Title

Extensive sequence variation in the 3′ untranslated region of the KRAS gene in lung and ovarian cancer cases

Journal

Cell Cycle, 13(6)

ISSN

1538-4101

Authors

Kim, Minlee
Chen, Xiaowei
Chin, Lena
et al.

Publication Date

2014-03-15

DOI

10.4161/cc.27941

Peer reviewed

Cell Cycle 13:6, 1030–1040; March 15, 2014; © 2014 Landes Bioscience

Extensive sequence variation in the 3' untranslated region of the *KRAS* gene in lung and ovarian cancer cases

Minlee Kim[1], Xiaowei Chen[1,2], Lena J Chin[1], Trupti Paranjape[3], William C Speed[4], Kenneth K Kidd[4], Hongyu Zhao[2,4], Joanne B Weidhaas[3], and Frank J Slack[1,*]

[1]Department of Molecular, Cellular, and Developmental Biology; Yale University; New Haven, CT USA; [2]Program in Computational Biology and Bioinformatics; Yale University School of Medicine; New Haven, CT USA; [3]Department of Therapeutic Radiology; Yale University School of Medicine; New Haven, CT USA; [4]Department of Genetics; Yale University School of Medicine; New Haven, CT USA

Keywords: KRAS, 3′ UTR, microRNA, *let-7*, sequence variant, SNP, non-small cell lung cancer, NSCLC, epithelial ovarian cancer, EOC

While cancer is a serious health issue, there are very few genetic biomarkers that predict predisposition, prognosis, diagnosis, and treatment response. Recently, sequence variations that disrupt microRNA (miRNA)-mediated regulation of genes have been shown to be associated with many human diseases, including cancer. In an early example, a variant at one particular single nucleotide polymorphism (SNP) in a *let-7* miRNA complementary site in the 3′ untranslated region (3′ UTR) of the *KRAS* gene was associated with risk and outcome of various cancers. The *KRAS* oncogene is an important regulator of cellular proliferation, and is frequently mutated in cancers. To discover additional sequence variants in the 3′ UTR of *KRAS* with the potential as genetic biomarkers, we resequenced the complete region of the 3′ UTR of *KRAS* in multiple non-small cell lung cancer and epithelial ovarian cancer cases either by Sanger sequencing or capture enrichment followed by high-throughput sequencing. Here we report a comprehensive list of sequence variations identified in cases, with some potentially dysregulating expression of *KRAS* by altering putative miRNA complementary sites. Notably, rs712, rs9266, and one novel variant may have a functional role in regulation of *KRAS* by disrupting complementary sites of various miRNAs, including *let-7* and miR-181.

Introduction

KRAS is a member of the RAS GTPase family known to regulate a variety of biological processes, including cellular proliferation and apoptosis.[1] The mutations in the coding region of the *KRAS* gene, which result in hyper-activation of the protein, are one of the major oncogenic mutation classes found in human cancer.[1] For example, about 17% of lung cancer, the leading cause of cancer mortality worldwide, and 14% of ovarian cancer, the most common cause of gynecologic cancer mortality, contain *KRAS* activating mutations (statistics from CDC, and ref. 2). However, limited treatment options and the heterogeneous nature of cancer pose a challenge in targeting this pathway. In addition, there are very few genetic biomarkers known to predict cancer risk, prognosis, diagnosis, and treatment sensitivity.

Mounting evidence in recent studies suggests that variants at single nucleotide polymorphisms (SNPs) in the 3′ untranslated region (3′ UTR) have the potential to be such genetic biomarkers.[3-6] microRNAs (miRNAs) negatively regulate gene expression at the post-transcriptional level by binding to complementary sites in the 3′ UTR of target messenger RNAs (mRNAs). A recent computational study identified 175 351 SNPs in human 3′ UTRs.[7] Due to strong negative selection, not only conserved miRNA binding sites,[8,9] but also the entire 3′ UTRs with miRNA target sites have fewer SNPs compared with non-miRNA target 3′ UTRs,[10] indicating the importance of 3′ UTR sequence integrity on the function of miRNAs. Despite this evolutional constraint, about 5% of SNPs that are predicted to reside in the seed region of validated miRNA binding sites[11] potentially disrupt miRNA-mediated regulation. As seen in earlier studies of Tourette syndrome[12] and muscular hypertrophy,[13] destroying and creating miRNA complementary sites by a rare and yet functional SNP can cause a deleterious outcome.

Since longer 3′ UTRs contain more miRNA complementary sites,[14] the 4.5 Kb 3′ UTR of *KRAS* may exhibit a greater complexity and dependence on regulation by miRNAs. *KRAS* is predicted to be targeted and regulated by many miRNAs, including the tumor-suppressive *let-7* family of miRNAs through 10 putative *let-7* complementary sites (LCSs) in its 3′ UTR.[15] An inverse relationship between *let-7* and KRAS protein levels in

*Correspondence to: Frank J Slack; Email: frank.slack@yale.edu
Submitted: 01/09/2014; Accepted: 01/21/2014; Published Online: 02/03/2014
http://dx.doi.org/10.4161/cc.27941

human lung tumor samples was confirmed.[15] In addition, in vivo delivery of *let-7* prevented and inhibited lung tumor growth in a mouse model of *Kras*-induced lung cancer by inhibiting cellular proliferation.[16-19]

Polymorphic miRNA-mediated regulation of *KRAS* in cancer has been shown through a rare SNP (rs61764370) identified in a *let-7* complementary site in its 3′ UTR.[20] The derived allele at rs61764370 was found to be associated with an increased risk of non-small cell lung cancer, epithelial ovarian cancer and triple-negative breast cancer.[20-22] Reduced survival was observed in oral cancer[23] with the variant, but not in NSCLC cases.[24] In addition, the variant was associated with adverse overall survival in metastatic colorectal cancer cases treated with anti-EGFR therapy[25] and better outcome in early-stage colorectal cancer cases.[26] **Table S1** summarizes additional associations between rs61764370 (or other known SNPs) in the *KRAS* 3′ UTR and diseases.

Since the 3′ UTR has often been neglected in previous searches for disease-associated variants, discovery of rare but functional variants in the 3′ UTR may provide a new insight to understanding initiation and progression of cancer. To identify additional variants in the 3′ UTR of *KRAS* that have the potential to be novel genetic biomarkers for non-small cell lung cancer (NSCLC) and epithelial ovarian cancer (EOC), we resequenced the entire 3′ UTR of *KRAS* using 2 different sequencing approaches. We performed direct Sanger sequencing of lung tumor DNA from NSCLC patients, and target enrichment and high-throughput sequencing of germline DNA of EOC patients. Here we report multiple sequence variants in the *KRAS* 3′ UTR identified in NSCLC and EOC patients. We found that 2 known SNPs, rs712 and rs9266, and 1 novel variant may play a role in the regulation of *KRAS*. Further analysis of these sequence variants will determine their potential functional roles in predicting cancer risk and outcome.

Results

The polymorphic 3′ UTR of the KRAS gene

The annotated 3′ UTR of the *KRAS* transcript variant B, which is the predominant isoform, is 4549 bp long (UCSC Genome Browser[36]), ranging from base pairs 25358180 to 25362728 on chromosome 12 in the reference human genome (GRCh37/hg19). NCBI dbSNP build135 (dbSNP135)[37] reports a total of 56 SNPs in this region (**Fig. 1**). Among 56 previously reported SNPs (known SNPs) in the 3′ UTR of *KRAS,* 17 SNPs reside in predicted high-confidence miRNA complementary sites (see "Materials an Methods"; **Table S2**).

New sequence variation identified in the 3′ UTR of *KRAS* in NSCLC cases

To examine the sequence variability in NSCLC cases, we amplified and sequenced the entire 3′ UTR of the *KRAS* gene from genomic DNA derived from lung tumor samples of a maximum of 70 NSCLC cases described previously[20] (see "Materials and Methods"). Due to the limited availability of DNA from the matching normal adjacent tissues, only tumor DNA was analyzed. Our 3′ UTR resequencing revealed a total of 44 sites with sequence variations, which included 22 known SNPs. The remaining 22 variants were not previously reported; therefore, they were presumed to be either novel germline SNPs or somatic mutations, as the genomic DNA used here was obtained from tumor samples (**Fig. 2A**; **Table S3**).

