



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova
Dipartimento di Scienze Chirurgiche e Gastroenterologiche

SCUOLA DI DOTTORATO DI RICERCA IN: Biologia e Medicina della
Rigenerazione

INDIRIZZO: Scienze Epatologiche e Gastroenterologiche

CICLO XXII

***ANALISI DEI PROFILI DI ESPRESSIONE DEI MICRORNA NELLA PROGRESSIONE
DELL'EPATOCARCINOMA E SVILUPPO DI METODI BIOINFORMATICI PER LA
PREDIZIONE DEI GENI TARGET***

Direttore della Scuola: Ch.mo Prof. Pier Paolo Parnigotto

Coordinatore d'indirizzo: Ch.mo Prof. Giacomo Carlo Sturniolo

Supervisore: Ch.mo Prof. Fabio Farinati

Dottorando: Enrico Lavezzo

SOMMARIO

I microRNA (miRNA) sono RNA non codificanti il cui ruolo di regolazione dell'espressione genica è stato scoperto e via via delucidato in anni recenti.

L'azione dei miRNA, non dissimilmente da quella di altre tipologie di RNA non codificanti, si esprime nell'interazione fisica specifica con trascritti genici (mRNA) e nell'impedimento o comunque nella riduzione quantitativa della produzione, a partire da questi ultimi, di proteine. L'effetto biologico del miRNA dipende pertanto dalla natura del suo bersaglio, chiamato target. Non è al momento disponibile una mappa completa dei target di ciascuno dei 721 miRNA umani noti ma numerosi studi hanno già dimostrato per molti di essi un coinvolgimento diretto nei meccanismi patogenetici di numerose patologie, fra cui il cancro.

I virus dell'epatite B (HBV) e C (HCV) si trasmettono ogni anno a centinaia di migliaia di persone nel mondo e sono causa di malattie necroinfiammatorie epatiche le quali, pur con percentuali e modalità differenti, possono portare allo sviluppo di una malattia cronica in grado di progredire nel lungo periodo provocando lo sviluppo di fibrosi, cirrosi ed epatocarcinoma (HCC); quest'ultimo è uno dei tumori umani più aggressivi e tanto la sua diagnosi precoce, quanto ancor più la sua terapia, permangono alquanto problematiche.

Molte ragioni, quindi, spingono la comunità medico-scientifica alla ricerca di nuove strategie che possano portare a effettivi miglioramenti nella gestione sanitaria di questa patologia.

Nel caso dei miRNA, sono già stati effettuati numerosi studi mirati a dimostrare l'esistenza di un preciso schema della loro espressione nel tessuto epatico canceroso, mentre il ruolo che essi possono ricoprire nelle fasi precedenti all'insorgenza del cancro rimane ancora largamente incompreso e, al momento, poco studiato.

Partendo da tali presupposti, gli obiettivi di questo studio sono:

- La definizione di un protocollo per l'analisi dei miRNA in campioni tissutali.
- Il monitoraggio dell'espressione dei miRNA e l'identificazione di quelli che risultino correlati con la progressione dalla patologia epatica ad eziologia virale all'HCC.
- L'individuazione del ruolo di tali miRNA attraverso la ricerca dei target specifici, tramite lo sviluppo e l'impiego di metodi bioinformatici.

I risultati dello studio si possono riassumere in due sezioni principali: in primo luogo la definizione dei profili di espressione dei miRNA nelle diverse fasi della progressione della

malattia epatica a partire dalla fibrosi lieve fino all'insorgenza dell'epatocarcinoma, conseguita grazie all'impiego di metodiche di Real-time PCR e microarray. In secondo luogo lo sviluppo di un metodo bioinformatico di predizione dei geni target, il quale è stato successivamente impiegato nella ricerca della funzione di tali miRNA attraverso la definizione dei loro bersagli.

In conclusione, questo studio ha permesso di individuare alcuni dei miRNA caratterizzati dal fatto di avere un'espressione correlata a fasi distinte di malattia epatica e a specifici parametri clinici (grading-staging; eziologia). Sono state inoltre investigate le *pathway* nelle quali essi potrebbero essere coinvolti, con l'obiettivo e l'auspicio che possano rivelarsi utili come strumenti diagnostici e prognostici e per lo sviluppo di nuove terapie basate sul trasferimento di miRNA o sull'inibizione della loro espressione.

ABSTRACT

MicroRNAs (miRNAs) are non-coding RNAs whose role in gene regulation has been discovered and gradually clarified in recent years.

The action of miRNAs, like that of other types of non-coding RNA, is based on their physical interaction with gene transcripts (mRNA), the inhibition or decrease of their translation and, as a consequence, of their protein production. As a result, miRNAs biological effect depends on the nature of their targets. Currently, there is not a complete map of the targets for each of the 721 human known miRNAs, but several studies have already shown a direct involvement of many of them in the pathogenetic mechanisms of a large number of diseases including cancer.

Hepatitis B (HBV) and C (HCV) viruses infect hundreds of thousands people worldwide every year and cause necroinflammatory diseases which, though in different ways and rates, can cause the development of a chronic disease able to progress and leading to the development of fibrosis, cirrosis and hepatic cancer (HCC); the latter is one of the most aggressive human cancers and its early diagnosis and therapy are still unsolved issues.

Thus, many reasons are pressing the scientific community towards the research of novel strategies useful to effectively improve the health management of the disease.

As regards miRNAs, several studies have been carried out showing the existence of a clear pattern of specific miRNA expression in hepatic cancer tissue, while their role in the onset and progression of hepatic disease remains largely unclarified and, at the moment, not much studied.

Starting from these assumptions, the aims of this study are:

- The definition of a protocol for the analysis of miRNAs in tissue samples.
- The monitoring of miRNA expression and the identification of those which correlate with hepatic disease progression from viral hepatitis to HCC.
- The identification of the roles of these molecules through the research of their specific targets, thanks to the development and the application of bioinformatic tools.

The results of this study can be summarized in two main sections: first, the definition of miRNAs expression profiles in the different phases of the hepatic disease progression, starting from slight fibrosis to the onset of hepatocellular carcinoma, obtained with realtime PCR and microarray

experiments. Second, the development of a bioinformatic tool able to predict miRNA targets, which has subsequently been applied in order to investigate their function.

