The tumor suppressor microRNA let-7 inhibits human LINE-1 retrotransposition

Pablo Tristán-Ramos[1,2], Alejandro Rubio-Roldan[1,5], Guillermo Peris[1,3,5], Laura Sánchez[1,5], Suyapa Amador-Cubero[1], Sebastien Viollet[4], Gael Cristofari[4], Sara R. Heras[1,2,*]

[1]GENYO, Centre for Genomics and Oncological Research: Pfizer/University of Granada/Andalusian Regional Government. PTS Granada, Av. de la Ilustración, 114, 18016 Granada, Spain.

[2]Dept. Biochemistry and Molecular Biology II, Faculty of Pharmacy, University of Granada, Campus Universitario de Cartuja, 18071 Granada, Spain

[3]Dept. of Computer Languages and Systems, Universitat Jaume I, Castellón de la Plana, 12071, Spain.

[4]University Cote d'Azur, Inserm, CNRS, IRCAN, Nice, France

[5] These authors contributed equally to this work.

[*]Correspondence should be addressed to S.R.H (sara.rodriguez@genyo.es).

ABSTRACT

Nearly half of the human genome is made of transposable elements (TEs) whose activity continues to impact its structure and function. Among them, Long INterspersed Element class 1 (LINE-1 or L1) elements are the only autonomously active TEs in humans. L1s are expressed and mobilized in different cancers, generating mutagenic insertions that could affect tumor's malignancy. Tumor suppressor microRNAs are ~22nt RNAs that post-transcriptionally regulate oncogene expression and are frequently downregulated in cancer. Here we explore whether they also influence L1 mobilization. We found that downregulation of let-7 correlates with accumulation of L1 insertions in human lung cancer. Furthermore, we demonstrate that let-7 binds to the L1 mRNA and impairs the translation of the second L1-encoded protein, ORF2p, reducing its mobilization. Overall, our data uncover a new role for let-7, one of the most relevant microRNAs, which is to maintain somatic genome integrity by restricting L1 retrotransposition.

Transposable elements (TEs) account for nearly half of the human genome[1]. However, the only TE that remains autonomously active nowadays is a non-Long Terminal Repeat (LTR) retrotransposon known as Long INterspersed Element class 1 (LINE-1 or L1), whose mobilization continues to impact our genome[2]. LINE-1s comprise >20% of our DNA[3] but only about 80-100 of the ~500,000 L1 copies present in the average human genome are full-length elements that retain the ability to mobilize and are thus called Retrotransposition-Competent L1s (RC-L1s)[4]. RC-L1s belong to the human-

specific L1Hs subfamily, are 6kb long and encode two proteins (L1-ORF1p and L1-ORF2p) that are indispensable for retrotransposition[5]. However, ORF2p is expressed at a significantly lower level than ORF1p[6,7], and these differences are thought to be controlled at the level of translation[8]. L1-ORF1p is a 40kDa RNA binding protein with nucleic acid chaperone activity[9,10], whereas L1-ORF2p is a 150 kDa protein with Endonuclease (EN) and Reverse Transcriptase (RT) activities[11,12]. RC-L1s mobilize by a 'copy-and-paste' mechanism, involving reverse transcription of an RNA intermediate and insertion of its cDNA copy at a new site in the genome (reviewed in [2]). Briefly, retrotransposition starts with the transcription of a full-length RC-L1 bicistronic mRNA, which is exported to the cytoplasm and translated, giving rise to L1-ORF1p and L1-ORF2p that bind preferentially to the same L1 mRNA to form a ribonucleoparticle (RNP)[13]. The RNP gains access to the nucleus where retrotransposition occurs by a mechanism known as Target Primed Reverse Transcription (TPRT)[14,15]. During TPRT, the endonuclease activity of L1-ORF2p nicks the genomic DNA, and its reverse transcriptase activity uses the L1 mRNA as a template to generate a new copy of the element in a different genomic location. L1 can target all regions of the genome, but integration is locally dictated by the presence of a consensus sequence 5'-A/TTTT-3', which is recognized by L1 endonuclease activity and allows annealing of L1 mRNA poly(A) to the target DNA[16-18]. Other non-autonomous retrotransposons such as Alu and SINE-R/VNTR/Alu (SVA) may hijack the L1-encoded proteins and be mobilized in trans[19,20]. Furthermore, L1-

encoded proteins can sporadically generate pseudogenes using cellular mRNAs as templates [21].

TEs can affect genome stability in several ways, including the accumulation of insertions and rearrangements[2,22]. Genomic alterations caused by L1 activity have resulted in several human disorders[23]. Among these alterations, new L1 insertions can disrupt a gene unit, induce changes in splicing patterns, or interfere with transcription (reviewed in [2] and [23]). Remarkably, new L1 insertions accumulate not only during early embryogenesis and in the germ line, being transmitted to the next generation[24], but also in cancer cells, which are characterized by genome instability[25-36] (thoroughly reviewed recently in [37,38]). In fact, L1s are highly expressed and mobilized in a wide range of human epithelial cancers[26,27], and high levels of L1 mobilization are found in lung and colorectal cancers[26,32]. Interestingly, several reports have shown that somatic L1 insertions can drive tumorigenesis and may even have initiated the tumor in normal cells[26,28,30,31,39]. Transcriptional control through methylation of the L1 promoter is one of the main defence mechanisms against L1 activity[40,41], and it has been demonstrated that hypomethylation of specific RC-L1s is associated with retrotransposition in early tumorigenesis[30,32]. However, additional post-transcriptional mechanisms that silence and reactivate L1 in somatic normal and tumor cells are not completely understood yet.

MicroRNAs (miRNAs) are small RNAs that are loaded into the Argonaute (AGO) proteins to form the RNA-Induced Silencing Complex (RISC) and post-transcriptionally repress gene expression[42,43]. Hundreds of *bona fide* miRNAs exist in humans and each of them is predicted to target many mRNAs[44]. Therefore, miRNAs could be influencing essentially all human developmental, physiological and pathological processes[42]. In particular, overall miRNA dysregulation has been described in cancer[45,46]. Interestingly, it was previously shown that mouse embryonic stem cells (mESCs) lacking mature miRNAs (DGCR8 or Dicer knockout) accumulate LINE-1 mRNA[47-49]. Whereas the increase in LINE-1 mRNA levels in the absence of DGCR8 was attributed to reduced non-canonical functions of the Microprocessor, which cleaves stem-loops present in L1 elements[48], it remains possible that miRNAs regulate L1 expression levels. Consistently, a previous study reported that miR-128 represses engineered L1 retrotransposition in cultured cells[50]. Thus, we hypothesized that some miRNAs could control L1 retrotransposition and that their misexpression in tumors could contribute to increased LINE-1 mobilization.

To test this possibility, we first analyzed whole genome sequencing data from a panel of human lung tumor/normal pairs and miRNA expression data from the same tumor samples. Notably, we found that samples containing tumor-specific L1 insertions express reduced levels of several members of the tumor

suppressor miRNA let-7 family, suggesting that this miRNA could influence retrotransposition *in vivo*. Indeed, we further demonstrate that let-7 binds directly to the L1-mRNA impairing L1-ORF2p translation, and reduces L1 retrotransposition in cultured tumor cells. Altogether, our results uncover a new role for let-7 in maintaining genome integrity and provide mechanistic insight into how downregulation of let-7 miRNAs in tumors may unleash L1 activity, causing genome instability and driving tumor genome evolution.

RESULTS

Downregulation of let-7 and miRNA-34a correlates with increased somatic L1 retrotransposition in human lung tumor samples

To identify potential miRNAs whose deregulation could produce a change in L1 retrotransposition in epithelial tumors, we focused on Non Small Cell Lung Cancer samples (NSCLC) from The Cancer Genome Atlas (TCGA), as endogenous L1s are known to retrotranspose efficiently in this tumor type[26,32,36]. We selected all the samples (45 patients) for which whole genome sequencing data from tumor and matched normal lung tissue, together with tumor miRNA-seq data, were available. We computationally identified tumor-specific somatic L1 retrotransposon insertions from whole-genome sequencing data using the MELT software[51]. Briefly, MELT detects Mobile Element Insertions (MEIs) by searching for discordant reads pairs and split reads that are enriched at genome positions containing new, non-referenced insertions[51]. First, to rule out possible biases produced by different coverage or quality of sample pairs, we analyzed the polymorphic germline L1 insertions identified by MELT. We selected the samples in which the number of polymorphic L1 insertions found in tumor/normal DNA pairs was similar and at least 63% of them were common to both DNAs, a total of 41 samples (**Supplementary Table I**). After exclusion of polymorphic L1s[52], we detected 413 putative *de novo* L1 insertions specific to cancer samples, which were absent in matched normal DNA from the same

patient (**Supplementary Table I** and **Supplementary Table II**). The low number of putative *de novo* insertions found in normal tissue but not in tumor tissue (3 in the 41 samples), expected to be zero, confirmed the specificity of the method. Consistent with previous studies, 409 of the 413 tumor specific *de novo* L1 insertions identified here occurred in intronic and intergenic regions (**Supplementary Table II**), likely representing passenger mutations[37,38].

To evaluate a possible correlation between L1 retrotransposition in lung cancer and miRNA expression, tumor samples were divided into two groups based on the presence (≥1) or absence (0) of tumor-specific L1 insertions (**Fig. 1a**). Using available miRNA-seq data across these samples in TCGA, we analyzed the expression of 26 miRNAs that have been previously associated with the development and/or progression of lung cancer, such as the let-7 family, the miR-34 family, or the miR-17-92 cluster[53] (**Fig. 1b**). Interestingly, we found that several members of the tumor suppressor let-7 family (let-7a, let-7e and let-7f) were significantly downregulated in the samples with ≥1 tumor-specific L1 insertions upon multiple t-testing adjusted with FDR<0.01 (**Fig. 1b and Supplementary Table III**). This correlation was also found for let-7a and let-7f using a different statistical analysis (Rank-sum test, **Supplementary Table IV**). Although all the members of the let-7 family have a similar mature sequence, and could potentially bind to the same RNA targets, their genomic location and timing of expression is markedly different[54]. Interestingly, reduced expression

of let-7a and let-7f has been observed in lung cancer samples[55,56]. Additionally, miR-34a, another tumor suppressor miRNA[57], was also significantly reduced in samples with tumor-specific L1 insertions (**Fig. 1b and Suppl. Table III**). As a control, analysis was repeated after L1 insertion counts were randomly reassigned to each sample. No significant correlation was found in any case (one example is shown in **Supplementary Fig. 1a and Supplementary Table V**). Notably, the differential expression of let-7a, let-7e, let-7f and miR-34a was also significant in a more restrictive analysis where all the miRNAs expressed in lung tumor samples (89 miRNAs) were considered (**Supplementary Table VI).** Thus, even though we cannot rule out a possible bias in the analysis due to sample variability and the limited number of cases available, these data suggest that let-7 and miR-34a might control the accumulation of new L1 insertions in human lung cancer samples. Next, we used SQuIRE (Software for Quantifying Interspersed Repeat Elements)[58] to quantify L1Hs expression in RNA-seq data from these tumor samples, available in TCGA. As expected, L1Hs RNA levels were significantly increased in samples with tumor-specific L1 insertions (**Supplementary Fig. 1b**). However, L1Hs expression negatively correlates with miR-34a but not with let-7 expression (**Supplementary Fig. 1b**).

To further corroborate our results, we analyzed the correlation between miRNA expression and the number of tumor-specific L1 insertions identified by Helman and collaborators in a group of 46 lung tumor samples using Transpo-

seq framework[26] (13 of them were also included in the previous analysis using MELT). Remarkably, the expression levels of let-7 family members (let-7a and let-7e) and miR-34a were again significantly reduced in those tumors containing tumor-specific L1 insertions when the 26 miRNAs related to lung cancer were analyzed (**Supplementary Figure 1c** and **Supplementary Table VII)** as well as when all the miRNAs expressed in lung were included (**Supplementary Table VIII**). Notably, the same analysis with the number of insertions randomly reassigned to each sample did not show any significant correlation with miRNA expression (**Supplementary Figure 1d** and **Supplementary Table IX**). Lastly, the same analysis was performed using 36 breast cancer samples which contain a notably smaller number of tumor-specific L1 insertions per sample as determined by Transpo-Seq [26]. No significant correlation was found for any of the 26 miRNAs related to lung cancer (**Supplementary Figure 1e and Supplementary Table X**) suggesting that the contribution of let-7 and mir-34a to L1 mobilization could be specific to some tumor types.

Overall, these results suggest that a downregulation of let-7 and/or miR-34 expression can influence the accumulation of tumor-specific L1 insertions in lung cancer.

Let-7 negatively regulates human LINE-1 retrotransposition *in vitro*

To investigate whether there is a causal relationship between the variation in let-7 and miR-34 expression levels and the accumulation of L1 insertions in

198 tumors, we tested the effect of these miRNAs on L1 mobilization using the
199 sRNA/L1 retrotransposition assay, recently developed in our lab[59]. This protocol
200 combines the previously described cell culture-based LINE-1 retrotransposition
201 reporter assay (reviewed in [2]) with microRNA mimics or inhibitors. Briefly, in
202 this assay, cells are transfected with a plasmid containing an RC-L1 tagged with
203 a reporter cassette (**Fig. 2a**). This cassette consists of a reporter gene (REP) in
204 antisense orientation relative to the L1, equipped with its own promoter and
205 polyadenylation signal, but interrupted by an intron located in the same
206 transcriptional orientation as L1. Thus, a functional reporter can only be
207 produced after a successful round of retrotransposition (**Fig. 2a**). For this assay,
208 we used cultured HeLa cells which express high levels of let-7a and almost
209 undetectable levels of miR-34a as analyzed by RT-qPCR (**Supplementary Fig.**
210 **2a**). We next analyzed L1 activity upon overexpressing let-7a and miR-34a,
211 using transfected synthetic miRNA mimics and a neomycin-resistance based
212 retrotransposition assay (using plasmid JM101/L1.3, **Fig. 2b** and Methods). As a
213 control, we performed a clonability assay co-transfecting the miRNA mimics
214 with a plasmid encoding a constitutively-expressed neomycin resistance gene
215 (pU6iNeo) to rule out possible effects of miRNA overexpression on cell growth
216 (**Fig. 2b**). In agreement with the above observation in lung tumor samples, we
217 reproducibly detected a significant decrease in L1 retrotransposition upon
218 overexpression of let-7a in HeLa cells without affecting the clonability of the
219 cells (**Fig. 2b**, top panel). Furthermore, as expected, overexpression of different

let-7 family members also reduced L1 retrotransposition (e.g. let-7b in **Supplementary Fig. 2b).** Strikingly, overexpression of miR-34 did not affect L1 mobilization or cell clonability in this assay (**Fig. 2b**, bottom panel). Similarly, let-7 overexpression inhibits L1 retrotransposition in HEK293T cells, which express lower endogenous levels of this miRNA as compared to HeLa cells (**Supplementary Fig. 2a**). We used a dual-luciferase reporter vector containing a different RC-L1, L1RP (pYX014, **Supplementary Fig. 2c**). This plasmid uses Firefly luciferase as retrotransposition indicator and encodes a Renilla luciferase in the backbone to normalize for transfection efficiency (**Supplementary Fig. 2c**). Notably, we observed a consistent decrease in L1 retrotransposition upon co-transfection of the let-7 mimic in HEK293T cells [measured as the relative luminescence ratio (L1-Fluc/Rluc)] (**Supplementary Fig. 2c**). As expected, an inactive L1RP containing two missense mutations in the ORF1-encoded protein did not show luciferase activity (plasmid pYX15, **Supplementary Fig. 2c**). Considering that miRNAs downregulate the expression of their targets, the decrease of L1 mobilization upon let-7 overexpression suggests that L1 mRNA could be a *bona fide* let-7 target. Conversely, miR-34 overexpression in HEK293T cells, where the endogenous levels are slightly higher than in HeLa cells (**Supplementary Fig. 2a**), led to an increase in L1 retrotransposition using the dual luciferase reporter vector pYX014 (**Supplementary Fig. 2d**). The different effects observed for miR-34 overexpression in HeLa (**Fig. 2b**) and HEK293T

cells (**Supplementary Fig. 2d**) suggest a potential indirect and cell-type specific effect of miR-34 on L1 mobilization.

To further investigate the role of let-7 on the control of L1 mobilization, we performed another panel of cell culture-based retrotransposition assays using a hairpin inhibitor to decrease intracellular let-7 levels. Although the inhibitor used was designed against let-7a, it has been shown to cross-react with other members of the family[60]. Consistent with our previous results, we found that depletion of let-7 in HeLa cells led to a two-fold increase in L1 retrotransposition without affecting the clonability of the cells using the neomycin-resistance cassette described above (**Fig. 2c**). A similar increase in L1 retrotransposition was observed in HEK293T cells upon let-7 depletion using an EGFP-based reporter cassette and a different human RC-L1, LRE3 (plasmid 99-UB-LRE3, **Supplementary Fig. 2e**). Furthermore, we confirmed that let-7 knock-down increased L1 retrotransposition in HEK293T using the luciferase reporter vectors pYX014 and pYX017 (**Supplementary Fig. 2f**). While both contain the same active human L1, L1RP, in pYX014 it is transcribed from the native promoter in the 5′UTR whereas in pYX017, it is highly transcribed from a CAG promoter.

Lastly, since our bioinformatic analysis showed an inverse correlation between let-7 expression and accumulation of L1 insertions in human lung tumor samples, we analyzed whether let-7 could regulate L1 retrotransposition in lung

cancer cells. To do that, we performed the luciferase-based retrotransposition assay in two lung cancer cell lines with significantly different endogenous levels of let-7, A549 and SK-MES-1 (**Supplementary Fig. 2g**). Interestingly, we observed that, in both cell lines, depletion of let-7 increased L1 retrotransposition by 2.5 times on average (**Fig. 2d**). Altogether, these data indicate that let-7 negatively regulates human L1 mobilization in a variety of cancer cell lines.

Let-7 binds directly to the coding sequence of L1 mRNA

The aforementioned regulation could occur either by a direct interaction between let-7-guided RISC and L1 mRNAs, or by an indirect effect, since let-7 could be regulating any host factor involved in the multiple steps of the retrotransposition cycle[61] or in L1 control[62]. Since a direct effect would be sequence-dependent, we performed a neomycin-resistance based retrotransposition assays in HeLa cells using non-human active LINEs, that differ in sequence from the human L1 but use the same target-primed reverse transcription mechanism for mobilization. Briefly, we used mouse TGF21 (L1GF subfamily) and zebrafish L2-1 and L2-2 (L2 clade). Structures of the different LINEs are shown in the left panel of **Fig. 3a**, and constructs are described in the Methods section. Interestingly, we observed that only human L1 mobilization was significantly affected by either the inhibition (**Fig. 3a**) or the overexpression

(**Supplementary Fig. 3a**) of let-7. These results suggested a direct, sequence-dependent interaction between let-7 and human L1 mRNA.

It is well established that miRNAs mostly bind their target mRNAs in their 3'UTRs[42], although 5'UTR and coding sequence binding sites have been described and validated[63-65]. Thus, to find out where the putative let-7 binding site was located in L1 mRNA, we performed the same retrotransposition assays but using an engineered human RC-L1 (L1.3) lacking either the 5´or the 3´UTR (**Fig. 3b** and **Supplementary Fig. 3b**). Notably, the effect of let-7 depletion or overexpression in engineered L1 mobilization was not abolished or reduced by the absence of either 5' or 3' UTR, suggesting that let-7 interacts with the coding sequence of human L1 mRNA (**Fig. 3b** and **Supplementary Fig. 3b**).

As a proof of concept, we investigated whether L1 mRNAs were bound directly by different human Argonaute (AGO) proteins, the main components of the RISC complex: AGO2, the only one with cleavage activity, and AGO1, specifically associated to miRNA function[66]. Interestingly, we observed by co-immunoprecipitation in HEK 293T cells that overexpressed FLAG-tagged AGO2 and AGO1 proteins interact with endogenous L1-ORF1p in an RNA-dependent manner (**Supplementary Fig. 3c**). This result suggests that L1 mRNA could be recognized by miRNAs and bound directly by Argonaute proteins. Consistently, it has previously been described that L1-ORF1p often

aggregates in cytoplasmic foci and colocalizes with L1 mRNA and AGO2 protein[67,68].

We further analyzed whether let-7 guides the RISC complex to L1 mRNAs by RNA-Immunoprecipitation (RIP) assay. For this, we used a human embryonic teratocarcinoma cell line (PA-1), characterized by high levels of endogenous LINE-1 mRNA and L1-ORF1p [69] and very low levels of let-7 miRNAs (**Supplementary Fig. 2g**). Briefly, we overexpressed FLAG-tagged AGO2, pulled it down, purified the endogenous bound RNAs, and analyzed them by RT-qPCR (**Fig. 3c**). We reasoned that if let-7 can bind L1 mRNA, let-7 overexpression should lead to an increase in the abundance of endogenous L1 mRNAs associated to AGO2. Strikingly, we observed an enrichment in the amount of L1 mRNA bound to AGO2 upon overexpression of let-7 resembling the behaviour of HMGA2 mRNA (**Fig. 3d**), a well-known target of let-7[70]. In contrast, none of the negative controls used, GAPDH and actin mRNAs, were enriched in the immunoprecipitation (**Fig. 3d**). Thus, these data suggest that let-7 guides Argonaute proteins to L1 mRNA, and that this interaction occurs within the L1 coding sequence.

A functional let-7 binding site is located in L1-ORF2

We next set out to predict and validate putative let-7 binding sites within the coding sequence of the L1 mRNA. We used two different software available online: miRanda[71] and RNA22[72]. The best predicted binding site for let-7 family

by each method were located in positions 2650-2671 (bs1) and 4596-4616 (bs2), respectively, in the consensus L1Hs sequence (top panel, **Supplementary Fig. 4a)**. In order to validate them, five tandem copies of each binding site (bs) were cloned in the 3'UTR region of the Renilla luciferase (Rluc) gene in the psiCHECK-2 vector, which also encodes a Firefly luciferase (Fluc) gene to correct for transfection efficiency (left panel, **Fig. 4b**). As controls, an unrelated sequence of the same length and a sequence with perfect complementarity to let-7 were cloned (no bs and perfect bs, respectively). Those constructs were co-transfected with let-7 mimic in HEK293T cells. The reporter constructs containing the RNA22-predicted binding site (bs2) and the positive control (perfect bs), but not the one with the miRanda-predicted binding site (bs1) or the negative control (no bs), showed a reduction of the relative luciferase ratio (RLuc/FLuc) upon let-7 overexpression (bottom panel, **Supplementary Fig.4a**). A deeper analysis of the residues in this region interacting with let-7 microRNA using RNAhybrids software[73] suggests that the functional 'bs2' is located within the coding sequence of L1-ORF2 (position 4587-4610 in L1.3, a commonly used human RC-L1, accession code # L19088.1[74]), between the RT and Cysteine-rich domains of this protein (**Fig. 4a, left panel**). Importantly, it is predicted to form a duplex with let-7 miRNA consisting of seven Watson–Crick pairings at positions 3-9 followed by an adenine at the mRNA nucleotide corresponding to the first nucleotide position of the miRNA, resembling a previously described functional noncanonical binding site termed offset 7-mer

(**Fig. 4a**) [75]. Altogether, these results suggest that this refined binding site, hereafter referred to as 'bs2rh' is a *bona fide* let-7 binding site. To further validate this binding site ('bs2rh') we generated a mutant sequence ('bs2rhmut', see **Fig. 4b,** left panel). Mutations introduced in 'bs2rh' are predicted to severely impede the duplex formation between L1 mRNA and let-7 (**Supplementary Fig. 4b**). Accordingly, the mutated sequence rescued the luciferase activity upon overexpression of let-7 (bs2rh vs bs2rhmut, **Fig. 4b**).

