Application of Computational Tools for Identification of miRNA and Their Target SNPs

George Priya Doss. C[1], Dike IP[2], Rao Sethumadhavan[1]*

[1]Bioinformatics Division, School of Biotechnology, Chemical and Biomedical Engineering, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India
[2] Department of Biological Sciences, Covenant Univeristy, Nigeria

*Corresponding author: Dr. Rao Sethumadhavan, Bioinformatics Division, School of Biotechnology, Chemical and Biomedical Engineering, Vellore Institute of Technology Vellore 632014, Tamil Nadu, India Tel: 0416-2202522, 2202523; Fax: 0416-2243092; E-mail: rsethumadhavan@vit.ac.in

Abstract

MicroRNAs (miRNAs) are a class of small non-protein-coding RNAs that play important regulatory roles by targeting for cleavage or translational repression and involved in diverse biological functions. Accumulation of large amount of biological data indicates that miRNAs can function as tumor suppressors and oncogenes. Mutation, misexpression, and altered mature miRNA processing are implicated in carcinogenesis and tumor progression. Common single-nucleotide polymorphisms (SNPs) in miRNAs may change their property through altering miRNA expression and/or maturation, and thus they may have an effect on thousands of target mRNAs, resulting in diverse functional consequences. In this work we used computational tools to predict the functional role of mRNAs targeted by miRNA in colon cancer genes. We have presented a method which allows the use of PupaSuite, UTRscan and miRBase as a pipeline for the prediction of miRNA and their target, and evaluated the functional role of mRNA in colon cancer.

Keywords: miRNA; SNPs; miRBase; PupaSuite; UTRscan

Introduction

Identifying the genes and mutations underlying phenotypic variation is one of the primary objectives of modern genetics, especially for traits of medical importance. Over the past decades, studies have analyzed and unveiled the genetic variants in the human genome, such as single-nucleotide polymorphisms (SNPs) (Stranger et al., 2007) which contribute to gene expression variation, and eventually to phenotypic variation in human populations (Rockman and Kruglyak, 2006). Single-nucleotide polymorphisms (SNPs) are the most frequent variation in the human genome, occurring once every several hundred base pairs throughout the genome. They have been studied extensively for defining the regions of disease candidate genes (Bernig and Chanock, 2006). The majority of SNPs in the genome occur in untranslated, intronic or intergenic regions. These SNPs could affect complex diseases through their effect on gene expression quantitatively. It is increasingly recognized that regulatory mutations could make a significant contribution to genetic variation, especially for complex traits, including disease susceptibility. Recent reports have shown that mutations in the coding region disrupting sequences recognized by splicing regulators such ESE or ESS can be considered an additional mutation mechanism leading to disease in humans (Pagani et al., 2003). This finding is particularly important for genetic counseling, where the pathogenicity assessment of any nucleotide substitution is crucial to correctly predict cancer risk. Taking in to this account,

our group recently published a review on the use of computational tools to identify deleterious SNPs in both coding and noncoding regions taking *TP53* as a pipeline gene (George et al., 2008).

