

***Arabidopsis* HAP2/GCS1 is a gamete fusion protein**

homologous to somatic and viral fusogens

Clari Valansi^{1, #}, David Moi^{2, #}, Evgenia Leikina³, Elena Matveev¹, Martín Graña⁴,
Leonid V Chernomordik³, Héctor Romero⁵, Pablo S. Aguilar^{2, *}, Benjamin
Podbilewicz^{1, *}

¹Department of Biology, Technion- Israel Institute of Technology, Haifa 32000, Israel

²Laboratorio de Biología Celular de Membranas, Instituto de Investigaciones
Biotecnológicas "Dr. Rodolfo A. Ugalde" (IIB), Universidad Nacional de San Martín
(UNSAM), Buenos Aires, CP(1650), Argentina

³Section on Membrane Biology, Laboratory of Cellular and Molecular Biophysics,
Eunice Kennedy Shriver National Institute of Child Health and Human Development,
National Institutes of Health, Bethesda MD 20892, USA

⁴Unidad de Bioinformática, Institut Pasteur Montevideo, Montevideo 11400, Uruguay

⁵Laboratorio de Organización y Evolución del Genoma, Unidad de Genómica
Evolutiva, Dpto. Ecología y Evolución, Facultad de Ciencias/C.U.R.E., Universidad
de la República, Montevideo, Uruguay

[#]These authors contributed equally to this work

*Correspondence to: paguilar@iib.unsam.edu.ar podbilew@technion.ac.il

Running title: HAP2/GCS1 is a gamete fusion protein

25 **Abstract**

26 Cell-cell fusion is inherent to any form of sexual reproduction. Loss of HAPLESS
 27 2/GENERATIVE CELL SPECIFIC 1 (HAP2/GCS1) proteins results in gamete fusion
 28 failure in different organisms but their exact role is unclear. Here we show that
 29 *Arabidopsis* HAP2/GCS1 expression in mammalian cells is sufficient to promote cell-
 30 cell fusion. Hemifusion and complete fusion depend on HAP2/GCS1 presence in
 31 both fusing cells. Furthermore, expression of HAP2 on the surface of pseudotyped
 32 vesicular stomatitis virus and on the target cells results in HAP2-dependent virus-cell
 33 fusion. This bilateral requirement can be bypassed by replacing the plant gene with
 34 *C. elegans* EFF-1 somatic cell fusogen in one of the fusing cells or the virus,
 35 indicating that HAP2/GCS1 and EFF-1 share a similar fusion mechanism. Structural
 36 modeling of the HAP2/GCS1 protein family predicts that they are homologous to
 37 EFF-1 and class II fusion proteins from enveloped viruses (e.g. dengue and Zika
 38 viruses). We name this superfamily FUSEXINS: FUSion proteins essential for sexual
 39 reproduction and EXoplasmic merger of plasma membranes. Thus, Fusexins unify
 40 the origin and evolution of sexual reproduction, enveloped virus entry into cells and
 41 somatic cell fusion.

42

43 **Introduction**

44 While proteins mediating cell-cell fusion in tissues have been demonstrated in the
 45 placenta of mammals (Syncytins) and in organs of invertebrates (e.g. EFF-1 in *C.*
 46 *elegans*), the machinery mediating sperm-egg fusion remains unknown (Aguilar et
 47 al., 2013; Bianchi et al., 2014; Maruyama et al., 2016). HAP2/GCS1 proteins have
 48 been implicated as potential gamete fusogens in *Arabidopsis* (Johnson et al., 2004;
 49 Mori et al., 2006; von Besser et al., 2006), *Chlamydomonas* (Liu et al., 2008),
 50 *Tetrahymena* (Cole et al., 2014), *Dictyostelium* (Okamoto et al., 2016) and
 51 *Plasmodium* (Liu et al., 2008). However the precise function of HAP2/GCS1 in
 52 gamete fusion is unknown. So far, there is no functional or structural evidence

53 indicating HAP2/GCS1 is directly involved in cell-cell fusion. Proteins may function as
54 direct fusogens, or alternatively, they may affect communication or intimate adhesion
55 before fusion takes place, as demonstrated for other gamete fusion candidates such
56 as Juno and Izumo receptors (Bianchi et al., 2014).

57

58 **Results and Discussion**

59 To determine whether HAP2/GCS1 is an authentic fusion protein we first tested
60 whether *Arabidopsis* HAP2 (AtHAP2) could fuse heterologous cells that normally do
61 not fuse. For this, we transfected Baby Hamster Kidney (BHK) cells with plasmids
62 encoding AtHAP2, EFF-1, or red (RFP) or green (GFP) fluorescent proteins as
63 negative controls, and assayed the extent of cell-cell fusion (**Fig. 1 A**). In controls,
64 when BHK cells were transfected with cytoplasmic RFP (RFPcyto-BHK) and mixed
65 with GFP-transfected BHK cells (GFP-BHK; **Fig. 1 B (i)**) about 5% of cells (red or
66 green, respectively) had two nuclei due to cell division and only 1.5% of the cells
67 expressed both GFP and RFPcyto out of the total GFP/RFPcyto expressing cells in
68 contact (**Fig. 1 C**). This apparent cytoplasmic content mixing could be due to
69 phagocytosis of fluorescent apoptotic bodies or background fusion. In contrast, when
70 AtHAP2 was transfected into BHK cells with either RFPcyto or GFP and the
71 transfected cells were co-incubated, we observed an average multinucleation of 33%
72 ± 3 and 41.3% ± 1.3 (green or red), and cytoplasmic content mixing in 11.3% ± 0.9 in
73 three independent experiments (**Fig. 1 B (ii, iv) and C**). Similar results were obtained
74 using the previously defined *C. elegans*' somatic cell fusogen EFF-1 (**Fig. 1B (iii)**
75 **and C**). To test whether some of the multinucleated cells resulted from faulty cell
76 division (e.g. nuclear division without cytokinesis) we incubated BHK-HAP2 cells with
77 fluorodeoxyuridine (FdUrd) to arrest the cell cycle at the G1/S transition. We found
78 that FdUrd failed to decrease multinucleation in HAP2-BHK cells, and actually
79 doubled the multinucleation, probably due to a block in cell division of multinucleated
80 cells formed by cell-cell fusion. In contrast, FdUrd did not significantly affect nuclei

81 content of BHK cells that were transfected with the GFP plasmid only (**Table S1**).
 82 Time-lapse microscopy experiments showed that mononucleated BHK-
 83 HAP2(RFPcyto) cells formed syncytia by merging their cytoplasm (Fig. 1 D; Videos
 84 1-4; Fig. S1 A). Thus, AtHAP2 expression in BHK cells is sufficient to promote cell-
 85 cell fusion, defining this protein as a bona fide fusogen.

86 We asked whether HAP2/GCS1 family members display any similarity to
 87 known fusogenic proteins. HAP2/GCS1 are type I membrane glycoproteins
 88 composed of an N-terminal signal peptide, a large ectodomain, and a C-terminal
 89 cytoplasmic tail (Fig. 2 A). The ectodomain contains a conserved 50 amino acid
 90 region (Pfam PF10699, H/G domain, Fig. 2 A) that is unique to the HAP2/GCS1
 91 protein family (Wong and Johnson, 2010). In search of structural similarities, we
 92 filtered and compiled ectodomains from all members of the HAP2/GCS1 family and
 93 subjected this dataset to two homology detection and structure prediction algorithms,
 94 HHblits (Remmert et al., 2012; Soding, 2005), and LOMETS (Wu and Zhang, 2007).
 95 We detected structural similarity to ectodomains of the EFF-1 protein from *C.*
 96 *elegans* (Pérez-Vargas et al., 2014) and class II viral fusogenic glycoproteins
 97 (Harrison, 2008; Igonet and Rey, 2012; Kielian et al., 2010; Podbilewicz, 2014; White
 98 et al., 2008) from three viral families, Flaviviridae, Togaviridae and Bunyaviridae (Fig.
 99 S2 A, B). Both methods detected similarity to class II fusogens throughout the entire
 100 HAP2/GCS1 family, consistent with their expected role in gamete fusion in
 101 evolutionary distant species (Fig. S2 C and D). To generate structural models of
 102 AtHAP2 we used I-TASSER (Yang et al., 2015). The overall fold of the AtHAP2
 103 ectodomain model showed the canonical architecture of class II fusogens with three
 104 domains: a β -barrel domain I (DI), a mostly β -stranded elongated domain (DII), and
 105 the Immunoglobulin (Ig)-like C2-set topology module (DIII), separated by a linker
 106 from DI (Fig. 2 B). Global protein structure comparisons between these proteins, as
 107 measured by Z-score (Holm et al., 2008) and TM-score (Wu et al., 2007), indicate
 108 high similarity, with scores typical for proteins belonging to the same fold (Fig. 2 C;

109 **Fig. S3 A**). Structural models with similar architecture were also obtained for
 110 HAP2/GCS1 proteins from several species (**Fig. S3 C**). An unrooted tree inferred
 111 from a structural dissimilarity matrix shows the structural relationship between
 112 families (**Fig. 2 D**). Although amino acid sequence conservation between AtHAP2,
 113 somatic and viral class II fusogens is low, most β -strands of AtHAP2 ectodomain are
 114 conserved and arranged in the same way as in *C. elegans* EFF-1 and viral fusogens
 115 (**Fig. 2 E**) (Pérez-Vargas et al., 2014). AtHAP2 also shares conserved cysteine
 116 residues with class II fusogens, and a prominent loop at the tip of DII (the cd loop,
 117 **Fig. 2 E**). In viral fusogens this loop is highly hydrophobic and necessary for
 118 anchoring at the host cell membrane (Podbilewicz, 2014), whereas in EFF-1 it is
 119 negatively charged (Pérez-Vargas et al., 2014). Unlike EFF-1, the AtHAP2 predicted
 120 cd loop is hydrophobic and flanked by charged residues (**Fig. S3 D**). Two additional
 121 loops neighbour the AtHAP2 cd loop (**Fig. 2**). One of these loops (between the i and j
 122 β -strands) forms part of the conserved H/G domain, is hydrophobic and point
 123 mutations at equivalent residues of *Chlamydomonas* HAP2/GCS1 block trafficking to
 124 the plasma membrane (Liu et al., 2015).

125 Class II fusion proteins mediate exoplasmic membrane fusion either
 126 unilaterally (e.g. from the virus envelope to the cell membrane) or bilaterally (e.g.
 127 EFF-1-mediated cell fusion (Podbilewicz, 2014; Podbilewicz et al., 2006)). HAP2 is
 128 essential in sperm for double fertilization in *Arabidopsis* (Johnson et al., 2004; Mori et
 129 al., 2006; von Besser et al., 2006). In *Chlamydomonas* and *Plasmodium* HAP2 is
 130 also active in male gametes (Liu et al., 2008). However, in the seven-sexed
 131 *Tetrahymena* efficient fertilization requires the presence of HAP2 in both fusing
 132 gametes (Cole et al., 2014). To test whether HAP2-mediated cell fusion requires
 133 HAP2 presence in either one or both fusing cells, we mixed HAP2-BHK-RFPcyto
 134 cells with BHK-GFP cells and determined the percentage of multinucleation in cells
 135 expressing RFPcyto, GFP or both. Whereas HAP2-BHK-RFPcyto cells had 35% $\pm$
 136 1.9 multinucleation, cytoplasmic content mixing between HAP2-BHK-RFPcyto and

137 BHK-GFP cells was only $2\% \pm 0.3$, undistinguishable from the negative control
138 transfections (**Fig. 3 A**). These findings indicate that AtHAP2 is required in both cells
139 for fusion to occur (**Fig. 3 B**).

140 We previously demonstrated that EFF-1 and its paralog AFF-1 can fuse BHK
141 cells in a heterotypic way (Avinoam et al., 2011). To test whether AtHAP2 can
142 promote cell fusion in trans with the nematode protein EFF-1, we mixed HAP2-BHK-
143 RFPcyto cells with EFF-1-BHK-GFP cells and looked for cells containing both
144 RFPcyto and GFP. We found that hybrids formed as efficiently as with EFF-1 in all
145 fusing cells ($10.7\% \pm 0.7$; **Fig. 3 A-C**). Thus, EFF-1 can substitute for HAP2 in one of
146 the cell membranes.

147 The extent of cell fusion is proportional to the level of surface expression of
148 fusion proteins (e.g. EFF-1 and AFF-1) (Avinoam et al., 2011; Gattegno et al., 2007;
149 Podbilewicz et al., 2006; Sapir et al., 2007). To determine whether AtHAP2 is
150 expressed on the plasma membrane we used surface biotinylation followed by
151 immunoblotting and immunofluorescence using an antibody against the extracellular
152 domain of HAP2 in intact and permeabilized BHK cells. Following surface
153 biotinylation we detected low amount of HAP2-V5 on the surface compared to EFF-
154 1-V5 (**Fig. 3 D**). Immunofluorescence showed that AtHAP2 localized to puncta on
155 the surface of 20% of the expressing BHK-AtHAP2-YFP cells (**Fig. 3 E; Fig. S1 B**
156 **and C**). Since lower surface expression of AtHAP2, compared to EFF-1, resulted in
157 similar fusogenic activity of both proteins, we conclude that AtHAP2 is a more
158 efficient fusion protein compared to EFF-1.

