Genomics 97 (2011) 71–76

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

OMiR: Identification of associations between OMIM diseases and microRNAs

Simona Rossi [a,1], Aristotelis Tsirigos [b,1], Antonio Amoroso [c], Nicoletta Mascellani [d], Isidore Rigoutsos [d,*], George A. Calin [a,*], Stefano Volinia [e,f,*]

[a] Dept. of Experimental Therapeutics, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, USA
[b] Bioinformatics & Pattern Discovery Group, Computational Biology Center, Deep Computing Institute, IBM Thomas J Watson Research Center, NY, 10598, USA
[c] Dept. of Genetics, Biology and Biochemistry, University of Torino, Via Santena 19, 10126 Torino, Italy
[d] Dept. of Pathology, Anatomy and Cell Biology, Thomas Jefferson University, 1020 Locust Street, PA, USA
[e] Dept. of Morphology and Embryology, University of Ferrara, Via Fossato di Mortara 64b, 44100 Ferrara, Italy
[f] Comprehensive Cancer Center, Ohio State University, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 24 June 2010
Accepted 16 October 2010
Available online 23 October 2010

Keywords:
miRNAs
OMIM diseases

ABSTRACT

A large number of loci for genetic diseases have been mapped on the human genome and a group of hereditary diseases among them have thus far proven unsuccessful to clone. It is conceivable that such "unclonable" diseases are not linked to abnormalities of protein coding genes (PCGs), but of non-coding RNAs (ncRNAs). We developed a novel approach termed OMiR (OMIM and miRNAs), to test whether microRNAs (miRNAs) exhibit any associations with mapped genetic diseases not yet associated with a PCG. We found that "orphan" genetic disease loci were proximal to miRNA loci more frequently than to loci for which the responsible protein coding gene is known, thus suggesting that miRNAs might be the elusive culprits. Our findings indicate that inclusion of miRNAs among the candidate genes to be considered could assist geneticists in their hunt for disease genes, particularly in the case of rare diseases.

1. Introduction

The Online Mendelian Inheritance in Man database (OMIM, http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM) contains information, including responsible PCGs, phenotypes, linkage data and references, on all human inherited or heritable disorders. Considered a phenotypic companion to the human genome project, and as of August 2010, OMIM comprised 4613 Mendelian diseases (which are prefixed using '#' (n = 2837) if the responsible gene is known and with '%' otherwise (n = 1776)), 367 genes with an associated phenotype (prefixed using '+'), and 13173 genes (including 183 microRNAs) with known sequence (prefixed using '*').

MicroRNAs (miRNAs) are small RNA molecu'les typically between 19 and 22 nucleotides in length that act as negative regulators of gene expression and are produced from larger primary transcripts (pri-miRNAs) [1]. In animal cells, miRNAs bind the RNA-induced silencing complex (RISC) and specific target messenger RNAs (mRNAs), which either remain untranslated or are degraded [2]. Many investigators have reported that miRNAs are critical in stem cell function [3] and tissue development [4], are involved in viral infections [5] and are associated with oncogenesis [6–9]. Furthermore, we previously reported that miRNAs are frequently located at human Cancer Associated Genomic Regions (CAGRs), such as amplified or deleted regions [10] or mouse modifier loci [11]. Mattick and colleagues recently reported that orthologous miRNA genes are located in CAGRs in human and mouse [12]. However, with the exception of X-linked mental retardation [13], miRNAs have not been thought of as possible causal of Mendelian diseases. To test the hypothesis that some miRNAs could be potentially responsible for a number of "orphan" Mendelian diseases, we developed the OMiR method and used it to conduct a bioinformatics study.

2. Methods

2.1. OMIM Morbid Map

The Morbid Map is an alphabetical list of diseases described in OMIM with their corresponding cytogenetic locations and related mapping studies and references (http://www.ncbi.nlm.nih.gov/Omim/getmorbid.cgi). We used Morbid Map to extract OMIM identifications (IDs) and corresponding disease names and loci. Information for

* Corresponding authors. I. Rigoutsos is to be contacted at Dept. of Pathology, Anatomy and Cell Biology, Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA 19107, USA. Fax: +1 215 955 5058. G.A. Calin, Department of Experimental Therapeutics, Unit 0036, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA. Fax: +1 731 745 4528. S. Volinia, Department of Morphology and Embryology, University of Ferrara, Via Fossato di Mortara 64b, 44100 Ferrara, Italy.
E-mail addresses: Isidore.Rigoutsos@jefferson.edu (I. Rigoutsos), gcalin@mdanderson.org (G.A. Calin), s.volinia@unife.it (S. Volinia).
[1] These authors contributed equally.