MicroRNAs (miRNAs) are small 21-25 nucleotide non-protein-coding RNAs that comprise an evolutionarily conserved class of ribo-regulators which modulate gene expression via the RNA interference pathway (Ambros, 2004). miRNAs are thought to regulate gene expression post-transcriptionally by forming Watson-Crick base pairs with target mRNAs. miRNAs use two distinct post-transcriptional mechanisms to down-regulate gene expression. They act by binding to the complementary sites on the 3' untranslated region (UTR) of the target gene to induce cleavage with near perfect complementarity or to repress productive translation (Brennecke et al., 2005) and also facilitate deadenylation, which leads to rapid mRNA decay (Wu et al., 2006). Accumulation of large amount of biological data suggests that a single miRNA could bind to hundreds of mRNA targets, and these targets could be implicated in the regulation of almost every biological process (Zamore and Helay, 2005). miRNAs assert their function as oncogenes or tumor suppressor genes via several potential mechanisms in various forms of cancers (Calin et al., 2002; Cimmino et al., 2005; Borkhardt et al., 2006). Literature survey shows that miRNAs are also involved in solid tumors such as lung cancer, breast cancer and colorectal cancer (CRC) (Iorio et al., 2005; Karube et al., 2005; Michael et al., 2003). Loss or amplification of miRNA genes has been reported in a variety of cancers, and altered patterns of miRNA expression may affect cell cycle and survival programs (Yanaihara et al., 2006). Because small variation in the quantity of miRNAs may have an effect on thousands of target mRNAs and result in diverse functional con-
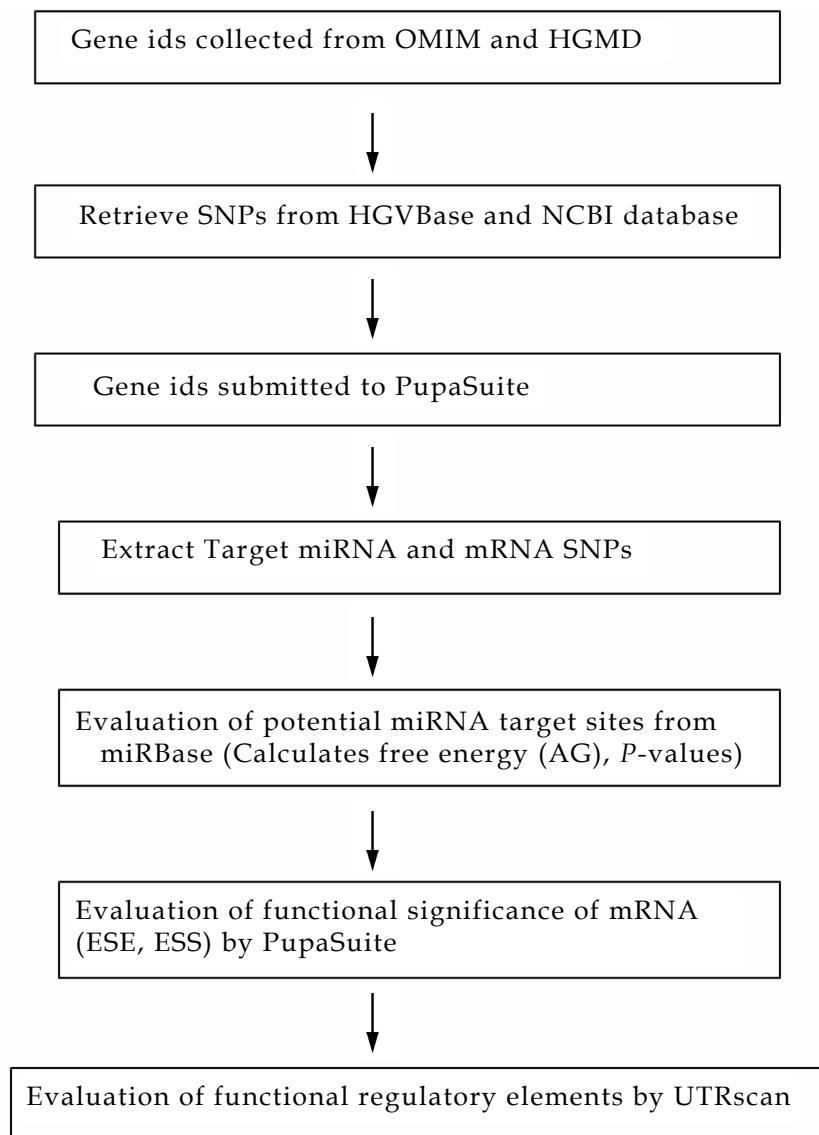
Gene ids collected from OMIM and HGMD

↓

Retrieve SNPs from HGVBase and NCBI database

↓

Gene ids submitted to PupaSuite

↓

Extract Target miRNA and mRNA SNPs

↓

Evaluation of potential miRNA target sites from miRBase (Calculates free energy (AG), *P*-values)

↓

Evaluation of functional significance of mRNA (ESE, ESS) by PupaSuite

↓

Evaluation of functional regulatory elements by UTRscan

Figure 1: Flow-chart of the proposed methodology for the prediction of miRNA and their targets.

sequences, the most common genetic variation, SNPs, in miRNA sequences may also be functional and therefore may represent ideal candidate biomarkers for cancer prognosis. In recent years, few studies have described role of naturally occurring human polymorphisms associated with miRNAs and their target sites (Iwai and Naraba, 2005; Chen and Rajewsky, 2006). Clop et al. (2006) stated that the SNP may affect organismal phenotype by altering a miRNA target site, leading to a significant alteration of protein expression. Given the wealth of data that is currently available in databases for human SNPs, such as the human genome variation database, HGVBase (Fredman et al., 2002) and the National Center for Biotechnology Information (NCBI) database, dbSNP (Smigielski et al., 2000), we can begin to identify naturally occurring variation associated with miRNAs and their targets using an *in silico* approach. SNPs in critical components of the miRNA system may have important phenotypic consequences, with implications for both evolutionary studies and biomedical research. In this study, we conducted a bioinformatics genome-wide survey on genes causing colon cancer and identified miRNA and their target mRNA using computational algorithims like PupaSuite (Reumers et al., 2008) which uses miRanda (John et al., 2005) and miRBase (Griffiths-jones et al., 2008). We searched for SNPs that would potentially affect novel target sites in humans and speculate that some of these variations may have functional effects using UTRscan and PupaSuite. These methods are based on different principles and merging them into a pipeline for predicting the function of target mRNA is therefore meaningful. Flow-chart of the proposed methodology is depicted in Figure 1.

