

Genetic variation and RNA structure regulate microRNA biogenesis

Noemi Fernandez^{1,3}, Ross A. Cordiner^{1,3}, Robert S. Young¹, Nele Hug¹, Sara Macias^{1,2} & Javier F. Cáceres¹

¹Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK.

² Present address: Institute of Infection and Immunology Research, University of Edinburgh, King's Buildings, Edinburgh, UK

Correspondence should be addressed to J.F.C. (Javier.Caceres@igmm.ed.ac.uk)

³These authors contributed equally to this work

Running title: Sequence and structural determinants of miRNA biogenesis in human cells

[Keywords: microRNAs; miR-30c-1; RNA secondary structure; Drosha; Microprocessor; SHAPE]

30

31 **MiRNA biogenesis is highly regulated at the post-transcriptional level; however,**
 32 **the role of sequence and secondary RNA structure in this process has not been**
 33 **extensively studied. A single G to A substitution present in the terminal loop of**
 34 **pri-mir-30c-1 in breast cancer patients had been previously described to result in**
 35 **increased levels of mature miRNA. Here, we report that this genetic variant**
 36 **directly affects Drosha-mediated processing of pri-mir-30c-1 *in vitro* and in**
 37 **cultured cells. Structural analysis of this variant revealed an altered RNA**
 38 **structure that facilitates the interaction with SRSF3, an SR protein family**
 39 **member that promotes pri-miRNA processing. Our results are compatible with a**
 40 **model whereby a genetic variant in pri-mir-30c-1 leads to a secondary RNA**
 41 **structure rearrangement that facilitates binding of SRSF3 resulting in increased**
 42 **levels of miR-30c. These data highlights that primary sequence determinants and**
 43 **RNA structure are key regulators of miRNA biogenesis.**

44

45 MicroRNAs (miRNAs) are short non-coding RNAs that negatively regulate the
46 expression of a large proportion of cellular mRNAs, thus affecting a multitude of
47 cellular and developmental pathways^{1,2}. The canonical miRNA biogenesis pathway
48 involves two sequential processing events catalyzed by RNase III enzymes. In the
49 nucleus, the Microprocessor complex, comprising the RNase III enzyme Drosha, the
50 double-stranded RNA-binding protein, DGCR8, and additional proteins carries out
51 the first processing event, which results in the production of precursor miRNAs (pre-
52 miRNAs)^{3,4}. These are exported to the cytoplasm, where a second processing event
53 carried out by another RNase III enzyme, Dicer, leads to the production of mature
54 miRNAs that are loaded into the RISC complex⁵.

55 Due to the central role of miRNAs in the control of gene expression, their levels
56 must be tightly controlled. As such, dysregulation of miRNA expression has been
57 shown to result in grossly aberrant gene expression and leads to human disease⁶⁻⁹. In
58 particular, the Microprocessor-mediated step of miRNA biogenesis has been shown to
59 be regulated by multiple signaling pathways, including the transforming growth
60 factor- β (TGF- β) pathway as well as the p53 response, leading to activation of subsets
61 of individual miRNAs^{10,11}. Furthermore, altered miRNA expression has been
62 associated with the progression of cancer^{12,13}, where a global downregulation of
63 miRNA expression is usually observed^{14,15}. It was recently shown that miRNA
64 biogenesis can also be regulated in a cell-density-dependent manner via the Hippo-
65 signaling pathway, and that the observed perturbation of this pathway in tumors may
66 underlie the widespread downregulation of miRNAs in cancer¹⁶. Thus, miRNA
67 production is tightly controlled at different levels during the biogenesis cascade.
68 Extensive evidence has shown that RNA-binding proteins (RNA-BPs) recognize the
69 terminal loop of miRNA precursors and influence either positively or negatively the

70 processing steps carried out by Drosha in the nucleus and/or Dicer in the cytoplasm.
71 These include the hnRNP proteins, hnRNP K and hnRNP A1, as well as the cold-shock
72 domain protein, Lin28 and the RNA helicases, p68/p72¹⁷. In the case of the
73 multifunctional RNA-binding protein hnRNP A1, we have previously shown that it
74 can act as an auxiliary factor by binding to the conserved terminal loop of pri-miR-
75 18a and promote its Microprocessor-mediated processing^{18,19}. Conversely, the same
76 protein can act as a repressor of let-7 production in differentiated cells²⁰.

77 Several studies have shown that there is a correlation between the presence of
78 polymorphisms in pri-miRNAs and the corresponding levels of mature miRNAs²¹;
79 however, a mechanistic understanding of how sequence variation and RNA structure
80 control miRNA biogenesis has not been explored in great detail. Screening of novel
81 genetic variants in human precursor miRNAs linked to breast cancer identified two
82 novel rare variants in the precursors of miR-30c and miR-17, resulting in
83 conformational changes in the predicted secondary structures and also leading to
84 altered expression of the corresponding mature miRNAs. These patients were non-
85 carriers of BRCA1 or BRCA2 mutations, suggesting the possibility that familial
86 breast cancer may be caused by variation in these miRNAs²². In particular, the single
87 G to A substitution in primary miR-30c-1 (pri-mir-30c-1) terminal loop, which was
88 also later observed in gastric cancer patients²³, results in an increase in the abundance
89 of the mature miRNA.

90 Here, we investigated the mechanism by which the pri-mir-30c-1 variant detected
91 in breast cancer patients results in an increased expression of this miRNA. We found
92 that this genetic variant directly affects the Microprocessor-mediated processing of
93 this miRNA. A combination of structural analysis with RNA chromatography coupled
94 to Mass spectrometry revealed changes in the pri-miRNA structure that lead to

differential binding of a protein factor, SRSF3, that has been previously reported to act as a miRNA biogenesis factor. These results provide a mechanism by which the pri-mir-30c-1 genetic variant results in an increased expression of the mature miR-30c. Altogether these data highlights that primary sequence as well as RNA structure have a crucial role in the post-transcriptional regulation of miRNA biogenesis.

RESULTS

The G/A substitution in pri-miR-30c-1 affects the Microprocessor-mediated processing of the primary miRNA

In order to understand the mechanism underlying miR-30c deregulation in breast and gastric cancers, we investigated how the reported G₂₇-to-A mutation observed in a Chinese population might affect miRNA biogenesis. It was previously shown that this substitution results in an increase in the abundance of the mature miRNA; however, the mechanism that leads to an increased expression is unknown^{22,23}. First, we transiently transfected MCF7 breast cancer cells with constructs encoding 380 nucleotides (nt) of primary hsa-pri-mir-30c-1 (pri-miRNA), either in a WT version or bearing the G/A variant (**Fig. 1a**). We observed that the G/A substitution resulted in increased levels of mature miR-30c (**Fig. 1b**), resembling the situation observed in patients with this mutation. Furthermore, this was not due to increased transcription of the G/A-harboring pri-miRNA, as shown by unchanged levels pri-miRNA levels (data not shown). In order to dissect the precise step of miRNA biogenesis pathway that is affected by the G/A substitution, we used an RNA version of pri-mir-30c-1 that has yet to undergo processing by the Microprocessor in the nucleus and by Dicer in the cytoplasm (pri-miRNA). As a counterpart, we transfected an RNA oligonucleotide that mimics the precursor miRNA (pre-miRNA), a sequence that arises upon

120 processing by the Microprocessor. Importantly, we observed approximately a two-
 121 three fold increase of miR-30c mature levels when transfecting the G/A sequence
 122 derived from the pri-miRNA sequence, whereas no changes were detected following
 123 transfection of the pre-miRNA sequence (**Fig. 1c**). This experiment demonstrates that
 124 the G/A substitution exclusively affects the Drosha-mediated processing of the pri-
 125 miRNA. Moreover, we could recapitulate this result in an *in vitro* reaction. We found
 126 that *in vitro* transcribed pri-mir-30c-1 was readily processed in the presence of MCF7
 127 total extracts, rendering a product of ~65 nts that corresponds to pre-mir-30c.
 128 Notably, the processing of the G/A variant was increased, when compared to the WT
 129 version, as was observed in living cells (**Fig. 1d**). The effect of the G/A variation in
 130 the processing of pri-mir-30c-1 was also recapitulated using a purified
 131 Microprocessor (FLAG-Drosha/FLAG-DGCR8 complex) and a shorter *in vitro*
 132 transcribed substrate (153nt) (**Supplementary Fig. 1**). Altogether, these
 133 complementary approaches indicate an enhanced Microprocessor-mediated
 134 processing of the G/A variant sequence and this recapitulates what was previously
 135 observed in breast and gastric cancer patients.

136

137 **The sequence of the terminal loop in pri-miR-30c-1 is highly conserved across** 138 **species**

139 Experiments described above confirmed a crucial role for the G₂₇ residue in
 140 influencing miR-30c biogenesis. We analyzed the genomic variation in the hsa-pri-
 141 mir-30c-1 sequence across vertebrates, and detected substantial evolutionary
 142 constraint across the entire locus, as indicated by positive GERP (Genomic
 143 Evolutionary Rate Profiling) scores (**Fig. 2a**). Constrained residues, which highlight
 144 regions under purifying selection, are located in the mature miRNA sequences in both

145 arms, as expected due to their effect in the regulation of gene expression.
146 Interestingly, part of the terminal loop (TL), where G₂₇ is embedded, has also a very
147 high level of constraint, which is suggestive of a role of this sequence in miRNA
148 biogenesis, as was previously described for a subset of miRNAs^{17,19,24}. In addition,
149 residues at the 5' end (nts -1 to -8; -11 to -16 and -20 to -23) and the 3' end (nts +1 to
150 +16 and +18 to +25) are also highly constrained. Indeed, several of these residues
151 were included as part of the stem in the *in silico* predicted RNA structure suggesting
152 their importance for maintaining the RNA secondary structure (**Fig. 2b**). These data
153 led us to focus our attention on these invariant sequences as potentially having a
154 crucial role in the regulation of miR-30c biogenesis.

