miR-SEA: miRNA Seed Extension based Aligner Pipeline for NGS Expression Level Extraction

miR-SEA: miRNA Seed Extension based Aligner Pipeline for NGS Expression Level Extraction

Gianvito Urgese[1], Giulia Paciello[1], Claudio Isella[2], Enzo Medico[2,3], Enrico Macii[1], Elisa Ficarra[1], and Andrea Acquaviva[1]

[1]Department of Control and Computer Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy
[2]Laboratory of Oncogenomics, Institute for Cancer Research at Candiolo (IRCC), Strada Provinciale 142, Candiolo, Italy
[3]Department of Oncology, Universita' di Torino, Candiolo, Turin, Italy
{gianvito.urgese,giulia.paciello,enrico.macii elisa.ficarra,andrea.acquaviva}@polito.it
{claudio.isella,enzo.medico}@ircc.it

Abstract. The advent of Next Generation Sequencing (NGS) technology has enabled a new major approach for micro RNAs (miRNAs) expression profiling through the so called RNA-Sequencing (RNA-Seq). Different tools have been developed in the last years in order to detect and quantify miRNAs, especially in pathological samples, starting from the big amount of data deriving from RNA sequencing. These tools, usually relying on general purpose alignment algorithms, are however characterized by different sensitivity and accuracy levels and in the most of the cases provide not overlapping predictions. To overcome these limitations we propose a novel pipeline for miRNAs detection and quantification in RNA-Seq sample, miRNA Seed Extension Aligner (miR-SEA), based on an experimental evidence concerning miRNAs structure. The proposed pipeline was tested on three Colorectal Cancer (CRC) RNA-Seq samples and the obtained results compared with those provided by two well-known miRNAs detection tools showing good ability in performing detection and quantification more adherent to miRNAs structure.

Keywords: miRNA, miRNA seed, RNA Sequencing, alignment algorithms, miRNA Expression Level, NGS, miR-SEA.

1 Introduction

Nowadays miRNAs are considered one of the most important regulatory class of non coding RNAs being involved in processes such as organismal development, establishment and maintenance of tissue differentiation [9] and capable to modulate in humans the expression of more than 60% of the known transcripts [10]. Furthermore, the impact of disregulated miRNAs expression levels has been recently investigated in different pathologies such as cardiovascular diseases [16], autoimmune disorders [17] and cancers [18] with remarkable impact related to diagnosis and therapeutic strategies.

The introduction of Next Generation Sequencing (NGS) technology has enabled a new major approach for miRNAs expression profiling through the so called RNA-Seq. The application of bioinformatics pipelines and tools to elaborate this data can account for the detection of both known and novel miRNAs, for their relative quantification by means of a digital approach (reads counts) and finally for their precise sequence identification [22] [24].
Different tools have been developed during the last years in order to detect known and novel miRNAs starting from RNA-Seq data. Among them miRDeep [24], miRanalyzer [26] [27], miRExpress [28], miRTRAP [29], DSAP [30], mirTools [31], MIReNA [32], miRNAkey [33] and mireap [34] have been recently compared showing quite different performances with refer to sensitivity and accuracy levels and not totally overlapping miRNAs output sets [25]. One common limitation of these tools is that they rely on general purpose alignment algorithms that does not take into account specific characteristics of miRNAs sequences.

To overcome this limitation, which impacts the accuracy of miRNA reads mapping process, in this paper we propose a novel pipeline for miRNAs identification and quantification in RNA-Seq samples based on a miRNA-specific alignment strategy. The whole work-flow is based on an experimental evidence concerning miRNAs structure extensively discussed in previous researches by Lewis et al. [35], Brennecke et al. [36] and Krek et al. [37].
These studies demonstrated the fundamental relevance of a perfect match between a region called *seed* centred on miRNAs nucleotides 2-8 and the 3'-UTR mRNA sequence for miRNA target recognition and transcript repression (*Fig. 1*). In particular miRNAs belonging to the same family were proven to present the same seed sequence and to share numerous conserved targets [10]. This distinctive feature is considered by almost all the target prediction tools that exploit the exact seed match during miRNAs targets search activity as discussed by Herrera et al. [11]. Exploiting these considerations, we developed a novel computational approach, named miRNA Seed Extension Aligner (miR-SEA), that can be used to detect and quantify miRNAs in biological samples. miR-SEA is essentially based on miRNAs seed sequence match, ungapped extension with up to two drops between miRNAs and *tags* (unique read sequences obtained from the initial set of reads), and minimum alignment size.

