

Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA

Ranhui Duan¹, ChangHui Pak^{1,2} and Peng Jin^{1,2,*}

¹Department of Human Genetics and ²Graduate Program in Genetics and Molecular Biology, Emory University School of Medicine, Atlanta, GA 30322, USA

Received January 25, 2007; Revised and Accepted March 13, 2007

MicroRNAs (miRNAs) are small non-coding RNAs that inhibit expression of specific target genes at the post-transcriptional level. Sequence variations in miRNA genes, including pri-miRNAs, pre-miRNAs and mature miRNAs, could potentially influence the processing and/or target selection of miRNAs. In this study, we have systematically identified single nucleotide polymorphisms (SNPs) associated with 227 known human miRNAs. Among 323 total SNPs that we identify, 12 are located within the miRNA precursor and one is at the eighth nucleotide (+8) of the mature miR-125a, which has been proposed to play a critical role in recognition of mRNA targets by miRNAs. Through a series of *in vivo* analyses, we unexpectedly find that this miR-125a SNP significantly blocks the processing of pri-miRNA to pre-miRNA, in addition to reducing miRNA-mediated translational suppression. Thus, our study reveals an additional structural requirement for pri-miRNA processing and emphasizes the importance of identifying new miRNA SNPs and their contributions to miRNA biogenesis and human genetic disease.

INTRODUCTION

MicroRNAs (miRNAs) are 18–25 nt, single-stranded non-coding RNAs that are generated from endogenous hairpin-shaped transcripts. miRNAs can suppress post-transcriptional gene expression by base pairing with their target messenger RNAs (mRNAs) and inducing either translational repression or mRNA degradation (1). miRNAs regulate a wide range of biological processes in animal development and human disease (1). They have been implicated in the regulation of chromosome architecture, development, growth control, apoptosis and stem cell maintenance (2). A strong link between miRNAs and cancer has been demonstrated recently, opening up a new avenue of investigation for cancer biology (3). miRNA expression profiling has also been shown to both faithfully reflect developmental lineages and classify disease states (4). To further elucidate the biological function of miRNAs, understanding their biogenesis and how this process is regulated is of utmost importance.

MiRNAs are initially transcribed as long primary transcripts (pri-miRNAs), which are then processed into ~65 nt hairpin-shaped precursor miRNAs (pre-miRNA) by the ‘Microprocessor’ complex. This nuclear complex consists of a member of the ribonuclease III family (RNaseIII), Drosha, and its cofactor,

DGCR8 (2). The pri-miRNA processing is a critical step in miRNA biogenesis, because it defines the miRNA sequences embedded in long pri-miRNAs by generating one end of the molecule (5). On the basis of both computational and biochemical analyses, a ‘ssRNA–dsRNA junction-anchoring’ model for the processing of pri-miRNA has been proposed recently (5). A typical pri-miRNA consists of a stem of ~33 bp, with a terminal loop and flanking segments. In this new model, the terminal loop is inessential, whereas the flanking ssRNA segments are critical for processing. The cleavage site is determined chiefly by the distance (~11 bp) from the stem-ssRNA junction (5). After processing by ‘Microprocessor’, pre-miRNAs are subsequently exported to cytoplasm by exportin-5 and cleaved by another RNaseIII, Dicer, to generate 18–25 nt mature miRNAs, which are then incorporated into RNA-induced silencing complexes to regulate gene expression (2). Furthermore, to recognize the target mRNA by each miRNA, a 7 nt seed sequence (at positions 2–8 from the 5′ end) in miRNAs was found to be crucial, and other nucleotide positions might contribute less dramatic yet important effects to the miRNA action (1).

Single nucleotide polymorphisms (SNPs) are the most abundant form of DNA variation in the human genome and contribute to human phenotypic differences. Polymorphisms

*To whom correspondence should be addressed at: Department of Human Genetics, Emory University School of Medicine, 615 Michael Street, Suite 301, Atlanta, GA 30322, USA. Tel: +1 4047273729; Fax: +1 4047275408; Email: pjin@genetics.emory.edu