More than 70% of NSCLC cases contain the derived alleles (non-ancestral alleles) at rs4597149, rs712, rs4963858, rs13096 rs1137188, rs8720, rs12587, and rs12245. A few novel variants, including novel variant 4, novel variant 17, and novel variant 20 also have the derived alleles that show a frequency of almost 50% (**Fig. 2A**; **Table S3**).

The sequence variants identified and enriched in the 3′ UTR of *KRAS* in EOC cases

We captured the 3′ UTR of the *KRAS* gene from germline DNA of 31 EOC cases of European descent using a NimbleGen Capture array and sequenced the captured *KRAS* 3′ UTR DNA using Illumina next generation technology (see "Materials and Methods"). We identified sequence variations at a total of 22 sites, which included 21 known SNPs and 1 novel variant (**Fig. 2B**; **Table S4A**). In order to minimize false positives, we only included novel variants that were identified by 2 different SNP calling methods, SAMtools and GATK. The 7 potential novel variants that were only identified by SAMtools are shown in **Table S4B**.

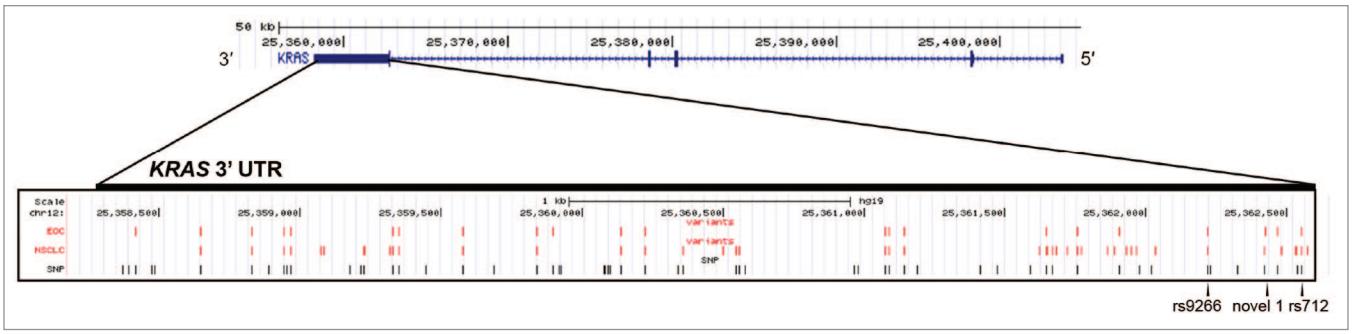

Figure 1. The known SNPs and variants identified in our cancer cases in the *KRAS* 3′ UTR. The *KRAS* gene is located on the minus strand of chromosome 12 from base pairs 25358180 to 25403864 (GRCh37/hg19). The 3′ UTR of the *KRAS* transcript variant B spans from base pairs 25358180 to 25362728. All known SNPs from dbSNP135 are included in this figure in black. The sites with sequence variation identified in our NSCLC and EOC cases are labeled in red. The figure was generated and modified from the UCSC Genome Browser.

To determine whether any particular sequence variants appeared more frequently in cases compared with controls, we measured a fold enrichment of a derived allele at each variable site. We compared the allele frequency of the derived allele at the variable site found in our cases to the allele frequency from European populations reported in the 1000 Genomes Project.[38] We were unable to measure the actual enrichment of the 3 known SNPs identified in EOC cases, rs61763589, rs191137453, and rs61764371, because the derived allele frequency was not found in European controls in the 1000 Genomes Project. For all other

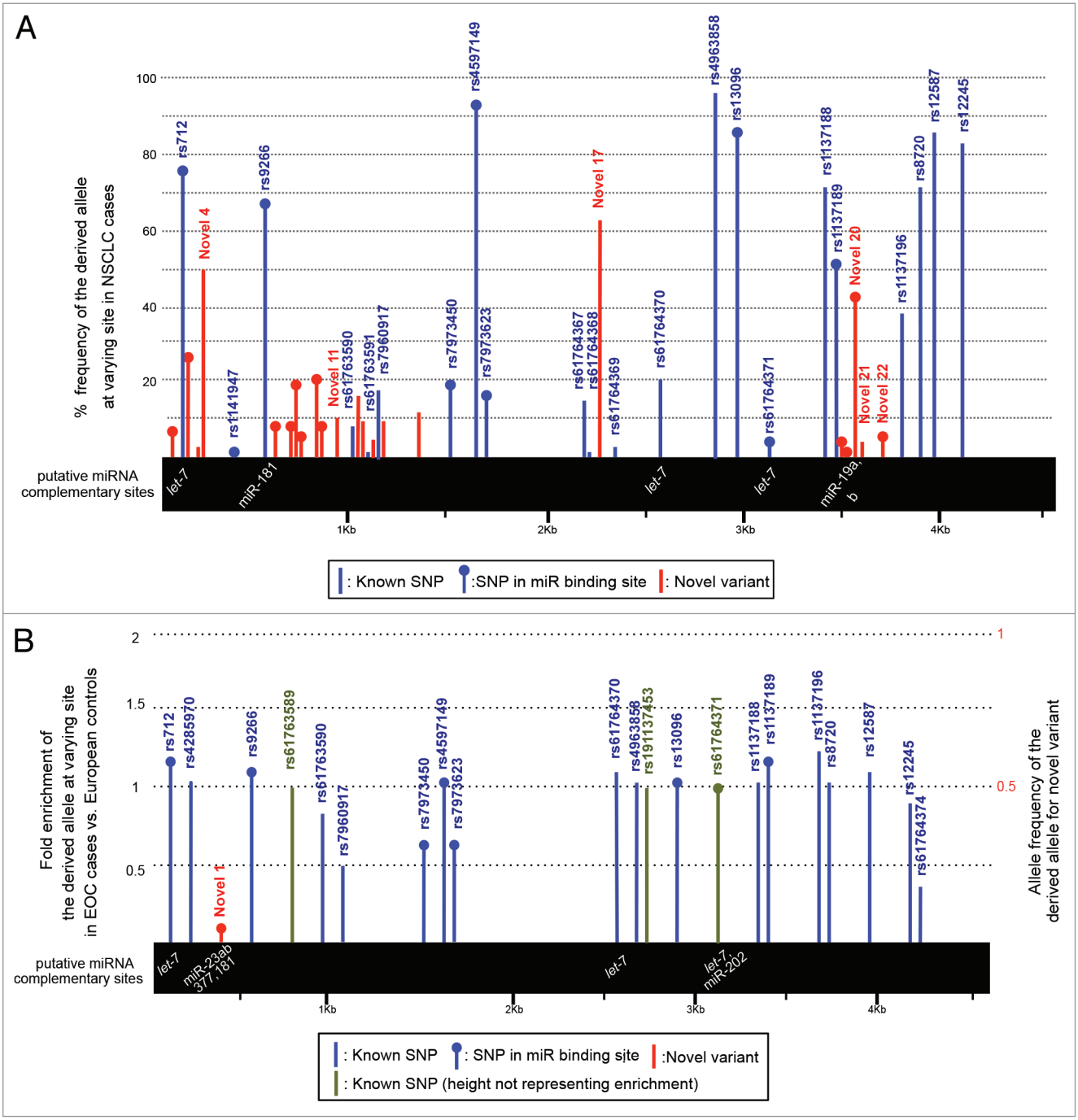

Figure 2. The sequence variation identified in our NSCLC and EOC cases. All sites with the sequence variation identified in cases were mapped to their relative positions in the 3′ UTR of *KRAS*. (**A**) The height of each bar represents the percent frequency of the derived allele at the variable sites in NSCLC cases. The blue bar represents known SNPs, the red bar represents novel variants (or somatic mutations), and the bar with a circular end represents a site within high confidence miRNA complementary sites. (**B**) The height of blue bars represents the fold enrichment of the derived allele at known SNP sites in our EOC cases vs. reference European controls from the 1000 Genomes Project. The height of red bar represents the derived allele frequency of a novel variant in cases. For 3 known SNPs in green bars, the allele frequencies of the derived alleles in European controls were 0; therefore, the height does not represent any numerical value.

sequence variants, no variant was highly enriched in our cases (**Table S4A**).

