

Minimum Free Energy Based Evaluation of mRNAs Secondary Structures Constructed by 18 Clinically Significant Exonic Single Nucleotide Polymorphisms (SNPs) and Haplotypes of 5 Missense SNPs of RB1 Gene

¹Sarder Nasir Uddin, ²Apurba Majumder,
³Khandker Khaldun Islam, ⁴Sk. Amir Hossain and ⁵Palash Kumar Sarker

^{1,3,4}Department of Biotechnology and Genetic Engineering, Khulna University, Bangladesh

²Department of Cell and Molecular Biology, University of Jyväskylä, Finland

⁵Department of Microbial Biotechnology, National Institute of Biotechnology, Bangladesh

Article history

Received: 27-05-2015

Revised: 11-08-2015

Accepted: 02-10-2015

Corresponding Author:

Sarder Nasir Uddin
Department of Biotechnology
and Genetic Engineering,
Khulna University, Bangladesh
E-mail: uddin@bge.ku.ac.bd;
opunasir@gmail.com

Abstract: Clinically significant 18 Single Nucleotide Polymorphisms (SNPs) from exon regions of *Retinoblastoma* gene (RB1) were analyzed to find out the structural variations in mRNAs. Online bioinformatic tools i.e., Vienna RNA, RNAfold were used for secondary structure analysis of mRNAs. Predicted minimum Free Energy Change (MFE) was calculated for mRNAs structures. It has been observed that the average of predicted MFE value from 13 nonsense mutations was higher (0.76 kcal/mol) in comparison to 5 missense mutations. Presumably, 13 nonsense mutations are responsible for Nonsense Mediated mRNA Decay (NMD), therefore, excluded from haplotype analysis. From the statistical analysis all the thermodynamic data obtained from four SNP haplotypes are significant ($p \leq 0.05$), followed by three-SNP haplotype data except Ensemble diversity ($p \leq 0.10$). Interestingly, MEF of Centroid Secondary Structure is highly significant ($p \leq 0.01$) in all the cases (Two-SNP haplotypes, Three-SNP haplotypes and Four-SNP haplotypes).

Keywords: RB1 Gene, Single Nucleotide Polymorphisms, Minimum Free Energy Change, Genomics, *Retinoblastoma*

Introduction

Role of RNAs in regulation of gene networks is now well characterized and the execution of the regulatory activities of the mRNA involves a wide range interaction with proteins, limiting RNA confirmation *in vivo* (Gregory, 2015; Shabalina *et al.*, 2014; 2013; Laederach, 2007). Thereby, DNA variants based structural changes of mRNA secondary structure likely affects the protein translation efficiency (Mita and Kuroiwa, 1989; Schmittgen *et al.*, 1994; Shalev *et al.*, 2002), while majority of mutations are transferred to the transcriptome (Morton, 2008). For example, Catecho-O-Mrthyltransferase (COMT) is a key regulator of pain, cognitive function and affective mood, modulate the protein expression by altering the mRNA secondary structure in presence of Single Nucleotide Polymorphism (SNP) (Nackley *et al.*, 2006). However, secondary structural analysis is focused on the functional motifs of mRNAs to determine the

thermodynamic properties, i.e., Minimum Free Energy (MFE). It is predicted due to stochastic molecular motion and other constraints. Therefore, the secondary structures of mRNAs could be analyzed based on the MFE and the ensemble optimal structures provided by different secondary structure prediction bioinformatic tools (Mathews *et al.*, 1999; Johnson *et al.*, 2011).

Different studies documented that DNA mutations causing variations in mRNA folding is associated with variation in gene expression. Kudla *et al.* (2009) found that various Green Fluorescent Protein (GFP) constructs those differ only in a single synonymous mutation, vary in their GFP expression significantly. The minimum free energy associated with the secondary structure of the first third of the mRNA construct correlates well with GFP expression. In consistent with this finding, Bartoszewski *et al.* (2010) reported a synonymous mutation that alters mRNA structures presumably causing cystic fibrosis. Additionally, it was hypothesized

that synonymous, nonsynonymous and UTR variants can potentially act in mildly deleterious and, in some cases, pathological fashion on pre- and post-translational levels through changes in RNA structure. Also, about 60% of the human exonic SNPs are predicted to alter RNA structure to some degree, as well as, haplotypes of SNP alleles possibly contribute such changes (Johnson *et al.*, 2011). Recently, single nucleotide induced changes of RNA conformation has been studied and it was recommended that SNPs could be used as a powerful tool study the impact on structural changes of RNAs as well as gene expression and function (Salari *et al.*, 2013). Summarizing these studies, we find scopes to identify putative functional gene-specific or exonic SNPs and their haplotypes within the positional candidate genes for desired traits or diseases those affect the secondary structures of mRNAs through predicted MFE.

