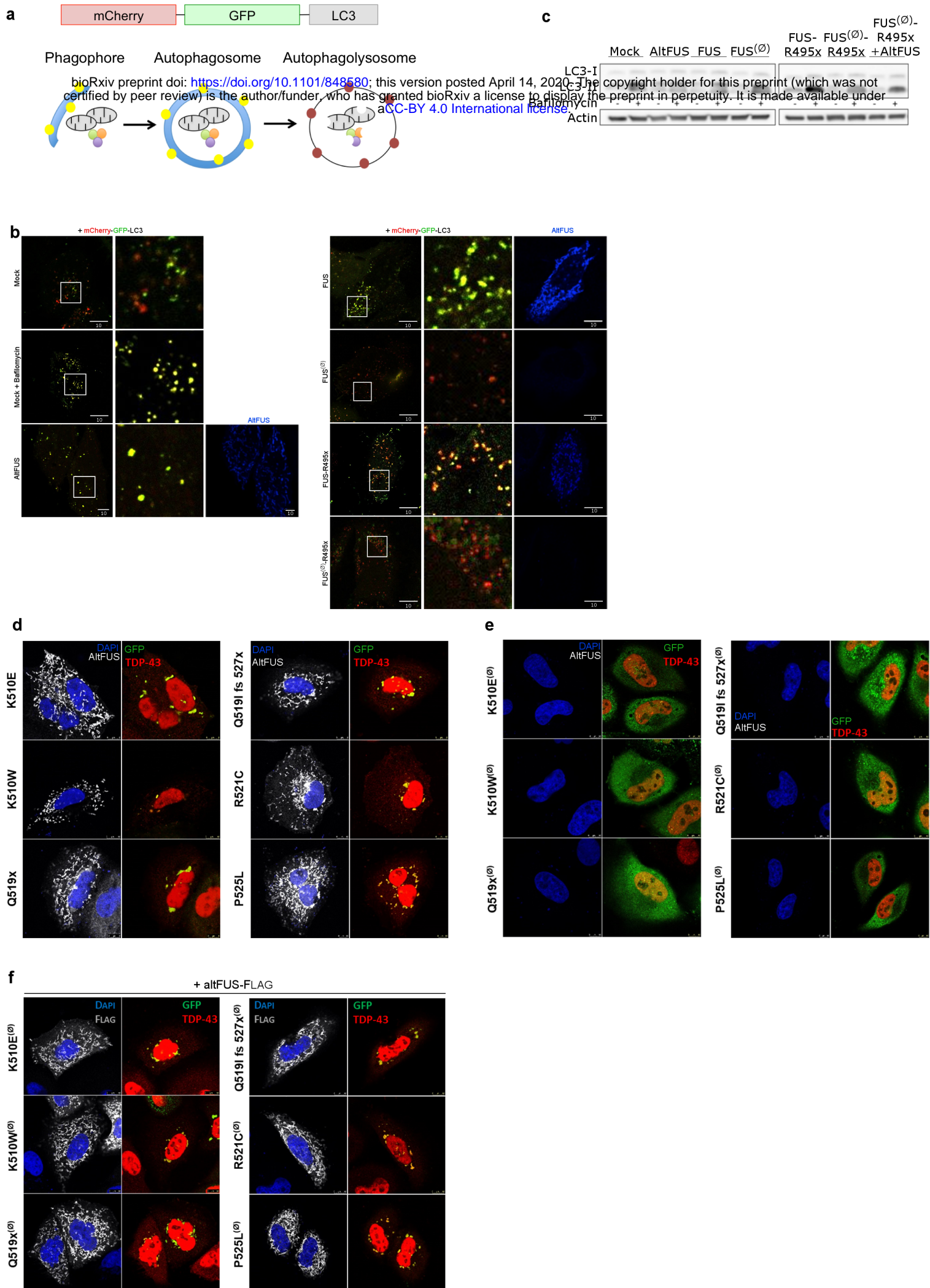


Figure EV6



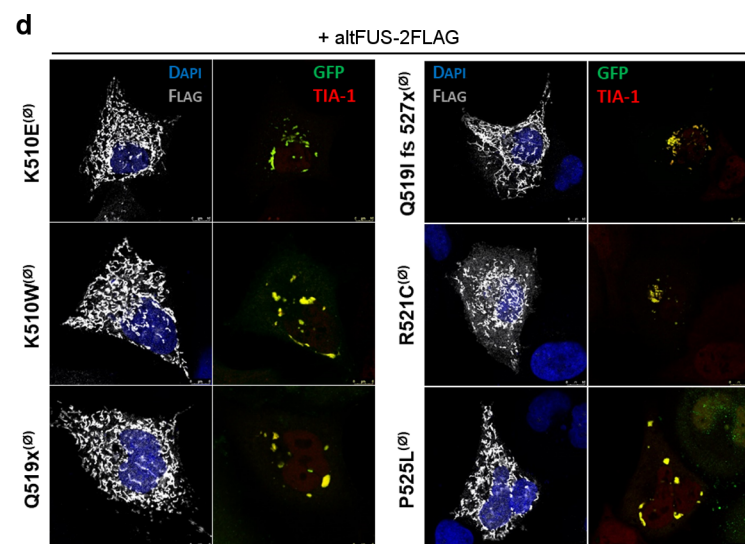
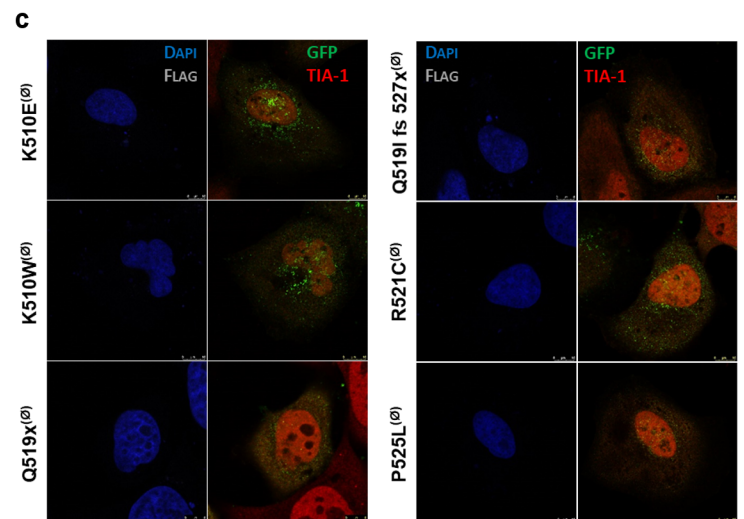
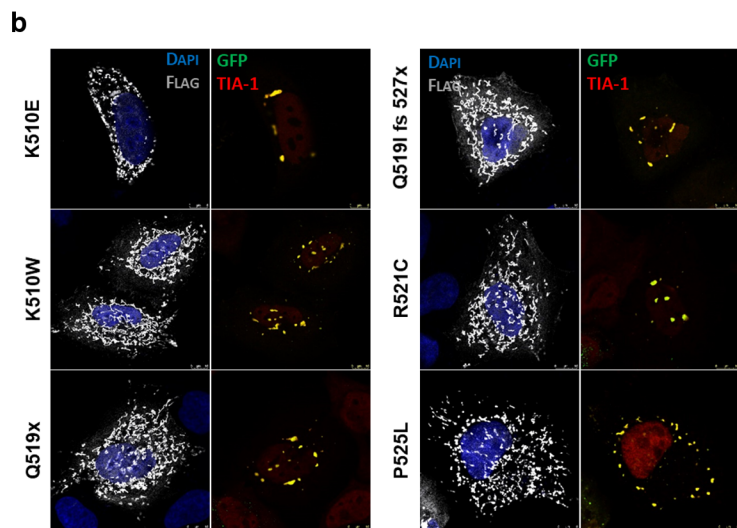
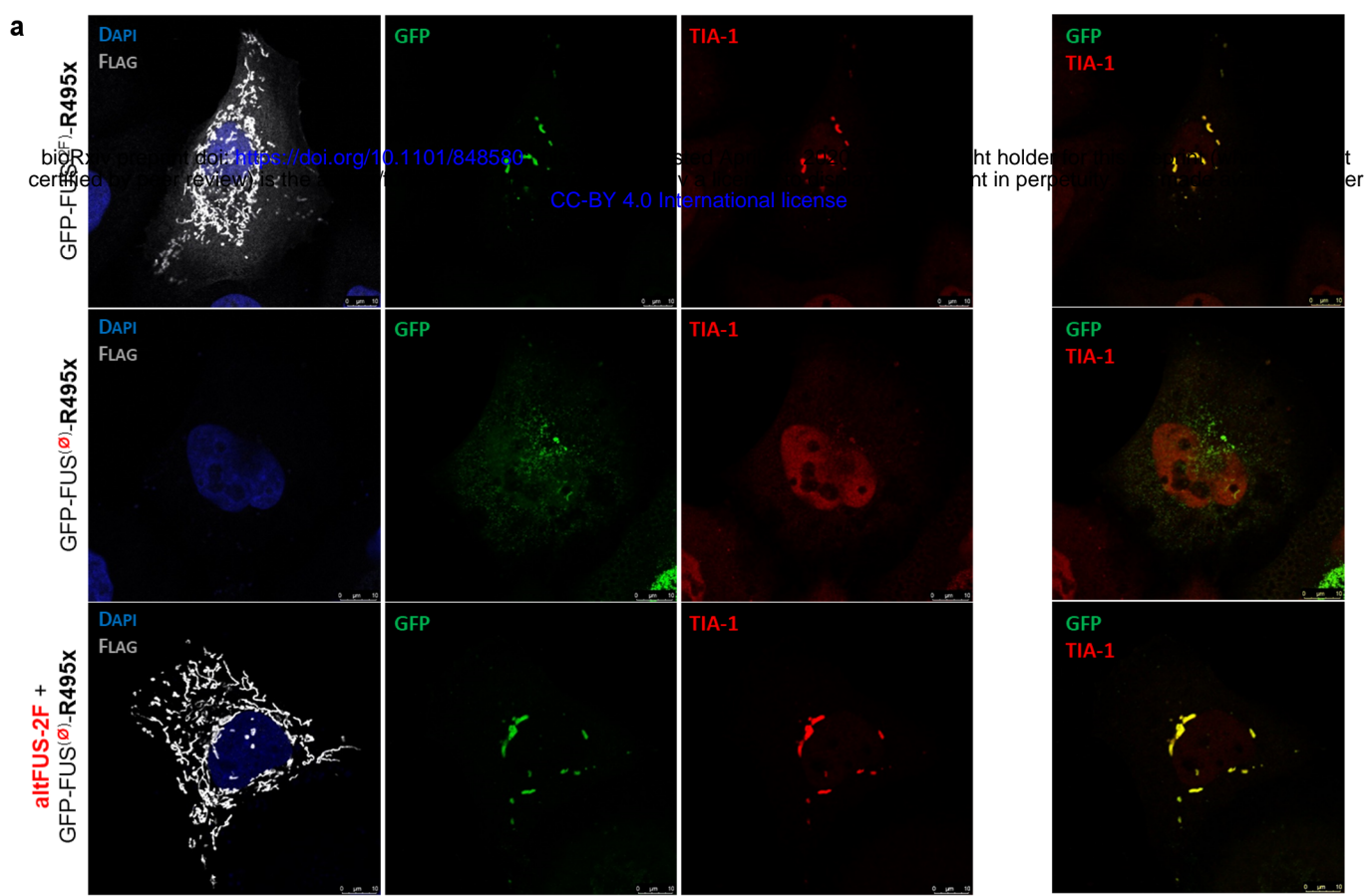
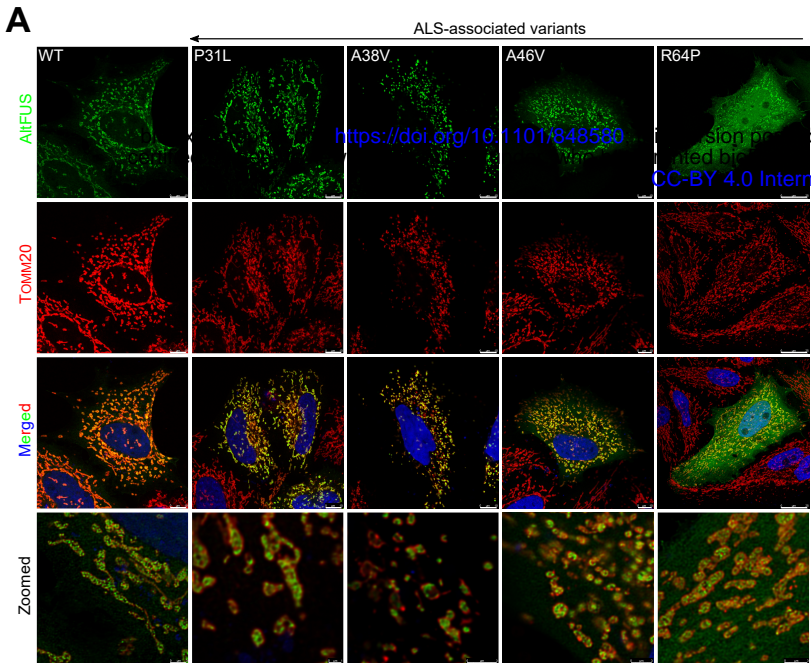