Table 1. SNPs associated with the precursor and mature forms of human miRNAs

miRNAs	Structure	Polymorphism
hsa-mir-125a	<pre> 5'-U U U C C CUGA C CCUU UA ACCUGUGA --- GGAC G CCAG C GUAGGU CCUGA CCUU UA ACCUGUGA --- GGAC C GGUC G GGUC G GGUCUCU GGAG - UGGACACU GGGA CCUA 3'-C U C A 5'-U G GA ACAUACUUCUUUUAU GC CCAUA - UGG A C C CUCU UGU UGAA GAA AUGUA - GGUAU AUC U 3'-A A A A 5'- CUG GG GC GGGCUUAGCUCUCU GUGAGCA GG C GAC CC CG CUUGAAUCCGGUA CACUUGU CU G C C C C A 5'-U GUGUCUC UUCUCU - A GUGGUUUUACCCU UGUUAGG UU ACG UC CACGGGG CAUAGGAC GG CACCAAGAUGGGA ACCAUCU UGU U 3'-C - - - - C U A C C 5'-C - - - - U U A U C C GAGAACUGAAUU CA GGUUU UG A C G GCUAC UGUUC UAUAGG GUCCA UUCUUGACUUAA GU UCCAG - AC U G 3'-U UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C ACCGGG G C C C G UAUUGUCGU G UG UU C C 3'-G A - UC UC - C CG GG ACC GGUCA C 5'-UG CUCG C AGCUGAU - UGUGGUUAGGUUUC UGUUUGU GAGC G UUGUCUG ACAUUGAGUC GUCAAAAG ACAACGGC 3'-CGG U CUU ACU A 5'-CCA GAGG - ACACC UCCA C UCC GUCUA CCAGGU UAGACUA U GUGU CA --- GG C CUCC CAGAUU C U U C U C - - U UAAA C 3'- GG U CA - - - C C - - U UAAA C 5'- GA U CA - - - C C - - U UAAA C CAGUG GG UUGU AGGG UCCA ACCGUAUC UGA CACUUU GGCG GGG GUCAU CU - GCA UCCU CAA GGUU UGGGUAG U ACU GUGGAA UCGU CCA 3'-GUA - A C CAA C U C U C G </pre>	G/T
hsa-mir-1-1	<pre> 3'-C U C A 5'-U G GA ACAUACUUCUUUUAU GC CCAUA - UGG A C C CUCU UGU UGAA GAA AUGUA - GGUAU AUC U 3'-A A A A 5'- CUG GG GC GGGCUUAGCUCUCU GUGAGCA GG C GAC CC CG CUUGAAUCCGGUA CACUUGU CU G C C C C A 5'-U GUGUCUC UUCUCU - A GUGGUUUUACCCU UGUUAGG UU ACG UC CACGGGG CAUAGGAC GG CACCAAGAUGGGA ACCAUCU UGU U 3'-C - - - - C U A C C 5'-C - - - - U U A U C C GAGAACUGAAUU CA GGUUU UG A C G GCUAC UGUUC UAUAGG GUCCA UUCUUGACUUAA GU UCCAG - AC U G 3'-U UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C ACCGGG G C C C G UAUUGUCGU G UG UU C C 3'-G A - UC UC - C CG GG ACC GGUCA C 5'-UG CUCG C AGCUGAU - UGUGGUUAGGUUUC UGUUUGU GAGC G UUGUCUG ACAUUGAGUC GUCAAAAG ACAACGGC 3'-CGG U CUU ACU A 5'-CCA GAGG - ACACC UCCA C UCC GUCUA CCAGGU UAGACUA U GUGU CA --- GG C CUCC CAGAUU C U U C U C - - U UAAA C 3'- GG U CA - - - C C - - U UAAA C 5'- GA U CA - - - C C - - U UAAA C CAGUG GG UUGU AGGG UCCA ACCGUAUC UGA CACUUU GGCG GGG GUCAU CU - GCA UCCU CAA GGUU UGGGUAG U ACU GUGGAA UCGU CCA 3'-GUA - A C CAA C U C U C G </pre>	G/A
hsa-mir-27a	<pre> 3'-A A A A 5'- CUG GG GC GGGCUUAGCUCUCU GUGAGCA GG C GAC CC CG CUUGAAUCCGGUA CACUUGU CU G C C C C A 5'-U GUGUCUC UUCUCU - A GUGGUUUUACCCU UGUUAGG UU ACG UC CACGGGG CAUAGGAC GG CACCAAGAUGGGA ACCAUCU UGU U 3'-C - - - - C U A C C 5'-C - - - - U U A U C C GAGAACUGAAUU CA GGUUU UG A C G GCUAC UGUUC UAUAGG GUCCA UUCUUGACUUAA GU UCCAG - AC U G 3'-U UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C ACCGGG G C C C G UAUUGUCGU G UG UU C C 3'-G A - UC UC - C CG GG ACC GGUCA C 5'-UG CUCG C AGCUGAU - UGUGGUUAGGUUUC UGUUUGU GAGC G UUGUCUG ACAUUGAGUC GUCAAAAG ACAACGGC 3'-CGG U CUU ACU A 5'-CCA GAGG - ACACC UCCA C UCC GUCUA CCAGGU UAGACUA U GUGU CA --- GG C CUCC CAGAUU C U U C U C - - U UAAA C 3'- GG U CA - - - C C - - U UAAA C 5'- GA U CA - - - C C - - U UAAA C CAGUG GG UUGU AGGG UCCA ACCGUAUC UGA CACUUU GGCG GGG GUCAU CU - GCA UCCU CAA GGUU UGGGUAG U ACU GUGGAA UCGU CCA 3'-GUA - A C CAA C U C U C G </pre>	A/G
hsa-mir-140	<pre> 3'- C G C C CG CUUGAAUCCGGUA CACUUGU CU G C C C C A 5'-U GUGUCUC UUCUCU - A GUGGUUUUACCCU UGUUAGG UU ACG UC CACGGGG CAUAGGAC GG CACCAAGAUGGGA ACCAUCU UGU U 3'-C - - - - C U A C C 5'-C - - - - U U A U C C GAGAACUGAAUU CA GGUUU UG A C G GCUAC UGUUC UAUAGG GUCCA UUCUUGACUUAA GU UCCAG - AC U G 3'-U UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C ACCGGG G C C C G UAUUGUCGU G UG UU C C 3'-G A - UC UC - C CG GG ACC GGUCA C 5'-UG CUCG C AGCUGAU - UGUGGUUAGGUUUC UGUUUGU GAGC G UUGUCUG ACAUUGAGUC GUCAAAAG ACAACGGC 3'-CGG U CUU ACU A 5'-CCA GAGG - ACACC UCCA C UCC GUCUA CCAGGU UAGACUA U GUGU CA --- GG C CUCC CAGAUU C U U C U C - - U UAAA C 3'- GG U CA - - - C C - - U UAAA C 5'- GA U CA - - - C C - - U UAAA C CAGUG GG UUGU AGGG UCCA ACCGUAUC UGA CACUUU GGCG GGG GUCAU CU - GCA UCCU CAA GGUU UGGGUAG U ACU GUGGAA UCGU CCA 3'-GUA - A C CAA C U C U C G </pre>	C/A
hsa-mir-146	<pre> 3'- C G C C CG CUUGAAUCCGGUA CACUUGU CU G C C C C A 5'-U GUGUCUC UUCUCU - A GUGGUUUUACCCU UGUUAGG UU ACG UC CACGGGG CAUAGGAC GG CACCAAGAUGGGA ACCAUCU UGU U 3'-C - - - - C U A C C 5'-C - - - - U U A U C C GAGAACUGAAUU CA GGUUU UG A C G GCUAC UGUUC UAUAGG GUCCA UUCUUGACUUAA GU UCCAG - AC U G 3'-U UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C ACCGGG G C C C G UAUUGUCGU G UG UU C C 3'-G A - UC UC - C CG GG ACC GGUCA C 5'-UG CUCG C AGCUGAU - UGUGGUUAGGUUUC UGUUUGU GAGC G UUGUCUG ACAUUGAGUC GUCAAAAG ACAACGGC 3'-CGG U CUU ACU A 5'-CCA GAGG - ACACC UCCA C UCC GUCUA CCAGGU UAGACUA U GUGU CA --- GG C CUCC CAGAUU C U U C U C - - U UAAA C 3'- GG U CA - - - C C - - U UAAA C 5'- GA U CA - - - C C - - U UAAA C CAGUG GG UUGU AGGG UCCA ACCGUAUC UGA CACUUU GGCG GGG GUCAU CU - GCA UCCU CAA GGUU UGGGUAG U ACU GUGGAA UCGU CCA 3'-GUA - A C CAA C U C U C G </pre>	C/G
hsa-mir-194-2	<pre> 3'-U UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C ACCGGG G C C C G UAUUGUCGU G UG UU C C 3'-G A - UC UC - C CG GG ACC GGUCA C 5'-UG CUCG C AGCUGAU - UGUGGUUAGGUUUC UGUUUGU GAGC G UUGUCUG ACAUUGAGUC GUCAAAAG ACAACGGC 3'-CGG U CUU ACU A 5'-CCA GAGG - ACACC UCCA C UCC GUCUA CCAGGU UAGACUA U GUGU CA --- GG C CUCC CAGAUU C U U C U C - - U UAAA C 3'- GG U CA - - - C C - - U UAAA C 5'- GA U CA - - - C C - - U UAAA C CAGUG GG UUGU AGGG UCCA ACCGUAUC UGA CACUUU GGCG GGG GUCAU CU - GCA UCCU CAA GGUU UGGGUAG U ACU GUGGAA UCGU CCA 3'-GUA - A C CAA C U C U C G </pre>	C/T
hsa-mir-196a-2	<pre> 3'- C G C C CG CUUGAAUCCGGUA CACUUGU CU G C C C C A 5'-U GUGUCUC UUCUCU - A GUGGUUUUACCCU UGUUAGG UU ACG UC CACGGGG CAUAGGAC GG CACCAAGAUGGGA ACCAUCU UGU U 3'-C - - - - C U A C C 5'-C - - - - U U A U C C GAGAACUGAAUU CA GGUUU UG A C G GCUAC UGUUC UAUAGG GUCCA UUCUUGACUUAA GU UCCAG - AC U G 3'-U UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C ACCGGG G C C C G UAUUGUCGU G UG UU C C 3'-G A - UC UC - C CG GG ACC GGUCA C 5'-UG CUCG C AGCUGAU - UGUGGUUAGGUUUC UGUUUGU GAGC G UUGUCUG ACAUUGAGUC GUCAAAAG ACAACGGC 3'-CGG U CUU ACU A 5'-CCA GAGG - ACACC UCCA C UCC GUCUA CCAGGU UAGACUA U GUGU CA --- GG C CUCC CAGAUU C U U C U C - - U UAAA C 3'- GG U CA - - - C C - - U UAAA C 5'- GA U CA - - - C C - - U UAAA C CAGUG GG UUGU AGGG UCCA ACCGUAUC UGA CACUUU GGCG GGG GUCAU CU - GCA UCCU CAA GGUU UGGGUAG U ACU GUGGAA UCGU CCA 3'-GUA - A C CAA C U C U C G </pre>	C/T
hsa-mir-199b	<pre> 3'- C G C C CG CUUGAAUCCGGUA CACUUGU CU G C C C C A 5'-U GUGUCUC UUCUCU - A GUGGUUUUACCCU UGUUAGG UU ACG UC CACGGGG CAUAGGAC GG CACCAAGAUGGGA ACCAUCU UGU U 3'-C - - - - C U A C C 5'-C - - - - U U A U C C GAGAACUGAAUU CA GGUUU UG A C G GCUAC UGUUC UAUAGG GUCCA UUCUUGACUUAA GU UCCAG - AC U G 3'-U UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C ACCGGG G C C C G UAUUGUCGU G UG UU C C 3'-G A - UC UC - C CG GG ACC GGUCA C 5'-UG CUCG C AGCUGAU - UGUGGUUAGGUUUC UGUUUGU GAGC G UUGUCUG ACAUUGAGUC GUCAAAAG ACAACGGC 3'-CGG U CUU ACU A 5'-CCA GAGG - ACACC UCCA C UCC GUCUA CCAGGU UAGACUA U GUGU CA --- GG C CUCC CAGAUU C U U C U C - - U UAAA C 3'- GG U CA - - - C C - - U UAAA C 5'- GA U CA - - - C C - - U UAAA C CAGUG GG UUGU AGGG UCCA ACCGUAUC UGA CACUUU GGCG GGG GUCAU CU - GCA UCCU CAA GGUU UGGGUAG U ACU GUGGAA UCGU CCA 3'-GUA - A C CAA C U C U C G </pre>	C/T
hsa-mir-220	<pre> 3'- C G C C CG CUUGAAUCCGGUA CACUUGU CU G C C C C A 5'-U GUGUCUC UUCUCU - A GUGGUUUUACCCU UGUUAGG UU ACG UC CACGGGG CAUAGGAC GG CACCAAGAUGGGA ACCAUCU UGU U 3'-C - - - - C U A C C 5'-C - - - - U U A U C C GAGAACUGAAUU CA GGUUU UG A C G GCUAC UGUUC UAUAGG GUCCA UUCUUGACUUAA GU UCCAG - AC U G 3'-U UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C UGUUC C C C C C CU GUAACAGCA AC CC AUG AGUGC C ACCGGG G C C C G UAUUGUCGU G UG UU C C 3'-G A - UC UC - C CG GG ACC GGUCA C 5'-UG CUCG C AGCUGAU - UGUGGUUAGGUUUC UGUUUGU GAGC G UUGUCUG ACAUUGAGUC GUCAAAAG ACAACGGC 3'-CGG U CUU ACU A 5'-CCA GAGG - ACACC UCCA C UCC GUCUA CCAGGU UAGACUA U GUGU CA --- GG C CUCC CAGAUU C U U C U C - - U UAAA C 3'- GG U CA - - - C C - - U UAAA C 5'- GA U CA - - - C C - - U UAAA C CAGUG GG UUGU AGGG UCCA ACCGUAUC UGA CACUUU GGCG GGG GUCAU CU - GCA UCCU CAA GGUU UGGGUAG U ACU GUGGAA UCGU CCA 3'-GUA - A C CAA C U C U C G </pre>	G/C