It has been found that exonic SNPs of human RB1 gene are clinically significant, those contribute to develop a monogenic genetic disease, Retinoblastoma (RB) (Friend *et al.*, 1986; Yandell *et al.*, 1989). Retinoblastoma is the most familiar primary intraocular malignancy of childhood (Mahoney *et al.*, 1990) which initiates from the mutation in RB1 gene that resides in chromosome number 13. It is also known as embryonic malignant neoplasm of retinal origin, often bilateral and almost always found in early childhood. The RB1 is the first tumor suppressor gene that has been identified and cloned (Weinberg, 1995) and also it found that mutational inactivation of both alleles of the RB1 tumor suppressor gene is responsible for developing retinal tumor causing disease. Retinoblastoma gene encodes a ubiquitously expressed 110-kDa nuclear phosphoprotein called pRb. It is notable that mRNA produced from RB1 gene consists of 4772 base pairs, having 27 exons and one of the long chain mRNA in human transcriptome (Burkhart and Sage, 2008; Leiderman *et al.*, 2007; www.ncbi.nlm.nih.gov/snp/?term=Retinoblastoma). In the current study, we have evaluated 18 different exonic SNPs and haplotypes of 5 missense SNPs of RB1 those potentially induce structural variations in localized mRNA structure through the estimation of MFE structure and centroid free structure of RNA. Subsequently, we compared the free energy fluctuations of different mRNAs sequence arises from point mutations and haplotypes to find out causative SNPs those contribute to form higher free energy structures, therefore, likely have effect on RB1 gene expression.

Materials and Methods

SNP Data Mining from Genomic Databases

We have accumulated data for clinically significant SNPs of RB1 gene from the different genomic databases through web search. Primarily we

were focused on the SNP database found in National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov/snp/?term=Retinoblastoma) which is a part of the United States National Library of Medicine (NLM). Subsequently, we have checked other databases, i.e., Human Gene Mutation Database (www.hgmd.org), Locus Specific Database (LSDB; www.hgvs.org/dblist/glsdb.html) for the gene sequence similarity and SNP details. These databases are considered as open source database and provide information about the variants of the specific locus on any gene across different species. The gene sequence similarity search was performed through Basic Local Alignment Search Tool (BLAST) (Duan *et al.*, 2003) and the sequences we found in these databases were completely matched. Finally, we used the RB1 gene sequence from NCBI having location number NC.000013.10 and base positions from 48877883 to 49056026. Accumulatively we have found about 3000 gene-specific SNPs for RB1 gene. Among all the SNPs we found 23 as clinically significant/LSDB submission (www.hgmd.cf.ac.uk/docs/oth_mut.html). However, based on literature review as potentially causative 18 exonic SNPs has been taken for mRNA secondary structure analysis in the current study. We have excluded intronic SNPs because it is not related to mRNA of RB1 gene.

Separation of Nonsense Mediated Decay (NMD) Causing SNPs and Haplotypes Reconstruction

Among 18 exonic SNPs we excluded 13 SNPs while constructing haplotypes for structural analysis. These 13 SNPs were predicted to be responsible for NMD, therefore, considered as less important in case of constructing haplotypes. Positions of the nonsense SNPs might cause NMD have been given in Table 3. While constructing haplotypes, no software was used because we didn't have any real data as or pedigree for this study. Five exonic SNPs were rs137853292, rs121908692, rs137853294, rs121913295 and rs137853296 used and coded as SNP1, SNP2, SNP3, SNP4 and SNP5, respectively. In total, 26 possible haplotypes were constructed using the alleles of 5 SNPs and all of them were subjected to mRNA structure analysis.

mRNA Secondary Structure Analysis

To obtain mRNA secondary structure from the RB1 gene sequence we have used RNAfold (Capon *et al.*, 2004; Shabalina *et al.*, 2006; Altschul *et al.*, 1990; Hofacker *et al.*, 1994; McCaskill, 1990; Zuker and Stiegler, 1981; Lorenz *et al.*, 2011; Turner and Mathews, 2009; Bompfunerwerer *et al.*, 2008; Hofacker and Stadler, 2006; Mathews *et al.*, 2004; Jia and Luo, 2006; Jia *et al.*, 2004) version 2.1.7 online tool (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) from Vienna RNA package 2.0 (Hofacker *et al.*, 1994). This web tool gives us information on RNA secondary structure prediction through energy minimization. It

provides three kinds of dynamic programming algorithms for structure prediction (i) the minimum free energy algorithm of which yields a single optimal structure, (ii) the partition function algorithm of which calculates base pair probabilities in the thermodynamic ensemble and (iii) the suboptimal folding algorithm of which generates all suboptimal structures within a given energy range of the optimal energy. For secondary structure comparison, the package contains several measures of distance (dissimilarities) using either string alignment or tree-editing. Finally, it uses algorithm to design sequences with a predefined structure (McCaskill, 1990; Zuker and Stiegler, 1981; Lorenz *et al.*, 2011; Turner and Mathews, 2009). To analyze the RB1 gene sequence and to obtain predicted mRNA secondary structure we have used the default parameters of RNAfold.

Results

From the literature and data bases we found natural and some disease causing variants in the RB1 gene. Although in the dbSNP of NCBI many SNPs has been reported for RB1 gene region, we identified 18 exonic SNPs to be included within our range of study as they are

clinically significant in LSDB submission with appropriate research findings. These 18 SNPs resided under the region of 12 exons out of total 27 exons of RB1 gene. Among these, exon 14, 17, 18, 19, 20 and 21 contain multiple SNPs having importance in pathogenesis. The detail of the SNPs information found from databases and used for this study is summarized in the Table 1.