Figure EV7



bioRxiv preprint doi: <https://doi.org/10.1101/848580>; this version posted April 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

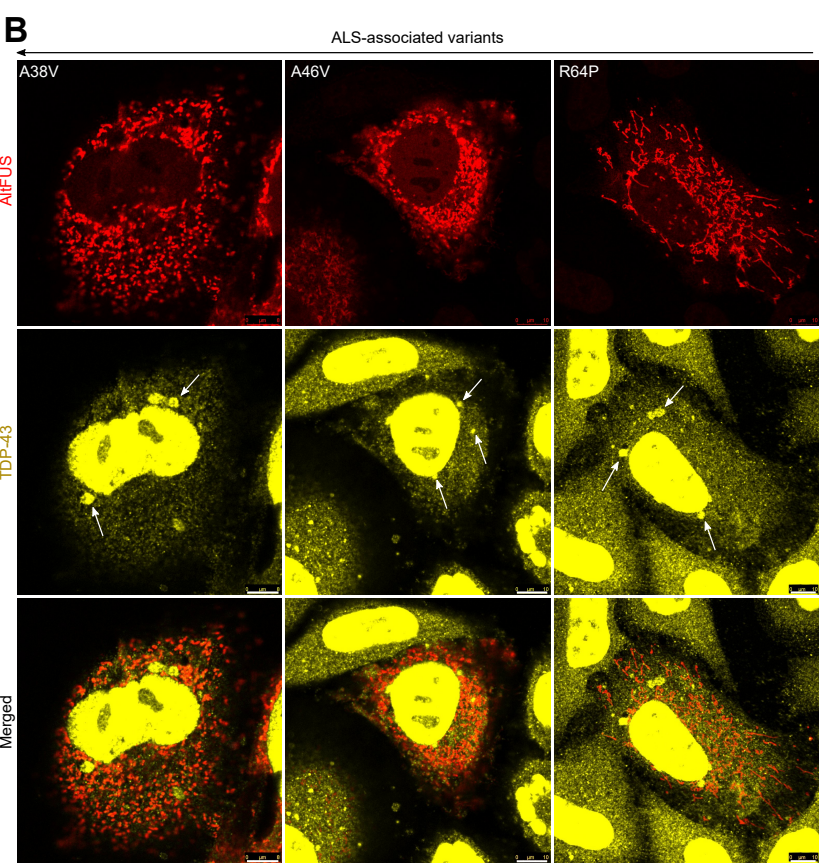


Figure EV8

<https://doi.org/10.1101/848580> this version posted April 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

|  |     |             |        |    |    |           |          |         |     |        |    |   |      |
|--|-----|-------------|--------|----|----|-----------|----------|---------|-----|--------|----|---|------|
| <i>Homo sapiens</i>                    | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Pan paniscus</i>                    | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Gorilla gorilla gorilla</i>         | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Protopithecus</i>                   | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Nomascus leucogenys</i>             | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Pongo abelii</i>                    | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Chlorocebus sabaeus</i>             | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Rhinopithecus roxellana</i>         | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Rhinopithecus bieti</i>             | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Macaca mulatta</i>                  | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Colobus angolensis palliatus</i>    | MGP | T           | PPS    | PV | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Macaca nemestrina</i>               | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Macaca fascicularis</i>             | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Mandrillus leucophaeus</i>          | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Cercocebus atys</i>                 | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Papio anubis</i>                    | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Callithrix jacchus</i>              | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Saimiri boliviensis boliviensis</i> | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Cebus capucinus imitator</i>        | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Sus scrofa</i>                      | MGP | T           | PPS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Physeter catodon</i>                | MGP | T           | QPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Tursiops truncatus</i>              | MGP | T           | RPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Orcinus orca</i>                    | MGP | T           | RPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Balaenoptera acutorostrata</i>      | MGP | T           | QPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Sorex araneus</i>                   | MGP | T           | QPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Aotus nancymae</i>                  | MGP | T           | PRS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Trichechus manatus latirostris</i>  | MGP | T           | PRS    | PG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Toxotrichus minima</i>              | MTL | L           | QAR    | H  |    |           |          |         | M   | ARAA   | MV | L | MDK  |
| <i>Ceratotherium simum simum</i>       | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Condylura cristata</i>              | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Galeopterus variegatus</i>          | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Odobenus rosmarus divergens</i>     | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Tarsius syrichta</i>                | MGP | T           | LPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Microcebus murinus</i>              | MGP | T           | PPS    | RG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Bison bison bison</i>               | MGP | T           | PPS    | RG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Propithecus coquereli</i>           | MGP | T           | PPS    | RG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Bos mutus</i>                       | MGP | T           | PPS    | RG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Pteropus vampyrus</i>               | MGL | T           | PPS    | LD | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | LAM | ARAA   | I  | L | LMAR |
| <i>Pteropus alecto</i>                 | MGL | T           | PPS    | LD | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | LAM | ARAA   | I  | L | LMAR |
| <i>Vicuqna pacos</i>                   | MGP | T           | LPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Rousettus aegyptiacus</i>           | MGL | T           | PPS    | LD | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Bos taurus</i>                      | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Equus caballus</i>                  | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Miniopterus natalensis</i>          | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Canis lupus familiaris</i>          | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Camelus ferus</i>                   | MGP | T           | LPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Camelus bactrianus</i>              | MGP | T           | LPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Equus przewalskii</i>               | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Equus asinus</i>                    | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Eptesicus fuscus</i>                | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Leptonyctotis weddellii</i>         | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Octodon degus</i>                   | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Otolemur gamettii</i>               | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Myotis davidii</i>                  | MGP | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Dipodomys ordii</i>                 | MGP | T           | LLS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Jaculus jaculus</i>                 | MGP | T           | PHS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Ochotona princeps</i>               | MGP | T           | PHS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Microtus ochrogaster</i>            | MGP | T           | LPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Mus musculus</i>                    | MGP | T           | LPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Peromyscus maniculatus bairdii</i>  | MGL | T           | LPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Erinaceus europaeus</i>             | MGR | T           | PPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Nannospalax galili</i>              | MGP | T           | LPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Panholops hodgsonii</i>             | MGP | T           | QLS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Drosophila melanogaster fusilli</i> | A   | T           | QLS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Rattus norvegicus</i>               | MGP | T           | LPS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Capra hircus</i>                    | MGP | T           | QLS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Brachypodium distachyon</i>         | MA  | T           | EGT    |    |    |           |          |         | RAM | ARAA   | I  | L | LMAR |
| <i>Ovis aries</i>                      | MGP | T           | QLS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Ovis aries musimon</i>              | MGP | T           | QLS    | LG | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Oreochromis niloticus</i>           | MVE | T           | A-V    | LA | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Haplochromis burtoni</i>            | MVE | T           | A-V    | LA | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Perkinsus marinus</i>               | MEV | T           | VGRVPS |    |    |           |          |         | QAM | ARAA   | I  | L | LMAR |
| <i>Eucalyptus grandis</i>              | MAS | PAAVAPPGTTT | SPA    | TT | AA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Gallus gallus</i>                   | MGL | T           | LLS    | LD | KV | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Sarcophilus hamsii</i>              | MGL | T           | LLS    | LD | KV | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Ictalurus punctatus</i>             | MVD | T           | EVS    | PV | RA | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Anolis carolinensis</i>             | MGP | T           | QLH    | PH | RV | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Trypanosoma</i>                     | MQ  | I           | VKT    | L  |    |           |          |         | ALE | VESSDI |    |   | EN   |
| <i>Xenopus tropicalis</i>              | MGP | T           | PPN    | QA | RV | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Danio rerio</i>                     | MIM | T           | PPN    | QA | RV | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Latimeria chalumnae</i>             | MV  | T           | LQ     | LG | RV | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |
| <i>Xenopus laevis</i>                  | MGP | T           | PPS    | QA | RV | IPSRVAVSP | TDSRVTVV | IASPRTL | QAM | ARAA   | I  | L | LMAR |

