BIOINFORMATICS ORIGINAL PAPER

Vol. 30 no. 3 2014, pages 343–352
doi:10.1093/bioinformatics/btt695

Structural bioinformatics

Advance Access publication November 30, 2013

Accessibility of microRNA binding sites in metastable RNA secondary structures in the presence of SNPs

Luke Day[1], Ouala Abdelhadi Ep Souki[1], Andreas A. Albrecht[2] and Kathleen Steinhöfel[1,*]

[1]Department of Informatics, King's College London, London WC2R 2LS and [2]Middlesex University London, School of Science and Technology, London NW4 4BT, UK

Associate Editor: Ivo Hofacker

ABSTRACT

Motivation: We study microRNA (miRNA) bindings to metastable RNA secondary structures close to minimum free energy conformations in the context of single nucleotide polymorphisms (SNPs) and messenger RNA (mRNA) concentration levels, i.e. whether features of miRNA bindings to metastable conformations could provide additional information supporting the differences in expression levels of the two sequences defined by a SNP. In our study, the instances [mRNA/3′UTR; SNP; miRNA] were selected based on strong expression level analyses, SNP locations within binding regions and the computationally feasible identification of metastable conformations.

Results: We identified 14 basic cases [mRNA; SNP; miRNA] of 3′ UTR-lengths ranging from 124 up to 1078 nt reported in recent literature, and we analyzed the number, structure and miRNA binding to metastable conformations within an energy offset δE above mfe conformations. For each of the 14 instances, the miRNA binding characteristics are determined by the corresponding `STarMir` output. Among the different parameters we introduced and analyzed, we found that three of them, related to the average depth and average opening energy of metastable conformations, may provide supporting information for a stronger separation between miRNA bindings to the two alleles defined by a given SNP.

Availability and implementation: At http://kks.inf.kcl.ac.uk/MSbind.html the `MSbind` tool is available for calculating features of metastable conformations determined by putative miRNA binding sites.

Contact: kathleen.steinhofel@kcl.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on June 12, 2013; revised on November 21, 2013; accepted on November 25, 2013

1 INTRODUCTION

Within the past few years, analyzing concentration levels of microRNAs (miRNAs) and their putative messenger RNA (mRNA) targets has become a major topic in miRNA research. Subkhankulova *et al*. (2008) experimentally evaluated a parameterized analytical expression that estimates the number of genes g having t transcripts present in a single cell. The parameters are adjusted based on microarray data for a large number of genes extracted from single embryonic mouse neural stem cells, where the actual aim is the comparison of transcript numbers in phenotypically identical cells. The authors conclude from observed data for ~13 000 genes that the typical number of gene copies lies between 5–20, with 85% of genes having <100 copies in a single cell. Although the analysis is carried out for a specific cell type, the authors expect similar distribution results for a wide range of cell types. Arvey *et al*. (2010) provide experimental evidence that short RNAs (miRNAs and siRNAs) having a higher number of target transcripts within a single cell will downregulate each individual target gene to a lesser extent than those with a lower number of targets, which implies that the competition between target genes for a limited number of small RNAs may determine the degree of downregulation; see also Salmena *et al*. (2011) for the concept of competing mRNAs associated with the number and distribution of multiple binding sites. Saito and Sætrom (2012) provide data supporting the assumption that endogenous miRNAs preferentially target mRNAs with long 3′UTRs. The authors also discuss—among other features—the interaction between exogenous and endogenous miRNAs and critically assess the value of microarray data in the context of miRNA target prediction. The sequencing method developed by Mullokandov *et al*. (2012) supports the assertion that only the most abundant miRNAs mediate target suppression. For example, deep sequencing of monocyte cells revealed the presence of ~310 miRNAs, with only ~40% of the miRNAs showing suppressing activity. For >80% of the targets, the corresponding miRNA was expressed >100 reads/million. For the miRNA target prediction tool `TargetScan`, Garcia *et al*. (2011) demonstrate how predictions may improve if target abundance is accounted for in binding scores. The impact of the life cycle of mRNAs on siRNA and miRNA efficacy is studied in Larsson *et al*. (2010). The authors draw the conclusion that miRNA target prediction could be improved if data about mRNA turnover rates are incorporated into prediction tools. Although Arvey *et al*. (2010) and Larsson *et al*. (2010) focus on mRNA concentration levels, Cuccato *et al*. (2011) propose different parameterized models of RNA interference that describe the effects of varying quantities of siRNAs. The models are derived from the basic equation $\mathrm{d}X_m/\mathrm{d}t = k_m - d_m X_m - \delta(X_m, X_s)$, where X_m and X_s are the mRNA and siRNA concentrations, k_m is the mRNA transcription rate, d_m the basal mRNA degradation rate and $\delta(X_m, X_s)$ is the siRNA-induced mRNA degradation rate. The models differ in the assumption about $\delta(X_m, X_s)$ and are fitted to experimental data obtained for a single siRNA targeting the coding region of the EGFP mRNA. The authors obtain the best fit for $\delta(X_m, X_s) = pX_m X_s^h/(q^h + X_s^h)$ with $h \sim 4.5$, $P \sim 0.008$/min and $q \sim 0.1$ pmol. Within a similar

*To whom correspondence should be addressed.

framework, Osella *et al*. (2011) study the so-called miRNA-mediated feedforward loop in which a master transcription factor regulates an miRNA together with a set of target genes, and the mathematical models studied by Loinger *et al*. (2012) additionally account for the concentration level of the argonaute protein complex. Baker *et al*. (2012) study analytically the impact of multiple small non-coding RNAs on the regulation of a single target mRNA and subsequently the dynamics of protein production. Ragan *et al*. (2011) combine the concept of miRNA binding site accessibility with miRNA and mRNA concentration levels. For $[S]$, $[T]$ and $[ST]$ denoting the equilibrium (final) concentrations of the miRNA, target mRNA and of the hybridized structure, respectively, the authors use the equilibrium condition $[ST]/([S][T]) = exp(-\Delta\Delta G/c)$, where $\Delta\Delta G$ is the energy score that accounts for making the binding site accessible and the free energy of the hybridized structure (the constant c stands for the product of the gas constant and the temperature). Combining the equilibrium condition with conservation of mass equations (for initial concentrations $[S_0]$ and $[T_0]$) eventually leads to an analytical expression for $[S]$ in terms of $[S_0], [T_0]$ and $\Delta\Delta G$, where the latter is calculated for a particular binding site. Marin and Vaniõek (2012) introduce a new accessibility-based algorithm that uses a statistical analysis of all putative binding sites for a given miRNA-3′UTR pair. Among the top 100 target predictions for 153 fruit fly miRNAs, the algorithm finds more than twice as many validated targets compared with other accessibility-based target prediction methods. Reviews of existing miRNA target prediction tools and information about latest developments can be found in Muniategui *et al*. (2012) and Shirdel *et al*. (2011). The target prediction tool `CoMeTa` designed by Gennarino *et al*. (2012) operates on the assumption that targets of a given miRNA are coexpressed with each other. The target prediction score is based on the evaluation of thousands of publicly available microarray data. For the 675 human miRNAs analyzed in the study, >90% of the validated targets fall within the first 50% of predicted targets (which, however, could be a large number). In a similar way, the tool `miRror` designed by Balaga *et al*. (2012) combines scores produced by an ensemble of established miRNA target prediction tools with rankings obtained from gene expression and HITS-CLIP data. Johnson and Srivastava (2013) consider the problem of selecting an antisense sequence that is able to effectively bind to a target mRNA and block protein synthesis. One of the key features of the authors' method is the presuppositions that mRNA secondary structures are in a constant state of flux and are assuming different suboptimal states, which also determines the approach taken in the present article. Johnson and Srivastava (2013) designed a tool that generates and compares suboptimal secondary structures of a given mRNA sequence. The comparison aims at identifying regions that are least similar among the set of folded structures, which indicates volatility in intramolecular hydrogen bonding. Such regions are seen as candidates for antisense binding. The method is evaluated on six mRNA sequences and compared with results produced by the `Soligo` application of `Sfold` (Ding *et al*., 2005).

