NAR Breakthrough Article

A pan-cancer atlas of somatic mutations in miRNA biogenesis genes

Paulina Galka-Marciniak, Martyna Olga Urbanek-Trzeciak, Paulina Maria Nawrocka and Piotr Kozlowski ⓘ*

Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

ABSTRACT

It is a well-known and intensively studied phenomenon that the levels of many miRNAs are differentiated in cancer. miRNA biogenesis and functional expression are complex processes orchestrated by many proteins cumulatively called miRNA biogenesis proteins. To characterize cancer somatic mutations in the miRNA biogenesis genes and investigate their potential impact on the levels of miRNAs, we analyzed whole-exome sequencing datasets of over 10 000 cancer/normal sample pairs deposited within the TCGA repository. We identified and characterized over 3600 somatic mutations in 29 miRNA biogenesis genes and showed that some of the genes are overmutated in specific cancers and/or have recurrent hotspot mutations (e.g. *SMAD4* in PAAD, COAD and READ; *DICER1* in UCEC; *PRKRA* in OV and *LIN28B* in SKCM). We identified a list of miRNAs whose level is affected by particular types of mutations in either *SMAD4*, *SMAD2* or *DICER1* and showed that hotspot mutations in the RNase domains in DICER1 not only decrease the level of 5p-miRNAs but also increase the level of 3p-miRNAs, including many well-known cancer-related miRNAs. We also showed an association of the mutations with patient survival. Eventually, we created an atlas/compendium of miRNA biogenesis alterations providing a useful resource for different aspects of biomedical research.

INTRODUCTION

Since the first reports of microRNA (miRNA) contributions to B-cell chronic lymphocytic leukemia (1), we have observed a substantial increase in reports describing the role of these small regulatory RNA molecules in different human diseases, including cancers (summarized in (2)). It was suggested that miRNAs are globally downregulated in cancer (3) and that upregulation or downregulation of certain miRNAs acting either as oncogenes or tumor suppressors may contribute to cancer development and progression (4,5). Numerous miRNA profiling studies have led to the identification of many miRNAs specifically altered in different types or subtypes of cancer. Many of these miRNAs play an important role in carcinogenesis and the regulation of different cancer-related processes, such as cell growth and differentiation, cell migration, apoptosis, and epithelial-to-mesenchymal transition (MET). Additionally, many miRNAs have been implicated as diagnostic and prognostic biomarkers and/or as potential therapeutic targets in cancer [e.g. (6–9)].

miRNAs are generated through a multistage process of miRNA biogenesis tightly controlled by various proteins consecutively nursing primary miRNA transcripts (pri-miRNAs) from their transcription to their cellular function within the miRNA-induced silencing complex (miRISC) (10–13). The major steps of canonical miRNA biogenesis include nuclear pri-miRNA processing by the microprocessor complex, whose core is formed by RNase DROSHA acting together with DGCR8 dimer and several other regulatory proteins, including P68 (DDX5) and P72 (DDX17), to release from the pri-miRNA hairpin-shaped secondary precursor (pre-miRNA). Next, pre-miRNA is exported to the cytoplasm by the Exportin-5(XPO5):Ran-GTP(RAN) complex, where it is intercepted by the multiprotein miRISC loading complex (RLC) containing the RNase DICER1, which cuts off the pre-miRNA apical loop to release an ∼22-bp-long miRNA duplex in assistance of partner proteins such as PRKRA (PACT) or TARBP1 (TRBP). Within the miRISC, the miRNA duplex is unwound (supported by, i.a., GEMIN4 and MOV10) to select a miRNA guide strand (mature miRNA) that recognizes mRNA targets by complementary interaction and silences them with the as-

sistance of AGO and TNRC6A (GW182) proteins by translation repression and/or RNA deadenylation and degradation. Each step of this process may be further regulated by additional mechanisms/proteins that either increase or decrease the miRNA biogenesis rate (11,14,15). For example, LIN28A/B binds to the apical loop of specific pre-miRNAs, including pre-let-7, and upon its uridylation by ZCCHC11 or ZCCHC6 TUTases leads to pre-miRNA degradation by DIS3L2 exonuclease (16,17). It should also be noted that there are alternative pathways of miRNA biogenesis, such as the generation of miRNAs (mirtrons) from specific short introns (DROSHA-independent) (18) or DICER1-independent processing of miR-451a by AGO2 (19,20).

There are currently >2600 human miRNAs deposited in miRBase (21,22). It is speculated that miRNAs regulate the expression of most protein-coding genes (23,24). miRNA levels and consequently levels of controlled genes may be affected by various factors and processes. First, the expression of miRNA genes, such as the expression of protein-coding genes, is regulated by various transcription factors, such as MYC, TP53 or SMAD4 (25–28). It was shown that many miRNA genes are located in copy number-variable (CNV) regions and are frequently amplified or deleted in cancer (29–31). miRNA genes may also be affected by aberrant DNA methylation and histone acetylation, leading to the silencing of the miRNA genes (32–38). Additionally, methylation of miRNA precursors may facilitate their processing (39) as well as impair the ability of miRNAs to downregulate their targets (40). It was also shown that the occurrence of single nucleotide polymorphisms (SNPs) (41–44) and germline or somatic mutations (45–49) may affect miRNA processing (level) and the ability of miRNAs to recognize their targets. Mature miRNAs may be captured and inactivated by cellular miRNA sponges such as lncRNAs, circRNAs, or pseudogenes (50,51). Finally, the deficiency or impairment of the components of miRNA gene transcription activators, miRNA processing machinery and the miRISC complex (for simplicity, cumulatively called miRNA biogenesis genes/proteins) described in the previous paragraph may affect miRNA levels and the effectiveness of miRNA gene silencing, respectively (52,53).

A body of evidence has indicated that deleterious germline mutations in the *DICER1* gene are responsible for DICER1 syndrome, an inherited disorder characterized by an increased frequency of various types of malignant and benign tumors that occur predominantly in infants and young children, the most common and most characteristic of which is pleuropulmonary blastoma (54,55). It was shown that in cancers associated with DICER1 syndrome as well as other early childhood cancers (e.g. Wilms' tumor), a specific pattern of somatic *DICER1* second-hit missense mutations occurs. All these mutations are located in or adjacent to metal-ion-binding residues (hotspots; predominantly D1709 and E1813) of the RNase IIIb domain (RIIIb) (54). Later, in similar types of childhood cancers, a similar pattern of somatic mutations was also identified in the corresponding residues (E1147, D1151) of the RIIIb in *DROSHA* (56–59). Functional analyses revealed that the mutations in *DICER1* lead to less effective generation of 5p-miRNAs (60–62), whereas mutations in *DROSHA* affect the generation of miRNAs from both pre-miRNA arms

(56), as reviewed in (54,63,64). Very recently, it was also shown that the recurring mutation in the RNase IIIa domain (RIIIa) of *DICER1* occurring predominantly in uterine carcinoma may cause the same effect as the mutations in RIIIb (65). Interestingly, the occurrence of *DROSHA* mutations in Wilms' tumor coincides with the occurrence of mutations in *SIX1* and *SIX2*, transcription factor genes that are also frequently mutated in the tumor (66). Another hotspot mutation commonly occurring in Wilms' tumor is E518K in the double-stranded RNA-binding domain (dsRBD) of DGCR8 (57,58,66). It was also shown that in cancers with a high rate of microsatellite instability (MSI), such as colon, gastric, and endometrial tumors, specific indel hotspots occur in *TRBP* and C-terminal positions of *XPO5* (67,68); however, these mutations were not further analyzed in other studies. Knowledge of the germline and somatic variation in miRNA biogenesis genes is summarized in (64). Additionally, *SMAD4*, encoding the SMAD4 transcription factor activating many genes in response to transforming growth factor beta (TGFB)/bone morphogenetic protein (BMP) signaling (69,70), is a well-known tumor suppressor gene that is highly mutated in many cancers, including pancreatic and colorectal cancers (71). Although SMAD4 was also implicated in the transcription of miRNA genes (25,72–74), the effect of *SMAD4* mutations has never been tested in the context of the activation of miRNA genes. Furthermore, it was shown that some SNPs in miRNA biogenesis genes are associated with the risk of various cancers. Examples include (i) the rs3742330 (A>C) SNP located in the 3′ UTR of *DICER1* that affects DICER1 mRNA stability and is associated with susceptibility and malignancy in gastric cancer (75,76), increased survival of T-cell lymphoma patients (77) and lower prostate cancer aggressiveness (78); (ii) the rs78393591 SNP in *DROSHA* and rs114101502 SNP in *ZCCHC11* (TUTase responsible for pre-miRNA uridylation and subsequent DICER1 cleavage inhibition) associated with the risk of breast cancer (79) and (iii) the rs11786030 and rs2292779 SNPs in *AGO2*, rs9606250 SNP in *DGCR8*, and rs1057035 SNP in *DICER1* associated with the survival of breast cancer patients (80). Additionally, the SNPs rs2740348 C>G and rs7813 C>T in *GEMIN4*, a gene involved in miRISC formation and miRNA-duplex unwinding, were implicated in the risk of several cancers, although the results were not conclusive (81–85).

In this study, we took advantage of the data generated within The Cancer Genome Atlas (TCGA) project to analyze the somatic mutations in miRNA biogenesis genes. As a result, in a wide panel of 33 cancer types consisting of over 10 000 samples, we identified hundreds of mutations and many recurrently mutated hotspot positions and showed that some of the genes are specifically overmutated in particular cancer types. We also confirmed the common occurrence of deleterious mutations in *SMAD4* and further characterized the specific hotspot mutations in *SMAD4*, *SMAD2* and *DICER1*, the last group of which were previously reported mostly in childhood cancers. We followed up on the consequences of some of the mutations and showed characteristic changes in miRNA profiles resulting from specific mutation types in *DICER1*, *SMAD4* and *SMAD2*. We also showed the associations of the mutations with can-

cer characteristics and patient survival. Additionally, the specific hotspot mutations in *DROSHA* and *DGCR8* commonly observed in Wilms' tumor and other childhood cancers were absent in adult cancers.

MATERIALS AND METHODS

Data resources

We used molecular and clinical data (Level 2) for 33 cancer types generated and deposited in the TCGA repository (http://cancergenome.nih.gov). These data included the results of somatic mutation calls in whole-exome sequencing (WES) datasets of 10 369 samples (later limited to 10 255) analyzed against matched normal (noncancer) samples with the use of the standard TCGA pipeline. Hypermutated samples were defined as samples with >10 000 mutations in the whole exome. As in general, the TCGA datasets include only one sample from each cancer specimen, in the analysis, we did not consider cancer stromal heterogeneity. Copy number data were obtained via Xena UCSC as a 'gene-level copy number (gistic2_thresholded)' dataset of the TCGA Pan-Cancer (PANCAN) cohort. The crystal structure of the phosphorylated SMAD2/SMAD4 heterotrimeric complex (PDB code: 1U7V) (86) was visualized with the use of PyMOL (Schrödinger, LLC, New York, NY, USA).

Data processing

We analyzed somatic mutations in 29 miRNA biogenesis genes (coding exons were extended by 2 nt on each side to enable identification of definitive splicing mutations). The genomic coordinates of the analyzed genes/regions are shown in Supplementary Table S1. From the WES data generated with the use of four different algorithms (MuSE, MuTect2, VarScan2 and SomaticSniper), we extracted somatic mutation calls with PASS annotation. The extraction was performed as described in (45) with a set of in-house Python scripts available at (https://github.com/martynaut/mirnaome_somatic_mutations). Briefly, the lists of somatic mutations detected by different algorithms were merged such that variants detected by more than one algorithm were not multiplicated. To further increase the reliability of the identified somatic mutations (and avoid the identification of uncertain mutations), we additionally removed those that did not fulfill the following criteria: (i) at least two alternative allele-supporting reads in a tumor sample (if no alternative allele-supporting read was detected in the corresponding normal sample); (ii) at least a 5× higher frequency of alternative allele-supporting reads in the tumor sample than in the corresponding normal sample; (iii) a somatic score parameter (SSC) > 30 (for VarScan2 and SomaticSniper) and (iv) a base quality (BQ) parameter for alternative allele-supporting reads in the tumor sample >20 (for MuSE and MuTect2). All mutations were designated according to HGVS nomenclature at the transcript and protein levels, and the effects of mutations were predicted using the Ensembl Variant Effect Predictor (VEP) tool (87). For visualization of mutations on the gene maps, we used ProteinPaint from St. Jude Children's Research Hospital – PeCan Data Portal (88). The protein domains visualized on gene maps were positioned according to UniProt (89)

miRNA expression analysis

We obtained miRNA expression data via Xena UCSC as batch-effects normalized data for TCGA Pan-Cancer data (note that due to normalization, some miRNA levels were below 0). Expression data from the set of ∼700 miRNAs (annotated in miRBase) were filtered to exclude miRNAs with undetectable signals (level = 0) in more than 30% of pan-cancer samples (or 10% when the analysis was performed for specific cancers). As high-confidence miRNAs, we considered those deposited in MirGeneDB v2.0 which were defined based on criteria that include careful annotation of the mature versus passenger miRNA strands and evaluation of evolutionary hierarchy (90,91). Next, to enable pan-cancer comparisons, we normalized the variation (range of −1 to 1) and median (median = 0) of miRNA levels to be equal in each cancer type. The normalized miRNA level changes (in pan-cancer) were calculated/expressed as differences, and the changes in raw (non-pan-cancer-normalized) miRNA levels (in individual cancers) were calculated/expressed as $\log_2$ fold changes. For the differential analysis of miRNA levels, normalized against the level of miR-451a (see Results) in UCEC, we additionally used miRNA level data expressed only as reads_per_million_miRNA_mapped (RPM) values (non-batch effects normalized), downloaded from FireBrowse (illuminahiseq_mirnaseq-miR_gene_expression). The analysis of isomiRs was performed with the use of the miRNA isoform dataset (illuminahiseq_mirnaseq-miR_isoform_expression) from FireBrowse.

Statistics

Unless stated otherwise, all statistical analyses were performed with statistical functions in the Python module scipy.stats. Particular statistical tests are indicated in the text, and if not stated otherwise, $P < 0.05$ was considered significant. If necessary, P-values were corrected for multiple tests. Mutation density was calculated as the number of detected mutations divided by the length of analyzed genes (total length of all coding exons). For patient survival analyses, we used a log-rank test (from the lifelines v0.24.8 library (Davidson-Pilon C *et al.*, https://zenodo.org/record/3833188#.X8fbMs1KhPY)). To determine the direction of mutation effects on survival, we used Cox's proportional hazard model. Survival plots were created using Kaplan-MeierFitter from the lifelines library. A comparison of the occurrence of mutations in miRNA biogenesis genes among cancer stages (with correction for cancer type) was performed with the Cochran–Mantel–Haenszel test for pan-cancer and Fisher's exact test for specific cancers.

