



Consiglio Nazionale delle Ricerche  
Istituto di Calcolo e Reti ad Alte Prestazioni

# **miRNAs differential expression analysis in healthy and BC patients subjected to a different dietary treatment, and investigation of their functional role by functional enrichment analysis.**

A. Fiannaca, M. La Rosa, L. La Paglia, R. Rizzo, A. Urso

***Rapporto Tecnico N.:***  
**RT-ICAR-PA-16-04**

**febbraio 2016**



Consiglio Nazionale delle Ricerche, Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR)  
– Sede di Cosenza, Via P. Bucci 41C, 87036 Rende, Italy, URL: [www.icar.cnr.it](http://www.icar.cnr.it)  
– Sede di Napoli, Via P. Castellino 111, 80131 Napoli, URL: [www.na.icar.cnr.it](http://www.na.icar.cnr.it)  
– Sede di Palermo, Viale delle Scienze ed.11, 90128 Palermo, URL: [www.pa.icar.cnr.it](http://www.pa.icar.cnr.it)



Consiglio Nazionale delle Ricerche  
Istituto di Calcolo e Reti ad Alte Prestazioni

# miRNAs differential expression analysis in healthy and BC patients subjected to a different dietary treatment, and investigation of their functional role by functional enrichment analysis.

A. Fiannaca<sup>1</sup>, M. La Rosa<sup>1</sup>, L. La Paglia<sup>1</sup>, R. Rizzo<sup>1</sup>, A. Urso<sup>1</sup>

**Rapporto Tecnico N.:**  
**RT-ICAR-PA-16-04**

**Data:**  
**febbraio 2016**

---

<sup>1</sup> Istituto di Calcolo e Reti ad Alte Prestazioni, ICAR-CNR, Sede di Palermo, Viale Delle Scienze ed.11, 90128 Palermo.

*I rapporti tecnici dell'ICAR-CNR sono pubblicati dall'Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche. Tali rapporti, approntati sotto l'esclusiva responsabilità scientifica degli autori, descrivono attività di ricerca del personale e dei collaboratori dell'ICAR, in alcuni casi in un formato preliminare prima della pubblicazione definitiva in altra sede.*

I microRNA (miRNA) sono piccole molecole di RNA a singolo filamento non codificante, della lunghezza di 22-25 nt, che regolano negativamente l'espressione genica a livello post-trascrizionale (1) mediante appaiamento con specifici RNA messaggeri (mRNA), inducendone la degradazione o impedendone la traduzione nel corrispondente prodotto proteico (2,3). Essi sono espressi in un numero molto elevato di organismi, incluso l'*Homo Sapiens*, sono filogeneticamente molto conservati, e possiedono un'elevata specificità di tessuto e dei vari stadi di sviluppo. E' stato di recente dimostrato il ruolo cruciale dei microRNA nella regolazione di importanti processi fisiologici, quali lo sviluppo, la proliferazione, il differenziamento, l'apoptosi e la risposta a differenti segnali extracellulari e di stress (4). Diverse evidenze mostrano inoltre una espressione differenziale tessuto-specifica, durante specifici stadi di sviluppo, o in alcune patologie (5) quali ad esempio quella tumorale. Essi infatti presentano una variazione di espressione nei tumori, e sembrerebbero essere coinvolti nel controllo di vie di segnalazione intracellulare che hanno un'influenza diretta sulla crescita cancerosa, come ad esempio la proliferazione e l'apoptosi e potrebbero dunque funzionare da oncogeni o da oncosoppressori. Queste evidenze lasciano supporre che i diversi livelli di espressione genica di uno specifico miRNA possano influenzare diversi pathways, sia in maniera positiva che negativa e spiegano anche il fatto che i livelli di espressione dell'mRNA e della proteina codificata da quel messaggero possano essere diversi (6-9).