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.10.004

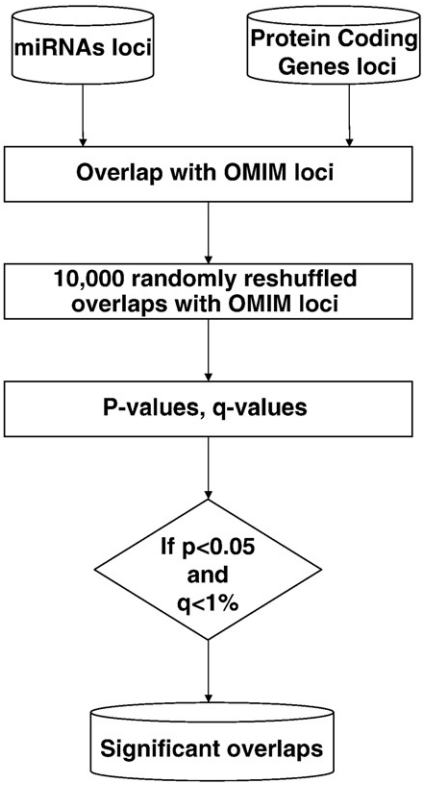

Fig. 1. The five steps of OMiR algorithm: 1) OMIM ID information extraction and conversion; 2) Overlap between miRNAs or Protein Coding Genes loci and OMIM loci; 3) 10,000 randomly reshuffled overlaps, 4) P-value and Q value calculation; and 5) Extraction of the significant overlaps.

diseases labeled using the OMIM IDs and prefixes was converted from cytogenetic location to base position using the UCSC Table Browser (available at http://genome.ucsc.edu/cgi-bin/hgTables, March 2006 release). We focused on: a) entries with a confirmed Mendelian phenotype or phenotypic locus for which the underlying molecular basis is not known (the diseases are prefixed with '%') and b) descriptive entries, usually of a phenotype with known molecular basis that does not represent a unique locus (prefixed with '#'). For statistical purposes, we considered as unknowns the OMIM diseases with a '%' prefix and as controls the '#' prefixed OMIM diseases because the genes responsible for them are known (and are not miRNAs). We examined also the OMIM diseases with prefix '*' (a gene of known sequence) and '+' (a known gene and corresponding phenotype) as controls. If an OMIM ID was associated to more than one locus, each locus was tested separately.

2.2. Extraction and management of miRNA information

The names, chromosome location, start and end positions of 939 miRNAs were downloaded from miRBase release 15.0 (ftp://mirbase.org/pub/mirbase/CURRENT/genomes/hsa.gff). Because each miRNA maps to a much smaller region (about 100 nt) than that mapped to the disease (millions of bp), if two or more miRNAs are significantly associated with an OMIM locus, then they are considered (a posteriori) as a relevant cluster.

2.3. Identifying significant associations of microRNAs and PCGs with OMIM disease loci

For a given OMIM locus, we computed its overlap (i.e. the number of common nucleotides between the miRNAs and the OMIM locus) with the loci of known miRNA transcripts. The probability (p-value) of the observed overlap was computed given the empirical null distribution of overlaps, which was estimated using 10,000 randomly reshuffled versions of the miRNA loci on each chromosome of the genome: each reshuffled set of loci yielded an overlap value, and all these overlaps comprise the empirical null distribution of overlaps. Finally, the false discovery rate (FDR) was calculated by comparing the distribution of p-values against the empirical null distribution of expected p-values, which was estimated using the same 10,000 randomly reshuffled miRNA loci. As before, each reshuffled version yields an overlap value, whose p-value can be computed using the entire set of 10,000 overlap values. All these p-values comprise the empirical null distribution of p-values. This method has been applied to miRNAs and PCGs' transcript loci independently in order to compare the relative number of statistically significant associations obtained. Human ENSEMBL release 54 was used to obtain the PCG loci and miRBase 15.0 to obtain miRNAs loci.

