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

Received July 31, 2020; Revised October 28, 2020; Editorial Decision November 30, 2020; Accepted December 04, 2020

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 3,600 somatic mutations in 29 miRNA biogenesis genes and showed that some of the genes are over-mutated 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-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 DGC8R 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-

1To whom correspondence should be addressed. Tel: +48 61 8528503; Fax: +48 61 8520532; Email: kozlowp@ibch.poznan.pl

© The Author(s) 2021. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
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 nonsense 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 DGC8 (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 (TGFβ)/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 DGC8, 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-
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 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 (illumihaiseq_mirnaseq-miR_gene_expression). The analysis of isomiRs was performed with the use of the miRNA isoform dataset (illumihaiseq_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 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, ∼30%) of the mutations were identified in a relatively small number (n = 114, ∼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 (∼3% of samples in pan-cancer), and the least mutated genes (∼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 (R² = 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 D387) occurring 13 times in OV, COAD, GBM and LUAD (Figure 3C). These hotspots were previously described 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).

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 mutations, i.e. p.P127L in the DRBM 2 domain of **PRKRA** (Figure 3D), occurring 13 times in OV, COAD, GBM and LUAD.
(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 occurrence of the mutation as an artifact 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 p.A603RfsTer71 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 dsRBDI 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 TGFβ/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, DDX17 and ZCCHC11, 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 abrogate transcription of SMAD-controlled genes in response to TGFβ/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.
samples 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 TGFβ/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 suppressor miRs 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 consistent 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 TGFβ/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 = 0.006 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 AAAs, 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
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 miRNAs (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 suppressor miR, 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 TGFβ 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 RIIa (N = 6) and RIIb (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 RIIa and RIIb 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 RIIa (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 5p-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 6B 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 non-hotspot 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 RIIla, 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).
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 TGFβ/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 TGFβ/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
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 knock-down 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 TGFβ signaling is critical for SMAD4(SMAD2), complex formation (93,134,135), the mutation precludes the SMAD complex formation and TGFβ/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 RIIB domain (E1705, D1709, D1810 and E1813) and 6 hotspot mutations in one AA residue (S1344) in RIIIA. These hotspot mutations are all highly conserved among species and are located in the conserved MH2 domain (138). The truncated fragment is important for efficient SMAD2 phosphorylation and includes two serine residues (S465 and S467) whose phosphorylation upon TGFβ signaling is critical for SMAD4(SMAD2), complex formation (93,134,135), the mutation precludes the SMAD complex formation and TGFβ/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 RIIB domain (E1705, D1709, D1810 and E1813) and 6 hotspot mutations in one AA residue (S1344) in RIIIA. These hotspot mutations are all highly conserved among species and are located in the conserved MH2 domain (138). The truncated fragment is important for efficient SMAD2 phosphorylation and includes two serine residues (S465 and S467) whose phosphorylation upon TGFβ signaling is critical for SMAD4(SMAD2), complex formation (93,134,135), the mutation precludes the SMAD complex formation and TGFβ/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 RIIB domain (E1705, D1709, D1810 and E1813) and 6 hotspot mutations in one AA residue (S1344) in RIIIA. These hotspot mutations are all highly conserved among species and are located in the conserved MH2 domain (138). The truncated fragment is important for efficient SMAD2 phosphorylation and includes two serine residues (S465 and S467) whose phosphorylation upon TGFβ signaling is critical for SMAD4(SMAD2), complex formation (93,134,135), the mutation precludes the SMAD complex formation and TGFβ/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 RIIB domain (E1705, D1709, D1810 and E1813) and 6 hotspot mutations in one AA residue (S1344) in RIIIA. These hotspot mutations are all highly conserved among species and are located in the conserved MH2 domain (138). The truncated fragment is important for efficient SMAD2 phosphorylation and includes two serine residues (S465 and S467) whose phosphorylation upon TGFβ signaling is critical for SMAD4(SMAD2), complex formation (93,134,135), the mutation precludes the SMAD complex formation and TGFβ/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).
Ila and R1IIb 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**

The study was based on the data acquired from TCGA (project ID: 16565).

**Author contributions:** P.G.-M. participated in conceiving the study, participated in writing the scripts, performed most of the statistical and computational analyses, drafted the manuscript, prepared figures, tables and supplementary materials; M.O.U.-T. participated in conceiving the study, participated in writing the scripts, discussed the study on all steps of analyses, participated in manuscript preparation; P.M.N. discussed the analyses on all steps of the study, participated in the creation of miRNA biogenesis genes panel; P.K. received financing, participated in conceiving the study, supervised and coordinated the study, drafted the manuscript (with P.G.-M.).

