CRNKL1 is a highly selective regulator of intron-retaining HIV-1 and cellular mRNAs

Han Xiao[1], Emanuel Wyler[2#], Miha Milek[2#], Bastian Grewe[3], Philipp Kirchner[4], Arif Ekici[4], Ana Beatriz Oliveira Villela Silva[1], Doris Jungnickl[1], Markus Landthaler[2,5], Armin Ensser[1], and Klaus Überla[1*]

[1] Institute of Clinical and Molecular Virology, University Hospital Erlangen, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany

[2] Berlin Institute for Medical Systems Biology, Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, Robert-Rössle-Strasse 10, 13125, Berlin, Germany

[3] Department of Molecular and Medical Virology, Ruhr-University, Bochum, Germany

[4] Institute of Human Genetics, University Hospital Erlangen, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany

[5] IRI Life Sciences, Institute für Biologie, Humboldt Universität zu Berlin, Philippstraße 13, 10115, Berlin, Germany

[#] these two authors contributed equally

[*]Corresponding author:
Prof. Dr. Klaus Überla
Institute of Clinical and Molecular Virology, University Hospital Erlangen
Friedrich-Alexander Universität Erlangen-Nürnberg
Schlossgarten 4, 91054 Erlangen
Germany
Tel: (+49) 9131-8523563
e-mail: klaus.ueberla@fau.de

Abstract

The HIV-1 Rev protein is a nuclear export factor for unspliced and incompletely-spliced HIV-1 RNAs. Without Rev, these intron-retaining RNAs are trapped in the nucleus. A genome-wide screen identified nine proteins of the spliceosome which all enhanced expression from the HIV-1 unspliced RNA after CRISPR/Cas knock-down. Depletion of DHX38, WDR70 and four proteins of the Prp19-associated complex (ISY1, BUD31, XAB2, CRNKL1) resulted in a more than 20-fold enhancement of unspliced HIV-1 RNA levels in the cytoplasm. Targeting of CRNKL1, DHX38, and BUD31 affected nuclear export efficiencies of the HIV-1 unspliced RNA to a much larger extent than splicing. Transcriptomic analyses further revealed that CRNKL1 also suppresses cytoplasmic levels of cellular mRNAs with selectively retained introns. Thus, CRNKL1 dependent nuclear retention seems to be a novel mechanism for the regulation of cytoplasmic levels of intron-retaining cellular mRNAs that is harnessed by HIV-1 to direct its complex splicing pattern.

Introduction

Intron retention is increasingly recognized as a regulatory mechanism of gene expression and diversification of the cellular proteome (1). Transcriptomic analyses indicate that a large percentage of eukaryotic genes encode transcripts retaining introns (2, 3). Intron-retaining transcripts may be trapped in the nucleus leading to their degradation. Alternatively, specialized nuclear export pathways may enable certain intron-retaining transcripts to reach the cytoplasm, where they can be translated or undergo nonsense-mediated decay. Retroviruses use intron retention to encode multiple proteins from one primary transcript. RNA-secondary structures or cellular or viral adaptor proteins link intron-retaining retroviral transcripts in the nucleus to the TAP/NXF1 or the CRM-1 nuclear export factors (4), leading to the export of intron-retaining RNAs. While a number of studies explored how nuclear retention can be overcome, the factors mediating nuclear retention of intron-containing mRNAs are mostly unknown, but are likely linked to partial assembly of splicing complexes on intron-containing RNA (5).

The first evidence for a regulated export and expression of intron-retaining RNAs in mammalian cells was obtained in studies investigating the replication of human immunodeficiency virus type 1 (HIV-1). Alternative splicing and intron retention leads to three types of transcripts: fully-spliced (FS) mRNAs, incompletely-spliced (IS) mRNAs, and the unspliced (US) genomic RNA (6). The process of viral gene expression is temporally regulated, resulting in early expression of regulatory proteins and late expression of structural proteins (7, 8). In the early phase, the US and IS transcripts are retained in the nucleus, and are subjected to the nuclear splicing or degradation machinery (9). Only FS transcripts are transported to the cytoplasm via the default mRNA export pathway, TAP/NXF1, leading to expression of Tat, Rev and Nef proteins (10, 11). The regulatory proteins Tat and Rev are nuclear cytoplasmic shuttle proteins and regulate viral gene expression. While Tat transactivates transcription by enhancing RNA elongation, Rev fine-tunes posttranscriptional processes and acts as a switch to initiate the late phase of the viral replication cycle (reviewed in Grewe and Uberla, 2010; Mousseau and Valente, 2017). It binds to the Rev-responsive element (RRE) which is present in the HIV-1 US and IS transcripts, but not in the FS transcripts, and multimerizes on the RRE. The protein exporting factor CRM1 and associated factors are recruited by the Rev-RRE complex, leading to the nuclear-cytoplasmic export of these two classes of RNA (14).

Nuclear retention of intron-retaining HIV-1 RNAs has been explained by their suboptimal splice sites, leading to the trapping of the intron-retaining RNAs in the spliceosome and subsequent degradation

70 (15). In addition, multiple inhibitory sequences (INS) were identified in the retained introns (16–20). The
71 presumed inactivation of these inhibitory sequences by codon-optimizing the AT-rich HIV-1 sequence
72 abolished nuclear retention (21) independent of the presence or absence of HIV-1 splice donor and
73 splice acceptor sites (22, 23). Why AT-rich intron-containing HIV-1 sequences, but not the codon-
74 optimized variants are retained remains elusive. It has been reported that knock-down of the
75 paraspeckle lncRNA NEAT1 could promote HIV Gag production through increased nucleocytoplasmic
76 export of INS-containing RNAs, implicating the nuclear paraspeckles in the nuclear retention of HIV US
77 and IS RNAs (24). However, this has been questioned by a study demonstrating that HIV-1 US RNA was
78 not actually colocalized with paraspeckles (25), despite that several paraspeckle proteins (PSF, Martin3
79 and RBM14) have been reported to associate with HIV-1 INS elements and/or promote the Rev-
80 dependent RNA export (26–28).

81 In addition to these viral determinants and lncRNA contributing to nuclear mRNA retention, depletion
82 of hnRNPA2/B1 has been shown to increase cytoplasmic levels of HIV-1 genomic RNA, but this
83 increase did not enhance Gag expression levels and the precise mechanism remains unknown (29).
84 With regard to nuclear retention of mRNA in general, a cellular complex named pre-mRNA retention
85 and splicing complex (RES) has been identified in yeast (30). The RES is composed of Bud13, Snu17p, and
86 Pml1P and inactivation of either of these genes in yeast reduces splicing of an intron with a weak
87 5´splice donor. In addition, Pml1P inactivation enhanced expression of a pre-mRNA reporter construct
88 without substantial reduction of the spliced mRNA suggesting a role of Pml1P and the RES in nuclear
89 retention. More recently, BUD13 was identified to bind to the poorly-spliced mRNA of mammalian IRF7,
90 a master regulator of the interferon response (31). Knock-down of BUD13 specifically enhanced
91 retention of intron 4 of IRF7 in stimulated cells thus reducing IRF7 protein levels and dampening the
92 interferon response. However, intron-containing IRF7 mRNA could not be detected in the cytoplasm (31),
93 indicating additional nuclear trapping mechanisms or rapid cytoplasmic degradation of intron-containing
94 IRF7 mRNA. HIV-1 introns also differ from typical RES-dependent introns (31–33) by their low GC
95 content, challenging the hypothesis that HIV-1 splicing may be regulated by the RES.

96 An alternative nuclear trapping mechanism could be a cellular quality control mechanism for mRNAs
97 located at the nuclear pore complex (NPC). Experimental or stress-induced interference with this
98 complex results in leakage of intron-retaining mRNAs into the cytoplasm (34–37). This finding suggests
99 that the NPC-associated quality control complex also has a role in nuclear retention of intron-containing
100 mRNAs, probably as a back-up mechanism not influencing the initial splicing events. Here we identified

101 cellular proteins involved in nuclear trapping of intron-retaining HIV-1 mRNAs by performing a genome-
102 wide screen for cellular factors, whose depletion resulted in enhanced expression from HIV-1 unspliced
103 transcripts in the absence of its export factor Rev. Furthermore, we examined the effect of identified
104 cellular factors on HIV US and FS RNAs localization and expression. This revealed that CRNKL1 is a
105 nuclear retention factor of the HIV-1 unspliced RNA and a highly selective regulator of cytoplasmic levels
106 of intron-retaining HIV-1 and cellular mRNAs.

107 **Results**

108 **Reporter cell lines for HIV-1 Rev-independent expression from the unspliced RNA**

109 The overall strategy to identify cellular proteins involved in the nuclear trapping of intron-retaining HIV-
110 1 mRNAs was to inactivate cellular genes by a genome-wide lentiviral CRISPR-Cas knock-out library and
111 to select knock-out cells that show enhanced Gag-expression from a proviral Rev-deficient HIV-1
112 reporter construct. To easily quantify HIV-Gag expression in living cells, its ORF was fused to the blue-
113 fluorescent protein (BFP) gene. The *rev* gene was inactivated by two point mutations and the DsRed
114 reporter gene was expressed in place of *nef* from a fully-spliced HIV-1 transcript in order to mark HIV
115 infected cells in the absence of Rev (Fig. 1A). Since we aimed to establish stable cell clones containing
116 such proviral reporter constructs, expression of *pol, vif, vpr, and vpu* was also blocked by point
117 mutations to avoid difficulties due to potential cytotoxic or cytostatic effects of these viral proteins (38–
118 40). The *env* gene was inactivated by a 4-bp deletion resulting in a frame shift. Although this reporter
119 construct, designated HIV-dual-GT only encodes the Gag-BPF fusion protein, DsRed, and Tat (for
120 sufficient transcriptional activation), all known *cis*-acting sequences potentially interacting with cellular
121 components should be maintained.

122 To characterize the expression pattern of HIV-dual-GT, HEK293T cells were transfected with plasmids
123 containing HIV-dual-GT or a *rev*-deficient *env*-deletion mutant of HIV in the presence or absence of a
124 Rev-expression plasmid. Western blot analyses revealed Rev-dependent expression of the Gag-BFP
125 fusion protein (Fig. 1B). Flow cytometric analysis of HEK293T cells transfected with HIV-dual-GT revealed
126 a major DsRed positive, but BFP negative cell population, that shifted to a double positive population by
127 co-transfection of a Rev-expression plasmid (Fig. 1C). DsRed expression seemed to decrease in the
128 presence of Rev consistent with reduced splicing and consequently reduced generation of fully-spliced
129 DsRed mRNA due to rapid export of unspliced transcripts (7).

130 To generate reporter cell lines harboring a properly integrated proviral DNA of HIV-dual-GT, it was
131 encapsidated and pseudotyped by cotransfection with HIV-gag-pol and VSV-G expression plasmids. The
132 Rev deficiency was also complemented by cotransfection of a Rev-expression plasmid. Titers of the HIV-
133 dual-GT vector were in the range of $5x10^5$ TU (transforming units)/ml. Flow cytometric analysis of Jurkat
134 cells transduced with this VSV-G pseudotyped HIV-dual-GT vector particles revealed a distinct
135 population of cells that was DsRed positive, but BFP negative (Fig. 1D). Transient transfection of this
136 transduced bulk culture with a Rev-expression plasmid shifted approximately 10% of the DsRed positive

137 population to a DsRed and BFP double-positive population. The magnitude of this effect was lower than
138 in the co-transfection experiments in 293T cells (Fig. 1C) consistent with the lower transfection
139 efficiency for Jurkat cells.

140 To obtain more homogenous reporter cells, Jurkat cell clones containing HIV-dual-GT were established
141 by transducing Jurkat cells at an MOI of 0.1, followed by flow cytometric sorting of single DsRed positive,
142 BFP negative cells into wells of a 96-well plate (Fig. 1E). Two of the expanded DsRed positive cell clones,
143 showing a rather homogenous population of DsRed positive and BPF-negative cells and designated J-
144 dual#3 and J-dual#6, were transduced at an MOI of 2 with VSV-G peudotyped *env* deletion mutants of
145 HIV containing or lacking a functional *rev* gene. In the presence of a functional *rev* gene the BFP
146 fluorescence was strongly induced in both reporter cell clones (Fig. 1F). Therefore, both clones were
147 selected for the screening for cellular genes that suppress BFP expression from the HIV-dual-GT reporter
148 virus.

149

150 **Genome-wide screening for cellular factors suppressing expression from intron-containing HIV-1**
151 **genomic RNA**

152 Poor expression levels of HIV-1 structural genes in the absence of Rev that can be overcome by
153 modifying the codon-usage, suggest that there is an active cellular suppressive mechanism acting on the
154 level of the viral transcript. To identify cellular proteins necessary for this postulated suppressive
155 mechanism, we used a GeCKOv2 genome-wide CRISPR-Cas knock-out library kindly provided by the
156 laboratory of Feng Zhang (41, 42). The plasmid library contains 122,411 different single guide RNA genes
157 (sgRNA) targeting 19,050 human genes (6 sgRNAs per gene) and 1,864 sgRNA genes targeting miRNA
158 genes (4 sgRNAs per miRNA gene), and includes 1,000 non-targeting (NT) sgRNA genes. Cas9, sgRNA and
159 puromycin-resistance genes are expressed from a single lentiviral vector construct, lentiCRISPRv2. After
160 amplification of the plasmid library and confirmation of its diversity by next generation sequencing (data
161 not shown), a stock of a VSV-G pseudotyped lentiviral vector library was prepared by transient
162 transfection of 293T cells and titrated on the Jurkat reporter cell clones, J-dual#3 and J-dual#6, and
163 parental Jurkat cells. All three cell lines were then transduced with the lentiviral CRISPR-library at an
164 MOI of around 2. Four days after transduction, the 0.01 to 0.02% DsRed positive cells showing the
165 highest BFP expression levels were sorted for next generation sequencing of the sgRNA genes delivered
166 to these cells by the lentiviral vector library. To be able to determine enrichment of sgRNA genes

167 enhancing Gag-BFP expression levels, the representation of each sgRNA gene from non-selected Jurkat
168 cells transduced with the lentiviral CRISPR-library was also determined. Selective enrichment of sgRNA
169 sequences in sorted J-dual#3 and J-dual#6 cells compared to non-sorted Jurkat cells was then
170 determined by MAGeCK, a computational tool developed for the analysis of CRISPR screens (43, 44). The
171 calculated robust ranking aggregation (RRA) score reflects enrichment of sgRNA sequences targeting the
172 same gene in the selected cells. The distribution of the RRA score for all targeted genes is plotted in Fig.
173 2C with the top 12 candidate genes highlighted. For eleven of these, at least two different sgRNA
174 sequences targeting the same gene were enriched in the selected cells arguing against confounding off-
175 target effects of single sgRNA (Fig. 2D). The probability of false discovery of the top 12 candidate genes
176 ranged from 0.0012 to 0.15.

177

178 **Analyzing interaction of validated hits shows a network of pre-mRNA splicing factors**

179 Based on the enrichment of the individual sgRNAs targeting each of the top 12 candidate genes in the
180 Gag-BFP positive cells selected by flow cytometry (Fig. S1), two sgRNA sequences were chosen for each
181 of the candidate genes (Table S1A) for further validation. Gene-specific sgRNA sequences and a non-
182 targeting sequence, named NT1, were cloned individually into the lentiCRISPRv2 vector plasmid. J-
183 dual#3 and J-dual#6 reporter cells were then transduced with lentiviral vector particles transferring
184 these sgRNA sequences. Upregulation of BFP expression was observed by flow cytometry for 10 of the
185 12 candidate genes. For nine of them, transfer of both sgRNA sequences upregulated BFP expression in
186 both reporter cell clones (Fig. S2A). For DUS2 and ZNF221 the screening results could not be confirmed.
187 Since long half-lifes of mRNAs and proteins may prolong the time from inactivation of the targeted gene
188 to the decline of protein levels, we continued to monitor the reporter cells transduced with sgRNA
189 genes targeting DUS2 and ZNF221 for up to 2 weeks post infection. However, upregulation of Gag-BFP
190 remained undetectable (data not shown). In addition, we confirmed sufficient transduction rates by the
191 lentiviral DUS2 and ZNF221 targeting vectors, excluding the possibility that reduced lentiviral vector
192 titers are responsible for their failure to upregulate BFP expression (Fig. S2B). Despite this, the entire
193 screening approach was highly specific showing a false discovery rate of only 0.166.

194 A first hint on potential mechanisms by which inactivation of the identified cellular genes could enhance
195 Gag-BFP expression in the reporter cell lines was obtained by searching for functional protein
196 association networks between the identified genes using the STRING online database (https://string-

197 db.org/). Strikingly, 9 of the 10 confirmed hits are involved in pre-mRNA splicing (Fig. 3). Among them is,
198 ESS2/DGCR14 for which no interactions with the other identified factors have been reported, but which
199 is associated with U6 snRNA as well as U1 and U4 snRNAs (45). WDR70 is not linked to splicing, however,
200 it contains a WD40 repeat domain which is shared by a number of splicing factors (46). The pre-mRNA
201 splicing process includes two excision-steps catalyzed by a large and dynamic RNP machine. The
202 confirmed hits DHX38, DHX8, SLU7 and DGCR14 are typical second-step factors. In contrast, XAB2, ISY1,
203 CRNKL1, CWC22 and BUD31 should be recruited at the first step, and intriguingly all belong to the
204 nuclear PPR19-associated complex (47–50). This complex is involved in diverse nuclear functions such as
205 DNA double strand break (DSB) repair, transcriptional elongation, and splicing (51). The core factor of
206 the complex, PRPF19 (human Prp19), however, was not identified in the screen. Consistently, targeting
207 PRPF19 by two sgRNAs in J-dual#3 did not affect Gag-BFP expression (Fig. S3A). Since most of the
208 confirmed hits are components of the spliceosome, we analyzed whether PRPF8, a ATPase/helicase at
209 the core of the spliceosome (52), could also regulate the HIV Gag expression. Moreover, BUD13 from
210 the RES complex and Tpr from the NPC, which were reported to have a nuclear retention function, were
211 tested as well by using two different CRISPR/Cas sgRNAs targeting each of the genes. None of these
212 targeting sgRNAs tested induced BFP expression (Fig. S3B-D).

213

214 **CRINKL1 and other cellular factors impact on HIV-1 unspliced and fully-spliced mRNA levels**

215 To follow-up on the reduction of splicing as the potential mechanism of enhanced Gag-BFP expression,
216 the top 5 hits and WDR70 were analyzed for their effect on HIV-1 mRNA splicing. Attempts to generate
217 J-dual#3 cells with a stable knock-out of the six selected target genes were not successful, presumably
218 due to the fact that the selected genes are essential for cell viability (53). We therefore analyzed HIV
219 RNA splicing and nuclear export after knock-down of these genes in short-term cultures. J-dual#3 cells
220 were transduced with the lentiviral vectors encoding the different guide RNA genes and Gag-BFP
221 positive cells were sorted to enrich for cells with the desired gene knock-down. Total RNA was then
222 extracted from the sorted cells and the copy numbers of unspliced HIV RNA per ng extracted total RNA
223 were determined and compared to the copy numbers of unspliced RNA in J-dual#3 cells transduced with
224 a negative control vector (NT1). Consistent with the stronger BFP expression, unspliced HIV RNA copy
225 numbers were also enhanced 3- to 9-fold (Fig. 4A). Fully-spliced HIV-1 RNAs were also quantified,
226 revealing a marginal decrease in cells transduced with XAB2 and WDR70 targeting sgRNAs, but not with
227 the other gene targeting sgRNA (Fig. 4A). Differential effects of the gene targeting on HIV-1 FS and HIV-1

228 US RNA also lead to decreased ratios of FS over US transcripts (Fig. 4A). For comparative reasons we also
229 transduced J-dual#3 cells in an independent experiment with a lentiviral vector expressing Rev, sorted
230 Gag-BFP positive cells and quantified HIV-1 US and FS transcripts. Rev enhanced the US HIV-1 RNA levels
231 approximately 5-fold, which coincided with a minor reduction in FS transcripts (Fig. 4A).