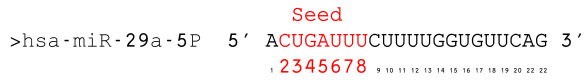

Fig. 1. miRNA seed position into hsa-miR-29a-5P.

We report the results of miRNAs detection and quantification experiments conducted on three RNA-seq CRC tissue samples. We compare the *tag counts* of miR-SEA against two well-known miRNA detection tools, highlighting its capability of performing detection and quantification more adherent to miRNAs structure.

2 Background

The discovery in 1993 by Lee et al. [1] that C-elegans lin-4 gene, known to have a role in the control of larval development, does not account for a protein but rather codify two transcripts approximately 22 nt and 61 nt long complementary to a repeated sequence element in the 3'-UTR of lin-41 mRNA, lay the foundations for a series of studies concerning this new class of novel RNAs called miRNAs involved in sequence-specific post transcriptional regulatory mechanisms [2].

miRNAs are about 22 nt long molecules highly conserved throughout evolution [3] found both in non-coding regions of genes or at intergenic loci [2] [4] [5]. They derive from longer transcripts called primary miRNAs (pri-miRNAs) with a typical hairpin structure that allows their recognition and cleavage by the Drosha enzyme inside the nucleus in a 70 nt long transcript, the precursor miRNA (pre-miRNA). Once exported to the cytoplasm pre-miRNA is processed by another RNase III enzyme, Dicer, resulting in a 22 bp long double stranded RNA molecule: One of these two strands originate the mature miRNA that after binding the RNA-induced silencing complex (RISC) [6] is known to interact with complementary mRNA targets [3].

Even if some cases of positive regulation have been recently discussed [7] [8], generally miRNAs are involved in mRNA genes expression downregulation. In particular, as proven by Hutvagner and al. [15], their perfect complementarity with a mRNA target can account for transcript cleavage whereas an imperfect complementarity is able to repress productive translation. With the main objective of detecting and quantifying known miRNAs in biological samples different molecular techniques have been applied over the past few years. Western blot analysis [21], polymerase chain reaction [20], microarray hybridization [19] and Sanger sequencing [23] have been widely investigated revealing not negligible pitfalls [22] such as the need for specific short primers in PCR experiments, the difficulty in assessing absolute quantifications with refer to microarrays or the inability in detecting miRNAs with low expression levels for what is concerning Sanger sequencing [12].

With respect to currently available tools, miR-SEA introduces six innovative features: i) Being based on seed perfect match miRNAs families can be identified with high accuracy; ii) The algorithm used to perform the alignment allows to distinguish among the different members of the same miRNA's family that are characterized, as said, by the same seed; iii) Being based on an accurate miRNA's model, already in the first phases of the method, it is possible to discard those *tags* not compatible with miRNA's structure, reducing the computational costs required for the analysis; iv) The pipeline is implemented in a customizable way in order to let users free to tune the parameters conveniently; v) No limitations on *tags* lengths have been introduced making the pipeline suitable for the analysis of different kind of datasets; vi) The standalone implementation of miR-SEA makes the proposed pipeline very user-friendly and independent from third parties tools. However, being developed in C++ language with the support of SeqAn library [38] miR-SEA is very fast and easy to customize.

3 Materials and Methods

3.1 miRNAs reference database and reads file elaboration

miRNAs sequences were downloaded from the microRNA Repository miRBase release 18.0 [39]. This set includes 21643 mature miRNAs from 168 different species among animals, plants and bacteria. Most of these miRNAs were experimentally identified or verified whereas others were computationally predicted. These miRNAs sequences represent our reference database that has to be accessed during *tags* mapping operations in order to correctly assign each *tag* to a specific miRNA. Three RNA-Seq Colorectal Cancer (CRC) datasets, produced using Illumina TruSeq technology [41], constituted the input on which we ran the proposed pipeline (*Fig. 2*). Firstly every set of reads, originally in fastq format, was converted in fasta format thanks to an ad hoc perl script and then processed taking advantage of the FASTX-Toolkit [40] in order to remove from the reads the adapter sequences introduced in the library preparation process. Once again using the FASTX-Toolkit [40] each set of reads conveniently clipped from the adapter, was elaborated to obtain the so called table of counts which reports all the unique read sequences detected in the sample, called *tags*, with their occurrence (*tag count*).

