A computational pipeline to identify new potential regulatory motifs in melanoma progression

A computational pipeline to identify new potential regulatory motifs in melanoma progression

Gianfranco Politano[1], Alfredo Benso[1,4], Stefano Di Carlo[1], Francesca Orso[2,3], Alessandro Savino[1], and Daniela Taverna[2,3]

[1] Politecnico di Torino,
Corso Duca degli Abruzzi 24, 10129, Torino, Italy
{alfredo.benso, stefano.dicarlo, gianfranco.politano, alessandro.savino}@polito.it
[2] Molecular Biotechnology Center (MBC),
Torino, Italy
{francesca.orso, daniela.taverna}@unito.it
[3] Center for Molecular Systems Biology,
Università di Torino, Torino, Italy
[4] Consorzio Interuniversitario Nazionale per l'Informatica
Verres (AO), Italy

Abstract. Molecular biology experiments allow to obtain reliable data about the expression of different classes of molecules involved in several cellular processes. This information is mostly static and does not give much clue about the causal relationships (i.e., regulation) among the different molecules. A typical scenario is the presence of a set of modulated mRNAs (up or down regulated) along with an over expression of one or more small non-coding RNAs molecules like miRNAs. To computationally identify the presence of transcriptional or post-transcriptional regulatory modules between one or more miRNAs and a set of target modulated genes, we propose a computational pipeline designed to integrate data from multiple online data repositories. The pipeline produces a set of three types of putative regulatory motifs involving coding genes, intronic miRNAs, and transcription factors. We used this pipeline to analyze the results of a set of expression experiments on a melanoma cell line that showed an over expression of miR-214 along with the modulation of a set of 73 other genes. The results suggest the presence of 27 putative regulatory modules involving *miR-214*, *NFKB1*, *SREBPF2*, *miR-33a* and 9 out of the 73 *miR-214* modulated genes (*ALCAM*, *POSTN*, *TFAP2A*, *ADAM9*, *NCAM1*, *SEMA3A*, *PVRL2*, *JAG1*, *EGFR1*). As a preliminary experimental validation we focused on 9 out of the 27 identified regulatory modules that involve *miR-33a* and *SREBF2*. The results confirm the importance of the predictions obtained with the presented computational approach.

Keywords: microRNA, miR-214, melanomas, biological pathways, gene regulation, post-transcriptional regulation

1 INTRODUCTION

Expression experiments on melanomas cell lines (as well as on tumor cell lines in general) often reveal an aberrant expression of coding and non-coding molecules, such

as microRNAs (miRNAs). miRNAs are 20 to 24 nucleotides long non-coding RNAs involved in the post-transcriptional down-regulation of protein-coding genes through imperfect base pairing with their target mRNAs. The peculiar characteristic of miRNAs is their ability to simultaneously affect the expression of several genes. Consequently, miRNAs have been implicated as possible key factors in several diseases [1,24,2,8,27]. Referring to melanomas, miRNAs such as *let-7a/b*, *miR-23a/b*, *miR-148*, *miR-155*, *miR-182*, *miR-200c*, *miR-211*, *miR214* and *miR-221/222* have been found to be differentially expressed in benign melanocytes versus melanoma cell lines or in benign melanocytic lesions versus melanomas in human samples. Targets of some of the above listed miRNAs are well-known melanoma-associated genes, like the oncogene *NRAS*, the microphthalmia-associated transcription factor (*MITF*), the receptor tyrosine kinase *c-KIT*, and the *AP-2* transcription factor (*TFAP2*). *miR-214* is known to control in vitro tumor cell movement and survival to anoikis, as well as in vivo malignant cell extravasation from blood vessels and lung metastasis formation. [20] and [21] show that *miR-214*, the product of an intron of the *Dynamin-3* gene on human chromosome 1, coordinates melanoma metastasis formation by modulating the expression of over 70 different genes, including two activating protein transcription factors (*TFAP2A* and *TFAP2C*) and the adhesion molecule *ALCAM*. The alteration in the expression level of some of these genes leads to important downstream effects on a number of key processes such as apoptosis, proliferation, migration and invasion. The static information about genes and miRNAs expression is very important but not sufficient to precisely understand the regulatory dynamics that cause the aberrant phenotype. A significant help in this direction can be provided by Systems Biology approaches designed to integrate the huge amount of biological data available online to infer possible regulatory motifs involving the molecules of interest highlighted by microarray expression experiments. In this work, we present a computational pipeline that allows biologists to identify the presence of different classes of regulatory modules between a miRNA and a set of target genes of interest. We used the pipeline to analyze the possible relationships between *miR-214* and the set of 73 proteins found modulated in melanoma expression experiments.

The pipeline is designed to automatically verify the presence of three different classes of regulatory modules, all characterized by an interplay between Transcription Factors (TFs) and miRNAs. Several studies (such as [28,19,7,11]) suggested that this type of interaction is particularly critical in cellular regulation during tumor genesis. The regulatory motifs that are analyzed (see Figure 1) are:

1. Type-0 (direct interactions), where the *miRNA* directly down-regulates one of the target proteins;
2. Type-1 (one-level indirect interactions), where the *miRNA* down-regulates a Transcription Factor, which eventually regulates one of the targets genes;
3. Type-2 (two-levels indirect interactions), where the *miRNA* targets a Transcription Factor (TF). The TF is then regulating a gene, which hosts an intragenic miRNA that acts as down-regulator of one of the target proteins.