159 Previously we demonstrated that eukaryotic fusogens can be assayed
160 efficiently by replacing the Vesicular Stomatitis Virus (VSV) fusion glycoprotein G
161 (VSVG) with a foreign fusogen and testing for the infectivity of the resulting
162 pseudovirus (Avinoam et al., 2011). To test whether AtHAP2 can functionally
163 substitute for VSVG, the fusogen glycoprotein G gene was deleted (VSVΔG) and
164 replaced by HAP2 (VSVΔG-HAP2; **Fig. 4 A**). We first infected BHK-HAP2 cells with

165 VSVΔG-G pseudoviruses and collected VSVΔG-HAP2 pseudoviruses from the
166 supernatants. As controls we generated VSVΔG-EFF-1 particles (Avinoam et al.,
167 2011). To test whether VSVΔG-HAP2 can infect cells we inoculated BHK cells, naïve
168 or expressing HAP2 or EFF-1 (**Fig. 4 B**). We observed a 7 to 70-fold increase in
169 infection efficiency in the fusogen-expressing cells compared to naïve cells (**Fig. 4**
170 **C**). Finally, to determine whether EFF-1 and HAP2 can interact through a bilateral
171 mechanism in this virus-cell fusion system, we infected BHK-EFF-1 with VSVΔG-
172 HAP2 and obtained similar infection efficiency to homotypic infection of BHK-HAP2.
173 Likewise, VSVΔG-EFF-1 pseudotyped viruses showed similar infection efficiencies
174 when infecting BHK-HAP2 or BHK-EFF-1 cells (**Fig. 4 C**). These finding reinforce our
175 conclusion from cell-cell fusion assays (**Fig. 3**) and further suggest that HAP2 and
176 EFF-1 directly interact to mediate heterotypic membrane fusion.

177 We then examined whether AtHAP2-mediated fusion proceeds via
178 hemifusion, a fusion intermediate in which outer leaflets of two membranes are
179 already merged but the inner leaflets remain distinct (Chernomordik and Kozlov,
180 2005). To this end, cells transfected with AtHAP2 were labeled with both cell tracker
181 as a cytosolic probe and Dil as a membrane probe, and were plated together with
182 transfected but unlabeled cells. In this experimental design, cell fusion caught at a
183 stage between hemifusion and fusion is expected to produce cells that acquired
184 membrane- but not content- probe (**Fig. 5 A**). Indeed we observed the appearance of
185 such cells (**Fig. 5 B; arrows**). Importantly, the number of cells labeled with only
186 membrane probe was much lower in experiments in which the labeled cells
187 expressing AtHAP2 were co-plated with unlabeled and non-transfected cells (**Fig. 5**
188 **C**). The apparent hemifusion in this control (~1.25%) could reflect background fusion
189 or overlapping cells. Our results indicate that HAP2-mediated fusion proceeds
190 through hemifusion intermediates. Our data are also the first evidence that a
191 homotypic fusogen has to be present in both membranes to mediate even

hemifusion, a fusion stage that is less energy demanding than opening and expansion of a fusion pore (Chernomordik and Kozlov, 2005).

Together, our findings indicate a somatic cell-like bilateral mechanism for AtHAP2 rather than a unilateral viral-like mechanism. Such bilateral mechanism is likely to operate not only in flowering plants but also in other organisms expressing members of the HAP2/GCS1 family (**Fig. 5 D, E and F**).

The observed structural and functional similarities between the gamete HAP2/GCS1 proteins, EFF-1 fusogen and class II viral fusion proteins could result from either common ancestry or convergence. Our data suggest a common ancestry based on the following observations: first, HAP2, EFF-1 and class II viral fusion proteins share a common domain architecture (**Fig. 2**). Second, proteins of the same fold are likely to be homologues with only a small fraction having different ancestries (Forslund et al., 2008). Third, comparisons of HMMs-based profiles of the complete ectodomain from different members of class II viral fusogens, HAP2 and FF (EFF-1/AFF-1) show low but significant amino acid sequence similarity between members of different groups (**Fig. S3 B**). Altogether, the combined conservation of structure, sequence and function makes a robust case for considering these proteins as diverged from a common ancestor.

Based on the ancient origin and function of exoplasmic membrane fusion we suggest naming this superfamily of fusion proteins the FUSEXINS (FUSion proteins essential for SEXual reproduction and EXoplasmic merger of plasma membranes). In conclusion, we have demonstrated the fusogenic activity of a fertilization protein and found a molecular link unifying sexual, somatic and viral membrane fusion. We hypothesize that Fusexins are ancient proteins and their emergence –either in viruses or cells— may have been a key innovation spurring sexual reproduction, exoplasmic somatic membrane fusion, and the rise of eukaryotes more than 2.5 billion years ago.

220 **Materials and Methods**

221 **Homology detection.** To detect structural homology in the HAP2/GCS1 sequences
 222 annotated in the InterPro database were taken as the HAP2 dataset (IPR018928). A
 223 total of 299 amino acid sequences were analyzed with TOPCONS (Tsirigos et al.,
 224 2015) to predict membrane topology. The resulting 299 ectodomain amino acid
 225 sequences were then used as queries against HHblits (Remmert et al., 2012) and
 226 LOMETS (Wu and Zhang, 2007). The Uniprot20 database (available on the Söding
 227 group server) was used to build Hidden Markov Models (HMMs) with one iteration of
 228 HHblits using default parameters. A subsequent search was performed using these
 229 HMMs against the PDB70 database (also available on the Söding group server) with
 230 default parameters to generate the final structural prediction using HMM-HMM
 231 alignment. To collect the results of the HHblits pipeline on all sequences, the output
 232 files were parsed with a script using CSB toolbox (Kalev et al., 2012). LOMETS, a
 233 meta algorithm that is part of the I-TASSER suite (Yang et al., 2015), was also used
 234 to predict structural homology for the 299 HAP2/GCS1 ectodomains. The standalone
 235 version uses six structural threading algorithms which are all variations of MUSTER
 236 (Wu and Zhang, 2008). LOMETS was run with default parameters for each HAP2
 237 ectodomain sequence. LOMETS output files were parsed with a custom parser
 238 written in python to compile alignment template codes, columns aligned and z-scores
 239 (online supplemental material). All structural predictions using LOMETS and HHblits
 240 were performed locally. TOPCONS topology predictions were run on the TOPCONS
 241 server. Complete lists of structural prediction results are available in **Table S2**.

242

243 **Models Generation.** Structural models were generated for a taxonomically diverse
 244 subset of ectodomain sequences using the I-TASSER server (Yang et al., 2015).
 245 EFF-1 and HAP2 ectodomains were prepared by removing signal sequences,
 246 transmembrane domains and C-terminal intracellular domains in the same way as for
 247 the structural prediction pipelines. Models were built from the ectodomains of

248 sequences from the HAP2/GCS1 family available in the Uniprot database for
 249 *Arabidopsis thaliana* (F4JP36), *Hydra vulgaris* (A3FEQ2), *Physarium polycephalum*
 250 (Q2PGG5), *Chlamydomonas reinhardtii* (A4GRC6), *Trypanosoma brucei gambiense*
 251 (D0A4L4) and *Tetrahymena thermophile* (A0A060A682). Additional models for the
 252 AFF-1/EFF-1 family were built from *Naegleria gruberi* (D2W008), *Strigamia maritima*
 253 (T1ISJ2), *Caenorhabditis elegans* (G5EGL9), and *Branchiostoma floridae* (C3YJ9)
 254 ectodomains. Structural constraints based on probable disulfide bonds were used as
 255 additional inputs on the I-TASSER server for subsequent models for the *Arabidopsis*
 256 HAP2 ectodomain. These bonds were inferred from residue conservation in MSAs of
 257 HAP2/GCS1 ectodomains and visual inspection of the preliminary model. I-TASSER
 258 was run a second time on the *Arabidopsis* ectodomain sequence using these
 259 structural constraints to generate the final model presented in this publication. All of
 260 the PDB models generated using I-TASSER are available in the online supplemental
 261 material.

262

263 **Structure-based multiple sequence alignment.** Structural alignments were used to
 264 generate high quality MSAs and pairwise superpositions of structures for use with
 265 structural similarity metrics. FATCAT (Ye and Godzik, 2004) pairwise alignments
 266 were run on all pairs of monomers in the I-TASSER models and crystal structures.
 267 Using a custom parser of the FATCAT output, pairwise alignments were generated in
 268 Fasta format. These alignments were iteratively merged using Clustal omega
 269 (Sievers et al., 2011) so that in each merger at least one sequence was shared
 270 between the Fasta files being merged until one a global alignment containing at least
 271 one of each of the sequences for the structures being aligned is produced. These
 272 final Fasta format alignments contain multiple sequence alignment possibilities for
 273 the set structures fed to the algorithm. Since all of the possibilities of sequence
 274 alignment are relatively similar, the final version of the alignment is chosen in Jalview
 275 (Waterhouse et al., 2009) by manual inspection.

276

277 **Structural similarity and structural phylogeny.** Structural similarity matrices were
 278 built using the Dali Z-score (Holm and Rosenstrom, 2010) and the TM-score (Zhang
 279 and Skolnick, 2005) using pairwise superpositions generated by FATCAT. It has
 280 been shown that Z-scores above 2 (Holm and Rosenstrom, 2010) and TM-scores
 281 over 0.5 (Xu and Zhang, 2010) are indicative of proteins belonging to the same
 282 Folds, usually homologous. TM-score matrix was transformed into a distance matrix
 283 with the metric $\text{dist} = 1 - \text{TM-score}$. This structure distance matrix was used to infer a
 284 phylogeny of ectodomains using FASTME (Lefort et al., 2015) on the default settings.
 285 The automation script made for the generation of structure based phylogenies is
 286 available in online supplemental material.

287

288 **Profile based phylogeny.** A taxonomically diverse set of ectodomain sequences of
 289 fusogens were used as search queries for HHblits against the Uniprot20 database to
 290 generate profiles, as hidden Markov models (HHM Format) incorporating also
 291 secondary structure information. For model construction realignment greediness
 292 (mact) was set to 10^{-1} and HHblits was run for 3 iterations on otherwise default
 293 parameters. These models were used in an all versus all search using HHsearch.
 294 Realignment greediness (mact) and secondary structure weight (ssw) in realignment
 295 in the All versus all HHsearch alignments were set to 10^{-4} and 3×10^{-1} , respectively.
 296 The automation script made for the generation of profile based phylogenies is
 297 available in online supplemental material.

298

299 **Installation and configuration of structural similarity analysis and profile-to-**
 300 **profile tools.** Python scripts used for structural and profile-to-profile analyses, are
 301 available in online supplemental material. They depend on SciPy, NumPy, Pandas,
 302 Biopython and the CSB toolbox python libraries. The structural phylogeny script
 303 requires the installation of TMalign (available on the Zhang lab server) and the PDB

304 protein comparison tool (available on the PDB server). The profile-to-profile analysis
305 script requires the installation of HHblits and HHsearch along with the uniprot20 and
306 PDB70 databases (available on the Söding group server). These scripts were built
307 for use in an Ubuntu 14.04LTS environment and have not been tested using other
308 operating systems.

309

310 **Electrostatic calculations.** To calculate electrostatic properties of the protein
311 surfaces, the Adaptive Poisson-Boltzmann Solver (Baker et al., 2001) was used with
312 default parameters, using atomic coordinates previously prepared with PDB2PQR
313 (Dolinsky et al., 2007). Solvent-accessible surfaces were coloured by electrostatic
314 potential and finally rendered in PyMol (<https://sourceforge.net/projects/pymol/>).

315

316 **Cells and reagents.** Baby Hamster Kidney cells (BHK) were BHK-21 (ATCC). BHK
317 cells were grown and maintained according to standard protocols using Dulbecco's
318 modified Eagle's medium (DMEM), supplemented with 10% fetal bovine serum (FBS,
319 Biological Industries, Kibbutz Beit Haemek, Israel), 100 U/ml penicillin and 100 µg/ml
320 streptomycin (Biological Industries), 2 mM L-glutamine (Biological Industries), 1 mM
321 sodium pyruvate (Gibco), 30 mM HEPES buffer pH 7.3 at 37°C (Biological
322 Industries). Cells were grown at 37° C with 5% CO₂. Cells were transfected using a
323 mixture of 8 µl Eugene HD (Promega) and 112 µl Opti-MEM (Gibco) with plasmids
324 specified below in 1 ml for every 35 mm plate (Nunc) without
325 replacing the medium (Avinoam et al., 2011).

