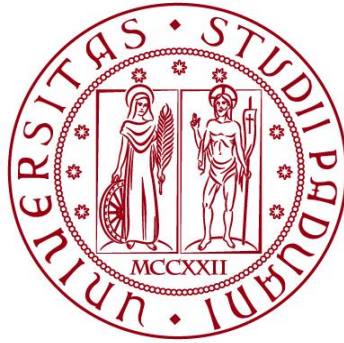


**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**DIPARTIMENTO DI BIOLOGIA**

**Corso di Laurea in BIOLOGIA MOLECOLARE**



**ELABORATO DI LAUREA**

**Analisi multi-specie di arricchimento di  
microRNA utilizzando sistemi di gestione  
automatizzata dei flussi di lavoro**

**Tutor: Prof.ssa Enrica Calura  
Dipartimento di Biologia**

**Laureanda: Alessia Mietto  
matricola:1161581**

**ANNO ACCADEMICO 2023/2024**



## **ABSTRACT**

L'analisi dell'arricchimento dei geni è ormai una delle applicazioni più utilizzate nella ricerca di biologia molecolare. Tuttavia, gli stessi principi statistici sono applicabili a tutti i tipi di omica. Con l'aumento della quantità di dati relativi alla quantificazione dell'espressione dei microRNA (miRNA), c'è anche un aumento della necessità di strumenti per la loro analisi per facilitarne l'annotazione e interpretazione attraverso le mappe della biologia dei sistemi.

Lo studio analizzato in questo elaborato, mira a presentare un nuovo strumento di analisi chiamato miEAA, approfondendone le sue caratteristiche di annotazione e analisi dei miRNA. miEAA è sia un pacchetto Python sia un'applicazione web che offre una varietà di test statistici combinati con la possibilità di eseguire le analisi su dati provenienti da dieci diversi organismi. Nell'articolo gli autori presentano diversi casi studio su tumore al rene e al seno e sulla malattia di Parkinson.

miEAA si basa su GeneTrail, uno strumento di analisi dell'arricchimento per insiemi genici, che consente l'analisi integrata di dataset trascrittomici, miRNomici, genomici e proteomici. Offre inoltre un gran numero di insiemi di riferimento predefiniti e una raccolta completa di categorie biologiche e consente confronti diretti tra i risultati calcolati.



## 1.INTRODUZIONE

### I microRNA e i sistemi biologici

I microRNA (o miRNA) sono speciali classi di RNA non codificanti a singolo filamento, lunghi circa 22 nucleotidi, coinvolti nella regolazione genica e identificati in molti organismi. Questi particolari RNA hanno una sequenza complementare a particolari regioni degli mRNA e ne regolano la loro abbondanza, tagliandoli o impedendone la traduzione. Il loro ruolo fondamentale è quello di regolare negativamente l'espressione genica a livello post-trascrizionale (*post-transcriptional gene silencing*, PTGS). I miRNA agiscono mediante il riconoscimento di specifici mRNA definiti "targets" al fine di determinarne la degradazione o la repressione della traduzione. La funzione di molti miRNA non è nota, ma per alcuni è stata provata la partecipazione a processi fisiologici e patologici (hanno un ruolo in proliferazione, apoptosi e differenziazione cellulare, possono essere deregolati in malattie umane e possono essere coinvolti nella tumorigenesi). I miRNA risultano essere evolutivamente molto conservati tanto che si pensa che siano elementi molto antichi dei meccanismi di regolazione dell'espressione genica. Le molecole costituenti i loro sistemi di espressione risultano essere simili nelle piante e negli animali, ma vi sono delle differenze inerenti ai loro meccanismi di espressione. Le alterazioni del normale pathway di espressione <sup>(1)</sup> dei miRNA possono avere conseguenze sulla normale fisiologia cellulare e portare a diversi tipi di patologie <sup>(2)</sup>, come cancro, neoplasie, malattie cardiache, diabete, danno d'organo e alterazioni dei processi regolativi del sistema nervoso (schizofrenia, disturbo bipolare e disturbo depressivo).

Decifrare i ruoli precisi di specifici miRNA in contesti fisiopatologici è impegnativo a causa degli alti livelli di complessità della rete delle loro interazioni. Infatti, la regolazione dell'espressione dell'mRNA da parte dei miRNA è spesso specifica di una cellula o un tessuto, è fortemente dipendente dallo stress e dallo stato metabolico dell'organismo e del tessuto; inoltre, può capitare che la loro attività regolatoria sia scarsamente correlata con i livelli di espressione del miRNA stesso. Questo perché la loro attività dipende da molti fattori di diversa natura. Vari meccanismi controllano non solo la loro espressione, ma anche la loro attività e/o biodisponibilità, quindi definire in silico le interazioni miRNA-gene attive nella cellula è molto complesso e va valutata caso per caso.

Per questo motivo, sono stati sviluppati strumenti web che permettono di gestire l'analisi di insiemi di miRNA.

Uno di questi strumenti è MiEAA (*MiRNA Enrichment Analysis and Annotation*), un'applicazione web che offre facili procedure per l'annotazione e una varietà di test statistici comunemente applicati come l'analisi della sovrarappresentazione

(ORA) e l'analisi dell'arricchimento del set di miRNA (GSEA), utili per rendere più facilmente comprensibili i dati degli studi su larga scala sull'espressione genica differenziale che spesso riportano centinaia, se non migliaia, di geni deregolati.

### **Le annotazioni dei geni**

Tutti i metodi di analisi trattate in questa tesi si basano sull'utilizzo di insiemi di geni (gene set) definiti *a priori*, ovvero gruppi di geni che condividono una caratteristica biologica: la posizione cromosomica o la regolazione dell'espressione, la funzione proteica o l'appartenenza ad una via di segnale della cellula. Esistono dei database di annotazione genica che forniscono questi insiemi di geni.

Sin dagli inizi dello sviluppo dell'analisi omica, diverse tipologie di database, contenenti informazioni sui percorsi di segnalazione cellulare, sono stati sviluppati in maniera tale da poter fornire informazioni in una sorta di lingua comune, che permettesse la condivisione di informazioni e lo sviluppo di diversi strumenti partendo da conoscenze diverse e che ha visto l'unione degli sforzi congiunti di comunità comprendenti molti campi della biologia. Nonostante il continuo impegno per mantenere le mappe sempre aggiornate rispetto le ultime scoperte biologiche, poiché necessitano di un'annotazione manuale accurata, risultano essere sempre un passo indietro rispetto le ultime scoperte e quindi incomplete.