Whereas detecting Type-0 and Type-1 interactions is a quite simple data integration task, the identification of all the necessary evidence to infer the existence of Type-2

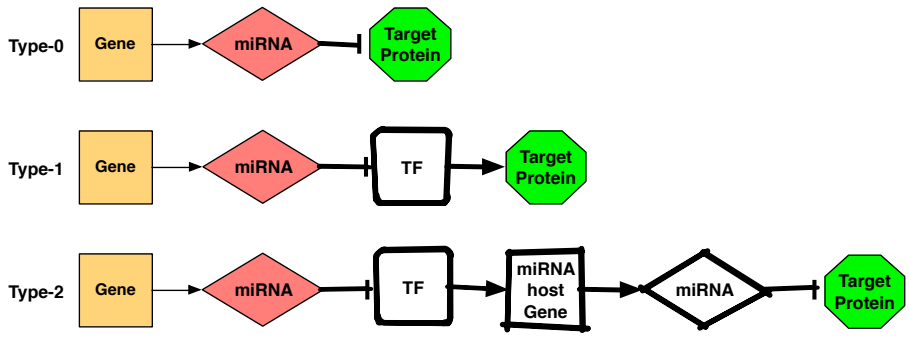

Fig. 1: Three classes of regulatory modules involving *miR-214* investigated in this paper: Type-0 (direct interactions), Type-1 (one-level indirect interactions), and Type-2 (two-levels indirect interactions)

interactions requires a more complex data integration process. Obviously, these three motifs are not the only interactions that may be investigated. Deeper interactions could be addresses, for example three-level integrations that involve two TFs before reaching a target protein (*miR-214* $\rightarrow$ TF1 $\rightarrow$ TF2 $\rightarrow$ Target Protein). At this stage these complex interactions have not been considered due to the difficulty to experimentally validate their existence.

2 Methods

The computational analysis to investigate the set of previously defined regulatory motifs (Figure 1) requires the successful integration of heterogeneous data sources. This section describes, at first, the set of public repositories included in the flow and, eventually, the computational pipeline that integrates and elaborates these data sources to search for the three classes of interactions.

2.1 Data sources

The following public repositories represent the main sources of information in our computational process:

– **microRNA.org** database [4] is used to search for miRNA target genes. MicroRNA.org uses the miRanda algorithm [12] for target predictions. The algorithm computes optimal sequence complementarity between a miRNA and a mRNA using a weighted dynamic programming algorithm. The overall database consists of 16,228,619 predicted miRNA target sites in 34,911 distinct 3'UTR from isoforms of 19,898 human genes. Each prediction is associated to a prediction confidence score, the mirSVR score. The mirSVR is a real number computed by a machine learning method for ranking miRNA target sites by a down-regulation score [3]. The lower (negative) is the score, the better is the prediction.

microRNA.org provides data organized in four different packages: (1) Good mirSVR score, Conserved miRNA, (2) Good mirSVR score, Non-conserved miRNA, (3) Non-good mirSVR score, Conserved miRNA, (4) Non-good mirSVR score, Non-conserved miRNA. They include prediction clustered in terms of mirSVR score (if considered good or not) and conservation (highly, low conserved). The computational pipeline includes a single databased, which unifies the four packages keeping all source information intact (stored in a specific field). The mirSVR score information is still included and it is used in order to be able to refine queries by filtering unreliable predictions out.

- **Transcription Factor Encyclopedia** (TFE) [25] and **Targetmine** [23,6] have been used to identify genes TFs. TFE provides details of Transcription Factor binding sites in close collaboration with Pazar, a public database of TFs and regulatory sequence information. While TFE includes both Upstream and Downstream Transcription Factors, Targetmine contains only Upstream Transcription Factors. For each gene, the database retrieves all upstream regulatory genes from the AMADEUS [13,14] and ORegAnno compiled TF-Target gene sets. AMADEUS contains TF and miRNA target sets in *human*, *mouse*, *D. Melanogaster*, and *C. Elegans*, collected from the literature. For each TF, it is reported its set of targets, given as a list of Ensembl gene IDs.
- **e-Utils programming utilities** [18] and **Mirbase.org** [16,10] allow for retrieving coordinates of precursor miRNAs and genes. miRBase is a searchable database of miRNA sequences and annotations already published. About 94.5% of the available mature miRNA sequences considered in this database have experimental evidence, thus representing a reliable source of information. Each miRNA entry in miRBase is annotated with the information on the location that is exploited to identify the host genes.

2.2 Computational Pipeline

The full pipeline has been developed in PHP language and coupled with a MySQL database, which mirrors an optimized subset of data coming from multiple online repositories. To better explain its functioning, we will refer to an actual analysis that we performed to search for *miR-214* mediated interactions. Nevertheless, the same analysis could be executed with any other miRNA and a set of target genes of interest. Figure 2 highlights the implemented computational pipeline. The modulator miRNA is *miR-214* and the list of target genes of interest is the set of 73 protein-coding genes reported in Table 1. Previous microarray experiments, [20], suggest that the *miR-214* modulates (directly or indirectly) these proteins. In Figure 2 these genes are named "Target Proteins".
The computational pipeline is organized into four main data integrations steps, which search for Type-0, Type-1 and Type-2 interactions.