In conclusion, this study has allowed the identification of some miRNAs whose expression was correlated to distinct phases of hepatic disease and to specific clinical parameters (grading-staging, etiology). Moreover, we investigated the pathways that could be involved by miRNAs differential expression, with the aim and the hope that they will be useful as diagnostic and prognostic instruments and will bring to the development of new therapeutic strategies based on miRNA transfer or on the inhibition of their expression.

INDICE

1. Introduzione	7
1.1. Le epatiti virali	7
1.1.1. Il virus dell'epatite B	7
1.1.2. Il virus dell'epatite C	8
1.1.3. Il ruolo del sistema immunitario nella patogenesi della malattia epatica	9
1.1.4. La risposta epatica al danno immuno-mediato	11
1.2. Fibrosi e cirrosi	11
1.2.1. Grading e staging	12
1.3. L'epatocarcinoma	13
1.4. I microRNA	14
1.4.1. miRNA e cancro	17
1.4.2. miRNA ed epatiti croniche-virali	18
1.5. Softwares di predizione dei geni target dei miRNA	19
1.5.1. Miranda	19
1.5.2. RNAhybrid	19
1.5.3. Pita	20
1.6. Scopo della tesi	21
2. Materiali e metodi	23
2.1. Pazienti	23
2.1.1. Pazienti affetti da epatite cronica	23
2.1.2. Pazienti affetti da HCC	24
2.2. Purificazione dell'RNA da tessuti	25
2.3. Quantificazione e analisi di qualità del RNA	25
2.4. Analisi dell'espressione dei miRNA mediante real-time PCR	26
2.5. Analisi dei risultati di real-time RT-PCR	29
2.6. Analisi dell'espressione dei miRNA mediante microarray	30
2.7. Analisi statistica dei microarray	32
2.8. Predizione bioinformatica dei geni target dei miRNA	35
3. Risultati	41
3.1. Parte I: Espressione dei microRNA nell'epatocarcinoma e nelle epatiti	

virali croniche	41
3.1.1. Espressione dei miRNA nell'HCC rispetto al tessuto non tumorale adiacente	41
3.1.2. Espressione del profilo globale di espressione dei miRNA nell'epatite virale cronica	47
3.1.2.1. Analisi in base all'eziologia virale	48
3.1.2.2. Confronto tra i gruppi di campioni con fibrosi o cirrosi	50
3.1.2.3. Analisi in base alla classificazione per grading	52
3.1.2.4. Analisi in base alla classificazione per staging	53
3.1.2.5. L'espressione di alcuni miRNA è associata a più parametri di progressione dell'epatite virale cronica	55
3.2. Parte II: Predizione dei target dei miRNA mediante analisi bioinformatica	56
4. Discussione	61
5. Appendice A: implementazione degli algoritmi in Perl	67
6. Bibliografia	81

1. INTRODUZIONE

1.1. Le epatiti virali

I virus dell'epatite B (HBV) e dell'epatite C (HCV) sono membri rispettivamente della famiglia degli hepadnavirus e dei flavivirus. Essi sono caratterizzati da uno spiccato tropismo per il fegato, non sono citopatici, si trasmettono per contatto con sangue e liquidi corporei infetti e sono causa di malattie necroinfiammatorie epatiche acute e croniche (Ganem et al., 2004; Pawlotsky, 2004). L'infezione da HBV in individui adulti immunocompetenti provoca, in più del 95% dei casi, una malattia epatica transiente e auto-limitante, che si risolve con la completa eliminazione del virus; nei neonati, invece, l'esposizione alla nascita si traduce in un'infezione persistente in più del 90% dei casi. L'infezione persistente da HBV è associata a vari gradi di malattia epatica cronica, e spesso progredisce con lo sviluppo di cirrosi e carcinoma epatocellulare (HCC) (Chisari et al., 1995). Più di 350 milioni di persone nel mondo sono affette da infezione cronica da HBV, e ogni anno ne muoiono circa un milione a causa delle complicanze a lungo termine della malattia (Chisari et al., 1995). In contrasto con la storia naturale di HBV, l'infezione acuta da HCV diventa persistente nel 70-90% dei casi; oltre 170 milioni di persone sono ad oggi infettate cronicamente da questo virus e soffrono delle stesse patologie a lungo termine provocate dall'infezione cronica da HBV (Hoofnagle, 2002). Da oltre vent'anni è disponibile un vaccino molto efficace contro l'infezione da HBV, mentre per HCV, nonostante gli sforzi della comunità medico-scientifica, non è stato ancora messo a punto un sistema efficace; per di più, gran parte dei pazienti affetti da infezione cronica da HBV e HCV non rispondono alle terapie attualmente disponibili basate principalmente su analoghi nucleosidici, interferone e ribavirina (Ganem et al., 2004; Pawlotsky, 2004).

1.1.1. Il virus dell'epatite B

Il virus dell'epatite B è un virus della famiglia degli *hepadnaviridae*, caratterizzato da un genoma circolare a DNA parzialmente a doppio filamento. HBV è classificato in 8 genotipi (A-H), ognuno dei quali ha una distinta distribuzione geografica. Il genoma, il cui filamento più lungo supera di poco le 3200 basi, è organizzato in modo compatto, con 4 *open reading frames* (ORF) parzialmente sovrapposte:

- C (core) codifica sia per il nucleocapside virale che per l'antigene HBeAg, a seconda del sito di inizio della traduzione. La funzione di HBeAg è ancora ampiamente sconosciuta, anche se sembra essere implicato nel promuovere un'infezione persistente.
- X è una proteina con funzioni multiple, implicata nella trasduzione del segnale, nell'attivazione trascrizionale, nella riparazione del DNA e nell'inibizione della degradazione proteica (Bouchard

et al., 2004). La funzione biologica e i meccanismi funzionali dell'antigene HBxAg rimangono però largamente sconosciuti.

- S (surface) codifica per proteine dell'envelope e può essere suddivisa, strutturalmente e funzionalmente, nelle regioni pre-S1, pre-S2 e S.

- P è la polimerasi virale, costituita da circa 800 aminoacidi e divisa funzionalmente in 3 domini: il dominio terminale, coinvolto nell'incapsidazione e nell'inizio della sintesi del filamento negativo; la trascrittasi inversa, che catalizza la sintesi del genoma; la ribonucleasi H, che degrada l'RNA pregenomico e facilita la replicazione.