155

156 **RNA structural analysis reveals the specific requirement of G₂₇ for maintaining** 157 **the pri-miRNA structure**

158 Next, in order to establish the importance of the G₂₇ to A substitution in RNA
159 structure, we performed structural analysis by Selective 2'-hydroxyl acylation
160 analyzed by primer extension (SHAPE)²⁵. This approach allows performing
161 quantitative RNA structural analysis at single nucleotide resolution and is mostly
162 independent of base composition. While highly reactive residues are located at single-
163 stranded regions, non-reactive nucleotides are involved in base pairs, non-Watson-
164 Crick base pairs, tertiary interactions or single stacking interactions in the C2'-endo
165 conformation²⁶. To this end, *in vitro* transcribed RNA comprising 380 nt of pri-miR-
166 30c-1 (either wild-type or the G/A variant sequence) was treated with N-methylisatoic
167 anhydride (NMIA), which reacts with the 2'-hydroxyl group of flexible nucleotides
168 (**Fig. 3a**). Gross modifications of SHAPE reactivity were observed in specific regions
169 of the G/A variant, when compared with the wild-type sequence. The resulting

170 profiles revealed a decreased SHAPE reactivity in the TL (residues 28-30), with a
 171 concomitant increase in the 5' region (-18,-16, and -15) as well as in the 3' end (nts
 172 +11, +16 and +19) (**Fig. 3a,b**). This result indicated the presence of different
 173 conformations in the pri-miRNA with the G-to-A substitution, as compared to its
 174 wild-type counterpart. In order to gain more information into the folding and tertiary
 175 structure of this pri-miRNA, we assessed the solvent accessibility of each nucleotide
 176 by hydroxyl radical cleavage footprinting, generated by reduction of hydrogen
 177 peroxide by iron (II)²⁷. Hydroxyl radicals break the accessible backbone of RNA with
 178 no sequence dependence. We defined buried regions, as zones with more than two
 179 consecutives nucleotides having a reactivity (R) smaller than the mean of all
 180 reactivity, whereas exposed regions are those with more than two consecutive
 181 nucleotides having R larger than the mean of all reactivity. We observed that the wild-
 182 type sequence presents two buried regions located between nts 8 to 40 and +9 to +25,
 183 as well as two exposed segments between nt -25 to 7 and nts 40 to +8
 184 (**Supplementary Fig. 2a**). The G to A substitution caused changes in the exposure to
 185 solvent, with both the TL (nts 28 to 40) and also the 3' end region (nts +17 to +22)
 186 becoming solvent accessible. By contrast, the 5' end (nts -7 to 7) and a small region in
 187 miRNA-3p (nts 55 to 58) are no longer solvent accessible.

188 Altogether, the SHAPE and radical hydroxyl data suggest that the G₂₇A
 189 substitution is indeed affecting the RNA flexibility of pri-miRNA-30c-1, modifying
 190 both base-pairing interaction as well as solvent accessibility of the nucleotides located
 191 in the TL and in the basal region of the stem. This could be a consequence of a long
 192 distance interaction disruption between those regions (**Fig. 3c** and **Supplementary**
 193 **Fig. 2b**), which could in turn modify the interaction with RNA-binding proteins
 194 important for miR-30c biogenesis.

195

196 **SRSF3 binds to a basal region of hsa-pri-mir-30c-1**

197 A working model that emerges from data described above is that either a repressor of
198 Microprocessor-mediated processing binds to the wild-type sequence or, alternatively,
199 the change in RNA structure induced by the G/A sequence variation could lead to the
200 binding of an activator. In order to identify RNA-BPs that differentially bind to either
201 the pri-miR-30c-1 WT or G/A sequence, we performed RNase-assisted
202 chromatography followed by mass spectrometry in MCF7 total cell extracts²⁸. This
203 resulted in the identification of twelve proteins that interact with the WT sequence
204 and eight that bind to the G/A variant, being 7 common between both substrates.
205 Significantly, several of the common proteins were previously implicated in miRNA
206 biogenesis and/or regulation, including the heat shock cognate 70 protein⁵, the hnRNP
207 proteins, hnRNP A1^{18,20} and hnRNP A2/B1²⁹, as well as the RNA helicase DDX17⁴,
208 and Poly [ADP-ribose] polymerase 1 (PARP)³⁰ (**Fig. 4a** and **Supplementary Fig. 3a**).

209 We decided to focus on those interactors that showed preferential interaction with
210 either the WT or the G/A variant sequence. In this category, the most likely candidate
211 that interacted with the WT sequence is the RNA-binding protein FUS/TLS (fused in
212 sarcoma/translocated in liposarcoma). Interestingly, this protein has been shown to
213 promote miRNA biogenesis by facilitating the co-transcriptional recruitment of
214 Drosha³¹; however, we could not validate the specific interaction between FUS and
215 miR-30c-1 WT sequence by Immunoprecipitation followed by Western blot analysis
216 (IP-WB) (**Supplementary Fig. 3b**). As for the G/A variant, there was a single
217 exclusive interactor, SRSF3, which is a member of the SR family of splicing
218 regulators. These family of proteins are involved in constitutive and alternative
219 splicing, but some of them have been shown to fulfil other cellular functions^{32,33}.

220 Importantly, SRSF3 has also been reported to be required for miRNA biogenesis³⁴.
 221 We could validate this interaction by RNA chromatography followed by Western-blot
 222 analysis with an antibody specific for SRSF3 (**Fig. 4b**). We also observed preferential
 223 binding of endogenous SRSF3 protein to the G/A variant of pri-mir-30c-1, as shown
 224 by immunoprecipitation of SRSF3 followed by RT-qPCR quantification of the
 225 associated pri-miRNA (**Fig. 4c**). In order to analyze the interaction of SRSF3 with
 226 pri-miR-30c-1 wild-type and G/A sequences, we carried out toeprint and SHAPE
 227 assays using SRSF3 protein purified from MCF7 cells. Toeprint analysis was
 228 performed with fluorescent labeled antisense primers and capillary electrophoresis³⁵.
 229 In this assay, bound SRSF3 will block the reverse transcriptase and will illuminate the
 230 site where SRSF3 is bound to RNA. Prominent toeprint of SRSF3 with the G/A
 231 sequence was observed around nt A₊₂₅-G₊₂₄ (RNA size 170nt, position A₊₂₅)(**Fig. 4d**).
 232 Similarly, analysis of SHAPE reactivity in the presence of added purified SRSF3,
 233 revealed a dose-dependent protection from NMIA attack upon addition of SRSF3 in
 234 specific RNA residues in a dose-dependent manner (nts -19, -18, +16,+18,+19) of the
 235 basal region of G/A (**Fig. 4f,e**). Of importance, a conserved CNNC motif (nts from
 236 +16 to +22), previously described as SRSF3 binding site^{34,36,37} is located within the
 237 recognition place. Together, we can conclude that the interaction of SRSF3 with pri-
 238 miR-30c-1 takes place at the CNNC motif at the basal region of the G/A variant.

239

240 **SRSF3 protein is responsible for the increased levels of the G/A variant of miR-** 241 **30c**

242 As previously described, SRSF3 was proposed to have a role in miRNA biogenesis by
 243 recognition of a CNNC motif located 17nts away from Drosha cleavage site³⁴. Pri-
 244 mir-30c-1 has two overlapping CNNC motifs (residues from +16 to +21). Notably,

accessibility around this region increased in the G/A variant, as determined by SHAPE and hydroxyl radical analysis (**Fig. 3** and **Supplementary Fig. 2**). Furthermore, toeprint and SHAPE assays in the presence of purified SRSF3 protein confirmed the specific recognition of the CNNC motif by SRSF3 in a dose-dependent manner (**Fig. 4**). Next, we addressed whether the preferential binding of SRSF3 to the pri-mir-30c G/A variant sequence was responsible for its increased expression by comparing the mature levels of miR-30c-1 wild-type or G/A variant under variable levels of SRSF3 expression. To this end, we co-transfected pri-mir-30c-1 constructs in MCF7 cells under transiently overexpression of SRSF3, or alternatively, transfected specific siRNAs to knock-down endogenous SRSF3 protein (**Supplementary Fig. 4**). Of interest, we observed that reduced levels of SRSF3 drastically decreased the levels of the G/A miR-30c variant, without affecting the levels of wild-type miR-30c (**Fig. 5a**, compare WT vs G/A panels). By contrast, transient overexpression of SRSF3 increased significantly the levels of wild-type mature miR-30c, but has a more modest effect on the G/A variant sequence. Altogether, these experiments suggest that SRSF3 binding is limiting in the WT scenario and that is essential to promote miRNA biogenesis in the G/A context.

To confirm the role of SRSF3 in the differential processing observed with pri-mir-30c-1 G/A variant sequence, we proceeded to mutate the two consecutive CNNC motifs that are the natural binding sites for SRSF3 (**Fig. 5b**). We generated a set of mutants that affected either the first or second CNNC motif (mut1 and mut2, respectively) or a deletion of both motifs (Δ CNNC). The CNNCmut1 carrying a double substitution C₊₁₅U₊₁₇ to AA, led to a severe reduction in the levels of miR-30c expression only with the G/A variant sequence (**Fig. 5b**). Similarly, a double substitution of the second CNNC motif G₊₂₀C₊₂₁ to AA (CNNCmut2) behaved

270 similarly, exclusively affecting the G/A variant. This experiment strongly suggests
271 that the binding of SRSF3 is an important determinant of miR-30c expression.
272 Finally, we could recapitulate the observation that SRSF3 binding is limiting for the
273 processing of wild-type pri-mir-30c-1 in an *in vitro* system, supplemented with
274 purified SRSF3 protein (Fig. 6). Firstly, we found that the FLAG-Drosha/FLAG-
275 DGCR8 complexes used for the *in vitro* processing assays contained residual levels of
276 SRSF3 protein (**Supplementary Fig. 5**). Thus, the relative higher processing of the
277 G/A variant can be explained by the preferential binding of SRSF3 present in the
278 reaction to the G/A variant. Importantly, addition of purified SRSF3 protein increases
279 the Microprocessor-mediated production of WT pre-mir-30c-1 (**Fig. 6a**), whereas
280 addition of purified SRSF3 to the G/A variant (**Fig. 6c**) or to Δ CNNC variants that
281 lack SRSF3 binding sites did not affect the processing activity (**Fig. 6b, d**). This is
282 reminiscent of what was observed in MCF7 cells under variable levels of SRSF3.

283

284 **DISCUSSION**

285 The central role of miRNAs in the regulation of gene expression requires that their
286 expression is tightly controlled. Indeed, the biogenesis of cancer-related miRNAs,
287 including those with a role as oncogenes ('oncomiRs), or those with tumor suppressor
288 functions is often dysregulated in cancer^{13,15,38}. Interestingly, some miRNAs have
289 been shown to display both tumor suppressor and also oncogenic roles, depending on
290 the cell type and the mRNA targets³⁹, as was described for miR-221, which exerts
291 oncogenic properties in the liver⁴⁰ but also acts as a tumor suppressor in erythroblastic
292 leukemias⁴¹. Furthermore, miR-375 has been shown to display a dual role in prostate
293 cancer progression, highlighting the importance of the cellular context on miRNA
294 function⁴².

295 Despite a more comprehensive knowledge on the role of RNA-BPs in the post-
 296 transcriptional regulation of miRNA production, there is only circumstantial evidence
 297 on how RNA sequence variation and RNA structure impact on miRNA processing.
 298 There are several reports showing that a single nucleotide substitution in the sequence
 299 of precursor miRNAs could have a profound effect in their biogenesis. Nonetheless,
 300 there is limited information about SNPs in the terminal loop region (TL) of pri-
 301 miRNAs. A bioinformatic approach led to the identification of 32 such SNPs in 21
 302 miRNA loop regions of human miRNAs⁴³. Some studies have found a correlation
 303 between the presence of polymorphisms in pri-miRNAs and expression levels of their
 304 corresponding mature miRNAs, affecting cancer susceptibility, as shown for miR-
 305 15/16 in chronic lymphocytic leukemia (CLL)^{44,45}, miR-146 in papillary thyroid
 306 carcinoma⁴⁶ and miR196a2 in lung cancer⁴⁷. Another example is the finding of a rare
 307 SNP in pre-miR-34a, which is associated with increased levels of mature miR-34a.
 308 This could be of biological significance since precise control of miR-34a expression is
 309 needed to maintain correct beta-cell function, thus this could affect type 2 diabetes
 310 susceptibility⁴⁸. The emerging picture is that human genetic variation could indeed not
 311 only have a role in miRNA function by affecting miRNA seed sequences and/or
 312 miRNA binding sites in the 3'UTRs of target genes, but it can also contribute
 313 significantly to modulation of miRNA biogenesis²¹.