Step 1 - detection of Type-0 interactions: Target genes that are directly regulated by *miR-214* represent the way Type-0 interactions manifest. microRNA.org provides information on *miR-214* direct targets. Thus, we queried microRNA.org database to search for all direct targets predicted for *miR-214*. Due to the computational

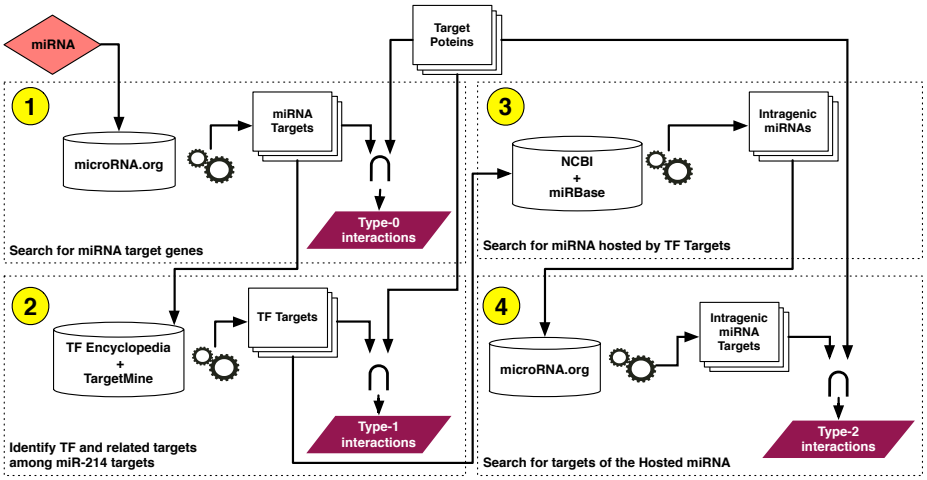

Fig. 2: The four steps pipeline to investigate the presence of transcriptional or post-transcriptional regulatory pathways.

approach used by microRNA.org to predict miRNA targets, query results can be affected by false positives. By restricting the query to "Good mirSVR score, Conserved miRNA" and to the "Good mirSVR score, Non-conserved miRNA" packages, a more reliable set of computed targets is represented, positively affecting all investigation results. Once the set of computed targets is retrieved, we applied a further filtering process according to the mirSVR score. In [3], the mirSVR score is shown to be meaningful with a cut-off of at most -0.1, derived from the empirical distribution of the extent of target down-regulation (measured as log-fold change) that is expected given a mirSVR score. For scores closer to zero, according to [15], the probability of meaningful down-regulation decreases while the number of predictions drastically increase. The final selection is done with mirSVR $<$ -0.3 in order to guarantee high reliable predictions.
Once the set of *miR-214* targets is defined, the pipeline intersects the full list with the set of 73 Target Proteins, in order to highlight the Type-0 interactions.

Step 2 - detection of Type-1 interactions: Step 2 starts from the full list of *miR-214* targets already computed during Step 1. In fact, the Type-1 interactions require to identify, among the miRNA targets, those that are also TFs for other genes.
The search is done for each *miR-214* target both in Transcription Factor Encyclopedia (TFE) and TargetMine. The output is the subset of *miR-214* targets that are also TFs. For each of them, the related targets are extracted and added to a list of TF Targets. The pipeline identifies all possible Type-1 interactions by intersecting this list with the set of 73 targets.
For each TF-target interaction we also check if TFE stores the information about its sign (either activation or inhibition). When available, this information is useful to have a better insight about the behavior of the regulatory module. When the

Table 1: List of 73 miR-214 modulated genes. In green and bold the set of proteins that result linked to miR-214 in the discovered regulatory modules. The arrow indicates if the gene was up regulated (up) or down regulated (down) in the microarray experiments; in red, proteins that do not show any connection

↑**ADAM9**	↓**JAM3**	↑**THY1**	CD44	ENG
↑**ALCAM**	↓**LRP6**	↑**TIMP3**	CD9	EPCAM
↓**BMPR1B**	↑**MET**	ADAM15	CDH1	ERBB2
↓**CD40**	↑**MMP2**	ADAM8	CDH11	ERBB3
↑**CD99**	↑**NCAM1**	APP	CDH2	EREG
↑**CEACAM1**	↓**POSTN**	ARHGAP12	CDH4	F2
↓**CEACAM5**	↑**PVRL2**	BCAM	CDHR5	FCER2
↓**EGFR**	↓**SEMA3A**	BSG	CLU	FLT1
↑**HBEGF**	↓**TFAP2A**	CD36	CTSD	HRG
↓**JAG1**	↓**TFAP2C**	CD40LG	CX3CL1	ICAM2
IL1R2	LCN2	TIMP1	IL8	LGALS3BP
TIMP2	ITGA3	MITF	VCAM1	ITGA6
PAK2	ITGAV	PODXL	ITGB1	PODXL2
ITGB3	PTEN	JAM1	PVR	JAM2
SELE	KDR	TGFBI		

sign of the TF is positive (i.e., activation), an increase in the expression of the miRNA is expected to lead to the down-regulation of its target. On the other side, a negative sign (i.e., inhibition) will result in a correlated expression between the miRNA and its target. If this prediction correlates with the actual expression of the target genes, this information can provide an additional biological validation of the detected motif.