|   | 80      | 90     | 100      | 110           | 120               | 130   | 140                  |
|---|---------|--------|----------|---------------|-------------------|-------|----------------------|
| <i>Homo sapiens</i>   | ART     | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APNRL | TGSSPP               |
| <i>Pan paniscus</i>   | ART     | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APNRL | TGSSPP               |
| <i>Gorilla gorilla gorilla</i>  | ART     | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APNRL | TGSSPP               |
| <i>bioRxiv preprint doi: <a href="https://doi.org/10.1101/848580">https://doi.org/10.1101/848580</a>; this version posted April 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.</i> | ART     | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APNRL | TGSSPP               |
| <i>Nomascus leucogenys</i>  | ART     | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APNRL | TGSSPP               |
| <i>Pongo abelii</i>   | ART     | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APNRL | TGSSPP               |
| <i>Chlorocebus sabaeus</i>  | ART     | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APNLL | MGSSPP               |
| <i>Rhinopithecus roxellana</i>  | ART     | QAME   | LSQLPRD  | MAQLAAMAVAR   |                   | APNLL | TGSSPP               |
| <i>Rhinopithecus bieti</i>  | ARTVLSQ | QAME   | LSQLPRD  | MAQLAAMAVAR   |                   | APNLL | TGSSPP               |
| <i>Macaca mulatta</i>   | ARA     | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APNLL | MGSSPP               |
| <i>Colobus angolensis palliatus</i>   | ART     | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APNLL | TGSSPP               |
| <i>Macaca nemestrina</i>  | ARA     | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APNLL | MGSSPP               |
| <i>Macaca fascicularis</i>  | ARA     | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APNLL | MGSSPP               |
| <i>Mandrillus leucophaeus</i>   | ARA     | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APNLL | MGSSPP               |
| <i>Cercocebus atys</i>  | ARA     | QAME   | LSQLPRG  | MARLAAMAVAR   |                   | APNLL | MGSSPP               |
| <i>Papio anubis</i>   | VRAVLSQ | QAME   | LSQLPRD  | MARLAAMAVAR   |                   | APSL  | MGSSPP               |
| <i>Callithrix jacchus</i>   |         | MAL    | LSQLPRD  | MAQLVAMAVAR   |                   | APNHL | TGSSPP               |
| <i>Saimiri boliviensis boliviensis</i>  | PRT     | QAMA   | LSQLPRD  | MAQLVAMAVAR   |                   | APNHL | TGSSPL               |
| <i>Cebus capucinus imitator</i>   | PRT     | QAMA   | LSQLPRD  | MAQLVAMAVAR   |                   | APNHL | TGSSPL               |
| <i>Sus scrofa</i>   | PRT     | QAMAHS | LSQLPRD  | MAQLVAMAVAR   |                   | VLNRL | TGSSLH               |
| <i>Physeter catodon</i>   | PRT     | QAMAHS | LSQLPRD  | MAQLVAMAVAK   |                   | VLSRL | MGSSPH               |
| <i>Tursiops truncatus</i>   | PRT     | QAMAHS | LSQLPRD  | MAQLVAMAVAK   |                   | VLSRL | TGSSPH               |
| <i>Orcinus orca</i>   | PRT     | QAMAHS | LSQLPRD  | MAQLVAMAVAK   |                   | VLSRL | TGSSPH               |
| <i>Balaenoptera acutorostrata</i>   | PRT     | QAMAHS | LSQLPRD  | MAQLVAMAVAK   |                   | VLSRL | TGSSPH               |
| <i>Sorex araneus</i>  |         | MALS   | PLPRD    | TAQLVMAAVR    |                   | VLSRP | MGSSPP               |
| <i>Aotus nancymae</i>   | PRT     | QAMA   | LSQLPRD  | MAQLVAIVVAR   |                   | APNHL | TGSSPP               |
| <i>Trichechus manatus latirostris</i>   | PRT     | QVMA   | PSQLPRD  | MAQLVAMAVAR   |                   | VPNHL | TGSSLR               |
| <i>Toxonia minima</i>   | PRT     | A      | PEMA     | PPRR          | TAR               |       |                      |
| <i>Ceratotherium simum simum</i>  | PRT     | QAME   | LSQLPKD  | MAQLVAMAVAR   |                   | VLSRL | TGSSPP               |
| <i>Condylura cristata</i>   | PRT     | QVMA   | LSQLPRD  | MAQLAAMAVAR   |                   | VLNRL | TGSSPP               |
| <i>Galeopterus variegatus</i>   | PRT     | QAMALN | QLPRD    | MAQPVMAMVVR   |                   | VLNRL | TGSSPP               |
| <i>Odobenus rosmarus divergens</i>  | PRT     | QAMA   | LSQVPRD  | MAQPVMAMVVR   |                   | VLSHL | MGNSPL               |
| <i>Tarsius syrichta</i>   | PRT     | QVIA   | LSQLHRD  | MAQVAMAVAR    |                   | VPSRL | MGSSPP               |
| <i>Microcebus murinus</i>   | LRT     | QAI    | VLSQLPRD | MAQLVAMAVAR   |                   | VPNRL | MGSSPL               |
| <i>Bison bison bison</i>  | PRT     | QATA   | HSQLPRD  | MAQLVAMAVAR   |                   | VLSHL | MGNSPR               |
| <i>Propithecus coquereli</i>  | PRT     | QAI    | LRQLPRD  | MAQLVAMAVAR   |                   | VPSRL | MGSSPP               |
| <i>Bos mutus</i>  | PRT     | QATA   | HSQLPRD  | IAQLVAMAVAR   |                   | VLSHL | MGNSPR               |
| <i>Pteropus vampyrus</i>  | PRT     | QAI    | TLSQLPRD | MAQQVAMA      | AR                | VLNLL | MGSSPP               |
| <i>Pteropus alecto</i>  | PRT     | QAI    | TLSQLPRD | MAQRVAMA      | AR                | VLNLL | MGSSPP               |
| <i>Vicuqna pacos</i>  | LRT     | QAMAHS | LSQLPRV  | MAQLVAMAAK    |                   | VRSRL | MGSSPH               |
| <i>Rousettus aegyptiacus</i>  | PRT     | QAI    | LSQLPRD  | MAQRVAMA      | AR                | VLNLL | MGSSPP               |
| <i>Bos taurus</i>   | PRT     | QATA   | HSQLPRD  | MAQLVAMAVAR   |                   | VLSHL | MGNSPR               |
| <i>Equus caballus</i>   | PRT     | QAMA   | LSQLPRD  | MAQLVDMAVAR   |                   | VLSRL | TGSSPP               |
| <i>Miniopterus natalensis</i>   | PRT     | QAMA   | LSQLPRD  | MAQLVAMAVDR   |                   | VLSRL | MGSSP                |
| <i>Canis lupus familiaris</i>   | PRT     | QAMA   | LSQLPRD  | MAQLVAMAVAR   |                   | VLSHL | MGNSPL               |
| <i>Camelus ferus</i>  | PRT     | QAMAHS | LSQLPRV  | MAQLVATAAK    |                   | VLSRL | MGSSPH               |
| <i>Camelus bactrianus</i>   | PRT     | QAMAHS | LSQLPRV  | MAQLVATAAK    |                   | VLSRL | MGSSPH               |
| <i>Equus przewalskii</i>  | PRT     | QAMA   | LSQLPRD  | MAQLVDMAVAR   |                   | VLSRL | TGSSPP               |
| <i>Equus asinus</i>   | LRT     | QAMA   | LSQLPRD  | MAQLVDMAVAR   |                   | VLSRL | TGSSPP               |
| <i>Eptesicus fuscus</i>   | PRT     | QAMA   | PSQLPRD  | MAQPVMAMVVR   |                   | VLSHL | MGSSP                |
| <i>Leptonychotes weddellii</i>  | PRT     | QAMA   | LSQVPRD  | MAQPVMAMVVR   |                   | VLSHL | MGNSPL               |
| <i>Octodon degus</i>  | LRT     | QVMA   | HSQLPRV  | MAPLVAMAVAR   |                   | VPSHL | MVSSPL               |
| <i>Otlemur gamettii</i>   | PRT     | QVIA   | LSQLPRD  | MAQLVAMAVAR   |                   | VLSHL | MGSSPL               |
| <i>Myotis davidii</i>   | PRT     | QAMA   | PSQLPRV  | MAQPVMAMVVR   |                   | VLSHL | MGSSP                |
| <i>Dipodomys ordii</i>  | LRT     | QVME   | LSQLPRG  | TAPLVGMAAAR   |                   | VPSRL | TGSSLP               |
| <i>Jaculus jaculus</i>  | PRT     | QVMA   | LSQPPRD  | MVPLVDMAVAR   |                   | VPN   | MGSSPH               |
| <i>Ochotona princeps</i>  | HRT     | QAMA   | LSQLPRV  | MARLAAMVAAR   |                   | VPSRL | TDNSPP               |
| <i>Microtus ochrogaster</i>   | PKT     | QAMA   | LSQLPRD  | MVPLEAMAAAR   |                   | VPNLL | MGSSPP               |
| <i>Mus musculus</i>   | PKT     | QAME   | LSQLPRD  | MVPLEAMAAAK   |                   | VHNL  | MGSSPP               |
| <i>Peromyscus maniculatus bairdii</i>   | PKT     | QAMA   | LSQLPRD  | MVPLEAMAAAK   |                   | VPSL  | MGSSPP               |
| <i>Erinaceus europaeus</i>  | PRT     | QAMV   | LSQPPRA  | TAPPAAMAAVR   |                   | VLS   | TGSSRP               |
| <i>Nannospalax galili</i>   | PRT     | QAMA   | LSRLPRD  | MVPLEAMAVAK   |                   | VPNLL | MDNSPP               |
| <i>Panthalops hodgsonii</i>   | PR I    | QATA   | HSQLPRD  | MAQLVAMAVAR   |                   | VLSRL | TGNSLR               |
| <i>Drosophila melanogaster fusilli</i>  | PRV     | MPSCYS | QTRRM    | R             | PRRW              |       | AGTVNPLG             |
| <i>Rattus norvegicus</i>  | PKT     | QAME   | LSQHPRD  | MVPLVAMAAAK   |                   | VPSLL | MGSSPP               |
| <i>Capra hircus</i>   | PR I    | QATA   | RSQLPRD  | MAQLVAMAVAR   |                   | VLSRL | TGNSLR               |
| <i>Brachypodium distachyon</i>  |         |        |          | VVQVLEMA      | LYL               | VVE   |                      |
| <i>Ovis aries</i>   | PR I    | QATA   | RSQLPRD  | MAQLVAMAVAR   |                   | VLSRL | TGNSLR               |
| <i>Ovis aries musimon</i>   | PR I    | QATA   | RSQLPRD  | MAQLVAMAVAR   |                   | VLSRL | TGNSLR               |
| <i>Oreochromis niloticus</i>  |         | AAPAP  |          | TIREAIAATAI   |                   | LSQVD | TGLHPP               |
| <i>Haplochromis burtoni</i>   |         | AAPAP  |          | TVREAIAATAI   |                   | LSQVD | TGLHPP               |
| <i>Perkinsus marinus</i>  |         |        |          |               |                   |       |                      |
| <i>Eucalyptus grandis</i>   |         |        |          |               |                   |       |                      |
| <i>Gallus gallus</i>  |         |        |          | MAAAEGTAVGR   |                   | AHSHP | MGRPPP               |
| <i>Sarcophilus hamsii</i>   | PRT     | QAMV   | PNRLPKD  | MVQPVAMEVAR   |                   | VPNLL | MVSSHP               |
| <i>Ictalurus punctatus</i>  | PHS     | PT     |          | ARVV LTA      | AAAVSLL           | RAE   | QRTISRPP             |
| <i>Anolis carolinensis</i>  | LKT     | LAMA   | LSPHRRP  | MAHQQDMAAAK   |                   | LLRHP | MDSSH                |
| <i>Trypanosoma</i>  | VKA     | KIQD   | KEGIPPD  | QQRLLIFAGKQL  |                   | E     | DGR                  |
| <i>Xenopus tropicalis</i>   | TRV     | LDIV   | LSQLHKD  | MDLVVME       | TPN               | HLSP  | MAAGSHSSPSP LTPA TPS |
| <i>Danio rerio</i>  | TAP     | TMDSP  | SQEDMD   | LKPLKATVSPVSR | IALVATVTPASLHQLKA |       | EATSSPL              |
| <i>Latimeria chalumnae</i>  | HKL     | LGME   | HSPRRV   | MAPVAME       | ALK               | HHR   | TPMNNLL              |
| <i>Xenopus laevis</i>   | IRA     | LGIV   | LSQLHNK  | GMDLEVME      | TPN               | HLSP  | MAA GS HSSRL         |