To further corroborate the functionality of ´bs2rh´ in the context of retrotransposition, we generated an allele mutated RC-L1 construct containing a mutated ´bs2rh´site (we introduced the same mutation described above, construct JM101/bs2rhmutL1.3, **Fig. 4c**). Intriguingly, the validated ´bs2rh´site is conserved through primate L1 evolution, being present in L1PA5 elements and containing a few mutations in older L1 subfamilies (**Supplementary Fig. 4c**). Accordingly, the introduction of the mutation contained in the 'bs2rhmut' sequence, which entails an amino acid change (P to G) in L1-ORF2p, leads to a reduction in RC-L1 mobility (**Fig. 4d**, right side graph). We observed that 'bs2rhmut' L1.3 retrotransposition was less affected by let-7 inhibition than wild-type L1.3 (**Fig. 4d**). Interestingly, this binding site is absent in zebrafish LINEs and relatively low conserved in mouse RC-L1s (**Supplementary Fig. 4d**) in agreement with the specific let-7 effect on human L1 retrotransposition showed above (**Fig. 3a** and **Supplementary Fig. 3a)**. However, the fact that

366 mutating this binding site reduced but not abolished the effect of let-7 in L1
367 mobilization suggest that additional mechanisms mediated by let-7 may work
368 to restrict human L1 retrotransposition. Overall, these results suggest that there
369 is at least a functional let-7 binding site in the ORF2 region of human L1 mRNA.

370 **Let-7 impairs L1-ORF2p translation**

371 The above experiments identified a functional let-7 binding site in L1-ORF2p,
372 and we next analyzed the functional consequences of let-7 binding to L1
373 mRNA. Since miRNAs can induce mRNA degradation[43], we analyzed the levels
374 of endogenous L1 mRNAs upon let-7 overexpression in HEK293T cells by RT-
375 qPCR. We found no significant changes in L1 mRNA levels at 24 and 48 hours
376 after let-7 overexpression, whereas those of other canonical let-7 targets (DICER
377 and HMGA2) were significantly reduced (**Fig. 5a**). Similarly, L1 mRNA levels
378 were not decreased upon let-7 overexpression (**Supplementary Fig. 5a**) or
379 increased upon let-7 depletion (**Supplementary Fig. 5b**) when L1 was
380 overexpressed in HEK293T cells. Thus, these data suggest that let-7 expression
381 does not trigger L1 mRNA degradation.

382 The other main effect of miRNAs on their target mRNAs is interference with
383 protein translation[76], so we analyzed the levels of endogenous L1-ORF1p upon
384 modulation of let-7 levels in HEK293T cells. We found significant changes in
385 HMGA2 but not in ORF1p expression upon let-7 overexpression (**Fig. 5b**) or
386 depletion (**Supplementary Fig. 5c**). We corroborated this results in a stable

HEK293T cell line that constitutively overexpresses a T7-tagged L1-ORF1p (from L1.3, and using a CMV promoter) (**Supplementary Fig. 5d**).

We next analyzed changes in L1-ORF2p levels. The translation of ORF2p occurs by a highly inefficient unconventional termination/reinitiation mechanism that, although could produce as few as one L1-ORF2p molecule per L1 mRNA[8], is enough to support efficient retrotransposition[5,8]. Consequently, it is technically challenging to detect endogenous L1-ORF2p. Thus, to study L1-ORF2p levels upon let-7 modulation, we generated a monocistronic construct expressing 3xFLAG-tagged ORF2p from a CMV promoter (L1-ORF2p from L1.3), pSA500. Strikingly, we observed an increase in ORF2p upon let-7 depletion and a decrease upon let-7 overexpression in HeLa cells (**Fig. 5c**) resembling the effect on DICER protein levels, a well described target of let-7 with several 8-mer sites[63] (**Fig. 5c**). To rule out that differences in L1-ORF2p expression were due to different transfection efficiencies, we took a fraction of each sample, extracted DNA, and quantified plasmid levels by qPCR using primers targeting the CMV promoter driving ORF2 expression or the EBNA-1 sequence in the plasmid backbone. We did not observe any significant differences in the amount of plasmid co-transfected with let-7 mimic (**Supplementary Fig. 5e**) or let-7 inhibitor (**Supplementary Fig. 5f**). Consistent with the data presented above, the difference at protein level neither correlates with changes in the levels of exogenous L1 ORF2-FLAG RNA (L1-ORF2-F RNA), as opposed to DICER

whose mRNA is also reduced (bottom panel, **Fig. 5c**). These data suggest that the differences in ORF2p levels are not due to variations in transfection or mRNA accumulation but to an effect of let-7 on ORF2-F translation.

To understand whether let-7 mediated translational repression of ORF2p is due to the specific interactions with the offset 7-mer site or to its location within the coding sequence, we generated three variants of pSA500 in which we introduced different sequences in its 3'UTR : a scrambled sequence ('scrb'), the binding site ('bs2rh') and a modified bs2 that contains a canonical 8-mer site for let-7 ('8mer') (**Supplementary Fig. 5g**). We co-transfected all these constructs in HeLa cells with let-7 mimic. First, by RT-qPCR we observed that similar levels of transfection (measuring constitutive EBNA expression from the plasmid backbone, **Supplementary Fig. 5h**) and let-7 overexpression (measuring the effect on endogenous DICER, **Supplementary Fig. 5h**) were achieved. The levels of ORF2 mRNA were not significantly affected in any case, although we observed a tendency towards a reduction on the RNA levels upon placement of the binding site ('bs2rh`) or the modified 8-mer binding site ('8mer') in the 3'UTR (**Supplementary Fig. 5h**). Furthermore, western blot analysis showed that placement of 'bs2rh' sequence in the 3'UTR of pSA500 slightly enhanced the reduction of ORF2-F protein upon let-7 overexpression (**Supplementary Fig. 5i**), an effect that was more prominent when the canonical site ('8mer') was tested. In agreement with previous studies[75], these results suggests that the

proficiency of ´bs2rh´, a noncanonical offset 7mer site, is weaker than that of a canonical let-7 binding site when they are located in 3`UTR. Moreover, we cannot rule out that the translational repression mediated by both biding sites located in 3`UTR could be attributed to mRNA destabilization. Additionally, using site directed mutagenesis we introduced two point-mutations in the ORF2 coding region to transform the offset 7-mer into a canonical let-7 8-mer site, generating pSA500-ORF2-8mer (**Supplementary Fig. 5j**). We co-transfected this construct in HeLa cells with let-7 mimic. Interestingly, the 8-mer site within the ORF does not affect the levels of mRNA (**Supplementary Fig. 5k**) and leads to a decrease in the protein level similar to that observed above for ´bs2rh´ (**Supplementary Fig. 5m**). Altogether, these results suggest that the translational repression mediated by bs2rh mainly depends on its location within the coding sequence, rather than on its non-canonical interaction.

To further characterize this effect, we next analyzed the impact of let-7 binding on L1-ORF2p translation in its natural context: a full-length bicistronic L1 RNA where L1-ORF2p is translated using the aforementioned termination/reinitiation mechanism[8]. In order to perform a more quantitative analysis, we combined the use of L1-encoded proteins with fluorescent tags and confocal microscopy or flow cytometry. Briefly, we generated a construct where L1-ORF1p and L1-ORF2p from a human L1.3 element were fused to EGFP and mCherry, respectively, at their C-terminus (plasmid pVan583, **Fig. 5d**). First, we

confirmed that ORF1p-GFP was expressed from this construct (**Supplementary Fig. 5n**) and that this tagged L1 was able to retrotranspose, although the addition of both florescent tags reduced its activity to ~30% of its untagged counterpart, JM101/L1.3 (**Supplementary Fig. 5o**). Next, by confocal microscopy we observed a reduction of L1-ORF2p-mCherry but not of L1-ORF1p-EGFP levels upon overexpression of let-7 in U2OS cells (**Supplementary Fig. 5q**). However, due to reduced transfection capacity of this construct (**Supplementary Fig. 5p)** and the inefficient translation of ORF2p-mCherry[8], we obtained an insufficient number of double positive cells to enable a quantitative analysis by microscopy. Therefore, we turned to a more sensitive and quantitative approach: flow cytometry. We co-transfected pVan583 with let-7 inhibitor in HeLa cells and analyzed EGFP+ cells (i.e. >3500 cells per sample). Notably, we found that depleting let-7 led to an increase in the number of mCherry+ cells in the EGFP+ population suggesting an increase in the synthesis of L1-ORF2p-mCherry (**Fig. 5d**).

Altogether, our results suggest that let-7 impairs L1-ORF2p translation, potentially altering the ratio between L1-ORF1p and L1-ORF2p, which we speculate could unbalance L1-RNP formation (**Fig. 6**).

DISCUSSION

Many studies have linked LINE-1 retrotransposons to cancer[25-32,34-36]. In particular, L1 insertions have been found to occur at high frequencies in lung cancer genomes[36]. L1 is also associated with genomic instability, since new insertions can potentially cause splicing alterations, exon disruptions, indel mutations or large genomic rearrangements [2,22,39]. How these elements are silenced and derepressed in somatic human tissues, and how these processes impact tumorigenesis is an open question. DNA methylation of the L1 promoter is an important inhibitor of L1 activity[40]. In fact, a consistent correlation between the number of somatic L1 insertions in lung cancer and hypomethylation of L1 promoters has been shown, both at a global and at a locus specific level [32,41]. However, considering the high level of somatic L1 activity in some of these patients, it is tempting to speculate that RC-L1s might also escape post-transcriptional restriction mechanisms[47,77]. On the other hand, among all tumor suppressor miRNAs, reduced let-7 expression occurs most frequently in cancer and typically correlates with poor prognosis[78]. Functionally, it is well known that a decrease in let-7 miRNAs leads to overexpression of their oncogenic targets such as MYC, RAS, HMGA2 among others[54].

Here we describe a new role for let-7 in controlling human L1 activity, which may contribute to its tumor suppressor function. First, we found a high

frequency of retrotransposition in NSCLC cancer, consistent with previous reports[26,36]. We further showed that human tumor samples with somatic L1 insertions present reduced let-7 expression. Additionally, we demonstrated that the mobilization of full-length L1s in cultured cells can be negatively regulated by let-7 in a variety of cell lines including lung cancer cells. It is worth noting that the expression of another tumor suppressor miRNA, miR-34a, is also reduced in lung tumors with L1 activity and correlates negatively with L1Hs RNA levels. However, we did not observe a consistent effect of the latter on L1 retrotransposition, under our experimental conditions. We speculate that mir-34 could indirectly regulate L1 mobilization, targeting a member of the epigenetic regulatory network controlling expression of active L1s in our genome.

AGO proteins are the main effectors of miRNA-guided gene silencing[79]. Our AGO2 RNA immunoprecipitation assay and the use of engineered retrotransposition constructs lacking 3'UTR or 5'UTR suggest that let-7 is actually guiding AGO2 to the human L1 mRNA, and that its binding occurs in L1 coding sequence (CDS). In fact, we have demonstrated that ORF2 contains a noncanonical offset 7-mer let-7 binding site previously described as functional for microRNA targeting[75]. Although it is widely accepted that most binding sites are located in the 3'UTR of target mRNAs[42], functional CDS binding sites for different miRNAs, including let-7, have also been described [63,64,80].

Furthermore, we demonstrate that mutations in this binding site reduce, but not abolish, the effect of let-7 modulation on human L1 mobility. These results suggest that additional let-7 binding sites may exist within the coding sequence of human L1 mRNA, or that let-7 might have redundant indirect effects. However, we failed to validate functional let-7 binding sites in a different L1-ORF2 region ('bs1', identified by miRanda, **Supplementary Fig 4a**) or a region within L1-ORF1, identified using local alignments (data not shown). Thus, alternative approaches might be needed to obtain an unbiased view of functional let-7 binding sites in the L1 mRNA. Importantly, the binding site for the only miRNA targeting human LINE-1 mRNA described so far, miR-128, is also located in L1-ORF2[50]. We speculate that L1-ORF2, the largest and most conserved region among human LINE-1 subfamilies, is preferentially targeted by miRNAs because of the restricted space in the short L1 3'UTR, as it has been suggested by computational analysis for other mRNAs [80]. Accordingly, it was previously shown that knocking down the Microprocessor complex increased retrotransposition of a LINE-1 lacking the 3'UTR to the same extent as that of a full-length element [48]. Moreover, consistent with the mechanism suggested for gene silencing mediated by miRNA binding sites located in CDS [64,81], we have demonstrated that let-7 impairs translation of L1 ORF2p without affecting mRNA stability. This conclusion is further supported by the fact that no correlation between the levels of let-7 and the expression of L1Hs RNA was observed in human lung tumor samples (**Supplementary Fig. 1b**). Interestingly,

the accumulation and translation of L1 an mRNA variant that have the natural ´bs2rh´ site substituted by a canonical let-7 8-mer site is similarly affected by let-7 overexpression as wild-type molecules (**Supplementary Fig. 5j-m**), suggesting that binding to the CDS region itself rather than the structure of the of base-pairing mediates translational repression, as previously described for other miRNA targeting CDS sites[64].

Since L1-ORF2p is expressed at a very low level and is essential for L1 retrotransposition[5], a small reduction in the abundance of this protein could unbalance RNP formation reducing human L1 mobilization. Besides binding L1-ORF2 sequences, let-7 could also be regulating other mRNAs encoding proteins that positively impact human L1 retrotransposition[62].

Let-7 is one of the most highly conserved families of miRNAs in the animal kingdom and is involved in multiple biological processes including differentiation, cell death, metabolism and cancer[54]. Here, all our findings support a model in which let-7 also guides the RISC to the mRNA of active L1s and impairs L1-ORF2p translation, altering the ratio between L1-ORF2p and L1-ORF1p in the L1-RNP and consequently reducing LINE-1 retrotransposition (**Fig. 6**). Mature let-7 is highly expressed in differentiated cells[54], where different mechanisms repress L1 activity to avoid somatic L1 insertions[41]. We hypothesize that alterations in let-7 expression in human cancer lead to an increased mobilization of actively transcribed L1s and, moreover L1- mediated

553 retrotransposition of non-autonomous transposable elements like Alu and SVA

554 [38], increasing genome instability and contributing to tumor progression.

ACKNOWLEDGEMENTS

The results shown here are in part based upon data generated by the TCGA Research Network: https://cancer.gov/TCGA. We thank Jose Luis Garcia-Perez (IGMM, Edinburgh, UK) for his valuable contribution to this project. We are also grateful to Marie-Jeanne Kempen, Jose Luis Garcia-Perez and Javier F. Caceres (IGMM, Edinburgh, UK) for comments and critical reading of the manuscript. We thank current members of the lab for useful discussions and methodological support. We are grateful to Drs. Jose Luis Garcia-Perez and John V. Moran (University of Michigan, US) for sharing numerous human L1 vectors and cell lines; Aleksandra Helwak for valuable input at the beginning of this project, Wenfeng An (Dakota State University, USA) for sharing L1-Fluc plasmids; Aurelien Doucet for sharing pAD500; Norihiro Okada and Masaki Kajikawa (Tokyo Institute of Technology, Japan) for sharing tagged zebrafish LINE-2 plasmids; Haig Kazazian and John Goodier (John Hopkins, US) for sharing tagged mouse LINE-1 plasmids; Tomo-ichiro Miyoshi (Kyoto University) for sharing pJCC5ORF1T7-ORF3xFLAG plasmid; Pedro Medina (Genyo, Spain) for the lung cancer cell lines and technical advises; Oliver Weichenrieder (MPI Tubingen, Germany) for the polyclonal L1-ORF1p antibody; Ian Adams (IGMM, Edinburgh, UK) for Flp-In-293 cells expressing T7-tagged ORF1p; Verónica Ayllón (Genyo, Spain) for support with flow cytometry analysis; and Guillermo Barturen (Genyo, Spain) for support with

bioinformatic and statistical analysis. We also thank Jennifer Parra for technical support, and Genyo's flow cytometry unit. This article is part of the doctoral thesis of P.T.-R., a graduate student in the PhD programme of Biomedicine at the University of Granada, who was supported by MINECO (PEJ-2014-A-31985 and SAF2015-71589-P). S.R.H. is funded by MINECO cofounding by European Regional Development Fund (SAF2015-71589-P), Ramón y Cajal (RYC-2016-21395) and a Career Integration Grant-Marie Curie (FP7-PEOPLE-2011-CIG-303812).

AUTHOR CONTRIBUTIONS

S.R.H conceived and supervised the study. P.T.-R. and S.R.H designed and interpreted the experiments. P.T.-R. performed most of the experiments and data analysis. A.R.-R and G.P. provided all the bioinformatic analysis, including identification of somatic L1 insertions by MELT. P.T.-R and G.P. performed statistical analysis. A.R.-R was responsible for data management. L.S. provided technical support and performed experiments. S.A.-C. generated pSA500 construct. S.V. and G. C. generated and characterized the pVan583 construct. P.T.-R., G.P. and S.R.H wrote the manuscript and all the authors contribute with a critical reading.

COMPETING INTERESTS

The authors declare no competing interests.

596 **DATA AVAILABILITY**

597 Raw data used to generate the figures is available upon request. Sequencing
598 data from tumor samples (used in Fig. 1 and Supplementary Fig. 1) is available
599 in TCGA (https://www.cancer.gov/tcga).

600 **ADDITIONAL INFORMATION**

601 Correspondence and request for materials and raw data should be addressed to
602 S.R.H. (sara.rodriguez@genyo.es)

603

METHODS

Sequencing data

Both WGS (aligned to HG19), miRNA expression quantification and RNAseq raw sequencing data files from TCGA were obtained from the Genomic Data Commons (GDC) Legacy Archive using the GDC Data Transfer Tool [82]. Cases of paired tumor-normal whole-genome sequencing (WGS) where tumor miRNAs expression data was available were retrieved for lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). High coverage (28-95x) WGS files aligned to hg19 from primary tumor and solid tissue normal samples, and miRNA gene quantification files from primary tumor were downloaded for LUAD (17 patients) and LUSC (28 patients).

WGS analysis

Putative somatic LINE-1 insertion calls for both normal tissue (NT) and primary tumor (PT) were obtained using MELT version 2.1.5 [51]. To discard possible sequencing artifacts, candidate somatic insertions were further filtered including calls supported with a minimum of three split-reads, with the highest accuracy assessment for breakpoint detection and passing all internal filters (MELT parameters ASSESS=5 and FILTER=PASS). Polymorphic insertion calls were found using a curated database included in TEBreak software (https://github.com/adamewing/tebreak) and excluded from final results. Several quality values were checked as a measure of filtering effects on original

(unfiltered) MELT results. First, somatic insertions found in NT alone, and NT and PT simultaneously were expected to be zero, and only a maximum of one insertion was allowed for these values. All samples passed this additional filtering.

Furthermore, polymorphic L1 insertions after MELT filtering were controlled, requiring that a similar number was found for PT and NT samples, and that this number were uncorrelated with sample coverage. Four samples in LUSC (TCGA-60-2695, TCGA-60-2722) and LUAD (TCGA-55-1594, TCGA-55-1596) were excluded from analysis because only a low number (<10%) of polymorphic insertions passed all filters.

Filtered LINE-1 calls were considered tumor somatic insertions if detected in primary tumor filtered results and absent in unfiltered solid normal tissue insertion set within a range of 100 bp.

Correlation with miRNA expression

Samples were divided in two groups depending on whether putative somatic insertions were or were not found in the primary tumor. Only miRNAs with medium-high expression (over 100 reads per million (RPM) mapped reads) were considered. For some of the analysis, expression of specific miRNAs known to be involved in the development and progression of lung cancer was analyzed.

For each miRNA, outliers were discarded (we considered outliers values deviating more than two standard deviation from the mean in each group).

Differentially expressed miRNAs were identified applying an unpaired two-tailed t test adjusted by FDR=1%. Results were confirmed using a rank-sum test. RPMs were normalized to the highest expression value of each miRNA to enable visualization of all miRNAs in the same graph. Tumor suppressor miRNAs and oncomiRs related to lung cancer used for this analysis were described in a recent revision [53]. As a control, L1 insertion numbers were randomly re-assigned to each sample and analysis was repeated. Moreover, analysis was done with L1 insertions determined by Helman and col. using Transpo-seq in lung and breast cancer samples obtained from TGCA as well [26]. Data processing and analysis was performed as described above.

L1Hs RNA expression

To analyze global TE expression in RNA-seq experiments, we use SQuIRE[58] (Software for Quantifying Interspersed Repeat Elements). SQuIRE quantifies expression at the subfamily level. It outputs read counts and fragments per kilobase transcript per million reads FPKM. Linear correlation between L1HsRNA levels and microRNAs expression was calculated using Pearson Correlation Coefficient in GraphPad Prism 6.

Cell culture

HEK293T, PA-1, HeLa and U2OS cells were originally obtained from ATCC and were provided by Drs Jose Luis Garcia-Perez (IGMM, Edinburgh, UK) and John V. Moran (University of Michigan, US). Lung cancer cell lines (A549, SK-MES-1)

were provided by Dr Pedro Medina (GENYO, Spain). Stable Flp-In-293 cells expressing T7-tagged L1-ORF1p were previously generated for a different study[83].

HEK293T, HeLa, U2OS, A549 and SK-MES1 cells were cultured in high-glucose Dulbecco's Modified Eagle's Media (DMEM, Gibco) supplemented with GlutaMAX, 10% foetal bovine serum (FBS, Hyclone) and 100 U/mL penicilin-streptomycin (Invitrogen).

PA-1 cells were cultured in Minimal Essential Medium (MEM, Gibco) supplemented with GlutaMAX, 10% heat-inactivated FBS (Gibco), 100 U/mL penicilin-streptomycin (Invitrogen) and 0.1 mM Non-Essential Amino Acids (Gibco).

All cells were maintained in humidified incubators at 37ºC with 5% CO_2. Absence of *Mycoplasma spp.* in cultured cells was confirmed at least once a month by a PCR-based assay (Minerva). Cell identity was confirmed at least once a year using STR genotype (at Lorgen, Granada).

Retrotransposition assays

Modified versions of previously established L1 retrotransposition assays [84-87] were performed and are described below[59]. In the Neo/Blast assays, $2x10^5$ HeLa JVM cells were plated in 6-well tissue culture plates. Within 24h, cells were co-transfected with 0.5-1µg of L1 plasmid and 60nM of let-7 mimic or 40nM of let-7

inhibitor and their respective controls scr and c- using Dharmafect DUO (Dharmacon) following manufacturer's instructions. For Neo assays, selection with 400 μg/mL of G418 (Life) was started 48h post-transfection. For Blast assays, selection with 10 μg/mL of blasticidin (Millipore) was started 5 days post-transfection. In both cases, medium was changed every two days. Between 12 and 14 days after transfection, cells were washed with 1X PBS (Gibco), fixed (2% formaldehyde, 0.2% glutaraldehyde in 1X PBS), and stained with 0.5% crystal violet. Colonies were manually counted. The number of antibiotic-resistant colonies was used to quantify retrotransposition levels in cultured cells. Clonability assay was performed with 0.5 μg pU6i Neo and $1x10^5$ HeLa cells[88].

Luciferase retrotransposition assays were performed as described[87]. pYX014, pYX015 or pYX017 were co-transfected with 60nM of let-7 mimic or 40nM of let-7 inhibitor using Lipofectamine 2000 (Life). Luciferase activity was measured 96h post-transfection using Dual-Luciferase Reporter Assay System (Promega), in a GloMax Luminometer (Promega).

EGFP-based retrotransposition assays were performed as described[85]. HEK293T cells were co-transfected with 99-UB-LRE3 and 40nM of let-7 inhibitor using Lipofectamine 2000 (Life). Retrotransposition (EGFP+ cells) was quantified 8 days post-transfection in a FACS Canto cytometer (BD).

MiRNAs mimics and inhibitors

Let-7a/b mimic (C-300473-05 and C-300476-05), miR-34 mimic (C-300551-07), let-7a hairpin inhibitor (IH-300474-07), and their respective controls scr and c- (CN-002000-01-05 and IN-002005-01-05), were purchased from Dharmacon. They were resuspended in 1x siRNA Buffer (Thermo) to a working concentration of 20 μM and kept at -80ºC.

RNA ImmunoPrecipitation (RIP)

$2x10^6$ PA-1 cells were transfected in 10cm tissue culture plates with 4 μg of FLAG-AGO2 and 25 nM of scr/mimic let-7 using lipofectamine 2000 (Life). Transfection with pBSKS (an empty plasmid) was used as a negative control for the IP. 48h post-transfection, cells were washed with ice-cold 1xPBS, scraped and transferred to a 1.5 ml tube. After centrifugation at 1500rpm for 2 minutes, cells were resuspended in 200 μl of cold resuspension buffer (20mM Tris (pH=7.5), 150mM NaCl, 1mM EDTA, 1mM EGTA) containing 1U/μL RNAsin Plus (Promega) and lysed adding 800 μl of cold lysis buffer (1% Triton X-100, 20mM Tris (pH=7.5), 150mM NaCl, 1mM EDTA, 1mM EGTA, 1mM phenylmethyl-sulfonyl fluoride (PMSF, Sigma), 1X cOmplete EDTA-free Protease Inhibitor cocktail (Roche) and incubating for 10-30 min on ice. After centrifugation (10000g for 10min at 4ºC), 10 μL of RQ1 Dnase (Promega) was added to the supernatant. Immunoprecipitation of FLAG-AGO2 was performed with Dynabeads Protein G (Life) and anti-FLAG M2 mouse (Sigma, F3165) for 3h at 4ºC with rotation. After five washes with lysis buffer, 10% of sample-

731 beads were used for protein extraction and western-blot, while 90% was
732 incubated with RQ1 DNAse for 30 min for later RNA extraction with Trizol LS
733 (Ambion).

