PIWIL2 downregulation in colon cancer promotes transposon activity and pro-tumorigenic phenotypes

Authors:

Alyssa Risner[1], Joyce Nair-Menon[1], Abhinav Cheedipudi[2], Joe R Delaney[3], Vamsi Gangaraju[3], Antonis Kourtidis[1]

Affiliations

[1]Department of Regenerative Medicine and Cell Biology, Medical University South Carolina, Charleston, SC

[2]University of South Carolina, Columbia, SC.

[3]Department of Biochemistry and Molecular Biology, Medical University South Carolina, Charleston, SC

Corresponding author: Antonis Kourtidis, kourtidi@musc.edu

ABSTRACT

Reactivation of transposable elements (TEs) in somatic tissues, particularly of LINE-1, is associated with disease by causing gene mutations and DNA damage. Previous work has shown that the PIWI pathway is crucial for TE suppression in the germline. However, the status and function of this pathway is not well characterized in differentiated somatic cells and there is lack of consensus on the role of the pathway in somatic tumorigenesis. To shed light on this conundrum, we examined the PIWI pathway in colon cancer through combining bioinformatic analyses and cell-based assays. Shifted Weighted Annotation Network (SWAN) analysis revealed that the pathway experiences significant allelic losses in colon cancer and that PIWIL2, the main catalytic component of the pathway responsible for TE silencing, experiences the highest percent deletions. PIWIL2 is downregulated in colon tumors of advanced stage, nodal metastasis, and in certain subtypes, correlating with poor survival, while it is also downregulated in ulcerative colitis, an inflammatory bowel disease that predisposes to colon cancer. PIWIL2 depletion in colon epithelial Caco2 cells leads to increased anchorage-independent growth, loss of transposon-targeting - non-canonical - piRNAs, increased LINE-1 levels and activity, and in DNA damage, altogether highlighting a tumor-suppressing role of PIWIL2 in the colon.

32 **SUMMARY STATEMENT**

33 This study investigates the PIWI-piRNA pathway in colon cancer using a nuanced bioinformatic and
34 cell-based approach, linking the downregulation of PIWIL2 to disease progression, transposable
35 element activation and DNA damage.

INTRODUCTION

Transposons, or transposable elements (TEs), are mobile DNA elements that make up approximately 45% of the human genome. Although they have been primarily studied in the germline and are overall considered inactive in differentiated somatic epithelial tissues, their reactivation has been shown to lead to a variety of diseases through the introduction of regulatory or protein coding changes when integrated into a new site (Beck et al., 2011, Elbarbary et al., 2016, Hancks and Kazazian, 2016). Additionally, recent studies have demonstrated that genomic instability, oncogene expression, and high mutation rates in approximately 50% of tumors are linked to increased TE activity (Tubio et al., 2014, Jang et al., 2019).

Among the several TE families existing in the human genome, the Long Interspersed Nuclear Element-1 (LINE-1) is the most abundant and the only autonomously mobilizing TE family in humans(Beck et al., 2011). While most copies of the LINE-1 TE are inactive due to truncations and mutations, 100-150 copies are functionally active in disease(Wang et al., 2024, Beck et al., 2010, Tubio et al., 2014, Rodriguez-Martin et al., 2020). The functional LINE-1 copies contain protein-coding open reading frames (ORFs), ORF1 and ORF2, that encode an RNA-binding protein and reverse transcriptase, respectively, to facilitate reverse transcription of the LINE1 RNA transcript(Beck et al., 2011, Elbarbary et al., 2016, Hancks and Kazazian, 2016). Reactivation of LINE-1 has been frequently observed in somatic tissues and particularly in the gastrointestinal tract, where it has been associated with colon cancer, the second most frequent and third deadliest cancer type in both sexes(Pitkanen et al., 2014, Solyom et al., 2012, Ewing et al., 2015, Tubio et al., 2014, Zhuo et al., 2015, Nam et al., 2023, Siegel et al., 2024). Indeed, recurrent LINE-1 insertions have been identified to cause mutations in known oncogenes or tumor suppressors such as APC, a key initiating event in colorectal tumorigenesis, indicating that LINE-1 retro-transposition events have the ability to serve as tumor-initiating events in colon cancer(Cajuso et al., 2019, Scott et al., 2016, Miki et al., 1992). In fact, colon tumors are significantly enriched with somatic retro-transpositions(Rodriguez-Martin et al., 2020). However, the reasons for increased TE activity in these tumors are still poorly understood.

Mechanisms of TE regulation previously described in the germline include DNA methylation and histone modifications of LINE-1 loci, as well as RNA interference by small RNAs through the PIWI pathway(Kuramochi-Miyagawa et al., 2008, Sigurdsson et al., 2012, Ariumi, 2016, Berrens et al., 2017, Walter et al., 2016). The PIWI pathway was first described in *Drosophila* to suppress TEs in the germline and to be required for spermatogenesis(Ku and Lin, 2014). PIWI proteins are members of the Argonaute family of endoribonucleases that bind PIWI interacting RNAs (piRNAs), a distinct class of small non-coding RNAs that are 24-32 nucleotides in length, altogether forming the piRNA-induced silencing complex (piRISC)(Siomi et al., 2011). Most piRNAs are derived from piRNA clusters within the genome,

which are also locations of a large number of truncated TEs (Siomi et al., 2011). piRISC recognizes piRNA - complementary RNA transcripts of TEs in their cytosolic phase and cleaves them, subsequently preventing retro-transposition and DNA mutagenesis(Ramat and Simonelig, 2021, Watanabe and Lin, 2014, Weick and Miska, 2014).

The PIWI-piRNA mechanism is conserved throughout evolution, including in mammals, where it has been found to be required for germline maintenance in mice(Unhavaithaya et al., 2009, Sun et al., 2022, Kuramochi-Miyagawa et al., 2008, Li et al., 2021). In humans, there are four PIWI protein members, namely PIWIL1, PIWIL2, PIWIL3 and PIWIL4(Zeng et al., 2020). In addition to PIWI proteins, the TE targeting function of piRISC is enabled by a number of additional protein components, including the Tudor family of proteins that act as protein scaffolds, as well as by RNA helicases, such as MOV10L1 and DDX4(Siomi et al., 2010, Mathioudakis et al., 2012, Reuter et al., 2009, Gao et al., 2024, Frost et al., 2010, Vourekas et al., 2015, Fu et al., 2019, Li et al., 2021, Wenda et al., 2017). Although it was previously thought that PIWI proteins were germline - specific in humans, a recent study that sought to characterize their expression patterns across multiple tissue types showed that PIWI proteins, namely PIWIL2 and PIWIL4, are also expressed substantially in somatic tissues(Meseure et al., 2020).

In addition to their expression in normal somatic tissues, a number of studies have suggested that PIWI proteins are overexpressed in somatic tumors and act as oncogenes, in a similar fashion to the overall scheme of re-expression of stem cell and germline markers in tumors; however, other studies have shown that PIWI proteins in fact act as tumor suppressors, altogether exposing a lack of consensus on the function of PIWI proteins in somatic tissues and tumors (Lee et al., 2006, Lu et al., 2012, Wang et al., 2015a, Zeng et al., 2017, Eckstein et al., 2018, Meseure et al., 2020, Kishani Farahani et al., 2023, Shi et al., 2020). This is particularly the case for PIWIL2, which is the key endoribonuclease that binds primary piRNAs in piRISC to target TE RNA transcripts for degradation, including LINE-1(Siomi and Siomi, 2015, De Fazio et al., 2011). Still, the function of PIWIL2 in differentiated epithelial tissues and its role in tumorigenesis has not been well characterized(Yan et al., 2011, Peng and Lin, 2013, Martinez et al., 2015, Jacobs et al., 2016, Gao, 2008, He et al., 2010, Liu et al., 2010, Ross et al., 2014). Previous *in vitro* work studying PIWIL2 has been primarily conducted in transformed cancer cell lines, or in commonly used cell lines such as HEK-293 cells, none of which are representative of the normal differentiated epithelium. Here, we aimed to elucidate the expression and function of the PIWI complex in colon cancer, through performing bioinformatic analyses and functional assays in colon epithelial cells.

RESULTS

103 **SWAN analysis identifies dysregulation of the PIWI pathway in colon cancer**

104 We first confirmed high expression levels of PIWIL2 in normal tissues, particularly of gastrointestinal
105 origin, including colon and rectal, through querying PIWIL2 expression levels both through NCBI and
106 through The Cancer Genome Atlas (TCGA), using the University of Alabama at Birmingham CANcer
107 data analysis portal (UALCAN) (**Fig. S1**). Then, to interrogate the status and potential role of the PIWI-
108 piRNA pathway in colon cancer, we initially used the Shifted Weighted Annotation Network (SWAN)
109 algorithm, which has the capability to interrogate pathway and copy-number alterations (CNAs) within
110 the TCGA data set(Bowers et al., 2022). Networks in the algorithm are scored based on the number of
111 interactions within the pathway and if haploinsufficiency data are available for the dataset. The Single-
112 pathway SWAN algorithm was used to perform Gene Ontology Biological Process Analysis on the PIWI-
113 piRNA metabolic pathway in the TCGA colon adenocarcinoma (COAD) dataset (**Fig. 1**). The pathway
114 was scored by the SWAN algorithm as overall haploinsufficient **(Fig. 1A**). Allelic losses or gains were
115 then mapped to each respective chromosome by the algorithm (**Fig. 1B**). This analysis showed that
116 there is not a single "hot spot" across the genome where all the observed allelic losses or gains are
117 concentrated, but that these are dispersed throughout the genome. Notably, PIWIL2, which is the main
118 catalytic component of the primary piRNA pathway involved in TE regulation, experienced the highest
119 percent deletion and allelic losses of the pathway queried (**Fig. 1C**). PIWIL1 and PIWIL4 were not
120 significantly enriched in the SWAN algorithm with gains or losses, and PIWIL3 was excluded from the
121 pathway by the algorithm. Proteins required for both primary and secondary piRNA processing
122 experienced allelic losses, including the nuclease EXD1; RNA helicase DDX4 (Vasa in *Drosophila*;
123 MVH in mice); PLD6 (Zucchini in *Drosophila*; mitoPLD in mice) that generates the 5' end of primary
124 piRNAs; RNA helicase MOV10L1 that is required for stabilizing the primary piRNA transcript; as well as
125 TDRD1 and TDRD9 that are required for stabilization of the complex and piRNA loading(Yang et al.,
126 2016, Wenda et al., 2017, Nishimasu et al., 2012, Han et al., 2015a, Vourekas et al., 2015, Zhang et
127 al., 2019, Mathioudakis et al., 2012). Interestingly, four components of the pathway experienced high
128 percentages of allelic amplifications and allelic gains in the COAD dataset, including TDRD12, TDRKH,
129 ASZ1 and FKBP6. TDRD12 is a binding partner of EXD1 while TDRKH is required for 3'end processing
130 of piRNAs(Pandey et al., 2013, Saxe et al., 2013). ASZ1 is required for phased piRNA biogenesis
131 through the stabilization of PIWIL2 and PIWIL4, while FKBP6 is a co-chaperone to facilitate piRNA
132 loading (Ma et al., 2009, Ikeda et al., 2022, Saxe et al., 2013, Xiol et al., 2012). Furthermore, we queried
133 whether PIWIL2 allelic loss correlates with patient survival using the cBioportal copy number alteration
134 (CNA) tool; indeed, allelic loss of PIWIL2 correlates with poorer patient survival compared to patients
135 without an allelic loss (**Fig. 1D**). The results from the SWAN and cBioportal analyses highlight the
136 overall loss of regulation of the entire piRNA biogenesis pathway in colon cancer and identify PIWIL2
137 as a top candidate for the loss of the pathway's regulation.

138

139 **PIWIL2 is downregulated in colon cancer in correlation with disease progression**

140 Then, to examine whether PIWIL2's allelic losses are also accompanied with downregulation of PIWIL2
141 in colon cancer, we interrogated the COAD (colon adenocarcinoma) dataset of TCGA using UALCAN,
142 to examine PIWIL2 mRNA expression stratified by normal vs: primary tumor; tumor stages; nodal
143 metastasis; p53 status; histological subtype (adenocarcinoma, mucinous adenocarcinoma);
144 microsatellite stability status; patient age; weight; race; and biological sex (**Fig. 2A-J**). Comparisons
145 between normal and primary tumors showed decreased PIWIL2 levels in tumors, as well as when
146 tumors were stratified by stage I-IV (**Fig. 2A,B**). Due to high variance in expression, this downregulation
147 of PIWIL2 only reached statistical significance in the transition to stage IV (**Fig. 2B**). We also did not
148 find any significant downregulation of PIWIL2 in adenomas through interrogating Gene Expression
149 Omnibus (GEO) repository dataset profiles (**Fig. S2**). However, downregulation of PIWIL2 in tumors
150 was significant when the tumors were stratified by lymph node metastasis, particularly in N0 tumors
151 compared to normal and between N0 and N2 tumors, denoting correlation of PIWIL2 both with early
152 transformation at N0 and with tumor aggressiveness at N2 (**Fig. 2C**). Interestingly, PIWIL2
153 downregulation does not correlate with p53 status (**Fig. 2D**); however, PIWIL2 was significantly
154 downregulated in adenocarcinomas compared to normal tissues, and even more downregulated in
155 mucinous adenocarcinomas (**Fig. 2E**). Mucinous adenocarcinomas possess a high degree of
156 microsatellite instability, which is a feature that is used to categorize colon tumors in MSI-H
157 (microsatellite instability - high), MSI-L (microsatellite instability - low), and MSS (microsatellite stable).
158 Since the UALCAN database does not include the option of stratifying tumors based on microsatellite
159 instability, we used GEPIA2(Tang et al., 2019) to examine expression of PIWIL2 in these sub-categories
160 of tumors. Interestingly, PIWIL2 is downregulated in MSI-L tumors and even more significantly in MSI-
161 H tumors, compared to MSS tumors (**Fig. 2F**). This result is in agreement with the downregulation of
162 PIWIL2 in the mucinous adenocarcinomas, which are microsatellite unstable. PIWIL2 also seems to be
163 progressively downregulated by age, where MSI-H tumors are also more prevalent (Guidoboni et al.,
164 2001) (**Fig. 2G**). Finally, there were no statistically significant differences in PIWIL2 expression by
165 weight, race, or biological sex (**Fig. 2H,I,J**). Low PIWIL2 expression in colon tumors was also correlated
166 with poor patient survival among all tumor stages, when stratified by stages 1-3 combined, and even
167 more pronounced in stage 4, using the Kaplan-Meier Plotter web-based tool (**Fig. 2K-M**), which is in
168 agreement with the CNA data (**Fig. 1D**).

169

170 Furthermore, examination of the methylation levels of the PIWIL2 promoter using UALCAN showed
171 that this was significantly elevated in colon tumors overall, indicated by a higher methylation beta-value,
172 as well as when these tumors were stratified by stage, nodal metastasis; p53 status; histological

173 subtype; age: biological sex; weight; and race (**Fig. 3**). These findings indicate that PIWIL2 expression
174 is suppressed across colon cancer possibly due to promoter methylation and associates with
175 microsatellite instability.
176
177 **PIWIL2 is also downregulated in ulcerative colitis**
178 Since we identified downregulation of PIWIL2 in colon tumors beginning from early forms of the disease,
179 such as in N0 tumors, we then sought to examine the status of PIWIL2 in precancerous conditions in
180 the colon, such as in inflammatory bowel disease that significantly increases the risk for colon
181 cancer(Kim and Chang, 2014, Gordon et al., 2018). To do this, we examined the expression levels of
182 members of the PIWI pathway that we identified using the SWAN analysis, using expression data from
183 patients with ulcerative colitis, one of the main manifestations of inflammatory bowel disease, available
184 through the PreMedIBD database(Linggi et al., 2021). This analysis showed that out of all the members
185 of the PIWI-piRNA pathway that were queried, PIWIL2 was the gene that was more predominantly
186 downregulated in ulcerative colitis (**Fig. 4**). This further denotes PIWIL2 as the member of the pathway
187 that is potentially most critical for tumor initiation. Altogether, the above analyses demonstrate that the
188 PIWI pathway is dysregulated in colon cancer, particularly its main component PIWIL2, and its
189 downregulation correlates with disease initiation and progression.
190
191 **PIWIL2 depletion in colon epithelial cells promotes pro-tumorigenic behavior**
192 Given that PIWIL2 exhibits the most allelic losses in the entire PIWI pathway and is also downregulated
193 in ulcerative colitis and colon cancer correlating with disease progression (**Fig. 1C, Fig. 2, Fig. 3, Fig.**
194 **4**), we sought to investigate whether loss of PIWIL2 may indeed promote pro-tumorigenic phenotypes.
195 To do this, we used Caco2 colon epithelial cells, a broadly used in vitro model of the well-differentiated
196 colonic epithelium(Grasset et al., 1984, Hidalgo et al., 1989, Sambuy et al., 2005). We first confirmed
197 that PIWI proteins were expressed in these cells, by examining our publicly available RNA-sequencing
198 data(Daulagala et al., 2024). The data confirmed expression of PIWIL2 and PIWIL4 in Caco2 cells, but
199 not of PIWIL1 and PIWIL3 (**Table S1**). We then generated a PIWIL2 CRISPR/Cas9 mediated knockout
200 (KO) in Caco2 cells (**Fig. S3**). The PIWIL2 KO was overall not well tolerated by these cells, since most
201 PIWIL2 CRISPR/Cas9 – transfected cells did not survive selection; however, we isolated a clone (**Fig.**
202 **S3**) where we confirmed 50% downregulation of PIWIL2 by qRT-PCR (**Fig. 5A**) and by western blot,
203 when assessing the full length PIWIL2 at the predicted molecular weight of 109 kDa and when
204 quantified over 3 biological replicates (**Fig. 5B, C**). We confirmed the specificity of the loss of the full
205 length PIWIL2 through introducing an LZRS-based retroviral expression vector with full length PIWIL2
206 FLAG-tagged. (**Fig. 5D**).
207

208 We then sought to assess the potential phenotypic effects of PIWIL2 depletion in these cells on cell
209 proliferation, since increased proliferation is an indication of pro-tumorigenic behavior. To do this, we
210 used the xCELLigence Real-Time Cell Analysis (RTCA) system, to monitor cell proliferation rates in
211 standard 2-dimensional conditions and in real time (**Fig. 5E**). Interestingly, the PIWIL2 KO cells had a
212 significantly lower proliferation rate indicated by a slower growth curve, than the Caco2 WT control
213 cells. Although we observed a similarly slower proliferation rate by partially depleting PIWIL2 using a
214 shRNA construct (**Fig. S4**), introduction of the PIWIL2-FLAG construct in the PIWIL2 KO cells not only
215 did not rescue the proliferation decrease, but it actually exacerbated it (**Fig. 5E**). This suggests that
216 although the full length PIWIL2 is responsible for proliferation suppression, there may be other isoforms
217 of PIWIL2 that act opposingly in promoting proliferation.
218
219 We then examined the ability of the PIWIL2 KO cells to grow in anchorage-independent conditions,
220 which is a hallmark of pro-tumorigenic transformation. To do this, we cultured cells using a low
221 attachment plate assay to determine if the PIWIL2 KO cells could form spheroids in these conditions
222 (**Fig. 5F-J**). Caco2 WT cells were able to form only small and very few spheroids after 7 days, which is
223 expected since these cells are well-differentiated epithelial cells and not expected to grow in anchorage-
224 independent conditions(Kourtidis et al., 2015). However, the PIWIL2 KO cells grew into much larger
225 spheroids that exhibited a significantly larger area, perimeter and diameter, and were relatively
226 symmetrical, when considering there were no significant changes in circularity. Introduction of the
227 PIWIL2-FLAG construct in the PIWIL2 KO cells reversed the increased anchorage-independent growth
228 and all the observed spheroid phenotypes, demonstrating the specificity of the results (**Fig 5. F-J**).
229 These data demonstrate that the full-length PIWIL2 suppresses proliferation and cell growth in an
230 anchorage independent manner, indicating a tumor-suppressing role of PIWIL2.
231
232 **PIWIL2 depletion results in loss of transposon-targeting piRNAs and in upregulation of the**
233 **LINE-1 transposable element**
234 Since PIWIL2 is the main enzymatic component of the PIWI pathway responsible for TE targeting
235 through piRNAs, we then investigated whether PIWIL2 depletion would indeed affect TE targeting. To
236 do this, we first performed RNA sequencing of Caco2 WT and PIWIL2 KO cells for short RNAs to
237 identify any effects on piRNAs upon PIWIL2 depletion. Intriguingly, our sequencing and subsequent
238 database search for canonical piRNAs using both the piRNA and the repBase databases did not yield
239 any hits, whereas the peak denoting the piRNA size at 25-31 nt was absent (**Fig. 6A**), both indicating
240 that canonical piRNAs are not expressed in Caco2 cells. However, we did detect a peak at the ~16 nt
241 size mark, which has been reported to be a product of piRNA processing and transposon targeting (**Fig.**
242 **6A**)(Wenda et al., 2017, Wang et al., 2015b, Xiol et al., 2012). These short RNAs are also enriched for