Materials and Methods

Searching of gene ids and their single nucleotide polymorphisms

A total of 20 genes responsible for causing colon cancer (Familial adenomatous polyposis and Lynch syndrome) as shown in Table 1 were collected from OMIM and HGMD and submitted to PupaSuite to extract miRNA and their target mRNA. Based on the results obtained from PupaSuite, we retrieved the information for UTR SNPs of the genes namely *APC*, *AXIN2*, *CTNNB1*, *MAP2K4*, *MLH1*, *MSH2*, *MSH3*, *MYH*, *TP53* and *STK11* from the Human genome variation database, HGVBase and National Center for Biotechnology Information (NCBI) database, dbSNP for our computational analysis.

Scanning for miRNA and their target sites

PupaSuite is a unique and more integrated interface of PupaSNP (Conde et al., 2004) and PupasView (Conde et al., 2005) are now synchronized to deliver annotations for both noncoding and coding SNP, as well as annotations for

the SwissProt set of human disease mutations. In this approach, the input consists of a list of genes (genes belonging to a given pathway, involved in a particular biological function, etc.) and the user must specify the type of gene identifiers by selecting either Ensembl or an external database (which include GenBank, Swissprot/TrEMBL and other gene ids supported by Ensembl). Pupasuite retrieves SNPs that could affect conserved regions that the cellular machinery uses for the correct processing of genes (intron/exon boundaries or exonic splicing enhancers) and uses miRanda an algorithm for the detection of potential microRNA target sites in genomic sequences.

Analyzing miRNA and their target sites

For the detection of potential microRNA target sites in genomic sequences, we applied miRBase to validate our

Gene	Chromosome location
APC	5q21
AXIN2	17q24.1
BMPR1A	10q22.3
CTNNB1	3p22-p21.3
DCC	18q21.3
EP300	22q13.2
KRAS2	12p12
MAP2K4	17p11
MLH1	3p21.3
MSH2	2p22-p21
MSH3	5q11-q12
MSH6	2p16
MYH	1p34.3-p32.1
TP53	17p13
PIK3CA	3q26.3
PMS1	2q31-33
PMS2	7p22.2
PTEN	10q23.3
SMAD4	18q21.1
STK11	19p13.3

Table 1: List of genes responsible for causing colon cancer.

prediction. miRBase scans one or more miRNA sequences against all sequences and potential target sites are reported. A dynamic programming local alignment is carried out between the query miRNA sequence and the reference sequence. The alignment procedure scores are based on sequence complementarity between A:U and G:C matches and not on sequence identity. Then, it takes the high-scoring alignments of those above as a score threshold, detected from local alignment and estimates the thermodynamic stability of RNA duplexes based on these alignments. For this method, miRanda utilizes folding routines from the RNAlib library, which is part of the ViennaRNA (Wuchty et al., 1999). The free energy (ΔG) of optimal strand-strand interaction between miRNA and UTR is computed by the

Vienna RNA folding routines and is a measure for the thermodynamic stability of a duplex. *P*-values for all target sites were calculated in miRinda based on the model proposed by Rehmsmeier et al., (2004). Base *P*-value is computed using distribution parameters derived from the genomic background of miRanda scores.

Scanning of noncoding SNPs

Functional significance of each SNP in untranslated region (UTR) was determined by UTRscan (Pesole and Liuni, 1999) available at (http://www.ba.itb.cnr.it/BIG/UTRScan). UTResource, which is an internet resource of sequence analysis of 5' and 3' UTR of eukaryotic mRNAs which are involved in many posttranscriptional regulatory pathways that control mRNA localization, stability, and translation efficiency (Sonenberg, 1994; Nowak, 1994). Briefly, two or three sequences of each UTR SNP that have a different nucleotide at an SNP position were analyzed by UTRscan, which looks for UTR functional elements by searching through user submitted sequence data for the patterns defined in the UTR site and UTR databases. If different sequences for each UTR SNP are found to have different functional patterns, this UTR SNP is predicted to have functional significance. The internet resources for UTR analysis were UTRdb and UTRsite. UTRdb contains experimentally proven biological activity of functional patterns of UTR sequence from eukaryotic mRNAs (Pesole et al., 2002). The UTRsite has the data collected from UTRdb and also is continuously enriched with new functional patterns.

