

In Silico Evidence of the Relationship Between miRNAs and siRNAs

L. Montanucci¹, P. Fariselli^{*1}, P.L. Martelli¹, I. Rossi^{1,2} and R. Casadio¹

¹*Biocomputing Group, Department of Biology, via Imerio 42, University of Bologna, Italy*

²*BioDec Srl, Via Calzavecchio 20/2, I-40033 Casalecchio di Reno (BO), Italy*

Abstract: Both short interfering RNAs (siRNAs) and microRNAs (miRNAs) mediate the repression of specific sequences of mRNA through the RNA interference pathway. In the last years several experiments have supported the hypothesis that siRNAs and miRNAs may be functionally interchangeable, at least in cultured cells. In this work we verify that this hypothesis is also supported by a computational evidence. We show that a method specifically trained to predict the activity of the exogenous siRNAs assigns a high silencing level to experimentally determined human miRNAs. This result not only supports the idea of the biological interchangeability of siRNAs and miRNAs but indicates that it is possible to use computational tools developed using synthetic small interference RNAs to study endogenous miRNAs.

INTRODUCTION

RNA interference (RNAi), first discovered in the nematode *Caenorhabditis elegans* [1], is an evolutionarily conserved endogenous pathway that mediates negative post-transcriptional gene regulation in eukaryotic cells [2-4]. The sequence-specific mediators of this regulatory pathway are small RNA molecules (~23 nt), that bind complementary regions of messenger transcripts and induce highly specific gene silencing [5].

Small RNAs belong to two major classes: microRNAs (miRNAs) and short interfering RNAs (siRNAs). MiRNAs originate from short hairpins (called microRNA precursors) endogenously produced through a multi-step process that starts in the nucleus [6]. On the contrary siRNAs can be either endogenous or exogenous and derive from double stranded RNA molecules (dsRNAs). miRNAs and siRNAs seem to differ only in their biogenesis [7-10].

Long dsRNAs and the miRNA hairpin-shaped precursors are initially processed by the cytoplasmic ribonuclease III-like enzyme Dicer which cleaves them producing dsRNAs of about 22 bp long with characteristic 2nt overhangs at the 3' terminus. Then they are incorporated, as single stranded RNA, into a ribonucleoprotein complex named RNA-induced silencing complex (RISC). RISC identifies target genes through an antisense complementarity between the mi/siRNA and the mRNA. When RISC binds the target messenger, mRNA can be inhibited and later reprocessed or degraded, depending on the degree of pair-matching [11, 12].

To study the effect of siRNAs on the gene expression, synthetic siRNAs are generally synthesized with perfect base complementarity to their target. Notwithstanding the perfect complementarity, different siRNAs have different knock-down efficacies and it is not completely known what determines the high effectiveness and specificity of some molecules or the weak functionality of others [13].

Similarly to siRNAs, plant microRNA were recognized to function through near-perfect base pairing with their targets [14-16]. On the contrary it was discovered that animal miRNAs/mRNA duplexes display only a partial base pair complementarity, since they can form complex secondary structures [17]. In particular a short sequence of 6 to 8 consecutive bases, located at the 5' of the mature miRNA, was found to be determinant in animal target recognition [18-21].

Summarizing, while animal miRNAs carry out their function through a partial complementarity, by construction synthetic siRNAs hybridize with perfect base pair matching to target mRNAs.

Many computational analyses have been carried and predictors based on different features have been developed to infer siRNAs effectiveness. Some were based on the siRNA sequence alone [22], others added thermodynamic properties [23-25], secondary structure information [26-28], compositional biases [29] or position specific properties [30-33].

Despite much knowledge has been achieved in the last few years, many aspects of the silencing mechanism are still not fully understood, including the reasons for the high specificity of the silencing mechanism and the exact relationship between small interference and micro RNAs functioning [14, 17, 34].

Recently, Huesken and coworkers [35] made available a conspicuous experimental data set on which they implemented a powerful predictor of siRNA activity. Using these data we show that a method specifically trained to predict the activity of synthetic siRNA, assigns a high silencing level to experimentally determined human miRNAs. This result not only supports the idea of siRNAs and miRNAs equivalence but indicates that it is possible to use computational tools developed using synthetic small interference RNAs to investigate endogenous miRNAs.