RESULTS

Distribution of somatic mutations in miRNA biogenesis genes across different cancer types

For analysis, we selected 29 miRNA biogenesis genes encoding proteins playing roles in (i) the transcription of primary miRNA precursors (pri-miRNAs), (ii) pri-miRNA to pre-miRNA processing in the nucleus, (iii) the export of pre-miRNA from the nucleus to the cytoplasm, (iv) pre-miRNA

Table 1. List and characteristics of the selected miRNA biogenesis genes

Gene/protein ID (alias)	miRNA-related function
Nuclear steps of miRNA biogenesis	
SMAD4	Activates miRNA precursor transcription upon TGFB/BMP activation (74,132,136)
FUS	Facilitates cotranscriptional DROSHA recruitment to pri-miRNAs; shown to bind a terminal loop of specific neuronal pri-miRNAs and to enhance their processing (141)
SRSF3 (SRP20)	Enhances mammalian pri-miRNA processing upon binding to the CNNC motif in the 3p flanking sequence (142)
DROSHA	RNase III; catalyzes pri-miRNA to pre-miRNA for processing/cleavage (11,143,144)
DGCR8	Cofactor of DROSHA; coordinates the recognition of pri-miRNA at the dsRNA-ssRNA junction; functions as a molecular anchor and direct DROSHA to cleave pri-miRNA ~11 bp before the junction (11,145)
DDX5 (P68)	Plays a role in recognition/binding of pri-miRNAs by the DROSHA complex; recruits DROSHA/DGCR8 to pri-miRNA (146,147)
DDX17 (P72)	Plays a role in recognition/binding of pri-miRNAs by the DROSHA complex (146,148)
GSK3B	Facilitates pri-miRNA binding by DROSHA and enhances DROSHA association with cofactors DGCR8 and P72 (149); DROSHA phosphorylation/stabilization (150)
SMAD2	Accelerates pri-miRNA processing by DROSHA; in complex with SMAD4 activates miRNA precursor transcription upon TGFB/BMP signaling (72,136,137)
Export to cytoplasm	
XPO5 (EXP5)	Plays a role in the nuclear export of pre-miRNA to the cytoplasm (151,152); facilitates the nuclear cleavage of clustered pri-miRNAs (153)
RAN	Interacts with XPO5; plays a role in the export of pre-miRNA to the cytoplasm (151)
Cytoplasmic steps of miRNA biogenesis	
DICER1 (DICER)	RNase III; catalyzes pre-miRNA to miRNA-duplex processing by cutting off the pre-miRNA terminal loop (11,154–156)
TARBP2 (TRBP)	Coordinates pre-miRNA recognition by DICER1 and the precision of DICER1 cleavage (157,158)
PRKRA (PACT)	Coordinates pre-miRNA cleavage by DICER1 and assures the precision of the cleavage; plays a role in miRISC assembly and thus participates in miRNA stabilization and accumulation in the cell (158,159)
ADAR	Double-stranded RNA-specific adenosine deaminase; plays a role in pri- and pre-miRNA stem editing (A to I), which makes miRNA precursors resistant to DICER1 cleavage (160)
KHSRP (FUBP2, KSRP)	Binds to the terminal loop sequence of a subset of miRNA precursors, promoting their maturation (161)
LIN28A and LIN28B	Bind to the terminal loops of specific pre-miRNAs (including pre-let-7 and pre-miR-9) and, upon recruitment of ZCCHC11 or ZCCHC6 that induces pre-miRNA uridylation, inhibit DICER1 processing (17,162,163)
ZCCHC11 (TUT4) and ZCCHC6 (TUT7)	Play a role in pre-miRNA uridylation and thus inhibit DICER1 processing (17,164)
DIS3L2	Exoribonuclease; targets the uridylated let-7 precursors (16)
miRNA functioning	
AGO1, AGO2, AGO3, and AGO4	Play a role in miRISC formation/loading (summarized in (165,166)); catalytically active AGO2 functions as an endonuclease upon complementary mRNA:miRNA interaction (167)
GEMIN4	Binds to the miRNA guide strand and facilitates the formation of a miRISC by unwinding the miRNA duplex (83,168)
MOV10	Upon interaction with the miRNA-loaded AGO-protein complexes, plays a role in mRNA degradation; present in P-bodies (169,170)
FMR1	Interacts with DICER1 and AGO1 during mRNA degradation (171); controls DROSHA expression (172)
TNRC6A (GW182)	Component of P-bodies; plays a role in mRNA degradation upon interaction with Argonaute proteins (173,174)

processing and miRNA maturation in the cytoplasm and (v) miRNA:target recognition/interaction and regulation of downstream silencing effects (Table 1 and Figure 1).

To identify somatic mutations in the miRNA biogenesis genes, we took advantage of WES datasets of 10 369 paired tumor/normal samples generated within the TCGA project. The collected samples cover 33 different cancer types (analyzed together as a pan-cancer cohort) (Supplementary Table S2). The list of all cancer types (full names and abbreviations) is shown in Figure 2 (to avoid confusion, we will use the abbreviations for the TCGA sample sets but not generally for particular types of cancer; in the latter case, we will use full cancer-type names). Applying the rigorous criteria described in the Materials and Methods, we identified a total of 5483 mutations in the pan-cancer cohort. However, a substantial fraction ($n = 1834$, $\sim$30%) of the mutations were identified in a relatively small number ($n = 114$, $\sim$1%) of hypermutated samples. As shown in Supplementary Figure S1, the number of mutations in hypermutated samples strongly depends on the general burden of mutations in these samples, implying enrichment of random, most likely passenger mutations in the hypermutated samples. Therefore, to reduce the proportion of confounding mutations, we removed hypermutated samples from subsequent analyses. The removed, hypermutated samples originated mostly from SKCM, UCEC and COAD (Supplementary Figure S1).

After removing the hypermutated samples, we continued analysis on 10 255 samples with 3649 mutations, including 2196 (60%) missense mutations, 774 (21%) synonymous mutations and 625 (17%) definitive deleterious mutations, consisting of 341 frameshift, 222 nonsense and 62 splice-site mutations (Supplementary Table S3). Other types of mutations, such as start or stop codon mutations or complex mutations, were also present in small fractions. At least one mutation was detected in 2104 (21%) samples, including numerous samples with more than one mutation (Figure 2A and B). The frequency of mutated samples differed substantially among cancer types, ranging from 45% (SKCM) to 0% (KICH) (Figure 2A, Supplementary Table S3), and roughly corresponded to the mutational burden in particular cancer types.

Frequency of somatic mutations in miRNA biogenesis genes across 33 cancers

As shown in Figure 2C, *TNRC6A*, *DICER1* and *SMAD4* are among the most highly mutated genes ($\sim$3% of samples in pan-cancer)*,* and the least mutated genes ($\sim$0.3%) are *LIN28A*, *RAN* and *SRSF3*. Although there is some correlation between the frequency of mutations in the particular genes and the length of their protein-coding sequences ($R^2 = 0.62$), the overall mutation frequency cannot be simply explained by the length of the genes. Additionally, none of the genes are well-known highly mutated genes or are located in late-replicating regions known to be overmutated in cancer (92).

Although there is a general correlation of mutation frequency in particular genes between cancers, there are also apparent striking exceptions of genes specifically overmutated in particular cancers (Figure 2C). This contrasts with

observations that over- and undermutated regions generally overlap between different cancer types (92) and suggests the nonrandom occurrence and potentially functional nature of the overmutations. To identify overmutated genes, we statistically compared the frequency of mutations (overmutation) in particular genes (versus all other genes) in particular cancers with corresponding frequencies in pan-cancer (Figure 2C, Supplementary Table S4). The most striking example of an overmutated gene is *SMAD4* overmutated in PAAD (22% of samples), READ (16%), COAD (14%), STAD (9%) and ESCA (8%). Other interesting examples included *AGO4*, *LIN28A* and *SMAD2* overmutated in BLCA (an otherwise extremely low-mutation cancer with no mutations in other genes); *LIN28B* overmutated in SKCM; *PRKRA* overmutated in OV; and *DDX5* overmutated in BRCA and KIRP. There are also other genes with increased mutation frequency, e.g. *TNRC6A* in STAD and *DICER1* in SKCM and UCEC (nominal $P > 0.005$), but these overmutations are only nominally significant. Consistent with the above findings, the overmutated genes are outliers in terms of the correlation of mutation frequencies in particular genes between particular cancer types and the remaining pan-cancer (Supplementary Figure S2).

Distribution of mutations in the miRNA biogenesis genes—identification of hotspot mutations

To illustrate the mutation distribution along the protein sequences, all the mutations were visualized in lollipop plots (Figure 3 and Supplementary Figure S3). As shown in the plots, although most of the mutations are quite evenly distributed along the genes, there are also some hotspot regions/mutations suggesting the functional nature of these changes. The most striking example is a cluster of eight hotspots of recurrently mutated amino acid (AA) residues (i.e. D351, G352, D355, P356, R361, H382, G386 and D537) occurring in the MH2 domain of SMAD4 (Figure 3A). The most prominent hotspot position is R361, which by itself acquired 37 missense mutations, accounting for 23% of all *SMAD4* missense mutations. There are also two recurrently mutated AA residues (i.e. P305 and R321) in the MH2 domain of SMAD2 (Figure 3B). The hotspot mutations in the MH2 domains (both in SMAD4 and SMAD2) likely affect SMAD4:SMAD2 heterotrimer formation. Additionally, there are two recurrent protein-truncating nonsense mutations in the MH2 domain of SMAD2, i.e. p.S306Ter and p.S464Ter, the latter of which occurs 13 times is localized in the last exon of *SMAD2* and likely truncates the last five AAs of the protein just before two phosphorylation sites (S465 and S467) critical for activation of SMAD2 upon TGFB/BMP signaling (93).

Two other clusters of missense hotspot mutations are located in metal ion-binding residues of the RIIIa (S1344) and RIIIb (E1705, D1709, D1810 and E1813) domains of *DICER1* (Figure 3C). These hotspots were previously detected and functionally characterized in various pediatric cancers (54,64), thyroid adenomas (94), and the TCGA cohort, mostly in UCEC samples (65).

It is also worth noting the hotspot missense mutation, i.e. p.P127L in the DRBM 2 domain of *PRKRA* (Figure 3D), occurring 13 times in OV, COAD, GBM and LUAD
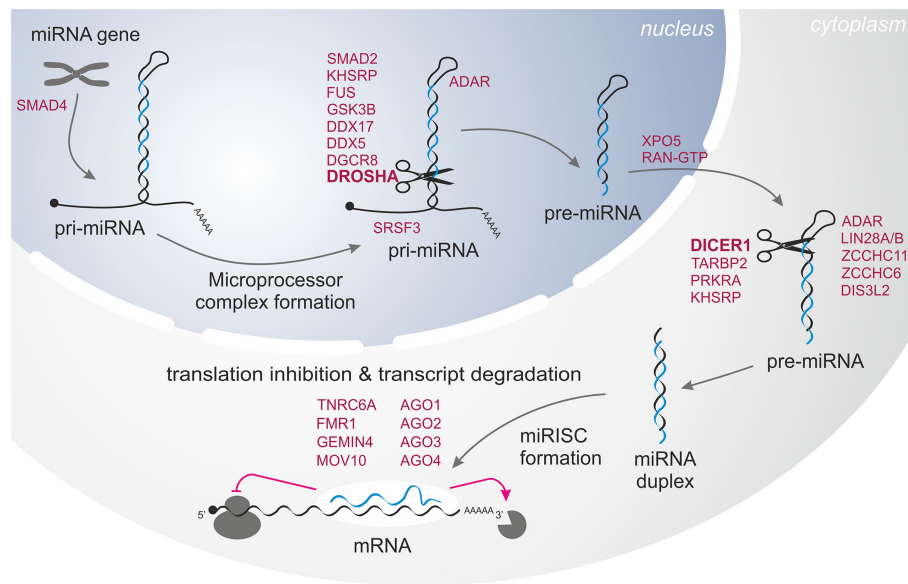
Figure 1. Schematic depiction of miRNA biogenesis and functions of the miRNA biogenesis proteins/genes involved in the subsequent steps indicated and described in Table 1.

(6, 3, 3 and 1 mutations, respectively). As the DRBM 2 domain plays a role in interaction with other proteins (e.g. protein kinase R, PKR), the mutation may affect the interactions. Consistently with the frequent occurrence of the mutation in OV, the changes in the PRKRA level were recently linked with the resistance of mucinous ovarian cancer to the miR-515-3p dependent platinum-based (oxaliplatin) treatment (95). Although the mutation overlaps with the SNP (rs75862065), the fact that it occurs predominantly in OV, which is otherwise a moderately mutated cancer, combined with the overall overmutation in PRKRA, argues against the accidental occurrences of the mutation as artifacts of the mutation calling process. There are also two hotspot synonymous mutations in FUS (p.G222= and p.G227=) (Supplementary Figure S3). A role of such hotspots cannot be excluded (96); however, we did not investigate them further in this study. Similarly, recurring missense mutations in the other genes may be important for specific cancers; however, they occur much less frequently.

There are also some hotspot indel mutations, e.g., p.N58IfsTer8 in ZCCHC11, p.Y948MfsTer16 in ZCCHC6, p.K121SfsTer8 and p.G163EfsTer20 in DDX17, p.W804GfsTer99 and p.R1183GfsTer7 in TNRC6A, p.L17CfsTer99 in AGO1, and pA603RfsTer71 in AGO2 (Figure 3 and Supplementary Figure S3). However, it must be noted that the exact position of indel hotspot mutations may not be necessarily driven by cancer advantage but by sequence properties (e.g., the presence of short tandem repeat motifs). In fact, many indel mutations and some of the indel hotspots occur at sequence motifs often mutated as a result of MSI. Additionally, as most indels in coding sequences result in frameshifts and premature termination of translation, triggering nonsense-mediated mRNA decay (NMD) and leading to the complete loss of mRNA, the exact position of indel hotspots may not be that important.

In the next step, we looked for hotspot mutations previously observed in the miRNA biogenesis genes, i.e. (i) E969 and E993 in RIIIa and E1147, D1151, Q1187 and E1222 in RIIIb in *DROSHA*, (ii) p.E518K in the dsRDB1 domain in *DGCR8* and (iii) p.R440Ter in *XPO5* observed in different pediatric cancers, especially in Wilms' tumor (56,57,59). Of these mutations, we found p.E518K in *DGCR8* occurring in two of 495 (0.4%) cases of THCA, p.R440Ter in *XPO5* in one case of UCEC and one case of SKCM, and p.D1151E in DROSHA in one case of COAD (Supplementary Figure S3). Of the indel hotspots in *XPO5* and *TARBP2* detected previously in colon, gastric, and endometrial tumors with MSI (67,68), we detected only two cases of the C insertion in the poly-C track (p.M145HfsTer13) in *TARBP2*: one in UCEC and one in STAD (both cancers often characterized by MSI).

