Genomic Med (2007) 1:29–33
DOI 10.1007/s11568-006-9000-3

RESEARCH ARTICLE

Searching for potential microRNA-binding site mutations amongst known disease-associated 3′ UTR variants

Nadia Chuzhanova · David N. Cooper · Claude Férec · Jian-Min Chen

Received: 15 September 2006 / Accepted: 14 December 2006 / Published online: 30 January 2007

Abstract The 3′ untranslated regions (3′ UTRs) of human protein-coding genes play a pivotal role in the regulation of mRNA 3′ end formation, stability/degradation, nuclear export, subcellular localisation and translation, and hence are particularly rich in *cis*-acting regulatory elements. One recent addition to the already large repertoire of known *cis*-acting regulatory elements are the microRNA (miRNA) target sites that are present in the 3′ UTRs of many human genes. miRNAs post-transcriptionally down-regulate gene expression by binding to complementary sequences on their cognate target mRNAs, thereby inducing either mRNA degradation or translational repression. To date, only one disease-associated 3′ UTR variant (in the *SLITRK1* gene) has been reported to occur within a bona fide miRNA binding site. By means of sequence complementarity, we have performed the first systematic search for potential miRNA-target site mutations within a set of 79 known disease-associated 3′ UTR variants. Since no variants were found that either disrupted or created binding sites for known human miRNAs, we surmise that miRNA-target site mutations are not likely to represent a frequent cause of human genetic disease.

Keywords *Cis*-acting regulatory elements · Human inherited disease · MicroRNA · MiRNA target site mutation · 3′ Untranslated region · 3′ UTR

Abbreviations

LAS	Left arm of the 'spacer' sequence between the upstream core polyadenylation signal and the pre-mRNA cleavage site
miRNA	MicroRNA
UCPAS	Upstream core polyadenylation signal
USS	Upstream sequence between the translational termination codon and the UCPAS
3′ UTR	3′ Untranslated region

N. Chuzhanova · D. N. Cooper
Institute of Medical Genetics, Cardiff University, Heath Park, Cardiff CF14 4XN, UK

C. Férec · J.-M. Chen
INSERM, U613, 29220 Brest, France

C. Férec · J.-M. Chen
Faculté de Médecine de Brest et des Sciences de la Santé, Université de Bretagne Occidentale, 29238 Brest, France

C. Férec · J.-M. Chen (✉)
Etablissement Français du Sang – Bretagne, 46 rue Félix Le Dantec, 29220 Brest, France
e-mail: Jian-Min.Chen@univ-brest.fr

C. Férec
CHRU Brest, Hôpital Morvan, Laboratoire de Génétique Moléculaire et d'Histocompatibilité, 29220 Brest, France

Introduction

The 3′ untranslated regions (3′ UTRs) of human protein-coding genes play a pivotal role in the regulation of mRNA 3′ end formation, stability/degradation, nuclear export, subcellular localisation and translation and are thus particularly rich in *cis*-acting regulatory elements (for recent reviews, see Chen et al. 2006a, b). One recent addition to the already large repertoire of *cis*-acting regulatory elements are the microRNA

(miRNA) target sites that are present in the 3′ UTRs of many human genes (e.g. John et al. 2004; Kiriakidou et al. 2004; Grun et al. 2005; Krek et al. 2005; Lewis et al. 2005; Robins and Press 2005; Xie et al. 2005). miRNAs are an abundant class of small (~22 nucleotide) non-coding RNAs. Upon binding to their cognate targets, they post-transcriptionally down-regulate gene expression by inducing either mRNA degradation or translational repression (for recent reviews, see Pillai 2005; Kim and Nam 2006; Valencia-Sanchez et al. 2006). To date, at least 800 distinct human miRNAs (~3% of the total number of human genes) have been identified (Bentwich et al. 2005) but many more probably still remain to be discovered. These miRNAs are each likely to be responsible for repressing the expression of a large number of different target genes (Lim et al. 2005).

The importance of miRNAs in gene regulation and their potential significance in both cancer biology and gene evolution (Farh et al. 2005; Lim et al. 2005; Zhang et al. 2006) suggests that mutations in miRNA-target sites might well also be important in the aetiology of human inherited disease. Indeed, one such lesion has recently been reported: a G → A transition (absent in 4,296 control chromosomes), which replaces a G:U wobble base-pair with an A:U Watson–Crick pairing in a binding site for human miRNA hsa-miR-189 within the 3′ UTR of the Slit and Trk-like 1 gene (*SLITRK1*; MIM# 609678), was identified in two unrelated patients with Tourette's syndrome and obsessive-compulsive symptoms (Abelson et al. 2005; Fig. 1). In vitro functional analysis demonstrated that, in the presence of hsa-miR-189, the mutant allele gave rise to increased repression of the reporter gene as compared with the wild-type allele.