734

735 **Co-immunoprecipitation**

736 PA-1 cells were transfected with FLAG-AGO1/2, or an empty plasmid ('mock').
737 48h post transfection, cells were lysed using lysis buffer and
738 immunoprecipitation with anti-FLAG M2 mouse (Sigma, F3165) as it was
739 described above. After the last wash with lysis buffer, the beads were treated
740 with 100 μg/ml RNAse A for 30 min. The western blot was perfomed with Anti-
741 ORF1p (1:1000) provided by Dr. Oliver Weichenrieder, (Max-Planck, Germany)
742 and subsequently, with anti-FLAG M2 mouse (Sigma, F3165).

743 **siCHECK luciferase assays**

744 In 24-well plates, $1x10^5$ HEK293T cells were seeded per well. Within 24h after
745 seeding, cells were co-transfected with 10ng of each siCHECK plasmid and 50-
746 80nM scr/let-7 mimic using Lipofectamine 2000 (Life). 24h post-transfection,
747 Firefly and Renilla luciferase measurements were performed in a GloMax
748 Luminometer (Promega) using Dual-Luciferase® Reporter Assay System
749 (Promega), following manufacturer's instructions.

750 **Site-directed mutagenesis**

751 Binding site mutant 'bs2rhmut'was generated using an established protocol [89].
752 Briefly, 2 sequential PCRs were performed, using an active L1.3 as a template.
753 First, two PCRs were performed using the following primers under standard
754 conditions: Let7-ORF2PCRafw/Let7-ORF2PCRa_PG2rv and Let7-
755 ORF2PCRarv/Let7-ORF2PCRa_PG2fw. The products of both reactions were
756 purified, mixed in equal amounts, and used as a template for a second PCR
757 using primers Let7-ORF2PCRafw/ Let7-ORF2PCRarv. Conditions for this PCR
758 were: 95ºC 5min, 10 cycles with (95ºC 15s, 50ºC 30s, 72ºC 60s), 25 cycles with
759 (95ºC 15s, 55ºC 30s, 72ºC 60s), 72ºC 10min. The resulting product contained the
760 mutated sequence in ORF2. This product was purified and cloned into a
761 plasmid containing an active L1 (pJCC5/L1.3) using EcoNI and BsaBI sites,
762 generating pJCC5/bs2mutL1.3. This mutant L1 was then cloned into pJM101
763 using NotI and BstZ17I sites, generating pJM101/bs2mutL1.3.

764 Binding site mutant 8-mer (pSA500 ORF2-8mer) was generated using the same
765 protocol described above. For the first two PCRs, primers used were Let7-Bcl1-
766 ORF2bs-PCRaFw/Let7-ORF2PCRa_8mer and Let7-
767 ORF2PCRb_8mer/pCEP4_Rv. The products of both reactions were purified,
768 mixed in equal amounts, and used as a template for a second PCR using
769 primers Let7-Bcl1-ORF2bs-PCRaFw/pCEP4_Rv. The product was purified and
770 cloned into pSA500 using BclI and BstZ17I sites.

771 Restriction enzymes were purchased from New England Biolabs (NEB).

Generation of 3'UTR variants of pSA500

Sequences were ordered as oligos flanked by a BstZ17I site (See Supplementary Table XI). First, 1μL of each Fw and Rv were annealed and phosphorylated with T4 Polynucleotide Kinase (PNK, NEB) using the following program: 30min at 37ºC, 5min at 95ºC, and then down to 25ºC at -5ºC/min. They were cloned in 3'UTR of pSA500 using a BstZ17I site, generating pSA500-3'UTR-scrb/bs2rh/8mer. Constructs were checked by digestion and Sanger sequencing.

Western Blot

Cells were washed with 1X PBS, trypsinised and pelleted at 200g for 4 minutes. To extract proteins, cell pellets were resuspended in 50-100 μL of RIPA buffer (Sigma) supplemented with 1X Complete EDTA-free Protease Inhibitor cocktail (Roche), PMSF (Sigma), 0.25% β-mercaptoethanol (Sigma) and incubated for 10 min on ice. Extracts were then centrifuged (13000rpm at 4ºC for 10 min) and debris-free supernatants were transferred to new tubes. Protein concentration was determined using the Micro BCA Protein Assay Kit (Thermo) following manufacturer's instructions.

To control the transfection efficiency in pSA500 western blots, three fractions of the cells were pelleted in different tubes after trypsinization, and DNA, RNA and protein extractions were performed simultaneously.

Proteins were resolved on an SDS-PAGE gel and transferred to a nitrocellulose membrane (BioRad). In L1-ORF2p western blots, proteins were resolved in a 4-15% Mini-PROTEAN TGX Precast Gels (BioRad), and transferred to a PVDF membrane using Trans-Blot Turbo Mini PVDF Transfer Packs (BioRad) and the Trans-Blot Turbo Transfer System (BioRad). For blotting we used the following antibodies: a polyclonal rabbit anti L1-ORF1p (1:1000, provided by Dr. Oliver Weichenrieder, Max-Planck, Germany), a monoclonal mouse anti L1-ORF1p (1:2000, Millipore), anti HMGA2 (1:1000, Abcam), anti DICER (1:1000, Cell Signalling), anti EBNA1 (1:500, Santa Cruz), anti tubulin (1:1000, Santa Cruz). For chemiluminescent detection we used anti rabbit HRP (1:1000, Cell Signaling) or anti mouse HRP (1:1000, Cell Signaling), and Clarity ECL Western Blotting Substrate (BioRad) or SuperSignal West Femto Maximum Sensitivity Substrate (Thermo). Images were acquired with an ImageQuant LAS4000 and quantified using ImageJ software. For Odyssey analysis, anti-rabbit and anti-mouse fluorescent antibodies (LI-COR) were used at 1:10000 dilution, and detection and quantification were performed in Odyssey (LI-COR).

qPCR and RT-qPCR

To control plasmid transfection in L1-ORF2p western blots, we used a method previously described[90]. Briefly, transfected cells were lysed in a buffer containing 10mM Tris pH=8.2, 10mM EDTA, 200mM NaCl, 0.5%SDS and 200µg/ml proteinase K, and incubating for 3h at 56ºC with agitation.

812 Afterwards, DNA was purified with phenol:chloroform:isoamyl alcohol
813 (25:24:1, Thermo) following standard procedures. 50ng of each DNA sample
814 were used per qPCR reaction (GoTaq qPCR Mix, Promega), and an
815 untransfected control was used to discard plasmid contamination. qPCR
816 method is as follows: 5min at 95º, and 40 cycles of 15s at 95º followed by 60s at
817 60ºC. Plasmid levels were quantified using CMV and EBNA-1 primers, and
818 normalization was performed using genomic GAPDH primers as described[90].

819 RNA was extracted from cells using Trizol (Invitrogen), following
820 manufacturer's instructions. 1µg of RNA was subsequently treated with RQ1
821 DNAse and purified by a phenol/chlorophorm extraction. cDNA was
822 synthesized using High-Capacity cDNA Reverse Transcription Kit (Applied
823 Biosystems), and used for qPCR (GoTaq qPCR Mix, Promega) using standard
824 protocols. Two controls were used to verify the absence of contaminating
825 gDNA: no-RT and no-template. Endogenous L1 mRNA was quantified using
826 N51 primers as described[91]. Transfected L1 mRNA was quantified using
827 NEOjunct2 primers designed to exclusively amplify the spliced NEO cDNA[90]
828 (in Supplementary Fig. 5a,b) or SV40 primers when the NEO cassette was
829 absent (Fig.5c and Supplementary Fig. 5h,k). GAPDH was used to normalize in
830 Fig. 5c because an additional qPCR was used to quantify plasmid levels and
831 discard differences in transfection efficiency. Elsewhere, EBNA mRNA,
832 expressed from the backbone of the plasmids, was used to normalize.

For mature miRNAs quantification, a RT-qPCR was used. 1µg of total RNA isolated with Trizol was polyadenylated and then cDNA was synthesised, using qScript miRNA cDNA synthesis Kit (QuantaBio). Quantitative PCR was performed using a universal primer against poly(A) and a miRNA-specific primer that allows the specific detection of polyadenylated mature miRNA and not its precursors (QuantaBio). qPCR was performed with primers Let-7a, Let-7b and miR-34a.

Flow cytometry

In 6-well plates, 2x105 HeLa cells were seeded per well. Next day, cells were transfected with 1µg of pVan583 and 40nM c-/let7inh using lipofectamine 2000 (Life). Seventy-two hours post-transfection, cells were washed with 1XPBS (Life), detached with TrypLE Express (Gibco) for 5-10 min at 37ºC, pelleted 4 min at 200g, resuspended in 1X PBS with 5% FBS and 5mM EDTA, and passed through a 70µM filter. After incubation with 10ug/mL 7AAD (Sigma) for 10min, fluorescence was quantified in a FACSAria (BD) cytometer. For each replicate, 10^5 cells were passed through the cytomerer. Only live and transfected cells (7AAD- and GFP+, between 3600 and 9300 cells) were used for %mCherry analysis, which was performed using FlowJo software (LLC). Controls were used to set the threshold for each fluorescent channel of detection: untransfected cells, and cells expressing either GFP only or mCherry only.

Confocal microscopy

2x105 U2OS per well were seeded in 6-well tissue culture plates. Next day, cells were transfected with 1μg of DNA and 60nM scr/mimic using lipofectamine 2000 (Life), following standard protocols. Twenty-four hours post-transfection, cells were detached and re-seeded in 24-well plates where a UV-sterilized glass slide had previously been placed. Forty-eight hours post-transfection, cells were washed in 1xPBS, fixed in paraformaldehyde (4% in 1xPBS) for 30 minutes at room temperature, and slides were mounted with Slow-Fade Gold Antifade reagent with DAPI (Life). Slides were imaged using a Zeiss LSM-710 confocal microscope (Leica).

Hybridization between let-7 and L1 prediction

The potential structure formed by let-7 and WT L1 or its binding site mutant was analyzed by RNAHybrid as described[73]. Briefly, the region of L1Hs identified as 'bs2' with RNA22 was paired to the mature sequence of let-7, using default parameters.

Plasmids

pYX014, pYX015 and pYX017 [87], JM101/L1.3[74] , JM105/L1.3[5], JJ101/L1.3 [92] and TAM102/L1.3 [84], 99-UB-LRE3[93], Tgf21-Neo [94], Zfl2-1-Neo and Zfl2-2-Neo [95] have been previously described. JJ101/L1.3Δ3'UTR is a derivative of JJ101/L1.3 but the L1 lacks the 3'UTR[48]. JM101/L1.3bs2mut was generated by cloning a bs2mut L1 into JM101/L1.3 (cloning strategy described in the 'Site-directed mutagenesis' section of the methods). FLAG-Ago1 and FLAG-Ago2 were a gift

875 from Edward Chan (Addgene plasmid# 21533 and # 21538)[96]. pSA500 is as
876 pAD500[67] where the TAP epitope was replaced by 3 consecutive copies of the
877 FLAG epitope tag obtained from PJCC5 ORF1T7 ORF3XFLAG using BclI and
878 BstZ17I restriction enzyme sites. pVan583 is a derivative of JM101/L1.3 where
879 EGFP and mCherry were cloned in frame with the last amino acid of L1-ORF1
880 and L1-ORF2, respectively. All psiCHECK2 constructs were generated by
881 cloning sequences synthesized and cloned in pUC57 (Genescript) into
882 psiCHECK2 using XhoI and NotI restriction enzyme sites (Promega).
883

REFERENCES

1. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
2. Richardson, S.R. et al. The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. *Microbiol Spectr* **3**, MDNA3-0061-2014 (2015).
3. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
4. Brouha, B. et al. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* **100**, 5280-5 (2003).
5. Moran, J.V. et al. High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917-27 (1996).
6. Ardeljan, D. et al. LINE-1 ORF2p expression is nearly imperceptible in human cancers. *Mob DNA* **11**, 1 (2020).
7. Taylor, M.S. et al. Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell* **155**, 1034-48 (2013).
8. Alisch, R.S., Garcia-Perez, J.L., Muotri, A.R., Gage, F.H. & Moran, J.V. Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev* **20**, 210-24 (2006).
9. Hohjoh, H. & Singer, M.F. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *Embo J* **16**, 6034-43 (1997).
10. Martin, S.L. & Bushman, F.D. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol Cell Biol* **21**, 467-75 (2001).
11. Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., Boeke, J.D. & Gabriel, A. Reverse transcriptase encoded by a human transposable element. *Science* **254**, 1808-10 (1991).
12. Feng, Q., Moran, J.V., Kazazian, H.H., Jr. & Boeke, J.D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905-16 (1996).
13. Kulpa, D.A. & Moran, J.V. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol* **13**, 655-60 (2006).
14. Luan, D.D., Korman, M.H., Jakubczak, J.L. & Eickbush, T.H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595-605 (1993).
15. Cost, G.J., Feng, Q., Jacquier, A. & Boeke, J.D. Human L1 element target-primed reverse transcription in vitro. *Embo J* **21**, 5899-910 (2002).
16. Flasch, D.A. et al. Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. *Cell* **177**, 837-851 e28 (2019).
17. Sultana, T. et al. The Landscape of L1 Retrotransposons in the Human Genome Is Shaped by Pre-insertion Sequence Biases and Post-insertion Selection. *Mol Cell* **74**, 555-570 e7 (2019).
18. Monot, C. et al. The specificity and flexibility of l1 reverse transcription priming at imperfect T-tracts. *PLoS Genet* **9**, e1003499 (2013).
19. Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35**, 41-8 (2003).
20. Hancks, D.C., Goodier, J.L., Mandal, P.K., Cheung, L.E. & Kazazian, H.H., Jr. Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum Mol Genet* **20**, 3386-400 (2011).
21. Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* **24**, 363-7 (2000).
22. Konkel, M.K. & Batzer, M.A. A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. *Semin Cancer Biol* **20**, 211-21 (2010).

23. Hancks, D.C. & Kazazian, H.H., Jr. Roles for retrotransposon insertions in human disease. *Mob DNA* **7**, 9 (2016).
24. Kazazian, H.H., Jr. & Moran, J.V. Mobile DNA in Health and Disease. *N Engl J Med* **377**, 361-370 (2017).
25. Ewing, A.D. et al. Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res* **25**, 1536-45 (2015).
26. Helman, E. et al. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res* **24**, 1053-63 (2014).
27. Lee, E. et al. Landscape of somatic retrotransposition in human cancers. *Science* **337**, 967-71 (2012).
28. Miki, Y. et al. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res* **52**, 643-5 (1992).
29. Rodic, N. et al. Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat Med* **21**, 1060-4 (2015).
30. Scott, E.C. et al. A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res* **26**, 745-55 (2016).
31. Shukla, R. et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* **153**, 101-11 (2013).
32. Tubio, J.M.C. et al. Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345**, 1251343 (2014).
33. Nguyen, T.H.M. et al. L1 Retrotransposon Heterogeneity in Ovarian Tumor Cell Evolution. *Cell Rep* **23**, 3730-3740 (2018).
34. Schauer, S.N. et al. L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. *Genome Res* **28**, 639-653 (2018).
35. Doucet-O'Hare, T.T. et al. LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc Natl Acad Sci U S A* **112**, E4894-900 (2015).
36. Iskow, R.C. et al. Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* **141**, 1253-61 (2010).
37. Scott, E.C. & Devine, S.E. The Role of Somatic L1 Retrotransposition in Human Cancers. *Viruses* **9**(2017).
38. Burns, K.H. Transposable elements in cancer. *Nature Reviews Cancer* **17**, 415-424 (2017).
39. Rodriguez-Martin, B. et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat Genet* (2020).
40. Bourc'his, D. & Bestor, T.H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96-9 (2004).
41. Sanchez-Luque, F.J. et al. LINE-1 Evasion of Epigenetic Repression in Humans. *Mol Cell* (2019).
42. Bartel, D.P. Metazoan MicroRNAs. *Cell* **173**, 20-51 (2018).
43. Fabian, M.R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* **79**, 351-79 (2010).
44. Friedman, R.C., Farh, K.K., Burge, C.B. & Bartel, D.P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92-105 (2009).
45. Lu, J. et al. MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-8 (2005).
46. Lujambio, A. & Lowe, S.W. The microcosmos of cancer. *Nature* **482**, 347-55 (2012).
47. Heras, S.R., Macias, S., Caceres, J.F. & Garcia-Perez, J.L. Control of mammalian retrotransposons by cellular RNA processing activities. *Mob Genet Elements* **4**, e28439 (2014).
48. Heras, S.R. et al. The Microprocessor controls the activity of mammalian retrotransposons. *Nat Struct Mol Biol* **20**, 1173-81 (2013).

49. Bodak, M., Cirera-Salinas, D., Yu, J., Ngondo, R.P. & Ciaudo, C. Dicer, a new regulator of pluripotency exit and LINE-1 elements in mouse embryonic stem cells. *FEBS Open Bio* **7**, 204-220 (2017).
50. Hamdorf, M. et al. miR-128 represses L1 retrotransposition by binding directly to L1 RNA. *Nat Struct Mol Biol* **22**, 824-31 (2015).
51. Gardner, E.J. et al. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome Res* **27**, 1916-1929 (2017).
52. Mir, A.A., Philippe, C. & Cristofari, G. euL1db: the European database of L1HS retrotransposon insertions in humans. *Nucleic Acids Res* **43**, D43-7 (2015).
53. Inamura, K. & Ishikawa, Y. MicroRNA In Lung Cancer: Novel Biomarkers and Potential Tools for Treatment. *J Clin Med* **5**(2016).
54. Roush, S. & Slack, F.J. The let-7 family of microRNAs. *Trends Cell Biol* **18**, 505-16 (2008).
55. Takamizawa, J. et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res* **64**, 3753-6 (2004).
56. Yanaihara, N. et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9**, 189-98 (2006).
57. He, L. et al. A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130-4 (2007).
58. Yang, W.R., Ardeljan, D., Pacyna, C.N., Payer, L.M. & Burns, K.H. SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Res* **47**, e27 (2019).
59. Tristan-Ramos, P. et al. sRNA/L1 retrotransposition: using siRNAs and miRNAs to expand the applications of the cell culture-based LINE-1 retrotransposition assay. *Philos Trans R Soc Lond B Biol Sci* **375**, 20190346 (2020).
60. Robertson, B. et al. Specificity and functionality of microRNA inhibitors. *Silence* **1**, 10 (2010).
61. Taylor, M.S. et al. Dissection of affinity captured LINE-1 macromolecular complexes. *Elife* **7**(2018).
62. Liu, N. et al. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553**, 228-232 (2018).
63. Forman, J.J., Legesse-Miller, A. & Coller, H.A. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci U S A* **105**, 14879-84 (2008).
64. Hausser, J., Syed, A.P., Bilen, B. & Zavolan, M. Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Res* **23**, 604-15 (2013).
65. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654-65 (2013).
66. Meister, G. et al. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**, 185-97 (2004).
67. Doucet, A.J. et al. Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet* **6**(2010).
68. Goodier, J.L., Zhang, L., Vetter, M.R. & Kazazian, H.H., Jr. LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Mol Cell Biol* **27**, 6469-83 (2007).
69. Garcia-Perez, J.L. et al. Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* **466**, 769-73 (2010).
70. Lee, Y.S. & Dutta, A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev* **21**, 1025-30 (2007).
71. John, B. et al. Human MicroRNA targets. *PLoS Biol* **2**, e363 (2004).
72. Miranda, K.C. et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* **126**, 1203-17 (2006).

73. Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**, 1507-17 (2004).
74. Sassaman, D.M. et al. Many human L1 elements are capable of retrotransposition. *Nat Genet* **16**, 37-43 (1997).
75. Kim, D. et al. General rules for functional microRNA targeting. *Nat Genet* **48**, 1517-1526 (2016).
76. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet* **16**, 421-33 (2015).
77. Pizarro, J.G. & Cristofari, G. Post-Transcriptional Control of LINE-1 Retrotransposition by Cellular Host Factors in Somatic Cells. *Front Cell Dev Biol* **4**, 14 (2016).
78. Nair, V.S., Maeda, L.S. & Ioannidis, J.P. Clinical outcome prediction by microRNAs in human cancer: a systematic review. *J Natl Cancer Inst* **104**, 528-40 (2012).
79. Meister, G. Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* **14**, 447-59 (2013).
80. Reczko, M., Maragkakis, M., Alexiou, P., Grosse, I. & Hatzigeorgiou, A.G. Functional microRNA targets in protein coding sequences. *Bioinformatics* **28**, 771-6 (2012).
81. Zhang, K. et al. A novel class of microRNA-recognition elements that function only within open reading frames. *Nat Struct Mol Biol* **25**, 1019-1027 (2018).
82. Grossman, R.L. et al. Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med* **375**, 1109-12 (2016).
83. MacLennan, M. et al. Mobilization of LINE-1 retrotransposons is restricted by Tex19.1 in mouse embryonic stem cells. *Elife* **6**(2017).
84. Morrish, T.A. et al. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet* **31**, 159-65 (2002).
85. Ostertag, E.M., Prak, E.T., DeBerardinis, R.J., Moran, J.V. & Kazazian, H.H., Jr. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res* **28**, 1418-23 (2000).
86. Wei, W., Morrish, T.A., Alisch, R.S. & Moran, J.V. A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. *Anal Biochem* **284**, 435-8 (2000).
87. Xie, Y., Rosser, J.M., Thompson, T.L., Boeke, J.D. & An, W. Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic Acids Res* **39**, e16 (2011).
88. Richardson, S.R., Narvaiza, I., Planegger, R.A., Weitzman, M.D. & Moran, J.V. APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *Elife* **3**, e02008 (2014).
89. Heckman, K.L. & Pease, L.R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc* **2**, 924-32 (2007).
90. Banuelos-Sanchez, G. et al. Synthesis and Characterization of Specific Reverse Transcriptase Inhibitors for Mammalian LINE-1 Retrotransposons. *Cell Chem Biol* (2019).
91. Macia, A. et al. Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Res* **27**, 335-348 (2017).
92. Kopera, H.C., Moldovan, J.B., Morrish, T.A., Garcia-Perez, J.L. & Moran, J.V. Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase. *Proc Natl Acad Sci U S A* **108**, 20345-50 (2011).
93. Coufal, N.G. et al. L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127-31 (2009).
94. Goodier, J.L., Ostertag, E.M., Du, K. & Kazazian, H.H., Jr. A novel active L1 retrotransposon subfamily in the mouse. *Genome Res* **11**, 1677-85 (2001).
95. Sugano, T., Kajikawa, M. & Okada, N. Isolation and characterization of retrotransposition-competent LINEs from zebrafish. *Gene* **365**, 74-82 (2006).

96. Lian, S.L. et al. The C-terminal half of human Ago2 binds to multiple GW-rich regions of GW182 and requires GW182 to mediate silencing. *RNA* **15**, 804-13 (2009).

FIGURE LEGENDS

Figure 1. Downregulation of let-7 and miR-34a miRNAs correlates with L1 retrotransposition accumulation in lung tumor samples. (a) Schematic representation of the bioinformatic analysis used to identify differentially expressed miRNAs in lung cancer samples with or without tumor specific L1 insertions. **(b)** A graph plot representing the expression levels of miRNAs previously associated with lung cancer[53] in lung tumor samples without (dark grey, N=14) and with (light grey, N=27) tumor specific L1 insertions identified by MELT. Differentially expressed miRNAs are marked with * and were identified applying an unpaired two-tailed t test adjusted by FDR<0.01. To enable representation of all miRNAs in one graph, expression in reads per million (rpm) was relative to the maximum value of each miRNA in each case. Whiskers were calculated using the Tukey method. Individual black dots represent outliers. Boxes extend from 25th to 75th percentiles, and lines in the middle of the boxes represent the median.