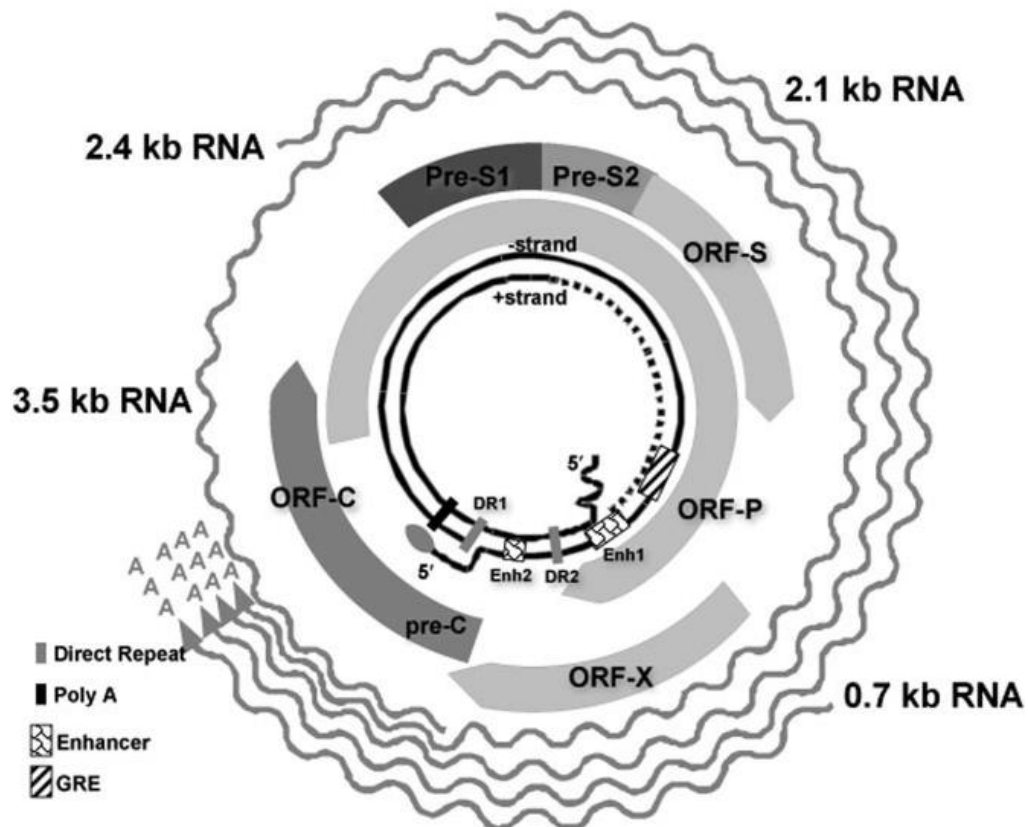


Fig. 1: Rappresentazione del genoma di HBV.

1.1.2. Il virus dell'epatite C

Il virus dell'epatite C è un virus della famiglia dei *flaviviridae*, il cui genoma è costituito da RNA a singolo filamento positivo, di circa 9600 basi.

Esistono 6 genotipi principali di HCV, a loro volta divisi in numerosi sottotipi. Mentre i genotipi 4, 5, e 6 sono decisamente più caratteristici delle popolazioni africane, arabe e asiatiche, i genotipi 1 (specialmente i sottotipi 1a e 1b), 2 e 3 sono i più frequenti nelle popolazioni autoctone europee, compresa quella italiana. Sul piano clinico, si è potuto constatare un differente esito delle

infezioni dovute a diversi genotipi (p.es., il genotipo 3 tende a cronicizzare molto meno frequentemente del genotipo 1) (Aronsohn et al., 2009).

Il genoma di HCV codifica un'unica poliproteina, trascritta a partire da un IRES (sito di legame interno per i ribosomi) posto nella sequenza 5'-UTR.

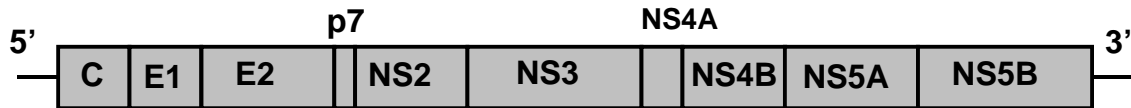


Fig. 2: Rappresentazione schematica del genoma di HCV riportante la suddivisione relativa alle proteine virali.

La poliproteina subisce in seguito una serie di eventi proteolitici che portano alla produzione di 10 distinte proteine a funzione perlopiù nota:

- C (core) è la proteina strutturale che costituisce il nucleocapside virale. Oltre a questo ruolo, core è una proteina largamente coinvolta nei meccanismi di interazione con la cellula ospite. Numerosi studi riportano la sua capacità di legare p53, p73, Rb, e di modulare l'attività di molte altre proteine cellulari coinvolte sia nella risposta immune che nell'apoptosi, oltre che di partecipare a importanti vie di regolazione cellulare (vie di wnt , MAPK e NFkB) (Fukutomi et al., 2005; Herzer et al., 2005).
- E1 ed E2 sono le glicoproteine dell'envelope virale e sono le principali responsabili della polimorficità genetica del virus.
- p7 è una proteina ancora in fase di caratterizzazione, un canale ionico transmembrana.
- NS2 (non structural 2) è una proteasi di incerta funzione.
- NS3 ha un dominio RNA-elicasi ed un dominio serin-proteasico coinvolto nel taglio delle successive proteine. NS3 interagisce con numerose proteine e vie di segnalazione cellulari ed è ritenuta responsabile sia di attività anti-immunitarie che trasformanti.
- NS4A è una proteasi che coopera con NS3.
- NS4B è una proteina integrale di membrana che induce la formazione di un compartimento specializzato chiamato "membranous web", il presunto sito di replicazione dell'RNA.
- NS5A è una proteina in grado di interagire sia con le vie di segnalazione citoplasmatiche che con le attività di trascrizione nucleare dell'ospite. Inoltre inducendo l'espressione di IL-8, contrasta la risposta interferonica cellulare.
- NS5B è la RNA-polimerasi RNA-dipendente che permette al virus di replicarsi.