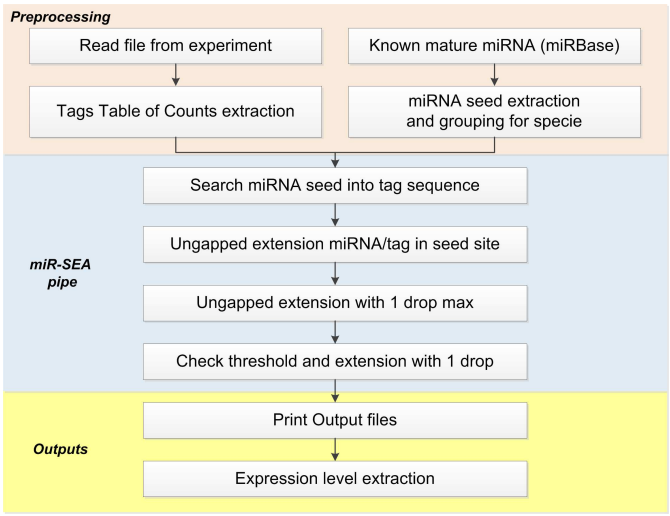

Fig. 2. miR-SEA pipeline representation.

3.2 miR-SEA workflow

miR-SEA is implemented in C++ and uses the SeqAn library that is distributed under the BSD license [38]. SeqAn library usage allows miR-SEA pipeline to take advantage of many optimized functions making the whole code modular from a structure point of view and fast in performing even computationally expensive

tasks. Input files of miR-SEA consist, as said, of a miRBase [39] fasta file and a *tags* table of counts file storing all the unique *tags* extracted from the biological sample. The pipeline, shown in *Fig. 2*, is essentially composed of three main steps that will be detailed in the following.

miR-SEA miRNAs reference database processing. In this phase seeds are searched and extracted from the miRNAs sequences contained in the reference database. All the miRNAs belonging to the same species and family sharing the same seed are here grouped and the collected information further elaborated in order to obtain a new file. This file reports for each detected seed different information such as the raw seed sequence, a unique identification code representing the species and finally a list of all the miRNAs characterized by the same seed sequence and so belonging to the same family.

Tags alignment with miR-SEA. Each extracted seed is searched, in this step, into the *tags* file. If the seed is detected in the *tag*, the alignment is extended without gaps or drops in both directions, on the matching miRNAs. The start match position into the *tag* sequence is so carefully evaluated by checking if compatible with the standard miRNA seed position as discussed in the following. *Tags* having a seed in incompatible positions with respect to miRNAs structure are discarded as shown in *Fig. 3.F*. All the compatible *tags* are then further extended in 3' direction allowing for the presence of a drop (mismatch). A threshold is imposed in order to consider an alignment: *Tags* not satisfying the threshold value will be not considered in the next phases of the proposed pipeline as depict in *Fig. 3.G*. whereas *tags* satisfying the aforementioned criteria will be once a time extended with the possibility to get another mismatch. With the main objective of clarify and deepen miR-SEA alignment policies we report in *Fig. 3* an overview of eight different situations that might occur during *tags* mapping operations, all involving *hsa-miR-29a-5P* miRNA. A common notation is used for all the examples: miRNA sequence is reported as first line, the *tag* as third line and between them, in the middle an alignment representation. The numbered line in each subfigure depicts instead the algorithm phase in which

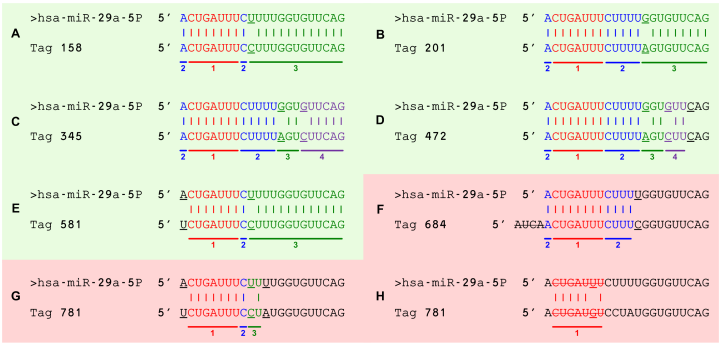

Fig. 3. Cases handled by miR-SEA.

