An assay to measure the activity of RNA elements and proteins that promote the expression of mRNA with retained introns.

Patrick E. H. Jackson[1,2], Jing Huang[2,3], Monika Sharma[2,4], Sara K. Rasmussen[2,4], Marie-Louise Hammarskjold[2,3] and David Rekosh[2,3*]

[1]Division of Infectious Diseases and International Health, Department of Medicine, University of Virginia, Charlottesville, Virginia, USA

[2]Myles H. Thaler Center for HIV and Human Retrovirus Research, University of Virginia, Charlottesville, Virginia, USA

[3]Department of Microbiology, Immunology and Cancer Biology, University of Virginia, Charlottesville, Virginia, USA

[4]Department of Surgery, University of Virginia, Charlottesville, Virginia, USA

*corresponding author

The ability to overcome cellular restrictions that exist for the export and translation of mRNAs with retained introns is a requirement for the replication of retroviruses and also for the expression of many mRNA isoforms transcribed from cellular genes. In some cases, RNA structures have been identified in the mRNA that directly interact with cellular factors to promote the export and expression of isoforms with retained introns. In other cases, a viral protein is also required to act as an adapter. In this report we describe a novel vector system that allows measurement of the ability of *cis*- and *trans*-acting factors to promote the export and translation of mRNA with retained introns.

One reporter vector used in this system is derived from an HIV proviral clone engineered to express two different fluorescent proteins from spliced and unspliced transcripts. The ratio of fluorescent signals is a measurement of the efficiency of export and translation. A second vector utilizes a third fluorescent protein to measure the expression of viral export proteins that interact with some of the export elements. Both vectors can be packaged into viral particles and be used to transduce cells, allowing expression at physiological levels from the integrated vector.

33 **Introduction.**

34 Intron retention is an important form of alternative splicing that is prevalent during
35 retroviral replication, and is also found in the regulation of cellular genes. Special mechanisms
36 have been described that promote the export and translation of mRNAs with retained introns[1].
37 Simple retroviruses, such as the Mason-Pfizer monkey virus (MPMV), rely on an RNA element
38 present in the transcript that retains an intron, which binds to a cellular protein complex
39 consisting of Nxf1 and the co-factor Nxt1 to promote export and translation[2-4]. This element is
40 referred to as a constitutive transport element (CTE) due to the lack of requirement for viral
41 factors. CTEs have also been found in cellular genes, including within intron 10 of *NXF1* where
42 a CTE functions to export an RNA that is translated into a short form of Nxf1[5-7].

43 A more complex retrovirus, HIV-1, requires the nucleo-cytoplasmic export of both
44 unspliced and incompletely spliced viral RNA transcripts for the translation of essential viral
45 proteins and for the packaging of progeny viral genomes[8, 9]. For these mRNAs, export and
46 translation is dependent on the viral Rev protein[10, 11] and an RNA secondary structure called the
47 Rev Response Element (RRE)[12-14]. Rev binding and multimerization on the RRE permits the
48 assembly of cellular factors, including Crm1 and Ran-GTP, to form an export-competent
49 ribonucleoprotein complex[15, 16]. In contrast, the completely spliced HIV transcripts can be
50 exported and translated in the absence of Rev. Other complex retroviruses, such as equine
51 infectious anemia virus (EIAV) (Rev and RRE)[17], HTLV (Rex and RexRE)[18], mouse mammary
52 tumor virus (Rem and RmRE)[19], and the youngest family of human endogenous retroviruses,
53 HERV-K (Rec and RcRE)[20, 21], use an analogous mechanism to accomplish the export and
54 translation of intron-containing transcripts.

55 HIV is notable for the high degree of sequence diversity exhibited during natural
56 infection[22] and the Rev-RRE system shows significant variation in functional activity between
57 different viral isolates from different hosts[23], and between isolates from the same host at
58 different time points during infection[24, 25]. While the role of Rev-RRE functional activity

differences in HIV pathogenesis has not been fully elucidated, there is evidence that it is an important factor in clinical disease. For example, high RRE activity has been shown to correlate with an increased rate of decline in CD4 count[26, 27]. Conversely, low Rev activity has been associated with prolonged survival in the pre-ART era[28] and Rev activity has been correlated with the sensitivity of HIV infected T-cells to CTL killing[29]. In experimental infection of ponies with a related virus, EIAV, variation in Rev functional activity was observed during the course of infection, and functional activity differences correlated with clinical disease state[17, 30].

Variations in the functional activity of the HIV Rev-RRE system have previously been assessed with subgenomic reporter assays[24, 31, 32] or lentiviral vector packaging assays[23]. Rev-dependent fluorescent reporter systems have also been developed for use in detecting HIV infection, but these have not been used to quantify differences in Rev-RRE activity[33]. Existing functional assays are limited by the multiple steps needed for sample preparation, which often leads to variation between experiments and low throughput. More importantly, nearly all existing assay systems have measured Rev-RRE function using transient transfection of non-lymphoid cell lines.

In order to further correlate the variation that is seen in the Rev-RRE system from different HIV viral isolates with clinical disease states, and to identify additional CTEs present in cellular genes, we developed a new assay system that quickly allows the functional evaluation of large numbers of putative export elements and trans-acting factors. The assay utilizes fluorescent proteins as reporters. A novel aspect of this system is that it uses packageable retroviral constructs, so that after packaging and transduction of target cells, expression can be measured from chromosomally integrated proviral sequences. The data in this report demonstrates the effectiveness of this system in evaluating the expression of mRNA with retained introns mediated by the HIV Rev-RRE axis, by elements from other viruses, and by cellular CTEs.

85 **Results.**

86

87 *A fluorescence-based high-throughput assay of HIV Rev-RRE functional activity.*

88

89 The HIV provirus is a transcription unit that produces several different mRNA isoforms
90 from a single promoter. Some of these isoforms are unspliced or incompletely spliced, retaining
91 one or more introns, and thus require Rev and the RRE for nucleocytoplasmic export and
92 expression of the proteins they encode. Other isoforms are fully spliced and do not require Rev
93 and the RRE for expression. An assay system to readily measure the functional activity of the
94 HIV Rev-RRE axis was designed by altering the full-length HIV proviral construct NL4-3[34] to
95 express two fluorescent proteins, one in a Rev-dependent fashion and the other in a Rev-
96 independent fashion (Figure 1). Briefly, the Rev-dependent signal was created by truncating the
97 *gag* gene, deleting the frameshift signal and inserting an in-frame cassette consisting of a *cis*-
98 acting hydrolase element (CHYSEL) sequence[35] and the *eGFP* gene. This alteration would be
99 expected to modify the provirus to express a truncated Gag protein and eGFP from the
100 unspliced Gag-Pol mRNA. As this mRNA contains two introns, it would require Rev and the
101 RRE for export and translation. The Rev-independent signal was created by deleting a portion
102 of *nef* and replacing it with either an *mCherry* or *TagBFP* sequence. Since Nef is made from a
103 fully spliced mRNA that does not require Rev for export and expression, expression of these
104 proteins would not be expected to require Rev. Additional modifications were made to the HIV-
105 derived construct to prevent Rev production from the reporter vector, to further render the
106 construct replication-incompetent, and to permit the easy exchange of RRE sequences within
107 *env.*

108 Next, we verified that the construct would express TagBFP or mCherry constitutively and
109 eGFP only in the presence of Rev. Both HIV-derived constructs were transfected individually
110 into 293T/17 cells along with varying masses of a plasmid that expressed Rev from the CMV

111 immediate early promoter (pCMV-Rev). Flow cytometry was performed to determine the mean
112 fluorescent intensity of the Rev-dependent signal (eGFP) and the Rev-independent signal
113 (mCherry or TagBFP). Gating was conducted as described in Methods and Figure 2 (a-c).
114 Relative Rev-RRE functional activity was calculated as the ratio of Rev-dependent to Rev-
115 independent signal for each experimental condition.

116 Only minimal eGFP signal was observed in the absence of Rev for both HIV-derived
117 constructs (Figure 3). The Rev-RRE activity for both constructs in the 0 ng Rev condition was
118 more than 30-fold less than when the lowest tested amount of Rev plasmid (6.25 ng) was
119 included in the transfection and more than 600-fold less than when the maximum tested amount
120 of Rev plasmid (100 ng) was included. While a low level of eGFP "leakiness" is apparent in the
121 flow cytometry dot plots (Figure 3, a and b), this occurred only when there was very high
122 expression of the Rev-independent signal (mCherry or TagBFP) and the system behaved in a
123 highly Rev-responsive fashion. For both constructs, the Rev-independent signal decreased
124 slightly between the 0 ng and 100 ng Rev conditions (2.1 fold for the eGFP-mCherry construct
125 and 2.6 fold for the eGFP-TagBFP construct), but the eGFP signal increased dramatically (313-
126 fold for eGFP-mCherry and 268-fold for eGFP-TagBFP) (Figure S1). The ratio of Rev-
127 dependent to Rev-independent signal was calculated for each condition and both constructs
128 displayed a strongly linear increase in functional activity with increasing amounts of Rev plasmid
129 over the tested range (eGFP-mCherry construct r^2=0.992, eGFP-TagBFP construct r^2=0.999)
130 (Figure 3 c).

131 To demonstrate that the CHYSEL sequence functioned to create a free eGFP protein
132 and a separate Gag fragment rather than a Gag-eGFP fusion product, the HIV-derived construct
133 was transfected into 293T/17 cells along with a plasmid that expressed Rev (pCMV-Rev). A
134 Western blot was performed from cell lysate that demonstrated high efficiency functioning of the
135 CHYSEL sequence to produce free eGFP (Figure S2).

136

137 *Measurement of differences in Rev and RRE activity.*

138

139 We have previously described Rev and RRE sequences that display significantly
140 different functional activity in prior assays[23, 24]. To assess whether the fluorescence-based assay
141 recapitulates these activity differences, assay constructs were created using previously tested
142 Revs and RREs.

143 We first tested a known high-activity RRE (SC3-M57A) along with a known low-activity
144 RRE (SC3-M0A) sequence that were both derived from naturally occurring viruses in a single
145 patient. These were previously determined to have significantly different activities in a study
146 using a sub-genomic GagPol reporter assay[24]. New eGFP-TagBFP HIV-derived constructs were
147 created by replacing the native RRE sequence with a 234-nt RRE derived from each virus. The
148 NL4-3 RRE was used for comparison. All three RRE-containing constructs were transfected into
149 293T/17 cells along with varying amounts of a CMV-Rev plasmid. All three constructs displayed
150 tight Rev responsiveness with minimal calculated activity in the absence of Rev (Figure 4 a). As
151 expected, the SC3-M57A RRE displayed substantially greater functional activity than the SC3-
152 M0A with all tested amounts of Rev plasmid.

153 We also previously described two naturally occurring subtype G HIV-1 Rev proteins with
154 substantially different activity as assessed with a sub-genomic lentiviral vector packaging
155 assay[23]. CMV-Rev plasmids were created expressing high-activity 9-G Rev, low activity 8-G
156 Rev, or NL4-3 Rev. A single eGFP-TagBFP HIV-derived construct containing an NL4-3 RRE
157 was transfected into 293T/17 cells along with varying amounts of each of the CMV-Rev
158 constructs. The fluorescent assay replicated the previously observed difference in activity, with
159 the expected high-activity 9-G Rev yielding substantially greater Rev-RRE functional activity
160 than the 8-G Rev (Figure 4 b).

161

162 *Evaluation of non-Rev/RRE nuclear export systems.*

163

164 To determine whether the fluorescence-based assay system could detect expression
165 accomplished using mechanisms other than the HIV-1 Rev-RRE system, constructs were
166 created containing elements from the “simple retrovirus” Mason-Pfizer monkey virus (MPMV),
167 the human *NXF1* gene, and the human endogenous retrovirus HERV-K which are known to
168 promote the export and translation of mRNAs with retained introns.