bioRxiv preprint doi: <https://doi.org/10.1101/848580>; this version posted April 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

|  |                  |    |    |         |            |      |          |       |       |             |      |
|--|------------------|----|----|---------|------------|------|----------|-------|-------|-------------|------|
| <i>Homo sapiens</i>                    | LAMASSQLPAAPR    | EV | T  | VAVLRA  | AA         | MGS  | PRVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Pan paniscus</i>                    | LVMASSQLPAAPR    | EV | T  | VAVLRA  | AA         | MGS  | PRVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Gorilla gorilla gorilla</i>         | LAMASSQLPAAPR    | EV | T  | VAVLRA  | AA         | MGS  | PRVGA    | TAS   | SLAMV | DSSL        |      |
| <i>Nomascus leucogerys</i>             | LAMASSQLPAAPR    | EV | T  | VAVLRA  | AA         | MGS  | PRVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Pongo abelii</i>                    | LAMASSQLPAAPR    | EV | T  | VAVLRA  | AA         | MGS  | PRVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Chlorocebus sabaeus</i>             | LAMASSQLPAAPR    | EV | T  | VAVLRP  | AA         | MGS  | PRVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Rhinopithecus roxellana</i>         | LVMASSQLPAAPR    | EV | T  | VAVLRP  | AA         | MGS  | PQVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Rhinopithecus bieti</i>             | LVMASSQLPAAPR    | EV | T  | VAVLRP  | AA         | MGS  | PQVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Macaca mulatta</i>                  | LAMASSQLPAAPR    | EV | T  | VAVLRP  | AA         | MGS  | PRVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Colobus angolensis palliatus</i>    | LVMASSQLPAAPR    | EV | T  | VAVLRP  | AA         | MGS  | PQVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Macaca nemestrina</i>               | LAMASSQLPAAPR    | EV | T  | VAVLRP  | AA         | MGS  | PRVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Macaca fascicularis</i>             | LAMASSQLPAAPR    | EV | T  | VAVLRP  | AA         | MGS  | PRVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Mandrillus leucophaeus</i>          | LAMASSQLPAAPR    | EV | T  | VAVLRP  | AA         | MGS  | RRVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Cercocebus atys</i>                 | LAMASSQLPAAPR    | EV | T  | VAVLRP  | AA         | MGS  | PRVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Papio anubis</i>                    | LAMASSQLPAAPR    | EV | T  | VAVLRP  | AA         | MGS  | PQVGA    | TAS   | SLAMV | DSSK        |      |
| <i>Callithrix jacchus</i>              | LAMVSSQLPAAPR    | EV | T  | GAVLRA  | AA         | MGS  | PRVAA    | TAS   | NLAMV | DSSK        |      |
| <i>Saimiri boliviensis boliviensis</i> | LAMASSQLPAAPR    | EV | M  | GAVLRA  | AA         | TGS  | PRVAA    | TAS   | SLAMV | DSSK        |      |
| <i>Cebus capucinus imitator</i>        | LAMASSQLPAAPR    | EV | T  | GAVLRA  | AA         | MGS  | PRVAA    | MAS   | SLAMV | DSSK        |      |
| <i>Sus scrofa</i>                      | LAMASSQLLAAPR    | EV | M  | VLVRA   | VA         | MGS  | PRVGAMAS | NLAMV | DSSK  |             |      |
| <i>Physeter catodon</i>                | LAMASSQLLAAPR    | EV | T  | VAVLRA  | AA         | MGS  | PRVGAMAS | SLAMV | DSSK  |             |      |
| <i>Tursiops truncatus</i>              | LAMASSQLLAAPR    | EV | T  | VAVLRA  | AA         | MGS  | PRVGAMAS | SLAMV | DSSK  |             |      |
| <i>Orcinus orca</i>                    | LAMASSQLLAAPR    | EV | T  | VAVLRA  | AA         | MGS  | PRVGAMAS | SLAMV | DSSK  |             |      |
| <i>Balaenoptera acutorostratus</i>     | LAMASSQLLAAPR    | EV | T  | VAVLRA  | AA         | MGS  | PRVGAMAS | SLAMV | DSSK  |             |      |
| <i>Sorex araneus</i>                   | LATASSQLPRAVPR   | EV | M  | AAVLRA  | AA         | MGS  | PRVAA    | MAS   | SLAMV | GSSK        |      |
| <i>Aotus nancymae</i>                  | LAMASSQLRAAPR    | EV | M  | GAVLRA  | AA         | MGS  | PRVAA    | TAS   | SLAMV | YSSK        |      |
| <i>Trichechus manatus latirostris</i>  | LAMASSQLPAAPR    | EA | M  | EVVLR   | AA         | MGS  | PRVEA    | MAN   | SLDMV | DSSQ        |      |
| <i>Toxopneustes purpurascens</i>       | PTPSL            |    |    |         |            |      |          |       |       |             |      |
| <i>Ceratotherium simum simum</i>       | LAMASNRLLAAPR    | EV | T  | AAVPR   | AA         | MGS  | PRVVA    | MAS   | SLAMV | DSSQ        |      |
| <i>Condylura cristata</i>              | LAMASSQLLAAPR    | EV | M  | GAVLRA  | AA         | TDS  | PRVEA    | MAS   | SLATV | DSSK        |      |
| <i>Galeopterus variegatus</i>          | LAMVNSQLPAVPR    | EV | M  | VAVLRA  | AA         | MGS  | PQVGAMAN | SLAMV | DSSK  |             |      |
| <i>Odobenus rosmarus divergens</i>     | LAMASSQLLAALR    | EV | T  | VVVLR   | AA         | TGS  | PRVGAMAS | SLP   | IV    | DSSQ        |      |
| <i>Tarsius syrichta</i>                | LAMVSNQLPAAPR    | EV | T  | VAVLRA  | AA         | MGS  | PRVGAMAN | SLAMV | DSSK  |             |      |
| <i>Microcebus murinus</i>              | LAMASSRLPAAPR    | EV | T  | VAVLRA  | AA         | TGS  | PRVGAMAN | SLAMV | DSSK  |             |      |
| <i>Bison bison bison</i>               | LAMASSQLLAAPQ    | EV | T  | VAVLKA  | AA         | MGS  | PRVEA    | TAS   | SLAMV | GSSK        |      |
| <i>Propithecus coquereli</i>           | LAMANSRLPAAPR    | EV | T  | VAVPRA  | AA         | MDS  | PQAGAMAS | SLAMV | DSSK  |             |      |
| <i>Bos mutus</i>                       | LAMASSQLLAAPQ    | EV | T  | VAVLKA  | AA         | MGS  | PRVEA    | TAS   | SLAMV | GSSK        |      |
| <i>Pteropus vampyrus</i>               | LAMASSQLPAAPR    | EV | M  | VAVLRA  | AA         | TMGS | PRVGMVS  | SLAMV | DSSK  |             |      |
| <i>Pteropus alecto</i>                 | LAMASSQLPAAPR    | EV | M  | VAVLRA  | AA         | TMGS | PRVGMVS  | SLAMV | DSSK  |             |      |
| <i>Vicuqna pacos</i>                   | LAMASSQLLVAPR    | EA | T  | VAVLRA  | AA         | MAS  | PRAGA    | TAS   | SLAMV | DSSR        |      |
| <i>Rousettus aegyptiacus</i>           | LAMASSQLLAAPR    | EV | M  | VAVLRA  | AA         | TMGS | PRVGMVS  | SLAMV | DSSK  |             |      |