Results

Prediction of miRNA and their target sites

Among the 20 genes which were submitted to PupaSuite, only 10 genes (50%) displayed miRNA and their targets. Pupasuite scans the whole genome to find SNPs located at miRNAs. PupaSuite uses miRanda an algorithm for the detection of potential microRNA target sites in genomic sequences. Four miRNAs namely hsa-miR-663, hsa-miR-328, hsa-miR-325 showed a target mRNA SNP rs10415095 and two miRNAs namely hsa-miR-638 and hsa-miR-572 showed a target mRNA SNP rs11552326 in *STK11* gene, two miRNAs namely hsa-miR-637 and hsa-miR-557 showed a target mRNA SNP rs35352891 and two miRNAs namely hsa-miR-539 and hsa-miR-431 showed a target mRNA SNP rs3219496 in *MYH* gene, two miRNAs namely hsa-miR-545 and hsa-miR-135b showed one target mRNA SNP rs17225060 in *MSH2* gene, three miRNAs namely hsa-miR-302c, hsa-miR-372 and hsa-miR-512-3p showed one target mRNA SNP rs1803985 in *MLH1* gene and two miRNAs namely hsa-miR-522 and hsa-miR-642 showed one target mRNA SNP in *APC* gene as shown in Table 2. PupaSiute uses miRanda, an algorithm for the detection of potential

microRNA target sites in genomic sequences, to localize all the SNPs situated in the region 3' UTR of these targets sites. miRanda calculated the scores based on the complementarity of nucleotides (A=U or G=C) and G=U wobble pairs, which are important for the accurate detection of RNA:RNA duplexes. The result is a score (S) for each detected complementarity match between a miRNA and a potential target gene and range is between 15.1679 to 19.0061. The minimum free energy (MFE) of the miRNA–target duplex was determined while predicting the miRNA target sites. The lower MFE values of the miRNAs and the target sites reveal the energetically more probable hybridizations between the miRNAs and the target genes. It can be seen from Table 1 that miRNAs predicted by PupaSuite exhibited a low minimum free energy range from -10.09 to -34.79. It is an important noteworthy finding in miRNA and their target mRNAs analysis using PupaSuite. Recent publications suggest that multiple potential binding sites of a miRNA in a single target are good evidence for the target being regulated by the miRNA. If we consider a potential binding site being a rare event in our random model, the number of binding sites can be approximated by a Poisson distribution (Hofacker et al., 1994). The provision of *P*-value and Base *P*-value for the miRNA allows the users to assess the confidence in the prediction. Based on the above observations depicted in Table 2 the miRNAs namely hsa-miR-622, hsa-miR-641, hsa-miR-560, hsa-miR-611, hsa-miR-637, hsa-miR-557, hsa-miR-663, hsa-miR-572, hsa-miR-560 and hsa-miR-638 predicted by PupaSuite by our investigation is also well documented by on experimental protocols (Bandres et al., 2006; Cummins et al., 2006). This is an important result from this work.

Predictions of potential phenotypic effect in SNPs

We further analyzed the miRNA targets and predicted the role of mRNA SNPs using PupaSuite. Besides indispensable *cis* regulatory motifs such as 5' and 3' splice sites and branch points, there are other *cis* regulatory sequences called exonic or intronic splicing enhancers and silencers (Fu, 2004). These sequences are recognized by a number of regulatory proteins, represented, for example, by serine/arginine-rich (SR) proteins, which bind RNA with limited sequence specificity (Liu et al., 1998). ESEs are common in alternative and constitutive exons, where they act as binding sites for Ser/Arg-rich proteins (SR proteins), a family of conserved splicing factors that participate in multiple steps of the splicing pathway (Graveley, 2000). ESSs are sequence elements that are known to regulate alternative splicing and also play a role in splice site selection (Fairbrother and Chasin, 2000). As a result of premRNA splicing, different combinations of exons may arise and may be a 'natural' cause of

