In Vitro and in Silico Analysis of miR-125a with rs12976445 Polymorphism in Breast Cancer Patients

Tomasz P. Lehmann [1,*] **, Joanna Miskiewicz** [2] **, Natalia Szostak** [3] **, Marta Szachniuk** [2,3,*] **, Sylwia Grodecka-Gazdecka** [4] **and Paweł P. Jagodziński** [1]

[1] Department of Biochemistry and Molecular Biology, Poznan University of Medical Sciences, Swiecickiego 6, 60-781 Poznan, Poland; pjagodzi@ump.edu.pl

[2] Institute of Computing Science, Poznan University of Technology, Piotrowo 2, 60-965 Poznan, Poland; jmiskiewicz@cs.put.poznan.pl

[3] Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland; nszostak@ibch.poznan.pl

[4] Department of Surgery, Chair and Clinic of Oncology Poznan University of Medical Sciences, Szamarzewskiego 82/84, 61-001 Poznan, Poland; sylwia.grodecka-gazdecka@oncology.am.poznan.pl

[*] Correspondence: tlehmann@ump.edu.pl (T.P.L.); mszachniuk@cs.put.poznan.pl (M.S.); Tel.: +48-618-546-513 (T.P.L.); +48-616-653-030 (M.S.); Fax: +48-854-6510 (T.P.L.); +48-618-771-525 (M.S.)

check for updates

Abstract: *Background*: Breast cancer affects over 2 million women yearly. Its early detection allows for successful treatment, which motivates to research factors that enable an accurate diagnosis. miR-125a is one of them, correlating with different types of cancer. For example, the miR-125a level decreases in breast cancer tissues; polymorphisms in the miR-125a encoding gene are related to prostate cancer and the risk of radiotherapy-induced pneumonitis. *Methods*: In this work, we investigated two variants of rs12976445 polymorphism in the context of breast cancer. We analyzed the data of 175 blood samples from breast cancer patients and compared them with the control data from 129 control samples. *Results*: We observed the tendency that in breast cancer cases TT genotype appeared slightly more frequent over CC and CT genotypes (statistically nonsignificant). The TT genotype appeared also to be more frequent among human epidermal growth factor receptor 2 (HER2) positive patients, compared to HER2 negative. In silico modelling showed that the presence of uridine (U) diminished the probability of pri-miR-125a binding to NOVA1 and HNRNPK proteins. We demonstrated that U and C -variants could promote different RNA folding patterns and provoke alternative protein binding. *Conclusions*: U-variant may imply a lower miR-125a expression in breast cancer.

Keywords: microRNA; RNA protein binding; RNA folding; breast cancer; polymorphism

1. Introduction

Constant investigation of the genetic background of breast cancer is a crucial endeavour since only 15–36% of hereditary breast cancers have known genetic background [1]. Many current studies refer to estimate the association of polymorphisms in genes encoding miRNA with cancer and to find how single nucleotide polymorphisms (SNP) modulate miRNA expression [2]. SNPs in genes encoding miRNAs have been revealed in several different types of cancer, including breast cancer [3–5]. It has been proven that SNPs in miRNA genes affect their expression in breast cancer and modulate the expression of miRNA target genes (e.g., miR-27a, miR-196a2, miR-559) [6–8]. SNPs contribute to abnormal expression of miRNAs in breast cancers either as upregulated oncomirs or downregulated suppressors [9]. Therefore, understanding of SNPs impact on miRNA levels might be helpful in the selection of breast cancer diagnostic markers.

One of such potentially useful miRNA marker is miR-125a-the gene is located on chromosome 19, in the proximity of miR-99b and let-7e. This miRNA cluster is situated within the first intron of the transcription unit of sperm acrosomal membrane-associated protein 6 (SPACA6) variant 1 [10]. The genomic DNA sequence located between -3875 bp and -3006 bp from pre-miR-99b drives the expression of SPACA6 gene and its intronic miR-99b/let-7e/miR-125a cluster [10]. miR-125a is involved in the development of various types of cancer, including lung, gastric, and breast cancers [6,11–13]. miR-125a functions as an oncomir or a tumour suppressor miRNA depending on the cellular context [6,14]. In breast cancer tissues, the miR-125a level is decreased when comparing to normal adjacent tissue. Several SNPs have been already described in pri-miR-125a and mature miR-125a: rs41275794 (A/G) and rs12976445 (C/T) SNPs in the pri-miR-125a; rs12975333 (G/T) SNP on the 5′ strand and -rs143525573 (A/G) SNP on the 3′ strand of mature miR-125a [15–17]. rs12976445 C and T haplotypes are correlated with cancerous and noncancerous diseases [18,19]. One of the SNPs in miR-125a, rs12976445, is considerably associated with lymph node metastasis, tumour stage (Classification of Malignant Tumors, TNM), estrogen receptor status, and progesterone receptor status [19]. The TT genotype of rs12976445 significantly increases the risk of mortality in breast cancer patients compared with those carrying the CC genotypes of the SNP [19]. Moreover, patients carrying the TT or CT genotype ofrs12976445 have a higher risk of radiotherapy-induced pneumonitis [20,21]. The CC-variant of the mentioned SNP has been shown to correlate with higher expression compared with C/T and T/T [20]. Homozygous CC-variant of rs12976445 was shown to increase the risk of prostate cancer [22]. In a study of autoimmune thyroid diseases, it was found that the C allele is significantly increased in Hashimoto's thyroiditis patients [18]. In the same study, a protecting role of the T allele was suggested to decrease the risk for Hashimoto's thyroiditis for about 31.6% [18].

Although it has been suggested that SNPs have an impact on miRNA expression and are associated with various diseases, the precise mechanism of the rs12976445 effect on miR-125a processing remains unknown. In vitro and in silico analysis of the SNP and the recognition of proteins interacting with rs12976445 SNP can reveal a detailed picture of the underlying mechanism. In this research, we analyzed the frequency of rs12976445 alleles and genotypes in breast cancer in Polish patients and estimated the genotype frequency with the receptor status. Following the methodology from our previous studies [23,24], we used in silico analysis and computational tools to generate and compare the secondary structure of miRNA of two alternative variants, C and U. Using in silico approach, we found potential proteins interacting with sequence variants of rs12976445 pre-miRNA.

2. Materials and Methods

2.1. Patients and Samples

27 Polish patients (women) were undergoing surgery for breast cancer in the Department of Surgery, Chair of Oncology of Poznan University of Medical Sciences (PUMS). The study protocol was approved by the bioethics board of PUMS (number 839/09 8 October 2009, 690/10 2 September 2010). Blood samples for experiments were collected from participants with their consent.

2.2. Blood Sampling and Measurements

A total 175 blood samples were obtained by antecubital vein puncture. Further, 129 control blood samples were collected from patients with the negative diagnoses of breast cancer, and not diagnosed with other types of cancer. All 304 samples were collected from 14–88 years old patients of the Department of Surgery, Chair of Oncology at PUMS.