169 The MPMV CTE directly mediates interaction between the viral mRNA with a retained
170 intron and cellular export factors without the requirement for an additional viral protein. An
171 eGFP-mCherry HIV-derived construct was created with the 162-nt MPMV CTE substituting for
172 the HIV RRE within *env*. A negative control construct was created for comparison by removing
173 the RRE and not substituting an alternative element. Both the MPMV CTE construct and the no
174 element construct were transfected into 293T/17 cells along with plasmids to overexpress the
175 cellular factors NXF1 and NXT1 that have been shown previously to enhance expression of
176 proteins from mRNA whose export has been mediated by the CTE function in 293T cells[36]. The
177 vector with the MPMV CTE showed clear functional activity that was 8-fold higher than
178 background (Figure 5 a).

179 A cellular CTE with a similar sequence and nearly identical structure to one of the 70-nt
180 degenerate repeats present in the MPMV CTE is found in the human NXF1 gene within its
181 intron 10 [5]. While the 162-nt MPMV CTE contains two internal RNA loops known to interact with
182 cellular export factors, the 96-nt NXF1 CTE contains only a single loop. We created constructs
183 that contained either one or two NXF1 CTEs to assess the functional implication of the number
184 of NXF1 binding loops for vector activity. These constructs were then transfected into 293T/17
185 cells along with plasmids that overexpress the cellular factors NXF1 and NXT1. The single CTE
186 construct displayed 3-fold increased activity relative to the no element background, while the
187 dual CTE construct displayed much greater activity, 33-fold over background (Figure 5 b).

188 Taken together, these results show that the vector system can be used to assess the function of
189 potential viral and cellular CTEs.

190 Like other complex retroviruses, the HERV-K (HML-2) provirus contains both an RNA
191 element RRE analogue called RcRE and a Rev viral protein analogue called Rec[37, 38]. To
192 assess whether the activity of this system could be scored in the fluorescent assay, HIV-derived
193 constructs were created in which the RRE was replaced with either one or two 433-nt RcRE
194 sequences in tandem. The RcRE containing HIV-derived constructs, and an RRE-containing
195 construct for reference, were transfected into 293T/17 cells along with varying amounts of a
196 CMV-Rec plasmid. As expected, the RRE-containing construct did not display significant
197 functional activity, as Rec does not function on the RRE[37] (Figure 5 c and d). Both the single
198 RcRE and dual RcRE constructs showed linear increases in functional activity with increasing
199 amounts of Rec plasmid (r^2=0.928 and 0.992, respectively) (Figure 5 c). The dual RcRE
200 construct showed substantially greater activity than the single RcRE construct in the presence
201 of functional Rec. Thus, it should be possible to use this vector to detect functional endogenous
202 Rec expression in cells that have active Herv-K expression.

203

204 *Specific inhibition of the Rev-RRE pathway by small molecules.*

205

206 Due to its central role in the HIV life cycle, the Rev-RRE regulatory axis is an attractive
207 therapeutic target. Small molecules that inhibit the Rev-RRE pathway have been developed,
208 though none have yet entered clinical use for the treatment of HIV[39]. A heterocyclic compound
209 designated 103833 and its derivatives have previously been shown to inhibit Rev-RRE
210 functional activity in luciferase and GagPol subgenomic reporter assays[40, 41]. We tested the
211 effect of this compound on both Rev-RRE activity and MPMV CTE activity in the fluorescent
212 assay. 293T/17 cells were pre-treated with varying concentrations of 103833 and the incubation
213 with the compound was continued after transfection with two different sets of plasmids. The

214 effect on the Rev-RRE pathway was tested by transfecting cells with the two color fluorescent
215 construct containing the NL4-3 RRE along with a separate CMV-Rev construct. The effect on
216 the CTE pathway was tested by transfecting cells with the two color fluorescent construct
217 containing the MPMV CTE along with plasmids to overexpress NXF1 and NXT1.

218 For both constructs, the signal intensity of the fluorescent protein expressed from the *nef*
219 position, mCherry, was not affected by the presence of 103833 (Figure 6 a, boxes). The signal
220 intensity of the protein expressed from the *gag* position, eGFP, showed an inverse relationship
221 with 103833 concentrations only in the case of the RRE-containing construct but not for the
222 CTE-containing construct (Figure 6 a, circles). Thus, the functional activity of the Rev-RRE
223 construct decreased with increasing 103833 concentrations, but the functional activity of the
224 CTE construct remained constant. This is consistent with specific inhibition of the Rev-RRE
225 pathway as previously described. Additionally, the fact that the mCherry signal was unaltered
226 shows that it is a good internal indicator for non-specific effects and toxicity. These results
227 demonstrate the utility of the assay for drug screening.

228

229 *Rev-RRE pathway inhibition by* trans-*dominant negative Rev M10.*

230

231 An alternative approach to Rev-RRE pathway inhibition is through use of a modified
232 *trans*-dominant negative Rev protein. Rev M10 is a Rev sequence with a mutation in the nuclear
233 export signal and is a well-characterized *trans*-dominant inhibitor of Rev function[42]. We decided
234 to test the effect of Rev M10 on the fluorescence-based functional assay. To do this, 293T/17
235 cells were transfected with the eGFP-TagBFP HIV-derived construct containing an NL4-3 RRE
236 together with a constant amount of the MSCV construct expressing a bicistronic transcript
237 consisting of the NL4-3 Rev and mCherry genes linked with an internal ribosomal entry site
238 (IRES). The MSCV construct permitted the measurement of mCherry fluorescence as a
239 surrogate for the level of NL4-3 Rev expression. Additionally, the cells were transfected with

240 variable amounts of a plasmid that expressed Rev M10 from a CMV immediate early promoter
241 (pCMV-RevM10). To compensate for the varying amount of Rev M10 plasmid added, an
242 appropriate amount of a pCMV plasmid without an insert was also added.
243 Figure 6 b demonstrates that as the amount of Rev M10 plasmid added to the
244 transfection increased, the Rev-dependent eGFP signal decreased. Notably, there was no
245 change in the mCherry signal, indicating that the amount of functional NL4-3 Rev remained
246 constant. In addition, the TagBFP signal also remained constant, indicating that Rev-
247 independent expression from the HIV construct was not inhibited. Thus, *trans*-dominant
248 inhibition by Rev M10 could be clearly demonstrated in this assay. This data, combined with the
249 drug inhibition data described above, demonstrates that this assay can easily measure
250 perturbations of Rev-RRE functional activity that are caused by different mechanisms.
251
252 *Measurement of Rev-RRE functional activity using packaged assay constructs.*
253
254 Both the HIV-derived RRE-containing constructs and the MSCV-derived Rev-containing
255 constructs were designed to be packageable as viral vectors. The ability to package these
256 constructs allows transduction, rather than transfection of target cells. This permits the assay of
257 Rev and RRE functional activity in different cell types, including primary cells, from an integrated
258 provirus. To demonstrate the ability of the fluorescent assay to assess Rev-RRE functional
259 activity after transduction, a construct carrying the NL4-3 RRE was tested in combination with
260 the three different Revs that were assayed above and known to have varying functional
261 activities.
262 Three MSCV constructs were created, each carrying a different Rev sequence and
263 expressing a bicistronic transcript encoding both Rev and an mCherry gene separated by an
264 IRES. The Rev sequences included the low-activity (8-G),the high-activity (9-G), and the NL4-3
265 Revs utilized above[23]. An HIV-derived RRE-containing construct carrying eGFP and TagBFP as

266 markers was used as the fluorescent reporter. The HIV-derived and MSCV-derived constructs
267 were packaged in 293T/17 cells using transient transfection systems with appropriate packaging
268 constructs and pseudotyped using VSV-G. The resulting viral stocks were then used to
269 transduce the CEM-SS lymphoid cell line, either with the HIV-derived construct alone or with the
270 HIV-derived and an MSCV-derived construct in combination.

271 Cells were successfully transduced with both constructs. Gating was performed on cells
272 expressing the fluorescent marker from the MSCV-derived construct, then an eGFP vs TagBFP
273 plot was constructed to show cells expressing the markers carried on the HIV-derived construct
274 (Figure 7 a). Rev-RRE functional activity (the ratio of eGFP to TagBFP fluorescence) was
275 calculated from the co-transduced population that expressed fluorescent markers from both
276 vectors (i.e. cells included in quadrants 1, 2, and 3 in Figure 7 a). The results showed that the
277 functional activity of NL4-3 Rev and 9-G Rev are both greater than the 8-G Rev (Figure 7d).
278 Thus, the functional activity of the Revs in the T cell line on integrated provirus was similar to
279 that found in the transient transfection assay in 293T/17 cells. Figure 7a-c also shows that a
280 minority of cells that received both the reporter plasmid and the Rev-expressing vector
281 (quadrant Q3) did not express eGFP. The existence of this population would be obscured in
282 assays dependent on bulk protein expression, such as p24 reporter assays, and is only
283 apparent due to the cell-level resolution of this system. Additional studies are underway to
284 further determine if Rev function is compromised in the population of cells that does not express
285 eGFP.

286

287 *Creation of a stable T-cell line containing a single integrated copy of the HIV-derived reporter*
288 *vector.*

289

290 To establish clonal stable cell lines, CEM-SS cells were transduced with the HIV-derived
291 vector containing mCherry in the *nef* position and eGFP in *gag*. Individual clones expressing

292 mCherry were then derived from the transduced cells by limiting dilution. Clones were analyzed
293 for mCherry and eGFP expression. Clone 5 showed a very uniform and strong mCherry
294 expression with virtually no eGFP or tag-BFP signal (Figure 8, left side plots). Integration of a
295 single copy of the HIV-derived construct within an intergenic region of chromosome 9
296 (downstream of gene RGP1 and upstream of long non-coding RNA AL1333410; chromosome 9:
297 35765800) was determined by inverse PCR (data not shown). The cells were then transduced
298 with the MSCV vector expressing NL4-3 Rev and TagBFP (Figure 8, right side plots). About 8%
299 of cells expressed TagBFP (Figure 8 a), demonstrating successful transduction with the MSCV
300 construct, and a similar percent expressed eGFP from the unspliced *gag* transcript (Figure 8 b).
301 It is likely that the different levels of eGFP and tag-BFP expression stem from the fact that the
302 integration site of the MSCV vector in each transduced cell is different, and this leads to
303 different levels of expression from the integrated proviral vector. Significantly there was a good
304 linear correlation of BFP versus eGFP expression levels, indicating that cells that expressed
305 higher Rev levels also expressed higher levels of the mRNA with the retained intron (Figure 8
306 c). These cells were subjected to cell sorting, allowing us to obtain cells that have different
307 levels of Rev expression that can be used for further studies.
308
309 **Discussion.**
310
311 In this paper, we describe a novel easily manipulated system that allows the detection
312 and characterization of RNA elements and trans-acting protein factors that promote the
313 nucleocytoplasmic export and translation of mRNAs with retained introns. Intron retention is
314 essential for retrovirus replication. Furthermore, intron retention in cellular genes is increasingly
315 recognized as playing an important role in development and differentiation, as well as in disease
316 states such as cancer and responses to cellular insults[1]. In the case of complex retroviruses,
317 such as HIV and EIAV, there is also considerable evidence that different viral isolates vary

318 significantly in their ability to express intron-containing viral RNA and that these differences are
319 important in pathogenesis and possibly also in the regulation of viral latency[17, 23, 24, 27, 30]. The
320 ability to detect and functionally characterize the RNA elements and proteins involved in the
321 export and translation of mRNAs with retained introns is essential for further insights into the
322 fate and regulation of these mRNAs.