243 T/U at their 5' end (**Fig. 6B**), which further corroborates that this peak represents processed piRNAs
244 and it is not a product of random degradation. Indeed, when we aligned those short RNA sequences to
245 TE sequences, we identified a significant population of those to align with TEs (**Table S2**). More
246 specifically, when we examined the "antisense" piRNA-like sequences, meaning the ones
247 complementary to TEs that would indicate the piRNAs targeting those TEs, we found that these are
248 almost entirely absent in the PIWIL2 KO cells, compared to the WT control cells, including many piRNAs
249 that target LINE-1 elements (**Fig. 6C**). Additionally, many of the "sense" short RNA fragments that would
250 represent the cleavage products of the same LINE-1 elements are also absent in the PIWIL2 KO cells
251 (**Fig. 6D**). This indicates that these LINE-1 elements are not targeted for degradation by piRNAs in the
252 PIWIL2 KO cells, which would result in increased overall levels of LINE-1. Indeed, qRT-PCR of the
253 L1RE1 RNA transcript, which includes the first ORF of the LINE-1 mRNA that encodes the ORF1p RNA
254 binding protein required for retro-transposition(Beck et al., 2011, Elbarbary et al., 2016), confirmed that
255 LINE-1 levels were increased in the PIWIL2 KO cells (**Fig. 6E)**. Altogether, these results demonstrate
256 that PIWIL2 depletion impairs piRNA-mediated targeting of TEs, especially of LINE-1 elements,
257 resulting in increased LINE-1 levels in these cells.
258
259 **PIWIL2 depletion increases activity of the LINE-1 transposable element**
260 Then, to examine whether the increased LINE-1 levels upon PIWIL2 depletion were also accompanied
261 by increased retro-transposition activity, we utilized a retro-transposition LINE1-GFP reporter
262 assay(Moran et al., 1996, Ostertag et al., 2000, Garcia-Perez et al., 2010, Kopera et al., 2016). In this
263 assay, a plasmid with a full-length retro-transposition competent LINE-1 element produces GFP only
264 when retro-transposition occurs and is an indicator of TE activity within the cell (denoted LINE1-GFP).
265 Additionally, the assay uses a negative control plasmid with an inactive LINE-1 element that contains a
266 defective ORF1p protein (LINE1 inactive (-)) that is unable to retro-transpose and produce GFP. Caco2
267 WT or PIWIL2 KO cells were transfected with either a constitutively active GFP reporter plasmid
268 (pCEP4+) to determine transfection efficiency, the LINE1-GFP retro-transposition reporter, or the LINE1
269 inactive (-) plasmid as a negative control. After 48 hours, transfected Caco2 WT and PIWIL2 KO cells
270 were placed in puromycin selection for 4 days and then imaged by fluorescence microscopy to identify
271 GFP positive cells (**Fig. 7A**). The average percentage of positive GFP expressing cells was normalized
272 to the transfection efficiency of each cell line for comparison (**Fig. 7B**). Caco2 WT + LINE1-GFP
273 fluorescence was minimal and not significantly different compared to Caco2 WT LINE-1 inactive (-)
274 cells, indicating there is minimal LINE-1 activity in the Caco2 WT cells. However, PIWIL2 KO + LINE1-
275 GFP exhibited significantly higher LINE-1 activity compared to the PIWIL2 KO + LINE1 inactive (-) cells,
276 indicating specific LINE-1 activation in the PIWIL2 KO cells; this increased activity was also significantly

277 higher compared to Caco2 WT + LINE1-GFP cells, demonstrating that PIWL2 depletion not only
278 increases LINE-1 mRNA transcript levels, but also LINE1 activity in colon epithelial Caco2 cells.
279
280 **PIWIL2 depletion promotes DNA damage**
281 The DNA insertional activity of TEs and especially of LINE-1 elements can cause DNA damage, which
282 is a hallmark of cancer (Gasior et al., 2006, Hanahan and Weinberg, 2011). Therefore, we examined
283 whether the increased LINE-1 activity upon PIWIL2 depletion in Caco2 cells (**Fig. 7A,B**) also results in
284 DNA damage in these cells (**Fig. 8**). Immunofluorescent staining followed by confocal microscopy of
285 Caco2 WT and PIWIL2 KO cells for γH2AX, a phosphorylated histone marker of DNA double stranded
286 breaks(Kuo and Yang, 2008), revealed that there was a significant increase in the number of γH2AX
287 foci per cell (**Fig. 8A,B**). We also confirmed increased levels of γH2AX compared to total H2AX by
288 immunoblotting (**Fig. 8C**). Furthermore, examination of γH2AX foci together with the LINE1-GFP retro-
289 transposition assay showed substantial overlap of γH2AX foci with transposon activity (**Fig. 8D,E**).
290 These results demonstrate that PIWIL2 depletion in somatic, well-differentiated colon epithelial Caco2
291 cells results in up regulation of transposon levels and activity, as well as in subsequent increased DNA
292 damage, which are all promoters of pro-tumorigenic transformation. These findings, taken together with
293 the dysregulation of the PIWI pathway and downregulation of PIWIL2 in colon tumors, as well as with
294 the significantly increased ability of the PIWIL2 KO cells to grow in anchorage-independent conditions,
295 underscore a tumor-suppressing role of PIWIL2 in the colon.
296
297 **DISCUSSION**
298 Prior studies have shown that the activation of TEs in tumors, including colon cancer, was frequent,
299 although little is known regarding the potential mechanisms of TE activation (Pitkanen et al., 2014,
300 Ewing et al., 2015, Solyom et al., 2012). Recently, the PIWI pathway, which is well-established to
301 suppress transposons in the germline, has also been shown to be expressed in the soma and be
302 implicated in cancer(Ross et al., 2014, Siomi et al., 2011). However, reports regarding the presence of
303 PIWI proteins in somatic tissues and the role of this pathway in tumorigenesis have been conflicting,
304 with some supporting a tumor-suppressing and others a tumor-promoting role of the pathway. Our
305 bioinformatic analysis, as well as our RNAseq, qRT-PCR, and protein expression data, confirm
306 expression of PIWIL2, the key regulator of the pathway, in the colon and in the well-differentiated colon
307 epithelial Caco2 cells (**Fig. 5, Fig. S1, Table S1**). Our results also show that the pathway experiences
308 significant allelic losses and its expression is downregulated in colon adenocarcinoma, with PIWIL2
309 having the highest percentage of deletions and allelic losses within the pathway (**Fig. 1**). Both PIWIL2
310 allelic loses and downregulation also correlate with poor patient survival (**Fig. 1, 2)**. Furthermore,
311 knockout of PIWIL2 in Caco2 cells resulted in increased expression and activity of the LINE-1 TE, the

most abundant and only autonomous TE in humans, as well as in DNA damage and in anchorage-independent growth, all hallmarks of pro-tumorigenic transformation (**Fig. 5,6,7,8**). Altogether, the data support the idea that PIWIL2 acts as a tumor suppressor in the colon.

Furthermore, when we interrogated PIWIL2 expression levels in the TCGA dataset by UALCAN analysis, PIWIL2 loss was more prevalent in the mucinous adenocarcinoma subtype of colon cancer (**Fig. 2E**), which is found in 10%-20% of colon cancer patients and is characterized by ≥50% of the tumor volume consisting of extracellular mucins(Fleming et al., 2012, Luo et al., 2019, Huang et al., 2021). Studies have also shown that mucinous adenocarcinoma tumors exhibit a higher frequency of microsatellite instability, CpG island methylation, and lower frequency of p53 mutations(Luo et al., 2019, Huang et al., 2021). Indeed, PIWIL2 downregulation in colon tumors seems to correlate with microsatellite instability, independently of p53 status (**Fig. 2D,F**). These findings further suggest the involvement of TE mobility in tumorigenesis, since retrotransposons like LINE-1 are known to give rise to de novo microsatellites during retro-transposition events due to internal proto-microsatellite sequences and poly-A mononucleotide repeats (Grandi and An, 2013). Since PIWIL2 downregulation seems to be independent of p53 status, which is key for maintaining genome integrity, it is likely that PIWIL2 dysregulation in microsatellite tumors is a potential mechanism of LINE-1 activation which would be interesting to be further interrogated.

Our database examination also showed that PIWL2 is downregulated in samples from patients with ulcerative colitis, which highlights a gap in current research, since the PIWI pathway has not been previously studied in colitis and in inflammatory bowel disease (**Fig. 4**). A few studies have indicated that there is upregulation of LINE-1 elements in inflammatory bowel disease, particularly in Crohn's Disease; however, the information on transposon expression and their regulation in these diseases is still very limited(Kanke et al., 2022, Wang et al., 2017). Given the above and that colitis and inflammatory bowel disease overall strongly predispose for colon cancer, it would be interesting to examine whether this PIWIL2 downregulation is a contributing factor to this predisposition, through upregulation of LINE elements in these diseases.

Interestingly, our *in vitro* studies showed that only a 50% knockout of PIWIL2 is tolerated by Caco2 cells, which correlates with the haploinsufficiency identified by the SWAN algorithm (**Fig. 1, 5A-C, Fig. S3**). In fact, this low tolerance for PIWIL2 loss may be due to the ensuing DNA damage caused by PIWIL2 downregulation (**Fig. 8**). However, we found that even a 50% depletion of PIWIL2 was able to significantly promote anchorage independent growth of Caco2 cells, which are otherwise well-differentiated and don't grow efficiently under these conditions (**Fig. 5F-J**). Acquisition of anchorage independent growth capability is a key event in epithelial cell pro-tumorigenic transformation. Furthermore, the increased DNA damage upon depletion of PIWIL2 was accompanied by increased

346 levels and activity of the LINE-1 TE, indicating that TE regulation and silencing is disrupted upon PIWIL2
347 depletion (**Fig. 7**). Both TE activation and DNA damage are driver events in tumorigenesis (Cajuso et
348 al., 2019, Scott et al., 2016, Miki et al., 1992, Gasior et al., 2006). Therefore, these findings, taken
349 together with our current results showing early downregulation of PIWIL2 in colon tumors or even in
350 pre-cancerous IBD, underscore a potential role of PIWIL2 in early colon tumorigenesis, which deems
351 further investigation.

352 Although our results in this study overall support a tumor suppressing role for PIWIL2 in colon
353 tumorigenesis, there have been conflicting reports regarding the role of PIWIL2 in cancer (Lee et al.,
354 2006, Lu et al., 2012, Wang et al., 2015a, Zeng et al., 2017, Eckstein et al., 2018, Meseure et al., 2020,
355 Kishani Farahani et al., 2023, Shi et al., 2020). Along these lines, both our CRISPR/Cas9-generated
356 PIWIL2 KO cells, as well as our PIWIL2 knockdown cells using a shRNA, exhibit slower proliferation
357 rate, compared to the control cells, which would imply a tumor-promoting role for PIWIL2 (**Fig. 5E, Fig.
358 S4**). However, unlike the anchorage-independent growth phenotype, we were not able to rescue the
359 effect of PIWIL2 depletion on proliferation, using our full length PIWIL2-FLAG construct, which actually
360 suppressed proliferation even more, further supporting the tumor suppressing role of PIWIL2 (**Fig. 5E)**.
361 It is likely that this discrepancy is due to the existence of additional PIWIL2 isoforms, which may have
362 opposing roles to the full length PIWIL2 protein. Such isoforms have been reported in the literature and
363 have been involved in tumorigenesis, or have been associated with certain cancer cells and tumor
364 types (Gainetdinov et al., 2014, Ye et al., 2010). These isoforms apparently lack parts of the PAZ or
365 PIWI domains, which are essential for piRNA processing and TE targeting, and seem to act through
366 downstream pathways other than the canonical piRNA-TE targeting pathway (Gainetdinov et al., 2014,
367 Ye et al., 2010). Although our PIWIL2 examination in this study shows an overall downregulation in
368 colon cancer and supports a tumor-suppressing role of the protein through piRNA-mediated TE
369 targeting, considering these additional isoforms in different cancer types could resolve any potentially
370 conflicting findings regarding the role of PIWIL2 in cancer.

371 PIWIL2 is part of a larger protein complex (piRISC) and the SWAN analysis identified additional
372 members of the PIWI-piRNA pathway and of piRISC as also significantly downregulated. For example,
373 other top targets from the SWAN analysis include the exonuclease EXD1 found in pi-bodies, the RNA
374 helicases DDX4 and MOV10L1, PLD6, and several members of the Tudor family of proteins that are
375 key for the functionality of piRISC. These proteins have also been identified to be dysregulated in other
376 cancer types, mainly breast and ovarian cancers, but they have not been studied in the colonic
377 epithelium, since this is still a novel and ongoing field of investigation(Olotu et al., 2024, Kim et al.,
378 2014, Lee et al., 2020). Therefore, it would be of particular interest to further investigate these additional
379 PIWI pathway components and their roles in suppressing TEs in colonic diseases. In addition, our

finding of the expression of non-canonical piRNAs in Caco2 cells is intriguing and warrants further investigation: it is likely that the PIWI pathway in somatic tissues acts through and processes different sets of short RNAs to target transposons, as opposed to the well-characterized piRNA substrates of the pathway in the germline. In summary, our work, through using a nuanced bioinformatic approach and by performing cell - based assays, sheds light into the currently conflicting views about the role of the PIWI pathway in somatic tissues and in cancer, opening future directions of investigation that will fully assess the tumor suppressing potential of this pathway.

METHODS

Bioinformatic Analysis

Multiple bioinformatic tools were used throughout this study to investigate changes in the PIWI protein pathway using data from The Cancer Genome Atlas (TCGA). The analytical tools are detailed below.

Shifted Weighted Annotation Network (SWAN): SWAN is an analytical tool used to interrogate pathway and copy-number alterations (CNAs) within the TCGA data set and can be found at the following webpage: https://www.delaneyapps.com/. The Single-pathway SWAN tool was used to perform Gene Ontology Biological Process Analysis on the piRNA metabolic pathway in COAD with a 0.001 significance threshold over 1000 permutations of the algorithm. RNA data was integrated with CNA data for this analysis and a Wilcoxon rank test was calculated as described (Bowers et al., 2022).

cBioportal: cBioportal is another open-source web-based tool to visualize multidimensional cancer genomics data sets, including data from TCGA Pan Cancer Atlas (Cerami et al., 2012, Gao et al., 2013, de Bruijn et al., 2023). This tool can be found at the following webpage: https://www.cbioportal.org. The TCGA COAD data set was selected and then queried for samples with shallow deletion CNAs (loss of one allele; HETLOSS) of PIWIL2 and a plot of Overall Survival was generated using the Comparison/Survival Tool.

UALCAN: UALCAN is a web-based tool from the University of Alabama at Birmingham that allows users to stratify TCGA data for more in-depth analysis of TCGA expression profiles. This tool can be found at the following webpage: https://ualcan.path.uab.edu/. A pan-tissue analysis of PIWIL2 expression was performed across multiple tissuetypes to compare overall expression in normalsamples. PIWIL2 expression and methylation was then queried in the TCGA colon adenocarcinoma (COAD) data set and stratified by normal vs tumor, tumor stage, nodal metastasis, p53 status, histological subtype, biological sex, age, weight, and race. The median,interquartile range, minimum and maximum values

414 are displayed, and statistical significance was calculated by the UALCAN algorithm using a Welch's T-
415 test as previously described (Chandrashekar et al., 2017, Chandrashekar et al., 2022). Results were
416 exported from UALCAN, and graphical displays were made in Prism 10 (Graphpad).
417
418 GEPIA2: GEPIA2 is an interactive web-based tool for analyzing RNA-seq data from the TCGA and
419 GTEx projects that can be found at the following webpage: http://gepia2.cancer-pku.cn/#index. The
420 single gene analysis search for PIWIL2 was used to plot box plots stratified by cancer subtype (MSS,
421 MSI-L, MSI-H) of gene expression. For this study, the median expression of colon adenocarcinoma
422 (COAD) was plotted from TCGA data available with the upper and lower quartiles. The GEPIA2
423 program calculated statistical significance using a one-way ANOVA. A $\log_2$-fold-change of 0.5 and p-
424 value of 0.5 were set as thresholds for the analysis(Tang et al., 2019).
425
426 Kaplan-Meier Plotter: This web-based tool is capable of correlating mRNA expression data and survival
427 data of colon cancer datasets from the Gene Expression Omibus (Gyorffy, 2024). A hazard ratio with
428 95% confidence intervals and logrank P value are calculated by the program. PIWIL2 was interrogated
429 to determine Regression Free Survival and stratified by all tumor stages, stages 1-3 combined and
430 stage 4 alone. This web-based tool can be found at the following webpage https://www.kmplot.com
431 (Gyorffy, 2024).
432
433 PreMedIBD: PreMedIBD is a web-based data mining tool for meta-analysis of ulcerative colitis tissue
434 gene expression. This tool can be found at the following webpage: https://premedibd.com/genes.html.
435 PIWIL2, PIWIL4, DDX4, PLD6, EXD1, TDRD1, TDRD9, TDRD12 and MOV10L1 were queried and the
436 log2-Fold gene expression changes between ulcerative colitis patients and non-IBD controls were
437 determined for each gene. The meta-analysis was calculated as previously described using a random
438 effects model and shows the meta-log2-Fold change and 95% confidence interval(Linggi et al., 2021).
439 A $\log_2$-fold-change of 0.5 was used as a threshold for significance. Datasets used are available in the
440 Gene Expression Omnibus: GSE13367, GSE9452, GSE53306, GSE38713, GSE47908, GSE73661,
441 GSE114527, GSE87466.
442
443 **Cell Culture**
444 Caco2 colon epithelial cells (ATCC, #HTB-37) were cultured in MEM (Corning, #10-010-CV) with 10%
445 fetal bovine serum (FBS) (Gibco,Fetal Bovine Serum Qualified One Shot #A31605-02), 1 mM sodium
446 pyruvate (Gibco, #11360-070), and 1x non-essential amino-acids (NEAA) solution (Gibco, #11140-050)
447 and were maintained at 37°C with 5% CO_2. Caco2 cells were authenticated by the University of Arizona
448 Genetics Core (via Science Exchange) and checked for misidentified, cross contaminated, or

449 genetically drifted cells and tested negative for mycoplasma contamination (LookOut Mycoplasma PCR
450 Detection Kit, Sigma Aldrich). The Caco2 PIWIL2 knockout (KO) cell line was generated using
451 CRISPR/Cas9 technology in collaboration with the Cell Models Core as part of the Medical University
452 of South Carolina COBRE in Digestive and Liver Disease following the protocol by Ran et al.(Ran et
453 al., 2013) The sgRNA sequence was designed to target exon 14 of PIWIL2
454 (sense:CACCGCCAATGAACTGATGCGTTGG/antisense: AAACCCAACGCATCAGTTCATTGGC)
455 using the online Synthego tool (www.synthego.com) and synthesized by Integrated DNA Technologies,
456 IDT. This PIWIL2-targeting sgRNA was cloned into the PX459 pSPCas9(BB)-2A-Puro vector and
457 Caco2 cells were transfected with 30 µg the sgRNA-containing vector using Lipofectamine 3000
458 (Invitrogen) according to the manufacturer's protocol. After 24 hours, cells were selected with 5 µg/ml
459 puromycin for 48 h. Successfully transfected cells were seeded to obtain single colonies, which were
460 picked and expanded for screening by western blot and qPCR. PCR using primers designed to amplify
461 the target region was performed (Forward1: CTCGAACACCTGGGCTCAAGCG and Reverse1:
462 ACCACGCACAGAGGTTCATACC and Forward2: ATCACATTCACCTTCCTAAGAGA and Reverse2:
463 GACCACGCACAGAGGTTCAT) and sequenced by Sanger sequencing (Eurofins Genomics).
464 Alignment of the PCR products was performed to characterize the indel of the PIWIL2 KO.
465

466 **RNA isolation and qPCR**

467 RNA was isolated by adding 1ml of Trizol (Invitrogen # 15596018), incubating the sample at RT for 5
468 minutes then using the PureLink RNA mini kit (Invitrogen #12183018A) for total RNA, using the whole
469 transcriptome protocol of the kit. Final RNA concentrations and purity were determined by using a
470 NanoDrop spectrophotometer. RNA was converted to cDNA using the High Capacity cDNA Reverse
471 Transcriptase Kit (Applied Biosystems #4368814). qPCR reactions were performed using the Taqman
472 FAST Universal PCR master mix (Applied Biosystems #4352042), in a BioRad CFX96 Touch/Connect
473 real time quantitative PCR machine. Data were analyzed and normalized to 18S ribosomal RNA by
474 ΔΔCt calculations to calculate fold change compared to control samples.
475

476 **RNA sequencing**

477 RNA from Caco2 WT and PIWIL2 KO cells was extracted as above and short RNA sequencing was
478 performed using the NovaSeq SE50 strategy. Sequenced reads were first processed using Cutadapt
479 (v1.15) to trim the 3′ adapter sequence (5′-AGATCGGAAGAGCACACGTCT-3′). Adapter-trimmed
480 reads were analyzed using the piPipes small RNA pipeline to quantify transposon-mapped reads and
481 normalize sense and antisense counts as counts per million (CPM) uniquely mapped to the hg19
482 genome(Han et al., 2015b). In parallel, reads were aligned to Repbase and known piRNA clusters using
483 Bowtie2(Langmead and Salzberg, 2012, Langmead et al., 2019) and only successfully aligned reads

were retained. These reads were then assessed for size distribution and 5′ nucleotide signature using the TBr2_basicanalyses.pl script (Rosenkranz et al., 2015). Results were summarized and plotted as shown in **Fig. 6.** The detailed list of hits is shown in **Table S2**.