MATERIALS AND METHODOLOGY

Databases

The data set of siRNAs is taken from Novartis [35] and consists of 2431 siRNA sequences whose inhibitory activi-

*Address correspondence to this author at the Biocomputing Group, Department of Biology, via Imerio 42, University of Bologna, Italy; E-mail: piero.fariselli@unibo.it

ties had been experimentally determined through a high-throughput reporter assay. The distribution of the values of the experimentally measured siRNA inhibitory activities is bell-shaped and ranges from 0 to 1.36. We compute our results using both the training/testing division indicated in Huesken *et al.* 2005 [35] (the same 2182 siRNA sequences were used for the training and the same 249 siRNAs were used as a test set) and a ten-fold cross-validation procedure. In terms of the final accuracy we did not notice any significant difference between the two different procedures. Furthermore, two databases of miRNAs were selected to investigate the siRNA/miRNA similarity. The first one consists of miRNAs extracted from TarBase that is a database of experimentally supported animal microRNA targets [36]. From TarBase version 3 (TarBase_V3), we kept the human miRNAs for which the sequences are reported. This set is composed of 36 different miRNAs.

A second dataset was derived from miRBase release 9.0 [37]. From miRBase we have filtered all human mature miRNAs whose evidence was reported as “experimental”. Mature sequences whose length was less than 21 nucleotides were removed from the set. The filtered test set contains 349 miRNAs.

To better highlight the predicted silencing activities of the miRNAs described above, we built a benchmark set, that in principle should contain both good and bad silencing candidates for their corresponding genes. The benchmark set consists of 1000 randomly-selected sequences of 23nt-long extracted from 1000 different human cDNAs.

The Predictor

Our predictor is based on a feed-forward neural network with standard error back propagation training algorithm. The dimension of the input window depends on the encoded information. Since in Huesken *et al.* 2005 [35] the authors designed siRNA sequences of 19 base pairs with two added nucleotides, we tested our method accordingly by considering both cases as input (namely, with 19 or 21 bases). We then added to the encoded sequence the possible nearest neighbor di-peptides and triplets as proposed before for other approaches [38]. This leads to different neural network architectures that vary for the numbers of the input nodes. In particular, if N is the number of the nucleotides (the first 19 or all the 21) we defined the following three types of input:

- Sequence encoding = $N*4$ input nodes.
- Sequence encoding + Dipeptides = $N*4 + 4*4$ input nodes.
- Sequence encoding + Dipeptides + Tripeptides = $N*4 + 4*4 + 4*4*4$ input nodes.

The largest input size consists of 164 neurons when we consider the Sequence encoding + Dipeptides + Tripeptides and $N=21$. All the trained networks have 8 hidden neurons and one output node that codes for the inhibition activity. The neural networks were trained to predict the values of the experimental inhibition activities of the 2431 siRNAs reported in Huesken *et al.* 2005 [35].

The evaluation of the performance of the prediction methods was carried using the Pearson correlation coefficient, defined as:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2 \right]^{1/2}}$$

RESULTS

Tuning the Predictor

In order to evaluate the accuracy of the different neural networks we adopted a 10-fold cross validation procedure. We then split the data set in 10 disjoint subsets and we evaluated the accuracy summing up the results obtained for the ten different testing sets. Other types of splitting such as the original division in two sets previously adopted [35] did not change the results. The results are reported in Table 1, where it is evident that an encoding that takes into account all the 21 siRNA nucleotides provides a better accuracy. Furthermore, as previously noticed with other methods, the inclusion of dipeptide and tripeptides information helps in improving the correlation with the experimental data.

The neural networks previously designed by the authors of the experimental data set [35] used only the sequence information and the reported Pearson correlation coefficient for the test set was 0.66, which is similar to the one obtained with our cross-validation procedure using the same encoding (0.65 in Table 1).

The predictions on the data set using our best performing method are also plotted in Fig. (1).

Table 1. Pearson Accuracy of on the Huesken DB

	Pearson Correlation Coefficient		
	Seq.	Seq+Dipep	Seq+Dipep+Tripep
19-nt	0.63	0.64	0.65
21-nt	0.65	0.66	0.68

siRNA and miRNA Comparison

Using our best performing predictor of Table 1 we tested the hypothesis of chemical-functional equivalence between siRNAs and endogenous miRNAs *in silico*. In particular, even if the computational method was trained to identify active siRNAs, it assigns high scores also to the endogenous miRNAs. However, the two types of small RNAs maintain a subtle difference: the designed siRNAs have a perfect base-matching with the corresponding mRNAs, while the miRNAs extracted from the data have only a partial base pairing.