314 In this study, we focused on a rare genetic variation found in the conserved
 315 terminal loop of pri-mir-30c-1 (G₂₇ to A variant) that was found in breast cancer and
 316 gastric cancer patients and leading to increased expression of miR-30c^{22,23}. There is
 317 circumstantial evidence that miR-30c is involved in many malignancies acting as
 318 tumour suppressor^{49,50} or as an oncogene⁵¹⁻⁵³. In order to understand the mechanism
 319 underlying miR-30c deregulation in breast cancer, we investigated how this mutation

affects miRNA biogenesis. We show that the G-A substitution in pri-mir-30c-1 directly affects Drosha-mediated processing both *in vitro* as well as in cultured cells (**Fig. 1, 5, 6 and Supplementary Fig. 1**). The conservation of pri-mir-30c-1 sequences across vertebrate species highlights the importance of the primary sequence in the TL, 5' and 3' regions (**Fig. 2**), suggesting a crucial role in miRNA biogenesis. Indeed, conserved sequences in TL have been shown to be important for recognition by auxiliary factors²⁴ as well as for DGCR8 binding⁵⁴, allowing efficient and accurate miRNA processing. It has also been shown that pri-miRNA tertiary structure is a major player in the regulation of miRNA biogenesis, as observed for the well characterized miR17-92 cluster⁵⁵⁻⁵⁷. Here, using SHAPE structural analysis, in conjunction with solvent accessibility analysis by hydroxyl cleavage, we found that the G/A sequence variation leads to a structural rearrangement of the apical region of the pri-miRNA affecting the conserved residues placed at the basal part of the stem (**Fig. 3 and Supplementary Fig. 2**). This demonstrates that pri-mir-30c-1 is organized as a complex and flexible structure, with the TL and the basal region of the stem potentially involved in a tertiary interaction. Further work is required to determine the existence of direct contacts between these regions.

Interestingly, we also observed that this RNA structure reorganization promotes the interaction with SRSF3, an SR protein family member that was demonstrated to facilitate pri-miRNA recognition and processing³⁴, by recognizing the CNNC motif located 17nts away from Drosha cleavage site. Pri-mir-30c-1 has two overlapping CNNC motifs (residues from +16 to +21) (**Fig. 5b**). Notably, accessibility around this region increased in G/A variant (**Fig. 3**). Furthermore, toeprint and SHAPE assays in the presence of purified SRSF3 protein clearly demonstrated that SRSF3 is specifically recognizing the CNNC motif in a dose-dependent manner (**Fig. 4**).

345 Altogether, data presented here suggest that binding of SRSF3 to the wild-type
346 sequence is limiting and that the structural reorganization induced by the G/A
347 substitution makes the SRSF3 binding sites more accessible. Taking everything into
348 account we propose a model whereby a genetic variant in a conserved region within
349 the TL of pri-mir-30c-1 causes a reorganization of the RNA secondary structure
350 promoting the interaction with SRSF3, which in turn enhances the Microprocessor-
351 mediated processing of pri-mir-30c-1 leading to increased levels of miR-30c (Fig. 7).
352 We conclude that primary sequence determinants and RNA structure are key
353 regulators of miRNA biogenesis.

354

355 **METHODS**

356 Methods and any associated references are available in the online version of the paper

357

358

359 **ACKNOWLEDGEMENTS**

360

361 We are grateful to David Fitzpatrick and Magdalena Maslon (MRC HGU) for
362 discussions and to Javier Martinez (IMBA, Vienna) and Encarnación Martínez-Salas
363 (CBM, Madrid) for critical reading of the manuscript. This work was supported by
364 Core funding from the Medical Research Council and by the Wellcome Trust (Grant
365 095518/Z/11/Z).

366

367 **AUTHOR CONTRIBUTIONS**

368

369 N.F., R.A.C, S.M. and J.F.C. conceived, designed, interpreted the experiments and
370 wrote the manuscript. N.F. performed most of the experiments and data analysis.
371 R.A.C. carried out *in vitro* pri-miRNA processing and N.H. performed biochemical

372 experiments. S.M. carried out some of the initial experiments. R.S.Y. provided
373 bioinformatics analysis. J.F.C. supervised the whole project.

374

375 **COMPETING INTERESTS STATEMENT**

376 The authors declare that they have no competing financial interests.

377

378 **SUPPLEMENTAL INFORMATION**

379 Supplemental information includes 5 figures, Supplemental Experimental Procedures,
380 and Supplemental References and can be found with this article online

381

382

383 **REFERENCES**

384

- 385 1. Ebert, M. S. & Sharp, P. A. Roles for microRNAs in conferring robustness to
386 biological processes. *Cell* **149**, 515–24 (2012).
- 387 2. Shenoy, A. & Blalock, R. H. Regulation of microRNA function in somatic
388 stem cell proliferation and differentiation. *Nat. Rev. Mol. Cell Biol.* **15**, 565–76
389 (2014).
- 390 3. Denli, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F. & Hannon, G. J.
391 Processing of primary microRNAs by the Microprocessor complex. *Nature*
392 **432**, 231–5 (2004).
- 393 4. Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of
394 microRNAs. *Nature* **432**, 235–40 (2004).
- 395 5. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell*
396 *Biol.* **15**, 509–24 (2014).
- 397 6. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA
398 biogenesis, function and decay. *Nat. Rev. Genet.* **11**, 597–610 (2010).
- 399 7. Winter, J., Jung, S., Keller, S., Gregory, R. I. & Diederichs, S. Many roads to
400 maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.*
401 **11**, 228–34 (2009).
- 402 8. Mendell, J. T. & Olson, E. N. MicroRNAs in stress signaling and human
403 disease. *Cell* **148**, 1172–87 (2012).
- 404 9. Finnegan, E. F. & Pasquinelli, A. E. MicroRNA biogenesis: regulating the
405 regulators. *Crit. Rev. Biochem. Mol. Biol.* **48**, 51–68
- 406 10. Davis, B. N., Hilyard, A. C., Lagna, G. & Hata, A. SMAD proteins control
407 DROSHA-mediated microRNA maturation. *Nature* **454**, 56–61 (2008).

- 408 11. Suzuki, H. I. *et al.* Modulation of microRNA processing by p53. *Nature* **460**,
409 529–33 (2009).
- 410 12. Garzon, R., Calin, G. A. & Croce, C. M. MicroRNAs in Cancer. *Annu. Rev.*
411 *Med.* **60**, 167–79 (2009).
- 412 13. Hata, A. & Lieberman, J. Dysregulation of microRNA biogenesis and gene
413 silencing in cancer. *Sci. Signal.* **8**, re3 (2015).
- 414 14. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**,
415 834–8 (2005).
- 416 15. Lin, S. & Gregory, R. I. MicroRNA biogenesis pathways in cancer. *Nat. Rev.*
417 *Cancer* **15**, 321–333 (2015).
- 418 16. Mori, M. *et al.* Hippo Signaling Regulates Microprocessor and Links Cell-
419 Density-Dependent miRNA Biogenesis to Cancer. *Cell* **156**, 893–906 (2014).
- 420 17. Castilla-Llorente, V., Nicastro, G. & Ramos, A. Terminal loop-mediated
421 regulation of miRNA biogenesis: selectivity and mechanisms. *Biochem. Soc.*
422 *Trans.* **41**, 861–5 (2013).
- 423 18. Guil, S. & Cáceres, J. F. The multifunctional RNA-binding protein hnRNP A1
424 is required for processing of miR-18a. *Nat. Struct. Mol. Biol.* **14**, 591–6 (2007).
- 425 19. Michlewski, G., Guil, S., Semple, C. A. & Cáceres, J. F. Posttranscriptional
426 regulation of miRNAs harboring conserved terminal loops. *Mol. Cell* **32**, 383–
427 93 (2008).
- 428 20. Michlewski, G. & Cáceres, J. F. Antagonistic role of hnRNP A1 and KSRP in
429 the regulation of let-7a biogenesis. *Nat. Struct. Mol. Biol.* **17**, 1011–8 (2010).
- 430 21. Hogg, D. R. & Harries, L. W. Human genetic variation and its effect on
431 miRNA biogenesis, activity and function. *Biochem. Soc. Trans.* **42**, 1184–9
432 (2014).
- 433 22. Shen, J., Ambrosone, C. B. & Zhao, H. Novel genetic variants in microRNA
434 genes and familial breast cancer. *Int. J. Cancer* **124**, 1178–82 (2009).
- 435 23. Mu, Y. & Su, X. Polymorphism in pre-miR-30c contributes to gastric cancer
436 risk in a Chinese population. *Med. Oncol.* **29**, 1723–32 (2012).
- 437 24. Choudhury, N. R. R. & Michlewski, G. Terminal loop-mediated control of
438 microRNA biogenesis. *Biochem. Soc. Trans.* **40**, 789–93 (2012).
- 439 25. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2'-hydroxyl
440 acylation analyzed by primer extension (SHAPE): quantitative RNA structure
441 analysis at single nucleotide resolution. *Nat. Protoc.* **1**, 1610–6 (2006).
- 442 26. Mortimer, S. a & Weeks, K. M. C2'-endo nucleotides as molecular timers
443 suggested by the folding of an RNA domain. *Proc. Natl. Acad. Sci. U. S. A.*
444 **106**, 15622–7 (2009).
- 445 27. Tullius, T. D. & Greenbaum, J. A. Mapping nucleic acid structure by hydroxyl
446 radical cleavage. *Curr. Opin. Chem. Biol.* **9**, 127–34 (2005).
- 447 28. Michlewski, G. & Cáceres, J. F. RNase-assisted RNA chromatography. *RNA*
448 **16**, 1673–8 (2010).