2.3. DNA Extraction, PCR, Restriction Analysis

DNA from blood was extracted using GenElute™ Mammalian Genomic DNA Miniprep kit (nr G1N70 Sigma-Aldrich, St. Luis, MO, USA). The quantity of obtained nucleic acid was assessed using a BioPhotometer™ (Eppendorf, Hamburg, Germany).

2.4. Polymerase Chain Reaction, Sequencing, and Restriction Analysis

DNA specimens were amplified using the standard PCR protocols. DreamTaq DNA Polymerase 0.02 U/μl (EP0701, Thermo Fisher Scientific, Waltham, MA, USA), dNTP Mix 1 mM (U151A, Promega, Madison, WI, USA), primers 0.5 μM were applied. The temperature profile was as follows: (1) the preheating phase, 95 °C for 10 min, (2) the amplification cycle of 35 repeats, 95 °C for 30 s and 55 °C for 30 s, and finishing the cycle with 72 °C for 30 s, (3) the final elongation process, 72 °C for 7 min. PCR primers: 5′-TTTTGGTCTTTCTGTCTCTGG-3′ and 5′-TGGAGGAAGGGTATGAGGAGT-3′, which were spanning the sequence of pre-miR-125a (accession number ENSG00000208008) were designed using the Oligo software (Molecular Biology Insights, Inc., DBA Oligo, Inc., Colorado Springs, CO, USA). This amplicon was used for the restriction analysis and MIR125A sequencing. PCR products of 27 samples collected from patients were purified using the gel-out purification kit (A&A Biotechnology, Gdynia, Poland) and sequenced at the DNA Sequencing and Oligonucleotide Synthesis Laboratory of the Institute of Biochemistry and Biophysics of the Polish Academy of Sciences (Warsaw, Poland). The sequencing results were analyzed using BioEdit Sequence Alignment Editor [25]. The rs12976445 SNP was genotyped using the BaeGI restriction enzyme (R0708S, New England Biolabs, Ipswich, MA, USA). The genotyping required 17.8 μL of the PCR product and 2 U of a restriction enzyme that were kept in 37 °C for 12 h. BaeGI cuts GKGCM|C K = G or T, M = A or C (T-variant TGTGCCTA, C-variant TGTGCCCA). Upon preliminary sequencing samples of 27 patients, we supposed that the uncut sequence would contain T-variant, although G- and A-variant could not be excluded [26]. The enzyme HpyF3I (DdeI) (ER 1881, Thermo Scientific, Waltham, MA, USA), which cuts C|TNAG, was used to analyze rs143525573 and rs12975333 SNP in the miR-125a amplicon. The applied procedure was analogous to the genotyping with the BaeGI restriction enzyme.

2.5. Preparing Sequence for in Silico Analysis

From the amplicon comprising pre-miR-125a sequence, we selected a fragment which consisted of the SNP nucleotide, the 25 nt upstream region of SNP, and the 25 nt downstream region of SNP. This subsequence was named as 51-miR-125a and it was analyzed in two variants, C and U, of the genotype.

2.6. Predicting Protein Interactions

RBPmap web server [27] and ATtRACT database [28] were used to find possible protein interactions within 51-miR-125a, for both U- and C-variants. RBPmap was designed in 2014 as a tool for finding potential binding sites of RNA-interacting proteins, especially in human, mouse, and *Drosophila melanogaster* genomes. Searching for potential binding sites of RNA-interacting proteins was performed using RBPmap server with three different stringency levels. ATtRACT is a database with an implemented search engine that applies a fast algorithm dedicated to finding motifs corresponding to RNA-interacting proteins. The database contains 370 RBPs and 1583 RBP consensus binding motifs that can be searched throughout different organisms (including humans).

2.7. 2D Structure Modelling

The secondary structure of 51 nt long RNA fragment of pri-miR-125a sequence (51-miR-125a) was predicted using bioinformatictools -RNAstructure [29] and RNAfold programs [30]. RNAfold is a part of the Vienna RNA package that collects tools specialised for analysing single-stranded nucleic acid sequences. RNAfold generates 2D models optimizing either minimum free energy (MFE mode) or minimal base-pair distance (centroid mode) parameters. RNAstructure is the second program that we used for predicting 2D structures. It can handle both RNA and DNA sequences and allows users to predict biomolecular secondary structures. Based on a given sequence, RNAstructure generates the lowest (minimum) free energy 2D structure model (MFE mode) and a 2D structure composed of highly probable base pairs (MaxExpect mode). Predicted models in this research for both RNAfold and RNAstructure were obtained using default settings.

2.8. Statistical Analysis

The expected genotype and allele frequencies for the observed variations were calculated for all 175 positive cases (Ca) and 129 control samples (Co). These frequencies were tested for Hardy-Weinberg equilibrium. Statistical analysis of genotypes frequency was conducted using the online tool SNPstats [31]. The association of rs12976445 in miR-125a in breast cancer cells was calculated using odds ratio (OR) at a 95% confidence interval (CI). The association of the genotypes with the receptor status was calculated using χ square test and the software GraphPadInStat version 7.00 (GraphPad Software, San Diego, California, USA).

3. Results

3.1. In Vitro Analysis

rs12976445 SNP in pri-miR-125a occurring in three genotypes TT, CT, and CC has been associated with cancer and other diseases [32–34]. We studied the pri-miR-125a amplicon (Figure 1) by sequencing and restriction analysis to determine the genotype frequency and its association with breast cancer.

```
Version U
uuuuggucuuucugucucuggcucucagaaugucucugugccuaucuccaucucugaccccc
acccagggucuaccgggccaccgcacaccauguUGCCAGUCUCUAGGUCCCUGAGACCCUU
UAACCUGUGAGGACAUCCAGGGUCACAGGUGAGGUUCUUGGGAGCCUGGCGUCUGGCCcaac
cacacaccuggggaauugcuggccugacuucugaccccugacuccucauaccuuccucca

Version C
uuuuggucuuucugucucuggcucucagaaugucucugugcccaucuccaucucugaccccc
acccagggucuaccgggccaccgcacaccauguUGCCAGUCUCUAGGUCCCUGAGACCCUU
UAACCUGUGAGGACAUCCAGGGUCACAGGUGAGGUUCUUGGGAGCCUGGCGUCUGGCCcaac
cacacaccuggggaauugcuggccugacuucugaccccugacuccucauaccuuccucca
```

Figure 1. Whole amplicon (247 nt) comprising pre-miR-125a sequence, the fragment of the sequence from Ensembl Gene ID: ENSG00000208008. Lower-case corresponds to pri-miR-125a fragments removed in pre-miR-125a processing; grey upper-case is pre-miR-125a; in bold upper-case we marked mature miR-125a 5p and 3p, respectively; bold underlined lower-case nucleotides are the analyzed U-, C-variants; bold italic lower-case is a 51 nt fragment (51-miR-125a) that was analyzed using computational methods. BaeGI restriction site is located directly before rs12976445 SNP, dividing sequence into 42 nt and 205 nt subsequences.