323 Our assay system improves on previous reporter assays in several important respects.
324 First, the use of three fluorescent proteins as reporters allows the simultaneous measurement of
325 expression from the mRNA with the retained intron, the overall transcription level of the reporter
326 vector, and the level of expression of the trans-acting protein involved in RNA export.
327 Furthermore, the use of flow cytometry to quantify expression levels provides for more rapid
328 acquisition of data compared with prior assays. Many previous reporter assays have required
329 collection of cells and subsequent quantification of non-fluorescent proteins such as p24 by
330 ELISA[24] or enzymatic activity[32]. In other cases, RNA export function has been measured by the
331 determination of vector titer[23]. The assay described here will permit more rapid determination of
332 the activity of larger numbers of Rev and RRE sequences derived from HIV patients. This will
333 allow the evaluation of the larger data sets that are needed to make good statistical correlations
334 of Rev-RRE activity with specific clinical outcomes. It will also more readily allow the
335 identification of novel RNA elements in cellular genes, and facilitate drug discovery targeting the
336 interaction of RNA signals with the protein factors that recognize them.

337 Second, nearly all previous reporter assays have measured the expression of mRNA
338 with retained introns from DNA plasmids that have been transiently expressed, despite the fact
339 that there is considerable evidence that mRNA export factors first encounter the mRNA when
340 they are in the process of being newly transcribed from chromatin[43, 44]. In the assay described
341 here, the vectors can be packaged into infectious retrovirus-vector particles which integrate into
342 the host cell genome. This allows measurement of expression of mRNAs with retained introns
343 that have been transcribed from chromatin, rather than transiently from transfected plasmids,

344 and is thus less artificial than previous assays. Overexpression of both the reporter mRNA and
345 export factor can also be avoided by manipulating the multiplicity of infection such that single
346 copies of reporter construct and the gene for the trans-acting protein factor are present in a cell.
347 This is important, as we know, for example, that in the case of HIV high levels of Rev protein
348 expression produce artifactual results and only low levels of Rev are expressed in an infected
349 cell[45]. The ability to package and pseudotype the vector constructs also permits the use of the
350 assay in diverse cell types and primary cells, rather than in only easily transfectable cell lines.
351 Thus, our assay system much more closely resembles the conditions of natural HIV infection
352 and the expression of normal levels of cellular mRNA.

353 Third, the ability to stably integrate assay constructs into the host cell chromosome also
354 permits the creation of cell lines containing one or both assay constructs. Using the CEM-SS T-
355 cell line, we created a clonal population containing a single integrated copy of an HIV-derived
356 assay construct within an intergenic region. This cell line constitutively expresses the Rev-
357 independent fluorescent marker (mCherry), but expresses the Rev-dependent marker (eGFP)
358 only after transfection or transduction with a Rev-containing construct. When this cell line is
359 transduced with an MSCV-derived construct carrying HIV Rev and TagBFP, eGFP expression
360 increases linearly with TagBFP expression. This demonstrates a close correlation between Rev
361 levels and eGFP expression, and the linear responsiveness of the system. Thus, the creation of
362 this cell line and others will permit more rapid drug screening studies and studies of the cellular
363 determinants involved in the export and translation of mRNA with retained introns during HIV
364 infection. Similar cell lines can also be created using these constructs containing cellular CTEs
365 to evaluate the role of cellular proteins in the export and translation of cellular mRNAs with
366 retained introns.

367 In summary, we have described a novel high-throughput fluorescence-based assay of
368 intron retention that permits the functional characterization of mechanisms that promote
369 expression from mRNAs with retained introns, including systems from complex and simple

370 retroviruses and cellular genes. This assay represents a substantial improvement from previous
371 assays in that studies can be performed at the level of integrated proviral constructs in a variety
372 of cell lines and primary cell types. Applications for this assay include drug screening for
373 compounds that inhibit or promote the export of mRNAs with retained introns, identification of
374 cellular CTEs, and investigation of the evolution of the Rev-RRE axis in clinical disease
375 progression during HIV infection.
376
377 **Methods.**
378
379 *Creation of assay constructs.*
380
381 HIV-derived constructs were created by modifying a plasmid containing the full-length
382 genome of laboratory strain NL4-3 (Genbank accession U26942) (Figure 1)[34, 46]. To create the
383 fluorescent signal from the unspliced transcript, *gag* was truncated at amino acid 407 and a
384 cassette was inserted in-frame containing a CHYSEL sequence derived from porcine
385 teschovirus-1[35] and the *eGFP* gene[47]. To create the fluorescent signal from the completely
386 spliced transcript, *nef* was deleted and replaced with either an *mCherry*[48] or *TagBFP* gene[49].
387 The construct was further modified to silence the native *rev* and to further render the construct
388 replication incompetent[50]. Finally, the 351-nt RRE sequence within *env* was flanked by an XmaI
389 and XbaI site to permit easy replacement.
390 To create the MSCV constructs, Rev sequences were cloned into MSCV-based
391 plasmids containing the genes for fluorescent proteins (Figure 1). *Rev* was cloned upstream of
392 an IRES-fluorescent protein cassette, such that both Rev and the fluorescent protein are
393 expressed from a bicistronic transcript. Additional non-MSCV Rev-containing plasmids were
394 created using an immediate early CMV promoter to drive Rev expression without a fluorescent
395 marker. Additional descriptions of the assay constructs are available in Supplemental Methods.

396 Versions of the HIV-derived construct containing different export elements were created
397 by replacing the RRE with alternative sequences. These constructs are denominated in the text
398 as pNL4-3(unspliced transcript fluorescent marker)(transport element)(spliced transcript
399 fluorescent marker), as in pNL4-3(eGFP)(NL4-3 RRE)(mCherry). Similarly, different Rev
400 sequences or other *trans*-acting proteins were substituted in the MSCV-derived construct.
401 These constructs are denominated pMSCV - *trans*-acting protein – IRES – fluorescent marker,
402 as in pMSCV-NL4-3 Rev-IRES-TagBFP. The CMV constructs are denominated as in CMV-NL4-
403 3 Rev.
404
405 *Western blot of Gag-CHYSEL-eGFP product.*
406
407 A Western blot was performed to test the cleavage efficiency of the CHYSEL sequence
408 incorporated in the *gag*-CHYSEL-*eGFP* cassette. 293T/17 cells were transfected with
409 constructs expressing eGFP alone, the *gag*-CHYSEL-*eGFP* cassette, a *gag-eGFP* fusion
410 sequence, and mCherry as a negative control. Western blotting was performed on cell lysate
411 using primary antibodies to eGFP and p24[51]. Additional details are available in the
412 Supplemental Methods.
413
414 *Transfection-based functional assays.*
415 The assay system permits determination of Rev-RRE functional activity at the level of
416 transfection. 293T/17 cells were co-transfected with an HIV-derived construct and either a
417 MSCV-derived construct or CMV-Rev using the polyethylenamine method[52]. Cells were
418 incubated for 24 hours after transfection. Flow cytometry was performed using the Attune NxT
419 flow cytometer with autosampler attachment (Thermo Fischer Scientific). Data analysis,
420 including color compensation, was performed using FlowJo v10 (FlowJo, LLC). Further details
421 are available in the Supplemental Methods.

422 The gating schemes used are shown in Figure 2. Cells were selected (panel a) and cell
423 aggregates excluded (b). Cells expressing the HIV-derived construct (bottom) were identified by
424 reference to a negative control (top) (c). Untransfected cells in the bottom left quadrant were
425 excluded from further analysis. When an MSCV-derived construct was used, a derived gate
426 from (b) was constructed to include only cells expressing the third fluorescent marker linked to
427 the *trans*-acting protein (d). Finally, cells expressing both constructs were included in a final
428 analysis (e).
429 To calculate the activity of the intron export mechanism, cells in the appropriate gate for
430 analysis (panel c excluding the bottom left quadrant or e) were considered. The arithmetic mean
431 fluorescence intensity (MFI) was determined for both fluorescent signals of the HIV-derived
432 construct, and the ratio of the unspliced transcript marker MFI (i.e. eGFP) to the spliced
433 transcript marker MFI (e.g. TagBFP or mCherry) was calculated. The resulting ratio was
434 adjusted by subtracting the activity of the 0 ng Rev condition (or equivalent) to compensate for
435 eGFP "leakiness." Finally, functional activity values within the experiment were expressed as a
436 proportion of the maximum for any condition, which is set as 100.
437
438 *Rev titration assay.*
439
440 The responsiveness of the HIV-derived construct to increasing amounts of Rev plasmid
441 was determined. 1000 ng of the HIV-derived construct pNL4-3(eGFP)(NL4-3 RRE)(TagBFP)
442 (pHR5564) was transfected into $8x10^5$ 293T/17 cells along with 0, 6.25, 12.5, 25, 50, or 100 ng
443 of CMV-NL4-3 Rev plasmid (pHR5186). The total mass of DNA in each transfection was kept
444 constant by the addition of varying masses of an empty CMV plasmid (pHR16). Flow cytometry
445 was performed 36 hours after transfection.
446
447 *Assay of different RRE sequences.*

448

449 Three HIV-derived constructs expressing eGFP and TagBFP were created by replacing
450 the native RRE sequence with alternative 234-nt RREs. The RREs were from NL4-3 (pHR5580)
451 as well as two patient-derived viruses: SC3-M0A and SC3-M57A (Genbank accession
452 KF559160.1 and KF559162.1; pHR5584 and pHR5586)[24]. In a previous assay system, SC3-
453 M57A was found to have higher activity than SC3-M0A[24, 53]. The constructs were tested
454 individually by transfecting them into 293T/17 cells along with 0, 25, 50, or 100 ng of CMV-SC3
455 Rev plasmid using a Rev sequence from the same patient as the RRE sequences (Genbank
456 accession KF559146; pHR4776)[24]. The total DNA mass in each condition was kept constant by
457 the addition of a variable mass of an empty CMV construct (pHR16).

458

459 *Assay of different Rev sequences.*

460

461 Three MSCV-derived constructs expressing mCherry and one of three different Revs
462 were created. The Rev sequences used were from NL4-3 (pHR5625) and two subtype G HIV
463 viruses: 8-G and 9-G (Genbank accession FJ389367[54] and JX140676[55]; pHR5629 and
464 pHR5627). In a previous assay system, 9-G was found to have higher activity than 8-G Rev[23].
465 293T/17 cells were transfected with 1000 ng pNL4-3(eGFP)(NL4-3 RRE)(TagBFP) (pHR5580)
466 along with 0, 12.5, 25, 50, 100, or 200 ng of each of the MSCV constructs. The total DNA mass
467 in each transfection was kept constant by addition of a variable mass of an empty CMV plasmid
468 (pHR16).