- 449 29. Villarroya-Beltri, C. *et al.* Sumoylated hnRNP A2B1 controls the sorting of
450 miRNAs into exosomes through binding to specific motifs. *Nat. Commun.* **4**,
451 2980 (2013).
- 452 30. Leung, A. K. L. *et al.* Poly(ADP-ribose) regulates stress responses and
453 microRNA activity in the cytoplasm. *Mol. Cell* **42**, 489–99 (2011).
- 454 31. Morlando, M. *et al.* FUS stimulates microRNA biogenesis by facilitating co-
455 transcriptional Drosha recruitment. *EMBO J.* **31**, 4502–10 (2012).
- 456 32. Long, J. C. & Cáceres, J. F. The SR protein family of splicing factors: master
457 regulators of gene expression. *Biochem. J.* **417**, 15–27 (2009).
- 458
- 459 33. Howard, J. M. & Sanford, J. R. The RNAissance family: SR proteins as
460 multifaceted regulators of gene expression. *Wiley Interdiscip. Rev. RNA* (2014).
461 doi:10.1002/wrna.1260
- 462 34. Auyeung, V. C., Ulitsky, I., McGeary, S. E. & Bartel, D. P. Beyond secondary
463 structure: primary-sequence determinants license pri-miRNA hairpins for
464 processing. *Cell* **152**, 844–58 (2013).
- 465 35. Gould, P. S., Bird, H. & Easton, A. J. Translation toeprinting assays using
466 fluorescently labeled primers and capillary electrophoresis. *Biotechniques* **38**,
467 397–400 (2005).
- 468 36. Hargous, Y. *et al.* Molecular basis of RNA recognition and TAP binding by the
469 SR proteins SRp20 and 9G8. *EMBO J.* **25**, 5126–37 (2006).
- 470 37. Ankö, M.-L., Morales, L., Henry, I., Beyer, A. & Neugebauer, K. M. Global
471 analysis reveals SRp20- and SRp75-specific mRNPs in cycling and neural
472 cells. *Nat. Struct. Mol. Biol.* **17**, 962–70 (2010).
- 473 38. Ohtsuka, M., Ling, H., Doki, Y., Mori, M. & Calin, G. A. MicroRNA
474 Processing and Human Cancer. *J. Clin. Med.* **4**, 1651–1667 (2015).
- 475 39. Kent, O. A. & Mendell, J. T. A small piece in the cancer puzzle: microRNAs as
476 tumor suppressors and oncogenes. *Oncogene* **25**, 6188–96 (2006).
- 477 40. Pineau, P. *et al.* miR-221 overexpression contributes to liver tumorigenesis.
478 *Proc. Natl. Acad. Sci. U. S. A.* **107**, 264–9 (2010).
- 479 41. Felli, N. *et al.* MicroRNAs 221 and 222 inhibit normal erythropoiesis and
480 erythroleukemic cell growth via kit receptor down-modulation. *Proc. Natl.*
481 *Acad. Sci. U. S. A.* **102**, 18081–6 (2005).
- 482 42. Costa-Pinheiro, P. *et al.* MicroRNA-375 plays a dual role in prostate
483 carcinogenesis. *Clin. Epigenetics* **7**, 1–14 (2015).
- 484 43. Xiong, X., Kang, X., Zheng, Y., Yue, S. & Zhu, S. Identification of loop
485 nucleotide polymorphisms affecting microRNA processing and function. *Mol.*
486 *Cells* **36**, 518–26 (2013).
- 487 44. Calin, G. A. *et al.* A MicroRNA signature associated with prognosis and
488 progression in chronic lymphocytic leukemia. *N. Engl. J. Med.* **353**, 1793–801
489 (2005).

- 490 45. Kminkova, J. *et al.* Identification of novel sequence variations in microRNAs
491 in chronic lymphocytic leukemia. *Carcinogenesis* **35**, 992–1002 (2014).
- 492 46. Jazdzewski, K. *et al.* Common SNP in pre-miR-146a decreases mature miR
493 expression and predisposes to papillary thyroid carcinoma. *Proc. Natl. Acad.*
494 *Sci. U. S. A.* **105**, 7269–7274 (2008).
- 495 47. Tian, T. *et al.* A functional genetic variant in microRNA-196a2 is associated
496 with increased susceptibility of lung cancer in Chinese. *Cancer Epidemiol.*
497 *Biomarkers Prev.* **18**, 1183–1187 (2009).
- 498 48. Locke, J. M., Lango Allen, H. & Harries, L. W. A rare SNP in pre-miR-34a is
499 associated with increased levels of miR-34a in pancreatic beta cells. *Acta*
500 *Diabetol.* **51**, 325–9 (2014).
- 501 49. Zhou, H. *et al.* microRNA-30c negatively regulates endometrial cancer cells by
502 targeting metastasis-associated gene-1. *Oncol. Rep.* **27**, 807–812 (2012).
- 503 50. Zhang, Q. *et al.* Role of microRNA-30c targeting ADAM19 in colorectal
504 cancer. *PLoS One* **10**, e0120698 (2015).
- 505 51. Garofalo, M. *et al.* EGFR and MET receptor tyrosine kinase–altered
506 microRNA expression induces tumorigenesis and gefitinib resistance in lung
507 cancers. *Nat. Med.* **18**, 74–82 (2011).
- 508 52. Xing, Y. *et al.* MicroRNA-30c contributes to the development of hypoxia
509 pulmonary hypertension by inhibiting platelet-derived growth factor receptor β
510 expression. *Int. J. Biochem. Cell Biol.* **64**, 155–166 (2015).
- 511 53. Moreno-Mateos, M. A. *et al.* Novel small RNA expression libraries uncover
512 hsa-miR-30b and hsa-miR-30c as important factors in anoikis resistance. *RNA*
513 **19**, 1711–25 (2013).
- 514 54. Nguyen, T. A. A. *et al.* Functional Anatomy of the Human Microprocessor.
515 *Cell* **161**, 1374–1387 (2015).
- 516 55. Chakraborty, S., Mehtab, S., Patwardhan, A. & Krishnan, Y. Pri-miR-17-92a
517 transcript folds into a tertiary structure and autoregulates its processing. *RNA*
518 **18**, 1014–28 (2012).
- 519 56. Chaulk, S. G., Xu, Z., Glover, M. J. N. & Fahlman, R. P. MicroRNA miR-92a-
520 1 biogenesis and mRNA targeting is modulated by a tertiary contact within the
521 miR-17–92 microRNA cluster. *Nucleic Acids Res.* **42**, 5234–5244 (2014).
- 522 57. Du, P., Wang, L., Sliz, P. & Gregory, R. I. A Biogenesis Step Upstream of
523 Microprocessor Controls miR-17–92 Expression. *Cell* (2015).
524 doi:10.1016/j.cell.2015.07.008

525

526 **Figure legends**

527 **Figure 1** A genetic variant in the terminal loop of hsa-pri-mir-30c-1 alters its normal
 528 expression. **(a)** Schematic representation of the hsa-pri-mir-30c-1 transcript indicating
 529 the G to A mutation observed in breast and gastric cancer patients. Nucleotides (nt)
 530 encompassing primary (pri-miRNA; nt -159 to +161) and precursor miRNA (pre-
 531 miRNA; nt 1 to 61) used in this experiment are indicated. Numbers are relative to the
 532 first nt of the mature miRNA. **(b)** The relative expression levels of mature miR-30c in
 533 MCF7 cells transfected with plasmids encoding either a wild-type sequence (WT) or
 534 one bearing the G/A mutation (n=12) **(c)** Levels of mature miR-30c in MCF7 cells
 535 transfected with *in vitro* transcribed pri-mir-30c-1 (pri-miRNA) or an RNA
 536 oligonucleotide that mimics pre-miRr-30c-1 (pre-miRNA) sequences, either in a
 537 wild-type or G/A version (n=5). Mann-Whitney U test was used to evaluate
 538 differences between WT and G/A samples. Error bars indicate SEM. ** p<0.01. **(d)**
 539 Representative *in vitro* processing of pri-mir-30c-1 (380 nt) in MCF7 total cell
 540 extracts. Quantification of pre-miRNA band intensities are shown below and
 541 expressed as the relative intensity normalized to pre-miR-30c-1 WT variant.

542

543 **Figure 2** Sequence conservation of pri-mir-30c-1 precursors **(a)** Nucleotide-level
 544 GERP scores across the locus, indicating extensive evolutionary constraint. A GERP
 545 score above zero indicates significant constraint, while a score below zero indicates
 546 an excess of nucleotide substitutions beyond the expected neutral rate. The purple
 547 bars display the total number of observed nucleotide substitutions found in aligned
 548 sequences from 98 vertebrates. Δ represents absence of the nucleotide from 98
 549 vertebrate sequences analyzed. The location of the modified G/A nucleotide is
 550 indicated by a green rectangle. **(b)** Model of the predicted secondary structure with

551 nucleotides colored as in (a) to reflect their variation. Nucleotides present in less than
552 50% of the species are indicated in lower case.

553

554 **Figure 3** Structural analysis reveals the importance of the G to A substitution in pri-
555 miR-30c-1 (a) Values of SHAPE reactivity (depicted using color coded bars) at each
556 individual nucleotide position correspond to the mean SHAPE reactivity (+/- SEM)
557 (n=10). (b) SHAPE differences plots of the G/A variant relative to the wild-type
558 sequence. Only nucleotides with absolute differences bigger than 0.25 and statically
559 significant are depicted (grey colors represent p-values). Bars above the basal line
560 indicate increased SHAPE reactivity for the G/A variant, whereas those below the
561 basal line indicate lower SHAPE reactivity. Mann-Whitney U test was used to
562 evaluate differences between WT and G/A samples. (c) Predicted secondary
563 structures of WT and G/A variant sequences, as determined by SHAPE reactivity
564 values.

565

566 **Figure 4** Identification of trans-acting factors binding to pri-mir-30c-1. (a) Venn -
567 diagram indicating the number of interactors observed for the WT sequence and the
568 G/A variant. (b) Validation of the interaction of SRSF3 with the G/A variant sequence
569 by RNA chromatography followed by Western-blot analysis with an antibody specific
570 for SRSF3. (c) Immunoprecipitation with a specific anti-SRSF3 antibody followed by
571 RT-qPCR quantification (+/- SEM) (n=4). (d) Representative toeprinting assay in the
572 presence or absence of 260nM of purified SRSF3 protein using fluorescent-labeled
573 primers and capillary electrophoresis. The unique RT stop in the G/A variant (n=2) is
574 indicated. (e) Volcano plots showing the SHAPE reactivity data in the presence of
575 260nM (left panel) or 500 nM (right panel) of purified SRSF3 protein. The red dots

576 indicate those positions protected by SRSF3 in a dose-dependent manner (n=5).
577 Mann-Whitney U test was used to evaluate differences (f) Schematic representation of
578 the basal stem of hsa-miR30c G/A illustrating the nucleotides protected by SRSF3 (in
579 red) as well as the RT stop detected by toeprinting assay.

580

581 **Figure 5** Role of SRSF3 in miR-30c expression. (a) The relative expression levels of
582 miR-30c WT and G/A in MCF7 cells with changing expression of SRSF3 protein.
583 Error bars indicate SEM (n=5) * p<0.05 ** p<0.01 (b) Mutational analysis of the
584 CNNC motif. Nucleotides substitutions present in the CNNC region are shown in
585 Red. Relative expression levels of miR-30c in MCF7 cells transfected with plasmids
586 encoding either a wild-type sequence (WT) or one bearing the G/A mutation, and also
587 including mutations in the CNNC motif, when indicated. Mann-Whitney U test was
588 used to evaluate differences. Error bars indicate SEM (n=6) * p<0.05 ** p<0.01.