In this study, we have made a first attempt to identify further potential miRNA-binding site mutations by systematically screening previously reported disease-associated 3′ UTR variants.

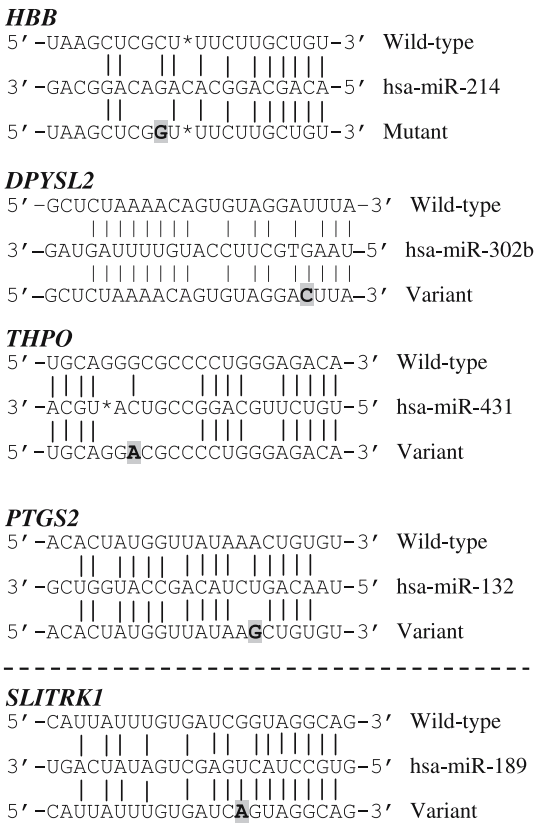

Fig. 1 Complementarity of several pairs of wild-type and variant alleles with different miRNAs. Watson–Crick base-pairs are depicted by bars. Gaps (indicated by asterisks) were introduced so as to maximise complementarity in the cases of *HBB* and *THPO*. The single nucleotide substitutions that distinguish the variant alleles from the wild-type alleles are highlighted in bold and grey. The only known pathological mutation in an miRNA-target site, a G > A transition in the 3′ UTR of the *SLITRK1* gene in two patients with Tourette's syndrome (Abelson et al. 2005), is included for the sake of comparison

Materials and methods

Data source

Of the previously collected 3′ UTR variants, only those that occurred within the sub-region known as the USS (upstream sequence between the translational termination codon and the UCPAS, upstream core polyadenylation signal) were included for analysis (Chen et al. 2006a, b). The variants that occurred within the UCPAS and LAS (left arm of the 'spacer' sequence between the UCPAS and the pre-mRNA cleavage site) of the 3′ UTR were excluded from evaluation for two reasons: (i) their number is very limited and (ii) almost all of these variants are explicable in terms of well-established pathological mechanisms.

Of the 83 previously identified USS variants, four isolated examples (i.e. Group 5 in Chen et al. (2006b)) were excluded owing to their complex nature. Thus, a total of 79 USS variants (viz. Groups 1–4 in Chen et al. (2006b)) were studied here.

Search for miRNA-binding sites spanning the sites of USS variants

Sequence spanning the 79 collated USS variants was searched for all known miRNA binding sites with *miRBase* software (http://microrna.sanger.ac.uk/sequences/search.shtml; Griffiths-Jones et al. 2006) using default parameters. For each variant, both the wild-type 3′ UTR

sequence and its mutated counterpart, each of total length ~50 bp flanking the site of mutation, were screened for the presence of miRNA binding sites, with all possible 25 bp fragments within these flanking sequences being examined sequentially.

Results and discussion

General considerations about miRNA target prediction

Whereas miRNAs have emerged as a new class of regulatory gene (reviewed in Alvarez-Garcia and Miska 2005; Yang et al. 2005; Kim and Nam 2006; Krutzfeldt et al. 2006), miRNA-target sites within the 3′ UTRs of human protein-coding genes constitute a new class of *cis*-acting regulatory elements. Although the mechanism underlying miRNA-target site interactions still remains to be elucidated, it is clear that the initiation of target site recognition often relies on short stretches (6–8 bp) of perfect (and consecutive) Watson–Crick miRNA-mRNA complementarity; this perfect match is typically located at the 5′ end of the miRNA (termed the 'seed' site; Lewis et al. 2003, 2005; Bentwich 2005; Brennecke et al. 2005; Rajewsky 2006).