In our article, we study miRNA bindings to sets of metastable secondary structures induced by 3′UTRs and their mutated counterparts, where the single nucleotide polymorphisms (SNPs) are located within the miRNA binding site. Long *et al*. (2007) briefly mention the analysis of miRNA bindings to metastable conformations. The authors compare the prediction performance over samples of 1000 secondary structures without constraints to samples of 100 metastable conformations. They report the sample of 1000 conformations is more favourable. In our approach, we discriminate between metastable conformations with regard to depth and average opening energies.

The impact of SNPs on minimum free energy mRNA conformations has been comprehensively studied in Johnson *et al*., (2011). The authors analyzed a total number of 34 557 SNPs in 12 450 genes. The minimum free energy conformations were calculated by using `RNAfold` (Gruber *et al*., 2008). The authors provide a great variety of data about the distribution of SNPs within mRNA transcripts and their effect on minimum free energy values of secondary structures as well as on the profile of the ensemble of suboptimal structures and structures with high Boltzmann probabilities, see also Johnson and Srivastava (2013), Martin *et al*. (2012) and Solomatin *et al*. (2010) for the impact of SNPs on the ensemble of secondary structures. The study (Johnson *et al*., 2011) also includes the analysis of various subsets of transcripts and SNPs, e.g. with respect to the transcript length and different SNP types.

Martin *et al*. (2012) study structural changes induced by SNPs in the 5′UTR of the human FTL gene in conjunction with associated SNPs that restore the overall wild-type ensemble of secondary structures, thus leading to the notion of structure-stabilizing haplotypes. The authors also analyzed the stabilizing effect of multiple structure-stabilizing haplotypes on binding sites of miRNAs and RNA binding proteins (nine cases of 3′UTRs and one case of 5′UTR). As pointed out by the authors, the findings suggest that certain SNP pairs are conserved in the human population because they stabilize ensembles of mRNA conformations.

Recently, databases have become available that report SNP data in the context of miRNA bindings, for example, `miRdSNP` (Bruno *et al*., 2012), `PolymiRTS` (Ziebarth *et al*., 2012), `MicroSNiPer` (Barenboim *et al*., 2010) and the related tool `RNAsnp` (Sabarinathan *et al*., 2013). Sethupathy and Collins (2008) critically assess several genetic association studies related miRNA bindings and the potential impact of SNPs in bindings regions. The authors highlight the importance of follow-up functional experiments for a deeper understanding of the real effect miRNA target site variations may have on the development of various human diseases. Consequently, we researched for our analysis the recent literature for studies on the impact of SNPs on miRNA bindings meeting the following four conditions: (i) the expression levels of both alleles involved are analyzed experimentally by SNP genotyping (for some instances, a combination of clinical association studies and strong *in silico* results) or related methods involving PCR experiments, (ii) the underlying allele information can be extracted as consistent data from the NCBI database, the dbSNP database and the Ensembl database (3′UTR transcripts), (iii) the miRNA bindings are predicted by the latest version of `STarMir` (Hammell *et al*., 2008) at least for the allele with the stronger inhibitory effect with the SNP position being inside the binding region and (iv) identifying all metastable conformations within an energy offset δE above mfe conformations is computationally feasible (3′UTR length $\leq$ 1000 nt). Meeting all four conditions at least to a certain extent

reduced the number of case studies we found to a small set of 14 RNA instances.

2 APPROACH

The present study assumes the existence of multiple active RNA conformations instead of a unique minimum free energy conformation as the single biologically active state, see Johnson and Srivastava (2013), Solomatin *et al.* (2010) and the literature therein. The second basic feature of our approach relates to the presence of multiple copies of each individual mRNA. In more detail, we proceed as follows for a given [mRNA; SNP; miRNA] instance (a flowchart of the approach is provided in Supplementary Fig. S1 iPart I, Section I–A):

(A) For both alleles (3′UTR and SNP), the sets MS(3′UTR, δE) and MS(SNP, δE) of metastable states (MS states) within an energy offset δE above the mfe conformation are identified by using `RNAsubopt` and `Barriers`; cf. (Gruber *et al.*, 2008; Wolfinger *et al.*, 2004; Wuchty *et al.*, 1999). Alternative tools are, e.g. (modified) version of `RNAlocopt` (Lorenz and Clote, 2011) and `Regliss` (Saffarian *et al.*, 2012).

(B) `STarMir` (Hammell *et al.*, 2008) is applied to both alleles for the given miRNA. Although the 14 basic cases [mRNA; SNP; miRNA] selected from recent literature were analyzed in the corresponding publications by prediction tools different from `STarMir`, we obtained for each of the cases at least for one allele a binding site predicted by `STarMir` with the SNP position inside.

(C) For the predicted binding site BS(miRNA), the elements of the sets MS(3′UTR, δE) and MS(SNP, δE) are examined with respect to a number of basic features, such as the number of base pair bindings within BS(miRNA) and the approximate free energy ΔG_{BS} of bindings within BS(miRNA) according to standard data of the nearest neighbour model (Chen *et al.*, 2012; Mathews *et al.*, 2004).

(D) Subsets of MS($\cdots$, δE) are analyzed in the context of the number of mRNA copies, as analyzed in Subkhankulova *et al.* (2008), in the same way as in (C), i.e. for $k = 20$, 60, 100 the sets MS($\cdots$, k) $\subset$ MS($\cdots$, δE) are examined, where k indicates the assumption about the number of mRNA copies. We note that the number of copies is different from concentration levels, which are measured, e.g. *per mol*, see Cuccato *et al.* (2011).

For a given set of metastable secondary structures, the application `MSbind` is available at http://kks.inf.kcl.ac.uk/MSbind.html for calculating features of metastable conformations determined by putative miRNA binding sites.

The data are sourced from the following publications:

[LIG3;rs4796030;miR-221]: The case is analyzed in Teo *et al.* (2012). The SNP rs4796030 is defined by A ↔ C at 3′UTR position 83 of NM_002311.4. The 3′UTR length is 124 nt. The authors studied bladder cancer cases by using clinical association studies combined with *in silico* target predictions. They conjecture a stronger inhibitory effect of miR-221 for the C-allele.

[CBR1;rs9024;miR-574-5p]: The case is analyzed in Kalabus *et al.* (2012). The SNP rs9024 is defined by G ↔ A at 3′UTR position 133 of NM_001757.2. The 3′UTR length is 284 nt. The authors used dual-luciferase assays to evaluate the miRNA binding to both alleles and observed a stronger inhibitory effect of miR-574-5p for the A-allele.

[HTR3E;rs56109847;miR-510-5p]: The case is analyzed in Kapeller *et al.* (2008), see also Sethupathy and Collins (2008). The SNP rs56109847 (rs62625044) is defined by G ↔ A at 3′UTR position 76 of NM_001256614.1. The 3′UTR length is 302 nt. The authors measure luciferase activity to evaluate the miRNA binding to both alleles and observe a stronger inhibitory effect of miR-510-5p for the G-allele, see Figure 1B in Kapeller *et al.* (2008).

[HLA-G;rs1063320;miR-148a-3p]: The case is analyzed in Tan *et al.* (2007). The SNP rs1063320 is defined by C ↔ G at 3′UTR position 233 of NM_002127.5. The 3′UTR length is 386 nt. The study aims at exploring factors affecting the asthma risk. The authors used real-time PCR and luciferase assays for measuring expression levels of miRNAs and C/G-alleles and found evidence for a stronger inhibitory effect of miR-148a-3p for the G-allele.

[PARP1;rs8679;miR-145-5p]: The case is analyzed in Teo *et al.* (2012). The SNP rs8679 is defined by T ↔ C at 3′UTR position 607 of NM_001618.3. The 3′UTR length is 769 nt. The methodology is the same as for [LIG3;rs4796030;miR-221], and the authors presume an additive effect of both instances on bladder cancer risk. For PARP1/rs8679, the authors conjecture a stronger inhibitory effect of miR-145-5p for the T-allele.