Functional consequences of *SMAD4 and SMAD2* mutations

From the visual investigation (Figure 2C and Figure 3A), it is apparent that *SMAD4* is the gene with the highest density of mutations (273 mutations, 160 mut/kbp) and the largest proportion of deleterious mutations (33%). As mentioned in the previous paragraph, there are also eight hotspots of missense mutations, seven of which are located in a relatively small region (D351 to G386) of the MH2 domain, playing a role in heterotrimer formation with other receptor-dependent SMAD proteins (R-SMADs, e.g. SMAD2) to mediate the TGFB/BMP transcriptional response (97,98). The collection of a relatively large number of mutations associated with different cancer types reveals that the proportion of hotspot and deleterious mutations [in pan-cancer, 71 (44%) versus 91 (56%), respectively] differs substantially between cancers (Figure 4A) and is the highest in READ (86% versus 14%; $P = 0.001$) and the lowest in PAAD (27% versus 73%; $P = 0.032$). This may suggest a
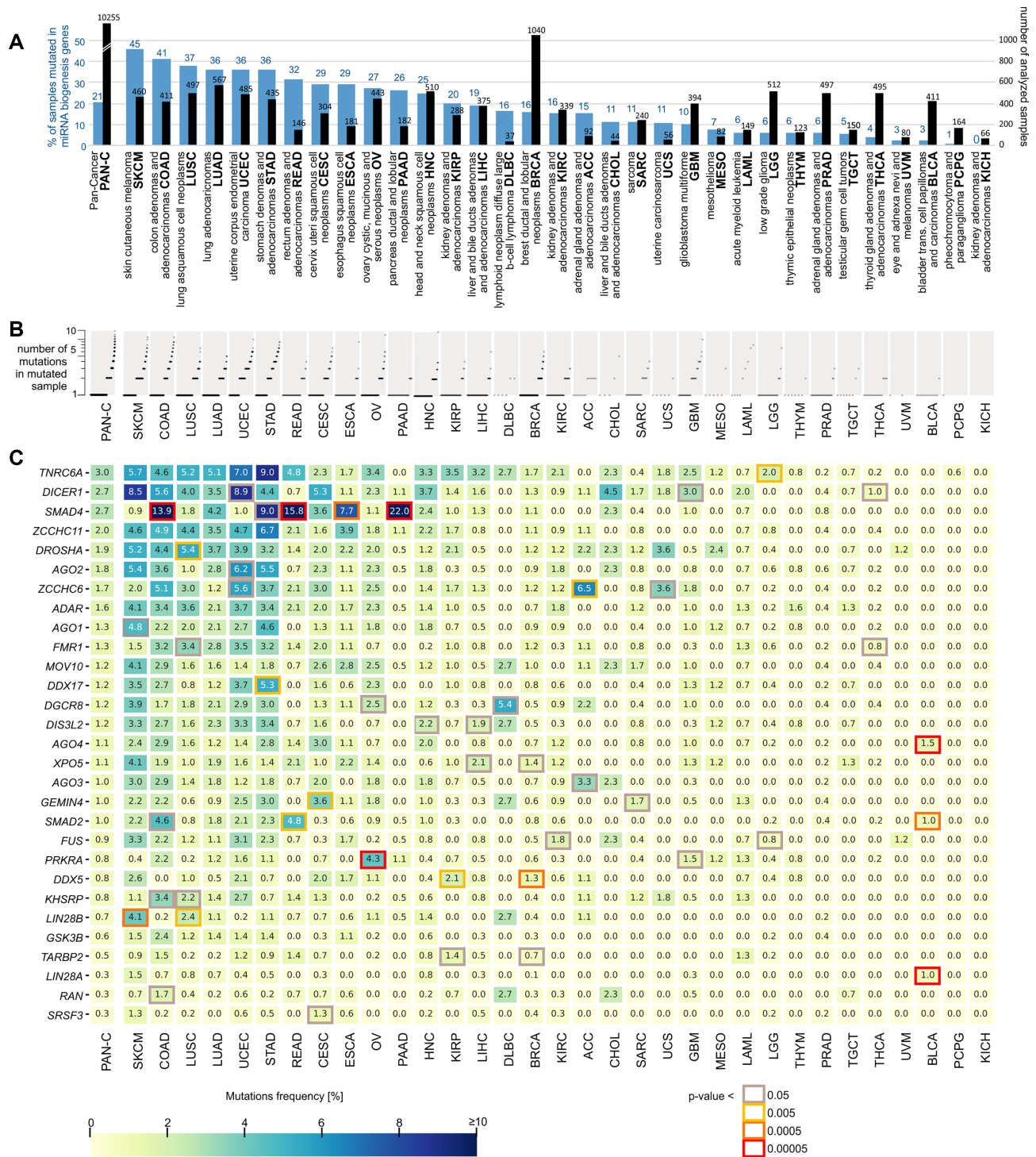
Figure 2. Mutation frequency in miRNA biogenesis genes across the analyzed cancer types. (**A**) The total number of samples (black bars) and percentage of samples with mutations (blue bars) in the panel of miRNA biogenesis genes. (**B**) The proportion of samples with different numbers of mutations. Each sample is shown as a dot (only samples with at least one mutation are shown). Due to the large number of samples with a particular number of mutations, dots overlap with each other. (**C**) Heatmap showing the frequency [%] of mutations within each of the miRNA biogenesis genes (y-axis) in different cancer types (x-axis). The genes are ordered by the frequency of mutations in pan-cancer (first column). The genes significantly overmutated in particular cancer types are marked with a color frame indicating the nominal *P*-value (Fisher's exact test; the *P*-value scale is indicated under the heatmap). A *P*-value <0.00005 (red frame) corresponds to significant enrichment after correction for multiple tests (adjusted *P* <0.05). Specific *P*-values are shown in Supplementary Table S4.
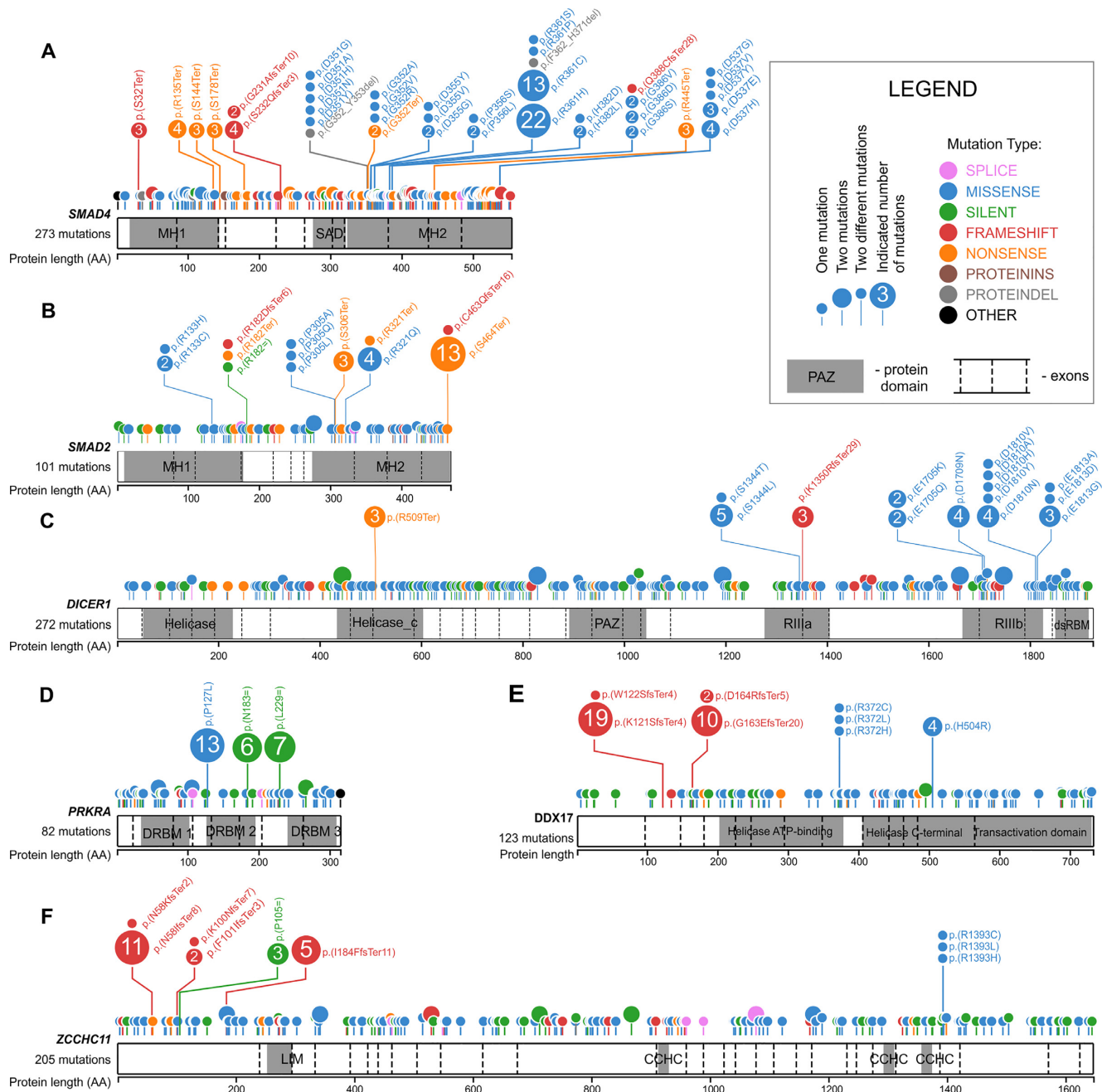
Figure 3. Distribution of the identified mutations in the miRNA biogenesis genes. (**A–F**) depict *SMAD4*, *SMAD2*, *DICER1*, *PRKRA*, *DOX17* and *ZC-CHC11*, respectively. The remaining miRNA biogenesis genes are shown in Supplementary Figure S3. Mutations are visualized in the form of lollipop plots along the gene coding sequences, with the exon structure and protein functional domains indicated. The size of a mutation symbol (circle) is proportional to the number of mutations, and the color indicates the type of mutation (as shown in the legend). All mutations were annotated according to HGVS nomenclature, and the effect of the mutations at the protein level was denoted with the VEP tool (Ensembl).

different role of SMAD4 in different cancers and different effects of these two types of mutations. Most of the hotspots (five of eight hotspots, 83% of all hotspot mutations) coincide with charged (basic or acidic) AAs, affecting the electrostatic properties of MH2 important for interaction with R-SMADs (86). To visualize the location of the hotspot AA residues, we marked them on a crystal structure of the heterotrimeric SMAD4:(SMAD2)$_2$ complex (86). As shown in Figure 4B and Supplementary S4, all hotspot residues

are located on the surfaces of the SMAD4:SMAD2 interaction, which is consistent with the notion that hotspot mutations prevent SMAD complex formation and thus avert transcription of SMAD-controlled genes in response to TGFB/BMP signaling. As in SMAD4, the SMAD2 hotspot residues localize at the SMAD4:SMAD2 interaction surfaces (Figure 4B and Supplementary S4).

The copy number analysis of *SMAD4* showed that *SMAD4* deletions are significantly more frequent in sam-
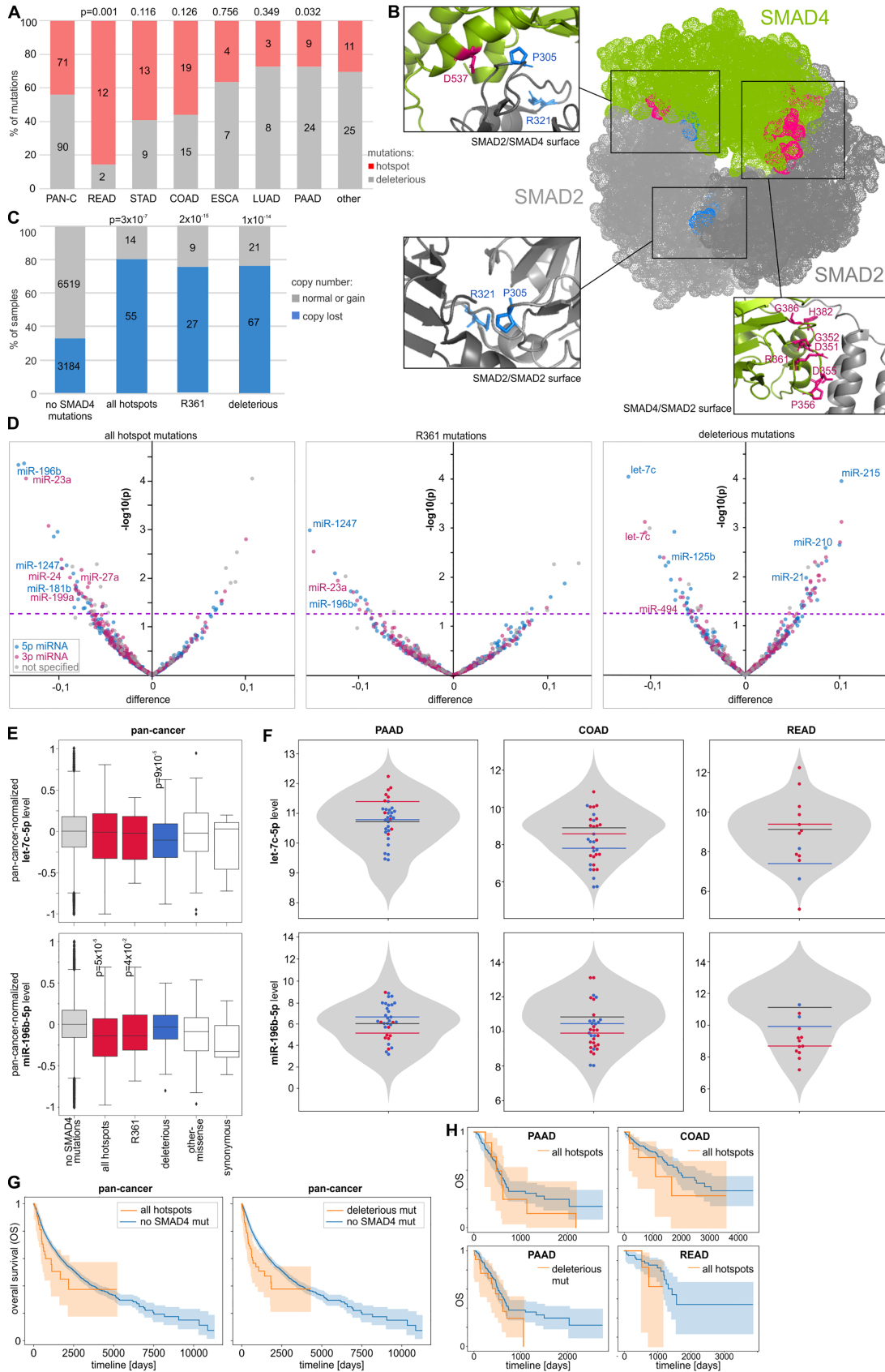
Figure 4. Characteristics of *SMAD4* mutations. (**A**) The proportions of hotspot and deleterious mutations (y-axis) in different cancer types (x-axis). (**B**) Localization of hotspot AA residues in the SMAD4:(SMAD2)$_2$ heterotrimeric complex in SMAD2/SMAD4, SMAD4/SMAD2 and SMAD2/SMAD2

ples with *SMAD4* mutations than in samples without such mutations, indicating frequent loss of the second allele [loss of heterozygosity (LOH) effect] in samples with both hotspot and deleterious mutations (Figure 4C). *SMAD4* mutations were previously shown to affect the expression of many genes. Although it was proposed that the TGFB/SMAD pathway also controls the expression of miRNAs (25,27,28,73,99,100), the effect of the mutations on miRNA levels has never been tested. To do so at the pan-cancer scale, we took into account only 522 miRNAs whose level was >0 in at least 70% of tested TCGA samples. We normalized miRNA levels to make the median (equal to 0) and variance of these levels comparable between cancers. As shown in Figure 4D, there is an excess of downregulated miRNAs in samples with hotspot mutations (e.g. at the level of *P*-value < 0.05, 78% and 22% of miRNAs are downregulated and upregulated, respectively), which is consistent with the expected impairment of SMAD complex formation and transcription factor activity. A similar effect was observed when the analysis was performed only for R361, the most frequently mutated hotspot residue (Figure 4D), and when the analysis was limited to the high-confidence miRNAs (annotated in MirGeneDB; Supplementary Figure S5A), which further strengthens the reliability of the observation. Among the downregulated miRNAs (Supplementary Table S5), there are many miRNAs with well-documented cancer-related functions, for example, the recently discovered suppressormiRs miR-23a-3p (reviewed in (101)), miR-196b-5p (102,103), and miR-1247-5p (104). Noteworthy, such an effect of excessive miRNA downregulation is not visible for deleterious mutations (Figure 4D and Supplementary Figure S5A). Among the most significantly downregulated or upregulated miRNAs in samples with deleterious mutations are let-7c-5p and miR-125b-5p or miR-215-5p, miR-21-5p and miR-210-5p, respectively, all well-known cancer-related miRNAs (e.g. (105)). The frequent observation of pairs of 5p and 3p-miRNAs (e.g. let-7c-5p and let-7c-3p) and groups of miRNAs generated from miRNA clusters (e.g. miR-99a/let-7c or miR-1/133a cluster) among the consistently altered miRNAs in samples with both hotspot and deleterious mutations indicates that at least some of the miRNA alterations result from aberrations at the transcriptional level. As shown in Figure 4E and F, the levels of let-7c-5p and miR-196-5p, which are examples of miRNAs most significantly altered in pan-cancer, often show similar alterations in the most commonly mutated cancers, i.e. PAAD, COAD and READ, although it must be noted that due to the very limited number of mutations, the results in individual cancers are not perfectly consis-

tent and have to be interpreted with caution. Nonetheless, among the miRNAs downregulated by either hotspot or deleterious *SMAD4* mutations are many miRNAs demonstrated before to be activated by TGFB/SMAD signaling. It includes miR-23a-3p, miR-27a and miR-24 constituting a miRNA cluster that is the first and most well-studied group of miRNAs regulated by SMAD4 (74), miR-181b-5p (106), 199a-3p (73) and miR-494-3p (107). To identify pathways/processes enriched in the genes regulated by the downregulated miRNAs, we performed KEGG pathway enrichment analysis with miRPath v3.0. The analysis showed that miRNAs downregulated both by the hotspot and deleterious *SMAD4* mutations are associated (adjusted $P < 0.01$) with similar cancer-related processes, including 'Pathways in cancer', 'ErbB signaling pathway', 'Glioma', and 'Proteoglycans in cancer' (Supplementary Table S6). It is worth noting that among the enriched pathways is also the 'TGF-beta signaling pathway' regulated by 32 of the 49 (65%, $P = 0.0002$) and 20 of the 29 (69%; $P = 0.00005$) miRNAs downregulated by the hotspot and deleterious mutations, respectively.