La presenza di queste lacune risulta ancora più evidente nel caso dei microRNA, in quanto, nonostante molti siano stati validati in maniera specifica tramite saggi con geni definiti *reporter*, la maggior parte dei target dei miRNA rimangono sconosciuti, quindi malgrado la grande varietà di dati sulla presenza fondamentale dei miRNA nella segnalazione cellulare, tutti i database mancano di annotazioni esaustive.

L'intricata quantità di miRNA e le relazioni tra i percorsi possono essere rappresentate utilizzando un network composto da interazioni miRNA-gene e gene-gene, ma l'analisi dei segnali biologici che percorrono questa rete, data la loro complessità, sono le maggiori sfide nella biologia integrativa.

Per questo motivo la comunità scientifica sfrutta la combinazione di diversi tipi di dati, come l'espressione dei geni e dei miRNA, mediante lo sviluppo di nuovi metodi statistici combinati con implementazioni algoritmiche per cercare di decifrarne nel dettaglio le connessioni.

## **I metodi statistici usati per l'analisi dei set di geni**

Lo scopo finale delle analisi dei gene set è trovare quei gruppi di geni che meglio descrivono il fenotipo del campione biologico che si sta studiando. Queste analisi supportano efficacemente l'interpretazione dei dati sperimentali genomici.

Possono essere suddivise in due principali categorie: ORA (*Over-Representation Analysis*) e Functional Class Scoring (FCS) methods come GSEA (*Gene Set Enrichment Analysis*)<sup>(4)</sup>.

I metodi ORA<sup>(5)</sup> si basano principalmente sull'analisi di tabelle di contingenza testando l'associazione tra il numero di geni differenzialmente espressi, ottenuti confrontando tra due condizioni sperimentali diverse (ad esempio sano vs malato, trattato vs non trattato) e l'appartenenza dei geni ad uno specifico gene set (ad esempio apoptosi, ciclo di Krebs). Alcuni classici esempi di test ORA sono il test dell'ipergeometrica o del Chi quadrato ( $\chi^2$ ). Gli approcci ORA sono utilizzati per determinare se specifiche funzioni o i processi biologici sono sovrarappresentati in liste di geni derivate sperimentalmente attraverso esperimenti di microarray o RNA-seq.

I metodi di FCS non hanno bisogno della lista dei geni differenzialmente espressi, ma usano tutti i geni annotati all'interno di un gene set. Spesso questi derivano da uno score del livello di espressione dei geni annotati nel gene set. L'ipotesi dei metodi FCS è che, sebbene grandi cambiamenti nei singoli geni possano avere effetti significativi sulle pathway, cambiamenti più deboli, ma coordinati in insiemi di geni funzionalmente correlati, possono anche avere effetti significativi. Pertanto, piuttosto che fissare una soglia arbitraria per identificare i geni significativi, tutti i geni vengono considerati nell'analisi. Uno dei metodi FCS più famoso e implementato in MiEAA è GSEA (Gene Set Enrichment Analysis) sviluppato da Efron et al.<sup>(6)</sup>

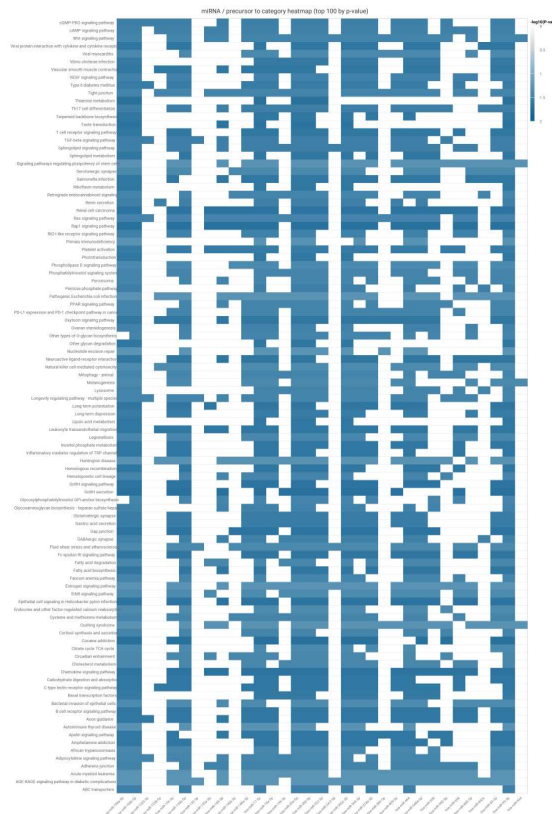
Il metodo GSEA<sup>(7)</sup> prevede un ordinamento decrescente per espressione dei geni e poi calcola un punteggio di arricchimento per ogni set di geni che rappresenta quanto i geni presenti nell'insieme si collocano nella parte superiore (alta espressione) o inferiore dell'elenco (bassa espressione). Successivamente, a ciascun punteggio, viene assegnata una significatività statistica attraverso un approccio permutazionale. L'applicazione di questi test ai dati genomici richiede la conoscenza di database di annotazione genica e la capacità di far dialogare tra loro diverse annotazioni geniche con le matrici di espressione e l'esecuzione di test statistici. Strumenti che ne permettono l'applicazione per l'analisi di dataset di espressione genica sono moltissimi, più rari sono quelli che si occupano di studiare i dati di espressione dei miRNA, tra questi abbiamo miEAA.

## 2. APPROCCIO SPERIMENTALE: DESCRIZIONE DELLE METODOLOGIE

### MiEAA: descrizione, funzionamento e principali caratteristiche

MiEAA è un'applicazione basata sul web che offre una varietà di test statistici comunemente applicati come l'analisi di sovrarappresentazione e l'analisi dell'arricchimento del set di miRNA. Questi test sono implementati similmente alle analisi usate per i geni e descritte nell'introduzione alla tesi. miEAA permette l'uso di diversi tipi di gene set. Il server web offre visualizzazioni come le mappe termiche e curve di somma con distribuzioni in background.

