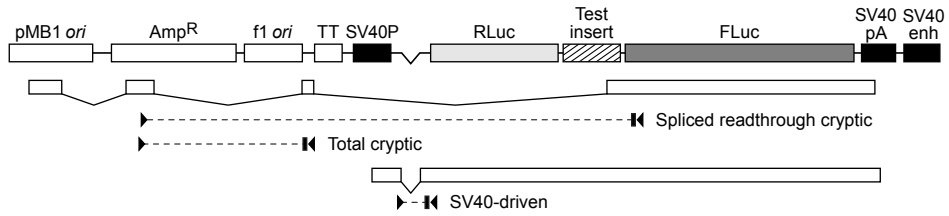
Plasmid	Spliced readthrough cryptic RNA	FLuc activity
pGL3-Basic	12	45
pGL3-Basic-RBG	313	444
pGL3-Enhancer	425	935
pGL3-Enhancer-RBG	34136	11982
pGL3-Promoter	9	1570
pGL3-Control	302	51402
pR-eIF4G-F	8185	3724
pR-XIAP-F	7114	3133
pR-E2A-F	3309	1405
pR-RBG-F	10000	4812
Promoterless pR-eIF4G-F	12351	3166
Promoterless pR-XIAP-F	10295	2989
Promoterless pR-E2A-F	4237	1384
Promoterless pR-RBG-F	10815	4706

Supplementary Table S1. Levels of spliced readthrough cryptic RNA and FLuc expression used in the regression analysis shown in Figure 6. Values are relative and the RNA level for pR-RBG-F was set to an arbitrary value of 10,000.

Plasmid or plasmid series	Supplier
pβgal series	Clontech
pBlue TOPO	Invitrogen
pCAT series	Promega
pCAT3 series	Promega
pCLuc series	New England Biolabs
pGeneBlazer	Invitrogen
pGL2 series	Promega
pGL3 series	Promega
pGL4 series	Promega
pGlow TOPO	Invitrogen
pGLuc series	New England Biolabs
pLacZ series	Clontech
pLightSwitch	SwitchGear Genomics
pRR series	Active Motif
pSEAP series	Clontech
pSEAP2 series	Clontech
pTL-Luc	Panomics

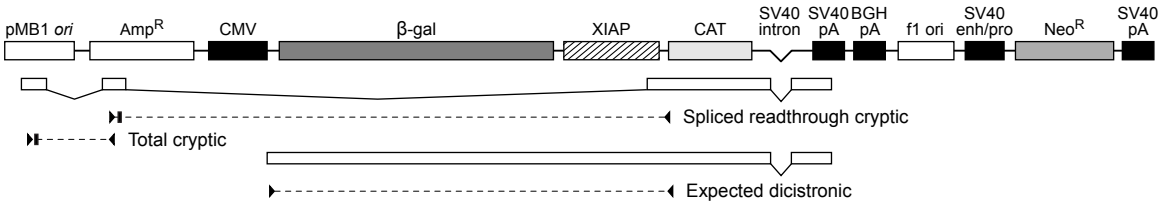
Supplementary Table S2. Commercially available reporter plasmids containing the cryptic promoter of the pMB1 *ori* in the same orientation as the reporter gene.

A



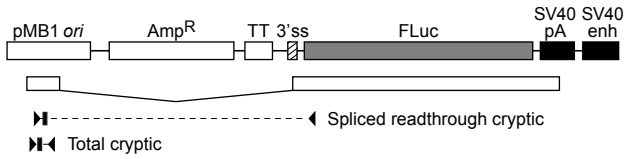
Target transcript	Primer or probe	Sequence
Spliced readthrough cryptic	Forward	TCTCAGTTCGGTGTAGGTCGTT
	Reverse	AGTTGCTCTCCAGCGGTTC
Total cryptic	Forward primer	TGCAACTTTATCCGCCTCCATC
	Probe	FAM-CGCACCGATCGCCCTTCCCAACAG-TAMRA
	Reverse primer	GGCTGGCGTAATAGCGAAGAG
Spliced readthrough cryptic	Forward primer	CGCCTCCATCCAGTCTATTAATTG
	Probe	FAM-TCCAGCGGTTCCATCTTCCAGCGG-TAMRA
	Reverse primer	GCGTATCTCTTCATAGCCTTATGC
SV40-driven	Forward primer	AGGCACTGGGCAGGTGTC
	Probe	FAM-AGTACCGGAATGCCAAGCTCTAGCC-TAMRA
	Reverse primer	CTTTCGAAGTCATGGTGGCTAGA