[WFS1;rs1046322;miR-668-3p]: The case is analyzed in Kovacs-Nagy *et al.* (2013). The SNP rs1046322 is defined by G ↔ A at 3′UTR position 253 of NM_001145853.1. The 3′UTR length is 779 nt. Expression levels of luciferase assays are measured for both alleles, with a stronger inhibitory effect of miR-668-3p for the G-allele.

[IL-23R;rs10889677;let-7e]: The case is analyzed in Zwiers *et al.* (2012). The SNP rs10889677 is defined by C ↔ A at 3′UTR position 309 of NM_144701.2. The 3′UTR length is 851 nt. The authors study risk factors for inflammatory bowel diseases. Wang *et al.* (2012b) associate [IL-23R;rs10889677] with breast cancer development. Zwiers *et al.* (2012) use real-time PCR and luciferase assays for measuring expression levels, and the authors conclude a stronger inhibitory effect of let-7e for the C-allele.

[RYR3;rs1044129;miR-367]: The case is analyzed in Zhang *et al.* (2011). The SNP rs1044129 is defined by A ↔ G at 3′UTR position 839 of NM_001036.3. The 3′UTR length is 880 nt. The authors study risk factors of breast cancer development. The authors use real-time PCR and luciferase assays for measuring expression levels. The authors observe a stronger inhibitory effect of miR-367 for the A-allele.

[AGTR1;rs5186;miR-155-5p]: The case is analyzed in Haas *et al.* (2012). The SNP rs5186 is defined by A ↔ C at 3′UTR position 86 of NM_032049.3. The 3′UTR length is 888 nt. The authors analyze luciferase assays for measuring expression levels and observe a stronger inhibitory effect of

miR-155-5p for the A-allele, see 'long case' in Figure 2C in Haas *et al*. (2012).

[FGF20;rs12720208;miR-433-3p]: The case is analyzed in Wang *et al*. (2008). The SNP rs12720208 is defined by C ↔ T at 3′UTR position 182 of NM_019851.2. The 3′UTR length is 903 nt. The authors analyze luciferase assays for measuring expression levels and observe a stronger inhibitory effect of miR-433-3p for the C-allele. We note that in this specific case we used the submission ss20399075 instead of the default dbSNP entry ss28476621 for consistency with the NCBI entry of NM_019851.2 and the corresponding 3′UTR transcript entry ENST00000180166 at ENSEMBL.

[HOXB5;rs9299;miR-7-5p]: The case is analyzed in Luo *et al*. (2012). The SNP rs9299 is defined by G ↔ A at 3′UTR position 141 of NM_002147.3. The 3′UTR length is 952 nt. The miRNA-3′UTR binding is studied in the context of bladder cancer development. Both real-time PCR and luciferase reporter assays are applied gene expression measurements. The authors observe a stronger inhibitory effect of miR-7-5p for the A-allele.

[RAD51;rs7180135;miR-197-3p]: The case is analyzed in Teo *et al*. (2012). The SNP rs7180135 is defined by G ↔ A at 3′UTR position 718 of NM_002875.4. The 3′UTR length is 978 nt. The methodology is the same as for [LIG3; rs4796030;miR-221] and [PARP1;rs8679;miR-145-5p]. The authors conjecture a stronger inhibitory effect of miR-197-3p for the G-allele.

[ORAI1;rs76753792;miR-519a-3p]: The case is analyzed in Chang *et al*. (2012). The SNP rs76753792 is defined by C ↔ T at 3′UTR position 86 of NM_032790.3. The 3′UTR length is 1034 nt. The authors study the susceptibility of atopic dermatitis in Japanese and Taiwanese populations. Among other methods, real-time PCR is applied to gene expression analysis. Chang *et al*. (2012) mention the impact of miRNAs as subject of future research, i.e. no specific miRNA is identified. Based on miRNA target predictions for ORAI1 and the SNP position, we selected miR-519a-3p for the present study. For miR-519a-3p, the binding prediction returned by `StarMir` is stronger for the C-allele.

[RAP1A;rs6573;miR-196a]: The case is analyzed in Wang *et al*. (2012a). The SNP rs6573 is defined by A ↔ C at 3′UTR position 366 of NM_002884.2. The 3′UTR length is 1078 nt. The authors study how rs6573 affects the risk of esophageal squamous cell carcinoma. The regulatory function of miR-196a is analyzed by luciferase reporter assays. The authors conclude a stronger inhibitory effect of miR-196a for the A-allele.

The 3′UTR- and SNP-sequences were identified by using NCBI databases and transcript information provided by ENSEMBL. A detailed description of all instances in terms of ENSEMBL ID codes is given in Supplementary Table S1 in Part I (Suppl-Mat-I), Section I-B.

3 RESULTS

When executing step (A), we applied the standard settings of `RNAsubopt` and `Barriers`, where isolated base pairings are not permitted and free energy values are discriminated with an accuracy of 0.1 kcal/mol. The setting of δE depends on the length of the 3′UTR and was selected in such a way that a sufficiently large number of metastable conformations is available for analyzing MS$(\cdots, k)$ with $k \leq 100$. Along with the energy offset δE, we tried to restrict metastable states to 'deep' local minima: The parameter D indicates the 'depth' of a local minimum or 'escape height' from a local minimum, which is taken from the barrier tree as the distance to the nearest saddle point. By $|\text{SecStruc}_{w/s}|$ (short for both cases of $|\text{SecStruc}_{weak}|$ and $|\text{SecStruc}_{strong}|$) we denote the number of secondary structures returned by `RNAsubopt` for an offset δE above the mfe conformation, where the index w indicates the allele with the weaker and s with the stronger miRNA inhibitory effect (binding prediction—for [ORAI1;rs76753792;miR-519a-3p]). Analogously, $|\text{MS}_{w/s}|$ is the number of local minima within the δE range with an 'escape height' larger or equal to D.

3.1 Number of metastable conformations

Results of step (A) with respect to $\delta\text{E}, D$ and $|\text{MS}_{w/s}|$ are summarized in Table 1. For $|\text{SecStruc}_{w/s}|$, the data are provided in Supplementary Table S2 in Part I (Suppl-Mat-I), Section I-C.

For 9 of the 14 instances, the values of $|\text{MS}_w|$ and $|\text{MS}_s|$ are relatively close or (much) larger for $|\text{MS}_s|$, see Table 1 and Figure 1. For the remaining instances, the ratio $|\text{MS}_w|/|\text{MS}_s|$ ranges from 1.31 (FGF20) up until 6.47 (AGTR1). Thus, the number of secondary structures classified as metastable conformations *per se* does not discriminate between the two cases of weaker and stronger binding to the associated miRNA.

Table 1. Data returned by `RNAsubopt` and `barriers`

	LIG3	CBR1	HTR3E	HLA-G	PARP1	WFS1	IL-23R	RYR3	AGTR1	FGF20	HOXB5	RAD51	ORAI1	RAP1A
L(3′UTR) nt	124	284	302	386	769	779	851	880	888	903	952	978	1034	1078
W-allele	A	G	A	C	C	A	A	G	C	T	G	A	T	C
S-allele	C	A	G	G	T	G	C	A	A	C	A	G	C	A
δEkcal/mol	6.0	6.0	6.0	4.0	3.0	2.7	2.0	2.3	2.3	2.3	2.0	2.2	2.0	2.0
Dkcal/mol	1.2	1.4	1.4	1.4	1.2	0.8	1.2	1.2	0.8	0.8	0.8	1.2	0.9	0.9
$\lvert\text{MS}_w\rvert$	349	7457	996	11 957	1709	79 273	1080	2628	14 943	7783	6481	6291	1332	238
$\lvert\text{MS}_s\rvert$	317	11 187	1173	10 473	14 281	79 577	964	19 936	2308	5936	3746	3850	577	238

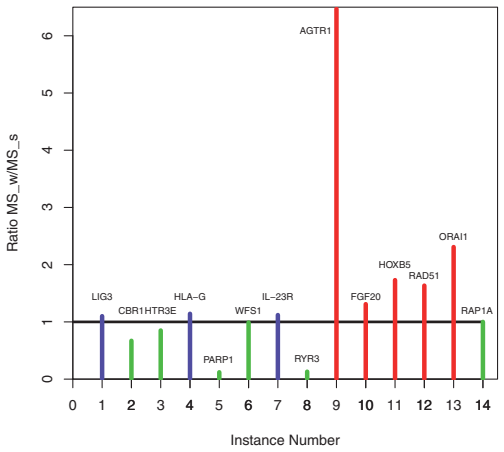

Fig. 1. $|MS_{weak}|/|MS_{strong}|$ ratios: The green colour indicates a larger or equal number of metastable conformations for the allele with the stronger miRNA binding. For the blue colour the ratio is larger but close to one, and the red colour indicates a much larger number of metastable conformations for the allele with the weaker miRNA binding

The correlation between $|MS_w|$ and $|MS_s|$ does not necessarily extend to $|SecStruc_{w/s}|$. For example, for LIG3 we have $|MS_w|/|MS_s| > 1$, whereas $|SecStruc_w|/|SecStruc_s| < 1$.