1.1.3. Il ruolo del sistema immunitario nella patogenesi della malattia epatica

Evidenze sperimentali indicano che gli esiti delle infezioni da HBV e HCV e la concomitante patogenesi delle malattie epatiche siano determinate da interazioni tra virus e ospite, mediate dal sistema immunitario (Guidotti et al., 2006). La difficoltà nella delucidazione dei meccanismi

patogenetici di questi virus risiede da un lato nella carenza di sistemi per la crescita in coltura e, dall'altro, nella mancanza di modelli animali (se si esclude lo scimpanzé) suscettibili all'infezione; ciononostante, un grande passo in avanti è stato compiuto in anni recenti, grazie all'analisi della storia naturale e dell'immunobiologia dei due virus nell'uomo e nello scimpanzé, e allo studio delle infezioni di altri virus delle famiglie degli *hepadnaviridae* (Seeger et al., 2000) e dei *flaviviridae* (Buckwold et al., 2003) in animali suscettibili. Un ulteriore progresso è stato ottenuto grazie a studi su linee cellulari e modelli murini esprimenti specifici geni virali oppure in grado di replicare a vari livelli il ciclo vitale e replicativo dei virus (Bartenschlager et al., 2003; Disson et al., 2004).

La relazione virus-ospite è un processo dinamico nel quale il virus cerca di diminuire la sua visibilità, mentre l'ospite tenta di prevenire ed eradicare l'infezione limitando il più possibile i danni a sé stesso. A seconda della natura delle cellule infettate e dell'agente infettante, l'infezione può innescare i meccanismi difensivi dell'immunità innata per limitare l'entità della diffusione del virus: questi meccanismi includono l'induzione di apoptosi, la produzione di citochine antivirali (come l'interferone di tipo I, α e β), l'innescamento di funzioni effettrici di componenti cellulari come le cellule *natural killer* (NK) e le cellule *natural killer T* (NKT). Esse possono essere rapidamente reclutate nel sito dell'infezione ed hanno il potenziale di riconoscere le cellule infettate prima dell'induzione dell'espressione del complesso maggiore di istocompatibilità (MHC) di classe I sulla superficie cellulare: cellule NK e NKT attivate possono partecipare alla patogenesi della malattia sia direttamente che indirettamente, producendo fattori solubili come citochine e chemochine ad azione antivirale, reclutando cellule infiammatorie nel tessuto infettato e modulando la risposta immunitaria adattativa. Sulla base di studi effettuati sullo scimpanzé è ormai appurato che HBV e HCV si replicano e si diffondono nel fegato in modo non citopatico e che i meccanismi difensivi dell'immunità innata non contribuiscono significativamente al controllo della viremia o alla patogenesi della malattia epatica (Thimme et al., 2003; Lanford et al., 2001).

Diverso è il discorso per quanto riguarda la risposta immunitaria adattativa: i linfociti T citotossici (CTL) CD8⁺ e i linfociti T-helper CD4⁺ giocano un ruolo chiave nell'immunità antivirale e, come le cellule dell'immunità innata descritte in precedenza, partecipano alla patogenesi virale sia direttamente che indirettamente. Le citochine e le chemochine prodotte dalle cellule T contribuiscono da un lato all'immunopatologia, in quanto promuovono le risposte immunitarie cellulari, ma dall'altro sono in grado di inibire la replicazione virale e l'espressione di antigeni virali senza produrre effetti citopatici, contenendo il danno tissutale e diminuendo il numero di cellule che devono essere uccise per eradicare l'infezione. Inoltre, esse partecipano alla modulazione delle risposte anticorpali che contribuiscono alla *clearance* del virus principalmente bloccandone l'ingresso nelle cellule suscettibili ed eliminandolo dal sangue (Guidotti et al., 2001).

Numerose evidenze sperimentali indicano che la risposta immunitaria mediata da CTL svolge un ruolo fondamentale nella patogenesi della malattia epatica durante l'infezione da HBV e HCV, primo perché l'inizio del danno epatico coincide con l'ingresso nel fegato delle cellule T CD8+ virus-specifiche (Thimme et al., 2003), e secondo perché è stata dimostrata una forte associazione tra l'intensità della risposta CTL virus-specifica e la severità del danno epatico non solo negli scimpanzé, ma anche in pazienti con infezioni acute e croniche da HBV e HCV (Shoukry et al., 2004). Indubbiamente, molto spesso i pazienti con epatiti virali acute che guariscono completamente sviluppano un danno epatocitario relativamente severo associato ad una vigorosa risposta CTL. Al contrario, il danno cellulare epatico è più contenuto nei pazienti con infezioni croniche, nei quali la risposta CTL è estremamente debole (Shoukry et al., 2004). Tutti insieme, questi risultati indicano che la risposta CTL è essenziale per indurre la maggior parte del danno epatico associato alle epatiti virali.

1.1.4. La risposta epatica al danno immuno-mediato

La risposta del fegato al danno immuno-mediato include la rigenerazione delle cellule epatocitarie e l'attivazione di processi infiammatori/fibrogenici che somigliano a quelli impiegati nella guarigione delle ferite. I meccanismi rigenerativi possono avere un effetto antivirale nel caso in cui il genoma virale non sia in grado di sopravvivere all'evento di divisione cellulare e le cellule figlie siano protette dalla re-infezione, mentre in presenza di una inefficiente risposta immunitaria, nei pazienti con infezione cronica, si instaura un processo ciclico di lieve distruzione tissutale e successiva riparazione che, nel lungo periodo, porta a fibrosi, cirrosi, steatosi e HCC (Fausto, 2004). Al processo di riparazione, che è caratterizzato dalla deposizione di componenti della matrice extracellulare (ECM), partecipano numerosi tipi cellulari, tra cui le cellule stellate (HCS), i miofibroblasti, i fibroblasti, le cellule di Kupffer, i monociti, le cellule NK e le cellule T (Fausto, 2004). Nelle epatiti virali croniche, però, il processo riparativo va incontro a deregolazione, facilitando una eccessiva e qualitativamente abnorme deposizione di ECM e aprendo la strada allo sviluppo della fibrosi e della cirrosi.

1.2. Fibrosi e cirrosi

La fibrosi e la cirrosi epatica sono conseguenze comuni di malattie epatiche croniche ad eziologia virale e non (es: abuso di alcool, accumulo di ferro o rame, etc.). Esse sono caratterizzate da uno sbilanciamento tra la fibrogenesi e la fibrolisi (Ramadori et al., 2004), che risulta nell'eccessiva deposizione intraepatica di ECM qualitativamente differente, nella sua composizione e organizzazione, da quella della normale rigenerazione epatica. La conseguenza di questo processo è la formazione di un denso reticolo di ECM, resistente alla degradazione enzimatica, intorno alle aree portali del fegato (fibrosi portale), che col passare del tempo si espande nei lobuli e forma dei

setti che possono arrivare a connettere la vena portale e quella centrale (fibrosi portocentrale) (Ramadori et al., 2004).