| <i>Bos taurus</i>                      | LAMASSQLLAAPQ    | EV | T  | VAVLKA  | AA         | MGS  | PRVEA    | TAS   | SLAMV | GSSK        |      |
| <i>Equus caballus</i>                  | LGMVSSQLLAAPR    | EV | M  | VAVLRA  | VA         | MDS  | PKVEA    | MAS   | SLAMV | DSSQ        |      |
| <i>Miniopterus natalensis</i>          | LAMASSQLLAAPL    | EV | T  | VAVLRA  | AA         | TGS  | PRVGAMAN | SLAMV | DSSK  |             |      |
| <i>Canis lupus familiaris</i>          | LAMASSQLLAAPR    | EV | M  | VVVLR   | AA         | TGS  | PRVGAMAN | SLAMV | DSSK  |             |      |
| <i>Camelus ferus</i>                   | LAMASSQLLVAPR    | EA | T  | VAVLRA  | AA         | MAS  | PRAGA    | TAS   | SLAMV | DSSR        |      |
| <i>Camelus bactrianus</i>              | LAMASSQLLVAPR    | EA | T  | VAVLRA  | AA         | MAS  | PRAGA    | TAS   | SLAMV | DSSR        |      |
| <i>Equus przewalskii</i>               | LGMVSSQLLAAPR    | EV | M  | VAVLRA  | VA         | MDS  | PKVEA    | MAS   | SLAMV | DSSQ        |      |
| <i>Equus asinus</i>                    | LGMVSSQLLAAPR    | EV | M  | VAVLRA  | VA         | MDS  | PKVEA    | MAS   | SLAMV | DSSQ        |      |
| <i>Eptesicus fuscus</i>                | LAMASSQLLAAPR    | EV | T  | VAVLRA  | AA         | TGS  | PRVGAMAN | SLAMV | DSSK  |             |      |
| <i>Leptonyx chotes weddellii</i>       | LAMASSQLLAAPR    | EV | M  | VVVLR   | AA         | TGS  | PRVGAMAN | SLAMV | DSSK  |             |      |
| <i>Octodon degus</i>                   | LAMVSSQLPAAPR    | EV | M  | VVVLR   | AA         | MGS  | PRVGAMAN | SLAMV | DSSK  |             |      |
| <i>Otolemur gamettii</i>               | LAMASSQLLAAPR    | EV | T  | AAVLRA  | AA         | MDS  | PQVGAMAN | SPAMA | DSSR  |             |      |
| <i>Myotis davidii</i>                  | PAMASSPLLAARR    | EV | T  | VAVLRA  | AA         | TGS  | PRVGAMAN | SLAMV | DSSK  |             |      |
| <i>Dipodomys ordii</i>                 | LAMASSQLPAAPL    | EV | M  | VAVPRA  | AA         | MGS  | LKVGA    | TAN   | SLAME | DSSR        |      |
| <i>Jaculus jaculus</i>                 | LAMASSQLLAAPQ    | EV | M  | VAVLRA  | AA         | MGS  | PRVEA    | MAN   | SPAMV | DSSK        |      |
| <i>Ochotona princeps</i>               | LAMGSSQLPAAPQ    | EV | T  | AAVLRA  | AA         | TGS  | PRAGAMAS | NLAMA | DSSK  |             |      |
| <i>Microtus ochrogaster</i>            | LAMANSQLLAAPR    | EV | M  | VAVLRA  | AA         | MGS  | PRVEA    | MAN   | SPAMV | DSSK        |      |
| <i>Mus musculus</i>                    | LAMANSQLLAAPQ    | EV | M  | VAVLRA  | AA         | MGN  | PRVEA    | MVN   | SLAMV | DSSK        |      |
| <i>Peromyscus maniculatus bairdii</i>  | LAMANSQLLAAPR    | EV | T  | VAVLRA  | AA         | MGS  | PRVEA    | MAN   | SPAMV | DSSK        |      |
| <i>Erinaceus europaeus</i>             | PAMASSRLLATPL    | EV | T  | VAAVLR  | AA         | MGS  | PRAGA    | TGS   | SRA   | TA          | DSSK |
| <i>Nannospalax galili</i>              | LAMANNQLLAALQ    | EV | T  | VGA LRA | AA         | MGS  | PRVEA    | MAN   | SPAMV | DSSK        |      |
| <i>Panthalops hodgsonii</i>            | LAMVSSQLLAAPQ    | EV | T  | VAVLKA  | AA         | MGS  | PRVEA    | MAS   | SLAMV | GSSK        |      |
| <i>Drosophila melanogaster fusilli</i> | SGTLSYSV         |    |    |         |            |      |          |       |       |             |      |
| <i>Rattus norvegicus</i>               | PVMANSQLLVAPQ    | EV | M  | VEVLR   | AA         | MGS  | PRVEA    | MVN   | SLAMV | DSSK        |      |
| <i>Capra hircus</i>                    | LAMASSQLLAAPQ    | EV | T  | VAVLKA  | AA         | MGS  | PRVEA    | MAS   | SLAMV | GSSK        |      |
| <i>Brachypodium distachyon</i>         | MGTGREAPAPH      | SV | E  | VEV     |            |      |          |       |       |             |      |
| <i>Ovis aries</i>                      | LAMASSQLLAAPQ    | EV | T  | VAVLKA  | AA         | MGS  | PRVEA    | MAS   | SLAMV | GSSK        |      |
| <i>Ovis aries musimon</i>              | LAMASSQLLAAPQ    | EV | T  | VAVLKA  | AA         | MGS  | PRVEA    | MAS   | SLAMV | GSSK        |      |
| <i>Oreochromis niloticus</i>           | RAV              |    |    |         |            |      |          |       |       |             |      |
| <i>Haplochromis burtoni</i>            | RVV              |    |    |         |            |      |          |       |       |             |      |
| <i>Perkinsus marinus</i>               |                  | R  |    |         |            |      |          |       |       |             |      |
| <i>Eucalyptus grandis</i>              |                  |    |    |         |            |      |          |       |       |             |      |
| <i>Gallus gallus</i>                   | PATTSRPPPPAPP    | EA | M  | GAAPRA  | PA         | TGS  | RPAPA    | TAS   | SPHTA | ASSP        |      |
| <i>Sarcophilus hamsii</i>              | LATANSQLLAALP    | EA | M  | VAVPRA  | VA         | MDS  | PRVGAMV  | R     | SLVMV | DNKV        |      |
| <i>Ictalurus punctatus</i>             | PTASSRRLQLQA     | S  | P  | ALG     | HQ         | TGS  | REEAG    |       |       |             |      |
| <i>Anolis carolinensis</i>             | PA TPSSQ - PVPLL | EV | M  | APALKP  | PA         | MGN  | PRVEA    | MAS   | NQAMA | ASSR        |      |
| <i>Trypanosoma</i>                     | LADYNIQ          |    |    |         |            |      |          |       |       |             |      |
| <i>Xenopus tropicalis</i>              | LQPVA TQEVAAAPS  | HP | A  | TSSSRV  | VD         | MGS  | NQV      | GVMEV | SRA   | AMEVNSNP    |      |
| <i>Danio rerio</i>                     | LAITS - LLLHLLP  | VA | IA | AAALS   | LA         | TVSS | SSSRV    | E     | G     | MEAVV - VSL |      |
| <i>Latimeria chalumnae</i>             | LV TASSRPLRPPQ   | PA | T  | EEVAPS  | LPA        | TGS  | PRVEA    | MDR   | I     | LLTA - SSSR |      |
| <i>Xenopus laevis</i>                  | LATRSSLQPVATQ    | EV | M  | VGAPSP  | LPTSSNKVVD | MDS  | RQV      | GVMEA | SRA   | AMEVNNNP    |      |

bioRxiv preprint doi: <https://doi.org/10.1101/848580>; this version posted April 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