326

327 **DNA constructs.** The *Arabidopsis* HAP2 plasmids were derived from
328 AtHAP2promoter::HAP2cds::YFP (PGL290; GenBank AF234315) (von Besser et al.,
329 2006) that fully rescues *hap2(-)* *Arabidopsis* mutants (Wong et al., 2010). HAP2 ORF
330 was amplified from PGL290 by PCR using primers

331 5'GCACTAGTATGGTGAACGCGATTTTAAT and
332 TCCCGCGGACTCTCACGTAGTCTTTGTT. The amplified AtHAP2 fragment was
333 subcloned into pIZT (Invitrogen) following SpeI and SacII digestion to obtain
334 pIZT::AtHAP2-V5-HIS used for expression in insect cells. AtHAP2-V5 was amplified
335 from pIZT::AtHAP2-V5-HIS using 5' CGTTATCGATATGGTGAACGCGATTTTAAT
336 and GCCCGGGTCACGTAGAATCGAGACCGA. The AtHAP2-V5 fragment was
337 subcloned into pCAGGS vector using ClaI and XmaI to obtain pCAGGS::AtHAP2-V5.
338 To subclone AtHAP2-V5 into pGENE B inducible system, pIZT::AtHAP2-V5 was
339 PCR amplified using 5'GAGGTACCATGGTGAACGCGATTTTAAT and
340 CCACTAGTTCACGTAGAATCGAGACCGA primers and cloned into pGENE B
341 between the KpnI and SpeI restriction sites to obtain pGENE B::AtHAP2-V5.
342 AtHAP2-YFP was subcloned into pCAGGS between the restriction sites ClaI and
343 XmaI. pGL290 was used as template and the insert was PCR amplified using
344 primers 5'CGTTATCGATATGGTGAACGCGATTTTAAT and
345 GCCCGGGTACTTGTACAGCTCGTCCA. pGENE B::AtHAP2::YFP was obtained
346 by PCR amplification of PGL290 using primers
347 5'GAGGTACCATGGTGAACGCGATTTTAAT and
348 CCACTAGTTCACGTAGAATCGAGACCGA. The fragment was cloned into pGENE
349 B following digestion with KpnI and SpeI.
350 EFF-1A::V5 was PCR amplified from pIZT:EFF-1A::V5 (Podbilewicz et al., 2006)
351 using primers OR54 (TTAATTGGTACCACTATGGAACCGCCGTTTGAGTGG) and
352 OR55 (AATTAAGCTAGCTCAACCGGTACGCGTAGAATCGAGACC) and cloned
353 into pGENE B using KpnI and SpeI.
354
355 For mifepristone-inducible expression in BHK cells we used the GeneSwitch System
356 (Wang et al., 1994) as recommended by Invitrogen (Avinoam et al., 2011). Cells
357 were transfected with pGENE B::EFF-1A-V5, pGENE B::AtHAP2-V5 or pGENE

358 B::AtHAP2::YFP. For transient expression of proteins we used pCAGGS::EFF-1-V5
359 (Avinoam et al., 2011), pCAGGS::AtHAP2-V5 or pCAGGS::AtHAP2-YFP.

360

361 To label nuclei and cytoplasm in mixing assays we used peGFP-N1 (Clontech)
362 (Addgene plasmid cat #6085-1) kindly provided by Maya Haus-Cohen and Yoram
363 Reiter. To label the cytoplasm we used pRFPnes (Hu et al., 2003) (pRFPcyto), kindly
364 provided by Claudio Giraudo and Jim Rothman. Oligonucleotides were synthesized
365 by IDT Sinteza, restriction enzymes were purchased from New England Biolabs
366 (NEB) or Fermentas, and ligations were performed using T4 ligase (Promega).

367

368 **Fusion assays.** For content mixing (fusion) assay, BHK cells at 70% confluence
369 were co-transfected (using Eugene HD; Promega) with 2µg of pGENE B::EFF-1-V5
370 (Pérez-Vargas et al., 2014), or pGENE B::AtHAP2-V5; 1 µg of pSWITCH to obtain
371 BHK-EFF-1 or BHK-HAP2 cells. As cotransfection markers we used RFP containing
372 a nuclear export signal (pRFPcyto (Avinoam et al., 2011)) or peGFP-N1 (Clontech).
373 Control cells were transfected with pRFPcyto or peGFP-N1. Four hours after
374 transfection the cells were washed twice with PBS⁻ in the plates (35mm Nunc
375 Surface) and detached using 0.05% EDTA solution (Biological Industries). BHK cells
376 were collected in Eppendorf tubes, and resuspended in DMEM with 10% FBS,
377 counted; equal amounts of cells were mixed and seeded in glass bottom plates (12
378 Well Black, Glass Bottom #1.5H; In Vitro Scientific). Four hours after mixing the EFF-
379 1 or AtHAP2 expression vectors were induced changing the medium with DMEM
380 containing 10% FBS and 10⁻⁷ M mifepristone. After 18 h the cells were fixed with 4%
381 paraformaldehyde in PBS and processed for immunofluorescence (see below). To
382 assay multinucleated cells and to detect the transfected proteins (AtHAP2-V5 or
383 EFF-1-V5), we stained cells with anti-V5 monoclonal antibody (Invitrogen) and nuclei
384 with 1 µg/ml DAPI (Podbilewicz et al., 2006). Micrographs were obtained using Wide
385 Field laser illumination using a Zeiss ELYRA system S.1 microscope at 20X (Plan-

Apochromat 20X/NA: 0.8) or 25X magnification (LCI Plan-Neofluar 25X/ NA: 0.8). We counted the number of nuclei in cells expressing RFPcyto or GFP. For cytoplasmic mixing we counted the number of GFP nuclei in RFPcyto positive cells. Transfection efficiency was evaluated as 13-16% based on RFPcyto, GFP and anti-V5 immunofluorescence. The fusion indexes (shown as percentage of multinucleation) were defined as the ratio between the number of nuclei in multinucleated cells (Nm) and the total number of nuclei in multinucleated cells and expressing cells that were in contact (Nc) but did not fuse using the following equation:

$$\% \text{ Multinucleation} = (Nm/Nc+Nm)100$$

The percentage of multinucleation was calculated independently for GFP or RFPcyto expressing cells. DAPI was used to confirm the number of nuclei in expressing cells. GFP (also green nuclei) and RFPcyto (nuclei not red; “black”).

The GFP+RFP mixing index was calculated using the following equation:

$$\% \text{ Mixing} = (Ngfp+rfp/nCrfp+gfp)100$$

Ngfp+rfp is the number of GFP nuclei in cells with cytoplasmic RFP. *nCrfp+gfp* is the number of contacts between RFPcyto and GFP expressing cells. Note that RFPcyto or GFP mononucleated cells that were not in contact in the micrographs were not counted. The multinucleation and mixing indexes are presented as means $\pm$ SEM of three independent experiments (N $\geq$ 400 for RFPcyto, GFP and RFPcyto+GFP calculations in each experiment). The images were obtained using wide field microscopy with laser illumination in a Zeiss ELYRA system S.1 microscope at 63X magnification (Plan-Apochromat NA: 1.4).

408

Hemifusion assay. BHK cells were grown in 35 mm tissue culture dishes to 75% confluence and transfected to express HAP2 as described above. 15 h later the cells were labeled with fluorescent lipid Vybrant Dil and CellTracker green (catalogue # V22885, and # C7025, ThermoFisher Sci.) at final concentrations of 4 μ M and 1 μ M,

413 respectively (donor cells). 1 h later (at 16 h post transfection) the cells were co-plated
 414 with either unlabeled transfected cells or with unlabeled non-transfected cells at 1-to-
 415 1 ratio (acceptor cells). The cells were fixed with phosphate-buffered 4%
 416 formaldehyde for 10 min at 22°C and, 8 h later, labeled with Hoechst 33342 and
 417 analyzed using fluorescence microscopy. To quantify hemifusion, the numbers of
 418 cells labeled only with the membrane probe in each field of view were normalized to
 419 the total number of cells labeled with both membrane and content probes. For
 420 transfected donor cells mixed with transfected acceptor cells, the total number of
 421 double labeled cells was 1451. For transfected donor cells that were mixed with non-
 422 transfected acceptor cells the total number of double labeled cells analyzed was
 423 1034. For non-transfected donor cells that were mixed with non-transfected acceptor
 424 cells the total number of double labeled cells analyzed was 1742.

425

426 **Immunofluorescence.** BHK cells were grown on tissue culture 35 mm plates with
 427 glass coverslips or in glass bottom plates (12 Well Black, Glass Bottom #1.5H; In
 428 Vitro Scientific). Non-permeabilized cells were placed on ice and fixed with 4%
 429 paraformaldehyde in PBS, washed twice with ice cold PBS followed by incubation in
 430 40 mM NH₄Cl to block free aldehydes and blocked in ice cold 1% FBS in PBS.
 431 Permeabilized cells were fixed with 4% paraformaldehyde in PBS, followed by
 432 incubation in 40 mM NH₄Cl to block free aldehydes, washed in PBS, permeabilized in
 433 0.1% tritonX-100 in PBS and blocked in 1% FBS in PBS. After fixation the
 434 plates/coverslips were incubated 1 h with mouse anti-V5 antibody (Invitrogen) 1:500
 435 or with rabbit anti At-HAP2 1:250 in PBS. The secondary antibodies were goat anti-
 436 mouse coupled to Alexa 643 or goat anti-rabbit coupled to Alexa 568 (Molecular
 437 Probes/Invitrogen) diluted 1:500 in PBS. Nuclei were visualized with 1 µg/ml DAPI
 438 (Avinoam et al., 2011). Permeabilized and non-permeabilized BHK cells were used
 439 as negative controls and gave background staining. The images were obtained using
 440 wide field microscopy with laser illumination in a Zeiss ELYRA system S.1

441 microscope at 63X magnification (Plan-Apochromat NA: 1.4).

442

443 **Live imaging of fusing cells.** Time-lapse microscopy to identify fusing cells was
 444 performed using a spinning disc confocal microscope (Yokogawa CSU-X) with a
 445 Nikon Eclipse Ti and a Plan Apo 20X (NA:0.75) objective. Six well plate with glass
 446 bottom (In vitro Scientific) were incubated in a Oko lab CO₂ and temperature
 447 controlled chamber at 37°C and 5% CO₂. Cells were co-transfected with pGENE
 448 B::AtHAP2-V5, pSWITCH and pRFPcyto to obtain BHK-HAP2 expressing RFPcyto.
 449 Six to 18 h post-induction images in DIC and red channels were obtained every 2-3
 450 min in different positions of the plate using high gain and minimum laser exposure.
 451 Images were captured with a iXon 3 EMCCD camera (Andor). Fusing cells were
 452 identified based on mixing of cytoplasms containing RFPcyto. Identified syncytia
 453 were imaged at higher resolution using Apo 60X (NA:1.4) objective. Confocal z-
 454 series were obtained to confirm the formation of multinucleated giant cells. Image
 455 analyses were done in Metamorph and Image J.

456

457 **Cell cycle inhibition.** To test whether multinucleation was as a result of failure of
 458 cytokinesis, we used 5-fluoro-2'-deoxyuridine (FdUrd) treatment to arrest BHK cells
 459 at G1/S phase (Dijkwel et al., 1986). BHK-21 cells were grown on tissue culture
 460 plates with glass coverslips (Marienfeld 1.5H). The cells at 70% confluence were co-
 461 transfected (using Fugene HD) with 2 µg/µl pCAGGS::AtHAP2 DNA and 1 µg/µl
 462 pEGFP-N1. For negative control BHK-21 were transfected with 1 µg/µl pEGFP-N1.
 463 Four hours after transfection the medium was changed with fresh DMEM with 10%
 464 FBS and 2 mM FdUrd (Sigma). 16-18 h after FdUdr addition the cells were fixed with
 465 4% PFA, permeabilized, stained with DAPI and incubated with mouse anti-V5,
 466 followed by anti-mouse ALEXA 568. We calculated multinucleation (see mixing
 467 assay) and compared the multinucleation index with control transfected cells without

468 FdUrd. The average results of two independent experiments are shown. N>800
 469 nuclei for each experiment.
 470
 471 **Surface biotinylation of proteins expressed on BHK cells.** BHK cells at 70%
 472 confluence were transfected with 2µg pCAGGS, pCAGGS::EFF-1-V5 or
 473 pCAGGS::AtHAP2-V5. 18-24 h post-transfection cells were washed twice with ice
 474 cold PBS++ and labeled with EZ-Link Sulfo NHS-Biotin (Pierce) for 30 min on ice.
 475 Surface biotinylation was followed by 4 washes with ice cold PBS++ and one wash
 476 with DMEM with 10% FBS to quench residual biotin followed by 2 washes with
 477 PBS++. After the washes, 300 µl of lysis buffer (50 mM Tris-HCl pH 8.0, 100 mM
 478 NaCl, 5 mM EDTA and 1% Triton X-100 supplemented with chymostatin, leupeptin,
 479 antipain and pepstatin and 10 mM Iodoacetamide) was added to each plate and the
 480 plates were incubated 15 minutes on ice. After removal of insoluble debris by
 481 centrifugation (10 min at 21,000 g), the lysate was mixed with NeutrAvidin Agarose
 482 Resin (Thermo) and SDS was supplemented to a final concentration of 0.3%. Affinity
 483 purification of biotinylated proteins was performed incubating for two to twelve hours
 484 at 4°C. The precipitated complex was then mixed with SDS-PAGE loading buffer with
 485 freshly added 50 mM DTT and incubated 5 minutes at 100°C. After pelleting by
 486 centrifugation the samples were separated on 4-12% SDS-PAGE gel (Bolt-
 487 Invitrogen) and analyzed by western blotting using anti-V5 mouse monoclonal
 488 antibodies (1:5000; Invitrogen) followed by 1:15000 HRP goat anti-mouse antibodies
 489 (Jackson). Loading was controlled using anti actin C4 monoclonal 1:2000 (MP,
 490 Biomedicals). Data shown are representative of at least three independent
 491 experiments (Podbilewicz et al., 2006).
 492
 493 **Pseudoviruses.** We prepared pseudoviruses based on the method originally used to
 494 analyze Ebola virus glycoprotein (Takada et al., 1997) and modified for AFF-1 and
 495 EFF-1 (Avinoam et al., 2011). BHK cells at 70% confluence were transfected with