Red color indicates mature mRNA and blue indicates the position of SNP.

in miRNA genes could potentially alter various biological processes by influencing the processing and/or target selection of miRNAs. In this study, we have systematically identified SNPs associated with 227 known human miRNAs, and we find multiple SNPs within both pri-miRNAs and pre-miRNAs. One of these SNPs in particular is located within the mature form of miR-125a, which drastically alters its processing from pri-miRNA to pre-miRNA. Our study reveals an additional, formerly unknown structural requirement for pri-miRNA processing and underscores the need for continued exploration of other new miRNA SNPs that may contribute to miRNA biogenesis and human disease.

RESULTS

Identification of SNPs associated with miRNA genes

To identify the SNPs associated with known human miRNAs, we began with a bioinformatics search of a human SNP database (<http://www.ncbi.nlm.nih.gov/SNP/>) based on 227 known human pre-miRNA sequences. Although the lengths of pri-miRNAs can vary significantly, it has been shown that an ~100 bp flanking sequence on either side of pre-miRNAs is sufficient for miRNA biogenesis. Thus, for each given miRNA, we focussed on spans of ~420 bp, which covers the mature miRNA and 200 bp of flanking sequence on each side. For the 227 known miRNAs examined, we identified 323 SNPs (Supplementary Table Online), and 12 of these are located within the miRNA precursors (Table 1). These

SNPs could potentially modulate the expression and processing of miRNAs.