|  |                  |            |          |                          |              |                           |        |
|--|------------------|------------|----------|--------------------------|--------------|---------------------------|--------|
| <i>Homo sapiens</i>                    | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTA          | AV-VV                     |        |
| <i>Pan paniscus</i>                    | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTA          | AV-V                      |        |
| <i>Gorilla gorilla gorilla</i>         | AMV DSS          | KA         | IP       | LRAMDSRTS                | TTA          | AV-V                      |        |
| <i>Protopithecus</i>                   | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTA          | AV                        |        |
| <i>Nomascus leucogenys</i>             | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTA          | AA                        |        |
| <i>Pongo abelii</i>                    | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTT          | AA-AV                     |        |
| <i>Chlorocebus sabaeus</i>             | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTA          | AA-VV                     |        |
| <i>Rhinopithecus roxellana</i>         | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTA          | AA-VV                     |        |
| <i>Rhinopithecus bieti</i>             | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTA          | AA-VV                     |        |
| <i>Macaca mulatta</i>                  | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTA          | AA-VV                     |        |
| <i>Colobus angolensis palliatus</i>    | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTA          | AA-VV                     |        |
| <i>Macaca nemestrina</i>               | AM-DSS           | KA         | IP       | LRVMDRSRTS               | TTA          | AA-VV                     |        |
| <i>Macaca fascicularis</i>             | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTA          | AA-VV                     |        |
| <i>Mandrillus leucophaeus</i>          | AM-DSS           | KA         | IP       | LRVMDRSRTS               | TTA          | AA-VV                     |        |
| <i>Cercocebus atys</i>                 | AM-DSS           | RA         | IP       | LRAMDSRTS                | TTA          | AA-VV                     |        |
| <i>Papio anubis</i>                    | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTA          | AA-VV                     |        |
| <i>Callithrix jacchus</i>              | AT-DSS           | KAP        | IR       | LRAMDSRTS                | TTA          | AA-AAA                    |        |
| <i>Saimiri boliviensis boliviensis</i> | PT-DSS           | KAP        | IR       | LRAMDSRTS                | TTA          | AA-AAA                    |        |
| <i>Cebus capucinus imitator</i>        | PT-DSS           | KAP        | IR       | LRAMDSRTS                | TTA          | AA-AAA                    |        |
| <i>Sus scrofa</i>                      | AM-DSS           | KA         | IP       | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Physeter catodon</i>                | AM-DSS           | KA         | IP       | LRAMGSRSTS               | TTA          | AV-E                      |        |
| <i>Tursiops truncatus</i>              | AM-DSS           | KA         | IP       | LRAMGSRSTS               | TTA          | AV-E                      |        |
| <i>Orcinus orca</i>                    | AM-DSS           | KA         | IP       | LRAMGSRSTS               | TTA          | AV-E                      |        |
| <i>Balaenoptera acutorostrata</i>      | AM-DSS           | KA         | IP       | LRA TGSRTS               | TTV          | AV-E                      |        |
| <i>Sorex araneus</i>                   | AM-DSN           | KAP        | IHL      | KVMDRSRTS                | TTA          | AV-EV                     |        |
| <i>Aotus nancymae</i>                  | PT-DSS           | KAP        | IR       | LRAMDSRTS                | TTA          | AA-AAA                    |        |
| <i>Trichechus manatus latirostris</i>  | AT-DSS           | KAP        | IP       | LRAMDSRTS                | TTV          | AV-GA                     |        |
| <i>Toxotrypana minima</i>              |                  |            |          | LRTPA TL                 |              | AV                        |        |
| <i>Ceratotherium simum simum</i>       | AM-DSS           | QAP        | IP       | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Condylura cristata</i>              | AM-GSS           | KAP        | ITP      | LKAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Galeopterus variegatus</i>          | AT-DSS           | KAP        | IP       | LRAMDSRTS                | TTV          | AA-EV                     |        |
| <i>Odobenus rosmarus divergens</i>     | AM-DSS           | KAP        | IP       | LKAMDSRTS                | TTV          | AV                        |        |
| <i>Tarsius syrichta</i>                | AT-DSS           | KAP        | ITP      | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Microcebus murinus</i>              | TT-GSS           | KAP        | IP       | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Bison bison bison</i>               | AM-DSS           | RA         | ITR      | LRAMGSRSTS               | TTV          | AV-A                      |        |
| <i>Propithecus coquereli</i>           | AT-GSS           | KAP        | IP       | LRAMDSRTS                | TTV          | AV-V                      |        |
| <i>Bos mutus</i>                       | AM-DSS           | RA         | ITR      | LRAMGSRSTS               | TTV          | AV-A                      |        |
| <i>Pteropus vampyrus</i>               | AM-DSS           | KAP        | ILL      | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Pteropus alecto</i>                 | AM-DSS           | KAP        | ILL      | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Vicuqna pacos</i>                   | AM-DSS           | RA         | ITP      | LRAMDSRTS                | TTA          | AV                        |        |
| <i>Rousettus aegyptiacus</i>           | AM-DSS           | KAP        | ILL      | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Bos taurus</i>                      | AM-GSS           | RA         | ITR      | LRAMGSRSTS               | TTV          | AV-A                      |        |
| <i>Equus caballus</i>                  | AM-DSS           | QAP        | ILL      | LRAMDSRTS                | TTA          | AV-EV                     |        |
| <i>Miniopterus natalensis</i>          | AM-DSS           | KAP        | ITL      | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Canis lupus familiaris</i>          | AM-DSS           | KAP        | IP       | LKAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Camelus ferus</i>                   | AM-DSS           | RA         | ITP      | LRAMDSRTS                | TTA          | AV                        |        |
| <i>Camelus bactrianus</i>              | AM-DSS           | RA         | ITP      | LRAMDSRTS                | TTA          | AV                        |        |
| <i>Equus przewalskii</i>               | AM-DSS           | QAP        | ILL      | LRVMDRSRTS               | TTA          | AV-EV                     |        |
| <i>Equus asinus</i>                    | AM-DSS           | QAP        | ILL      | LRAMDSRTS                | TTA          | VV-EV                     |        |
| <i>Eptesicus fuscus</i>                | AM-DSS           | RAP        | ILL      | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Leptonychotes weddellii</i>         | AM-DSS           | KAP        | IP       | LKAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Octodon degus</i>                   | AT-DSS           | KAP        | ITP      | LRAMDSRTS                | TTA          | AV-A                      |        |
| <i>Otolemur gamettii</i>               | AT-GNS           | KA         | IP       | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Myotis davidii</i>                  | AM-DSS           | KVP        | ILL      | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Dipodomys ordii</i>                 | AM-DSS           | KAL        | IP       | LRAMDSRTS                | TTV          | AV-AE                     |        |
| <i>Jaculus jaculus</i>                 | AM-DSS           | KVP        | ITP      | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Ochotona princeps</i>               | AM-GSS           | KAP        | ITP      | LRA TGSRTS               | TTA          | AA-EVA                    |        |
| <i>Microtus ochrogaster</i>            | AM-DNN           | KAP        | ITH      | LRVMDNRS                 | TTA          | AV-EV                     |        |
| <i>Mus musculus</i>                    | AT-DNN           | KAP        | ITH      | LRVMDNRS                 | TTV          | AV-EV                     |        |
| <i>Peromyscus maniculatus bairdii</i>  | AM-DNS           | KAP        | ITH      | LRVMDNRS                 | TTA          | AV-EV                     |        |
| <i>Erinaceus europaeus</i>             | AT-DSS           | KAL        | ITP      | LRRA TGSRTS              | TTA          | AA-EVA                    |        |
| <i>Nannospalax galili</i>              | AM-DSN           | RAP        | IP       | LRAMDSRTS                | TTV          | AV-EV                     |        |
| <i>Panholops hodgsonii</i>             | AM-ASS           | RA         | ITR      | LRA TGSRTS               | TVV          | VV-AA                     |        |
| <i>Drosophila melanogaster fusilli</i> |                  |            |          | SRCSTGR                  |              | WT                        |        |
| <i>Rattus norvegicus</i>               | AT-DSN           | KAP        | ITH      | LRVMDNRS                 | TTA          | AV-EV                     |        |
| <i>Capra hircus</i>                    | AM-GSS           | RA         | ITR      | LRA TGSRTS               | TVV          | VV-AA                     |        |
| <i>Brachypodium distachyon</i>         |                  |            |          | GHLGE IMSA TV            |              | VL                        |        |
| <i>Ovis aries</i>                      | AM-GSS           | RA         | ITR      | LRA TGSRTS               | TAV          | VV-AA                     |        |
| <i>Ovis aries musimon</i>              | AM-GSS           | RA         | ITR      | LRA TGSRTS               | TAV          | VV-AA                     |        |
| <i>Oreochromis niloticus</i>           |                  |            |          | VGMV TQVSP TA            |              | PE                        |        |
| <i>Haplochromis burtoni</i>            |                  |            |          | VGMV TQVSP TA            |              | PV                        |        |
| <i>Perkinsus marinus</i>               | IT               | EG         | TISL     | G INMI IV-VTTMTTTTTRPTTR |              | TAL TML                   |        |
| <i>Eucalyptus grandis</i>              |                  |            |          | ETLLP                    |              | IMV                       |        |
| <i>Gallus gallus</i>                   | PT-ASS           | PH         | TARRRA   | TASRAS                   | TAT          | AA-P                      |        |
| <i>Sarcophilus hamsii</i>              | PM-DSS           | KAP        | ITP      | LKDMASRTS                | ITA          | AV-GA                     |        |
| <i>Ictalurus punctatus</i>             |                  |            |          |                          |              |                           |        |
| <i>Anolis carolinensis</i>             | AP-MGS           | S          | LPT      | THLRAMASKVSM             | MAV          | AV-AV                     |        |
| <i>Trypanosoma</i>                     | VC-RRC           | YA         | R        | LPVRASNC                 | RKKAC        |                           |        |
| <i>Xenopus tropicalis</i>              | MGASRVRVA        | MAVSS      | QALM-DSS | SNQA                     | TASPRAMAS    | SKHSIVVELVD IRTLLPWA EAMV |        |
| <i>Danio rerio</i>                     | EARAVDLGVAEANTSP | LSLEEGLTAH | LLT      | TA                       | PLHHKAMGNKVS | TDR                       |        |
| <i>Latimeria chalumnae</i>             |                  | PS         | RART     | GSS                      | RA           | TLLSRAMDS                 | SNLMAA |
| <i>Xenopus laevis</i>                  | MGASRARVA        | TAVSSSS    | QALM-DSS | NQA                      | TTS LRA      | IASSKH                    | IV     |