L'importanza del ruolo che i miRNA svolgono nella regolazione dell'espressione genica è dimostrata anche dal fatto che pochi miRNA possono agire su molteplici mRNA (targets) lasciando dedurre il potenziale ruolo di queste piccole molecole come regolatori dei processi biologici, e in diversi contesti come nel cancro. Infatti, essendo un singolo miRNA in grado di regolare l'espressione genica di centinaia di geni bersaglio, la sua alterata espressione cellulare porterà alla deregolazione di numerosi pathways molecolari che potrebbero contribuire, separatamente o sinergicamente, all'instaurarsi di un fenotipo cellulare di natura maligna (10-12).

Sebbene la ricerca si sia focalizzata negli ultimi decenni sul potenziale ruolo regolativo dei miRNA nel processo tumorale, più di recente si è ipotizzato che queste molecole potessero caratterizzare coorti diverse di pazienti affetti e non da tumore trattati con diversi trattamenti alimentari. Ciò perché recenti studi mirano a legare diversi regimi alimentari ad outcome e prognosi diversi nella patologia tumorale.

Diversi lavori hanno evidenziato alcuni miRNA coinvolti nello stress sensibile ai nutrienti nelle piante (13) Come detto precedentemente i miRNA sono importanti elementi regolativi nella risposta allo stress ambientale ed intervengono nel ripristinare l'omeostasi in caso di cambiamenti ambientali repentini o in caso di severe condizioni di stress di lunga durata. Essi dunque rappresentano un meccanismo di riprogrammazione dell'espressione genica e di adattamento cellulare (14-15).

Inoltre è stata evidenziata una differente espressione di specifici miRNA legata a specifici nutrienti come aminoacidi, carboidrati, acidi grassi, vitamine ecc. Queste piccole molecole potrebbero essere dunque dei potenziali agenti terapeutici nelle patologie legate alla dieta e potrebbero essere usati come markers nella valutazione dello stato nutrizionale in studi di intervento alimentare (16)

Da tutte queste evidenze, nell'ambito del progetto MIRCO si è voluto valutare l'assetto di microRNA in coorti di pazienti affette e non da carcinoma della mammella, prima e dopo intervento alimentare per evidenziare la presenza di specifici miRNA dieta-correlati, e dunque per valutare se un diverso regime alimentare potesse

influire sull'assetto genomico e se le diverse coorti (affette e non affette da Ca. Ma.) potessero essere caratterizzate da un punto di vista molecolare da diversi tipi di marcatori. A tal fine oltre alla valutazione dell'espressione differenziale dei microRNA, si è voluta effettuare una valutazione di tipo funzionale attraverso un'analisi di "gene enrichment".

Lo studio proposto è stato suddiviso in diverse fasi:

- **Valutazione analitica qualitativa e quantitativa dei dati – analisi di espressione differenziale**
- **Analisi di enrichment funzionale sui dati analizzati**

Per ciò che riguarda la valutazione qualitativa e quantitativa dei dati, inizialmente è stata effettuata un'analisi di espressione differenziale di miRNA in coorti diverse di pazienti (affette e non affette da carcinoma della mammella (Ca. Ma.); trattate e non con diverso regime alimentare). L'analisi è stata effettuata su campioni provenienti da metodica di next generation sequencing (NGS) e open array. E' stata effettuata analisi di clustering mediante algoritmo DESeq, il quale permette di effettuare un'analisi differenziale di espressione genica basandosi sulla distribuzione binomiale negativa al fine di associare in maniera specifica miRNAs e specifiche caratteristiche cliniche patologiche dei soggetti analizzati (per le coorti di soggetti affetti da tumore mammario).

Nello specifico è stata effettuata l'analisi del miRnoma di pazienti sane controllo (non trattate con trattamento alimentare) rispetto a pazienti sane trattate, e un'analisi del miRnoma di pazienti affette da carcinoma della mammella controllo (non trattate con trattamento alimentare) rispetto a pazienti affette da tumore mammario trattate. Dall'analisi è stata rilevata la over-espressione e down-espressione di particolari miRNAs in specifiche coorti. I dati relativi all'analisi sono riportati in tabella 1.