232 The splicing process is closely linked to nuclear mRNA trafficking (54, 55). Therefore, we also determined
233 the effect of targeting the six selected candidate genes on the cytoplasmic levels of HIV-1 US and FS RNA
234 (Fig. 4B). Proper separation of the cytoplasm from nuclear components was controlled by examining
235 GAPDH pre-mRNA levels and a nuclear marker protein in the cytoplasmic fraction from Jurkat cells (56,
236 57) . GAPDH pre-mRNA levels per ng total RNA extracted from the cytoplasm were approximately 50-
237 fold lower than GAPDH pre-mRNA levels per ng total RNA extracted from whole cells. Consistently, the
238 nuclear LaminB protein could not be detected in the cytoplasmic fractions (Fig. S4).

239 After transduction with the different targeting vectors, HIV-1 US RNA in the cytoplasm of Gag-BFP
240 positive cells was enhanced approximately 22 to 116-fold, while the cytoplasmic FS RNA levels were
241 affected only 0.6 to 2.7-fold. The magnitude of the enhancement of cytoplasmic levels of HIV-1 US RNA
242 exceeded the enhancement of HIV-1 US RNA in whole cell extracts, indicating that knocking-down the
243 selected target genes not only reduced splicing but also enhanced nuclear export of the HIV-1 US RNA.
244 Strong enhancement of cytoplasmic HIV-1-US RNA export was also observed in Rev-expressing cells.
245 However, in contrast to the cells transduced with the targeting vectors, cytoplasmic FS RNA levels in
246 Rev-transduced cells were significantly reduced, consistent with the previously reported suppression of
247 the TAP/NXF1 mediated RNA export by Rev (58).

248 To dissect the relative contributions of inhibition of splicing and enhancement of nuclear export of HIV-1
249 US RNA, we compared the ratio of HIV-1 FS RNA to HIV-1 US RNA in whole cell extracts as a measure of
250 splicing efficiency and the ratio of HIV-1 US RNA in cytoplasmic extracts to the HIV-1 US RNA in whole
251 cell extracts as a measure of nuclear export efficiency. This revealed two different patterns of responses.
252 The fold-reduction of the splicing efficiencies in cells transduced with ISY1, XAB2 and WDR70 targeting
253 vectors mirrored the fold-enhancement of the respective nuclear export efficiencies. Expression of Rev
254 induced a similar pattern of response. In contrast, transduction with DHX38, BUD31 or CRNKL1 targeting
255 vectors predominantly enhanced the nuclear export efficiency with at least three-fold lower effects on
256 splicing efficiency (Fig. 4C). In particular, knock-down of CRNKL1 enhanced the ratio of HIV-1 US RNA in
257 cytoplasmic extracts to HIV-1 US RNA in total cell extracts more than 34-fold, while the ratio of HIV-1 FS
258 RNA to HIV-1 US RNA in whole cell extracts was only decreased 4-fold. Targeting of CRNKL1 also led to

259 the strongest enhancement of cytoplasmic levels of HIV-1 US RNAs without reducing cytoplasmic HIV-1
260 FS RNA levels (Fig. 4C), indicating that CRNKL1 is a major nuclear retention factor of the genomic HIV-1
261 US RNA.

262 **CRNKL1 knock-down shifts cytoplasmic levels HIV-1 splice variants**

263 To confirm the role of CRNKL1 in the regulation of cytoplasmic HIV RNA levels and to explore potential
264 effects of CRNKL1 on cytoplasmic levels of cellular RNAs a transcriptomic analysis was performed after a
265 short-term knock-down of CRNKL1. J-dual#3 reporter cells were again transduced with the CRNKL1
266 targeting vector and Gag-BFP-positive cells were sorted by flow cytometry. RNA was extracted from the
267 cytoplasmic fraction of sorted cells and J-dual#3 cells transduced with a control vector. The extracted
268 RNAs from 4 biological replicates each were then enriched for mRNAs and sequenced by the Illumina
269 High-seq procedure. In addition to this, cytoplasmic mRNA was extracted from J-dual#3 cells expressing
270 Gag-BFP after transduction with the Rev-encoding lentiviral vector and sequenced.

271 To confirm the effectiveness of Cas9/sgRNA-mediated knock-down by transcriptomic analysis, a core
272 target region in exon 3 of the CRNKL1 gene was defined as nucleotides -3 to +3 relative to the SpCas9
273 cleavage site. The number of RNA-seq reads that contained 10 nucleotides upstream and downstream
274 of the core target region were then determined (Fig. S5). While 55 reads on average mapped to the
275 region flanking the core target region in the control cells, only a mean of 5.8 reads were detected in cells
276 transduced with the CRNKL1 knock-down vector (Fig. S5). Larger deletions and substitutions outside the
277 core target region probably reduce the number of reads that can be mapped to the flanking region.
278 Comparison of the reads from CRNKL1 knock-down cells that can be mapped to the flanking region with
279 the wild type CRNKL1 core target region revealed insertions, deletions and point mutations in 22 out of
280 23 reads. In control cells, only 2 out of 219 reads contained single point mutations clearly indicating
281 efficient inhibition of CRNKL1 expression after Cas9/sgRNA targeting (Fig. S5).

282 HIV reads were mapped to the proviral DNA of HIV-dual-GT revealing more reads mapping to intronic
283 regions after CRNKL1 knock-down and Rev expression than in the controls (Fig. 5A). The percentage of
284 reads mapping to HIV among all the reads uniquely mapping to either the human or the viral genome
285 (total uniquely mapped reads) was enhanced two fold in CRNKL1 knock-down cells (Fig. 5B). This was
286 not due to an enhanced transcriptional activity since the percentage of reads mapping to Exon 1, which
287 is shared by all HIV transcripts, was only affected marginally(Fig. 5C). Changes in HIV-1 US RNA were
288 assessed by determining the percentage of reads mapping to the first intron of HIV-1 between SD1–SA1

289 (5D) or mapping to the unspliced SD1 site (5E) revealing 4 and 12-fold increases under CRNKL1 knock-
290 down conditions, respectively. The percentage of reads mapping to the unspliced SD4 site and therefore
291 containing the *env* intron also increased by a factor of 9 (Fig. 5F). The change in the percentage of all
292 spliced HIV reads was negligible (Fig. 5G) indicating that knock-down of CRNKL1 does not reduce splicing
293 of HIV RNAs in general.

294 To further explore differences in cytoplasmic levels of differentially spliced HIV-1 transcripts the precise
295 splice site usage was also systematically analyzed by calculating the percentage of reads covering
296 specific SD-SA pairs among all spliced HIV reads (Fig. S6A,B). This revealed that knock-down of CRNKL1
297 reduced cytoplasmic HIV transcripts spliced from SD1 to SA1 approximately two-fold, while transcripts
298 spliced from SD1 to SA3 were enhanced three-fold (Fig. 5H). To further assess the effect of the CRNKL1
299 knock-down on the major Env encoding transcript, the percentage of reads spliced to SA5 was
300 determined (Fig. 5I). Independent of the SD sites, usage of SA5 trended to be reduced, which is similar
301 to an effect recently reported for the splicing factor DDX17 (59). Moreover, the level of fully spliced
302 transcripts seemed unaffected by the CRNKL1 knock-down since the percentages of reads spliced either
303 from SD1 to any splice acceptor site or from SD4 to SA7 were nearly the same (Fig. 5J).

304

305 **CRNKL1 is a regulator of cytoplasmic cellular mRNA levels**

306 The influence of the CRNKL1 knock-down on cytoplasmic mRNA levels was also determined (Fig. S7A,B).
307 To display the results the log2-transformed expression differences were plotted against the expression
308 level for each gene in a MA-plot (Fig. 6A). To reduce noise and arbitrarily large fold changes from low
309 expressed genes the log2 fold changes are shrunk putting less weight on low expressing genes.
310 Significant differences after correction for multiple testing revealed the upregulation of 1249 and
311 downregulation of 2569 mRNAs, respectively. Two hundred and ten mRNAs even showed a more than
312 4-fold (log2 fold change >2) increase after knock down of CRNKL1. Interestingly, upregulated mRNAs
313 were enriched for the functional categories involved in transcription (Fig. 6B) while downregulated
314 mRNAs were more frequently associated with functions in cell cycle and organelle organization (Fig. 6B).

315 Given the close link between RNA export and splicing, a more general role of CRNKL1 on cytoplasmic
316 levels of splice variants was analyzed upon CRNKL1 knock-down. Individual splicing events (including
317 skipped exons, alternative splice sites and retained introns) were quantified using the percent spliced-in
318 metric (PSI, Schafer et al., 2015) and differential analysis was carried out by SUPPA (61). Of all quantified

319 splicing events in cytoplasmic mRNAs upon CRNKL1 depletion, 2.4 % were upregulated and 1.6%
320 downregulated (Fig. S7C,D). Introns were more frequently retained upon CRNKL1 depletion compared to
321 wild-type cells (Fig. 6C). Similarly, alternative 3' and 5' splice sites were more frequently detected upon
322 CRNKL1 knock-down, suggesting that CRNKL1 inhibits usage of alternative splice sites (Fig. 6C).

323 A total of 295 introns were more frequently retained in cytoplasmic mRNAs after CRNKL1 knock-down.
324 Further analyses revealed that not all introns of the same transcript were affected in the same way (Fig.
325 S8), but rather that intron retention was highly specific for one or a few introns of the same gene.
326 Enhanced cytoplasmic levels of intron-retaining mRNAs could be confirmed by real-time PCR with
327 primers spanning the respective exon – intron junctions (Fig. S8).

328 Genes containing introns or alternative splice sites that were upregulated after CRNKL1 knock-down
329 were not enriched for particular functional categories (data not shown). Enhanced detection of introns
330 or alternative splice sites in cytoplasmic mRNAs after CRNKL1 knock-down indicates that CRNKL1
331 suppresses either these splicing events or the nuclear export of the respective splice variants. To gain
332 insight into the principles of regulation of these CRNKL1 knock-down dependent introns and alternative
333 splice sites, we analyzed common features of these splicing events, including splice site strength, branch
334 point and polypyrimidine tract score, GC content and intron length. CRNKL1 knock-down-dependent
335 alternative 3' splice sites showed significantly lower branch point scores than constitutive 3' splice sites
336 (Fig. 6D), but a similar length of polypyrimidine tracts (Fig. S7E). We also found that CRNKL1-dependent
337 introns were significantly longer than CRNKL1-independent introns within the same transcript (Fig. 6E)
338 and trended to have longer polypyrimidine tracts (Fig. S7F) but showed no difference in GC content (Fig.
339 S7G).

340

341 **Discussion**

342 Using a genome-wide screen for cellular factors repressing expression of an HIV-1 structural protein
343 from the unspliced RNA, we identified ten target proteins linked to mRNA metabolism, with nine of
344 them associated with the spliceosome. Interestingly, five of them (ISY1, BUD31, XAB2, CRNKL1 and
345 CWC22) are found in the Prp19-associated complex. This complex is important for stabilizing the
346 spliceosome, which is also involved in DNA repair, transcriptional elongation, and RNA export (51). Given
347 that the yeast homolog of WDR70 was reported to interact with Prp19 (62) and is now shown to affect
348 RNA splicing, it could be a new member of the Prp19-associated complex. Due to the close linkage of

349 splicing with RNA stability and nuclear RNA export, it is difficult to disect the precise mechanism by
350 which deletion of the identified nuclear proteins enhance cytoplasmic RNA levels of HIV-1-US RNA.

351 Theoretically, enhanced transcription, reduced nuclear or cytoplasmic degradation, decreased splicing,
352 and/or enhanced nuclear export can increase cytoplasmic levels of HIV-1 US RNA ($US_{cyt.}$) (Fig. 7).
353 Enhanced transcription and decreased nuclear degradation seem unlikely since the percentage of reads
354 mapping to the first HIV-1 Exon present on all viral transcripts was not significantly enhanced (Fig. 5B).
355 The absence of enhanced cytoplasmic levels of FS RNAs and DsRed expression levels from the dual
356 reporter construct (Fig. 4B, Fig. S2A) also argues against such a mechanism. Assuming a single pool of
357 nuclear HIV-1 US RNA that could follow either of the two remaining pathways, differential effects on the
358 nuclear export efficiency ($US_{cyt.}/US_{whole}$) and splicing efficacy (FS_{whole}/US_{whole}) are expected (Fig. 7).
359 Enhancing nuclear export of HIV-1 US RNA should strongly enhance the ratio of $US_{cyt.}/US_{whole}$. Since rapid
360 export may reduce splicing and degradation, the ratio of FS_{whole}/US_{whole} may responsively decrease,
361 although to a lower extent. Reduced splicing should primarily decrease the ratio of FS_{whole}/US_{whole}, but
362 not the ratio of $US_{cyt.}/US_{whole}$ (Fig. 7). Similar to an enhanced nuclear export, reduced degradation of
363 cytoplasmic HIV-1 US RNA should also lead to enhanced $US_{cyt.}/US_{whole}$ and decreased FS_{whole}/US_{whole} ratios
364 (Fig. 7), but the association of the identified proteins with the spliceosome localized in the nucleus
365 argues against such a cytoplasmic effector mechanism.

366 Clearly, ablation of all 6 target genes tested reduced the splicing efficacy (FS_{whole}/US_{whole}) indicating that
367 the targeted genes are enrolled in splicing and/or nuclear retention. The relative contribution of
368 reduced splicing versus enhanced nuclear export of the HIV-1 US RNA differed between the different
369 target genes. The XAB2 knock-down decreased the splicing efficacy 12.5-fold and the nuclear export
370 efficacy of the HIV-1 US RNA 11.7-fold. Since splicing and nuclear export efficacy were affected to a
371 similar degree, it is not possible to conclude whether XAB2 predominantly acts as a splicing or nuclear
372 retention factor. This is consistent with previous observations that depleting splicing factors (e.g. SF1,
373 U2AF65 and Prp19) can cause pre-mRNA leakage into the cytoplasm (30, 63–65). In contrast to the
374 effect of the XAB2 knock-down, the splicing efficacy after knock-down of CRNKL1 is only decreased 4.0-
375 fold, while the nuclear export efficiency based of the HIV-1 US RNA increases 34.6-fold. Since the ratio of
376 HIV-1 FS to US transcripts in the cytoplasmic RNA is also reduced to a larger extent than in the whole cell
377 RNA, enhanced nuclear export of the HIV US RNA after inactivation of the target gene seems to play a
378 dominant role. This indicates that CRNKL1 is primarily required for retention of the HIV-1 US RNA in the
379 nucleus and not for its splicing. The precise molecular mechanism by which the CRNKL1 knock-down

380 leads to a more than 100-fold enhancement of cytoplasmic HIV US RNA levels remains to be clarified.
381 CRNKL1 has been previously reported to regulate splicing and assist the export of cellular intron-less
382 mRNAs in cooperation with the TREX (TRanscription-EXport) complex, U2AF65 and other NTC members
383 (66–68). Our observation that targeting of CRNKL1 did not enhance cytoplasmic levels of HIV-1 FS RNA,
384 which uses the default TAP/NXF1 export pathway, indicates that CRNKL1 is not a global nuclear mRNA
385 retention factor. Our transcriptomic analysis also indicates a highly selective differential regulation of
386 cytoplasmic levels of HIV-1 splice variants. The only spliced HIV-1 transcripts that were clearly
387 upregulated by CRNKL1 knock-down have spliced SD1 to the SA3 site (Fig. 5G) and may encode Tat.
388 Whether this represents an upregulation of *env* intron retaining transcripts or the tat-encoding fully-
389 spliced transcript is unclear. The enhanced percentage of cytoplasmic HIV transcripts harbouring an
390 unspliced SD4 site (Fig. 5F) and the unchanged percentage of transcripts harbouring a spliced SD4 site
391 (Fig. 5J) argue for the former.

392 The transcriptomic analyses after CRNKL1 knock-down also revealed a selective regulation of
393 cytoplasmic levels of cellular mRNAs. We found that 11.0% and 20.1% of expressed mRNAs in the
394 cytoplasm were up- and downregulated, respectively. Functionally, the upregulated mRNAs are
395 associated with transcription, while downregulated mRNAs are more frequently associated with the cell
396 cycle and organelle organization suggesting that CRNKL1 participates as a master regulator in these
397 cellular programs. Since altered nuclear export after CRNKL1 depletion could result in differences in the
398 abundance of different cytoplasmic mRNA isoforms, we investigated alternatively spliced cytoplasmic
399 isoforms upon CRNKL1 knock-down. Upon CRNKL1 depletion a much higher number of included introns
400 and alternative 3' and 5' splice sites than the excluded ones were detected, suggesting that CRNKL1
401 promotes splicing of these events or highly selective nuclear retention. Intron retention in the absence
402 of CRNKL1 was highly restricted and even varied among introns of the same gene (Fig. S8). Searching for
403 common properties of introns that were retained revealed that their mean length was significantly
404 greater compared to introns with unchanged inclusion upon CRNKL1 depletion. In addition, usage of 3'
405 splice sites with lower branch point scores seems to be blocked in the presence of CRNKL1. Therefore,
406 the sequence features of CRNKL1-dependent splicing events seem to reside at the 3' rather than at the
407 5' end of the intron.

408 Mechanistically, one plausible hypothesis based on the nuclear retention observed for the HIV-1
409 unspliced RNA is that a nuclear RNP complex binds in a CRNKL1 dependent manner to RNA motifs
410 present in a highly selective subset of cellular and viral introns leading to their nuclear retention and

411 splicing. CRNKL1 dependent binding to binding motives on cellular transcripts leading to nuclear
412 retention and degradation may also explain the increase in cytoplasmic levels of cellular mRNAs upon
413 CRNKL1 knock-down.

414 In summary, our study identifies three spliceosomal proteins that are required for nuclear retention of
415 the HIV-1 unspliced RNA and indicates that CRNKL1-dependent nuclear retention is a novel mechanism
416 for the regulation of cytoplasmic levels of intron-retaining cellular mRNAs that is hijacked by HIV-1.

417

418 **Materials and Methods**

419 **Plasmids** Plasmids expressing viral proteins, HIV-1 Rev (pcRev), HIV-1 Tat (pcTat), HIV-1 codon-
420 optimized Gag-Pol (Hgp^{syn}) and VSV-G (pHiT-G) were described previously (23, 56). Lentiviral packaging
421 plasmid pPAX2 and pMD2.G were obtained from Addgene (#12260 and #12259). HIV-1 proviral plasmid
422 HIVΔe is based on the HIV-1 molecular clone NL4-3 and contains a 4-nucleotide deletion at its NheI site
423 leading to inactivation of *env*. The plasmid HIVΔr-e contains an additional mutation of the start codon of
424 *rev* and a premature stop codon in the first *rev* exon (57).

425 Proviral reporter plasmid HIV-dual-GT was constructed based on HIVΔr-e. A cassette containing
426 mTagBFP2, a PEST coding sequence (destabilized BFP) and two stop codons were ligated in frame to the
427 3' end of *gag*, forming a new ORF for Gag-BFP. The Slippery Sequence TTTTTTA (nt 2085 to nt 2091 of
428 the GenBank entry AF324493.2) was mutated to CTTCCTG to prevent Gag-Pol translation (69). The
429 coding sequence of Nef close to its 5´end (nt 8796 to nt 8880 of the GenBank entry AF324493.2) was
430 replaced by a cassette consisting of a 3 x Ala linker sequence and the DsRed max coding region followed
431 by two stop codons. The sequence (nt 5131 to nt 6297 of the GenBank entry AF324493.2) ranging from
432 *vif* to *vpu* was replaced by the corresponding region of NL4-3-Δ6-drEGFP (70) harboring inactivating
433 point mutations in *vif*, *vpr* and *vpu*. The *rev* ORF was inactivated as described above for HIVΔr-e by PCR
434 mutagenesis. The plasmid HIV-dual-GT including the final sequence is available from AddGene (plasmid
435 ID 122696). Individual CRISPR knock-out plasmids were constructed based on lentiCRISPRv2 (#52961,
436 Addgene) and GeCKOv2 sgRNA target sequences. All sgRNA sequences of the GeCKO v2 library can be
437 found at www.addgene.org/pooled-library/zhang-human-gecko-v2/. The different sgRNA sequences
438 selected for validation are provided in Supplementary Table S1A. Each of the sgRNA sequences were
439 cloned into the vector following the instructions on the Addgene webpage (www.addgene.org/52961/).