Gene ids	miRNA	Target UTR SNPs	Allele frequency	Functional significance	Score	Mfe	P-value	Base P-value
APC	hsa-miR-522	**rs397768**	G:132.00	ESE	16.8787	-10.09	0.0101492	3.680990e-02
	hsa-miR-642	rs397768	G:161.00	-	18.161	-21.73	0.0298284	3.028230e-02
AXIN2	hsa-miR-622	**rs10438779**	T: 112.00	ESE	16.3782	-27.41	0.00466768	5.088980e-02
CTNNB1	hsa-miR-768-3p	rs34854135	A: 169.00	-	15.8323	-24.71	0.00213437	4.151550e-02
	hsa-miR-185	rs4135388	C: 155.00	-	16.7094	-21.79	0.0173813	5.877850e-02
	hsa-miR-641	**rs1722851**	T: 150.00 G: 143.00	ESE	18.0451	-23.86	0.00108233	1.163210e-02
MAP2K4	hsa-miR-487b	rs1049443	A: 86.00	ESE	14.2729	-17.17	0.00145788	7.669690e-02
MLH1	hsa-miR-302c	rs1803985	G: 161.00	ESS	18.8929	-17.57	0.00354918	4.574460e-03
	hsa-miR-372	rs1803985	G: 159.00	ESS	19.0061	-21.39	0.00458024	7.808160e-03
	hsa-miR-512-3p	rs1803985	T: 153.00	ESS	18.1925	-18.63	0.00421638	1.326750e-02
MSH2	hsa-miR-545	rs17225060	A: 146.00	-	16.8281	-20.72	0.0090846	4.087860e-02
	hsa-miR-135b	rs17225060	G: 122.00	-	16.9697	-12.56	0.00122064	2.499990e-02
MSH3	hsa-miR-560	**rs394592**	G: 159.00 A: 123.00	-	16.8712	-32.69	0.00978912	9.837350e-03
	hsa-miR-373	rs34505312	C: 111.00	-	16.7582	-18.84	0.000533874	7.606520e-02
MYH	hsa-miR-539	rs3219494	C: 96.00	-	17.6012	-20.04	0.0242289	2.452720e-02
	hsa-miR-611	rs34961970	T: 99.00	-	15.6309	-27.88	0.0384867	3.924690e-02
	hsa-miR-637	rs35352891	G: 114.00	-	17.3388	-26.19	0.0423881	4.331270e-02
	hsa-miR-611	rs36053993	C: 76.00 T: 80.00	ESE	15.6309	-27.88	0.0384867	3.924690e-02
	hsa-miR-539	rs3219496	G: 135.00 T: 142.00	ESS	17.6012	-20.04	0.0242289	2.452720e-02
	hsa-miR-431	rs3219496	G: 153.00 T: 157.00	ESS	16.5305	-24.33	0.0106976	2.958700e-02
	hsa-miR-557	rs35352891	G: 99.00	-	17.3708	-20.98	0.0268101	2.717600e-02
TP53	hsa-miR-150	rs2543552	A: 137.00 G: 134.00	-	15.969	-23.79	0.000272854	6.649060e-02
	hsa-miR-453	**rs35919705**	G: 118.00	-	15.8994	-22.81	0.00418259	5.098810e-02
	hsa-miR-375	rs1794293	C: 104.00	ESE	15.1804	-18.45	0.000406331	5.230040e-02
STK11	hsa-miR-663	**rs10415095**	T: 164.00 C: 188.00	ESE	17.0637	-34.79	0.000105978	3.906140e-02
	hsa-miR-328	rs10415095	C: 102.00	ESE	16.2596	-22.31	0.0068421	8.633920e-02
	hsa-miR-572	rs11552326	G: 245.00 C: 252.00	ESS	15.1679	-24.79	0.00716065	7.014360e-02
	hsa-miR-560	rs11552326	G: 252.00 C: 257.00	ESS	16.6124	-32.43	0.00215883	1.170500e-02
	hsa-miR-325	rs10415095	T: 110.00	ESE	15.2574	-21.72	0.0291086	7.981930e-02
	hsa-miR-638	rs10415095	T: 124.00 C: 125.00	ESE	16.3958	-27.83	0.0337589	3.434190e-02

ESE- Exon Splicing Enhancer; ESS- Exon Splicing Silencer; SNPs highlighted in bold show functional pattern change of 15 LOX-DICE by UTRscan

Table 2: miRNA and their target mRNA in colon cancer.

errors in gene expression by introducing premature termination codons or altering protein structure leading to changes in spectrum of interacting proteins, intracellular localization, protein stability or posttranslational modification (Stamm et al., 2005). Seven SNPs rs397768, rs10438779, rs1722851, rs1049443, rs36053993, rs1794293 and rs10415095 with ids were predicted to disrupt the exonic splicing enhancers and three SNPs rs1803985, rs3219496 and rs11552326 with ids were predicted to disrupt the exonic splicing silencers by PupaSuite.

Discussion

A revolution is underway in the approach to studying the genetic basis of cancer. In the past, most studies have focused on protein coding genes and their regulation at the transcriptional level. The recent explosion of miRNA research and discovery further underscores the importance of these regulatory molecules in many key biological processes, such as development, cellular differentiation, cell cycle control and apoptosis. Polymorphisms and mutations in the corresponding sequence space (machinery, miRNA precursors and target sites) are likely to make a significant contribution to phenotypic variation, including disease susceptibility. The mutations in miRNAs or polymorphisms in the mRNAs targeted by miRNAs may also contribute to cancer predisposition and progression (Saunders et al., 2007). Their expression profiles can be used for the classification, diagnosis, and prognosis of human malignancies. Most of the mutations identified till date lead to alter in primary sequence and hence alter in protein structure (missense or nonsense, insertion–deletions in the open reading frame, or mutations causing splicing errors). Recent studies show that regulatory mutations could make a significant contribution to genetic variation, including disease susceptibility lead to the identification of the mutations in regulatory variants (rSNPs) affecting transcript levels in cis (Pastinen et al., 2006). The most common interpretation of such cis effects is that the corresponding variants are modulating the activity of regulatory elements, including promotors and enhancers.