|  |    |              |  |    |     |              |               |                |
|--|----|--------------|--|----|-----|--------------|---------------|----------------|
| <i>Homo sapiens</i>                    |    |              |  |    | VE  | V - E        | V             |                |
| <i>Pan paniscus</i>                    |    |              |  |    | AE  |              | V             |                |
| <i>Gorilla gorilla gorilla</i>         |    |              |  |    | AE  |              | V             |                |
| <i>Pan troglodytes</i>                 |    |              |  |    | AE  |              | V             |                |
| <i>Nomascus leucogenys</i>             |    |              |  |    | VE  | V - E        | V             |                |
| <i>Pongo abelii</i>                    |    | V            |  |    | VE  | V - E        | V             |                |
| <i>Chlorocebus sabaeus</i>             |    |              |  |    | VE  |              | V             |                |
| <i>Rhinopithecus roxellana</i>         |    |              |  |    | VE  |              | V             |                |
| <i>Rhinopithecus bieti</i>             |    |              |  |    | VE  |              | V             |                |
| <i>Macaca mulatta</i>                  |    |              |  |    | VE  |              | V             |                |
| <i>Colobus angolensis palliatus</i>    |    |              |  |    | VE  |              | V             |                |
| <i>Macaca nemestrina</i>               |    |              |  |    | VE  |              | V             |                |
| <i>Macaca fascicularis</i>             |    |              |  |    | VE  |              | V             |                |
| <i>Mandrillus leucophaeus</i>          |    |              |  |    | VE  |              | V             |                |
| <i>Cercocebus atys</i>                 |    |              |  |    | VE  |              | V             |                |
| <i>Papio anubis</i>                    |    |              |  |    | VE  |              | V             |                |
| <i>Callithrix jacchus</i>              |    | AVG          |  |    | VE  | G - G        | V             |                |
| <i>Saimiri boliviensis boliviensis</i> |    | V            |  |    | VE  | V - G        | V             |                |
| <i>Cebus capucinus imitator</i>        |    | VVV          |  |    | EE  | V - G        | V             |                |
| <i>Sus scrofa</i>                      |    |              |  |    | VE  | G - V        | A             |                |
| <i>Physeter catodon</i>                |    |              |  |    | VE  | G - V        | V             |                |
| <i>Tursiops truncatus</i>              |    |              |  |    | VE  | G - V        | V             |                |
| <i>Orcinus orca</i>                    |    |              |  |    | VE  | G - V        | V             |                |
| <i>Balaenoptera acutorostrata</i>      |    |              |  |    | VE  | G - V        | V             |                |
| <i>Sorex araneus</i>                   |    |              |  |    | VA  | G - V        | V             |                |
| <i>Aotus nancymae</i>                  |    | VVV          |  |    | VE  | V - G        | V             |                |
| <i>Trichechus manatus latirostris</i>  |    |              |  |    | AE  | E - V        | VE            |                |
| <i>Toxonia minima</i>                  |    |              |  |    |     | V            | V             |                |
| <i>Ceratotherium simum simum</i>       |    |              |  |    | AE  | G - V        | V             |                |
| <i>Condylura cristata</i>              |    |              |  |    | AE  | G - V        | V             |                |
| <i>Galeopterus variegatus</i>          |    |              |  |    | AE  | G - V        | V             |                |
| <i>Odobenus rosmarus divergens</i>     |    |              |  |    |     |              |               |                |
| <i>Tarsius syrichta</i>                |    |              |  |    | AE  | G - A        | V             |                |
| <i>Microcebus murinus</i>              |    |              |  |    | AE  | G - A        | V             |                |
| <i>Bison bison bison</i>               |    |              |  |    | AE  | G - V        | A             |                |
| <i>Propithecus coquereli</i>           |    |              |  |    | AE  | G - V        | V             |                |
| <i>Bos mutus</i>                       |    |              |  |    | AE  | G - V        | A             |                |
| <i>Pteropus vampyrus</i>               |    |              |  |    | AE  | G - V        | V             |                |
| <i>Pteropus alecto</i>                 |    |              |  |    | AE  | G - V        | V             |                |
| <i>Vicuqna pacos</i>                   |    |              |  |    | VE  | G - V        | V             |                |
| <i>Rousettus aegyptiacus</i>           |    |              |  |    | AE  | G - V        | V             |                |
| <i>Bos taurus</i>                      |    |              |  |    | AE  | G - V        | A             |                |
| <i>Equus caballus</i>                  |    |              |  |    | AE  | G - V        | V             |                |
| <i>Miniopterus natalensis</i>          |    |              |  |    | AE  | G - V        | V             |                |
| <i>Canis lupus familiaris</i>          |    | V            |  |    | AE  | G - V        | V             |                |
| <i>Camelus ferus</i>                   |    |              |  |    | VE  | G - V        | V             |                |
| <i>Camelus bactrianus</i>              |    |              |  |    | VE  | G - V        | V             |                |
| <i>Equus przewalskii</i>               |    |              |  |    | AE  | G - V        | V             |                |
| <i>Equus asinus</i>                    |    |              |  |    | AE  | G - V        | V             |                |
| <i>Eptesicus fuscus</i>                |    |              |  |    | AE  | G - V        | E             |                |
| <i>Leptonychotes weddellii</i>         |    | V            |  |    | AE  | G - V        | V             |                |
| <i>Octodon degus</i>                   |    |              |  |    | VE  | G - V        | V             |                |
| <i>Otolemur gamettii</i>               |    | V            |  |    | VE  | G - V        | V             |                |
| <i>Myotis davidii</i>                  |    |              |  |    | E   | G - V        | E             |                |
| <i>Dipodomys ordii</i>                 |    |              |  |    | AE  | G - V        | V             |                |
| <i>Jaculus jaculus</i>                 |    |              |  |    | VE  | G - V        | V             |                |
| <i>Ochotona princeps</i>               |    |              |  |    | E   | G - V        | A             |                |
| <i>Microtus ochrogaster</i>            |    |              |  |    | VE  | G - V        | V             |                |
| <i>Mus musculus</i>                    |    |              |  |    | VE  | G - V        | V             |                |
| <i>Peromyscus maniculatus bairdii</i>  |    |              |  |    | VE  | G - V        | V             |                |
| <i>Erinaceus europaeus</i>             |    |              |  |    | E   | G - V        | A             |                |
| <i>Nannospalax galili</i>              |    |              |  |    | E   | G            | V             |                |
| <i>Panthalops hodgsonii</i>            |    |              |  |    | AE  | G - V        | A             |                |
| <i>Drosophila melanogaster fusilli</i> |    |              |  |    | P   | R - TM       |               |                |
| <i>Rattus norvegicus</i>               |    |              |  |    | VE  | G - V        | V             |                |
| <i>Capra hircus</i>                    |    |              |  |    | AE  | G - V        | A             |                |
| <i>Brachypodium distachyon</i>         |    |              |  |    | AE  |              |               |                |
| <i>Ovis aries</i>                      |    |              |  |    | AE  | G - V        | A             |                |
| <i>Ovis aries musimon</i>              |    |              |  |    | AE  | G - V        | A             |                |
| <i>Oreochromis niloticus</i>           |    |              |  |    |     |              |               |                |
| <i>Haplochromis burtoni</i>            |    |              |  |    |     |              |               |                |
| <i>Perkinsus marinus</i>               |    | EM           |  |    |     | I - LV       | IDS - S       |                |
| <i>Eucalyptus grandis</i>              |    |              |  |    | EE  | G - V        | V             |                |
| <i>Gallus gallus</i>                   |    |              |  |    | VE  | G - V        | V             |                |
| <i>Sarcophilus hamsii</i>              |    |              |  |    | VE  | G - V        | V             |                |
| <i>Ictalurus punctatus</i>             |    |              |  |    |     |              |               |                |
| <i>Anolis carolinensis</i>             |    |              |  |    | EE  | AAV          | VD            |                |
| <i>Trypanosoma</i>                     |    |              |  |    |     | G - H        | C             |                |
| <i>Xenopus tropicalis</i>              |    | EG           |  |    |     | L - IRVAMEAK | IEEGDVEDLVVVV |                |
| <i>Danio rerio</i>                     |    |              |  |    |     |              |               |                |
| <i>Latimeria chalumnae</i>             | VV | GMVAKSRVALGV |  | SR | ITA | AE           | G - E         | VDLAVE - DL IE |
| <i>Xenopus laevis</i>                  |    |              |  |    |     |              |               |                |

<https://doi.org/10.1101/848580>; this version posted April 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