Interestingly, a G-U polymorphism was identified at the eighth nucleotide of mature miR-125a (miR-125a-U) (Fig. 1A). In the human genome, miR-125a is located on chromosome 19q13.41. To further confirm the presence of this SNP in the general population, we designed a SNP TaqMan assay to distinguish these two alleles. We screened the CEPH panel (1200 individuals with diverse ethnic backgrounds) and identified one individual who was heterozygous at this site (major allele, G; minor allele, U). As the second to seventh or eighth nucleotides of mature miRNAs have been proposed to act as critical seed sequences for miRNA-mediated translational control, this polymorphism could principally alter the selection of mRNA targets (1).

miR-125a SNP blocks the processing of pri-miRNA to miRNA precursor

To explore the functional significance of this SNP, we generated expression vectors for the major and minor alleles of miR-125a (miR-125a-G and miR-125a-U). Upon transfection into human HEK293 cells, which do not express miR-125a endogenously, we detected the mature form of miR-125a-G by northern blot or quantitative RT-PCR (Figs 1B and 2B); however, we failed to detect the mature form of miR-125a-U (Figs 1B and 2B). This could be due to either altered miRNA processing or a defect in miRNA biogenesis machinery in the transfected cells. To exclude the possibility that the expression of

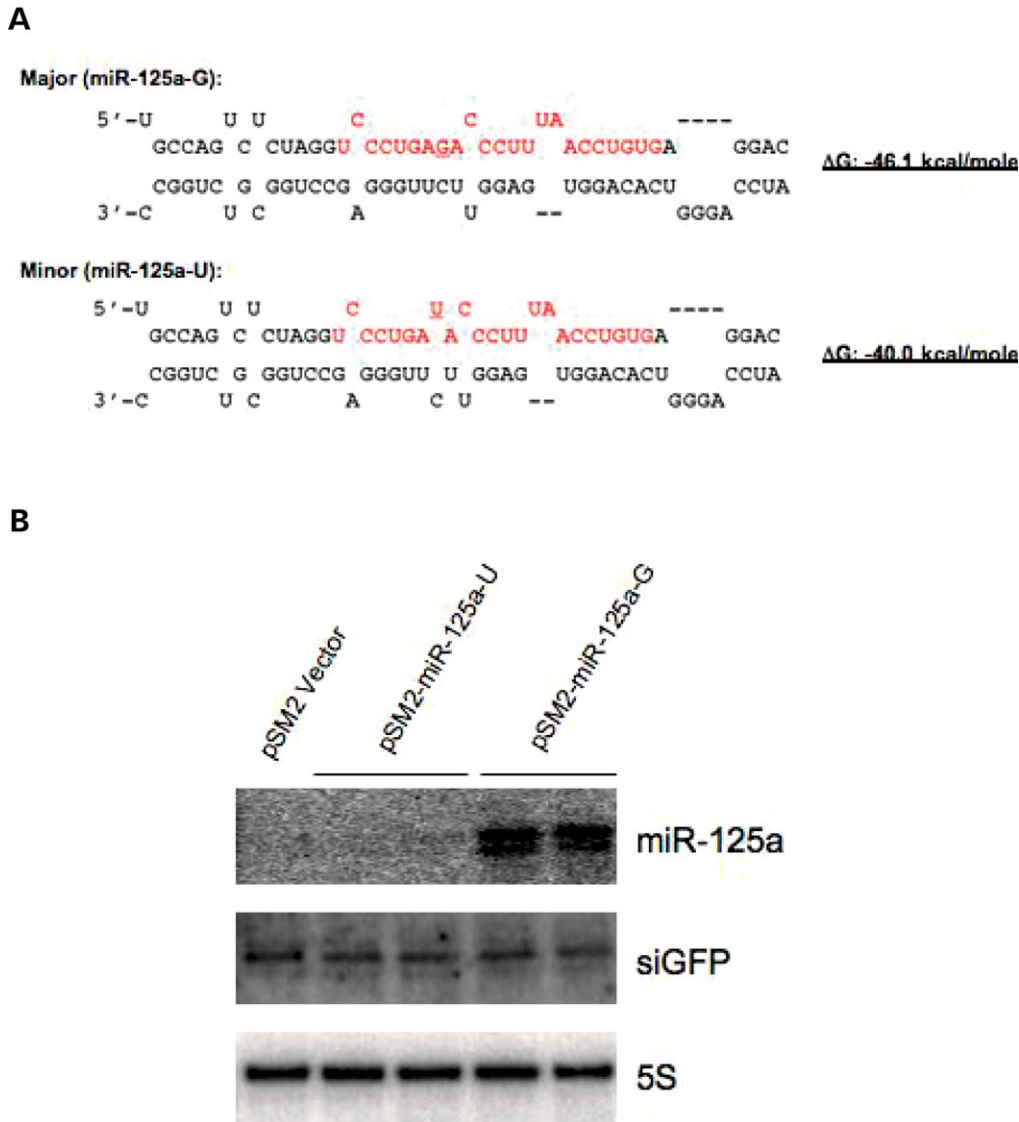


Figure 1. The SNP associated with mature miR-125a alters miRNA biogenesis. (A) The SNP associated with mature miR-125a. The hairpin structures of both major and minor alleles, as predicted by Mfold, are shown (6). Mature miRNAs are highlighted in red, and the polymorphic nucleotide is underlined. The free energy calculated by Mfold is indicated on the right. (B) The minor allele of miR-125a fails to produce mature miRNAs. Human HEK293 cells were transfected with the miR-125a expression vector, indicated on the top, along with a control pSM2-siGFP expression vector. The total RNAs (5 mg) isolated from transfected cells were used for northern blots as shown. 5S RNA was used as a loading control.

miR-125a-U interferes with the miRNA processing machinery, we co-transfected a siRNA-expressing vector (pSM2-siGFP) with the miR-125a expression vectors. We observed comparable siGFP expression in the cells transfected with either miR-125a construct (Fig. 1B). These results suggest that the miR-125a SNP alters the biogenesis of miR-125a.

To determine which step of the miRNA biogenesis is being altered by this SNP, we have designed specific primers to recognize either pri-miRNA alone or both pri-miRNA and pre-miRNA (Fig. 2A). Using these primers and the mature miRNA TaqMan assay, we performed quantitative real-time PCR to determine the relative expression levels of the pri-, pre- and mature forms of miR-125a-G and miR-125a-U in transfected cells. Whereas the level of pri-miR-125a-U was increased by

~20% compared with pri-miR-125a-G, the overall level of pri- and pre-miR-125a-U was only 50% of that of miR-125a-G (Fig. 2B). These data indicate that the SNP within the mature miR-125a sequence alters the processing of pri-miRNA to pre-miRNA.