Figure 2. Let-7 regulates engineered human LINE-1 retrotransposition. (a) Left panel: rationale of the retrotransposition assay in cultured cells. From left to right: transcription start site in the 5'UTR (black arrow), the two L1 open reading frames ORF1 (brown rectangle) and ORF2 (blue rectangle), the antisense-oriented reporter cassette (white rectangles, backward REP) interrupted by an intron, and the reporter gene promoter (inverted black

arrow). Black lollipops represent poly(A) signals. TSD: Target Site Duplications. SD: Splicing Donor. SA: Splicing Acceptor. Right panel: reporter cassettes used in this study: neomycin (NEO or *mneoI*) and blasticidin (BLAST or *mblastI*) resistance, enhanced green fluorescent protein (EGFP or *megfpI*), and firefly luciferase (FLUC or *mflucI*). **(b and c)** Structures of pJM101/L1.3 and pU6iNeo constructs are shown. **(b)** HeLa cells were cotransfected with one of the plasmids and let-7a/miR-34a mimic and their control (scr). **(c)** HeLa cells were cotransfected with one of the plasmids and let-7 inhibitor and their control (c-). In **(b-c),** a representative well of three replicates is shown. Quantification of each experiment is shown at the right as average of three replicates ± s.d. **(d)** Structures of pYX017 and pYX015 are shown. CAG: Chicken Actin Globin promoter. The '*' symbol in pYX015 indicates the two point-mutations in L1-ORF1p that abolish retrotransposition. Lung cancer cell lines A549 and SK-MES-1 were co-transfected with pYX017 or pYX015 and let-7 inhibitor or its control (c-). Luciferase activity was measured 96h post-transfection, and Firefly luciferase signal was normalized to Renilla luciferase to correct for differences in transfection efficiency or cell survival. In both cases, averages of three replicates ± s.d. are shown. * in graphs, denotes statistical significance (p value < 0.05). RLU: Relative Luminescence Units.

Figure 3. Let-7 binds directly to the coding sequence of L1 mRNA. (a) Cell culture based retrotransposition assay using *mneoI* reporter cassette. HeLa cells

1135 were co-transfected with LINEs from different species and let-7 inhibitor or its
1136 control (c-). All constructs have an exogenous CMV promoter to normalized
1137 transcription. Black arrows represent transcription start sites. Grey triangles in
1138 mouse and zebrafish LINEs illustrate the presence of monomers in the 5'UTR of
1139 these elements. Stem loop (grey) pictures the hairpin structures present in the 3'
1140 UTR of the zebrafish LINE-2s, required for retrotransposition. White stripes are
1141 included to remark the differences in sequence of zebrafish L2-2 and L2-1 with
1142 respect to the human L1.3 and mouse $L1G_F$. A representative well of three
1143 biological replicates is shown in the middle panel. Quantification is shown on
1144 the right as average ± s.d. **(b)** Cell culture-based retrotransposition assay with
1145 blasticidicin resistance cassette. HeLa cells were co-transfected with LINEs
1146 lacking either the 5' or the 3' UTR (structures shown in the left) and let-7
1147 inhibitor or its control (c-). A representative well of three replicates is shown.
1148 Quantification is shown at the right as average of three replicates ± s.d. In **(a-b)**,
1149 '*' indicates $p<0.05$. **(c)** Scheme of RNA Immunoprecipitation (RIP) of AGO2-
1150 FLAG and RT-qPCR analysis of endogenous mRNA enrichment upon let-7
1151 overexpression. Embryonic teratocarcinoma cells (PA-1) were co-transfected
1152 with a plasmid to overexpress AGO2-FLAG and let-7 mimic. AGO2-FLAG
1153 (orange circle with green flag) was immunoprecipitated with a FLAG antibody
1154 (black circle and lines), and the RNA bound to AGO2 (L1 mRNA is shown in
1155 blue) was purified and analyzed by RT-qPCR. Transfection with pBSKS (empty
1156 vector) was used as a negative control. **(d)** Real-time RT-qPCR analysis of

endogenous L1 mRNA upon immunoprecipitation of AGO-2-FLAG of one representative experiment of three replicates. Left panel: loading controls are shown for input and IP. Right panel: mRNA relative enrichment upon let-7 overexpression: LINE-1 (blue), HMGA2 (yellow), GAPDH (red), and ACTIN (grey) are shown.

Figure 4. L1 mRNA contains a functional let-7 binding site located in L1-ORF2. (a) RNAhybrid prediction of the best binding site for let-7 located in the L1 sequence. Base-pairing between this region and let-7b is shown (green rectangle). Localization of the putative binding site ('bs2rh') within L1 sequence is shown (green line). Structure of LINE-1 is shown: transcription start site (black arrow), 5' untranslated region (UTR), ORF1, ORF2 with its three domains endonuclease (EN), reverse transcriptase (RT), and cysteine-rich (C), and 3' UTR. Folding energy of the predicted binding site is shown below. **(b)** psiCHECK2 assay with 'bsrh' and 'bs2rhmut'. Left panel: scheme of psiCHECK2 plasmid containing SV40 promoter, Renilla luciferase gene (RLuc, orange rectangle), different sequences cloned in the 3'UTR of the RLuc gene (grey rectangle), and Firefly luciferase gene (FLuc, brown rectangle). Comparison of the sequences cloned in psiCHECK2 as 'bs2rh' and 'bs2rhmut' is shown below, with the nucleotides interacting with let-7 in capital letters and the different nucleotides highlighted in red. Right panel: HEK293T cells were

co-transfected with the different psiCHECK2 constructs and scr or let-7 mimic. Error bars represent s.d. of three replicates. RLU: Relative Luminescence Units. **(c)** Scheme of the wild type L1.3 and the binding site mutant 'bs2rhmut L1.3' generated by site-directed mutagenesis. Red thunder indicates location of the mutated binding site. **(d)** HeLa cells were co-transfected with JM101/L1.3 or JM101/bs2rhmut L1.3 and let-7 inhibitor or its control (c-). A representative well of three replicates is shown. Quantification is shown as retrotransposition rate (relative to c-) and raw colony count, in both cases as average of three replicates ± s.d. * denotes $p<0.05$ after applying a t-test.

Figure 5. Let-7 impairs L1-ORF2p translation. (a) RT-qPCR analysis of endogenous LINE-1 (blue bar), HMGA2 (purple bar), DICER (green bar), and ACTIN (orange bar) mRNAs upon let-7 overexpression in HEK293T. Cells were transfected with let-7 mimic or its control (scr), and RNA was extracted at 24h or 48h post-transfection. GAPDH was used to normalize. Error bars indicate s.d. of three replicates. **(b)** Western blot analyses of endogenous L1-ORF1p and HMGA2 protein levels in HEK293T cells upon let-7 overexpression. Cells were transfected with let-7 mimic. Representative well and quantification of the western blot are shown. Error bars indicate s.d. of four biological replicates. **(c)** Western blot analyses of L1-ORF2p-FLAG upon let-7 overexpression or

depletion in HeLa cells. A scheme of construct pSA500 is shown. HeLa cells were co-transfected with pSA500 and let-7 mimic or inhibitor and their controls (scr or c- respectively). L1-ORF2p was detected using a FLAG antibody. DICER, a known let-7 target, was used as a positive control. Representative well and quantification of the western blot are shown. Error bars represent s.d of three replicates. Below, RT-qPCR analyses of the levels of DICER and L1-ORF2-3xFLAG mRNA upon overexpression or depletion of let-7. GAPDH was used to normalize. Error bars represent s.d of three replicates. **(d)** Flow cytometry quantification of L1-ORF2p-mCherry levels upon let-7 depletion in HeLa cells. The structure of construct pVan583 is shown. pVan583 is a derivative of JM101/L1.3 with L1-ORF1p fused to EGFP and L1-ORF2p fused to mCherry both at the C-terminus. HeLa cells were co-transfected with pVan583 and let-7 inhibitor or its control (c-), and fluorescence was measured by flow cytometry. Graph shows the percentage of mCherry+ cells in the EGFP+ (transfected) population. Error bars indicate s.d. of three replicates. A representative FACS histogram of three replicates in each condition is shown (the percentage of ORF1p-GFP positive cells expressing ORF2p-Cherry protein).

Figure 6. Model for the control of LINE-1 retrotransposition by the tumor suppressor microRNA let-7. Besides the well-known regulation of oncogenes

(lower part of the scheme in grey), we propose a novel tumor-suppressor role for let-7 microRNAs (upper part of the scheme in colour). Once LINE-1 RNA is transcribed from an active L1 located in the genome and is exported into the cytoplasm, let-7 (drawn in red) binds and guides the RISC complex (grey circle) to the L1 mRNA (blue line). This binding leads to the inhibition of ORF2p translation (blue circle) and consequently, impairs the formation of the ribonucleoparticle (ORF1p (brown circle) and ORF2p (blue circle)). The reduction in ORF2p levels results in a decrease in the reverse transcriptase activity in the nucleus and the number of new L1 copies integrated in the genome, consequently, reduces the L1-associated genomic instability.

Supplementary Figure 1. Controls to corroborate the correlation between let-7 and miR-34a expression levels and accumulation of tumor-specific L1 insertion in lung tumor samples. (a) Analysis in Figure 1 was repeated after randomly re-assigning the value of tumor-specific L1 insertions to the samples, showing no significant correlation with miRNA levels. **(b)** RNA-seq analysis showed that L1Hs is overexpressed in lung tumor samples with tumor-specific L1 insetions (left panel). Pair-wise correlations between expression levels of L1Hs and let-7a, let-7e, let-7f-2 and miR-34a (Pearson' *r*). *P*-value was considered significant if <0.05. **(c)** Correlation between miRNA expression and tumor-specific L1 insertions identified by Helman and col. using Transpo-seq was analyzed as in Figures 1 and Supplementary 1a. Reduced expression of

members of the let-7 family and miR-34a correlated with increased number of tumor-specific L1 insertions. **(d)** Analysis of Supplementary Fig 1c was repeated after randomly re-assigning the value of tumor-specific L1 insertions to the samples. **(e)** Analysis in **(c)** was repeated using breast samples, in which tumor-specific L1 insertions had been identified by Helman and col. using Transpo-seq. **(a, c, d, e)** A schematic representation of the bioinformatic analysis used is represented on a left panel. Differentially expressed miRNAs were identified applying an unpaired two-tailed t test adjusted by FDR<0.01. To enable representation of all miRNAs in one graph, expression (rpm) was relative to the maximum value of each miRNA in each case. Whiskers were calculated using the Tukey method. Individual black dots represent outliers. Boxes extend from 25^{th} to 75^{th} percentiles, and lines in the middle of the boxes represent the median.

Supplementary Figure 2. Engineered human L1 retrotransposition increases and decreases upon let-7 depletion or overexpression, respectively, in a variety of cell lines. (a) RT-qPCR quantification of mature let-7a and mir-34a in HeLa and HEK293T cells. Averages of three replicates are shown. Error bars indicate s.d. **(b)** The structure of pJJ101/L1.3 is shown. HeLa cells were cotransfected with L1 plasmid and let-7b mimic or its control (scr). A representative well of three replicates is shown. Quantification is shown at the right as average of three replicates ± s.d. * indicates p<0.05 after applying two-

tailed t test. **(c-f)** Cell culture based retrotransposition assays with luciferase (c, d, f) or enhanced GFP (e) cassettes. In all cases, error bars indicate s.d., and * indicates $p<0.05$. In **(c)**, **(d)** and **(f)**, RLU: Relative Luminescence Units. **(c)** Structure of pYX014 and pYX015 are shown. HEK293T cells were cotransfected with one of them and let-7 mimic or its control (scr). Averages of three replicates are shown. Error bars indicate s.d. (d) HEK293T cells were cotransfected with pYX014 or pYX015 and miR-34 mimic or its control (scr). Averages of three replicates are shown. Error bars indicate s.d.**(e)** The structure of 99-UB-LRE3 is shown. HEK293T cells were cotransfected with L1 plasmid and let-7 inhibitor or its control (c-). Average of three replicates is shown. **(f)** Structure of pYX017 and pYX014 are shown. See **(c)** for pYX015 structure. HEK293T cells were cotransfected with L1 plasmid and let-7 inhibitor or its control (c-). Average of three replicates is shown. Luciferase signal is shown relative to pYX014 c-. **(g)** RT-qPCR analysis of the mature let-7a and let-7b miRNA expression levels in all cell lines used in this study. Averages of three replicates are shown. Error bars indicate s.d.

Supplementary Figure 3. Let-7 guides AGO to the coding sequence of human L1 mRNA. (a) Cell culture-based retrotransposition assay with *mneoI* tagged constructs. HeLa cells were co-transfected with LINEs from different species and let-7 mimic. For details of the constructs used see **Fig. 3a**. A representative well of three independent biological replicates is shown. Quantification is

shown in the right as average of two replicates ± s.d. **(b)** Cell culture-based retrotransposition assay using *mblastI* tagged constructs. HeLa cells were co-transfected with LINEs lacking either the 5′ or the 3′ UTR (detailed structures shown in **Fig. 3b**) and let-7 mimic or its control (scr). A representative well of four replicates is shown. Quantification is shown in the right as average of four replicates ± s.d. **(c)** Co-immunoprecipitation (co-IP) of FLAG-AGO1/2 and endogenous ORF1p in the presence/absence of RNase. PA-1 cells were co-transfected with a plasmid to overexpress AGO1 or AGO2 tagged with FLAG epitope. 'mock' condition refers to cells transfected with an empty plasmid. AGO1/2 was immunoprecipitated using a FLAG antibody in the presence or absence of RNase A. Loading controls (input) and western blot (IP) are shown. Scheme in the right illustrates the model this experiment suggests: interaction between AGO1/2 and ORF1p is RNA-dependent, therefore both proteins interact directly with L1 mRNA and are probably guided by miRNAs.

Supplementary Figure 4. Identification of a let-7 binding site in human L1-ORF2. (a) The best predicted binding site of two different algorithms were tested with psiCHECK2. Location, method used for its prediction and ΔG of its predictive binding to L1.3 are summarized in the table. Predicted pairing of these sequences with let-7a and let-7b is shown on the left. Graphs below show the results of the psiCHECK2 assays with each of them (see Fig. 4b for a detailed structure of this vector). HEK293T cells were co-transfected with three

different psiCHECK2 constructs and let-7 mimic or its control (scr). 'no bs' is a negative control (a sequence without complementarity to let-7) and 'perfect bs' is a positive control (with perfect complementarity to let-7). RLU: Relative Luciferase Units. Error bars indicate s.d. of three replicates. **(b)** RNAhybrid prediction of 'bs2rhmut' and let-7 interaction. Base-pairing between this region and let-7b is shown (green rectangle). Folding energy is shown on the right. **(c)** Alignment of the consensus sequence of L1PA1 to L1PA16 families showing conservation of the let-7 binding site ´bs2rh´ (in blue). **(d)** Alignment of the predicted binding site region in all L1s used in this study: human L1.3/L1RP/LRE3, mouse $L1G_F$, and zebrafish L2.1 and L2.2. Alignment below was performed with ORF2 protein sequences to localize the binding site region within each LINE (asterisks denote conserved aminoacids). Human and mouse L1s nucleotide sequences were further analyzed. Blue nucleotides represent the 'bs2rh' region that was predicted by RNAhybrid and validated as a binding site in psiCHECK2. Red nucleotides are those which differ in mouse $L1G_F$ respect to human L1.3, LRE3 or L1RP.

Supplementary Figure 5: Let-7 affects exogenous L1-ORF2p translation but not L1-ORF1p or L1 mRNA levels. (a-b) RT-qPCR analysis of L1 mRNA levels transcribed from a plasmid. HEK293T were transfected with JM101/L1.3 and **(a)** let-7 mimic or **(b)** let-7 inhibitor. Exogenous L1 mRNA was specifically detected using primers against the spliced neomycin-resistance cassette. HMGA2 was

used as a positive control. EBNA-1, expressed constitutively from the plasmid backbone, was used to normalize. Error bars indicate s.d. *p<0.05. **(c)** Western blot analyses of endogenous L1-ORF1p and HMGA2 protein levels upon let-7 depletion. HEK293T cells were transfected with let-7 inhibitor or its control (c-). Error bars indicate s.d. of two biological replicates. *p<0.05. **(d)** Western blot analyses of stably-expressed T7-tagged L1-ORF1p upon depletion or overexpression of let-7. Stable Flp-In-293 cells expressing T7-tagged L1-ORF1p were transfected with let-7 inhibitor or mimic, or their controls (c- and scr, respectively). HMGA2 was used as a positive control. Western blot (left) and its quantification by Odyssey (right) are shown. Error bars indicate s.d. of two replicates. *p<0.05. **(e-f)** Transfection control for western blot shown in Fig. 5c. Plasmid levels (pSA500) upon let-7 overexpression **(e)** or depletion **(f)** were analyzed by qPCR using two different pairs of primers: CMV and EBNA. Genomic GAPDH was used to normalize. Error bars represent s.d of three replicates. **(g)** Scheme of the different derivatives of pSA500 generated by insertion of different sequences in its 3'UTR: 'scrb', 'bs2rh' and '8mer'. Different nucleotides in 8mer compared to bs2rh are shown in red. **(h)** RT-qPCR analysis of ORF2-F, EBNA and DICER mRNA levels upon let-7 overexpression. HeLa cells were transfected with one of the three pSA500 3'UTR derivatives shown in **(g)** and let-7 mimic. Graph shows the RNA levels of the different mRNAs upon let-7 overexpression, relative to the scr condition, in each case. EBNA was used to normalize ORF2-F levels, and GAPDH was used to normalize EBNA and

DICER. **(i)** Western blot analyses of L1-ORF2p-FLAG upon let-7 overexpression. A fraction of cells from **(h)** were used. L1-ORF2p was detected using a FLAG antibody. DICER, a known let-7 target, was used as a positive control. Representative well and quantification of the western blot are shown. Error bars represent s.d of three replicates. P values are shown. **(j)** Scheme of pSA500 ORF2-8mer generated by mutation of two nucleotides in the sequence of bs2rh. Different nucleotides and aminoacids in 8mer compared to bs2rh are shown in red. **(k)** RT-qPCR analysis of ORF2-F, EBNA and DICER mRNA levels upon let-7 overexpression. HeLa cells were transfected with pSA500 and pSA500ORF2-8mer showed in **(j)** and let-7 mimic. Graph shows the RNA levels of the different mRNAs upon let-7 overexpression, relative to the scr condition, in each case. EBNA was used to normalize ORF2-F levels, and GAPDH was used to normalize EBNA and DICER. **(i)** Western blot analyses of L1-ORF2p-FLAG upon let-7 overexpression. A fraction of cells from **(k)** were used. L1-ORF2p was detected using a FLAG antibody. DICER, a known let-7 target, was used as a positive control. Representative well and quantification of the western blot are shown. Error bars represent s.d. of three replicates. P values are shown. **(n)** Western blot analysis of ORF1p-GFP expression from different plasmids. pCEP4 was used as a negative control. Two blots were incubated in parallel with anti-ORF1p and anti-GFP antibodies. **(o)** Cell culture-based retrotransposition assay comparing retrotransposition efficiency of different constructions. HeLa cells were transfected with either JM101/L1.3, pVan583 or

JM105/L1.3 and selected with neomycin. Quantification (corrected for transfection efficiency, shown in **(p**) is shown in the right. **(p)** Transfection efficiency comparison between JM101/L1.3 and pVan583. HeLa cells were transfected with one of the constructs, and plasmid levels were quantified by qPCR. **(o,p)** As negative control a mutant construct containing a missense mutation in the RT domain (D702A) was used (JM105/L1.3) **(q)** U2-OS cells were co-transfected with pVAN583 and let-7 mimic or its control (scr), and fluorescence was analyzed by confocal microscopy. Arrows indicate L1-ORF1p-EGFP or L1-ORF2p-mCherry foci.

Supplementary Figure 6. Uncropped versions of the western blots shown in this study.

Supplementary Table I. Summary of L1 insertions found by MELT.

Supplementary Table II. Description of tumor-specific L1 insertions found by MELT.

Supplementary Table III. Multiple t-tests of lung cancer-related miRNA expression in samples with/without insertions (found by MELT).

Supplementary Table IV. Rank-sum test of lung cancer-related miRNA expression in samples with/without insertions (found by MELT).

Supplementary Table V. Multiple t-tests of lung cancer-related miRNA expression in samples after randomization of the number of insertions (found by MELT).

Supplementary Table VI. Multiple t-tests of miRNA expression in samples with/without insertions (found by MELT), including all miRNAs expressed in lung cancer.

Supplementary Table VII. Multiple t-tests of lung cancer-related miRNA expression in samples with/without insertions (found by Transpo-Seq).

Supplementary Table VIII. Multiple t-tests of miRNA expression in samples with/without insertions (found by Transpo-Seq), including all miRNAs expressed in lung cancer.

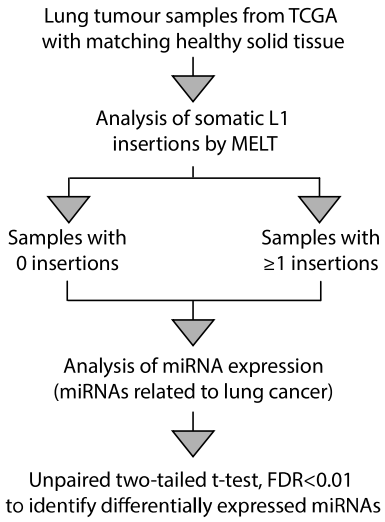
Supplementary Table IX. Multiple t-tests of lung cancer-related miRNA expression in samples after randomization of the number of insertions (found by Transpo-Seq).

Supplementary Table X. Multiple t-tests of lung cancer-related miRNA expression in breast cancer samples with/without insertions (found by Transpo-Seq).

Supplementary Table XI. Primers used in this study.