A comparison of *SMAD4* mutations with clinical cancer data did not show an association of mutations with tumor staging but showed an association with decreased overall survival (OS) at the pan-cancer level (log-rank test: $P = 0.049$ and $P = 0.0006$ for hotspot and deleterious mutations, respectively) (Figure 4G). As the analyses performed at the pan-cancer level may be affected by the unequal distribution of mutations among cancer types, we repeated the survival analysis with the most frequently mutated individual cancers. Although the analyses of individual cancer types were of very limited statistical power because of a low number of mutations, some cancer types also showed trends toward decreased survival of patients with the mutations, i.e. COAD and READ (Figure 4H).

Although *SMAD2*, with a total of 101 mutations, including 35 deleterious mutations, is much less densely mutated (70 mut/kb) than *SMAD4* (Figure 3B), it also contains recurrently mutated hotspot AA residues in the MH2 domain. The hotspots include two AA residues, i.e. P305 and R321 mutated 3 and 4 times, respectively, and the nonsense mutation p.S464Ter truncating the protein by five AAs, which is the most common *SMAD2* mutation (13 occurrences in cancers such as COAD, STAD, and BRCA). As in SMAD4, the SMAD2 hotspot residues localize at the SMAD4:SMAD2 interaction surfaces (Figure 4B and Supplementary Figure S4).

We separately analyzed miRNA level profiles in samples with hotspot missense mutations ($N = 7$), the p.S464Ter

surfaces (as indicated in the figure). Mutated positions in *SMAD4* are indicated in pink, whereas *SMAD2* mutations are indicated in blue. (**C**) The proportion of copy number alterations of *SMAD4* (y-axis) in samples with different types of *SMAD4* mutations. (**D**) Volcano plots depicting miRNA level alterations in samples with different types of *SMAD4* mutations (indicated above the graphs) compared to samples without any *SMAD4* mutations. Positive and negative values on the x-axis indicate increased and decreased miRNA levels in mutated samples, respectively. The y-axis indicates the *P*-value on a $\log_{10}$ scale, and the $P = 0.05$ threshold is indicated by a pink dashed line. The y-axis indicates the $\log_{10}$ *P*-values. Blue, pink, and gray dots indicate 5p-miRNAs, 3p-miRNAs and miRNAs whose arm was not specified, respectively. The IDs of selected miRNAs are indicated on the graphs. (**E**) Boxplots showing the distribution of the selected miRNA levels (y-axes) in samples with different types of mutations versus samples without any *SMAD4* mutations (x-axes). The *P*-values <0.05 are indicated on the graphs. (**F**) Violin plots showing the distribution of non-pan-cancer-normalized levels (y-axis) of the miRNAs shown in E in samples without *SMAD4* mutations and specific samples with different types of *SMAD4* mutations (dots of different colors). Horizontal lines indicate median miRNA levels in samples with different types of mutations. (**G**) and (**H**) Kaplan–Meier plots indicating the OS of patients without and with a specified type of *SMAD4* mutation in pan-cancer and specific cancer types, respectively.

mutation ($N = 13$), and deleterious mutations (not including p.S464Ter; N = 22). As in samples with *SMAD4* mutations, samples with p.S464Ter and deleterious mutations showed a substantial excess of downregulated miR-NAs [95% (34 of 36 altered miRNAs at $P < 0.05$) and 70% (16 of 23), respectively]) (Figure 5A and Supplementary Table S5). A similar effect was also observed in the group of high-confidence miRNAs (Supplementary Figure S5B). The excess of downregulations is not visible for the hotspot missense mutations; however, it must be noted that due to a very low number of hotspot mutations, the analysis of miRNA levels for this type of mutation is of very low statistical power. Nonetheless, we observed one miRNA, miR-329-3p, playing a role in different cancers (108–111)), which was consequently downregulated in all three mutational groups. Other downregulated miRNAs playing a role in cancer include (i) miR-1247-5p and miR-1247-3p (in samples with p.S464Ter), recently identified as tumor suppressor miRNAs (104), and observed as highly expressed in embryonic/placenta cells which may indicate their role in quickly dividing cells (112); (ii) miR-7-5p (in samples with deleterious mutations), a well-recognized suppressormiR, playing a role in downregulation of the growth, metastasis, and prognosis of various tumors (reviewed in (113)); and (iii) miR-380-5p (in samples with hotspot missense mutations), downregulating TP53 to control cellular survival (114). As shown in Figure 5B and C, the examples of miR-329-3p and miR-380-5p (both members of the MIR-154 family playing a role in pulmonary fibrosis and being a target of TGFB signaling (115)) illustrate that miRNA level changes identified in the pan-cancer analysis are also reflected in the most frequently mutated cancers, i.e. COAD and READ. SMAD2, similar to other R-SMADs, may also posttranscriptionally increase the level of a specific group of miRNAs, facilitating the processing of their pri-miRNAs by DROSHA (72). Among the 44 miRNAs shown to be upregulated by R-SMADs or containing specific R-SMAD-interacting sequence motifs (23 of them tested in this study), miR-421 (a member of the MIR-95 family) ($P = 0.01$), miR-188-5p ($P = 0.03$), and miR-877-5p ($P = 0.04$) were downregulated in samples with the p.S464Ter mutation. A comparison of the SMAD2 mutations with clinical characteristics of cancers did not reveal an association of mutations with cancer stages, but p.S464Ter showed a trend toward decreased OS of patients with the mutations (Figure 5D) in pan-cancer and in the individual cancers with >1 p.S464Ter occurrence in informative samples.

As shown in Figure 5E, although there is some overlap between miRNAs altered by the hotspot missense mutations in *SMAD2* and *SMAD4*, generally the overlaps between the effects of different mutation types are small. The small overlap, may mainly result from the low statistical power of the analysis and the fact that different mutations predominate in different cancer types (with different miRNA profiles), but it may also result from various involvement of the proteins in distinct regulatory processes, i.e. (i) SMAD4 plays a role in the activation of miRNA transcription and in this process it may interact with different R-SMADs (not only SMAD2; different mutations may differentially affect interactions with different R-SMADs) whereas (ii) SMAD2 (but not SMAD4) plays a role in post-

transcriptional regulation of miRNA precursors. Additionally, secondary regulatory processes may also play a role.

Functional consequences of *DICER1* mutations

Another highly mutated gene with several characteristic hotspots is *DICER1,* with a total of 272 mutations (46 mut/kb), including 38 deleterious mutations and 27 missense mutations located in the aforementioned hotspots in the RIIIa ($N = 6$) and RIIIb ($N = 21$) domains (Figure 3C). As shown in Figure 6A, the occurrence of DICER1 mutations does not correlate with gene copy number alterations, and only a small fraction of the *DICER1* mutations coincide with gene deletion or amplification. The miRNA expression analysis at the pan-cancer level showed that the hotspot mutations in both the RIIIa and RIIIb domains were associated with the global downregulation of 5p-miRNAs with simultaneous upregulation of 3p-miRNAs (Figure 6B). This effect of 5p-miRNA downregulation was observed before and was explained by the fact that the hotspot mutations in RIIIb but also in RIIIa (65) affect the activity of the RIIIb domain, preventing cleavage of miRNA precursors at the 5p arm and the release of 5p-miRNAs. The increase in 3p-miRNAs was considered an artifact, i.e. an apparent effect counterbalancing the global decrease in 5-miRNAs, resulting from the standard miRNA level normalization procedure that standardized the amount of each miRNA against the total number of miRNA reads. Surprisingly, we observed a similar effect of decreased 5p-miRNAs and increased 3p-miRNAs in samples with deleterious mutations that do not affect RIIIb but are assumed to lead to complete loss of DICER1 (Figure 6 B and C; for comparison, please see the effect of *SMAD4* mutations in which 5p- and 3p-miRNAs are more or less equally distributed between the decreased and increased miRNAs). The effect of the asymmetrical distribution of altered miRNAs was not observed for other nonhotspot missense mutations outside the RNase domains and synonymous mutations for which, as expected, miRNA level changes were very low (Figure 6C). To avoid potential biases associated with pan-cancer normalization, we performed expression analysis separately for UCEC, which is the cancer type with the most frequent mutations in *DICER1* (∼9%), with 2 mutations in RIIIa, 10 mutations in RIIIb and 9 deleterious mutations. Although of lower statistical power, the UCEC analysis showed a similar effect of the hotspot and deleterious mutations on globally decreased levels of 5p-miRNAs and increased levels of 3p-miRNAs (Figure 6B). To directly check whether the increase in 3p-miRNAs in samples with the *DICER1* mutations is an effect of normalization against the total number of miRNA-specific reads, we normalized the miRNA levels against the level of miR-451a used as a reference gene. miR-451a is a miRNA whose biogenesis is not dependent on DICER1 processing (116); therefore, its level should not be affected by DICER1 mutations. As shown in Figure 6C, normalization against miR-451a in samples with hotspot mutations abolished the effects of neither 5p-miRNA decreases nor 3p-miRNA increases. A consistent effect was observed when the analysis was performed on RPM (non-batch-effects normalized) miRNA levels and when the analysis was narrowed down to the high-confidence miRNAs
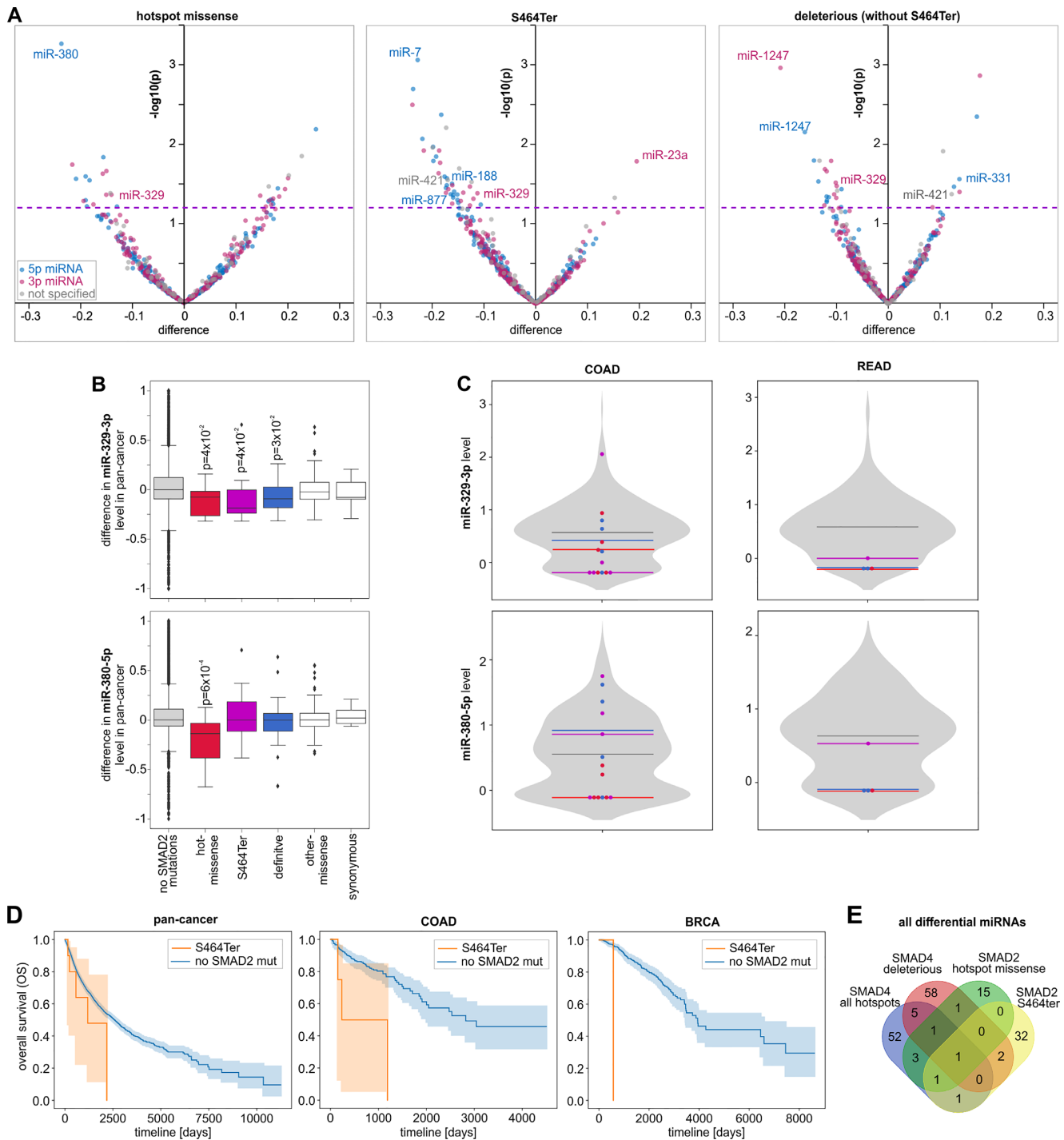
Figure 5. Characteristics of SMAD2 mutations. (**A**) Volcano plots depicting miRNA level alterations in samples with the particular types of *SMAD2* mutations (indicated above the graphs) compared to samples without any *SMAD2* mutations (graph scheme as in Figure 4D). (**B**) Boxplots showing the distribution of the selected miRNA levels (x-axes) in samples with different types of *SMAD2* mutations vs. samples without *SMAD2* mutations. (**C**) Violin plots showing the distribution of non-pan-cancer-normalized levels (y-axis) of the miRNAs shown in B in samples without *SMAD2* mutations and specific samples with different types of *SMAD2* mutations (graph scheme as in Figure 4F). (**D**) Kaplan–Meier plots showing the OS of patients with p.S464Ter and without any SMAD2 mutations in pan-cancer, COAD and BRCA. (**E**) Venn diagram showing the overlap between miRNAs altered by the indicated types of SMAD2 and SMAD4 mutations.
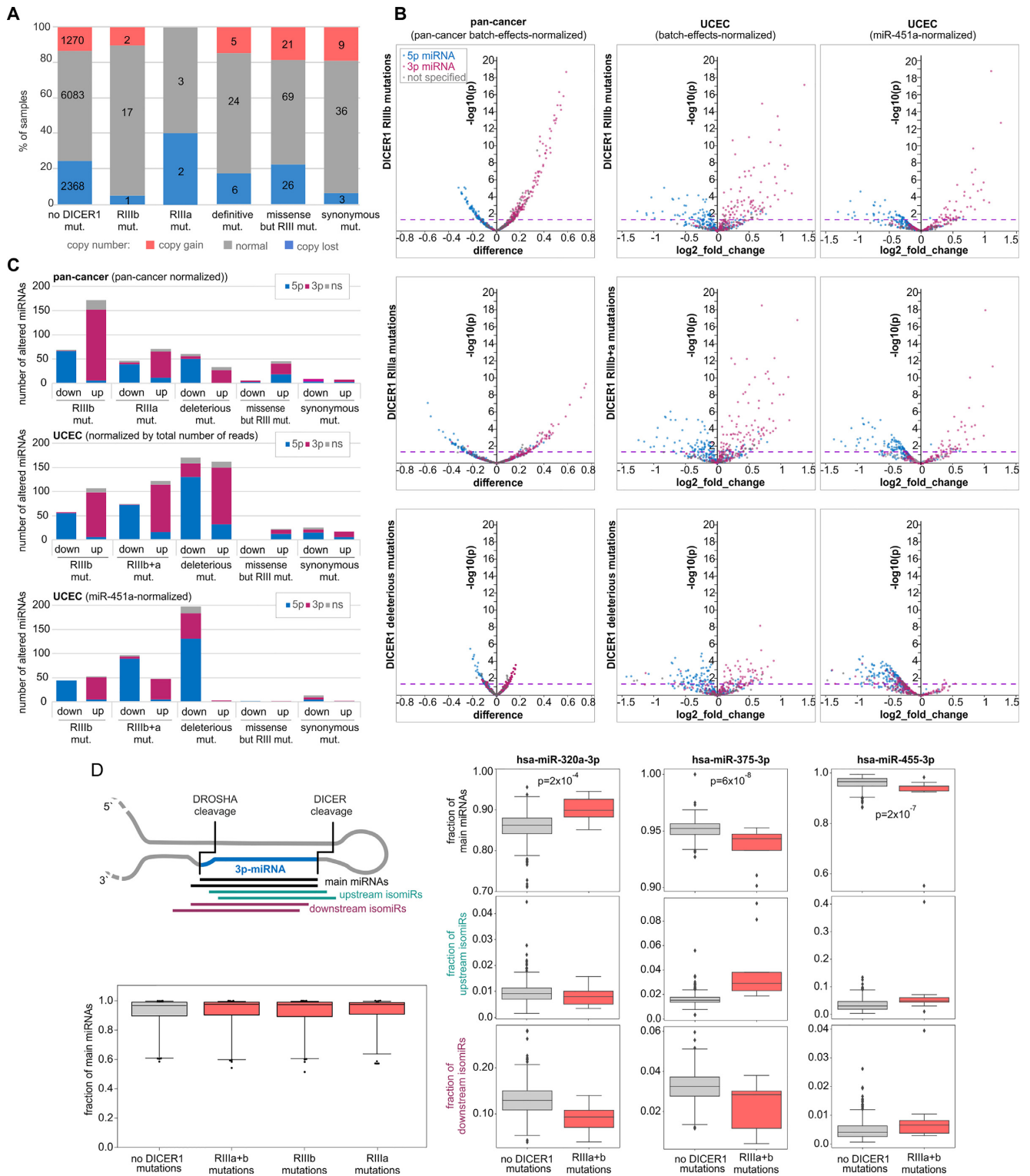
Figure 6. Functional consequences of *DICER1* mutations. (**A**) The proportion of copy number alterations of *DICER1* (y-axis) in samples with different types of *DICER1* mutations. (**B**) Volcano plots depicting miRNA level alterations in pan-cancer (first column) and UCEC before and after normalization against the miR-451a level (second and third columns, respectively) in samples with different types of *DICER1* mutations (indicated on the left). Blue, pink, and gray dots indicate 5p-miRNAs, 3p-miRNAs, and miRNAs whose arm was not specified, respectively. Other graph details as in Figure 4D. (**C**) Each graph shows the proportions of 3p- and 5p-miRNAs (y-axis) among the miRNAs downregulated and upregulated in samples with specified types of *DICER1* mutations (x-axis; at $P < 0.05$). The graphs (from the top) show the alterations in pan-cancer, and UCEC before and after normalization against the miR-451a level. (**D**) Above; a schematic representation of the analyzed isomiR fractions. Below; a box plot showing distributions of the main miRNA fractions of 49 3p-miRNAs highly expressed in UCEC (y-axis) in samples with the RIIIa and RIIIb mutations vs. samples with no *DICER1* mutations (x-axis). On the right; examples of miRNAs whose main fraction was altered in samples with the *DICER1* mutations, against changes in the levels of upstream and downstream isomiRs (please note the different scale in the graphs).