469

470 *Assay of HERV-K RcRE and Rec activity.*

471

472 An HIV-derived eGFP-mCherry construct was modified by replacing the HIV RRE
473 sequence with a 433-nt HERV-K RcRE sequence (Genbank accession AF179225.1[37],

474 pHR5476). An additional HIV construct was created by inserting a second identical RcRE
475 sequence in tandem (pHR5631). Both RcRE-containing constructs and an HIV-derived RRE-
476 containing construct (pHR5460), were tested by transfecting them individually into 293T/17 cells
477 along with 0, 50, 100, 200, and 400 ng of a CMV plasmid containing the HERV-K Rec gene
478 (Genbank accession AY395522.1[38]), CMV-Rec (pHR5313). Flow cytometry was performed 48
479 hours after transfection.
480
481 *Assay of CTE activity.*
482
483 A 96-nt CTE derived from intron 10 of *NXF1* was incorporated into an eGFP-mCherry
484 HIV-derived construct in place of the RRE (see Genbank reference sequence NM_001081491.1
485 bases 1928 to 2023, pHR5781)[5]. An additional construct was created from this vector by
486 inserting a second *NXF1* CTE within *env* at the NheI site (pHR5783). Another HIV-derived
487 construct was created by replacing the RRE with a single copy of a 162-nt CTE from MPMV
488 (Genbank accession AF033815.1 bases 7388 to 7549, pHR5608)[2, 3].
489 Each of the resulting CTE constructs were tested by transfecting 1000 ng of each
490 individually into 293T/17 cells along with 133 ng CMV-Nxf1 plasmid (pHR3704) and 33 ng CMV-
491 Nxt1 plasmid (pHR2415) to overexpress these cellular factors. Flow cytometry was performed
492 48 hours after transfection.
493
494 *Small molecule Rev-RRE pathway inhibitor assay.*
495
496 A previously described heterocyclic compound, 3-amino-5-ethyl-4,6-dimethylthieno[2,3-
497 b]pyridine-2-carboxamide (103833), inhibits Rev-RRE function[40]. 293T/17 cells were pre-treated
498 with varying concentrations of 103833 and the incubation with the compound was continued
499 after transfection. To test the response of the Rev-RRE system to the inhibitor, cells were

500 transfected with 1000 ng of pNL4-3(eGFP)(NL4-3 RRE)(mCherry) (pHR5460) and 100 ng of
501 pMSCV-NL4-3 Rev-IRES-TagBFP (pHR5420). To test the response of the CTE system, cells
502 were transfected with 1000 ng of pNL4-3(eGFP)(MPMV CTE)(mCherry) (pHR5608), 133 ng
503 CMV-Nxf1 (pHR3704), and 33 ng CMV-NxT1 (pHR2415). Data analysis was conducted by
504 gating on cells successfully transduced with the HIV-derived two-color constructs (see Figure 2
505 panel c). Additional details are available in the Supplemental Methods.
506
507 Trans-*dominant negative Rev M10 assay.*
508
509 To measure the effect of the co-expression of Rev M10[42] on Rev-RRE activity, $4x10^5$
510 293T/17 cells were transfected with 1000 ng pNL4-3(eGFP)(NL4-3 RRE)(TagBFP) (pHR5580),
511 75 ng pMSCV-NL4-3 Rev-IRES-mCherry (pHR5625), and either 600 ng, 400 ng, 200 ng, 100
512 ng, 50 ng, or 0 ng of CMV-Rev M10 (pHR636). Variable amounts of an empty CMV plasmid
513 (pHR16) were included to maintain a constant DNA mass in each transfection. Data analysis
514 was performed by gating on cells successfully transduced with the HIV-derived construct (see
515 Figure 2 panel c). The arithmetic MFI of mCherry was determined to measure relative
516 expression of NL4-3 Rev.
517
518 *Packaging and titering HIV and MSCV constructs.*
519
520 The HIV-derived construct pNL4-3(eGFP)(NL4-3 RRE)(TagBFP) (pHR5580) was
521 packaged and VSV-G pseudotyped using a transient second generation packaging system
522 including psPAX2 (Addgene plasmid 12260, pHR5691) and pMD2.G (Addgene plasmid 12259,
523 pHR5693). Three MSCV constructs of the form pMSCV-Rev-IRES-mCherry carrying the NL4-3,
524 8-G, or 9-G Revs (pHR5625, 5629, 5627 respectively) were packaged as well in a transfection
525 system that included pHIT60-CMV-GagPol[56] (pHR1854) and pMD2.G (pHR5693). Vector stock

526 titer was determined in 293T/17 cells using constitutively expressed fluorescent proteins as the
527 marker of transduction. Additional details are available in the Supplemental Methods.
528
529 *Assay of Rev activity in transduced cells.*
530
531 The relative functional activity of Rev-RRE pairs was assayed by co-transducing CEM-
532 SS cells with viral vectors derived from pNL4-3(eGFP)(NLR-3 RRE)(TagBFP) and the different
533 pMSCV-Rev-IRES-mCherry constructs[57, 58]. As controls, the vectors were also added to the
534 CEM-SS cultures individually. Seventy-two hours after transduction, cells were harvested and
535 flow cytometry was performed. Functional activity was determined by gating on cells transduced
536 with both constructs (see Figure 2 panel e). See the Supplemental Methods for additional
537 details.
538
539 *Creation of T-cell line containing the integrated two color reporter construct.*
540
541 To create a cell line with a stably integrated HIV-derived proviral construct carrying
542 eGFP and mCherry, CEM-SS cells were transduced with a VSV-G pseudotyped pNL4-
543 3(eGFP)(NL4-3 RRE)(mCherry) viral vector (pHR5604). A limiting dilution was performed in 96
544 well plates to generate monoclonal populations, and clones expressing the HIV-derived
545 construct were identified by looking for mCherry expression via fluorescent microscopy. The
546 presence of a single integrated copy of the HIV construct was confirmed by a modified inverse
547 PCR procedure[59]. Additional details are available in the Supplemental Methods.
548 After identifying a clone with a single integrated HIV-derived construct, intron export was
549 demonstrated by transducing cells with the VSV-G pseudotyped pMSCV-NL4-3 Rev-IRES-
550 TagBFP (pHR5420) construct as above. Seventy-two hours after transduction, eGFP, mCherry,
551 and TagBFP expression was quantified by flow cytometry.

552

553 **Data Availability.**

554 No datasets were generated or analyzed during the current study. Sequences of the
555 evaluated intron-retention elements are available from Genbank at the accession numbers
556 included in the text. Plasmids and cell lines utilized in this study are available from the authors
557 upon request.

558

559 **References.**

560 1. Rekosh, D. & Hammarskjold, M. Intron retention in viruses and cellular genes: Detention,
561 border controls and passports. *Wiley Interdisciplinary Reviews: RNA* **9**, e1470 (2018).
562 2. Petropoulos, C. Retroviral taxonomy, protein structures, sequences, and genetic maps.
563 *Retroviruses.Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY*, 757-805 (1997).
564 3. Bray, M. *et al*. A small element from the Mason-Pfizer monkey virus genome makes human
565 immunodeficiency virus type 1 expression and replication Rev-independent. *Proc. Natl. Acad.*
566 *Sci. U. S. A.* **91**, 1256-1260 (1994).
567 4. Grüter, P. *et al*. TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export
568 from the nucleus. *Mol. Cell* **1**, 649-659 (1998).
569 5. Li, Y. *et al*. An intron with a constitutive transport element is retained in a Tap messenger
570 RNA. *Nature* **443**, 234 (2006).
571 6. Li, Y. *et al*. An NXF1 mRNA with a retained intron is expressed in hippocampal and
572 neocortical neurons and is translated into a protein that functions as an Nxf1 cofactor. *Mol. Biol.*
573 *Cell* **27**, 3903-3912 (2016).
574 7. Bor, Y. C. *et al*. The Wilms' tumor 1 (WT1) gene (+KTS isoform) functions with a CTE to
575 enhance translation from an unspliced RNA with a retained intron. *Genes Dev.* **20**, 1597-1608
576 (2006).

577 8. Fernandes, J. D., Booth, D. S. & Frankel, A. D. A structurally plastic ribonucleoprotein
578 complex mediates post-transcriptional gene regulation in HIV-1. *Wiley Interdisciplinary Reviews:*
579 *RNA* **7**, 470-486 (2016).
580 9. Frankel, A. D. & Young, J. A. HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **67**, 1-
581 25 (1998).
582 10. Pollard, V. W. & Malim, M. H. The HIV-1 rev protein. *Annual Reviews in Microbiology* **52**,
583 491-532 (1998).
584 11. Frankel, A. D. & Young, J. A. HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **67**,
585 1-25 (1998).
586 12. Hadzopoulou-Cladaras, M. *et al*. The rev (trs/art) protein of human immunodeficiency virus
587 type 1 affects viral mRNA and protein expression via a cis-acting sequence in the env region. *J.*
588 *Virol.* **63**, 1265-1274 (1989).
589 13. Hammarskjold, M. L. *et al*. Regulation of human immunodeficiency virus env expression by
590 the rev gene product. *J. Virol.* **63**, 1959-1966 (1989).
591 14. Malim, M. H., Hauber, J., Le, S., Maizel, J. V. & Cullen, B. R. The HIV-1 rev trans-activator
592 acts through a structured target sequence to activate nuclear export of unspliced viral mRNA.
593 *Nature* **338**, 254-257 (1989).
594 15. Bai, Y., Tambe, A., Zhou, K. & Doudna, J. A. RNA-guided assembly of Rev-RRE nuclear
595 export complexes. *Elife* **3**, e03656 (2014).
596 16. Neville, M., Stutz, F., Lee, L., Davis, L. I. & Rosbash, M. The importin-beta family member
597 Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear
598 export. *Current Biology* **7**, 767-775 (1997).
599 17. Baccam, P. *et al*. Subpopulations of equine infectious anemia virus Rev coexist in vivo and
600 differ in phenotype. *J. Virol.* **77**, 12122-12131 (2003).

601 18. Hidaka, M., Inoue, J., Yoshida, M. & Seiki, M. Post-transcriptional regulator (rex) of HTLV-1
602 initiates expression of viral structural proteins but suppresses expression of regulatory proteins.
603 *EMBO J.* **7**, 519-523 (1988).
604 19. Mertz, J. A., Simper, M. S., Lozano, M. M., Payne, S. M. & Dudley, J. P. Mouse mammary
605 tumor virus encodes a self-regulatory RNA export protein and is a complex retrovirus. *J. Virol.*
606 **79**, 14737-14747 (2005).
607 20. Magin, C., Lower, R. & Lower, J. cORF and RcRE, the Rev/Rex and RRE/RxRE
608 homologues of the human endogenous retrovirus family HTDV/HERV-K. *J. Virol.* **73**, 9496-9507
609 (1999).
610 21. Magin-Lachmann, C. *et al*. Rec (formerly Corf) function requires interaction with a complex,
611 folded RNA structure within its responsive element rather than binding to a discrete specific
612 binding site. *J. Virol.* **75**, 10359-10371 (2001).
613 22. Smyth, R. P., Davenport, M. P. & Mak, J. The origin of genetic diversity in HIV-1. *Virus Res.*
614 **169**, 415-429 (2012).
615 23. Jackson, P. E., Tebit, D. M., Rekosh, D. & Hammarskjold, M. Rev–RRE Functional Activity
616 Differs Substantially Among Primary HIV-1 Isolates. *AIDS Res. Hum. Retroviruses* (2016).
617 24. Sloan, E. A. *et al*. Limited nucleotide changes in the Rev response element (RRE) during
618 HIV-1 infection alter overall Rev-RRE activity and Rev multimerization. *J. Virol.* **87**, 11173-
619 11186 (2013).
620 25. Sherpa, C. *et al*. Evolution of the HIV-1 RRE during natural infection reveals key nucleotide
621 changes that correlate with altered structure and increased activity over time. *bioRxiv*, 483511
622 (2018).
623 26. Phuphuakrat, A. & Auewarakul, P. Functional variability of Rev response element in HIV-1
624 primary isolates. *Virus Genes* **30**, 23-29 (2005).

625 27. Phuphuakrat, A., Paris, R. M., Nittayaphan, S., Louisirirotchanakul, S. & Auewarakul, P.
626 Functional variation of HIV-1 Rev response element in a longitudinally studied cohort. *J. Med.*
627 *Virol.* **75**, 367-373 (2005).
628 28. Hua, J., Caffrey, J. J. & Cullen, B. R. Functional consequences of natural sequence
629 variation in the activation domain of HIV-1 Rev. *Virology* **222**, 423-429 (1996).
630 29. Bobbitt, K. R. *et al*. Rev activity determines sensitivity of HIV-1-infected primary T cells to
631 CTL killing. *Immunity* **18**, 289-299 (2003).
632 30. Belshan, M. *et al*. Genetic and biological variation in equine infectious anemia virus Rev
633 correlates with variable stages of clinical disease in an experimentally infected pony. *Virology*
634 **279**, 185-200 (2001).
635 31. Daelemans, D. *et al*. A synthetic HIV-1 Rev inhibitor interfering with the CRM1-mediated
636 nuclear export. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14440-14445 (2002).
637 32. Hope, T. J., Huang, X. J., McDonald, D. & Parslow, T. G. Steroid-receptor fusion of the
638 human immunodeficiency virus type 1 Rev transactivator: mapping cryptic functions of the
639 arginine-rich motif. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 7787-7791 (1990).
640 33. Wu, Y., Beddall, M. H. & Marsh, J. W. Rev-dependent indicator T cell line. *Curr. HIV. Res.* **5**,
641 394-402 (2007).
642 34. Adachi, A. *et al*. Production of acquired immunodeficiency syndrome-associated retrovirus in
643 human and nonhuman cells transfected with an infectious molecular clone. *J. Virol.* **59**, 284-291
644 (1986).
645 35. Kim, J. H. *et al*. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1
646 in human cell lines, zebrafish and mice. *PloS one* **6**, e18556 (2011).
647 36. Jin, L., Guzik, B. W., Bor, Y. C., Rekosh, D. & Hammarskjold, M. L. Tap and NXT promote
648 translation of unspliced mRNA. *Genes Dev.* **17**, 3075-3086 (2003).
649 37. Yang, J. *et al*. An ancient family of human endogenous retroviruses encodes a functional
650 homolog of the HIV-1 Rev protein. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13404-13408 (1999).