B



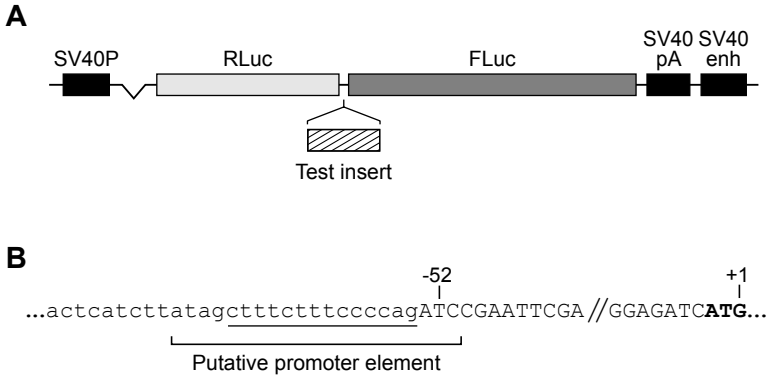
Target transcript	Primer or probe	Sequence
Intended dicistronic	Forward	CTGCTTACTGGCTTATCGAAATTA
	Reverse	CTTTACGATGCCATTGGGATA
Spliced readthrough	Forward	TCTCAGTTCGGTGTAGGTCGTT
	Reverse	CTTTACGATGCCATTGGGATA
Total cryptic	Forward primer	CGCCTTTCTCCCTTCGGGAA
	Probe	FAM-TGGCGCTTTCTCATAGCTCACGCTG-TAMRA
	Reverse primer	AGCACTGGGGCCAGATGGTA
Spliced readthrough cryptic	Forward primer	AATAAACCAGCCAGCCGGAA
	Probe	FAM-TGGTCCTGCAACTTTATCCGCCTCC-TAMRA
	Reverse primer	TCGCCAAGCTCAGATCCTCTA