Nonetheless, the results reported in Fig. (2), confirm the hypothesis, since the neural-network predicts high inhibition activities for most of the experimental miRNAs. This is true for both human microRNAs derived from TarBase and miRBase databases (Fig. 2). In Fig. (2), where we report the distribution of the predicted efficacies for miRNAs, it is evident that the vast majority of miRNAs are predicted with an inhibitory level higher than 0.7. Moreover, the random selection of the human cDNA sequences behaves as expected, showing a normal-like distribution of the predicted activity with median value around 0.7 (Fig. 2). It is worth noticing

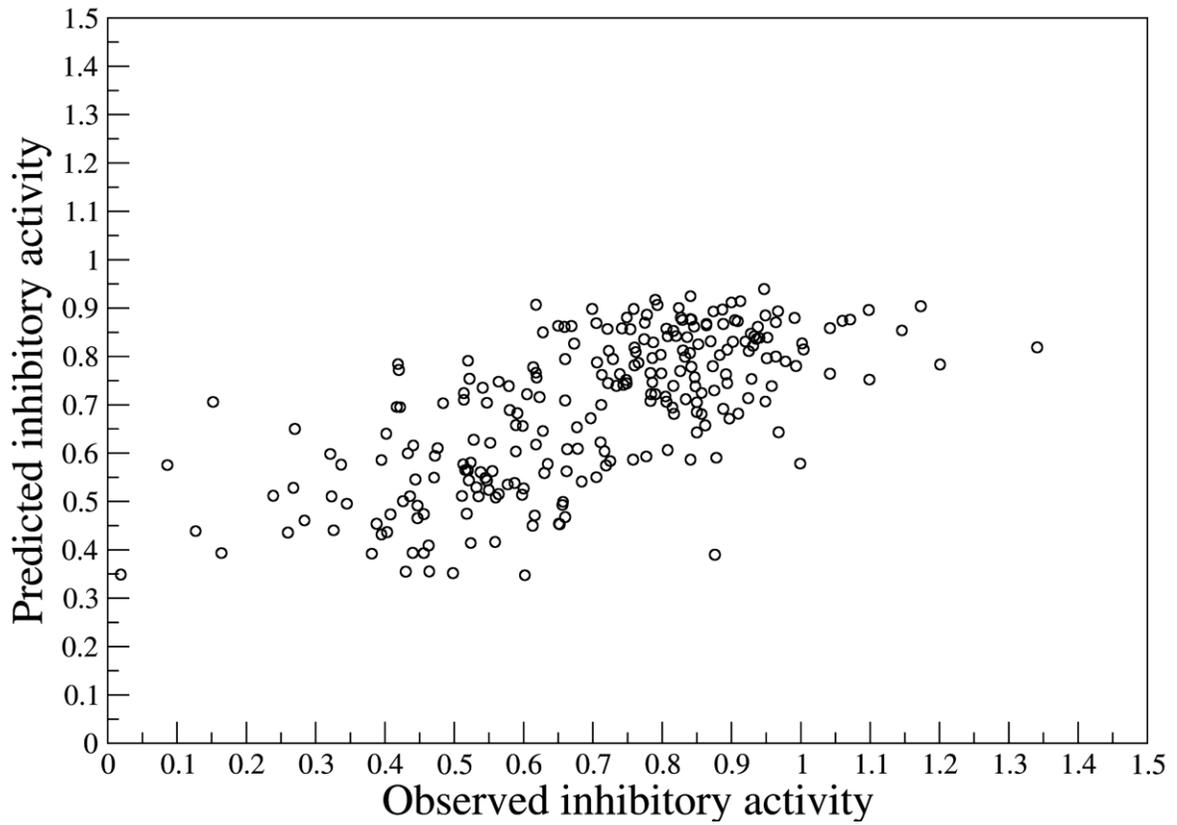


Fig. (1). Correlation between observed and predicted inhibitory activities for the test set of 249 siRNAs. The Pearson correlation value is 0.68.