440 **GeCKO v2 library** The pooled Human GeCKO v2 CRISPR knockout plasmid library was a gift from Feng
441 Zhang (41) and obtained from Addgene (#1000000048 and #1000000049). It is supplied as two sub-
442 libraries, each containing 3 different sgRNA constructs targeting each gene in the genome. Amplification
443 of the lentiviral vector plasmid library was performed essentially as described by the contributors'
444 protocol (41). The complexity of the library was confirmed by HiSeq NGS of the pooled plasmid library
445 (data not shown).

446 **Cell culture** HEK293T (DSMZ) and TZM-bl (71) (NIH AIDS reagent program) cell lines were maintained in
447 DMEM medium supplemented with 10% FCS and 1% penicillin/streptomycin. Jurkat, Clone E61 (ATCC)
448 and reporter cell lines J-dual#3 and J-dual#6 were maintained in RPMI 1640 supplemented with 10% FCS,
449 1% penicillin/streptomycin. After expansion, J-dual#3 and J-dual#6 stocks were stored frozen in aliquots
450 at -80℃. A fresh aliquot was thawed and expanded in each assay.

451 **Cell transfection** HEK293T cells were routinely transfected by the polyethylenimine (PEI)-precipitation
452 method (56). In general, 3.5 million cells per T25-flask (CELLSTAR) were seeded 24 h before transfection.
453 For each transfection, plasmid DNA was mixed with calf thymus carrier DNA (Thermo Fisher Scientific) to
454 give a total of 10 µg DNA. The 10 µg DNA were mixed with 15 µl PEI (1µg/µl, Sigma Aldrich) and added
455 to the cells. After 8 h of incubation with the cells, the transfection medium was exchanged with fresh
456 culture medium. Supernatants or cells were harvested 48 h post transfection. To produce pseudotyped
457 HIVΔe, HIVΔr-e and HIV-dual-GT vector particles, 3 µg proviral plasmid, 3 µg Hgp^{syn}, 2 µg pHiT-G, 1 µg
458 pcTat and 1 µg pcRev were used. To produce individual CRISPR-KO lentiviral vector particles, 5 µg
459 lentiCRISPRv2 plasmid, 3.75 µg pPAX2 and 1.25µg pMD2.G were used. To produce the pooled human
460 GeCKO v2 lentiviral vector library, 24 millions of 293T cells were seeded into a T175 tissue culture flask.
461 24 h later, 20 µg GeCKO v2 library plasmids, 15 µg pPAX2, 5 µg pMD2.G, and 120 µl GenJet transfection
462 reagent (SignaGen) were used for transfection. HIV-dual-GT-transduced Jurkat cells were transfected by
463 the electroporation method described elsewhere (72). In brief, 5.0 millions of stably transduced cells
464 were electroporated with 50 µg plasmids by the Gene Pulser X cell™ Electroporation System (Bio-Rad) at
465 250 V and 1500 microfarads. 48 h post transfection, cells were harvested for flow cytometry analysis.
466 The transfection efficiency after expression of a GFP reporter plasmid was estimated to be between 10%
467 - 20%.

468 **Western blot analyses** Immunoblotting for HIV-1 Gag/CA p24 was described previously (57).
469 Immunoblotting for Rev was performed with primary antibody Sheep Anti-Rev (1:4,500, #H6006, US

470 Biological Life Sciences) and secondary antibody Rabbit Anti-Goat Ig/HRP (1:5,000, #P0160, Dako).
471 Immunoblotting for α-Tubulin was performed using polyclonal Rabbit Anti-α-Tubulin antiserum as
472 primary antibody (1:4,000, #600-401-880, Rockland), and Swine Anti-Rabbit Ig/HRP (1:5,000, #P0217,
473 Dako) as secondary antibody. In the experiment of Figure 1B, anti-α-Tubulin blotting was conducted
474 after stripping the blotted anti-CA membrane by stripping buffer (#2504, Millipore). The membrane was
475 then re-blocked, and subsequently re-blotted with antibodies as described above. Immunoblotting for
476 LaminB was performed with primary antibody Goat anti-LaminB (C-20) (1:500, #sc-6216, Santa Cruz
477 Biotechnology) and secondary antibody Rabbit anti-goat Ig HRP (1:5,000, #P0160, Dako).

478 **Lentiviral vector preparation, infection and titration** Lentiviral vector particles were produced by
479 transfection of 293T cells with the plasmids mentioned above. HIV-1 proviruses were packaged with a
480 3rd generation packaging system (Hgp^{syn}, pcRev, pcTat and pHiT-G) (23, 73, 74). CRISPR-KO constructs
481 and a pool of the two GeCKO v2 plasmid A and B sublibraries were packaged with a 2nd generation
482 packaging system (pPAX2 and pMD2.G) (42, 75) for higher infectious titers. 48 h after transfection, the
483 supernatants from transfected 293T cells were centrifuged at 1,500 rpm for 5 min to remove cell debris,
484 and afterwards filtered through 0.45 µm filters. Viral stocks were aliquoted and stored under -80℃.
485 Two batches of the GeCKO v2 lentiviral vector library were prepared. HiSeq NGS analyses of sgRNA
486 representation in cells transduced with both batches showed an almost 100% identification of the
487 sgRNAs encoded by the two GeCKO v2 plasmid sublibraries.

488 Jurkat, J-dual#3 and J-dual#6 cells were infected with the lentiviral vectors or the lentiviral vector library
489 by spinoculation. 1.5 million of cells were centrifuged with 1 ml vector preparation, in the presence of 8
490 µg/ml polybrene (Sigma-Aldrich), at 2,000 rpm, for 3 h, at 33℃. After spinoculation, the supernatants
491 were removed and the cells were continued to be cultured.

492 VSV-G-pseudotyped HIVΔe, HIVΔr-e and HIV-dual-GT vector particles were titrated on TZM-bl cells as
493 previously described (57). The titers were $5.8x10^6$ and $5.1x10^6$ TU/ml for two independent titrations of
494 HIVΔe, $4.3x10^6$ and $5.7x10^6$ TU/ml for two independent titrations of HIVΔr-e and $6.4x10^5$ and $5.0x10^5$
495 TU/ml for two independent titrations of HIV-dual-GT. VSV-G-pseudotyped HIV-dual-GT vector particles
496 were also titrated on Jurkat cells. The lentiviral reporter vector was serially diluted to 1:1, 1:2, 1:4, 1:8
497 and 1:16. The infection was done by spinoculation method. At 48 hpi, cells were analyzed by flow
498 cytometry to determine the percentage of DsRed positive cells in each dilution. Vector titers on Jurkat
499 cells were calculated as follows for all dilutions: (initial cell number) X (percentage of $DsRed^+$ cells) X

500 (dilution factor). The highest value resulting from these dilutions was taken as vector titer on Jurkat cells.
501 The determined titers for two independent titrations of HIV-dual-GT were $4.9x10^5$ and $6.9x10^5$ TU/ml.
502 The individual CRISPR-KO vectors, as well as the GeCKO v2 lentiviral vector library, were titrated on J-
503 dual reporter cells and parental Jurkat cells. The cells infected by the serially diluted virus were grown
504 for 24 h, afterwards divided into two equal aliquots. One aliquot was then cultured in 2 µg/ml
505 puromycin, the other without puromycin. After 30 h of growth, viable cells were counted by trypan blue
506 staining. The transduction percentage was calculated in the following way: (number of living cells in
507 treated population) x100/ (number of living cells in untreated population). For the GeCKO v2 lentiviral
508 vector library the transduction percentage was multiplied by the number of cells exposed to calculate
509 the titer for the different dilutions of the library. The highest value obtained from the different dilutions
510 of the same batch of the lentiviral vector library was taken as its titer. The titers were $1.3x10^7$ TU/ml for
511 the first batch and $2.2x10^6$ TU/ml for the second batch.

512 **Flow cytometry analysis and Cell Sorting** Flow cytometry analyses were done using the BD™ LSRII flow
513 cytometer (BD Biosciences). FACS data were analyzed by the Flowjo software. Cell sorting was done on a
514 MoFlo XDP (Beckman Coulter) at the FACS core facility of the medical faculty.

515 **Next-Generation Sequencing of sgRNA libraries** Genomic DNA was extracted from the sorted or the
516 control cells either using the Quick-DNA Mini Kit or the Midi Plus Kit (Zymo Research). The integrated
517 LentiCRISPRv2 vector sgRNA sequences were amplified through three rounds of PCR using the Taq DNA
518 Polymerase S (high specificity) (Genaxxon Bioscience). All obtained genomic DNA was used as template
519 in parallel reactions in PCR1, with the cycling conditions 94℃ 2 min, then 18 cycles' 95-55-70℃, 15-20-
520 45 s, final extension 70℃ for 3 min. PCR1 reactions of respective cell samples were then combined and
521 10 µl of the pooled PCR1 product was used as template in PCR2, consisting of 95℃ 2 min, then 15
522 cycles' 95-60-72℃, 15-20-30 s, final extension 72℃ for 3 min. In the indexing PCR3, 10 ng of purified
523 PCR2 product were subjected to 95℃ 2 min, then 8 cycles' 95-59-72℃, 15-20-30 s, final extension 72℃
524 for 3 min. The primer sequences and primer combinations used for each sample at each step are
525 presented in Tables S1B,C. After PCR2 and PCR3, PCR products were purified by the AMPure XP PCR-
526 Cleanup Kit (Beckman Coulter); DNA concentrations were measured by the Qubit Quant-iT dsDNA HS
527 Assay Kit (Thermo Fischer Scientific). Manipulations followed the manufacturers' instructions. DNA
528 libraries were sequenced on Illumina MiSeq (for enriched samples) or Ilumina HiSeq 2500 (for
529 unselected control) instruments. NGS readouts were subjected to MAGeCK analysis (43).

530 **Cell fractionation, RNA isolation and RNA quantification** The cell fractionation protocol has been
531 described previously (76). The plasma membrane was lysed with chilled NP40-Buffer consisting of 10
532 mM Hepes (pH 7.8 adjusted by KOH); 10 mM KCl; 20% Glycerol; 0.25% NP40. DTT was added to a final
533 concentration of 1 mM just before use. After incubation on ice for a maximum of two minutes, the
534 supernatant consisting of the cytoplasmic fractions was collected after centrifugation at 400 g, for 5 min,
535 at 4℃. Cytoplasmic RNA was extracted by the TRIzol (Thermo Fisher Scientific) -Chloroform (Sigma-
536 Aldrich) method (77) or by the Direct-zol RNA miniprep plus kit (Zymo Research). RNA from the whole
537 cell extracts was purified using the RNA-Easy Kit (Qiagen). RNA extracted by TRIzol-Chloroform or the
538 RNA-Easy kit was treated with DNaseI (NEB) to remove DNA contamination. Under the protection of 5
539 mM EDTA, DNaseI was inactivated at 75℃ for 10 min. For the Direct-zol RNA miniprep plus kit purified
540 RNA, DNA contamination was removed by in-column DNA digestion reagents supplied. Total RNA
541 concentration was measured by the Qubit Quant-iT RNA HS Assay Kit (Thermo Fisher Scientific)
542 according to the manufacturer's instructions.

543 Reverse-transcription PCRs (RT-qPCR) for quantification of HIV-1 US, FS RNAs and cellular GAPDH pre-
544 mRNAs were done with the QuantiTect SYBR Green RT-PCR Kit (Qiagen), and performed essentially as
545 described previously (56, 78). Using the applied Biosystems 7500 Real-Time PCR machine cycling
546 conditions were as follows: HIV-1 US RNA: 95℃ 10 sec for denaturation, 65℃ 60 sec for annealing, 72℃
547 30 sec for elongation and fluorescence detection; for HIV-1 FS RNA: 95℃ 10 sec for denaturation, 64℃
548 15 sec for annealing, 72℃ 15 sec for elongation, and 81℃ 30 sec for fluorescence detection.

549 RT-qPCRs for relative quantification of Intron 6-retaining RPL10 mRNAs and Intron 1-retaining C19orf53
550 mRNAs were done with primers illustrated in Table S1D. Cycling conditions for both: 95℃ 10 sec for
551 denaturation, 50℃ 60 sec for annealing, 72℃ 60 sec for elongation and fluorescence detection.

552 **Calculation of the ratios of HIV FS and US RNA levels and fold changes** To measure the effect of
553 candidate genes and Rev on RNA splicing efficacy, we first determined the ratio between FS_{whole}:US_{whole}
554 by dividing the FS copy numbers/ng extracted RNA by the US copy numbers/ng extracted RNA for each
555 sample of each treatment group (knock-down or +Rev) and then calculated the mean of the ratios of
556 replicates of each treatment group. The mean of the ratios of each treatment group were then divided
557 by the mean of the ratios observed in matched control cells transduced with the NT1 or the GFP-
558 expression vector to calculate the fold changes given in Table 2.

559 To measure the effect of candidate genes and Rev on the nuclear export efficacy (35) for HIV-1 US RNA
560 for each treatment group, the mean US copy numbers/ng RNA extracted from the replicates of the
561 cytoplasmic fractions were divided by the mean US copy numbers/ng RNA extracted from replicates of
562 the whole cell lysates. The ratio of each treatment group were then again divided by the ratios observed
563 in controls cells transduced with the NT1 or the GFP-expression vector to calculate the fold changes
564 given in Table 2.

565 Fold changes for cytoplasmic levels of US or FS RNA were calculated by dividing the mean US or FS copy
566 numbers per ng of extracted RNA of each treatment group by the mean derived from the control groups.

567 **RNA-seq** Cytoplasmic RNAs were extracted from the sorted BFP+ or unsorted control J-dual#3 cells
568 with the Direct-zol RNA miniprep plus kit. RNA qualities were controlled by the 2100 Bioanalyzer
569 instrument (Agilent) (data not shown). For each sample, a DNA fragment library was generated from the
570 total RNA input using the Illumina TrueSeq Stranded mRNA kit and performed essentially as described
571 by the manufacturer's protocol (https://support.illumina.com/downloads/truseq-stranded-mrna-
572 reference-guide-1000000040498.html). The libraries were sequenced on an Illumina HiSeq 2500
573 platform and the reads were saved together with a quality score. Raw data were converted to reads
574 including a quality score by Software bcl2fastq v2.17. Sequences matching Illumina TrueSeq adapters
575 were masked. Low quality bases, poly-A or poly-T stretches as well as masked bases were trimmed by
576 Software fqtrim v0.9.5. Read quality was checked after sequencing and after base trimming by Software
577 fastqc v0.11.8. For the viral genome, reads were aligned to HIV-Dual-GT using hisat2 (79). Read counting
578 for the different features was performed using samtools (80) and regtools (81).

579 For splicing and differential expression analysis, reads were aligned to the human transcriptome using
580 RSEM v1.2.20 (82) and Bowtie v1.1.2 (82) as read alignment program using default parameters.
581 Differential expression analysis was performed by DESeq2 v1.18.1 (82) using the RSEM output for read
582 counts per gene. We considered genes that showed an absolute log2-transformed fold change greater
583 than 1 and a Padj value of less than 0.1 as differentially expressed. To exclude lowly expressed genes, we
584 only considered those with mean TPM greater than 1 in either untreated or knockdown samples.
585 Gencode v19 annotation was used in all steps of the analysis, including in the definition of protein
586 coding genes. The reads mapping profiles were exhibited via the Integrative Genomics Viewer (IGV) v2.3
587 (83). Differential splicing analysis was performed with SUPPA (61) using TPM values from the RSEM
588 transcript isoform output. For PSI calculation, events were required to pass the expression filter of TPM
589 greater than 1 (-f 1 parameter). We considered splicing events with Padj values lower than 0.01 and

590 absolute ΔPSI greater than 0.25 as differentially included or excluded. Polypyrimidine track length, and
591 branch point scores were calculated using BPfinder (84). Splice-site strength was calculated using
592 MaxENT (85) (hollywood.mit.edu/burgelab/software.html).

593 **Data availability**

594 The raw and processed data of RNA-seq are accessible via the GEO (Gene Expression Omnibus)
595 Accession Number GSE144347.

596 **Acknowledgements**

597 H.X. was supported by a fellowship from the China Scholarship Council. The authors would like to
598 acknowledge support from the Medical Faculty´s core facilities "Cell Sorting and Immunomonitoring"
599 and "Microarray / Next Generation Sequencing". Plasmids psPAX2 and pMD2.G are gifts from Dr. Didier
600 Trono. CRISPR reagents and libraries were a gift from Feng Zhang obtained from Addgene. Plasmid NL4-
601 3-Δ6-drEGFP was kindly provided by Dr. Robert Siliciano.

602 **Author Contributions**

603 K. Ü., A. E., B. G. and M.L. designed and supervised the study. H. X. and D. J. performed the experiments.
604 H. X., E. W., M. M, P. K., A. B. O. V. S., A. E. and K. Ü. analyzed the data. H. X. and K. Ü. wrote the
605 manuscript. A.E., B.G., E. W., M. M, P. K., A. E., A.B.O.V.S., M.L. and D.J. reviewed the manuscript.

606 **Declaration of Interests**

607 The authors declare that they have no competing interests.

608 **References**

609 1. Jacob AG, Smith CWJ. 2017. Intron retention as a component of regulated gene expression
610 programs. Hum Genet 136:1043–1057.

611 2. Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis
612 T, Frey B, Irimia M, Blencowe BJ. 2014. Widespread intron retention in mammals functionally
613 tunes transcriptomes. Genome Res 24:1774–1786.

614 3. Zhang S-J, Wang C, Yan S, Fu A, Luan X, Li Y, Sunny Shen Q, Zhong X, Chen J-Y, Wang X, Chin-Ming
615 Tan B, He A, Li C-Y. 2017. Isoform Evolution in Primates through Independent Combination of
616 Alternative RNA Processing Events. Mol Biol Evol 34:2453–2468.

617 4. Hammarskjold M-L, Rekosh D. 2017. SR proteins: To shuttle or not to shuttle, that is the question.
618 J Cell Biol 216:1875–1877.

619 5. Galganski L, Urbanek MO, Krzyzosiak WJ. 2017. Nuclear speckles: molecular organization,
620 biological function and role in disease. Nucleic Acids Res 45:10350–10368.

621 6. Cochrane AW, McNally MT, Mouland AJ. 2006. The retrovirus RNA trafficking granule: from birth
622 to maturity. Retrovirology 3:18.

623 7. Ahmad N, Maitra RK, Venkatesan S. 1989. Rev-induced modulation of Nef protein underlies
624 temporal regulation of human immunodeficiency virus replication. Proc Natl Acad Sci U S A
625 86:6111–5.

626 8. Karn J, Stoltzfus CM. 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene
627 expression. Cold Spring Harb Perspect Med 2:a006916.

628 9. Lemieux C, Marguerat S, Lafontaine J, Barbezier N, Bähler J, Bachand F. 2011. A Pre-mRNA
629 Degradation Pathway that Selectively Targets Intron-Containing Genes Requires the Nuclear
630 Poly(A)-Binding Protein. Mol Cell 44:108–119.

631 10. Sandri-Goldin RM. 2004. Viral regulation of mRNA export. J Virol 78:4389–96.

632 11. Cullen BR. 2000. Nuclear RNA export pathways. Mol Cell Biol 20:4181–7.

633 12. Grewe B, Uberla K. 2010. The human immunodeficiency virus type 1 Rev protein: menage a trois
634 during the early phase of the lentiviral replication cycle. J Gen Virol 91:1893–1897.