C

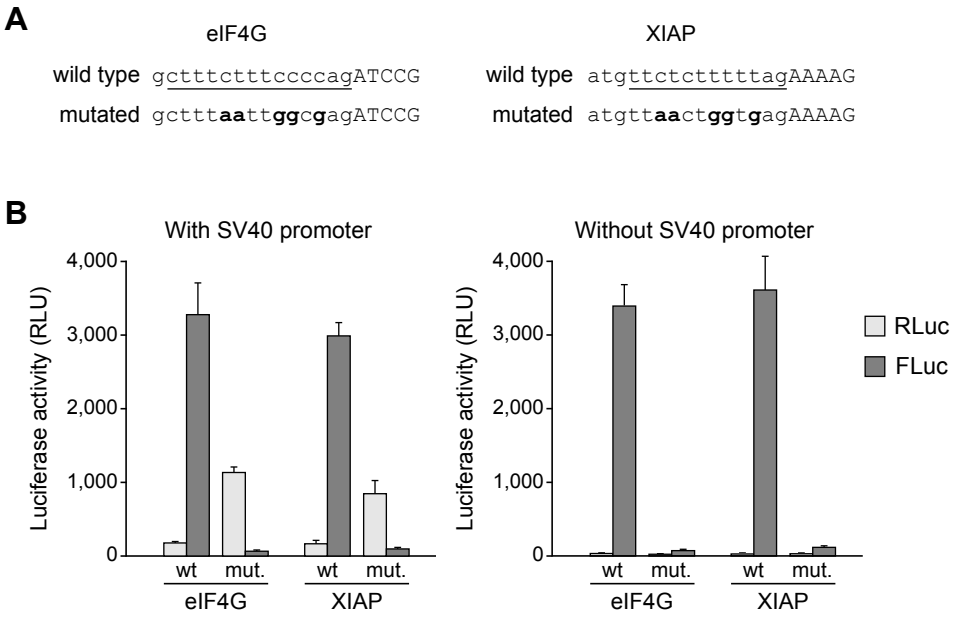


Target transcript	Primer or probe	Sequence
Total cryptic	Forward primer	CGTGCGCTCTCCTGTTCC
	Probe	FAM-CCTGCCGCTTACCGGATACCTGTCC-TAMRA
	Reverse primer	CAGCGTGAGCTATGAGAAAGC
Spliced readthrough cryptic	Forward primer	CGCCTTTCTCCCTTCGGGAA
	Probe	FAM-TGGCGCTTTCTCATAGCTCACGCTG-TAMRA
	Reverse primer	GCGCTGGGCCCTTCTTAATG

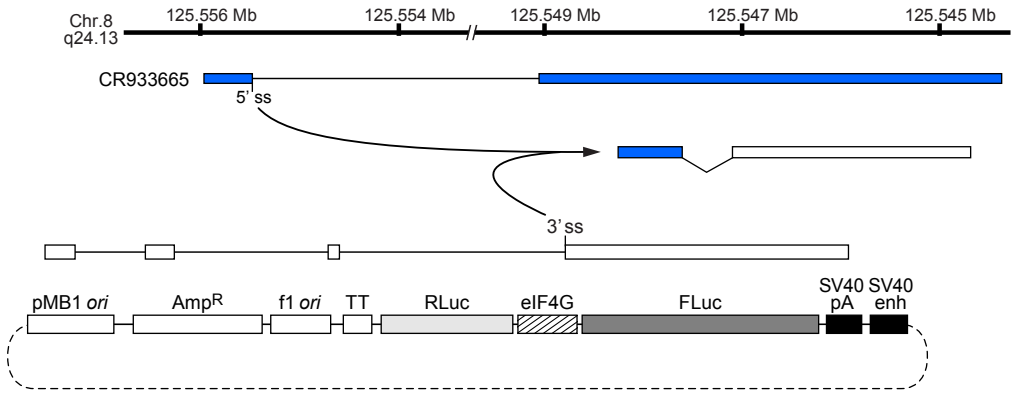
Supplementary Figure S1. Primers and probes used in this study. The binding locations of oligonucleotides on target transcripts are depicted schematically and their sequences are given in the tables. Arrowheads represent primers and rectangles represent probes. Entries in green are for primers used in standard RT-PCR and those in blue are for primers and probes used in quantitative RT-PCR. (A) Oligonucleotides for RNAs of pRF and pGL3 plasmids. The forward primer for detection of SV40-driven transcripts straddles the junction formed by splicing of the chimeric intron. (B) Oligonucleotides for RNAs of pβgal/CAT plasmids. (C) Oligonucleotides for RNAs of pGL4 plasmids. Sequences are 5' to 3'. FAM, 6-carboxyfluorescein; TAMRA, 6-carboxytetramethylrhodamine.