The derived alleles at the variable sites in cancer cases may alter predicted miRNA complementary sites

To examine the effects of the derived alleles at the variable sites on miRNA complementary sites, we first determined whether the sequence variants identified reside in predicted high-confidence miRNA complementary sites. We first utilized 2 web-based databases, Patrocles[39] and miRNASNP,[40] to predict whether a given sequence variant could result in gain or loss of miRNA complementary sites. However, since these databases cataloged only a few known SNPs, we computationally predicted the effects of the derived allele at all the variable sites identified using the miRanda target prediction algorithm[41] with a selected target/query sequence (see "Materials and Methods"). **Tables S6 and S7** list all miRNA complementary sites that are predicted to be gained and lost by the derived alleles at the variable sites by the miRanda algorithm. Since not all miRNA complementary sites predicted by the miRanda algorithm are in our list of high-confidence miRNA complementary sites, we first selected common miRNA complementary sites predicted by the miRanda algorithm with the ancestral alleles and in our high confidence miRNA complementary sites. To specifically determine whether the derived alleles were predicted to disrupt miRNA complementary sites, we compared those common miRNA complementary sites to miRNA complementary sites with the derived allele predicted by the miRanda algorithm.

Among the 44 variable sites identified in NSCLC cases, 21 reside within miRNA complementary sites, and for 8 of those, the derived alleles may alter the predicted sites (**Table S6**). In EOC cases, among a total of 22 variable sites identified, 9 reside in miRNA complementary sites (**Table S7**). The derived alleles at 2 of those 9 sites may disrupt predicted sites (**Table S7**). There were 17 common known SNPs found in both cancer sets, and 8 of them reside in miRNA complementary sites (**Table 1**). The derived alleles at 2 SNPs, rs712 and rs9266, out of 8 are predicted to disrupt the putative miRNA complementary sites, including *let-7* and miR-181. Notably, all patients who had the derived allele at rs9266 also had the derived allele at rs712.

rs712*G and rs9266*C are widely distributed in world populations

To determine the prevalence of the derived alleles at rs712 (rs712*G) and at rs9266 (rs9266*C) across general world populations, we combined allele frequency data for rs712 and rs9266 previously entered in the Allele Frequency Database (ALFRED).[42] Data exist for a total of 2385 healthy individuals from 45 populations representing several different regions of the world (**Fig. 3A**). The frequency of the derived alleles at rs712 and rs9266 varied among different populations, but both exhibited a very similar general trend: the frequency of this

Table 1. The common SNPs identified in both NSCLC and EOC cases

	Chr12 hg19	SNP ID	% freq of the derived allele in NSCLC	freq of the derived allele in EOC	Putative miRNA complementary sites	LCS	Ago binding
1	25362552	rs712	75.4	0.597	miR-422a, miR-330–5p, miR-3125, miR-877, miR-378bc, miR-1299, miR-193b, miR-200b,c, miR-429	*	*
2	25362217	rs9266	68.5	0.581	miR-181abcd, miR-4262, miR-132		*
3	25361756	rs61763590	7.1	0.081	miR-133ab, miR-421		
4	25361646	rs7960917	16.7	0.113			
5	25361142	rs7973450	20.0	0.129	miR-143, miR-302c*		*
6	25361091	rs4597149	90.9	1	miR-410, miR-340, miR-376c		*
7	25361074	rs7973623	16.4	0.129	miR-501-5p		*
8	25360224	rs61764370	21.4	0.97	miR-18ab	*	
9	25360138	rs4963858	98.5	1	miR-599, miR-148a,b, miR-152, miR-199a,b-5p		
10	25359841	rs13096	85.9	0.565	miR-101, miR-144, miR-493*, let-7a-2*, let-7g*		*
11	25359577	rs61764371	3.3	0.032	let-7 family, miR-202	*	*
12	25359352	rs1137188	71.2	0.581	miR-129–5p, miR-876–5p, miR-421, miR-511, miR-513b, miR-624, miR-541*, miR-1290		
13	25359328	rs1137189	51.7	0.597	miR-32*, miR-380		*
14	25358969	rs1137196	37.7	0.516	miR-129–5p		
15	25358943	rs8720	73.8	0.581	miR-2115*		
16	25358828	rs12587	85.9	0.597	miR-425		
17	25358650	rs12245	84.4	0.435	miR-421, miR-143		

Seventeen known SNPs were commonly identified in both NSCLC and EOC cases. LCS, let-7 complementary sites

allele increased across Eurasia with distance from Africa but was lower in Native American populations. East Asian populations generally exhibited the highest frequency of the derived alleles. Since we have previously genotyped another *let-7* disrupting SNP, rs61764370, from world populations,[20] we compared the frequencies of the derived alleles at 2 *let-7*-disrupting SNPs,

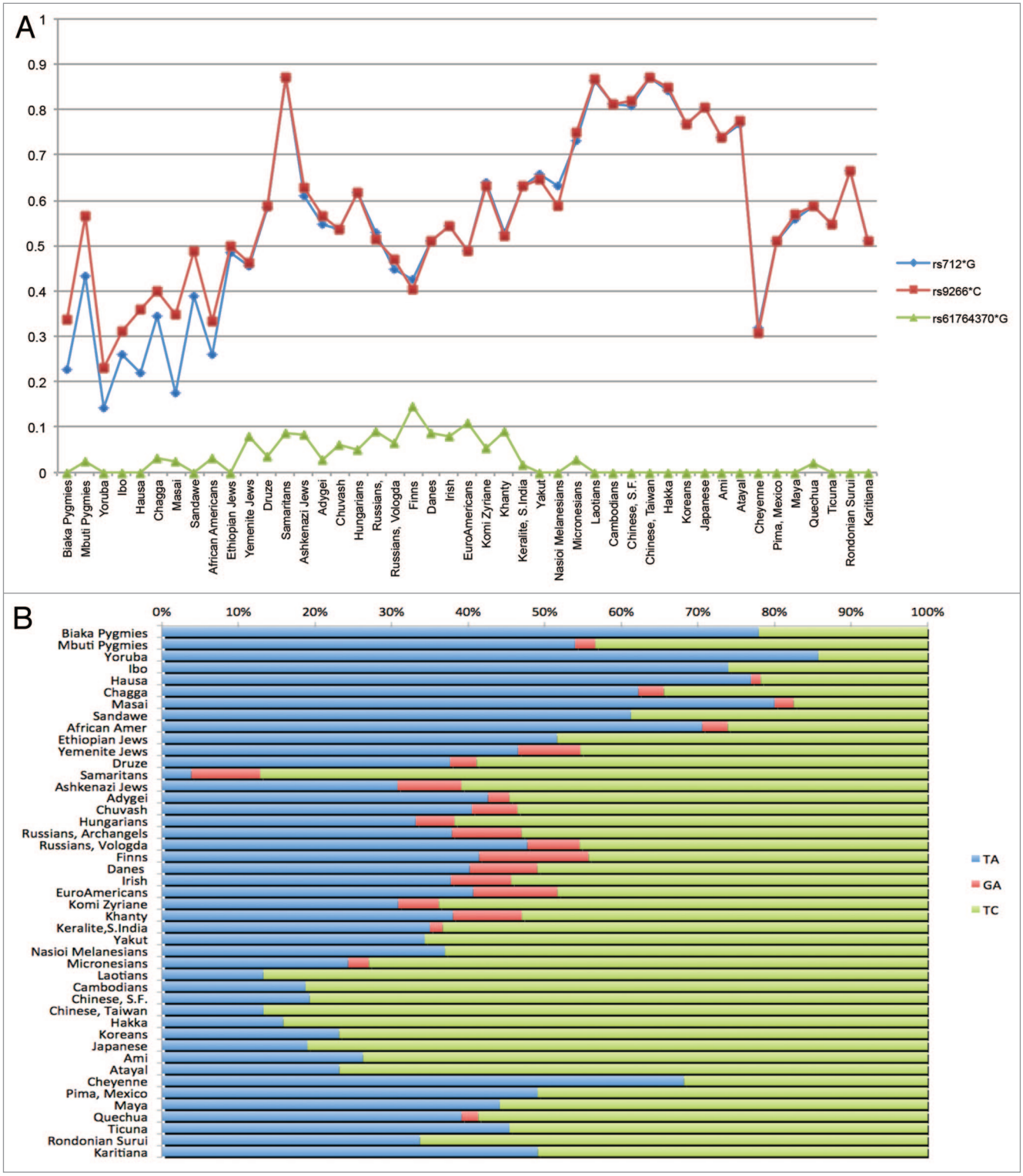

Figure 3. Frequencies of the derived alleles at rs712, rs9266 and rs61764370. (**A**) Individual allele frequencies for the derived alleles at rs712 (blue), rs9266 (red), and rs61764370 (green) in samples of normal individuals from 45 populations are combined from ALFRED. (**B**) The same data for 2 *let-7*-disrupting SNPs, rs712 and rs61764370, as in (**A**), are presented in stacked bar format as the frequencies of the 3 haplotypes inferred to be present in these 45 populations. There is no evidence for the presence of the GG haplotype comprised of the derived alleles at both SNPs.