We amplified DNA obtained from 27 breast cancer tumours to assess the frequency of several known SNPs in pri-miR-125a sequence. The sequencing of 27 pri-miR-125a amplicons showed that only nrs12976445 polymorphic site was highly variable in our samples. Subsequently, we estimated the association of rs12976445 with breast cancer using DNA from 304 blood samples, 175 positive (cases, Ca) and 129 negative (controls, Co) samples. We digested the pri-miR-125a amplicon using BeaGI restriction enzyme producing 42 and 205 fragments if C or non-digesting if T variant was present. We performed statistical analysis with SNPStats on-line software. The allele and genotype frequencies analysis are presented in Table 1. We found that rs12976445 in pri-miR-125a followed Hardy-Weinberg equilibrium in breast cancer cases (Ca) set (Table 1).

Table 1. The allele, genotype frequencies, the exact test for Hardy–Weinberg equilibrium analysis, and the single nucleotide polymorphism (SNP) association with breast cancer status of rs12976445 miR-125a gene polymorphism. Ca—cases, Co—controls, OR—odds ratio, CI—confidence interval.

SNP Allele Frequencies ($n = 304$)						
	All subjects		Ca		Co	
Allele	Count	Proportion	Count	Proportion	Count	Proportion
T	414	0.68	241	0.69	173	0.67
C	194	0.32	109	0.31	85	0.33
SNP Genotype Frequencies ($n = 304$)						
	All subjects		Ca		Co	
Genotype	Count	Proportion	Count	Proportion	Count	Proportion
C/C	24	0.08	14	0.08	10	0.08
T/C	146	0.48	81	0.46	65	0.5
T/T	134	0.44	80	0.46	54	0.42
SNP Exact Test for Hardy–Weinberg Equilibrium ($n = 304$)						
	TT	TC	CC	T	C	*p*-value
All subjects	134	146	24	414	194	0.086
Ca	80	81	14	241	109	0.38
Co	54	65	10	173	85	0.16
SNP Association with Response STATUS ($n = 304$, Crude Analysis)						
Model	Genotype	Ca	Co	OR (95% CI)		*p*-value
Codominant	T/T	80 (45.7%)	54 (41.9%)	1.00		
	C/T	81 (46.3%)	65 (50.4%)	1.19 (0.74–1.91)		0.77
	C/C	14 (8%)	10 (7.8%)	1.06 (0.44–2.56)		
Dominant	T/T	80 (45.7%)	54 (41.9%)	1.00		0.5
	C/T+ C/C	95 (54.3%)	75 (58.1%)	1.17 (0.74–1.85)		
Recessive	T/T+ C/T	161 (92%)	119 (92.2%)	1.00		0.94
	C/C	14 (8%)	10 (7.8%)	0.97 (0.41–2.25)		
Over-dominant	T/T+ C/C	94 (53.7%)	64 (49.6%)	1.00		0.48
	C/T	81 (46.3%)	65 (50.4%)	1.18 (0.75–1.86)		
Log-additive	n/a	n/a	n/a	1.10 (0.76–1.58)		0.62

The statistical analysis, comprising odds ratio (OR) at 95% confidence interval (CI), are presented in Table 1. In the co-dominant model (TT vs. TC vs. CC), the heterozygous CT genotype of rs12976445 SNP was slightly more frequent in control (Co) set (50.4%) compared to cases (Ca) (46.3%) with OR Co/Ca = 1.19, 95%, CI = 0.74–1.91 and $p = 0.77$ (Table 1). In the dominant model (TT vs. TC + CC), we observed slightly higher level of CT + CC in controls over cases with OR = 1.17, 95% (CI = 0.74–1.85) and $p = 0.5$. These calculations indicated the tendency that in breast cancer cases, TT genotype appeared slightly more frequent than CC and CT genotypes.

Additionally, we analyzed two other known SNPs, rs143525573 and rs12975333, located in the mature miR-125a coding sequence. The DdeI enzyme was applied to assess these two SNPs. We did not observe the variability of rs143525573 and rs12975333 in the studied group of patients with breast cancer.

Since SNP in pri-miR-125a could modulate the level of mature miRNA, we assessed the relationship of rs12976445 variability with the status of human epidermal growth factor receptor 2 (HER2), ER1α and PR receptors. mRNA encoding receptor HER2 (human epidermal growth factor receptor 2, receptor tyrosine-protein kinase erbB-2, encoded by ERBB2 gene) is a target of miR-125a [33]. HER2 together with ER1α and PR receptors are used in the diagnosis of breast cancer. Therefore, we decided to perform tests of the relationship of rs12976445 variation with the status of these receptors. HER2 receptor status 0, 1, and 2 without amplification was classified as the negative group (N), while status

2 with amplification and status 3 were ranked as the positive group (P). Patients were also divided into two groups according to ER1α or PR status, 0–10% as the negative group and 10–75% as the positive group. We analyzed data using χ square test. The lowest *p*-value, 0.0606, was obtained for TT, CT, CC genotypes analysis for HER2 receptor. Our analysis revealed a tendency that T allele predominated in HER2 positive samples (Table 2). These results are concomitant with genotype frequency, which TT-variant slightly predominates in cases over controls. To determine how C- and T(U)-variants could regulate miR-125a expression, we decided to analyze the 2D structure of pre-miRNA usingin silicomethods.

Table 2. The association analysis between miR-125a rs12976445 polymorphism and HER2, ER and PR receptors status in breast cancer patients. The association was calculated using χ square test. The status of receptors was divided into two groups-positive (P) and negative (N).

Receptor	Case	Variant	Group Status		*p*-Value
			P	N	
HER2	Genotype	TT	14	30	0.0606
		CT	7	48	
		CC	1	6	
		TT+CT	21	78	0.2609
		CT+CC	8	54	
	Alleles frequency	T	35	108	0.0814
		C	9	60	
ER	Genotype	TT	36	8	0.2001
		CT	37	17	
		CC	6	1	
		TT+CT	73	25	0.2891
		CT+CC	33	18	
	Alleles frequency	T	109	33	0.5702
		C	49	19	
PR	Genotype	TT	30	14	0.4922
		CT	35	20	
		CC	6	1	
		TT+CT	65	34	1.0000
		CT+CC	41	21	
	Alleles frequency	T	95	48	0.9297
		C	47	22	

3.2. In Silico Analysis of RNA Binding Proteins (RBPs)

To find potential binding proteins to the pri-miR-125a containing C- and U-variants, we analyzed the 51-nucleotide fragment of pri-miR-125a using RBPmap web server and ATtRACT database [27,28]. Both, the C- and U-variants were analyzed using RBPmap with three different stringency levels: low, medium, and high. Low stringency level corresponds to significant *p*-value < 0.01 and suboptimal *p*-value < 0.02; medium stringency level thresholds were at significant *p*-value < 0.005 and suboptimal *p*-value < 0.01; high stringency level was set to <0.001 and <0.01 for significant *p*-value and suboptimal *p*-value, respectively. We searched for any available human motif stored in RBPmap. Low stringency algorithm mapped SRSF5, PTBP1, and PCBP1 proteins in both C-/U-variants. BRUNOL4 and BRUNOL5 proteins were only mapped in U-variant, whereas SRSF3, HNRNPK, and NOVA1 proteins were only

found in C-variant. In medium and high stringency levels, RBPmap found only PTBP1 protein motif in U-variant. For the C-variant, medium stringency algorithm found following potentially binding proteins: SRSF3, HNRNPK, NOVA1, and PTBP1. High stringency level for C-variant found the same RBPs as the medium algorithm, excluding HNRNPK protein. All RBPmap-predicted proteins that potentially bind to the region containing C-/U-variant in 51-miR-125a sequence are shown in Table 3.