Fig 1

a



b

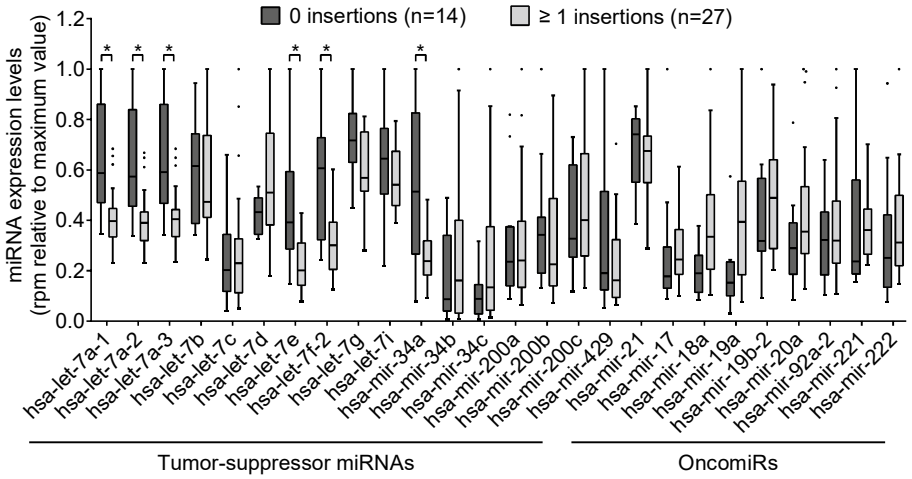
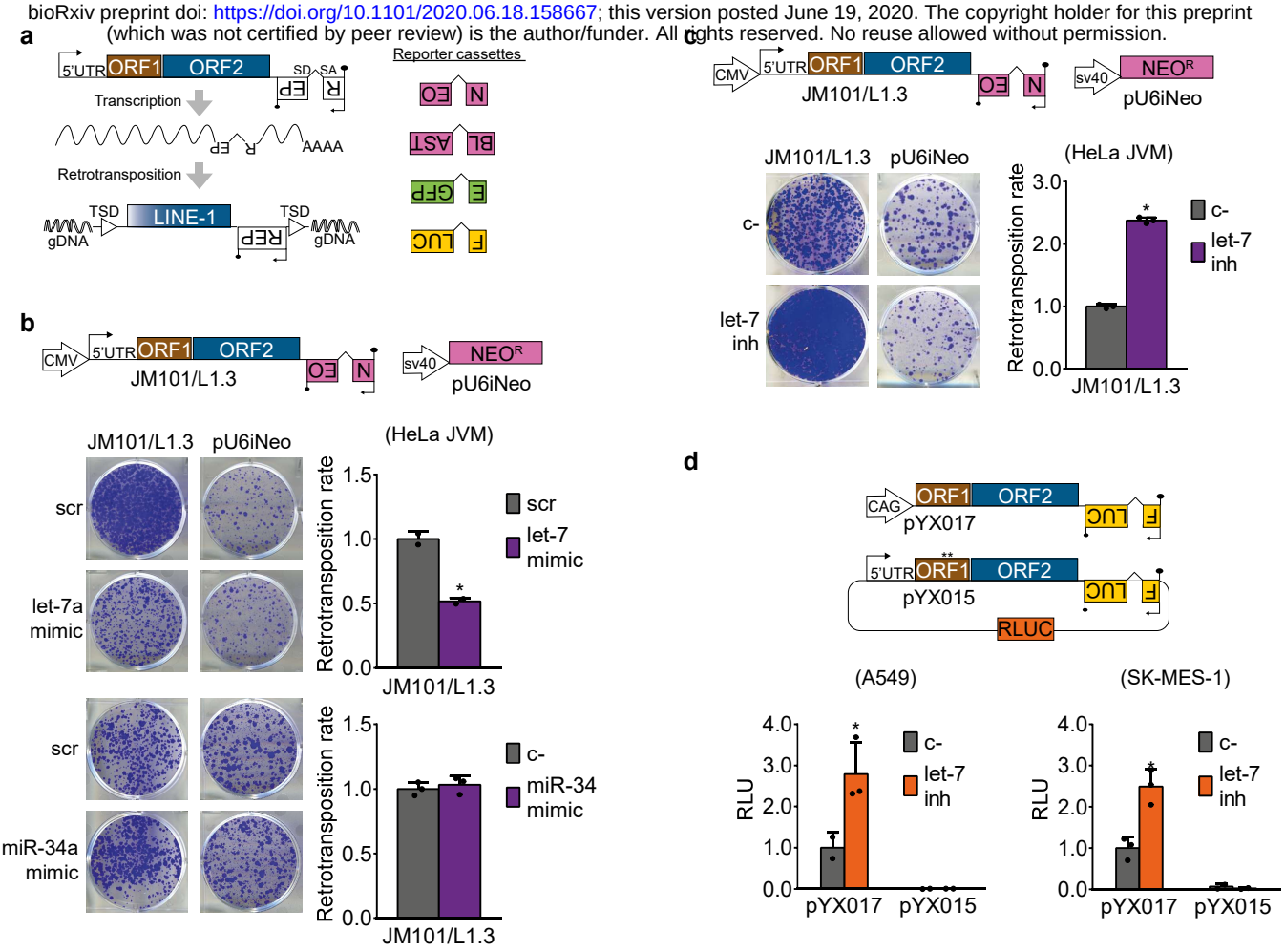
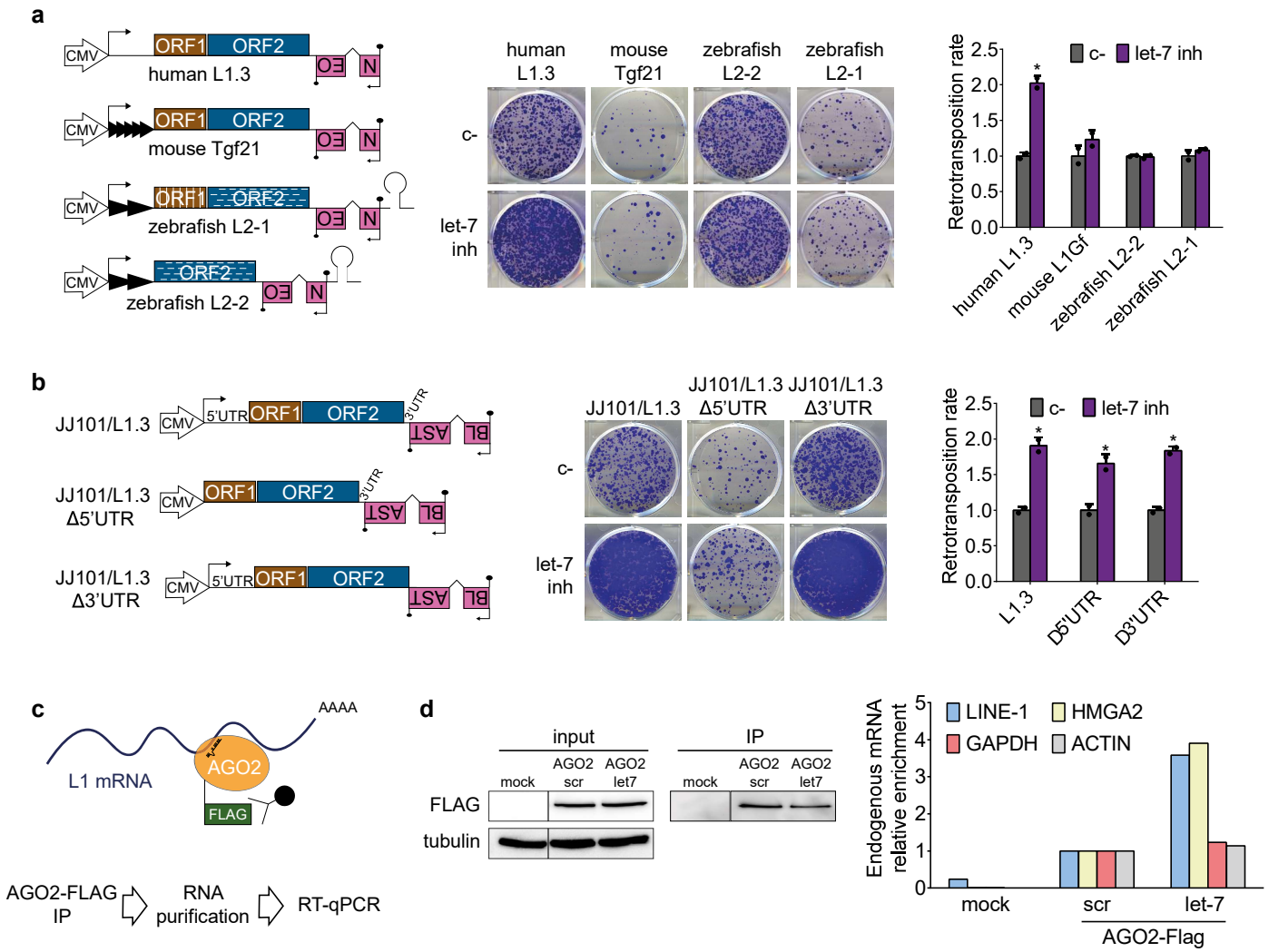
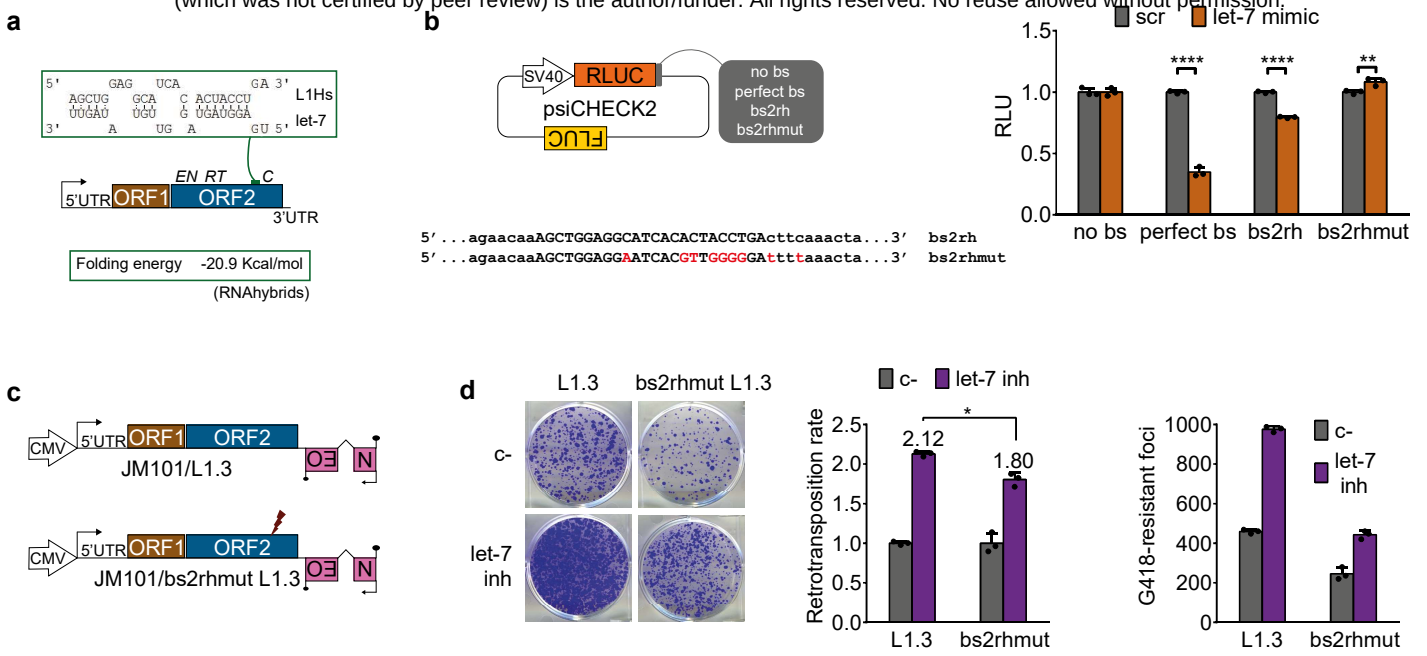


Fig 2







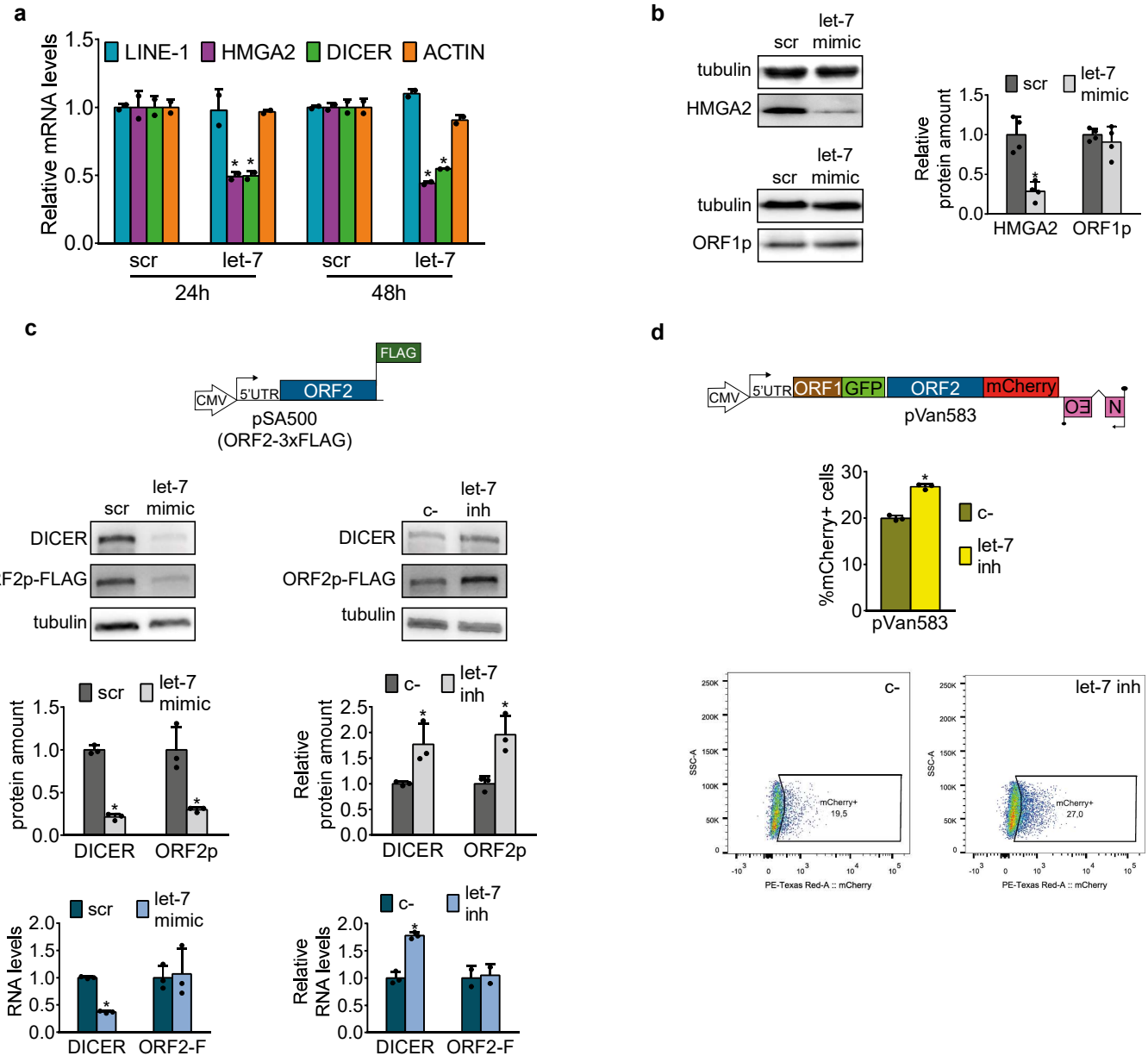
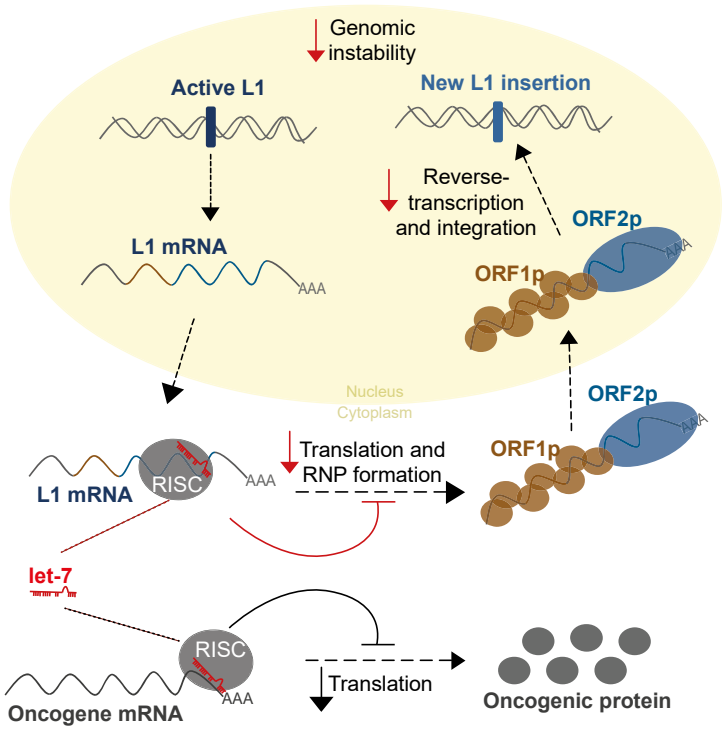
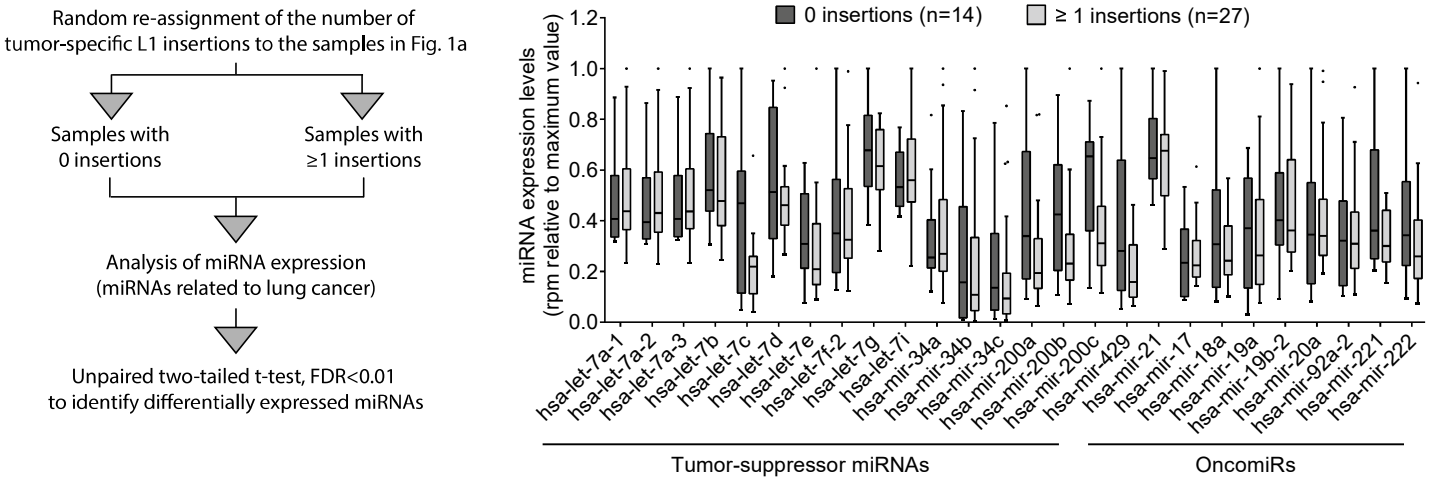