**FUNDING**

Polish National Science Centre [2016/22/A/N2Z/00184 to P.K.]. Funding for open access charge: Polish National Science Centre [2016/22/A/N2Z/00184].

Conflict of interest statement. None declared.

**REFERENCES**

1. Calif,G.A., Dumitru,C.D., Shimizu,M., Bichi,R., Zupo,S., Noch,E., Aldier,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., Calif,G.A., Liu,C.G., Amb,s,B., Cinmino,A., Petrocca,F., Visonc,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-Kersch, 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. Krutzfeldt,V.A. and Herceg,Z. (2010) Oncogenic microRNAs (Oncomirs) as a new class of cancer biomarkers. *BioEssays* *Rev. Mol. Cell. Dev. Biol.*, 32, 894–904.

8. Kasinski,A.L. and Slack,F.J. (2011) Epigenetics and genes. *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.*, 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., Hradzilova,K., Pachernik,J., Cetkovska,K., Uldrijan,S., Zdrahal,Z. and Vanacova,S. (2013) Mammalian Dsil2 exoribonuclease targets the uridylated precursors of let-7 miRNAs. *RNA N. Y.*, 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., Suayman,S., Dijkstra,K.K., El Andaloussi,S., Weinberg,M.S. and Wood,M.J.A. (2012) The biogenesis and characterization of mammalian miRNAs 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) miBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, 42, D68–73.

22. Kozomara,A. and Birgaoanu,M. and Griffiths-Jones,S. (2019) miBase: 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,Q.J. (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.*, 21, 1317–1325.

29. Czubak, A., Lewandowska, M.A., Klonekowska, 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. *Oncoarget*, 6, 23399–23416.

30. Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F. and 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., Racanelli, S., Zardo, G., Sterne, L.M., Maneci, 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. *Cell Cancer Genet.*, 5, 2202–2222.

33. Ali Syeda, Z., Langden, S.S.S., Munkhzul, C., Lee, M. and Song, S.J. (2016) SomamiR 2.0: a database of somatic mutations in miRNA genes in lung cancer-potential prioritization using human population and disease datasets. *Hum. Mutat.*, 47, 163–172.

34. Saito, Y. and Jones, P.A. (2006) Epigenetic activation of tumor suppressor microRNAs in human cancer cells. *Cell Cycle Georget.*, 5, 2202–2222.

35. Ali Syeda, Z., Langden, S.S.S., Munkhzul, C., Lee, M. and Song, S.J. (2016) SomamiR 2.0: a database of somatic mutations in miRNA genes in lung cancer-potential prioritization using human population and disease datasets. *Hum. Mutat.*, 47, 163–172.

36. Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F. and 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.

37. Lujambio, A., Esteller, M. (2009) How epigenetics can explain the human genome. *Nat. Rev. Genet.*, 105, 23399–23416.

38. Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M., Ludwig, N. et al. (2015) Mutations in the SIX1/2 pathway and the DROSHA/DGCR8 miRNA microprocessor complex underlie high-risk blastemal type ovarian cancers. *N. Engl. J. Med.*, 377–382.

39. Auvil, J.M., Meerzaman, D., Chen, Q.-R., Hsu, C.H., Yan, C. et al. (2019) A microRNA DNA methylation signature for human cancer metastasis. *Cell Cycle Georget.*, 10, 13556–13561.

40. Kloviczeckiewicz, J., Sobolewska, A., Lejnowski, D., Cullawn, J.F. and Bartoszewski, R. (2018) microRNA single polynucleotide polymorphism influences on microRNA biogenesis and mRNA target specificity. *Gene*, 610, 482–485.

41. Kloviczeckiewicz, J., Sobolewska, A., Lejnowski, D., Cullawn, J.F. and Bartoszewski, R. (2018) microRNA single polynucleotide polymorphism influences on microRNA biogenesis and mRNA target specificity. *Gene*, 610, 482–485.

42. Kloviczeckiewicz, J., Sobolewska, A., Lejnowski, D., Cullawn, J.F. and Bartoszewski, R. (2018) microRNA single polynucleotide polymorphism influences on microRNA biogenesis and mRNA target specificity. *Gene*, 610, 482–485.