Table 3. Motif sequences of possible protein-binding sites in C- variant and U-variant found by RBPmap tool. SNPs are marked as bold.

Variant	Protein	Mode			Motif Containing Variant
		Low	Medium	High	
C	HNRNPK	✓	✓		C**C**AUCUC
	NOVA1	✓	✓	✓	C**C**AU
	PCBP1	✓			**C**AUCUCC
	PTBP1	✓	✓	✓	C**C**AUCU
	SRSF3	✓	✓	✓	CC**C**AUCU
	SRSF5	✓			**C**AUCUCC
U	BRUNOL4	✓			UGUGCC**U**
	BRUNOL5	✓			UGUGCC**U**
	PCBP1	✓			CC**U**AUCU
	PTBP1	✓	✓	✓	C**U**AUCU
	SRSF5	✓			CC**U**AUCU, **U**AUCUCC

We focused on proteins that were variant-specific. Less frequent C-variant interacts potentially with HNRNPK and NOVA1—the proteins involved in the RNA processing [35,36]. The C-variant was also mapped with SRSF3 protein, which was not found by RBPmap in U-variant on any stringency level of the algorithm execution and appeared on every stringency level of C-variant. On the contrary, BRUNOL4 and BRUNOL5, proteins that regulate alternative splicing of pre-mRNA and are possibly connected with mRNA editing and translation, were only mapped in U-variant of the analyzed sequence.

In the next step, we used ATtRACT database to search for potential RBP motifs in 51-miR-125a sequence. In the U-variant, ATtRACT predicted PTBP1 motif corresponding to the results from RBPmap, where PTBP1 was mapped in all three stringency levels of the algorithm. In the C-variant, ATtRACT predicted YBX1 and NOVA1 binding motifs. YBX1 protein is involved in cellular processes, including pre-mRNA splicing, transcriptional and translational regulation. It is also potentially involved in miRNA processing [37]. What brings particular interest, is the mapping of the NOVA1 protein in the C-variant. It complies to the results from RBPmap presented earlier. Similarly, as PTBP1, RBPmap predicted NOVA1 binding to 51-miR-125a for all stringency levels (low, medium, high). The results suggest that C-variant in 51-miR-125a is connected with different RBP (PTBP1) than the U-variant (NOVA1). Binding motifs for C-/U-variants from ATtRACT database are presented in Table 4.

Table 4. Motif sequences of possible protein-binding sites in C-variant and U-variant found in ATtRACT database. SNPs are marked as bold.

Variant	Protein	Motif Containing Variant
C	NOVA1	C**C**AU
	YBX1	**C**AUC
U	PTBP1	C**U**AU, CC**U**AU