Thermodynamic parameters have been found for mRNAs of RB1 gene. In presence of minor alleles of these 18 exonic SNPs, subsequent secondary structures of mRNAs were obtained through RNAfold analysis. In each case, construct of the secondary structure by mutant allele of the SNPs produces the details of the mRNA structures in the output file of the analysis tool. These included (i) plain secondary structure, (ii) secondary structure with base pair probability, (iii) secondary structure with positional entropy and (iv) mountain plot. In addition to this, thermodynamic information of mRNA secondary structures was found as MFE prediction followed other thermodynamic ensemble prediction. Figure 1 shows the wild type mRNA structure with thermodynamic parameters of RB1 and other structures for mutant alleles have been included in supplemental files (Sup 1).

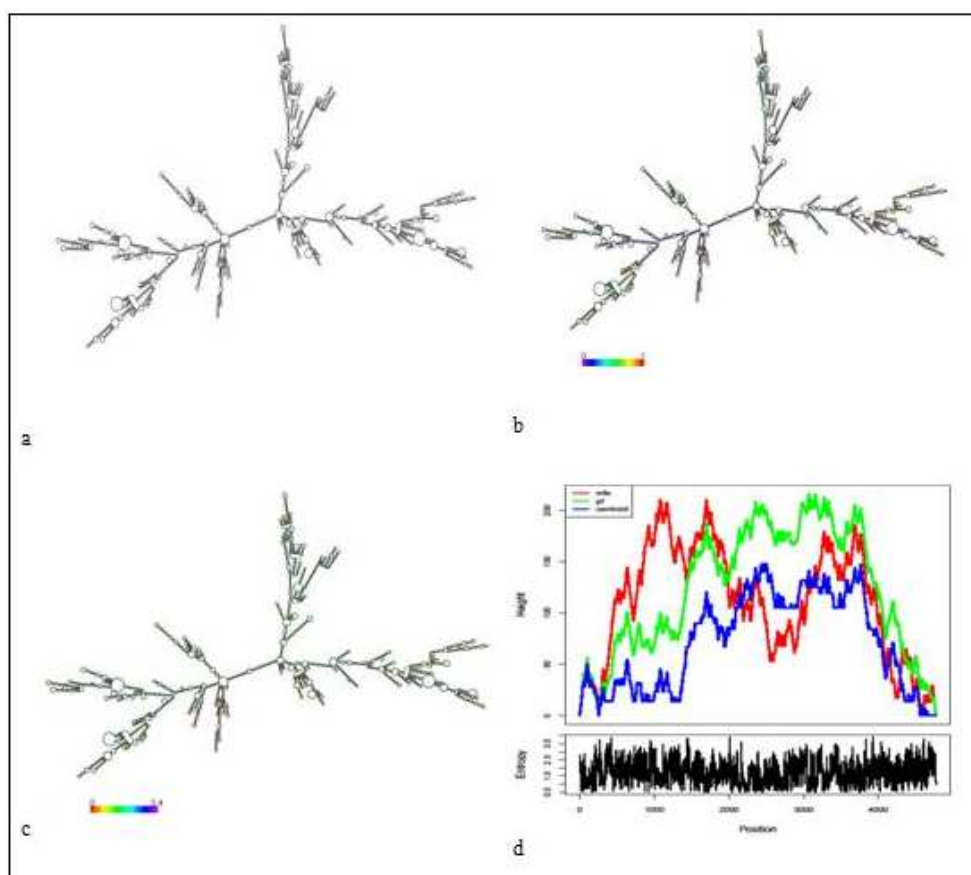


Fig. 1. mRNA secondary structure of RB1 wild type. (a) plain secondary structure (b) secondary structure with base pair probability (c) secondary structure with positional entropy (d) mountain Plot