496 2µg pCAGGS::EFF-1-V5 or pCAGGS::AtHAP2-V5. After 24 h at 37°C in 5% CO₂,
 497 cells were infected with VSVG-complemented VSVΔG recombinant virus (VSVΔG-G)
 498 at a multiplicity of infection (MOI) of 2-5 for 1 h at 37°C in 5% CO₂ in serum free
 499 medium (DMEM). Following six washes to remove unabsorbed VSVΔG-G we
 500 incubated the cells for 24 h at 37°C. The supernatants containing VSVΔG-HAP2 and
 501 VSVΔG-EFF-1 were collected and centrifuged at 600g for 10 min at 4°C to clear cell
 502 debris. Virions were collected and concentrated by two consecutive centrifugations at
 503 100,000 g through a 20%, and 10% sucrose cushion and resuspended in 130 mM
 504 NaCl, 25 mM Hepes pH 7.4.

505

506 **Determination of the titers of VSV pseudotype viruses.** To determine the titer of
 507 each pseudovirus preparation, 10⁴ BHK cells were plated in each well of a 96 well
 508 tissue culture plate. BHK cells were transfected with 1 µg/ml DNA of pCAGGS::EFF-
 509 1-V5, pCAGGS::HAP2-V5 or non transfected. We performed six serial dilutions of the
 510 viruses that were used to inoculate the cells. Following 18-24 h of incubation, GFP
 511 expressing cells were counted in at least two dilutions using a Zeiss ELYRA system
 512 S.1 microscope at 20X (Plan-Apochromat 20X/NA: 0.8). Each experiment was
 513 repeated at least three times. All the infections were done in the presence of anti-G
 514 monoclonal antibody to prevent the activity of any residual VSVΔG-G (Avinoam et
 515 al., 2011).

516

517 **Data analysis.** We did not predetermine sample size using statistical methods. The
 518 experiments were not randomized. We were not blinded to allocation during
 519 experiments and assessment of outcomes. Interobserver error was estimated for
 520 counting of multinucleated cells, cells in contact and content mixing experiments. To
 521 estimate inter-observer variation two investigators counted cells in contact,
 522 multinucleation and mixing of contents for the same 20 micrographs of different fields

(examples of the micrographs used can be seen in Fig. 3 B). The differences between percentages of multinucleation and content mixing obtained by the two observers was less than 10%. Figures were prepared with Adobe Photoshop CS5, Adobe Illustrator CS6 and Image J.

Statistical tests. The results are expressed as means $\pm$ SEM. For each experiment at least three independent biological repetitions were performed. We evaluated the significance of differences between mean values by using the unpaired t test function (Graph Pad software).

Online supplemental material

Fig. S1 shows confocal sections of fused HAP2-BHK cells and low surface localization of AtHAP2 in mammalian cells. Fig. S2 shows graphic summaries of HHblits and LOMETS structural prediction results for proteins of the HAP2/GSC1 family. Fig. S3 shows global structural comparisons and HMMs profile to profile analyses of fusexins. Videos 1 to 4 show time lapse imaging of BHK-HAP2 cells undergoing cell fusion. Table S1 shows that FdUrd treatment increases HAP2-mediated cell fusion. Table S2 is a dataset including HHblits and LOMETS structural prediction results. ITASSERectodomains.pdb file contains the structures shown in Fig. S3 C. profilePhylogeny.py contains the automation script for generation of profile based phylogenies using HHblits and HHsearch. StructuresToMSAandPhylo.py contains the automation script for the generation of structure based phylogenies using TAlign and Fatcat.

547 References

- 548 Aguilar, P.S., M.K. Baylies, A. Fleissner, L. Helming, N. Inoue, B. Podbilewicz, H.
549 Wang, and M. Wong. 2013. Genetic basis of cell-cell fusion mechanisms.
550 *Trends in genetics : TIG*. 29:427-437.
- 551 Avinoam, O., K. Fridman, C. Valansi, I. Abutbul, T. Zeev-Ben-Mordehai, U.E. Maurer,
552 A. Sapir, D. Danino, K. Gruenewald, J.M. White, and B. Podbilewicz. 2011.
553 Conserved Eukaryotic Fusogens Can Fuse Viral Envelopes to Cells. *Science*.
554 332:589-592.
- 555 Baker, N.A., D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon. 2001.
556 Electrostatics of nanosystems: application to microtubules and the ribosome.
557 *Proc Natl Acad Sci U S A*. 98:10037-10041.
- 558 Bianchi, E., B. Doe, D. Goulding, and G.J. Wright. 2014. Juno is the egg Izumo
559 receptor and is essential for mammalian fertilization. *Nature*. 508:483-487.
- 560 Chernomordik, L.V., and M.M. Kozlov. 2005. Membrane hemifusion: crossing a
561 chasm in two leaps. *Cell*. 123:375-382.
- 562 Cole, E.S., D. Cassidy-Hanley, J. Fricke Pinello, H. Zeng, M. Hsueh, D. Kolbin, C.
563 Ozzello, T. Giddings, Jr., M. Winey, and T.G. Clark. 2014. Function of the
564 male-gamete-specific fusion protein HAP2 in a seven-sexed ciliate. *Curr Biol*.
565 24:2168-2173.
- 566 Dijkwel, P.A., P.W. Wenink, and J. Poddighe. 1986. Permanent attachment of
567 replication origins to the nuclear matrix in BHK-cells. *Nucleic Acids Res*.
568 14:3241-3249.
- 569 Dolinsky, T.J., P. Czodrowski, H. Li, J.E. Nielsen, J.H. Jensen, G. Klebe, and N.A.
570 Baker. 2007. PDB2PQR: expanding and upgrading automated preparation of
571 biomolecular structures for molecular simulations. *Nucleic Acids Res*.
572 35:W522-525.
- 573 Forslund, K., A. Henricson, V. Hollich, and E.L. Sonnhammer. 2008. Domain tree-
574 based analysis of protein architecture evolution. *Mol Biol Evol*. 25:254-264.
- 575 Gattegno, T., A. Mittal, C. Valansi, K.C. Nguyen, D.H. Hall, L.V. Chernomordik, and
576 B. Podbilewicz. 2007. Genetic control of fusion pore expansion in the
577 epidermis of *Caenorhabditis elegans*. *Mol Biol Cell*. 18:1153-1166.
- 578 Harrison, S.C. 2008. Viral membrane fusion. *Nature structural & molecular biology*.
579 15:690-698.
- 580 Holm, L., S. Kaariainen, P. Rosenstrom, and A. Schenkel. 2008. Searching protein
581 structure databases with DaliLite v.3. *Bioinformatics*. 24:2780-2781.
- 582 Holm, L., and P. Rosenstrom. 2010. Dali server: conservation mapping in 3D.
583 *Nucleic Acids Res*. 38:W545-549.
- 584 Hu, C., M. Ahmed, T.J. Melia, T.H. Sollner, T. Mayer, and J.E. Rothman. 2003.
585 Fusion of cells by flipped SNAREs. *Science*. 300:1745-1749.
- 586 Igonet, S., and F.A. Rey. 2012. SnapShot: Viral and eukaryotic protein fusogens.
587 *Cell*. 151:1634-1634 e1631.
- 588 Johnson, M.A., K. von Besser, Q. Zhou, E. Smith, G. Aux, D. Patton, J.Z. Levin, and
589 D. Preuss. 2004. Arabidopsis hapless mutations define essential
590 gametophytic functions. *Genetics*. 168:971-982.
- 591 Kalev, I., M. Mechelke, K.O. Kopec, T. Holder, S. Carstens, and M. Habeck. 2012.
592 CSB: a Python framework for structural bioinformatics. *Bioinformatics*.
593 28:2996-2997.
- 594 Kielian, M., C. Chanel-Vos, and M. Liao. 2010. Alphavirus Entry and Membrane
595 Fusion. *Viruses*. 2:796-825.
- 596 Lefort, V., R. Desper, and O. Gascuel. 2015. FastME 2.0: A Comprehensive,
597 Accurate, and Fast Distance-Based Phylogeny Inference Program. *Mol Biol*
598 *Evol*. 32:2798-2800.
- 599 Liu, Y., J. Pei, N. Grishin, and W.J. Snell. 2015. The cytoplasmic domain of the
600 gamete membrane fusion protein HAP2 targets the protein to the fusion site

in Chlamydomonas and regulates the fusion reaction. *Development*. 142:962-971.

Liu, Y., R. Tewari, J. Ning, A.M. Blagborough, S. Garbom, J. Pei, N.V. Grishin, R.E. Steele, R.E. Sinden, W.J. Snell, and O. Billker. 2008. The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in Chlamydomonas and Plasmodium gametes. *Genes Dev*. 22:1051-1068.

Maruyama, D., M. Ohtsu, and T. Higashiyama. 2016. Cell fusion and nuclear fusion in plants. *Seminars in cell & developmental biology*.

Mori, T., H. Kuroiwa, T. Higashiyama, and T. Kuroiwa. 2006. GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nature cell biology*. 8:64-71.

Okamoto, M., L. Yamada, Y. Fujisaki, G. Bloomfield, K. Yoshida, H. Kuwayama, H. Sawada, T. Mori, and H. Urushihara. 2016. Two HAP2-GCS1 homologs responsible for gamete interactions in the cellular slime mold with multiple mating types: Implication for common mechanisms of sexual reproduction shared by plants and protozoa and for male-female differentiation. *Dev Biol*. 415:6-13.

Pérez-Vargas, J., T. Krey, C. Valansi, O. Avinoam, A. Haouz, M. Jamin, H. Raveh-Barak, B. Podbilewicz, and F.A. Rey. 2014. Structural basis of eukaryotic cell-cell fusion. *Cell*. 157:407-419.

Podbilewicz, B. 2014. Virus and Cell Fusion Mechanisms. *Annual Review of Cell and Developmental Biology*. 30:111-139.

Podbilewicz, B., E. Leikina, A. Sapir, C. Valansi, M. Suissa, G. Shemer, and L.V. Chernomordik. 2006. The C. elegans developmental fusogen EFF-1 mediates homotypic fusion in heterologous cells and in vivo. *Dev Cell*. 11:471-481.

Remmert, M., A. Biegert, A. Hauser, and J. Soding. 2012. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*. 9:173-175.

Sapir, A., J. Choi, E. Leikina, O. Avinoam, C. Valansi, L.V. Chernomordik, A.P. Newman, and B. Podbilewicz. 2007. AFF-1, a FOS-1-Regulated Fusogen, Mediates Fusion of the Anchor Cell in C. elegans. *Dev Cell*. 12:683-698.

Sievers, F., A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J.D. Thompson, and D.G. Higgins. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*. 7:539.

Soding, J. 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics*. 21:951-960.

Takada, A., C. Robison, H. Goto, A. Sanchez, K.G. Murti, M.A. Whitt, and Y. Kawaoka. 1997. A system for functional analysis of Ebola virus glycoprotein. *Proc Natl Acad Sci U S A*. 94:14764-14769.

Tsirigos, K.D., C. Peters, N. Shu, L. Kall, and A. Elofsson. 2015. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res*. 43:W401-407.

von Besser, K., A.C. Frank, M.A. Johnson, and D. Preuss. 2006. Arabidopsis HAP2 (GCS1) is a sperm-specific gene required for pollen tube guidance and fertilization. *Development*. 133:4761-4769.

Wang, Y., B.W. O'Malley, Jr., S.Y. Tsai, and B.W. O'Malley. 1994. A regulatory system for use in gene transfer. *Proc Natl Acad Sci U S A*. 91:8180-8184.

Waterhouse, A.M., J.B. Procter, D.M. Martin, M. Clamp, and G.J. Barton. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 25:1189-1191.

White, J.M., S.E. Delos, M. Brecher, and K. Schornberg. 2008. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol*. 43:189-219.