The Caco2 total mRNA sequencing that was used to extract the PIWI1-4 expression data shown in **Table S1** have been previously published and are publicly available in the GEO public database, under the accession number GSE156860(Daulagala et al., 2024).

Immunoblotting

Whole-cell protein lysates were obtained using RIPA buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 1% NP-40, 0.5% deoxycholic acid, and 0.1% SDS) supplemented with protease inhibitor (RPI Research Products, #P50750-1) and phosphatase inhibitor (Thermo Fisher Scientific, Halt Phosphatase inhibitor cocktail, #1862495) at 1:100 dilution. Cell culture plates were scraped with a cell scraper, lysates were homogenized through a syringe and then cleared by centrifugation at 15,000 rpm for 5 minutes. Protein quantification was performed using a Pierce BCA Protein Assay (Thermo Fisher Scientific, #23227). Protein extracts were mixed with Laemmli sample buffer and separated by SDS–PAGE using 4–20% Tris-Glycine eXtended (TGX) gels (Bio-Rad) and transferred to 0.2 µm nitrocellulose membranes (Bio-Rad) with the Trans-Blot Turbo Transfer System (Bio-Rad). Transfer efficiency was determined using a Ponceau S stain (Thermo Fisher Scientific, #A40000279). Membranes were blocked in 3% milk for 45 minutes and incubated overnight with primary antibody rotating at 4°C. The membranes were washed with 1x TBST, incubated with secondary antibody for 45 min at 4°C, followed by room temperature for 1 hour, and then washed with 1x TBST. Signals were detected by luminescence using Pierce ECL (Thermo Fisher Scientific, #32209) and a ChemiDoc Imaging System (Bio-Rad). Blots were quantified using the Analyze gels program in FIJI/ImageJ.

PIWIL2-FLAG Expression Constructs

For PIWIL2 expression and rescue experiments, a C-terminal PIWIL2-3xFLAG tagged construct was designed and synthesized using the LZRS-MS-NEO retroviral backbone with the PIWIL2-3xFLAG (PIWIL2-FLAG) CDNA sequence inserted at the EcoRI site (Genscript). The retrovirus was generated using HEK Pheonix cells according to standard protocols; and Caco2 wild type (WT) and PIWIL2 KO cells were infected with either the empty vector (LZRS) or LZRS-PIWIL2-FLAG. Retroviral infection was followed by antibiotic selection using 250 µg/ml G418 disulfate (Thermo Fisher Scientific, #J63871.AD) for 12 days, then maintaining the cells in 125 µg/ml G418 disulfate. PIWIL2 and FLAG expression was confirmed through immunoblotting.

519 **shRNA Knockdown**

520 shRNAs were derived from the pLKO.1-based TRC1 (Sigma-Aldrich/RNAi Consortium) shRNA library
521 (pLKO.1-puro Non-Target shRNA Control, SHC016; PIWIL2 #7882 TRCN0000007882). Caco2 cells
522 were transfected with 1.5 µg of the shRNA-containing vector using Lipofectamine 3000 (Invitrogen)
523 according to the manufacturer's protocol and cell culture media was changed 16 hours post-
524 transfection. 48 hours post transfection, protein and RNA was isolated, and cells were used for the
525 xCelligence proliferation assay in Supplemental Figure 4.

526

527 **xCELLigence Proliferation Assay**

528 16 well E-plates (Agilent, # 300 600 890) were used for the experiment on an Agilent xCELLigence
529 Real Time Cell Analysis (RTCA) Dual Purpose (DP) machine and RTCA Basic software (Agilent). Pre-
530 warmed media was added to each well and used to take a background measurement for 1 minute.
531 Cells were grown until 80% confluency, washed with 1xPBS, trypsinized, centrifuged at 1000 rpm for 5
532 min and then resuspended in cell culture media for a manual cell count using trypan blue. 20,000
533 cells/well were then plated for each cell line in a volume of 100 µl/well for a total volume of 150 µl/well.
534 Each plate included blank wells with 150 µl of media for a negative control. Cells were allowed to settle
535 and attach to the plate at room temperature in the cell culture hood for 30 minutes, after which the plate
536 was inserted into the machine. The Xcelligence machine was set to take a measurement every hour
537 for 50 hours. The experiment was completed with n = 3 biological replicates and 4 technical replicates
538 per cell line that were averaged to determine the growth curve. Results were exported from the RTCA
539 software and graphical displays were made in Prism 10 (Graphpad). The cell index from each run was
540 normalized to the final end point of the Caco2 WT cell line to determine fold change and a two-way
541 repeated measures ANOVA with Greenhouse–Geisser correction and Bonferroni correction.

542

543 **Low Attachment (Anchorage-Independent Growth) Assay**

544 Cells were trypisinzed with 0.25% Trypsin-EDTA (Gibco, #25200-056) and counted using a manual cell
545 count and trypan blue (Corning, #25-900-CI). Cells were then suspended in culture media appropriate
546 for the cell line that contained 2.5% basement membrane substrate (Cultrex Basement Membrane
547 Extract, Pathclear, #3432-001-01) and plated on an ultra-low attachment 96-well plate (Corning, #3474)
548 at a seeding density of 1000 cells per well. Plates were cultured for a total of 7 days, adding 50 ul of
549 media to the wells every other day and were fixed using 4% paraformaldehyde (PFA) (Electron
550 Microscopy Sciences, # 157-4) diluted to a final concentration of 1% PFA. The experiment was
551 completed with n = 3 biological replicates and 9 technical replicates per cell line. Following fixation,
552 plates were imaged on a Keyence BZ-X810 microscope at 2x magnification. Images were analyzed in

553 FIJI/ImageJ by thresholding and using the analyze particles function to measure the area, perimeter,
554 Feret's diameter, and circularity of the spheroids.
555
556 **LINE1 Reporter Assay**
557 The LINE1 reporter constructs used in this study were a gift from Eline Luning Prak. The plasmids used
558 in this study contain the indicated modifications to the pCEP4 vector backbone as previously
559 described(Moran et al., 1996, Ostertag et al., 2000, Garcia-Perez et al., 2010, Kopera et al., 2016). In
560 detail:
561 *pCEP4puroeGFP* (pCEP4+) The vector backbone is the pCEP4 plasmid with the hygromycin
562 resistance gene replaced with puromycin for selection. It contains the coding sequence for enhanced
563 GFP (eGFP) and constitutively expresses eGFP with puromycin selection. This is the positive control
564 for the retrotransposition assay and is used to determine the transfection efficiency (Addgene
565 #177922).
566 *pJM111L1eGFP (LINE1 inactive)* The vector backbone is the pCEP4 plasmid with the hygromycin
567 resistance gene replaced with puromycin for selection. This vector contains a retro-transposition
568 defective allele of LINE1 that contains two missense mutations ($RR_{261-262}AA$) in the LINE1 ORF1-
569 encoded protein (ORF1p) and also contains the eGFP indicator cassette. This plasmid cannot express
570 eGFP and is a negative control for the retro-transposition assay (Addgene #42941).
571 *p99LINE1RP-GFP (LINE1-GFP reporter)* The vector backbone is the pCEP4 plasmid with the
572 hygromycin resistance gene replaced with puromycin for selection. This vector lacks the CMV promoter
573 and contains a full-length retro-transposition competent LINE1 element and the eGFP indicator
574 cassette subcloned into the 3'UTR of LINE1 (Addgene #42940).
575 Caco2 wild type or PIWIL2 KO cells were seeded onto 18mm circular glass coverslips in 12 well plates
576 for immunofluorescence staining and allowed to grow until 70% confluency. Cells were transfected
577 using Lipofectamine 3000 (Thermo Fisher #L3000-001) according to manufacturer's instructions with
578 1µg of pCEP4puroeGFP, or 1µg pJM111L1eGFP or 1µg p99LINE1RP-GFP(Day 0). 16 hours post
579 transfection, the media was changed with fresh cell culture media (Day 1). At 48 hours post-transfection,
580 cells were placed in puromycin selection at a concentration of 5 µg/ml (Day 2). 48 hours later, media
581 was changed with fresh, and cells were kept in puromycin selection (Day 4). After 4 days of puromycin
582 selection (Day 6 post-transfection), the cells were then fixed with 4% paraformaldehyde, washed with
583 1xPBS with 10mM glycine followed by permeabilization with 0.02% Triton-X 100 and stained with DAPI.
584 For the co-localization experiments, the coverslips were fixed and stained for EGFP (Abcam
585 ab184601,1:100 dilution) and γH2AX (Cell Signaling 20 E3 (9718T), 1:400 dilution) and DAPI. Three
586 fields per condition were captured and the experiment was completed with n = 4 biological replicates
587 for Figure 7 and n = 3 biological replicates for Figure 8 and imaged using a Keyence BZ-X810

588 microscope. Images were analyzed by FIJI/ImageJ by thresholding and using the analyze particles
589 function. The average number of retro-transposition events was divided by the transfection efficiency
590 of the pCEP4+ positive control plasmid for each cell line.
591
592 **Immunofluorescence**
593 Cells were grown in 12-well plates on 18 mm circular sterile glass coverslips (Fisher Scientific #72222-
594 01) until they reached 80% confluency. Cells were washed once with 1x PBS and fixed with 100%
595 methanol (Thermo Fisher Scientific) at –20°C for 7 min followed by 1x PBS wash. Coverslips were then
596 blocked with serum free Protein Block reagent (Dako, #X0909) at RT for 1hr and stained with primary
597 antibodies diluted in Antibody Diluent (Dako, #S3022) as described in the antibodies table overnight at
598 4°C. Cells were then washed with 1x PBS and stained with the fluorescent-labeled secondary
599 antibodies for 1 hour at room temperature. Coverslips were then washed with 1x PBS, co-stained with
600 DAPI (Sigma-Aldrich, #D8417) and mounted with Aqua-Poly/Mount (Polysciences, #18606-20).
601
602 **Confocal Imaging and Image processing**
603 Confocal images of γH2AX foci were acquired using a Leica SP8 confocal microscope with a 63x Plan-
604 Apochromat 1.4NA DIC oil immersion objective (Leica) and 405 nm and 633 nm lasers. Image
605 acquisition was completed using the Leica Application Suite software at 2048 × 2048 resolution and
606 with 0.3 μm intervals along the z-axis. The same imaging parameters were used across conditions for
607 each set of experiments to allow for comparisons. Z-stacks of all samples were analyzed using the
608 FIJI/ImageJ analysis software and a max projection of each image was used to visualize the nucleus
609 and account for uneven cell thickness over the entire field. Three fields of each sample were captured,
610 and the experiment was performed in n = 3 biological replicates. Total γH2AX foci were quantified by
611 thresholding the images and using the analyze particles function, then normalized to the total number
612 of cells in the field. Nuclei on the borders were excluded from the cell count calculation in FIJI.
613
614 **Analysis and Statistics**
615 For all measurements, sample size and related statistics are indicated in the respected figure legends
616 and were analyzed and graphed using Prism 10 (GraphPad). All experiments were performed in at
617 least three independent experiments, unless otherwise noted, with alpha = 0.05 and representative
618 images are shown. Statistical tests are noted in each figure legend along with p-values. For tests that
619 required multiple comparisons, the multiplicity adjusted p-value is reported.
620
621 **Antibodies**
622 <u>Primary Antibodies</u>

Antibody Name	Company	Catalogue Number	Animal	WB dilution	IF dilution
PIWIL2*	Sigma Aldrich	SAB3500749	Rb	1 to 1000	-
β-actin (Actin)	Cell Signaling	4967L	Rb	1 to 2000	-
Anti-FLAG M2	Sigma Aldrich	F1804	Ms	1 to 2000	-
H2AX total	Santa Cruz	sc517336	Ms	1 to 1000	-
γH2AX	Cell Signaling	20 E3 (9718T)	Rb	1 to 1000	1 to 400 (MeOH and PFA Fixed)
EGFP	Abcam	ab184601	Ms	-	1:100 (PFA Fixed)

* The PIWL2 antibody was validated by using the PIWIL2-FLAG construct in Fig 5D. All other antibodies are standard, commercially available and vendor-validated antibodies.

Secondary Antibodies

Antibody Name	Company	Catalogue Number	Animal	WB dilution	IF dilution
HRP-anti-rabbit	Jackson ImmunoResearch	711-035-152	Rb	1 to 2000	-
HRP-anti-mouse	Jackson ImmunoResearch	715-035-150	Ms	1 to 2000	-
AlexaFluor 647 anti-mouse	Invitrogen	A21236	Ms	-	1 to 500 (MeOH and PFA Fixed)
AlexaFluor 488 anti-rabbit	Invitrogen	A11034	Rb	-	1 to 500 (PFA Fixed)

qPCR Probes

Probe name	Company	Assay ID
18S	Thermo Fisher	Hs99999901_s1
PIWIL2	Thermo Fisher	Hs01032719_m1
L1RE1(APFVNJ2)	Thermo Fisher	custom sequence

ACKNOWLEDGMENTS

We would like to thank Drs. Christiana Kappler and Stephen Duncan, Cell Models Core, Medical University of South Carolina Center of Biomedical Research Excellence (COBRE) in Digestive & Liver Disease (CDLD; National Institutes of Health P20 GM130457) for support with CRISPR/Cas9 reagents.

COMPETING INTERESTS

No competing interests declared.

FUNDING

This work was supported by National Institute of Health grants R01 DK124553, R01 DK136658, R21 CA246233, P20 GM130457 (COBRE in Digestive & Liver Disease, MUSC), P30 DK123704 (Digestive Disease Research Center, MUSC), to AK; R01 GM160815 to VG; and DP2 CA280626 to JRD. A.R. was supported by National Institutes of Health training grants TL1 TR001451, UL1 TR001450, T32 5T32DK124191, and 1F31DK138780-01.

DATA AND RESOURCE AVAILABILITY

All data are included within the manuscript, Figures, Supplementary Figures and Tables. The piRNA sequencing raw data shown in Fig 6 and Table S2 are being deposited to GEO and an accession number is pending and will be provided upon acceptance of the manuscript. The PIWIL2-FLAG construct will be deposited on Addgene and will be available upon acceptance of the manuscript.

REFERENCES

ARIUMI, Y. 2016. Guardian of the Human Genome: Host Defense Mechanisms against LINE-1 Retrotransposition. *Front Chem,* 4**,** 28.

BECK, C. R., COLLIER, P., MACFARLANE, C., MALIG, M., KIDD, J. M., EICHLER, E. E., BADGE, R. M. & MORAN, J. V. 2010. LINE-1 retrotransposition activity in human genomes. *Cell,* 141**,** 1159-70.

BECK, C. R., GARCIA-PEREZ, J. L., BADGE, R. M. & MORAN, J. V. 2011. LINE-1 elements in structural variation and disease. *Annu Rev Genomics Hum Genet,* 12**,** 187-215.

BERRENS, R. V., ANDREWS, S., SPENSBERGER, D., SANTOS, F., DEAN, W., GOULD, P., SHARIF, J., OLOVA, N., CHANDRA, T., KOSEKI, H., VON MEYENN, F. & REIK, W. 2017. An endosiRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells. *Cell Stem Cell,* 21**,** 694-703 e7.

BOWERS, R. R., JONES, C. M., PAZ, E. A., BARROWS, J. K., ARMESON, K. E., LONG, D. T. & DELANEY, J. R. 2022. SWAN pathway-network identification of common aneuploidy-based oncogenic drivers. *Nucleic Acids Res,* 50**,** 3673-3692.

CAJUSO, T., SULO, P., TANSKANEN, T., KATAINEN, R., TAIRA, A., HANNINEN, U. A., KONDELIN, J., FORSSTROM, L., VALIMAKI, N., AAVIKKO, M., KAASINEN, E., RISTIMAKI, A., KOSKENSALO, S., LEPISTO, A., RENKONEN-SINISALO, L., SEPPALA, T., KUOPIO, T., BOHM, J., MECKLIN, J. P., KILPIVAARA, O., PITKANEN, E., PALIN, K. & AALTONEN, L. A. 2019. Retrotransposon insertions can initiate colorectal cancer and are associated with poor survival. *Nat Commun,* 10**,** 4022.

CERAMI, E., GAO, J., DOGRUSOZ, U., GROSS, B. E., SUMER, S. O., AKSOY, B. A., JACOBSEN, A., BYRNE, C. J., HEUER, M. L., LARSSON, E., ANTIPIN, Y., REVA, B., GOLDBERG, A. P., SANDER, C. & SCHULTZ, N. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov,* 2**,** 401-4.

CHANDRASHEKAR, D. S., BASHEL, B., BALASUBRAMANYA, S. A. H., CREIGHTON, C. J., PONCE-RODRIGUEZ, I., CHAKRAVARTHI, B. & VARAMBALLY, S. 2017. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia,* 19**,** 649-658.

CHANDRASHEKAR, D. S., KARTHIKEYAN, S. K., KORLA, P. K., PATEL, H., SHOVON, A. R., ATHAR, M., NETTO, G. J., QIN, Z. S., KUMAR, S., MANNE, U., CREIGHTON, C. J. & VARAMBALLY, S. 2022. UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia,* 25**,** 18-27.

DAULAGALA, A. C., CETIN, M., NAIR-MENON, J., JIMENEZ, D. W., BRIDGES, M. C., BRADSHAW, A. D., SAHIN, O. & KOURTIDIS, A. 2024. The epithelial adherens junction component PLEKHA7 regulates ECM remodeling and cell behavior through miRNA-mediated regulation of MMP1 and LOX. *bioRxiv 10.1101/2024.05.28.596237*.

DE BRUIJN, I., KUNDRA, R., MASTROGIACOMO, B., TRAN, T. N., SIKINA, L., MAZOR, T., LI, X., OCHOA, A., ZHAO, G., LAI, B., ABESHOUSE, A., BAICEANU, D., CIFTCI, E., DOGRUSOZ, U., DUFILIE, A., ERKOC, Z., GARCIA LARA, E., FU, Z., GROSS, B., HAYNES, C., HEATH, A., HIGGINS, D., JAGANNATHAN, P., KALLETLA, K., KUMARI, P., LINDSAY, J., LISMAN, A., LEENKNEGT, B., LUKASSE, P., MADELA, D., MADUPURI, R., VAN NIEROP, P., PLANTALECH, O., QUACH, J., RESNICK, A. C., RODENBURG, S. Y. A., SATRAVADA, B. A., SCHAEFFER, F., SHERIDAN, R., SINGH, J., SIROHI, R., SUMER, S. O., VAN HAGEN, S., WANG, A., WILSON, M., ZHANG, H., ZHU, K., RUSK, N., BROWN, S., LAVERY, J. A., PANAGEAS, K. S., RUDOLPH, J. E., LENOUE-NEWTON, M. L., WARNER, J. L., GUO, X., HUNTER-ZINCK, H., YU, T. V., PILAI, S., NICHOLS, C., GARDOS, S. M., PHILIP, J., AACR PROJECT GENIE BPC CORE TEAM, A. P. G. C., KEHL, K. L., RIELY, G. J., SCHRAG, D., LEE, J., FIANDALO, M. V., SWEENEY, S. M., PUGH, T. J., SANDER, C., CERAMI, E., GAO, J. & SCHULTZ, N. 2023. Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Res,* 83**,** 3861-3867.

DE FAZIO, S., BARTONICEK, N., DI GIACOMO, M., ABREU-GOODGER, C., SANKAR, A., FUNAYA, C., ANTONY, C., MOREIRA, P. N., ENRIGHT, A. J. & O'CARROLL, D. 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature,* 480**,** 259-63.

ECKSTEIN, M., JUNG, R., WEIGELT, K., SIKIC, D., STOHR, R., GEPPERT, C., AGAIMY, A., LIEB, V., HARTMANN, A., WULLICH, B., WACH, S. & TAUBERT, H. 2018. Piwi-like 1 and -2 protein expression levels are prognostic factors for muscle invasive urothelial bladder cancer patients. *Sci Rep,* 8**,** 17693.

ELBARBARY, R. A., LUCAS, B. A. & MAQUAT, L. E. 2016. Retrotransposons as regulators of gene expression. *Science,* 351**,** aac7247.

EWING, A. D., GACITA, A., WOOD, L. D., MA, F., XING, D., KIM, M. S., MANDA, S. S., ABRIL, G., PEREIRA, G., MAKOHON-MOORE, A., LOOIJENGA, L. H., GILLIS, A. J., HRUBAN, R. H., ANDERS, R. A., ROMANS, K. E., PANDEY, A., IACOBUZIO-DONAHUE, C. A., VOGELSTEIN, B., KINZLER, K. W., KAZAZIAN, H. H., JR. & SOLYOM, S. 2015. Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res,* 25**,** 1536-45.