589

590 **Figure 6** Representative *in vitro* processing assays of radiolabeled shorter pri-mir-
591 30c-1 variants in the presence of immunopurified FLAG-Drosha/FLAG-DGCR8
592 complexes (+) or FLAG immunoprecipitates (-), with addition of purified SRSF3
593 protein (1x corresponds to 65nM of protein). (a, b) Processing of WT pri-mir-30c-1
594 comprising or lacking SRSF3 binding sites (WT and WT ΔCNC, respectively). (c,d)
595 Processing of pri-mir-30c-1 G/A variant comprising or lacking SRSF3 binding sites
596 (G/A and G/A ΔCNC, respectively). Quantification of pre-miRNA band intensities
597 are shown below and expressed as the relative intensity normalized to *in vitro*
598 processing assay in the absence of purified SRSF3 protein (lane 3 in each case).

599

600 **Figure 7** Cartoon depicting a model whereby the genetic variant identified in pri-mir-
601 30c-1 leads to a secondary RNA structure rearrangement that facilitates binding of
602 SRSF3 resulting in increased Microprocessor-mediated processing of pri-miR-30c-1.

603

604 **Supplementary Figure 1** Representative *in vitro* processing of shorter pri-mir-30c-1
605 substrates (153nt) comprising either a wild-type sequence (WT) or bearing the G/A
606 mutation in the presence of immunopurified FLAG-Drosha/FLAG-DGCR8
607 complexes (+) or FLAG immunoprecipitates (-). Quantification of pre-miRNA band
608 intensities are shown below and expressed as the relative intensity normalized to pre-
609 mir-30c-1 WT variant (lane 3). Intervening lanes were removed from the gel image,
610 which is indicated by the black vertical line.

611

612 **Supplementary Figure 2 (a)** Histograms of cleavage intensity versus nucleotide
613 position determined using *in situ*-generated hydroxyl radicals values corresponding to
614 the mean cleavage (+/- SEM) (n=6). **(b)** Predicted secondary structures of WT and
615 G/A variant sequences, as determined by hydroxyl radical reactivity values.

616

617 **Supplementary Figure 3** Identification of proteins interacting with WT and G/A pri-
618 miR30c-1 sequences. **(a)** Table shows proteins identified in two independent
619 experiments. **(b)** Validation of the interaction of SRSF3 with the G/A variant
620 sequence by RNA immunoprecipitation followed by Western Blot analysis with
621 specific antibodies.

622

623 **Supplementary Figure 4** SRSF3 levels following knock-down or overexpression in
624 MCF7 cells. **(a)** Analysis of SRSF3 mRNA levels. Error bars indicate SEM (n=5). **(b)**
625 Representative Western blot of SRSF3 protein levels.

626

627 **Supplementary Figure 5** Detection of SRSF3 protein in FLAG-Drosha/FLAG-
628 DGCR8 complexes used in the *in vitro* processing assays shown on Fig 6 and
629 Supplementary Figure 1. FLAG immunoprecipitates of control cells or cells
630 transfected with FLAG-Drosha + FLAG-DGCR8 were analyzed in a Western blot
631 assay with an anti-SRSF3 specific antibody.

632

633 **ONLINE METHODS**

634 **Plasmids constructions**

635 Pri-mir-30c-1 construct was amplified from human genomic DNA by PCR with
636 specific primers 30c1s (5'-CAAGTGGTTCTGTGTTTTTATTG-3') and 30c1a (5'-
637 GTACTTAGCCACAGAAGCGCA-3') The PCR product was digested with *EcoRI*
638 and was subsequently cloned into the pCDNA3.1 vector (ThermoFisher). The G₂₇ to
639 A mutation was generated by a two-step PCR strategy. First, pri-miR-30c-1 was
640 amplified with a 30Cmut1 oligo (5'-CCTTGAGCTTACAGCTGAGAG-3') and
641 30c1s and also with 30Cmut2 oligo (5'-CTCTCAGCTGTAAGCTCAAGG-3') and
642 30c1a. Both PCR products, were purified (Qiagen), pooled and used as a template for
643 amplification with 30c1s and 30c1a primers. The resulting PCR product was cloned in
644 pGEMt (Promega). pGEMt G/A plasmid was digested with *EcoRI* to clone in
645 pCDNA3.1. The CNNC motifs were subjected to site-specific mutagenesis by PCR
646 amplification, as previously described⁵⁸. Briefly, 10 ng of pri-miR30c-1 (WT or G/A)
647 was PCR amplified with the desired mutagenic oligonucleotide

648 (mut1: 5'-CTTCATTTGATGTTTTCCATGGC-3',
649 mut2: 5'-CTTCTTTTTTTTTTTTCCATGGC-3' or
650 CNNC 5'-CTTCAGATGTTTTCCATGGC-3')
651 and CNNCs primer (5'-CTGCTTACTGGCTTATCG-3'). The PCR product was
652 cleaned (Qiagen) and used as the 5'-flanking primer in a second PCR with an equal
653 molar amount of primer CNNCa (5'-GATATCTGCAGAATTCCTAG-3'). The
654 product of the second PCR was digested with *EcoRI* (New England Biolabs), purified
655 by agarose gel electrophoresis and ligated to the large *EcoRI* fragments of primiR-
656 30c-1 to produce the desired constructs (CNNCmut1, mut2 and ΔCNNC). All the
657 sequences were confirmed by automatic sequencing.

658 **Cell Culture**

659 MCF7 and HEK 293T cells were grown in high glucose Dulbecco's modified Eagle's
660 medium (Invitrogen) supplemented with 10% (v/v) fetal calf serum (Invitrogen) and
661 penicillin-streptomycin (Invitrogen) and incubated at 37°C in the presence of 5%
662 CO₂. Cells were tested for mycoplasma contamination.

664 **Transfections**

665 MCF7 cells grown in 24-well plates were transfected with either pri-mir-30c-1 (WT
666 or G/A) constructs, *in vitro* transcribed RNA (0.33μg/10⁵ cells), or oligonucleotides
667 encoding pre-mir-30c (Sigma Aldrich)
668 (5'UGUAAACAUCUACACUCUCAGCUGUGAGCUCAAGGUGGCUGGGAGA
669 GGGUUGUUUACUCC-3') using Attractene (Qiagen), following manufacturer's
670 instruction. pCDNA-3, pri-miR-30a or oligo30a
671 (5'UGUAAACAUCUACGACUGGAAGCUGUGAAGCCACAAAUGGGCUUUA
672 GUCGGAUGUUUGCAGC-3') were used as negative controls in DNA or RNA

673 transfections, respectively. Cells extracts were prepared at 8h or 48h (after RNA/DNA
674 addition) by direct lysis using 100 µl of lysis buffer (50mM Tris-HCl at pH 7.8, 120
675 mM NaCl, 0.5% NP40). For SRSF3 gene silencing/overexpression MCF7 cells grown
676 in 15 cm dishes were transfected with ON-TARGETplus siRNA (Dharmacon) or
677 pCG. As negative controls, cells were transfected with ON-Targetplus Non-targeting
678 siRNAs and pCG-T7 plasmid, respectively. Cells were split in 24-well dishes 24h
679 after transfection and 24h later transfected with different versions of pri-miRNA
680 constructs. HEK 293Ts were grown to 70% confluency in 6-well plates and then
681 transiently co-transfected with 3µg of FLAG-Drosha and 1µg of FLAG-DGCR8, or
682 4µg FLAG-empty vector per well. Cells were expanded for 24h, then split to 10cm
683 plates and expanded for a further 24hr before cells were scraped, collected and snap
684 frozen until required.

685

686 **SRSF3 purification**

687 Purification of SRSF3 from MCF7 cells was performed as previously described⁵⁹,
688 with minor modifications. Briefly, MCF7 cells were grown to confluence in 15 cm
689 dishes and transient transfected with 15 µg of pCG-T7-SRSF3⁶⁰. Forty eight hours
690 after transfection cell pellets were resuspended in 20ml of ice-cold lysis buffer (50
691 mM NaP buffer, pH8, 0.5 M Na Cl, 5 mM b-glycerophosphate, 5 mM KF, 0.1 %
692 Tween 20, and 1x protease inhibitor cocktail) and sonicated 5 times 30s followed by
693 centrifugation at maximum speed for 20 min at 4°C. After centrifugation the
694 supernatant was loaded into a chromatography column (Biorad) previously prepared
695 with T7 Tag antibody Agarose (Novagen). The flow-through was collected and
696 loaded a second time. The column was washed two times with 10 ml of lysis buffer.
697 Then, eluates were eluted with 10 serial 0.8 ml volumes of elution buffer (0.1 M citric

acid, pH 2.2, 5 μ M β -glycerophosphate and 5mM KF) and collected in microcentrifuges tubes containing 200 μ l of 1M Tris pH 8.8 mixed and stored at 4°C. The fractions were analysed on SDS page (Invitrogen) followed by Coomassie blue staining. The elutes containing the protein were dialyzed overnight against BC100 buffer (20 mM tris, pH8, 100 mM KCl, 0.2 mM EDTA pH8 and 20% glycerol), and stored in aliquots at -80°C.

704

705 **Western blot analysis**

Equal amounts of total protein (determined by Bradford assay) were loaded in 12% NUPAGE gels (Invitrogen) and transferred to cellulose membranes using IBLOT system (Invitrogen). Identification of SRSF3 was performed with a rabbit polyclonal antibody (RN080PW, MBL, Dilution 1:500), followed by a secondary horseradish peroxidase-conjugated antibody and ECL detection (Pierce). Other primary antibodies used in this study were: mouse monoclonal anti-PARP-1 antibody (E-8): sc-74469, Santa Cruz Biotechnology, Dilution 1:500); Rabbit polyclonal anti-DDX17 antibody ((S-17): sc-86409, Santa Cruz Biotechnology, Dilution 1:500); Rabbit polyclonal anti-hnRNP A1 antibody (PA5-19431, Invitrogen antibodies, Dilution 1:500); Rabbit polyclonal anti-TLS/FUS antibody (ab23439, Abcam, Dilution 1:1000).

716

717 ***In vitro* transcription of pri-miRNA substrates**

Prior to RNA synthesis, pri-mir-30c-1 plasmids (wild-type and G/A variant) were linearized using *ApaI* (New England Biolabs). In addition, shorter pri-mir-30c-1 probes were PCR amplified from pri-mir-30c-1 plasmids (wild-type, G/A, and their respective Δ CNNC) for *in vitro* processing assays (shown on **Fig. 6** and **Supplementary Fig. 1**) using a forward primer harboring a T7-promoter sequence

723 fused to a 19nt sequence complementary to pri-mir-30c-1 (5'-
724 AATACGACTCACTATAGGGCTGATCAACCCTGGACC-3') and a reverse primer
725 (5'- AGTGGAGACTGTTCTTCTTC-3'), which when *in vitro* transcribed generated
726 a 153nt (CNNC) or 146nt (Δ CNNC) and then subsequently PCR purified (Qiagen) for
727 *in vitro* transcription. 1 μ g of DNA or 400ng of PCR product were *in vitro* transcribed
728 for 1-2h at 37°C using 50 U of T7 RNA polymerase (Roche) in the presence of 0.5
729 mM rNTPs and 20 U of RNasin (Promega). When needed, RNA transcripts were
730 labeled using (α -³²P)-UTP (800 Ci/mmol, Perkin Elmer) and following DNase
731 treatment, unincorporated ³²P-UTP was eliminated by exclusion chromatography in
732 TE equilibrated columns (GE Healthcare) or PAGE gel purification, followed by
733 phenol/chloroform and ethanol precipitation.