La cirrosi rappresenta la fase finale della fibrosi, nella quale i setti fibrosi arrivano a circondare noduli di epatociti in rigenerazione, causando profonde modificazione dell'architettura del fegato, insufficienza funzionale e riversamento di sangue venoso contenente tossine nella circolazione sistemica (Ramadori et al., 2004). La velocità della progressione della fibrosi e della cirrosi durante le infezioni croniche da HBV e HCV è accelerata dalla severità e dalla durata della malattia epatica, e fattori noti per essere associati ad un più severo danno epatico (es: età avanzata al momento dell'infezione, sesso maschile, abuso di alcool, obesità, diabete e coinfezione da HIV) sono anch'essi implicati in una progressione più rapida di queste patologie (Ramadori et al., 2004). Sulla base dei risultati sintetizzati in precedenza, è ormai chiaro che la patogenesi di fibrosi e cirrosi durante le epatiti virali sia multifattoriale, richiedendo la presenza di danno epatico ricorrente immuno-mediato, infiammazione sostenuta e attivazione di cellule produttrici di ECM.

1.2.1. Grading e staging

A partire dagli anni '90 è avvenuta una rivoluzione per quanto riguarda la classificazione delle epatiti virali croniche, dovuta alla consapevolezza dell'inadeguatezza degli allora esistenti parametri di categorizzazione per la valutazione dei cambiamenti istologici durante i trial clinici (Scheuer, 1991; Ishak et al., 1995): i patologi hanno quindi introdotto i concetti del *grading* e dello *staging*, tutt'ora utilizzati per classificare i campioni biotici. Questi nuovi parametri hanno la funzione di dividere il continuum della progressione dell'epatite cronica in categorie discrete, meno prone a variabilità soggettive del patologo e più adatte a studi comparativi. In particolare, il *grading* è usato per descrivere l'intensità dell'attività necroinfiammatoria, mentre lo *staging* è una misura della fibrosi e delle alterazioni dell'architettura epatica. Esistono diversi indici proposti per classificare le epatiti, basati su *staging* e *grading*, e in tabella 1 è riportata la classificazione di Sheuer che è stata presa come riferimento per questo studio.

<i>Grade</i>	<i>Portal/periportal activity</i>	<i>Lobular activity</i>
0	None	None
1	Portal inflammation	Inflammation but no necrosis
2	Mild piecemeal necrosis	Focal necrosis or acidophil bodies
3	Moderate piecemeal necrosis	Severe focal cell damage
4	Severe piecemeal necrosis	Damage includes bridging necrosis
<i>Stage</i>	<i>Fibrosis</i>	
0	None	
1	Enlarged, fibrotic portal tracts	
2	Periportal or portal-portal septa, but intact architecture	
3	Fibrosis with architectural distortion, but no obvious cirrhosis	
4	Probable or definite cirrhosis	

Tab. 1: classificazione di Sheuer delle epatiti croniche.

1.3. L'epatocarcinoma

La forma di cancro definita come epatocarcinoma (in seguito abbreviato in HCC, dal nome inglese) è una fra le più diffuse al mondo, e nonostante gli sforzi notevoli della comunità medica e scientifica, permane una delle forme più letali. Nonostante la grande maggioranza dei casi di HCC mondiali si concentri in aree geografiche molto lontane dall'Italia (Africa sub-sahariana ed estremo oriente), l'aumento registrato nel numero di casi, unitamente all'aumento della popolazione a rischio, ne fanno un'autentica minaccia sanitaria anche nel nostro Paese.

I fattori di rischio principali per l'insorgenza di questa neoplasia appaiono essere, ad oggi, i medesimi della cirrosi epatica (con l'aggiunta, evidentemente limitata ad alcune realtà geografiche, dell'assunzione di alimenti contaminati con aflatossina B1 di *Aspergillus flavus*).

In particolare, circa l'80% dei casi di HCC risulta insorgere su fegato cirrotico, ed evidentemente la progressione cirrosi-cancro è al momento oggetto di studio da parte della comunità scientifica. I fattori di rischio per HCC più diffusi nel nostro paese appaiono pertanto essere: infezione da HCV, infezione da HBV, abuso di alcool, patologie metaboliche e tesaurismi (emocromatosi, morbo di Wilson). Sebbene un persistente danno epatico immuno-mediato possa essere sufficiente per causare l'insorgenza di HCC, altri fattori virali possono contribuire alla carcinogenesi durante le infezioni croniche come, ad esempio, l'integrazione del genoma di HBV in prossimità di geni pro-carcinogenici (es: oncogeni, recettori per fattori di crescita, telomerasi, etc.). Va detto, tuttavia, che l'integrazione di HBV non dovrebbe essere possibile in epatociti quiescenti e, dunque, se questo processo contribuisce alla carcinogenesi, è probabile che sia secondario ad altri eventi procarcinogenici che innescano la rigenerazione epatocitaria. Anche l'espressione di polipeptidi virali, come la proteina X di HBV, è stata implicata nel processo di trasformazione maligna degli epatociti infettati: tali prodotti virali possono contribuire alla carcinogenesi grazie alla loro capacità di attivare fattori di trascrizione e interagire con numerose vie di trasduzione del segnale (Brechot, 2004). HCV, invece, non si integra nel genoma ospite in quanto è un virus ad RNA, ma produce alcune proteine in grado di regolare la proliferazione e la vitalità cellulare (es: le proteine C, NS3 e NS5A) (Liang et al., 2004).

Complessivamente, si può a buona ragione concludere che, partendo da differenti fenomeni (mutazione inserzionale, danno immuno-mediato, azione di oncogeni virali, repressione di oncosoppressori cellulari), si arrivi alla deregolazione di elementi chiave per il controllo della stabilità genetica, della vitalità e della proliferazione cellulare i quali, non diversamente da quanto avviene in molti altri tipi di cancro, attivano il processo che, gradualmente e in tempi relativamente lunghi, porta all'insorgenza dell'epatocarcinoma.