Fig 6

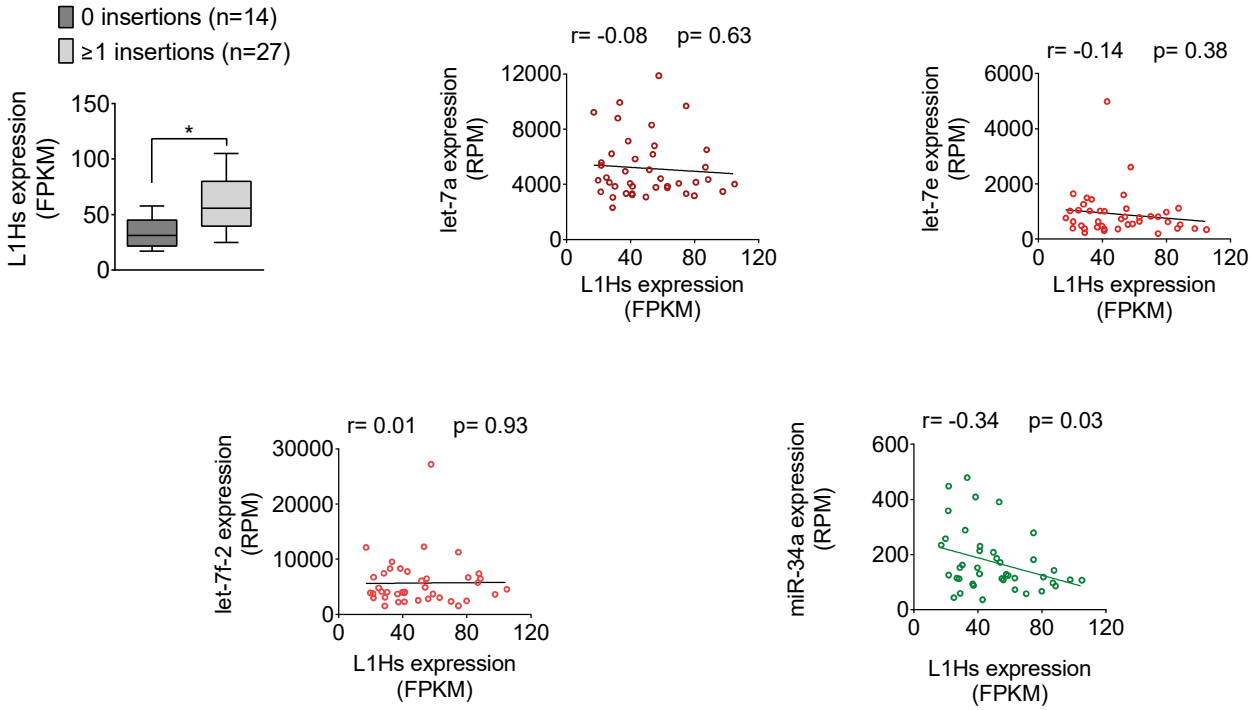


Supp Fig 1

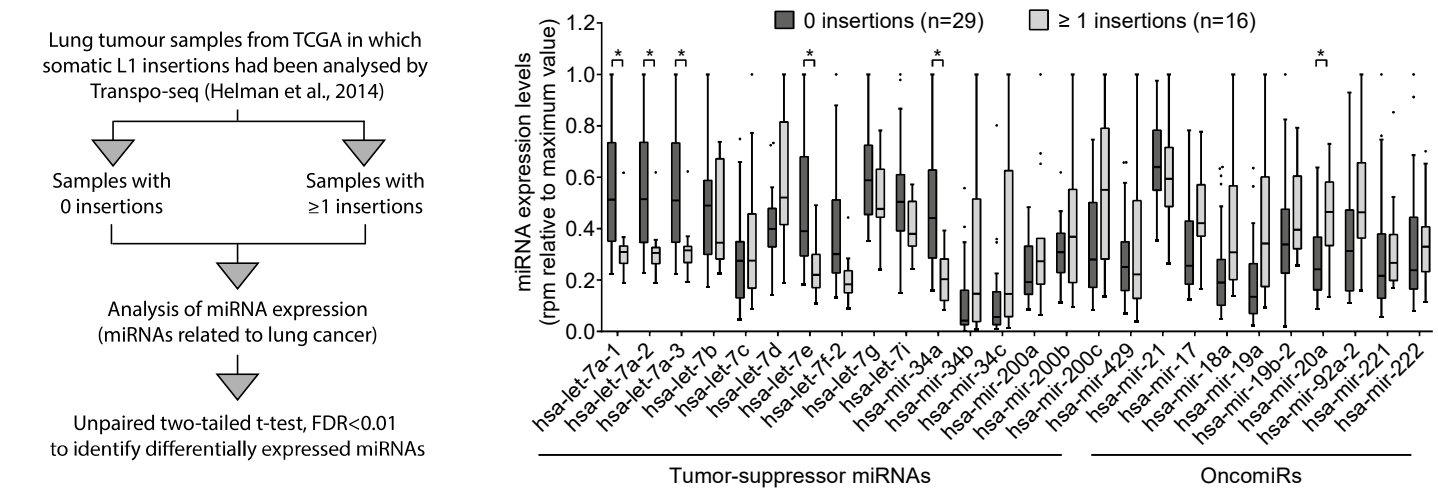
a



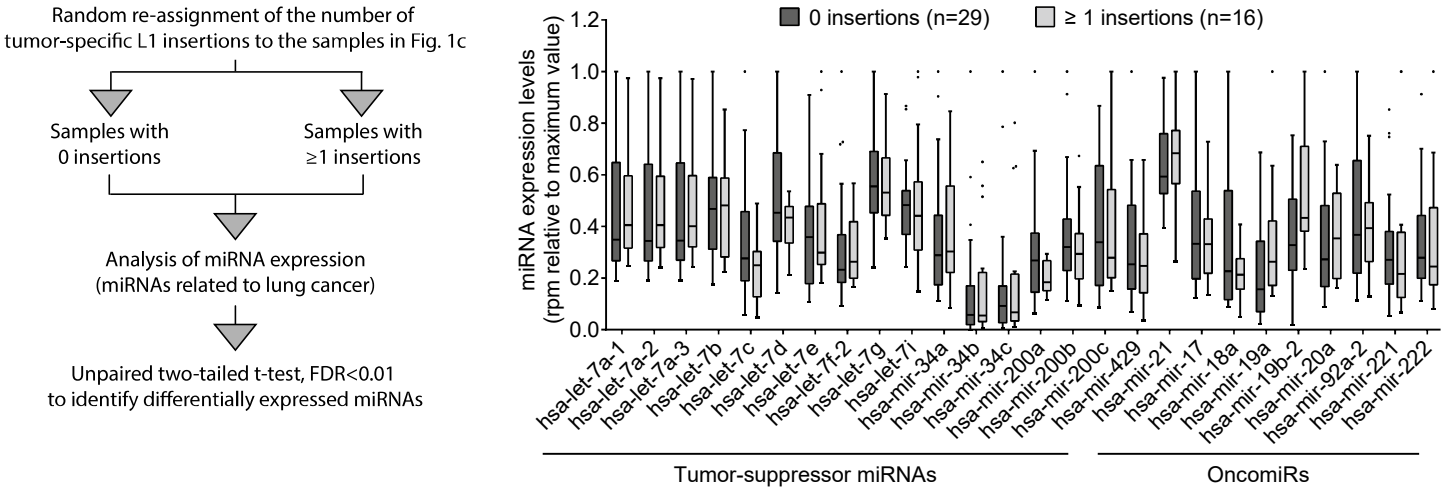
b



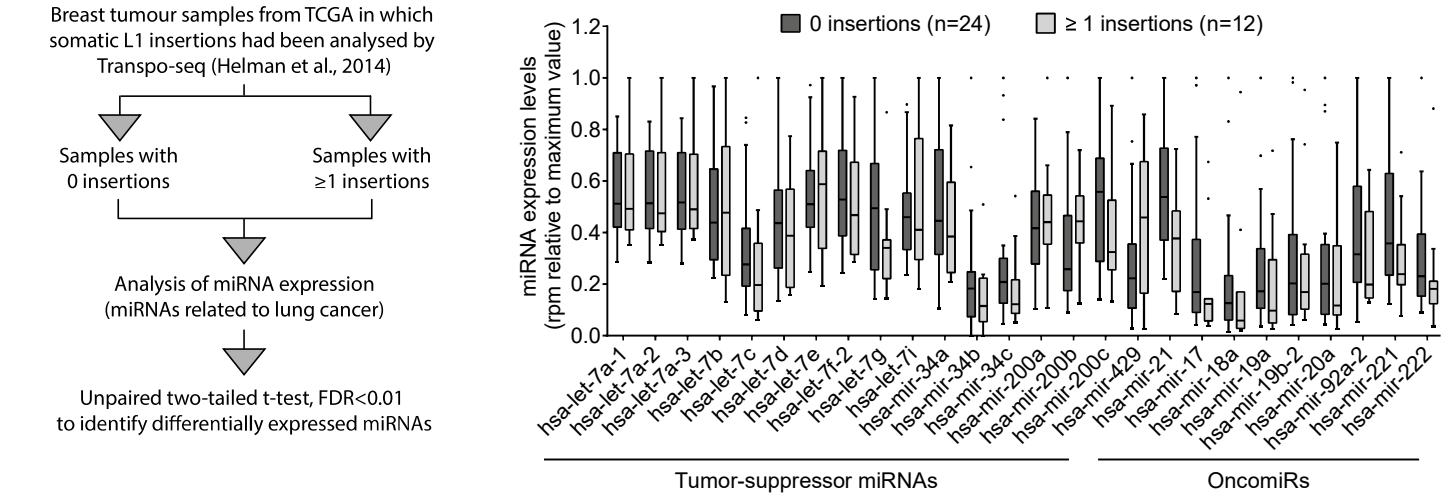
c



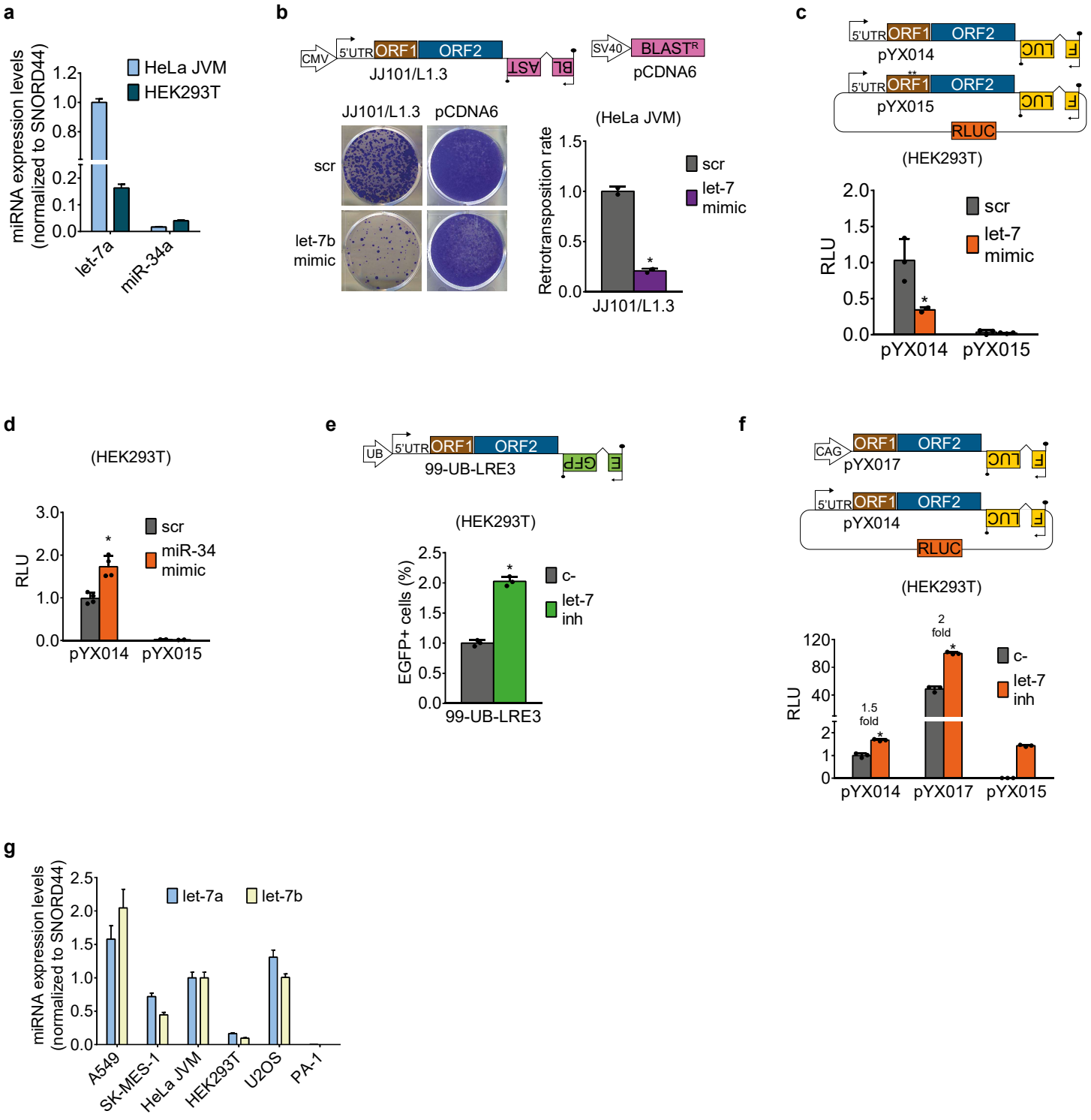
d



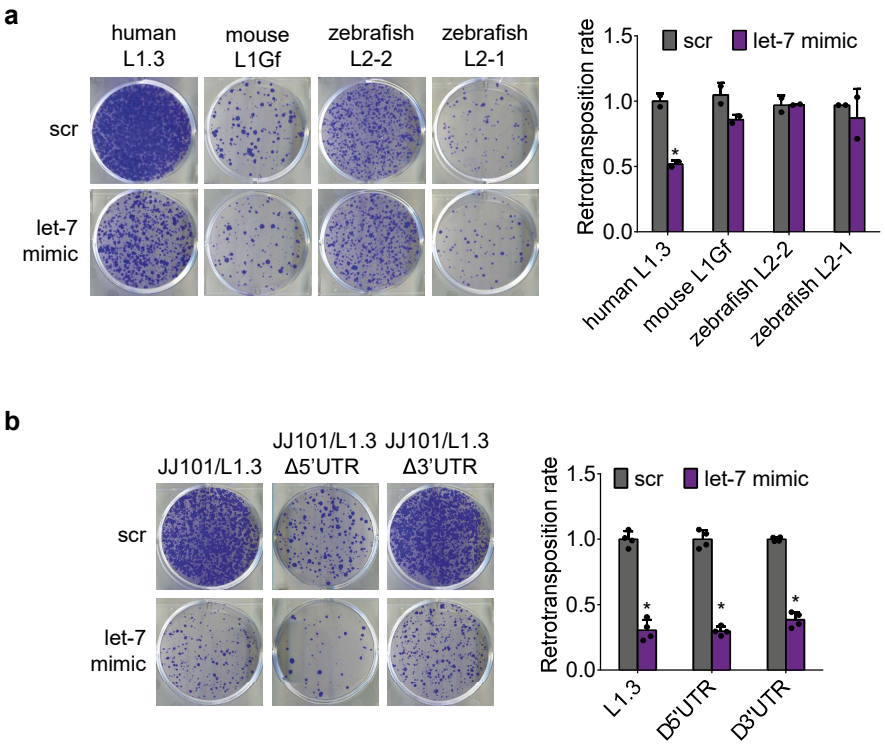
e



Supp Fig 2

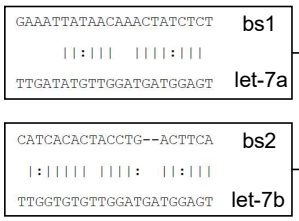


Supp Fig 3

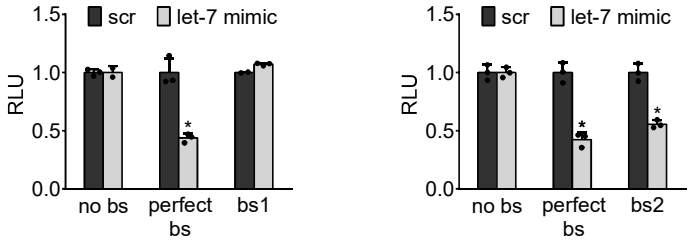


Supp Fig 4

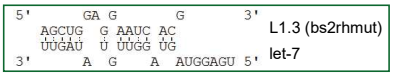
a



Predicted binding site	Location (Pos in L1Hs)	Prediction method/software	dG	Validated by siCHECK?
bs (1)	ORF2 (2650-2671)	mIRanda	-17.5	NO
bs (2)	ORF2 (4596-4616)	RNA22	-19.4	**YES**



b



Folding energy -9.7 Kcal/mol
(RNAhybrids)

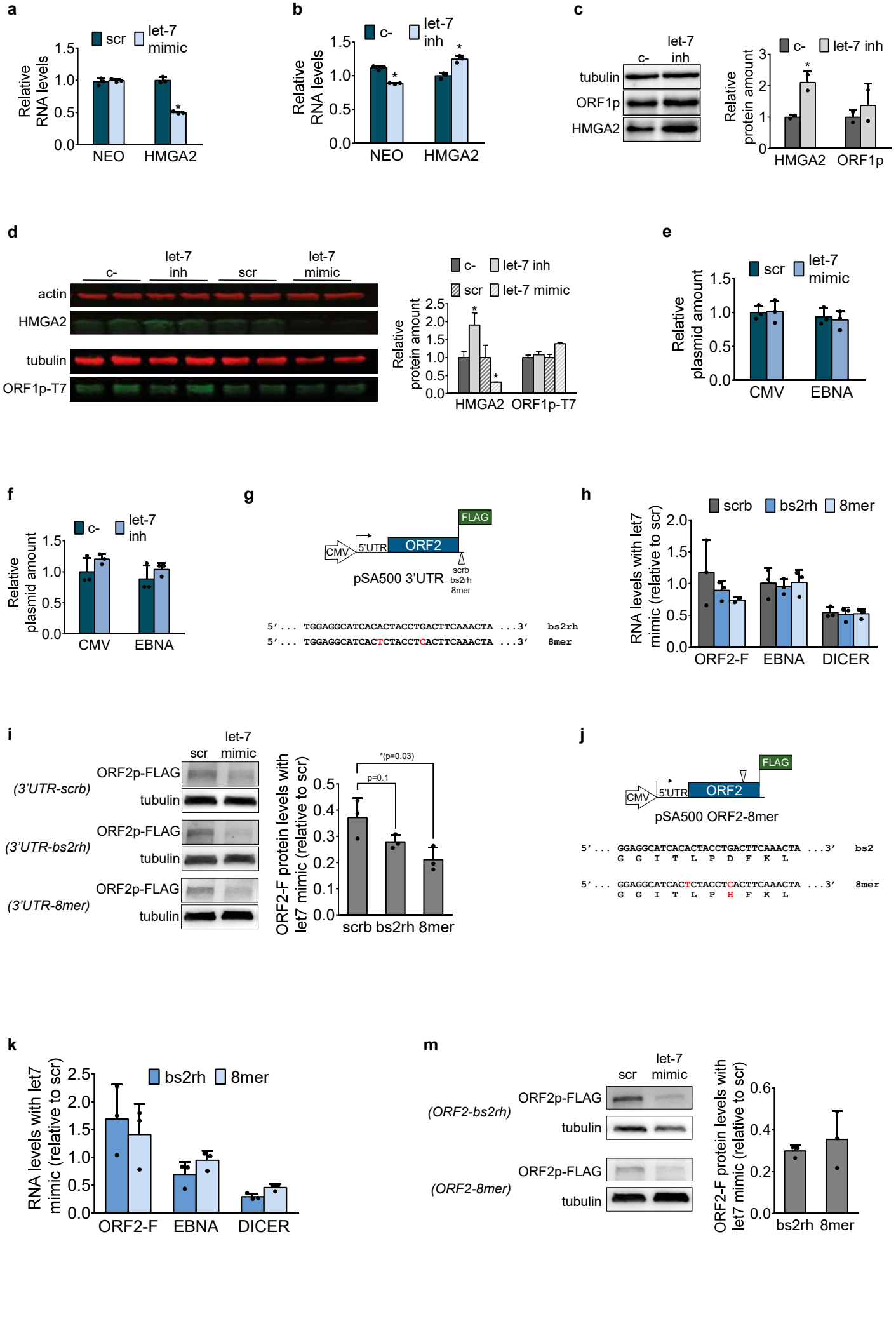
c

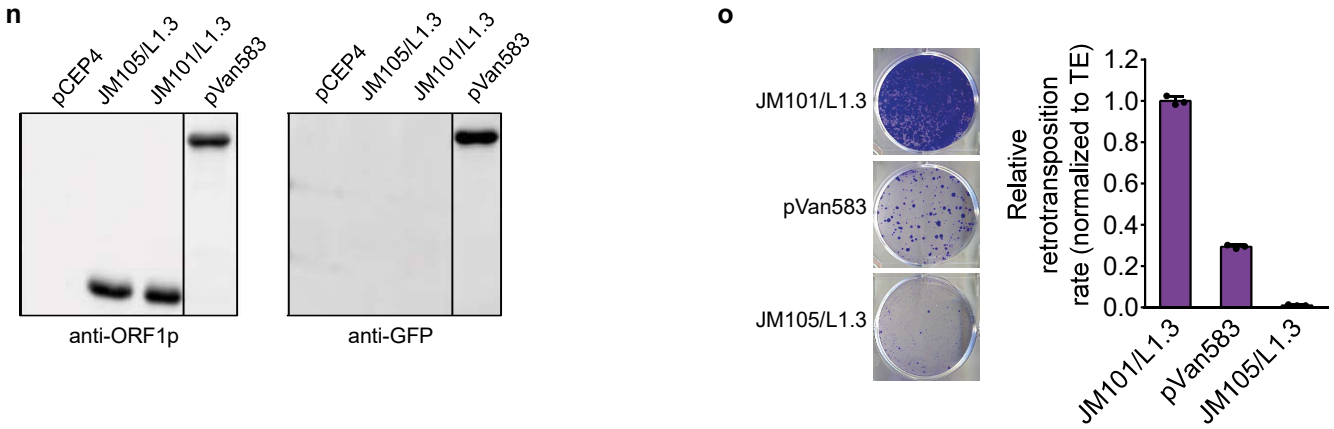
```
L1PA1       AAAAAGAACAAAGCTGGAGGCATCACACTACCTGACTTCAAACTAT 4618
L1PA2       AAAAAGAACAAAGCTGGAGGCATCACACTACCTGACTTCAAACTAT 4619
L1PA3       AAAAAGAACAAAGCTGGAGGCATCACACTACCTGACTTCAAACTAT 4749
L1PA4       AAAAAGAACAAAGCTGGAGGCATCACACTACCTGACTTCAAACTAT 4746
L1PA5       AAAAAGAACAAAGCTGGAGGCATCACACTACCTGACTTCAAACTAT 4742
L1PA6       AAAAAGAACAAAGCTGGAGGCATCACGCTACCTGACTTCAAACTAT 4733
L1PA7       AAAAAGAACAAAGCTGGAGGCATCACGCTACCTGACTTCAAACTAT 5065
L1PA8       AAAAAGAACAAAGCTGGAGGCATCACGCTACCTGACTTCAAACTAT 5059
L1PA8A      AAAAAGAACAAAGCTGGAGGCATCACGCTACCTGACTTCAAACTAT 5047
L1PA10      AAAAAGAACAAAGCTGGAGGCATCACGCTACCCGACTTCAAACTAT 4987
L1PA11      AAAAAGAACAAAGCTGGAGGCATCACGCTACCTGACTTCAAACTAT 5118
L1PA12      AAAAAGAACAAAGCTGGAGGCATCACGCTACCCGACTTCAAACTAT 6380
L1PA13A     AAAAAGAACAAAGCTGGAGGCATCACGCTACCTGACTTCAAACTAT 5336
L1PA13B     AAAAAGAACAAAGCTGGAGGCATCACACTACCTGACTTCAAACTAT 5100
L1PA14      AAAAAGAACAAAGCTGGAGGCATCACATTACCTGACTTCAAACTAT 5896
L1PA15A     AAAAAGAACAAAGCTGGAGGCATCACATTACCTAACTTCAAACTAT 4796
L1PA15B     AAAAAGAACAAAGCTGGAGGCATCACATTACCTAACTTCAAACTAT 5203
L1PA16      AAAAAGAACAAAGCTGGAGGCATCACATTACCCGACTTCAAACTAT 5431
            ************************** **** ************
```

d

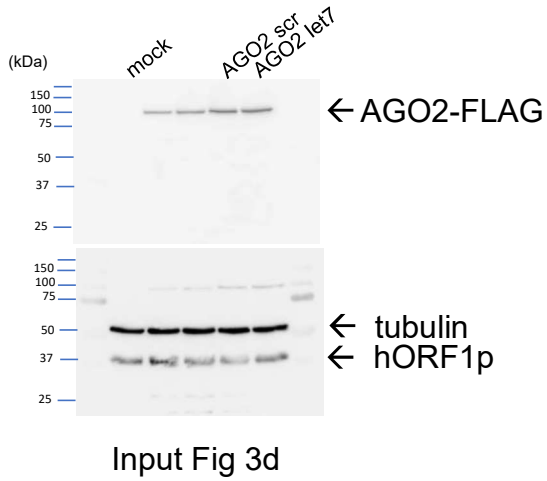
```
Human_L1.3    AAGAACAAAGCTGGAGGCATCACACTACCTGACTTCAAACTATACTAC
Human_LRE3    AAGAACAAAGCTGGAGGCATCACACTACCTGACTTCAAACTATACTAC
Human_L1RP    AAGAACAAAGCTGGAGGCATCACACTACCTGACTTCAAACTATACTAC
Mouse_L1Gf    AAAAGAACCTCTGGTGGAATCACCATGCCTGACCTAAAGCTTTACTAC
              ** *  *   **** ** *****  * ****** * ** ** ******
```

```
Human_L1.3-ORF2p      ELEKTTLKFIWNQKRARIAKSILSQKNKAGGITLPDFKLYYKATVTKTAWYWYQNRDIDQ 898
Human_LRE3-ORF2p      ELEKTTLKFIWNQKRARIAKSILSQKNKAGGITLPDFKLYYKATVTKTAWYWYQNRDIDQ 898
Human_L1RP-ORF2p      ELEKTTLKFIWNQKRARIAKSILSQKNKAGGITLPDFKLYYKATVTKTAWYWYQNRDIDQ 898
Mouse_L1Gf-ORF2p      ELEGAICKFIWNNKKPRIAKTLLKDKRTSGGITMPDLKLYYRAIVIKTAWYWYRDRQVDQ 905
Zebrafish_L2.1-ORF2p  HITPILSSLHWLPVKFRIEFKIL--------------LLTYKALNNLAPVYLTNL----- 797
Zebrafish-L2.2-ORFp   HVTPLLVRLHWLPVAARIKFKTL--------------MFAYKVTSGLAPSYLHSL----- 893
                      .:      : *     **  . *               : *:.    :  *
```

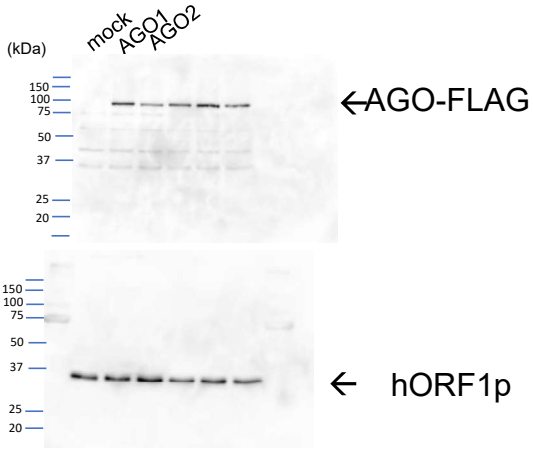




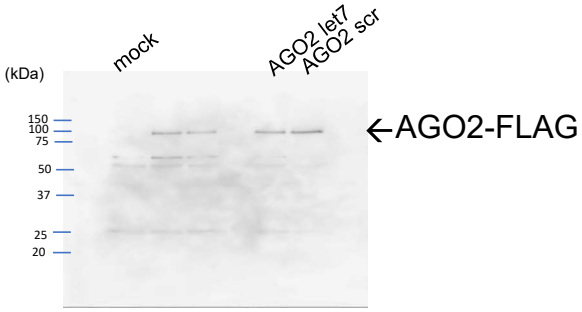
Supp Fig 6



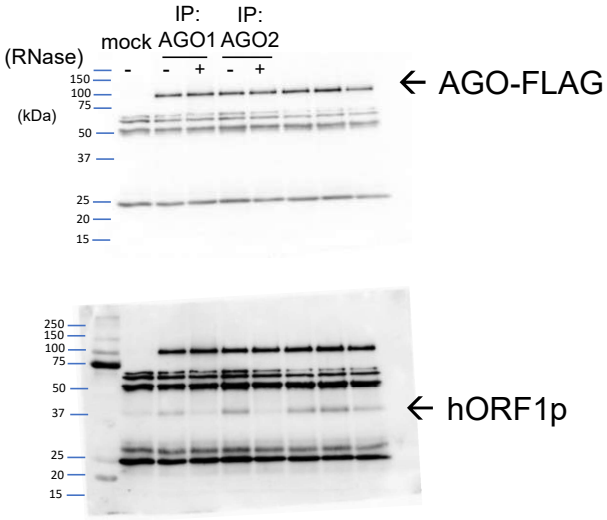
Input Fig 3d



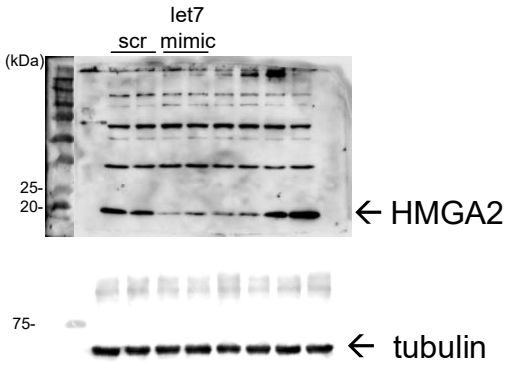
Input Supp Fig 3c



IP Fig 3d



IP Supp Fig 3c



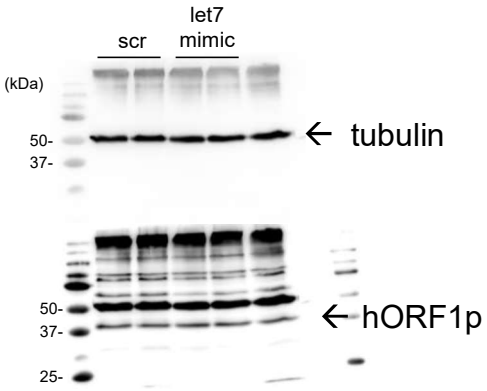
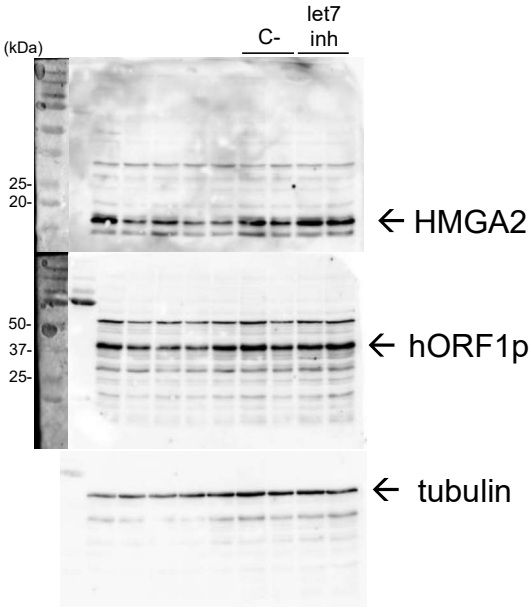
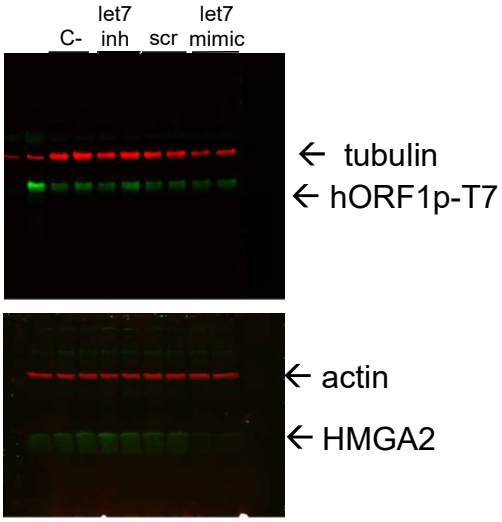