Table 1. Details of the clinically significant 18 exonic SNPs found in RB1 gene

Given SNP ID	Exonic and base position in RB1	NCBI SNP ID and alleles (wild>mutant)	SNP Position and codon changes in RB1 mRNA	SNP Position and changes in pRB protein	Author/Article described the clinical significance
c.575G>T	Exon 4 4.9E+07	rs121913296 G>T	575 GAA⇒TAA	137 E[Glu]⇒Ter[OCH]	Lohmann <i>et al.</i> (1997)
c.762T>A	Exon 6 4.9E+07	rs121913298 T>A	762 TTA⇒TAA	199 L[Leu]⇒Ter[OCH]	Dommering <i>et al.</i> (2012)
c.1125C>T	Exon 10 4.9E+07	rs121913300 C>T	1125 CGA⇒TGA	320 R[Arg]⇒Ter[OPA]	Dommering <i>et al.</i> (2012)
c.1238C>T	Exon 11 4.9E+07	rs121913301 C>T	1238 CGA⇒TGA	358 R[Arg]⇒Ter[OPA]	Yandell <i>et al.</i> (1989)
c.1499C>T	Exon 14 4.9E+07	rs3092891 C>T	1499 CGA⇒TGA	445 R[Arg]⇒Ter[OPA]	Yandell <i>et al.</i> (1989)
c.1529C>T	Exon 14 4.9E+07	rs121913302 C>T	1529 CGA⇒TGA	455 R[Arg]⇒Ter[OPA]	dbSNP
c.1820C>T	Exon 17 4.9E+07	rs121913303 C>T	1820 CGA⇒TGA	552 R[Arg]⇒Ter[OPA]	Dommering <i>et al.</i> (2012)
c.1832C>T	Exon 17 4.9E+07	rs121913304 C>T	1832 CGA⇒TGA	556 R[Arg]⇒Ter[OPA]	Onadim <i>et al.</i> (1997)
c.1901C>T	Exon 18 4.9E+07	rs121913305 C>T	1901 CGA⇒TGA	579 R[Arg]⇒Ter[OPA]	Dommering <i>et al.</i> (2012)
c.1866C>T	Exon 18 4.9E+07	rs137853292 C>T	1866 TCA⇒TTA	567 S[Ser]⇒L [Leu]	Yandell <i>et al.</i> (1989)
c.1984T>A	Exon 19 4.9E+07	rs137853297 T>A	1984 TAT⇒TAA	606 Y[Tyr]⇒Ter[OCH]	de Jong <i>et al.</i> (2006)
c.2109C>T	Exon 19 4.9E+07	rs121908692 C>T	2109 TCA⇒T	648 S [Ser] ⇒Y [Tyr]	Lohmann and Gallie (2004)
c.2147C>T	Exon 20 4.9E+07	rs137853294 C>T	2147 CGG⇒TGG	661 R[Arg]⇒W[Trp]	Yandell <i>et al.</i> (1989)
c.2189G>T	Exon 20 4.9E+07	rs137853295 G>T	2189 GAA⇒TAA	675 E[Glu]⇒Ter[OCH]	Yandell <i>et al.</i> (1989)
c.2283G>T	Exon 21 4.9E+07	rs121913295 G>T	2283 TGT⇒TTT	706 C[Cys]⇒F [Phe]	Kaye <i>et al.</i> (1990)
c.2300T>C	Exon 21 4.9E+07	rs137853296 T>C	2300 TGC⇒CGC	712 C[Cys]⇒R[Arg]	Otterson <i>et al.</i> (1999)
c.2408G>T	Exon 22 4.9E+07	rs121913297 G>T	2408 GAG⇒TAG	748 E[Glu]⇒Ter[AMB]	Yandell <i>et al.</i> (1989)
c.2525C>T	Exon 23 4.9E+07	rs137853293 C>T	2525 CGA⇒TGA	787 R[Arg]⇒Ter[OPA]	Yandell <i>et al.</i> (1989)

For single SNP genotypes, we found MFE values of optimal secondary structures and centroid secondary structures (Table 2), while ensembles data were obtained as thermodynamic free energy and ensemble diversity. The wild type structure contain the major alleles of all SNPs has the MEF -1154.64 (kcal/mol) and -827.99(kcal/mol) for optimal and centroid secondary structure, respectively. Interestingly the students *t*-test (Table 3) showed that, the mean value of the secondary structures of SNPs examined with the minor alleles in the mRNA structure has the mean value -1154 (kcal/mol) showing no significant ($p = 0.40$) difference with the wild type. In the contrary, centroid secondary structures mean value -783.37 (kcal/mol) was highly significant ($p = 0.004$) to wild type. The results were difficult to interpret because we found larger standard deviation (64.35 kcal/mol) and larger standard error (15.17 kcal/mol) of mean during centroid secondary structure analysis, which indicate the significance of this structural

variation might be false positive. Irrespective of this drawback, the strong level of significance of the centroid secondary structure leads us to proceed with haplotype secondary structure analysis. Additionally, no significant variation of mRNA secondary structure due to mutations has been found for thermodynamic ensemble and ensemble diversity of mRNAs.

The multiple minor alleles of different SNPs constructed the haplotype effects in nascent mRNAs. In total 10 two-SNPs, 10 three-SNPs, 5 four-SNPs and 1 five-SNPs haplotypes has been analyzed similarly as single SNPs and the data presented in Table 4. Five-SNPs haplotype was complete mutant and had distinct values different than wild type and was not considered under statistical analysis. All of other haplotypes containing minor alleles of SNPs were compared with the wild type structure which contains major alleles for all SNPs, therefore, haplotypes as well.

Table 2. Thermodynamic information of the secondary structure of mRNAs constructed by 18 exonic SNPs of RB1 gene

RB1 SNP genotypes (minor alleles represent mutants)	MFE of optimal secondary structure (kcal/mol)	MFE of centroid secondary structure (kcal/mol)	Free energy of the thermodynamic ensemble	Ensemble Diversity of mRNAs
Wild type	-1154.6	-827.99	-1244.0	1370.30
c.575G>T	-1153.0	-827.14	-1242.0	1375.07
c.762T>A	-1156.0	-826.89	-1244.0	1359.64
c.1125C>T	-1154.6	-824.22	-1244.1	1375.66
c.1238C>T	-1154.6	-756.63	-1244.0	1475.85
c.1499C>T	-1153.0	-712.94	-1242.4	1482.10
c.1529C>T	-1155.1	-869.42	-1243.8	1318.78
c.1820C>T	-1153.8	-699.29	-1242.7	1483.01
c.1832C>T	-1153.8	-699.29	-1242.7	1483.01
c.1901C>T	-1156.4	-683.77	-1245.3	1473.74
c.1866C>T	-1154.0	-792.83	-1243.5	1390.67
c.1984T>A	-1155.4	-842.81	-1245.0	1309.07
c.2109C>T	-1154.0	-837.57	-1243.1	1292.25
c.2147C>T	-1155.8	-829.39	-1245.4	1376.05
c.2189G>T	-1153.1	-816.31	-1243.5	1272.07
c.2283G>T	-1150.0	-790.71	-1241.2	1336.22
c.2300T>C	-1156.0	-769.73	-1244.6	1462.81
c.2408G>T	-1155.8	-669.36	-1243.6	1489.54
c.2525C>T	-1156.9	-852.45	-1244.5	1317.53