We applied computational tools like PupaSuite, miRBase and UTRscan to validate miRNA and their targets using colon cancer genes. PupaSuite uses miRanda algorithm for the identification of miRNAs and their target mRNA. Out of twenty genes retrieved from OMIM causing colon cancer, PupaSuite predicted miRNA and their target mRNA for only ten genes (50%). A total of thirty miRNAs and mRNAs were obtained from PuaSuite for further analysis. Of those predicted target genes in causing colon cancer (Table 2), miRNAs in genes namely *APC*, *MLH1*, *MSH2*, *MYH* and *STK11* had more than one predicted target interaction site. These results suggest that 3' UTRs with more

than one predicted target site for a given miRNA are more reliable than those with a single site. This concept of multiple miRNAs binding sites with target mRNA is well supported by experimental analysis in drosophila (Enright et al., 2003). Out of the predicted thirty miRNAs, ten miRNAs (33%) namely hsa-miR-622, hsa-miR-641, hsa-miR-560, hsa-miR-611, hsa-miR-637, hsa-miR-557, hsa-miR-663, hsa-miR-572, hsa-miR-560 and hsa-miR-638 predicted by PupaSuite by our investigation is also well documented by on experimental protocols (Bandres et al., 2006; Cummins et al., 2006). Functional role of target mRNA SNPs were validated using PuaSuite and UTRscan. *In-silico* methods provide a useful tool for an initial approach to any mutation suspected of causing aberrant RNA processing. These mutations can result in either complete skipping of the exon, retention of the intron or the introduction of a new splice site within an exon or intron. In rare cases, mutations that do not disrupt or create a splice site, activate preexisting pseudo splice sites consistent with the proposal that introns contain splicing inhibitory sequences (Baralle and Baralle, 2005). Recent studies showed that the mutations in *cis* splicing regulating sequences, which might shift production to mRNA with cancer-prone potential (Scholzova et al., 2007). Yang et al. (2003) showed, however, that this mutation disrupts an exonic splicing enhancer and leads to production of null protein due to aberrant splicing. Among the twenty one targeted SNPs by miRNAs, seven SNPs (38%) rs397768, rs10438779, rs1722851, rs1049443, rs36053993, rs1794293 and rs10415095 with ids were predicted to disrupt the exonic splicing enhancers and three SNPs rs1803985, rs3219496 and rs11552326 with ids were predicted to disrupt the exonic splicing silencers, whereas thirteen SNPs (62%) showed no functional significance by PupaSuite. Varied levels of alternative splicing have been detected for some of the splicing mutations in colon cancer genes (Lastella et al., 2006; Gorlov et al., 2003). By UTRscan, six SNPs (29%) showed a functional pattern change 15-LOX-DICE. Founding members of miRNAs were discovered by genetic screening approaches, experimental approaches were limited by their low efficiency, time consuming, and high cost. As a consequence, several web-based or non-web-based computer software programs are publicly available for predicting miRNAs and their targets have been devised in order to predict targets for follow up experimental validation. Even though many computational methods for the identification of miRNA may have its own limitations, but there is no other option now other than to use computational methods for miRNA predictions. The next step in miRNA research is to identify and experimentally validate their mRNA targets. Since direct experimental methods for

discovering miRNA targets are lacking, a large number of target prediction algorithms have been developed. Our results from this study suggests that the application of computational algorithms, PupaSuite and UTRscan analysis might provide an alternative approach to select target SNPs by understanding the effect of SNPs on the functional attributes or molecular phenotype of a protein. Our result also endorses a study with an *in vivo* experimental protocol. Studies using SNPs to probe the genetic basis of human disease can provide insights into susceptibility to a disease, modification of the phenotype of a monogenic disease, and response to pharmacologic treatment. The functional analysis in this study may be a good model for further research in genetically inherited disease.

Conclusion

We have presented computational tools for the identification of miRNA and their target mRNA in colon cancer. We tried to predict the functional roles of SNPs in mRNA region.

Based on this, we derived at the following conclusions:

Among the twenty genes selected for our analysis, miRNAs and their target mRNA are exhibited only by ten genes. Of these, only five genes showed multiple miRNA

interactive sites for single mRNA.

Out of thirty targeted SNPs by miRNAs, only seven SNPs disrupted the exonic splicing enhancers, three SNPs disrupted the exonic splicing silencers while thirteen SNPs showed no functional significance. Six SNPs exhibited functional pattern change of 15-LOX-DICE in un-translated regions.

We emphasize that our approach in selecting miRNAs and their target mRNA in colon cancer using computational tools is of significant importance and the same methodology could be adapted to other types of cancer genes also. Evaluation of target mRNA functional role will be a major challenge of future studies in the field of cancer biomarker research and other types of disease.