Secondary structure has also been employed to predict miRNA targets in two recent studies (Robins et al. 2005; Zhao et al. 2005). However, as opined by Rajewsky (2006), ''it is not yet clear if the proposed algorithms help to improve specificity or sensitivity of algorithms that do not take into account 3′ UTR secondary structure''. Rajewsky also raised two problems related to secondary structure prediction viz. the unreliability of the *mfold* program (Zuker et al. 2003) in cases where input length exceeds a given limit, and the question of whether the predicted secondary structure can be extrapolated to the in vivo situation. In this regard, it is pertinent to note that using 'local' rather than 'global' secondary structure prediction, we identified consistent patterns of secondary structural change that potentiated the discrimination of functional USS variants from their non-functional counterparts (Chen et al. 2006b). We should however emphasise that these predictions were made in the context of interactions between mRNAs and their cognate *trans*-acting proteins. *Trans*-acting RNA molecules may well bind/influence a given mRNA secondary structure in rather different ways than *trans*-acting proteins. Further, the free energy of the miRNA/mRNA duplex, which has often been evaluated alongside sequence complementarity, has been found to be a poor predictor of miRNA-target sites (Rajewsky and Socci 2004; Rajewsky 2006).

Consequently, in seeking potential miRNA-target site mutations amongst the 79 known 3′ UTR variants (Chen et al. 2006a, b), we relied entirely upon the criterion of sequence complementarity at the 'seed' site. Under stringent conditions viz. the requirement of at least a 7-nt consecutive match to the first eight nucleotides at the 5′ end of the miRNA (Bentwich 2005; Brennecke et al. 2005), none of the analysed USS variants were found to disrupt or create binding sites for known human miRNAs. However, under more relaxed conditions, viz. the requirement of at least a 5-nt consecutive match to the 'seed' region of the miRNA, four such cases were found (Fig. 1). The observed *E*-values for these 'hits' were found to be >1 but <10 (default parameters); hence the formal probabilities of finding these alignments by chance alone were relatively high (*HBB*, 0.9257; *DPYSL2*, 0.2289; *THPO*, 0.7759; *PTGS2*, 0.8891). The utility of such a statistical analysis should probably not be overstated since, in the case of the Tourette's syndrome-associated mutation located within the hsa-miR-189-binding site in the 3′ UTR of the *SLITRK1* gene (Abelson et al. 2005), we calculate that the corresponding *P*-value for finding this sequence similarity by chance alone is 0.8647.

The secondary structural changes associated with these four variants have been previously predicted by Chen et al. (2006b). Although one (*DPYSL2*), which also had the lowest *P*-value (0.2289) of the four candidate mutations, was predicted to generate a type 1 secondary structural change with unknown functional significance, the remaining three [as well as the G → A mutation in the *SLITRK1* gene (Abelson et al. 2005)] were predicted to generate a pattern I secondary structural change (Chen et al. 2006b). It is nevertheless impossible to draw any firm conclusions at this stage. The four variants in question (Fig. 1) will therefore now be discussed individually in the context of their putative disease-associations and potential biological significance.

HBB: Predicted functional consequence inconsistent with the presumed causative role

The C>G mutation, 6 nucleotides downstream of the translational termination codon of the *HBB* gene (MIM# 141900; encoding *β*-globin), was reported in *β*-thalassaemia intermedia patients (Jankovic et al. 1991; Maragoudaki et al. 1998). The mutant G allele displays decreased complementarity to miRNA hsa-miR-214 as compared with the wild-type C allele

(Fig. 1). Thus, it might be predicted that the mRNA product of the *HBB* G allele would be less subject to miRNA-mediated gene repression. Were this to be true, the *HBB* G variant allele might be expected to be associated with increased expression of the gene as compared with the wild-type C allele; this would however be inconsistent with the variant's putative role in causing β-thalassaemia intermedia. This notwithstanding, it still remains possible that the variant G allele constitutes a miRNA target mutation. However, in this case, any ensuing functional consequences would have been concealed by a concomitant effect on mRNA stability, an effect already demonstrated by in vitro functional analysis (Sgourou et al. 2002).