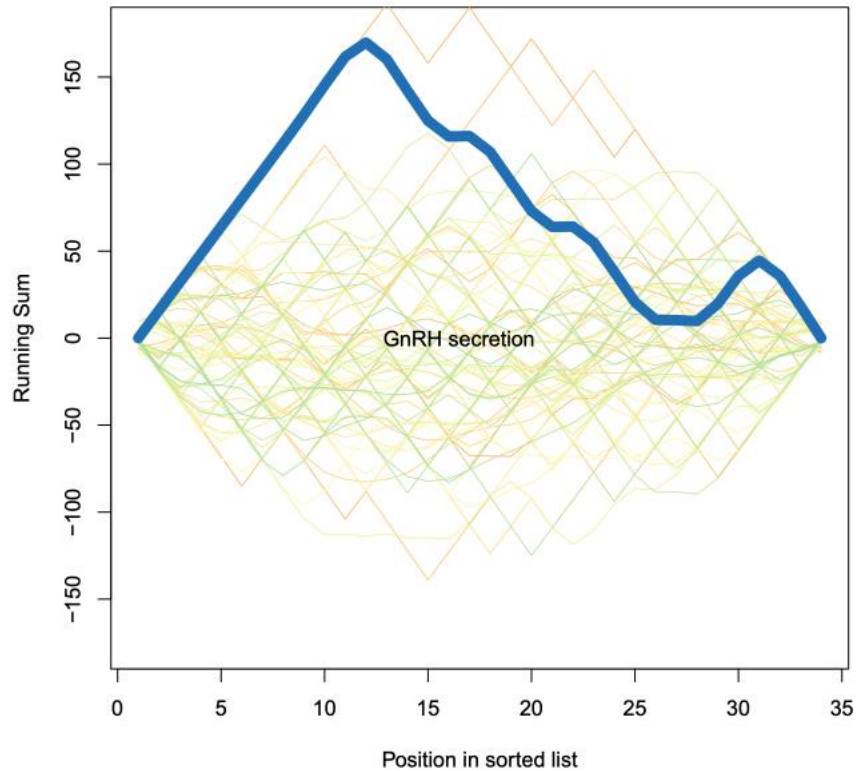
Le mappe di calore (Fig.1) sono rappresentazioni grafiche interattive che mostrano i punteggi di arricchimento per un insieme selezionato di miRNA e pathways provenienti da un database specifico. I punteggi di arricchimento visualizzati sono calcolati dai rispettivi p-value. Maggiore è il punteggio di arricchimento, più scuro è il riquadro e più significativo è l'arricchimento del rispettivo miRNA nei rispettivi pathways. Passando il mouse sopra le celle della mappa termica si evidenziano i miRNA e i pathways corrispondenti con il rispettivo punteggio. In questo modo risulta evidente se i pathways di targeting dei miRNA sono specifici o se un miRNA sta prendendo di mira quasi tutte le categorie. In quest'ultimo caso i risultati significativi possono essere, per esempio, il risultato di una distorsione nell'analisi dell'arricchimento.



**Figura 1:** mappa di calore dei pathways mirati in modo significativo da almeno un miRNA.



Nelle curve di somma con distribuzioni in background (Fig.2) vengono rappresentati i miRNAs che mostrano un punteggio di arricchimento statisticamente significativo e che possono essere considerati come potenziali regolatori nelle condizioni analizzate sovrapposti ai risultati attesi per l'analisi che si sta svolgendo.



**Figura 2:** curva di somma con distribuzioni in background. La linea blu rappresenta i risultati effettivi dell'analisi svolta sui campioni, mentre le linee sottili verdi e blu rappresentano i risultati attesi per l'analisi.

miEAA riporta anche gli ID PubMed dei manoscritti che hanno contribuito all'annotazione dei miRNA in miRBase.

Oltre ai diversi test statistici, miEAA fornisce anche una ricca funzionalità in termini di categorie di miRNA e queste informazioni possono essere applicate per analizzare sia i precursori di miRNA sia i miRNA maturi. Inoltre, per rendere lo strumento il più utile possibile, sono stati implementati gli strumenti di supporto come i convertitori tra diverse versioni miRBase e convertitori dai nomi miRNA ai nomi dei precursori. In combinazione con i convertitori miRBase, miEAA supporta le impostazioni di ricerca in cui sono in uso versioni precedenti di miRBase. La versione aggiornata, miEAA 2.0, supporta l'input di miRNA da dieci organismi comunemente studiati.

L'utilizzo della pagina web richiede pochi passaggi:

- Selezionare il tipo di set di miRNA, scegliendo tra precursori e miRNA maturi.
- Selezionare la tipologia di analisi (GSEA o ORA)
- Selezionare la specie su cui eseguire l'analisi e caricare l'elenco dei miRNA/precursori da analizzare. I nomi dei miRNA / precursori devono corrispondere alla specie selezionata e i dati dovrebbero essere caricati come semplice elenco, in cui ogni riga corrisponde ad un singolo ID.
- Aggiungere, nel caso in cui si selezioni il metodo ORA, un set di riferimento, ovvero un set di dati con cui confrontare il set preso in analisi. Ad esempio, nel caso in cui si sia eseguito un esperimento di microarray da cui sono stati selezionati solo determinati miRNA da analizzare, si mette come set di riferimento tutti i miRNA individuati sull'array. Se il set di riferimento non viene aggiunto tutti i miRNA/precursori raccolti vengono utilizzati come set di riferimento.
- Selezionare eventuali parametri riguardanti le categorie (posizione cromosomica, geni bersaglio, pathways, organi, malattie, etc.), aggiustamenti del p-value per i test multipli (nessun aggiustamento, aggiustamento FDR, regolazione Bonferroni, etc), dettagli sulla correzione del p-value (correzione per ogni categoria o collettivamente per tutte le categorie). È possibile inoltre indicare il livello di significatività e il numero di riscontri minimi richiesti per sottocategoria, ovvero il numero minimo di corrispondenze evidenziabili per sottocategoria.

Al termine, infine, vengono rappresentati i risultati.