2.4. MiRNA target analysis

A miRNA target analysis was performed to determine if the identified associations could reveal whether the functions of mRNA targets could be related to the associated disease. Using the David Bioinformatics Resources system (http://david.abcc.ncifcrf.gov), we compared the list of terms related to the predicted targeted mRNAs. The terms were evaluated by Bonferroni p-value correction ($P<0.05$). We evaluated Gene Ontology (biological process, molecular function and cellular component, http://www.geneontology.org/), INTERPRO (http://www.ebi.ac.uk/interpro/), PATHWAY (http://www.biocarta.com/ and http://www.genome.jp/kegg/), UP_TISSUE (http://david.abcc.ncifcrf.gov/) and SP_PIR_KEYWORDS (functional category) and OMIM disease (http://www.ncbi.nlm.nih.gov/omim/). The predicted targeted mRNAs were retrieved using miRGen (http://www.diana.pcbi.upenn.edu/data/

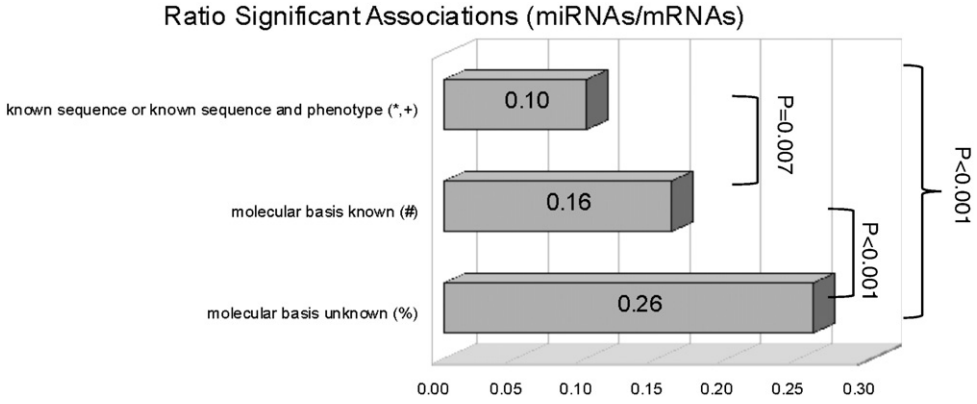

Fig. 2. Ratios between the percentage of significant OMIM::miRNAs associations and OMIM::mRNAs associations, for each OMIM ID type studied: '%', molecular basis unknown, '#' molecular basis known and '*' and '+', known sequence and/or phenotype.

Table 1
Six significant association between % OMIM IDs and miRNAs.

OMIM-ID	Description and locus	Expected genes	Associated genes	P-value	Q-value	miRNA symbols	RNA22 and miRGen target terms (David, Bonferroni P<0.05)
%609572	Photoparoxysmal response 2; PPR2; photoparoxysmal response with or without idiopathic generalized epilepsy; gene map locus 13q31.3;	0.92	7	0	0	mir-17 mir-18a mir-19a mir-19b-1 mir-20a mir-622 mir-92a-1	GO: cell death, apoptosis, cell differentiation, cell development, regulation of cellular metabolic process, transmembrane receptor protein tyrosine kinase signaling pathway, anatomical structure morphogenesis, regulation of gene expression, regulation of transcription from RNA; INTERPRO: serine/threonine protein kinase; SP_PIR_KEYWORDS: serine/threonine-protein kinase; TISSUE: brain, fetal brain.
%300509	Dyslexia, susceptibility to, 9; DYX9; gene map locus Xq27.3; DXS8043	2.88	21	<0.0001	0	mir-506 mir-507 mir-508 mir-509-1 mir-509-2 mir-509-3 mir-510 mir-513a-1 mir-513a-2 mir-513b mir-513c mir-514-1 mir-514-2 mir-514-3 mir-888 mir-890 mir-891a mir-891b mir-892a mir-892b	GO: developmental process, anatomical structure development, multicellular organismal development, system development, nervous system development; TISSUE: brain.
%300244	Terminal osseous dysplasia and pigmentary defects; TODPD; gene map locus Xq27.3-q28	7.47	34	<0.0001	0	mir-105-1 mir-105-2 mir-1184 mir-224 mir-452 mir-506 mir-507 mir-508 mir-509-1 mir-509-2 mir-509-3 mir-510 mir-513a-1 mir-513a-2 mir-513b mir-513c mir-514-1 mir-514-2 mir-514-3 mir-767 mir-888 mir-890 mir-891a mir-891b mir-892a mir-892b	GO: developmental process, anatomical structure development, multicellular organismal development, nervous system development, cellular developmental process, cell differentiation, anatomical structure morphogenesis; UP_SEQ_FEATURE: mutagenesis site; TISSUE: brain, epithelium, fetal brain.
%300388	Polymicrogyria, Bilateral Perisylvian; BPP; gene map locus Xq27.2-q28; DXS8103 and Xqter	8.52	35	<0.0001	0	mir-105-1 mir-105-2 mir-1184 mir-224 mir-452 mir-506 mir-507 mir-508 mir-509-1 mir-509-2 mir-509-3 mir-510 mir-513a-1 mir-513a-2 mir-513b mir-513c mir-514-1 mir-514-2 mir-514-3 mir-767 mir-888 mir-890 mir-891a mir-891b mir-892a mir-892b	GO: cellular component organization and biogenesis, developmental process, anatomical structure development, regulation of biological process, system development, multicellular organismal development, nervous system development, cellular developmental process, cell differentiation, anatomical structure morphogenesis, cell development, synapse; INTERPRO: SRC homology-3, Pleckstrin-like, Pleckstrin homology-type, serine/yhreonine protein kinase; KEGG_PATHWAY: calcium signaling pathway; SP_PIR_KEYWORDS: disease mutation, proto-oncogene, serine/threonine-protein kinase, cytoskeleton, magnesium; UP_SEQ_FEATURE: mutagenesis site; TISSUE: brain, epithelium, fetal brain, melanoma.
%608423	Muscular dystrophy, limb-girdle, type 1F; LGMD1F; gene map locus 7q32.1-q32.2; D7S2544	0.99	6	0	0.01	mir-129-1 mir-182 mir-183 mir-335 mir-593 mir-96	GO: anatomical structure development, developmental process, system development, multicellular organismal development, anatomical structure morphogenesis, organ development, nervous system development, organ morphogenesis; INTERPRO: plekstrin homology-type; SO_PIR_KEYWORDS: cytoskeleton; TISSUE: brain, muscle.
%300614	Deafness, X-linked 5; DFNX5; gene map locus Xq23-q27.3; DXS1220 and DXS8084	22.03	48	<0.0001	0	mir-106a mir-1264 mir-1277 mir-1298 mir-18b mir-1911 mir-1912 mir-19b-2 mir-20b mir-220a mir-320d-2 mir-363 mir-424 mir-448 mir-450a-1 mir-450a-2 mir-450b mir-503 mir-504 mir-505 mir-506 mir-507 mir-508 mir-509-1 mir-509-2 mir-509-3 mir-510 mir-513a-1 mir-513a-2 mir-513b mir-513c mir-514-1 mir-514-2 mir-514-3 mir-542 mir-766 mir-888 mir-890 mir-891a mir-891b mir-892a mir-892b mir-92a-2 mir-934	GO: developmental process, anatomical structure development, system development, multicellular organismal development, cellular developmental process, organ development, cell death, nervous system development, anatomical structure morphogenesis, programmed cell death, apoptosis, regulation of programmed cell death, regulation of apoptosis, regulation of gene expression, dephosphorylation; INTERPRO: serine/threonine protein kinase, serine/threonine protein kinase, active site; SP_PIR_KEYWORDS: disease mutation; Tissue: fetal brain.