DPYSL2: Predicted functional consequence partially consistent with the presumed causative role

The 2236T>C polymorphism in the 3′ UTR of the dihydropyrimidinase-like 2 gene (*DPYSL2*; MIM# 602463) was reported to reduce the susceptibility to schizophrenia in a Japanese patient sample (Nakata et al. 2003). Based upon the observation that the expression of DPYSL2 is significantly decreased in the frontal cortex of people with schizophrenia and affective disorder (Johnston-Wilson et al. 2000), were the C allele to be 'disease-protective', it might reasonably be expected to give rise to increased expression of the gene. This would contrast with the prediction that the C allele should be more subject to translational repression by hsa-miR-302b than the T allele (Fig. 1). However, in a North American Caucasian population, the C allele has been reported to increase susceptibility to schizophrenia (Hong et al. 2005). If this later finding were not simply spurious, the predicted functional consequence of the T>C polymorphism would be consistent with its presumed causative role.

THPO and *PTGS2*: Predicted functional consequences consistent with presumed causative roles

If the rs6141(+24)G>A polymorphism in the *THPO* gene were indeed to be associated with a low platelet count (Garner et al. 2005), it should result in the increased expression of the gene because circulating plasma levels of THPO are generally inversely correlated with platelet levels (Garner et al. 2005 and references therein). If the 9850A>G polymorphism in the *PTGS2* gene were indeed to confer colorectal cancer risk (Cox et al. 2004), it should also lead to overexpression of the gene (see Campa et al. 2004 for relevant information). The reduced complementarity of the *THPO* A variant allele to hsa-miR-431 and of the *PTGS2* G variant allele to hsa-miR-132, as compared with their respective wild-type alleles (Fig. 1), is therefore consistent with such functional consequences.

Conclusions

We have taken a set of known disease-associated 3′ UTR variants and performed the first systematic search for potential miRNA-target site mutations. Under fairly stringent conditions that required at least a 7-nt consecutive match to the first eight nucleotides at the 5′ end of the miRNA (Bentwich 2005; Brennecke et al. 2005), no variants were found that either disrupted or created binding sites for known human miRNAs. The 3′ UTR of the *HBB* gene, which displays a 6-nt consecutive match with the seed region of hsa-miR-214, could potentially serve as a functional miRNA target. However, not only does the C>G mutation not occur within the seed region, but its predicted functional consequence is also inconsistent with its presumed causative role, making it unlikely to be biologically significant in the context of miRNA-target interactions. As for the remaining three variants, it appears unlikely that 5-mers would be sufficient to serve as seeds for the initiation of miRNA-target site recognition. In summary, our analysis has suggested that miRNA-target site mutations are not a frequent cause of human genetic disease.

Acknowledgements This work was performed during early 2006 while JMC was a visiting Professor of Genetics supported by the Ministère de la Jeunesse, de l'Éducation Nationale et de la Recherche, France. This work was supported by the INSERM (Institut National de la Santé et de la Recherche Médicale), France.

References

Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, Mathews CA, Pauls DL, Rasin MR, Gunel M, Davis NR, Ercan-Sencicek AG, Guez DH, Spertus JA, Leckman JF, Dure LS 4th, Kurlan R, Singer HS, Gilbert DL, Farhi A, Louvi A, Lifton RP, Sestan N, State MW (2005) Sequence variants in *SLITRK1* are associated with Tourette's syndrome. Science 310:317–320

Alvarez-Garcia I, Miska EA (2005) MicroRNA functions in animal development and human disease. Development 132:4653–4662

Bentwich I (2005) Prediction and validation of microRNAs and their targets. FEBS Lett 579:5904–5910

Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, Sharon E, Spector Y,

Bentwich Z (2005) Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 37:766–770

Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of microRNA-target recognition. PLoS Biol 3:e85

Campa D, Zienolddiny S, Maggini V, Skaug V, Haugen A, Canzian F (2004) Association of a common polymorphism in the cyclooxygenase 2 gene with risk of non-small cell lung cancer. Carcinogenesis 25:229–235

Chen JM, Férec C, Cooper DN (2006a) A systematic analysis of disease-associated variants in the 3′ regulatory regions of human protein-coding genes I: general principles and overview. Hum Genet 120:1–21

Chen JM, Férec C, Cooper DN (2006b) A systematic analysis of disease-associated variants in the 3′ regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3′ UTR variants. Hum Genet 120:301–333

Cox DG, Pontes C, Guino E, Navarro M, Osorio A, Canzian F, Moreno V, Bellvitge Colorectal Cancer Study Group (2004) Polymorphisms in prostaglandin synthase 2/cyclooxygenase 2 (*PTGS2/COX2*) and risk of colorectal cancer. Br J Cancer 91:339–343

Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP (2005) The widespread impact of mammalian microRNAs on mRNA repression and evolution. Science 310:1817–1821