sano C vs sano L	sano C vs sano H	tumor c vs tumor L	tumor c vs tumor H
Novel:10_535	Novel:3_12120	Novel:10_594	Novel:17_5440
Novel:14_3049	Novel:X_17965	Novel:18_6499	Novel:7_16124
Novel:2_11051	hsa-mir-1270	Novel:3_12824	hsa-mir-1197
Novel:5_14080	hsa-mir-1301	Novel:7_15579	hsa-mir-1277
hsa-mir-3135a	hsa-mir-219b	hsa-mir-452	hsa-mir-141
hsa-mir-6782	hsa-mir-28	hsa-mir-5188	hsa-mir-19a
	hsa-mir-3127	hsa-mir-99b	hsa-mir-19b-1
	hsa-mir-3614		hsa-mir-19b-2
	hsa-mir-3688-1		hsa-mir-340
	hsa-mir-3688-2		hsa-mir-3620
	hsa-mir-3928		hsa-mir-374a
	hsa-mir-3939		hsa-mir-452
	hsa-mir-4775		hsa-mir-4732
	hsa-mir-543		hsa-mir-580
	hsa-mir-551b		
	hsa-mir-651		
	hsa-mir-671		
	hsa-mir-6802		
	hsa-mir-6820		
	hsa-mir-708		
	hsa-mir-769		

In tabella sono riportati i profili di quattro diversi miRNomi in quanto il trattamento alimentare prevedeva l'uso di due diversi tipi di olio (Low and High) a ridotto ed elevato contenuto di polifenoli.

Legenda: C= controllo non trattato, L= low poliphenol concentration, H= high poliphenol concentration.

Al fine di dare un significato funzionale a tale risultato si è proceduto ad effettuare un'analisi di "gene enrichment" delle coorti oggetto dello studio. Sono stati utilizzati diversi tools computazionali quali target-interaction tools, risorse di Gene Ontology e analisi dei pathways.

Queste metodiche mirano rispettivamente alla ricerca di targets putativi dei miRNAs differenzialmente espressi, e di verificare se alcuni di questi targets sono correlati in maniera forte a qualche stato o processo cellulare o a delle specifiche funzioni. Ciò permetterebbe di valutare il significato prognostico o predittivo di specifici miRNAs legati alla patologia tumorale, dipendenti da diverso regime alimentare ed eventualmente di trovare nuovi markers e nuove strategie di intervento terapeutico mirate.

Come si evince dalla tabella 1, per tutte le categorie di soggetti analizzate non sono stati rilevati miRNA comuni tranne che per il miR-452. Esso è presente in entrambe le coorti di soggetti affetti da Ca. Ma. A prescindere dal trattamento alimentare e risulta essere per entrambe over-espresso. Tutti gli altri miRNA sono diversi, alcuni down- altri over-espressi. Per il dettaglio vedi tabella n.2.

DE miRNAs	SC vs SL	SC vs SH	TC vs TL	TC vs TH
Down	3 hsa-miR-378, hsa-miR-520, hsa-miR-590-5p	6 hsa-miR-151-3p, hsa-miR-29c, hsa-miR-338-5p, hsa-miR-367, hsa-miR-520e, hsa-miR-551b,	5 hsa-let-7b, hsa-miR-125b, hsa-miR-320, hsa-miR-520e, hsa-miR-636,	12 hsa-miR-16, hsa-miR-17, hsa-miR-223, hsa-miR-30a-5p, hsa-miR-30b, hsa-miR-320, hsa-miR-551b, hsa-miR-572, hsa-miR-618, hsa-miR-720, hsa-miR-766, hsa-miR-99a
up	11 hsa-miR-150, hsa-miR-151-3p, hsa-miR-155, hsa-miR-302c, hsa-miR-30b, hsa-miR-30c, hsa-miR-367, hsa-miR-520c-3p, hsa-miR-572, hsa-miR-625, hsa-miR-657	14 hsa-miR-150, hsa-miR-191, hsa-miR-223, hsa-miR-302c, hsa-miR-30b, hsa-miR-30c, hsa-miR-320, hsa-miR-483-5p, hsa-miR-484, hsa-miR-519b-3p, hsa-miR-572, hsa-miR-601, hsa-miR-618, hsa-miR-92	3 hsa-miR-1274B, hsa-miR-30, hsa-miR-942,	7 hsa-miR-132, hsa-miR-140-3p, hsa-miR-151-3p, hsa-miR-192, hsa-miR-324-3p, hsa-miR-324-5p, hsa-miR-409-3p