3.2 MiRNA binding sites and energy predictions

The results obtained in step (B) are summarized in Table 2. For a given input [3′UTR/SNP;miRNA], `STarMir` returns a large number of data items and graphical representations of miRNA binding patterns. We focus on the binding regions and four energy values:

(a) ΔG_{nucl} relates to the assumption that the initial stage of base pairing (nucleation) requires a gain in free energy that is greater than the energy cost for the translational and rotational entropy loss when both miRNA and mRNA are fixed in a conformation by intermolecular base pairing. The value of ΔG_{nucl} is calculated by using a sample of 1000 structures computed by `Sfold`, where the calculation is restricted to short base pair blocks within a 4 nt single-stranded segment of a putative binding site. The energy cost for the translational and rotational entropy loss is called initiation energy ΔG_{init}, and the standard setting in the `STarMir` tool is $\Delta G_{\text{init}} = 4.09 > \text{kcal/mol}$, i.e. $\Delta G_{\text{nucl}} + \Delta G_{\text{initiation}} < 0\text{kcal/mol}$ can be seen as a basic requirement for miRNA–mRNA interaction.

(b) $\Delta G_{\text{disrupt}}$ is the energy needed for the disruption of base pairs that are present within a putative binding site in a given mRNA secondary structure. We set $\Delta\Delta G_{\text{dis}} = \Delta G^{w}_{\text{disrupt}} - \Delta G^{s}_{\text{disrupt}}$.

(c) ΔG_{hybrid} is the energy gained by the hybridization of the miRNA with the particular binding site.

(d) ΔG_{total} is the basic `STarMir` score defined by $\Delta G_{\text{total}} = \Delta G_{\text{hybrid}} - \Delta G_{\text{disrupt}}$. We set $\Delta\Delta G_{\text{tot}} = \Delta G^{s}_{\text{total}} - \Delta G^{w}_{\text{total}}$.

The binding regions and energy values shown in Table 2 are determined by the `STarMir` prediction with the strongest seed

Table 2. MiRNA binding predictions by `STarMir`

	LIG3	CBR1	HTR3E	HLA-G	PARP1	WFS1	IL-23R	RYR3	AGTR1	FGF20	HOXB5	RAD51	ORAI1	RAP1A
L(3′UTR) nt	124	284	302	386	769	779	851	880	888	903	952	978	1034	1078
W-allele	A	G	A	C	C	A	A	G	C	T	G	A	T	C
S-allele	C	A	G	G	T	G	C	A	A	C	A	G	C	A
miR-	221	574	510	148a	145	668	let-7e	367	155	433	7	197	519a	196a
SNP pos	83	133	76	233	607	253	309	839	86	182	141	718	86	366
BindSite-w	77–98	121–62	50–80	221–39	592–614	234–58	291–308	830–57	57–90	166–87	126–54	707–25	81–102	348–70
BindSite-s	80–98	121–62	50–80	221–39	592–614	234–58	291–310	835–57	57–90	166–87	126–54	707–25	69–88	348–70
$\Delta G^{w}_{\text{total}}$	−8.53	−19.72	−14.24	−14.62	1.70	−6.40	0.70	4.35	11.28	−2.18	−8.29	−7.00	5.62	−11.30
$\Delta G^{w}_{\text{disrupt}}$	−11.77	−8.09	−8.26	−9.28	−20.90	−14.20	−21.40	−19.65	−27.89	−10.12	−13.61	−15.00	−23.52	−5.40
$\Delta G^{w}_{\text{hybrid}}$	−20.30	−27.80	−22.50	−23.90	−19.20	−20.60	−20.70	−15.30	−16.60	−12.30	−21.90	−22.00	−17.90	−16.70
$\Delta G^{w}_{\text{nucl}}$	−3.39	−5.36	−2.46	−4.55	−0.65	−1.72	−0.05	−0.19	−2.27	−4.66	−0.49	−1.20	−0.02	−3.97
$\Delta G^{s}_{\text{total}}$	−8.64	−21.84	−16.05	−19.49	−7.51	−15.96	1.12	−3.38	7.66	−5.23	−11.20	−14.31	−7.21	−16.25
$\Delta G^{s}_{\text{disrupt}}$	−15.96	−5.96	−12.65	-11.01	−12.19	−10.64	−25.82	−11.42	−28.56	−9.27	−11.20	−14.29	−9.49	−5.05
$\Delta G^{s}_{\text{hybrid}}$	−24.60	−27.80	−28.70	−30.50	−19.70	−26.60	−24.70	−14.80	−20.90	−14.50	−22.40	−28.60	−16.70	−21.30
$\Delta G^{s}_{\text{nucl}}$	−0.20	−5.22	−5.05	−3.24	−0.93	−4.42	0.00	−1.22	−0.58	−2.36	−1.39	−0.56	−4.32	−6.97
$\Delta\Delta G_{\text{tot}}$	−0.11	−2.12	−1.81	−4.87	−9.21	−9.56	0.42	−7.73	−3.62	−3.05	−2.91	−7.31	−12.83	−4.95
$\Delta\Delta G_{\text{dis}}$	4.19	−2.13	4.39	−12.89	−8.71	−3.56	4.42	−8.23	0.67	−0.85	−2.41	−0.71	−14.03	−0.35
$\lvert BP_w \rvert$	9	14	9	9	22	12	15	14	24	10	15	6	16	12
$\lvert BP_s \rvert$	7	12	15	11	11	12	17	7	22	9	15	6	6	12

match among all predictions having the SNP position inside, leading in some cases to weaker ΔG_{total} values. In case of a missing strong seed match, the target predictions provided by the PITA tool (Kertesz *et al*., 2007) are taken into account.

By $|BP_w|$ and $|BP_s|$, we denote the number of base pairs in the corresponding mfe conformation within the miRNA binding region predicted by STarMir for an individual allele (w = *weak* and s = *strong* allele with respect to miRNA binding).

Detailed information about the STarMir output is provided in Supplementary Material Part II (Suppl-Mat-II), which also includes PITA predictions and, where available, data provided by FindTar (Ye *et al*., 2008); for target predictions by other tools, please see Supplementary Material Part I (Suppl-Mat-I), Section I-D. Except for IL-23 R (L = 851), the ΔG_{total} predictions by STarMir are stronger for the allele identified for stronger miRNA bindings in the corresponding publication (cf. the row for $\Delta\Delta G_{\text{tot}}$ and description of instances in Section 2; we recall that for ORAI1 no particular miRNA is mentioned). Except for LIG3 (L = 124), HTR3E (L = 302), HLA-G (L = 386), IL-23R (L = 851) and AGTR1 (L = 888), the absolute value of $\Delta G_{\text{disrupt}}$ is smaller for the allele with the stronger miRNA binding stated in the corresponding publication (s-allele indicated in Table 2, see also the row for $\Delta\Delta G_{\text{dis}}$). For IL-23R and AGTR1, the total STarMir score is positive for both alleles, for IL-23R even with $\Delta G^{s}_{\text{total}} > \Delta G^{w}_{\text{total}}$.