651 38. Mayer, J. *et al*. Human endogenous retrovirus HERV-K (HML-2) proviruses with Rec protein
652 coding capacity and transcriptional activity. *Virology* **322**, 190-198 (2004).
653 39. Heguy, A. Inhibition of the HIV Rev transactivator : a new target for therapeutic intervention.
654 *Front. Biosci.* **2**, d283-97 (1997).
655 40. Shuck-Lee, D. *et al*. Heterocyclic compounds that inhibit Rev-RRE function and human
656 immunodeficiency virus type 1 replication. *Antimicrob. Agents Chemother.* **52**, 3169-3179
657 (2008).
658 41. Nakamura, R. L. *et al*. Identification and Optimization of Thienopyridine Carboxamides as
659 Inhibitors of HIV Regulatory Complexes. *Antimicrob. Agents Chemother.* **61**,
660 10.1128/AAC.02366-16. Print 2017 Jul (2017).
661 42. Malim, M. H., Böhnlein, S., Hauber, J. & Cullen, B. R. Functional dissection of the HIV-1 Rev
662 trans-activator—derivation of a trans-dominant repressor of Rev function. *Cell* **58**, 205-214
663 (1989).
664 43. Iacampo, S. & Cochrane, A. Human immunodeficiency virus type 1 Rev function requires
665 continued synthesis of its target mRNA. *J. Virol.* **70**, 8332-8339 (1996).
666 44. Viphakone, N. *et al*. TREX exposes the RNA-binding domain of Nxf1 to enable mRNA
667 export. *Nature communications* **3**, 1006 (2012).
668 45. Jayaraman, B., Fernandes, J., Yang, S., Smith, C. & Frankel, A. D. Highly Mutable Linker
669 Regions Regulate HIV-1 Rev Function and Stability. *bioRxiv*, 424259 (2018).
670 46. Salminen, M. O. *et al*. Recovery of virtually full-length HIV-1 provirus of diverse subtypes
671 from primary virus cultures using the polymerase chain reaction. *Virology* **213**, 80-86 (1995).
672 47. Zhang, G., Gurtu, V. & Kain, S. R. An enhanced green fluorescent protein allows sensitive
673 detection of gene transfer in mammalian cells. *Biochem. Biophys. Res. Commun.* **227**, 707-711
674 (1996).
675 48. Shaner, N. C. *et al*. Improved monomeric red, orange and yellow fluorescent proteins
676 derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567 (2004).

677 49. Subach, O. M. *et al*. Conversion of red fluorescent protein into a bright blue probe. *Chem.*
678 *Biol.* **15**, 1116-1124 (2008).
679 50. Bryant, M. & Ratner, L. Myristoylation-dependent replication and assembly of human
680 immunodeficiency virus 1. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 523-527 (1990).
681 51. Wehrly, K. & Chesebro, B. p24 antigen capture assay for quantification of human
682 immunodeficiency virus using readily available inexpensive reagents. *Methods* **12**, 288-293
683 (1997).
684 52. Boussif, O. *et al*. A versatile vector for gene and oligonucleotide transfer into cells in culture
685 and in vivo: polyethylenimine. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7297-7301 (1995).
686 53. Sherpa, C., Rausch, J. W., Le Grice, S. F., Hammarskjold, M. L. & Rekosh, D. The HIV-1
687 Rev response element (RRE) adopts alternative conformations that promote different rates of
688 virus replication. *Nucleic Acids Res.* **43**, 4676-4686 (2015).
689 54. Yamaguchi, J. *et al*. HIV type 1 group M subtype G in Cameroon: five genome sequences.
690 *AIDS Res. Hum. Retroviruses* **25**, 469-473 (2009).
691 55. Sanchez, A. M. *et al*. Development of a contemporary globally diverse HIV viral panel by the
692 EQAPOL program. *J. Immunol. Methods* **409**, 117-130 (2014).
693 56. Soneoka, Y. *et al*. A transient three-plasmid expression system for the production of high
694 titer retroviral vectors. *Nucleic Acids Res.* **23**, 628-633 (1995).
695 57. Foley, G. E. *et al*. Continuous culture of human lymphoblasts from peripheral blood of a
696 child with acute leukemia. *Cancer* **18**, 522-529 (1965).
697 58. Nara, P. *et al*. Simple, rapid, quantitative, syncytium-forming microassay for the detection of
698 human immunodeficiency virus neutralizing antibody. *AIDS Res. Hum. Retroviruses* **3**, 283-302
699 (1987).
700 59. Chun, T. *et al*. In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to
701 stable latency. *Nat. Med.* **1**, 1284 (1995).

Acknowledgements.

The pMSCV-IRES-mCherry and pMSCV-IRES-TagBFP vectors were gifts from Dario Vignali obtained through Addgene (unpublished). The plasmids psPAX2 and pMD2.G were gifts from Didier Trono obtained through Addgene. Technical assistance and guidance on data interpretation was provided by the University of Virginia Flow Cytometry Core Facility. The p24 hybridoma (183-H12-5C) (Cat# 1513) was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH from Dr. Bruce Chesebro.

Author Contributions.

P. E. H. J wrote the first draft of the manuscript text, prepared all figures, and performed many of the experiments. M. S. and S. R. performed the experiments relating to HERV-K Rec and the RcRE. J. H. was involved in the development of the two color vector and performed many of the initial experiments not reported here to characterize the system, as well as some of the experiments reported here. D. R. and M.-L. H. conceptualized the assay system, contributed to study design, manuscript revisions, and project oversight. All authors reviewed the manuscript.

Competing Interests.

The authors declare no competing interests.

Tables and Figures.

726 Figure 1. Schematic of retroviral vector constructs used in this study. The two-color constructs
727 (left boxes) are based on the HIV NL4-3 genome with the following modifications from left to
728 right: mutation of the myristoylation signal in *gag*, truncation of *gag*, addition of a CHYSEL
729 sequence and an *eGFP* gene in the *gag* reading frame, deletion of the frameshifting site,
730 missense mutation of the *rev* start codon, insertion of a stop codon in the first exon of rev,
731 introduction of a frameshift mutation in *env*, placement of an XmaI and an XbaI site flanking the
732 RRE within env, and deletion of the beginning of *nef* with the substitution of a second
733 fluorescent protein in this position. The single-color constructs (right boxes) are MSCV vectors
734 modified with a *rev* gene flanked by two restriction enzyme sites as well as an in-frame IRES
735 and a fluorescent protein gene. Two construct sets were used in this study. (a) Construct set A
736 has an eGFP in the *gag* reading frame, mCherry in the *nef* reading frame, and TagBFP in the
737 Rev vector. (b) Construct set B has an eGFP in the *gag* reading frame, TagBFP in the *nef*
738 reading frame, and mCherry in the Rev vector. Myr = myristoylation site, CHYSEL = cis-acting
739 hydrolase element, f.s. = frameshift site, LTR = long terminal repeat, GFP = green fluorescent
740 protein, BFP = blue fluorescent protein, RFP = red fluorescent protein, MSCV = murine stem
741 cell virus, IRES = internal ribosomal entry site, ψ = packaging signal.
742
743 Figure 2. Assay gating strategies. The top panels (A) show untransfected cells. The bottom
744 panels (B) show cells transfected with the vector constructs shown in Figure 1B. 293T/17 cells
745 for analysis were identified by gating first on a forward-scatter versus side-scatter plot (a)
746 followed by an additional gate for single cells constructed on a forward-scatter area versus
747 forward-scatter height plot (b). The fluoresence intensity (eGFP versus TagBFP) of each
748 individual cell is displayed, representing the two markers carried on the HIV-derived construct.
749 (c). Untransfected cells were used to define the double-negative population which was excluded
750 from further analysis. Cells which expressed the HIV-derived construct (i.e. Q1, Q2 and Q3 -
751 eGFP-positive or TagBFP-positive or both) were used for analysis. In other experiments utilizing

752 the MSCV-derived construct carrying a third fluorescent marker, an alternative analysis could be
753 used. In addition to gating on single cells expressing the HIV-derived construct, a gate was
754 created including cells expressing the MSCV-derived construct as well (i.e. mCherry positive)
755 (d). Finally, the population of cells expressing both the HIV-derived construct (i.e. Q1, Q2, or Q3
756 from c) and the MSCV-derived construct (i.e. mCherry positive from d) was used for analysis
757 (e). Red boxes denote gates used for calculation of functional activity.
758
759 Figure 3. Response of the two color reporter vector to increasing amounts of Rev plasmid. 1000
760 ng of an eGFP-mCherry (a) or eGFP-TagBFP (b) HIV-derived construct was transfected into
761 $8x10^5$ 293T/17 cells along with a variable amount of CMV-Rev plasmid. Single cells were
762 plotted with Rev-dependent eGFP signal coming from the unspliced mRNA along the y-axis and
763 the Rev-independent signal coming from the fully spliced mRNA along the x-axis (mCherry in a
764 and TagBFP in b). The amount of Rev plasmid transfected along with the two-color construct is
765 listed below each plot. (c) For each construct, the ratio of Rev-dependent (eGFP) to Rev-
766 independent (mCherry or TagBFP) signal MFI was calculated for each experimental condition
767 as a measure of Rev-RRE functional activity (see gating strategy in Figure 2 panel c). The
768 activity in the 100 ng condition for each construct was set as 100. N=2, vertical error bars
769 represent SEM, which are too small to be seen in this figure.
770
771 Figure 4. Differential activity of RRE and Rev sequences. (a) Three different 234-nt RREs were
772 cloned into the eGFP/TagBFP HIV-derived construct, including RREs from NL4-3 and two
773 derived from patient isolates (SC3-M0A and SC3-M57A)[24]. 1000 ng of the HIV-derived construct
774 was transfected into $8x10^5$ 293T/17 cells along with 0, 25, 50, or 100 ng of pCMV-SC3-(M0-
775 B/M57-A) Rev[24]. Twenty-four hours after transfection, flow cytometry was performed and the
776 ratio of eGFP/TagBFP signal was determined as a measure of Rev-RRE functional activity. The

777 maximum measured activity was set as 100. (b) Three Rev sequences from NL4-3 or two
778 different primary isolates[54, 55] were cloned into the MSCV vector carrying an mCherry
779 fluorescent marker. 1000 ng of an eGFP/TagBFP HIV-derived construct containing an NL4-3
780 RRE sequence was transfected into $8x10^5$ 293T/17 cells along with 12.5, 25, 50, 100, or 200 ng
781 of a pMSCV-Rev-IRES-mCherry construct. Flow cytometry was performed 24 hours after
782 transfection and the data were analyzed as above. N=2, bars, which are too small to be seen in
783 this figure, represent SEM for both graphs.
784
785 Figure 5. Activity measurements of heterologous viral and cellular RNA export elements in the
786 HIV-derived reporter construct. The eGFP-mCherry HIV-derived construct was modified to carry
787 a variety of RNA transporting elements (or no element) instead of the RRE seen in Figure 1.
788 The modified vectors were then transfected into 293T cells and eGFP and mCherry
789 fluorescence was measured. (a). An HIV-derived construct containing an MPMV CTE was
790 tested along with a construct containing no element. The activity of the MPMV CTE construct
791 was set as 100 and the other values were normalized accordingly. (b). HIV-derived constructs
792 containing either one or two copies of the CTE from *NXF1* intron 10 were tested along with a
793 construct containing no element. Activity of the dual *NXF1* CTE construct was set as 100 and
794 the other values were normalized accordingly. In both (a) and (b), Nxf1 and Nxt1 were co-
795 expressed plasmids under the control of the CMV-IE promoter. N=2, bars represent SEM. (c)
796 HIV-derived constructs containing single or dual HERV-K RcRE elements or an NL4-3 RRE
797 were co-transfected together with either 0, 50, 100, 200, or 400 ng CMV-Rec plasmid. The
798 activity of the dual RcRE construct in the 400 ng CMV-Rec condition was set as 100 and the
799 other values were normalized accordingly. (d) The 200 ng data point from C is presented in bar
800 graph form for comparison. GFP = green fluorescent protein, MPMV = Mason-Pfizer monkey
801 virus, CTE = constitutive transport element, RcRE= Rec Response Element from HERV-K.
802