Garner C, Best S, Menzel S, Rooks H, Spector TD, Thein SL (2005) Two candidate genes for low platelet count identified in an Asian Indian kindred by genome-wide linkage analysis: glycoprotein IX and thrombopoietin. Eur J Hum Genet 14:101–108

Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34:D140–D144

Grun D, Wang YL, Langenberger D, Gunsalus KC, Rajewsky N (2005) microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. PLoS Comput Biol 1:e13

Hong LE, Wonodi I, Avila MT, Buchanan RW, McMahon RP, Mitchell BD, Stine OC, Carpenter WT Jr, Thaker GK (2005) Dihydropyrimidinase-related protein 2 (*DRP-2*) gene and association to deficit and nondeficit schizophrenia. Am J Med Genet B Neuropsychiatr Genet 136:8–11

Jankovic L, Dimovski AJ, Kollia P, Karageorga M, Loukopoulos D, Huisman TH (1991) A C → G mutation at nt position 6 3′ to the terminating codon may be the cause of a silent beta-thalassemia. Int J Hematol 54:289–293

John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS (2004) Human microRNA targets. PLoS Biol 2:e363

Johnston-Wilson NL, Sims CD, Hofmann JP, Anderson L, Shore AD, Torrey EF, Yolken RH (2000) Disease-specific alterations in frontal cortex brain proteins in schizophrenia, bipolar disorder, and major depressive disorder. The Stanley Neuropathology Consortium. Mol Psychiatry 5:142–149

Kim VN, Nam JW (2006) Genomics of microRNA. Trends Genet 22:165–173

Kiriakidou M, Nelson PT, Kouranov A, Fitziev P, Bouyioukos C, Mourelatos Z, Hatzigeorgiou A (2004) A combined computational-experimental approach predicts human microRNA targets. Genes Dev 18:1165–1178

Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N (2005) Combinatorial microRNA target predictions. Nat Genet 37:495–500

Krutzfeldt J, Poy MN, Stoffel M (2006) Strategies to determine the biological function of microRNAs. Nat Genet 38(Suppl 1):S14–S19

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. Cell 115:787–798

Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120:15–20

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433:769–773

Maragoudaki E, Vrettou C, Kanavakis E, Traeger-Synodinos J, Metaxotou-Mavrommati A, Kattamis C (1998) Molecular, haematological and clinical studies of a silent beta-gene C → G mutation at 6 bp 3′ to the termination codon (+1480 C → G) in twelve Greek families. Br J Haematol 103:45–51

Nakata K, Ujike H, Sakai A, Takaki M, Imamura T, Tanaka Y, Kuroda S (2003) The human dihydropyrimidinase-related protein 2 gene on chromosome 8p21 is associated with paranoid-type schizophrenia. Biol Psychiatry 53:571–576

Pillai RS (2005) MicroRNA function: multiple mechanisms for a tiny RNA? RNA 11:1753–1761

Rajewsky N (2006) microRNA target predictions in animals. Nat Genet 38(Suppl 1):S8–S13

Rajewsky N, Socci ND (2004) Computational identification of microRNA targets. Dev Biol 267:529–535

Robins H, Press WH (2005) Human microRNAs target a functionally distinct population of genes with AT-rich 3′ UTRs. Proc Natl Acad Sci USA 102:15557–15562

Robins H, Li Y, Padgett RW (2005) Incorporating structure to predict microRNA targets. Proc Natl Acad Sci USA 102:4006–4009

Sgourou A, Papachatzopoulou A, Psiouri L, Antoniou M, Zoumbos N, Gibbs R, Athanassiadou A (2002) The beta-globin C → G mutation at 6 bp 3′ to the termination codon causes beta-thalassaemia by decreasing the mRNA level. Br J Haematol 118:671–676

Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. Genes Dev 20:515–524

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature 434:338–345

Yang M, Li Y, Padgett RW (2005) MicroRNAs: small regulators with a big impact. Cytokine Growth Factor Rev 16:387–393

Zhang L, Huang J, Yang N, Greshock J, Megraw MS, Giannakakis A, Liang S, Naylor TL, Barchetti A, Ward MR, Yao G, Medina A, O'brien-Jenkins A, Katsaros D, Hatzigeorgiou A, Gimotty PA, Weber BL, Coukos G (2006) MicroRNAs exhibit high frequency genomic alterations in human cancer. Proc Natl Acad Sci USA 103:9136–9141

Zhao Y, Samal E, Srivastava D (2005) Serum response factor regulates a muscle-specific microRNA that targets *Hand2* during cardiogenesis. Nature 436:214–220

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415