(Supplementary Table S5 and Supplementary Figure S5C and D, respectively). This indicates that hotspot mutations, as reported previously (60–62,65,94), decrease 5p-miRNA levels, but contrary to previous reports, they also increase the levels of numerous 3p-miRNAs. KEGG pathway enrichment analysis (Supplementary Table S6) showed that upregulated 3p-miRNAs, among others, are strongly associated with various cancer-related processes, such as 'TGF-beta signaling pathway', 'Pathways in cancer', 'ErbB signaling pathway' and 'Hippo signaling pathway', which were found among the top ten most significant associations (adjusted $P <0.00001$). Consistent with the predicted loss-of-function effect of the deleterious mutations, after normalization against miR-451a, the deleterious mutations are associated almost exclusively with a decrease in miRNA levels (Figure 6C). Although an excess of 5p-miRNAs is observed among the decreased miRNAs, there is also a substantial fraction ($n = 52$, 26%) of 3p-miRNAs.

Finally, we checked whether the *DICER1* hotspot mutations affect the 5′end heterogeneity of 3p-miRNAs and proportions of generated isomiRs. The heterogeneity of many miRNAs has been documented before (117–120) and it was shown that the proportion of isomiRs may differ substantially in different conditions, including cancer (117,121). To reliably estimate fractions of isomiRs, for analysis we took only 3p-miRNAs ($n = 132$) undergoing high expression ($\geq$100 RPM) in UCEC. For each of the miRNAs, we compared the level of their isoforms with the most commonly generated 5′end (main miRNAs) versus level of minor (either upstream or downstream) isomiRs. As shown in Figure 6D the mutations do not affect the global level of main miRNAs, which fraction among the analyzed miRNAs ranged from ~0.6 to 1 with a median in all cases >0.95. However, in some of the individual miRNAs, the level of major miRNAs is significantly altered (mostly decreased) in samples with the *DICER1* hotspot mutations (Supplementary Table S8 and Figure 6D (right panel)). Although there is striking concordance between the effects of the mutations in RIIIa and RIIIb (Supplementary Table S8) the overall changes in fractions of the main miRNAs are very small. It is unlikely that such small changes in fractions of main miRNAs may cause any meaningful biological effect, however that may result in more substantial changes in some of the minor isomiRs (Figure 6D (right panel)).

The lists of miRNAs differentiated in pan-cancer and UCEC with different types of *DICER1* mutations are shown in Supplementary Table S5, it includes a substantial number of the high-confidence miRNAs. Among the altered miRNAs are many miRNAs whose function is well recognized in cancer and many upregulated 3p-miRNAs are passengers of such miRNAs, examples of which are listed in Supplementary Table S7. A comparison of the *DICER1* mutations with clinical characteristics of cancers did not reveal any significant associations.

DISCUSSION

In this study, we provide a comprehensive pan-cancer analysis of somatic mutagenesis in a panel of 29 miRNA biogenesis genes in 33 adult-onset cancers. In total, we identified 3,649 mutations, composed of 60% missense and 17%

deleterious mutations. Overall, ~21% of the samples had at least one mutation, but the percentage differed substantially between the cancer types, reaching >40% in SKCM or COAD. The most frequently mutated genes were *TNRC6A*, *DICER1*, *SMAD4* and *ZCCHC11*, with mutations in 2–3% of pan-cancer samples. The frequency of mutations in the miRNA biogenesis genes is similar to that in cancer-specific drivers, such as *MYC*, *ALK*, *CACNA1A*, *POLE*, *BCL2*, *NOTCH2*, *MET*, *HRAS*, *FGFR* and *PIK3R1,* with pan-cancer mutation frequencies up to ~3% (122). Consistent with the potential cancer-specific role of the genes, some of them are much more frequently mutated in particular cancers, for example, *SMAD4* in PAAD (22%), READ (16%), COAD (14%), and ESCA (8%); *DICER1* in SKCM and UCEC (~9%); *TNRC6A* in STAD (9%) and UCEC (7%); ZCCHC6 in ACC (6.5%); *SMAD2* in COAD and READ (~5%); and *PRKRA* in OV (4.3%). The frequencies of mutations in some of the genes are specifically and significantly enriched in particular cancer types.

The high frequency of mutations in *SMAD4* is consistent with the findings of many previous studies in which *SMAD4* was analyzed as a transcription factor playing a role in the activation of many cancer-related genes in response to TGFB/BMP signaling (123–126). Comprehensive analysis of a large number of samples allowed us to define a profile of *SMAD4* mutations in most common human cancers covered by TCGA. It confirmed the high frequency of *SMAD4* mutations in cancers such as PAAD, COAD and READ but also revealed a relatively high frequency of the mutations in STAD, LUAD and ESCA and lower frequencies in many other cancers (e.g. CESC, LUSC, HNC and CHOL) that are much less studied in the context of *SMAD4* deficiency. With the large collection of mutations, we have identified 8 hotspot residues in the MH2 domain, including some not reported before, and revealed that the proportions of hotspot and deleterious *SMAD4* mutations differ substantially between cancers, with the fraction of hotspot mutations significantly enriched in READ (86%) and the lowest in PAAD (27%). This may suggest different mechanisms and effects of SMAD4 inactivation in different cancer types. We showed that most of the hotspots localize in residues of charged AAs, which may affect the electrostatic interactions of the MH2 domain with corresponding domains in R-SMADs (e.g. SMAD2) critical for SMAD complex formation (86). Such an effect of both loss-of-charge and gain-of-charge *SMAD4* mutations was computationally predicted in a previous study (127). Consistently, we showed that all the *SMAD4* hotspots as well as much less frequently mutated hotspot residues in the corresponding MH2 domain in SMAD2 localize at surfaces of the SMAD4:SMAD2 interaction, confirming the role of hotspot mutations in destabilizing the SMAD complex and hence revealing an inhibitory role of these mutations in downstream TGFB/BMP signaling. We showed that both the deleterious and hotspot *SMAD4* mutations frequently (~80%) coincide with gene deletions, which indicates inactivation of the second allele and is consistent with previous observations of frequent LOH in the region (128–131). Although it was shown that SMAD4 plays a role in the activation of miRNA gene transcription (summarized in (132)), the effect of *SMAD4* mutations on miRNA expression

has never been tested. We identified numerous miRNAs (predominantly high-confidence miRNAs) differentially expressed in samples with *SMAD4* mutations and showed different patterns of miRNA level changes induced by hotspot and deleterious mutations. While hotspot mutations predominantly downregulate miRNAs, deleterious mutations both increase and decrease miRNA levels. This finding is consistent with the results obtained for *SMAD4* knockdown by RNAi in hepatic stellate cells (73) and again shows a different effect of deleterious mutations that, as a result of NMD, most likely leads to complete loss of SMAD4 and hotspot mutations that modify SMAD4 structure, affecting its interactions with R-SMADs and other proteins. Therefore, the effect of hotspot mutations may result not only from impeded R-SMADs:SMAD4 complex formation but also from a shifted balance in the binding of competing coactivators and corepressors (such as p300/CBP and Ski or SnoN, respectively), whose contribution determines the outcome of signaling events (reviewed in (125,133)).

Similar effects of predominant downregulation of miRNA levels were observed for p.S464Ter, the most frequent *SMAD2* hotspot mutation, as well as for deleterious *SMAD2* mutations. The p.S464Ter mutation is localized in the last exon of *SMAD2* and truncates the last 5 AAs of the protein but most likely does not activate NMD. As the truncated fragment is important for efficient SMAD2 phosphorylation and includes two serine residues (S465 and S467) whose phosphorylation upon TGFB signaling is critical for SMAD4(SMAD2)$_2$ complex formation (93,134,135), the mutation precludes the SMAD complex formation and TGFB/BMP signaling (123). We showed that the mutation is associated with decreased survival of cancer patients. Of note, two BRCA patients with the mutations showed a strikingly short OS. Regardless of its role in complexes with SMAD4, SMAD2 (along with other R-SMADs) may directly interact with a specific set of miRNA precursors, accelerating their processing by DROSHA (72,136). This posttranscriptional regulation of miRNA processing may also be affected by SMAD2 mutations, as the MH2 domain of R-SMADs was shown to interact with the P68 RNA helicase participating in the recruitment of DROSHA and DGCR8 to pri-miRNAs (137).

Another gene with characteristic hotspot mutations is *DICER1*. We found 21 hotspot mutations affecting 4 metal-ion-binding AA residues in the RIIIb domain (E1705, D1709, D1810 and E1813) and 6 hotspot mutations in one AA residue (S1344) in RIIIa. These hotspots were previously detected and functionally characterized in various pediatric cancers, including cancers associated with DICER1 syndrome (e.g. pleuropulmonary blastoma) and Wilms' tumor, as summarized in (54,64). More recently, they were also investigated in thyroid adenomas (94) and the TCGA cohort, mostly in UCEC samples (65). It was shown that RIIIb hotspot mutations affect the RIIIb cleavage of the 5p-arm of pre-miRNAs, resulting in inefficient production of 5p-miRNAs (60–62,94). In a very recent study, it was also shown that the hotspot mutations of the S1344 residue, although located in RIIIa, spatially interfere with RIIIb, resulting in the same effect as that observed for the mutations in RIIIb (65). The miRNA profiling performed in our

study showed a global decrease in 5p-miRNAs and a global increase in 3p-miRNAs in samples with both hotspot and deleterious *DICER1* mutations. A similar effect has been observed before (61,62,65,94,138,139). The global decrease in 5p-miRNAs has been interpreted as the result of inefficient 5-miRNA processing, but the increase in 3p-miRNAs was considered an artifact of miRNA level normalization (against the total number of miRNA reads) that, as a reflection of the global 5p-miRNA deficit, resulted in a relative increase in unaffected 3p-miRNAs (65). To eliminate this potential bias, we normalized the miRNA levels against the level of miR-451a, which is a DICER1-independent miRNA whose level is not affected by DICER1 deficiency (116). However, normalization against the miR-451a level did not eliminate the asymmetrical effect of the hotspot mutations also visible in the high-confidence miRNAs, showing that both the 5p-miRNA decreases and 3p-miRNA increases are real. We have additionally shown that the increased 3p-miRNAs strongly associate with cancer-related terms/pathways and include many miRNAs well recognized in cancer (Supplementary Table S6). The increase in 3p-miRNAs may result from a lack of competition with 5p-miRNAs during transfer to and loading onto the RISC, observed as 5p/3p arm shifting or switching (140). However, this would require some alternative mechanism of releasing 3p-miRNAs from partially processed (nicked only at 3p-arms) pre-miRNAs and transferring them to the RISC, bypassing the miRNA duplex stage. Although it was previously speculated that AGO2 may play some role in such a process (19,20), in our opinion, this step warrants further investigation. Unlike the hotspot mutations, the deleterious mutations (after normalization against miR-451a) almost exclusively decrease miRNA levels, which confirms their loss-of-function nature. Although the effect of deleterious mutations is more profound for 5p-miRNAs, the mutations also affect 3p-miRNAs (∼30%).