rs712 and rs61764370 (**Fig. 3A**). While the derived allele at rs712 was generally frequent in Asian populations, the derived allele at rs61764370 (rs61764370*G) was very rarely found outside of Europe and Africa.

rs712 is identified in a haplotype of SNPs across the *KRAS* gene

To examine haplotypes of SNPs across the *KRAS* gene in world populations, we first generated haplotypes that contain rs712 and rs61764370, which are 2328 bp apart. Only 3 genotypes were identified, and a haplotype that contained the derived alleles at both rs712 and rs61764370 in cis was absent (**Fig. 3B**). To examine the haplotypes with additional SNPs in the *KRAS* gene, we included 11 other SNPs that we had previously genotyped (**Fig. S1**). These 11 SNPs include 5 SNPs in an adjacent intron, 1 SNP in the 3′ UTR of the *KRAS* transcript variant A, 4 SNPs in the 3′ UTR of the *KRAS* transcript variant B, and 1 SNP in an intergenic region of the 3′ UTR of *KRAS*. In all of the 8 common haplotypes that were identified, the derived alleles at rs712 and rs61764370 were never found together in cis, confirming that rs712*G is not a marker for rs61764370*G. In our NSCLC and EOC cases, we also noticed that rs712*G and rs61764370*G were never definitively found together in cis. Approximately 10% of our cancer cases were heterozygous at both SNP sites, and the presence of the derived alleles at both sites in cis would be unlikely according to the haplotype inference from genotypes based on an EM algorithm implemented in PLINK (http://pngu.mgh.harvard.edu/purcell/plink).[43]

The derived alleles at rs712 and rs9266 alter luciferase reporter expression

To test whether the presence of the derived alleles at rs712 and rs9266 have an effect on *KRAS* reporter expression, we generated 4 luciferase reporter constructs containing a fragment of the 3′ UTR of *KRAS* with: (1) no derived alleles (KRAS-WT); (2) the derived allele at rs712 (KRAS-rs712*G); (3) the derived allele at rs9266 (KRAS-rs9266*C), and (4) the derived alleles at both rs712 and rs9266 (KRAS-rs712*G/rs9266*C). We did not observe statistically significant differences in luciferase expression when the 4 reporters were individually transfected in Caov3 (ovarian adenocarcinoma epithelial cell line) and A549 (lung adenocarcinoma epithelial cell line) (data not shown). However, we reproducibly observed ~15% and 9% derepression in luciferase expression with KRAS-rs712*G and KRAS-rs9266*C, respectively, compared with KRAS-WT in HeLa (cervical adenocarcinoma epithelial cell line) (**Fig. 4A**). In addition, the presence of the derived alleles at both varying sites exhibited an additive effect, an approximately 30% derepression in luciferase expression.

To determine how the derived alleles at rs712 and rs9266 disrupt the miRNA complementary sites, the minimum free energy (mfe) required to form a duplex between a complementary site and a miRNA and predicted RNA structures were examined by RNAhybrid[44] and Mfold,[45] respectively. Specifically, the high confidence complementary sites for miR-378, miR-3125, and miR-877, predicted to be disrupted by rs712*G, and such sites for miR-181 and miR-4262, predicted to be disrupted by rs9266*C were examined (**Table S6**). In addition, since rs712 affects the last nucleotide of *let-7* complementary site 1 (LCS1), and *let-7d* is the only *let-7* family member miRNA that is predicted to form a base pair at this position with the ancestral allele, the LCS1/*let-7d* duplex was also examined. While the presence of the derived allele disrupted the base paring specifically at the SNP site in all target/miRNA duplexes, the derived alleles did not appear to substantially alter the minimum free energy required to form these duplexes (**Fig. S2**). When a 74-nt sequence with the derived allele or the ancestral allele at each varying site was individually folded into a secondary structure of mRNA, the predicted RNA structure was altered by rs712*G, but not by rs9266*C (Mfold)[45] (**Fig. S3A and B**). Especially, the seed region of LCS1 shifted from a stem to a loop by rs712*G (**Fig. S3A**). Since we observed the additive derepression on reporter expression when both rs712 and rs9266 had the derived alleles, the secondary structures of a 657-nt sequence, which was the actual length of the 3′ UTR of *KRAS* in the reporters, were examined using the UEA sRNA tools.[46] As seen in the predicted structures with the 74-nt sequence, only rs712*G altered the secondary structure significantly (**Fig. S3C**).

The novel variant 1 from EOC cases affects 3′ UTR function in luciferase reporter assays

Since 1 variant was identified as novel in EOC cases by 2 different stringent SNP calling methods and resided in an Argonaute binding site, the effect of the derived allele at this varying site was determined by luciferase reporter assays. We observed ~15% derepression in luciferase reporter expression with the variant (KRAS-novel 1*T) as compared with its non-variant reporter (KRAS-WT) in HeLa cells (**Fig. 4B**). The predicted secondary structures of a 657-nt sequence with and without the variant remained unchanged (data not shown). However, the variant may weaken the formation of a duplex between miR-23a or the miR-181 family and the target region with the varying site. (**Fig. S4**).

Discussion

Using 2 different sequencing methods, we have sequenced the entire 3′ UTR of *KRAS*, which spans approximately 4.5 Kb, and identified a wide range of sequence variations in multiple NSCLC and EOC cases. Although our conclusions were hampered by some limitations in the type of DNA sequenced in NSCLC cases and a small number of samples in EOC cases, we found that the derived allele at 2 known SNPs, rs712 and rs9266, identified in both NSCLC and EOC cases, and 1 novel variant in EOC cases could disrupt regulation of *KRAS*. In addition, although no variant was found to be enriched in EOC cases with statistical significance, detecting the derived allele at rs61763589, rs191137453, and rs61764371 exclusively in our EOC cases and not in European populations in the 1000 Genomes Project suggests that these variants have the potential to be enriched in cancer and potentially alter KRAS expression. All 3 SNPs reside in complementary sites of miRNAs, such as miR-580,[47,48] miR-26ab,[49,50] the *let-7* family, and miR-202, that play an important role in tumorigenesis. In particular, rs61763589 is within tumor-suppressive *let-7*[16-19,49] and miR-202[51-53] complementary sites, which are also in our list of high confidence miRNA complementary sites.

For rs712, rs9266, and novel variant 1, we observed statistically significant differences in luciferase reporter expression between one with the derived allele at the SNP site and the other with the ancestral allele in HeLa (cervical adenocarcinoma cell line), but not in Caov3 (ovarian adenocarcinoma cell line) and A549 (lung adenocarcinoma cell line). Since each cell line expresses different miRNA profiles, we can expect changes in reporter expression only if the cognate miRNAs that bind to regions with the varying site are expressed in the cell line. Since HeLa cells are known to have high *let-7*, miR-181, and miR-23 levels,[54-56] disruption of these miRNA complementary sites may account for the dysregulation of *KRAS* reporter expression. Notably, *let-7*, miR-181, and miR-23 have been implicated in the development of various cancers.[57-59]

Although we cannot exclude the possibility that the derived alleles may also disrupt binding of another endogenous miRNA or multiple miRNAs in HeLa, our result agrees with the previous findings that reported the role of *let-7* and miR-181 in *KRAS* regulation.[58,60] When several point mutations were introduced to *let-7* complementary site 1 (LCS1) of the murine *Kras* 3′ UTR, derepression of luciferase reporter expression was observed compared with its non-mutated counterpart.[60] Since LCS1 is highly conserved across different vertebrate species, altering one or multiple base pairs in LCS1 likely disrupts binding of the *let-7* family, specifically *let-7d*, to this site and eventually leads to dysregulation of *KRAS* in both humans and mice. In addition, a mutation in the miR-181 complementary site where rs9266 is located prevented binding of miR-181 and resulted in derepression of luciferase reporter expression in 2 glioblastoma cell lines, U251 and LN229.[58] Moreover, the RNA secondary structure may also be implicated in cis-regulation of *KRAS*, as rs712*G resulted in distinct secondary structures compared with the predicted structures with rs712*T. Altered RNA structure may affect the accessibility of endogenous miRNAs and, thus, miRNA-dependent regulation of *KRAS*.