(continued on next page)

Table 1 (*continued*)

OMIM-ID	Description and locus	Expected genes	Associated genes	P-value	Q-value	miRNA symbols	RNA22 and miRGen target terms (David, Bonferroni P<0.05)
%138900	Uric acid concentration, serum, quantitative trait locus 1; UAQTL1; gout susceptibility 1; gene map locus 4q25; D4S2623	1.14	8	0.0001	0	mir-297 mir-302a mir-302b mir-302c mir-302d mir-367 mir-576	
%152430	Longevity 1; gene map locus 4q25; D4S1564	1.14	8	0.0001	0	mir-297 mir-302a mir-302b mir-302c mir-302d mir-367 mir-576	
%611494	Atrial fibrillation, familial, 5; ATFB5; gene map locus 4q25; rs2200733; rs2200733 and rs10033464	1.14	8	0.0001	0	mir-297 mir-302a mir-302b mir-302c mir-302d mir-367 mir-576	
%610840	Mitral valve prolapse, myxomatous 3; MMVP3; gene map locus 13q31.3-q32.1; D13S132	1.5	7	0	0	mir-17 mir-18a mir-19a mir-19b-1 mir-20a mir-622 mir-92a-1	GO: cell death, apoptosis, cell differentiation, cell development, regulation of cellular metabolic process, transmembrane receptor protein tyrosine kinase signaling pathway, anatomical structure morphogenesis, regulation of gene expression, regulation of transcription from RNA; INTERPRO: serine/ threonine protein kinase; SP_PIR_KEYWORDS: disease mutation, serine/threonine-protein kinase, magnesium, transcription regulation; TISSUE: brain, eye, fetal brain.

public/PublicData/miRGen/v3/Targets/) and RNA22 (http://cbcsrv.watson.ibm.com/rna22.html) with the following parameters: minimum number of paired-up bases in heteroduplex = 12, maximum folding energy for heteroduplex = −25 kcal/mol, maximum number of allowed UN-paired bases = 0 in seed/nucleus of 6 nucleotides.