635 13. Mousseau G, Valente ST. 2017. Role of Host Factors on the Regulation of Tat-Mediated HIV-1
636 Transcription. Curr Pharm Des 23:4079–4090.

637 14. Suhasini M, Reddy TR. 2009. Cellular proteins and HIV-1 Rev function. Curr HIV Res 7:91–100.

638 15. Pollard VW, Malim MH. 1998. THE HIV-1 REV PROTEIN. Annu Rev Microbiol 52:491–532.

639 16. Cochrane AW, Jones KS, Beidas S, Dillon PJ, Skalka AM, Rosen CA. 1991. Identification and
640 characterization of intragenic sequences which repress human immunodeficiency virus structural
641 gene expression. J Virol 65:5305–13.

642 17. Mikaélian I, Krieg M, Gait MJ, Karn J. 1996. Interactions of INS (CRS) elements and the splicing

643 machinery regulate the production of Rev-responsive mRNAs. J Mol Biol 257:246–64.

644 18. Brighty DW, Rosenberg M, Steitz JA. 1994. A cis-acting repressive sequence that overlaps the
645 Rev-responsive element of human immunodeficiency virus type 1 regulates nuclear retention of
646 env mRNAs independently of known splice signals. Proc Nati Acad Sci USA 91:8314–8318.

647 19. Black A, Luo J, Chun S, Bakker A, Fraser J, Rosenblatt J. 1996. Specific binding of polypyrimidine
648 tract binding protein and hnRNP A1 to HIV-1 CRS elements. Virus Genes 12:275–285.

649 20. Maldarelli F, Martin MA, Strebel K. 1991. Identification of posttranscriptionally active inhibitory
650 sequences in human immunodeficiency virus type 1 RNA: novel level of gene regulation. J Virol
651 65:5732–43.

652 21. Haas J, Park E-C, Seed B. 1996. Codon usage limitation in the expression of HIV-1 envelope
653 glycoprotein. Curr Biol 6:315–324.

654 22. Graf M, Bojak A, Deml L, Bieler K, Wolf H, Wagner R. 2000. Concerted action of multiple cis-acting
655 sequences is required for Rev dependence of late human immunodeficiency virus type 1 gene
656 expression. J Virol 74:10822–6.

657 23. Wagner R, Graf M, Bieler K, Wolf H, Grunwald T, Foley P, Überla K. 2000. Rev-Independent
658 Expression of Synthetic *gag-pol* Genes of Human Immunodeficiency Virus Type 1 and Simian
659 Immunodeficiency Virus: Implications for the Safety of Lentiviral Vectors. Hum Gene Ther
660 11:2403–2413.

661 24. Zhang Q, Chen C-Y, Yedavalli VSRK, Jeang K-T. 2013. NEAT1 long noncoding RNA and paraspeckle
662 bodies modulate HIV-1 posttranscriptional expression. MBio 4:e00596-12.

663 25. Kula A, Gharu L, Marcello A. 2013. HIV-1 pre-mRNA commitment to Rev mediated export through
664 PSF and Matrin 3. Virology 435:329–340.

665 26. Zolotukhin AS, Michalowski D, Bear J, Smulevitch S V., Traish AM, Peng R, Patton J, Shatsky IN,
666 Felber BK. 2003. PSF Acts through the Human Immunodeficiency Virus Type 1 mRNA Instability
667 Elements To Regulate Virus Expression. Mol Cell Biol 23:6618.

668 27. Yedavalli VSRK, Jeang K-T. 2011. Matrin 3 is a co-factor for HIV-1 Rev in regulating post-
669 transcriptional viral gene expression. Retrovirology 8:61.

670 28. Fong K-W, Li Y, Wang W, Ma W, Li K, Qi RZ, Liu D, Songyang Z, Chen J. 2013. Whole-genome
671 screening identifies proteins localized to distinct nuclear bodies. J Cell Biol 203:149–64.

672 29. Gordon H, Ajamian L, Valiente-Echeverrìa F, Lévesque K, Rigby WF, Mouland AJ. 2013. Depletion
673 of hnRNP A2/B1 overrides the nuclear retention of the HIV-1 genomic RNA. RNA Biol 10:1714–
674 1725.

675 30. Dziembowski A, Ventura A-P, Rutz B, Caspary F, Faux C, Halgand F, Laprévote O, Séraphin B.
676 2004. Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and
677 splicing. EMBO J 23:4847–56.

678 31. Frankiw L, Majumdar D, Burns C, Vlach L, Moradian A, Sweredoski MJ, Baltimore D. 2019. BUD13
679 Promotes a Type I Interferon Response by Countering Intron Retention in Irf7. Mol Cell 73:803–

680 814.e6.

681 32. Amit M, Donyo M, Hollander D, Goren A, Kim E, Gelfman S, Lev-Maor G, Burstein D, Schwartz S,
682 Postolsky B, Pupko T, Ast G. 2012. Differential GC Content between Exons and Introns Establishes
683 Distinct Strategies of Splice-Site Recognition. Cell Rep 1:543–556.

684 33. Fernandez JP, Moreno-Mateos MA, Gohr A, Miao L, Chan SH, Irimia M, Giraldez AJ. 2018. RES
685 complex is associated with intron definition and required for zebrafish early embryogenesis. PLoS
686 Genet 14:e1007473.

687 34. Galy V, Gadal O, Fromont-Racine M, Romano A, Jacquier A, Nehrbass U. 2004. Nuclear Retention
688 of Unspliced mRNAs in Yeast Is Mediated by Perinuclear Mlp1. Cell 116:63–73.

689 35. Coyle JH, Bor Y-C, Rekosh D, Hammarskjold M-L. 2011. The Tpr protein regulates export of
690 mRNAs with retained introns that traffic through the Nxf1 pathway. RNA 17:1344–56.

691 36. Bonnet A, Bretes H, Palancade B. 2015. Nuclear pore components affect distinct stages of intron-
692 containing gene expression. Nucleic Acids Res 43:4249–61.

693 37. Zander G, Hackmann A, Bender L, Becker D, Lingner T, Salinas G, Krebber H. 2016. mRNA quality
694 control is bypassed for immediate export of stress-responsive transcripts. Nature 540:593–596.

695 38. Blanco R, Carrasco L, Ventoso I. 2003. Cell killing by HIV-1 protease. J Biol Chem 278:1086–93.

696 39. Malim MH, Emerman M. 2008. HIV-1 accessory proteins--ensuring viral survival in a hostile
697 environment. Cell Host Microbe 3:388–98.

698 40. Stewart SA, Poon B, Jowett JB, Xie Y, Chen IS. 1999. Lentiviral delivery of HIV-1 Vpr protein
699 induces apoptosis in transformed cells. Proc Natl Acad Sci U S A 96:12039–43.

700 41. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE,
701 Doench JG, Zhang F. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells.
702 Science 343:84–7.

703 42. Sanjana NE, Shalem O, Zhang F. 2014. Improved vectors and genome-wide libraries for CRISPR
704 screening. Nat Methods 11:783–784.

705 43. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M, Liu XS. 2014. MAGeCK
706 enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout
707 screens. Genome Biol 15:554.

708 44. Shifrut E, Carnevale J, Tobin V, Roth TL, Woo JM, Bui CT, Li PJ, Diolaiti ME, Ashworth A, Marson A.
709 2018. Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune
710 Function. Cell 175:1958–1971.e15.

711 45. Takada I, Tsuchiya M, Yanaka K, Hidano S, Takahashi S, Kobayashi T, Ogawa H, Nakagawa S,
712 Makishima M. 2018. Ess2 bridges transcriptional regulators and spliceosomal complexes via
713 distinct interacting domains. Biochem Biophys Res Commun 497:597–604.

714 46. Schapira M, Tyers M, Torrent M, Arrowsmith CH. 2017. WD40 repeat domain proteins: a novel
715 target class? Nat Rev Drug Discov 16:773–786.

716 47. Yan C, Wan R, Bai R, Huang G, Shi Y. 2017. Structure of a yeast step II catalytically activated
717 spliceosome. Science 355:149–155.

718 48. Zhan X, Yan C, Zhang X, Lei J, Shi Y. 2018. Structure of a human catalytic step I spliceosome.
719 Science 359:537–545.

720 49. Hou S, Qu D, Li Y, Zhu B, Liang D, Wei X, Tang W, Zhang Q, Hao J, Guo W, Wang W, Zhao S, Wang
721 Q, Azam S, Khan M, Zhao H, Zhang L, Lei H. 2019. XAB2 depletion induces intron retention in
722 POLR2A to impair global transcription and promote cellular senescence. Nucleic Acids Res
723 47:8239–8254.

724 50. Zhang X, Yan C, Hang J, Finci LI, Lei J, Shi Y. 2017. An Atomic Structure of the Human Spliceosome.
725 Cell 169:918–929.e14.

726 51. Chanarat S, Sträßer K. 2013. Splicing and beyond: The many faces of the Prp19 complex. Biochim
727 Biophys Acta 1833:2126–2134.

728 52. Grainger RJ, Beggs JD. 2005. Prp8 protein: at the heart of the spliceosome. RNA 11:533–57.

729 53. Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, Lander ES, Sabatini DM. 2015.
730 Identification and characterization of essential genes in the human genome. Science 350:1096–
731 101.

732 54. Reed R. 2003. Coupling transcription, splicing and mRNA export. Curr Opin Cell Biol 15:326–331.

733 55. Valencia P, Dias AP, Reed R. 2008. Splicing promotes rapid and efficient mRNA export in
734 mammalian cells. Proc Natl Acad Sci U S A 105:3386–91.

735 56. Blissenbach M, Grewe B, Hoffmann B, Brandt S, Uberla K. 2010. Nuclear RNA Export and
736 Packaging Functions of HIV-1 Rev Revisited. J Virol 84:6598–6604.

737 57. Grewe B, Hoffmann B, Ohs I, Blissenbach M, Brandt S, Tippler B, Grunwald T, Uberla K. 2012.
738 Cytoplasmic utilization of human immunodeficiency virus type 1 genomic RNA is not dependent
739 on a nuclear interaction with gag. J Virol 86:2990–3002.

740 58. Taniguchi I, Mabuchi N, Ohno M. 2014. HIV-1 Rev protein specifies the viral RNA export pathway
741 by suppressing TAP/NXF1 recruitment. Nucleic Acids Res 42:6645–58.

742 59. Sithole N, Williams CA, Vaughan AM, Kenyon JC, Lever AML. 2018. DDX17 Specifically, and
743 Independently of DDX5, Controls Use of the HIV A4/5 Splice Acceptor Cluster and Is Essential for
744 Efficient Replication of HIV. J Mol Biol 430:3111–3128.

745 60. Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N. 2015. Alternative Splicing Signatures
746 in RNA-seq Data: Percent Spliced in (PSI). Curr Protoc Hum Genet 87:11.16.1-11.16.14.

747 61. Alamancos GP, Pagès A, Trincado JL, Bellora N, Eyras E. 2015. Leveraging transcript quantification
748 for fast computation of alternative splicing profiles. RNA 21:1521–31.

749 62. Ren L, McLean JR, Hazbun TR, Fields S, Vander Kooi C, Ohi MD, Gould KL. 2011. Systematic two-
750 hybrid and comparative proteomic analyses reveal novel yeast pre-mRNA splicing factors
751 connected to Prp19. PLoS One 6:e16719.

752 63. Katahira J, Yoneda Y. 2009. Roles of the TREX complex in nuclear export of mRNA. RNA Biol
753 6:149–52.

754 64. Takemura R, Takeiwa T, Taniguchi I, McCloskey A, Ohno M. 2011. Multiple factors in the early
755 splicing complex are involved in the nuclear retention of pre-mRNAs in mammalian cells. Genes
756 to Cells 16:1035–1049.

757 65. Wysoczanski P, Zweckstetter M. 2016. Retention and splicing complex (RES) - the importance of
758 cooperativity. RNA Biol 13:128–33.

759 66. Minocha R, Popova V, Kopytova D, Misiak D, Hüttelmaier S, Georgieva S, Sträßer K. 2018. Mud2
760 functions in transcription by recruiting the Prp19 and TREX complexes to transcribed genes.
761 Nucleic Acids Res 46:9749–9763.

762 67. Edenfeld G, Volohonsky G, Krukkert K, Naffin E, Lammel U, Grimm A, Engelen D, Reuveny A, Volk
763 T, Klämbt C. 2006. The Splicing Factor Crooked Neck Associates with the RNA-Binding Protein
764 HOW to Control Glial Cell Maturation in Drosophila. Neuron 52:969–980.

765 68. Lei H, Zhai B, Yin S, Gygi S, Reed R. 2013. Evidence that a consensus element found in naturally
766 intronless mRNAs promotes mRNA export. Nucleic Acids Res 41:2517–25.

767 69. Biswas P, Jiang X, Pacchia AL, Dougherty JP, Peltz SW. 2004. The human immunodeficiency virus
768 type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target
769 for antiviral therapy. J Virol 78:2082–7.

770 70. Yang H-C, Xing S, Shan L, O'Connell K, Dinoso J, Shen A, Zhou Y, Shrum CK, Han Y, Liu JO, Zhang H,
771 Margolick JB, Siliciano RF. 2009. Small-molecule screening using a human primary cell model of
772 HIV latency identifies compounds that reverse latency without cellular activation. J Clin Invest
773 119:3473–86.

774 71. Platt EJ, Wehrly K, Kuhmann SE, Chesebro B, Kabat D. 1998. Effects of CCR5 and CD4 cell surface
775 concentrations on infections by macrophagetropic isolates of human immunodeficiency virus
776 type 1. J Virol 72:2855–64.

777 72. de Jong SJ, Albrecht J-C, Schmidt M, Müller-Fleckenstein I, Biesinger B. 2010. Activation of
778 noncanonical NF-kappaB signaling by the oncoprotein Tio. J Biol Chem 285:16495–503.

779 73. Malim MH, Hauber J, Fenrick R, Cullen BR. 1988. Immunodeficiency virus rev trans-activator
780 modulates the expression of the viral regulatory genes. Nature 335:181–183.

781 74. Fouchier RA, Meyer BE, Simon JH, Fischer U, Malim MH. 1997. HIV-1 infection of non-dividing
782 cells: evidence that the amino-terminal basic region of the viral matrix protein is important for
783 Gag processing but not for post-entry nuclear import. EMBO J 16:4531–9.

784 75. Naldini L, Trono D, Verma IM. 2016. Lentiviral vectors, two decades later. Science 353:1101–
785 1102.

786 76. Thomas M, Sonntag E, Müller R, Schmidt S, Zielke B, Fossen T, Stamminger T. 2015. pUL69 of
787 Human Cytomegalovirus Recruits the Cellular Protein Arginine Methyltransferase 6 via a Domain
788 That Is Crucial for mRNA Export and Efficient Viral Replication. J Virol 89:9601–15.

789 77. Rio DC, Ares M, Hannon GJ, Nilsen TW. 2010. Purification of RNA using TRIzol (TRI reagent). Cold
790 Spring Harb Protoc 2010:pdb.prot5439.

791 78. Brandt S, Blißenbach M, Grewe B, Konietzny R, Grunwald T, Überla K. 2007. Rev Proteins of
792 Human and Simian Immunodeficiency Virus Enhance RNA Encapsidation. PLoS Pathog 3:e54.

793 79. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and
794 genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37:907–915.

795 80. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and
796 population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–93.

797 81. Feng Y-Y, Ramu A, Cotto KC, Skidmore ZL, Kunisaki J, Conrad DF, Lin Y, Chapman WC, Uppaluri R,
798 Govindan R, Griffith OL, Griffith M. 2018. RegTools: Integrated analysis of genomic and
799 transcriptomic data for discovery of splicing variants in cancer. bioRxiv 436634.

800 82. Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or
801 without a reference genome. BMC Bioinformatics 12:323.

802 83. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011.
803 Integrative genomics viewer. Nat Biotechnol 29:24–6.

804 84. Corvelo A, Hallegger M, Smith CWJ, Eyras E. 2010. Genome-wide association between branch
805 point properties and alternative splicing. PLoS Comput Biol 6:e1001016.

806 85. Yeo G, Burge CB. 2004. Maximum Entropy Modeling of Short Sequence Motifs with Applications
807 to RNA Splicing Signals. J Comput Biol 11:377–394.

808

809 **Figure legends**

810 **Figure 1. Establishment of HIV-dual-GT reporter cell clones.** (A) Map of the HIV-dual-GT proviral
811 reporter construct. (B) Western blot analysis of lysates of HEK293T cells transfected with HIV-dual-GT or
812 HIVΔr-e in the presence or absence of an Rev expression plasmid (pcRev) with antibodies to the p24
813 capsid protein, αTublin or Rev. (C) 293T cells transfected with HIV-dual-GT and pcRev or pcDNA3.1 were
814 analyzed by flow cytometry for Gag-BFP and DsRed expression. (D) Jurkat cells transduced with the HIV-
815 dual-GT reporter vector (MOI 0.3) and transiently transfected with pcRev or pcDNA control plasmid
816 were analyzed by flow cytometry. (E) Scheme for the generation of reporter cell clones. (F) HIV-dual-GT-
817 transduced Jurkat reporter cell clones J-dual#3 and J-dual#6 were superinfected with VSV-G-
818 pseudotyped HIV-1 viruses only differing in the expression of Rev and analyzed by flow cytometry. All
819 assays were performed at least twice and results of one representative experiment are shown.

820 **Figure 2. Genome-wide CRISPR/Cas screen.** (A) Outline of screening strategy. (B) Flow cytometric
821 analysis of J-dual#3 and J-dual#6 reporter cells without transduction and four days after transduction by
822 the GeCKO v2 lentiviral vector library (MOI around 2) immediately prior to the sorting process. Total
823 number of events shown in the graph, total number of cells analyzed in the screen and total numbers of
824 cells sorted are indicated. (C) Enrichment analysis for guide RNAs in transduced and selected J-dual#3
825 and J-dual#6 reporter cells compared to transduced but unselected Jurkat cells by the MAGeck program.
826 The robust ranking aggregation (RRA) score reflects the enrichment of guide RNA targeting the same
827 gene in cells selected from both reporter cell clones. The distribution of the negative log10 of the RRA
828 score for all targeted genes in alphabetical order on the X-axis is shown with the top 12 candidate genes
829 highlighted. (D) Additional MAGeck ranking parameters of the top 12 candidate genes. fdr, false
830 discovery rate; # sgRNA, number of enriched guide RNAs targeting the same gene.

831 **Figure 3. Protein interaction network of confirmed screening hits.** The 10 validated hits were analyzed
832 by the online database STRING for protein-protein interactions. The graph shows known and predicted
833 protein-protein interactions for the most enriched pathway, the "mRNA splicing" pathway (pathway ID:
834 GO 0000398; false discovery rate: 5.07e-14; minimum required interaction score: 0.400 (medium
835 confidence)).

836 **Figure 4. Role of confirmed screening hits in HIV-1 RNA splicing and export.** (A) J-dual#3 cells were co-
837 infected at an MOI of 3 with a lentiviral vector encoding Cas9 and a lentiviral vector expressing the guide
838 RNAs indicated. In addition, J-dual#3 cells were transduced by a lentiviral vector encoding GFP or Rev at
839 an MOI of 1. At 4 days after infection, BFP positive cells were sorted by flow cytometry from cells
840 transduced with lentiviral vectors encoding Rev or the guide RNAs targeting the confirmed hits. Matched
841 controls expressing NT1 or GFP were used without sorting. Total RNA was purified from whole cell
842 extracts. HIV-1 US and FS RNAs copy numbers in these samples were quantified by RT-qPCR and are
843 shown as log copies per ng total RNA extracted from whole cells. Log copies from at least three
844 experiments, as well as their mean and SEM are shown. The mean of the ratio of FS_{whole}/US_{whole} of each
845 experiment is shown above the x-axis for each of the treatment groups. (B) J-dual#3 knockout cells and
846 control cells were prepared as described in (A), but total RNA was extracted from the cytoplasmic
847 fraction of each cell sample. Log copies per ng total cytoplasmic RNA and mean ratios of $FS_{cyt.}/US_{cyt.}$ were
848 calculated and presented as in (A). (C) The fold changes in the different treatment groups for the splicing
849 efficacy (FS_{whole}/US_{whole}), the nuclear export efficacy ($US_{cyt.}/US_{whole}$), and cytoplasmic levels of unspliced

850 (US $_{cyt.}$) and fully spliced (FS $_{cyt.}$) transcriptes were calculated by dividing the values in the treatment
851 groups by the ones obtained from the matched control cells (NT1, GFP) as described in the Materials
852 and Methods. Paired t-test was performed on the log copy numbers/ng total RNA between each of the
853 treatment groups and the matched control. * $p<0.05$, ** $p<0.01$, ***$p<0.001$. Unless otherwise
854 indicated, differences were not significant. Dotted lines separate experiments that were not performed
855 in parallel and therefore contain independent controls. WDR70.KD cells were sorted on day 7 after
856 transduction, while other cells were sorted on day 4.