803 Figure 6. The effect of inhibitors of Rev function on fluorescent protein expression from the
804 reporter vectors. (a) Compound 103833 decreases Rev-RRE functional activity in a dose
805 dependent manner but does not affect CTE function. Two vector constructs were tested in the
806 presence of compound 103833, a Rev-pathway small molecule inhibitor. The RRE construct
807 was identical to that in Figure 1a. The CTE construct was created by replacing the RRE with an
808 MPMV CTE. The constructs were transfected into 293T/17 cells with Rev (RRE construct) or
809 Nxf1 and NxT1 (CTE construct). Cells were pre-treated with the indicated concentrations of
810 103833 for 24 hours before transfection and remained in the medium until 24 hours after
811 transfection when the cells were harvested for analysis. Both eGFP and mCherry signals were
812 quantified for both constructs. The MFI for each fluorescent signal in the absence of inhibitor
813 was set as 100 and the other values were normalized accordingly. (b) *Trans*-dominant negative
814 Rev M10 inhibits Rev-RRE functional activity in a dose dependent manner. $4x10^5$ 293T/17 cells
815 were transfected with 1000 ng of an eGFP-TagBFP HIV-derived vector along with 75 ng of
816 pMSCV-Rev-IRES-mCherry vector (see Figure 1B). Additionally, a variable mass of a plasmid
817 expressing *trans*-dominant negative Rev-M10 plasmid (pCMV-TD-Rev) was transfected into the
818 system along with an empty vector to equalize total DNA mass of the pCMV plasmids to a total
819 of 600 ng. The MFI of the Rev-dependent (eGFP), Rev-independent (TagBFP), and Rev-linked
820 (mCherry) signals was determined for each ratio of Rev plasmid to Rev-M10 plasmid. For each
821 fluorescent protein, the MFI in the absence of Rev-M10 was set at 100 and the other values
822 were normalized accordingly. N=2, bars, which are too small to be seen in this figure, represent
823 SEM for both graphs. MFI = arithmetic mean fluorescence intensity.
824
825 Figure 7. Measurement of Rev-RRE functional activity from an integrated chromatin associated
826 provirus. HIV- and MSCV-derived constructs as in Figure 1 b were packaged and VSV-G
827 pseudotyped in 293T/17 cells. A single HIV-derived construct with an NL4-3 RRE was used
828 along with three MSCV constructs with 9-G, 8-G, or NL4-3 Rev sequences. CEM-SS cells were

829 transduced with both the HIV-derived construct and one of the MSCV-derived constructs at a
830 multiplicity of infection of 0.5. After 72 hours, fluorescence was measured via flow cytometry and
831 gating for analysis was performed as in Figure 2 e (i.e. analyzed cells fluoresce with mCherry,
832 and with eGFP and/or TagBFP). To generate dot plots, single cells were plotted with Rev-
833 dependent eGFP signal along the y-axis and the Rev-independent TagBFP signal along the x-
834 axis with the (A) NL4-3 Rev, (B) 8-G Rev, and (C) 9-G Rev. Flow cytometry plots represent all
835 single cells measured in three experimental replicates prior to analysis gating. (D) Relative Rev-
836 RRE functional activity was calculated for each Rev construct as the ratio of eGFP:TagBFP MFI
837 in the set of cells expressing fluorescent markers from both transducing constructs. The 9-G
838 Rev activity was set as 100 and the other values were normalized accordingly. N=3, bars
839 represent SEM.
840
841 Figure 8. Induction of eGFP expression from a CEM-SS cell line containing a stably integrated
842 single copy of the reporter plasmid. CEM-SS cells were transduced with an eGFP-mCherry HIV-
843 derived construct and a clonal cell line was created by limiting dilution. Thus, almost all cells
844 show a uniformity of mCherry expression (Plot a-left). The cell line was subsequently
845 transduced by a pMSCV-Rev-IRES-TagBFP and fluorescence was measured for all three
846 markers. Untransduced cells are shown on the left, cells transduced with the MSCV-derived
847 Rev/BFP expressing vector are shown on the right. Plot a (right) shows that transduction
848 caused about 8.27% of the cells to express TagBFP. Plot b (right) shows that transduction
849 caused about 8.79% of the cells to express eGFP. Plot c (right) shows that there is a linear
850 relationship between the BFP and eGFP signal, indicating a strong correlation between Rev
851 levels in a cell (BFP) and expression of the unspliced mRNA isoform (eGFP).
852
853 Figure S1. Independent measurement of fluorescent markers from Rev-dependent and Rev-
854 independent transcripts. 293T/17 cells were transfected with either eGFP-mCherry or eGFP-

855 TagBFP HIV-derived constructs along with varying amounts of CMV-Rev plasmid, as in Figure
856 3. For each HIV-derived construct and each experimental condition, the mean fluorescence
857 intensity of both the Rev-independent (TagBFP or mCherry) and Rev-dependent (eGFP)
858 fluorescent markers was measured. The maximum MFI for each marker on each transcript was
859 set as 100, and the other values for that marker were normalized accordingly.
860
861 Figure S2. The P2A CHYSEL sequence allows the production of separate Gag and eGFP
862 proteins. To produce a fluorescent signal in a Rev-dependent fashion, *gag* was truncated and a
863 cassette was inserted in frame consisting of a P2A CHYSEL sequence and the eGFP gene. The
864 CHYSEL was expected to result in the production of two separate polypeptides – a Gag
865 fragment and a free eGFP. Western blotting was performed to demonstrate that this cleavage
866 was efficient. 293T cells were transfected with constructs producing eGFP alone (lane 1), an
867 HIV vector similar to Figure 1A with Gag-CHYSEL-eGFP (lane 2), an HIV vector the with a Gag-
868 eGFP fusion sequence without a CHYSEL (lane 3), and an HIV vector producing only mCherry
869 (lane 4). Blotting was performed using an antibody directed against eGFP. Free eGFP is seen in
870 lanes 1 and 2. The 70 kDa band in lane 3 corresponds to the Gag-eGFP fusion product.

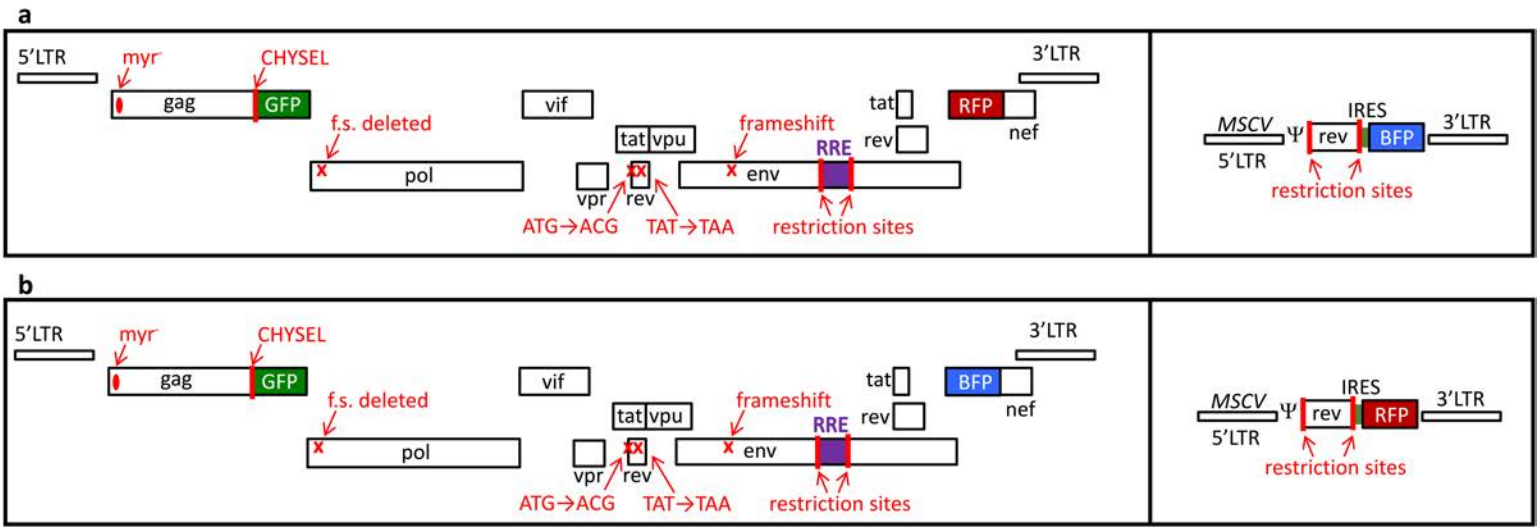
a
5'LTR
myr
CHYSEL
gag
GFP
f.s. deleted
pol
vif
vpr
tat
vpu
rev
ATG→ACG
TAT→TAA
frameshift
RRE
env
restriction sites
tat
rev
RFP
nef
3'LTR
MSCV
Ψ
rev
IRES
BFP
3'LTR
5'LTR
restriction sites
b
5'LTR
myr
CHYSEL
gag
GFP
f.s. deleted
pol
vif
vpr
tat
vpu
rev
ATG→ACG
TAT→TAA
frameshift
RRE
env
restriction sites
tat
rev
BFP
nef
3'LTR
MSCV
Ψ
rev
IRES
RFP
3'LTR
5'LTR
restriction sites