3.3. In Silico Modelling of pri-miR-125a Folding

Computational modelling of the 2D structure of 51-miR-125a was performed using RNAfold (Figure 2) and RNAstructure (Figure 3) [29,30].
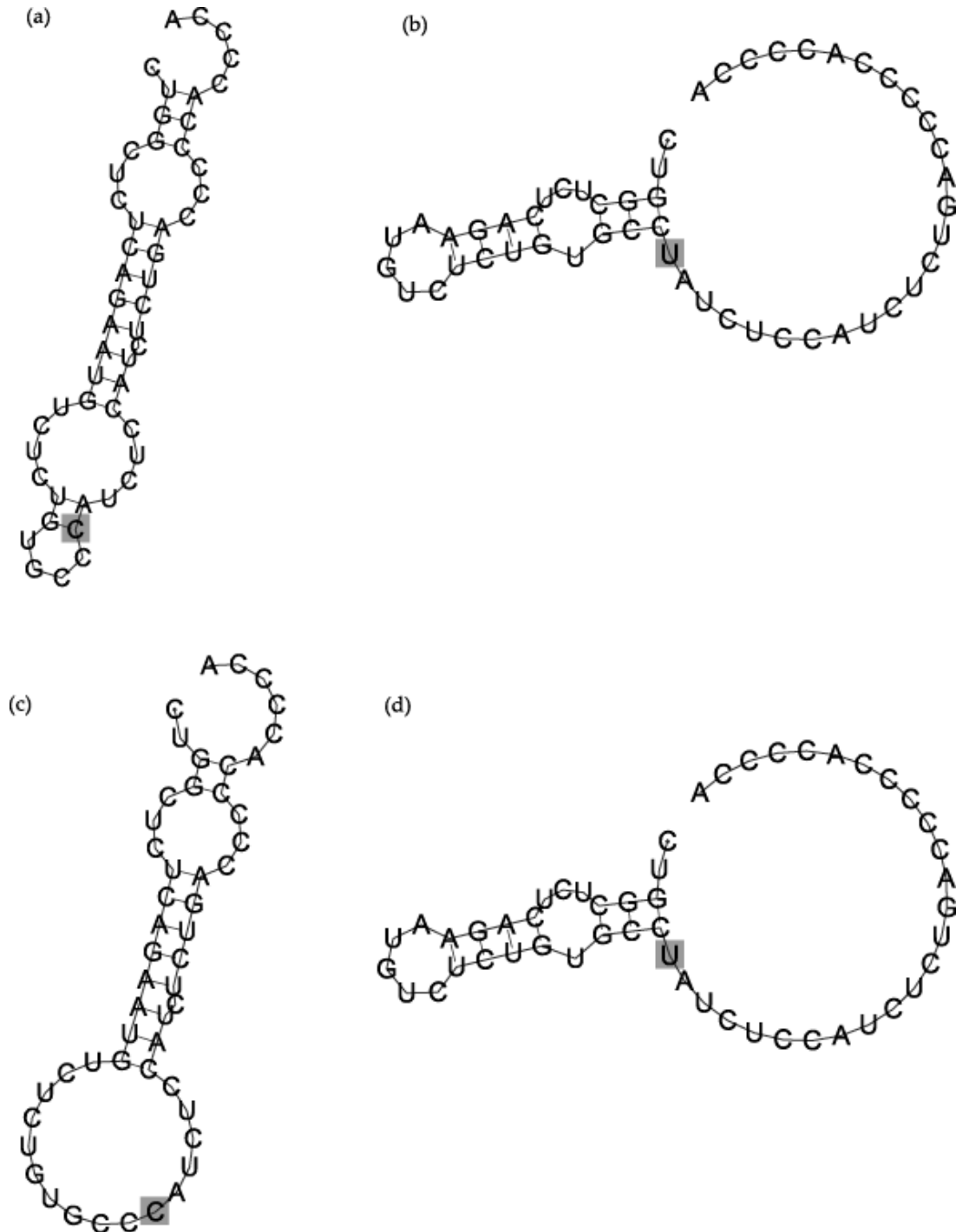
Figure 2. The 2D structures of 51-miR-125a with C-variant and U-variant predicted by RNAfold. In grey squares we marked SNPs. (**a**) C-variant MFE model (minimizing free energy), (**b**) U-variant MFE model (minimizing free energy) (**c**) C-variant variant centroid model (minimizing base-pair distance) and (**d**) U-variant centroid model (minimizing base-pair distance).
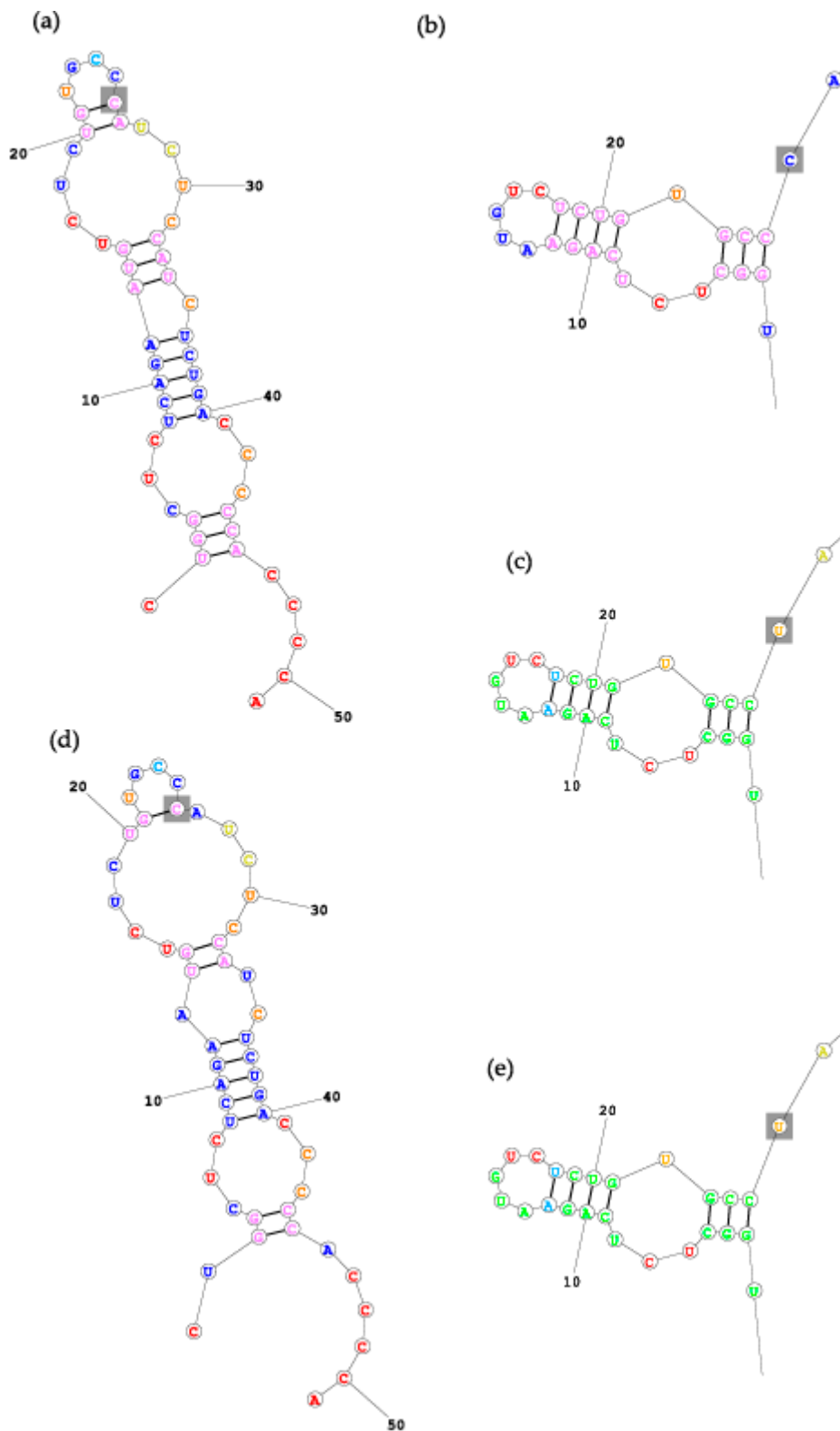
Figure 3. The 2D structures of 51-miR-125a with C-variant (**a,b,d**) and U-variant (**c,e**) predicted by RNAstructure. In grey squares we marked SNPs. (**a**) C-variant MFE model (minimizing free energy), (**b**) C-variant MFE model, and (**c**) U-variant MFE model, (**d**) C-variant MaxExpect model (most probable base-pairing) (**e**) U-variant MaxExpect model. (**b,c,e**) models were cut for better visibility of SNP fragment.

Predicted models revealed variant-dependent folding of the 51 nucleotides long pri-miR- 125a region comprising rs12976445. The modelling of C-variant 51-miR-125a with RNAfold resulted in two different structures when we applied two modes of analysis: minimizing free energy (MFE) and minimizing base-pair distance (centroid). For the U-variant using MFE and centroid modes, we obtained the same model of RNA structure. In the RNAfold MFE model of the C-variant (Figure 2a), SNP created a base-pairing near the hairpin loop. In the RNAfold centroid model of the C-variant structure, SNP was unpaired within the hairpin loop. In both RNAfold modes MFE and centroid from U-variant, SNP was the first unpaired nucleotide of a long unpaired subsequence located on 3′ end of 51-miR-125a.

Predicting the 2D structure of 51-miR-125a using RNAstructure resulted in two MFE structures for C-variant (Figure 3a,b) and a single model for U-variant. In both, MFE and centroid (Figure 3a,d, respectively), structures predicted for 51-miR-125a C-variant by RNAstructure (Figure 3), we observed the same location of SNP as in the model generated by the RNAfold in the MFE mode (Figure 2a). Moreover, all predicted U-variant foldings (Figure 2b,d and Figure 3c,e) have the same structure, independent of the 2D structure prediction software used or mode. In both structures generated with RNAstructure MFE (Figure 3a) and centroid (Figure 3d) modes, SNP was a part of base-pairing near the hairpin loop. In the remaining three models, RNAstructure MFE C-variant (in Figure 3b) and U-variant RNAstructure MFE and centroid (3c,e), SNP is a part of a long unpaired region starting with SNP (26th nucleotide) and moving along till the last (51st) nucleotide. The differences observed between U- and C- variant models in both RNAfold and RNAstructure may suggest SNP-dependent folding of the structure.