Fig 5b

Supp Fig 6



Supp Fig 5c



Supp Fig 5d

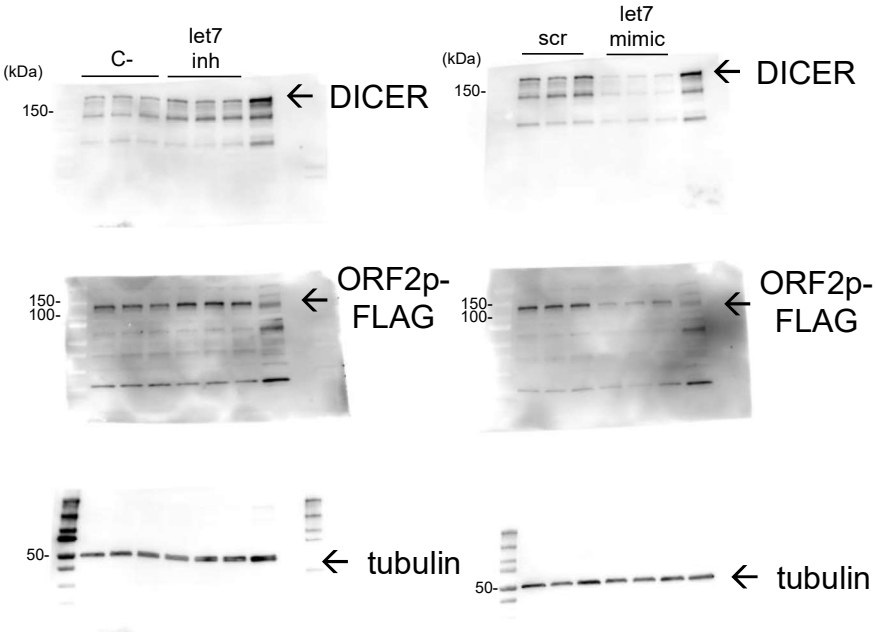
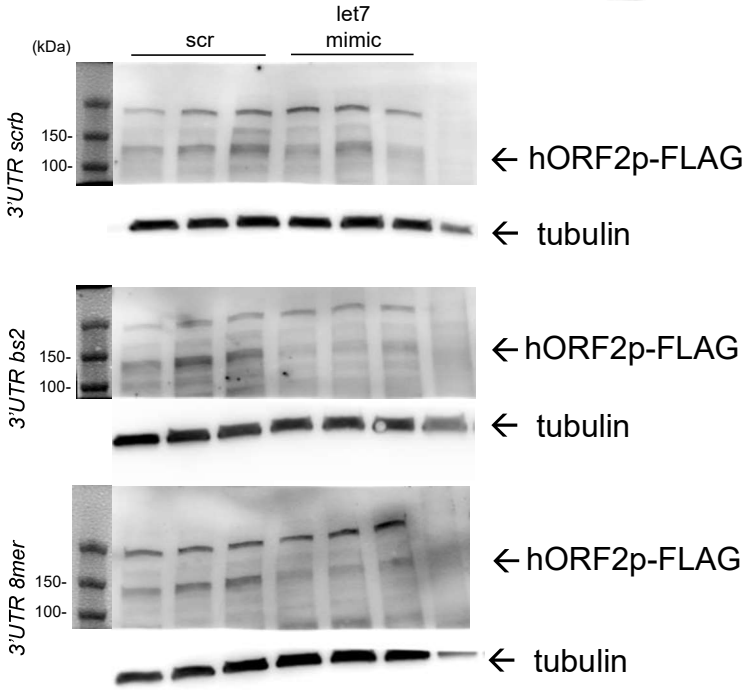
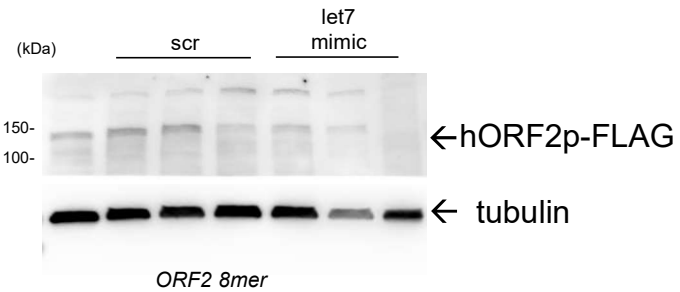


Fig 5c



Supp Fig 5i



	Types of tumor	Coverages		Unfiltered Poly L1			MELT insertion calls					
Sample		*PT*	*NT*	*PT*	*NT*	*Common*	*PT*	*NT*	*NT-PT*	*Poly PT*	*Poly NT*	*Poly Common*
TCGA-18-3408	*LUSC*	70.3	34.5	138	136	98	9	0	0	48	43	43
TCGA-18-3415	*LUSC*	76.0	38.0	141	134	95	26	0	0	35	34	31
TCGA-18-4721	*LUSC*	69.5	44.6	138	141	102	16	0	1	56	57	54
TCGA-21-1076	*LUSC*	54.0	38.3	133	126	93	5	0	0	29	28	27
TCGA-21-1078	*LUSC*	44.4	45.3	118	120	78	0	1	0	14	15	14
TCGA-21-1083	*LUSC*	89.3	65.4	129	130	105	29	0	1	50	50	48
TCGA-22-5477	*LUSC*	56.0	44.2	126	131	94	17	0	0	45	47	42
TCGA-22-5485	*LUSC*	95.4	41.1	133	136	97	19	0	0	53	55	50
TCGA-22-5492	*LUSC*	81.0	32.8	136	140	103	20	0	0	45	47	42
TCGA-33-4586	*LUSC*	70.3	37.9	134	144	91	12	0	0	38	42	35
TCGA-34-2596	*LUSC*	51.8	45.6	213	136	84	55	0	1	37	41	32
TCGA-34-2600	*LUSC*	37.9	62.2	110	132	70	6	0	0	27	33	23
TCGA-43-3394	*LUSC*	47.2	46.4	138	147	108	5	0	0	26	30	23
TCGA-43-3920	*LUSC*	71.9	35.2	110	113	81	8	0	0	29	30	26
TCGA-56-1622	*LUSC*	44.3	41.9	111	124	78	3	0	0	17	25	16
TCGA-60-2698	*LUSC*	63.6	34.2	125	124	89	34	0	1	27	28	24
TCGA-60-2711	*LUSC*	47.1	52.9	143	137	99	5	0	0	29	28	27
TCGA-60-2713	*LUSC*	46.6	51.9	122	119	80	3	1	0	15	19	13
TCGA-60-2719	*LUSC*	62.5	29.0	125	123	90	4	0	0	11	12	9
TCGA-60-2724	*LUSC*	62.8	49.7	125	131	93	3	0	0	24	26	23
TCGA-66-2744	*LUSC*	66.3	51.2	130	128	99	4	0	0	38	36	36
TCGA-66-2759	*LUSC*	63.9	51.6	127	129	92	39	0	0	52	53	49
TCGA-66-2766	*LUSC*	46.8	44.1	133	125	82	9	0	0	15	11	11
TCGA-66-2789	*LUSC*	83.6	37.0	130	127	89	0	0	0	37	33	32
TCGA-66-2793	*LUSC*	73.9	35.4	129	143	91	54	0	0	42	43	41
TCGA-66-2795	*LUSC*	61.8	34.8	122	126	82	14	0	0	28	27	24
TCGA-38-4628	*LUAD*	84.5	34.6	125	121	75	0	0	0	13	12	10
TCGA-38-4630	*LUAD*	48.5	36.2	117	127	82	6	0	0	20	21	19
TCGA-49-4486	*LUAD*	34.7	35.3	106	113	70	0	0	0	10	11	10
TCGA-49-4510	*LUAD*	41.0	30.8	132	122	83	0	0	0	11	11	8
TCGA-49-4512	*LUAD*	55.8	40.5	125	126	92	0	0	0	31	28	27
TCGA-49-6742	*LUAD*	64.2	68.2	117	137	85	3	0	0	22	26	20
TCGA-50-5930	*LUAD*	46.3	45.6	127	122	87	0	0	0	24	23	23
TCGA-50-5932	*LUAD*	40.0	45.3	109	117	86	0	0	0	29	33	29
TCGA-50-6591	*LUAD*	45.1	51.2	114	131	80	0	0	0	22	26	21
TCGA-55-6972	*LUAD*	44.8	38.3	116	108	80	5	0	0	31	33	26
TCGA-55-6982	*LUAD*	64.8	33.6	139	138	92	0	0	0	21	16	16
TCGA-55-6984	*LUAD*	33.3	41.3	109	115	75	0	1	0	32	32	26
TCGA-55-6986	*LUAD*	39.7	33.4	131	130	89	0	0	0	20	20	17
TCGA-73-4659	*LUAD*	36.4	35.7	108	115	72	0	0	0	14	15	14
TCGA-91-6847	*LUAD*	46.2	42.6	101	125	68	0	0	1	22	30	19
TOTAL INSERTIONS							**413**	**3**	**5**	**1189**	**1230**	**1080**

		Description
Sample:		Case submitter ID in GDC database
Type of tumor		
Coverage:		Average depth of coverage in bam files
	PT	Primary tumor sample coverage
	NT	Normal tissue sample coverage
Unfiltered Poly L1:		Polymorphic L1 insertions found in MELT output before filtering
	PT	Number of primary tumor L1 polymorphic insertion calls
	NT	Number of normal tissue L1 polymorphic insertion calls
	Common	Number of L1 polymorphic insertion calls found both in tumor and normal tissue
MELT insertion calls:		L1 somatic insertion calls after filtering
	PT	Number of L1 insertion calls found in primary tumor
	NT	Number of L1 insertion calls found in normal tissue
	PT-NT	Number of L1 insertion calls found both in tumor and normal tissue
	Poly PT	Number of primary tumor L1 polymorphic insertion calls after filtering
	Poly NT	Number of normal tissue L1 polymorphic insertion calls after filtering
	Poly Common	Number of L1 polymorphic insertion calls found both in tumor and normal tissue after filtering

Supp Table I

	Discovery?	P value	Mean 0 insertions	Mean >1 insertions	Difference	SE of difference	t ratio	df
hsa-let-7a-1	*	< 0,0001	6389,63	4112,88	2276,76	508,865	4,47419	37
hsa-let-7a-2	*	< 0,0001	12754,3	8199,58	4554,68	1035,16	4,4	37
hsa-let-7a-3	*	< 0,0001	6448,99	4180,23	2268,76	514,737	4,40761	37
hsa-let-7b		0,700569	15437,2	14717,5	719,724	1857,49	0,387472	38
hsa-let-7c		0,337504	1257,44	1654,59	-397,152	408,714	0,971711	37
hsa-let-7d		0,0762467	690,492	897,586	-207,094	113,452	1,8254	36
hsa-let-7e	*	< 0,0001	1202,24	584,67	617,569	133,397	4,62955	37
hsa-let-7f-2	*	< 0,0001	7173,84	3887,56	3286,28	739,198	4,44574	37
hsa-let-7g		0,00515672	711,604	562,65	148,954	50,0942	2,97348	37
hsa-let-7i		0,175339	600,312	531,649	68,6635	49,6936	1,38174	37
hsa-mir-34a	*	< 0,0001	258,375	121,504	136,871	30,2276	4,52802	38
hsa-mir-34b		0,291543	14,1297	22,0518	-7,92218	7,4038	1,07002	37
hsa-mir-34c		0,0658459	40,1709	102,484	-62,3128	32,8725	1,89559	37
hsa-mir-200a		0,864283	1326,75	1388,6	-61,851	359,354	0,172117	37
hsa-mir-200b		0,602682	895,754	794,176	101,578	193,418	0,525174	36
hsa-mir-200c		0,538162	11985,1	13472,9	-1487,86	2393,83	0,621538	36
hsa-mir-429		0,0562723	353,184	195,655	157,529	79,863	1,97249	36
hsa-mir-21		0,51946	307470	290656	16814,2	25852,6	0,650389	37
hsa-mir-17		0,815847	837,164	882,253	-45,089	192,193	0,234603	36
hsa-mir-18a		0,00951988	12,7872	24,2019	-11,4146	4,16723	2,73914	36
hsa-mir-19a		0,0076794	24,9402	50,9265	-25,9862	9,21607	2,81966	37
hsa-mir-19b-2		0,332557	135,853	160,875	-25,0219	25,4848	0,981837	37
hsa-mir-20a		0,117302	212,595	298,805	-86,2105	53,7244	1,60468	36
hsa-mir-92a-2		0,294047	7299,51	9238,63	-1939,11	1821,07	1,06482	36
hsa-mir-221		0,769015	234,95	223,054	11,8957	40,1947	0,295953	35
hsa-mir-222		0,147941	94,1347	123,918	-29,7829	20,1298	1,47954	35

Supp Table III

	Down-regulated NO INSERTIONS (FDR)	Up-regulated NO INSERTIONS (FDR)
hsa-let-7a-2	0.999670833493728	0.009919197811194 **
hsa-let-7a-1	0.999670833493728	0.011404852379541 *
hsa-let-7f-2	0.999670833493728	0.011404852379541 *
hsa-let-7a-3	0.999670833493728	0.022693746426405 *
hsa-let-7e	0.999670833493728	0.078094490857035
hsa-let-7g	0.9951132961285	0.872206669613474
hsa-mir-34a	0.9951132961285	0.872206669613474
hsa-mir-21	0.998834967282168	0.872206669613474
hsa-let-7b	0.772942472861957	0.95466322269098
hsa-let-7c	0.423601380637498	0.95466322269098
hsa-let-7d	0.423601380637498	0.95466322269098
hsa-let-7i	0.772942472861957	0.95466322269098
hsa-mir-34b	0.429905740326531	0.95466322269098
hsa-mir-34c	0.423601380637498	0.95466322269098
hsa-mir-200a	0.608911585117655	0.95466322269098
hsa-mir-200b	0.665217880956605	0.95466322269098
hsa-mir-200c	0.434085457694848	0.95466322269098
hsa-mir-429	0.639258429507727	0.95466322269098
hsa-mir-17	0.423601380637498	0.95466322269098
hsa-mir-18a	0.484865179133113	0.95466322269098
hsa-mir-19a	0.423601380637498	0.95466322269098
hsa-mir-19b-2	0.428260752603725	0.95466322269098
hsa-mir-20a	0.428260752603725	0.95466322269098
hsa-mir-92a-2	0.639258429507727	0.95466322269098
hsa-mir-221	0.428260752603725	0.95466322269098
hsa-mir-222	0.423601380637498	0.95466322269098

** FDR<0.01; * FDR <0.05

Supp Table IV

	Discovery?	P value	Mean 0 insertions	Mean >1 insertions	Difference	SE of difference	t ratio	df
hsa-let-7a-1		0.811259	0.480698	0.495713	-0.0150151	0.0624291	0.240515	37
hsa-let-7a-2		0.789314	0.469172	0.485841	-0.0166693	0.0619347	0.269144	37
hsa-let-7a-3		0.816877	0.48102	0.495479	-0.0144588	0.0619966	0.233219	37
hsa-let-7b		0.87881	0.584209	0.573199	0.01101	0.0717229	0.153508	38
hsa-let-7c		0.00640231	0.407236	0.213361	0.193875	0.0671812	2.88585	38
hsa-let-7d		0.27441	0.562443	0.484549	0.0778941	0.0701826	1.10988	36
hsa-let-7e		0.335279	0.341726	0.280913	0.060813	0.0622923	0.976252	37
hsa-let-7f-2		0.820621	0.416792	0.40006	0.0167322	0.0732697	0.228364	37
hsa-let-7g		0.45017	0.681717	0.630438	0.0512785	0.0672057	0.763008	38
hsa-let-7i		0.607081	0.563465	0.591099	-0.0276344	0.0532797	0.518667	37
hsa-mir-34a		0.648227	0.330608	0.366547	-0.0359388	0.0781491	0.459875	38
hsa-mir-34b		0.85133	0.244194	0.226401	0.0177933	0.0942753	0.188737	37
hsa-mir-34c		0.291671	0.269225	0.176315	0.09291	0.0868538	1.06973	37
hsa-mir-200a		0.0505448	0.414823	0.258322	0.156501	0.0774322	2.02114	37
hsa-mir-200b		0.0739224	0.43227	0.291067	0.141203	0.0767126	1.84068	36
hsa-mir-200c		0.020224	0.544989	0.367096	0.177893	0.0732138	2.42978	36
hsa-mir-429		0.00789936	0.382171	0.191162	0.191009	0.0677867	2.81779	35
hsa-mir-21		0.296335	0.693197	0.633363	0.0598333	0.0564844	1.05929	37
hsa-mir-17		0.437549	0.24329	0.2907	-0.0474101	0.0603897	0.78507	36
hsa-mir-18a		0.217464	0.371112	0.28663	0.0844819	0.0673011	1.25528	36
hsa-mir-19a		0.940266	0.34279	0.348968	-0.0061786	0.0818928	0.075447	37
hsa-mir-19b-2		0.821321	0.444056	0.461259	-0.0172033	0.0756331	0.227457	37
hsa-mir-20a		0.771917	0.384745	0.408165	-0.0234204	0.0801902	0.29206	36
hsa-mir-92a-2		0.703678	0.348132	0.377677	-0.0295447	0.0770598	0.383399	36
hsa-mir-221		0.0398273	0.463609	0.328545	0.135065	0.0632584	2.13513	35
hsa-mir-222		0.298158	0.402576	0.324532	0.0780442	0.0739587	1.05524	37

Supp Table V

	Discovery?	P value	Mean 0 insertions	Mean >1 insertions	Difference	SE of difference	t ratio	df
hsa-let-7a-1	*	< 0,0001	6389,63	4112,88	2276,76	508,865	4,47419	37
hsa-let-7a-2	*	< 0,0001	12754,3	8199,58	4554,68	1035,16	4,4	37
hsa-let-7a-3	*	< 0,0001	6448,99	4180,23	2268,76	514,737	4,40761	37
hsa-let-7b		0,700569	15437,2	14717,5	719,724	1857,49	0,387472	38
hsa-let-7c		0,337504	1257,44	1654,59	-397,152	408,714	0,971711	37
hsa-let-7d		0,0762467	690,492	897,586	-207,094	113,452	1,8254	36
hsa-let-7e	*	< 0,0001	1202,24	584,67	617,569	133,397	4,62955	37
hsa-let-7f-2	*	< 0,0001	7173,84	3887,56	3286,28	739,198	4,44574	37
hsa-let-7g		0,00515672	711,604	562,65	148,954	50,0942	2,97348	37
hsa-let-7i		0,175339	600,312	531,649	68,6635	49,6936	1,38174	37
hsa-mir-100		0,0308682	6353,47	3619,02	2734,45	1218,3	2,24448	37
hsa-mir-101-1		0,143803	11381,9	9328,17	2053,72	1375,16	1,49344	37
hsa-mir-103-1		0,345597	14274,3	16349,1	-2074,85	2171,02	0,955702	36
hsa-mir-106b		0,0270753	586,445	851,139	-264,695	114,994	2,30181	37
hsa-mir-10a		0,0143427	28188,1	20543,5	7644,66	2974,95	2,56967	37
hsa-mir-10b		0,821102	11268,3	12411,2	-1142,88	5017,36	0,227785	36
hsa-mir-1247		0,796181	18,0916	16,502	1,58957	6,1087	0,260214	36
hsa-mir-125a		0,0389207	829,867	596,371	233,496	109,055	2,14109	37
hsa-mir-125b-1		0,35568	530,09	443,163	86,9269	92,9365	0,935336	37
hsa-mir-126		0,698115	6045,68	5611,75	433,924	1110,07	0,390897	37
hsa-mir-128-1		0,0021347	64,9002	126,617	-61,7164	18,6523	3,30879	36
hsa-mir-128-2		0,00226531	50,6081	94,2402	-43,6321	13,2741	3,28701	36
hsa-mir-1307		0,111592	1452,57	1988,9	-536,335	329,256	1,62893	38
hsa-mir-140		0,109876	609,357	477,695	131,662	80,4261	1,63705	38
hsa-mir-141		0,0305471	2313,91	3802,62	-1488,71	661,921	2,24908	37
hsa-mir-142		0,00516305	3184,57	6203,09	-3018,52	1016,98	2,96812	38
hsa-mir-143		0,285539	58195,3	46484,4	11710,9	10802,6	1,08407	36
hsa-mir-145		0,299229	1448,57	1234,46	214,114	203,363	1,05287	37
hsa-mir-146b		0,545521	996,727	1113,16	-116,43	190,836	0,610105	37
hsa-mir-148a		0,161206	61209,2	79392,5	-18183,2	12732	1,42816	39
hsa-mir-148b		0,0384939	267,314	365,963	-98,6481	45,9197	2,14828	36
hsa-mir-152		0,012338	282,767	542,564	-259,797	98,4552	2,63873	35
hsa-mir-155		0,0602033	323,66	484,118	-160,458	82,7688	1,93863	37
hsa-mir-16-1	*	0,00096899	830,11	1434,27	-604,164	168,916	3,57672	38
hsa-mir-17		0,815847	837,164	882,253	-45,089	192,193	0,234603	36
hsa-mir-181a-1	*	< 0,0001	2348,29	1221,24	1127,06	240,133	4,69347	37
hsa-mir-181a-2		0,323303	1000,45	1157,06	-156,611	156,387	1,00143	36
hsa-mir-181b-1	*	0,00011176	633,475	276,068	357,406	82,4227	4,33626	36
hsa-mir-182		0,835261	35813,4	34373,7	1439,75	6873,23	0,209472	36
hsa-mir-183		0,0268516	13969	24303,7	-10334,7	4472,2	2,31088	35
hsa-mir-186		0,0517296	348,563	474,994	-126,431	62,8911	2,01032	37
hsa-mir-18a		0,00951988	12,7872	24,2019	-11,4146	4,16723	2,73914	36
hsa-mir-191		0,00430778	807,206	570,062	237,144	77,8223	3,04725	36
hsa-mir-192		0,383406	600,689	1471,79	-871,105	987,182	0,882416	36
hsa-mir-199a-1		0,655153	1438,37	1583,45	-145,085	322,227	0,450257	37
hsa-mir-199a-2		0,444745	2366,91	2801,86	-434,948	563,065	0,772466	37
hsa-mir-199b		0,243662	2893,91	3769,63	-875,726	739,151	1,18477	37
hsa-mir-19a		0,0076794	24,9402	50,9265	-25,9862	9,21607	2,81966	37
hsa-mir-19b-2		0,332557	135,853	160,875	-25,0219	25,4848	0,981837	37
hsa-mir-200a		0,545175	1641,77	1388,6	253,172	414,706	0,610485	38
hsa-mir-200b		0,203681	1111,96	794,176	317,786	245,582	1,29401	37
hsa-mir-200c		0,538162	11985,1	13472,9	-1487,86	2393,83	0,621538	36
hsa-mir-203		0,0158299	11718,8	32772,2	-21053,4	8324,57	2,52907	37
hsa-mir-20a		0,117302	212,595	298,805	-86,2105	53,7244	1,60468	36
hsa-mir-21		0,51946	307470	290656	16814,2	25852,6	0,650389	37
hsa-mir-210		0,0172087	1947,49	3770,29	-1822,8	730,746	2,49443	37
hsa-mir-22		0,895489	75762,6	74727,7	1034,97	7824,83	0,132268	37
hsa-mir-221		0,769015	234,95	223,054	11,8957	40,1947	0,295953	35
hsa-mir-222		0,147941	94,1347	123,918	-29,7829	20,1298	1,47954	35
hsa-mir-23a		0,00199196	3344,79	5408,62	-2063,83	622,717	3,31423	39
hsa-mir-23b		0,00407049	1654,61	2647,63	-993,021	323,595	3,06871	36
hsa-mir-24-2		0,0272395	2236,27	3016,98	-780,713	339,172	2,30182	36
hsa-mir-25		0,00378113	6370,21	12381,6	-6011,35	1941,31	3,09654	36
hsa-mir-26a-2		0,00253955	2939,4	1919,34	1020,06	314,355	3,24494	36
hsa-mir-26b		0,0544779	1172,98	917,991	254,989	128,503	1,98431	38
hsa-mir-27a		0,137614	1486,71	1963,04	-476,33	313,678	1,51853	36
hsa-mir-27b		0,346726	2006,64	2317,66	-311,022	326,327	0,953098	37
hsa-mir-28		0,2459	4913,7	5658,42	-744,722	631,623	1,17906	37
hsa-mir-29a	*	< 0,0001	12806,3	4979,3	7827,01	1627,22	4,81005	38
hsa-mir-29b-1		0,00373056	632,847	353,857	278,991	90,2772	3,09038	38
hsa-mir-29b-2		0,00271523	684,27	379,423	304,847	95,0333	3,20779	38
hsa-mir-29c		0,00780712	2796,16	1535,74	1260,42	448,71	2,80898	38
hsa-mir-30a		0,435523	20155,8	17960,5	2195,26	2785,49	0,788107	38
hsa-mir-30b		0,00191122	711,898	415,752	296,146	88,4233	3,34919	36
hsa-mir-30c-2		0,213525	526,853	623,005	-96,152	75,9661	1,26572	37
hsa-mir-30d		0,00791168	9850,62	6235,47	3615,15	1285,41	2,81245	36
hsa-mir-30e		0,173315	15094,5	17274,6	-2180,11	1569,36	1,38917	36
hsa-mir-34a	*	< 0,0001	258,375	121,504	136,871	30,2276	4,52802	38
hsa-mir-34b		0,291543	14,1297	22,0518	-7,92218	7,4038	1,07002	37
hsa-mir-34c		0,0658459	40,1709	102,484	-62,3128	32,8725	1,89559	37
hsa-mir-361		0,0210504	384,11	283,146	100,964	41,8974	2,4098	37
hsa-mir-374a		0,417523	875,585	954,074	-78,4886	95,7578	0,819657	38
hsa-mir-375	*	0,00092376	21228,1	3623,56	17604,6	4887,95	3,60163	37
hsa-mir-429		0,0562723	353,184	195,655	157,529	79,863	1,97249	36
hsa-mir-532		0,11409	698,079	927,46	-229,381	141,741	1,61831	37
hsa-mir-92a-2		0,294047	7299,51	9238,63	-1939,11	1821,07	1,06482	36
hsa-mir-93		0,0120629	3794,53	6767,78	-2973,25	1124,56	2,64392	36
hsa-mir-99b		0,265917	37791,9	31273,6	6518,25	5770,44	1,12959	37
hsa-mir-99b		0,701176	27724	29320,3	-1596,38	4127,95	0,386724	37
hsa-mir-99b		0,34135	30202,9	26313,6	3889,3	4040,07	0,962683	41