Altered processing by the miR-125a SNP is RNA secondary structure-dependent

We used the Mfold program to examine how this SNP could alter the RNA secondary structure predicted for the major allele of miR-125a (6). We found that this SNP could introduce a base-pairing mismatch, alter free energy values and

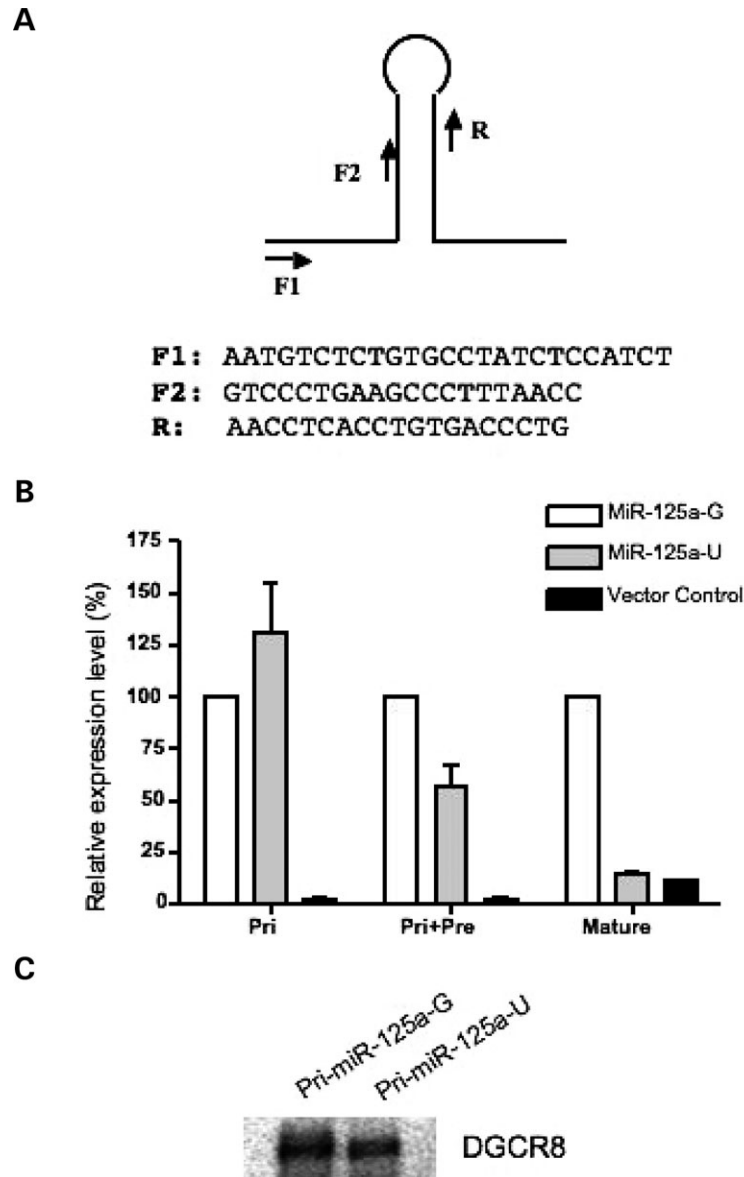


Figure 2. miR-125a SNP blocks the processing of pri-miRNA to pre-miRNA. (A) The positions and sequences of the primers used for real-time PCR to quantify the level of pri-miRNA and pre-miRNA are indicated. Note that the primer for pre-miRNA (F2) also recognizes the corresponding pri-miRNA and therefore measures both forms. F1 and R are for pri-miRNA, whereas F2 and R for both pri- and pre-miRNA. (B) Quantitative RT-PCR was used to measure levels of the pri-, pre- and mature forms of miR-125a-G and miR-125a-U in transfected HEK293 cells. The relative expression levels as determined by $\Delta\Delta C_t$ analyses are shown. Transfection was repeated three times. (C) RNA-binding assay with DGCR8 protein and pri-miR-125a (both major and minor alleles). The same amounts of DGCR8 protein and labeled pri-miR-125a were used for RNA-binding reactions.

create an enlarged RNA bulge in the predicted structure (Fig. 1A). It has previously been shown that recognition of the ssRNA-dsRNA junction and adjacent ~ 11 bp stem by DGCR8 is critical for the processing of pri-miRNA (5). However, the miR-125a SNP does not alter this junction structure. We performed RNA-binding assays using DGCR8 protein and pri-miR-125a RNA and found that the miR-125a SNP did not alter the interaction between DGCR8 and pri-miRNA (Fig. 2C). This suggests that additional sequences or structures may be required for efficient processing of

pri-miRNA. To test this possibility, we generated a double mutant (miR-125a-U-M) that contained both the G-U polymorphism from miR-125a-U and a complementary C-A base change that would preserve the predicted secondary structure of pri- and pre-miR-125a (Fig. 3A). We performed transfection using this construct and found that miR-125a-U-M could be processed normally, as determined by northern blot and quantitative real-time PCR. The ratio between mature and pri-miRNA using miR-125a-U-M construct is similar to the ratio using construct expressing major