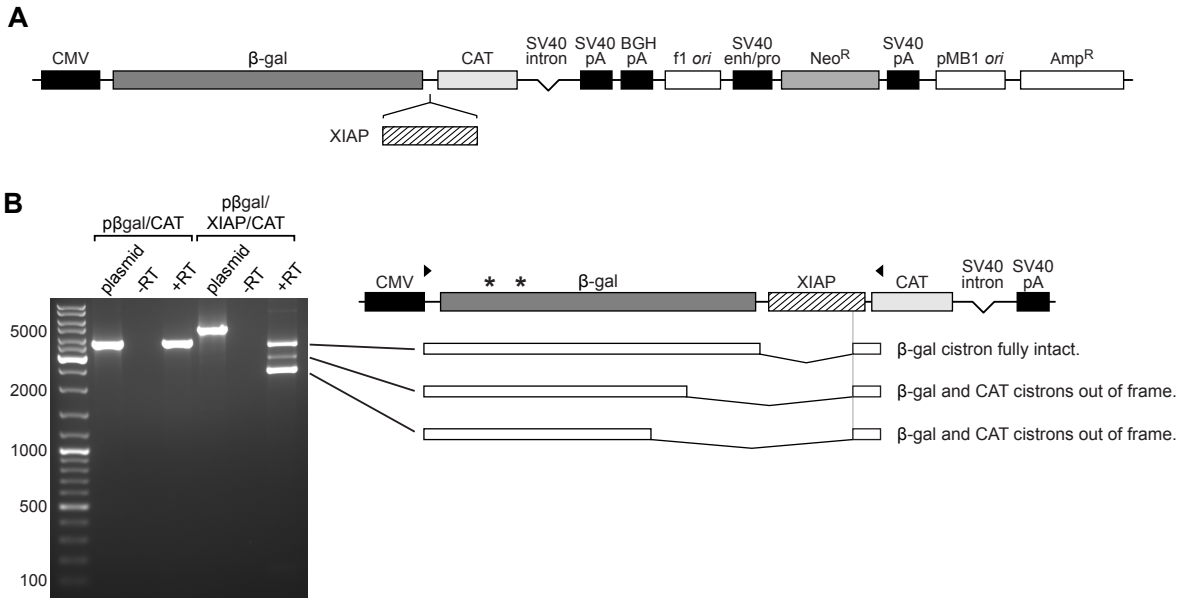
Supplementary Figure S2. Dicistronic plasmid pRF and segment of the putative eIF4G IRES reported to possess cryptic promoter function. (A) Schematic of the transcriptional unit of pRF. SV40P, proximal promoter of SV40; RLuc, *Renilla* luciferase gene; FLuc, firefly luciferase gene; pA, polyadenylation signal; enh, enhancer. As indicated, a chimeric intron is located between the SV40P and the RLuc cistron. (B) Region of the eIF4G sequence reported to contain a cryptic promoter element essential for second-cistron expression from promoterless pRF. The element, located between -69 and -49 relative to the start codon (bold) in the native eIF4G mRNA, precisely encompasses the polypyrimidine tract and AG dinucleotide (underlined) of an authentic 3' ss of eIF4G pre-mRNA. Lower and upper case nucleotides represent intronic and exonic sequences, respectively.