Step 3 - Intragenic miRNAs enrichment: The last two steps of the proposed computational pipeline are used to identify Type-2 interactions, which represent the most complex considered motif.
The previous list of TF Targets needs to be enriched in order to identify a set of candidate intragenic miRNAs. Intragenic miRNAs represent around 50% of the mammalian miRNAs. Most of these intragenic miRNAs are located within introns of protein coding genes (miRNA host genes) and are referred to as intronic miRNAs, whereas the remaining miRNAs are overlapping with exons of their host genes and are thus called exonic miRNAs. Moreover, the majority of intragenic miRNAs are sense strand located, while only a very small portion is anti-sense strand located.

In this step we want to consider both intronic and exonic miRNAs, either sense or anti-sense strand located. For each TF Target, its set of candidate intragenic miRNAs are retrieved by querying the miRBase database. To be able to correctly identify intragenic miRNAs of a given host gene, we use e-Utils to resolve the genomic coordinates of the gene. The gene coordinates are the right input to query the miRBase database for all miRNAs whose coordinates are embraced within the ones of the gene. The enrichment feeds the last step to complete the detection of Type-2 interactions.

Step 4 - detection of Type-2 interactions: So far, the *miR-214* has been correlated to all its targets annotated as TF (Step 2), and each of them with the intragenic miRNAs (if any) of the gene they target (Step 3). At this point, with the full list of possibly modulated Intragenic miRNAs, the pipeline searches microRNA.org to obtain the list of their possible targets. By filtering this list and keeping only those targets that correspond to any of the 73 gene of interest, the identification of the possible Type-2 interactions is complete.

2.3 Data integration: strategy, bottlenecks, and optimizations

To efficiently design and reliably use the data produced by a pipeline of this type, it is necessary to take into account a number of bottlenecks related to data integration and data reliability.
Given the large amount of available databases and their lack of standardization rules for data consistency and interoperability [9], the integration of information coming from heterogeneous sources is often a hard and computationally expensive task. To properly collect all necessary data, the pipeline needs to interrogate multiple databases; to integrate this data, it is necessary that records that refer to the same item (gene, TF, miRNA) share a set of compatible fields that can be used to link these data together. Field compatibility is based upon synonyms and accession numbers across the selected information sources. In this work we specifically designed and populated a custom database to cope with this issue. The database contains information from the databases listed in Section 2.1 enhanced with additional meta data. Such meta data are collections of accession numbers (i.e., Entrez GeneID and approved common symbol) whose presence simplifies SQL join operations on queries, both reducing the execution time and avoiding either multiple or ambiguous matches.
From a data reliability point of view, the type of analysis and data integration performed by this pipeline is obviously affected by the reliability of the source data extracted by online repositories. In the presented pipeline there are two critical and potential sources of unreliable or inconsistent data:

1. miRNA and TF targets;
2. TF regulatory action (enhancer, silencer).

In the case of miRNA and TF targets, it is well known how the different algorithms used to predict the potential targets can lead to very different results. In this case a possible way to approach the problem is to define a policy to merge similar information coming from different sources (i.e. TF targets extracted from TFE

and TargetMine, or miRNA targets extracted from different databases). In this scenario, the pipeline has to handle potential incoherencies between the data sources. In the presented work, to handle TF targets we designed two user selectable policies: (i) intersection or (ii) union of data. In the first case, only targets present in both databases are selected, whereas in the second case the union of the targets of both databases is considered. Both policies offer drawbacks and advantages: while intersecting data among sources seems to offer a more reliable dataset but possibly a higher rate of false negatives, merging multiple sources results in a larger dataset is able to provide a more investigative outcome but also a higher probability of false positives. Although more keen to false-positives, the results presented in this work are obtained using the "union of data" policy in order to include in the analysis as much information as possible. This choice is also justified by the high level of abstraction of the proposed work together with its aim to infer putative relations that will have in any case to be biologically validated.
For what concerns miRNA targets, the presented pipeline uses, for now, a single database (microRNA.org) and relies on a mirSVR score threshold to exclude target predictions that are too unreliable. Nevertheless, also in this case an approach based on data integration policies could be more reliable and give the user more flexibility. In particular the availability of several different miRNA target prediction databases could allow the implementation of a "majority voting" policy, where a target is considered reliable if it appears in at least *k out of n* databases, with *k* selectable by the user.
Finally, concerning TF regulatory action, it is important to point out here that the proposed computational analysis cannot always predict the sign of the resulting differential expression (up or down regulation). In fact, following the Type-2 regulatory chain, if *miR-214* is silenced the expression of the target protein is very likely inhibited. If, instead, *miR-214* is over expressed, the regulatory module "removes" the inhibition and allows the target gene expression to possibly change. The only realistic way to experimentally verify the presence of the Type-2 regulatory module is to correlate the over expression of *miR-214* with the under expression of the cascade TF $\rightarrow$ gene $\rightarrow$ miRNA that follows *miR-214* (see Figure 1). This is obviously true unless the transcription factor acts as a repressor of its own target, which is statistically unlikely to happen. The type of regulatory action of a TF on its target is an information that is very difficult to find and the only database that offers it (Transcription Factor Encyclopedia) does so for a very limited set of TFs. Consequently, for most motifs it is not possible to predict if its presence is or not compatible with the actual modulation detected in the expression experiment.