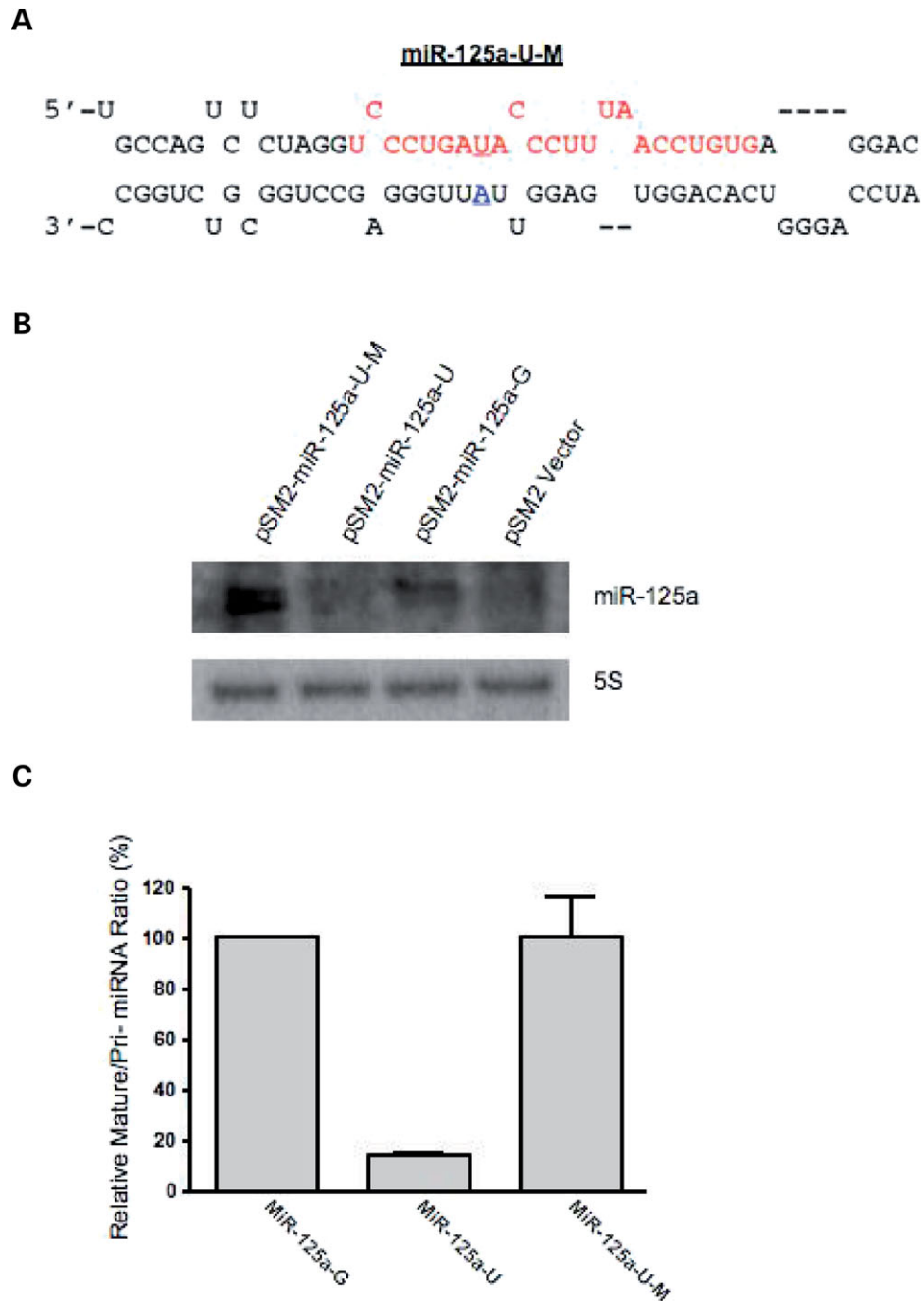


Figure 3. Altered processing by the miR-125a SNP is RNA secondary structure-dependent. (A) The hairpin structure of double-mutant miR-125a (miR-125a-U-M), which will preserve the predicted secondary structure as predicted by Mfold, is shown. Mature miRNA is highlighted in red, and the mutation on the opposite strand is indicated in blue. Both polymorphic and mutated nucleotides are underlined. (B) The double-mutant miR-125a produces mature miRNAs normally. Human HEK293 cells were transfected with the miR-125a expression vectors indicated on the top. The total RNAs (5 mg) isolated from transfected cells were used for northern blots as shown. 5S RNA was used as a loading control. (C) Quantitative RT-PCR was performed to calculate the mature miRNA to pri-miRNA ratio in the cells expressing miR-125a-G, miR-125a-U and miR-125a-U-M.

allele, which suggests that the processing of pri-miRNA has returned to normal (Fig. 3B and C). These data indicate that in addition to the ssRNA-dsRNA junction and ~11 bp stem, a structure of ~22 bp stem could also be critical for pri-miRNA processing.

miR-125a SNP attenuates miRNA-mediated translational suppression

To determine the functionality of this SNP, we examined whether it could modulate translational regulation mediated by

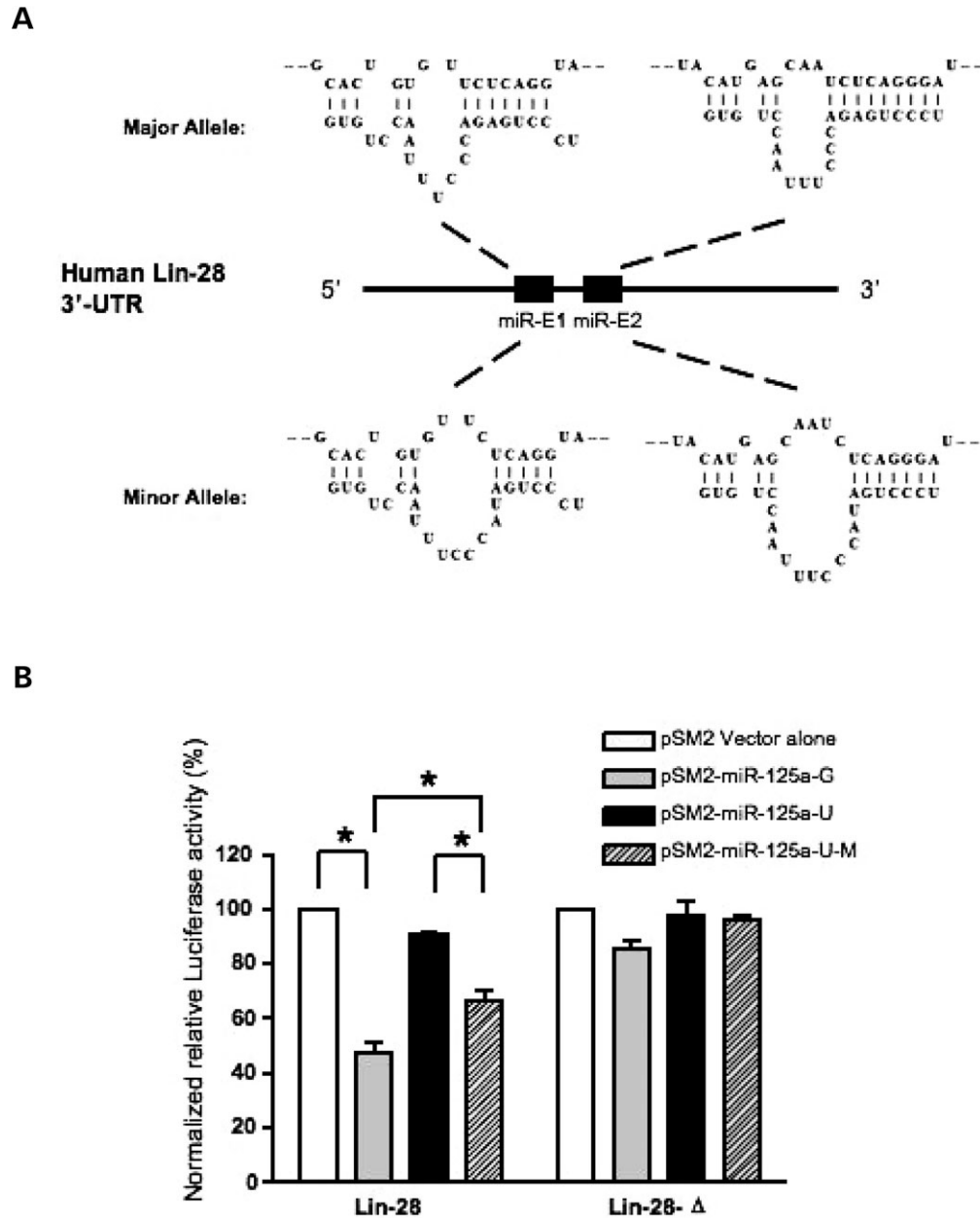


Figure 4. miR-125a SNP reduces miRNA-mediated translational suppression. (A) A diagram showing the predicted base pairing between both major and minor alleles of miR-125a and the 3'-UTR of human Lin-28 mRNA. miR-E1 and miR-E2 are the target sites of miR-125a within the 3'-UTR of human Lin-28 mRNA. (B) The SNP associated with mature miR-125a alters its translational suppression efficiency. Luciferase constructs either containing (Lin-28) or lacking (Lin-28-Δ) miR-125a-targeting sequences were co-transfected with the indicated miR-125a expression vectors into HEK293 cells. Normalized relative luciferase activities are shown. * $P < 0.001$.