FLEMING, M., RAVULA, S., TATISHCHEV, S. F. & WANG, H. L. 2012. Colorectal carcinoma: Pathologic aspects. *J Gastrointest Oncol,* 3**,** 153-73.

FROST, R. J., HAMRA, F. K., RICHARDSON, J. A., QI, X., BASSEL-DUBY, R. & OLSON, E. N. 2010. MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. *Proc Natl Acad Sci U S A,* 107**,** 11847-52.

FU, K., TIAN, S., TAN, H., WANG, C., WANG, H., WANG, M., WANG, Y., CHEN, Z., WANG, Y., YUE, Q., XU, Q., ZHANG, S., LI, H., XIE, J., LIN, M., LUO, M., CHEN, F., YE, L. & ZHENG, K. 2019. Biological and RNA regulatory function of MOV10 in mammalian germ cells. *BMC Biol,* 17**,** 39.

GAINETDINOV, I. V., SKVORTSOVA, Y. V., STUKACHEVA, E. A., BYCHENKO, O. S., KONDRATIEVA, S. A., ZINOVIEVA, M. V. & AZHIKINA, T. L. 2014. Expression profiles of PIWIL2 short isoforms differ in testicular germ cell tumors of various differentiation subtypes. *PLoS One,* 9**,** e112528.

GAO, J., AKSOY, B. A., DOGRUSOZ, U., DRESDNER, G., GROSS, B., SUMER, S. O., SUN, Y., JACOBSEN, A., SINHA, R., LARSSON, E., CERAMI, E., SANDER, C. & SCHULTZ, N. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal,* 6**,** pl1.

GAO, J., JING, J., SHANG, G., CHEN, C., DUAN, M., YU, W., WANG, K., LUO, J., SONG, M., CHEN, K., CHEN, C., ZHANG, T. & DING, D. 2024. TDRD1 phase separation drives intermitochondrial cement assembly to promote piRNA biogenesis and fertility. *Dev Cell,* 59**,** 2704-2718 e6.

GAO, J. X. 2008. Cancer stem cells: the lessons from pre-cancerous stem cells. *J Cell Mol Med,* 12**,** 67-96.

GARCIA-PEREZ, J. L., MORELL, M., SCHEYS, J. O., KULPA, D. A., MORELL, S., CARTER, C. C., HAMMER, G. D., COLLINS, K. L., O'SHEA, K. S., MENENDEZ, P. & MORAN, J. V. 2010. Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature,* 466**,** 769-73.

GASIOR, S. L., WAKEMAN, T. P., XU, B. & DEININGER, P. L. 2006. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol,* 357**,** 1383-93.

GORDON, I. O., AGRAWAL, N., WILLIS, E., GOLDBLUM, J. R., LOPEZ, R., ALLENDE, D., LIU, X., PATIL, D. Y., YERIAN, L., EL-KHIDER, F., FIOCCHI, C. & RIEDER, F. 2018. Fibrosis in ulcerative colitis is directly linked to severity and chronicity of mucosal inflammation. *Aliment Pharmacol Ther,* 47**,** 922-939.

GRANDI, F. C. & AN, W. 2013. Non-LTR retrotransposons and microsatellites: Partners in genomic variation. *Mob Genet Elements,* 3**,** e25674.

GRASSET, E., PINTO, M., DUSSAULX, E., ZWEIBAUM, A. & DESJEUX, J. F. 1984. Epithelial properties of human colonic carcinoma cell line Caco-2: electrical parameters. *Am J Physiol,* 247**,** C260-7.

GUIDOBONI, M., GAFA, R., VIEL, A., DOGLIONI, C., RUSSO, A., SANTINI, A., DEL TIN, L., MACRI, E., LANZA, G., BOIOCCHI, M. & DOLCETTI, R. 2001. Microsatellite instability and high content of activated cytotoxic lymphocytes identify colon cancer patients with a favorable prognosis. *Am J Pathol,* 159**,** 297-304.

GYORFFY, B. 2024. Integrated analysis of public datasets for the discovery and validation of survival-associated genes in solid tumors. *Innovation (Camb),* 5**,** 100625.

HAN, B. W., WANG, W., LI, C., WENG, Z. & ZAMORE, P. D. 2015a. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science,* 348**,** 817-21.

HAN, B. W., WANG, W., ZAMORE, P. D. & WENG, Z. 2015b. piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics,* 31**,** 593-5.

HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell,* 144**,** 646-74.

HANCKS, D. C. & KAZAZIAN, H. H., JR. 2016. Roles for retrotransposon insertions in human disease. *Mob DNA,* 7**,** 9.

HE, G., CHEN, L., YE, Y., XIAO, Y., HUA, K., JARJOURA, D., NAKANO, T., BARSKY, S. H., SHEN, R. & GAO, J. X. 2010. Piwil2 expressed in various stages of cervical neoplasia is a potential complementary marker for p16. *Am J Transl Res,* 2**,** 156-69.

HIDALGO, I. J., RAUB, T. J. & BORCHARDT, R. T. 1989. Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal epithelial permeability. *Gastroenterology,* 96**,** 736-49.

755 HUANG, A., YANG, Y., SHI, J. Y., LI, Y. K., XU, J. X., CHENG, Y. & GU, J. 2021. Mucinous adenocarcinoma: A unique
756 clinicopathological subtype in colorectal cancer. *World J Gastrointest Surg,* 13**,** 1567-1583.
757 IKEDA, S., TANAKA, K., OHTANI, R., KANDA, A., SOTOMARU, Y., KONO, T. & OBATA, Y. 2022. Disruption of piRNA
758 machinery by deletion of ASZ1/GASZ results in the expression of aberrant chimeric transcripts in
759 gonocytes. *J Reprod Dev,* 68**,** 125-136.
760 JACOBS, D. I., QIN, Q., LERRO, M. C., FU, A., DUBROW, R., CLAUS, E. B., DEWAN, A. T., WANG, G., LIN, H. & ZHU, Y.
761 2016. PIWI-Interacting RNAs in Gliomagenesis: Evidence from Post-GWAS and Functional Analyses.
762 *Cancer Epidemiol Biomarkers Prev,* 25**,** 1073-80.
763 JANG, H. S., SHAH, N. M., DU, A. Y., DAILEY, Z. Z., PEHRSSON, E. C., GODOY, P. M., ZHANG, D., LI, D., XING, X., KIM,
764 S., O'DONNELL, D., GORDON, J. I. & WANG, T. 2019. Transposable elements drive widespread expression
765 of oncogenes in human cancers. *Nat Genet,* 51**,** 611-617.
766 KANKE, M., KENNEDY NG, M. M., CONNELLY, S., SINGH, M., SCHANER, M., SHANAHAN, M. T., WOLBER, E. A.,
767 BEASLEY, C., LIAN, G., JAIN, A., LONG, M. D., BARNES, E. L., HERFARTH, H. H., ISAACS, K. L., HANSEN, J. J.,
768 KAPADIA, M., GUILLEM, J. G., FESCHOTTE, C., FUREY, T. S., SHEIKH, S. Z. & SETHUPATHY, P. 2022. Single-
769 Cell Analysis Reveals Unexpected Cellular Changes and Transposon Expression Signatures in the Colonic
770 Epithelium of Treatment-Naive Adult Crohn's Disease Patients. *Cell Mol Gastroenterol Hepatol,* 13**,** 1717-
771 1740.
772 KIM, E. R. & CHANG, D. K. 2014. Colorectal cancer in inflammatory bowel disease: the risk, pathogenesis,
773 prevention and diagnosis. *World J Gastroenterol,* 20**,** 9872-81.
774 KIM, K. H., KANG, Y. J., JO, J. O., OCK, M. S., MOON, S. H., SUH, D. S., YOON, M. S., PARK, E. S., JEONG, N., EO, W.
775 K., KIM, H. Y. & CHA, H. J. 2014. DDX4 (DEAD box polypeptide 4) colocalizes with cancer stem cell marker
776 CD133 in ovarian cancers. *Biochem Biophys Res Commun,* 447**,** 315-22.
777 KISHANI FARAHANI, R., SOLEIMANPOUR, S., GOLMOHAMMADI, M. & SOLEIMANPOUR-LICHAEI, H. R. 2023.
778 PIWIL2 Regulates the Proliferation, Apoptosis and Colony Formation of Colorectal Cancer Cell Line. *Iran J*
779 *Biotechnol,* 21**,** e3176.
780 KOPERA, H. C., LARSON, P. A., MOLDOVAN, J. B., RICHARDSON, S. R., LIU, Y. & MORAN, J. V. 2016. LINE-1 Cultured
781 Cell Retrotransposition Assay. *Methods Mol Biol,* 1400**,** 139-56.
782 KOURTIDIS, A., NGOK, S. P., PULIMENO, P., FEATHERS, R. W., CARPIO, L. R., BAKER, T. R., CARR, J. M., YAN, I. K.,
783 BORGES, S., PEREZ, E. A., STORZ, P., COPLAND, J. A., PATEL, T., THOMPSON, E. A., CITI, S. &
784 ANASTASIADIS, P. Z. 2015. Distinct E-cadherin-based complexes regulate cell behaviour through miRNA
785 processing or Src and p120 catenin activity. *Nat Cell Biol,* 17**,** 1145-57.
786 KU, H. Y. & LIN, H. 2014. PIWI proteins and their interactors in piRNA biogenesis, germline development and gene
787 expression. *Natl Sci Rev,* 1**,** 205-218.
788 KUO, L. J. & YANG, L. X. 2008. Gamma-H2AX - a novel biomarker for DNA double-strand breaks. *In Vivo,* 22**,** 305-9.
789 KURAMOCHI-MIYAGAWA, S., WATANABE, T., GOTOH, K., TOTOKI, Y., TOYODA, A., IKAWA, M., ASADA, N., KOJIMA,
790 K., YAMAGUCHI, Y., IJIRI, T. W., HATA, K., LI, E., MATSUDA, Y., KIMURA, T., OKABE, M., SAKAKI, Y., SASAKI, H.
791 & NAKANO, T. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI
792 and MIWI2 in murine fetal testes. *Genes Dev,* 22**,** 908-17.
793 LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.
794 LANGMEAD, B., WILKS, C., ANTONESCU, V. & CHARLES, R. 2019. Scaling read aligners to hundreds of threads on
795 general-purpose processors. *Bioinformatics,* 35**,** 421-432.
796 LEE, E., LOKMAN, N. A., OEHLER, M. K., RICCIARDELLI, C. & GRUTZNER, F. 2020. A Comprehensive Molecular and
797 Clinical Analysis of the piRNA Pathway Genes in Ovarian Cancer. *Cancers (Basel),* 13.
798 LEE, J. H., SCHUTTE, D., WULF, G., FUZESI, L., RADZUN, H. J., SCHWEYER, S., ENGEL, W. & NAYERNIA, K. 2006.
799 Stem-cell protein Piwil2 is widely expressed in tumors and inhibits apoptosis through activation of
800 Stat3/Bcl-XL pathway. *Hum Mol Genet,* 15**,** 201-11.
801 LI, Y., ZHANG, Y. & LIU, M. 2021. Knockout Gene-Based Evidence for PIWI-Interacting RNA Pathway in Mammals.
802 *Front Cell Dev Biol,* 9**,** 681188.
803 LINGGI, B., JAIRATH, V., ZOU, G., SHACKELTON, L. M., MCGOVERN, D. P. B., SALAS, A., VERSTOCKT, B.,
804 SILVERBERG, M. S., NAYERI, S., FEAGAN, B. G. & VANDE CASTEELE, N. 2021. Meta-analysis of gene
805 expression disease signatures in colonic biopsy tissue from patients with ulcerative colitis. *Sci Rep,* 11**,**
806 18243.

807 LIU, J. J., SHEN, R., CHEN, L., YE, Y., HE, G., HUA, K., JARJOURA, D., NAKANO, T., RAMESH, G. K., SHAPIRO, C. L.,
808 BARSKY, S. H. & GAO, J. X. 2010. Piwil2 is expressed in various stages of breast cancers and has the
809 potential to be used as a novel biomarker. *Int J Clin Exp Pathol,* 3**,** 328-37.
810 LU, Y., ZHANG, K., LI, C., YAO, Y., TAO, D., LIU, Y., ZHANG, S. & MA, Y. 2012. Piwil2 suppresses p53 by inducing
811 phosphorylation of signal transducer and activator of transcription 3 in tumor cells. *PLoS One,* 7**,** e30999.
812 LUO, C., CEN, S., DING, G. & WU, W. 2019. Mucinous colorectal adenocarcinoma: clinical pathology and
813 treatment options. *Cancer Commun (Lond),* 39**,** 13.
814 MA, L., BUCHOLD, G. M., GREENBAUM, M. P., ROY, A., BURNS, K. H., ZHU, H., HAN, D. Y., HARRIS, R. A., COARFA,
815 C., GUNARATNE, P. H., YAN, W. & MATZUK, M. M. 2009. GASZ is essential for male meiosis and suppression
816 of retrotransposon expression in the male germline. *PLoS Genet,* 5**,** e1000635.
817 MARTINEZ, V. D., VUCIC, E. A., THU, K. L., HUBAUX, R., ENFIELD, K. S., PIKOR, L. A., BECKER-SANTOS, D. D.,
818 BROWN, C. J., LAM, S. & LAM, W. L. 2015. Unique somatic and malignant expression patterns implicate
819 PIWI-interacting RNAs in cancer-type specific biology. *Sci Rep,* 5**,** 10423.
820 MATHIOUDAKIS, N., PALENCIA, A., KADLEC, J., ROUND, A., TRIPSIANES, K., SATTLER, M., PILLAI, R. S. & CUSACK,
821 S. 2012. The multiple Tudor domain-containing protein TDRD1 is a molecular scaffold for mouse Piwi
822 proteins and piRNA biogenesis factors. *RNA,* 18**,** 2056-72.
823 MESEURE, D., VACHER, S., BOUDJEMAA, S., LAE, M., NICOLAS, A., LECLERE, R., CHEMLALI, W., CHAMPENOIS,
824 G., SCHNITZLER, A., LESAGE, L., DUBOIS, T. & BIECHE, I. 2020. Biopathological Significance of PIWI-piRNA
825 Pathway Deregulation in Invasive Breast Carcinomas. *Cancers (Basel),* 12.
826 MIKI, Y., NISHISHO, I., HORII, A., MIYOSHI, Y., UTSUNOMIYA, J., KINZLER, K. W., VOGELSTEIN, B. & NAKAMURA, Y.
827 1992. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer*
828 *Res,* 52**,** 643-5.
829 MORAN, J. V., HOLMES, S. E., NAAS, T. P., DEBERARDINIS, R. J., BOEKE, J. D. & KAZAZIAN, H. H., JR. 1996. High
830 frequency retrotransposition in cultured mammalian cells. *Cell,* 87**,** 917-27.
831 NAM, C. H., YOUK, J., KIM, J. Y., LIM, J., PARK, J. W., OH, S. A., LEE, H. J., PARK, J. W., WON, H., LEE, Y., JEONG, S. Y.,
832 LEE, D. S., OH, J. W., HAN, J., LEE, J., KWON, H. W., KIM, M. J. & JU, Y. S. 2023. Widespread somatic L1
833 retrotransposition in normal colorectal epithelium. *Nature,* 617**,** 540-547.
834 NISHIMASU, H., ISHIZU, H., SAITO, K., FUKUHARA, S., KAMATANI, M. K., BONNEFOND, L., MATSUMOTO, N.,
835 NISHIZAWA, T., NAKANAGA, K., AOKI, J., ISHITANI, R., SIOMI, H., SIOMI, M. C. & NUREKI, O. 2012. Structure
836 and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature,* 491**,** 284-7.
837 OLOTU, O., KOSKENNIEMI, A. R., MA, L., PARAMONOV, V., LAASANEN, S., LOURAMO, E., BOURGERY, M.,
838 LEHTINIEMI, T., LAASANEN, S., RIVERO-MULLER, A., LOYTTYNIEMI, E., SAHLGREN, C., WESTERMARCK, J.,
839 VENTELA, S., VISAKORPI, T., POUTANEN, M., VAINIO, P., MAKELA, J. A. & KOTAJA, N. 2024. Germline-
840 specific RNA helicase DDX4 forms cytoplasmic granules in cancer cells and promotes tumor growth. *Cell*
841 *Rep,* 43**,** 114430.
842 OSTERTAG, E. M., PRAK, E. T., DEBERARDINIS, R. J., MORAN, J. V. & KAZAZIAN, H. H., JR. 2000. Determination of L1
843 retrotransposition kinetics in cultured cells. *Nucleic Acids Res,* 28**,** 1418-23.
844 PANDEY, R. R., TOKUZAWA, Y., YANG, Z., HAYASHI, E., ICHISAKA, T., KAJITA, S., ASANO, Y., KUNIEDA, T.,
845 SACHIDANANDAM, R., CHUMA, S., YAMANAKA, S. & PILLAI, R. S. 2013. Tudor domain containing 12
846 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice. *Proc Natl Acad Sci U S A,* 110**,**
847 16492-7.
848 PENG, J. C. & LIN, H. 2013. Beyond transposons: the epigenetic and somatic functions of the Piwi-piRNA
849 mechanism. *Curr Opin Cell Biol,* 25**,** 190-4.
850 PITKANEN, E., CAJUSO, T., KATAINEN, R., KAASINEN, E., VALIMAKI, N., PALIN, K., TAIPALE, J., AALTONEN, L. A. &
851 KILPIVAARA, O. 2014. Frequent L1 retrotranspositions originating from TTC28 in colorectal cancer.
852 *Oncotarget,* 5**,** 853-9.
853 RAMAT, A. & SIMONELIG, M. 2021. Functions of PIWI Proteins in Gene Regulation: New Arrows Added to the piRNA
854 Quiver. *Trends Genet,* 37**,** 188-200.
855 RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the
856 CRISPR-Cas9 system. *Nat Protoc,* 8**,** 2281-2308.
857 REUTER, M., CHUMA, S., TANAKA, T., FRANZ, T., STARK, A. & PILLAI, R. S. 2009. Loss of the Mili-interacting Tudor
858 domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat*
859 *Struct Mol Biol,* 16**,** 639-46.

RODRIGUEZ-MARTIN, B., ALVAREZ, E. G., BAEZ-ORTEGA, A., ZAMORA, J., SUPEK, F., DEMEULEMEESTER, J., SANTAMARINA, M., JU, Y. S., TEMES, J., GARCIA-SOUTO, D., DETERING, H., LI, Y., RODRIGUEZ-CASTRO, J., DUESO-BARROSO, A., BRUZOS, A. L., DENTRO, S. C., BLANCO, M. G., CONTINO, G., ARDELJAN, D., TOJO, M., ROBERTS, N. D., ZUMALAVE, S., EDWARDS, P. A., WEISCHENFELDT, J., PUIGGROS, M., CHONG, Z., CHEN, K., LEE, E. A., WALA, J. A., RAINE, K. M., BUTLER, A., WASZAK, S. M., NAVARRO, F. C. P., SCHUMACHER, S. E., MONLONG, J., MAURA, F., BOLLI, N., BOURQUE, G., GERSTEIN, M., PARK, P. J., WEDGE, D. C., BEROUKHIM, R., TORRENTS, D., KORBEL, J. O., MARTINCORENA, I., FITZGERALD, R. C., VAN LOO, P., KAZAZIAN, H. H., BURNS, K. H., GROUP, P. S. V. W., CAMPBELL, P. J., TUBIO, J. M. C. & CONSORTIUM, P. 2020. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat Genet,* 52**,** 306-319.

ROSENKRANZ, D., HAN, C. T., ROOVERS, E. F., ZISCHLER, H. & KETTING, R. F. 2015. Piwi proteins and piRNAs in mammalian oocytes and early embryos: From sample to sequence. *Genom Data,* 5**,** 309-13.

ROSS, R. J., WEINER, M. M. & LIN, H. 2014. PIWI proteins and PIWI-interacting RNAs in the soma. *Nature,* 505**,** 353-359.

SAMBUY, Y., DE ANGELIS, I., RANALDI, G., SCARINO, M. L., STAMMATI, A. & ZUCCO, F. 2005. The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics. *Cell Biol Toxicol,* 21**,** 1-26.

SAXE, J. P., CHEN, M., ZHAO, H. & LIN, H. 2013. Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. *EMBO J,* 32**,** 1869-85.

SCOTT, E. C., GARDNER, E. J., MASOOD, A., CHUANG, N. T., VERTINO, P. M. & DEVINE, S. E. 2016. A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res,* 26**,** 745-55.

SHI, S., YANG, Z. Z., LIU, S., YANG, F. & LIN, H. 2020. PIWIL1 promotes gastric cancer via a piRNA-independent mechanism. *Proc Natl Acad Sci U S A,* 117**,** 22390-22401.

SIEGEL, R. L., GIAQUINTO, A. N. & JEMAL, A. 2024. Cancer statistics, 2024. *CA Cancer J Clin,* 74**,** 12-49.