857 **Figure 5. Effect of CRNKL1 depletion on HIV RNA splicing efficiency, accuracy, and export.** (A)
858 Alignments of reads from the transcriptomic analysis to HIV-dual-GT. Shown are coverage profiles for
859 the individual RNA-seq samples, colored, from top to bottom the non-targeting control (NT1, red),
860 CRNKL1 knock-down (cyan), and rev overexpression (purple). Splice sites are shown in orange (donors)
861 and dark yellow (acceptors). Coding regions and LTRs are indicated in dark blue. (B) Percentage of all HIV
862 reads among the total (HIV + human) uniquely mapped reads. (C) Percentage of reads mapping to HIV
863 Exon 1 (from the transcription start site to SD1) among the total uniquely mapped reads. (D) Percentage
864 of reads in the first intron between SD1 and SA1 among the total uniquely mapped reads. (E) Percentage
865 of reads across the unspliced SD1 site among the total uniquely mapped reads. (F) Percentage of reads
866 across the unspliced SD4 site among the total uniquely mapped reads. (G) Percentage of reads spanning
867 SD and SA sites (spliced reads) among the total uniquely mapped reads. (H) Percentages of reads
868 spanning SD1 and the indicated splice acceptor sites among all spliced HIV reads. For SD1-SA4abc, reads
869 spliced from SD1 to SA4a, SA4b and SA4c are summed up. (I) Percentages of reads spanning the
870 indicated splice donor sites and SA5 among all spliced HIV reads. (J) Percentage of reads spliced from
871 SD1 to any SA site or from SD4 to SA7 among all spliced HIV reads. (B-J) Shown are averages and
872 standard deviations for the control and CRNKL1 knock-down samples. Individual values are shown as
873 black dots. For the rev overexpression sample (purple), the individual value is shown.

874

875 **Figure 6. Changes in the host transcriptome upon CRNKL1 knock-down.** (A) RNA-seq differential
876 expression analysis of mRNA abundance upon CRNKL1 KD. MA plot shows differences in the abundance
877 of protein-coding transcripts in CRNKL1 KD vs. control (NT1) conditions. mRNAs with significant
878 differences are depicted in red (Padj <0.1). (B) Gene Ontology enrichment analysis of mRNAs with
879 upregulated and downregulated expression upon CRNKL1 KD. Top enriched GO Biological Process terms
880 are shown. (C) Upregulated (included) (ΔPSI (KD-NT1) >0.25, Padj <0.01) and downregulated (excluded)
881 (ΔPSI (KD-NT1) <-0.25, Padj<0.01) splicing events were quantified from the RNA-seq CRNKL1 KD
882 experiment. Bar plot shows absolute numbers of splicing events that are up or down regulated upon
883 CRNKL1 knock-down. (D) Branch point scores were calculated by BPfinder and alternative 3'splice sites
884 with higher usage in CRNKL1 KD condition were compared to 3'splice sites of the same genes with
885 unchanged (constitutive) usage. Box plots and data points are shown. Wilcoxon rank sum test was used
886 to test for significance. Absolute numbers of 3' splice sites included in the analysis for each group are
887 indicated. (E) Length of retained introns with higher inclusion upon KD (CRNKL1-dependent) was
888 compared to the length of unchanged introns upon KD (CRNKL1-independent), which were located in
889 the same transcripts as CRNKL1-dependent ones. Box plots and data points are shown. Wilcoxon rank

890 sum test was used to test for significance. Absolute numbers of retained introns analyzed in both groups
891 are indicated. PSI, percent spliced-in for splicing events.

892 **Figure 7. Potential mechanisms leading to increased cytoplasmic levels of unspliced HIV-1 RNA.**
893 Graphical representation of HIV-1 RNA metabolism and expected consequences of up- (↑) or down- (↓)
894 regulation or no change (→) of the respective steps on the nuclear export efficacy ($US_{cyt.}/US_{whole}$) and
895 splicing efficacy (FS_{whole}/US_{whole}). The nuclear export pathway is highlighted in bold.

896

897 **Supplementary figure legends**

898 **Figure S1. Enrichment of individual sgRNAs targeting the top 12 candidate genes.** NGS data from
899 unselected Jurkat cells (control) and selected J-dual#3 and J-dual#6 cells were analyzed by MAGeCK.
900 Results are shown as normalized counts for each of the six guide RNAs targeting the indicated candidate
901 genes.

902 **Figure S2. Validation of screening hits.** (A) J-dual#3 and J-dual#6 reporter cells were co-infected with a
903 lentiviral vector encoding Cas9 and a lentiviral vector expressing the indicated guide RNAs at an MOI of
904 1. Flow cytometric analyses were performed 7 days after infection. Each of the 12 candidate genes was
905 targeted by two different sgRNAs. A non-targeting sgRNA (NT1) was used as a control. Representative
906 results from at least 2 experiments are shown. (B) Percent transduction by individual lentiviral knock-out
907 vectors**.** All individual lentiCRISPR-KO constructs were packaged into VSV-G pseudotyped lentiviral
908 vector particles under the same conditions. The same volume of individual lentiCRISPR vector
909 preparation was used to infect J-dual#3 cells. The percent transduction was calculated as described in
910 Materials and Methods.

911 **Figure S3. Influence of selected proteins reported to be involved in pre-mRNA splicing and nuclear**
912 **retention on Gag-BFP expression.** Two sgRNAs for each gene were picked from the GeCKO v2 library to
913 target PRPF8, PRPF19, TPR and BUD13, respectively. The individual lentiCRISPR knockout vectors were
914 constructed and used to transduce the J-dual#3 reporter cell line. Transduction efficiencies were
915 confirmed (data not shown) and cells were analyzed by flow cytometry 7 days after transduction (A-D).

916 **Figure S4. Validation of fractionation.** (A) Representative Western blot analyses for LaminB (nuclear)
917 and αTublin (cytoplasmic) in whole cell extracts and nuclear and cytoplasmic fractions from $1x10^6$ Jurkat
918 cells. (B) Differences in GAPDH pre mRNA levels. CT values obtained by real time PCR from whole cell or
919 cytoplasmic extracts from three experiments are shown. Numbers indicate the fold difference between
920 GAPDH pre mRNA levels/weight of extracted RNA from whole cell and cytoplasmic extracts, assuming 2-
921 fold differences in the targeted RNA copy numbers per ΔCT and adjusting for minor differences in the
922 amounts of RNA added to the PCR reactions.

923 **Figure S5. Number of reads and deviations from the wild type CRNKL1 target region after knock down**
924 **of CRNKL1**. The CRNKL1 sgRNA target region with the predicted Cas9 cleavage site (arrow) and the PAM
925 region (shaded in orange) are aligned with reads from the transcriptome from four control (NT1 exp1-4)

926 and four CRNKL1 knock down experiments (KD exp1-4). Underlined sequences were used to map reads
927 to the target region. Point mutations, deletions (Δ) and insertions (I with number of nucleotides inserted
928 given above) are indicated. "n": number of reads with the respective sequence. "m": percent mutation
929 rate calculated as [100 x number of mutated reads / number of total reads mapped to the CRNKL1
930 target region].

931 **Figure S6. Changes in splice site usage of HIV RNAs upon CRNKL1 depletion.** (A) Heat map representing,
932 for every row, the relative frequency of indicated HIV-1 splice donor-splice acceptor pairs. Colors
933 represent the log10 transformed percentages relative to all spliced HIV reads, as shown in the color bar
934 underneath the heat map. Rows are sorted according to the splice donor sites. (B) As in A, but sorted for
935 splice acceptor sites.

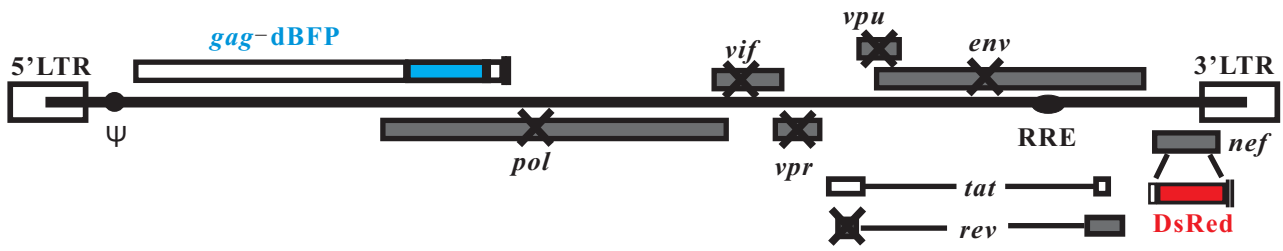
936 **Figure S7. Quality control analyses for the transcriptomic data and additional analyses on differential**
937 **splicing and intronic features**. (A) Principal component analysis (PCA) of exonic read counts. Variance
938 was assessed between biological replicates (1-4) and conditions (NT1, KD, REV). (B) Pairwise assessment
939 of Pearson correlation coefficients between all exonic read counts from all RNA-seq samples. (C)
940 Differential analysis of splicing events upon CRNKL1 KD. Volcano plot of ΔPSI (KD-NT1) values vs.
941 negative log10-transformed Padj values is shown. To select up- and downregulated splicing events upon
942 CRNKL1 KD (blue and red data points) the following cutoffs were used, respectively: ΔPSI (KD-NT1) >0.25,
943 Padj <0.01; and ΔPSI (KD-NT1) <-0.25, Padj <0.01. (D) Pairwise assessment of Pearson correlation
944 coefficients between all splicing event PSI values from all RNA-seq samples. (E) The length of
945 polypyrimidine tracts of alternative 3'splice sites with higher usage under CRNKL1 KD condition were
946 compared to those of constitutive 3'splice sites present in the same mRNAs. Wilcoxon rank sum test was
947 used to test for significance. Absolute numbers of 3'splice sites in both groups are indicated. (F)
948 Polypyrimidine track length was obtained by BPfinder for retained introns with higher inclusion upon KD
949 (CRNKL1-dependent) and compared to that of unchanged retained introns upon KD (CRNKL1-
950 independent), which were located in the same transcripts as CRNKL1-dependent ones. Box plots and
951 data points are shown. Wilcoxon rank sum test was used to test for significance. Absolute numbers of
952 retained introns in both groups are indicated. (G) GC content of retained introns with higher inclusion
953 upon KD (CRNKL1-dependent) was compared to the GC content of unchanged introns upon KD (CRNKL1-
954 independent), which were located in the same transcripts as CRNKL1-dependent ones. Box plots and
955 data points are shown. Wilcoxon rank sum test was used to test for significance. Absolute numbers of
956 retained introns in both groups are indicated.

957 **Figure S8. Examples and validation of intron retention in cellular transcripts after CRNKL1 knock-down.**
958 Mapping of the reads to genes with enhanced intron retention (RPL10, C19orf53 CD320), an alternative
959 5´and 3´splice site usage (5th intron of AURKB) and an unaffected gene (EIF6) are shown for the four
960 replicates of control (NT1) and CRNKL1 knock-down cells and the rev expressing cells. For RPL10 and
961 C19orf53, intron retention was confirmed by RT-qPCR. Binding sites for specific primers spanning the
962 exon-intron junctions are indicated by green arrows. RT-qPCRs were performed for relative
963 quantification of respective intron-retaining transcripts using the cytoplasmic RNAs from the RNA-seq
964 experiment. ΔCT(NT1-KD) ±SEM values obtained from the 4 replicates are indicated.

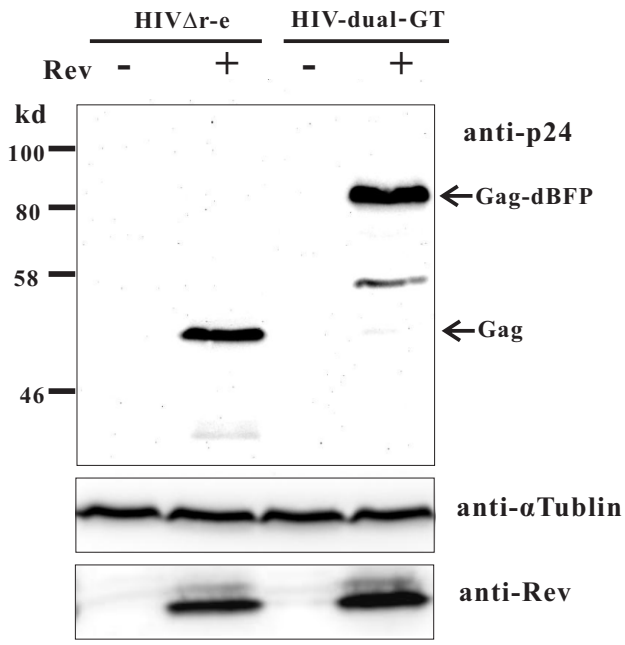
965 **Table S1. Primer sequences and combinations.** (A) Sequences of sgRNAs used for screening hits
966 validation. (B) Sequences of primers used for NGS libraries construction. (C) Primer combinations for
967 NGS libraries construction. (D) Sequences of primers used for validation of retained introns.

Figure 1

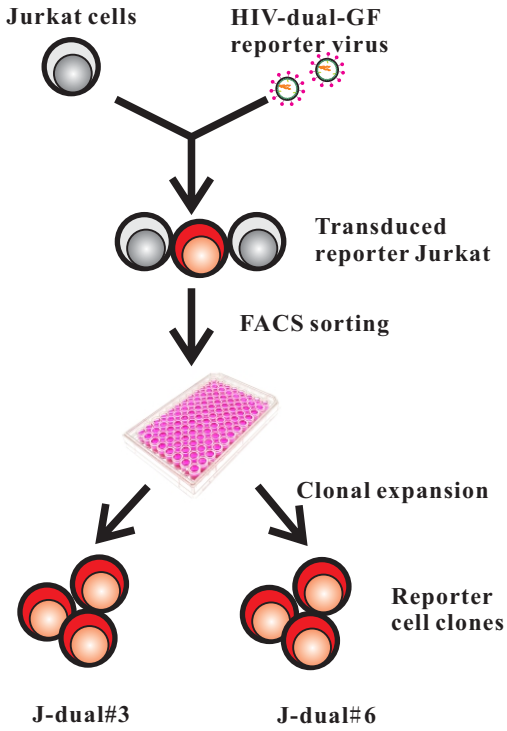
A HIV-dual-GT



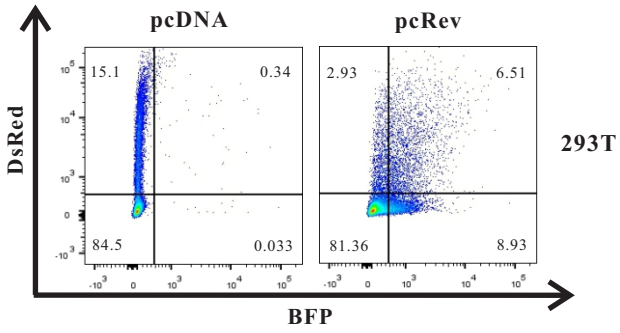
B



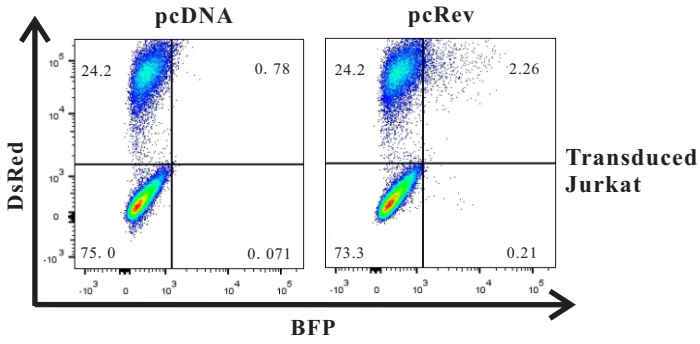
E



C



D



F

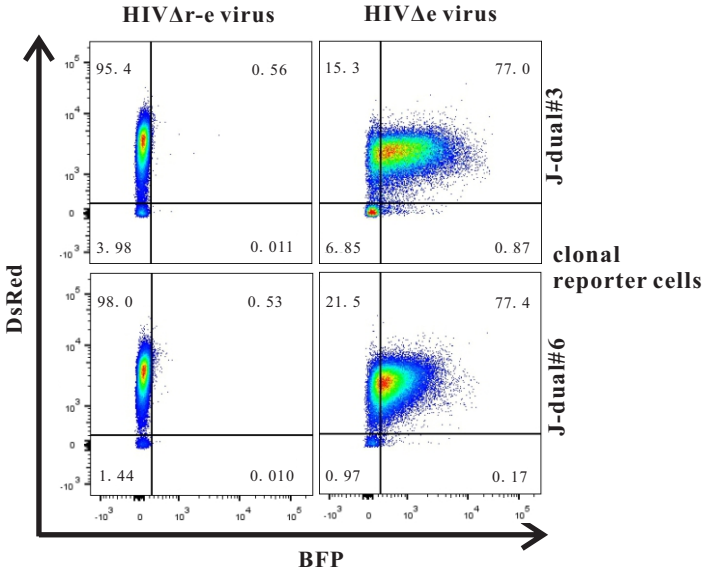
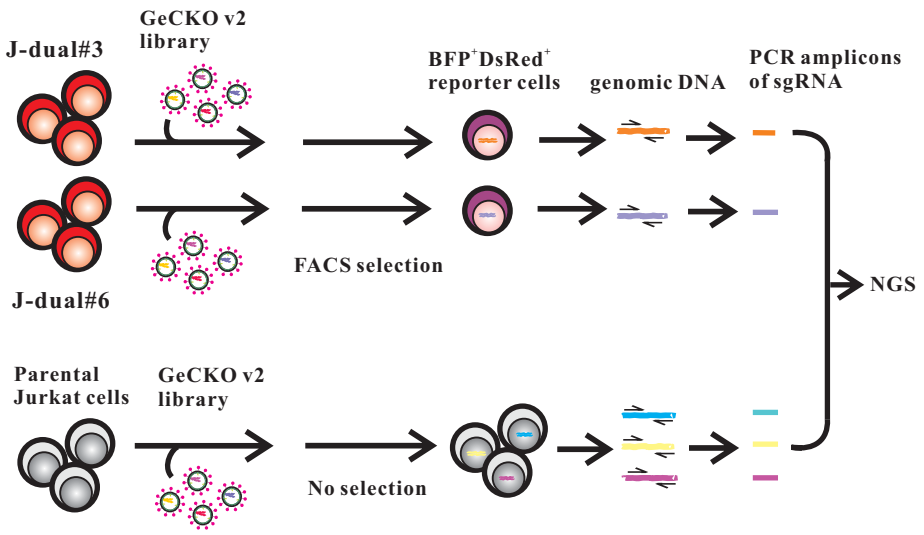
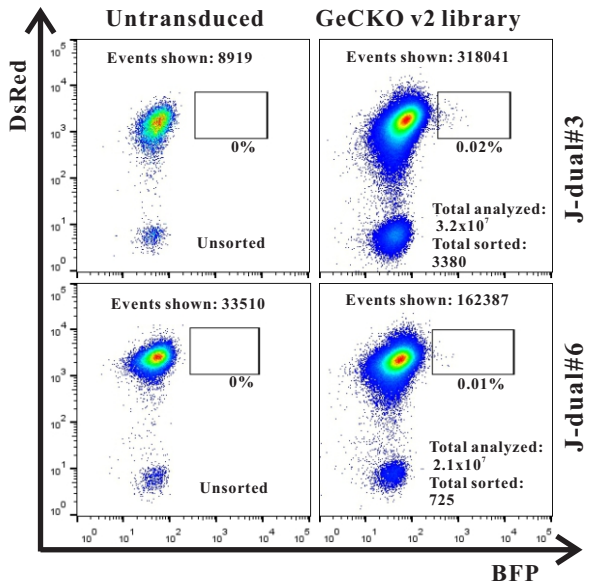


Figure 2

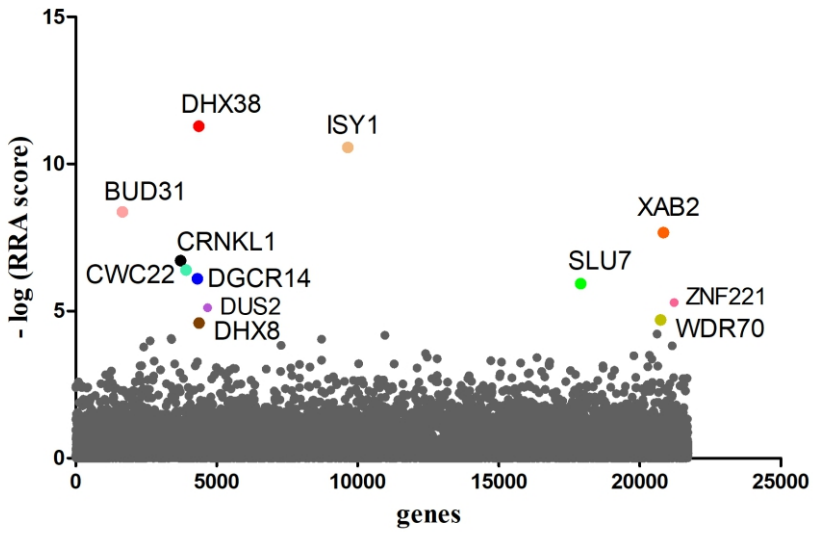
A



B



C



D Table 1. Ranking parameters for the top 12.