734

735 ***In vitro* processing assays**

736 The *in vitro* processing reactions were performed as previously described⁶¹ with
737 minor modifications. Radiolabeled *in vitro* transcribed pri-mir-30c-1 (50,000 c.p.m.)
738 was incubated with 650 μ g of MCF7 total cell extract (Fig. 1d), or incubated with
739 FLAG-tagged complexes immunopurified from HEK293T cells transiently co-
740 expressing FLAG-Drosha & FLAG-DGCR8 or FLAG-empty vector control, in the
741 absence or RNase⁶² (**Fig. 6** and **Supplementary Fig. 1**). Additionally, *in vitro*
742 processing reactions were supplemented with increasing concentrations of
743 immunopurified T7-SRSF3 (**Fig. 6**). The *in vitro* processing reactions were performed
744 in the presence of buffer A (0.5 mM ATP, 20 mM creatine phosphate and 6.4 mM
745 MgCl₂). Reactions were incubated for 1h at 37°C and treated with proteinase K. RNA
746 was extracted by phenol/chloroform and ethanol precipitation. Samples were resolved
747 in an 8% 1 $\times$ TBE polyacrylamide urea gel.

748

749 **miRNA qRT-PCR**

750 The relative amounts of miR-30c present in total RNA samples were measured using
751 Exiqon miRCURY LNATM following manufacturer's instructions. Total RNA from
752 cytoplasmic lysates was isolated using RNazol (Invitrogen) and RT was carried out
753 with Universal RT microRNA PCR using 500 ng of RNA. A 1/20 dilution of the RT
754 reaction was used for qPCR with a specific microRNA LNATM primers set and the
755 LightCycler system with the FastStart DNA Master Green II (Roche). The amount of
756 miR-30c detected in the reaction was normalized by a parallel reaction performed
757 with U6 and SNOR48 primers.

758

759 **Quantitative RNA co-immunoprecipitation (qRNAco-IP)**

760 The qRNAco-IP was performed as previously described⁶³, with several modifications.
761 Briefly, lysates from pri-miR-30c-1 (WT and G/A) transfected cells were pre-cleared
762 with mouse IgG beads followed by incubation with a polyclonal rabbit anti-SRSF3
763 antibody (MLB). The complexes were pulled-down using protein G beads
764 (Amershan), then treated with proteinase K (Sigma) and RNA was extracted and
765 purified using Trizol (Invitrogen). qRT-PCR was carried out with superscriptIII
766 (Invitrogen) RT using 500ng of RNA as a template and the TL
767 (5'-GGAGTAAACAACCCTCTCCCAGC-3')
768 and actin (5'-GGTCTCAAACATGATCTGGG-3') primers. Then a 1/20 dilution of
769 the RT reaction was used for qPCR with appropriate pairs of specific primers
770 (TL: 5'-CATCCTACACTCTC-3' and
771 5'-GGAGTAAACAACCCTCTCCCAGC-3',
772 actin: 5'-GGGTCAGAAGGATTCCTATG-3' and

5'-GGTCTCAAACATGATCTGGG-3'). qRNAco-IP values were calculated according to the formula: qPCR value TL mRNA IP/qPCR value TLmRNA total)/(qPCRvalue control mRNA IP/qPCR value control mRNA total)

776

777 **Evolutionary constraint**

Evolutionary constraint was quantified for individual nucleotide positions as the number of rejected substitutions, as calculated by the GERP++ algorithm. This data was extracted from the UCSC genome browser, where they had been calculated over their 36-way mammalian genome alignments.

782

783 **Phylogenetic conservation**

The alignment of 98 vertebrate sequences of pri-mir-30c-1 were retrieved from USCS genome browser. Each nucleotide frequency was plotted in a heat map. RNA structure models were performed using RNAstructure software (<http://rna.urmc.rochester.edu/RNAstructureWeb/>). The most energetically stable RNA structure according to the program was used to depict the model of pri-mir-30c-1.

790

791 **SHAPE analysis**

Pri-mir-30c-1 RNA was treated with NMIA (Invitrogen), as previously described⁶⁴. For primer extension, equal amounts of NMIA treated and untreated RNA (0.5 pmols) were incubated with 2 pmol of 5'-end fluorescently-labeled primer (5'-CTAGATGCATGCTCGAGCG-3'). NED fluorophore was used for both, treated and not treated RNAs while VIC fluorophore was used for the sequencing ladder, as described previously.cDNA products were resolved by capillary electrophoresis⁶⁴.

797

798 Pri-miR-30c-1-SRSF3 complexes were assembled in folding buffer (100 mM HEPES
799 pH 8, 6 mM MgCl₂) using 170 nM RNA in the presence of increasing amounts of
800 purified SRSF3 protein (260 and 500 nM). Then, RNA alone or pre-incubated with
801 SRSF3 was treated with NMIA. RNA was phenol extracted and ethanol precipitated
802 and then subjected to primer extension analysis.

803

804 **Hydroxyl radical footprinting**

805 Pri-mir-30c-1 RNA was subjected to Hydroxyl radical footprinting, as previously
806 described⁶⁵. Briefly, 1.7 pmol of RNA was denatured and folded in folding buffer
807 (40 mM MOPS pH 8.0, 80 mM KOAc, and 0, 0.5 or 6 mM MgCl₂). Samples were
808 incubated with 1 µL of the Fe(II)–EDTA complex, 1 µL of sodium ascorbate and 1 µL
809 of hydrogen peroxide for 30 s at 37 °C. Fe(II)–EDTA
810 (7.5 mM Fe(SO₄)₂(NH₄)₂·6H₂O and 11.25 mM EDTA, pH 8.0), 0.3% hydrogen
811 peroxide and 150 mM sodium ascorbate solutions were freshly prepared. As a control
812 a lacking Fe(II)–EDTA reaction was performed. Samples were quenched and
813 precipitated by addition of one-third of 75% glycerol, 1 µL of 20 mg·mL⁻¹ glycogen,
814 1 µL of 3 M NaCl, 2 µL of 0.5 M EDTA and 2.5 volumes of ice-cold ethanol. RNAs
815 were re-suspended and reverse-transcribed using fluorescent primers as described
816 for SHAPE reactivity. cDNA products were resolved by capillary electrophoresis.

817

818 **SHAPE reactivity and Hydroxyl radical cleavage data analysis**

819 SHAPE electropherograms of each RT were analyzed using the quSHAPE software⁶⁶.
820 Then, data from 10 independent assays were used to calculate the mean of SHAPE
821 reactivity. To minimize the technical variation, each RT stop quantitative value was
822 normalized to the total intensity of the average of 10 highly reactive nucleotides in the

823 reaction. Grubbs' test was used to identify outliers. To obtain SHAPE differences
824 between pri-mir-30c WT and GA or Free RNA and RNA-SRSF3 complexes, the
825 SHAPE reactivity values obtained in WT or free RNA were subtracted from the
826 reactivity values obtained in GA or in the presence of the protein. The statistical
827 analysis was performed by Mann–Whitney *U* test. Only nucleotides with absolute
828 differences larger than 0,25 and statistically significant were considered. Hydroxyl
829 radical cleavage intensities of each reaction were also analyzed using quSHAPE
830 software⁶⁶. Then, data from 6 independent assays were used to calculate the mean of
831 hsa-pri-mR-30c WT or GA cleavage. Grubbs' test was used to identify outliers.
832 Buried regions were defined as zones with more than two consecutive nucleotides
833 having a reactivity (R) smaller than the mean of all reactivity, whereas exposed
834 regions are those with more than two consecutive nucleotides having R larger than the
835 mean of all reactivity.

836

837 **Toeprint assays**

838 This assay carried out following the protocol previously described³⁵, with minor
839 modifications. Briefly, pri-mir-30c-1: SRSF3 complexes were assembled as described
840 for SHAPE analysis. After protein incubation samples were subsequently subject to
841 primer extension using fluorescent primer (NED) 5'CTAGATGCATGCTCGAGCG-
842 3.pri-miR-30c. Primer extension products were extracted with phenol, and ethanol
843 precipitated pellets were resuspended in 5 µl of HI-DiTM formamide
844 (ThermoFisher), which included 0.5 µl of GeneScanTM 500 LizTM dye size standards.
845 The products were separated by capillary electrophoresis and analyzed by
846 GeneMarker software.

847 **RNA chromatography**

848 RNase-assisted RNA chromatography with RNase A/T1 was performed, as
849 described²⁸, using *in vitro* transcribed pri-mir-30c-1 and total MCF7 cell extracts.
850 RNA-bound proteins were separated using 4-12% NUPAGE bis-tris system
851 (Invitrogen) and individual lanes were subjected to mass spectroscopy (BSRC Mass
852 Spectrometry and Proteomics Facility, University of St Andrews). Results were
853 confirmed by Western blot analysis with anti-SRSF3 (MLB), anti-PARP-1 (Santa
854 Cruz), anti-DDX17 (Santa cruz), anti-hnRNPA1 (Thermo Scientific) and anti-FUS
855 (abcam) specific antibodies.