Acknowledgements

References

1. Ambros V (2004) The functions of animal microRNAs. Nature 431:350–355.» CrossRef » Pubmed » Google Scholar

2. Bandres E, Cubedo E, Agirre X, Malumbres R, Zarate R, et al. (2006) Identification by Real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. Mol Cancer 5: 29. »CrossRef » Pubmed » Google Scholar

3. Baralle D, Baralle M (2005) Splicing in action: assessing disease causing sequence Changes. J Med Genet 42: 737-748. »CrossRef » Pubmed » Google Scholar

4. Bernig T, Chanock SJ (2006) Challenges of SNP genotyping and genetic variation: its future role in diagnosis and treatment of cancer. Expert Rev Mol Diagn. 6:319–331. » CrossRef » Pubmed » Google Scholar

5. Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of micro- RNA-target recognition. PLOS Biol 3: e85.»CrossRef » Pubmed » Google Scholar

6. Borkhardt A, Fuchs U, Tuschl T (2006) MicroRNA in chronic lymphocytic Leukemia. N Engl J Med 354: 524-5. » CrossRef » Pubmed » Google Scholar

7. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, et al. (2002) Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci 99: 15524–9. » CrossRef » Pubmed » Google Scholar

8. Chen K, Rajewsky N (2006) Natural selection on human microRNA binding sites inferred from SNP data. Nat Genet 38:1452-1456.» CrossRef » Pubmed » Google Scholar

9. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, et al. (2005) miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci 102:13944–9. » CrossRef » Pubmed » Google Scholar

10. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, et al. (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet 38:813-818.   » CrossRef » Pubmed » Google Scholar

11. Conde L, Vaquerizas JM, Santoyo J, Al-Shahrour F, Ruiz-Llorente S, et al. (2004) PupaSNP Finder: a web tool for finding SNPs with putative effect at transcriptional level. Nucleic Acids Res 32: W242–W248. » CrossRef » Pubmed » Google Scholar

12. Conde L, Vaquerizas JM, Ferrer CC, Cruz X, Orozco

M, et al. (2005) Pupas View: a visual tool for selecting suitable SNPs, with putative pathological effect in genes, for genotyping purposes. Nucleic Acids Res 33: W501-W505. » CrossRef » Pubmed » Google Scholar

13. Cummins JM, He Y, Leary RJ, Pagliarini R, et al. (2006) The colorectal microRNAome. Proc Natl Acad Sci USA 103:3687-3692.» CrossRef » Pubmed » Google Scholar

14. van Deventer SJ (2000) Cytokine and cytokine receptor polymorphisms in infectious Disease, Intensive Care Med 26: S98-S102. » CrossRef » Pubmed » Google Scholar

15. Enright AJ, John B, Gaul U, Tuschl T, Sander C, et al. (2003) MicroRNA targets in Drosophila. Genome Biol 5: R1.» CrossRef » Pubmed » Google Scholar

16. Fairbrother WG, Chasin LA (2000) Human genomic sequences that inhibit splicing. Mol Cell Biol 20:6816–6825. » CrossRef » Pubmed » Google Scholar

17. Fredman D, Siegfried M, Yuan YP, Bork P, Lehvslaiho H, et al. (2002) HGVbase: a human sequence variation database emphasizing data quality and a broad spectrum of data sources. Nucleic Acids Res 30: 387–391. » CrossRef » Pubmed » Google Scholar

18. Fu XD (2004) Towards a splicing code. Cell 119: 736–738.» CrossRef » Pubmed » Google Scholar

19. George PDC, Sudandiradoss C, Rajasekaran R, Choudhury P, Sinha P, et al. (2008) Applications of computational algorithm tools to identify functional SNPs, Funct Integr Genomics 8: 309-316.

20. Graveley BR (2000) Sorting out the complexity of SR protein functions. RNA 6:1197–1211. » CrossRef » Pubmed » Google Scholar

21. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36: 154-158.» CrossRef » Pubmed » Google Scholar

22. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer S, Tacker M, et al. (1994) Fast Folding and Comparison of RNA Secondary Structures. Monatshefte f. Chemie 125: 167-188. » CrossRef » Google Scholar

23. Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, et al. (2005) MicroRNA gene expression deregulation in human breast cancer. Cancer Res 65: 7065–70.» CrossRef » Pubmed » Google Scholar

24. Gorlov IP, Gorlova OY, Frazier ML, Amos CI (2003) Missense Mutations in hMLH1 and hMSH2 Are Associated with Exonic Splicing Enhancers. Am J Hum Genet 73:1157–1161.» CrossRef » Pubmed » Google Scholar

25. Iwai N, Naraba H (2005) Polymorphisms in pre-

microRNAs. Biochem Biophys Res Commun 331:1439-1444. » CrossRef » Pubmed » Google Scholar