. Tot campioni: 86  
. Totale miRNAs: 758

L'analisi di enrichment effettuata sui quattro cluster di miRNA differenzialmente espressi ha evidenziato interessanti risultati sia per le coorti sane che affette da carcinoma della mammella legate però all'uso del solo olio H. Sia miRNomi di soggetti sani che trattati con olio H presentano infatti targets legati alle stesse classi funzionali. Le categorie funzionali evidenziate sono legate nello specifico a termini di

Gene Ontology (GO) quali regolazione della trascrizione, di processi metabolici, espressione genica, regolazione trascrizionale, ecc. (per il dettaglio vedi tabella n. 3). Le coorti trattate con olio L non presentano invece risultati rilevanti.

Tabella 3:

SC vs SH		TC vs TH	
Annotation Cluster 1	Enrichment Score: 4.689737628820823	Annotation Cluster 1	Enrichment Score: 6.652878501753605
Category	Term	Category	Term
GOTERM_BP_FAT	GO:0045935~positive regulation of nucleobase, nucleoside, nucleotide metabolic process	GOTERM_BP_FAT	GO:0045893~positive regulation of transcription, DNA-dependent
GOTERM_BP_FAT	GO:0045941~positive regulation of transcription	GOTERM_BP_FAT	GO:0051254~positive regulation of RNA metabolic process
GOTERM_BP_FAT	GO:0010557~positive regulation of macromolecule biosynthetic process	GOTERM_BP_FAT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter
GOTERM_BP_FAT	GO:0045893~positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	GO:0045941~positive regulation of transcription
GOTERM_BP_FAT	GO:0010604~positive regulation of macromolecule metabolic process	GOTERM_BP_FAT	GO:0010628~positive regulation of gene expression
GOTERM_BP_FAT	GO:0051173~positive regulation of nitrogen compound metabolic process	GOTERM_BP_FAT	GO:0045935~positive regulation of nucleobase, nucleoside, nucleotide metabolic process
GOTERM_BP_FAT	GO:0051254~positive regulation of RNA metabolic process	GOTERM_BP_FAT	GO:0010557~positive regulation of macromolecule biosynthetic process
GOTERM_BP_FAT	GO:0010628~positive regulation of gene expression	GOTERM_BP_FAT	GO:0031328~positive regulation of cellular biosynthetic process
GOTERM_BP_FAT	GO:0009891~positive regulation of biosynthetic process	GOTERM_BP_FAT	GO:0051173~positive regulation of nitrogen compound metabolic process
GOTERM_BP_FAT	GO:0031328~positive regulation of cellular biosynthetic process	GOTERM_BP_FAT	GO:0009891~positive regulation of biosynthetic process
GOTERM_BP_FAT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	GOTERM_BP_FAT	GO:0010604~positive regulation of macromolecule metabolic process
Annotation Cluster 2	Enrichment Score: 4.654875303904787	Annotation Cluster 2	Enrichment Score: 5.451555414272936
Category	Term	Category	Term
SP_PIR_KEYWORDS	transcription regulation	GOTERM_CC_FAT	GO:0005626~insoluble fraction
SP_PIR_KEYWORDS	Transcription	GOTERM_CC_FAT	GO:0005624~membrane fraction
GOTERM_BP_FAT	GO:0006350~transcription	GOTERM_CC_FAT	GO:0000267~cell fraction
		Annotation Cluster 3	Enrichment Score: 4.577699614059783
		Category	Term
		SP_PIR_KEYWORDS	transcription regulation
		SP_PIR_KEYWORDS	Transcription
		GOTERM_BP_FAT	GO:0045449~regulation of transcription
		Annotation Cluster 4	Enrichment Score: 4.266684140782111
		Category	Term
		SP_PIR_KEYWORDS	zinc-finger
		SP_PIR_KEYWORDS	zinc
		GOTERM_MF_FAT	GO:0046914~transition metal ion binding
		Annotation Cluster 5	Enrichment Score: 4.068528438825615
		Category	Term
		GOTERM_BP_FAT	GO:0010558~negative regulation of macromolecule biosynthetic process
		GOTERM_BP_FAT	GO:0009890~negative regulation of biosynthetic process
		GOTERM_BP_FAT	GO:0016481~negative regulation of transcription
		GOTERM_BP_FAT	GO:0031327~negative regulation of cellular biosynthetic process
		GOTERM_BP_FAT	GO:0010629~negative regulation of gene expression
		GOTERM_BP_FAT	GO:0045892~negative regulation of transcription, DNA-dependent
		GOTERM_BP_FAT	GO:0051253~negative regulation of RNA metabolic process
		GOTERM_BP_FAT	GO:0045934~negative regulation of nucleobase, nucleoside, nucleotide metabolic process
		GOTERM_BP_FAT	GO:0010605~negative regulation of macromolecule metabolic process
		GOTERM_BP_FAT	GO:0051172~negative regulation of nitrogen compound metabolic process