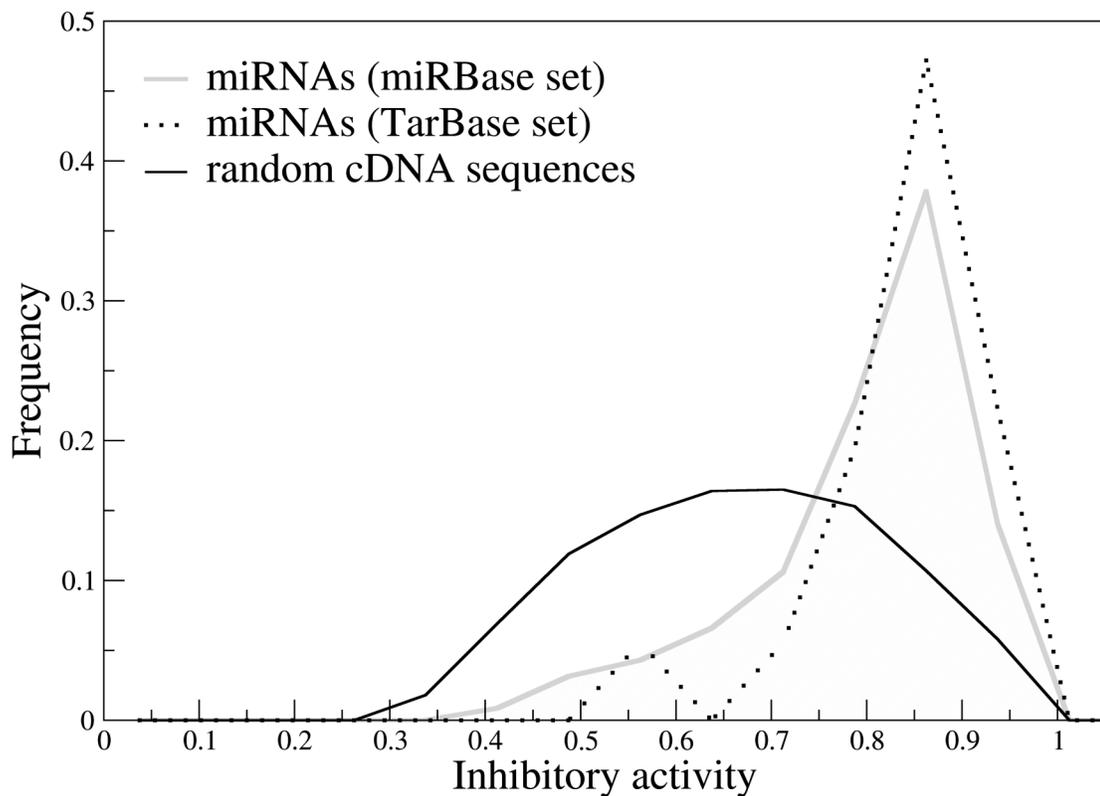


Fig. (2). Distribution of predicted inhibitory activities for the two miRNA data sets of (from TarBase: dotted line; from miRBase: gray line) compared to the random selection of cDNA candidates (black solid line).

that the distribution of the inhibitory activity values of the siRNAs experimentally obtained [35] has a similar shape and the same median value. This findings not only support the idea of siRNAs and miRNAs equivalence but indicate that it is possible to adopt computational tools developed using synthetic small interference RNAs to investigate endogenous miRNAs.

CONCLUSION

In this paper we developed a new method for predicting the inhibitory activity of small interference RNAs starting from their sequence. Our results indicate that an input encoding that takes into account di-peptides and tri-peptides information achieves higher accuracy with respect to simpler input encoding based only on the sequence representation [35]. Furthermore, with a computational approach we showed that siRNAs and miRNAs can be functionally interchangeable. Although several experimental evidences support this hypothesis, to the best of our knowledge this is the first time that a computational method confirms the relationships between siRNAs and miRNAs.

ACKNOWLEDGEMENTS

This work was supported by the following grants: LIBI-Laboratorio Internazionale di BioInformatica and the Biosapiens Network of Excellence project within the European Union's VI Framework Programme (contract number LSGH-CT-2003-503265), delivered to RC. IR acknowledges the EU FP6 Specific Targeted Research Project TargetHerpes (LSHG-CT-2006-037517) for support.