sub-sequences have been evaluated: i) With numbered 1 red line is represented miRNA seed sequence searching activity in the *tag*; ii) With numbered 2 blue line the seed extension step in 5' and 3' directions until a first mismatch is found; iii) With numbered 3 green line the alignment extension in 3' direction until the second mismatch; iv) With numbered 4 violet line the new extension step performed if a minimum alignment length is reached in the previous phases, until a last allowed mismatch. In the reported examples miR-SEA parameters were setted as $Size_{min} = 13$ (minimum alignment length for last extension) and $Begin\_Tag_{min} < 4$ (start *tag* position alignment on miRNA seed). The choice of these values led to discard alignments represented with light red background in *Fig. 3* and to consider those characterized by the light green one. In *Fig. 3.A* is reported a *tag* alignment with a a mismatch in position 10 that lead the algorithm to complete the mapping procedure during its second phase because no other mismatches are detected between miRNA and *tag* sequences. A mismatch in the *tag* position 14 conducted miR-SEA to assume the same behaviour of *Fig. 3.A* as depicted in *Fig. 3.B*. *Fig. 3.C* shows an alignment executed trough all the four phases of the proposed algorithm because two mismatches are found in the *tag* sequence in positions compatible with the imposed parameters. In *Fig. 3.D* alignment is instead completed in correspondence of the third gap at position 20. A mismatch in *tag* position 1 allows miR-SEA with the imposed settings to proceed in the alignment as depicts *Fig. 3.E*. According to $Begin\_Tag_{min} < 4$, *tag* reported in *Fig. 3.F* is discarded after the second algorithm step since *tag* alignment start position exceeds the third nucleotide; this condition was carefully evaluated because a high number of nucleotides flanking seed 5' position on the *tag* could account for a different kind of RNA structure such as a pre-miRNA since it is known that the seed start position on mature miRNAs is generally located at its second nucleotide as stated by Bartel et al. [42]. *Fig. 3.G* shows a case in which the *tag* is discarded because the third phase of the algorithm detected a second mismatch in position 12 not compatible with the imposed $Size_{min}$. Also this choice derived from experimental observations: As proven by Bartel et al. [42] indeed sequences having two mismatches too close with respect to seed sequence and to their relative positions are not compatible with miRNAs structure. Finally in *Fig. 3.H* is reported the case in which a mismatch occurs in the seed sequence leading to the immediate *tag* discarding.

miR-SEA output files generation. Finally all the *tags* satisfying the previously described criteria are saved in two files according with the kind of alignment detected. In particular if a *tag* is aligned with only a miRNA, it is saved into a unique *tag* file (that share the same format as miRanalyzer [27] output file), instead if the *tag* is aligned with multiple miRNAs then it is saved into an ambiguous *tag* file containing a complete list of *tag* informations and the miRNAs sequences on which the *tag* has been mapped. Concerning the miRNAs expression file we decided to be compliant with miRanalyzer [27] output file with some added features. Two miRNAs expression files have been produced: One reporting the information related to the ambiguous matches (*tags* mapped on more than one miRNA) and one with data related to the unique matches (*tags* mapped solely on a miRNA).

4 Results and Discussion

In order to evaluate miR-SEA performances on correctly assigning *tags* sequences to known miRNAs, an initial analysis of the predictions provided by two well known miRNAs alignment tools that are miRanalyzer [27] and miRExpress [28] was performed on the aforementioned three CRC RNA-Seq datasets. All the tests were executed imposing default running parameters and considering as reference mirBase release 18 database [39]. In *Fig. 4* are reported respectively for Sample 1, 2 and 3 on the y-axis the logarithmic read counts differences between miRanalyzer [27] and miRExpress [28] observed for the miRNAs on the x-axis. As it is possible to note *tag* counts differences range for the detected miRNAs among samples. The maximum values are equal to 1319967 for Sample 1, 761623 for Sample 2 and 360285 for Sample 3, whereas the averages are respectively 3712, 1237 and 1037.

As discussed in the Introduction section miRNAs detection tools are indeed built on top of different alignment tools or methods that lead to different sensitivity and specificity levels of the detection, determining predictions that generally poorly overlap. To deepen this aspect evaluating the effective ability of