3. Results

OMiR was implemented in perl and C++. First, we converted the OMIM, miRNAs and PCGs information from cytogenetic regions in base positions corresponding to various locations of OMIM diseases

Table 2
RNA22 folding energy between genes related to %138900, %606460 and %611494 and OmiR associated miRNA clusters.

mRNA	miRNA	Offset	Folding energy (kcal/mol)
gi|146260266|ref|NM_001455.3| FOXO3	miR-297	3797 to 3817	−27.00
gi|146260266|ref|NM_001455.3| FOXO3	miR-297	6785 to 6805	−26.50
gi|146260269|ref|NM_201559.2| FOXO3	miR-297	3770 to 3790	−27.00
gi|146260269|ref|NM_201559.2| FOXO3	miR-297	6758 to 6778	−26.50
gi|152963645|ref|NM_000325.5| PITX2 3	miR-297	1831 to 1851	−26.40
gi|24234707|ref|NM_153426.1| PITX2 2	miR-297	1754 to 1774	−26.40
gi|24234710|ref|NM_153427.1| PITX2 1	miR-297	1616 to 1636	−26.40
gi|27894305|ref|NM_000576.2| IL1B	miR-297	427 to 447	−30.00
gi|27894331|ref|NM_000877.2| IL1R1	miR-297	672 to 692	−25.20
gi|27894331|ref|NM_000877.2| IL1R1	miR-297	4830 to 4850	−29.80
gi|27894331|ref|NM_000877.2| IL1R1	miR-297	1036 to 1056	−26.50
gi|27894331|ref|NM_000877.2| IL1R1	miR-297	4224 to 4244	−25.50
gi|73622112|ref|NM_001223.3| CASP1	miR-297	223 to 243	−25.10
gi|73622114|ref|NM_033292.2| CASP1	miR-297	223 to 243	−25.10
gi|146260266|ref|NM_001455.3| FOXO3	miR-302a	5946 to 5968	−25.70
gi|146260269|ref|NM_201559.2| FOXO3	miR-302a	5919 to 5941	−25.70
gi|208879433|ref|NM_183395.2| NLRP3	miR-302a	1807 to 1829	−25.00
gi|208879433|ref|NM_183395.2| NLRP3	miR-302a	231 to 253	−26.40
gi|208879434|ref|NM_004895.4| NLRP3	miR-302a	1807 to 1829	−25.00
gi|208879434|ref|NM_004895.4| NLRP3	miR-302a	231 to 253	−26.40
gi|208879435|ref|NM_001127462.2| NLRP3	miR-302a	1807 to 1829	−25.00
gi|208879435|ref|NM_001127462.2| NLRP3	miR-302a	231 to 253	−26.40
gi|208879436|ref|NM_001127461.2| NLRP3	miR-302a	1841 to 1863	−25.00
gi|208879436|ref|NM_001127461.2| NLRP3	miR-302a	265 to 287	−26.40
gi|208879437|ref|NM_001079821.2| NLRP3	miR-302a	1199 to 1221	−25.00
gi|27502389|ref|NM_001562.2| IL18	miR-302a	997 to 1019	−29.50
gi|73622111|ref|NM_033294.2| CASP1	miR-302a	371 to 393	−34.80
gi|73622112|ref|NM_001223.3| CASP1	miR-302a	587 to 609	−34.80
gi|73622114|ref|NM_033292.2| CASP1	miR-302a	650 to 672	−34.80
gi|73622118|ref|NM_033293.2| CASP1	miR-302a	371 to 393	−34.80
gi|146260266|ref|NM_001455.3| FOXO3	miR-302b	5946 to 5968	−25.70
gi|146260269|ref|NM_201559.2| FOXO3	miR-302b	5919 to 5941	−25.70
gi|208879433|ref|NM_183395.2| NLRP3	miR-302b	1807 to 1829	−25.80
gi|208879433|ref|NM_183395.2| NLRP3	miR-302b	231 to 253	−26.40
gi|208879434|ref|NM_004895.4| NLRP3	miR-302b	1807 to 1829	−25.80
gi|208879434|ref|NM_004895.4| NLRP3	miR-302b	231 to 253	−26.40