856

857 SUPPLEMENTARY REFERENCES

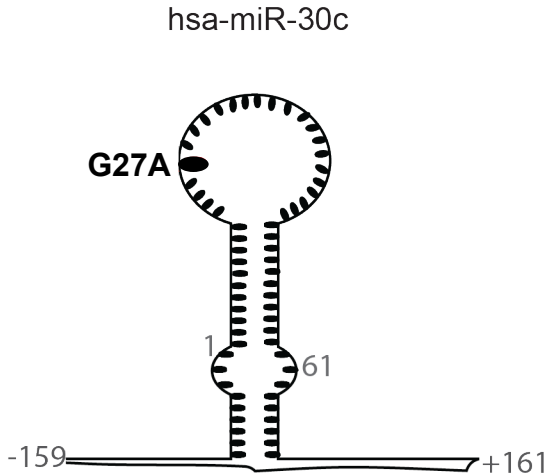
- 858 58. Perrin, S. & Gilliland, G. Site-specific mutagenesis using asymmetric
859 polymerase chain reaction and a single mutant primer. *Nucleic Acids Res.* **18**,
860 7433–8 (1990).
- 861 59. Cazalla, D., Sanford, J. R. & Cáceres, J. F. A rapid and efficient protocol to
862 purify biologically active recombinant proteins from mammalian cells. *Protein*
863 *Expr. Purif.* **42**, 54–8 (2005).
- 864 60. Cáceres, J. F., Misteli, T., Sreaton, G. R., Spector, D. L. & Krainer, A. R. Role
865 of the modular domains of SR proteins in subnuclear localization and
866 alternative splicing specificity. *J. Cell Biol.* **138**, 225–38 (1997).
- 867 61. Heras, S. R. *et al.* The Microprocessor controls the activity of mammalian
868 retrotransposons. *Nat. Struct. Mol. Biol.* **20**, 1173–81 (2013).
- 869 62. Lee Y, Kim VN. In vitro and in vivo assays for the activity of Drosha complex.
870 *Methods Enzymol* **427**, 89–106 (2007).
- 871 63. Tenenbaum, S. a., Lager, P. J., Carson, C. C. & Keene, J. D. Ribonomics:
872 Identifying mRNA subsets in mRNP complexes using antibodies to RNA-
873 binding proteins and genomic arrays. *Methods* **26**, 191–198 (2002).
- 874 64. Fernández, N., García-Sacristán, A., Ramajo, J., Briones, C. & Martínez-Salas,
875 E. Structural analysis provides insights into the modular organization of
876 picornavirus IRES. *Virology* **409**, 251–261 (2011).
- 877 65. Lozano, G., Fernandez, N. & Martinez-Salas, E. Magnesium-dependent folding
878 of a picornavirus IRES element modulates RNA conformation and eIF4G
879 interaction. *FEBS Journal* **281**, 3685–700 (2014).
- 880 66. Karabiber, F., McGinnis, J. L., Favorov, O. V & Weeks, K. M. QuShape: rapid,
881 accurate, and best-practices quantification of nucleic acid probing information,
882 resolved by capillary electrophoresis. *RNA* **19**, 63–73 (2013).

883

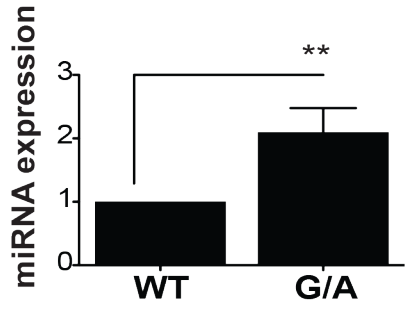
884

Fig 1

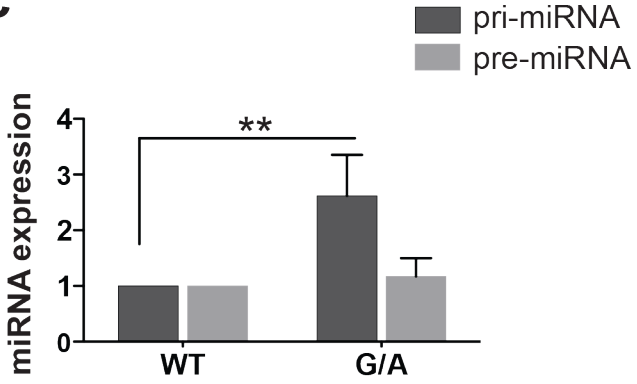
a



b



c



d

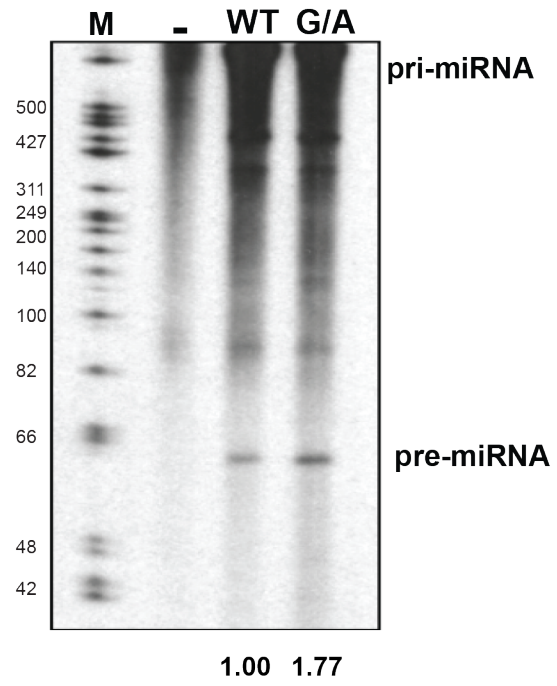
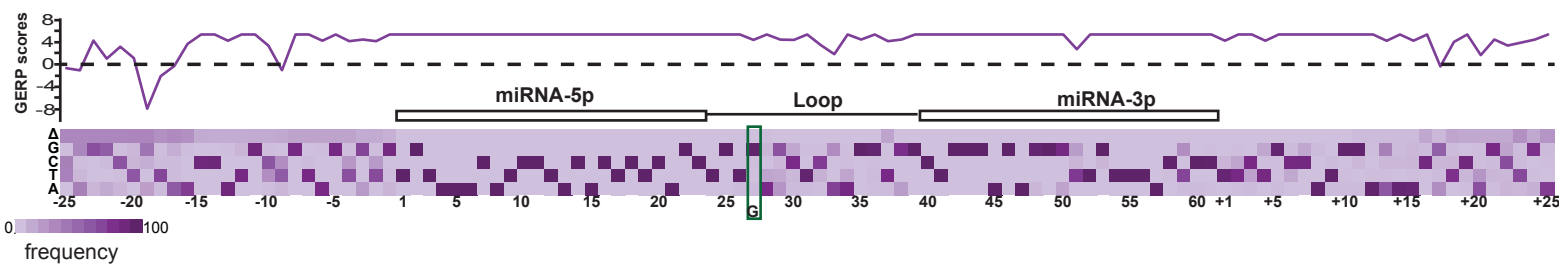


Fig 2

bioRxiv preprint doi: <https://doi.org/10.1101/093179>; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

a



b

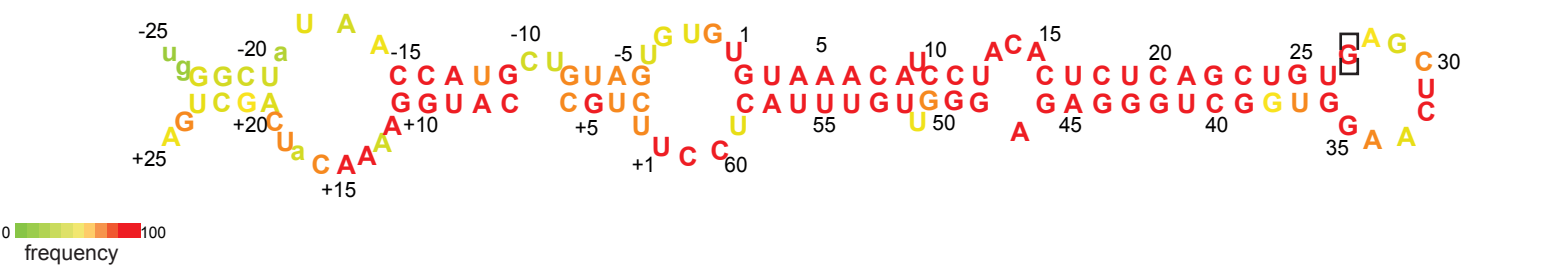


Fig 3 bioRxiv preprint doi: <https://doi.org/10.1101/093179>; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