4. Discussion

The analysis of our results concerning rs12976445 SNP in miR-125a revealed that the TT genotype was slightly more frequent in breast cancer patients and HER2 positive patients. Our study revealed only a tendency, and we obtained *p*-values above 0.05. One reason for this is the participation of miR-125a with numerous other factors in the development of breast cancer. The correlation of such multigenic diseases with single SNP requires a greater number of individuals and controls. Our findings suggested that rs12976445 has the potential to be a predictive biomarker for cancer risk, but a meta-analysis of a greater number of cases is required. Several studies have described the association of rs12976445 genotypes TT, CT, and CC in miR-125a with cancer and other diseases. In the Chinese study of Jiao et al. TT genotype has been significantly related to increased risk of mortality in breast cancer patients compared with those carrying the CC genotypes [19]. miR-125a rs12976445 was significantly associated with survival in codominant, recessive, and dominant models. However, only an association under the codominant model remained significant after adjustment for lymph node metastasis, TNM stage, estrogen receptor, and progesterone receptor [19]. It has been shown in previous studies that the miR-125a level is decreased in breast cancer [32–34], however, the level of miR-125a-5p is significantly higher in younger patients than in the older ones [38]. It has been shown that SNPs located in miR-125a are associated with breast cancer tumorigenesis [39] and rs12976445 SNP in miR-125a may serve as a prognostic biomarker for breast cancer [19]. Although rs12976445 is associated with breast and prostate cancer, the impact of this SNP has been rarely studied in a functional assay. The prominent study of the effect of rs12976445 on miR-125a expression was evaluated in the context of recurrent pregnancy loss [15,40]. In embryonic kidney cells, HEK293T miR-125a expression level of C haplotype was nearly four-fold higher than T haplotype [15,40]. The question remained how the rs12976445 genotype impacts the miR-125a level.

In the previous studies, rs12976445 SNP in miR-125a has been associated with the risk of pneumonitis [20]. The expression level of miR-125a mRNA has been significantly downregulated in the CT and TT groups, and CC genotype samples demonstrated upregulated miR-125a expression [20,21]. Rs12976445 polymorphism, also associated with the risk of diabetic nephropathy, showed that the expression levels of miR-125a were approximately three times lower in patients carrying TT and

CT than in the CC [41]. Other studies showed that concomitantly with miR-125a the expression of miR-205 in breast cancer is also downregulated due to SNP variations in the miR-205 sequence [42]. Studies in various cell lines identified the differential expression of miR-205 in breast cancer cell lines in correlation with the missing number of AGC repeats [42].

Herein, we decided to analyze the association of rs12976445 genotypes with breast cancer. Moreover, we performed computational research to find potential protein interacting with the SNP region and to predict and compare the 2D structure of C-/U-variant sequence. In both cases, we used two independent tools, RBPmap and ATtRACT for RNA binding protein (RBP) search, and RNAstructure and RNAfold programs to build 2D models. We observed different RBP mapped to C- and U-variants of the sequence. Both programs that we used indicated the interaction between U-variant and PTBP1 protein. In C-variant, NOVA1 protein appears in the results obtained from both ATtRACT and RBPmap. It should be underlined that NOVA1 was found only in the C-variant sequence, whereas connection with PTBP1 protein was established in U- and C-variant by RBPmap program. Another interesting fact was the appearance of HNRNPK protein in the results for C-variant obtained from RBPmap for low and medium stringency level of the algorithm search criteria. This protein was not recognized as a possible binding protein to the U-variant. This indicated that RNA processing proteins have a different affinity to C-variant and U-variant potentially modulating the probability of pre-miR-125a RNA maturation.

We also predicted the 2D structures of 51 nt sequence with C-/U-variant. We applied the same RNAfold system as Hu et al. [15]. These authors have modelled the 2D structures of rs12976445 allele T, revealing that the rare allele T can neither change the predicted secondary structure nor the predicted ΔG [15]. Hu used 1016 nt fragment of RNA, whereas we used RNAfold with 51 nt, revealing two different 2D structures and different hairpin structures in U-variant and C-variant. The shorter sequence used in our study strengthened the obtained results of modelling as in silico methods deal better with shorter sequences. In a later paper, Hu et al. admitted the difference in 2D structure between variants of rs12976445 rare allele T using the same software as previously and the same 1016 nt RNA fragment [40]. For RNAfold-predicted structures we received different models per each variant. C-variants predicted by this tool were either a part of a hairpin loop (as an unpaired nucleotide) or as a base pair near the hairpin loop. On the contrary, U-variants in RNAfold were part of a long unpaired region on the 3' end of the sequence. RNAstructure program predicted 3 models for the C-variant sequence and 2 for the U-variant. In the Figure 3b,c,e structures were identical (except for the SNP in A1.2). Let us notice, that in Figures 2a and 3a are the same structures, and Figure 3c,d models are identical, despite using different predicting tools.

Using two computational programs, we revealed the potential differences in RNA-binding proteins between the analyzed C- and U-variant. We found that NOVA1 and HRNPK RNA-binding protein may interact with the C-variant and PTBP1 with the U-variant. Polypyrimidine tract binding protein 1 (PTBP1) binds to mRNA and regulates alternative splicing patterns [43]. In the previous reports, it has been shown that PTBP1 enhances miR-101-guided AGO2 (Argonaute) interaction with MCL1, thereby regulating miR-101-induced apoptosis and cell survival [44]. NOVA1 stimulates miRNA function by different mechanisms that converge on Argonaute proteins, a core component of the miRNA-induced silencing complex (miRISC). NOVA1 physically interacts with Ago proteins, and control neuronal miRISC function at the level of Ago proteins, with possible implications for the regulation of synapse development and plasticity [35,45]. Heterogeneous nuclear ribonucleoprotein K (hnRNPK), a ubiquitously occurring RNA-binding protein (RBP), can interact with many nucleic acids and various proteins and is involved in several cellular functions including transcription, translation, splicing, chromatin remodelling, etc. [36].

The limitation of the current study is the number of cases and controls. In the future study, a larger group of breast cancer patients should be analyzed to confirm the tendency of TT genotype association with breast cancer and to test the significance of the observed associations.

5. Conclusions

Understanding of the miRNA processing in cancer cells is an important step towards global fight against cancer. Among others, it could explain the reason why miRNA-125a level is decreased in breast cancer. Moreover, it can also help to focus the investigation on solving the upstream cascade of factors participating in abnormal regulation of miRNA in cancer. We were attempting to reveal potential features of rs12976445 SNPs of miRNA-125aby using all available resources, in vitro analysis on experimental data and using bioinformatics resources. The combination of these two approaches, in vitro and in silico, is more efficient and allows for a broader research perspective. Our analysis showed that the TT genotype of miRNA-125a was slightly more frequent in breast cancer patients and HER2 positive patients. We also demonstrated that the U-variant of rs12976445 diminished the probability of pri-miR-125a binding to NOVA1 and HNRNPK proteins. Our in silico analysis revealed that C- and U-variants could promote different RNA folding patterns that may further affect protein binding. Altogether, these may imply a lower miR-125a expression in breast cancer. These results may not only be useful for diagnostic purposes but can also contribute to the research into novel therapies for breast cancer. In a wider perspective, our experimental protocol and findings may be extended into different types of cancer or other diseases, where miRNA expression level is affected. Following this path, in future work, we would like to consider the other molecular types of cancer to extend the understanding of the impact of rs12976445 on miR-125a expression and verified the obtained in silico results by in vivo methods