The instance [IL-23R;rs10889677;let-7e] has been analyzed in Zwiers *et al*. (2012), with strong experimental evidence for an inhibitory effect of let-7e on the C-allele. The binding patterns for the C allele provided by STarMir (see Suppl-Mat-II) and in Figure 3A of Zwiers *et al*. (2012) differ only slightly towards the 5′ end of the C allele. We note that for the A allele we have the only case in Table 2 where the SNP position is not within the 'binding site' predicted by STarMir. The PITA tool returns (the equivalent of) $\Delta G_{\text{total}} = -14.12$ kcal/mol for the C allele, and $\Delta G_{\text{total}} = -10.02$ kcal/mol for the A allele (see Suppl-Mat-II), which is in line with the experimental data from Zwiers *et al*. (2012). The absolute value of $\Delta G_{\text{disrupt}}$ is also slightly smaller for the C allele.

For [LIG3;rs4796030;miR-221], the absolute value of $\Delta G_{\text{disrupt}}$ is larger for the C allele by 4.19 kcal/mol, and there is no PITA prediction with the SNP position inside the reported binding region. The PITA predictions close to the SNP position favour the C allele with respect to the total score and ΔG_{hybrid}, which complies with Teo *et al*. (2012). The selection of positions 77–98 for the A allele and 80–98 for the C allele is motivated by the seed-like bindings predicted by STarMir, see Suppl-Mat-II.

For ORAI1, the positions 69–88 were selected for the C allele due to the larger value of $\Delta G^{s}_{\text{nucl}} = -4.32$kcal/mol.

For [HTR3E;rs56109847;miR-510-5p], the ΔG_{hybrid} values is stronger for the G allele (*s*-case) by 6.2 kcal/mol, and the stronger binding prediction is supported by PITA and FindTar (no binding prediction for A allele—*w*-case).

For [HLA-G;rs1063320;miR-148a-3p], PITA returns a stronger total score for the G allele, which is in line with the STarMir total score and experimental data from Tan *et al*. (2007). The absolute value of $\Delta G_{\text{disrupt}}$ is larger for the G allele (by 1.12 kcal/mol for PITA and 1.73 kcal/mol for STarMir). FindTar also strongly supports the miRNA binding to the G allele.

For [AGTR1;rs5186;miR-155-5p], the ΔG_{hybrid} values is stronger for the A allele (*s*-case) by 4.3 kcal/mol. The stronger binding prediction for the A allele is supported by PITA and FindTar (no binding prediction for C allele—*w*-case).

For the 14 instances under consideration, the PITA predictions either support (or improve for IL-23R) the STarMir predictions displayed in Table 2, or no predictions with the SNP position inside the reported binding region are returned.

In summary, if STarMir and PITA (for IL-23R) are taken together, the ΔG_{total} predictions are stronger for the corresponding allele (*s*-case) identified in the description of the 14 instances in Section 2.

3.3 Analysis of metastable conformations

STarMir and PITA operate on sequences as input, not on representations of secondary structures. Therefore, we use RNAeval (Gruber *et al*., 2008) for the energy evaluation of metastable conformations within binding regions predicted by STarMir, which also complies with the data generated by RNAsubopt and Barriers, see Table 1. To facilitate a coherent analysis of energy values, we use energy values calculated for binding regions within the corresponding mfe structure as templates for comparisons, instead of using the associated $\Delta G_{\text{disrupt}}$ values reported in Table 2.

Let S denote a secondary structure (either mfe conformation or metastable conformation) for a 3′UTR (wild-type or SNP-type) listed in Table 1. By S_{open} we denote the associated secondary structure where all base pair bindings within the miRNA binding region reported in Table 2 are removed. For example, if S is the mfe conformation of the 3′UTR of the C allele of LIG3, then S has seven base pair bindings in positions 80–98 (see Table 2), and the seven base-pair bindings are removed in S_{open}. For S being a metastable conformation, the number of base pair bindings within the miRNA binding region can be larger, the same or smaller in comparison with the corresponding value reported in Table 2 for the mfe conformation. We then define

$$\Delta G^{\text{ind}}_{\text{open}} = \texttt{RNAeval}(\text{S}) - \texttt{RNAeval}(\text{S}_{\text{open}}) \tag{1}$$

$$\Delta\Delta G_{\text{ind}} = \Delta G^{\text{ind:}s}_{\text{open}} - \Delta G^{\text{ind:}w}_{\text{open}} \tag{2}$$

The index 'ind' specifies the different cases we consider, and the different values in (1) and (2) relate either to individual structures or to average values (according to the value assigned to 'ind') over sets of metastable conformations:

(i) ind = mfe indicates the single mfe conformation.
(ii) ind = tot indicates the average value over all metastable conformations as counted in Table 1. For example, $\Delta G^{\text{tot:}s}_{\text{open}}$ stands for the average value over 317 metastable conformations in case of the C allele of LIG3, see Table 1.
(iii) ind = N + indicates for N = 20, 60 and 100 the N + metastable conformations S with the N + lowest free energy values calculated by RNAeval(S) that are above the mfe conformation. Because each energy level usually adds more than a single conformation, the notion 'N+' indicates that the highest energy level involved covers N conformations above the mfe structure, plus in most cases

some more structures, i.e. the actual number N+ of conformations can be slightly larger than N.
There are a few exceptions for 60+, where the number is between 50 and 60, because the next energy level results already in a conformation number above 100. For example, for PARP1, 60+ means 55, because 55 conformations are accumulated at energy level −186.4 kcal/mol, whereas level −186.3 kcal/mol adds 48 conformations, which leads to 100+ = 103.

Furthermore, we order the metastable conformations S_{ms} with respect to the absolute value of $|\Delta G_{open}|$ and the depth $D(S_{ms})$ (deepest first), respectively. As in (2), we define

$$\Delta\Delta G_{asc:N+} = \Delta G_{open}^{asc:N+:s} - \Delta G_{open}^{asc:N+:w} \quad (3)$$

$$R_{N+} = D_{N+}^{w}/D_{N+}^{s} \quad (4)$$

where $D_{N+}^{w/s}$ denotes the average depth of the N+ deepest metastable conformations. The index 'asc' in (3) indicates that, unlike in (2), the N+ metastable conformations are ranked in ascending order with respect to $|\Delta G_{open}|$, and the average value is taken over N+ conformations.

Finally, we combine 'opening energies', as defined in (2) and (3), with the depth $D(S_{ms})$ of metastable conformations, as defined by $D_{N+}^{w/s}$ and used in (4): We look at the average depth of structures S_{ms} with the N+ smallest values of $|\Delta G_{open}|$. Let $D_{open}^{N+:s}$ and $D_{open}^{N+:w}$ denote the average depth of metastable conformations S_{ms} counted in the calculation of $\Delta G_{open}^{asc:N+:s}$ and $\Delta G_{open}^{asc:N+:w}$. We then set

$$\Delta D_{open}^{N+} = D_{open}^{N+:s} - D_{open}^{N+:w} \quad (5)$$

Comprehensive information about the distribution of metastable conformations and their respective energy values is provided in Supplementary Material Part III. (Suppl-Mat-III). In Table 3 we report representative data that are useful for discriminating between the two alleles involved for each instance. Positive values of $\Delta\Delta G_{mfe}$, $\Delta\Delta G_{tot}$, $\Delta\Delta G_{N+}$, $\Delta\Delta G_{asc:N+}$ and ΔD_{open}^{N+} are interpreted as being in favour of the allele with the stronger miRNA binding stated in the underlying literature source, which is also the case for $R_{N+} < 1$.

For LIG3 (L = 124), HTR3E (L = 302), HLA-G (L = 386), IL-23R (L = 851) and RAD51 (L = 978), the data for $\Delta\Delta G_{mfe}$, $\Delta\Delta G_{tot}$, $\Delta\Delta G_{N+}$ and $\Delta\Delta G_{asc:N+}$ shown in Table 3 are not in favour of the allele with the stronger miRNA binding stated in the underlying literature (HOXB5 with L = 952 not included here, please see later in the text).