With exception of hotspot mutations found in *DICER1*, we found only a few mutations in other miRNA biogenesis genes, i.e. *DROSHA*, *DGCR8* and *XPO5*, which have been recently observed in different childhood cancers associated with DICER1 syndrome and in Wilms' tumor (56–59,66). This may indicate specific functions of the mutations/genes characteristic of childhood cancers but not playing a role in adult-onset cancers. We also excluded the frequent occurrence of specific indel hotspots in *TRBP* and *XPO5* previously reported at high frequencies in cancers associated with MSI (67,68).

In summary, in this study, we present a comprehensive pan-cancer analysis of somatic mutations accumulated in genes involved in miRNA biogenesis and function. We showed that some of these genes are specifically mutated in particular cancers. We identified many hotspot mutations, including some recurring in specific cancer types. We extended knowledge about the types and distribution of *SMAD4* mutations and showed their effect on the expression of miRNA genes. We also showed that all hotspot mutations in *SMAD4* and *SMAD2* affect AA residues located at the surface of the SMAD4:SMAD2 interaction. Moreover, we distinguished and further characterized the effects of deleterious and hotspot missense mutations in *DICER1,* among others, showing that hotspot mutations in the RI-

IIa and RIIIb domains not only decrease the levels of 5p-miRNAs but also increase the levels of 3p-miRNAs. As a result, we have identified numerous miRNAs that are significantly increased or decreased in samples with particular mutation types, including many well-known cancer-related miRNAs. We also linked some of the mutations with the patients' clinical outcomes. Furthermore, we created a compendium of information presented as an atlas and maps of mutations in miRNA biogenesis genes that may be useful resources of information for studying a particular gene or cancer type.

DATA AVAILABILITY

The results published here are based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. The set of in-house Python scripts is available at (https://github.com/martynaut/mirnaome_somatic_mutations).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Calin,G.A., Dumitru,C.D., Shimizu,M., Bichi,R., Zupo,S., Noch,E., Aldler,H., Rattan,S., Keating,M., Rai,K. *et al.* (2002) Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 15524–15529.
2. Peng,Y. and Croce,C.M. (2016) The role of MicroRNAs in human cancer. *Signal Transduct. Target. Ther.*, **1**, 15004.
3. Lu,J., Getz,G., Miska,E.A., Alvarez-Saavedra,E., Lamb,J., Peck,D., Sweet-Cordero,A., Ebert,B.L., Mak,R.H., Ferrando,A.A. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.
4. Volinia,S., Calin,G.A., Liu,C.-G., Ambs,S., Cimmino,A., Petrocca,F., Visone,R., Iorio,M., Roldo,C., Ferracin,M. *et al.* (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 2257–2261.
5. Esquela-Kerscher,A. and Slack,F.J. (2006) Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer*, **6**, 259–269.
6. Tan,W., Liu,B., Qu,S., Liang,G., Luo,W. and Gong,C. (2018) MicroRNAs and cancer: key paradigms in molecular therapy. *Oncol. Lett.*, **15**, 2735–2742.
7. Krutovskikh,V.A. and Herceg,Z. (2010) Oncogenic microRNAs (OncomiRs) as a new class of cancer biomarkers. *BioEssays News Rev. Mol. Cell. Dev. Biol.*, **32**, 894–904.
8. Kasinski,A.L. and Slack,F.J. (2011) Epigenetics and genetics. MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Nat. Rev. Cancer*, **11**, 849–864.
9. Cui,M., Wang,H., Yao,X., Zhang,D., Xie,Y., Cui,R. and Zhang,X. (2019) Circulating microRNAs in cancer: potential and challenge. *Front. Genet.*, **10**, 626.
10. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
11. Ha,M. and Kim,V.N. (2014) Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.*, **15**, 509–524.
12. Kim,V.N., Han,J. and Siomi,M.C. (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, **10**, 126–139.
13. Bartel,D.P. (2018) Metazoan microRNAs. *Cell*, **173**, 20–51.
14. Michlewski,G. and Cáceres,J.F. (2019) Post-transcriptional control of miRNA biogenesis. *RNA N. Y. N*, **25**, 1–16.
15. Treiber,T., Treiber,N. and Meister,G. (2019) Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. *Nat. Rev. Mol. Cell Biol.*, **20**, 5–20.
16. Ustianenko,D., Hrossova,D., Potesil,D., Chalupnikova,K., Hrazdilova,K., Pachernik,J., Cetkovska,K., Uldrijan,S., Zdrahal,Z. and Vanacova,S. (2013) Mammalian DIS3L2 exoribonuclease targets the uridylated precursors of let-7 miRNAs. *RNA N. Y. N*, **19**, 1632–1638.
17. Heo,I., Joo,C., Kim,Y.-K., Ha,M., Yoon,M.-J., Cho,J., Yeom,K.-H., Han,J. and Kim,V.N. (2009) TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell*, **138**, 696–708.
18. Sibley,C.R., Seow,Y., Saayman,S., Dijkstra,K.K., El Andaloussi,S., Weinberg,M.S. and Wood,M.J.A. (2012) The biogenesis and characterization of mammalian microRNAs of mirtron origin. *Nucleic Acids Res.*, **40**, 438–448.
19. Cheloufi,S., Dos Santos,C.O., Chong,M.M.W. and Hannon,G.J. (2010) A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature*, **465**, 584–589.
20. Cifuentes,D., Xue,H., Taylor,D.W., Patnode,H., Mishima,Y., Cheloufi,S., Ma,E., Mane,S., Hannon,G.J., Lawson,N.D. *et al.* (2010) A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science*, **328**, 1694–1698.
21. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–73.
22. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
23. Chou,C.-H., Shrestha,S., Yang,C.-D., Chang,N.-W., Lin,Y.-L., Liao,K.-W., Huang,W.-C., Sun,T.-H., Tu,S.-J., Lee,W.-H. *et al.* (2018) miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **46**, D296–D302.
24. Friedman,R.C., Farh,K.K.-H., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
25. Kong,W., Yang,H., He,L., Zhao,J., Coppola,D., Dalton,W.S. and Cheng,J.Q. (2008) MicroRNA-155 is regulated by the transforming growth factor beta/Smad pathway and contributes to epithelial cell plasticity by targeting RhoA. *Mol. Cell. Biol.*, **28**, 6773–6784.
26. Tong,Z., Cui,Q., Wang,J. and Zhou,Y. (2019) TransmiR v2.0: an updated transcription factor-microRNA regulation database. *Nucleic Acids Res.*, **47**, D253–D258.
27. Kato,M., Putta,S., Wang,M., Yuan,H., Lanting,L., Nair,I., Gunn,A., Nakagawa,Y., Shimano,H., Todorov,I. *et al.* (2009) TGF-beta activates Akt kinase through a microRNA-dependent amplifying circuit targeting PTEN. *Nat. Cell Biol.*, **11**, 881–889.

28. Chung,A.C.K., Huang,X.R., Meng,X. and Lan,H.Y. (2010) miR-192 mediates TGF-beta/Smad3-driven renal fibrosis. *J. Am. Soc. Nephrol. JASN*, **21**, 1317–1325.

29. Czubak,K., Lewandowska,M.A., Klonowska,K., Roszkowski,K., Kowalewski,J., Figlerowicz,M. and Kozlowski,P. (2015) High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer. *Oncotarget*, **6**, 23399–23416.

30. Calin,G.A., Sevignani,C., Dumitru,C.D., Hyslop,T., Noch,E., Yendamuri,S., Shimizu,M., Rattan,S., Bullrich,F., Negrini,M. *et al.* (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 2999–3004.

31. Zhang,L., Huang,J., Yang,N., Greshock,J., Megraw,M.S., Giannakakis,A., Liang,S., Naylor,T.L., Barchetti,A., Ward,M.R. *et al.* (2006) microRNAs exhibit high frequency genomic alterations in human cancer. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 9136–9141.

32. Fazi,F., Racanicchi,S., Zardo,G., Starnes,L.M., Mancini,M., Travaglini,L., Diverio,D., Ammatuna,E., Cimino,G., Lo-Coco,F. *et al.* (2007) Epigenetic silencing of the myelopoiesis regulator microRNA-223 by the AML1/ETO oncoprotein. *Cancer Cell*, **12**, 457–466.

33. Han,L., Witmer,P.D., Casey,E., Valle,D. and Sukumar,S. (2007) DNA methylation regulates MicroRNA expression. *Cancer Biol. Ther.*, **6**, 1284–1288.

34. Saito,Y. and Jones,P.A. (2006) Epigenetic activation of tumor suppressor microRNAs in human cancer cells. *Cell Cycle Georget. Tex*, **5**, 2220–2222.

35. Ali Syeda,Z., Langden,S.S.S., Munkhzul,C., Lee,M. and Song,S.J. (2020) Regulatory mechanism of MicroRNA expression in cancer. *Int. J. Mol. Sci.*, **21**, 1723.

36. Lujambio,A., Calin,G.A., Villanueva,A., Ropero,S., Sánchez-Céspedes,M., Blanco,D., Montuenga,L.M., Rossi,S., Nicoloso,M.S., Faller,W.J. *et al.* (2008) A microRNA DNA methylation signature for human cancer metastasis. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 13556–13561.

37. Lujambio,A. and Esteller,M. (2009) How epigenetics can explain human metastasis: a new role for microRNAs. *Cell Cycle Georget. Tex*, **8**, 377–382.

38. Glaich,O., Parikh,S., Bell,R.E., Mekahel,K., Donyo,M., Leader,Y., Shayevitch,R., Sheinboim,D., Yannai,S., Hollander,D. *et al.* (2019) DNA methylation directs microRNA biogenesis in mammalian cells. *Nat. Commun.*, **10**, 5657.

39. Alarcón,C.R., Lee,H., Goodarzi,H., Halberg,N. and Tavazoie,S.F. (2015) N6-methyl-adenosine (m6A) marks primary microRNAs for processing. *Nature*, **519**, 482–485.

40. Konno,M., Koseki,J., Asai,A., Yamagata,A., Shimamura,T., Motooka,D., Okuzaki,D., Kawamoto,K., Mizushima,T., Eguchi,H. *et al.* (2019) Distinct methylation levels of mature microRNAs in gastrointestinal cancers. *Nat. Commun.*, **10**, 3888.

41. Saunders,M.A., Liang,H. and Li,W.-H. (2007) Human polymorphism at microRNAs and microRNA target sites. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 3300–3305.

42. Marcinkowska,M., Szymanski,M., Krzyzosiak,W.J. and Kozlowski,P. (2011) Copy number variation of microRNA genes in the human genome. *BMC Genomics*, **12**, 183.

43. Króliczewski,J., Sobolewska,A., Lejnowski,D., Collawn,J.F. and Bartoszewski,R. (2018) microRNA single polynucleotide polymorphism influences on microRNA biogenesis and mRNA target specificity. *Gene*, **640**, 66–72.

44. Sun,G., Yan,J., Noltner,K., Feng,J., Li,H., Sarkis,D.A., Sommer,S.S. and Rossi,J.J. (2009) SNPs in human miRNA genes affect biogenesis and function. *RNA N. Y.*, **15**, 1640–1651.

45. Galka-Marciniak,P., Urbanek-Trzeciak,M.O., Nawrocka,P.M., Dutkiewicz,A., Giefing,M., Lewandowska,M.A. and Kozlowski,P. (2019) Somatic mutations in miRNA genes in lung cancer-potential functional consequences of non-coding sequence variants. *Cancers*, **11**, 793.

46. Oak,N., Ghosh,R., Huang,K.-L., Wheeler,D.A., Ding,L. and Plon,S.E. (2019) Framework for microRNA variant annotation and prioritization using human population and disease datasets. *Hum. Mutat.*, **40**, 73–89.

47. Bhattacharya,A. and Cui,Y. (2016) SomamiR 2.0: a database of cancer somatic mutations altering microRNA-ceRNA interactions. *Nucleic Acids Res.*, **44**, D1005–D1010.

48. Hornshøj,H., Nielsen,M.M., Sinnott-Armstrong,N.A., Świtnicki,M.P., Juul,M., Madsen,T., Sallari,R., Kellis,M., Ørntoft,T., Hobolth,A. *et al.* (2018) Pan-cancer screen for mutations in non-coding elements with conservation and cancer specificity reveals correlations with expression and survival. *NPJ Genomic Med.*, **3**, 1.

49. Urbanek-Trzeciak,M.O., Galka-Marciniak,P., Nawrocka,P.M., Kowal,E., Szwec,S., Giefing,M. and Kozlowski,P. (2020) Pan-cancer analysis of somatic mutations in miRNA genes. *EBioMedicine*, **61**, 103051.

50. Tay,Y., Rinn,J. and Pandolfi,P.P. (2014) The multilayered complexity of ceRNA crosstalk and competition. *Nature*, **505**, 344–352.

51. Poliseno,L., Salmena,L., Zhang,J., Carver,B., Haveman,W.J. and Pandolfi,P.P. (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, **465**, 1033–1038.

52. Joyce,B.T., Zheng,Y., Zhang,Z., Liu,L., Kocherginsky,M., Murphy,R., Achenbach,C.J., Musa,J., Wehbe,F., Just,A. *et al.* (2018) miRNA-Processing gene methylation and cancer risk. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.*, **27**, 550–557.

53. Merritt,W.M., Lin,Y.G., Han,L.Y., Kamat,A.A., Spannuth,W.A., Schmandt,R., Urbauer,D., Pennacchio,L.A., Cheng,J.-F., Nick,A.M. *et al.* (2008) Dicer, drosha, and outcomes in patients with ovarian cancer. *N. Engl. J. Med.*, **359**, 2641–2650.

54. Foulkes,W.D., Priest,J.R. and Duchaine,T.F. (2014) DICER1: mutations, microRNAs and mechanisms. *Nat. Rev. Cancer*, **14**, 662–672.

55. Robertson,J.C., Jorcyk,C.L. and Oxford,J.T. (2018) DICER1 syndrome: DICER1 mutations in rare cancers. *Cancers*, **10**, 143.

56. Torrezan,G.T., Ferreira,E.N., Nakahata,A.M., Barros,B.D.F., Castro,M.T.M., Correa,B.R., Krepischi,A.C.V., Olivieri,E.H.R., Cunha,I.W., Tabori,U. *et al.* (2014) Recurrent somatic mutation in DROSHA induces microRNA profile changes in Wilms tumour. *Nat. Commun.*, **5**, 4039.

57. Walz,A.L., Ooms,A., Gadd,S., Gerhard,D.S., Smith,M.A., Guidry Auvil,J.M., Meerzaman,D., Chen,Q.-R., Hsu,C.H., Yan,C. *et al.* (2015) Recurrent DGCR8, DROSHA, and SIX homeodomain mutations in favorable histology Wilms tumors. *Cancer Cell*, **27**, 286–297.

58. Wegert,J., Ishaque,N., Vardapour,R., Geörg,C., Gu,Z., Bieg,M., Ziegler,B., Bausenwein,S., Nourkami,N., Ludwig,N. *et al.* (2015) Mutations in the SIX1/2 pathway and the DROSHA/DGCR8 miRNA microprocessor complex underlie high-risk blastemal type Wilms tumors. *Cancer Cell*, **27**, 298–311.