- 655 Wong, J.L., and M.A. Johnson. 2010. Is HAP2-GCS1 an ancestral gamete fusogen?
656 *Trends Cell Biol.* 20:134-141.
- 657 Wong, J.L., A.R. Leydon, and M.A. Johnson. 2010. HAP2(GCS1)-dependent gamete
658 fusion requires a positively charged carboxy-terminal domain. *PLoS Genet.*
659 6:e1000882.
- 660 Wu, S., J. Skolnick, and Y. Zhang. 2007. Ab initio modeling of small proteins by
661 iterative TASSER simulations. *BMC biology.* 5:17.
- 662 Wu, S., and Y. Zhang. 2007. LOMETS: a local meta-threading-server for protein
663 structure prediction. *Nucleic Acids Res.* 35:3375-3382.
- 664 Wu, S., and Y. Zhang. 2008. MUSTER: Improving protein sequence profile-profile
665 alignments by using multiple sources of structure information. *Proteins.*
666 72:547-556.
- 667 Xu, J., and Y. Zhang. 2010. How significant is a protein structure similarity with TM-
668 score = 0.5? *Bioinformatics.* 26:889-895.
- 669 Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang. 2015. The I-TASSER
670 Suite: protein structure and function prediction. *Nat Methods.* 12:7-8.
- 671 Ye, Y., and A. Godzik. 2004. FATCAT: a web server for flexible structure comparison
672 and structure similarity searching. *Nucleic Acids Res.* 32:W582-585.
- 673 Zhang, Y., and J. Skolnick. 2005. TM-align: a protein structure alignment algorithm
674 based on the TM-score. *Nucleic Acids Res.* 33:2302-2309.
- 675

676 **Acknowledgements**

677 We thank Julian Wong and Mark Johnson for providing an anti-AtHAP2 ectodomain
678 antibody and a clone containing the AtHAP2 coding sequence (pgl290). We thank
679 Daniela Megrian, Agustina Olivera-Couto and Fernan Agüero for help with
680 databases. Ofer Katz and Boaz Gildor for preliminary experiments and molecular
681 biology. Liliana Avila Ospina for AtHAP2 subcloning and preliminary multinucleation
682 assays. Eduard Baquero for advice on biochemistry. Rony Chanoch and Meytal
683 Landau for preliminary structural bioinformatics studies. We thank Alejandro Colman-
684 Lerner, Alberto Kornblihtt, Diego Alvarez, Ori Avinoam, Amir Sapir, Meital Oren-
685 Suissa, Javier Matias Hernandez, Mark Johnson, Peter Walter, Gustavo Caetano-
686 Anollés and Dan Cassel for helpful discussions and for critically reading the
687 manuscript. This work was supported by the International Centre for Genetic
688 Engineering and Biotechnology (grant CRP/URU11-01 to P.S.A.), the Fondo para la
689 Investigación Científica y Tecnológica, Argentina (grant PICT-2014-3034, to P.S.A.),
690 Israel Science Foundation, (grant 826/08, to B.P.), the European Research Council
691 (Advanced grant ELEGANSFUSION 268843 to B.P.), and the United States-Israel
692 Binational Science (grant 2013151 to L.V.C. and B.P.). The research in the L.V.C.

laboratory was also supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health. D.M. is a fellow from CONICET, Argentina. M.G. and H.R. acknowledge support from Agencia Nacional de Investigación e Innovación and Programa de Desarrollo de las Ciencias Básicas.

The authors declare no competing financial interests.

Author contributions. B.P. and P.S.A. designed the experiments. C.V. designed and performed fusion assays and fluorescence microscopy experiments. D.M. performed most of the bioinformatics data gathering and data processing. E.L. and L.V.C. designed, performed and interpreted the hemifusion experiments. E.M. performed surface biotinylation. B.P. performed live imaging of cell fusion and counted some fusion assays. D.M., M.G., H.R. and P.S.A. designed the structural and evolutionary bioinformatic strategies. C.V., D.M., H.R., M.G., B.P. and P.S.A. analyzed the data. P.S.A. and B.P. supervised the work and wrote the paper with input from all the other authors.

Author Information. Correspondence and requests for materials should be addressed to P.S.A. (paguilar@iib.unsam.edu.ar) and B.P. (podbilew@technion.ac.il).

Footnotes

Abbreviations used:

AtHAP2	<i>Arabidopsis thaliana</i> HAP2
BHK	Baby Hamster Kidney cells
EFF-1	Epithelial Fusion Failure 2
FUSEXINS	FUSion proteins essential for sexual reproduction and EXoplasmic merger of plasma membranes

721	GCS1	Generative Cell Specific 1
722	HAP2	Happless 2
723	HMMs	hidden Markov Models
724	VSV	Vesicular Stomatitis Virus
725	VSVΔG	VSV pseudoviruses in which the glycoprotein G gene was deleted

726

727 Figure legends

728

729 Figure 1.

730 **Arabidopsis HAP2 is sufficient to fuse mammalian BHK cells. (A)** BHK cell-cell
731 fusion assay: After discarding a possible failure in cell division (**Table S1**) cell-cell
732 fusion is measured by the appearance of **i**) multinucleated cells labelled with either
733 cytoplasmic RFP (RFPcyto; magenta) or nuclear and cytoplasmic GFP (green).
734 Fusion is also indicated by the appearance **ii**) of multinucleated cells containing
735 nuclear GFP and fluorescence from both RFPcyto and GFP in the cytoplasm. **iii**)
736 Nuclei are labelled with DAPI (blue) after fixation and permeabilization of the cells.
737 **(B) i) RFPcyto + GFP:** Negative control shows mononucleated cells expressing
738 RFPcyto (magenta) or nuclear and cytoplasmic GFP (green). **ii) HAP2(RFPcyto) +**
739 **HAP2(GFP):** BHK cells were transfected with AtHAP2 and GFP (green) or RFPcyto
740 (magenta); merged image of hybrid cell that contains mixed cytoplasm and three
741 nuclei. **iii) EFF-1(RFPcyto) + EFF-1(GFP):** Hybrid binucleate cell emerged after
742 EFF-1 expression and mixing of magenta and green cells (arrow). EFF-1(RFPcyto)
743 binucleate cells (arrowhead). **iv) HAP2(RFPcyto) + HAP2(GFP):** Heterokaryons
744 (hybrids) express magenta cytoplasm and green nuclei and cytoplasm (arrows).
745 Multinucleate green cells (arrowheads). Scale bars, i and ii 10 μm; (iii and iv) 20 μm.
746 **(C)** Quantification of multinucleation and content mixing experiments. Magenta and
747 green bars represent the fraction of multinucleated cells (two nuclei or higher) out of
748 all the cells in contact (magenta or green respectively). Black bars represent the
749 RFPcyto and GFP content mixing index. The fusion and mixing indexes are
750 presented as means ± SEM of three independent experiments. Total number of
751 nuclei counted in multinucleated cells and in cells in contact N ≥ 1000 for each
752 experimental condition. Used unpaired t-test comparing each color (RFPcyto, GFP,
753 mixed) for EFF-1 and HAP2 to the negative control (RFPcyto+GFP). * p<0.01; **
754 p<0.005; *** p<0.001; **** p<0.0005. **(D)** Still images from time-lapse experiments
755 reveal merging of two mononucleated (**i**) and three (**ii**) cells expressing RFPcyto and
756 HAP2 (arrows and arrowheads). Time indicated in h:min (see **Video 1** and **2** for **i** and
757 **ii** respectively). Note that the top nucleus (arrow in **i**) disappears due to defocus at
758 2:34. Two nuclei are out of focus at 4:57 (**ii**; see **Fig. S1 A**). Scale bars, 20 μm.

759

760 Figure 2.

761 **Model of the 3D structure of Arabidopsis HAP2 ectodomain indicates it is a**
762 **class II fusogen. (A)** Diagram of *A. thaliana* HAP2 protein coloured by domains
763 according to the ectodomain modelled structure: signal peptide (sp) and
764 transmembrane domain (TM), black; domain I, red; domain II, yellow, domain III,
765 blue, domain I-III linker, cyan; stem, magenta and intracellular domain, grey. **(B)**
766 Cartoon of AtHAP2 ectodomain modelled structure (residues 41-494), alongside with
767 *C. elegans* EFF-1 (PDB 4ojc), Dengue virus E glycoprotein (DV1 E, PDB 4gsx),
768 Semliki Forest Virus E1 glycoprotein (SFV E1, PDB 1rer) and Severe Fever with
769 Thrombocytopenia Syndrome Virus glycoprotein Gc (SFTSV Gc, PDB 5g47) class II

fusion proteins. Structures are coloured according to domains as in **A** and **E**; cd loops are shown in orange (facing up). **(C)** Structural similarity scores between experimental and modelled fusogens computed after flexible structural alignment. Blue matrix: TM-scores (values >0.5, are indicative of significant structural similarity), orange matrix: z-scores (values >2 are considered as indicative of significant structural similarity). **(D)** Unrooted tree inferred using a distance matrix. Boldface grey M indicates modelled structure, colors are HAP2, blue; EFF-1/AFF-1/FF, green; class II viral fusogens, red; see **Fig. S3 A**. **(E)** Structure-based alignment of AtHAP2 ectodomain with fusion proteins shown in **B**. Background colors indicate the domains organization (as in **A** and **B**), arrows in color and rounded rectangles denote beta sheets and alpha helices, respectively. The black box in AtHAP2 sequence marks the HAP2/GCS1 domain, the cd loops are marked in orange. Cysteines involved in conserved disulfide bonds stabilizing the cd loop are denoted in bold italics.

Figure 3.

HAP2 is a bilateral fusogen. **(A)** Quantification of content mixing experiments showing that HAP2 and EFF-1 are required in both interacting cells to form hybrids. Control RFPcyto + GFP is the same as in **Fig. 1 B**. Total number of nuclei counted and unpaired Student's t-test were performed as in **Fig. 1 C**. *** p<0.001; **** p<0.0005. Non-significant, p>0.05. **(B)** Representative fields used to determine percentages of multinucleation and content mixing. **RFPcyto + GFP**: Mixed control cells; Nuclear staining DAPI (blue). Dividing green cell with two nuclei (arrowhead). Scale bar, 20 µm. **EFF-1(RFPcyto) + GFP**: BHK-EFF-1 (magenta) do not mix with GFP-expressing cells, revealing that EFF-1-mediated fusion is bilateral (homotypic). Multinucleate cells expressing RFPcyto (arrowheads). **HAP2(RFPcyto) + GFP**: BHK-HAP2 multinucleation (arrowheads) and failure to mix with GFP-expressing cells revealing HAP2-mediated fusion is bilateral. **HAP2(RFPcyto) + EFF-1(GFP)**: Hybrids (arrows) between BHK-EFF-1 (green) and BHK-HAP2 (magenta) reveal heterotypic merger of cells. Multinucleated green or red cells (arrowheads). Some hybrids are mononucleate probably due to cell division or nuclear fusion following merger. Scale bar, 20 µm. **(C)** Examples of multinucleate cells containing 2-4 nuclei are marked with arrows. The cells are able to divide after fusion therefore the number of nuclei per cells is usually smaller than 6. **EFF-1(RFPcyto) + GFP** and **HAP2(RFP) + GFP** images show no mixing indicating that HAP2-mediated fusion is bilateral (homotypic) in BHK cells. Scale bars, 10 µm. **(D)** Immunoblot of vector (negative control), EFF-1 and HAP2 proteins carrying a V5 epitope fused to the cytoplasmic tail. Surface biotinylation of BHK cells expressing the different proteins. S, surface expression after affinity purification using neutravidin agarose beads; L, lysate. The amount of sample of HAP2 "S" is 600 times higher than EFF-1. The amount of lysate for HAP2 "L" is 12 times higher than for EFF-1. **(E)** Immunofluorescence of HAP2-YFP shows surface expression in cells revealed by anti-HAP2Extracellular (HAP2EC) polyclonal antibody. Without permeabilization 20% of the cells expressing HAP2-YFP showed punctate expression on the surface; merged image (magenta). Untransfected BHK cells showed no immunoreactivity. Following permeabilization with detergent there is colocalization between immunostaining with anti-HAP2EC and the YFP signal revealing the specificity of the antibody that recognizes HAP2 in the reticular ER and perinuclear localization; merged image (yellow). Scale bar, 20 µm. Immunofluorescence in permeabilized BHK-HAP2-V5 cells using anti-V5 antibody shows similar localization (see **Fig. S1 B and C**).

Figure 4

HAP2 can fuse the VSVΔG pseudovirus to target cells. **(A)** Cartoon illustrates the generation of VSVΔG-HAP2 pseudoviruses. Transfected BHK cells express HAP2 protein on the surface and were infected with G-complemented VSVΔG recombinant

825 virus (VSVΔG-G). The viral genome encodes GFP replacing VSVG. Infection results
826 in viral-induced expression of GFP by target cells (green cytoplasm). VSVΔG-HAP2
827 pseudoviruses were collected from the supernatant. **(B)** The activity of VSVΔG-HAP2
828 was tested on BHK (untransfected, naïve), BHK-HAP2 and BHK-EFF-1 cells.
829 Cartoon modified from (Avinoam et al., 2011). **(C)** Titers of VSVΔG pseudoviruses.
830 The type of protein on the viral membrane (EFF-1 or HAP2) and on the BHK target
831 cells (naïve, EFF-1 or HAP2) is indicated. Titers in infectious units (IU) represent the
832 number of cells expressing GFP per microliter 24 h after virus inoculation. Data are
833 mean ± SEM (n=3 experiments). We found no significant difference between
834 infection with VSVΔG-HAP2 and VSVΔG-EFF-1 of the different BHK target cells
835 (Two-tailed paired t test).