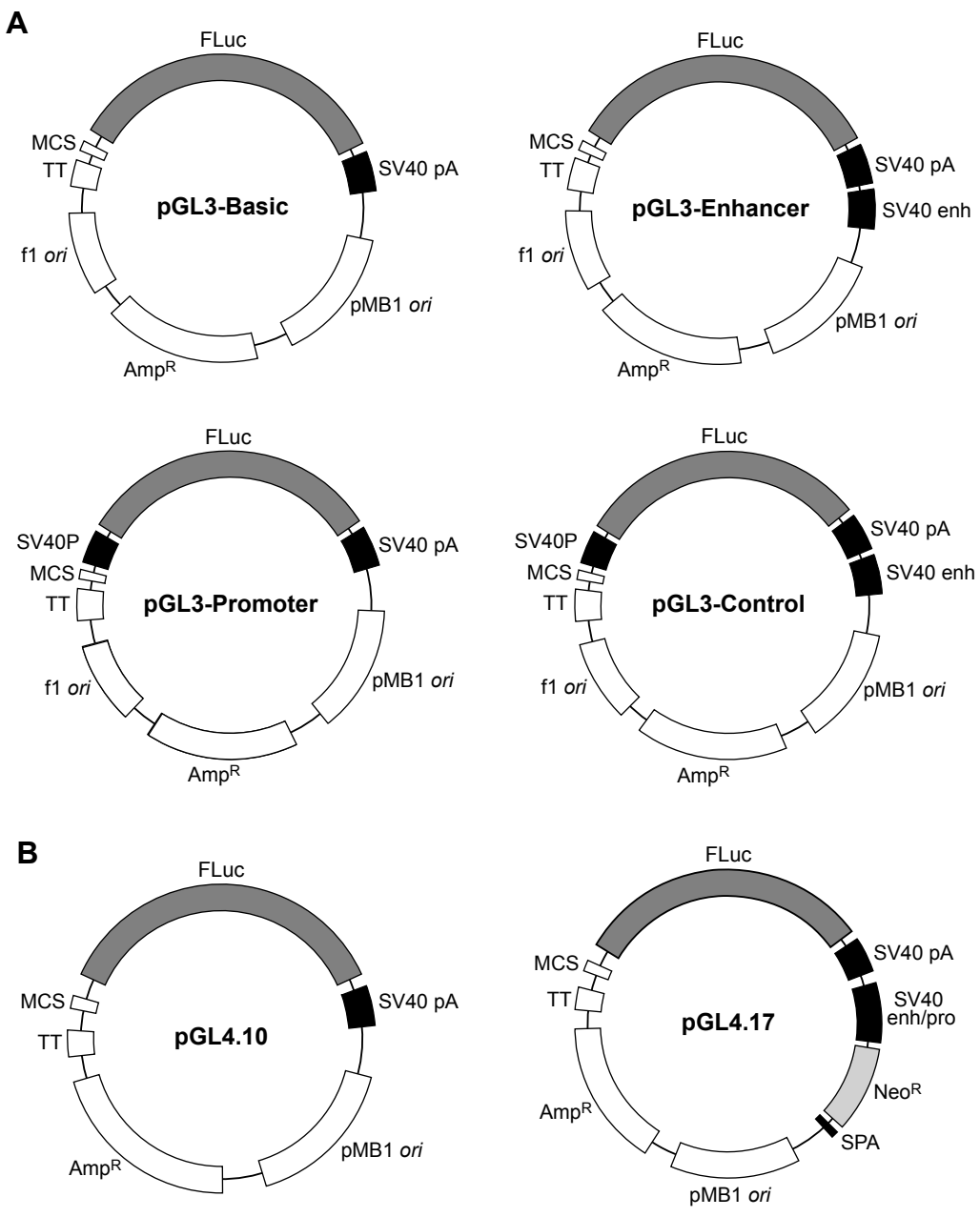
Supplementary Figure S3. Mutation of the eIF4G and XIAP sequences at their 3' ss greatly reduces second-cistron expression from promoterless and promoter-containing pRF. (A) Sequences of the wild type 3' ss and the mutated variants tested. Nucleotides in lower and upper case are intronic and exonic, respectively. The polypyrimidine tracts and AG dinucleotides of the 3' ss are underlined in the wild type sequences and the altered bases in the mutated forms are in bold. (B) Results of luciferase assays with plasmids containing the wild type and 3' ss-mutated putative IRES/promoter elements. Shown are mean values ± SD (n = 3) The increase in RLuc expression from the mutated plasmids in the left panel is likely a result of a reduction in mis-splicing of the dicistronic RNA.