59. Rakheja,D., Chen,K.S., Liu,Y., Shukla,A.A., Schmid,V., Chang,T.-C., Khokhar,S., Wickiser,J.E., Karandikar,N.J., Malter,J.S. *et al.* (2014) Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. *Nat. Commun.*, **2**, 4802.

60. Heravi-Moussavi,A., Anglesio,M.S., Cheng,S.-W.G., Senz,J., Yang,W., Prentice,L., Fejes,A.P., Chow,C., Tone,A., Kalloger,S.E. *et al.* (2012) Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers. *N. Engl. J. Med.*, **366**, 234–242.

61. Anglesio,M.S., Wang,Y., Yang,W., Senz,J., Wan,A., Heravi-Moussavi,A., Salamanca,C., Maines-Bandiera,S., Huntsman,D.G. and Morin,G.B. (2013) Cancer-associated somatic DICER1 hotspot mutations cause defective miRNA processing and reverse-strand expression bias to predominantly mature 3p strands through loss of 5p strand cleavage. *J. Pathol.*, **229**, 400–409.

62. Gurtan,A.M., Lu,V., Bhutkar,A. and Sharp,P.A. (2012) In vivo structure-function analysis of human Dicer reveals directional processing of precursor miRNAs. *RNA N. Y.*, **18**, 1116–1122.

63. Kurzynska-Kokorniak,A., Koralewska,N., Pokornowska,M., Urbanowicz,A., Tworak,A., Mickiewicz,A. and Figlerowicz,M. (2015) The many faces of dicer: the complexity of the mechanisms regulating Dicer gene expression and enzyme activities. *Nucleic Acids Res.*, **43**, 4365–4380.

64. Lin,S. and Gregory,R.I. (2015) MicroRNA biogenesis pathways in cancer. *Nat. Rev. Cancer*, **15**, 321–333.

65. Vedanayagam,J., Chatila,W.K., Aksoy,B.A., Majumdar,S., Skanderup,A.J., Demir,E., Schultz,N., Sander,C. and Lai,E.C. (2019) Cancer-associated mutations in DICER1 RNase IIIa and IIIb domains exert similar effects on miRNA biogenesis. *Nat. Commun.*, **10**, 3682.

66. Gadd,S., Huff,V., Walz,A.L., Ooms,A.H.A.G., Armstrong,A.E., Gerhard,D.S., Smith,M.A., Auvil,J.M.G., Meerzaman,D., Chen,Q.-R. *et al.* (2017) A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. *Nat. Genet.*, **49**, 1487–1494.

67. Melo,S.A., Moutinho,C., Ropero,S., Calin,G.A., Rossi,S., Spizzo,R., Fernandez,A.F., Davalos,V., Villanueva,A., Montoya,G. *et al.* (2010) A genetic defect in Exportin-5 traps precursor microRNAs in the nucleus of cancer cells. *Cancer Cell*, **18**, 303–315.

68. Melo,S.A., Ropero,S., Moutinho,C., Aaltonen,L.A., Yamamoto,H., Calin,G.A., Rossi,S., Fernandez,A.F., Carneiro,F., Oliveira,C. *et al.* (2009) A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. *Nat. Genet.*, **41**, 365–370.

69. Yang,G. and Yang,X. (2010) Smad4-mediated TGF-β signaling in tumorigenesis. *Int. J. Biol. Sci.*, **6**, 1–8.

70. Massagué,J. and Chen,Y.G. (2000) Controlling TGF-beta signaling. *Genes Dev.*, **14**, 627–644.

71. Zhao,M., Mishra,L. and Deng,C.-X. (2018) The role of TGF-β/SMAD4 signaling in cancer. *Int. J. Biol. Sci.*, **14**, 111–123.

72. Davis,B.N., Hilyard,A.C., Nguyen,P.H., Lagna,G. and Hata,A. (2010) Smad proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha. *Mol. Cell*, **39**, 373–384.

73. Davoodian,P., Ravanshad,M., Hosseini,S.Y., Khanizadeh,S., Almasian,M., Nejati Zadeh,A. and Esmaiili Lashgarian,H. (2017) Effect of TGF-β/smad signaling pathway blocking on expression profiles of miR-335, miR-150, miR-194, miR-27a, and miR-199a of hepatic stellate cells (HSCs). *Gastroenterol. Hepatol. Bed Bench*, **10**, 112–117.

74. Huang,S., He,X., Ding,J., Liang,L., Zhao,Y., Zhang,Z., Yao,X., Pan,Z., Zhang,P., Li,J. *et al.* (2008) Upregulation of miR-23a approximately 27a approximately 24 decreases transforming growth factor-beta-induced tumor-suppressive activities in human hepatocellular carcinoma cells. *Int. J. Cancer*, **123**, 972–978.

75. Song,X., Zhong,H., Wu,Q., Wang,M., Zhou,J., Zhou,Y., Lu,X. and Ying,B. (2017) Association between SNPs in microRNA machinery genes and gastric cancer susceptibility, invasion, and metastasis in Chinese Han population. *Oncotarget*, **8**, 86435–86446.

76. Xie,Y., Wang,Y., Zhao,Y. and Guo,Z. (2015) Single-nucleotide polymorphisms of microRNA processing machinery genes are associated with risk for gastric cancer. *OncoTargets Ther.*, **8**, 567–571.

77. Li,X., Tian,X., Zhang,B., Zhang,Y. and Chen,J. (2012) Variation in dicer gene is associated with increased survival in T-cell lymphoma. *PLoS One*, **7**, e51640.

78. Nikolić,Z., Savić Pavićević,D., Vučić,N., Cerović,S., Vukotić,V. and Brajušković,G. (2017) Genetic variants in RNA-induced silencing complex genes and prostate cancer. *World J. Urol.*, **35**, 613–624.

79. Qian,F., Feng,Y., Zheng,Y., Ogundiran,T.O., Ojengbede,O., Zheng,W., Blot,W., Ambrosone,C.B., John,E.M., Bernstein,L. *et al.* (2016) Genetic variants in microRNA and microRNA biogenesis pathway genes and breast cancer risk among women of African ancestry. *Hum. Genet.*, **135**, 1145–1159.

80. Sung,H., Jeon,S., Lee,K.-M., Han,S., Song,M., Choi,J.-Y., Park,S.K., Yoo,K.-Y., Noh,D.-Y., Ahn,S.-H. *et al.* (2012) Common genetic polymorphisms of microRNA biogenesis pathway genes and breast cancer survival. *BMC Cancer*, **12**, 195.

81. Yang,H., Dinney,C.P., Ye,Y., Zhu,Y., Grossman,H.B. and Wu,X. (2008) Evaluation of genetic variants in microRNA-related genes and risk of bladder cancer. *Cancer Res.*, **68**, 2530–2537.

82. Liang,D., Meyer,L., Chang,D.W., Lin,J., Pu,X., Ye,Y., Gu,J., Wu,X. and Lu,K. (2010) Genetic variants in MicroRNA biosynthesis pathways and binding sites modify ovarian cancer risk, survival, and treatment response. *Cancer Res.*, **70**, 9765–9776.

83. Liu,J., Liu,Z., Wei,M., He,Y., Liao,B., Liao,G., Li,H. and Huang,J. (2012) Genetic variants in the microrna machinery gene gemin4 are associated with risk of prostate cancer: a case-control study of the chinese han population. *DNA Cell Biol.*, **31**, 1296.

84. Mullany,L.E., Herrick,J.S., Wolff,R.K., Buas,M.F. and Slattery,M.L. (2016) Impact of polymorphisms in microRNA

biogenesis genes on colon cancer risk and microRNA expression levels: a population-based, case-control study. *BMC Med. Genomics*, **9**, 21.

85. Fang,X., Yin,Z., Li,X., Xia,L. and Zhou,B. (2016) Polymorphisms in GEMIN4 and AGO1 genes are associated with the risk of lung cancer: a case-control study in chinese female Non-Smokers. *Int. J. Environ. Res. Public. Health*, **13**, 939.

86. Chacko,B.M., Qin,B.Y., Tiwari,A., Shi,G., Lam,S., Hayward,L.J., De Caestecker,M. and Lin,K. (2004) Structural basis of heteromeric smad protein assembly in TGF-beta signaling. *Mol. Cell*, **15**, 813–823.

87. McLaren,W., Gil,L., Hunt,S.E., Riat,H.S., Ritchie,G.R.S., Thormann,A., Flicek,P. and Cunningham,F. (2016) The ensembl variant effect predictor. *Genome Biol.*, **17**, 122.

88. Zhou,X., Edmonson,M.N., Wilkinson,M.R., Patel,A., Wu,G., Liu,Y., Li,Y., Zhang,Z., Rusch,M.C., Parker,M. *et al.* (2016) Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nat. Genet.*, **48**, 4–6.

89. Consortium,U.P. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.

90. Fromm,B., Domanska,D., Høye,E., Ovchinnikov,V., Kang,W., Aparicio-Puerta,E., Johansen,M., Flatmark,K., Mathelier,A., Hovig,E. *et al.* (2020) MirGeneDB 2.0: the metazoan microRNA complement. *Nucleic Acids Res.*, **48**, D132–D141.

91. Fromm,B., Billipp,T., Peck,L.E., Johansen,M., Tarver,J.E., King,B.L., Newcomb,J.M., Sempere,L.F., Flatmark,K., Hovig,E. *et al.* (2015) A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human micrornaome. *Annu. Rev. Genet.*, **49**, 213–242.

92. Lawrence,M.S., Stojanov,P., Polak,P., Kryukov,G.V., Cibulskis,K., Sivachenko,A., Carter,S.L., Stewart,C., Mermel,C.H., Roberts,S.A. *et al.* (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, **499**, 214–218.

93. Abdollah,S., Macías-Silva,M., Tsukazaki,T., Hayashi,H., Attisano,L. and Wrana,J.L. (1997) TbetaRI phosphorylation of Smad2 on Ser465 and Ser467 is required for Smad2-Smad4 complex formation and signaling. *J. Biol. Chem.*, **272**, 27678–27685.

94. Poma,A.M., Condello,V., Denaro,M., Torregrossa,L., Elisei,R., Vitti,P. and Basolo,F. (2019) DICER1 somatic mutations strongly impair miRNA processing even in benign thyroid lesions. *Oncotarget*, **10**, 1785–1797.

95. Hisamatsu,T., McGuire,M., Wu,S.Y., Rupaimoole,R., Pradeep,S., Bayraktar,E., Noh,K., Hu,W., Hansen,J.M., Lyons,Y. *et al.* (2019) PRKRA/PACT expression promotes chemoresistance of mucinous ovarian cancer. *Mol. Cancer Ther.*, **18**, 162–172.

96. Sharma,Y., Miladi,M., Dukare,S., Boulay,K., Caudron-Herger,M., Groß,M., Backofen,R. and Diederichs,S. (2019) A pan-cancer analysis of synonymous mutations. *Nat. Commun.*, **10**, 2569.

97. Macias,M.J., Martin-Malpartida,P. and Massagué,J. (2015) Structural determinants of SMAD function in TGF-β signaling. *Trends Biochem. Sci.*, **40**, 296–308.

98. Holtzhausen,A., Golzio,C., How,T., Lee,Y.-H., Schiemann,W.P., Katsanis,N. and Blobe,G.C. (2014) Novel bone morphogenetic protein signaling through Smad2 and Smad3 to regulate cancer progression and development. *FASEB J.*, **28**, 1248–1267.

99. Qin,W., Chung,A.C.K., Huang,X.R., Meng,X.-M., Hui,D.S.C., Yu,C.-M., Sung,J.J.Y. and Lan,H.Y. (2011) TGF-β/Smad3 signaling promotes renal fibrosis by inhibiting miR-29. *J. Am. Soc. Nephrol. JASN*, **22**, 1462–1474.

100. Wu,C.-W. and Storey,K.B. (2018) Regulation of Smad mediated microRNA transcriptional response in ground squirrels during hibernation. *Mol. Cell. Biochem.*, **439**, 151–161.

101. Wang,N., Tan,H.-Y., Feng,Y.-G., Zhang,C., Chen,F. and Feng,Y. (2019) microRNA-23a in human cancer: its roles, mechanisms and therapeutic relevance. *Cancers*, **11**, 7.

102. Rawat,V.P.S., Götze,M., Rasalkar,A., Vegi,N.M., Ihme,S., Thoene,S., Pastore,A., Bararia,D., Döhner,H., Döhner,K. *et al.* (2020) The microRNA miR-196b acts as tumor suppressor in Cdx2 driven acute myeloid leukemia. *Haematologica*, **105**, 285–289.

103. Liang,G., Meng,W., Huang,X., Zhu,W., Yin,C., Wang,C., Fassan,M., Yu,Y., Kudo,M., Xiao,S. *et al.* (2020) miR-196b-5p-mediated downregulation of TSPAN12 and GATA6 promotes tumor progression in non-small cell lung cancer. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 4347–4357.

104. Liang,J., Zhou,W., Sakre,N., DeVecchio,J., Ferrandon,S., Ting,A.H., Bao,S., Bissett,I., Church,J. and Kalady,M.F. (2018) Epigenetically regulated miR-1247 functions as a novel tumour suppressor via MYCBP2 in methylator colon cancers. *Br. J. Cancer*, **119**, 1267–1277.

105. Di Leva,G., Garofalo,M. and Croce,C.M. (2014) MicroRNAs in cancer. *Annu. Rev. Pathol.*, **9**, 287–314.

106. Wang,B., Hsu,S.-H., Majumder,S., Kutay,H., Huang,W., Jacob,S.T. and Ghoshal,K. (2010) TGFbeta-mediated upregulation of hepatic miR-181b promotes hepatocarcinogenesis by targeting TIMP3. *Oncogene*, **29**, 1787–1797.

107. Li,L., Li,Z., Kong,X., Xie,D., Jia,Z., Jiang,W., Cui,J., Du,Y., Wei,D., Huang,S. *et al.* (2014) Down-regulation of microRNA-494 via loss of SMAD4 increases FOXM1 and β-catenin signaling in pancreatic ductal adenocarcinoma cells. *Gastroenterology*, **147**, 485–497.

108. Zhou,J., Li,W., Guo,J., Li,G., Chen,F. and Zhou,J. (2016) Downregulation of miR-329 promotes cell invasion by regulating BRD4 and predicts poor prognosis in hepatocellular carcinoma. *Tumour Biol. J. Int. Soc. Oncodev. Biol. Med.*, **37**, 3561–3569.

109. Wang,X., Lu,X., Zhang,T., Wen,C., Shi,M., Tang,X., Chen,H., Peng,C., Li,H., Fang,Y. *et al.* (2016) mir-329 restricts tumor growth by targeting grb2 in pancreatic cancer. *Oncotarget*, **7**, 21441–21453.

110. Kang,H., Kim,C., Lee,H., Rho,J.G., Seo,J.-W., Nam,J.-W., Song,W.K., Nam,S.W., Kim,W. and Lee,E.K. (2016) Downregulation of microRNA-362-3p and microRNA-329 promotes tumor progression in human breast cancer. *Cell Death Differ.*, **23**, 484–495.

111. Li,W., Liang,J., Zhang,Z., Lou,H., Zhao,L., Xu,Y. and Ou,R. (2017) MicroRNA-329-3p targets MAPK1 to suppress cell proliferation, migration and invasion in cervical cancer. *Oncol. Rep.*, **37**, 2743–2750.