I risultati ORA contengono categorie significative, sottocategorie, valori p-value grezzi e aggiustati (q-value), tipo di arricchimento, conteggi dei miRNA attesi e osservati e i miRNA/ precursori risultanti dall'analisi e contenuti nella sottocategoria corrispondente. I valori di arricchimento possono essere "sovrarappresentati" o "sottorappresentati" e sono visualizzati rispettivamente da frecce rosse verso l'alto o verdi verso il basso (Fig. 3)

Categoria	Sottocategoria	Arricchimento	Valore P	Pregolato	Valore Q	Previsto	Osservato	miRNA/precursori
			Da A	Da A	Da A	Da A	Da A	
Tessuti con Interazioni miRNA-TF (TransmiR)	Polmone	↓ sottorappresentati	0.0014987	0.0095761	0.0095761	16.2386	10	has-mir-199b; has-mir-181a-2; has-mir-150; has-mir-155; has-mir-17; has-mir-19a; has-mir-20a; has-mir-221; has-mir-222; has-mir-508
Tessuti con Interazioni miRNA-TF (TransmiR)	Utero	↓ sottorappresentati	0.001596	0.0095761	0.0095761	14.6803	8	has-mir-181a-2; has-mir-150; has-mir-155; has-mir-17; has-mir-19a; has-mir-20a; has-mir-221; has-mir-222
Malattie (HMDD)	Carcinoma a cellule squamose, orale	↑ sovrarappresentati	0.0024723	0.0098017	0.0098017	2.63147	8	has-mir-181a-2; has-mir-144; has-mir-155; has-mir-17; has-mir-19a; has-mir-20a; has-mir-221; has-mir-222
Malattie (HMDD)	Osteoporosi	↑ sovrarappresentati	0.0025193	0.0098142	0.0098142	0.616379	4	has-mir-451a; has-mir-181a-2; has-mir-150; has-mir-144
Malattie (HMDD)	Prolattinoma	↑ sovrarappresentati	0.0025193	0.0098142	0.0098142	0.616379	4	has-mir-451a; has-mir-199b; has-mir-144; has-mir-17

**Figura 3:** rappresentazione risultati ORA

Anche i risultati del GSEA contengono le categorie significative, eventuali sottocategorie, un grafico della somma parziale, che può essere scaricata in diversi formati, in cui vengono rappresentati i miRNAs che mostrano un punteggio di arricchimento statisticamente significativo e che possono essere considerati come potenziali regolatori nelle condizioni analizzate; p-value, q-value, i dettagli sull'arricchimento, i conteggi dei miRNA osservati e i miRNA/ precursori risultanti dall'analisi e contenuti nella sottocategoria corrispondente. Nella tabella dei risultati viene riportato un grafico che è colorato di rosso quando la categoria in analisi risulta arricchita di miRNAs con alta espressione e di verde se la categoria è "depleta", cioè presenta la maggior parte di miRNAs con bassa espressione (Fig. 4).



La mappa termica mostra i p-value (sottoforma di  $-\log_{10}(\text{p-value})$ , punteggi alti p-value bassi e viceversa) delle top 100 sottocategorie più significative riportando il p-value per ogni miRNA incluso nella categoria sulle colonne. Passando il mouse sulle etichette delle righe verrà visualizzata la categoria di origine (Fig.6).