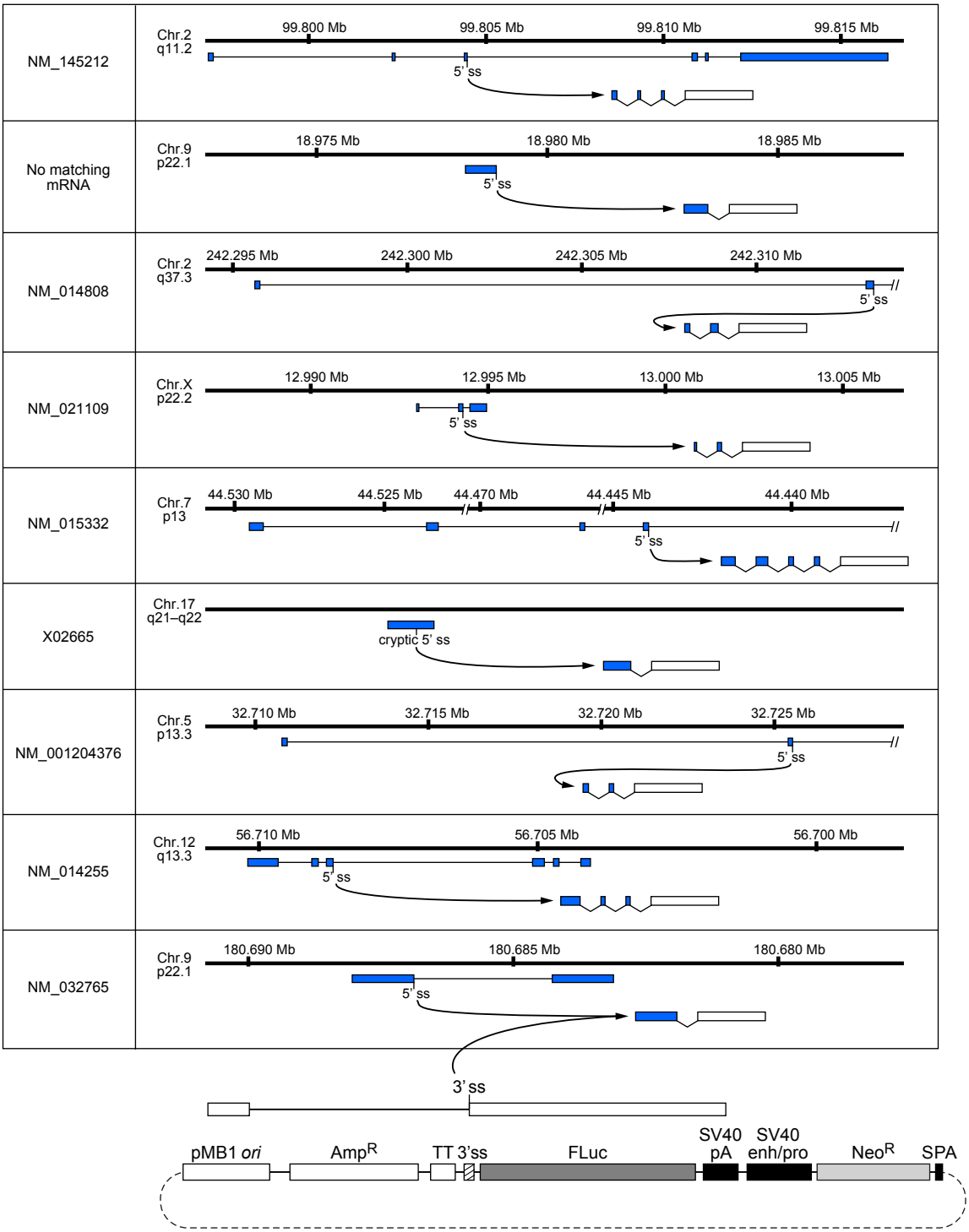
Supplementary Figure S4. *Trans*-splicing between a cryptic RNA from promoterless pR-eIF4G-F and an endogenous human RNA. Shown is a schematic illustration of the splicing event that lead to formation of the chimeric transcript. At top is the genomic locus encoding the endogenous RNA, aligned to a matching cDNA sequence in GenBank (accession CR933665). Splicing between the pre-mRNA of this cDNA and a cryptic transcript of the transfected plasmid resulted in a chimeric RNA encoding FLuc.