Table 3. Single SNP analysis results considering four different assessment criteria

Assessment criteria of 18 single SNPs	Mean±SE	SD	t-Stat	pvalue
MFE of optimal secondary structure (kcal/mol)	-1154.54±0.39	1.66	0.253	0.402
MFE of centroid secondary structure (kcal/mol)	-783.38±15.17	64.35	2.942	0.005***
Free energy of the thermodynamic ensemble (kcal/mol)	-1243.63±0.27	1.15	1.469	0.080*
Ensemble diversity	-1392.95±18.08	76.69	1.253	0.115

The *t-test* has been performed with the wild type values mentioned in table 2. Level of significance has been determined as * $p \leq 0.10$ = slightly significant, ** $p \leq 0.05$ = significant, *** $p \leq 0.01$ = highly significant

Table 4. Thermodynamic information of the secondary structure of mRNAs constructed by haplotypes of 5 nonsense SNPs of RB1 gene

RB1 Haplotypes	Haplotype structure of the mRNA	MFE of optimal secondary structure (kcal/mol)	MFE of centroid secondary structure (kcal/mol)	Free energy of the thermodynamic ensemble (kcal/mol)	Ensemble diversity
Wild type	C_C_C_G_T	-1154.60	-827.99	-12440.00	1370.30
SNPs 1_2	T_T	-1152.50	-824.27	-1242.60	1304.52
SNPs 1_3	T_T	-1155.20	-791.84	-1244.90	1396.63
SNPs 1_4	T_T	-1149.50	-736.86	-1240.60	1314.93
SNPs 1_5	T_C	-1155.40	-735.40	-1244.00	1460.28
SNPs 2_3	T_T	-1155.24	-838.77	1244.48	1296.78
SNPs 2_4	T_T	-1148.80	-802.58	-1240.40	1299.94
SNPs 2_5	T_C	-1155.40	-840.89	-1243.50	1383.19
SNPs 3_4	T_T	-1151.20	-790.41	-1242.50	1339.85
SNPs 3_5	T_C	-1157.20	-762.03	-1245.90	1467.58
SNPs 4_5	T_C	-1150.60	-772.57	-1241.70	1348.66
SNPs 1_2_3	T_T_T	-1153.70	-825.47	-1243.90	1309.30
SNPs 1_2_4	T_T_T	-1148.30	-743.92	-1239.80	1285.99
SNPs 1_2_5	T_T_C	-1153.30	-804.50	-1243.00	1403.43
SNPs 1_3_4	T_T_T	-1150.70	-739.76	-1242.00	1318.38
SNPs 1_3_5	T_T_C	-1156.60	-729.65	-1245.40	1463.62
SNPs 1_4_5	T_T_C	-1150.10	-721.82	-1241.10	1316.28
SNPs 2_3_4	T_T_T	-1150.00	-803.78	-1241.80	1303.65
SNPs 2_3_5	T_T_C	-1156.60	-843.39	-1244.90	1388.25
SNPs 2_4_5	T_T_C	-1150.20	-809.02	-1240.90	1286.39
SNPs 3_4_5	T_T_C	-1151.80	-775.97	-1243.10	1352.77
SNPs 1_2_3_4	T_T_T_T	-1149.50	-745.12	-1241.20	1289.69
SNPs 1_2_3_5	T_T_T_C	-1154.50	-802.70	-1244.40	1408.71
SNPs 1_2_4_5	T_T_T_C	-1149.70	-751.56	-1240.40	1269.46
SNPs 1_3_4_5	T_T_T_C	-1151.30	-733.55	-1242.50	1319.95
SNPs 2_3_4_5	T_T_T_C	-1151.40	-810.22	-1242.30	1290.08
SNPs 1_2_3_4_5	T_T_T_T_C	-1150.94	-752.76	-1241.76	1273.10

SNP 1, 2, 3, 4, 5 was coded to c.1866C>T (rs137853292), c.2109C>T (rs121908692), c.2147C>T (rs137853294), c.2283G>T (rs121913295), c.2300T>C (rs137853296), respectively

Table 5. Haplotype analysis results considering four different assessment criteria

Assessment criteria of haplotypes SNPs	Mean±SE	SD	t-Stat	P value
Two-SNPs haplotypes				
MFE of optimal secondary structure (kcal/mol)	-1153.13±0.93	2.94	1.625	0.069*
MFE of centroid secondary structure (kcal/mol)	-789.56±12.11	38.29	3.174	0.007***
Free energy of the thermodynamic ensemble (kcal/mol)	-1243.06±0.58	1.83	1.676	0.064*
Ensemble diversity	1361.25±20.20	68.89	-0.448	0.332
Three-SNPs haplotypes				
MFE of optimal secondary structure (kcal/mol)	-1152.16±0.89	2.84	2.759	0.011**
MFE of centroid secondary structure (kcal/mol)	-779.73±13.72	43.39	3.517	0.003***
Free energy of the thermodynamic ensemble (kcal/mol)	-1242.59±0.57	1.79	2.535	0.016**
Ensemble diversity	1370.30±18.50	58.50	-1.486	0.086*
Four-SNPs haplotypes				
MFE of optimal secondary structure (kcal/mol)	-1151.31±0.89	1.99	3.745	0.010***
MFE of centroid secondary structure (kcal/mol)	-768.73±15.76	35.23	3.767	0.009***
Free energy of the thermodynamic ensemble (kcal/mol)	-1242.15±0.67	1.51	2.784	0.025**
Ensemble diversity	1370.30±24.64	55.09	-2.221	0.045**