Supp Table VI

	Discovery?	P value	Mean 0 insertions	Mean >1 insertions	Difference	SE of difference	t ratio	df
hsa-let-7a-1	*	0,0010389	6649,04	3861,02	2788,03	789,586	3,531	41
hsa-let-7a-2	*	0,00103993	13280,8	7674,08	5606,71	1588,01	3,53066	41
hsa-let-7a-3	*	0,00119799	6711,58	3936,7	2774,88	797,016	3,48159	41
hsa-let-7b		0,459978	16965,3	15325	1640,27	2198,5	0,746085	40
hsa-let-7c		0,252202	1359,77	1733,47	-373,697	321,769	1,16139	41
hsa-let-7d		0,00400422	628,825	895,523	-266,698	87,5773	3,04528	42
hsa-let-7e	*	0,00080871	1011,89	520,1	491,788	135,959	3,61717	41
hsa-let-7f-2		0,00451808	6594,19	3489,51	3104,68	1030,17	3,01375	39
hsa-let-7g		0,037565	703,605	572,997	130,608	60,817	2,14756	42
hsa-let-7i		0,0897111	618,008	491,085	126,923	72,9825	1,73909	40
hsa-mir-34a	*	0,00013916	247,358	106,943	140,415	33,3277	4,21318	40
hsa-mir-34b		0,0383327	12,3378	28,8924	-16,5546	7,73492	2,14025	41
hsa-mir-34c		0,0360055	46,9052	110,132	-63,227	29,1619	2,16813	41
hsa-mir-200a		0,0779602	1035,25	1478,11	-442,86	244,957	1,80791	41
hsa-mir-200b		0,152301	696,547	900,843	-204,296	139,999	1,45927	40
hsa-mir-200c		0,0046207	9935,68	16573,6	-6637,9	2215,26	2,99644	41
hsa-mir-429		0,43293	178,86	211,756	-32,8956	41,5255	0,792179	40
hsa-mir-21		0,320546	329966	302639	27326,7	27176,8	1,00552	41
hsa-mir-17		0,0236508	633,971	932,884	-298,913	127,18	2,35031	41
hsa-mir-18a		0,0193843	15,7268	26,9796	-11,2528	4,62348	2,43382	41
hsa-mir-19a		0,00416747	23,5588	46,7878	-23,229	7,64293	3,03928	40
hsa-mir-19b-2		0,181101	108,289	135,447	-27,1586	19,9443	1,36172	39
hsa-mir-20a	*	0,00223585	177,453	300,195	-122,742	37,5713	3,2669	40
hsa-mir-92a-2		0,0399419	5822,92	8259,04	-2436,12	1148,13	2,12182	41
hsa-mir-221		0,711901	235,589	255,839	-20,2506	54,4554	0,371875	41
hsa-mir-222		0,677125	106,974	116,9	-9,92579	23,6633	0,419459	40

Supp Table VII

	Discovery?	P value	Mean 0 insertions	Mean >1 insertions	Difference	SE of difference	t ratio	df
hsa-let-7a-1	*	0,0010389	6649,04	3861,02	2788,03	789,586	3,531	41
hsa-let-7a-2	*	0,00103993	13280,8	7674,08	5606,71	1588,01	3,53066	41
hsa-let-7a-3	*	0,00119799	6711,58	3936,7	2774,88	797,016	3,48159	41
hsa-let-7b		0,459978	16965,3	15325	1640,27	2198,5	0,746085	40
hsa-let-7c		0,252202	1359,77	1733,47	-373,697	321,769	1,16139	41
hsa-let-7d		0,00400422	628,825	895,523	-266,698	87,5773	3,04528	42
hsa-let-7e	*	0,00080871	1011,89	520,1	491,788	135,959	3,61717	41
hsa-let-7f-2		0,00451808	6594,19	3489,51	3104,68	1030,17	3,01375	39
hsa-let-7g		0,037565	703,605	572,997	130,608	60,817	2,14756	42
hsa-let-7i		0,0897111	618,008	491,085	126,923	72,9825	1,73909	40
hsa-mir-100		0,119421	6957,06	3528,24	3428,82	2154,76	1,59128	40
hsa-mir-101-1		0,0111306	12970,7	8765,45	4205,22	1579,69	2,66205	40
hsa-mir-103-1		0,0227826	11917,9	16054,6	-4136,74	1748,31	2,36613	41
hsa-mir-106b	*	0,00041307	554,632	935,879	-381,247	99,1694	3,8444	41
hsa-mir-10a		0,555026	22781,2	20612,4	2168,74	3643,37	0,595255	40
hsa-mir-10b		0,249851	8534,46	12088,6	-3554,18	3044,9	1,16726	41
hsa-mir-1247		0,793903	18,0602	16,3879	1,67225	6,35844	0,262997	40
hsa-mir-125a		0,0718557	620,823	489,48	131,343	71,0811	1,84779	41
hsa-mir-125b-1		0,523782	408,383	346,686	61,6967	95,9255	0,643174	40
hsa-mir-126		0,340332	4188,06	5023,98	-835,918	866,216	0,965023	40
hsa-mir-128-1	*	0,00087143	66,7548	129,475	-62,7207	17,4292	3,59859	40
hsa-mir-128-2	*	0,00015327	47,5223	96,1291	-48,6068	11,5932	4,19269	39
hsa-mir-1307		0,00254383	1228,86	2080,89	-852,034	265,008	3,21512	41
hsa-mir-140		0,142376	585,044	465,254	119,791	80,0871	1,49576	41
hsa-mir-141		0,00718043	2461,99	3893,56	-1431,57	505,895	2,82978	41
hsa-mir-142		0,565181	4554,63	5139,64	-585,008	1008,88	0,579862	41
hsa-mir-143		0,023165	69682,2	42261,3	27420,9	11623,5	2,3591	41
hsa-mir-145		0,267181	1237,3	1028,65	208,648	185,42	1,12527	40
hsa-mir-146b		0,0820425	1494,59	1074,8	419,793	235,759	1,78061	43
hsa-mir-148a		0,152025	84403,9	65727,2	18676,6	12807,1	1,4583	43
hsa-mir-148b		0,467341	328,614	369,769	-41,155	56,0836	0,733815	40
hsa-mir-151a		0,459035	2993,58	3238,22	-244,637	327,12	0,747852	39
hsa-mir-152		0,0639914	331,083	458,198	-127,115	66,7281	1,90497	40
hsa-mir-155		0,130193	341,27	449,531	-108,26	70,0283	1,54595	39
hsa-mir-16-1		0,0736175	958,08	1238,03	-279,951	152,479	1,836	41
hsa-mir-17		0,0236508	633,971	932,884	-298,913	127,18	2,35031	41
hsa-mir-18a		0,0193843	15,7268	26,9796	-11,2528	4,62348	2,43382	41
hsa-mir-181a-1		0,00537774	2042,15	1319,19	722,96	245,594	2,94372	40
hsa-mir-181a-2		0,906397	1125,88	1107,3	18,5828	157,04	0,118332	40
hsa-mir-181b-1	*	0,00072474	581,012	321,778	259,234	70,9332	3,65463	41
hsa-mir-182		0,0841456	27737,5	36189,3	-8451,82	4774,75	1,77011	41
hsa-mir-183	*	< 0,0001	12057,1	24025,6	-11968,5	2717,66	4,40398	41
hsa-mir-186		0,239244	382,762	458,933	-76,1709	63,7813	1,19425	41
hsa-mir-19a		0,00416747	23,5588	46,7878	-23,229	7,64293	3,03928	40
hsa-mir-19b-2		0,181101	108,289	135,447	-27,1586	19,9443	1,36172	39
hsa-mir-191		0,191718	724,401	576,374	148,027	111,516	1,32741	41
hsa-mir-192		0,803246	838,475	905,982	-67,5067	269,11	0,250852	39
hsa-mir-199a-1		0,598369	1303,68	1451,13	-147,45	277,695	0,530979	40
hsa-mir-199a-2		0,581062	2277,46	2554,76	-277,308	498,43	0,556363	40
hsa-mir-199b		0,605234	3043,83	3402,08	-358,256	687,621	0,521008	40
hsa-mir-20a	*	0,00223585	177,453	300,195	-122,742	37,5713	3,2669	40
hsa-mir-200a		0,0779602	1035,25	1478,11	-442,86	244,957	1,80791	41
hsa-mir-200b		0,152301	696,547	900,843	-204,296	139,999	1,45927	40
hsa-mir-200c		0,0046207	9935,68	16573,6	-6637,9	2215,26	2,99644	41
hsa-mir-203	*	0,00201005	9763,98	30921,1	-21157,2	6401,4	3,30508	40
hsa-mir-21		0,320546	329966	302639	27326,7	27176,8	1,00552	41
hsa-mir-210	*	0,00074486	1540,32	3259,13	-1718,81	471,515	3,64529	41
hsa-mir-22		0,266497	64120,7	71701,1	-7580,31	6728,99	1,12652	41
hsa-mir-221		0,711901	235,589	255,839	-20,2506	54,4554	0,371875	41
hsa-mir-222		0,677125	106,974	116,9	-9,92579	23,6633	0,419459	40
hsa-mir-23a		0,00290393	3738,53	5197,75	-1459,22	461,418	3,16248	42
hsa-mir-23b		0,0844826	2083,79	2694,8	-611,002	345,567	1,76811	41
hsa-mir-24-2		0,632854	2617,8	2810,27	-192,466	399,879	0,481311	41
hsa-mir-25	*	< 0,0001	6493,3	14823,1	-8329,81	1705,08	4,88529	41
hsa-mir-26a-2	*	0,00042198	3311,49	1779,25	1532,24	399,306	3,83726	41
hsa-mir-26b		0,341009	955,44	828,891	126,549	131,361	0,963372	41
hsa-mir-27a		0,877617	1709,25	1664,89	44,3625	286,248	0,154979	40
hsa-mir-27b		0,348142	2458,61	2079,79	378,828	399,265	0,948815	42
hsa-mir-28		0,0853609	4880,43	6318,45	-1438,02	816,168	1,76192	42
hsa-mir-29a	*	< 0,0001	11772,1	4468,46	7303,65	1575,09	4,63696	41
hsa-mir-29b-1	*	0,00102067	584,082	317,214	266,869	75,448	3,53712	41
hsa-mir-29b-2	*	0,00125344	626,645	360,234	266,411	77,0026	3,45977	42
hsa-mir-29c	*	0,00048511	2710,04	1343,48	1366,56	359,751	3,79863	40
hsa-mir-30a		0,98941	17710,5	17669,1	41,4525	3103,71	0,0133558	40
hsa-mir-30b		0,011888	903,381	409,363	494,018	187,644	2,63274	41
hsa-mir-30c-2		0,589226	540,557	577,704	-37,1467	68,241	0,544345	40
hsa-mir-30d		0,0458337	11205,3	6985,72	4219,61	2045,56	2,06282	39
hsa-mir-30e		0,562658	17124,9	16152	972,928	1666,65	0,583762	40
hsa-mir-34a	*	0,00013916	247,358	106,943	140,415	33,3277	4,21318	40
hsa-mir-34b		0,0383327	12,3378	28,8924	-16,5546	7,73492	2,14025	41
hsa-mir-34c		0,0360055	46,9052	110,132	-63,227	29,1619	2,16813	41
hsa-mir-361		0,269269	318,97	271,97	47	41,9532	1,1203	40
hsa-mir-374a		0,804948	973,355	948,577	24,7778	99,672	0,248593	40
hsa-mir-375	*	< 0,0001	28638,7	1739,27	26899,5	6135,86	4,38397	39
hsa-mir-429		0,43293	178,86	211,756	-32,8956	41,5255	0,792179	40
hsa-mir-532		0,298194	798,969	969,966	-170,998	162,284	1,05369	41
hsa-mir-92a-2		0,0399419	5822,92	8259,04	-2436,12	1148,13	2,12182	41
hsa-mir-93	*	< 0,0001	3183,31	7875,17	-4691,87	863,43	5,43399	41
hsa-mir-99b		0,34135	30202,9	26313,6	3889,3	4040,07	0,962683	41

Supp Table VIII

	Discovery?	P value	Mean 0 insertions	Mean >1 insertions	Difference	SE of difference	t ratio	df
hsa-let-7a-1		0,98099	0,463243	0,46501	-0,0017664	0,0736816	0,0239734	41
hsa-let-7a-2		0,987214	0,462162	0,463356	-0,00119429	0,0740704	0,0161237	41
hsa-let-7a-3		0,958626	0,463335	0,467166	-0,00383084	0,0733935	0,0521959	41
hsa-let-7b		0,876557	0,469361	0,459521	0,00983987	0,0629417	0,156333	40
hsa-let-7c		0,0697986	0,351242	0,232458	0,118784	0,0637979	1,86188	41
hsa-let-7d		0,0759944	0,506158	0,395171	0,110987	0,0610043	1,81933	42
hsa-let-7e		0,7281	0,384512	0,409737	-0,0252252	0,0720638	0,35004	41
hsa-let-7f-2		0,780127	0,330059	0,310725	0,0193339	0,0687819	0,28109	39
hsa-let-7g		0,738525	0,585207	0,565964	0,0192428	0,057266	0,336025	42
hsa-let-7i		0,857094	0,478507	0,489858	-0,0113513	0,0626413	0,181212	41
hsa-mir-34a		0,576568	0,357295	0,399707	-0,0424121	0,0753306	0,563013	40
hsa-mir-34b		0,843279	0,150189	0,1639	-0,0137105	0,0689114	0,198959	41
hsa-mir-34c		0,722209	0,162988	0,190226	-0,0272382	0,0760927	0,357961	41
hsa-mir-200a		0,066558	0,304624	0,198382	0,106242	0,0563675	1,8848	41
hsa-mir-200b		0,406833	0,367692	0,313634	0,0540582	0,0644839	0,83832	40
hsa-mir-200c		0,937121	0,41371	0,40718	0,00652934	0,0822601	0,0793743	41
hsa-mir-429		0,525664	0,314785	0,273339	0,0414462	0,0647349	0,640245	40
hsa-mir-21		0,61918	0,638485	0,665902	-0,027417	0,054745	0,500814	41
hsa-mir-17		0,707371	0,390812	0,363903	0,0269092	0,071186	0,378012	41
hsa-mir-18a		0,211808	0,332274	0,236385	0,0958888	0,0755979	1,26841	41
hsa-mir-19a		0,0964777	0,220507	0,334859	-0,114352	0,0671789	1,70221	40
hsa-mir-19b-2		0,0194079	0,357518	0,522364	-0,164846	0,0676035	2,43842	39
hsa-mir-20a		0,740829	0,337519	0,359804	-0,0222848	0,0669096	0,333059	40
hsa-mir-92a-2		0,629718	0,424938	0,388063	0,036875	0,0759108	0,485767	41
hsa-mir-221		0,610987	0,309385	0,275763	0,0336223	0,0655927	0,512593	41
hsa-mir-222		0,835716	0,321192	0,335416	-0,0142234	0,0681416	0,208733	40

Supp Table IX

	Discovery?	P value	Mean 0 insertions	Mean >1 insertions	Difference	SE of difference	t ratio	df
hsa-let-7a-1		0,639935	0,543149	0,573305	-0,0301565	0,0638536	0,472276	32,0
hsa-let-7a-2		0,641723	0,539759	0,569778	-0,0300192	0,0639054	0,469744	32,0
hsa-let-7a-3		0,610677	0,545227	0,577777	-0,0325507	0,0633092	0,514155	32,0
hsa-let-7b		0,856758	0,473164	0,48882	-0,0156556	0,0860574	0,181921	33,0
hsa-let-7c		0,559296	0,338033	0,286321	0,0517125	0,0876397	0,590057	32,0
hsa-let-7d		0,611078	0,434326	0,393702	0,0406238	0,0791002	0,513574	32,0
hsa-let-7e		0,722865	0,558457	0,586251	-0,0277937	0,0776863	0,357769	32,0
hsa-let-7f-2		0,704482	0,550936	0,522865	0,0280716	0,0733327	0,382798	31,0
hsa-let-7g		0,16717	0,476528	0,35792	0,118608	0,0839122	1,41348	32,0
hsa-let-7i		0,804942	0,481651	0,50096	-0,0193089	0,0775417	0,249013	32,0
hsa-mir-34a		0,311754	0,515715	0,425365	0,0903507	0,0879076	1,02779	32,0
hsa-mir-34b		0,298895	0,223496	0,147123	0,0763732	0,0723623	1,05543	33,0
hsa-mir-34c		0,210877	0,28382	0,176961	0,106859	0,0837481	1,27596	33,0
hsa-mir-200a		0,629332	0,420852	0,46082	-0,0399678	0,0819848	0,487503	31,0
hsa-mir-200b		0,347076	0,354562	0,432957	-0,0783956	0,0821086	0,95478	31,0
hsa-mir-200c		0,216411	0,512961	0,405404	0,107558	0,0852929	1,26104	32,0
hsa-mir-429		0,132548	0,28543	0,433004	-0,147574	0,0955307	1,54478	31,0
hsa-mir-21		0,0212201	0,5502	0,369336	0,180864	0,0746397	2,42317	32,0
hsa-mir-17		0,337206	0,279988	0,187778	0,0922098	0,0946403	0,974319	32,0
hsa-mir-18a		0,770611	0,209792	0,181271	0,0285214	0,0969651	0,294141	31,0
hsa-mir-19a		0,500835	0,258057	0,200665	0,057392	0,084289	0,680896	32,0
hsa-mir-19b-2		0,757504	0,30268	0,271613	0,0310671	0,0997864	0,311336	33,0
hsa-mir-20a		0,589296	0,273441	0,221664	0,0517761	0,0949707	0,54518	33,0
hsa-mir-92a-2		0,15013	0,405995	0,277911	0,128084	0,0868001	1,47563	31,0
hsa-mir-221		0,10285	0,435371	0,29448	0,140891	0,0839777	1,67772	33,0
hsa-mir-222		0,341062	0,300849	0,227858	0,0729905	0,0755236	0,966459	32,0

Supp Table X

Primer name	Sequence (5' to 3')
Let7aAAA	TGAGGTAGTAGGTTGTATAGTTAAA
Let7bAAA	TGAGGTAGTAGGTTGTGTGGTTAAA
miR34aAAA	TGGCAGTGTCTTAGCTGGTTGTAAA
Let7-ORF2PCRafw	CACAAGCATTCTTATACACC
Let7-ORF2PCRbrv	TATGGCTAGCCAGTTTTCCC
Let7-ORF2PCRa_PG2rv	AAAATCCCCCAACGTGATTCCTCCAGCTTTGTTC
Let7-ORF2PCRb_PG2fw	GAGGAATCACGTTGGGGGATTTTAAACTATACTAC
N51 Fw	GAATGATTTTGACGAGCTGAGAGAA
N51 Rv	GTCCTCCCGTAGCT CAGAGTAATT
SV40 Fw	TGGACAAACCACAACTAGAATGC
SV40 Rv	TTGCAGCTTATAATGGTTAC
HMGA2 Fw	TTGCTGCCTTTGGGTCTTCC
HMGA2 Rv	CAGCGCCTCAGAAGAGAGGACG
DICER Fw	AGTGGTAGGCTTTCACACAG
DICER Rv	AGAAAGGACCCATTGGTGAG
GAPDH Fw	TGCACCACCAACTGCTTAGC
GAPDH Rv	GGCATGGACTGTGGTCATGAG
NEOjunct2 Fw	TGCCTCGTCCTGAAGCTC
NEOjunct2 Rv	CAATCGGCTGCTCTGATG
CMV Fw	ACTGCCAAGTAGGAAAGTCCCA
CMV Rv	ATGCCAAGTACGCCCCCTAT
EBNA-1 Fw	CGTCATCTCCGTCATCACC
EBNA-1 Rv	AGATTTGCCTCCCTGGTTTC
genomicGAPDH Fw	CGTTTCCCAAAGTCCTCCTGT
genomicGAPDG Rv	AGGTGATCGGTGCTGGTTC
scrb Fw	TACAGTTGCGTTGTAGAACGATATAGAGGAACTACGCAGTAAGGTA
scrb Rv	TACCTTACTGCGTAGTTCCTCTATATCGTTCTACAACGCAACTGTA
bs2 Fw	TACGAACAAAGCTGGAGGCATCACACTACCTGACTTCAAACTAGTA
bs2 Rv	TACTAGTTTGAAGTCAGGTAGTGTGATGCCTCCAGCTTTGTTCGTA
8mer Fw	TACGAACAAAGCTGGAGGCATCACTCTACCTCACTTCAAACTAGTA
8mer Rv	TACTAGTTTGAAGTGAGGTAGAGTGATGCCTCCAGCTTTGTTCGTA
Let7-ORF2PCRa_8mer	GAAGTGAGGTAGAGTGATGCCTCCAGCTTTGTTC
Let7-ORF2PCRb_8mer	GAGGCATCACTCTACCTCACTTCAAACTATACTAC
Let7-Bcl1-ORF2bs-PCRaFw	TGGATTCACAGCCGAATTCTACC

Supp Table XI