Al momento, la diagnosi viene effettuata principalmente su risultanze istologiche (biopsia epatica), che fanno seguito ad indicazioni provenienti da esami di tipo radiologico (risonanza magnetica, TAC) o ecografico, mentre non esiste ancora una terapia farmacologica in grado di

ridurre in modo sostanziale la mortalità di questo cancro. L'unica terapia con effetto sostanziale sulla sopravvivenza dei pazienti è quella chirurgica (Llovet et al., 1999). Tale terapia può essere costituita dalla resezione epatica, ma anche da interventi loco-regionali quali termoablazione (RFA) e iniezione percutanea di etanolo (PEI), e dalla chemoembolizzazione (TACE). Tuttavia, la terapia chirurgica si può applicare solo ad una parte dei casi, segnatamente a quelli in cui le condizioni generali del paziente e quelle del fegato non canceroso siano sufficientemente buone da consentire l'intervento. Inoltre, specialmente nei casi legati ad infezione virale persistente, la probabilità di recidive è estremamente alta. La possibilità di ricorrere al trapianto di fegato (OLT) è evidentemente la migliore opzione terapeutica possibile, ma naturalmente non può essere applicata che a pochissimi soggetti per l'esiguità degli organi disponibili e per le numerose restrizioni applicate per l'inserimento dei pazienti con HCC nelle liste d'attesa.

Tornando alle terapie farmacologiche, sono state introdotte nell'uso clinico, talvolta con buon successo, oltre alla chemioterapia tradizionale, ormonoterapia (Tamoxifene), immunoterapia (IFN), terapia antiangiogenetica (Sorafenib). Anche la terapia genica è stata testata nel recente passato con risultati promettenti (Gerolami et al., 2003; Habib et al., 2002). Indubbiamente, comunque, in questo campo sono attesi ed auspicati miglioramenti sostanziali nel prossimo futuro.

1.4. I microRNA

Quando nel secolo scorso furono conseguite le grandi scoperte sugli acidi nucleici, venne stabilito il celebre principio secondo il quale si procedeva in natura dal DNA all'RNA alle proteine, fondando su di esso praticamente l'intera biologia molecolare. Il nuovo secolo ha portato alla sostanziale demolizione di queste certezze, peraltro ormai da anni notevolmente messe in discussione, attraverso la scoperta di un notevole mondo di 'eccezioni' alle regole precedentemente stabilite.

Una di queste eccezioni è costituita dai cosiddetti RNA non-codificanti, ovvero un gruppo piuttosto eterogeneo e numeroso di famiglie di molecole di RNA contraddistinte dal fatto di non venire tradotte in proteine, ma di avere una qualche funzione autonoma, non sempre conosciuta. Fra gli RNA non codificanti, senza voler fare un elenco esaustivo, troviamo l'RNA ribosomiale (rRNA) e il transfer RNA (tRNA), ben noti per la loro funzione legata alla produzione di proteine, ma anche altri RNA con attività catalitica-regolatoria come snRNA (small nuclear RNA), SRP (signal recognition particle) RNA, snoRNA (small nucleolar RNA), shRNA (short hairpin RNA), siRNA (small interfering RNA) e, per finire, miRNA (microRNA).

Qualunque sia il ruolo evolutivo di queste molecole, a partire dalla loro scoperta si sono aperti orizzonti del tutto nuovi per la ricerca biologica, impegnata da un lato nella loro scoperta, analisi, caratterizzazione e contestualizzazione funzionale, e dall'altro nella valutazione del loro

impiego come strumenti sperimentali e come possibili agenti terapeutici. Di fondamentale importanza a tale riguardo è stata la scoperta dell'RNA-interferenza (RNAi), un raffinato meccanismo cellulare di regolazione dell'espressione genica che sfrutta la capacità dei siRNA, brevi molecole di RNA a doppio filamento, di legare molecole di mRNA con alta omologia di sequenza e di provocare, attraverso l'assemblaggio di un complesso proteico, la degradazione del mRNA bersaglio.

Molti elementi accomunano i siRNA ai miRNA, a partire dal meccanismo di generazione e da quello di azione, parzialmente sovrapponibili.

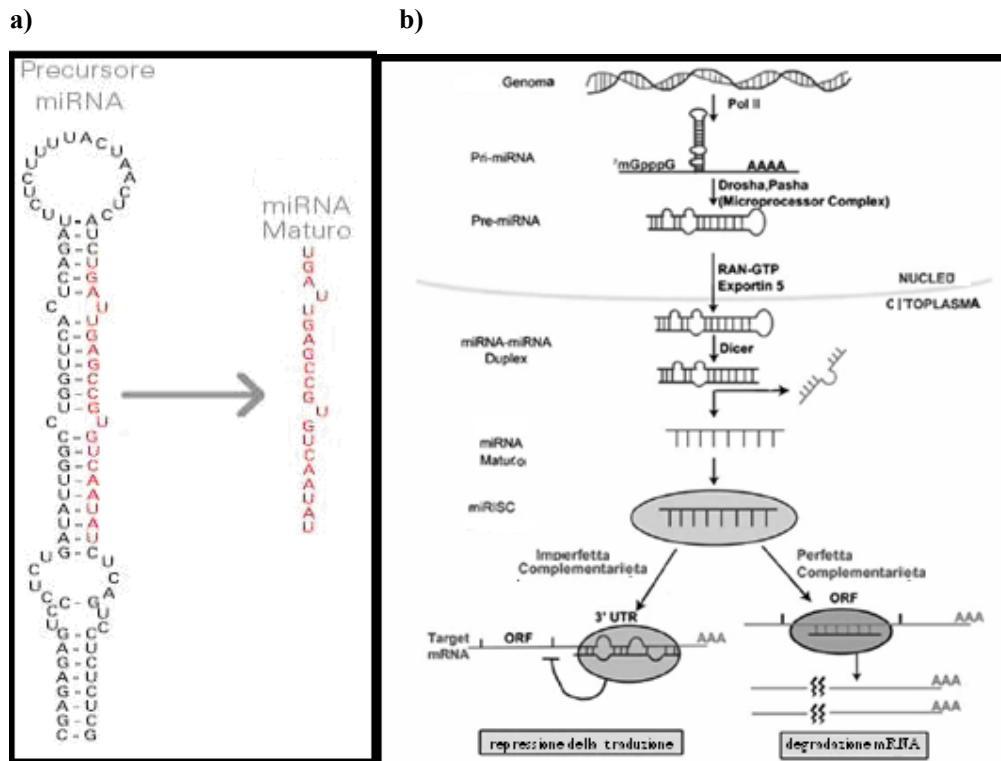


Fig. 3: a) Rappresentazione schematica di una molecola precursore e del corrispondente miRNA maturo. b) Schema del ciclo biologico dei miRNA: a partire dal genoma, vengono trascritti dalla polimerasi Pol III i pri-miRNA, molecole con caratteristica struttura secondaria ad appaiamento interno. Successivamente, il complesso Drosha-Pasha taglia la parte di RNA appaiata, che costituisce il pre-miRNA (70 nucleotidi circa). Dopo l'esportazione nel citoplasma da parte di exportina 5, il pre-miRNA è processato da Dicer a formare un doppio filamento (simile ai siRNA) che poi si denatura dando origine ad un miRNA maturo. Esso potrà complessarsi con RISC (RNA-induced silencing complex) e agire da agente di legame specifico per i mRNA bersaglio. Se il legame avviene con alta complementarietà, si avrà la degradazione del mRNA, altrimenti (più frequentemente) sarà repressa la traduzione del mRNA.