Table 2 (*continued*)

mRNA	miRNA	Offset	Folding energy (kcal/mol)
gi|208879435|ref|NM_001127462.2| NLRP3	miR-302b	1807 to 1829	−25.80
gi|208879435|ref|NM_001127462.2| NLRP3	miR-302b	231 to 253	−26.40
gi|208879436|ref|NM_001127461.2| NLRP3	miR-302b	1841 to 1863	−25.80
gi|208879436|ref|NM_001127461.2| NLRP3	miR-302b	265 to 287	−26.40
gi|208879437|ref|NM_001079821.2| NLRP3	miR-302b	1199 to 1221	−25.80
gi|27502389|ref|NM_001562.2| IL18	miR-302b	997 to 1019	−30.40
gi|73622111|ref|NM_033294.2| CASP1	miR-302b	371 to 393	−35.30
gi|73622112|ref|NM_001223.3| CASP1	miR-302b	587 to 609	−35.30
gi|73622114|ref|NM_033292.2| CASP1	miR-302b	650 to 672	−35.30
gi|73622118|ref|NM_033293.2| CASP1	miR-302b	371 to 393	−35.30
gi|146260266|ref|NM_001455.3| FOXO3	miR-302c	7212 to 7234	−26.60
gi|146260269|ref|NM_201559.2| FOXO3	miR-302c	7185 to 7207	−26.60
gi|208879433|ref|NM_183395.2| NLRP3	miR-302c	811 to 833	−25.80
gi|208879433|ref|NM_183395.2| NLRP3	miR-302c	1807 to 1829	−28.32
gi|208879433|ref|NM_183395.2| NLRP3	miR-302c	231 to 253	−25.50
gi|208879434|ref|NM_004895.4| NLRP3	miR-302c	811 to 833	−25.80
gi|208879434|ref|NM_004895.4| NLRP3	miR-302c	1807 to 1829	−28.32
gi|208879434|ref|NM_004895.4| NLRP3	miR-302c	231 to 253	−25.50
gi|208879435|ref|NM_001127462.2| NLRP3	miR-302c	811 to 833	−25.80
gi|208879435|ref|NM_001127462.2| NLRP3	miR-302c	1807 to 1829	−28.32
gi|208879435|ref|NM_001127462.2| NLRP3	miR-302c	231 to 253	−25.50
gi|208879436|ref|NM_001127461.2| NLRP3	miR-302c	845 to 867	−25.80
gi|208879436|ref|NM_001127461.2| NLRP3	miR-302c	1841 to 1863	−28.32
gi|208879436|ref|NM_001127461.2| NLRP3	miR-302c	265 to 287	−25.50
gi|208879437|ref|NM_001079821.2| NLRP3	miR-302c	203 to 225	−25.80
gi|208879437|ref|NM_001079821.2| NLRP3	miR-302c	1199 to 1221	−28.32
gi|27502389|ref|NM_001562.2| IL18	miR-302c	997 to 1019	−32.60
gi|73622111|ref|NM_033294.2| CASP1	miR-302c	371 to 393	−33.20
gi|73622112|ref|NM_001223.3| CASP1	miR-302c	587 to 609	−33.20
gi|73622114|ref|NM_033292.2| CASP1	miR-302c	650 to 672	−33.20
gi|73622118|ref|NM_033293.2| CASP1	miR-302c	371 to 393	−33.20
gi|146260266|ref|NM_001455.3| FOXO3	miR-302d	1759 to 1781	−26.30
gi|146260266|ref|NM_001455.3| FOXO3	miR-302d	4768 to 4790	−29.50
gi|146260266|ref|NM_001455.3| FOXO3	miR-302d	5946 to 5968	−27.30
gi|146260269|ref|NM_201559.2| FOXO3	miR-302d	1732 to 1754	−26.30
gi|146260269|ref|NM_201559.2| FOXO3	miR-302d	4741 to 4763	−29.50
gi|146260269|ref|NM_201559.2| FOXO3	miR-302d	5919 to 5941	−27.30
gi|208879433|ref|NM_183395.2| NLRP3	miR-302d	2443 to 2465	−25.10
gi|208879433|ref|NM_183395.2| NLRP3	miR-302d	3797 to 3819	−26.50
gi|208879433|ref|NM_183395.2| NLRP3	miR-302d	231 to 253	−26.50
gi|208879434|ref|NM_004895.4| NLRP3	miR-302d	2443 to 2465	−25.10
gi|208879434|ref|NM_004895.4| NLRP3	miR-302d	4139 to 4161	−26.50
gi|208879434|ref|NM_004895.4| NLRP3	miR-302d	231 to 253	−26.50
gi|208879435|ref|NM_001127462.2| NLRP3	miR-302d	2443 to 2465	−25.10
gi|208879435|ref|NM_001127462.2| NLRP3	miR-302d	3968 to 3990	−26.50
gi|208879435|ref|NM_001127462.2| NLRP3	miR-302d	231 to 253	−26.50
gi|208879436|ref|NM_001127461.2| NLRP3	miR-302d	2477 to 2499	−25.10
gi|208879436|ref|NM_001127461.2| NLRP3	miR-302d	4002 to 4024	−26.50
gi|208879436|ref|NM_001127461.2| NLRP3	miR-302d	265 to 287	−26.50
gi|208879437|ref|NM_001079821.2| NLRP3	miR-302d	1835 to 1857	−25.10
gi|208879437|ref|NM_001079821.2| NLRP3	miR-302d	3531 to 3553	−26.50
gi|27502389|ref|NM_001562.2| IL18	miR-302d	997 to 1019	−28.50
gi|73622111|ref|NM_033294.2| CASP1	miR-302d	371 to 393	−33.30
gi|73622112|ref|NM_001223.3| CASP1	miR-302d	587 to 609	−33.30
gi|73622114|ref|NM_033292.2| CASP1	miR-302d	650 to 672	−33.30
gi|73622118|ref|NM_033293.2| CASP1	miR-302d	371 to 393	−33.30
gi|27894331|ref|NM_000877.2| IL1R1	miR-576-3p	2803 to 2824	−25.20