In addition to the functional role of rs712 in *KRAS* regulation, one study reported that the genotype of G/G, which corresponds to the derived alleles at rs712, is associated with an increased risk of oral squamous cell carcinoma compared with the genotypes of T/T or G/T.[61] However, the ancestral allele at rs712 was also found to be associated with an increased risk of gastric and colorectal cancers.[62,63] Since different types of cells and tissues have different miRNA profiles, the role of a specific SNP may vary depending on the types of cancer. Therefore, further case-control studies need to be performed to validate whether a variant at rs712 can be a predictor for the risk of NSCLC and EOC as well.

Since previous studies reported associations between rs61764370 and various cancer risk and outcome, we

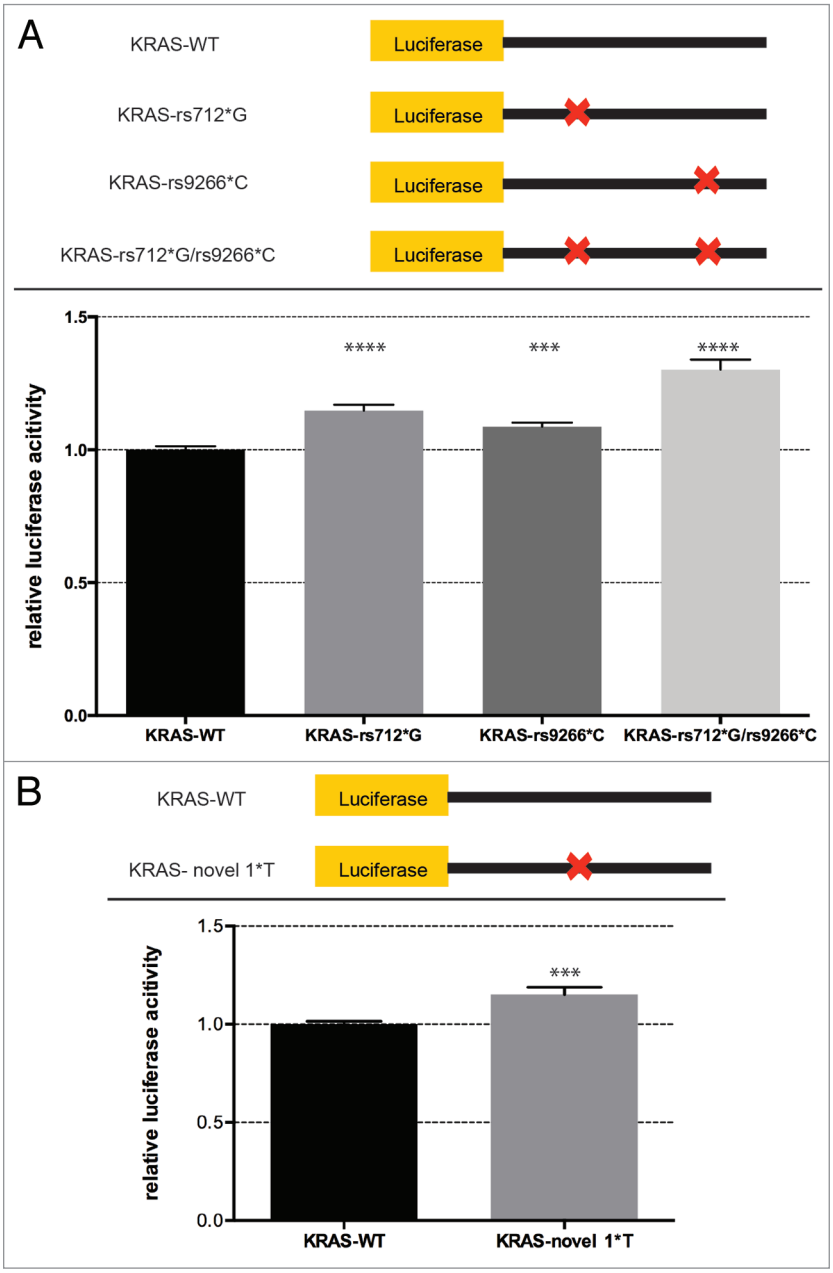

Figure 4. The effect of the derived alleles at rs712 and rs9266, and one novel variant on luciferase expression. (**A**) Four luciferase reporter constructs were generated with a fragment of the 3′ UTR of *KRAS* in psi-Check2. KRAS-WT: the ancestral alleles (T) at both rs712 and rs9266. KRAS-rs712*G: the derived allele (G) at 712. KRAS-rs9266*C: the derived allele (C) at rs9266. KRAS-rs712*G/rs9266*C: the derived alleles at both rs712 and rs9266. Luciferase reporter assays in HeLa cells indicated ~15% derepression in relative luciferase activity in KRAS-rs712*G and ~9% depression in KRAS-rs9266*C compared with KRAS-WT. In the presence of both derived alleles (KRAS-rs712*G/rs9266*C), ~30% derepression was observed. Error bars: standard error of the mean. ***P value < 0.001; ****P value < 0.0001. (**B**) KRAS-novel 1*T: the derived allele (T) at the varying site. Luciferase reporter assays in HeLa cells indicated ~15% derepression in relative luciferase activity in KRAS-novel 1*T compared with KRAS-WT.

hypothesized that a combination of the derived alleles at 2 *let-7*-disrupting SNPs, rs712 and rs61764370, may be a novel risk marker. Although approximately 10% of our cancer cases were heterozygous at both rs712 and rs61764370, it would be very unlikely to detect the derived alleles at both sites in cis. We will need to examine the haplotypes of rs712 with rs61764370 or other SNPs in a larger case-control study to test this hypothesis that a combination of the derived alleles at 2 SNP sites may be a novel biomarker for cancer risk.

This study has provided an exhaustive list of sequence variations identified in the complete region of the 3′ UTR of *KRAS* in multiple NSCLC and EOC cases. Although the list presents promising candidate biomarkers for cancer risk and progression, experimental validation is required to understand the biological role of the individual SNP as well as the combination of multiple common SNPs. Investigating the effects of SNPs on RNA structure may also reveal an additional mechanism by which SNPs contribute to disease phenotype as suggested by an earlier study.[64] In addition, a larger case-control study will be required to assess the association between a SNP and cancer risk or outcome. Any functionally validated variants could serve as a powerful and beneficial prognostic and diagnostic predictor in the clinic.

Materials and Methods

Clinical study populations

Lung tissue samples from patients with a diagnosis of NSCLC were collected following Yale University Human Investigation Committee approval. Cases were chosen based on the availability of frozen stored tissue from lung tumor resections from 1994 through 2003 and from recent cases with extra tissue available. A maximum of 70 NSCLC patients were included in the analysis from a study described previously.[20] Thirty-one EOC patients were drawn from a study described previously.[21] GYN59, GYN125, and GYN151 from the EOC study populations were known to carry the derived allele at rs61764370[21] and were included as technical positive controls for the high-throughput sequencing. All patients in the study provided their written informed consent as approved by Yale University Human Investigation Committee. Patient data were collected including age, ethnicity, and family history of cancer.