2.4 Biological methods

Computational predictions have been validated against the following biological setup.

Cell culture MA-2 cells were provided by R.O. Hynes [26] and maintained as described in [20].

Transient transfections of pre-miRs To obtain transient *miR-214* over expression, cells were plated in 6-well plates at 30-50% confluency and transfected 24h later using RNAiFect (QIAGEN, Stanford, CA) reagent, according to manufacturers instructions, with 75 nM Pre-miR™ miRNA Precursor Molecules-Negative Control (a non-specific sequence) or Pre-miR-214.

RNA isolation and qRT-PCR for miRNA or mRNA detection Total RNA was isolated from cells using TRIzol® Reagent (Invitrogen Life Technologies, Carlsbad, CA). qRT-PCRs for miR detection were performed with TaqMan® MicroRNA Assays *hsa-miR-33a* assay ID 002306, U6 snRNA assay ID001973 (all from Applied Biosystems, Foster City, CA) on 10 ng total RNA according to the manufacturer's instructions. For mRNA detection, 1 ug of DNAse-treated RNA (DNA-free™ kit, Ambion, Austin, TX) was retrotranscribed with RETROscript™ reagents (Ambion, Austin, TX) and qRT-PCRs were carried out using SREBPF2 gene-specific primers (FW:gccctggaagtgacagagag, RV: tgctttcccagggagtga) and the Probe #21 of the Universal Probe Library (Roche, Mannheim, GmbH) using a 7900HT Fast Real Time PCR System. The expression of the U6 small nucleolar RNA or of 18S was quantitatively normalized for miR or mRNA detection, respectively. The relative expression levels between samples were calculated using the comparative delta CT (threshold cycle number) method (2-DDCT) with a control sample as the reference point [5].

3 Results and discussion

Using the presented pipeline we were able to identify 27 Type-2 interactions. No Type-0 or Type-1 have been identified. All results have been manually verified and the pipeline has also been tested against very simple (and known) examples of each interaction type. Thus, the fact that no Type-0 and Type-1 interactions were found has no biological meaning: it only is the proof that, in the available database, there is no evidence of their presence.

The 27 Type-2 interactions target 22 out of the 73 considered *miR-214* potential interacting proteins, which have been marked in green in Table 1. The full list of the 27 identified regulatory modules is shown in Table 2.

From our predictions, *miR-214* influences two transcription factors: *NFKB1* and *TP53* (average mirSVR = -0.4). Seven of the genes regulated by these two TFs were identified as host genes for miRNAs targeting at least one of the 73 *miR-214* modulated proteins: *APOLD1*, *BBC3*, *C11orf10*, *GDF15*, *NFATC2*, *SREBF2*, and *SVIL*. The hosted miRNAs are: *hsa-mir-33a*, *hsa-mir-604*, *hsa-mir-611*, *hsa-mir-613*, *hsa-mir-3189*, *hsa-mir-3191*, and *hsa-mir-3194*. The average mirSVR score is significantly low (average mirSVR $<$ -0.71). The high significance of the mirSVR scores, resulting from interactions between the intragenic miRNAs and their target proteins, is particularly evident for *TFAP2A*, which outperforms the others with a mirSVR score of -1.3043.

In this work, as a preliminary experimental validation, we focused our attention on the first 9 identified regulatory modules involving *miR-214*, *NFKB1*, *SREBF2*, *miR-33a* and 9 of the 73 considered proteins (*ALCAM*, *POSTN*, *TFAP2A*, *ADAM9*, *NCAM1*,

Table 2: The 27 Type-2 regulatory modules related to miR-214 as obtained by the pipeline after data scraping. The set of final targets (surface protein in the table) is limited to the 73 genes listed in Table 1. The first 9 modules have been experimentally validated.