and genes. Then, we calculated the associations between miRNAs or clusters of miRNAs or mRNAs and OMIM diseases loci and finally validated these through permutations. The five steps of OMiR are presented in the Fig. 1. Out of the total number of associations, at 1% FDR (false discovery rate) we found 5.06%, 8.46% and 10.15% significant associations ($P<0.05$) between miRNAs and OMIM IDs '*' or '+', between miRNAs and the '#' prefixed OMIM diseases and miRNAs and the OMIM diseases with a '%' prefix, respectively.

The percentages of associations with mRNAs, in the same OMIM IDs categories were 48.25% ('*' or '+'), 52.45% ('#') and 38.55% ('%') respectively. Based on the generalized hypothesis of 10,000 available associations, the Fisher exact test between the categories '#' and '%' was $P<0.001$, while using all three categories was $P<0.001$ (Fig. 2). Table 1 reports the significant associations between 10 "orphan" diseases and 6 miRNAs clusters containing 63 miRNAs in total (6.7% of the human miRNAs from miRBase 15.0), (Supplemental file MiRNAs_OMIM_Associations.xls reports all found 88 significant associations).

Then we explored for 6 of the 88 significant '%' associated miRNAs the mRNAs target terms using DAVID (http://david.abcc.ncifcrf.gov/) (Table S1; the complete lists of significant target terms are reported in Supplemental Materials (Table S2–S7 for RNA22 predicted targets and file miRGen_DAVID.zip for miRGen predicted targets)). In several instances, the related Gene Ontology, INTERPRO, PATHWAY, UP_TISSUE and SP_PIR_KEYWORDS functional terms were indeed relevant for the claimed associated disease. Table 1 shows the significant associations obtained for miRNAs associated with '%' by statistical validation using 10,000 permutations ($P<0.05$, FDR<1%). For example, the cluster miR-297, miR-302a, miR-302b, miR-302c, miR-302d, miR-367, miR-376 is significantly associated by OMiR to UAQTL1 (uric acid concentration serum, quantitative trait locus 1), gout

susceptibility 1, longevity 1 and ATFB5 (atrial fibrillation familial 5) loci, all located in 4q25 chromosomal region. Experimentally, several PCGs were already found to be involved in the pathogenesis of these disorders, but none in a causal association: CPPD, CASP1, NALP3, IL-1B and IL1R for UAQTL1 disease, FOXO3A for longevity and PITX2 for atrial fibrillation. RNA22 predicted that miR-297 targets FOXO3, PITX2, IL1B, IL1R, IL18 and CASP1; miR-302a/b/c/d targets FOXO3, NLRP3 and CASP1 and miR-576-3p targets IL1R (Table 2).

4. Discussion

Traditionally, Mendelian disorders have been classified as recessive or dominant and by OMIM based on type of known information about them. However, there is increasing recognition that these classifications can be explored to find possible candidate for "orphan" diseases. Here, we investigated the possible association of miRNAs with human "orphan" diseases. We found 43 miRNA clusters consisting of about 37.2% (n = 349) of all known miRNAs and located in 80 OMIM '%' loci. The diseases we found significantly associated with the clusters of miRNAs are either associated to a locus, one or more markers and for few of them, several involved genes are known. For six of them we presented a functional study as a result of the integrated analysis of different functional databases (e.g. GO, KEGG pathways, SP-PIR keywords) provided interesting biological support to OMiR findings. As the number of individuals affected by inherited Mendelian disorders, excluding cancers, are usually small and therefore the available samples limited, OMIR can be a useful companion to the scientists for focusing on interesting culprit genes. Furthermore, with the identification of other categories of ncRNAs OMiR algorithm can be re-run for identification of further new possible "protein-less" non-coding culprits.