43. Kloviczeckiewicz, J., Sobolewska, A., Lejnowski, D., Cullawn, J.F. and Bartoszewski, R. (2018) microRNA single polynucleotide polymorphism influences on microRNA biogenesis and mRNA target specificity. *Gene*, 610, 482–485.

44. Kloviczeckiewicz, J., Sobolewska, A., Lejnowski, D., Cullawn, J.F. and Bartoszewski, R. (2018) microRNA single polynucleotide polymorphism influences on microRNA biogenesis and mRNA target specificity. *Gene*, 610, 482–485.
104. Liang,J., Zhou,W., Sakre,N., DeVecchio,J., Ferrandon,S., Ting,A.H., Bao,S., Bissett,J., 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, 287–316.

106. Wang,B., Hsu,S.H., Majumder,S., Kutay,H., Huang,W., Jacob,S.T. and Ghoshal,K. (2010) TGFB-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-404 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,L., 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., Liang,J., Zhou,W., Sakre,N., DeVecchio,J., Ferrandon,S., Bailey,M.H., Tokheim,C., Porta-Pardo,E., Bertrand,D., Weerasinha,A., Colaprico,A., Wendel,M.C., Kim,J., Reardon,B. et al. (2018) Comprehensive characterization of cancer driver genes and mutations. *Cell*, 173, 371–385.
biogenesis by facilitating co-transcriptional Drosha recruitment. *EMBO J.*, **31**, 4502–4510.

142. Auyeung,V.C., Ulitsky,I., McGearry,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., Anu,C., Han,J., Choi,H., Kim,J., Yim,J., Lee,J., Provost,P., Rädmark,O., Kim,S. et al. (2003) The nuclear R-Nase 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., Yan,K.-P., Amuthan,G., Chendrimada,T., Suzuki,H., Yamagata,K., Fujiyama,S., Matsumoto,T., Koshida,I., Yokomori,T., Iwamoto,T., Kato,S. and Fukuda,T., Yamagata,K., Ikeda,N., Koshida,I., Kato,S. and Mochida,K. (2017) Molecular mechanisms of Dicer: RNA interference enzyme Drosha at S300 and S302. *Biochem. Biophys. Res. Commun.*, **489**, 621–629.

148. Gregory,R.I., Yan,K.-P., Amuthan,G., Chendrimada,T., Suzuki,H., Yamagata,K., Fujiyama,S., Matsumoto,T., Koshida,I., Yokomori,T., Iwamoto,T., Kato,S. and Mochida,K. (2017) Molecular mechanisms of Dicer: RNA interference enzyme Drosha at S300 and S302. *Biochem. Biophys. Res. Commun.*, **489**, 621–629.

149. trout,T., Cooch,N., Shiekhattar,R. (2004) The microprocessor complex mediates the genesis of microRNAs. *Nature*, **429**, 529–533.

150. 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*, **429**, 529–533.

151. Lund,E., Güttinger,S., Calado,A., Dahlberg,J.E. and Kutay,U. (2001) A cellular function for the RNAase III enzyme Drosha at S300 and S302. *EMBO J.*, **20**, 740–744.

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

153. 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.

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

155. Meister,G., Landthaler,M., Peters,L., Chen,P.Y., Urlaub,H., Lührmann,R. and Tuschl,T. (2005) Identification of novel Argonaute-associated proteins. *Genes Dev.*, **19**, 2269–2271.

156. Mourelatos,Z., Dostie,J., Paushkin,S., Sharma,A., Charroux,B., Dupuy,A., Michalski,A., Calado,A., Dahlberg,J.E., Kennedy,T. et al. (2003) CARM1 recruits Dicer to Ago2 for microRNA processing and gene silencing. *Nature*, **425**, 711–715.

157. 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.*, **15**, 2149–2155.

158. 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.*, **15**, 2149–2155.

159. 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 regulates miR-9 production. *Int. J. Mol. Sci.*, **15**, 7102–7120.

160. 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.

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

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

163. Nowak,J., Lai,C., Maller,J.L. and Zamore,P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small RNA. *Science*, **293**, 1729–1741.

164. 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.

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

166. 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.*, **15**, 2149–2155.

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. Meister,G., 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.

169. Meister,G., 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.

170. Meister,G., 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.

171. Meister,G., 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.

172. Meister,G., 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.

173. Meister,G., 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.

174. Meister,G., 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.

175. Meister,G., 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.

176. Meister,G., 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.

177. Meister,G., 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.

178. Meister,G., 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.

179. Meister,G., 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.

180. Meister,G., 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.