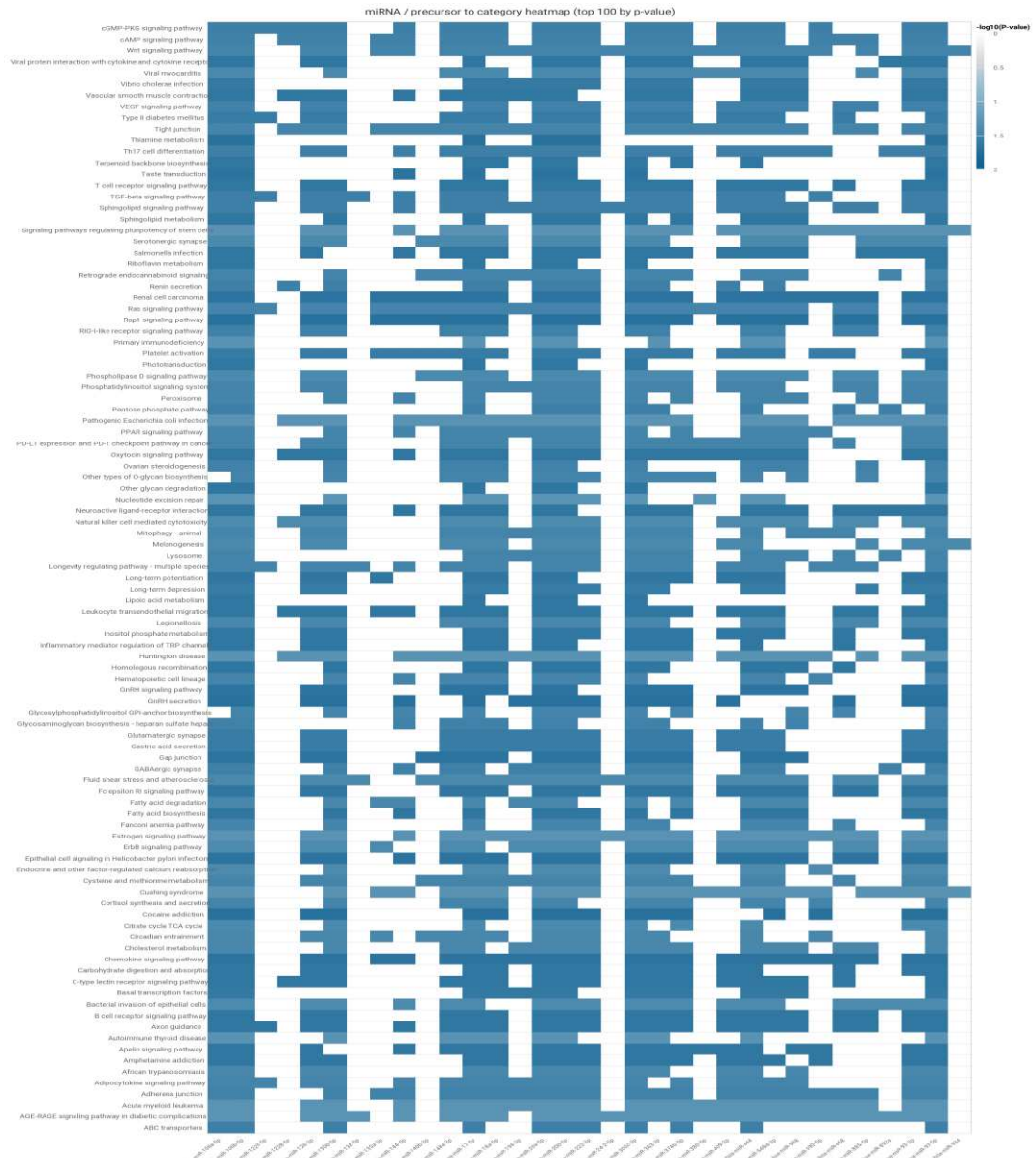


Figura 6: mappa di calore

### 3.RISULTATI E DISCUSSIONE

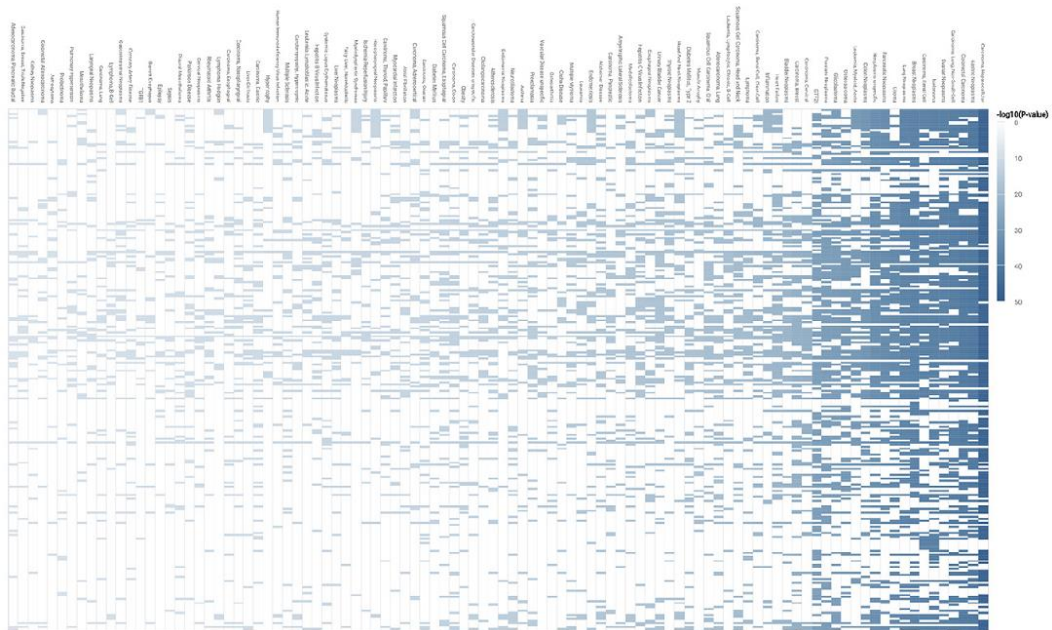
#### Presentazione dei casi di studio

Nell'articolo gli studiosi mostrano, mediante tre casi di studio differenti, il funzionamento di miEAA.

#### Caso di studio 1: Carcinoma renale a cellule chiare

Come primo caso di studio sono stati acquisiti 591 campioni di miRNA umani dal progetto *Kidney Renal Clear Cell Carcinoma*, di cui 520 campioni di tumore primario (PT) e 71 campioni normali di tessuto solido (STN). Questi campioni fanno parte del *The Cancer Genome Atlas (TCGA)* catalogue. Sono stati presi in considerazione 1881 precursori e 321 sono stati rilevati in almeno il 50% dei campioni di ogni biogruppo (PT o STN), e tra questi, 282 risultano essere differenzialmente espressi tra PT e STN.

L'analisi di sovrarappresentazione dei precursori, con una soglia di significatività fissata a 0,05, ha portato come risultato che 541 precursori sono significativamente arricchiti e sette significativamente depleti come si può osservare nella Fig.7. Un sottoinsieme di precursori è sempre presente in tutte le categorie significative, mentre altri sembrano essere più specifici. Le prime 10 categorie, ordinate secondo l'aumento *del P-value*, sono associate al cancro, incluso il carcinoma a cellule renali. Inoltre, il rapporto osservato sopra le attese (123/48,6) indica un forte arricchimento dei precursori deregolamentati nei reni e in altri tipi di cancro.



**Figura 7:** visualizzazione Web dei risultati del primo caso di studio mediante mappa termica con categorie lungo l'asse X e il precursore corrispondente sull'asse Y.

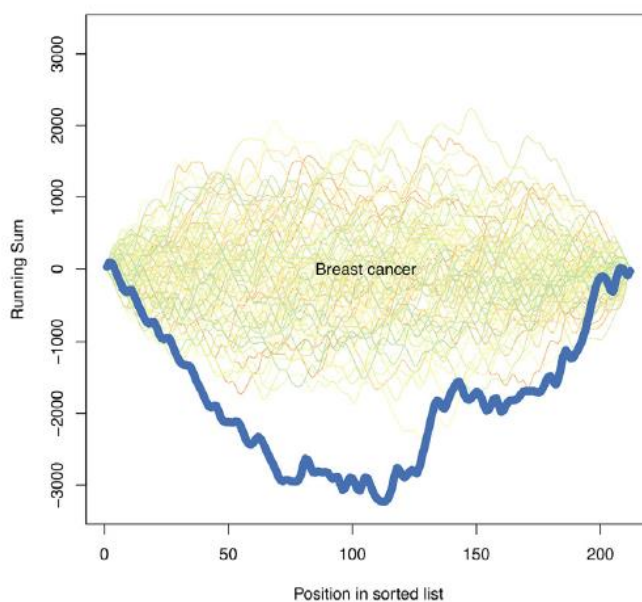
Caso di studio 2: Modello di topo per lo studio della progressione del cancro al seno

Nel secondo caso di studio, per mostrare le numerose categorie di analisi presenti, hanno selezionato un set di dati dal *Gene Expression Omnibus* (GEO), un archivio di dati genomici funzionali pubblico, relativi ai microRNA circolanti da un modello di topo con tumore al seno, misurati con microarray. Il set di dati comprende 36 campioni, di cui una parte aventi mutazione (NeuT+) e una parte *wild type* (NeuT-) di pari età, raccolti in tre diversi stadi della malattia: premaligno, preinvasivo e invasivo. Mantenendo la soglia di significatività simile a quella del primo caso, sono stati rilevati 212 miRNA dall'analisi di differenziale espressione. A questi 212 miRNA, inseriti in un elenco ordinato partendo da quelli rilevati nello stadio precanceroso e posizionando per ultimi quelli risultati negli stadi invasivi, è stato applicato il metodo GSEA sia ai campioni aventi la mutazione sia ai campioni *wild type*.