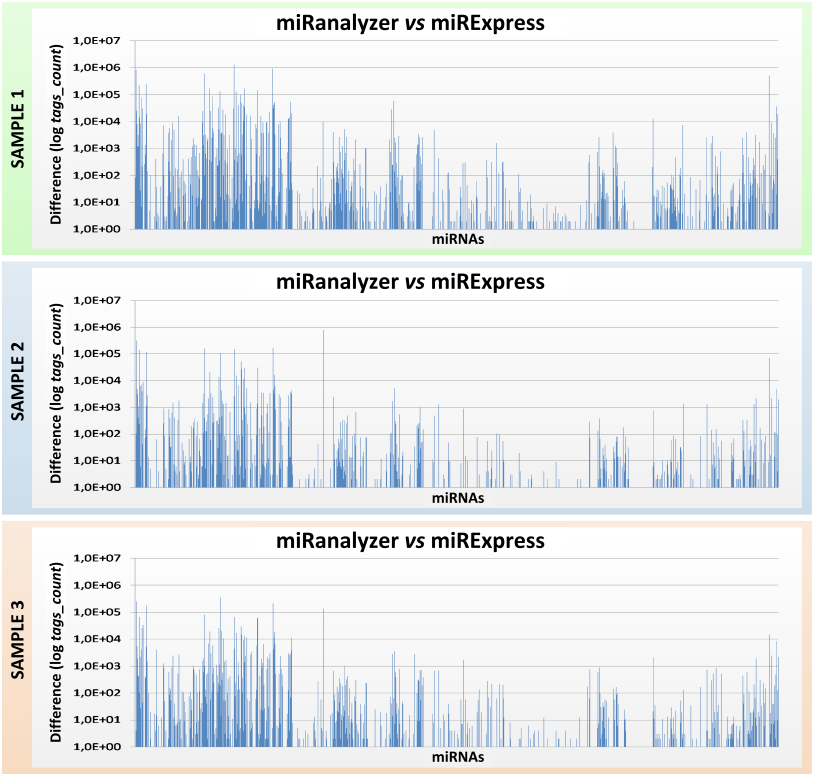

Fig. 4. Logarithmic difference in *tag counts* between miRanalyzer and miRExpress.

the different algorithms in attribute the correct miRNA to a data *tag* a new experimental setup has been built. The alignment of mirBase 20 [39] miRNAs sequences has been performed on the same mirBASE miRNAs sequences and a *tag* considered properly mapped if univocally assigned to itself. In this phase different tools and algorithms have been tested, both specific for miRNAs detection and for general purpose alignments. The ability in assigning the correct miRNA to a specific *tag* was assessed by running Blast with default parameters setting and with word size equal to 16, miRExpress and mirAnalyzer with default parameters and miR-SEA with different configurations. These analyses led to quite different results in relation to the correct miRNAs assignments as shown in *Fig. 5*. The best outcome was reached by miRExpress [28] that correctly assigned 2503 *tags* among the human 2578 of the miRBase [39] database, whereas ambigue predictions were detected especially by miRanalyzer [27].

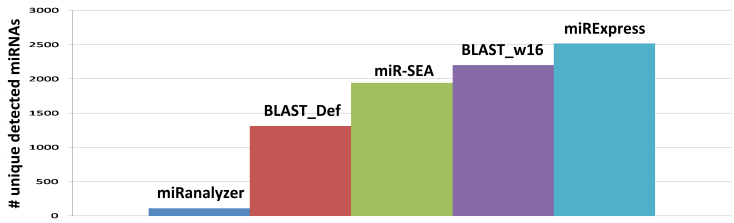

Fig. 5. Performance comparison of different aligner tools in correctly assigning mirBase *tags* to mirBase reference.

The results presented in *Fig. 6* for what is concerning miR-SEA were obtained by selecting, among the different tested configurations, that capable to provide the better prediction, in other terms a higher number of *tags* correctly assigned to itself. In particular 32 run were performed by attributing different values to two selected parameters that are $Size_{min}$ and $Begin\_Tag_{min}$, leading to the worst prediction of 1573 correctly assigned *tags* over the total 2578 in case of values respectively equal to 10 and 8 and to the best of 1946 if $Size_{min}$ equal to 16 and $Begin\_Tag_{min}$ to 2.
Starting from these considerations all data presented in the following have been obtained by selecting 16 for the $Size_{min}$ parameter and 2 for the $Begin\_Tag_{min}$.

In *Fig. 6* are shown respectively for Sample 1, 2 and 3 on the different rows, the logarithmic cross-correlation values between miR-SEA *tags* counts on the x-axis and miRanalyzer or miRExpress *tags* counts on y-axis. They can be explained by sectioning the respective areas in five main portions. In particular: i) Dots located on the x-axis account for miR-SEA *tags* assignments to specific miRNAs and at the same time for no *tags* assignments performed by the other tool (i.e. miRExpress or miRanalyzer) to the same miRNAs; ii) Dots placed on the y-axis explain the opposite prediction, that is *tags* mapped on a given miRNA detected only by the compared tool; iii) The 45^o line reports those miRNAs identified as equally expressed by both miR-SEA and the compared tool;