For LIG3, IL-23R and RAD51 the values of R_{N+} are either close to or smaller than 1.00, and the values of ΔD_{open}^{N+} are all positive, suggesting more stable local minima for structures with the smallest absolute value of opening energies for the allele with the stronger miRNA binding.

For HLA-G, the values of R_{N+} are equal or close to 1.00, and the values of ΔD_{open}^{60+} and ΔD_{open}^{100+} are positive. Moreover, the $\Delta\Delta G_{asc:N+}$-values shown in Table 3 are close to zero, and the values of $\Delta\Delta G_{N+}$ are in the range of $\Delta G_{disrupt}^{s} - \Delta G_{disrupt}^{w} = -1.73$ kcal/mol from Table 2. Thus, for HLA-G the values of $\Delta G_{hybrid}^{s/w}$ shown in Table 2 seem to be decisive for an assessment of a putative miR-148a-3p ↔ HLA-G/rs1063320 binding (based on prediction tools).

For HTR3E, only the values of ΔD_{open}^{N+} are in favour of the allele with the stronger miRNA binding stated in the underlying literature. Unlike the case of HLA-G, the negative values of $\Delta\Delta G_{N+}$ and $\Delta\Delta G_{asc:N+}$ are more substantial, i.e. in terms of absolute values above the range of corresponding values from Table 2. Thus, only the values of ΔG_{hybrid}^{s}, ΔG_{nucl}^{s} (both Table 2) and ΔD_{open}^{N+} (Table 3) support the binding to the G allele (*s*-case).

For HOXB5 (L = 952), the $\Delta\Delta G_{N+}$ values are negative yet close to zero, and the $\Delta\Delta G_{asc:N+}$ values support the stronger miRNA binding to the A-allele (*s*-case), which makes the HOXB5 instance different from the five instances discussed

Table 3. Energy values calculated by `RNAeval`

	LIG3	CBR1	HTR3E	HLA-G	PARP1	WFS1	IL-23R	RYR3	AGTR1	FGF20	HOXB5	RAD51	ORAI1	RAP1A
L(3′UTR) nt	124	284	302	386	769	779	851	880	888	903	952	978	1034	1078
W-allele	A	G	A	C	C	A	A	G	C	T	G	A	T	C
S-allele	C	A	G	G	T	G	C	A	A	C	A	G	C	A
$\Delta\Delta G_{mfe}$	−2.20	1.00	−13.51	−0.80	13.02	0.00	−4.20	0.75	2.70	3.20	−0.70	−1.90	21.81	0.00
$\Delta\Delta G_{tot}$	−1.95	1.31	−11.18	−0.80	17.98	0.03	−7.44	0.82	2.74	2.83	−0.56	−1.85	21.80	0.00
$\Delta\Delta G_{100+}$	−2.60	2.19	−11.26	−1.18	12.92	0.00	−4.20	0.58	2.70	3.91	−0.65	−1.90	21.81	0.00
$\Delta\Delta G_{60+}$	−3.20	2.10	−11.39	−1.73	11.48	0.00	−4.20	0.66	2.70	4.19	−0.65	−1.90	21.81	0.00
$\Delta\Delta G_{20+}$	−3.07	3.56	−10.28	−1.61	7.03	0.00	−4.20	0.75	2.70	4.67	−0.70	−1.90	21.81	0.00
$\Delta\Delta G_{asc:100+}$	−1.90	1.04	−6.86	−0.69	21.84	7.42	−13.40	2.00	2.72	1.84	1.50	−2.18	21.65	0.00
$\Delta\Delta G_{asc:60+}$	−1.58	0.98	−4.70	−0.78	21.36	7.48	−13.40	2.13	2.72	1.93	3.38	−2.37	21.54	0.00
$\Delta\Delta G_{asc:20+}$	−0.91	0.93	−4.38	−0.69	21.04	7.62	−13.40	2.61	2.70	2.00	2.36	−2.61	21.00	0.00
R_{100+}	0.96	0.90	1.14	1.00	0.92	1.00	0.98	0.92	1.25	1.06	1.06	1.00	1.14	1.00
R_{60+}	0.93	0.92	1.11	1.00	0.93	1.00	1.00	0.91	1.32	1.05	1.05	1.00	1.10	1.00
R_{20+}	0.88	0.91	1.20	1.01	0.93	1.00	1.00	0.93	1.09	1.04	1.04	1.00	1.07	1.00
ΔD_{open}^{100+}	0.05	−0.05	0.15	0.02	0.35	−0.08	0.30	0.04	1.13	−0.07	0.17	0.23	−0.05	0.00
ΔD_{open}^{60+}	0.05	−0.11	0.26	0.49	0.53	−0.10	0.40	0.00	1.13	−0.11	−0.14	0.39	0.08	0.00
ΔD_{open}^{20+}	0.04	0.00	0.13	−0.02	0.64	−0.16	0.40	−0.26	1.13	−0.15	−0.03	1.05	0.75	0.00

earlier in the text. The R_{N+} values are close to 1.00, and ΔD_{open}^{100+} is positive.

For RAP1A (L = 1078), the `STarMir` predictions for $\Delta G_{disrupt}$ and ΔG_{hybrid}, respectively, are close for both alleles. The SNP at position 366 is located in the middle of a loop (positions 363–369) in both mfe secondary structures, which leads to identical values of $\Delta G_{open}^{ind:s/w}$ and related features. Therefore, similar to HTR3E, ΔG_{hybrid}^{s} and ΔG_{nucl}^{s} from Table 2 appear to determine the evaluation of miR-196a ↔ RAP1A/rs6573 bindings.

For CBR1 (L = 284), WFS1 (L = 779), RYR3 (L = 880), AGTR1 (L = 888), FGF20 (L = 903) and ORAI1 (L = 1034), the data for $\Delta\Delta G_{mfe}$, $\Delta\Delta G_{tot}$, $\Delta\Delta G_{N+}$ and $\Delta\Delta G_{asc:N+}$ (Table 3) are all in favour of the allele with the stronger miRNA binding stated in the underlying literature. However, for each of the five instances at least one of the two parameters R_{N+} and ΔD_{open}^{N+} does not fully support the predicted binding.

For RYR3, only $\Delta D_{open}^{20+} = -0.26$ kcal/mol is clearly not in favour of the predicted binding. Similarly, the values of ΔD_{open}^{N+} do not support the predicted stronger binding for CBR1. For ORAI1, this is the case for the R_{N+} values and $\Delta D_{open}^{100+} = -0.05$ kcal/mol, but with a relatively strong value of $\Delta D_{open}^{20+} = +0.75$ kcal/mol. For AGTR1, all three values of ΔD_{open}^{N+} are clearly in favour of the predicted binding pattern. The instances WFS1 and FGF20 are the only two cases where for all N + values both parameters do not support the predicted stronger binding.

Finally, for PARP1 (L = 769) all energy values shown in Table 3 support the stronger miRNA binding to the T-allele (*s*-case).

We recall that for the instance ORAI1 from Chang *et al.* (2012) no individual miRNA is identified in the literature source. For ORAI1 and miRNA-519a-3p, the two binding sites predicted by `STarMir` intersect only by 8 nt, with a positive value $\Delta G_{total} = 5.62$ kcal/mol for the T-allele (*w*-case), which suggests that no binding occurs. Although $R_{N+} > 1.00$ for all cases of N+ considered, we obtain strong positive values for $\Delta\Delta G_{mfe}$, $\Delta\Delta G_{tot}$, $\Delta\Delta G_{N+}$ and $\Delta\Delta G_{asc:N+}$, along with the `STarMir` predictions $\Delta G_{total} = -7.21$ kcal/mol and $\Delta G_{nucl} = -4.32$ kcal/mol for the C-allele (*s*-case). Moreover, the data for ΔD_{open}^{N+} are $\Delta D_{open}^{10+} = 0.89$ kcal/mol, $\Delta D_{open}^{20+} = 0.75$ kcal/mol and $\Delta D_{open}^{60+} = 0.08$ kcal/mol, which support a binding of miR-519a-3p to the 3′UTR of ORAI1.