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

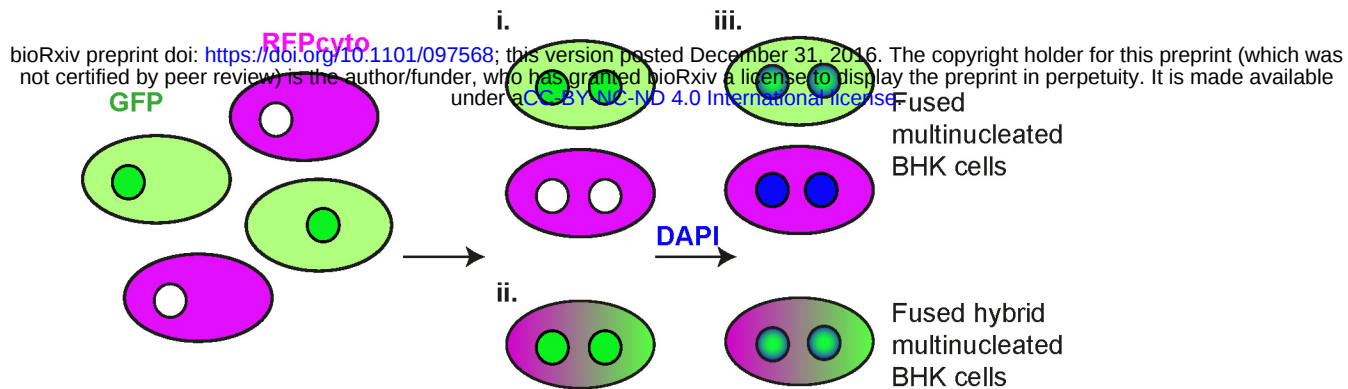
861

862

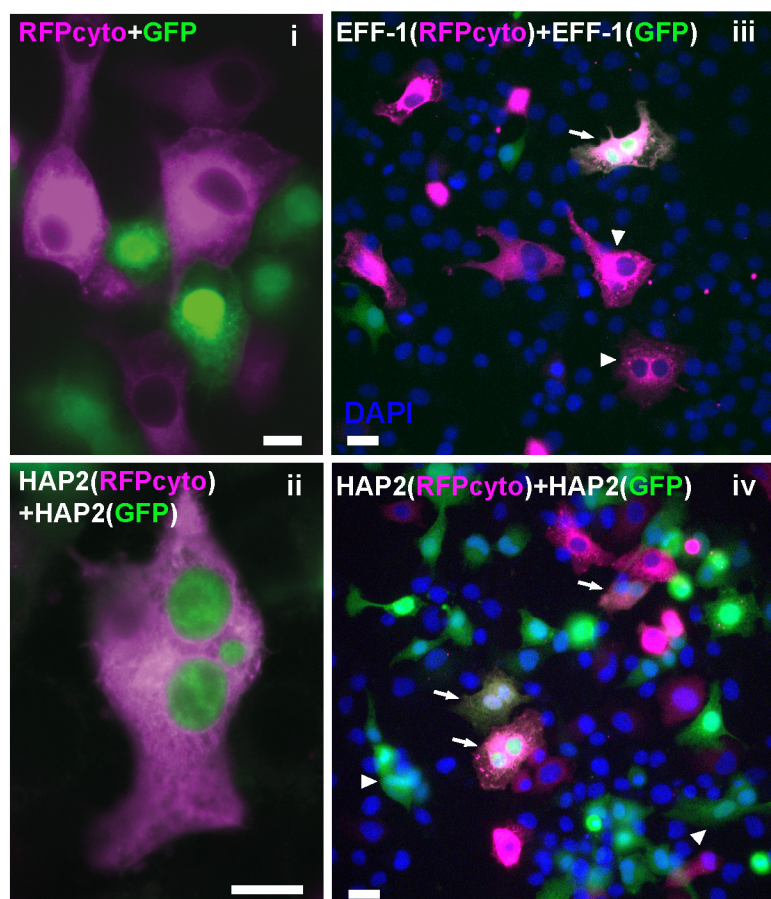
Figure 5.

Hemifusion in HAP2-mediated fusion is bilateral and Fusexins use divergent mechanisms. **(A)** Cartoon illustrates the hemifusion assay where one cell type, labeled with two fluorescent probes, acts as the “donor” cell and the unlabeled cell acts as “acceptor”. Due to internalization of fluorescent lipid from the plasma membrane, by the time we score fusion this probe mostly labels intracellular membranes. **(B)** Fluorescence microscopy images of AtHAP2 transfected cells labeled with both cell tracker (green, content probe) and Vybrant Dil (red, membrane probe) co-plated with unlabeled AtHAP2 transfected cells. Top panel: green cell tracker, Dil (red), Hoechst 33342 (blue). Bottom panel: green cell tracker and Hoechst 33342 (blue). Hemifusion event is detected as an appearance of a cell (marked by an arrow) that acquired only membrane probe apparently from an adjacent double-labeled cell (arrowhead). Scale bar, 20 μm. **(C)** Hemifusion extents quantified as the ratio between numbers of cells labeled only with membrane probe and numbers of cells labeled with both membrane and content probes. The results are means ± SEM (n=3). **(D)** Viral class II trimeric fusion proteins (viral Fusexins) have a unilateral fusion mechanism and the Fusexin is present only in the virus’ envelope or in one cell during cell-cell fusion. **(E)** Somatic cellular Fusexins (e.g. EFF-1 and AFF-1; on green cells) use a bilateral homotypic mechanism and the model proposes Fusexin has to be in both cells forming trans-trimeric complexes (Pérez-Vargas et al., 2014; Podbilewicz et al., 2006). Our results suggest a mechanistic model where sperm AtHAP2 fuses animal cells using a homotypic design (bilateral; red cells). HAP2 and EFF-1 can fuse cells in trans using a heterotypic mechanism. **(F)** We hypothesize that the egg of Arabidopsis (the central cell too) expresses an unidentified Fusexin that interacts with sperm HAP2.

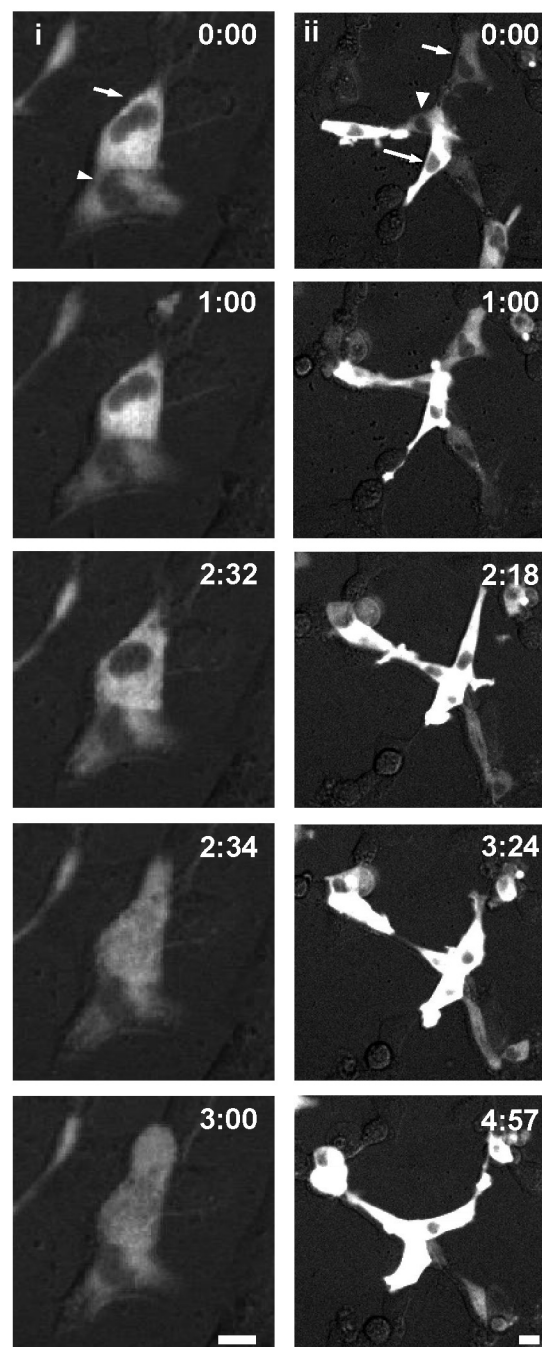
A



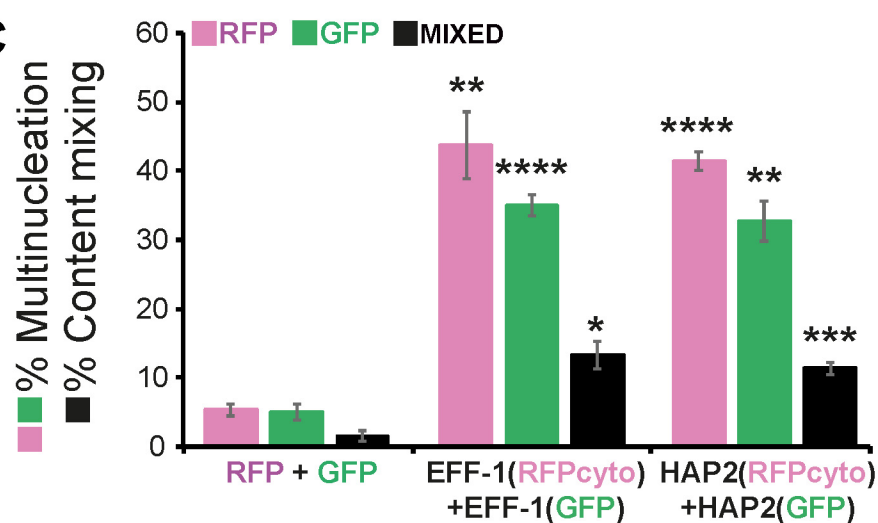
B

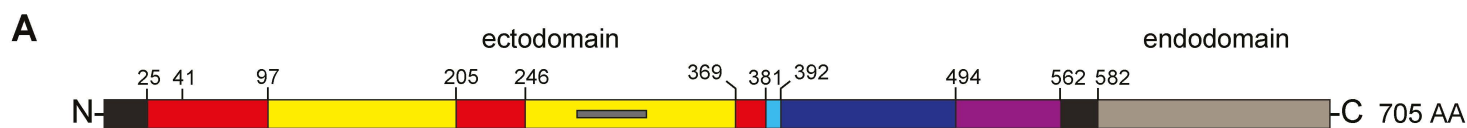


D



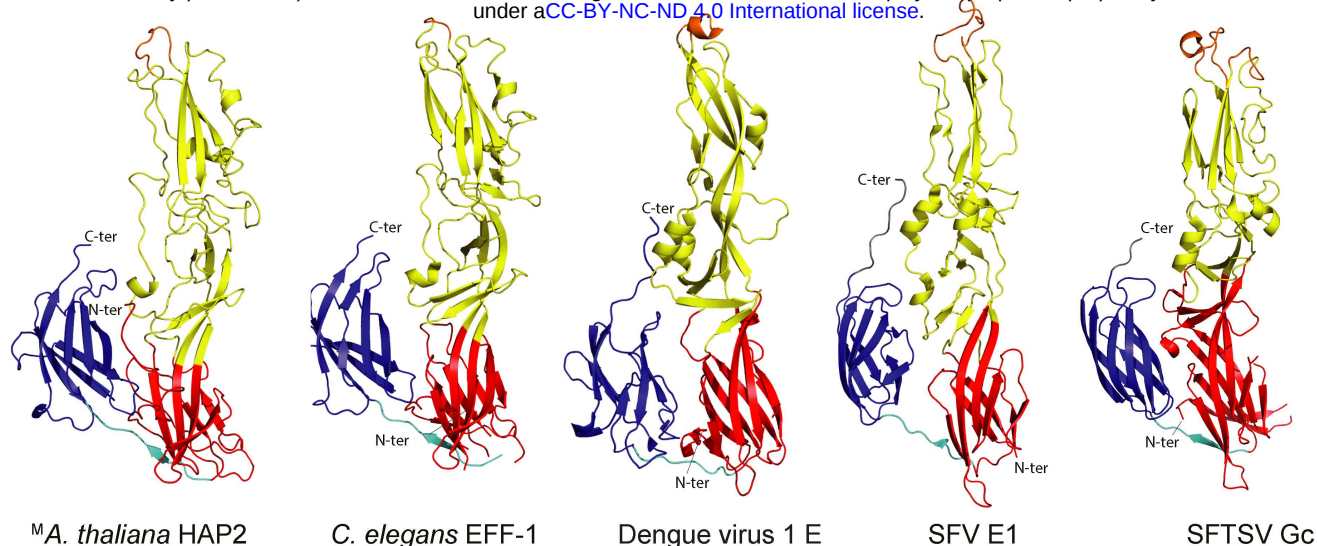
C





bioRxiv preprint doi: <https://doi.org/10.1101/097565>; this version posted December 31, 2016. 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-NC-ND 4.0 International license.