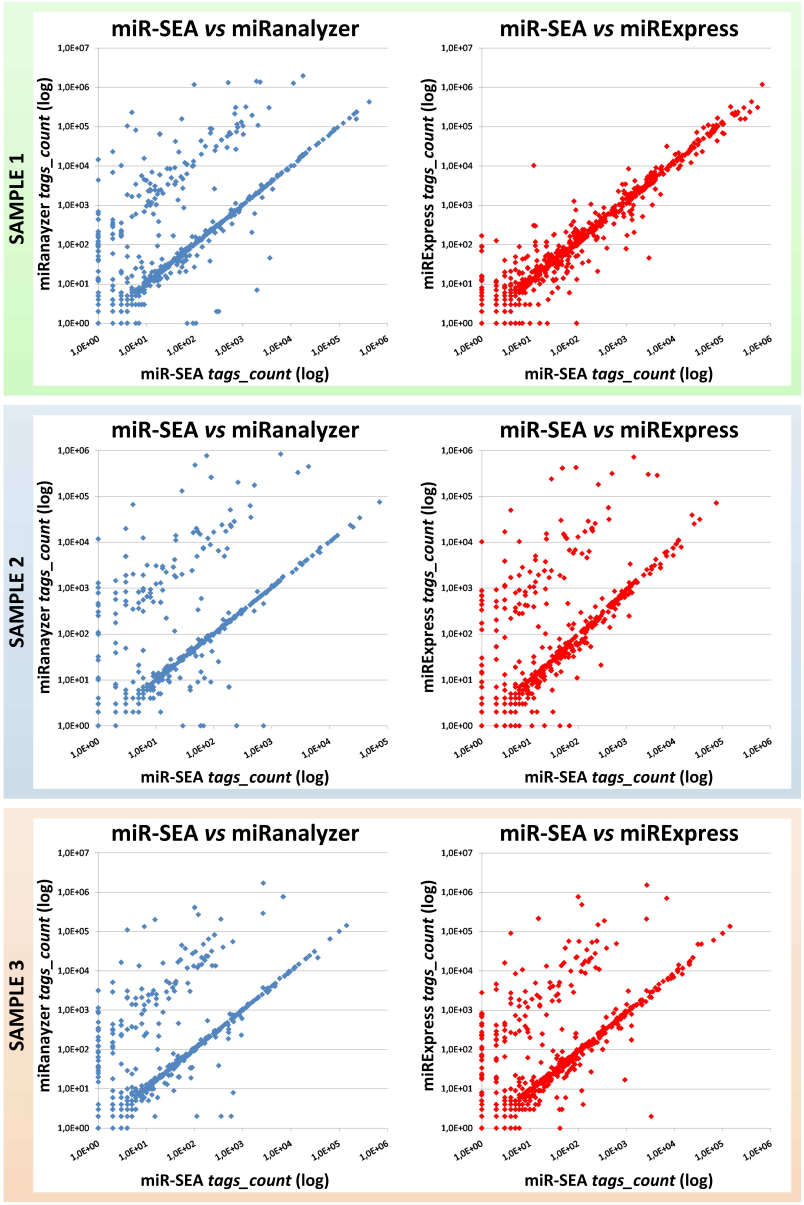

Fig. 6. Logarithmic cross-correlation between miR-SEA vs miRanalyzer and miR-SEA vs miRExpress *tags* counts.

iv) The area above the 45^o line includes all the miRNAs detected as more expressed (*tags* scored) by the competitor tool, probably because the more relaxed *tags* mapping approaches implemented by the different algorithms, don't requiring for example a perfect seed sequence match; v) The zone under the 45^o line reports those miRNAs that miR-SEA detected as more *tags scored* than its op-

posite tool. Predictions belonging to this area can be considered very accurate because of the deterministic manner of detecting in the *tags* miRNA seed sequence in the correct position and the minimum *tag* alignment size required.
It is worth noting however from these graphs that miR-SEA with respect to miRanalyzer account for lower miRNA expression levels attributing a minor number of *tags* to the same miRNAs. Since at the days we could not found totally experimentally validated RNA-Seq datasets we only speculate that this behaviour is a consequence of the higher specificity of miR-SEA that derives from an accurate miRNA model implementation based on experimental evidences. In relation to miR-SEA and miRExpress [28] cross-correlation values, the predictions are quite similar if Sample 1 is considered whereas a trend analogous to those observed between miRanalyzer [27] and miR-SEA is evident in Sample 2 and 3. This behaviour can be attributed once a time to the different alignment algorithms implemented in the tools.