SIGURDSSON, M. I., SMITH, A. V., BJORNSSON, H. T. & JONSSON, J. J. 2012. The distribution of a germline methylation marker suggests a regional mechanism of LINE-1 silencing by the piRNA-PIWI system. *BMC Genet,* 13**,** 31.

SIOMI, H. & SIOMI, M. C. 2015. RNA. Phased piRNAs tackle transposons. *Science,* 348**,** 756-7.

SIOMI, M. C., MANNEN, T. & SIOMI, H. 2010. How does the royal family of Tudor rule the PIWI-interacting RNA pathway? *Genes Dev,* 24**,** 636-46.

SIOMI, M. C., SATO, K., PEZIC, D. & ARAVIN, A. A. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol,* 12**,** 246-58.

SOLYOM, S., EWING, A. D., RAHRMANN, E. P., DOUCET, T., NELSON, H. H., BURNS, M. B., HARRIS, R. S., SIGMON, D. F., CASELLA, A., ERLANGER, B., WHEELAN, S., UPTON, K. R., SHUKLA, R., FAULKNER, G. J., LARGAESPADA, D. A. & KAZAZIAN, H. H., JR. 2012. Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res,* 22**,** 2328-38.

SUN, Y. H., LEE, B. & LI, X. Z. 2022. The birth of piRNAs: how mammalian piRNAs are produced, originated, and evolved. *Mamm Genome,* 33**,** 293-311.

TANG, Z., KANG, B., LI, C., CHEN, T. & ZHANG, Z. 2019. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res,* 47**,** W556-W560.

TUBIO, J. M., LI, Y., JU, Y. S., MARTINCORENA, I., COOKE, S. L., TOJO, M., GUNDEM, G., PIPINIKAS, C. P., ZAMORA, J., RAINE, K., MENZIES, A., ROMAN-GARCIA, P., FULLAM, A., GERSTUNG, M., SHLIEN, A., TARPEY, P. S., PAPAEMMANUIL, E., KNAPPSKOG, S., VAN LOO, P., RAMAKRISHNA, M., DAVIES, H. R., MARSHALL, J., WEDGE, D. C., TEAGUE, J. W., BUTLER, A. P., NIK-ZAINAL, S., ALEXANDROV, L., BEHJATI, S., YATES, L. R., BOLLI, N., MUDIE, L., HARDY, C., MARTIN, S., MCLAREN, S., O'MEARA, S., ANDERSON, E., MADDISON, M., GAMBLE, S., GROUP, I. B. C., GROUP, I. B. C., GROUP, I. P. C., FOSTER, C., WARREN, A. Y., WHITAKER, H., BREWER, D., EELES, R., COOPER, C., NEAL, D., LYNCH, A. G., VISAKORPI, T., ISAACS, W. B., VAN'T VEER, L., CALDAS, C., DESMEDT, C., SOTIRIOU, C., APARICIO, S., FOEKENS, J. A., EYFJORD, J. E., LAKHANI, S. R., THOMAS, G., MYKLEBOST, O., SPAN, P. N., BORRESEN-DALE, A. L., RICHARDSON, A. L., VAN DE VIJVER, M., VINCENT-SALOMON, A., VAN DEN EYNDEN, G. G., FLANAGAN, A. M., FUTREAL, P. A., JANES, S. M., BOVA, G. S., STRATTON, M. R., MCDERMOTT, U. & CAMPBELL, P. J. 2014. Mobile DNA in cancer. Extensive

912 transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science,* 345**,**
913 1251343.
914 UNHAVAITHAYA, Y., HAO, Y., BEYRET, E., YIN, H., KURAMOCHI-MIYAGAWA, S., NAKANO, T. & LIN, H. 2009. MILI, a
915 PIWI-interacting RNA-binding protein, is required for germ line stem cell self-renewal and appears to
916 positively regulate translation. *J Biol Chem,* 284**,** 6507-19.
917 VOUREKAS, A., ZHENG, K., FU, Q., MARAGKAKIS, M., ALEXIOU, P., MA, J., PILLAI, R. S., MOURELATOS, Z. & WANG,
918 P. J. 2015. The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. *Genes Dev,* 29**,**
919 617-29.
920 WALTER, M., TEISSANDIER, A., PEREZ-PALACIOS, R. & BOURC'HIS, D. 2016. An epigenetic switch ensures
921 transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife,* 5.
922 WANG, H. L., CHEN, B. B., CAO, X. G., WANG, J., HU, X. F., MU, X. Q. & CHEN, X. B. 2015a. The clinical significances
923 of the abnormal expressions of Piwil1 and Piwil2 in colonic adenoma and adenocarcinoma. *Onco Targets*
924 *Ther,* 8**,** 1259-64.
925 WANG, L., NORRIS, E. T. & JORDAN, I. K. 2017. Human Retrotransposon Insertion Polymorphisms Are Associated
926 with Health and Disease via Gene Regulatory Phenotypes. *Front Microbiol,* 8**,** 1418.
927 WANG, W., HAN, B. W., TIPPING, C., GE, D. T., ZHANG, Z., WENG, Z. & ZAMORE, P. D. 2015b. Slicing and Binding by
928 Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms. *Mol Cell,* 59**,** 819-30.
929 WANG, Z. Y., GE, L. P., OUYANG, Y., JIN, X. & JIANG, Y. Z. 2024. Targeting transposable elements in cancer:
930 developments and opportunities. *Biochim Biophys Acta Rev Cancer,* 1879**,** 189143.
931 WATANABE, T. & LIN, H. 2014. Posttranscriptional regulation of gene expression by Piwi proteins and piRNAs. *Mol*
932 *Cell,* 56**,** 18-27.
933 WEICK, E. M. & MISKA, E. A. 2014. piRNAs: from biogenesis to function. *Development,* 141**,** 3458-71.
934 WENDA, J. M., HOMOLKA, D., YANG, Z., SPINELLI, P., SACHIDANANDAM, R., PANDEY, R. R. & PILLAI, R. S. 2017.
935 Distinct Roles of RNA Helicases MVH and TDRD9 in PIWI Slicing-Triggered Mammalian piRNA Biogenesis
936 and Function. *Dev Cell,* 41**,** 623-637 e9.
937 XIOL, J., CORA, E., KOGLGRUBER, R., CHUMA, S., SUBRAMANIAN, S., HOSOKAWA, M., REUTER, M., YANG, Z.,
938 BERNINGER, P., PALENCIA, A., BENES, V., PENNINGER, J., SACHIDANANDAM, R. & PILLAI, R. S. 2012. A
939 role for Fkbp6 and the chaperone machinery in piRNA amplification and transposon silencing. *Mol Cell,* 47**,**
940 970-9.
941 YAN, Z., HU, H. Y., JIANG, X., MAIERHOFER, V., NEB, E., HE, L., HU, Y., HU, H., LI, N., CHEN, W. & KHAITOVICH, P.
942 2011. Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res,* 39**,** 6596-607.
943 YANG, Z., CHEN, K. M., PANDEY, R. R., HOMOLKA, D., REUTER, M., JANEIRO, B. K., SACHIDANANDAM, R.,
944 FAUVARQUE, M. O., MCCARTHY, A. A. & PILLAI, R. S. 2016. PIWI Slicing and EXD1 Drive Biogenesis of
945 Nuclear piRNAs from Cytosolic Targets of the Mouse piRNA Pathway. *Mol Cell,* 61**,** 138-52.
946 YE, Y., YIN, D. T., CHEN, L., ZHOU, Q., SHEN, R., HE, G., YAN, Q., TONG, Z., ISSEKUTZ, A. C., SHAPIRO, C. L.,
947 BARSKY, S. H., LIN, H., LI, J. J. & GAO, J. X. 2010. Identification of Piwil2-like (PL2L) proteins that promote
948 tumorigenesis. *PLoS One,* 5**,** e13406.
949 ZENG, G., ZHANG, D., LIU, X., KANG, Q., FU, Y., TANG, B., GUO, W., ZHANG, Y., WEI, G. & HE, D. 2017. Co-
950 expression of Piwil2/Piwil4 in nucleus indicates poor prognosis of hepatocellular carcinoma. *Oncotarget,* 8**,**
951 4607-4617.
952 ZENG, Q., WAN, H., ZHAO, S., XU, H., TANG, T., OWARE, K. A. & QU, S. 2020. Role of PIWI-interacting RNAs on cell
953 survival: Proliferation, apoptosis, and cycle. *IUBMB Life,* 72**,** 1870-1878.
954 ZHANG, X., YU, L., YE, S., XIE, J., HUANG, X., ZHENG, K. & SUN, B. 2019. MOV10L1 Binds RNA G-Quadruplex in a
955 Structure-Specific Manner and Resolves It More Efficiently Than MOV10. *iScience,* 17**,** 36-48.
956 ZHUO, C., LI, Q., WU, Y., LI, Y., NIE, J., LI, D., PENG, J., LIAN, P., LI, B., CAI, G., LI, X. & CAI, S. 2015. LINE-1
957 hypomethylation in normal colon mucosa is associated with poor survival in Chinese patients with
958 sporadic colon cancer. *Oncotarget,* 6**,** 23820-36.

959

960

961 **FIGURE LEGENDS**

962 **Figure 1. The PIWI-piRNA pathway experiences copy number alterations (CNAs) in colon cancer.**
963 Shifted Weighted Annotation Network (SWAN) analysis was performed to interrogate pathway and
964 CNAs within The Cancer Genome Atlas (TCGA) dataset. The Single-pathway SWAN tool was used to
965 perform Gene Ontology Biological Process Analysis on the piRNA metabolic pathway in colon
966 adenocarcinoma (COAD). A) SWAN pathway scoring for each member of the piRNA pathway where
967 blue represents negatively scored allelic losses and red represents positively scored copy number
968 gains within the network with an overall trend of haploinsufficiency as indicated by the grey line
969 (Wilcoxon rank sum $p < 1.7E-10$). B) Circos plot mapping the frequency of allelic losses (blue) or gains
970 (red) to each respective chromosome labeled on the outside ring. Gene labels are included if a score
971 was beyond one standard deviation from a zero-change value. C) Table of the percentages of CNAs
972 for each member of the piRNA pathway ranked by percent biallelic deletions. D) The cBioportal
973 database was used to interrogate TCGA COAD dataset for Overall survival of PIWIL2 allelic loss
974 compared to unaltered tumors. Red indicates PIWIL2 shallow deletions while blue indicates samples
975 with no CNAs. (Logrank Test $p = 0.0204$)

976

977 **Figure 2. PIWIL2 is differentially expressed in colon tumors of advanced stage and nodal**
978 **metastasis, histological subtype, and age.** PIWIL2 expression was queried in the TCGA colon
979 adenocarcinoma (COAD) data set and stratified by comparing normal colon tissue vs A) primary tumor,
980 B) tumor stage, C) tumor nodal metastasis, D) tumor p53 status, E) tumor histological subtype, F) tumor
981 microsatellite stability status, as well as G) age, H) weight, I) race and J) biological sex of colon cancer
982 patients. Normal tissues were defined as adjacent non-cancerous tissue. For panels A-E and G-I the
983 UALCAN web-based tool was used to stratify TCGA RNA-seq data for more in-depth analysis of
984 expression profiles. The box and whisker plots display the median, interquartile range, minimum and
985 maximum values of PIWIL2 expression in transcript per million (TPM) for each dataset, and $*p<0.05$,
986 $**p<0.01$, $***p<0.001$. For panel F The GEPIA2 tool was utilized to further stratify PIWIL2 expression
987 in the TCGA COAD dataset by: micro-satellite stable (MSS), micro-satellite instable low (MSI-L), and
988 micro-satellite instable high (MSI-H). The GEPIA2 program calculated statistical significance using one-
989 way ANOVA ($*p<0.05$). K-M) The Kaplan-Meier Plotter tool was used to determine if PIWIL2 expression
990 impacted patient regression free survival for all stages combined (K), Stages 1+2+3 (L) and Stage 4(M).

991

992 **Figure 3. The promoter of PIWIL2 is highly methylated in colon cancer.** PIWIL2 promoter
993 methylation was queried in the TCGA colon adenocarcinoma (COAD) data set and stratified by
994 comparing normal colon tissue vs A) primary tumor, B) tumor stage, C) tumor nodal metastasis, D)
995 tumor p53 status, E) tumor histological subtype, as well as F) biological sex, G) age, H) weight, and I)

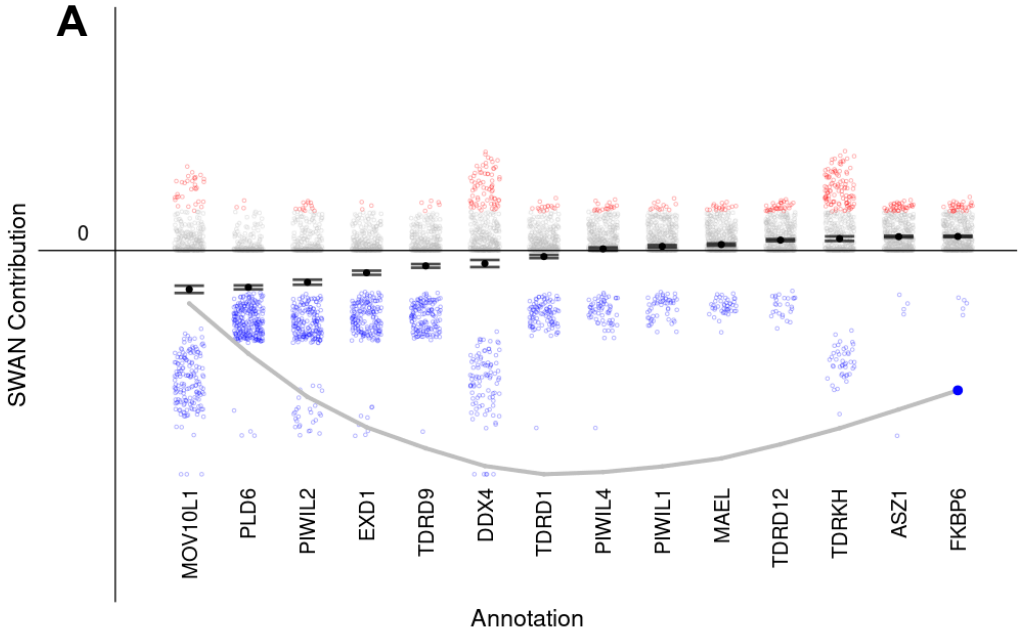
996 race of colon cancer patients. Normal tissues were defined as adjacent non-cancerous tissue. The
997 median, interquartile range, minimum and maximum values of the methylation beta value is displayed
998 and *p <0.05, **p<0.01, ***p<0.001. A beta value closer to 1 indicates higher levels of methylation.
999
000 **Figure 4. PIWIL2 is downregulated in ulcerative colitis patients compared to healthy, non-IBD**
001 **patients.** A meta-analysis of ulcerative colitis tissue gene expression was performed using the
002 PreMedIBD Gene Tool. A) PIWIL2, B) PIWIL4, C) MOV10L1, D) TDRD1, E) TDRD9, F) TDRD12, G)
003 DDX4, H) PLD6 and I) EXD1 were queried and the $\log_2$-fold gene expression changes between
004 ulcerative colitis patients and non-IBD controls were determined for each gene. The meta-analysis was
005 calculated using a random effects model and shows the meta-log2-Fold change and 95% confidence
006 interval. Ulcerative Colitis datasets used are available in the Gene Expression Omnibus: GSE13367,
007 GSE9452, GSE53306, GSE38713, GSE47908, GSE73661, GSE114527, GSE87466 and are
008 represented by grey dots. The summary fold change is displayed with a blue line indicating
009 downregulation, while a red line indicates upregulation. A $\log_2$-fold-change of 0.5 was used as a
010 threshold for significance.
011
012 **Figure 5. PIWIL2 suppresses pro-tumorigenic phenotypes.** A) qPCR of PIWIL2 comparing PIWIL2
013 KO to Caco2 wild type (WT). Data were analyzed and normalized to 18S ribosomal RNA by ΔΔCt
014 calculations to calculate fold change. (n=3 biological replicates, mean ± SE, t-test *p=0.0309). B)
015 Quantification of the upper band of PIWIL2 by western blot to confirm the knockout (n=3 biological
016 replicates, mean ± SE, t-test **p = 0.0025). C) Representative western blot of total PIWIL2 and β-actin
017 (Actin) loading control in Caco2 WT and PIWIL2 KO cells. The band at the expected molecular weight
018 of full length PIWIL2 of 109 kDa was quantified and indicated by a red arrow. D) Caco2 WT and PIWIL2
019 KO cells were infected with an LZRS empty vector (LZRS), and PIWIL2 KO cells were infected with a
020 FLAG-tagged PIWIL2 construct (PIWIL2-FLAG) to show rescue of the upper band of PIWIL2 by western
021 blot (red arrow), FLAG, and β-actin (Actin) E) xCELLigence cellular impedance assay to examine the
022 proliferation rates of Caco2+LZRS, PIWIL2 KO+LZRS and PIWIL2 KO+PIWIL2-FLAG cells in 2-
023 dimensions (2D) (n=3 biological replicates, two-way repeated measures ANOVA with Geisser-
024 Greenhouse correction and Bonferroni correction for multiple comparisons (Caco2+LZRS vs PIWIL2
025 KO+LZRS **p<0.01, Caco2+LZRS vs PIWIL2 KO+PIWIL2-FLAG *p<0.05, PIWIL2 KO+LZRS vs
026 PIWIL2 KO+PIWIL2-FLAG n.s). F) Representative images of Caco2+LZRS, PIWIL2 KO+LZRS and
027 PIWIL2 KO+PIWIL2-FLAG cells grown in anchorage independent conditions after 7 days. G) Area, H)
028 Perimeter I) Diameter and J) Circularity of Caco2+LZRS, PIWIL2 KO+LZRS and PIWIL2 KO+PIWIL2-
029 FLAG spheroids (*p<0.05). For the low attachment assay n=3 biological replicates, mean ± SE is

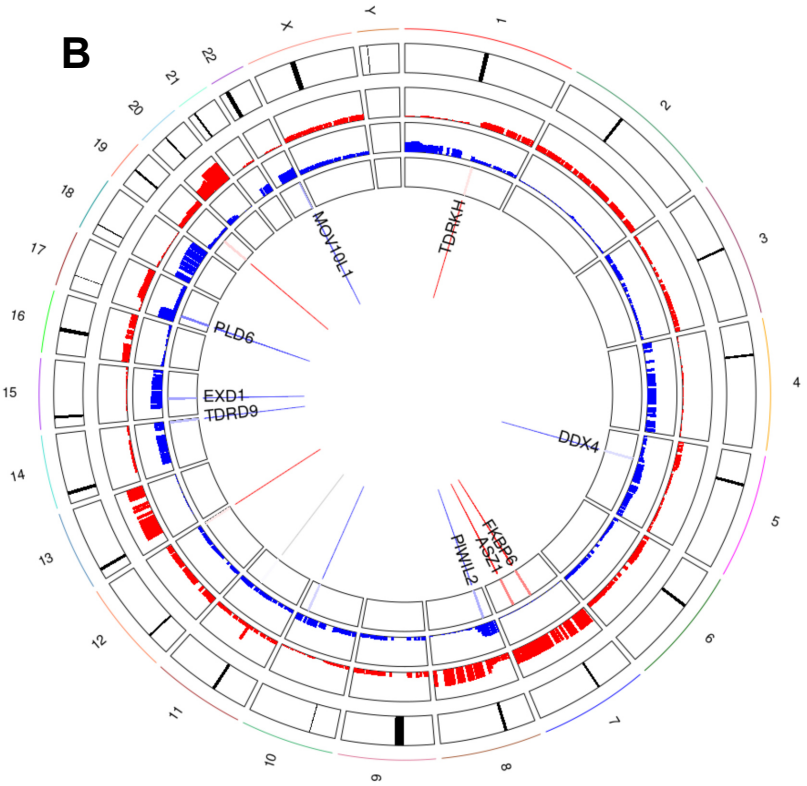
030 displayed, and a one-way ANOVA with a Bonferroni correction for multiple comparisons was used for
031 statistical analysis between groups.
032
033 **Figure 6. RNA sequencing analysis shows that PIWIL2 regulates a set of non-canonical piRNAs**
034 **targeting transposons**. A) Size distribution of RNA reads obtained through short RNA sequencing of
035 Caco2 WT and PIWIL2 KO cells, after aligning reads to both the piRNA cluster and the repBase
036 databases. Nucleotide lengths is shown on the x axis, whereas the fraction of the respected reads
037 against the total sequencing reads are shown on the y axis. B) Nucleotide enrichment of the 5' end of
038 the short RNA reads identified in the piRNA cluster and the repBase databases. C-D) Scatter plots of
039 the antisense (meaning complementary to the respected TEs indicated in the graph) and of the sense
040 (meaning the same as the respected TEs indicated in the graph) short RNA (piRNA-like) sequence
041 reads as counts per million (CPM) from Caco2 WT (x axis) and PIWIL2 KO (y axis) cells. In red are
042 examples of specific LINE-1 element targets that are present in both the antisense and sense piRNA
043 sequences in Caco2 WT cells but absent in PIWIL2 KO cells. All sequencing results are normalized
044 average reads from n=3 biological replicates. Source data can be found in Table S2. E) qPCR of L1RE1
045 comparing PIWIL2 KO to Caco2 WT. Data were analyzed and normalized to 18S ribosomal RNA by
046 ΔΔCt calculations to calculate fold change. (n=3 biological replicates, mean ± SE, t-test *p = 0.0267)
047
048 **Figure 7. PIWIL2 depletion increases LINE1 activity.** A) Representative fluorescent images of Caco2
049 WT or PIWIL2 KO cells that were transfected with either a constitutively active GFP reporter (pCEP4+),
050 a LINE1 retro-transposition defective reporter (LINE1 inactive (-)) or a LINE1-GFP reporter plasmid that
051 expresses GFP only when LINE1 retro-transposition occurs. B) Quantification of the average % positive
052 GFP expressing cells normalized to transfection efficiency for each cell line. (n = 4 biological replicates
053 with 3 fields taken per condition and averaged, mean ± SE, two-way ANOVA with Bonferroni correction
054 for multiple comparisons, PIWIL2 KO LINE1 inactive (-) vs. PIWIL2 KO LINE1 GFP *p = 0.0212,
055 Caco2 WT LINE1 inactive (-) vs. PIWIL2 KO LINE1 GFP **p = 0.0076, Caco2 WT LINE1 GFP vs.
056 PIWIL2 KO LINE1 GFP **p = 0.0091).
057
058 **Figure 8. PIWIL2 depletion promotes DNA damage via LINE-1 transposon activation.** A) γH2AX
059 increases in PIWIL2 KO cells compared to Caco2 WT cells by immunofluorescence. Representative
060 confocal images shown B) Quantification of γH2AX foci per cell (n = 9 fields (3 biological replicates, 3
061 fields per replicate), mean ± SE, t-test *p = 0.0151) C) Representative western blot of total H2AX,
062 phosphorylated γH2AX, and β-actin (Actin) loading control in Caco2 and PIWIL2 KO cells. D) LINE-1
063 GFP reporter assay staining for eGFP and γH2AX co-localization as indicated by white arrows. E)
064 Quantification of the average % positive GFP and γH2AX expressing cells normalized to transfection