4 CONCLUSION

Of the 14 instances we analyzed, 13 instances are sensitive to the parameters $\Delta\Delta G_{ind}$, $\Delta\Delta G_{asc:N+}$, R_{N+} and ΔD_{open}^{N+} we introduced in Equation (1) until Equations (5). For RAP1A (L = 1078), slightly larger values of δE did not create differences between basic parameters for both alleles and eventually led to an unmanageable size of data for standard desktop computer configurations.

The absence of experimental data about copy numbers of mRNA transcripts considered in the present study prevents the selection of a particular value of ΔD_{open}^{N+} (or of the other two highlighted parameters), which is why we considered three representative values of N + simultaneously, without calculating *P*-values. The upper bound of N = 100+ is motivated by the data provided in Subkhankulova *et al.* (2008), see Figure 3B there.

The data provided in Table 3 indicate that $\Delta\Delta G_{mfe}$, $\Delta\Delta G_{tot}$, $\Delta\Delta G_{N+}$ not necessarily contribute to a deeper insight into miRNA binding patterns to different alleles, see also the discussion of instances in Section 3.3. In particular, $\Delta\Delta G_{tot}$ and $\Delta\Delta G_{N+}$ are related only to free energy values of metastable structures, which is why a further discrimination by the depth of metastable conformations and the opening energy of binding regions was introduced. Consequently, we focus in the summary of findings presented in Table 4 on the values calculated for $\Delta\Delta G_{asc:N+}$, R_{N+} and ΔD_{open}^{N+}.

In Table 4, the row '`StarMir`(total)' indicates by '+' that the ΔG_{total} score (see Table 2) supports the allele with the stronger miRNA binding stated in the underlying literature; '+/–' indicates $0 < \Delta G_{total}^{s} < \Delta G_{total}^{w}$. As aforementioned, data about estimations of copy numbers are not available for the mRNA transcripts we consider in the present study. To avoid the inclusion of irrelevant data (by averaging or thresholding), we consider a 'best case scenario' for each instance: We select a value of N+ in such a way that the support of the miRNA binding to the allele identified in the underlying literature source (*s*-case) is maximized. In case of multiple N+ values (for 11 instances same pattern as in Table 4 for at least two N+), the smallest N+ is selected and named in Table 4 in the row 'N+ selection'.

For the selected N+, the Table 4 entry is labelled as positive '+', if $\Delta\Delta G_{asc:N+}$ and ΔD_{open}^{N+} are positive, respectively, or $R_{N+} < 1.00$. If the data are inconclusive (equal or close to 0.00 or 1.00), we use '+/–'. For example, for HLA-G we select

Table 4. Summary of discrimination between WT and SNP alleles

	LIG3	CBR1	HTR3E	HLA-G	PARP1	WFS1	IL-23R	RYR3	AGTR1	FGF20	HOXB5	RAD51	ORAI1	RAP1A
L(3′UTR) nt	124	284	302	386	769	779	851	880	888	903	952	978	1034	1078
W-allele	A	G	A	C	C	A	A	G	C	T	G	A	T	C
S-allele	C	A	G	G	T	G	C	A	A	C	A	G	C	A
`StarMir` (total)	+	+	+	+	+	+	−	+	±	+	+	+	+	+
N + selection	20+	20+	20+	60+	20+	20+	100+	100+	20+	20+	100+	20+	20+	20+
$\Delta\Delta G_{asc:N+}$	±	+	−	±	+	+	−	+	+	+	+	−	+	±
R_{N+}	+	+	−	±	+	±	+	+	−	−	−	±	−	±
ΔD_{open}^{N+}	+	±	+	+	+	−	+	+	+	−	+	+	+	±

$N+ = 60+$ and obtain from Table 3 the entries for Table 4 as follows: $\Delta\Delta G_{\text{asc:N+}} = +/-$, $R_{N+} = +/-$, and $\Delta D^{N+}_{\text{open}} = +$.

Table 4 demonstrates that the combined measure $\Delta D^{N+}_{\text{open}}$ defined in (5) is the best match to the binding predictions, with two inconclusive and two negative values. The inconclusive value of $\Delta D^{20+}_{\text{open}}$ for CBR1 is accompanied by two positive values of the other two parameters, and RAP1A is a special instance due to the SNP location, as discussed in Section 3.3. The negative values of $\Delta D^{20+}_{\text{open}}$ for WFS1 and FGF20 are accompanied by relatively strong positive values of $\Delta\Delta G_{\text{asc:20+}}$. Future research will include to establish a suitable total score that takes into account weighted partial scores provided by $\Delta\Delta G_{\text{asc:N+}}$, R_{N+} and $\Delta D^{N+}_{\text{open}}$. For example, the simple weight selection $-1 \equiv -$, $0 \equiv \pm$ and $1 \equiv +$ applied to Table 4 results in a total score of -1 for HTR3E and FGF20, and a total score of 0 for WFS1, RAD51 and RAP1A. However, WFS1 has a strong $\Delta\Delta G_{\text{asc:N+}}$ score, RAD51 has a strong $\Delta D^{N+}_{\text{open}}$ score and RAP1A is a special case due to the location of the SNP. The relatively strong score of $\Delta\Delta G_{\text{asc:N+}}$ could outweigh the other two scores. HTR3E remains an ambiguous case, see also the discussion of the HTR3E–miR-510-5p instance in Sethupathy and Collins, (2008).

We hypothesize that an in-depth analysis of metastable conformations based on parameters such as $\Delta\Delta G_{\text{asc:N+}}$, R_{N+} and $\Delta D^{N+}_{\text{open}}$ can provide useful information for the assessment of putative miRNA–mRNA bindings in the context of SNPs. In the literature sources, we researched for the current study, the number of genes and miRNAs exposed to experimental analysis is relatively small, yet each analysis is time-consuming and costly. Examining features of metastable conformations in a preprocessing step of wet lab experiments may improve the confidence about expected miRNA–mRNA bindings. We emphasize that for ORAI1 no specific miRNA is identified by Chang *et al.* (2012). The data we presented support the binding of miR-519a-3p to the 3′UTR of ORAI1 in the region of position 86.

In the present study, the energy values associated with the accessibility of binding sites are approximated by using `RNAeval`, and the range of the binding sites is directly adapted from `STarMir`. In future research, we aim at individual miRNA–mRNA binding predictions over samples of metastable conformations taken from the sets of secondary structures defining the values of $\Delta\Delta G_{\text{asc:N+}}$, R_{N+} and $\Delta D^{N+}_{\text{open}}$.

The approach presented in the current study is clearly limited by the vast amount of secondary structures returned by `RNAsubopt` for longer sequences. Because we are aiming at the applicability of `MSbind` by using standard hardware facilities, we are seeking for alternative ways of identifying metastable conformations, with preference given to approximation methods, where at the cost of completeness (within an acceptable error range) metastable conformations are calculated within a feasible time. Future research will also include to investigate haplotypes and miRNA bindings, as discussed in Martin *et al.* (2012), and the analysis of SNPs in miRNAs, see (Johnson *et al.*, 2011). Furthermore, we plan to examine data provided by eQTL studies as an additional source of experimental data bearing information about the effect of SNPs on miRNA bindings to putative targets. For example, Westra *et al.* (2013) report that *trans*-eQTL SNPs are enriched within miRNA binding sites in blood cell lines studied in the article, see Figure 1 in Westra *et al.* (2013). Although in the present study we focus on miRNA binding within 3′UTRs, we also plan to look into the case where miRNA binding sites near the terminal coding region along with the presence of SNPs may affect the overall 2D or 3D structure (The authors are grateful to an anonymous referee for pointing out the potential of eQTL studies as well as the importance of miRNA binding sites near terminal coding regions in the context of miRNA bindings to metastable RNA conformations.).

ACKNOWLEDGMENTS

The authors are grateful to the anonymous referees for their careful reading of the manuscript and the multitude of valuable suggestions that helped them to improve and extend the presentation of their results.

Conflict of Interest: none declared.

REFERENCES

Arvey,A. *et al.* (2010) Target mRNA abundance dilutes microRNA and siRNA activity. *Mol. Syst. Biol.*, **6**, 363.