La prima analisi ha restituito 311 categorie significative, mentre la seconda non ne ha restituite nessuna. Nel complesso, molte più categorie sembravano essere deplete che arricchite, suggerendo una regolazione su larga scala delle vie molecolari man mano che i miRNA sono down regolati in NeuT+. È stato riscontrato inoltre che la differenziazione dei macrofagi, lo sviluppo vascolare e la via di segnalazione del VEGF sono depleti, il che potrebbe essere un segnale per il carico tumorale, ovvero la quantità di cellule cancerose presenti nel corpo, di topi NeuT+ allo stadio di carcinoma mammario invasivo.

Infine, sono state anche valutate le GSEA per il confronto tra NeuT+ e NeuT- in tutte e tre le fasi. Le prime due impostazioni hanno restituito un insieme piuttosto aspecifico di categorie con tutti i P-value situati vicino al limite di significatività, l'ultimo confronto ha prodotto invece risultati interessanti:

- I. Tutte le categorie sono deplete, vale a dire che non sono presenti categorie con p-value significativi. Inoltre
- II. I risultati mostrano che diverse dozzine di miRNA conservati sono down-regolati nel modello NeuT+ nella fase invasiva (Fig.8)



**Figura 8:** Grafico GSEA con distribuzioni simulate dello sfondo (linee da verde a arancione) e esaurimento effettivo per il cancro al seno (linea blu scuro) osservato durante la valutazione del secondo caso di studio.

### Caso di studio 3: Biomarcatori del Parkinson da PPMI e confronto con TAM2.0

In ultimo è stato fornito un caso di studio non oncologico, valutando le prestazioni del miEAA su un set di dati ad alta risoluzione di piccoli RNA non codificanti nel sangue intero. Questo set di dati è disponibile gratuitamente dal portale dati Parkinson's Progression Markers Initiative (PPMI). I campioni sono stati acquisiti e sequenziati a partire da 1600 individui, raccogliendo fino a cinque campioni di sangue per un periodo di oltre tre anni per gli snCRNA. I conteggi grezzi sono stati normalizzati in letture per milione (rpm) e i precursori sono stati filtrati in modo analogo ai criteri definiti per il caso di studio TCGA. Successivamente, sono stati confrontati i profili dei precursori dei miRNA di 2337 campioni di Parkinson con 1538 controlli abbinati per età.

Sono stati confrontati i P-value grezzi degli strumenti per escludere un'influenza delle dimensioni delle categorie disponibili. Un confronto diretto ha evidenziato 72 riscontri da entrambi gli strumenti (altri 70 segnalati solo da TAM e 144 solo da miEAA). Nomi di categoria molto simili, ma non esattamente corrispondenti, sono stati abbinati manualmente, ma sono rimaste comunque diverse categorie definite in modo ambiguo. Sono stati selezionati poi i 32 insiemi miEAA più significativi e i 30 set TAM più significativi e si è osservata una sovrapposizione di 23 categorie, suggerendo infatti risultati migliori comparabili per le categorie più significative. Il morbo di Parkinson, risulta coperto da tre miRNA nel TAM e da sei miRNA nel miEAA.



#### 4.CONCLUSIONI

Come si evince da questo articolo, gli strumenti statistici per l'analisi dell'arricchimento biologico sono fondamentali per comprendere i dati omici di diversa provenienza. Tuttavia, la parte davvero fondamentale, sono la qualità delle annotazioni riguardanti i microRNA di diverso tipo che possono essere fornite all'utente.

Tra queste, lo studio analizzato sottolinea come i nuovi sviluppi nel campo della ricerca sui miRNA abbiano fornito una serie di categorie biologiche che ricoprono la maggior parte degli aspetti, delle proprietà e delle funzioni dei miRNA. Questi ultimi sono stati ottenuti anche grazie all'analisi comparata tra specie, che ha permesso di mettere a disposizione studi su larga scala sull'espressione genica differenziale con centinaia, se non migliaia, di geni deregolati. La gestione di questo grande quantitativo di informazioni ha richiesto l'utilizzo di strumenti specifici per l'analisi di grandi insiemi di dati e la loro gestione agevolata per utenti senza competenze informatiche specifiche.

Lo studio presentato prende in analisi una nuova versione di MiEAA (*MiRNA Enrichment Analysis and Annotation*), e copre molti degli aspetti necessari per poter ottenere una buona analisi dei dati di microRNA, migliorando in senso assoluto e rispetto alla versione precedente l'insieme delle categorie disponibili per l'annotazione funzionale dei microRNA sia a livello quantitativo che a livello qualitativo, sia offrendo approcci per la correzione del P-value. Uno dei problemi presi in considerazione dalla nuova versione di MiEAA riguarda incompatibilità di annotazione degli identificativi di microRNA tra quelli usati nelle categorie funzionali e quelli generalmente usati per l'analisi di espressione.

In MiEAA gli autori forniscono strumenti di supporto per la conversione dei nomi di precursori e miRNA in diverse versioni miRBase e per la mappatura dei precursori in miRNA o viceversa. Stabilire una nomenclatura standardizzata è sicuramente un punto fondamentale per queste analisi.