miR-125a in a reporter gene system. In this previously established system, miR-125a could suppress the translation of a luciferase reporter containing the 3'-UTR of the human Lin-28 mRNA (7). The suppression depends on the presence of miR-125a targeting sequences in the 3'-UTR. Compared with the major allele, the minor allele reduces the stability of the pairing between miRNA and mRNA (Fig. 4A). There was no translational suppression with pSM2-miR-125a-U vector, as no significant mature miR-125a was produced (Fig. 4B). We performed

co-transfection experiments using luciferase constructs either containing (Lin-28) or lacking (Lin-28-Δ) miR-125a target sequences and the various miR-125a expression vectors. We found that despite the higher expression level of mature miR-125a-U in the cells transfected with pSM2-miR-125a-U-M expression vector, the miR-125a-mediated translational suppression was substantially more compromised, suggesting that this SNP could also alter the efficiency of miR-125a-mediated translational suppression (Fig. 4B and data not shown).

DISCUSSION

MiRNAs comprise a growing class of non-coding RNAs that are believed to regulate gene expression via translational repression (2). Sequence variations within the miRNA genes, including pri-miRNAs, pre-miRNAs and mature miRNAs, could potentially influence the processing and/or target selection of miRNAs. In this study, we perform a systematic analysis of SNPs associated with all known miRNAs in the human genome to date. Among 323 total SNPs that we identify, 12 are located within the miRNA precursor and one is at the eighth nucleotide (+8) within the mature form of miR-125a, which has been proposed to play a critical role in the recognition of mRNA targets by miRNA. Through a series of *in vivo* analyses, we demonstrate that this miR-125a SNP blocks the processing of pri-miRNA to pre-miRNA significantly in addition to reducing miRNA-mediated translational suppression. These data suggest that SNPs that reside within the miRNA genes could indeed regulate miRNA biogenesis and alter target selection, thereby potentially having profound biological effects.

Both segmental copy number variations (CNVs) and SNPs in the human genome can contribute greatly to the genetic basis of human phenotypic differences. Among the identified CNVs, 14 of the high-frequency CNV loci were found to harbor 21 human miRNAs, indicating that these CNVs could lead to different expression levels of those miRNAs (8). The biological consequences of these miRNA CNVs remain to be determined. SNPs are the most abundant form of DNA variation in the human genome. Polymorphisms in miRNA genes could potentially impact various biological processes by influencing the processing and/or target selection of miRNAs. This notion is supported by our current study of the SNP associated with mature miR-125a. Our findings also illustrate the importance of further research into the biological roles of other miRNA-associated SNPs, many of which have yet to be identified.

Our results suggest that the miR-125a SNP alters the processing of pri-miRNA to pre-miRNA in an RNA structure-dependent manner. In our transfection experiments, we failed to detect the mature miRNA when this SNP is introduced into the pri-miRNA. Furthermore, we observed an accumulation of pri-miRNAs and a substantial reduction of pre-miRNAs when this minor allele was expressed. Interestingly, this alteration in miRNA biogenesis is RNA secondary structure-dependent. Although this SNP introduces a base-pairing mismatch, alters free energy values and creates an enlarged RNA bulge in the predicted structure, it nonetheless does not change the structure deemed essential for pri-miRNA processing by the 'ssRNA-dsRNA junction-anchoring' model proposed recently (5). In such a junction-anchoring model, DGCR8 could play a major role in substrate recognition by directly anchoring at the ssRNA-dsRNA junction. DGCR8 also interacts with the stem of dsRNA and the terminal loop for full activity, although the terminal loop structure is not critical for the DGCR8 binding and cleavage reaction. After the initial recognition step, Droscha may transiently interact with the substrate for catalysis. The processing center of Droscha is placed at ~11 bp from the basal segments (5). Then, the question becomes how the miR-125a SNP alters the processing of pri-miRNA, and there could be two possibilities. One is that there

may be additional structural requirements for the efficient processing of pri-miRNA by the Droscha-DGCR8 complex. Alternatively, there may be additional proteins capable of binding to pri-miRNAs and modulating the processing of pri-miRNAs. These interactions could contribute to either specific or differential expressions of miRNAs, and the miR-125a SNP would be a candidate modulator of such interactions.

In the human genome, miR-125a is located at chromosome 19q13.41, a region that is frequently deleted in primary gliomas, especially oligodendrogliomas (9). Identified targets of miR-125a include Lin-28, Lin-41, ERBB2 and ERBB3 (7,10,11). MiR-125a was found to be downregulated in one breast cancer miRNA profiling study (12). Interestingly, a more recent study has revealed that both miR-125a and miR-125b were significantly downregulated in ERBB2-amplified and over-expressing breast cancers, which clinically matched against ERBB2-negative human breast cancers (11). Furthermore, it has also been demonstrated that enforced expression of miR-125 could suppress the expression of both ERBB2 and ERBB3. Consistent with the suppression of both ERBB2 and ERBB3 signaling, breast cancer cells overexpressing miR-125a or miR-125b were impaired in their anchorage-dependent growth and exhibited reduced migration and invasion capacities (11). These observations raise the possibility that miR-125 could play important roles in the pathogenesis of breast cancer. Given that the SNP associated with miR-125a identified here could block the processing of pri-miR-125a and abolish the production of mature miR-125a, further exploration of the relationship between this SNP and development of ERBB2-dependent human breast cancer could prove fruitful.

In summary, we have systematically identified SNPs associated with 227 known human miRNAs. Through this analysis, we identified a particular SNP in miR-125a that revealed an additional, formerly unknown structural requirement for the processing of pri-miRNA and that reduced translational suppression by miR-125a. Further characterization of these miRNA-associated SNPs would likely improve our understanding of miRNA biogenesis and the potential contribution of these SNPs to normal human variation and disease pathogenesis.