Balaga,O. *et al.* (2012) Toward a combinatorial nature of microRNA regulation in human cells. *Nucleic Acid Res.*, **40**, 9404–9416.

Baker,C. *et al.* (2012) Stochastic modeling of regulation of gene expression by multiple small RNAs. *Phys. Rev.*, **85**, 061915.

Barenboim,M. *et al.* (2010) MicroSNiPer: a web tool for prediction of SNP effects on putative microRNA targets. *Hum. Mutat.*, **31**, 1223–1232.

Bruno,A. *et al.* (2012) miRdSNP: a database of disease-associated SNPs and microRNA target sites on 3′UTRs of human genes. *BMC Bioinformatics*, **13**, 44.

Chang,W. *et al.* (2012) ORAI1 genetic polymorphisms associated with the susceptibility of atopic dermatitis in Japanese and Taiwanese populations. *PLoS One*, **7**, e29387.

Chen,J.L. *et al.* (2012) Testing the nearest neighbor mModel for canonical RNA base pairs: revision of GU parameters. *Biochemistry*, **51**, 3508–3522.

Cuccato,G. *et al.* (2011) Modeling RNA interference in mammalian cells. *BMC Syst. Biol.*, **5**, 19.

Ding,Y. *et al.* (2005) RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA*, **11**, 1157–1166.

Garcia,D.M. *et al.* (2011) Weak seed-pairing stability and high target-site abundance decrease the proficiency of *lsy-6* and other microRNAs. *Nat. Struct. Mol. Biol.*, **18**, 1139–1147.

Gennarino,V. *et al.* (2012) Identification of microRNA-regulated gene networks by expression analysis of target genes. *Genome Res.*, **22**, 1163–1172.

Gruber,A.R. *et al.* (2008) The Vienna RNA Web suite. *Nucleic Acids Res.*, **36**, W70–W74.

Haas,U. *et al.* (2012) MicroRNA-mediated regulation of gene expression is affected by disease-associated SNPs within the 3′–UTR via altered RNA structure. *RNA Biol.*, **9**, 924–937.

Hammell,M. *et al.* (2008) mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat. Methods*, **5**, 813–819.

Johnson,E. and Srivastava,R. (2013) Volatility in mRNA secondary structure as a design principle for antisense. *Nucleic Acids Res.*, **41**, e43.

Johnson,A.D. *et al.* (2011) RNA structures affected by single nucleotide polymorphisms in transcribed regions of the human genome. *Web Med. Cent. Bioinf.*, **2**, WMC001600.

Kalabus,J.L. *et al.* (2012) MicroRNAs differentially regulate carbonyl reductase 1 (CBR1) gene expression dependent on the allele status of the common polymorphic variant rs9024. *PLoS One*, **7**, e48622.

Kapeller,J. *et al.* (2008) First evidence for an association of a functional variant in the microRNA-510 target site of the serotonin receptor type 3E gene with diarrhea predominant irritable bowel syndrome. *Hum. Mol. Genet.*, **17**, 2967–2977.

Kertesz,M. *et al.* (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
Kovacs-Nagy,R. *et al.* (2013) Association of aggression with a novel microRNA binding site polymorphism in the wolframin gene. *Am. J. Med. Genet. B*, **162B**, 404–412.
Larsson,E. *et al.* (2010) mRNA turnover rate limits siRNA and microRNA efficacy. *Mol. Syst. Biol.*, **6**, 433.
Loinger,A. *et al.* (2012) Competition between small RNAs: a quantitative view. *Biophys. J.*, **102**, 1712–1721.
Long,D. *et al.* (2007) Potent effect of target structure on microRNA function. *Nat. Struct. Mol. Biol.*, **14**, 287–294.
Lorenz,W.A. and Clote,P. (2011) Computing the partition function for kinetically trapped RNA secondary structures. *PLoS One*, **6**, e16178.
Luo,J. *et al.* (2012) A microRNA-7 binding site polymorphism in HOXB5 leads to differential gene expression in bladder cancer. *PLoS One*, **7**, e40127.
Marin,R.M. and Vaniček,J. (2012) Optimal use of conservation and accessibility filters in microRNA target prediction. *PLoS One*, **7**, e32208.
Martin,J.S. *et al.* (2012) Structural effects of linkage disequilibrium on the transcriptome. *RNA*, **18**, 77–87.
Mathews,D. *et al.* (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA*, **101**, 7287–7292.
Mullokandov,G. *et al.* (2012) High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat. Methods*, **9**, 840–849.
Muniategui,A. *et al.* (2012) Joint analysis of miRNA and mRNA expression data. *Brief. Bioinform.*
Osella,M. *et al.* (2011) The role of incoherent microRNA-mediated feedforward loops in noise buffering. *PLoS Comput. Biol.*, **6**, e1001101.
Ragan,C. *et al.* (2011) Quantitative prediction of miRNA–mRNA interaction based on equilibrium concentrations. *PLoS Comput. Biol.*, **7**, e1001090.
Sabarinathan,R. *et al.* (2013) RNAsnp: Efficient detection of local RNA secondary structure changes induced by SNPs. *Hum. Mutat.*, **34**, 546–556.
Saffarian,A. *et al.* (2012) RNA locally optimal secondary structures. *J. Comput. Biol.*, **19**, 1120–1133.
Saito,T. and Sætrom,P. (2012) Target gene expression levels and competition between transfected and endogenous microRNAs are strong confounding factors in microRNA high-throughput experiments. *Silence*, **3**, 3.
Salmena,L. *et al.* (2011) A ceRNA hypothesis: The Rosetta Stone of a hidden RNA language? *Cell*, **146**, 353–358.
Sethupathy,P. and Collins,F.S. (2008) MicroRNA target site polymorphisms and human disease. *Trends Genet.*, **24**, 489–497.
Shirdel,E. *et al.* (2011) NAViGaTing the micronome - using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs. *PLoS One*, **6**, e17429.
Solomatin,S. *et al.* (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature*, **463**, 681–686.
Subkhankulova,T. *et al.* (2008) Modelling and measuring single cell RNA expression levels find considerable transcriptional differences among phenotypically identical cells. *BMC Genomics*, **9**, 268.
Tan,Z. *et al.* (2007) Allele-specific targeting of microRNAs to HLA-G and risk of asthma. *Am. J. Hum. Genet.*, **81**, 829–834.
Teo,M. *et al.* (2012) 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*, **33**, 581–586.
Wang,G. *et al.* (2008) Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of alpha-synuclein. *Am. J. Hum. Genet.*, **82**, 283–289.
Wang,K. *et al.* (2012a) MiR-196a binding-site SNP regulates RAP1A expression contributing to esophageal squamous cell carcinoma risk and metastasis. *Carcinogenesis*, **33**, 2147–2154.
Wang,L. *et al.* (2012b) A miRNA binding site single-nucleotide polymorphism in the 3'-UTR region of the IL23R gene is associated with breast cancer. *PLoS One*, **7**, e49823.
Wang,H.J. *et al.* (2013) Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nature Gen.*, **45**, 1238–1245.
Wolfinger,M.T. *et al.* (2004) Efficient computation of RNA folding dynamics. *J. Phys. A Math. Gen.*, **37**, 4731–4741.
Wuchty,S. *et al.* (1999) Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, **49**, 145–164.
Ye,W. *et al.* (2008) The effect of central loops in miRNA:MRE duplexes on the efficiency of miRNA-mediated gene regulation. *PLoS One*, **3**, e1719.
Zhang,L. *et al.* (2011) Functional SNP in the microRNA-367 binding site in the 3′UTR of the calcium channel ryanodine receptor gene 3 (RYR3) affects breast cancer risk and calcification. *Proc. Natl Acad. Sci. USA*, **108**, 13653–13658.
Ziebarth,J. *et al.* (2012) PolymiRTS database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. *Nucleic Acids Res.*, **40**, 216–221.
Zwiers,A. *et al.* (2012) A variant of the IL-23R gene associated with inflammatory bowel disease induces loss of microRNA regulation and enhanced protein production. *J. Immunol.*, **188**, 1573–1577.