|  |  |
|--|--|
| <i>Homo sapiens</i>                    | EV TM  |
| <i>Pan paniscus</i>                    | EV TM  |
| <i>Gorilla gorilla gorilla</i>         | EV TM  |
| <i>Pan troglodytes</i>                 | EV TM  |
| <i>Nomascus leucophaeus</i>            | EV TM  |
| <i>Pongo abelii</i>                    | EV TM  |
| <i>Chlorocebus sabaeus</i>             | EV AM  |
| <i>Rhinopithecus roxellana</i>         | EV TM  |
| <i>Rhinopithecus bieti</i>             | EV TM  |
| <i>Macaca mulatta</i>                  | EV AM  |
| <i>Colobus angolensis palliatus</i>    | EV TM  |
| <i>Macaca nemestrina</i>               | EV AM  |
| <i>Macaca fascicularis</i>             | EV AM  |
| <i>Mandrillus leucophaeus</i>          | EV AM  |
| <i>Cercocebus atys</i>                 | EV AM  |
| <i>Papio anubis</i>                    | EV AM  |
| <i>Callithrix jacchus</i>              | EV AM  |
| <i>Saimiri boliviensis boliviensis</i> | EV AT  |
| <i>Cebus capucinus imitator</i>        | EV TM  |
| <i>Sus scrofa</i>                      | EV VM  |
| <i>Physeter catodon</i>                | EV TM  |
| <i>Tursiops truncatus</i>              | EV TM  |
| <i>Orcinus orca</i>                    | EV TM  |
| <i>Balaenoptera acutorostrata</i>      | EV TM  |
| <i>Sorex araneus</i>                   | EV TM  |
| <i>Aotus nancymaae</i>                 | EV AM  |
| <i>Trichechus manatus latirostris</i>  | EV TM  |
| <i>Toxopneustes</i>                    | PLVP   |
| <i>Ceratotherium simum simum</i>       | EV AM  |
| <i>Condylura cristata</i>              | EV AM  |
| <i>Galeopterus variegatus</i>          | EV AM  |
| <i>Odobenus rosmarus divergens</i>     | EV TM  |
| <i>Tarsius syrichta</i>                | EV VM  |
| <i>Microcebus murinus</i>              | EV TM  |
| <i>Bison bison bison</i>               | EV AM  |
| <i>Propithecus coquereli</i>           | EV TM  |
| <i>Bos mutus</i>                       | EV AM  |
| <i>Pteropus vampyrus</i>               | EV TT  |
| <i>Pteropus alecto</i>                 | EV TT  |
| <i>Vicuña pacos</i>                    | EV AM  |
| <i>Rousettus aegyptiacus</i>           | EV TT  |
| <i>Bos taurus</i>                      | EV AT  |
| <i>Equus caballus</i>                  | EV AM  |
| <i>Miniopterus natalensis</i>          | EV TM  |
| <i>Canis lupus familiaris</i>          | EV TM  |
| <i>Camelus ferus</i>                   | EV AM  |
| <i>Camelus bactrianus</i>              | EV AM  |
| <i>Equus przewalskii</i>               | EV AM  |
| <i>Equus asinus</i>                    | EV AM  |
| <i>Eptesicus fuscus</i>                | EV TM  |
| <i>Leptonychotes weddellii</i>         | EV TM  |
| <i>Octodon degus</i>                   | EA IT  |
| <i>Otolemur gamettii</i>               | EV TM  |
| <i>Myotis davidii</i>                  | EV AM  |
| <i>Dipodomys ordii</i>                 | EA TM  |
| <i>Jaculus jaculus</i>                 | EA TM  |
| <i>Ochotona princeps</i>               | EG AM  |
| <i>Microtus ochrogaster</i>            | EA TM  |
| <i>Mus musculus</i>                    | EA TM  |
| <i>Peromyscus maniculatus bairdii</i>  | EA TM  |
| <i>Erinaceus europaeus</i>             | EV TM  |
| <i>Nannospalax galili</i>              | EA VM  |
| <i>Panholops hodgsonii</i>             | EV AM  |
| <i>Drosophila melanogaster fusilli</i> | SRAV   |
| <i>Rattus norvegicus</i>               | EA TM  |
| <i>Capra hircus</i>                    | EV AM  |
| <i>Brachypodium distachyon</i>         |  |
| <i>Ovis aries</i>                      | EV AM  |
| <i>Ovis aries musimon</i>              | EV AM  |
| <i>Oreochromis niloticus</i>           | AI TA  |
| <i>Haplochromis burtoni</i>            | AI TA  |
| <i>Perkinsus marinus</i>               | NSSS   |
| <i>Eucalyptus grandis</i>              | VGPM   |
| <i>Gallus gallus</i>                   | PAAM   |
| <i>Sarcophilus hamsii</i>              | EV AM  |
| <i>Ictalurus punctatus</i>             |  |
| <i>Anolis carolinensis</i>             | LV AM  |
| <i>Trypanosoma</i>                     | SNLR   |
| <i>Xenopus tropicalis</i>              | VVVEEEQVALTAAEGAAVVAEAAWAVVNGEDSVNLVDHVKAADPPDMKWPNK T T L T T T QSSCKDLERMLQLSQL          |
| <i>Danio rerio</i>                     | E - V T  |
| <i>Latimeria chalumnae</i>             | - - VEEEA - - VGDEEA - - AVAVAVAVVNVV - - - - - ASVNLVDPG - - - - -                        |
| <i>Xenopus laevis</i>                  | - - - - - GGAPA - - - - - VAEEAWAVVNGEDSVNLVDHMKVADPPPEMKWLNK T T L T T T QSLCKDLERMLQLSQL |

bioRxiv preprint doi: <https://doi.org/10.1101/848580>; this version posted April 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



|  |                          |      |
|--|--------------------------|------|
| <i>Homo sapiens</i>                    | AKI                      | NPP  |
| <i>Pan paniscus</i>                    | AKI                      | NPP  |
| <i>Gorilla gorilla gorilla</i>         | AKI                      | NPP  |
| <i>Pan troglodytes</i>                 | AKI                      | NPP  |
| <i>Nomascus leucogenys</i>             | AKI                      | NPP  |
| <i>Pongo abelii</i>                    | AKI                      | NPP  |
| <i>Chlorocebus sabaeus</i>             | AKI                      | NPP  |
| <i>Rhinopithecus roxellana</i>         | AKI                      | NPP  |
| <i>Rhinopithecus bieti</i>             | AKI                      | NPP  |
| <i>Macaca mulatta</i>                  | AKI                      | NPP  |
| <i>Colobus angolensis palliatus</i>    | AKI                      | SPP  |
| <i>Macaca nemestrina</i>               | AKI                      | NPP  |
| <i>Macaca fascicularis</i>             | AKI                      | NPP  |
| <i>Mandrillus leucophaeus</i>          | AKI                      | NPP  |
| <i>Cercocebus atys</i>                 | AKI                      | NPP  |
| <i>Papio anubis</i>                    | AKI                      | NPP  |
| <i>Callithrix jacchus</i>              | AKI                      | SPP  |
| <i>Saimiri boliviensis boliviensis</i> | AKI                      | SPP  |
| <i>Cebus capucinus imitator</i>        | AKI                      | SPP  |
| <i>Sus scrofa</i>                      | AKI                      | SPP  |
| <i>Physeter catodon</i>                | AKI                      | SPP  |
| <i>Tursiops truncatus</i>              | AKI                      | SPP  |
| <i>Orcinus orca</i>                    | AKI                      | SPP  |
| <i>Balaenoptera acutorostrata</i>      | AKI                      | SPP  |
| <i>Sorex araneus</i>                   | AKI                      | SPP  |
| <i>Aotus nancymae</i>                  | AKI                      | SPP  |
| <i>Trichechus manatus latirostris</i>  | AKI                      | SPP  |
| <i>Toxopneustes minima</i>             | TMM                      | MPP  |
| <i>Ceratotherium simum simum</i>       | AKI                      | SPP  |
| <i>Condylura cristata</i>              | AKI                      | NPP  |
| <i>Galeopterus variegatus</i>          | AKI                      | SPP  |
| <i>Odobenus rosmarus divergens</i>     | AKI                      | SPP  |
| <i>Tarsius syrichta</i>                | AKI                      | SLP  |
| <i>Microcebus murinus</i>              | AKI                      | SPP  |
| <i>Bison bison bison</i>               | AKI                      | SPP  |
| <i>Propithecus coquereli</i>           | AKI                      | SPP  |
| <i>Bos mutus</i>                       | AKI                      | SPP  |
| <i>Pteropus vampyrus</i>               | VKI                      | NPQ  |
| <i>Pteropus alecto</i>                 | VKI                      | NPQ  |
| <i>Vicuña pacos</i>                    | AKT                      | SPP  |
| <i>Rousettus aegyptiacus</i>           | VKI                      | NPP  |
| <i>Bos taurus</i>                      | AKI                      | SPP  |
| <i>Equus caballus</i>                  | AKI                      | SPP  |
| <i>Miniopterus natalensis</i>          | AKI                      | SPP  |
| <i>Canis lupus familiaris</i>          | AKI                      | SPP  |
| <i>Camelus ferus</i>                   | AKT                      | SPP  |
| <i>Camelus bactrianus</i>              | AKT                      | SPP  |
| <i>Equus przewalskii</i>               | AKI                      | SPP  |
| <i>Equus asinus</i>                    | AKI                      | SPP  |
| <i>Eptesicus fuscus</i>                | AKI                      | SPP  |
| <i>Leptonychotes weddellii</i>         | AKI                      | SPP  |
| <i>Octodon degus</i>                   | AKI                      | SPP  |
| <i>Otolemur gamettii</i>               | AKI                      | SPP  |
| <i>Myotis davidii</i>                  | AKI                      | SPP  |
| <i>Dipodomys ordii</i>                 | AKI                      | SPP  |
| <i>Jaculus jaculus</i>                 | VKI                      | SPP  |
| <i>Ochotona princeps</i>               | AKT                      | SPL  |
| <i>Microtus ochrogaster</i>            | AKI                      | SPP  |
| <i>Mus musculus</i>                    | AKI                      | SPL  |
| <i>Peromyscus maniculatus bairdii</i>  | AKI                      | SPP  |
| <i>Erinaceus europaeus</i>             | AKT                      | SPP  |
| <i>Nannospalax galili</i>              | AKI                      | SPP  |
| <i>Panholops hodgsonii</i>             | AKI                      | SPP  |
| <i>Drosophila melanogaster fusilli</i> | GTA                      | SRR  |
| <i>Rattus norvegicus</i>               | AKI                      | SPL  |
| <i>Capra hircus</i>                    | AKI                      | SPP  |
| <i>Brachypodium distachyon</i>         |                          | ALG  |
| <i>Ovis aries</i>                      | AKI                      | SPP  |
| <i>Ovis aries musimon</i>              | AKI                      | SPP  |
| <i>Oreochromis niloticus</i>           | ATS                      | RPA  |
| <i>Haplochromis burtoni</i>            | ATS                      | RPA  |
| <i>Perkinsus marinus</i>               | RTT                      | ITP  |
| <i>Eucalyptus grandis</i>              | V - AGV LMV LDW LGHSEVNP |      |
| <i>Gallus gallus</i>                   | GRS                      | RPR  |
| <i>Sarcophilus hamsii</i>              | AKI                      | NHL  |
| <i>Ictalurus punctatus</i>             |                          | TAE  |
| <i>Anolis carolinensis</i>             | ARI                      | SPL  |
| <i>Trypanosoma</i>                     | MKK                      | KLR  |
| <i>Xenopus tropicalis</i>              | LTTSNKLESL - R - LTRRL   | GS L |
| <i>Danio rerio</i>                     | TKI                      | AHQ  |
| <i>Latimeria chalumnae</i>             | IKEDQG                   | MMQ  |
| <i>Xenopus laevis</i>                  | LRTTSNRLA                | SSR  |

bioRxiv preprint doi: <https://doi.org/10.1101/848580>; this version posted April 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.