Supplementary Figure S5. Insertion of the XIAP sequence into pβgal/CAT causes the dicistronic transcript to mis-splice, but does not create isoforms that appear to be competent for efficient CAT translation. (A) Complete structure of pβgal/CAT, showing the insertion site of the XIAP sequence. CMV, cytomegalovirus immediate early promoter; enh/pro, enhancer and promoter; NeoR, neomycin resistance gene. (B) Results of RT-PCR analysis to detect splicing of the dicistronic transcript of pβgal/CAT, with and without the XIAP insert. Left, agarose gel of reaction products. Right, schematic detailing structure of products amplified from pβ gal/XIAP/CAT RNA as determined by sequencing. Arrowheads indicate the locations of the primers used for amplification, and the asterisks indicate the location of targets for siRNAs previously reported to knock down β-gal but not CAT expression from pβgal/XIAP/CAT.



Supplementary Figure S6. Structure of the parental pGL3 and pGL4 plasmids used in this study.

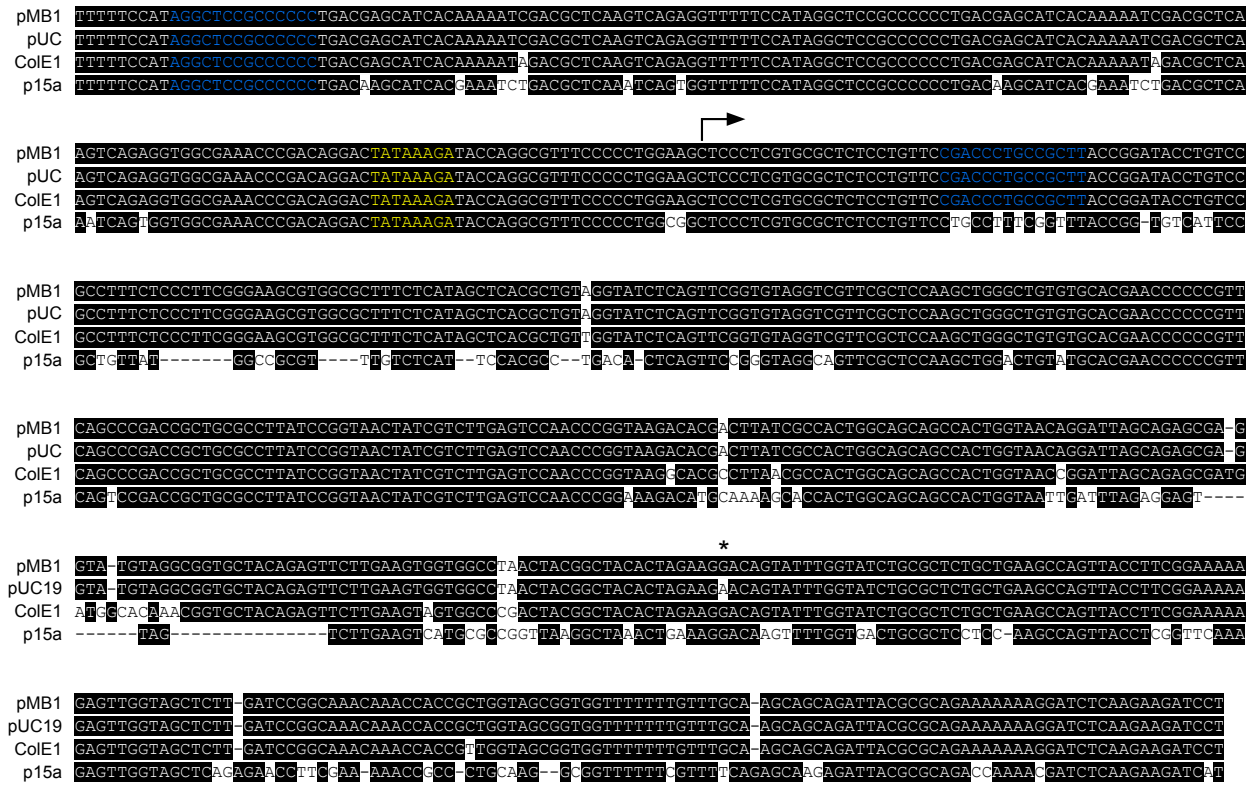


Supplementary Figure S7. Formation of chimeric RNA species by *trans*-splicing between cellular RNAs and cryptic RNA of pGL4.17 containing the globin test sequence. Shown are individual splicing events identified by 5' RACE. Each event was detected only once and all 5' ss utilized were authentic except as noted and contained canonical GT donor dinucleotides. The left column lists GenBank accession numbers corresponding to the cellular transcripts involved. The right column shows the genomic locus and exon-intron structure of the corresponding pre-mRNA (exons in blue) and the structure of the chimeric transcripts. The plasmid-derived terminal exon in the chimeric RNAs is represented by an unfilled rectangle.

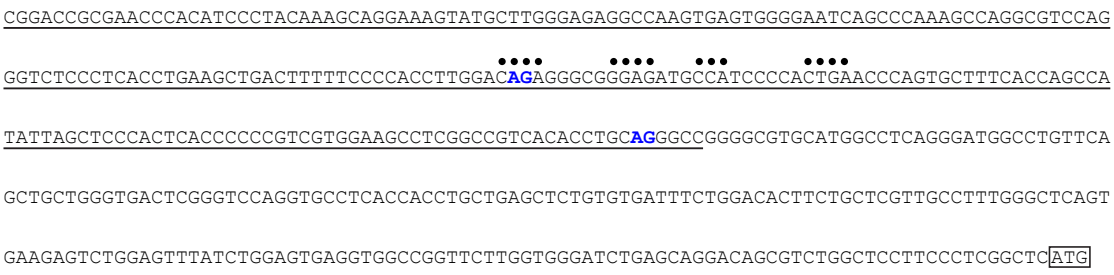
A

Replication origin	BLAST hits	Percentage of all hits
pMB1/pUC	4,801	94.9
ColE1	20	0.4
p15a	144	2.8
pSC101	96	1.9

B