Genes	RRA score	p-value	fdr	rank	# sgRNA
DHX38	5.17E-12	2.28E-07	0.001238	1	4
ISY1	2.71E-11	2.28E-07	0.001238	2	4
BUD31	4.28E-09	2.28E-07	0.001238	3	4
XAB2	2.13E-08	2.28E-07	0.001238	4	4
CRNKL1	1.92E-07	1.60E-06	0.006931	5	2
CWC22	3.99E-07	2.05E-06	0.007426	6	4
DGCR14	7.95E-07	4.33E-06	0.013437	7	2
SLU7	1.18E-06	6.16E-06	0.016708	8	2
ZNF221	5.08E-06	2.08E-05	0.04505	9	2
DUS2	7.68E-06	1.66E-05	0.040154	10	3
WDR70	2.00E-05	6.77E-05	0.133663	11	3
DHX8	2.51E-05	8.42E-05	0.152228	12	1

Figure 3

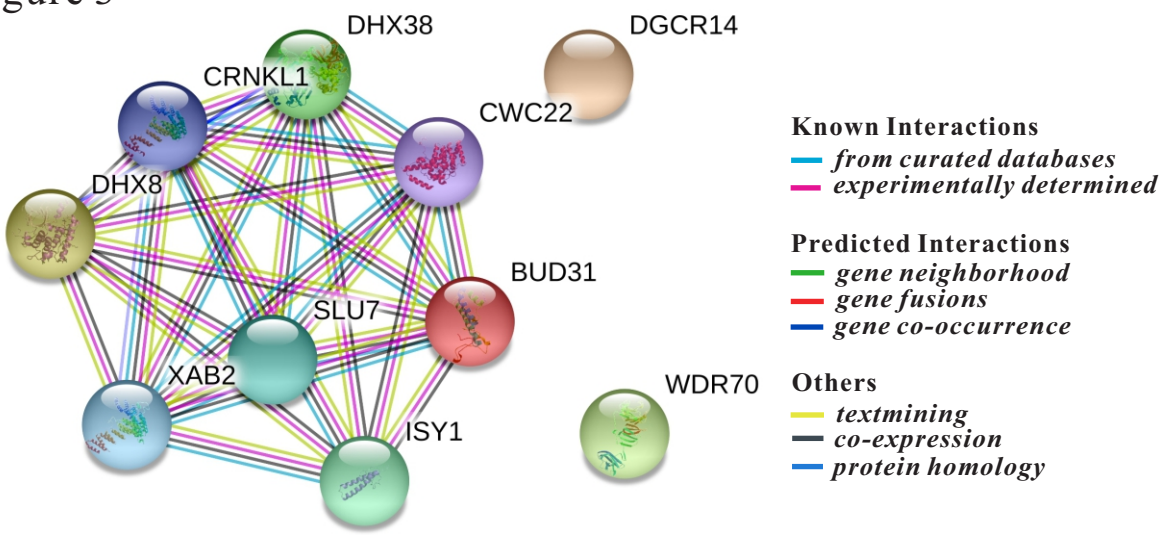
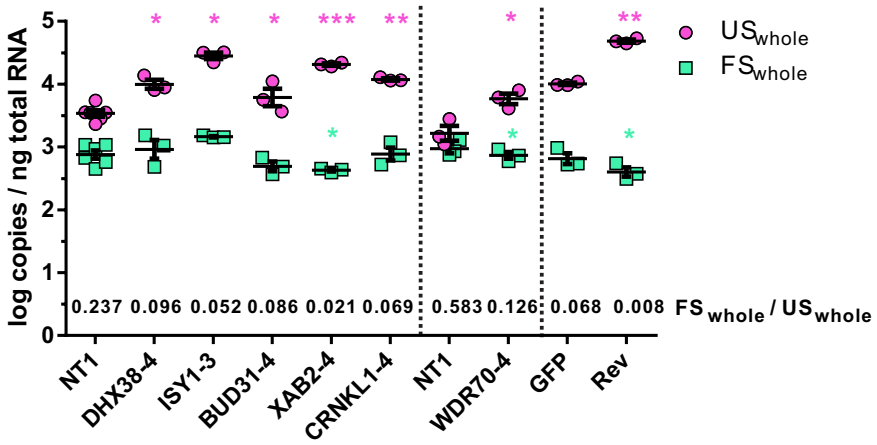
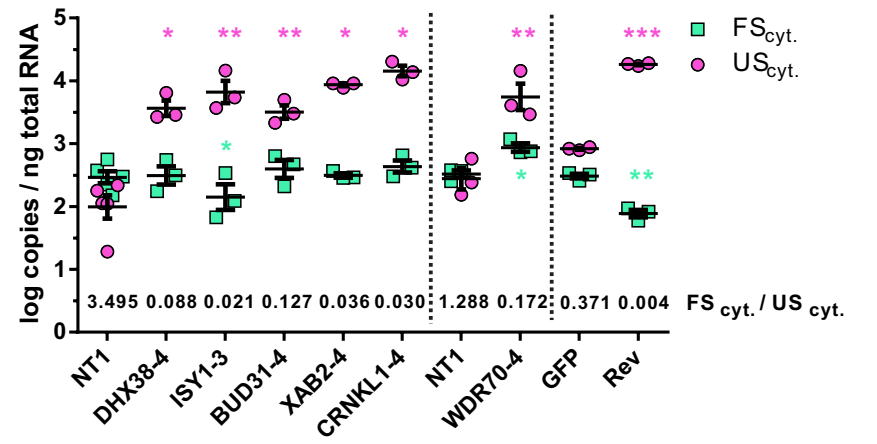

DHX38
DGCR14
CRNKL1
CWC22
DHX8
BUD31
SLU7
XAB2
WDR70
ISY1
Known Interactions
from curated databases
experimentally determined
Predicted Interactions
gene neighborhood
gene fusions
gene co-occurrence
Others
textmining
co-expression
protein homology

Figure 4

A



B



C Table 2. Fold changes of RNA levels & ratios.

Knock -down	Fold change of			
	splicing efficiency	export efficiency	$US_{cyt.}$	$FS_{cyt.}$
-DHX38	1/2.9	10.8	31.4	1.1
-ISY1	1/3.7	7.7	62.0	0.6
-BUD31	1/2.2	13.8	26.5	1.4
-XAB2	1/12.5	11.7	68.0	1.0
-CRNKL1	1/4.0	34.6	116.4	1.4
-WDR70	1/4.6	6.5	22.1	2.7
+Rev	1/7.8	4.6	22.0	0.3

Figure 5

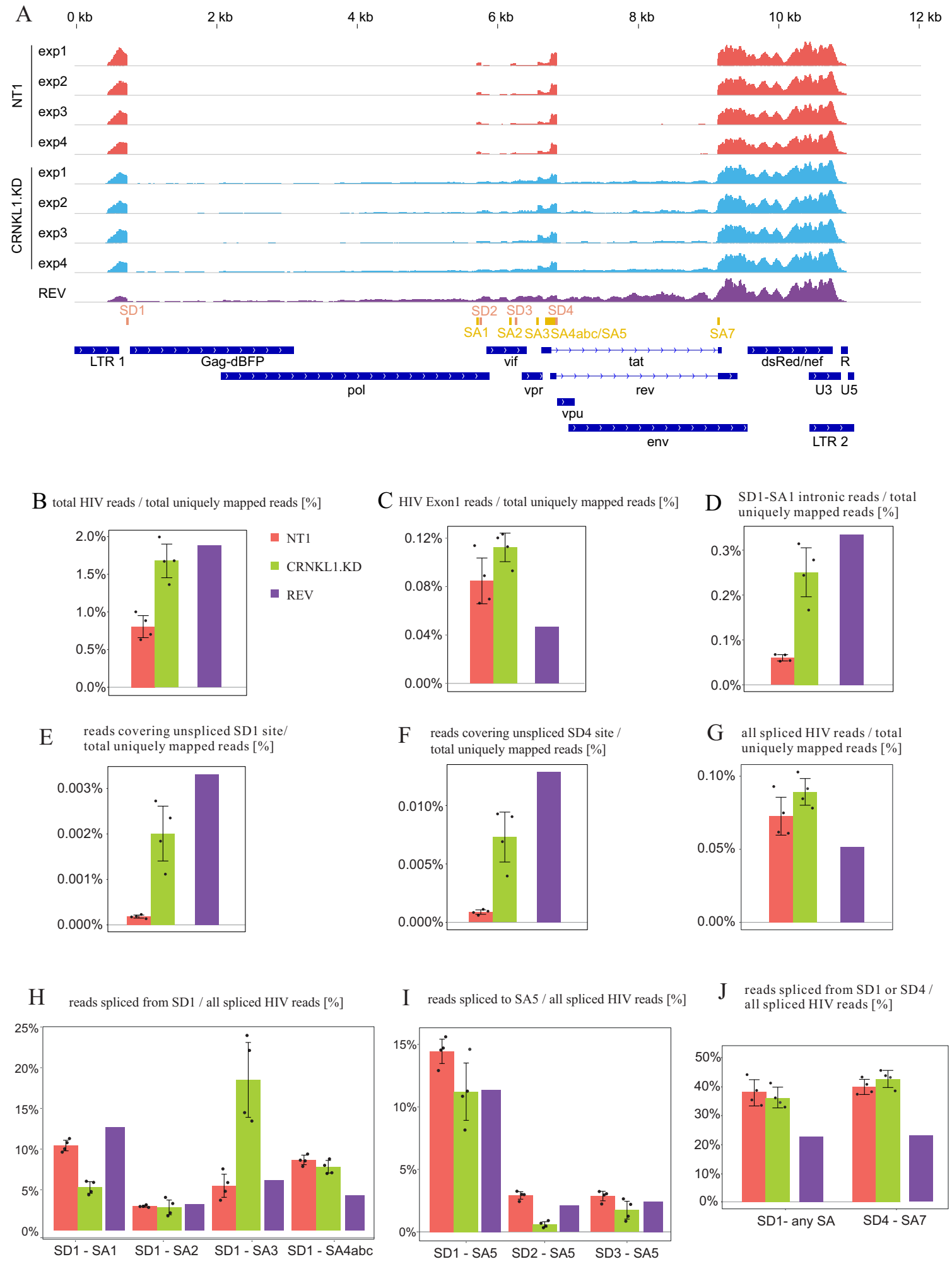

A
0 kb
2 kb
4 kb
6 kb
8 kb
10 kb
12 kb
NT1
exp1
exp2
exp3
exp4
CRNKL1.KD
exp1
exp2
exp3
exp4
REV
SD1
SD2
SD3
SD4
SA1
SA2
SA3
SA4abc/SA5
SA7
LTR 1
Gag-dBFP
vif
tat
dsRed/nef
R
pol
vpr
rev
U3
U5
vpu
env
LTR 2
B total HIV reads / total uniquely mapped reads [%]
2.0%
1.5%
1.0%
0.5%
0.0%
NT1
CRNKL1.KD
REV
C HIV Exon1 reads / total uniquely mapped reads [%]
0.12%
0.08%
0.04%
0.00%
D SD1-SA1 intronic reads / total uniquely mapped reads [%]
0.3%
0.2%
0.1%
0.0%
E reads covering unspliced SD1 site/ total uniquely mapped reads [%]
0.003%
0.002%
0.001%
0.000%
F reads covering unspliced SD4 site / total uniquely mapped reads [%]
0.010%
0.005%
0.000%
G all spliced HIV reads / total uniquely mapped reads [%]
0.10%
0.05%
0.00%
H reads spliced from SD1 / all spliced HIV reads [%]
25%
20%
15%
10%
5%
0%
SD1 - SA1
SD1 - SA2
SD1 - SA3
SD1 - SA4abc
I reads spliced to SA5 / all spliced HIV reads [%]
15%
10%
5%
0%
SD1 - SA5
SD2 - SA5
SD3 - SA5
J reads spliced from SD1 or SD4 / all spliced HIV reads [%]
50%
40%
30%
20%
10%
0%
SD1- any SA
SD4 - SA7

Figure 6

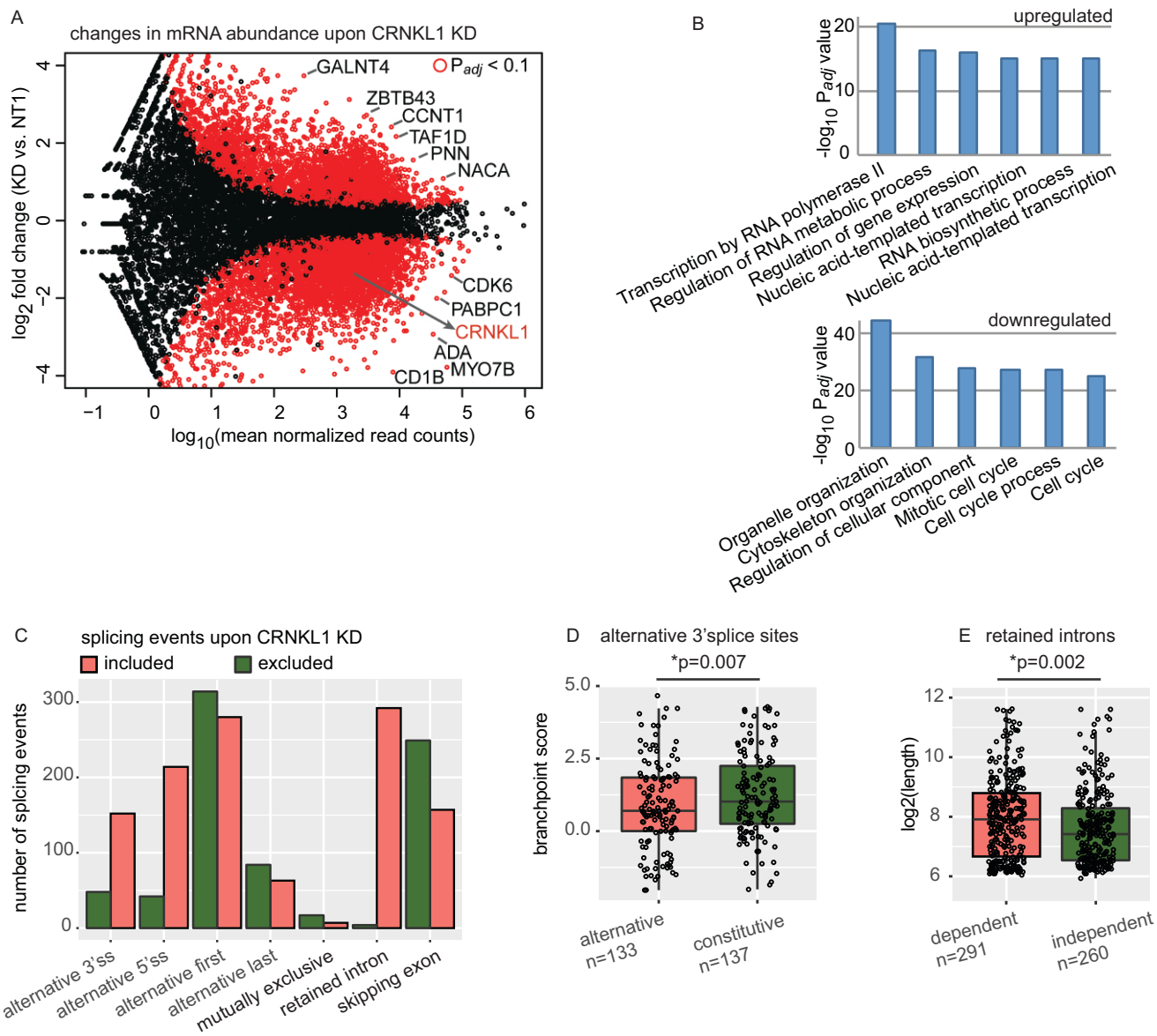

A
changes in mRNA abundance upon CRNKL1 KD
GALNT4
P_{adj} < 0.1
ZBTB43
CCNT1
TAF1D
PNN
NACA
CDK6
PABPC1
CRNKL1
ADA
CD1B
MYO7B
$\log_2$ fold change (KD vs. NT1)
$\log_{10}$(mean normalized read counts)
B
upregulated
$-\log_{10}$ P_{adj} value
Transcription by RNA polymerase II
Regulation of RNA metabolic process
Regulation of gene expression
Nucleic acid-templated transcription
RNA biosynthetic process
Nucleic acid-templated transcription
downregulated
Organelle organization
Cytoskeleton organization
Regulation of cellular component
Mitotic cell cycle
Cell cycle process
Cell cycle
C
splicing events upon CRNKL1 KD
included
excluded
number of splicing events
alternative 3'ss
alternative 5'ss
alternative first
alternative last
mutually exclusive
retained intron
skipping exon
D
alternative 3'splice sites
*p=0.007
branchpoint score
alternative
n=133
constitutive
n=137
E
retained introns
*p=0.002
log2(length)
dependent
n=291
independent
n=260