26. John B, Enright AJ, Aravin A, Tuschl T, Sander C, et al. (2005) Human MicroRNA targets, PLoS Biol 3: 264. » CrossRef » Pubmed » Google Scholar

27. Karube Y, Tanaka H, Osada H, Tomida S, Tatematsu Y, et al. (2005) Reduced expression of Dicer associated with poor prognosis in lung cancer patients. Cancer Sci 96: 111–5.» CrossRef » Pubmed » Google Scholar

28. Lastella P, Surdo NC, Resta N, Guanti G, Stella A (2006) In silico and in vivo splicing analysis of MLH1 and MSH2 missense mutations shows exon- and tissue-specific effects. BMC Genomics 7: 243. » CrossRef » Pubmed » Google Scholar

29. Liu HX, Zhang M, Krainer AR (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes Dev 12: 1998–2012. » CrossRef » Pubmed » Google Scholar

30. Michael MZ, O' Connor SM, van Holst Pellekaan NG, Young GP, James RJ (2003) Reduced accumulation of specific microRNAs in colorectal neoplasia. Mol Cancer Res 1882-91. » CrossRef » Pubmed » Google Scholar

31. Nowak R (1994) Mining treasures from 'junk DNA'. Science 263:608–610. » CrossRef » Pubmed » Google Scholar

32. Pagani F, Stuani C, Tzetis M, Kanavakis E, Efthymiadou A, et al. (2003) New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. Hum Mol Genet 12: 1111-20.» CrossRef » Pubmed » Google Scholar

33. Pastinen T, Ge B, Hudson TJ (2006) Influence of human genome polymorphism on gene expression. Hum Mol Genet 15:R9-R16.» CrossRef » Pubmed » Google Scholar

34. Pesole G, Liuni S (1999) Internet resources for the functional analysis of 5' and 3' untranslated regions of eukaryotic mRNAs. Trends Genet 15: 378. » CrossRef » Pubmed » Google Scholar

35. Pesole G, Liuni S, Grillo G, Licciulli F, Mignone F, et al. (2002) UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. Nucleic Acids Res 30:335–340. » CrossRef » Pubmed » Google Scholar

36. Reumers J, Conde L, Medina I, Maurer-Stroh S, Durme JV, et al. (2008) Joint annotation of coding and non-coding single nucleotide polymorphisms and mutations in the SNPeffect and PupaSuite databases. Nucleic Acids Res 36: D825-D829. » CrossRef » Pubmed » Google Scholar

37. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes. RNA 10: 1507-1517. » CrossRef » Pubmed » Google Scholar

38. Rockman MV, Kruglyak L (2006) Genetics of global gene expression. Nat Rev Genet 7: 862-872. » CrossRef » Pubmed » Google Scholar

39. Saunders MA, Liang H, Li WH (2007) Human polymorphism at microRNAs and microRNA target sites. Proc Natl Acad Sci USA 104: 3300-3305. » CrossRef » Pubmed » Google Scholar

40. Scholzova E, Malik R, Sevcik J, Kleibl Z (2007) RNA regulation and cancer development. Cancer Lett 246:12-23. » CrossRef » Pubmed » Google Scholar

41. Smigielski EM, Sirotkin K, Ward M, Sherry ST (2000) dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res 28:352–355. » CrossRef » Pubmed » Google Scholar

42. Sonenberg N (1994) mRNA translation: influence of the 5' and 3' untranslated regions. Curr Opin Genet Dev 4:310-315. » Pubmed » Google Scholar

43. Stamm S, Ben AS, Rafalska I, Tang Y, Zhang Z, et al. (2005) Function of alternative splicing. Gene 344: 1-20. » CrossRef » Pubmed » Google Scholar

44. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, et al. (2007) Population genomics of human gene expression. Nat Genet 39:1217-1224. » CrossRef » Pubmed » Google Scholar

45. Wuchty S, Fontana W, Hofacker IL, Schuster P (1999) Complete suboptimal folding of RNA and the stability of secondary structures. Biopolymers 49:145-65. » CrossRef » Pubmed » Google Scholar

46. Wu L, Fan J, Belasco JG (2006) MicroRNAs direct rapid deadenylation of mRNA. Proc Natl Acad Sci USA 103: 4034-4039. » CrossRef » Pubmed » Google Scholar

47. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, et al. (2006) Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell 9:189-198. » CrossRef » Pubmed » Google Scholar

48. Yang Y, Swaminathan S, Martin BK, Sharan SK (2003) Aberrant splicing induced by missense mutations in BRCA1: clues from a humanized mouse model. Hum Mol Genet 12: 2121–2131. » CrossRef » Pubmed » Google Scholar

49. Zamore PD, Haley B (2005) Ribo-gnome: the big world of small RNAs. Science 309:1519-1524. » CrossRef » Pubmed » Google Scholar