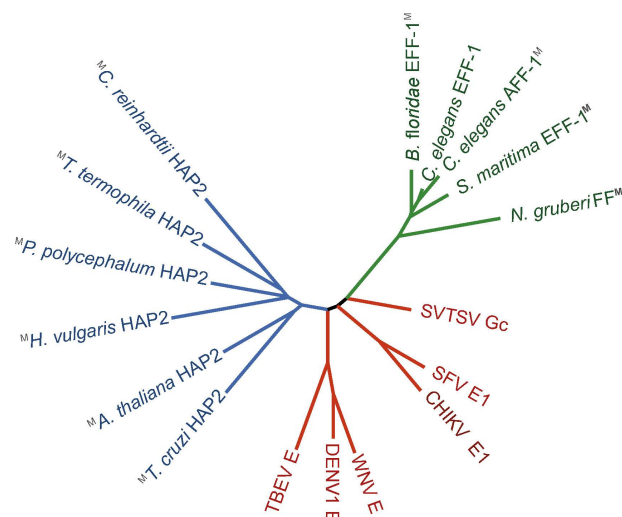
B



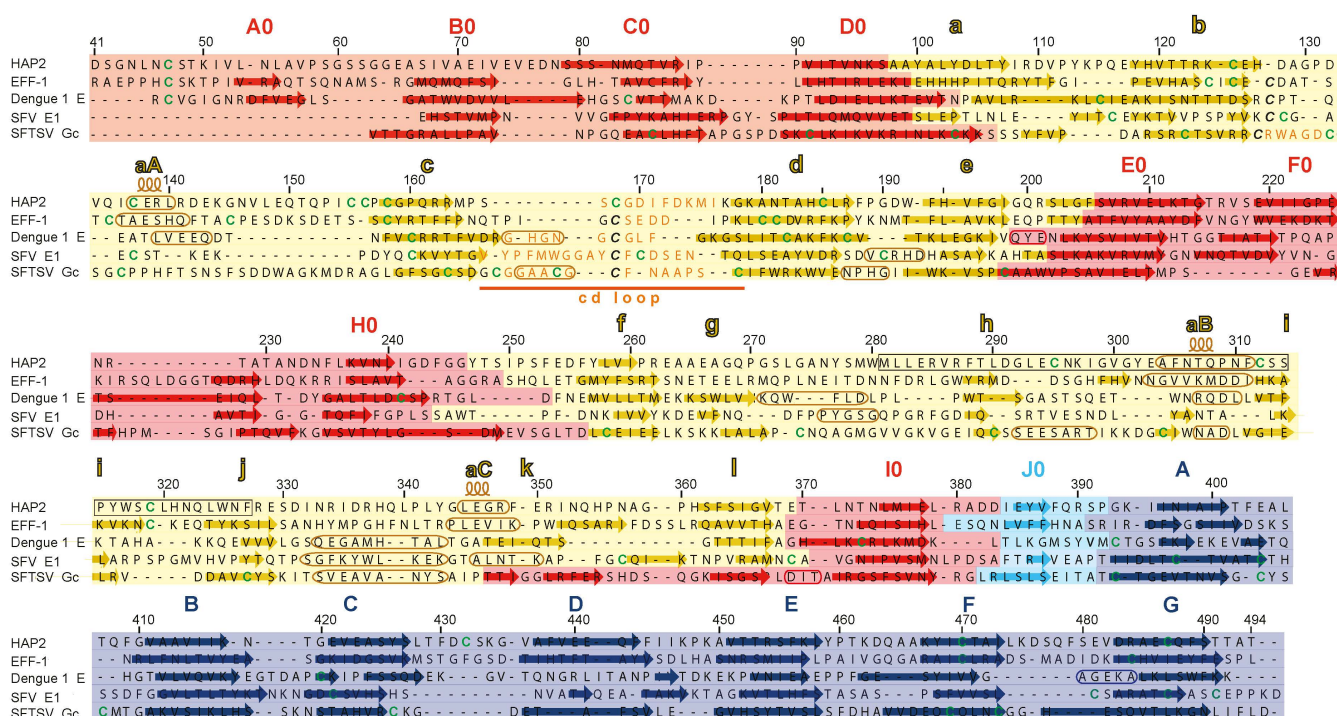
C

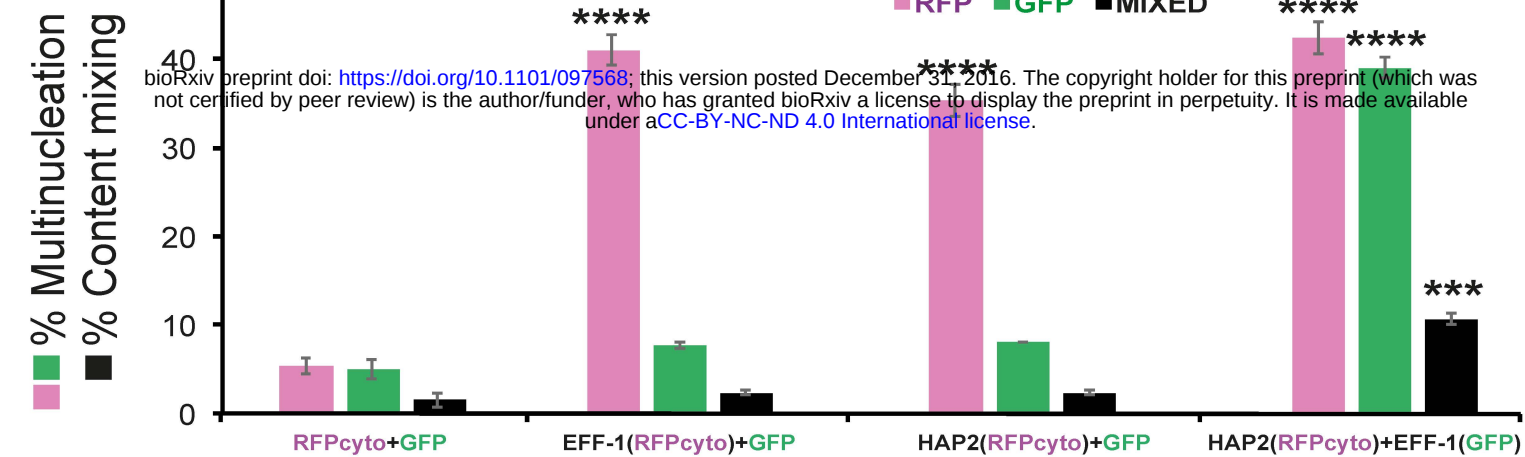
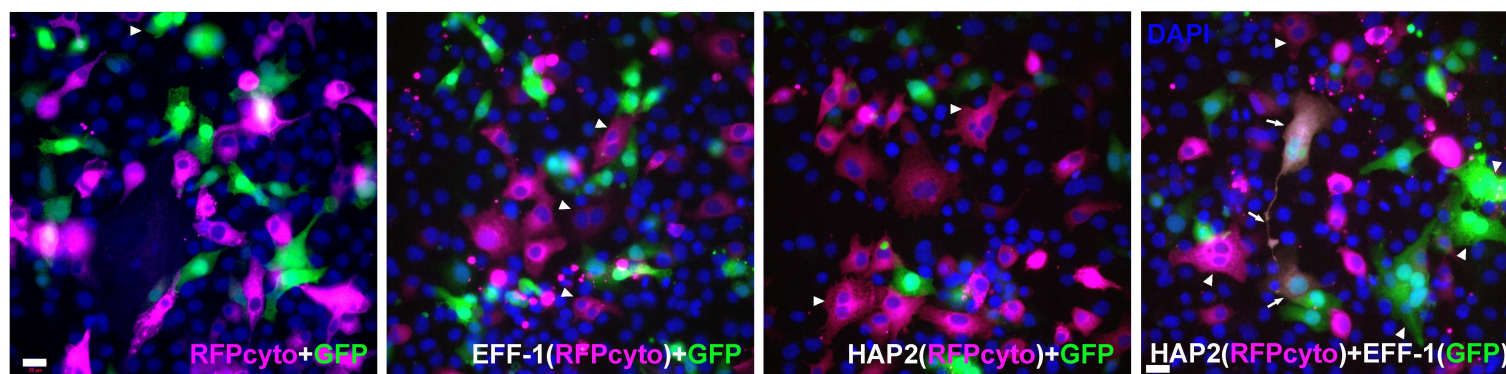
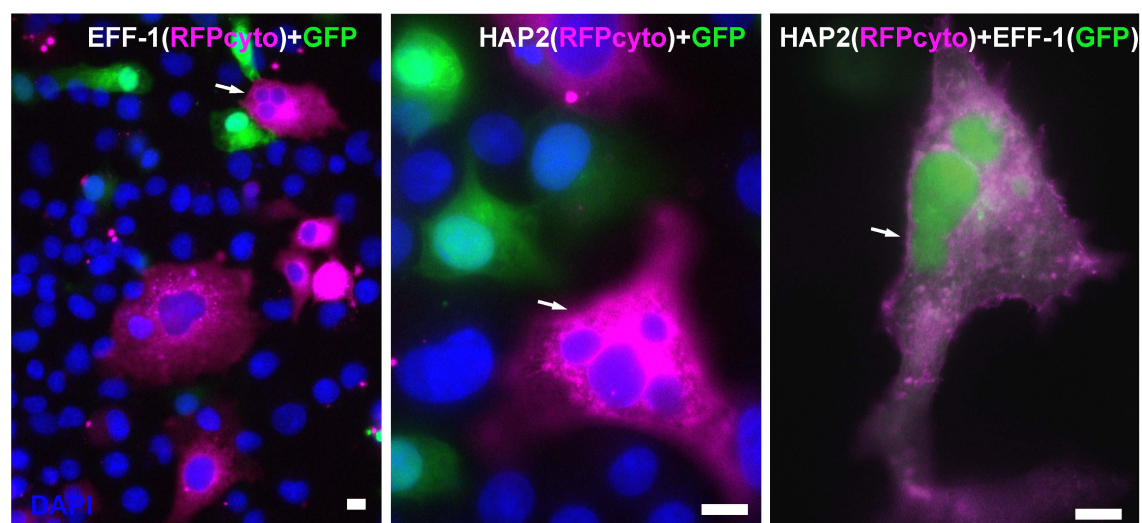
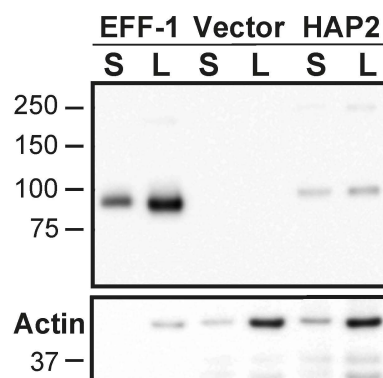
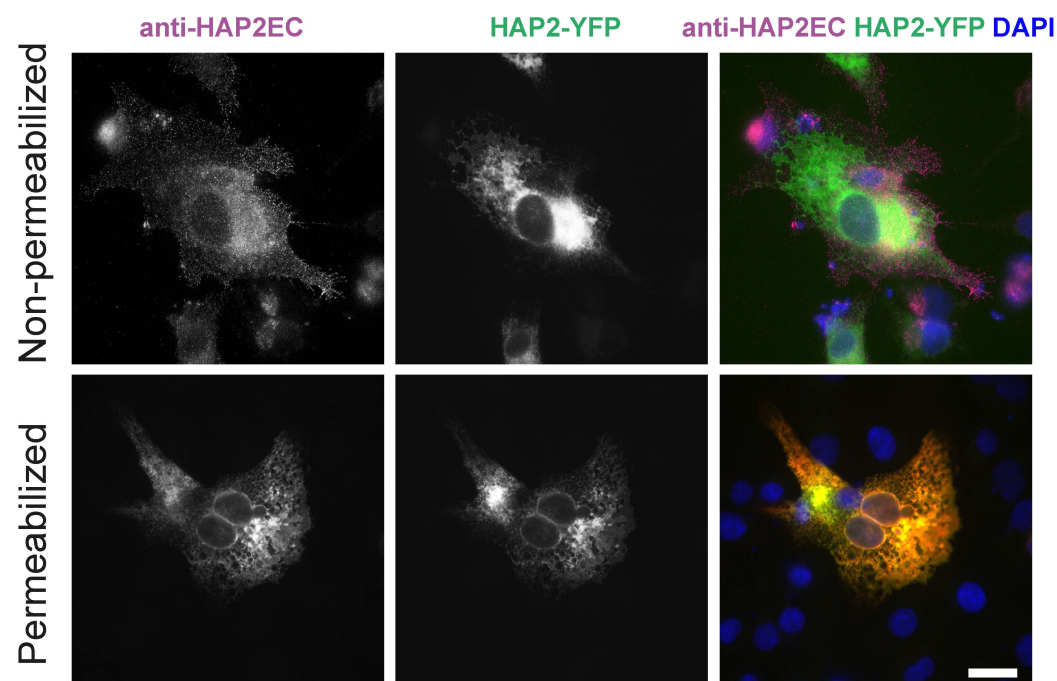
EFF-1			0.68	0.76	0.66	0.67
SFV E1	11			0.8	0.72	0.74
SFTSV Gc	16	14			0.74	0.77
Dengue virus 1 E	11	10	12			0.73
HAP2	8	10	12	8		