Figure 1

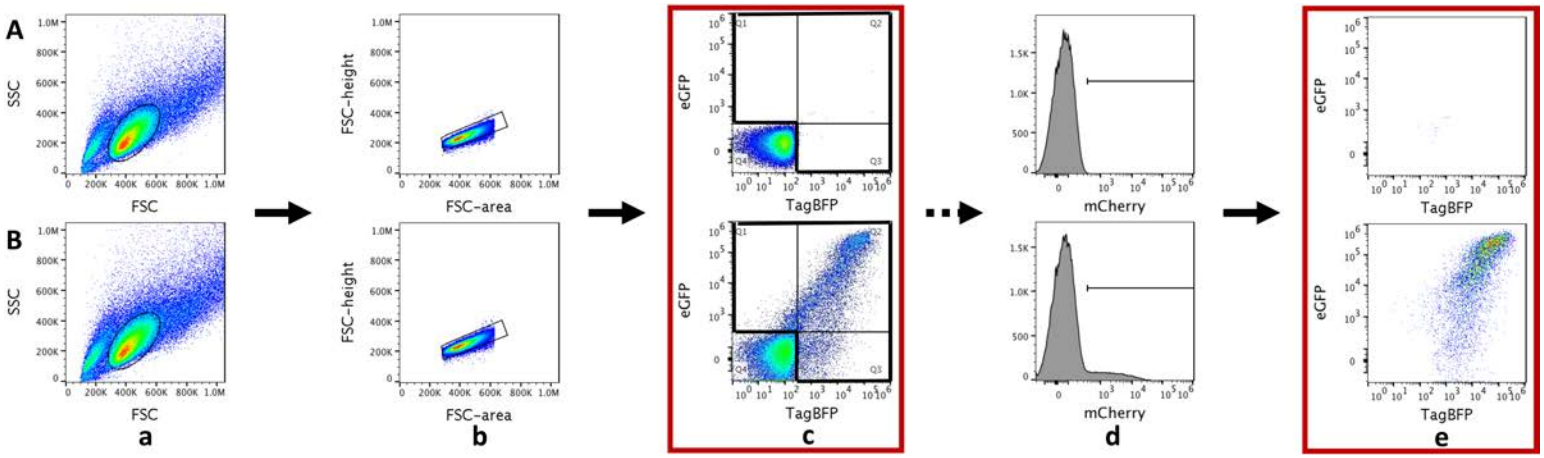

Figure 2

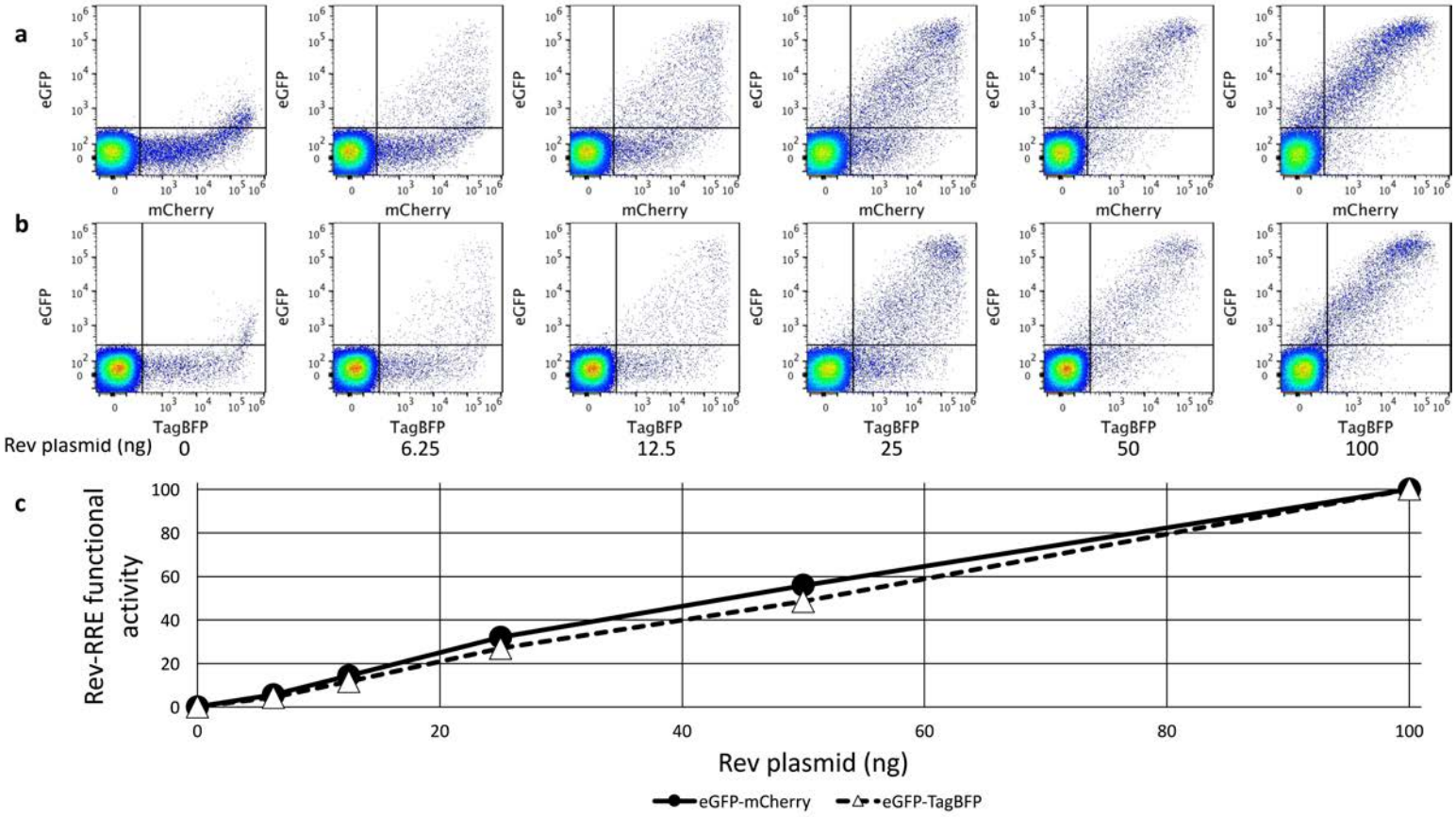
a
b
c
eGFP
mCherry
TagBFP
Rev plasmid (ng)
0
6.25
12.5
25
50
100
Rev-RRE functional activity
Rev plasmid (ng)
eGFP-mCherry
eGFP-TagBFP

Figure 3

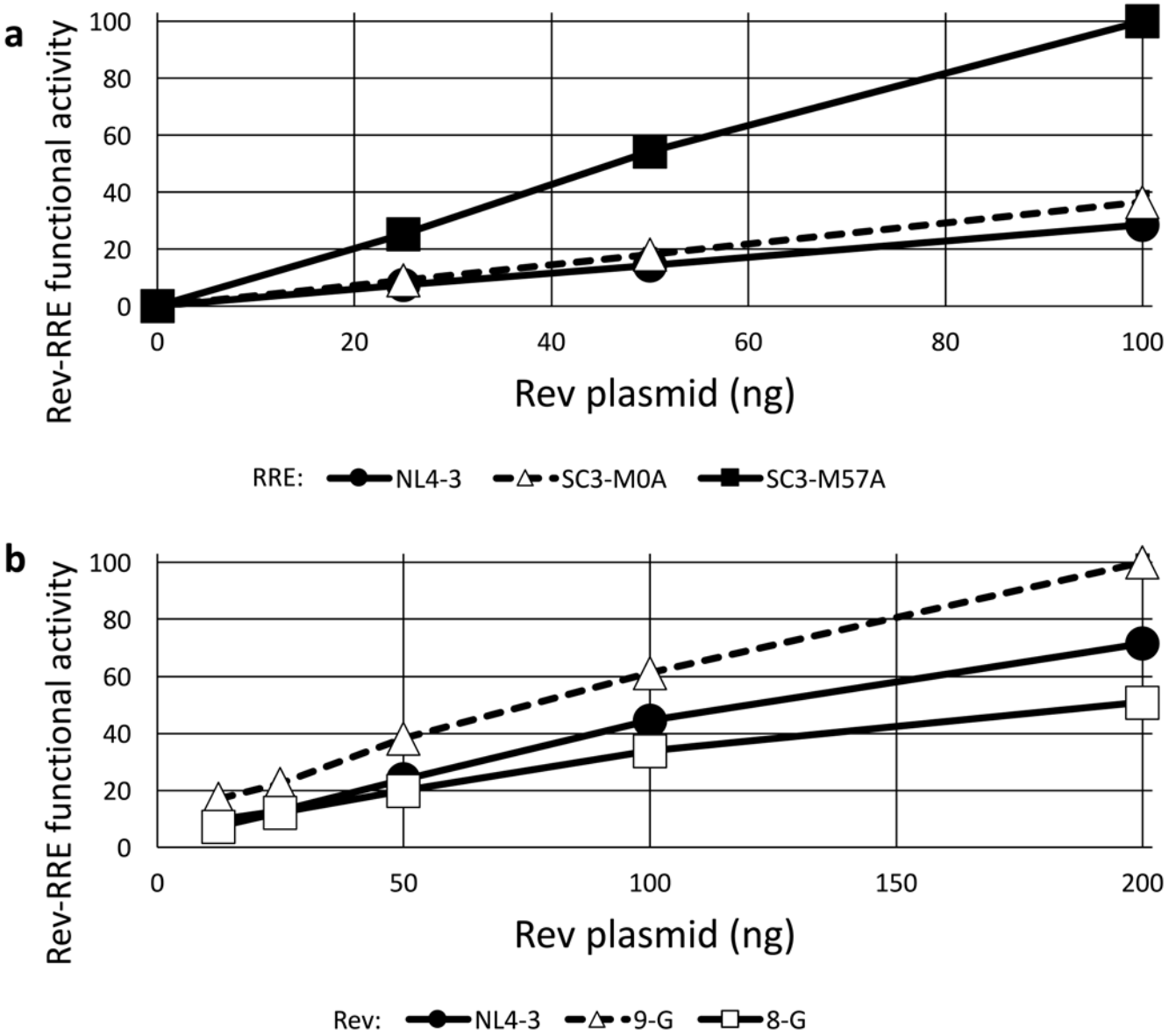
a
Rev-RRE functional activity
100
80
60
40
20
0
0
20
40
60
80
100
Rev plasmid (ng)
RRE:
NL4-3
SC3-M0A
SC3-M57A
b
Rev-RRE functional activity
100
80
60
40
20
0
0
50
100
150
200
Rev plasmid (ng)
Rev:
NL4-3
9-G
8-G

Figure 4

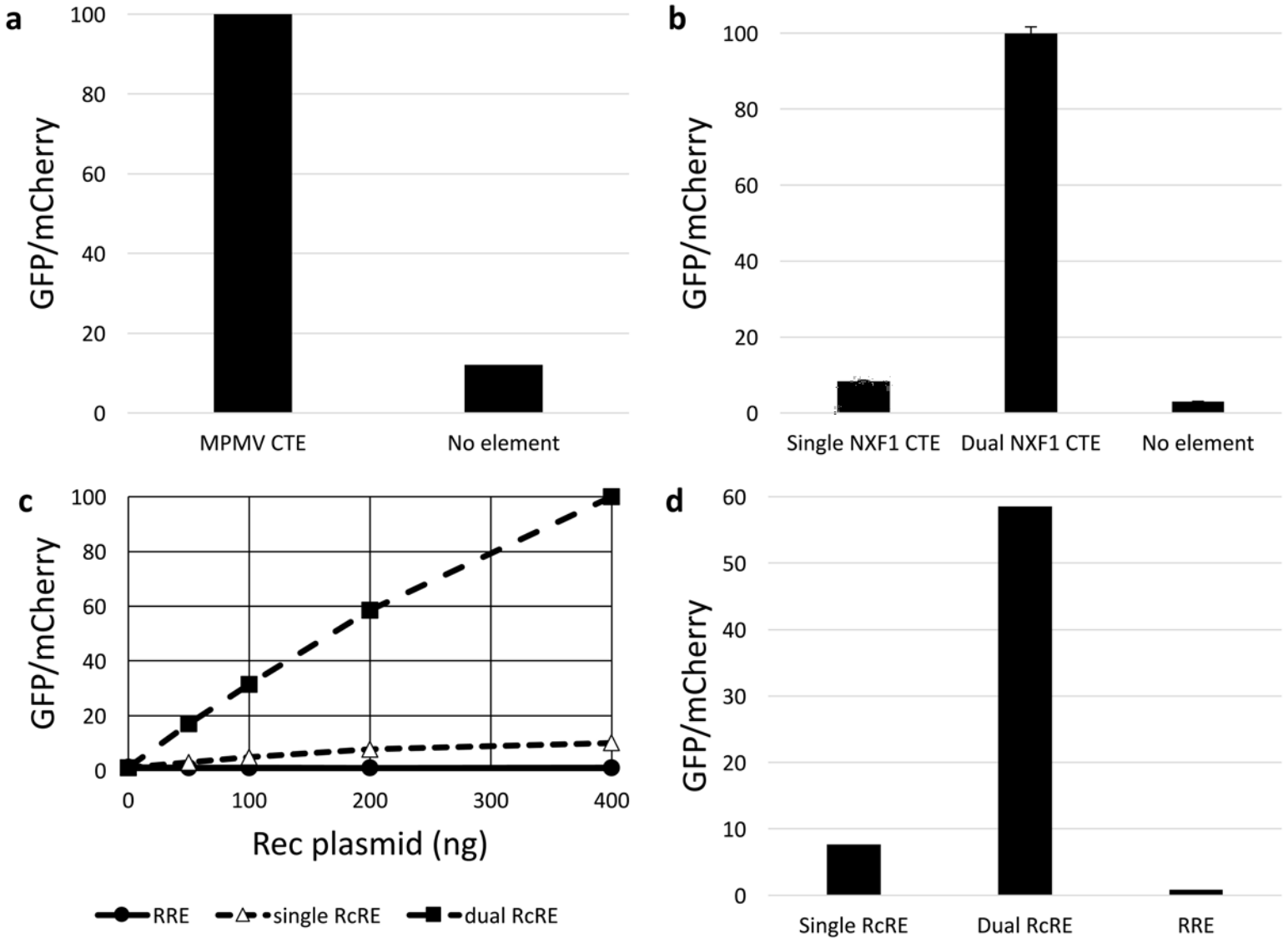
a
100
80
60
40
20
0
GFP/mCherry
MPMV CTE
No element
b
100
80
60
40
20
0
GFP/mCherry
Single NXF1 CTE
Dual NXF1 CTE
No element
c
100
80
60
40
20
0
GFP/mCherry
0
100
200
300
400
Rec plasmid (ng)
RRE
single RcRE
dual RcRE
d
60
50
40
30
20
10
0
GFP/mCherry
Single RcRE
Dual RcRE
RRE