Entrando nel dettaglio, i miRNA, diffusi ubiquitariamente e molto conservati negli eucarioti superiori (ma rintracciabili anche in altre forme viventi, fra cui numerosi virus), sono molecole di RNA a singolo filamento di 19-25 nucleotidi, scoperte negli anni 90 nel nematode *C.elegans* (in

particolare i primi miRNA identificati furono lin-4 e let-7) (Lee et al., 1993). Essi sono trascritti dal genoma sia singolarmente, sia come gruppi di miRNA, sia congiuntamente ad altri geni (p.es. molti miRNA sono codificati da sequenze introniche).

Caratteristica fondamentale utile alla definizione del miRNA è la molecola precursore, con la sua peculiare struttura secondaria. Proprio sull'identificazione dei precursori dei miRNA nel genoma si è basata la scoperta, ottenuta principalmente per via bioinformatica, degli oltre 700 miRNA umani attualmente conosciuti, oltre alle centinaia noti per le altre specie.

Naturalmente, uno dei principali elementi di interesse scientifico collegato alla scoperta dei miRNA è risultato sin da subito la ricerca della loro funzione. Si è già accennato della stretta vicinanza del processo di produzione dei miRNA con quello che porta alla produzione di siRNA coinvolti nella cosiddetta 'RNA-interference'. In quest'ultimo meccanismo, ormai ben delucidato, le piccole molecole di RNA ottenute hanno la possibilità di interagire fisicamente con i trascritti cellulari, appaiandosi in modo specifico e innescando una risposta cellulare di difesa che porta alla distruzione del trascritto stesso. Per quanto riguarda invece i miRNA, sebbene risultino evidenze di organismi (es. vegetali) nei quali si assiste al medesimo comportamento, appare più frequente tuttavia un fenomeno leggermente diverso, nel quale il miRNA partecipa, in veste di elemento di riconoscimento semi-specifico (appaiamento imperfetto), ad un complesso ribo-proteico che lega i trascritti-bersaglio nel citoplasma cellulare e, senza attivare la risposta di degradazione dei medesimi, ne inibisce (in modo al momento non ancora esattamente definito) la traduzione da parte dei ribosomi. In sintesi, l'azione del miRNA più frequentemente osservata negli organismi animali consiste nel legame alla sequenza 3'-UTR di un mRNA, il quale non potrà essere efficacemente tradotto in proteina. Come per l'RNAi, anche verso i miRNA si è manifestato un grande interesse legato sia alla scoperta e caratterizzazione del loro ruolo nei processi fisiopatologici, sia ad un potenziale impiego terapeutico (in questo caso legato ai non molti miRNA con bersagli molecolari noti) (Tong et al., 2008).

Come sarà illustrato nel paragrafo successivo, una notevole mole di pubblicazioni scientifiche si sono in particolare concentrate sul ruolo dei miRNA nei processi di tumorigenesi. Tuttavia, altri studi hanno avuto come oggetto i miRNA di origine virale (tra cui i miRNA di CMV, EBV, KSHV, SV40) (Cullen, 2006; Pfeffer et al., 2005), l'interazione virus-ospite mediata dai miRNA prodotti da entrambe le parti e l'effetto di miRNA umani sul ciclo vitale dei virus. Di particolare interesse ai fini della nostra ricerca sono risultate le osservazioni dell'apparente interazione positiva fra un miRNA cellulare (mir-122) e il virus HCV in cellule umane infettate (Chang et al., 2008; Jopling et al., 2005).

1.4.1. miRNA e cancro

Al pari di ogni altro gene, un micro RNA può essere associato, in ragione della sua funzione biologica e della sua espressione specifica, ad un preciso stato fisiopatologico. Sin dalla loro scoperta, i miRNA sono oggetto di studio a tale proposito, e numerosissime pubblicazioni scientifiche negli ultimi anni hanno evidenziato correlazioni significative fra l'espressione di alcuni gruppi di miRNA e la presenza di determinate patologie, spesso neoplastiche (Blenkiron et al., 2007; Caldas et al., 2005; Calin et al., 2004; Cho et al., 2007; Stahlhut Espinosa et al., 2006).

Il meccanismo alla base della capacità dei miRNA di avere un ruolo attivo nello sviluppo del cancro può essere sintetizzato come segue: se un miRNA ha tra i suoi bersagli un determinato trascritto genico, e quel trascritto codifica per un oncosoppressore, un aumento dell'espressione locale del miRNA avrà l'effetto di favorire l'insorgenza o lo sviluppo di una neoplasia. Simile sarà chiaramente l'effetto della ridotta espressione di un miRNA che invece abbia tra i suoi bersagli un oncogene.

L'osservazione dell'espressione dei vari miRNA nelle neoplasie ha fornito la possibilità di individuare nell'espressione alterata di uno o più micro RNA, indipendentemente dalla conoscenza della loro funzione specifica, un marcatore identificativo di una determinata patologia potenzialmente utilizzabile come indicatore diagnostico con possibile impiego clinico. Nella maggior parte dei casi, questi studi sfruttano tecnologie di analisi su scala genomica, ovvero testano contemporaneamente l'espressione di tutti i miRNA conosciuti, utilizzando la tecnologia dei microarrays. Altri studi si basano invece sulla ricerca più circoscritta di gruppi ristretti di miRNA, con tecnologie standard come realtime PCR e Northern Blot.

La combinazione delle due tipologie di metodiche appare, al momento, la via di elezione per studi che intendano evidenziare, attraverso la determinazione dell'espressione dei miRNA, un loro ruolo diretto o indiretto nello sviluppo e nella progressione di patologie neoplastiche.