miR_214	mirSVR	TF	miRNA_Host	Intragenic_miRNA	Surface Protein	mirSVR
miR-214	**-0.4056**	**NFKB1**	**SREBF2**	**hsa-mir-33a**	**ALCAM**	**-0.504**
miR-214	**-0.4056**	**NFKB1**	**SREBF2**	**hsa-mir-33a**	**POSTN**	**-0.9944**
miR-214	**-0.4056**	**NFKB1**	**SREBF2**	**hsa-mir-33a**	**TFAP2A**	**-1.3043**
miR-214	**-0.4056**	**NFKB1**	**SREBF2**	**hsa-mir-33a**	**ADAM9**	**-0.8819**
miR-214	**-0.4056**	**NFKB1**	**SREBF2**	**hsa-mir-33a**	**NCAM1**	**-1.1293**
miR-214	**-0.4056**	**NFKB1**	**SREBF2**	**hsa-mir-33a**	**SEMA3A**	**-1.0884**
miR-214	**-0.4056**	**NFKB1**	**SREBF2**	**hsa-mir-33a**	**PVRL2**	**-0.3633**
miR-214	**-0.4056**	**NFKB1**	**SREBF2**	**hsa-mir-33a**	**JAG1**	**-0.7951**
miR-214	**-0.4056**	**NFKB1**	**SREBF2**	**hsa-mir-33a**	**EGFR**	**-0.5771**
miR-214	-0.4056	NFKB1	SVIL	hsa-mir-604	MMP2	-0.5526
miR-214	-0.4056	NFKB1	SVIL	hsa-mir-604	CEACAM5	-0.6373
miR-214	-0.4056	NFKB1	C11orf10	hsa-mir-611	THY1	-0.3774
miR-214	-0.4056	NFKB1	C11orf10	hsa-mir-611	NCAM1	-0.4402
miR-214	-0.4056	NFKB1	APOLD1	hsa-mir-613	MET	-0.8579
miR-214	-0.4056	NFKB1	APOLD1	hsa-mir-613	ALCAM	-0.5254
miR-214	-0.4056	NFKB1	APOLD1	hsa-mir-613	TIMP3	-0.582
miR-214	-0.4056	NFKB1	APOLD1	hsa-mir-613	CEACAM1	-0.9242
miR-214	-0.4056	NFKB1	APOLD1	hsa-mir-613	BMPR1B	-0.7156
miR-214	-0.4056	NFKB1	APOLD1	hsa-mir-613	TFAP2C	-0.6921
miR-214	-0.4056	NFKB1	APOLD1	hsa-mir-613	JAG1	-0.4012
miR-214	-0.4056	NFKB1	NFATC2	hsa-mir-3194	CD99	-0.8366
miR-214	-0.4056	NFKB1	NFATC2	hsa-mir-3194	CD40	-0.7136
miR-214	-0.3966	TP53	GDF15	hsa-mir-3189	JAM3	-0.8858
miR-214	-0.3966	TP53	GDF15	hsa-mir-3189	PVRL2	-0.5146
miR-214	-0.3966	TP53	GDF15	hsa-mir-3189	HBEGF	-0.3806
miR-214	-0.3966	TP53	GDF15	hsa-mir-3189	LRP6	-0.6945
miR-214	-0.3966	TP53	BBC3	hsa-mir-3191	HBEGF	-0.8502

SEMA3A, *PVRL2*, *JAG1* and *EGFR1*). We evaluated *miR-33a* and *SREBPF2* expression levels following *miR-214* over expression in MA-2 melanoma cells and we observed a decrease in *miR-33a* and *SREBF2* expression as shown in Figure 3.

The observed co-regulation of *miR-33a* and *SREBPF2* is in agreement with literature data published in [17], thus supporting our computational predictions. The down-regulation of *miR-33a* following *miR-214* over expression could contribute to *miR-214*-mediated cell invasion, in fact it has been demonstrated that an enforced expression of *miR-33a* inhibits the motility of lung cancer cells [22].

This regulatory module resulted to be very interesting also because *SREBPF2* and *miR-33a* act in concert to control cholesterol homeostasis [17]. In fact, *SREBPF2* controls the expression of many cholesterogenic and lipogenic genes, such as low-density lipoprotein (*LDL*) receptor, 3-hydroxy-3-methylglutaryl coenzyme A reductase, and fatty acid synthase. Instead, *miR-33a* targets the adenosine triphosphate-binding cassette A1 (*ABCA1*) cholesterol transporter, a key mediator of intracellular cholesterol efflux from liver to apolipoprotein A-I (*apoA-I*) to obtain high-density lipoprotein (*HDL*). Considering that the lipogenic pathway is a metabolic hallmark of cancer cells, these preliminary data suggest a potential role of *miR-214* in this aspect of cancer formation and progression. Our hypothesis is further supported by experimental results ([20]), obtained from microarray analysis in a context of *miR-214* over expression.