5 Conclusions and Future Works

miR-SEA is a software pipeline that aim to detect and quantify miRNAs in RNA-Seq samples. Being based on strong experimental evidences about miRNAs structure, miR-SEA is capable to guarantee high accuracy in miRNAs detection and expression evaluation. In particular the perfect seed matching requirement for the assignment of a *tag* to a specific miRNA, allows a reliable identification of the miRNAs family represented by the *tag*. Furthermore this activity lead to discard, a conspicuous number of not significant *tags*, allowing a reduction of the computational costs required for the following steps execution. For what is concerning its implementation, the use of SeqAn library with optimized integrated functions permit to reduce once a time the computational costs. The totally customizable structure of miR-SEA therefore let users free to tune conveniently the running parameters and to perform analysis on *tags* of any length. Moreover since not built on top of any tools it can be used without installing other softwares. We would like to underline finally that the predictions provided by miR-SEA can also be combined with those deriving from other different miRNAs detection tools with the main objective of attribute an higher level of confidence to the identified miRNAs being miR-SEA pipeline, as already highlighted, implemented considering strong biological evidence for what is concerning miRNA structure. This evaluation can be easily done by calculating the cross-correlation between miR-SEA (x-axis) and the other tool (y-axis) *tags* counts for a specific miRNA as previously shown in *Fig. 6*. In particular if the values deriving from this comparison lay on the 45^o line it is possible to affirm that also the other tool performed *tags* assignments to miRNAs by considering biological miRNAs features. If they are instead located upper with respect to this line, miRNAs *tags* assignments is not performed by considering the miRNAs biological evidences we evaluated in the proposed pipeline. Predictions obtained running miR-SEA on the three CRC datasets before presented will be however experimentally proven by performing in lab analyses, in order to assess carefully miR-SEA performances (in terms of accuracy and specificity) in detecting miRNAs in RNA-Seq samples.

References

1. Lee, R.C., Feinbaum, R.L., Ambros, V.: The C. elegans Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to lin-14. Cell. 75, 843–854 (1993)
2. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T.: Identification of Novel Genes Coding for Small Expressed RNAs. SCIENCE. 294, 853–858 (2001)
3. Bartel, D.P.: MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell. 116, 281–297 (2004)
4. Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P.: An Abundant Class of Tiny RNAs with Probable Regulatory Roles in Caenorhabditis elegans. SCIENCE. 294, 858–862 (2001)
5. MacFarlane, L.A., Murphy, P.R.: MicroRNA: Biogenesis, Function and Role in Cancer. Science. 297, 2056–2060 (2002)
6. Mourelatos, Z., Dostie, J: miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs.. Genes Dev. 16, 720–728 (2002)
7. Vasudevan, S., Tong, Y., Steitz, J.A.: Switching from Repression to Activation: MicroRNAs Can Up-Regulate Translation. Science. 318, 1931–1934 (2007)
8. Orom, U.A., Nielsen, F.C., Lund, A.H.: MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. Mol Cell. 30, 460–471 (2008)
9. Alvarez-Garcia, I., Miska, E.A.: MicroRNA functions in animal development and human disease. Development. 132, 4653–4662 (2005)
10. Friedman, R.C., Kai-How Farh, K., Burge, C.B., Bartel, D.P.: Most mammalian mRNAs are conserved targets of microRNAs. Genome Research, 19, 92–105 (2009)
11. Reyes-Herrera, P.H., Ficarra, E.: One Decade of Development and Evolution of MicroRNA Target Prediction Algorithms. Genomics Proteomics Bioinformatics, 10, 254–263 (2012)
12. Mendes, N.D., Freitas, A.T., Sagot, M.F.: Current tools for the identification of miRNA genes and their targets. Nucleic Acids Res., 37, 2419–2433 (2009)
13. Wienholds, E., Kloosterman, W.P., Miska, E., Alvarez-Saavedra E., Berezikov E., de Bruijn E., Horvitz, H.R., Kauppinen S., Plasterk R.H.: MicroRNA expression in zebrafish embryonic development. Science. 309 (2005)
14. Krol, J., Loedige, I., Filipowicz, W.: The widespread regulation of microRNA biogenesis, function and decay. Nature Reviews Genetics. 11, 597–610 (2010)
15. Hutvgner G., Zamore P.D.: A microRNA in a multiple-turnover RNAi enzyme complex. Current Genomic. 11, 537–561 (2001)
16. Small, E.M., Frost, R.J.A., Olson, E.N.: MicroRNAs Add a New Dimension to Cardiovascular Disease. Circulation, 1022–1032 (2010)
17. Tili, E., Michaille, J.J., Costinean, S., Croce, C.M.: MicroRNAs, the immune system and rheumatic disease. Nat Clin Pract Rheumatol., 4, 534–541 (2008)
18. Iorio, M.V., Croce, C.M.: MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. EMBO Mol.Med. 4, 143-159 (2012)
19. Barad, O., Meiri, E., Avniel, A., Aharonov, R., Barzilai, A. et al. : MicroRNA expression detected by oligonucleotide microarrays: System establishment and expression profiling in human tissues. Genome Research, 14, 2486–2494 (2004)
20. Peltier, H.J., Latham, G.J.: Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. RNA, 14, 844–852 (2008)
21. Turchinovich, A., Weiz, L., Langheinz, A., Burwinkel, B.: Characterization of extracellular circulating microRNA. Nucleic Acids Research, 39, 7223–7233 (2011)