065 efficiency for each cell line. (n = 3 biological replicates with 3 fields taken per condition and averaged,
066 mean ± SE, two-way ANOVA with Bonferroni correction for multiple comparisons, Caco2 WT LINE1
067 inactive (-) vs. PIWIL2 KO LINE1 GFP *p = 0.0351.
068
069 **Figure S1. PIWIL2 is highly expressed in normal somatic tissues particularly of gastrointestinal**
070 **origin.** A) RNA-seq data of normal tissues (project PRJEB4337) for PIWIL2, derived from NCBI and
071 from the related PIWIL2 entry (Gene ID: 55124). Mean RPKMs (Reads Per Kilobase of exon per Million
072 mapped reads) and error bars are shown from n=2 (duodenum, ovary, pancreas, urinary bladder), n=3
073 (adrenal, appendix, brain, endometrium, esophagus, fat, gall bladder, liver, salivary gland, skin,
074 stomach), n=4 (bone marrow, heart, kidney, placenta, prostate, small intestine, spleen, thyroid), n=5
075 (colon, lung, lymph node) samples per indicated tissue. B) The UALCAN web-based tool was used to
076 perform analysis of TCGA RNA-seq data to determine PIWIL2 expression across multiple normal tissue
077 samples. Box plots show median log2TPM (transcripts per kilobase million), upper-lower quartile, and
078 maximum-minimum distribution. Tumor types and number of samples queried: bladder urothelial
079 carcinoma (BLCA; n=19), breast invasive carcinoma (BRCA; n=114), cervical squamous cell carcinoma
080 and endocervical adenocarcinoma (CESC; n=3), cholangiocarcinoma (CHOL; n=9), colon
081 adenocarcinoma (COAD; n=41), esophageal carcinoma (ESCA; n=11), glioblastoma multiforme (GBM;
082 n=5), head and neck squamous cell carcinoma (HNSC; n=44), kidney chromophobe (KICH; n=25),
083 kidney renal clear cell carcinoma (KIRC; n=72), kidney renal papillary cell carcinoma (KIRP; n=32),
084 liver hepatocellular carcinoma (LIHC; n=50), lung adenocarcinoma (LUAD; n=59), lung squamous cell
085 carcinoma (LUSC; n=52), pancreatic adenocarcinoma (PAAD; n=4), prostate adenocarcinoma (PRAD;
086 n=52), pheochromocytoma and paraganglioma (PCPG, n=3), rectum adenocarcinoma (READ; n=10),
087 sarcoma (SARC; n=2), skin cutaneous melanoma (SKCM; n=1), thyroid carcinoma (THCA; n=59),
088 thymoma (THYM; n=2), stomach adenocarcinoma (STAD; n=34), and uterine corpus endometrial
089 carcinoma (UCEC; n=35). Tissues showing the highest expression in both datasets are highlighted in
090 red boxes and are of gastrointestinal origin.
091
092 **Figure S2. PIWIL2 is not differentially expressed in colon adenomas compared to normal**
093 **samples.** The colon adenoma datasets used are available from the Gene Expression Omnibus (GEO):
094 GDS2947/239303_at, GDS2947/220686_s_at and GDS2947/217421_at. N=32 samples normal and
095 adenoma were included in all cases; a Welch's t-test was used to compare normal vs colon adenoma
096 samples within a dataset. The box and whisker plots show the median, interquartile range, minimum
097 and maximum values of PIWIL2 expression for each dataset, displayed in arbitrary units between
098 datasets.
099

Figure S3. PIWIL2 KO clone selection and mutation characterization. A) Representative western blots of the screening process to identify PIWIL2 knockout clones, after transfection of Caco2 cells with CRISPR/Cas9 and a gRNA targeting PIWIL2. Clone 11 (indicated by *) was selected for subsequent studies as it was the only clone to survive selection and maintenance in cell culture. B) Schematic of the genomic region including exon 14 of PIWIL2 that was targeted by the gRNA, where highlighted are the genomic alterations that we identified after DNA sequencing of clone 11: a nucleotide substitution (G to A), altering the splicing target site and a large deletion within the intron. The PCR primers used to amplify and sequence the specific region are also indicated. C) Alignment of the DNA sequencing results between the WT cells and the PIWIL2 KO cells, where we identified both an unaffected allele (allele 2), as well as the actual modified allele (allele 1); modifications are indicated in red boxes.

Figure S4. Depletion of PIWIL2 by transient shRNA knockdown exhibits slowed growth in 2-dimensions (2D). A) qPCR of PIWIL2 comparing Caco2 cells treated with a Non-Target shRNA or shRNA targeting PIWIL2. Expression was normalized to 18S ribosomal RNA and fold change was calculated by ΔΔCt calculations. (n=3 biological replicates, mean ± SE, t-test *p=0.0243). B) Representative western blot of Caco2 cells treated with a Non-Target shRNA or shRNA targeting PIWIL2. C) xCELLigence cellular impedance assay to examine the proliferation rates of Caco2 cells treated with a Non-Target shRNA or shRNA targeting PIWIL2 (n=3 biological replicates, two-way repeated measures ANOVA with a Bonferroni correction for multiple comparisons, *p=0.0247).





C

Gene	Percent Deleted	Percent Allelic Loss	Percent Normal	Percent Extra Allele	Percent Amplified
PIWIL2	5.9	38.8	37	18.3	0
EXD1	1.8	36.6	58	3.6	0
DDX4	1.1	20.5	69.3	8.8	0.2
PLD6	0.9	54	42	3.2	0
MOV10L1	0.7	35.7	60.7	2.9	0
ASZ1	0.2	0.9	49.2	49.2	0.5
PIWIL4	0.2	14.2	72.9	12.4	0.2
TDRD1	0.2	21.9	72.9	4.7	0.2
TDRD9	0.2	32.3	60.3	7.2	0
FKBP6	0	1.1	49.7	49	0.2
MAEL	0	9.7	68.8	21	0.5
PIWIL1	0	12.2	68.4	19	0.5
TDRD12	0	6.1	73.4	19.4	1.1
TDRKH	0	9.9	70	19.6	0.5

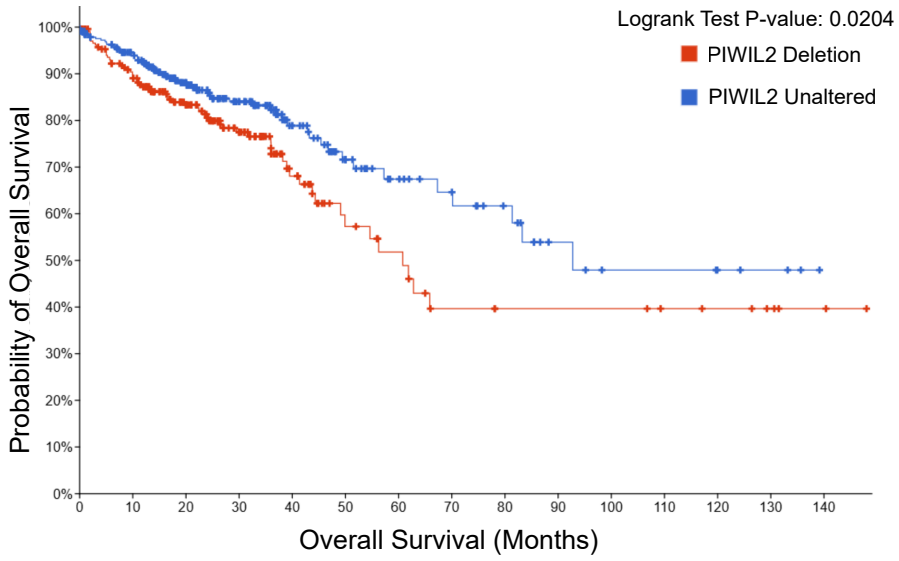


Figure 1

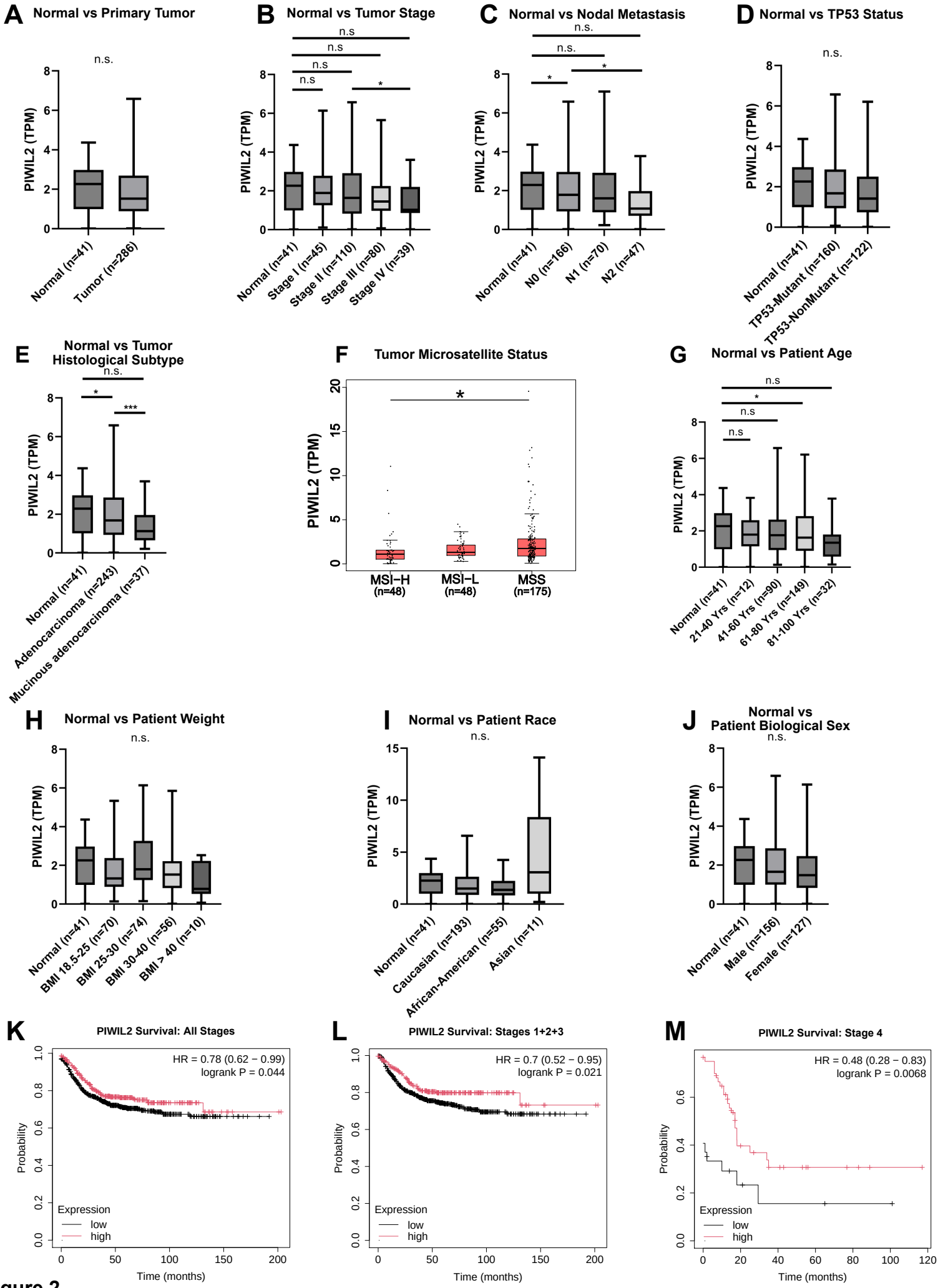

A
Normal vs Primary Tumor
n.s.
PIWIL2 (TPM)
Normal (n=41)
Tumor (n=286)
B
Normal vs Tumor Stage
n.s
n.s
n.s
n.s
*
Normal (n=41)
Stage I (n=45)
Stage II (n=110)
Stage III (n=80)
Stage IV (n=39)
C
Normal vs Nodal Metastasis
n.s.
n.s.
*
*
Normal (n=41)
N0 (n=166)
N1 (n=70)
N2 (n=47)
D
Normal vs TP53 Status
n.s.
Normal (n=41)
TP53-Mutant (n=160)
TP53-NonMutant (n=122)
E
Normal vs Tumor Histological Subtype
n.s.
*
***
Normal (n=41)
Adenocarcinoma (n=243)
Mucinous adenocarcinoma (n=37)
F
Tumor Microsatellite Status
*
MSI-H (n=48)
MSI-L (n=48)
MSS (n=175)
G
Normal vs Patient Age
n.s
*
n.s
n.s
Normal (n=41)
21-40 Yrs (n=12)
41-60 Yrs (n=90)
61-80 Yrs (n=149)
81-100 Yrs (n=32)
H
Normal vs Patient Weight
n.s.
Normal (n=41)
BMI 18.5-25 (n=70)
BMI 25-30 (n=74)
BMI 30-40 (n=56)
BMI > 40 (n=10)
I
Normal vs Patient Race
n.s.
Normal (n=41)
Caucasian (n=193)
African-American (n=55)
Asian (n=11)
J
Normal vs Patient Biological Sex
n.s.
Normal (n=41)
Male (n=156)
Female (n=127)
K
PIWIL2 Survival: All Stages
HR = 0.78 (0.62 − 0.99)
logrank P = 0.044
Probability
Expression
low
high
Time (months)
L
PIWIL2 Survival: Stages 1+2+3
HR = 0.7 (0.52 − 0.95)
logrank P = 0.021
Probability
Expression
low
high
Time (months)
M
PIWIL2 Survival: Stage 4
HR = 0.48 (0.28 − 0.83)
logrank P = 0.0068
Probability
Expression
low
high
Time (months)

Figure 2

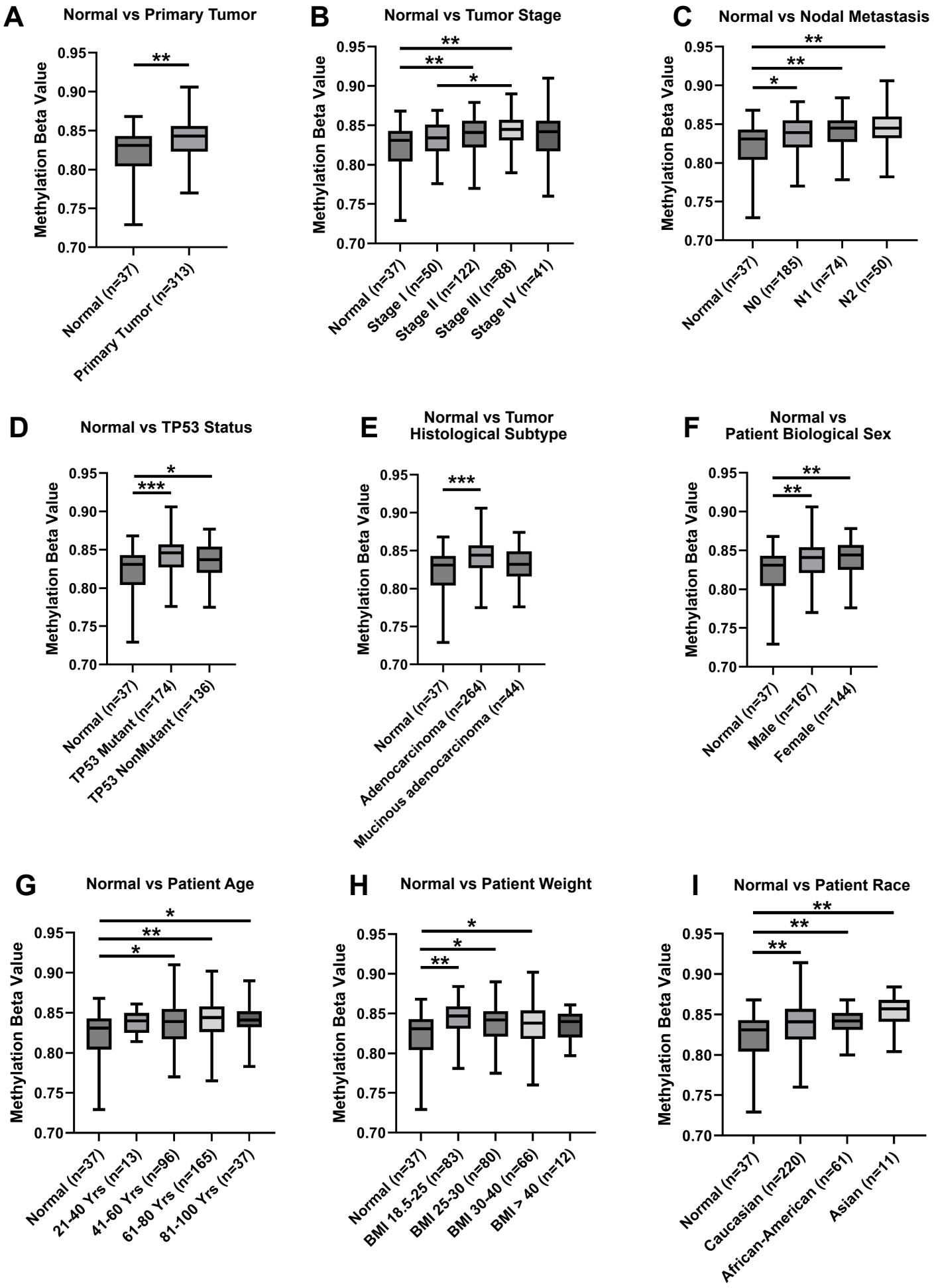

A
Normal vs Primary Tumor
Methylation Beta Value
0.95
0.90
0.85
0.80
0.75
0.70
**
Normal (n=37)
Primary Tumor (n=313)
B
Normal vs Tumor Stage
Methylation Beta Value
**
**
*
Normal (n=37)
Stage I (n=50)
Stage II (n=122)
Stage III (n=88)
Stage IV (n=41)
C
Normal vs Nodal Metastasis
Methylation Beta Value
**
**
*
Normal (n=37)
N0 (n=185)
N1 (n=74)
N2 (n=50)
D
Normal vs TP53 Status
Methylation Beta Value
*
***
Normal (n=37)
TP53 Mutant (n=174)
TP53 NonMutant (n=136)
E
Normal vs Tumor Histological Subtype
Methylation Beta Value
***
Normal (n=37)
Adenocarcinoma (n=264)
Mucinous adenocarcinoma (n=44)
F
Normal vs Patient Biological Sex
Methylation Beta Value
**
**
Normal (n=37)
Male (n=167)
Female (n=144)
G
Normal vs Patient Age
Methylation Beta Value
*
**
*
Normal (n=37)
21-40 Yrs (n=13)
41-60 Yrs (n=96)
61-80 Yrs (n=165)
81-100 Yrs (n=37)
H
Normal vs Patient Weight
Methylation Beta Value
*
*
**
Normal (n=37)
BMI 18.5-25 (n=83)
BMI 25-30 (n=80)
BMI 30-40 (n=66)
BMI > 40 (n=12)
I
Normal vs Patient Race
Methylation Beta Value
**
**
**
Normal (n=37)
Caucasian (n=220)
African-American (n=61)
Asian (n=11)