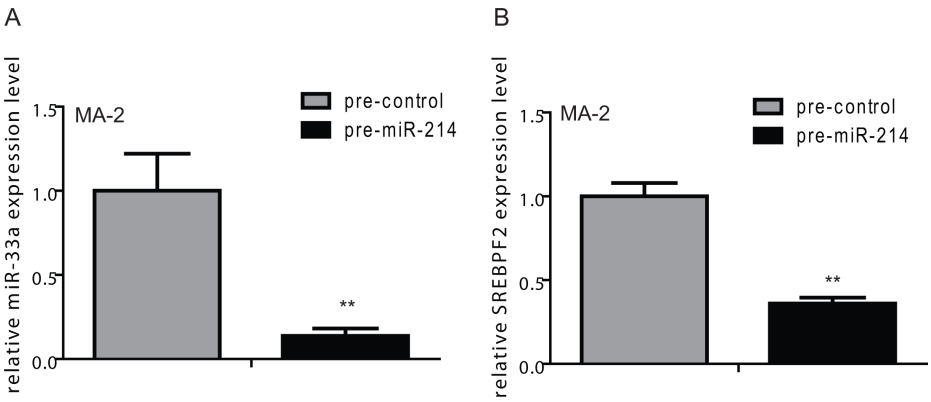

Fig. 3: miR-33a, and SREBPF2 expression modulations. (A) miR-33a expression levels tested by qRT-PCR in the MA-2 melanoma cell line following transfection with miR-214 precursors or their negative controls (pre-miR-214 or control). (B) SREBPF2 mRNA expression levels were evaluated in MA-2 cells by Real Time PCR analysis 72h following transient transfection with miR-214 precursors or their negative controls (pre-miR-214 or control). Results are shown as fold changes (meanSD of triplicates) relative to controls, normalized on U6 RNA level and 18S, respectively. All experiments performed in our work were tested for miR-214 modulations; representative results are shown here.

We applied an Ingenuity Functional Analysis in order to look for molecular and cellular functions associations within the almost 500 differentially expressed genes detected by microarray analysis comparing cells over expressing *miR-214* versus controls. The Ingenuity Pathways Knowledge Base (http://www.ingenuity.com/) is currently the world largest database of knowledge on biological networks, with annotations performed by experts. The significance value obtained with the Functional Analysis for a dataset is a measure of the likelihood that the association between a set of Functional Analysis molecules in our experiment and a given process or pathway is due to random chance. The p-value is calculated using the right-tailed Fisher Exact Test and it considers both the number of functional analysis molecules that participate in that function and the total number of molecules that are known to be associated with that function in the Ingenuity Knowledge Base. In our case, the most significant functions associated to our dataset resulted to be Cellular Assembly and Organization (7.08E-04 ÷ 3.95E-02, 25 molecules) and Lipid Metabolism (9.54E-04 ÷ 4.23E-02, 18 molecules).

4 Conclusions

In this paper we presented a computational pipeline created for investigating possible regulatory pathways between a miRNA and a set of target genes.

The pipeline identified 27 putative regulatory pathways that link together *miR-214* and a set of 73 proteins already annotated as co-regulated with the miRNA in melanomas. A preliminary experimental validation performed on 9 out of the 27 pathways provided interesting insights about the regulatory mechanisms involving *miR-*

214 in the considered disease. The analysis suggests the involvement of *miR-214* in metabolic pathways that could control metastatization. Moreover, the study highlights the relevance of specific *miR-214* modulated genes, such as *ALCAM*, *HBEGF*, *JAG1*, *NCAM1*, and *PVRL2*, which correspond to surface proteins redundantly regulated by multiple pathways. Further laboratory experiments are under way to complete the validations of the full set of identified regulatory modules. Nevertheless, the preliminary results presented in this work already represent a significant achievement that seems to confirm the overall quality of the predictions obtained with the proposed computational approach.

Acknowledgments

This work has been partially supported by grants from Regione Valle d'Aosta (for the project: "Open Health Care Network Analysis" - CUP B15G13000010006), from the Italian Ministry of Education, University.

References

1. Beezhold, K.J., Castranova, V., Chen, F.: Review microprocessor of microRNAs: regulation and potential for therapeutic intervention. Molecular Cancer 9(134) (2010)
2. Benso, A., Di Carlo, S., Politano, G., Savino, A.: A new mirna motif protects pathways' expression in gene regulatory networks. In: Proceedings IWBBIO 2013: International Work-Conference on Bioinformatics and Biomedical Engineering. pp. 377–384 (2013)
3. Betel, D., Koppal, A., Agius, P., Sander, C., Leslie, C.: Comprehensive modeling of microrna targets predicts functional non-conserved and non-canonical sites. Genome Biol 11(8), R90 (2010)
4. Betel, D., Wilson, M., Gabow, A., Marks, D.S., Sander, C.: The microRNA.org resource: targets and expression. Nucleic Acids Res 36(Database issue), D149–53 (Jan 2008)
5. Bookout, A.L., Mangelsdorf, D.J.: Quantitative real-time pcr protocol for analysis of nuclear receptor signaling pathways. Nuclear receptor signaling 1 (2003)
6. Chen, Y.A., Tripathi, L.P., Mizuguchi, K.: Targetmine, an integrated data warehouse for candidate gene prioritisation and target discovery. PLoS ONE 6(3), e17844 (03 2011)
7. Delfino, K.R., Rodriguez-Zas, S.L.: Transcription factor-microrna-target gene networks associated with ovarian cancer survival and recurrence. PLoS ONE 8(3), e58608 (03 2013), http://dx.doi.org/10.1371%2Fjournal.pone.0058608
8. Di Carlo, S., Politano, G., Savino, A., Benso, A.: A systematic analysis of a mi-RNA inter-pathway regulatory motif. Journal of clinical bioinformatics 3(20) (2013)
9. Gaudet, P., Bairoch, A., Field, D., Sansone, S.A., Taylor, C., Attwood, T.K., Bateman, A., Blake, J.A., Bult, C.J., Cherry, J.M., Chisholm, R.L., Cochrane, G., Cook, C.E., Eppig, J.T., Galperin, M.Y., Gentleman, R., Goble, C.A., Gojobori, T., Hancock, J.M., Howe, D.G., Imanishi, T., Kelso, J., Landsman, D., Lewis, S.E., Karsch Mizrachi, I., Orchard, S., Ouellette, B.F., Ranganathan, S., Richardson, L., Rocca-Serra, P., Schofield, P.N., Smedley, D., Southan, C., Tan, T.W., Tatusova, T., Whetzel, P.L., White, O., Yamasaki, C., on behalf of the BioDBCore working group: Towards biodbcore: a community-defined information specification for biological databases. Database 2011 (2011), http://database.oxfordjournals.org/content/2011/baq027.abstract