22. Pritchard, C.C., Cheng, H.H., Tewari, M.: MicroRNA profiling: approaches and considerations. Nature, 13, 358–369 (2012)
23. Aravin, A., Tuschl, T.: Identication and characterization of small RNAs involved in RNA silencing. FEBS Letters, 579, 5830–5840 (2005)
24. Friedlander, M.R., Chen, W., Adamidi, C., Maaskol, J., Einspanier, R., Knespel1, S., Rajewsky1, N.: Discovering microRNAs from deep sequencing data using miRDeep. Nature Biotechnology, 26, 407–415 (2008)
25. Li, Y., Zhang, Z., Liu, F., Vongsangnak, W., Jing, Q., Shen, B.: Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis. Nucleic Acids Research, 40, 4298–4305 (2012)
26. Hackenberg, M., Sturm, M., Langenberge, D., Falcn-Prez, J.M., Aransay, A.M.: miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Research, 37, W68–W76 (2009)
27. Hackenberg, M., Rodriguez-Ezpeleta, N., Aransay, A.M.: miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. Nucleic Acids Research, 39, W132–W138 (2011)
28. Wang, W.C., Lin, F.M, Chang, W.M., Lin, K.Y., Huang, H.D., Lin, N.S.: miRExpress: Analyzing high-throughput sequencing data for profiling microRNA expression. Nucleic Acids Research, 10, 328 (2009)
29. Hendrix, D., Levine, M., Shi, W.: miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. Genome Biology, 11, R39 (2010)
30. Huang, P.J., Liu, Y.C., Lee, C.C., Lin, W.C., Gan, R.R., Lyu, P.C., Tang, P.: DSAP: deep-sequencing small RNA analysis pipeline. Nucleic Acids Res., 38, W385-91 (2010)
31. Zhu, E., Zhao, F., Xu, G., Hou, H., Zhou, L., Li, X., Sun, Z., Wu, J.: mirTools: microRNA profiling and discovery based on high-throughput sequencing. Nucleic Acids Res., 38, W392-97 (2010)
32. Mathelier, A., Carbone, A.: MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. Bioinformatics, 26, 2226-34 (2010)
33. Ronen, R., Gan, I., Modai, S., Sukacheov, A., Dror, G., Halperin, E., Shomron, N.: miRNAkey: a software for microRNA deep sequencing analysis. Bioinformatics, 26, 2615-6 (2010)
34. mireap Tool, `http://sourceforge.net/projects/mireap/`
35. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Barte, D.P., Burge, C.B.: Prediction of mammalian microRNA targets. Cell, 26, 787-98 (2003)
36. Brennecke, J., Stark, A., Russell, R.B., Cohen, S.M.: Principles of MicroRNA-Target Recognition. PLOS Biology, 3, 404-418 (2005)
37. Krek, A.: Combinatorial microRNA target predictions. Nature Genetics, 37, 495-500 (2005)
38. Doring, A., Weese, D., Rausch, T., Reinert, K.: SeqAn an efficient, generic C++ library for sequence analysis. BMC Bioinformatics, 9-11 (2008)
39. Bartel, D.P.: MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, no. 2, 281-297 (2004) `http://www.mirbase.org/`
40. FASTX-Toolkit, `http://hannonlab.cshl.edu/fastx_toolkit/index.html`
41. Truseq, `http://www.illumina.com/truseq.ilmn`
42. Bartel, David P. "MicroRNAs: target recognition and regulatory functions." Cell 136, no. 2, 215-233 (2009)