I risultati ottenuti sembrano promettenti seppur limitati ad un numero esiguo di campioni.

Dall'enrichment funzionale è stata evidenziata una lista di geni comuni ad entrambe le categorie analizzate che necessiterà di ulteriori indagini ed approfondimenti (tabella 4).

Tabella4:

Lista di geni comuni alle categorie CS vs CH e TS vs TH

PPARA	SIX4	TCF4
MITF	PPARGC1B	MAF
PAX6	NCOA3	KLF6
RORA	HIPK2	SMAD4
FOXO3	ONECUT2	CREB5
ZNF148	SOX6	NR4A3
AR	POU2F1	WWTR1
ATXN1	MDM2	SELE
PKNOX1	BMPR2	BBX
CSRNP5	REST	NR2E1
ATXN7	CNOT4	PGR
MIER1	RCOR3	TRIM33
MLL10	VGLL3	TFCP2L1
MYBL1	TFDP2	PLAG1
ZFX4	TRPS1	JAZF1
BACH2	ZNF81	ZNF280B
PURB	LCOR	SCML2
NR3C2	ETV1	CHD6
ZNF24	NFIA	

**Metodiche utilizzate:**

- **Progettazione delle procedure per l'elaborazione dei dati - analisi di espressione differenziale**

Questa attività è stata mirata all'identificazione di modelli innovativi di analisi di dati di tipo High-throughput al fine di ottenere il riconoscimento e la classificazione dei dati stessi. E' nota la difficoltà di analizzare un dataset di una condizione reale, contenente una mole di dati elevata, come quelli provenienti dalle metodiche di Next Generation Sequencing (NGS). Dato l'esiguo numero di campioni analizzato, è stato proposto come modello di analisi l'algoritmo DESeq.

Nello specifico del progetto, l'algoritmo ha permesso di associare cluster di microRNA differenzialmente espressi a particolari cluster di soggetti sottoposti a trattamenti differenti e/o con specifiche caratteristiche cliniche.

Infine, sono stati effettuati diversi esperimenti e test dei dataset virtuali sull'algoritmo prescelto.

Sono state inoltre definite le procedure software attraverso l'utilizzo di specifiche funzioni di calcolo e loro interazione mediante uso di programmazione strutturata e procedurale.

Nelle prove di simulazione effettuate sono stati caratterizzati dei cluster di microRNA strettamente associati a specifiche caratteristiche cliniche dei dataset utilizzati. Tali features sono le stesse che si ritrovano all'interno del dataset biologico oggetto dello studio.

Una volta acquisiti i dati reali, dato l'esiguo numero di campioni pervenuti, abbiamo deciso di applicare un algoritmo adeguato al numero di campioni, al fine di dare un migliore valore statistico all'analisi dei dati. L'algoritmo prescelto è il DESeq, il quale permette di effettuare un'analisi differenziale di espressione genica basandosi sulla

distribuzione binomiale negativa. E' stata effettuata inoltre la messa a punto delle variabili e thresholds relative a tale algoritmo.