10. Griffiths-Jones, S., Grocock, R.J., Van Dongen, S., Bateman, A., Enright, A.J.: miRBase: microRNA sequences, targets and gene nomenclature. Nucleic acids research 34(suppl 1), D140–D144 (2006)
11. Iwama, H.: Coordinated networks of micrornas and transcription factors with evolutionary perspectives. In: Schmitz, U., Wolkenhauer, O., Vera, J. (eds.) MicroRNA Cancer Regulation, Advances in Experimental Medicine and Biology, vol. 774, pp. 169–187. Springer Netherlands (2013), http://dx.doi.org/10.1007/978-94-007-5590-1_10
12. John, B., Enright, A., Aravin, A., Tuschl, T., Sander, C., Marks, D.: Human microRNA targets. PLoS biology 2(11), e363 (2004)
13. Linhart, C., Halperin, Y., Shamir, R.: Amadeus. http://acgt.cs.tau.ac.il/amadeus/download.html (2013)
14. Linhart, C., Halperin, Y., Shamir, R.: Transcription factor and microrna motif discovery: the amadeus platform and a compendium of metazoan target sets. Genome research 18(7), 1180–1189 (2008)
15. MicroRNA.org: Microrna.org - release notes. http://www.microrna.org/microrna/releaseNotes.do (2013)
16. mirbase.org: mirbase.org. http://www.mirbase.org (2013)
17. Najafi-Shoushtari, S.H., Kristo, F., Li, Y., Shioda, T., Cohen, D.E., Gerszten, R.E., Näär, A.M.: Microrna-33 and the srebp host genes cooperate to control cholesterol homeostasis. Science 328(5985), 1566–1569 (2010)
18. NCBI: Entrez programming utilities help. http://www.ncbi.nlm.nih.gov/books/NBK25501 (2013)
19. Peng, C., Wang, M., Shen, Y., Feng, H., Li, A.: Reconstruction and analysis of transcription factor–mirna co-regulatory feed-forward loops in human cancers using filter-wrapper feature selection. PLoS ONE 8(10), e78197 (10 2013), http://dx.doi.org/10.1371%2Fjournal.pone.0078197
20. Penna, E., Orso, F., Cimino, D., Tenaglia, E., Lembo, A., Quaglino, E., Poliseno, L., Haimovic, A., Osella-Abate, S., De Pittà, C., et al.: microrna-214 contributes to melanoma tumour progression through suppression of tfap2c. The EMBO journal 30(10), 1990–2007 (2011)
21. Penna, E., Orso, F., Cimino, D., Vercellino, I., Grassi, E., Quaglino, E., Turco, E., Taverna, D.: mir-214 coordinates melanoma progression by upregulating alcam through tfap2 and mir-148b downmodulation. Cancer research 73(13), 4098–4111 (2013)
22. Rice, S.J., Lai, S.C., Wood, L.W., Helsley, K.R., Runkle, E.A., Winslow, M.M., Mu, D.: Microrna-33a mediates the regulation of high mobility group at-hook 2 gene (hmga2) by thyroid transcription factor 1 (ttf-1/nkx2–1). Journal of Biological Chemistry 288(23), 16348–16360 (2013)
23. The Mizuguchi Laboratory: Targetmine. http://targetmine.nibio.go.jp/ (2013)
24. Tu, K., Yu, H., Hua, Y.J., Li, Y.Y., Liu, L., Xie, L., Li, Y.X.: Combinatorial network of primary and secondary microrna-driven regulatory mechanisms. Nucleic acids research 37(18), 5969–5980 (2009)
25. Wasserman Lab: Transcription factor encyclopedia (TFe). http://www.cisreg.ca/cgi-bin/tfe/home.pl (2012)
26. Xu, L., Shen, S.S., Hoshida, Y., Subramanian, A., Ross, K., Brunet, J.P., Wagner, S.N., Ramaswamy, S., Mesirov, J.P., Hynes, R.O.: Gene expression changes in an animal melanoma model correlate with aggressiveness of human melanoma metastases. Molecular Cancer Research 6(5), 760–769 (2008)
27. Yuan, X., Liu, C., Yang, P., He, S., Liao, Q., Kang, S., Zhao, Y.: Clustered micrornas' coordination in regulating protein-protein interaction network. BMC systems biology 3(65) (2009)

28. Zhao, M., Sun, J., Zhao, Z.: Synergetic regulatory networks mediated by oncogene-driven microrna and transcription factors in serous ovarian cancer. Molecular BioSystems 9(12), 3187–3198 (2013)