112. Enquobahrie,D.A., Abetew,D.F., Sorensen,T.K., Willoughby,D., Chidambaram,K. and Williams,M.A. (2011) Placental microRNA expression in pregnancies complicated by preeclampsia. *Am. J. Obstet. Gynecol.*, **204**, 178.

113. Zhao,J., Tao,Y., Zhou,Y., Qin,N., Chen,C., Tian,D. and Xu,L. (2015) MicroRNA-7: a promising new target in cancer therapy. *Cancer Cell Int.*, **15**, 103.

114. Swarbrick,A., Woods,S.L., Shaw,A., Balakrishnan,A., Phua,Y., Nguyen,A., Chanthery,Y., Lim,L., Ashton,L.J., Judson,R.L. *et al.* (2010) miR-380-5p represses p53 to control cellular survival and is associated with poor outcome in MYCN amplified neuroblastoma. *Nat. Med.*, **16**, 1134–1140.

115. Milosevic,J., Pandit,K., Magister,M., Rabinovich,E., Ellwanger,D.C., Yu,G., Vuga,L.J., Weksler,B., Benos,P.V., Gibson,K.F. *et al.* (2012) Profibrotic Role of miR-154 in Pulmonary Fibrosis. *Am. J. Respir. Cell Mol. Biol.*, **47**, 879–887.

116. Yang,J.-S., Maurin,T., Robine,N., Rasmussen,K.D., Jeffrey,K.L., Chandwani,R., Papapetrou,E.P., Sadelain,M., O'Carroll,D. and Lai,E.C. (2010) Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 15163–15168.

117. Telonis,A.G., Magee,R., Loher,P., Chervoneva,I., Londin,E. and Rigoutsos,I. (2017) Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types. *Nucleic Acids Res.*, **45**, 2973–2985.

118. Starega-Roslan,J., Krol,J., Koscianska,E., Kozlowski,P., Szlachcic,W.J., Sobczak,K. and Krzyzosiak,W.J. (2011) Structural basis of microRNA length variety. *Nucleic Acids Res.*, **39**, 257–268.

119. Starega-Roslan,J., Witkos,T.M., Galka-Marciniak,P. and Krzyzosiak,W.J. (2015) Sequence features of Drosha and Dicer cleavage sites affect the complexity of isomiRs. *Int. J. Mol. Sci.*, **16**, 8110–8127.

120. Desvignes,T., Loher,P., Eilbeck,K., Ma,J., Urgese,G., Fromm,B., Sydes,J., Aparicio-Puerta,E., Barrera,V., Espín,R. *et al.* (2020) Unification of miRNA and isomiR research: the mirGFF3 format and the mirtop API. *Bioinforma. Oxf. Engl.*, **36**, 698–703.

121. Telonis,A.G. and Rigoutsos,I. (2018) Race disparities in the contribution of miRNA isoforms and tRNA-derived fragments to triple-negative breast cancer. *Cancer Res.*, **78**, 1140–1154.

122. Bailey,M.H., Tokheim,C., Porta-Pardo,E., Sengupta,S., Bertrand,D., Weerasinghe,A., Colaprico,A., Wendl,M.C., Kim,J., Reardon,B. *et al.* (2018) Comprehensive characterization of cancer driver genes and mutations. *Cell*, **173**, 371–385.

123. Fleming,N.I., Jorissen,R.N., Mouradov,D., Christie,M., Sakthianandeswaren,A., Palmieri,M., Day,F., Li,S., Tsui,C., Lipton,L. *et al.* (2013) SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer. *Cancer Res.*, **73**, 725–735.

124. Korkut,A., Zaidi,S., Kanchi,R.S., Rao,S., Gough,N.R., Schultz,A., Li,X., Lorenzi,P.L., Berger,A.C., Robertson,G. *et al.* (2018) A pan-cancer analysis reveals High-Frequency genetic alterations in mediators of signaling by the TGF-β superfamily. *Cell Syst.*, **7**, 422–437.

125. Massagué,J. and Wotton,D. (2000) Transcriptional control by the TGF-beta/Smad signaling system. *EMBO J.*, **19**, 1745–1754.

126. Massagué,J. (2012) TGFβ signalling in context. *Nat. Rev. Mol. Cell Biol.*, **13**, 616–630.

127. Nolan,B.E., Levenson,E. and Chen,B.Y. (2017) Influential mutations in the SMAD4 trimer complex can be detected from disruptions of electrostatic complementarity. *J. Comput. Biol. J. Comput. Mol. Cell Biol.*, **24**, 68–78.

128. Woodford-Richens,K.L., Rowan,A.J., Gorman,P., Halford,S., Bicknell,D.C., Wasan,H.S., Roylance,R.R., Bodmer,W.F. and Tomlinson,I.P. (2001) SMAD4 mutations in colorectal cancer probably occur before chromosomal instability, but after divergence of the microsatellite instability pathway. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 9719–9723.

129. Miyaki,M., Iijima,T., Konishi,M., Sakai,K., Ishii,A., Yasuno,M., Hishima,T., Koike,M., Shitara,N., Iwama,T. *et al.* (1999) Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis. *Oncogene*, **18**, 3098–3103.

130. Koyama,M., Ito,M., Nagai,H., Emi,M. and Moriyama,Y. (1999) Inactivation of both alleles of the DPC4/SMAD4 gene in advanced colorectal cancers: identification of seven novel somatic mutations in tumors from Japanese patients. *Mutat. Res.*, **406**, 71–77.

131. Salovaara,R., Roth,S., Loukola,A., Launonen,V., Sistonen,P., Avizienyte,E., Kristo,P., Järvinen,H., Souchelnytskyi,S., Sarlomo-Rikala,M. *et al.* (2002) Frequent loss of SMAD4/DPC4 protein in colorectal cancers. *Gut*, **51**, 56–59.

132. Blahna,M.T. and Hata,A. (2012) Smad-mediated regulation of microRNA biosynthesis. *FEBS Lett.*, **586**, 1906–1912.

133. Tecalco-Cruz,A.C., Ríos-López,D.G., Vázquez-Victorio,G., Rosales-Alvarez,R.E. and Macías-Silva,M. (2018) Transcriptional cofactors Ski and SnoN are major regulators of the TGF-β/Smad signaling pathway in health and disease. *Signal Transduct. Target. Ther.*, **3**, 15.

134. Wu,J.W., Hu,M., Chai,J., Seoane,J., Huse,M., Li,C., Rigotti,D.J., Kyin,S., Muir,T.W., Fairman,R. *et al.* (2001) Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling. *Mol. Cell*, **8**, 1277–1289.

135. Souchelnytskyi,S., Tamaki,K., Engström,U., Wernstedt,C., ten Dijke,P. and Heldin,C.H. (1997) Phosphorylation of Ser465 and Ser467 in the C terminus of Smad2 mediates interaction with Smad4 and is required for transforming growth factor-beta signaling. *J. Biol. Chem.*, **272**, 28107–28115.

136. Davis,B.N., Hilyard,A.C., Lagna,G. and Hata,A. (2008) SMAD proteins control DROSHA-mediated microRNA maturation. *Nature*, **454**, 56–61.

137. Warner,D.R., Bhattacherjee,V., Yin,X., Singh,S., Mukhopadhyay,P., Pisano,M.M. and Greene,R.M. (2004) Functional interaction between Smad, CREB binding protein, and p68 RNA helicase. *Biochem. Biophys. Res. Commun.*, **324**, 70–76.

138. Wu,M.K., Vujanic,G.M., Fahiminiya,S., Watanabe,N., Thorner,P.S., O'Sullivan,M.J., Fabian,M.R. and Foulkes,W.D. (2018) Anaplastic sarcomas of the kidney are characterized by DICER1 mutations. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc*, **31**, 169–178.

139. Chen,J., Wang,Y., McMonechy,M.K., Anglesio,M.S., Yang,W., Senz,J., Maines-Bandiera,S., Rosner,J., Trigo-Gonzalez,G., Grace Cheng,S.W. *et al.* (2015) Recurrent DICER1 hotspot mutations in endometrial tumours and their impact on microRNA biogenesis. *J. Pathol.*, **237**, 215–225.

140. Kern,F., Amand,J., Senatorov,I., Isakova,A., Backes,C., Meese,E., Keller,A. and Fehlmann,T. (2020) miRSwitch: detecting microRNA arm shift and switch events. *Nucleic Acids Res.*, **48**, W268–W274.

141. Morlando,M., Dini Modigliani,S., Torrelli,G., Rosa,A., Di Carlo,V., Caffarelli,E. and Bozzoni,I. (2012) FUS stimulates microRNA

biogenesis by facilitating co-transcriptional Drosha recruitment. *EMBO J.*, **31**, 4502–4510.

142. Auyeung,V.C., Ulitsky,I., McGeary,S.E. and Bartel,D.P. (2013) Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell*, **152**, 844–858.

143. Lee,Y., Ahn,C., Han,J., Choi,H., Kim,J., Yim,J., Lee,J., Provost,P., Rådmark,O., Kim,S. *et al.* (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415–419.

144. Kim,Y.-K., Kim,B. and Kim,V.N. (2016) Re-evaluation of the roles of DROSHA, Export in 5, and DICER in microRNA biogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E1881–E1889.

145. Han,J., Lee,Y., Yeom,K.-H., Kim,Y.-K., Jin,H. and Kim,V.N. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.*, **18**, 3016–3027.

146. Fukuda,T., Yamagata,K., Fujiyama,S., Matsumoto,T., Koshida,I., Yoshimura,K., Mihara,M., Naitou,M., Endoh,H., Nakamura,T. *et al.* (2007) DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nat. Cell Biol.*, **9**, 604–611.

147. Suzuki,H.I., Yamagata,K., Sugimoto,K., Iwamoto,T., Kato,S. and Miyazono,K. (2009) Modulation of microRNA processing by p53. *Nature*, **460**, 529–533.

148. Gregory,R.I., Yan,K.-P., Amuthan,G., Chendrimada,T., Doratotaj,B., Cooch,N. and Shiekhattar,R. (2004) The microprocessor complex mediates the genesis of microRNAs. *Nature*, **432**, 235–240.

149. Fletcher,C.E., Godfrey,J.D., Shibakawa,A., Bushell,M. and Bevan,C.L. (2017) A novel role for GSK3β as a modulator of Drosha microprocessor activity and MicroRNA biogenesis. *Nucleic Acids Res.*, **45**, 2809–2828.

150. Tang,X., Li,M., Tucker,L. and Ramratnam,B. (2011) Glycogen synthase kinase 3 beta (GSK3β) phosphorylates the RNAase III enzyme Drosha at S300 and S302. *PLoS One*, **6**, e20391.

151. Lund,E., Güttinger,S., Calado,A., Dahlberg,J.E. and Kutay,U. (2004) Nuclear export of microRNA precursors. *Science*, **303**, 95–98.

152. Bohnsack,M.T., Czaplinski,K. and Görlich,D. (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA*, **10**, 185–191.

153. Wang,J., Lee,J.E., Riemondy,K., Yu,Y., Marquez,S.M., Lai,E.C. and Yi,R. (2020) XPO5 promotes primary miRNA processing independently of RanGTP. *Nat. Commun.*, **11**, 1845.

154. Hutvágner,G., McLachlan,J., Pasquinelli,A.E., Bálint,E., Tuschl,T. and Zamore,P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, **293**, 834–838.

155. Knight,S.W. and Bass,B.L. (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. *Science*, **293**, 2269–2271.

156. Song,M.-S. and Rossi,J.J. (2017) Molecular mechanisms of Dicer: endonuclease and enzymatic activity. *Biochem. J.*, **474**, 1603–1618.

157. Chendrimada,T.P., Gregory,R.I., Kumaraswamy,E., Norman,J., Cooch,N., Nishikura,K. and Shiekhattar,R. (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, **436**, 740–744.

158. Lee,H.Y., Zhou,K., Smith,A.M., Noland,C.L. and Doudna,J.A. (2013) Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing. *Nucleic Acids Res.*, **41**, 6568–6576.

159. Lee,Y., Hur,I., Park,S.-Y., Kim,Y.-K., Suh,M.R. and Kim,V.N. (2006) The role of PACT in the RNA silencing pathway. *EMBO J.*, **25**, 522–532.

160. Tomaselli,S., Bonamassa,B., Alisi,A., Nobili,V., Locatelli,F. and Gallo,A. (2013) ADAR enzyme and miRNA story: a nucleotide that an make the difference. *Int. J. Mol. Sci.*, **14**, 22796–22816.

161. Trabucchi,M., Briata,P., Filipowicz,W., Ramos,A., Gherzi,R. and Rosenfeld,M.G. (2011) KSRP promotes the maturation of a group of miRNA precuresors. *Adv. Exp. Med. Biol.*, **700**, 36–42.

162. Nowak,J.S., Choudhury,N.R., de Lima Alves,F., Rappsilber,J. and Michlewski,G. (2014) Lin28a regulates neuronal differentiation and controls miR-9 production. *Nat. Commun.*, **5**, 3687.

163. Viswanathan,S.R., Daley,G.Q. and Gregory,R.I. (2008) Selective blockade of microRNA processing by Lin28. *Science*, **320**, 97–100.

164. Thornton,J.E., Du,P., Jing,L., Sjekloca,L., Lin,S., Grossi,E., Sliz,P., Zon,L.I. and Gregory,R.I. (2014) Selective microRNA uridylation by Zcchc6 (TUT7) and Zcchc11 (TUT4). *Nucleic Acids Res.*, **42**, 11777–11791.

165. Höck,J. and Meister,G. (2008) The Argonaute protein family. *Genome Biol.*, **9**, 210.

166. Meister,G. (2013) Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.*, **14**, 447–459.

167. Rand,T.A., Petersen,S., Du,F. and Wang,X. (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*, **123**, 621–629.

168. Mourelatos,Z., Dostie,J., Paushkin,S., Sharma,A., Charroux,B., Abel,L., Rappsilber,J., Mann,M. and Dreyfuss,G. (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.*, **16**, 720–728.

169. Meister,G., Landthaler,M., Peters,L., Chen,P.Y., Urlaub,H., Lührmann,R. and Tuschl,T. (2005) Identification of novel argonaute-associated proteins. *Curr. Biol. CB*, **15**, 2149–2155.

170. Kenny,P.J., Zhou,H., Kim,M., Skariah,G., Khetani,R.S., Drnevich,J., Arcila,M.L., Kosik,K.S. and Ceman,S. (2014) MOV10 and FMRP regulate AGO2 association with microRNA recognition elements. *Cell Rep.*, **9**, 1729–1741.

171. Jin,P., Zarnescu,D.C., Ceman,S., Nakamoto,M., Mowrey,J., Jongens,T.A., Nelson,D.L., Moses,K. and Warren,S.T. (2004) Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nat. Neurosci.*, **7**, 113–117.

172. Wan,R.-P., Zhou,L.-T., Yang,H.-X., Zhou,Y.-T., Ye,S.-H., Zhao,Q.-H., Gao,M.-M., Liao,W.-P., Yi,Y.-H. and Long,Y.-S. (2017) Involvement of FMRP in primary MicroRNA processing via enhancing drosha translation. *Mol. Neurobiol.*, **54**, 2585–2594.

173. Liu,J., Rivas,F.V., Wohlschlegel,J., Yates,J.R., Parker,R. and Hannon,G.J. (2005) A role for the P-body component GW182 in microRNA function. *Nat. Cell Biol.*, **7**, 1261–1266.

174. Jakymiw,A., Lian,S., Eystathioy,T., Li,S., Satoh,M., Hamel,J.C., Fritzler,M.J. and Chan,E.K.L. (2005) Disruption of GW bodies impairs mammalian RNA interference. *Nat. Cell Biol.*, **7**, 1267–1274.