References

1. Keeney, M.G.; Couch, F.J.; Visscher, D.W.; Lindor, N.M. Non-BRCA familial breast cancer: Review of reported pathology and molecular findings. *Pathology* **2017**, *49*, 363–370. [CrossRef] [PubMed]
2. Malhotra, P.; Read, G.H.; Weidhaas, J.B. Breast Cancer and miR-SNPs: The Importance of miR Germ-Line Genetics. *Noncoding RNA* **2019**, *5*, 27. [CrossRef] [PubMed]
3. Mosallaei, M.; Simonian, M.; Esmaeilzadeh, E.; Bagheri, H.; Miraghajani, M.; Salehi, A.R.; Mehrzad, V.; Salehi, R. Single nucleotide polymorphism rs10889677 in miRNAs Let-7e and Let-7f binding site of IL23R gene is a strong colorectal cancer determinant: Report and meta-analysis. *Cancer Genet.* **2019**, *239*, 46–53. [CrossRef]
4. Yan, Z.; Zhou, Z.; Li, C.; Yang, X.; Yang, L.; Dai, S.; Zhao, J.; Ni, H.; Shi, L.; Yao, Y. Polymorphisms in miRNA genes play roles in the initiation and development of cervical cancer. *J. Cancer* **2019**, *10*, 4747–4753. [CrossRef]
5. Yin, Z.; Cui, Z.; Ren, Y.; Xia, L.; Wang, Q.; Zhang, Y.; He, Q.; Zhou, B. Association between polymorphisms in pre-miRNA genes and risk of lung cancer in a Chinese non-smoking female population. *Lung Cancer* **2016**, *94*, 15–21. [CrossRef] [PubMed]
6. Bahreini, F.; Ramezani, S.; Shahangian, S.S.; Salehi, Z.; Mashayekhi, F. miR-559 polymorphism rs58450758 is linked to breast cancer. *Br. J. Biomed. Sci.* **2019**, *77*, 29–34. [CrossRef]
7. Linhares, J.J.; Azevedo, M., Jr.; Siufi, A.A.; de Carvalho, C.V.; WolgienMdel, C.; Noronha, E.C.; Bonetti, T.C.; da Silva, I.D. Evaluation of single nucleotide polymorphisms in microRNAs (hsa-miR-196a2 rs11614913 C/T) from Brazilian women with breast cancer. *BMC Med. Genet.* **2012**, *13*, 119. [CrossRef] [PubMed]

8. Zhang, N.; Huo, Q.; Wang, X.; Chen, X.; Long, L.; Jiang, L.; Ma, T.; Yang, Q. A genetic variant in pre-miR-27a is associated with a reduced breast cancer risk in younger Chinese population. *Gene* **2013**, *529*, 125–130. [CrossRef]

9. Loh, H.Y.; Norman, B.P.; Lai, K.S.; Rahman, N.; Alitheen, N.B.M.; Osman, M.A. The Regulatory Role of MicroRNAs in Breast Cancer. *Int. J. Mol. Sci.* **2019**, *20*, 4940. [CrossRef]

10. Potenza, N.; Panella, M.; Castiello, F.; Mosca, N.; Amendola, E.; Russo, A. Molecular mechanisms governing microRNA-125a expression in human hepatocellular carcinoma cells. *Sci. Rep.* **2017**, *7*, 10712. [CrossRef]

11. Li, G.; Ao, S.; Hou, J.; Lyu, G. Low expression of miR-125a-5p is associated with poor prognosis in patients with gastric cancer. *Oncol. Lett.* **2019**, *18*, 1483–1490. [CrossRef] [PubMed]

12. Sun, C.; Zeng, X.; Guo, H.; Wang, T.; Wei, L.; Zhang, Y.; Zhao, J.; Ma, X.; Zhang, N. MicroRNA-125a-5p modulates radioresistance in LTEP-a-2 non-small cell lung cancer cells by targeting SIRT7. *Cancer Biomark* **2020**, *27*, 39–49. [CrossRef] [PubMed]

13. Tang, L.; Zhou, L.; Wu, S.; Shi, X.; Jiang, G.; Niu, S.; Ding, D. miR-125a-5p inhibits colorectal cancer cell epithelial-mesenchymal transition, invasion and migration by targeting TAZ. *Onco Targets Ther.* **2019**, *12*, 3481–3489. [CrossRef] [PubMed]

14. Ma, J.; Zhan, Y.; Xu, Z.; Li, Y.; Luo, A.; Ding, F.; Cao, X.; Chen, H.; Liu, Z. ZEB1 induced miR-99b/let-7e/miR-125a cluster promotes invasion and metastasis in esophageal squamous cell carcinoma. *Cancer Lett.* **2017**, *398*, 37–45. [CrossRef] [PubMed]

15. Hu, Y.; Liu, C.M.; Qi, L.; He, T.Z.; Shi-Guo, L.; Hao, C.J.; Cui, Y.; Zhang, N.; Xia, H.F.; Ma, X. Two common SNPs in pri-miR-125a alter the mature miRNA expression and associate with recurrent pregnancy loss in a Han-Chinese population. *RNA Biol.* **2011**, *8*, 861–872. [CrossRef]

16. Morales, S.; De Mayo, T.; Gulppi, F.A.; Gonzalez-Hormazabal, P.; Carrasco, V.; Reyes, J.M.; Gomez, F.; Waugh, E.; Jara, L. Genetic Variants in pre-miR-146a, pre-miR-499, pre-miR-125a, pre-miR-605, and pri-miR-182 Are Associated with Breast Cancer Susceptibility in a South American Population. *Genes* **2018**, *9*, 427. [CrossRef]

17. Peterlongo, P.; Caleca, L.; Cattaneo, E.; Ravagnani, F.; Bianchi, T.; Galastri, L.; Bernard, L.; Ficarazzi, F.; Dall'olio, V.; Marme, F.; et al. The rs12975333 variant in the miR-125a and breast cancer risk in Germany, Italy, Australia and Spain. *J. Med. Genet.* **2011**, *48*, 703–704. [CrossRef]

18. Cai, T.; Li, J.; An, X.; Yan, N.; Li, D.; Jiang, Y.; Wang, W.; Shi, L.; Qin, Q.; Song, R.; et al. Polymorphisms in MIR499A and MIR125A gene are associated with autoimmune thyroid diseases. *Mol. Cell. Endocrinol.* **2017**, *440*, 106–115. [CrossRef]

19. Jiao, L.; Zhang, J.; Dong, Y.; Duan, B.; Yu, H.; Sheng, H.; Huang, J.; Gao, H. Association between miR-125a rs12976445 and survival in breast cancer patients. *Am. J. Transl. Res.* **2014**, *6*, 869–875.