Reference world population data

Normal population data were collected using DNA from the anonymous samples that are part of the Kidd collection of population samples.[27] These samples of healthy adults were collected with informed consent for studies such as this. SNPs were typed using TaqMan assays obtained from Applied Biosystems using standard protocols. Haplotype estimation was done using PHASE.[28] Additional specifics on all aspects of the reference populations, data collection, and analyses of the haplotypes are given in Pelletier et al.[29]

DNA isolation and direct Sanger sequencing for NSCLC cases

DNA was isolated from frozen and FFPE lung tissue using the Dneasy Blood and Tissue kit (Qiagen). Depending on the availability of genomic DNA from tumor samples, 50–70 genomic DNA samples were sequenced. Segments of the *KRAS* 3′ UTR were amplified using *Pfu* Turbo DNA polymerase (Stratagene) or Phusion high-fidelity DNA polymerase (NEB) and DNA primers specific to the sequence (**Table S8**). PCR products were purified using the QIAquick PCR purification kit, 96-PCR purification kit (Qiagen), or at the core sequencing laboratory at the W.M. Keck Laboratory at Yale University and sequenced using the same primers. Sequences of the each segment were aligned to the annotated 3′ UTR of the *KRAS* transcript variant B from NCBI36/hg18 refseq (NM_004985.3, chr12: 25249447–25253995) using the SeqMan program (DNA Star). Once GRCh37/hg19 refseq was released, we confirmed that there were no sequence differences and updated our sequence to GRCh37/hg19 refseq (chr12: 25358180–25362728).

DNA isolation and capture enrichment and high-throughput sequencing for EOC cases

Germline DNA was isolated from either blood or saliva of EOC patients using the QiaAmp DNA Kit (Qiagen) or the Oragene DNA Kit (DNA Genotek), respectively. The coordinates of the *KRAS* 3′ UTR was identified and submitted to Roche Diagnostics for custom array design using the 2.1M (2.1M probes) HX1 NimbleGen sequence capture array. Genomic DNA was separately sheared to fragments compatible with the Solexa sequencing system. Next, the individual genomic DNA was denatured and hybridized to our custom NimbleGen DNA array carrying sequences complementary to the 3′ UTR. Finally, the captured DNA was eluted and processed through standard high-throughput sequencing on an Illumina platform at the core-sequencing laboratory at the W.M. Keck Laboratory at Yale. Individual samples were run per well of the flow cell along with a standard positive control using a read length of 74 bp. All samples were sequenced by paired-end sequencing. The raw sequencing data have been deposited in the European Nucleotide Archive (ENA) under accession number PRJEB4945. Individual targeted sequences were mapped to the reference genome using BWA, and then BWA mapping results were processed by SAMtools/Maq to obtain the genotype for the target regions with parameters -d8-D10000-11e-5-20-41e-7. The sequencing gave about 500× read depth for each sample with a very low error rate. To increase the accuracy of SNP calling, we used GATK (http://www.ncbi.nlm.nih.gov/pubmed?term=20644199) to perform multi-sample SNP calling after duplicate marking and local realignment. GATK SNP calling was then compared with SAMtools output. The overlapping calling set was considered to have higher chance to be true positive SNPs. Among the SNPs identified as novel only in SAMtools, those with allele frequencies less than 0.016 were removed.

Predicting putative miRNA complementary sites

To predict miRNA complementary sites in the 3′ UTR of *KRAS,* all miRNA complementary sites that are conserved and have a good mirSVR score from microRNA.org release August 2010,[30] and all conserved miRNA complementary sites from TargetScan release 6.1[31] were combined. In addition, putative *let-7* complementary sites were obtained from 2 previous publications.[15,20] Predicted miRNA complementary sites that

overlapped with Argonaute binding sites were selected as high-confidence miRNA complementary sites. To search for Argonaute binding sites in the *KRAS* gene, sets from AGO 1–4 PAR-CLIP (Hafner 2010),[32] AGO2 CLIP-SEQ (Kisher 2011),[33] and AGO2 MNASE PAR-CLIP (Kishore 2011)[33] were selected in the doRiNA database.[34]

Predicting the effects of the derived allele at SNPs on miRNA complementary sites

As a target sequence, a 49-nucleotide sequence that contained the alternative alleles at the variable site and surrounding 24-nucleotide sequences from both ends were used. All known human miRNAs found in miRBase release 18[35] were screened for their complementary sites in each given target sequence using the miRanda algorithm with default command line arguments, -sc 140-go-9-ge-4-en 1 (no energy filtering), as stated in the August 2010 release note.[41]

Luciferase reporter constructs and luciferase reporter assays

To make a luciferase reporter with the ancestral allele at rs712 and the derived allele at rs9266 (KRAS-rs9266*C), first 657 bp of the *KRAS* 3′ UTR (NCBI36/hg18, chr12: 25253338–25253995) was amplified from a modified version of pGL3-KRAS wild-type[20] using the forward primer MK1: 5′-CCCGCTCGAG ATACAATTTG TACTTTTTTC TTAAGGCATA C-3′ and reverse primer MK2: 5′-ATAAGAATGC GGCCGCGGCC TTATAATAGT TTCCATTGCC TTG-3′. Since the original pGL3-KRAS wild-type[20] contained the derived alleles at rs712 and rs9266, the reporter had been previously modified by reverting the derived allele (the G allele) at rs712 to the ancestral allele (the T allele) using site-directed mutagenesis (Stratagene). The amplified product was cloned into psi-Check2 (Promega) that has been digested with *Xho*I and *Not*I. The luciferase reporter with the derived alleles at rs712 and rs9266 (KRAS-rs712*G/rs9266*C) was generated using site-directed mutagenesis with the forward primer LCS1R: 5′-GAGGAAAAAA AAACCTTCCA CTGTCATTT-3′ and the reverse primer MK59: 5′-GGAAGTTTTT TTTTCCTCGA AGTGCCAGT-3′. The luciferase reporters with the ancestral allele at rs9266 (KRAS-rs712*G and KRAS-WT) were generated using the forward primer MK73: 5′-GTTTGTCATC CCTGATGAAT GTAAAGTTAC-3′ and the reverse primer MK74: 5′-GTAACTTTAC ATTCATCAGG GATGACAAAC-3′. KRAS-novel 1 was generated using the forward primer MK102: 5′-GATTACTTCT TATTTTTCTT ACTAATTGTG AATGTTGGTG TG-3′ and the reverse primer MK103: 5′-CACACCAACA TTCACAATT**A** GTAAGAAAAA TAAGAAGTAA TC-3′. HeLa cells were cultured in RPMI (Gibco) with 10% fetal bovine serum (Sigma-Aldrich) and 1X penicillin/streptomycin (Invitrogen). HeLa cells were transfected with 5 ng of each luciferase reporter using Lipofectamine 2000 (Invitrogen) and Opti-MEM (Gibco) for 24 h. Activities of *Renilla* and Firefly luciferase were measured using the Dual-Luciferase Reporter Assay System (Promega) and Wallac Victor 1420[2] (Perkin Elmer). Two-tailed *t* tests were used to measure statistical significance of differences in reporter expression.

Note Added in Proof

NCBI dbSNP has been updated to dbSNP138. Since our analysis was done using dbSNP135, we have confirmed that the SNPs identified as novel in our study are not in the updated database and remain novel.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Supplemental Materials

Supplemental materials may be found here:
www.landesbioscience.com/journals/cc/article/27941