- **Analisi di enrichment funzionale sui dati analizzati**

L'analisi di enrichment effettuata sui quattro cluster di miRNA differenzialmente espressi è stata effettuata utilizzando diversi tools computazionali:

- 1) Per la valutazione e la predizione di targets di miRNA differenzialmente espressi sono stati utilizzati 7 diversi algoritmi di predizione quali miRWalk 2.0, miRDB, MicroT4, Rna22, miRanda, RNAhybrid and Targetscan . Ciò al fine di dare maggiore potere statistico all'analisi
- 2) Il Database for Annotation, Visualization and Integrated Discovery (DAVID) è stato utilizzato per effettuare l'analisi di enrichment funzionale dei targets predetti, associati ai miRNA differenzialmente espressi. Sono state prese in considerazione tutte e tre le sottocategorie relative alle annotazioni di "gene ontology" (funzioni molecolari, componenti cellulari, pathway biologici). Il servizio web permette di visualizzare i termini biologici over-rappresentati più rilevanti associati ad una data lista genica.
- 3) La risorsa Reactome è un tool di "pathway analysis" che permette di analizzare dataset di geni, di effettuare "pathway assignment", analisi di enrichment e di over-rappresentazione. Esso è stato utilizzato per visualizzare la lista di geni target all'interno di pathways specifici ed eventualmente rilevanti nel contesto della patologia tumorale.

## Bibliografia

1. He L, Hannon GJ, "MicroRNAs: small RNAs with a big role in gene regulation", *Nat Rev Genet.* Jul; 5(7):522-31, 2004
2. Kim VN, " MicroRNA biogenesis: coordinated cropping and dicing" ,*Nat Rev Mol Cell Biol.* May;6(5):376-85; \_2005;
3. Petersen CP, Bordeleau ME, Pelletier J, Sharp PA, " Short RNAs repress translation after initiation in mammalian cells". *Mol Cell.* Feb 17; 21(4):533-42; 2006
4. Du T, Zamore P., "microPrimer: the biogenesis and function of microRNA", *Development*; 132:4645–52, 2005
5. Kristina Schee, Øystein Fodstad, and Kjersti Flatmark, "Mini-Review, MicroRNAs as Biomarkers in Colorectal Cancer ", *The American Journal of Pathology*, Vol. 177, No. 4, October 2010
6. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, et al, "MicroRNA expression profiles classify human cancers", *Nature*, 435:834–838, 2005
7. Esquela-Kerscher A, Slack FJ, " Oncomirs - microRNAs with a role in Cancer", *Nat Rev Cancer*, 6:259–269, 2006

8. He L, Hannon GJ, “ MicroRNAs: small RNAs with a big role in gene Regulation”, *Nat Rev Genet*, 5:522–531, 2004
9. Mitchell PS, Parkin RK, Kroh EM, et al., “ Circulating microRNAs as stable blood-based markers for cancer detection”, *Proc Natl Acad Sci USA*, 105:10513–10518, 2008
10. Marson A, Levine S, Cole M, et al. “Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells”, *Cell*;134:521–33, 2008
11. Du T, Zamore P., “ microPrimer: the biogenesis and function of microRNA”, *Development*;132:4645–52, 2005
12. Brennecke J, Stark A, Russell R, et al., “ Principles of MicroRNA-Target Recognition”, *PLoS Biol*;3:e85, 2005
13. Chiou TJ: The role of microRNAs in sensing nutrient stress. *Plant Cell Environ* 2007; 30: 323–332.
14. Holtz J, Pasquinelli AE: Uncoupling of lin-14 mRNA and protein repression by nutrient deprivation in *Caenorhabditis elegans*. *RNA* 2009; 15: 400–405. 10
15. Leung AK, Sharp PA: MicroRNA functions in stress responses. *Mol Cell* 2010; 40: 205–215. 1
16. Laura García-Segura Martha Pérez-Andrade Juan Miranda-Ríos The Emerging Role of MicroRNAs in the Regulation of Gene Expression by Nutrients *J Nutrigenet Nutrigenomics* 2013, ;6:16–31