Figure 5

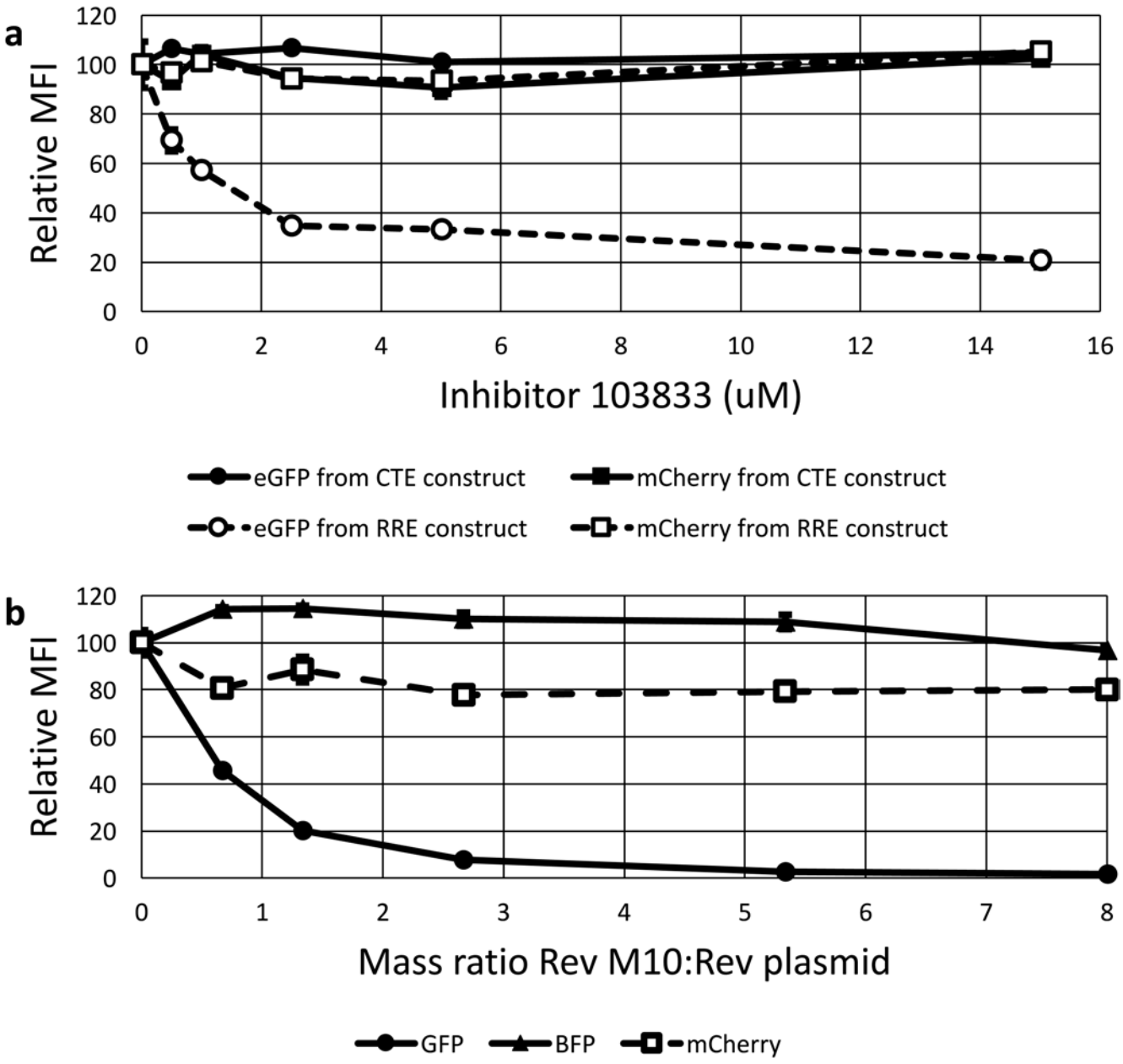
a
Relative MFI
120
100
80
60
40
20
0
0
2
4
6
8
10
12
14
16
Inhibitor 103833 (uM)
eGFP from CTE construct
mCherry from CTE construct
eGFP from RRE construct
mCherry from RRE construct
b
Relative MFI
120
100
80
60
40
20
0
0
1
2
3
4
5
6
7
8
Mass ratio Rev M10:Rev plasmid
GFP
BFP
mCherry

Figure 6

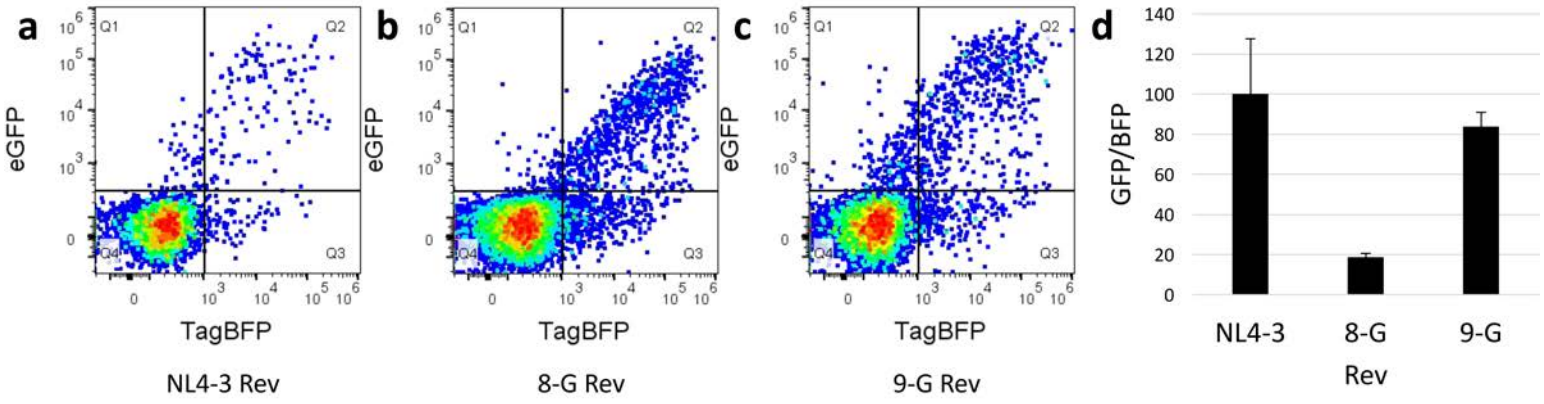
a
b
c
d
eGFP
TagBFP
Q1
Q2
Q3
Q4
NL4-3 Rev
8-G Rev
9-G Rev
GFP/BFP
140
120
100
80
60
40
20
0
NL4-3
8-G
9-G
Rev

Figure 7

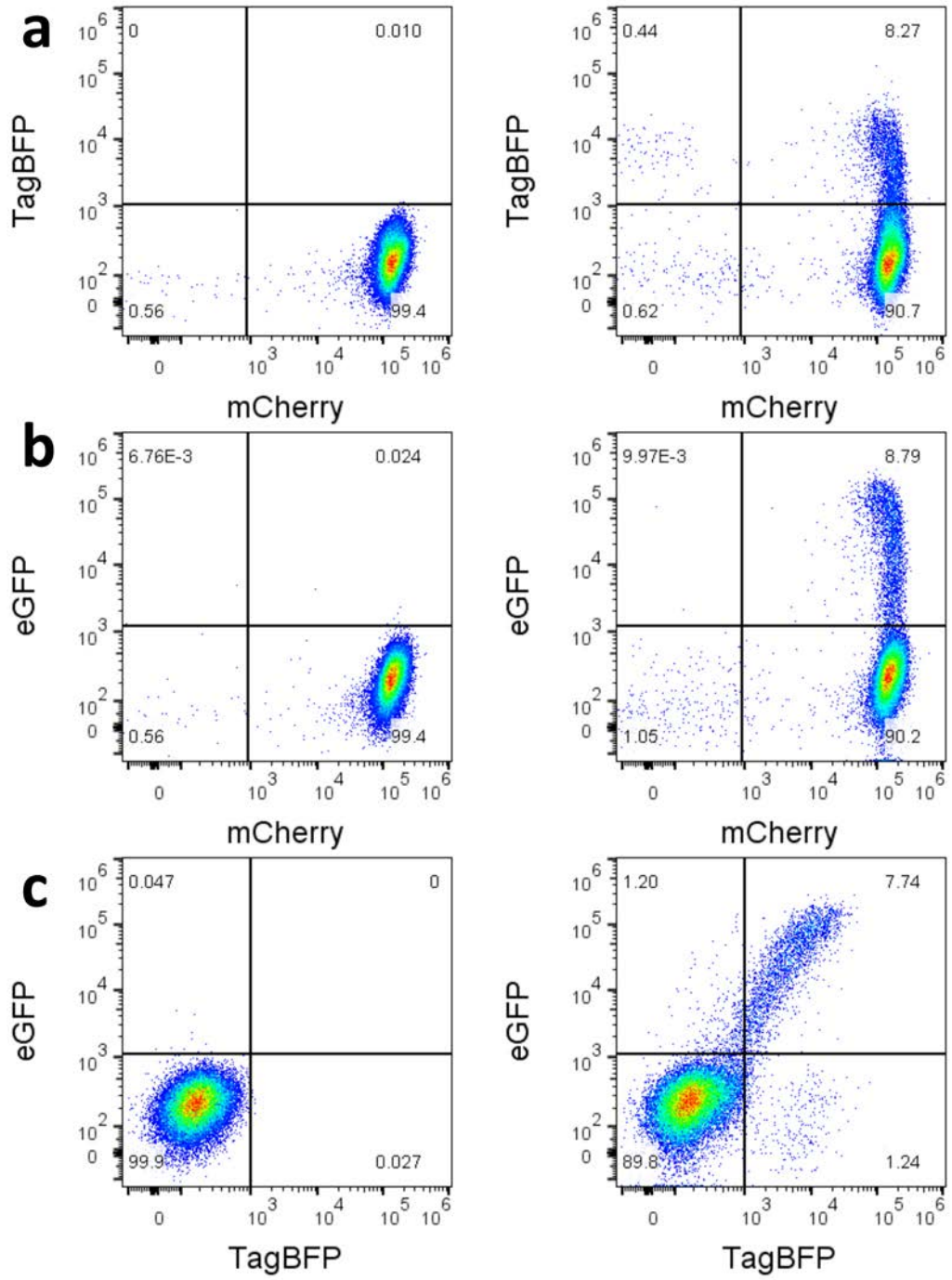
a
TagBFP
mCherry
0
0.010
0.56
99.4
0.44
8.27
0.62
90.7
b
eGFP
mCherry
6.76E-3
0.024
0.56
99.4
9.97E-3
8.79
1.05
90.2
c
eGFP
TagBFP
0.047
0
99.9
0.027
1.20
7.74
89.8
1.24

Figure 8