Figure 3

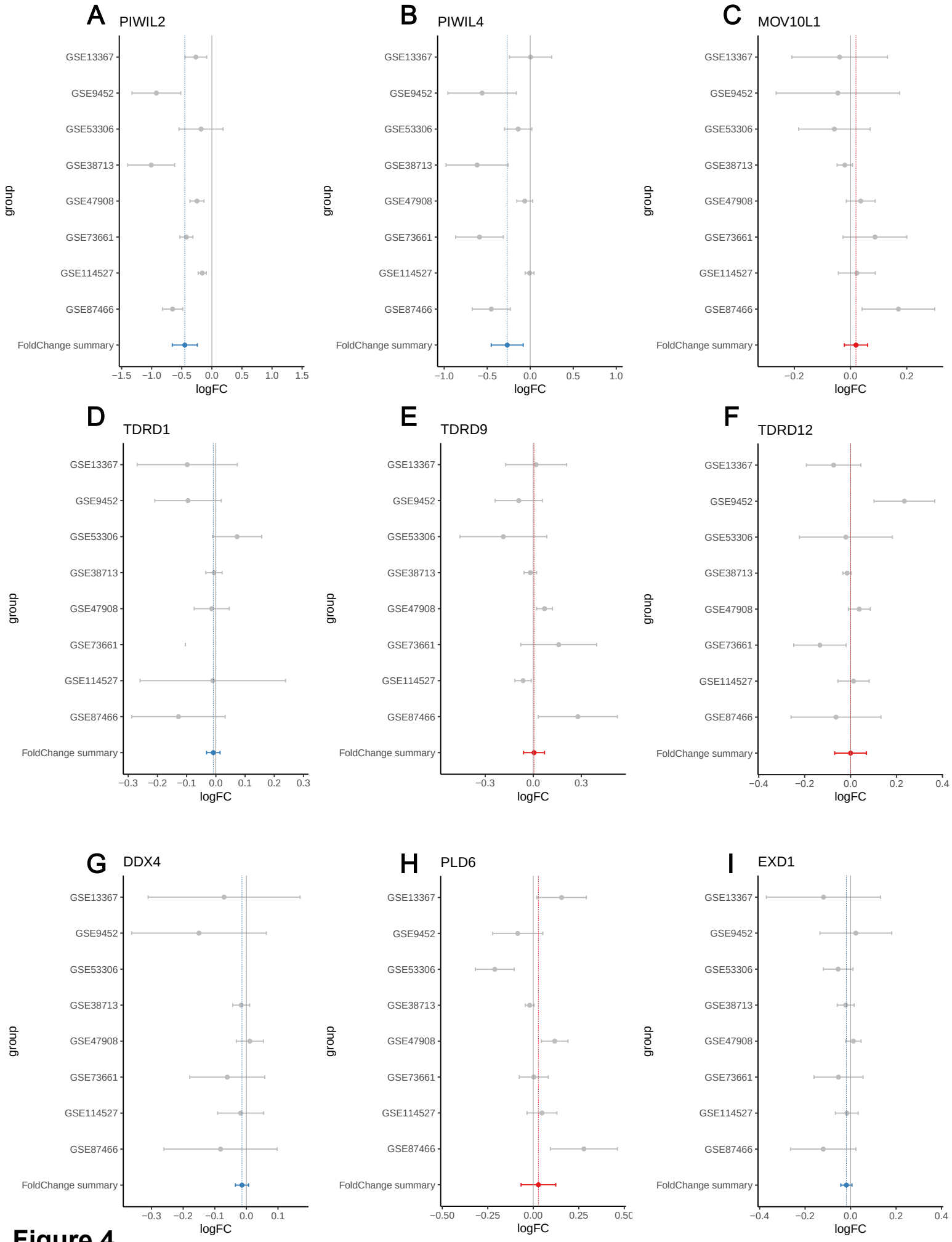

A
PIWIL2
B
PIWIL4
C
MOV10L1
D
TDRD1
E
TDRD9
F
TDRD12
G
DDX4
H
PLD6
I
EXD1
GSE13367
GSE9452
GSE53306
GSE38713
GSE47908
GSE73661
GSE114527
GSE87466
FoldChange summary
group
logFC

Figure 4

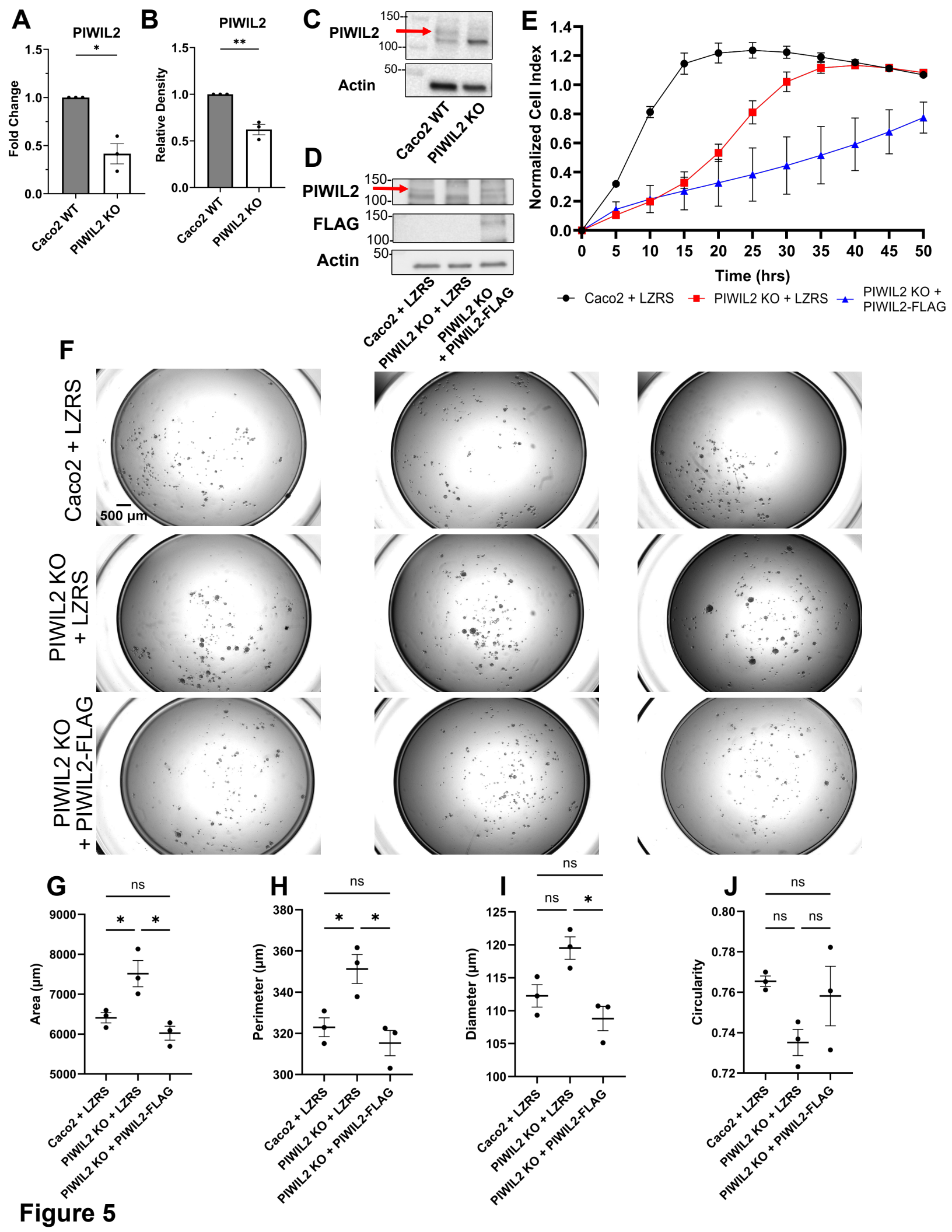

A
PIWIL2
Fold Change
Caco2 WT
PIWIL2 KO
B
PIWIL2
Relative Density
C
PIWIL2
Actin
D
PIWIL2
FLAG
Actin
Caco2 + LZRS
PIWIL2 KO + LZRS
PIWIL2 KO + PIWIL2-FLAG
E
Normalized Cell Index
Time (hrs)
F
500 µm
G
Area (µm)
H
Perimeter (µm)
I
Diameter (µm)
J
Circularity
ns

Figure 5

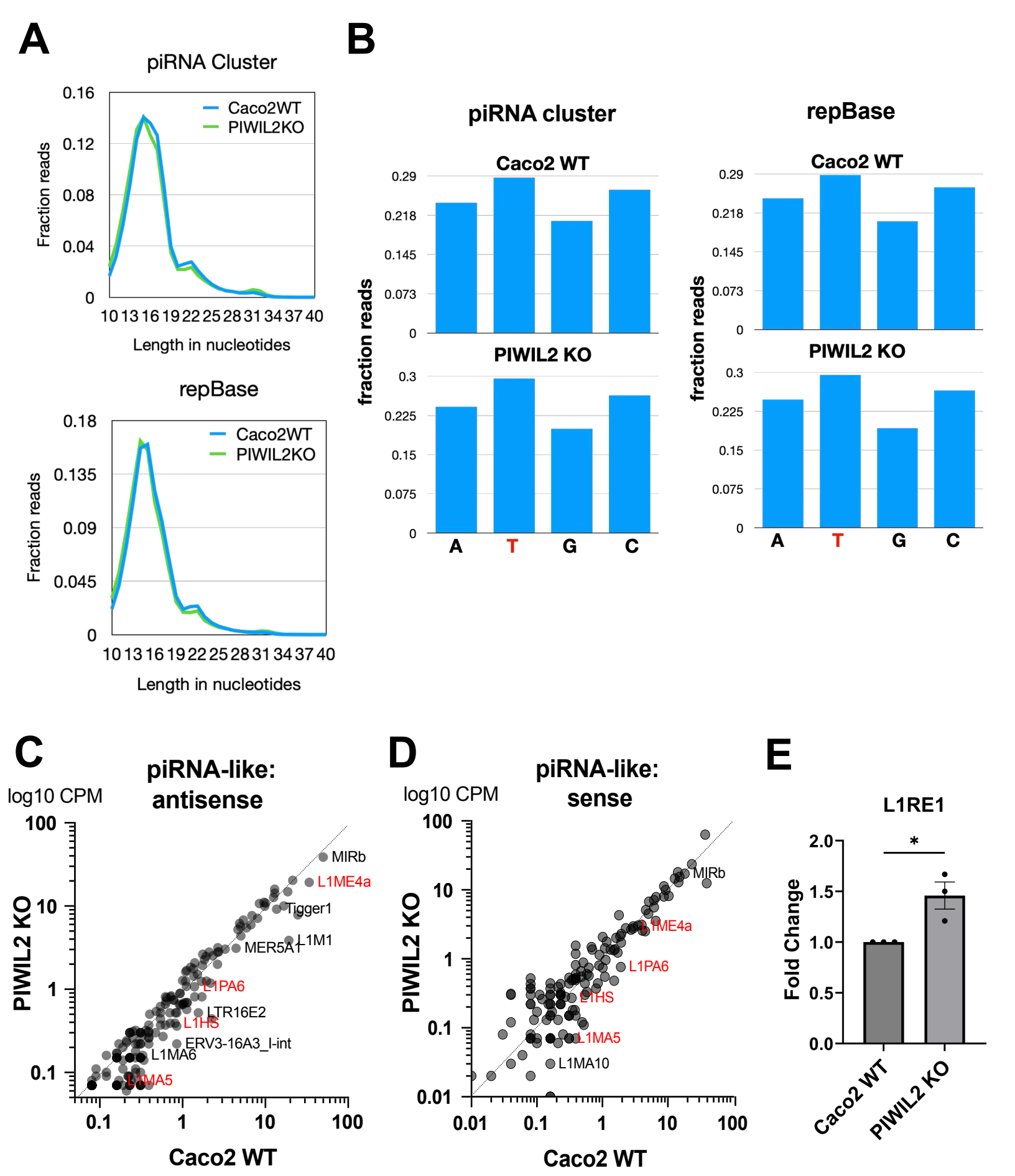

A
piRNA Cluster
Caco2WT
PIWIL2KO
Fraction reads
0.16
0.12
0.08
0.04
0
10 13 16 19 22 25 28 31 34 37 40
Length in nucleotides
repBase
Caco2WT
PIWIL2KO
Fraction reads
0.18
0.135
0.09
0.045
0
10 13 16 19 22 25 28 31 34 37 40
Length in nucleotides
B
piRNA cluster
repBase
Caco2 WT
PIWIL2 KO
fraction reads
0.29
0.218
0.145
0.073
0
0.3
0.225
0.15
0.075
A T G C
C
piRNA-like: antisense
log10 CPM
PIWIL2 KO
Caco2 WT
MIRb
L1ME4a
Tigger1
L1M1
MER5A1
L1PA6
LTR16E2
L1HS
ERV3-16A3_I-int
L1MA6
L1MA5
D
piRNA-like: sense
log10 CPM
PIWIL2 KO
Caco2 WT
MIRb
L1ME4a
L1PA6
L1HS
L1MA5
L1MA10
E
L1RE1
Fold Change
2.0
1.5
1.0
0.5
0.0
*
Caco2 WT
PIWIL2 KO

Figure 6

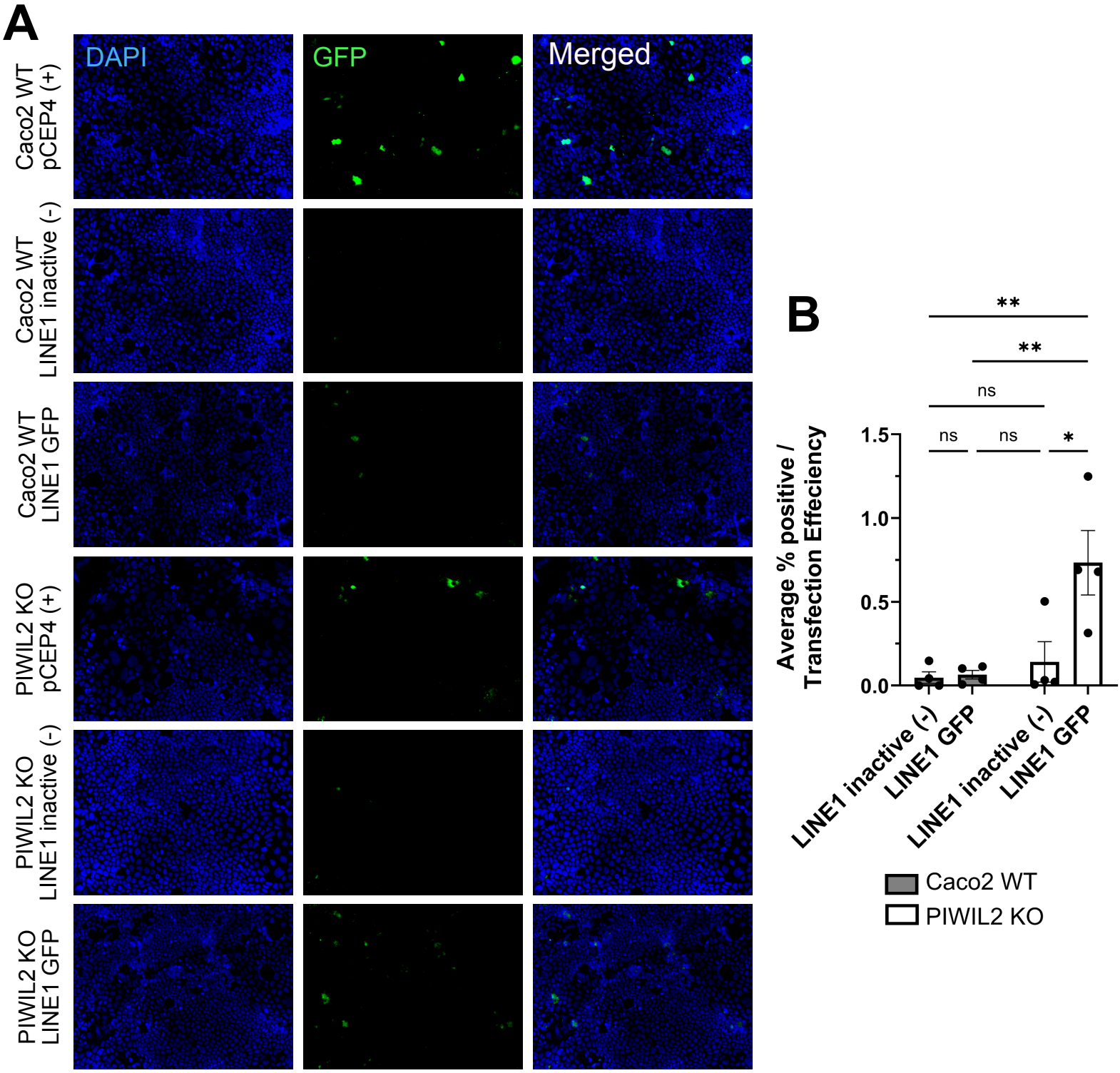

A
DAPI
GFP
Merged
Caco2 WT pCEP4 (+)
Caco2 WT LINE1 inactive (-)
Caco2 WT LINE1 GFP
PIWIL2 KO pCEP4 (+)
PIWIL2 KO LINE1 inactive (-)
PIWIL2 KO LINE1 GFP
B
Average % positive / Transfection Effeciency
1.5
1.0
0.5
0.0
**
**
ns
ns
ns
*
LINE1 inactive (-)
LINE1 GFP
LINE1 inactive (-)
LINE1 GFP
Caco2 WT
PIWIL2 KO