REFERENCES

- [1] A. Fire, S. Xu, M. K. Montgomery, S. A. Kosta, S. E. Driver, and C. Mello, "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*", *Nature*, vol. 391, pp. 806-811, 1998.
- [2] G. J. Hannon, "RNA interference", *Nature*, vol. 418, pp. 244-250, 2002.
- [3] G. J. Hannon, and J. J. Rossi, "Unlocking the potential of the human genome with RNA interference", *Nature*, vol. 431, pp. 371-378, 2004.
- [4] C. Mello, and D. Conte, "Revealing the world of RNA interference", *Nature*, vol. 431, pp. 338-342, 2004.
- [5] D. P. Bartel, "MicroRNAs: Genomics, Biogenesis, Mechanism and Function", *Cell*, vol. 116, pp. 281-297, 2004.
- [6] M. A. Carmell, and G. J. Hannon, "RNase III enzymes and the initiation of gene silencing", *Nat. Struct. Mol. Biol.*, vol. 11, pp. 214-218, 2004.
- [7] S. M. Elbashir, W. Lendeckel, and T. Tuschl, "RNA interference is mediated by a 21- 22-nucleotide RNAs", *Gene Dev.*, vol. 15, pp. 88-200, 2001.
- [8] V. Ambros, B. Bartel, D. P. Bartel, et al., "A uniform system for microRNA annotation", *RNA*, vol. 9, no. 3, pp. 277-279, March 2003.
- [9] Y. Zeng, R. Yi, and B. R. Cullen, "MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms", *Proc. Natl. Acad. Sci.*, vol. 100, no. 17, pp. 9779-9784, August 2003.
- [10] J. G. Doench, C. P. Petersen, and P. A. Sharp, "siRNAs can function as miRNAs", *Gene Dev.*, vol. 17, pp. 438-42, 2003.
- [11] M. A. Valencia-Sanchez, J. Liu, G. J. Hannon, and R. Parker, "Control of translation and mRNA degradation by miRNAs and siRNAs", *Gene Dev.*, vol. 20, pp. 515-524, July 2006.
- [12] T. M. Rana, "Illuminating the silence: understanding the structure and function of small RNAs", *Nat. Rev. Mol. Cell. Biol.*, vol. 8, no. 1, pp. 23-36, January 2007.
- [13] A. Reynolds, D. Leake, Q. Boese, S. Scaringe, W.S. Marshall, and A. Khvorova, "Rational siRNA design for RNA interference", *Nat. Biotechnol.*, vol. 22, pp. 326-330, 2004.
- [14] M. W. Rhoades, B. J. Reinhart, L. P. Lim, C. B. Burge, B. Bartel, and D. P. Bartel, "Prediction of Plant MicroRNA Targets", *Cell*, vol. 110, no. 4, pp. 513-520, August 2002.
- [15] C. Llave, Z. Xie, K. D. Kasschau, and J. C. Carrington, "Cleavage of Scarecrow-like mRNA Targets Directed by a Class of Arabidopsis miRNA", *Science*, vol. 297, no. 5589, pp.2053-2056, September 2002.
- [16] G. Tang, B. J. Reinhart, D. P. Bartel, and P. D. Zamore, "A biochemical framework for RNA silencing in plants" *Gene Dev.*, vol. 17, no. 1 pp. 49-63, January 2003.
- [17] M. Rehmsmeier, P. Steffen, M. Höchsmann, and R. Giergerich, "Fast and effective prediction of microRNA/target duplexes", *Bioinform.*, vol. 10, no. 10, pp. 1507-1517, October 2004.
- [18] B. P. Lewis, I. Shih, M. W. Jones-Rhoades, D. P. Bartel, and C. B. Burge, "Prediction of Mammalian MicroRNA Targets", *Cell*, vol. 115, no. 7, pp. 787-789, December 2003.
- [19] B. P. Lewis, C. B. Burge, and D. P. Bartel, "Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets", *Cell*, vol. 120, no. 1, pp.15-20, January 2005.
- [20] J. Brennecke, A. Stark, R. B. Russel, and S. M. Cohen, "Principles of MicroRNA-Target Recognition", *PLoS Biol.*, vol. 3, no. 3, pp. e85-e85, March 2005.
- [21] K. K. Farh, A. Grimson, C. Jan, et al., "The widespread impact of mammalian MicroRNAs on mRNA repression and evolution", *Science*, vol. 310, no. 5755, pp. 1817-1821, December 2005.
- [22] P. Sætrom and O. Snøve, "A comparison of siRNA efficacy predictors", *Biochem. Biophys. Res. Commun.*, vol. 321, no. 1, pp. 247-253, 2004.
- [23] A. Khvorova, A. Reynolds and S. D. Jayasena, "Functional siRNAs exhibit strand bias", *Cell*, vol. 115, no. 2, pp. 209-216, 2003.
- [24] O. Matveeva, Y. Nechipurenko, L. Rossi, et al., "Comparison of approaches for rational siRNA design leading to a new efficient and transparent method", *Nucleic Acids Res.*, vol. 35, no. 8, pp. e63, January 2007.
- [25] A. M. Chalk, C. Wahlestedt, and E. L. Sonnhammer, "Improved and automated prediction of effective siRNA", *Biochem. Biophys. Res. Commun.*, vol. 319, no. 1, pp. 264-74, June 2004.
- [26] K. Q. Luo and D. C. Chang, "The gene-silencing efficiency of siRNA is strongly dependent on the local structure of mRNA at the targeted region", *Biochem. Biophys. Res. Commun.*, vol. 318, no. 1, pp. 303-310, May 2004.
- [27] S. M. Yiu, P. W. Wong, T. W. Lam, et al., "Filtering of ineffective siRNAs and improved siRNA design tool", *Bioinformatics*, vol. 21, no. 2, pp. 144-51, January 2005.
- [28] B. S. E. Heale, H. S. Soifer, C. Bowers, and J. J. Rossi, "SiRNA target site secondary structure predictions using local stable substructures", *Nucleic Acids Res.*, vol. 33, pp. e30-e30, 2005.
- [29] A. Henschel, F. Buchholz, and B. Habermann, "DEQOR: a web-based tool for the design and quality control of siRNAs", *Nucleic Acids Res.*, vol. 32, pp. W113-120, July 2004.
- [30] A. C. Hsieh, R. Bo, J. Manola, et al., "A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens", *Nucleic Acids Res.*, vol. 32, no. 3, pp. 893-901, February 2004.
- [31] K. Takayuki and T. Suzuki, "Specific residues at every third position of siRNA shape its efficient RNAi activity", *Nucleic Acids Res.*, vol. 35, no. 4, pp. e27-e27, February 2007.
- [32] W. Gong, Y. Ren, Q. Xu, et al., "Integrated siRNA design based on surveying of features associated with high RNAi effectiveness", *BMC Bioinform.*, vol. 7, pp. 516, November 2006.
- [33] A. S. Peek, "Improving model predictions for RNA interference activities that use support vector machine regression by combining and filtering features", *BMC Bioinform.*, vol. 8, pp. 182, June 2007.