The *t*-test has been performed with the wild type values mentioned in Table 2. Level of significance has been determined as * $p \leq 0.10$ = slightly significant, ** $p \leq 0.05$ = significant, *** $p \leq 0.01$ = highly significant

Discussion

In this study, our concentration of analysis and observations particularly focused on the structural diversity of RB1 mRNAs caused by SNPs and their haplotypes. Our findings suggest that, genetic mutations can contribute to the structural diversity of mRNAs. It was predicted that the mutations of RB1 gene are clinically significant but the question remain to be addressed how the gene specific SNPs could be causative. The answer could be found at the stage of mRNA synthesis and a subsequent analysis of structural variations in protein structure could prove it (Bompfunewerer *et al.*, 2008). It is mentionable that the mRNAs and protein secondary structures have been found in other scientific studies (Hofacker and Stadler, 2006).

RB1 mRNAs is unstructured and this structural diversity causes significant variation to the stability of the structure. Here, in this study we found energy dot plot of predicted structure which is widely considered as an efficient structure prediction system (Mathews *et al.*, 2004). These entire energy dot plots have been derived from the mRNA sequences having alternative alleles of point mutations extracted from different exons. Our analysis results are in agreement of earlier studies where it was found that native mRNAs exerts lower free energy in comparison to random sequences (Jia and Luo, 2006). For single SNP and haplotype analysis, the mean free energy values of the wild type mRNA structures of RB1 (Table 3), were lower comparing random variation arises from mutational effects, which confirms more stable structure for wild type. Previous studies reported that exonic SNPs affect mRNA translation (Jia *et al.*, 2004; Zuker and Jacobson, 1995; Seffens and Digby, 1999) and thereby supports our idea to conduct the study on

synonymous, no synonymous and one frame shift mutation for assessment for mRNA structural effects. Moreover, functional role of these SNPs could be the topic of future research while we create foundation for future studies through bioinformatics analysis of free energy structures.

For haplotype study we have excluded the SNPs causes nonsense mediated decay because in mRNA decay rates has proven link with mutations and translation is affected in the downstream (Ikemura, 1985). We think that it might not be worthy to accommodate NMD SNPs to study haplotypes, instead to keep the scope for protein structure analysis in further research it is very likely to find out mRNA secondary structures using non-NMD SNPs. However, it is mentionable that, these NMD SNPs might be highly causative (Sorensen *et al.*, 1989; Sharp *et al.*, 1995; Duan and Antezana, 2003) and may have implications to create retinoblastoma.

A transcriptome-wide *in silico* analysis of mRNA folding (Pickering and Willis, 2005) signifies the involvement mutations and genetic codes to control the stability and periodicity of mRNA secondary structures in human. Single gene analysis for secondary structures using clinically significant SNPs could be more worthy to find the causative one. Our study could be a good example which includes not only the single SNPs but also the haplotypes of the SNPs for RB1 gene. Findings of our study is limited to the prediction using bioinformatics tools, however, we can explore the protein structures for mutations in further analysis which could prove the value of our results. Conclusively, our study provides footsteps to explore the genetic basis of retinoblastoma and could give confidence to the researchers conducting association analysis of SNPs with retinoblastoma phenotype.

Conclusion

The study gives us comprehensible portrait of the consequence of SNPs on mRNA structure and corresponding free energy changes. Here, structural deterioration and statistical analysis of the data opens a potential explanation on the stability of the mRNA molecules, which plays major role in retinoblastoma formation and initiation. The identified crucial SNPs on the coding region can be further incorporated for proteomic level research. So it requires further expedition to figure out the dynamics as well as to defeat the current limitations to comprehend the biological processes.

Acknowledgement

The authors would like to thank Biotechnology and Genetic Engineering Discipline, Khulna University, Bangladesh for logistic support.

Funding Information

There is no recognized funding institute/organization supporting this research.

Author's Contributions

Sarder Nasir Uddin: Experimental design, Data analysis, Manuscript preparation and Coordinate the study.

Apurba Majumder: Experimental work, Data analysis, Manuscript draft preparation.

Khandker Khaldun Islam: Experimental design, Manuscript writing, Data analysis, Supervising 2nd author (AM).

Sk. Amir Hossain: SNPs analysis.

Palash Kumar Sarker: SNPs analysis.

Ethics

There is no an ethical issue or conflict of interest.