20. Huang, X.; Zhang, T.; Li, G.; Guo, X.; Liu, X. Regulation of miR-125a expression by rs12976445 single-nucleotide polymorphism is associated with radiotherapy-induced pneumonitis in lung carcinoma patients. *J. Cell. Biochem.* **2019**, *120*, 4485–4493. [CrossRef]

21. Quan, H.Y.; Yuan, T.; Hao, J.F. A microRNA125a variant, which affects its mature processing, increases the risk of radiationinduced pneumonitis in patients with nonsmallcell lung cancer. *Mol. Med. Rep.* **2018**, *18*, 4079–4086.

22. Damodaran, M.; Paul, S.F.D.; Venkatesan, V. Genetic Polymorphisms in miR-146a, miR-196a2 and miR-125a Genes and its Association in Prostate Cancer. *Pathol. Oncol. Res.* **2018**, *26*, 193–200. [CrossRef] [PubMed]

23. Miskiewicz, J.; Tomczyk, K.; Mickiewicz, A.; Sarzynska, J.; Szachniuk, M. Bioinformatics Study of Structural Patterns in Plant MicroRNA Precursors. *Biomed. Res. Int.* **2017**, *2017*, 6783010. [CrossRef]

24. Miskiewicz, J.; Szachniuk, M. Discovering Structural Motifs in miRNA Precursors from the Viridiplantae Kingdom. *Molecules* **2018**, *23*, 1367. [CrossRef] [PubMed]

25. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **1999**, *41*, 95–98.

26. Xu, K.; Cai, H.; Shen, Y.; Ni, Q.; Chen, Y.; Hu, S.; Li, J.; Wang, H.; Yu, L.; Huang, H.; et al. [Management of corona virus disease-19 (COVID-19): The Zhejiang experience]. *Zhejiang Da Xue Xue Bao Yi Xue Ban* **2020**, *49*, 147–157. [PubMed]

27. Paz, I.; Kosti, I.; Ares, M., Jr.; Cline, M.; Mandel-Gutfreund, Y. RBPmap: A web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.* **2014**, *42*, W361–W367. [CrossRef]

28. Giudice, G.; Sanchez-Cabo, F.; Torroja, C.; Lara-Pezzi, E. ATtRACT-a database of RNA-binding proteins and associated motifs. *Database (Oxford)* **2016**, *2016*. [CrossRef]

29. Reuter, J.S.; Mathews, D.H. RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinform.* **2010**, *11*, 129. [CrossRef]

30. Gruber, A.R.; Lorenz, R.; Bernhart, S.H.; Neubock, R.; Hofacker, I.L. The Vienna RNA websuite. *Nucleic Acids Res.* **2008**, *36*, W70–W74. [CrossRef]

31. Sole, X.; Guino, E.; Valls, J.; Iniesta, R.; Moreno, V. SNPStats: A web tool for the analysis of association studies. *Bioinformatics* **2006**, *22*, 1928–1929. [CrossRef]

32. Guo, X.; Wu, Y.; Hartley, R.S. MicroRNA-125a represses cell growth by targeting HuR in breast cancer. *RNA Biol.* **2009**, *6*, 575–583. [CrossRef]

33. Scott, G.K.; Goga, A.; Bhaumik, D.; Berger, C.E.; Sullivan, C.S.; Benz, C.C. Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA miR-125a or miR-125b. *J. Biol. Chem.* **2007**, *282*, 1479–1486. [CrossRef] [PubMed]

34. Serpico, D.; Molino, L.; Di Cosimo, S. microRNAs in breast cancer development and treatment. *Cancer Treat. Rev.* **2014**, *40*, 595–604. [CrossRef] [PubMed]

35. Xin, Y.; Li, Z.; Zheng, H.; Ho, J.; Chan, M.T.V.; Wu, W.K.K. Neuro-oncological ventral antigen 1 (NOVA1): Implications in neurological diseases and cancers. *Cell Prolif.* **2017**, *50*, e12348. [CrossRef]

36. Xu, Y.; Wu, W.; Han, Q.; Wang, Y.; Li, C.; Zhang, P.; Xu, H. Post-translational modification control of RNA-binding protein hnRNPK function. *Open Biol.* **2019**, *9*, 180239. [CrossRef] [PubMed]

37. Wu, S.L.; Fu, X.; Huang, J.; Jia, T.T.; Zong, F.Y.; Mu, S.R.; Zhu, H.; Yan, Y.; Qiu, S.; Wu, Q.; et al. Genome-wide analysis of YB-1-RNA interactions reveals a novel role of YB-1 in miRNA processing in glioblastomamultiforme. *Nucleic Acids Res.* **2015**, *43*, 8516–8528. [CrossRef]

38. He, H.; Xu, F.; Huang, W.; Luo, S.Y.; Lin, Y.T.; Zhang, G.H.; Du, Q.; Duan, R.H. miR-125a-5p expression is associated with the age of breast cancer patients. *Genet. Mol. Res.* **2015**, *14*, 17927–17933. [CrossRef]

39. Li, W.; Duan, R.; Kooy, F.; Sherman, S.L.; Zhou, W.; Jin, P. Germline mutation of microRNA-125a is associated with breast cancer. *J. Med. Genet.* **2009**, *46*, 358–360. [CrossRef]

40. Hu, Y.; Huo, Z.H.; Liu, C.M.; Liu, S.G.; Zhang, N.; Yin, K.L.; Qi, L.; Ma, X.; Xia, H.F. Functional study of one nucleotide mutation in pri-miR-125a coding region which related to recurrent pregnancy loss. *PLoS ONE* **2014**, *9*, e114781. [CrossRef]

41. Li, C.; Lei, T. Rs12976445 Polymorphism is Associated with Risk of Diabetic Nephropathy Through Modulating Expression of MicroRNA-125 and Interleukin-6R. *Med. Sci. Monit.* **2015**, *21*, 3490–3497. [CrossRef]

42. Zhang, J.; Wei, B.; Hu, H.; Liu, F.; Tu, Y.; He, F. The association between differentially expressed micro RNAs in breast cancer cell lines and the micro RNA-205 gene polymorphism in breast cancer tissue. *Oncol. Lett.* **2018**, *15*, 2139–2146. [CrossRef]

43. Keppetipola, N.; Sharma, S.; Li, Q.; Black, D.L. Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2. *Crit. Rev. Biochem. Mol. Biol.* **2012**, *47*, 360–378. [CrossRef] [PubMed]

44. Cui, J.; Placzek, W.J. PTBP1 enhances miR-101-guided AGO2 targeting to MCL1 and promotes miR-101-induced apoptosis. *Cell Death Dis.* **2018**, *9*, 552. [CrossRef] [PubMed]

45. Storchel, P.H.; Thummler, J.; Siegel, G.; Aksoy-Aksel, A.; Zampa, F.; Sumer, S.; Schratt, G. A large-scale functional screen identifies Nova1 and Ncoa3 as regulators of neuronal miRNA function. *EMBO J.* **2015**, *34*, 2237–2254. [CrossRef] [PubMed]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.