Supplementary Figure S8. Relative pervasiveness and sequence of the pMB1 *ori* compared to other replication origins used in cloning vectors. (A) Comparison of the frequency of replication origins in GenBank. BLAST searches of GenBank were performed with query sequences specific for the pMB1/pUC, ColE1, p15a and pSC101 origins as detailed in Materials and Methods. (B) Alignment of the pMB1, pUC, ColE1 and p15a replication origins. The bent arrow indicates the transcription start site of the cryptic promoter. The TATA box homology is in yellow, the upstream GC box homology, and a second downstream sequence closely matching the GC box consensus are in blue. The asterisk denotes the nucleotide that distinguishes the variant of the pMB1 *ori* present in the pUC plasmids. Sequences are from accession numbers J01749 (pMB1), M77789 (pUC), V00270 (ColE1), and V00309 (p15a).



Supplementary Figure S9. Splicing of cryptic readthrough transcripts potentially explains apparent IRES and cryptic promoter activity and putative transcription start sites within the MYEOV 5' UTR. Shown is the sequence of the human MYEOV 5' UTR found to exhibit apparent IRES and promoter activity in plasmid reporter assays. Circles represent putative transcription start sites as determined previously by S1 nuclease protection analysis of RNA generated in cells transfected with pGL3-Enhancer containing the MYEOV 5' UTR. The portion of the sequence complementary to the probe used is underlined. Blue AG dinucleotides indicate the location of the two 3' ss within the UTR unanimously predicted by the splice site prediction programs NetGene2, NNSplice, MaxEntScan and HSF. If the RNA analyzed included cryptic readthrough transcript spliced at the upstream predicted ss, the S1 analysis could be expected to indicate initiation at the sites marked by the circles, with the variation in the exact initiation site being explained by "breathing" at the end of the RNA-probe duplex that can occur with this methodology.