References

- Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D. Lipman, 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215: 403-410. DOI: 10.1016/S0022-2836(05)80360-2
- Bartoszewski, R.A., M. Jablonsky, S. Bartoszevska, L. Stevenson and Q. Dai *et al.*, 2010. A synonymous single nucleotide polymorphism in $\Delta F508$ CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J. Biol. Chem.*, 285: 28741-28748. DOI: 10.1074/jbc.M110.154575
- Bompmfunewerer, A.F., R. Backofen, S.H. Bernhart, J. Hertel and I.L. Hofacker *et al.*, 2008. Variations on RNA folding and alignment: Lessons from benasque. *J. Math. Biol.*, 56: 129-144. DOI: 10.1007/s00285-007-0107-5
- Burkhardt, D.L. and J. Sage, 2008. Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nat. Rev. Cancer*, 8: 671-682. DOI: 10.1038/nrc2399
- Capon, F., M.H. Allen, M. Ameen, A.D. Burden and D. Tillman *et al.*, 2004. A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups. *Hum. Mol. Genet.*, 13: 2361-2368. DOI: 10.1093/hmg/ddh273
- de Jong, P.T.V.M., C.M. Mooy, G. Stoter, W.M.H. Eijkenboom and G.P.M. Luyten, 2006. Late-onset retinoblastoma in a well-functioning fellow eye. *Ophthalmology*, 113: 1040-1044. PMID: 16631255
- Dommering, C.J., T. Marees, A.H. van der Hout, S.M. Imhof and H. Meijers-Heijboer *et al.*, 2012. RB1 mutations and second primary malignancies after hereditary retinoblastoma. *Familial Cancer*, 11: 225-233. PMID: 22205104
- Duan, J. and M.A. Antezana, 2003. Mammalian mutation pressure, synonymous codon choice and mRNA degradation. *J. Mol. Evolution*, 57: 694-701. DOI: 10.1007/s00239-003-2519-1
- Duan, J., M.S. Wainwright, J.M. Comeron, N. Saitou and A.R. Sanders *et al.*, 2003. Synonymous mutations in the human Dopamine Receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum. Mol. Genet.*, 12: 205-216. DOI: 10.1093/hmg/ddg055
- Friend, S.H., R. Bernards, S. Rogelj, R.A. Weinberg and J.M. Rapaport *et al.*, 1986. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature*, 323: 643-646. DOI: 10.1038/323643a0
- Gregory, B.D., 2015. Global analysis of RNA-protein interactions and RNA secondary structure in eukaryotes. *Proceedings of the Plant and Animal Genome 23rd Conference*, Jan. 10-14, San Diego, CA, Philadelphia, PA.
- Hofacker, I.L. and P.F. Stadler, 2006. Memory efficient folding algorithms for circular RNA secondary structures. *Bioinformatics*, 22: 1172-1176. DOI: 10.1093/bioinformatics/btl023
- Hofacker, I.L., W. Fontana, P.F. Stadler, S. Bonhoeffer and M. Tacker *et al.*, 1994. Fast folding and comparison of RNA secondary structures. *Chem. Monthly*, 125: 167-188. DOI: 10.1007/BF00818163
- Ikemura, T., 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evolution*, 2: 13-34. PMID: 3916708
- Jia, M. and L. Luo, 2006. The relation between mRNA folding and protein structure. *Biochem. Biophys. Res. Commun.*, 343: 177-182. DOI: 10.1016/j.bbrc.2006.02.135
- Jia, M., L. Luo and C. Liu, 2004. Statistical correlation between protein secondary structure and messenger RNA stem-loop structure. *Biopolymers*, 73: 16-26. DOI: 10.1002/bip.10496