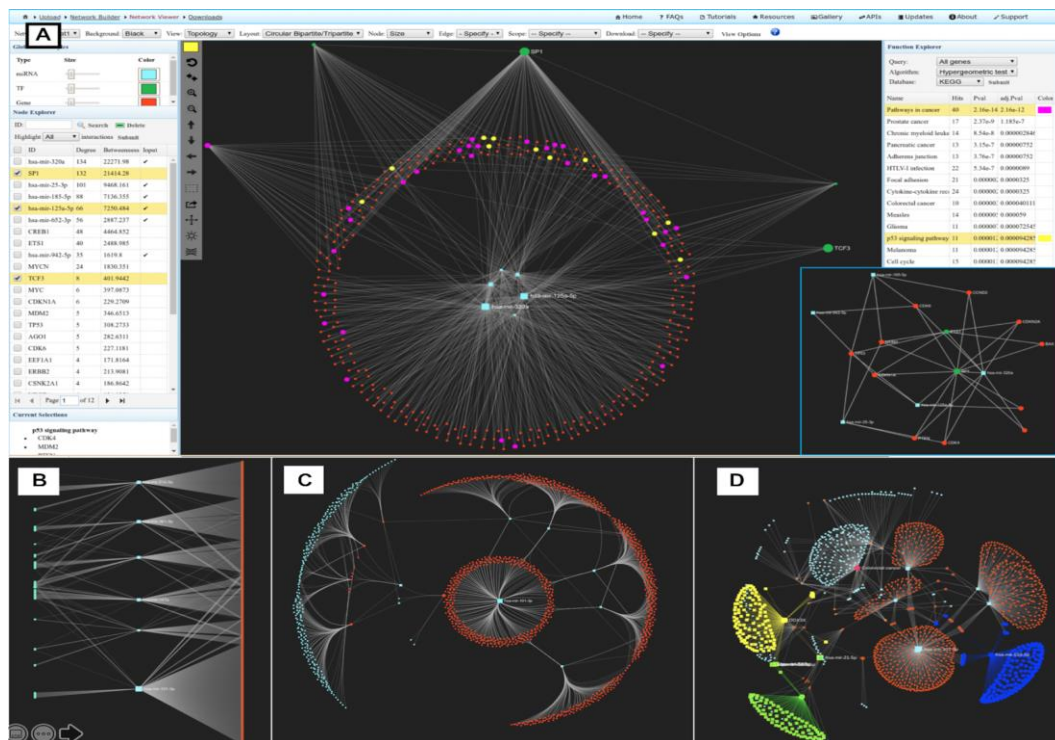
In secondo luogo, visto che spesso l'analisi dei dati ad alta velocità di trasmissione del miRNA comporta il confronto di più biogruppi, l'utilizzo della nuova interfaccia di MiEAA, fa sì che sia possibile eseguire più analisi riducendo al minimo il tempo impiegato per la configurazione e l'aggregazione dei risultati. Da questa analisi risulta quindi che miEAA 2.0 è uno strumento flessibile, completo, altamente accessibile e veloce per l'annotazione e l'analisi dell'arricchimento del miRNA.

Ad oggi, sono stati sviluppati diversi strumenti per l'analisi funzionale dei miRNA, e durante il mio periodo di tirocinio, svolto presso il Dipartimento di Biologia nel laboratorio della Dott.ssa Calura, ne ho applicati alcuni.

Oltre che miEAA, ho avuto modo di analizzare altri tre strumenti basati sul web, mirNET2.0, DIANA-miRPath 3.0 e mirPathDB 2.0, che risultano essere tra gli strumenti più recenti e aggiornati per l'implementazione delle analisi dei segnali biologici.

mirNET2.0 <sup>(7)</sup> è una piattaforma basata sul Web di facile utilizzo progettata per aiutare a chiarire le funzioni dei microRNA integrando i dati degli utenti con le conoscenze esistenti tramite l'analisi visiva basata sulla rete. Prevede tre passaggi principali: l'immissione dei dati, la creazione della rete e l'analisi visiva della rete. Questa piattaforma è stata sviluppata per aiutare a illustrare le relazioni complesse come quelle che intercorrono tra miRNA e geni ( un miRNA può regolare più geni e un gene può essere regolato da più miRNA) attraverso la visualizzazione basata sulle reti miRNA-target accoppiate con una migliore analisi funzionale. Sono supportate dieci specie diverse e vengono utilizzati undici database diversi per annotare i miRNA. I moduli di maggiore interesse sono quelli per l'analisi e l'esplorazione di polimorfismi a singolo nucleotide, quelli per includere i fattori di trascrizione e valutare la loro cooperazione con i miRNA nella sintonizzazione dell'espressione genica attraverso cicli di feedback e quelli per selezionare predizioni di legame su elementi non codificanti come lunghi RNA non codificanti e RNA circolari. miRNet 2.0 fornisce quattro tipi di query (tutti i geni, i geni evidenziati, tutti i miRNA, i miRNA evidenziati), due algoritmi di arricchimento (test ipergeometrici e campionamento empirico), nove librerie di annotazioni (tre librerie di geni-set e sei librerie di miRNA-set), che rappresentano il supporto più completo per comprendere le funzioni collettive dei miRNA.

Partendo da questi input, lo strumento identifica le caratteristiche significative, seleziona i loro target previsti o validati, identifica le categorie funzionali arricchite e visualizza i risultati in un ambiente di networking di base efficiente e personalizzabile (Fig.9).

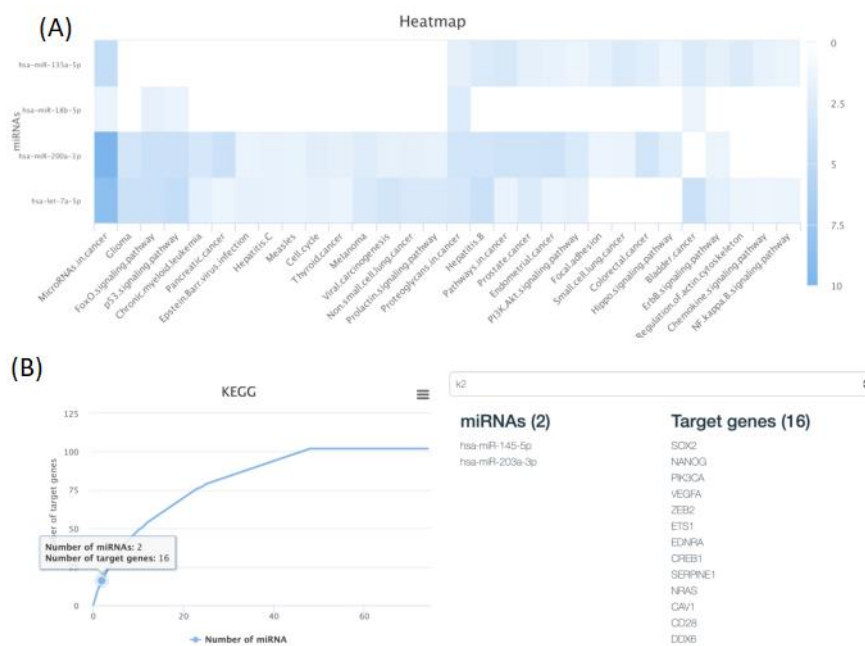


**Figura 9:** Caratteristiche principali e diversi layout di rete. (A) Visualizzazione tipica della pagina. Il pannello centrale mostra una rete in layout circolare-tripartito e i pannelli circostanti forniscono funzioni per l'analisi e la personalizzazione della rete. Un modulo di rete estratto è stato visualizzato in basso a destra. (B) Disposizione lineare-tripartita. (C) Layout concentrico con raggruppamento bordi. (D) Layout backbone con diversi moduli evidenziati in colori diversi.