References

1. Karnoub AE, Weinberg RA. Ras oncogenes: split personalities. Nat Rev Mol Cell Biol 2008; 9:517-31; PMID:18568040; http://dx.doi.org/10.1038/nrm2438
2. Pylayeva-Gupta Y, Grabocka E, Bar-Sagi D. RAS oncogenes: weaving a tumorigenic web. Nat Rev Cancer 2011; 11:761-74; PMID:21993244; http://dx.doi.org/10.1038/nrc3106
3. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer 2010; 10:389-402; PMID:20495573; http://dx.doi.org/10.1038/nrc2867
4. Godshalk SE, Paranjape T, Nallur S, Speed W, Chan E, Molinaro AM, Bacchiocchi A, Hoyt K, Tworkoski K, Stern DF, et al. A Variant in a microRNA complementary site in the 3′ UTR of the KIT oncogene increases risk of acral melanoma. Oncogene 2010; 30:1542-50; PMID:21119596
5. Teo MTW, Landi D, Taylor CF, Elliott F, Vaslin L, Cox DG, Hall J, Landi S, Bishop DT, Kiltie AE. The role of microRNA-binding site polymorphisms in DNA repair genes as risk factors for bladder cancer and breast cancer and their impact on radiotherapy outcomes. Carcinogenesis 2012; 33:581-6; PMID:22166496; http://dx.doi.org/10.1093/carcin/bgr300
6. Campayo M, Navarro A, Viñolas N, Tejero R, Muñoz C, Diaz T, Marrades R, Cabanas ML, Gimferrer JM, Gascon P, et al. A dual role for KRT81: a miR-SNP associated with recurrence in non-small-cell lung cancer and a novel marker of squamous cell lung carcinoma. PLoS One 2011; 6:e22509; PMID:21799879; http://dx.doi.org/10.1371/journal.pone.0022509
7. Bruno AE, Li L, Kalabus JL, Pan Y, Yu A, Hu Z. miRdSNP: a database of disease-associated SNPs and microRNA target sites on 3'UTRs of human genes. BMC Genomics 2012; 13:44; PMID:22276777; http://dx.doi.org/10.1186/1471-2164-13-44
8. Chen K, Rajewsky N. Natural selection on human microRNA binding sites inferred from SNP data. Nat Genet 2006; 38:1452-6; PMID:17072316; http://dx.doi.org/10.1038/ng1910
9. Saunders MA, Liang H, Li W-H. Human polymorphism at microRNAs and microRNA target sites. Proc Natl Acad Sci USA 2007; 104:3300-5; PMID:17360642; http://dx.doi.org/10.1073/pnas.0611347104
10. Hu Z, Bruno AE. The Influence of 3'UTRs on microRNA Function Inferred from Human SNP Data. Comp Funct Genomics 2011; 2011:910769; PMID:22110399; http://dx.doi.org/10.1155/2011/910769
11. Richardson K, Lai C-Q, Parnell LD, Lee Y-C, Ordovas JM. A genome-wide survey for SNPs altering microRNA seed sites identifies functional candidates in GWAS. BMC Genomics 2011; 12:504; PMID:21995669; http://dx.doi.org/10.1186/1471-2164-12-504
12. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, Mathews CA, Pauls DL, Rasin MR, Gunel M, et al. Sequence variants in SLITRK1 are associated with Tourette syndrome. Science 2005; 310:317-20; PMID:16224024; http://dx.doi.org/10.1126/science.1116502
13. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibé B, Bouix J, Caiment F, Elsen J-M, Eychenne F, et al. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet 2006; 38:813-8; PMID:16751773; http://dx.doi.org/10.1038/ng1810
14. Cheng C, Bhardwaj N, Gerstein M. The relationship between the evolution of microRNA targets and the length of their UTRs. BMC Genomics 2009; 10:431; PMID:19751524; http://dx.doi.org/10.1186/1471-2164-10-431
15. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ. RAS is regulated by the let-7 microRNA family. Cell 2005; 120:635-47; PMID:15766527; http://dx.doi.org/10.1016/j.cell.2005.01.014

16. Trang P, Medina PP, Wiggins JF, Ruffino L, Kelnar K, Omotola M, Homer R, Brown D, Bader AG, Weidhaas JB, et al. Regression of murine lung tumors by the let-7 microRNA. Oncogene 2010; 29:1580-7; PMID:19966857; http://dx.doi.org/10.1038/onc.2009.445
17. Trang P, Wiggins JF, Daige CL, Cho C, Omotola M, Brown D, Weidhaas JB, Bader AG, Slack FJ. Systemic delivery of tumor suppressor microRNA mimics using a neutral lipid emulsion inhibits lung tumors in mice. Mol Ther 2011; 19:1116-22; PMID:21427705; http://dx.doi.org/10.1038/mt.2011.48
18. Kumar MS, Erkeland SJ, Pester RE, Chen CY, Ebert MS, Sharp PA, Jacks T. Suppression of non-small cell lung tumor development by the let-7 microRNA family. Proc Natl Acad Sci USA 2008; 105:3903-8; PMID:18308936; http://dx.doi.org/10.1073/pnas.0712321105
19. Esquela-Kerscher A, Trang P, Wiggins JF, Patrawala L, Cheng A, Ford L, Weidhaas JB, Brown D, Bader AG, Slack FJ. The let-7 microRNA reduces tumor growth in mouse models of lung cancer. Cell Cycle 2008; 7:759-64; PMID:18344688; http://dx.doi.org/10.4161/cc.7.6.5834
20. Chin LJ, Ratner E, Leng S, Zhai R, Nallur S, Babar I, Muller R-U, Straka E, Su L, Burki EA, et al. A SNP in a let-7 microRNA complementary site in the KRAS 3′ untranslated region increases non-small cell lung cancer risk. Cancer Res 2008; 68:8535-40; PMID:18922928; http://dx.doi.org/10.1158/0008-5472.CAN-08-2129
21. Ratner E, Lu L, Boeke M, Barnett R, Nallur S, Chin LJ, Pelletier C, Blitzblau R, Tassi R, Paranjape T, et al. A KRAS-variant in ovarian cancer acts as a genetic marker of cancer risk. Cancer Res 2010; 70:6509-15; PMID:20647319; http://dx.doi.org/10.1158/0008-5472.CAN-10-0689
22. Paranjape T, Heneghan H, Lindner R, Keane FK, Hoffman A, Hollestelle A, Dorairaj J, Geyda K, Pelletier C, Nallur S, et al. A 3′-untranslated region KRAS variant and triple-negative breast cancer: a case-control and genetic analysis. Lancet Oncol 2011; 12:377-86; PMID:21435948; http://dx.doi.org/10.1016/S1470-2045(11)70044-4
23. Christensen BC, Moyer BJ, Avissar M, Ouellet LG, Plaza SL, McClean MD, Marsit CJ, Kelsey KT. A let-7 microRNA-binding site polymorphism in the KRAS 3′ UTR is associated with reduced survival in oral cancers. Carcinogenesis 2009; 30:1003-7; PMID:19380522; http://dx.doi.org/10.1093/carcin/bgp099
24. Nelson HH, Christensen BC, Plaza SL, Wiencke JK, Marsit CJ, Kelsey KT. KRAS mutation, KRAS-LCS6 polymorphism, and non-small cell lung cancer. Lung Cancer 2010; 69:51-3; PMID:19854534; http://dx.doi.org/10.1016/j.lungcan.2009.09.008
25. Graziano F, Canestrari E, Loupakis F, Ruzzo A, Galluccio N, Santini D, Rocchi M, Vincenzi B, Salvatore L, Cremolini C, et al. Genetic modulation of the Let-7 microRNA binding to KRAS 3′-untranslated region and survival of metastatic colorectal cancer patients treated with salvage cetuximab-irinotecan. Pharmacogenomics J 2010; 10:458-64; PMID:20177422; http://dx.doi.org/10.1038/tpj.2010.9
26. Smits KM, Paranjape T, Nallur S, Wouters KAD, Weijenberg MP, Schouten LJ, van den Brandt PA, Bosman FT, Weidhaas JB, van Engeland M. A let-7 microRNA SNP in the KRAS 3'UTR is prognostic in early-stage colorectal cancer. Clin Cancer Res 2011; 17:7723-31; PMID:21994416; http://dx.doi.org/10.1158/1078-0432.CCR-11-0990
27. Pakstis AJ, Fang R, Furtado MR, Kidd JR, Kidd KK. Mini-haplotypes as lineage informative SNPs and ancestry inference SNPs. Eur J Hum Genet 2012; 20:1148-54; PMID:22535184; http://dx.doi.org/10.1038/ejhg.2012.69
28. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001; 68:978-89; PMID:11254454; http://dx.doi.org/10.1086/319501
29. Pelletier C, Speed WC, Paranjape T, Keane K, Blitzblau R, Hollestelle A, Safavi K, van den Ouweland A, Zelterman D, Slack FJ, et al. Rare BRCA1 haplotypes including 3'UTR SNPs associated with breast cancer risk. Cell Cycle 2011; 10:90-9; PMID:21191178; http://dx.doi.org/10.4161/cc.10.1.14359
30. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. Nucleic Acids Res 2008; 36:D149-53; PMID:18158296; http://dx.doi.org/10.1093/nar/gkm995
31. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 2009; 19:92-105; PMID:18955434; http://dx.doi.org/10.1101/gr.082701.108
32. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr., Jungkamp A-C, Munschauer M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 2010; 141:129-41; PMID:20371350; http://dx.doi.org/10.1016/j.cell.2010.03.009
33. Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nat Methods 2011; 8:559-64; PMID:21572407; http://dx.doi.org/10.1038/nmeth.1608
34. Anders G, Mackowiak SD, Jens M, Maaskola J, Kuntzagk A, Rajewsky N, Landthaler M, Dieterich C. doRiNA: a database of RNA interactions in post-transcriptional regulation. Nucleic Acids Res 2012; 40:D180-6; PMID:22086949; http://dx.doi.org/10.1093/nar/gkr1007
35. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 2011; 39:D152-7; PMID:21037258; http://dx.doi.org/10.1093/nar/gkq1027
36. Dreszer TR, Karolchik D, Zweig AS, Hinrichs AS, Raney BJ, Kuhn RM, Meyer LR, Wong M, Sloan CA, Rosenbloom KR, et al. The UCSC Genome Browser database: extensions and updates 2011. Nucleic Acids Res 2012; 40:D918-23; PMID:22086951; http://dx.doi.org/10.1093/nar/gkr1055
37. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001; 29:308-11; PMID:11125122; http://dx.doi.org/10.1093/nar/29.1.308
38. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA; 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature 2010; 467:1061-73; PMID:20981092; http://dx.doi.org/10.1038/nature09534
39. Hiard S, Charlier C, Coppieters W, Georges M, Baurain D. Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. Nucleic Acids Res 2010; 38:D640-51; PMID:19906729; http://dx.doi.org/10.1093/nar/gkp926
40. Gong J, Tong Y, Zhang H-M, Wang K, Hu T, Shan G, Sun J, Guo A-Y. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat 2012; 33:254-63; PMID:22045659; http://dx.doi.org/10.1002/humu.21641
41. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. microRNA targets in Drosophila. Genome Biol 2003; 5:R1; PMID:14709173; http://dx.doi.org/10.1186/gb-2003-5-1-r1
42. Rajeevan H, Soundararajan U, Kidd JR, Pakstis AJ, Kidd KK. ALFRED: an allele frequency resource for research and teaching. Nucleic Acids Res 2012; 40:D1010-5; PMID:22039151; http://dx.doi.org/10.1093/nar/gkr924
43. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559-75; PMID:17701901; http://dx.doi.org/10.1086/519795
44. Krüger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res 2006; 34:W451-4; PMID:16845047; http://dx.doi.org/10.1093/nar/gkl243
45. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003; 31:3406-15; PMID:12824337; http://dx.doi.org/10.1093/nar/gkg595
46. Moxon S, Schwach F, Dalmay T, Maclean D, Studholme DJ, Moulton V. A toolkit for analysing large-scale plant small RNA datasets. Bioinformatics 2008; 24:2252-3; PMID:18713789; http://dx.doi.org/10.1093/bioinformatics/btn428
47. Nairismägi M-L, Vislovukh A, Meng Q, Kratassiouk G, Beldiman C, Petretich M, Groisman R, Füchtbauer E-M, Harel-Bellan A, Groisman I. Translational control of TWIST1 expression in MCF-10A cell lines recapitulating breast cancer progression. Oncogene 2012; 31:4960-6; PMID:22266852; http://dx.doi.org/10.1038/onc.2011.650
48. Almog N, Ma L, Schwager C, Brinkmann BG, Beheshti A, Vajkoczy P, Folkman J, Hlatky L, Abdollahi A. Consensus micro RNAs governing the switch of dormant tumors to the fast-growing angiogenic phenotype. PLoS One 2012; 7:e44001; PMID:22952847; http://dx.doi.org/10.1371/journal.pone.0044001
49. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci USA 2004; 101:2999-3004; PMID:14973191; http://dx.doi.org/10.1073/pnas.0307323101
50. Kota J, Chivukula RR, O'Donnell KA, Wentzel EA, Montgomery CL, Hwang H-W, Chang T-C, Vivekanandan P, Torbenson M, Clark KR, et al. Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. Cell 2009; 137:1005-17; PMID:19524505; http://dx.doi.org/10.1016/j.cell.2009.04.021
51. Wang Q, Huang Z, Guo W, Ni S, Xiao X, Wang L, Huang D, Tan C, Xu Q, Zha R, et al. microRNA-202-3p Inhibits Cell Proliferation by Targeting ADP-Ribosylation Factor-Like 5A in Human Colorectal Carcinoma. Clin Cancer Res 2013; PMID:24327274; http://dx.doi.org/10.1158/1078-0432.CCR-13-1023
52. Zhao Y, Li C, Wang M, Su L, Qu Y, Li J, Yu B, Yan M, Yu Y, Liu B, et al. Decrease of miR-202-3p expression, a novel tumor suppressor, in gastric cancer. PLoS One 2013; 8:e69756; PMID:23936094; http://dx.doi.org/10.1371/journal.pone.0069756
53. Buechner J, Tømte E, Haug BH, Henriksen JR, Løkke C, Flægstad T, Einvik C. Tumour-suppressor microRNAs let-7 and mir-101 target the proto-oncogene MYCN and inhibit cell proliferation in MYCN-amplified neuroblastoma. Br J Cancer 2011; 105:296-303; PMID:21654684; http://dx.doi.org/10.1038/bjc.2011.220

54. Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP. The microRNAs of Caenorhabditis elegans. Genes Dev 2003; 17:991-1008; PMID:12672692; http://dx.doi.org/10.1101/gad.1074403
55. Mayr C, Hemann MT, Bartel DP. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. Science 2007; 315:1576-9; PMID:17322030; http://dx.doi.org/10.1126/science.1137999
56. Ruike Y, Ichimura A, Tsuchiya S, Shimizu K, Kunimoto R, Okuno Y, Tsujimoto G. Global correlation analysis for micro-RNA and mRNA expression profiles in human cell lines. J Hum Genet 2008; 53:515-23; PMID:18465083; http://dx.doi.org/10.1007/s10038-008-0279-x
57. Zhang H, Hao Y, Yang J, Zhou Y, Li J, Yin S, Sun C, Ma M, Huang Y, Xi JJ. Genome-wide functional screening of miR-23b as a pleiotropic modulator suppressing cancer metastasis. Nat Commun 2011; 2:554; PMID:22109528; http://dx.doi.org/10.1038/ncomms1555
58. Wang XF, Shi ZM, Wang XR, Cao L, Wang YY, Zhang JX, Yin Y, Luo H, Kang CS, Liu N, et al. MiR-181d acts as a tumor suppressor in glioma by targeting K-ras and Bcl-2. J Cancer Res Clin Oncol 2012; 138:573-84; PMID:22207524; http://dx.doi.org/10.1007/s00432-011-1114-x
59. Pekarsky Y, Santanam U, Cimmino A, Palamarchuk A, Efanov A, Maximov V, Volinia S, Alder H, Liu C-G, Rassenti L, et al. Tcl1 expression in chronic lymphocytic leukemia is regulated by miR-29 and miR-181. Cancer Res 2006; 66:11590-3; PMID:17178851; http://dx.doi.org/10.1158/0008-5472.CAN-06-3613
60. Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nat Genet 2007; 39:673-7; PMID:17401365; http://dx.doi.org/10.1038/ng2003
61. Wang W-Y, Chien Y-C, Wong Y-K, Lin Y-L, Lin J-C. Effects of KRAS mutation and polymorphism on the risk and prognosis of oral squamous cell carcinoma. Head Neck 2012; 34:663-6; PMID:21688344; http://dx.doi.org/10.1002/hed.21792
62. Pan X-M, Sun R-F, Li Z-H, Guo X-M, Zhang Z, Qin H-J, Xu G-H, Gao L-B. A let-7 KRAS rs712 polymorphism increases colorectal cancer risk. Tumour Biol 2014; 35:831-5; PMID:23975373
63. Li Z-H, Pan X-M, Han B-W, Guo X-M, Zhang Z, Jia J, Gao L-B. A let-7 binding site polymorphism rs712 in the KRAS 3′ UTR is associated with an increased risk of gastric cancer. Tumour Biol 2013; 34:3159-63; PMID:23729275; http://dx.doi.org/10.1007/s13277-013-0885-x
64. Halvorsen M, Martin JS, Broadaway S, Laederach A. Disease-associated mutations that alter the RNA structural ensemble. PLoS Genet 2010; 6:e1001074; PMID:20808897; http://dx.doi.org/10.1371/journal.pgen.1001074