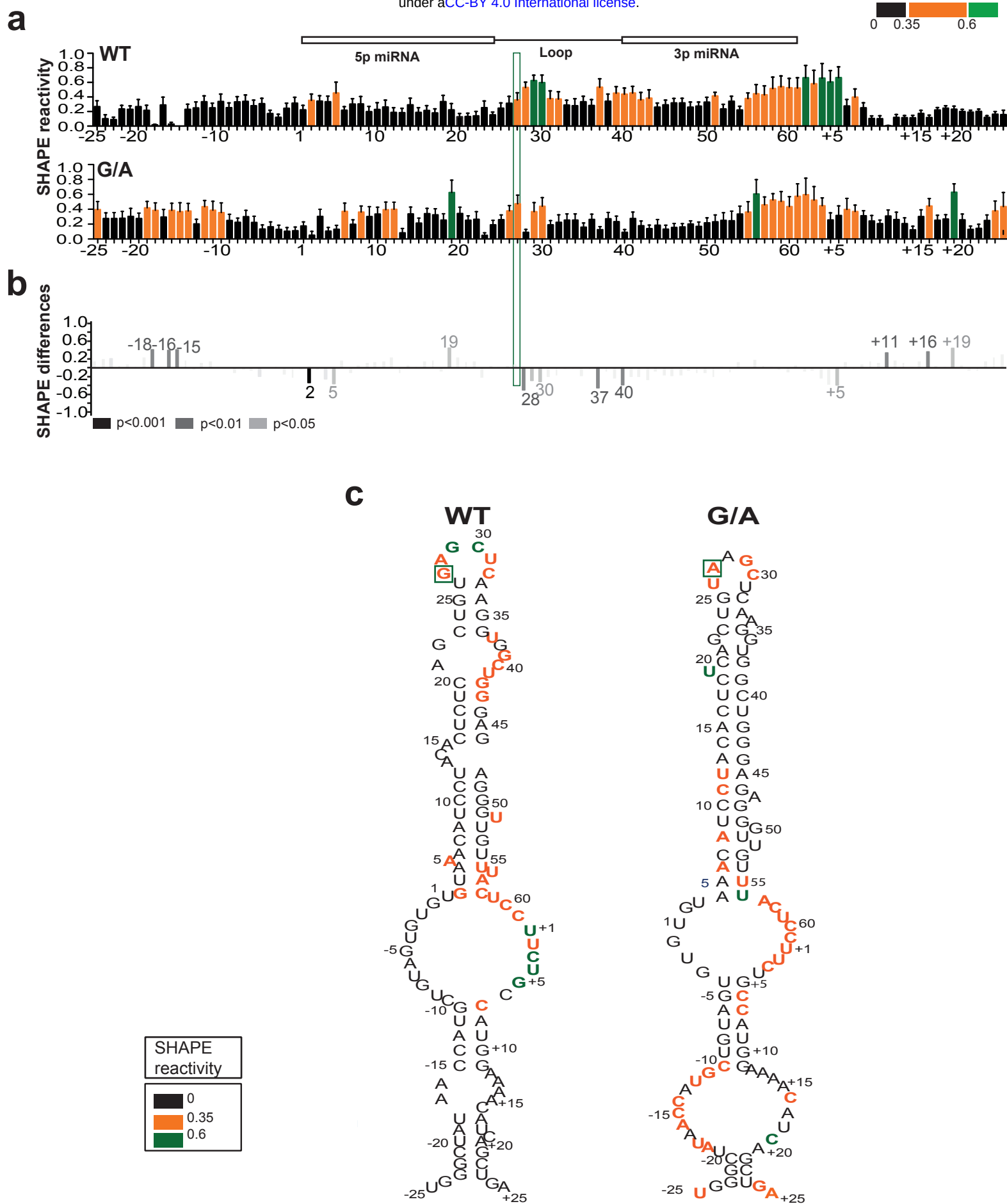


Fig 4

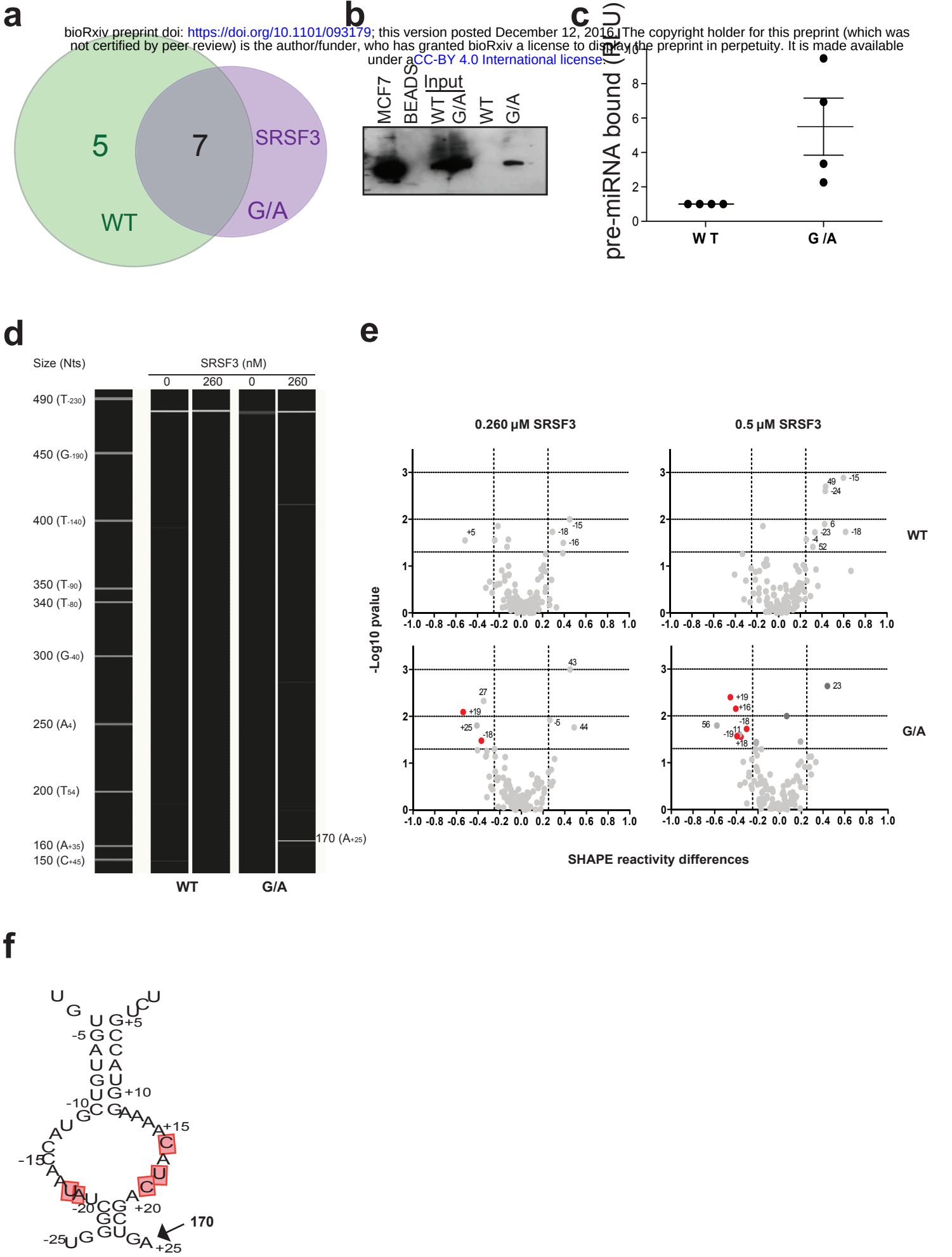
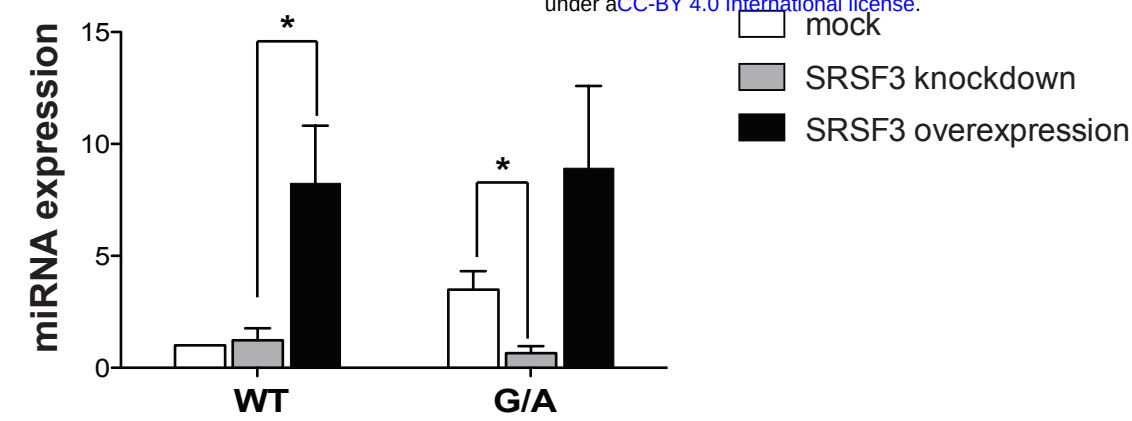


Fig 5

a bioRxiv preprint doi: <https://doi.org/10.1101/093179>; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



b

+15 CAUCAGC +21 CNNC
A - A - - - - CNNCmut1
- - - - - AA CNNCmut2
- - - - - Δ CNNC

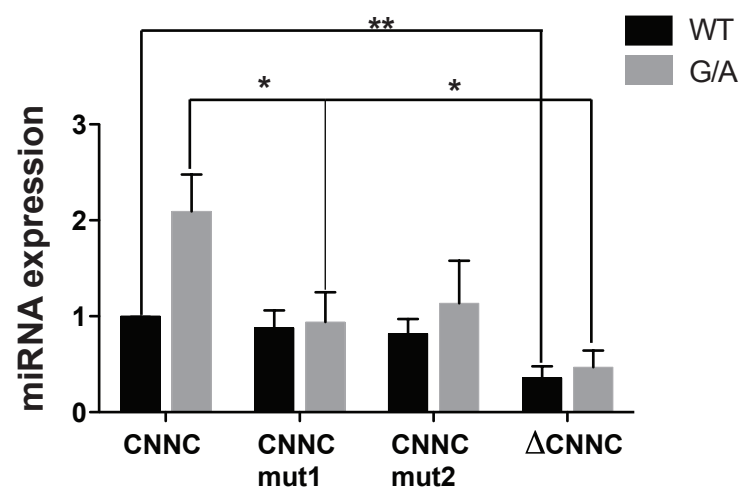


Fig 6

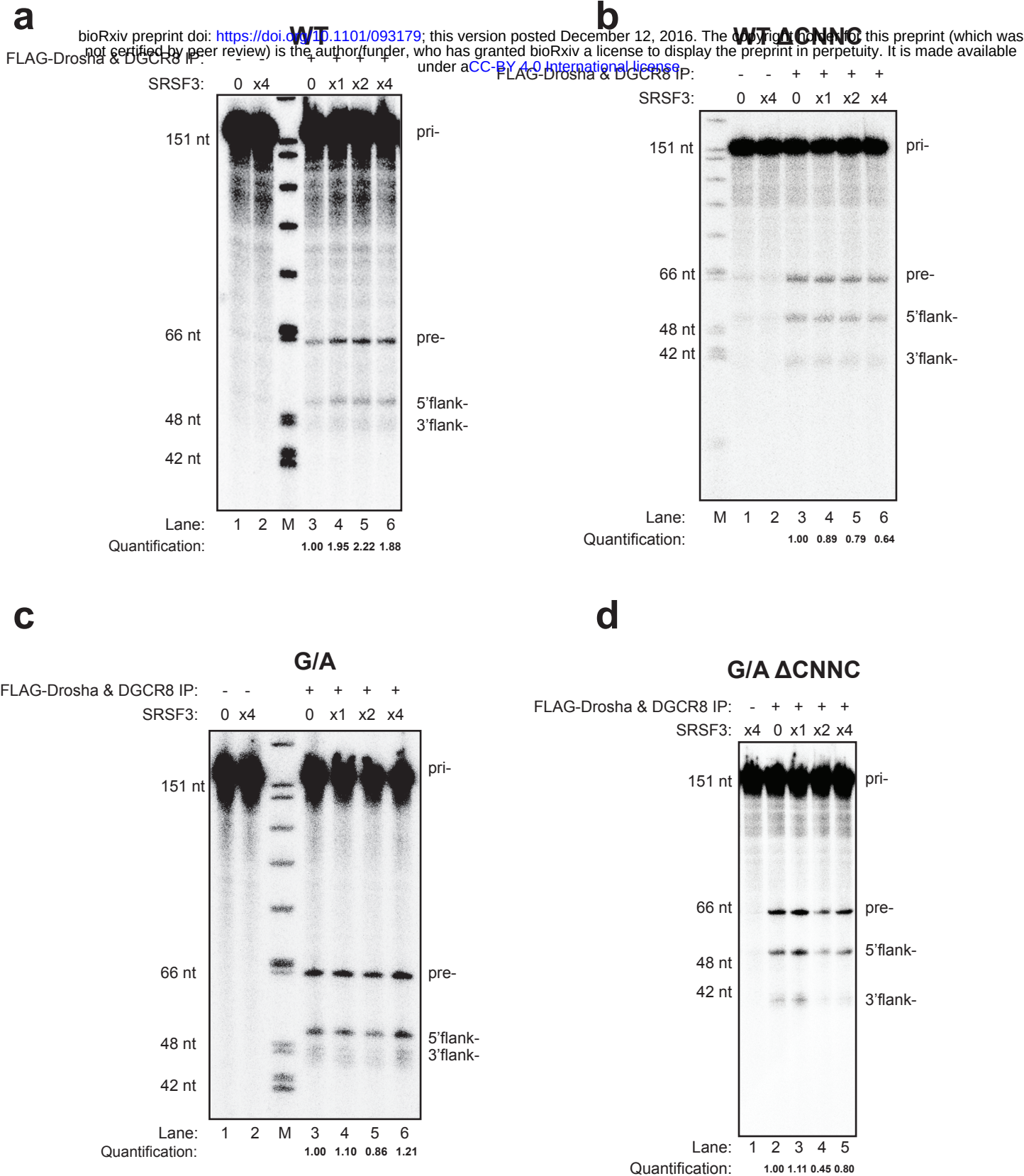


Fig 7

bioRxiv preprint doi: <https://doi.org/10.1101/093179>; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