MATERIALS AND METHODS

DNA plasmids

To generate pSM2-miR-125a-G, sequence corresponding to the miR-125a precursor and 125 bp flanking region on each side was amplified from normal human genomic DNA and inserted into the pSM2 vector (Open Biosystems). Plasmid pSM2-miR125a-U was generated by introducing a G-T substitution at the sequence corresponding to the eighth position of the mature miR-125a (Stratagene, La Jolla, CA, USA). A double mutant (pSM2-miR125a-U-M) contained the G-T polymorphism in addition to a complementary C-A base change, as described in the text. Lin-28 reporter constructs were reported previously (7).

Cell culture, transfection and luciferase assay

Human HEK293 cells were grown in Dulbecco's modified Eagle's medium (GIBCO) supplemented with 10% FBS and

penicillin/streptomycin. Plasmids were transfected into cells by Lipofectamine 2000 (Invitrogen) following the manufacturer's protocol. Reporter plasmids and pRL-TK plasmid were used for transfection, and each sample was transfected in triplicate. Luciferase assays were performed in 48 h using the Dual-Luciferase reporter 1000 assay systems (Promega, Madison, WI, USA) and the Luminometer TD-20/20 (Turner Designs).

RNA-binding assay

PCK-FLAG-DGCR8 plasmid was transfected into human HEK293 grown in a 100 mm dish using Lipofectamine 2000 (Invitrogen) following the manufacturer's protocol. Purification of the immunoprecipitated FLAG-DGCR8 protein was done as previously described (5). To generate the probe for RNA-binding assay, 100 ng of pSM2-miR-125a-G and pSM2-miR-125a-U was used as templates, with the forward primer containing T7 promoter sequences at their 5' ends. Primer sequences are 5'-TAATACGACTCACTATAGGGTG CCTATCTCCATCTC-3' (forward primer) and 5'-TGCTCT GGAGGAAGGGTATG-3' (reverse primer). The PCR products were then gel-purified using QIAquick gel extraction kit (Qiagen) and used as templates (109 ng) for *in vitro* transcription using RNAMaxx high-yield transcription kit (Stratagene) with ³²P-UTP. After phenol/chloroform extraction and ethanol precipitation, the labeled probes were incubated at 95°C for 1 min and 37°C for 1 h for annealing to form a hairpin structure. Then, 1 × 10⁶ c.p.m. of the labeled probes were incubated with 1 μg of immunoprecipitated recombinant protein for binding and UV crosslinking as previously described (5). Reaction was loaded on 7.5% SDS-polyacrylamide gel and visualized by PhosphorImager (Amersham Biosciences).

Northern blot of small RNAs

RNAs were isolated with Trizol (Invitrogen) and separated on 15% TBE urea gel, then transferred and UV crosslinked to nylon membrane (GE Osmonics Lab Store, Minnetonka, MN, USA). The ³²P-UTP-labeled probe was prepared with the Ambion mirVana miRNA Probe Construction Kit. Both anti-miR-125a and anti-miR-125a-SNP were chemically synthesized and mixed at a 1:1 ratio as templates for *in vitro* transcription. The membranes were pre-hybridized at 65°C for 1 h and hybridized for 12–16 h at room temperature. Membranes were then washed three times at RT and twice at 42°C. Membranes were exposed and scanned with a Typhoon 9200 PhosphorImager (Amersham Biosciences).

Quantitative RT-PCR

Total RNAs from cultured cells were isolated with Trizol (Invitrogen). RNA samples were reverse-transcribed into cDNA with the ThermoScript First-Strand Synthesis System (Invitrogen). Real-time PCR was performed with gene-specific primers and Power SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA) using the 7500 Standard Real-Time PCR System (Applied Biosystems). The primers were 5'-AATGTCTCTGTGCCTATCTCCATCT-3' (F1),

5'-GTCCCTGAAGCCCTTTAACC-3' (F2) and 5'-AACCT CACCTGTGACCTG-3' (R). Relative quantities of mature miRNAs were determined using Applied Biosystems TaqMan® microRNA Assays. Relative quantities of miRNA were determined using the ΔΔCt method, as provided by the manufacturer. Individual assays were performed using primers and probes provided by Applied Biosystems.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We thank J. Belasco for Lin-28 reporter constructs, N. Kim for DGCR8 plasmid and C. Strauss and K. Garber for critical reading of the manuscript. P.J. is supported by NIH grants R01 NS051630 and R01 MH076090. P.J. is a recipient of the Beckman Young Investigator Award and the Basil O'Connor Scholar Research Award, as well as an Alfred P Sloan Research Fellowship in Neuroscience. We would like to dedicate this paper to the memory of our friend and mentor, Dr. Dean Danner.

Conflict of Interest statement. None declared.

REFERENCES

- Plasterk, R.H. (2006) Micro RNAs in animal development. *Cell*, **124**, 877–881.
- Zamore, P.D. and Haley, B. (2005) Ribo-gnome: the big world of small RNAs. *Science*, **309**, 1519–1524.
- Croce, C.M. and Calin, G.A. (2005) miRNAs, cancer, and stem cell division. *Cell*, **122**, 6–7.
- 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.
- Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T. and Kim, V.N. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, **125**, 887–901.
- Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
- Wu, L. and Belasco, J.G. (2005) Micro-RNA regulation of the mammalian lin-28 gene during neuronal differentiation of embryonal carcinoma cells. *Mol. Cell. Biol.*, **25**, 9198–9208.
- Wong, K.K., deLeeuw, R.J., Dosanjh, N.S., Kimm, L.R., Cheng, Z., Horsman, D.E., MacAulay, C., Ng, R.T., Brown, C.J., Eichler, E.E. *et al.* (2007) A comprehensive analysis of common copy-number variations in the human genome. *Am. J. Hum. Genet.*, **80**, 91–104.
- Law, M.E., Templeton, K.L., Kitange, G., Smith, J., Misra, A., Feuerstein, B.G. and Jenkins, R.B. (2005) Molecular cytogenetic analysis of chromosomes 1 and 19 in glioma cell lines. *Cancer Genet. Cytogenet.*, **160**, 1–14.
- Schulman, B.R., Esquela-Kerscher, A. and Slack, F.J. (2005) Reciprocal expression of lin-41 and the microRNAs let-7 and mir-125 during mouse embryogenesis. *Dev. Dyn.*, **234**, 1046–1054.
- Scott, G.K., Goga, A., Bhaumik, D., Berger, C.E., Sullivan, C.S. and Benz, C.C. (2007) Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA miR-125a or miR-125b. *J. Biol. Chem.*, **282**, 1479–1486.
- Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M. *et al.* (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.*, **65**, 7065–7070.