Figure 7

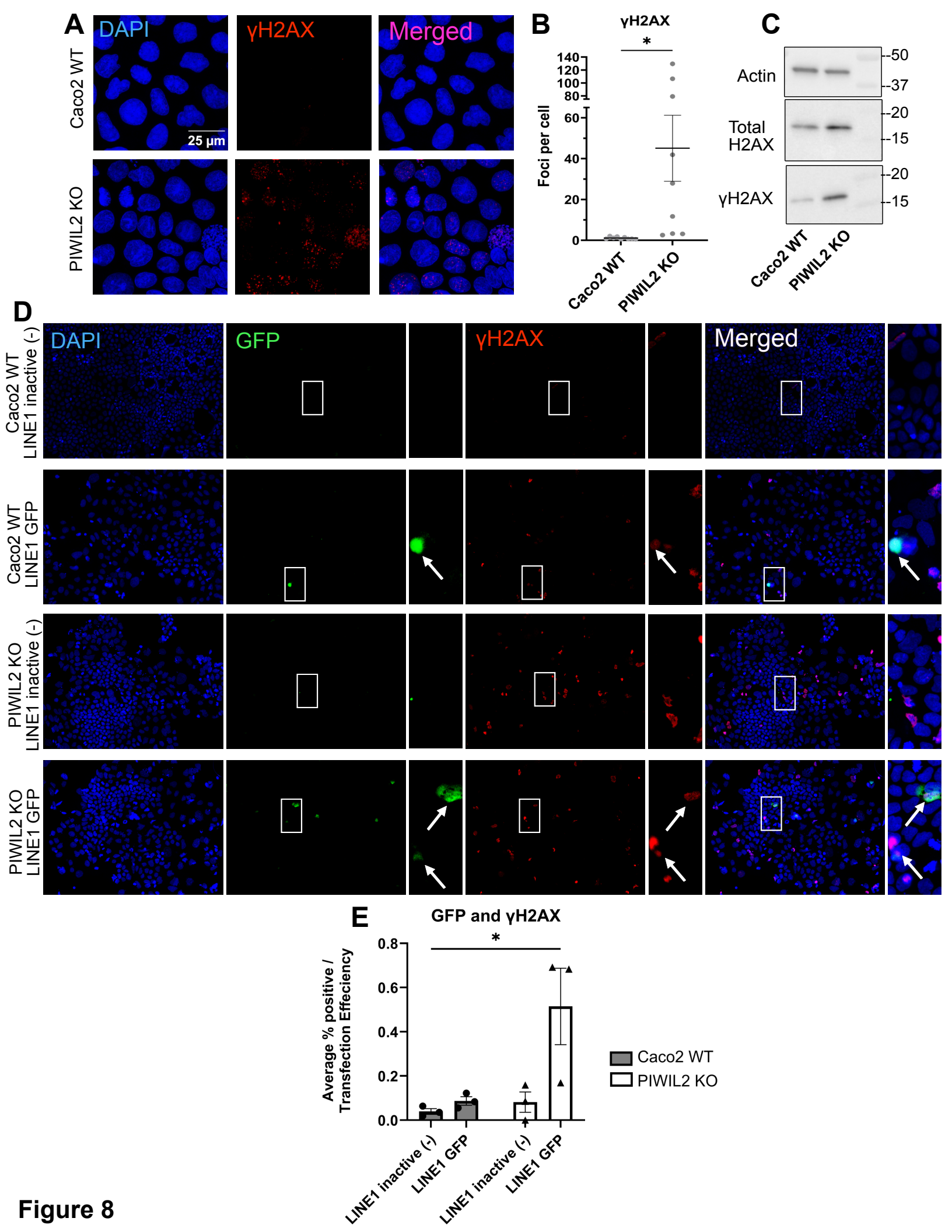
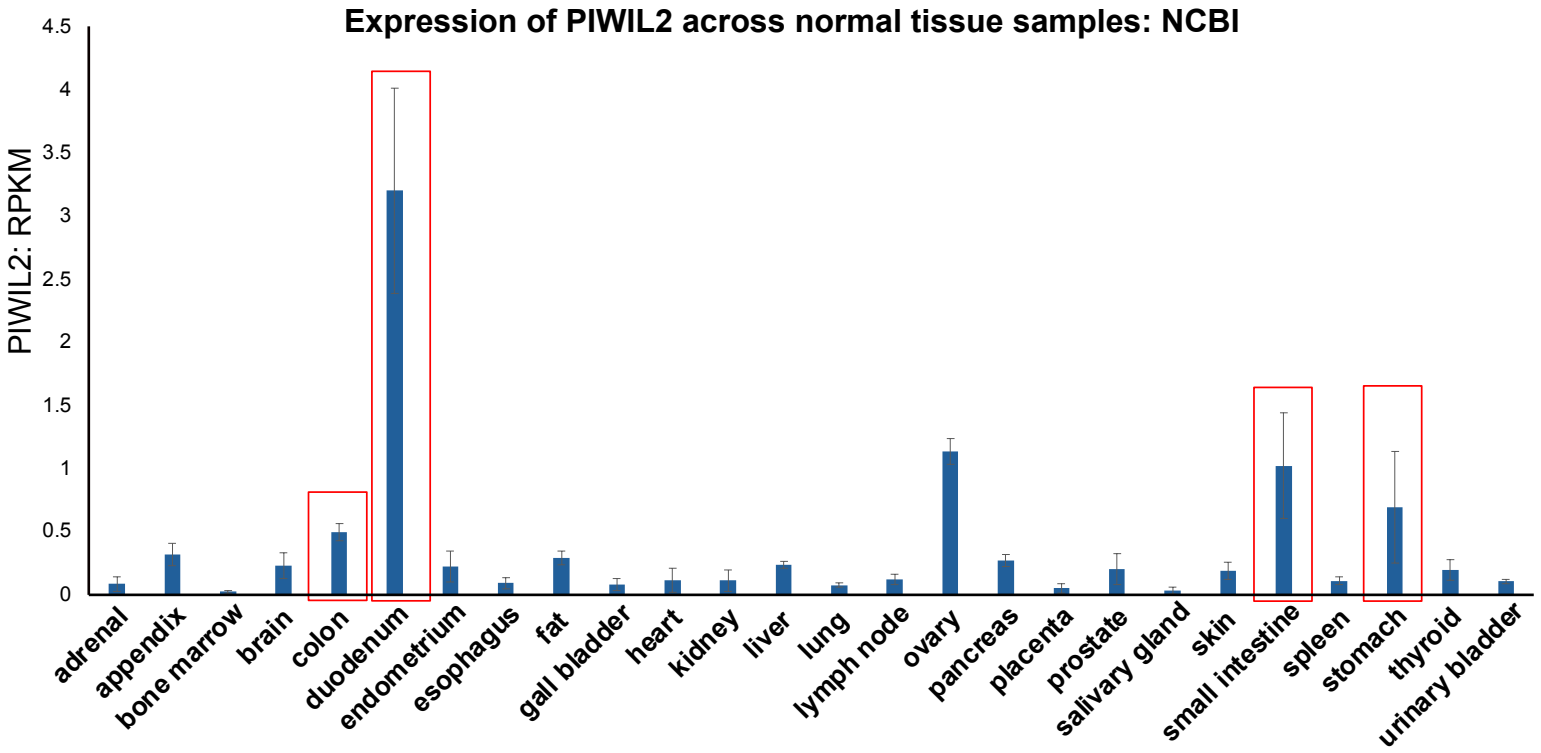

A
DAPI
γH2AX
Merged
Caco2 WT
PIWIL2 KO
25 µm
B
γH2AX
*
Foci per cell
140
120
100
80
60
40
20
0
Caco2 WT
PIWIL2 KO
C
Actin
Total H2AX
γH2AX
50
37
20
15
20
15
Caco2 WT
PIWIL2 KO
D
DAPI
GFP
γH2AX
Merged
Caco2 WT LINE1 inactive (-)
Caco2 WT LINE1 GFP
PIWIL2 KO LINE1 inactive (-)
PIWIL2 KO LINE1 GFP
E
GFP and γH2AX
*
Average % positive / Transfection Effeciency
0.8
0.6
0.4
0.2
0.0
LINE1 inactive (-)
LINE1 GFP
LINE1 inactive (-)
LINE1 GFP
Caco2 WT
PIWIL2 KO

Figure 8

A


Expression of PIWIL2 across normal tissue samples: NCBI
PIWIL2: RPKM
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
adrenal
appendix
bone marrow
brain
colon
duodenum
endometrium
esophagus
fat
gall bladder
heart
kidney
liver
lung
lymph node
ovary
pancreas
placenta
prostate
salivary gland
skin
small intestine
spleen
stomach
thyroid
urinary bladder

B

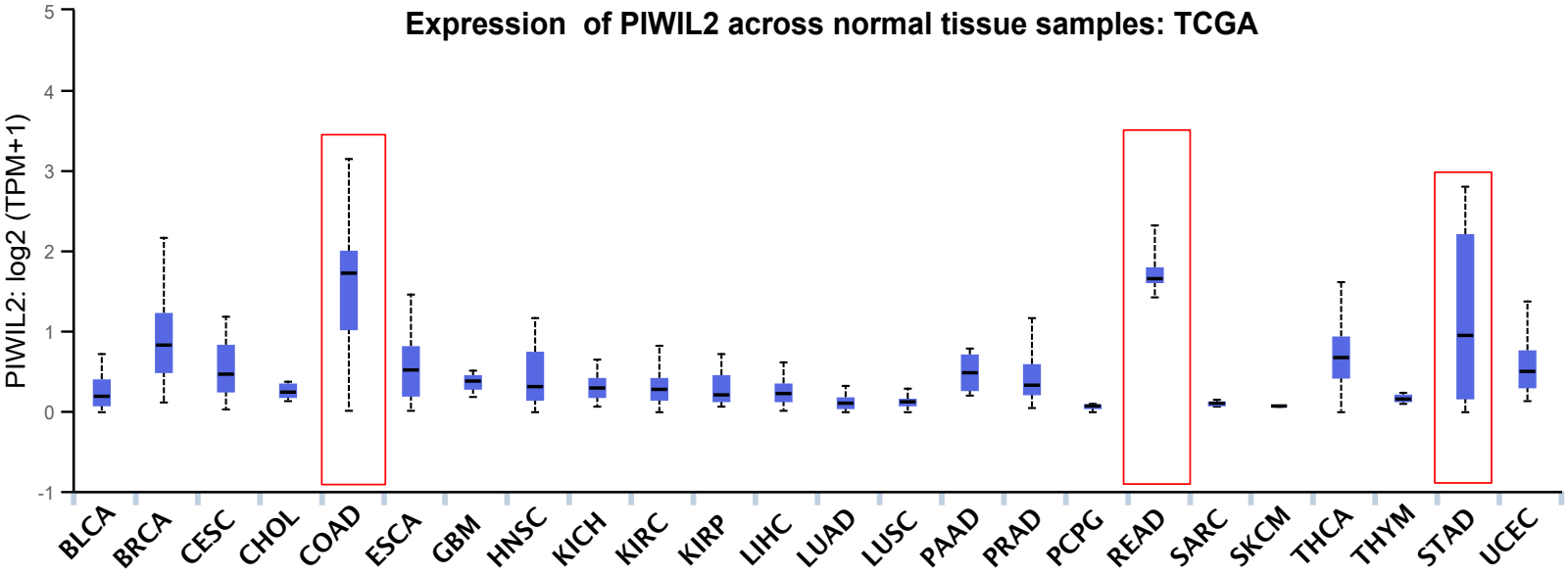

Expression of PIWIL2 across normal tissue samples: TCGA
PIWIL2: log2 (TPM+1)
5
4
3
2
1
0
-1
BLCA
BRCA
CESC
CHOL
COAD
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LIHC
LUAD
LUSC
PAAD
PRAD
PCPG
READ
SARC
SKCM
THCA
THYM
STAD
UCEC

Figure S1

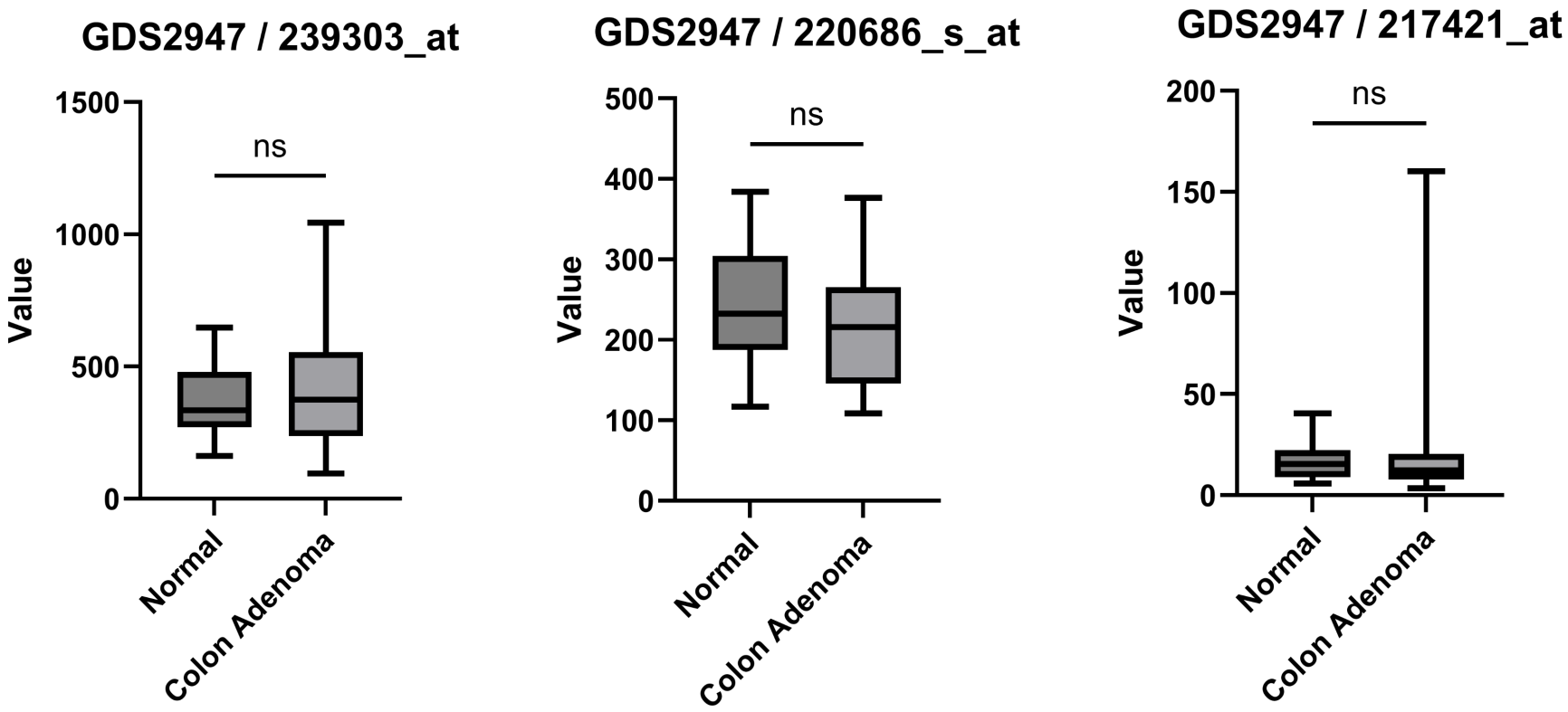

GDS2947 / 239303_at
1500
1000
500
0
Value
ns
Normal
Colon Adenoma
GDS2947 / 220686_s_at
500
400
300
200
100
0
Value
ns
Normal
Colon Adenoma
GDS2947 / 217421_at
200
150
100
50
0
Value
ns
Normal
Colon Adenoma

Figure S2

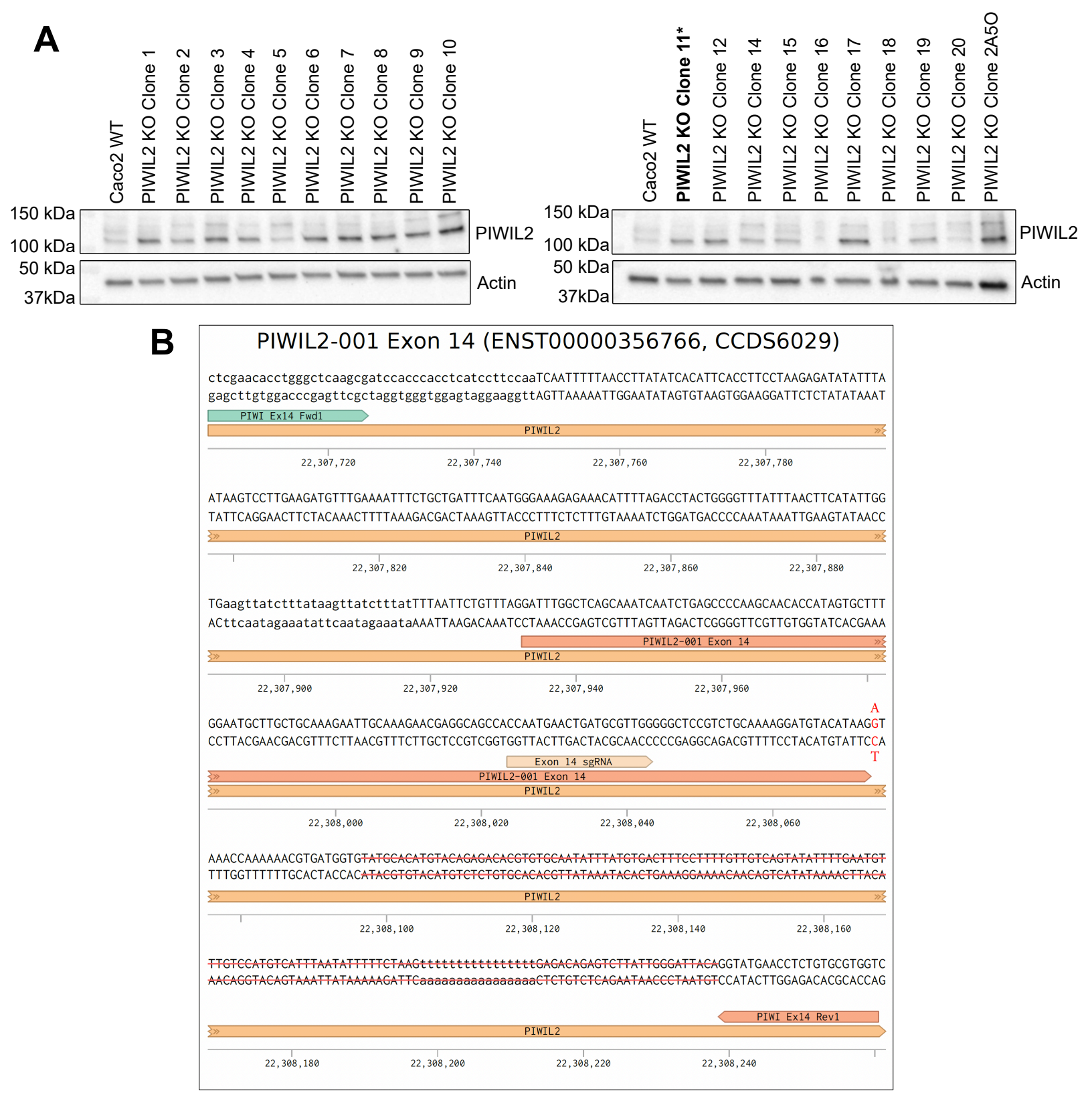

A
Caco2 WT
PIWIL2 KO Clone 1
PIWIL2 KO Clone 2
PIWIL2 KO Clone 3
PIWIL2 KO Clone 4
PIWIL2 KO Clone 5
PIWIL2 KO Clone 6
PIWIL2 KO Clone 7
PIWIL2 KO Clone 8
PIWIL2 KO Clone 9
PIWIL2 KO Clone 10
150 kDa
100 kDa
50 kDa
37kDa
PIWIL2
Actin
Caco2 WT
PIWIL2 KO Clone 11*
PIWIL2 KO Clone 12
PIWIL2 KO Clone 14
PIWIL2 KO Clone 15
PIWIL2 KO Clone 16
PIWIL2 KO Clone 17
PIWIL2 KO Clone 18
PIWIL2 KO Clone 19
PIWIL2 KO Clone 20
PIWIL2 KO Clone 2A5O
150 kDa
100 kDa
50 kDa
37kDa
PIWIL2
Actin
B
PIWIL2-001 Exon 14 (ENST00000356766, CCDS6029)
PIWI Ex14 Fwd1
PIWIL2
22,307,720
22,307,740
22,307,760
22,307,780
22,307,820
22,307,840
22,307,860
22,307,880
PIWIL2-001 Exon 14
22,307,900
22,307,920
22,307,940
22,307,960
Exon 14 sgRNA
22,308,000
22,308,020
22,308,040
22,308,060
22,308,100
22,308,120
22,308,140
22,308,160
PIWI Ex14 Rev1
22,308,180
22,308,200
22,308,220
22,308,240
C
Consensus
PIWIL2 KO allele 1
PIWIL2 KO allele 2
WT
461 bases
2 features
exon 14
gRNA

Figure S3

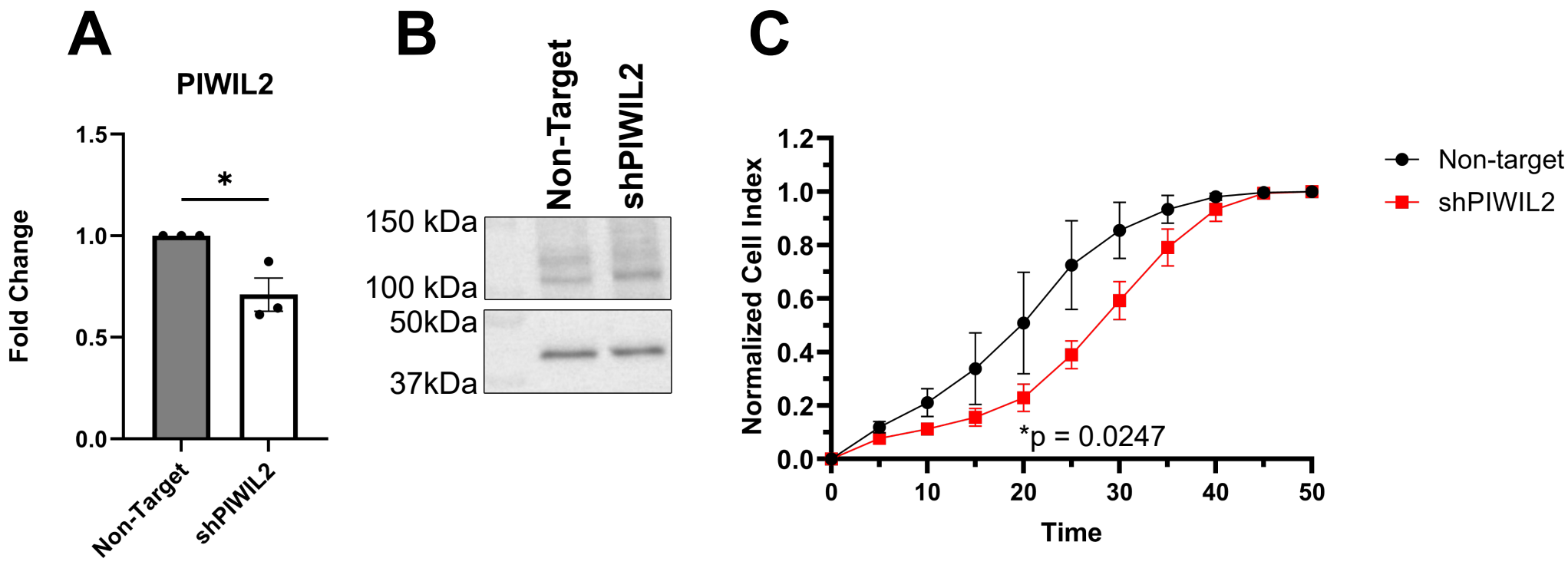

A
PIWIL2
Fold Change
1.5
1.0
0.5
0.0
*
Non-Target
shPIWIL2
B
Non-Target
shPIWIL2
150 kDa
100 kDa
50kDa
37kDa
C
Normalized Cell Index
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
Time
*p = 0.0247
Non-target
shPIWIL2

Figure S4