- [34] Y. Zeng, R. Yi, and B. R. Cullen, "MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms" *Proc. Natl. Acad. Sci.*, vol. 100, no. 17, pp. 9779-9784, August 2003.
- [35] D. Huesken, J. Lange, C. Mikanin, *et al.*, "Design of a genome-wide siRNA library using an artificial neural network" *Nat. Biotechnol.*, vol. 23, no. 8, pp. 995-1001, July 2005.
- [36] P. Sethupathy, B. Corda, and A.G. Hatzigeorgiou, "TarBase: A comprehensive database of experimentally supported animal microRNA targets", *RNA*, vol. 12, no. 2, pp. 192-197, February 2006.
- [37] S. Griffiths-Jones, J. Grocock, S. Van Dongen, A. Bateman, and A. J. Enright, "miRBase: microRNA sequences, targets and gene nomenclature", *Nucleic Acids Res.*, vol. 34, pp. D140-144, January 2006.
- [38] J. P. Vert, N. Foveau, C. Lajaunie, and Y. Vandenbrouck, "An accurate and interpretable model for siRNA efficacy prediction", *BMC Bioinform.*, vol. 7, no. 520, November 2006.

Received: February 25, 2007

Revised: April 2, 2008

Accepted: April 5, 2008

© Montanucci *et al.*; Licensee *Bentham Open*.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.5/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.