Figure 7

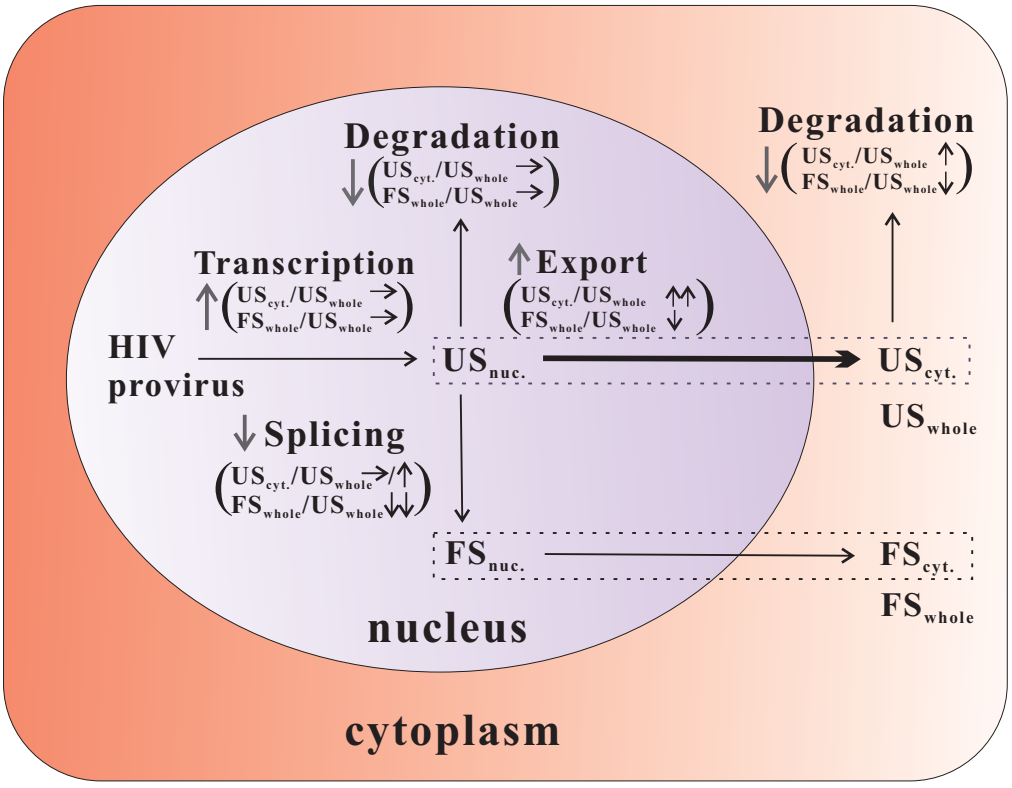

Degradation
↓(US cyt./US whole →
FS whole/US whole →)
Degradation
↓(US cyt./US whole ↑
FS whole/US whole ↓)
Transcription
↑(US cyt./US whole →
FS whole/US whole →)
↑Export
(US cyt./US whole ↑↑
FS whole/US whole ↓)
HIV
provirus
US nuc.
US cyt.
US whole
↓Splicing
(US cyt./US whole →/↑
FS whole/US whole ↓↓)
FS nuc.
FS cyt.
FS whole
nucleus
cytoplasm

Figure S1

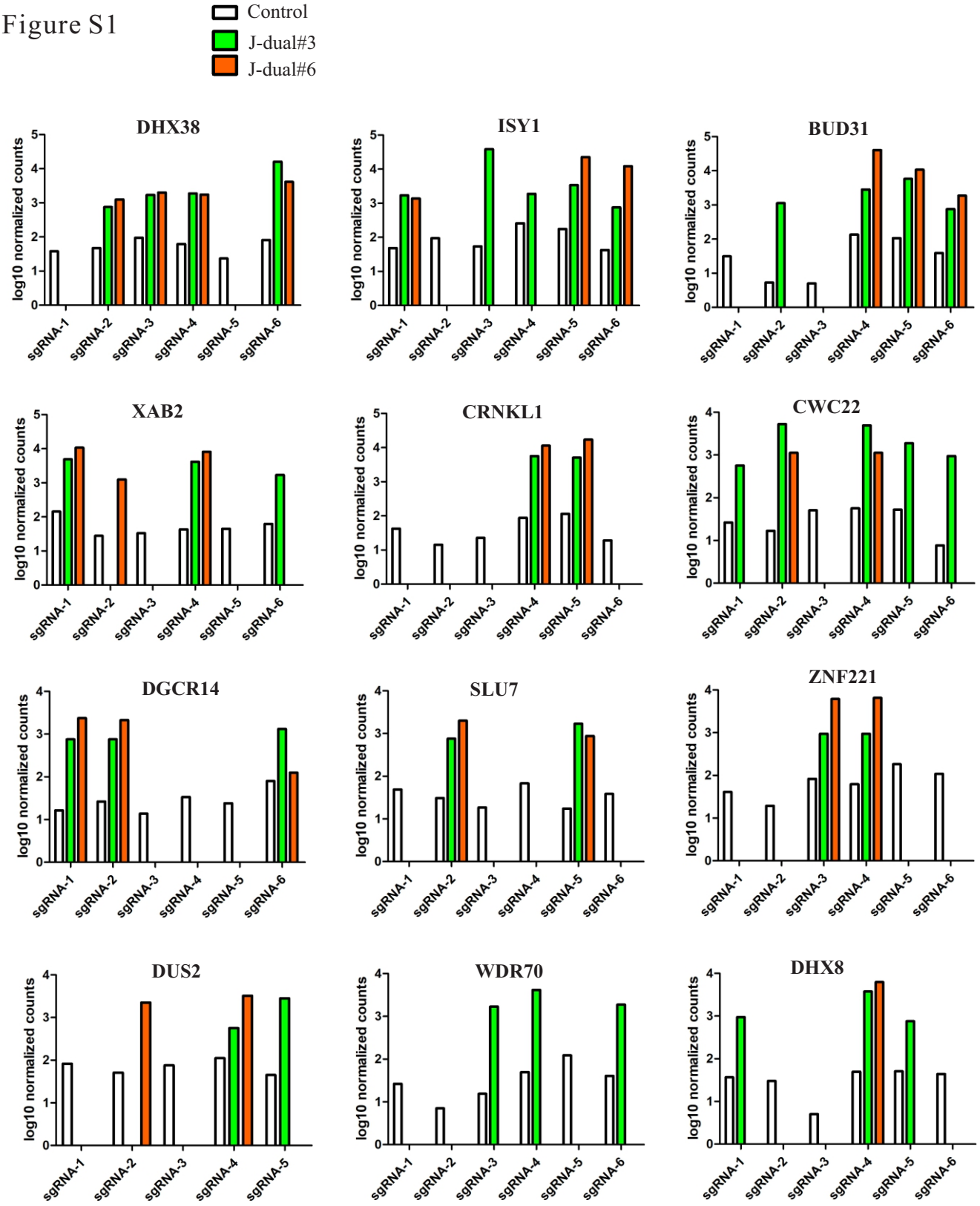
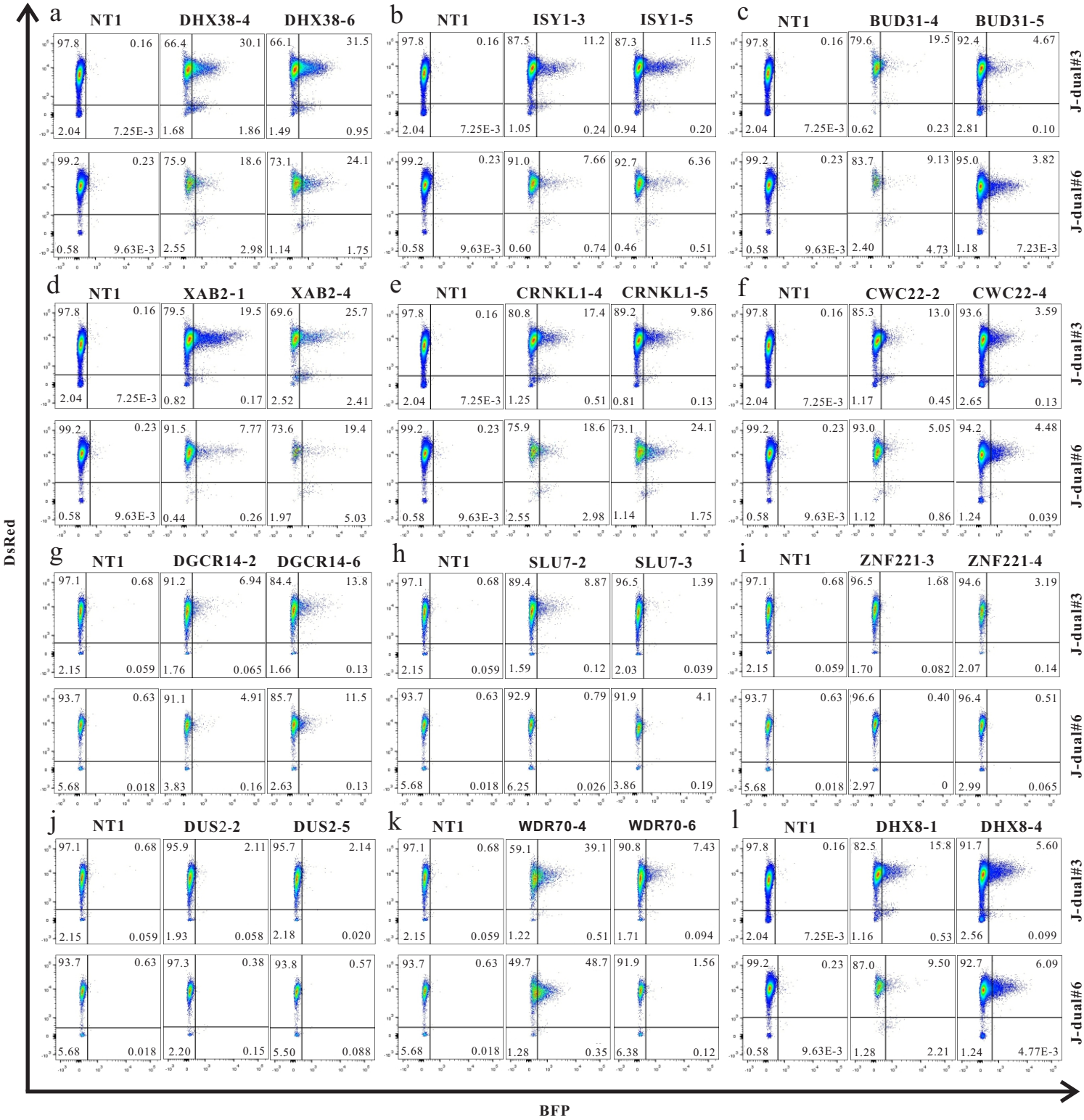

Control
J-dual#3
J-dual#6
DHX38
ISY1
BUD31
XAB2
CRNKL1
CWC22
DGCR14
SLU7
ZNF221
DUS2
WDR70
DHX8
log10 normalized counts
sgRNA-1
sgRNA-2
sgRNA-3
sgRNA-4
sgRNA-5
sgRNA-6

Figure S2

A


a NT1 DHX38-4 DHX38-6
97.8 0.16 2.04 7.25E-3
66.4 30.1 1.68 1.86
66.1 31.5 1.49 0.95
99.2 0.23 0.58 9.63E-3
75.9 18.6 2.55 2.98
73.1 24.1 1.14 1.75
b NT1 ISY1-3 ISY1-5
97.8 0.16 2.04 7.25E-3
87.5 11.2 1.05 0.24
87.3 11.5 0.94 0.20
99.2 0.23 0.58 9.63E-3
91.0 7.66 0.60 0.74
92.7 6.36 0.46 0.51
c NT1 BUD31-4 BUD31-5
97.8 0.16 2.04 7.25E-3
79.6 19.5 0.62 0.23
92.4 4.67 2.81 0.10
99.2 0.23 0.58 9.63E-3
83.7 9.13 2.40 4.73
95.0 3.82 1.18 7.23E-3
d NT1 XAB2-1 XAB2-4
97.8 0.16 2.04 7.25E-3
79.5 19.5 0.82 0.17
69.6 25.7 2.52 2.41
99.2 0.23 0.58 9.63E-3
91.5 7.77 0.44 0.26
73.6 19.4 1.97 5.03
e NT1 CRNKL1-4 CRNKL1-5
97.8 0.16 2.04 7.25E-3
80.8 17.4 1.25 0.51
89.2 9.86 0.81 0.13
99.2 0.23 0.58 9.63E-3
75.9 18.6 2.55 2.98
73.1 24.1 1.14 1.75
f NT1 CWC22-2 CWC22-4
97.8 0.16 2.04 7.25E-3
85.3 13.0 1.17 0.45
93.6 3.59 2.65 0.13
99.2 0.23 0.58 9.63E-3
93.0 5.05 1.12 0.86
94.2 4.48 1.24 0.039
g NT1 DGCR14-2 DGCR14-6
97.1 0.68 2.15 0.059
91.2 6.94 1.76 0.065
84.4 13.8 1.66 0.13
93.7 0.63 5.68 0.018
91.1 4.91 3.83 0.16
85.7 11.5 2.63 0.13
h NT1 SLU7-2 SLU7-3
97.1 0.68 2.15 0.059
89.4 8.87 1.59 0.12
96.5 1.39 2.03 0.039
93.7 0.63 5.68 0.018
92.9 0.79 6.25 0.026
91.9 4.1 3.86 0.19
i NT1 ZNF221-3 ZNF221-4
97.1 0.68 2.15 0.059
96.5 1.68 1.70 0.082
94.6 3.19 2.07 0.14
93.7 0.63 5.68 0.018
96.6 0.40 2.97 0
96.4 0.51 2.99 0.065
j NT1 DUS2-2 DUS2-5
97.1 0.68 2.15 0.059
95.9 2.11 1.93 0.058
95.7 2.14 2.18 0.020
93.7 0.63 5.68 0.018
97.3 0.38 2.20 0.15
93.8 0.57 5.50 0.088
k NT1 WDR70-4 WDR70-6
97.1 0.68 2.15 0.059
59.1 39.1 1.22 0.51
90.8 7.43 1.71 0.094
93.7 0.63 5.68 0.018
49.7 48.7 1.28 0.35
91.9 1.56 6.38 0.12
l NT1 DHX8-1 DHX8-4
97.8 0.16 2.04 7.25E-3
82.5 15.8 1.16 0.53
91.7 5.60 2.56 0.099
99.2 0.23 0.58 9.63E-3
87.0 9.50 1.28 2.21
92.7 6.09 1.24 4.77E-3
J-dual#3
J-dual#6
DsRed
BFP

B

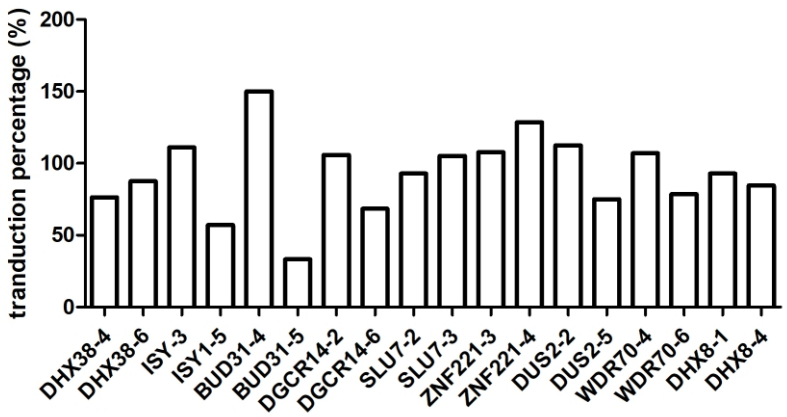

tranduction percentage (%)
200
150
100
50
0
DHX38-4
DHX38-6
ISY-3
ISY1-5
BUD31-4
BUD31-5
DGCR14-2
DGCR14-6
SLU7-2
SLU7-3
ZNF221-3
ZNF221-4
DUS2-2
DUS2-5
WDR70-4
WDR70-6
DHX8-1
DHX8-4