Our efforts to identify miRNAs important for diseases without known pathogeny are not isolated. Recently two papers have been published with the aim to evaluate human diseases and miRNAs genomic region associations. Lu and colleagues by manually curating the published associations, built a human miRNA associated disease network, suggesting that the applied method can be used in the future to discover novel disease associations [14]. By using a panel of single-nucleotide polymorphisms in miRNAs, Muiños-Gimeno and colleagues identified two miRNA regions showing geographical allelic frequency variation among four HapMAp populations [15]. None of these studies applied and compared the same method to a large scale of diseases already associated with a responsible gene or miRNA and "orphan" disease as we present with OmiR.

5. Conclusion

We devised and implemented OMiR, an algorithm to compute and study the associations between miRNAs and genetic diseases. We took advantage of the OMIM database to retrieve disease information. Additional association options can be implemented to extend the reach of OMiR, such as the possibility of using markers known by the user that have not yet been published. Moreover, updating OMiR with new OMIM IDs and new miRNAs will increase the robustness of the assay. Our work represents a novel way to extract information on the relationships between miRNAs and diseases with OMIM IDs not associated with a specific disease gene.

Supplementary materials related to this article can be found online at doi:10.1016/j.ygeno.2010.10.004.

Acknowledgments

We thank National Institutes of Health and Online Mendelian Inheritance in Man, OMIM™, URL: http://www.ncbi.nlm.nih.gov/omim/ for the available information at ftp://ftp.ncbi.nih.gov/repository/OMIM/. We thank Angelique Geehan, ESL, Department of Scientific Publications, The University of Texas MD Anderson Cancer Center, for expert editorial assistance.

Funding: GAC was supported as a Fellow of The University of Texas M. D. Anderson Research Trust, a Fellow of The University of Texas System Regents Research Scholar, and by the Ladjevardian Regents Research Scholar Fund. Work in Dr Calin's laboratory is supported in part by NIH/NCI, DOD and by 2009 Seena Magowitz — Pancreatic Cancer Action Network — AACR Pilot Grant.

References

[1] D.P. Bartel, MicroRNAs: target recognition and regulatory functions, Cell 136 (2009) 215–233.
[2] I. Rigoutsos, New tricks for animal microRNAS: targeting of amino acid coding regions at conserved and nonconserved sites, Cancer Res. 69 (2009) 3245–3248.
[3] Y. Tay, et al., MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation, Nature 455 (2008) 1124–1128.
[4] A.M. Krichevsky, et al., A microRNA array reveals extensive regulation of microRNAs during brain development, RNA 9 (2003) 1274–1281.
[5] J.L. Umbach, et al., Identification of viral microRNAs expressed in human sacral ganglia latently infected with herpes simplex virus 2, J. Virol. 84 (2009) 1189–1192.
[6] R. Spizzo, et al., SnapShot: microRNAs in cancer, Cell 137 (2009) 586-586.e1.
[7] M.S. Nicoloso, R. Spizzo, M. Shimizu, S. Rossi, G.A. Calin, MicroRNAs—the micro steering wheel of tumour metastases, Nat. Rev. Cancer 9 (2009) 293–302.
[8] A. Esquela-Kerscher, F.J. Slack, OncomicroRNAs — microRNAs with a role in cancer, Nat. Rev. Cancer 6 (2006) 259–269.
[9] S. Volinia, et al., A microRNA expression signature of human solid tumors defines cancer gene targets, Proc. Natl Acad. Sci. USA 103 (2006) 2257–2261.
[10] G.A. Calin, C. Sevignani, C.D. Dumitru, T. Hyslop, E. Noch, et al., Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers, Proc. Natl Acad. Sci. USA 101 (2004) 2999–3004.
[11] C. Sevignani, et al., Mammalian microRNAs: a small world for fine-tuning gene expression, Mamm. Genome 17 (2006) 189–202.
[12] I.V. Makunin, et al., Orthologous microRNA genes are located in cancer-associated genomic regions in human and mouse, PLoS ONE 2 (2007) e1133.
[13] W. Chen, et al., Mutation screening of brain-expressed X-chromosomal microRNA genes in 464 patients with nonsyndromic X-linked mental retardation, Eur. J. Hum. Genet. 15 (2007) 375–378.
[14] M. Lu, et al., An analysis of human microRNA and disease associations, PLoS ONE 3 (2008) e3420.
[15] M. Muiños-Gimeno, et al., Design and evaluation of a panel of single-nucleotide polymorphisms in microRNA genomic regions for association studies in human disease, Eur. J. Hum. Genet. 18 (20) (2010) 218–226.