D

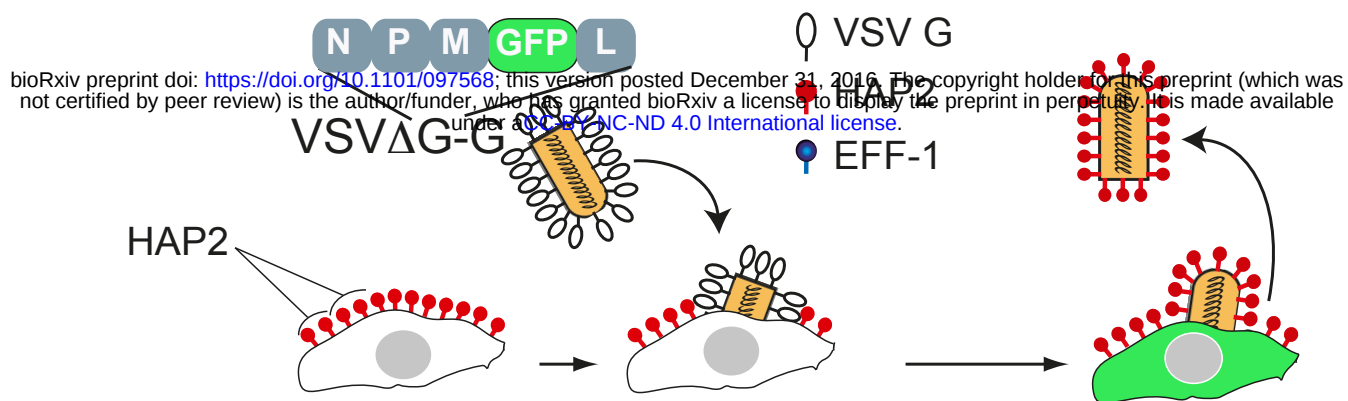


E

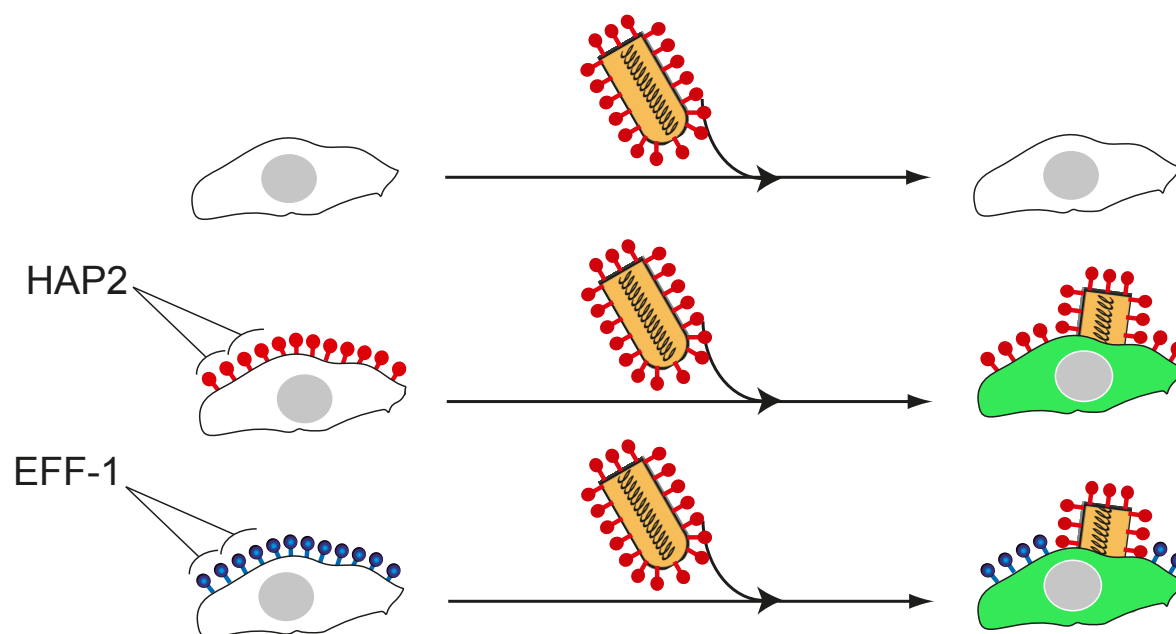


A**B****C****D****E**

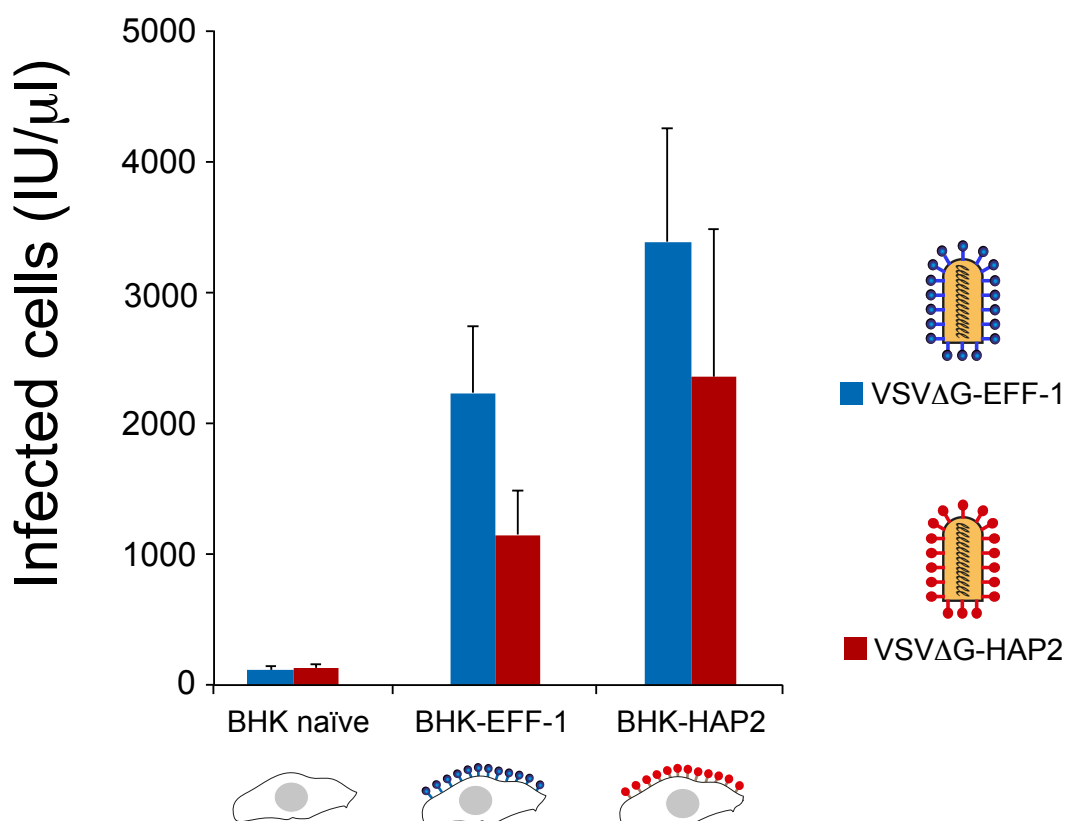
A



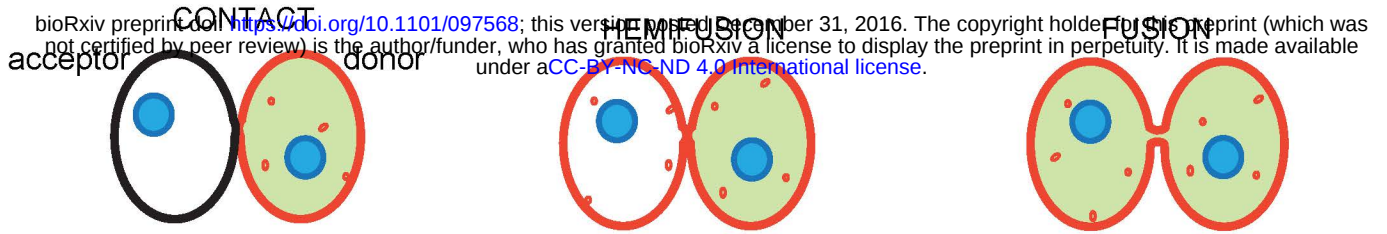
B



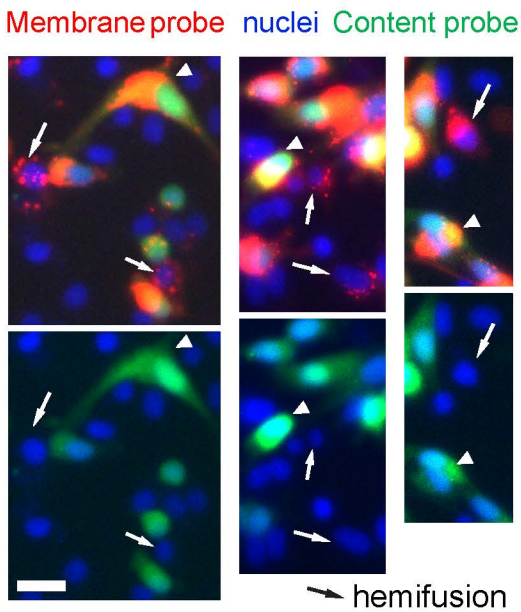
C



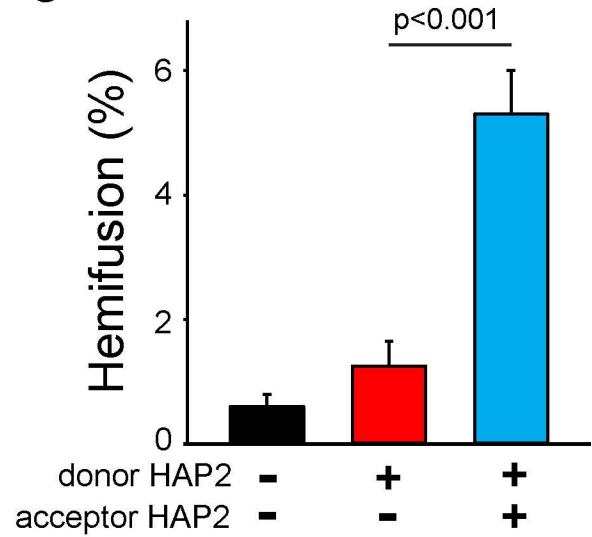
A



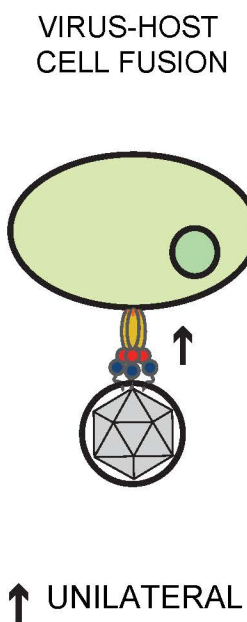
B



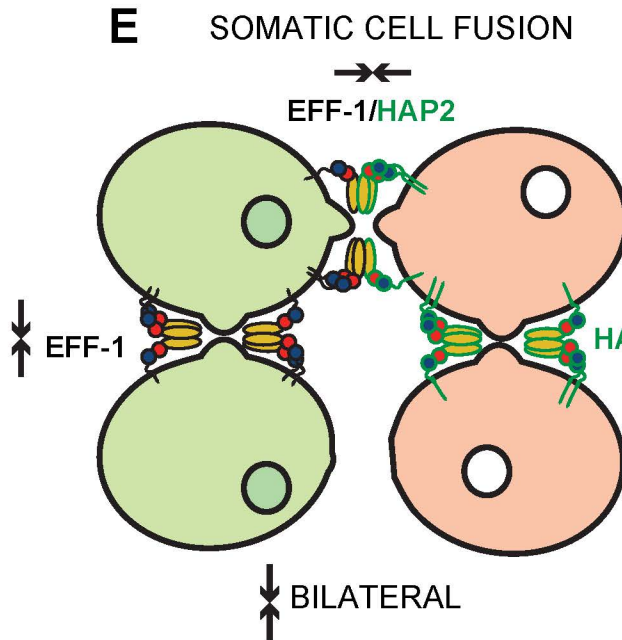
C



D



E



F