Figure S3

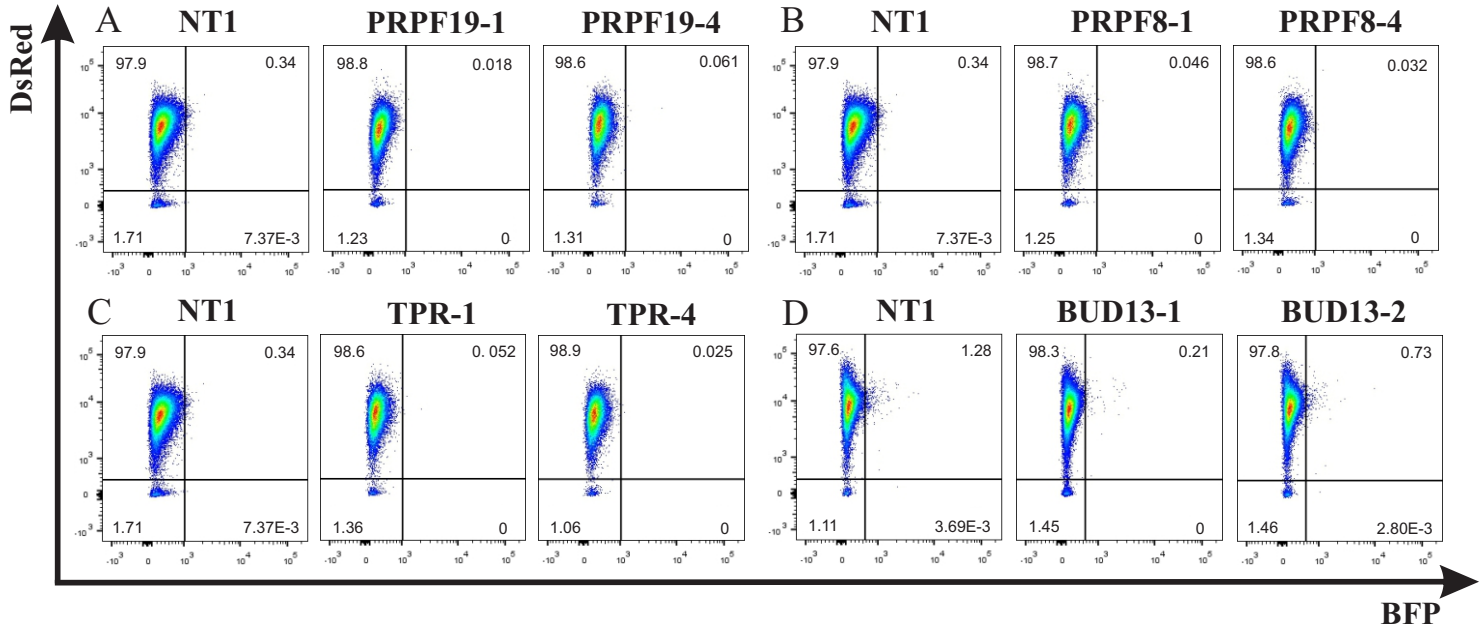


Figure S4

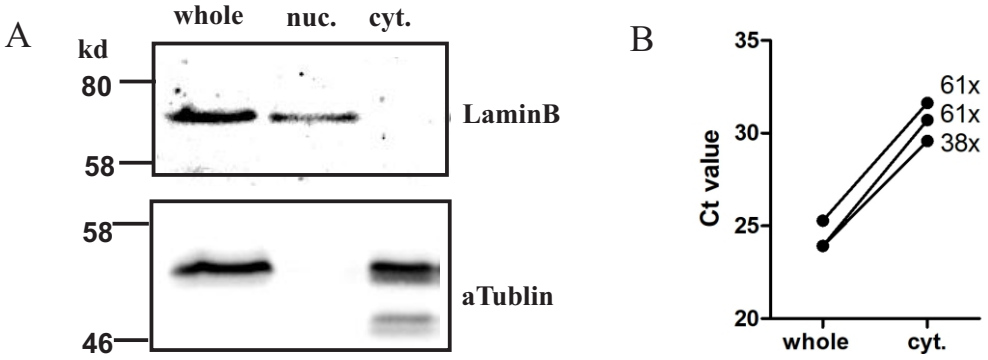


Figure S5

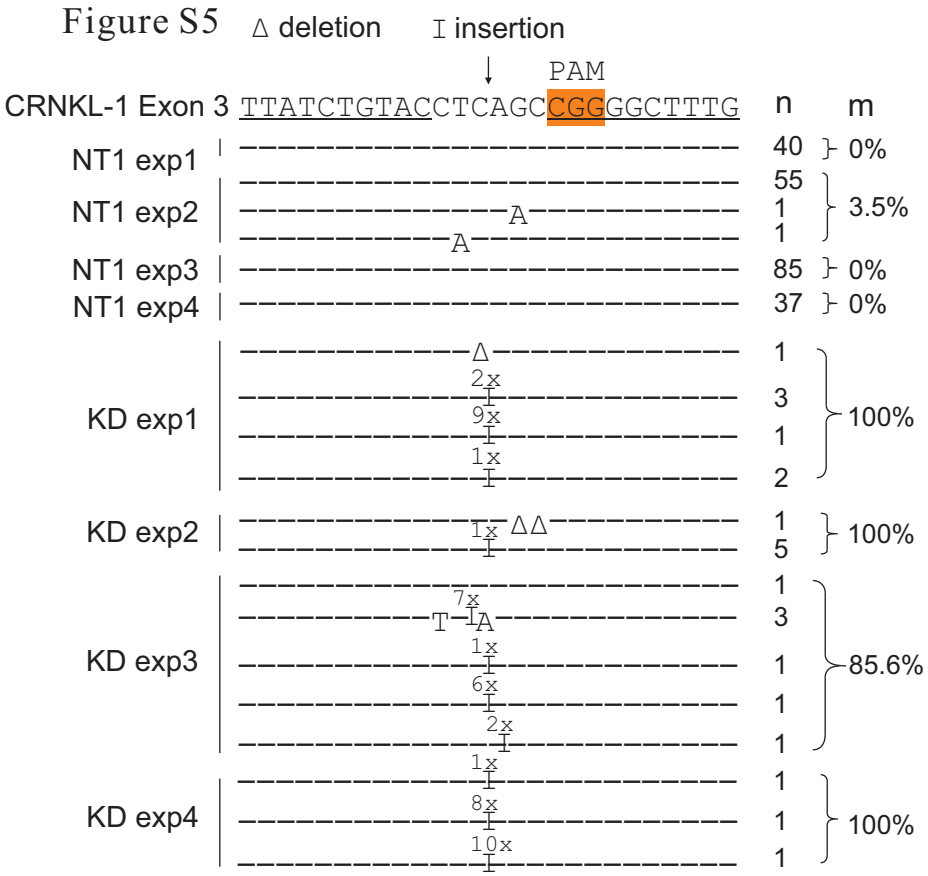


Figure S6

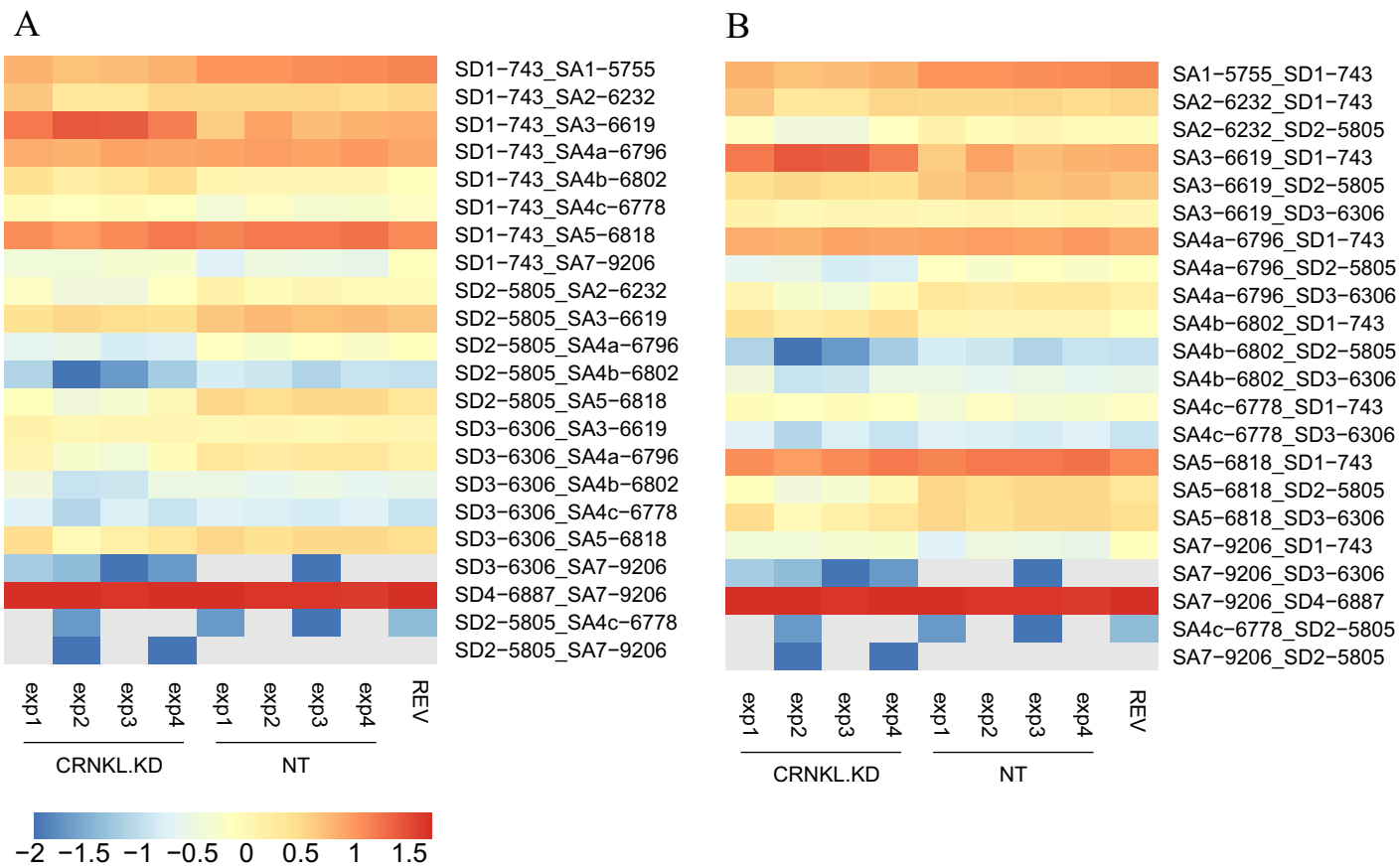


Figure S7

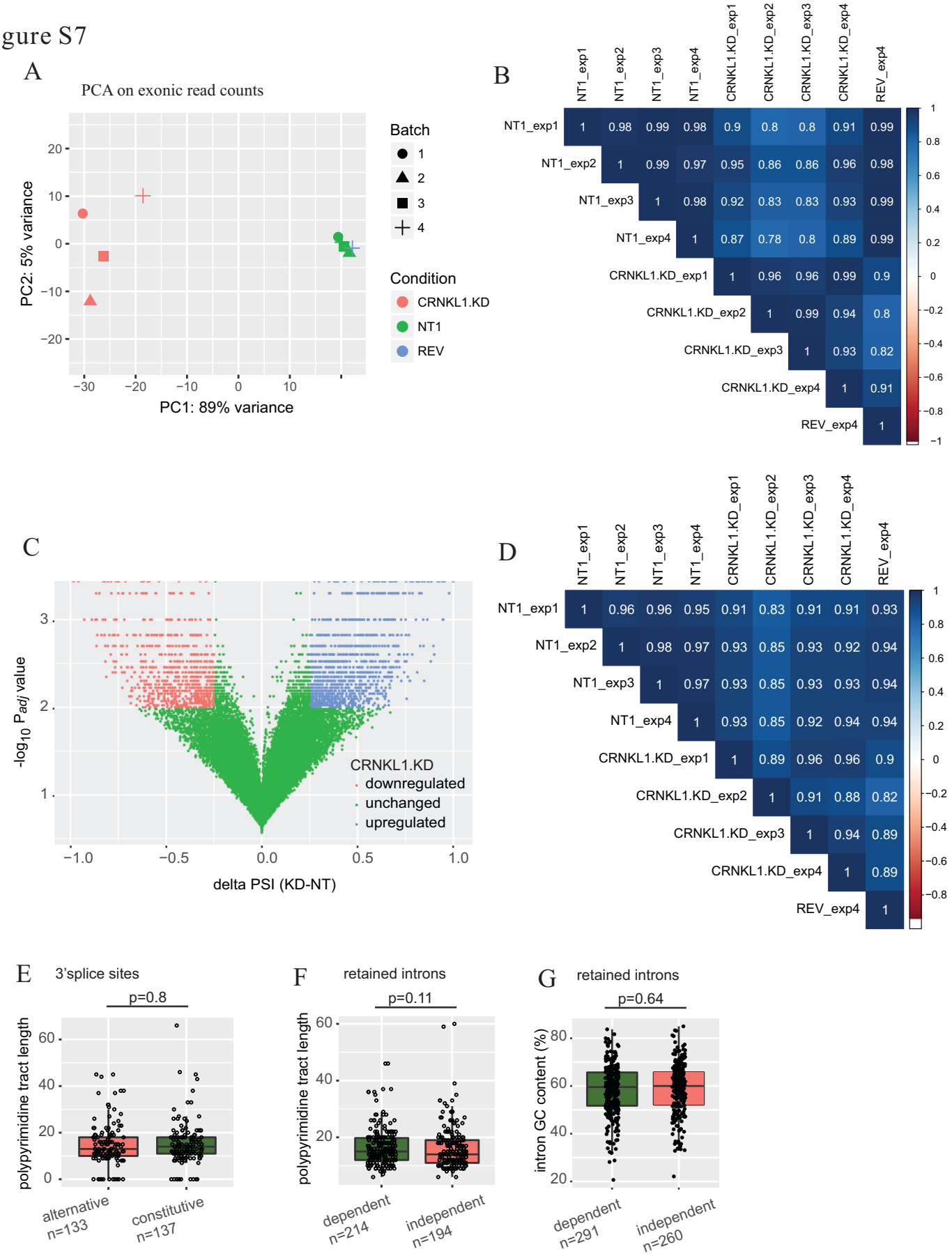

A
PCA on exonic read counts
PC2: 5% variance
PC1: 89% variance
Batch
1
2
3
4
Condition
CRNKL1.KD
NT1
REV
B
NT1_exp1 NT1_exp2 NT1_exp3 NT1_exp4 CRNKL1.KD_exp1 CRNKL1.KD_exp2 CRNKL1.KD_exp3 CRNKL1.KD_exp4 REV_exp4
NT1_exp1 1 0.98 0.99 0.98 0.9 0.8 0.8 0.91 0.99
NT1_exp2 1 0.99 0.97 0.95 0.86 0.86 0.96 0.98
NT1_exp3 1 0.98 0.92 0.83 0.83 0.93 0.99
NT1_exp4 1 0.87 0.78 0.8 0.89 0.99
CRNKL1.KD_exp1 1 0.96 0.96 0.99 0.9
CRNKL1.KD_exp2 1 0.99 0.94 0.8
CRNKL1.KD_exp3 1 0.93 0.82
CRNKL1.KD_exp4 1 0.91
REV_exp4 1
C
-log10 Padj value
delta PSI (KD-NT)
CRNKL1.KD
downregulated
unchanged
upregulated
D
NT1_exp1 NT1_exp2 NT1_exp3 NT1_exp4 CRNKL1.KD_exp1 CRNKL1.KD_exp2 CRNKL1.KD_exp3 CRNKL1.KD_exp4 REV_exp4
NT1_exp1 1 0.96 0.96 0.95 0.91 0.83 0.91 0.91 0.93
NT1_exp2 1 0.98 0.97 0.93 0.85 0.93 0.92 0.94
NT1_exp3 1 0.97 0.93 0.85 0.93 0.93 0.94
NT1_exp4 1 0.93 0.85 0.92 0.94 0.94
CRNKL1.KD_exp1 1 0.89 0.96 0.96 0.9
CRNKL1.KD_exp2 1 0.91 0.88 0.82
CRNKL1.KD_exp3 1 0.94 0.89
CRNKL1.KD_exp4 1 0.89
REV_exp4 1
E
3'splice sites
p=0.8
polypyrimidine tract length
alternative n=133
constitutive n=137
F
retained introns
p=0.11
polypyrimidine tract length
dependent n=214
independent n=194
G
retained introns
p=0.64
intron GC content (%)
dependent n=291
independent n=260

Figure S8 primer

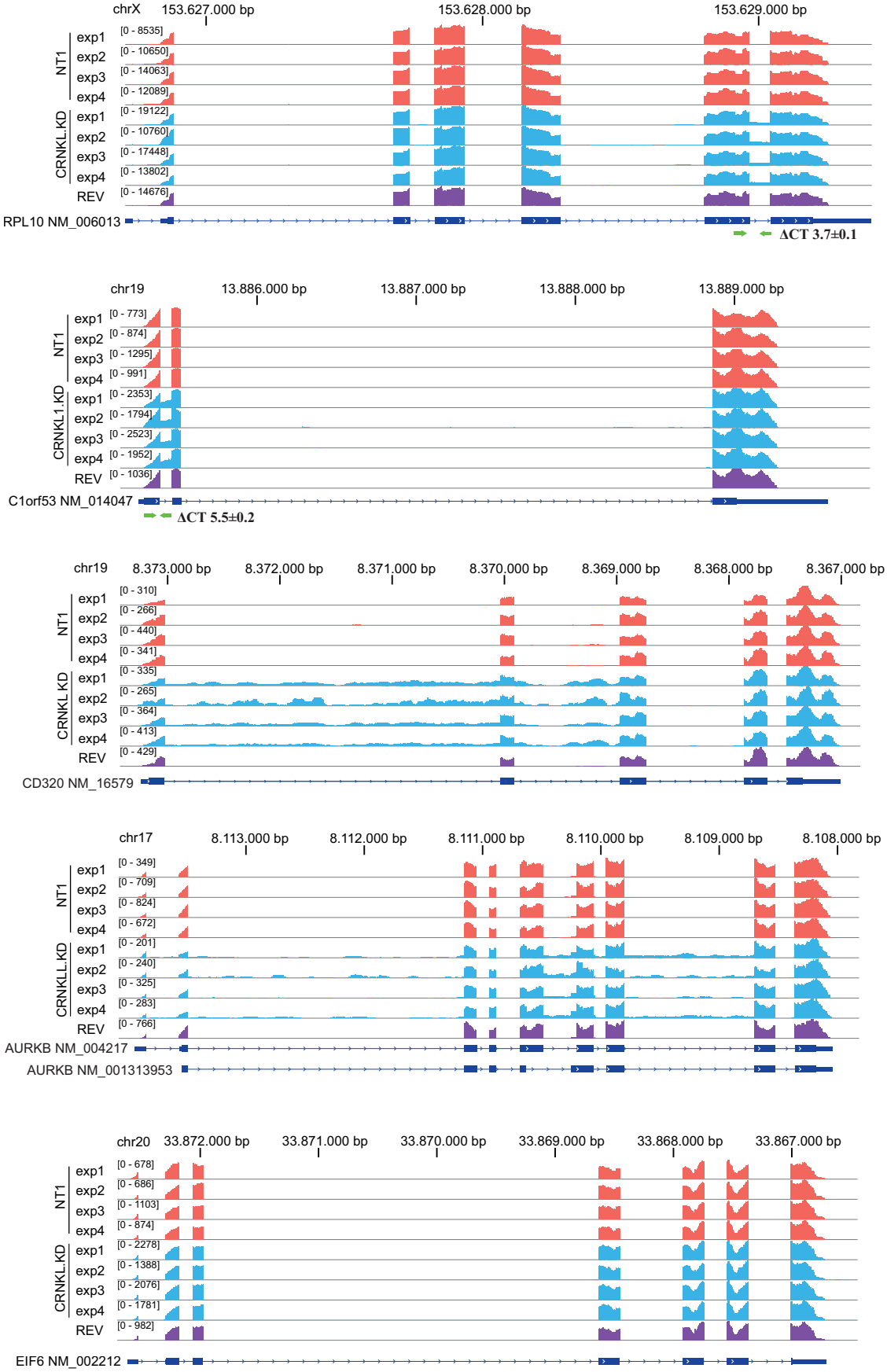


Table S1A. Sequences of sgRNAs

sgRNAs	sequences
NT1	ATCGTTTCCGCTTAACGGCG
DHX38-4	TGCCGGCTCGCCCTGACCGC
DHX38-6	CCAATGCCAACCAGCGGTCA
ISY1-5	TTCCCTCTTCCAGCTGAGCC
ISY1-3	CCAGGCCCGAAATGCAGAAA
BUD31-4	ATCATCGAGTGCACACACTG
BUD31-5	AGCTGAAACAGAACCGCATG
XAB2-1	TGACAGCTACGCCCAGTTCG
XAB2-4	GGCGCAGCAAGACGCTGTTG
CRNKL1-4	GCAGTTATCTGTACCTCAGC
CRNKL1-5	TCATCTGTGATCTTCTGTTG
CWC22-2	AAGAACAAGAACGATCCCCC
CWC22-4	CTTCTGTCATGACCAGAAGA
DGCR14-2	GGACGAGGAAGAGTATATCG
DGCR14-6	AGCCACATTTGAAACCCCTG
SLU7-2	ATAATTACTAAGTACCGCAA
SLU7-3	GCTAGAAGAACAGCGAAAAT
ZNF221-3	TCACTTGAACTGCTTCATTC
ZNF221-4	CTTCATCAACAATCACACTC
DUS2-2	CGCAGAGCGAGCCCTTGCTG
DUS2-5	CCTGTTCTGCTCTCTTTCAC
WDR70-4	TCATTACAGTATAGTAACAC
WDR70-6	GATCCCTTCAGCCCTGTGAG
DHX8-1	GCCTGGCTCCGACCCGATTA
DHX8-4	GCGAAGCCTTCCACTAGCAA
BUD13-1	GTCGCAAGCGTCGCAAAAAG
BUD13-2	GGCAGATGCCGGCGTCGACC
PRPF19-1	GAGACGGGCAATGACACGGC
PRPF19-4	TACAGTCTCTAACGAAGTGC
PRPF8-1	GATGTCGACTCTCTACCGCC
PRPF8-4	TCATAGAACCAGTCCAACAC
TPR-1	GCCGCAAGCTTTGACACTCT
TPR-4	CAAAGAACTTGAAATTGCTC

Table S1B. Sequences of NGS primers

primers	sequences
GECKO-R1v2	TGTGGGCGATGTGCGCTCTG
GECKO-F1v1v2	AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG
GECKO-R2v2-Nextera	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTCCCA CTCCTTTCAAGACCTAG
GECKO-F2-Nextera	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTTGTGG AAAGGACGAAACACCG
GECKO-F2-Nextera+2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTCTTGT GGAAAGGACGAAACACCG
GECKO-F2-Nextera+3	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCTCTTG TGGAAAGGACGAAACACCG
GECKO-F2-Nextera+6	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTAGCTC TTGTGGAAAGGACGAAACACCG
N712	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGG GCTCGG
N511	AATGATACGGCGACCACCGAGATCTACACTCTCTCCGTCGT CGGCAGCGTC
N716	CAAGCAGAAGACGGCATACGAGATTAGCGAGTGTCTCGTGG GCTCGG
N701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGG GCTCGG
N501	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGT CGGCAGCGTC

Table S1C. Primer pairs used in NGS PCRs

Step	Sample	Primers
PCR1	all samples	GECKO-R1v2, GECKO-F1v1v2
PCR2	J-dual#3	GECKO-R2v2-Nextera, GECKO-F2-Nextera+3
	J-dual#6	GECKO-R2v2-Nextera, GECKO-F2-Nextera+6
	half of the control	GECKO-R2v2-Nextera, GECKO-F2-Nextera
	half of the control	GECKO-R2v2-Nextera, GECKO-F2-Nextera+2
PCR3	J-dual#3	N712, N511
	J-dual#6	N716, N511
	Control	N701, N501

Table S1D. Sequences of RT-qPCR primers

primers	sequences
RPL10-IN6s	CAAGTTCAAGTTTCCTG
RPL10-IN6as	TAAGGAGCAATGGAAAG
C19orf53-IN1s	GGCAGCCTCTGAAAAGA
C19orf53-IN1as	CTGGGGAAAGATGTGC