DIANA-miRPath 3.0 <sup>(8)</sup> è una suite software online dedicata alla valutazione dei ruoli normativi del miRNA e all'identificazione di percorsi controllati e rende possibile l'annotazione funzionale di uno o più miRNA utilizzando statistiche standard (distribuzioni ipergeometriche), distribuzioni empiriche imparziali e/o statistiche di meta-analisi. DIANA-miRPath 3.0 annota i miRNA di sette specie in base alla loro associazione con i percorsi KEGG e le categorie GO e verifica la loro associazione funzionale utilizzando approcci ORA con correzione del bias. Come output fornisce mappe KEGG aggiornate con inclusione dei miRNA i cui target appartengono al pathway.

mirPathDB 2.0 <sup>(9)</sup> è un database incentrato sull'effetto dei miRNA nei percorsi. Insieme alle categorie GO, mirPathDB è l'unico strumento che raccoglie le annotazioni dei percorsi da diverse fonti e, in questa nuova versione aggiornata, grazie alle sue nuove funzionalità personalizzate, risulta essere una delle risorse più complete e avanzate per lo studio dei miRNA e i relativi percorsi di destinazione. *mirPathDB* memorizza informazioni sui miRNA (candidati), sui loro geni bersaglio e sulle loro vie bersaglio. Per accedere a queste informazioni, il database offre agli utenti due rappresentazioni distinte: una vista incentrata sull'miRNA e una vista incentrata sul percorso. I moduli in analisi possono essere

analizzati sotto una vista miRNAcentrica, dove per ogni miRNA viene riportato un elenco di pathway significativamente mirati; sotto una visione percorso-centrica, dove per ogni percorso viene riportato l'elenco dei miRNA che prendono di mira in modo significativo quel particolare percorso e in una visualizzazione multi-miRNA, dove data una lista di miRNA è riportata la mappa di calore dei percorsi mirati in modo significativo da almeno un miRNA. mirPathDB 2.0 si differenzia dalle altre piattaforme sul web, grazie ad alcune caratteristiche peculiari quali l'analisi di similarità dei miRNA, che definisce gruppi simili di miRNA in base alla loro sequenza di base, sulla sequenza matura, geni target e percorsi bersaglio e l'analisi di copertura massima, che trova il minor numero di miRNA sufficienti per regolare un determinato percorso (Fig.10)



**Figura 10:** (A) Esempio di visualizzazione della mappa termica personalizzata. Le righe rappresentano i risultati dell'arricchimento per le destinazioni dei quattro miRNA. Le colonne rappresentano tutti i percorsi KEGG significativi per i diversi miRNA. I colori più scuri indicano associazioni più significative tra miRNA e percorso di destinazione. (B) Esempio di visualizzazione interattiva per un'analisi di copertura massima specifica dell'utente. La curva a sinistra indica quanti dei geni bersaglio specificati possono essere presi di mira da un numero crescente di miRNA. L'asse x mostra il numero crescente di miRNA e l'asse Y il numero di geni bersaglio coperti.

Qualunque sia il sistema di analisi utilizzato, il punto fondamentale degli studi su larga scala sull'espressione genica differenziale è avere a disposizione degli strumenti di facile utilizzo che permettano di rendere più facilmente comprensibile l'interpretazione dei dati, e che permettano anche agli studiosi meno esperti dell'ambito bioinformatico di poter svolgere le analisi necessarie.

## 5. BIBLIOGRAFIA

1. Svetlana A. Shabalina, Eugene V. Koonin;  
***Origins and evolution of eukaryotic RNA interference.*** Trends Ecol Evol. Author manuscript; available in PMC 2009 Jun 11. Published in final edited form as: Trends Ecol Evol. 2008 Oct; 23(10): 578–587. Published online 2008 Aug 18. doi: 10.1016/j.tree.2008.06.005 (PMCID: PMC2695246)
2. Marta Correia de Sousa, Monika Gjorgjieva, Dobrochna Dolicka, Cyril Sobolewski, Michelangelo Foti;  
***Deciphering miRNAs' Action through miRNA Editing.*** Int J Mol Sci. 2019 Dec; 20(24): 6249. Published online 2019 Dec 11. doi: 10.3390/ijms20246249 (PMCID: PMC6941098)
3. Enrica Calura and Chiara Romualdi  
***miRNA bioinformatics and pathway analysis.***
4. Tarca, Bhatti, Romero  
***A Comparison of Gene Set Analysis Methods in Terms of Sensitivity.*** Prioritization and Specificity, PLoS One. 2013
5. Hui Sun Leong, David Kipling;  
***Text-based over-representation analysis of microarray gene lists with annotation bias.*** Nucleic Acids Res. 2009 Jun; 37(11): e79. Published online 2009 May 8. doi: 10.1093/nar/gkp310 (PMCID: PMC2699530)
6. Efron B, Tibshirani R (2007)  
***On testing the significance of sets of genes.*** Ann Appl Stat 1: 107–129. [[Google Scholar](#)]
7. Le Chang, Guangyan Zhou, Othman Soufan, Jianguo Xia,  
***miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology.*** Nucleic Acids Research, Volume 48, Issue W1, 02 July 2020, (<https://doi.org/10.1093/nar/gkaa467>)

8. Ioannis S. Vlachos, Konstantinos Zagganas, Maria D. Paraskevopoulou, Georgios Georgakilas, Dimitra Karagkouni, Thanasis Vergoulis, Theodore Dalamagas, Artemis G. Hatzigeorgiou,  
***DIANA-miRPath v3.0: deciphering microRNA function with experimental support.*** *Nucleic Acids Research*, Volume 43, Issue W1, 1 July 2015, (<https://doi.org/10.1093/nar/gkv403>)
9. Tim Kehl, Fabian Kern, Christina Backes, Tobias Fehlmann, Daniel Stöckel, Eckart Meese, Hans-Peter Lenhof, Andreas Keller  
***miRPathDB 2.0: a novel release of the miRNA Pathway Dictionary Database.*** *Nucleic Acids Res.* 2020 Jan 8; 48(D1): D142–D147. Published online 2019 Nov 6. (doi: 10.1093/nar/gkz1022) (PMCID: PMC7145528)

## 6.APPENDICE

[miEAA 2.0 integrating multi-species microRNA.pdf](#)