- Johnson, A.D., H. Trumbower and W. Sadee, 2011. RNA structures affected by single nucleotide polymorphisms in transcribed regions of the human genome. DOI: 10.9754/journal.wmc.2011.001600
- Kaye, F.J., R.A. Kratzek, J.L. Gerster and J.M. Horowitz, 1990. A single amino acid substitution results in a retinoblastoma protein defective in phosphorylation and oncoprotein binding. Proc. Natl. Acad. Sci., 87: 6922-6926. PMID: 2168563
- Kudla, G., A.W. Murray, D. Tollervey and J.B. Plotkin, 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. Science, 324: 255-258. DOI: 10.1126/science.1170160
- Laederach, A., 2007. Informatics challenges in structured RNA. Brief Bioinform., 8, 294-303. DOI: 10.1093/bib/bbm026
- Leiderman, Y.I., S. Kiss and S. Mukai, 2007. Molecular genetics of RB1-the retinoblastoma gene. Seminars Ophthalmol., 22: 247-254. DOI: 10.1080/08820530701745165
- Lohmann, D.R. and B.L. Gallie, 2004. Retinoblastoma: Revisiting the model prototype of inherited cancer. Am. J. Med. Genet C Semin. Med. Genet, 129: 23-28. PMID: 15264269
- Lohmann, D.R., M. Gerick, B. Brandt, U. Oelschläger and B. Lorenz *et al.*, 1997. Constitutional RB1-gene mutations in patients with isolated unilateral retinoblastoma. Am. J. Hum. Genet, 61: 282-294. PMID: 9311732
- Lorenz, R., S.H. Bernhart, C.H. zu Siederdisen, H. Tafer and C. Flamm *et al.*, 2011. ViennaRNA Package 2.0. Algorithms Mol. Biol., 6: 26-26. DOI: 10.1186/1748-7188-6-26
- Mahoney, M.C., W.S. Burnett, A. Majerovics and H. Tanenbaum, 1990. The epidemiology of ophthalmic malignancies in New York State. Ophthalmology, 97: 1143-1147. DOI: 10.1016/S0161-6420(90)32445-4
- Mathews, D.H., J. Sabina, M. Zuker and D.H. Turner, 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol., 288: 911-940. DOI: 10.1006/jmbi.1999.2700
- Mathews, D.H., M.D. Disney, D. Matthew, J.L. Childs and S.J. Schroeder *et al.*, 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. National Acad. Sci., 101: 7287-7292. DOI: 10.1073/pnas.0401799101
- McCaskill, J.S., 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structures. Biopolymers, 29: 1105-1119. DOI: 10.1002/bip.360290621
- Mita, T. and T. Kuroiwa, 1989. Division of plastids by a plastid-dividing ring in *Cyanidium caldarium*. Protoplasma, 1: 133-152. DOI: 10.1007/978-3-7091-9008-1_16
- Morton, N.E., 2008. Into the Post-HapMap Era. Adv. Genet., 60: 727-742. DOI: 10.1016/S0065-2660(07)00425-7
- Nackley, A.G., S.A. Shabalina, I.E. Tchivileva, K. Satterfield and O. Korchynskyi *et al.*, 2006. Human catechol-o-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. Science, 314: 1930-1933. DOI: 10.1126/science.1131262
- Onadim, Z., A.J. Woolford, J.E. Kingston and J.L. Hungerford, 1997. The RB1 gene mutation in a child with ectopic intracranial retinoblastoma. Brit. J. Cancer, 76: 1405-1409. PMID: 9400934
- Otterson, G.A., S. Modi, K. Nguyen, A.B. Coxon and F.J. Kaye, 1999. Temperature-sensitive RB mutations linked to incomplete penetrance of familial retinoblastoma in 12 families. Am. J. Hum. Genet, 65: 1040-1046. PMID: 10486322
- Pickering, B.M. and A.E. Willis, 2005. The implications of structured 5' untranslated regions on translation and disease. Seminars Cell Develop. Biol., 16: 39-47. DOI: 10.1016/j.semcdb.2004.11.006
- Salari, R., C. Kimchi-Sarfaty, M.M. Gottesman and T.M. Przytycka, 2013. Sensitive measurement of single-nucleotide polymorphism-induced changes of RNA conformation: Application to disease studies. Nucleic Acids Res., 47: 44-53. DOI: 10.1093/nar/gks1009
- Schmittgen, T., J.M. Weaver, R.A. Badalament, M.G. Wientjes and E.A. Klein *et al.*, 1994. Correlation of human bladder tumor histoculture proliferation and sensitivity to mitomycin C with tumor pathobiology. J. Urol., 152: 1632-1636. PMID: 7933219
- Seffens, W. and D. Digby, 1999. mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. Nucleic Acids Res., 27: 1578-1584. DOI: 10.1093/nar/27.7.1578
- Shabalina, S.A., A.Y. Ogurtsov and N.A. Spiridonov, 2006. A periodic pattern of mRNA secondary structure created by the genetic code. Nucleic Acids Res., 34: 2428-2437. DOI: Spiridonov 10.1093/nar/gkl287
- Shabalina, S.A., A.Y. Ogurtsov, N.A. Spiridonov and E.V. Koonin, 2014. Evolution at protein ends: Major contribution of alternative transcription initiation and termination to the transcriptome and proteome diversity in mammals. Nucleic Acids Res., 42: 7132-7144. DOI: 10.1093/nar/gku342
- Shabalina, S.A., N.A. Spiridonov and A. Kashina, 2013. Sounds of silence: Synonymous nucleotides as a key to biological regulation and complexity. Nucleic Acids Res., 41: 2073-2094. DOI: 10.1093/nar/gks1205
- Shalev, U., J.W. Grimm and Y. Shaham, 2002. Neurobiology of relapse to heroin and cocaine seeking: A review. Pharmacol. Rev., 54: 1-42. PMID: 11870259

- Sharp, P.M., M. Averof, A.T. Lloyd, G. Matassi and J.F. Peden, 1995. DNA sequence evolution: The sounds of silence. *Philosophical Trans. Royal Society B: Biol. Sci.*, 349: 241-247.
DOI: 10.1098/rstb.1995.0108
- Sorensen, M.A., C.G. Kurland and S. Pedersen, 1989. Codon usage determines translation rate in *Escherichia coli*. *J. Mol. Biol.*, 207: 365-377.
DOI: 10.1016/0022-2836(89)90260-X
- Turner, D.H. and D.H. Mathews, 2009. NNDB: The nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.*, 38: 280-282. DOI: 10.1093/nar/gkp892
- Weinberg, R.A., 1995. The retinoblastoma protein and cell cycle control. *Cell*, 81: 323-330.
DOI: 10.1016/0092-8674(95)90385-2
- www.ncbi.nlm.nih.gov/snp/?term=Retinoblastoma
- Yandell, D.W., T.A. Campbell, S.H. Dayton, R. Petersen and D. Walton *et al.*, 1989. Oncogenic point mutations in the human retinoblastoma gene: Their application to genetic counseling. *New England J. Med.*, 321: 1689-1695.
DOI: 10.1056/NEJM198912213212501
- Zuker, M. and P. Stiegler, 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, 9: 133-148. DOI: 10.1093/nar/9.1.133
- Zuker, M. and A.B. Jacobson, 1995. 'Well-determined' regions in RNA secondary structure prediction: Analysis of small subunit ribosomal RNA. *Nucleic Acids Res.*, 23: 2791-2798.
DOI: 10.1093/nar/23.14.2791