Naturalmente anche l'epatocarcinoma è stato oggetto di studio per quanto concerne l'espressione ed il ruolo dei miRNA, e dagli studi su scala genomica condotti con la metodica dei microarrays si può ricavare una lista di quelli che, fino ad oggi, sono risultati differenzialmente espressi nell'HCC rispetto al fegato non canceroso (Jiang et al., 2008; Li et al., 2008; Murakami et al., 2006; Shah et al.; 2007; Varnholt et al., 2008). Tra di essi citiamo mir-9, mir-17, mir-18, mir-21, mir-33, mir-135 e mir-222, che tendono a risultare più spesso sovraespressi nell'HCC, mentre let-7, mir-122, mir-130, mir-145, mir-150, mir-181, mir-195, mir-199a, mir-199b, mir-223 risultano più frequentemente sottoespressi. La lista è ovviamente solo parziale, considerata anche la mole di dati presenti in letteratura, ma propone dei candidati molto forti, in quanto confermati da studi effettuati da diversi gruppi di ricerca.

Va tuttavia segnalato come studi di espressione, come questi, non possano stabilire nessi causali fra i fenomeni, non possono cioè spiegare se l'alterazione dell'espressione di uno o più miRNA

sia causa o conseguenza della condizione tumorale. Sicuramente studi più focalizzati, e con una componente di caratterizzazione funzionale, possono essere di maggiore aiuto in questo senso.

Uno dei miRNA più spesso presi in esame è mir-122 (Elmén et al., 2008; Girard et al., 2008), espresso specificamente nel fegato, e indicato da più studi come promotore della crescita *in vitro* di HCV. Questo miRNA viene indicato in modo quasi univoco come nettamente sottoespresso nell'HCC, e sono stati proposti, e in parte validati, numerosi suoi bersagli come la Ciclina G1 (Gramantieri et al., 2007), regolatrice del ciclo cellulare, il trasportatore di membrana per aminoacidi CAT-1, enzimi metabolici come AldoA e Bckdk, e Cd320 (il recettore per VLDL). Nell'insieme, quindi, si propone per mir-122 un ruolo sia nella regolazione del ciclo cellulare che nella risposta allo stress e nel metabolismo lipidico.

Anche mir-21, che è invece sovraespresso in HCC (e negli altri tipi di cancro), è stato associato ad un preciso bersaglio, l'oncosoppressore PTEN (Meng et al., 2007) (che si aggiunge a PDCD4 e TPM1 visti in precedenza), mentre alcuni membri della famiglia let-7, sottoespressi in HCC, sono associati agli oncogeni Ras e c-Myc, e mir-145, sottoespresso in HCC (come ad esempio nel cancro del colon-retto) ha per possibile bersaglio la MAP-chinasi ERK5.

In conclusione, come per gli altri tipi di cancro studiati, nell'epatocarcinoma esiste una base di partenza di dati raccolti nel corso di numerosi studi che indicano la presenza di un gruppo di miRNA tendenti ad avere espressione deregolata rispetto al fegato non canceroso, e per alcuni di essi sono presenti indizi concreti di una possibile azione, diretta o indiretta, di interferenza con oncogeni ed oncosoppressori cellulari tale da giustificare un loro coinvolgimento diretto nella patogenesi del cancro.

1.4.2. miRNA ed epatiti croniche-virali

Se per l'HCC esiste una letteratura specifica, che si sta consolidando in questi ultimi anni, il ruolo dei miRNA nel corso delle epatiti virali croniche non è ancora stato delucidato. Infatti non esistono, al momento, studi sui profili di espressione dei miRNA atti ad identificare dei candidati che possano essere associati alla progressione della fibrosi e della cirrosi, né tantomeno lavori mirati a valutare la funzione specifica di alcuni miRNA e dei geni che essi modulano con la loro attività di repressione della traduzione. Pertanto si rende estremamente necessario uno sforzo in tale direzione, sia per aumentare il grado di comprensione dei meccanismi alla base della progressione della malattia epatica, sia per scopi terapeutici, dato che i profili di espressione dei miRNA potrebbero essere impiegati come marcatori predittivi di risposta ai trattamenti oppure, più direttamente, per lo sviluppo di terapie mirate al ripristino di valori di espressione nella norma dei miRNA deregolati, inibendo eventuali miRNA sovraespressi o aumentando l'espressione di quelli che risultino maggiormente sottoespressi.

1.5. Softwares di predizione dei geni target dei miRNA

L'identificazione dei geni target dei miRNA è un passo fondamentale per la comprensione del ruolo di queste molecole nei *network* di regolazione genica. Gli algoritmi computazionali, sviluppati sulla base di regole evidenziate sperimentalmente, sono una parte importante dello sforzo messo in atto dalla comunità scientifica per rivelare le interazioni tra miRNA e geni target. Esistono, ad oggi, numerosi algoritmi sviluppati per la predizione di geni target, con caratteristiche leggermente diverse tra loro, ma fondamentalmente accomunati dal fatto di basarsi principalmente sulla complementarità di sequenza tra il miRNA e il mRNA, l'unico vero meccanismo di interazione noto. Tuttavia, tutti i programmi esistenti generano una certa quantità di falsi positivi e, in minor numero, falsi negativi: questo problema è dovuto al fatto che, per la maggior parte dei miRNA noti, vengono predetti un gran numero di trascritti target, a causa della brevità della sequenza di interazione, e, inoltre, molti trascritti sono bersaglio di miRNA diversi. Se a questo si aggiunge che un singolo miRNA può avere molteplici trascritti target, si può intuire facilmente come una fetta importante di geni sia modulata dall'azione di queste molecole, stimata intorno al 50% dell'intero trascrittoma (Gusev, 2008). Non essendo questa la sede appropriata per una completa presentazione di tutti gli algoritmi attualmente disponibili, saranno presentati in seguito soltanto quelli impiegati per l'attività illustrata in questo lavoro di tesi.

1.5.1. Miranda

MiRanda (John et al., 2004) è uno dei primi algoritmi sviluppati per la predizione di geni target dei miRNA e la sua pubblicazione risale al 2004. L'algoritmo di scansione è basato principalmente sulla complementarità di sequenza tra miRNA maturo e sito di target, e in secondo luogo sull'energia di legame del duplex miRNA-target e sulla conservazione evolutiva del sito target in geni omologhi. Essendo uno dei primi programmi pubblicati, MiRanda è anche uno dei meno recenti e le sue predizioni potrebbero quindi risultare meno accurate rispetto ad altri software sviluppati successivamente.

1.5.2. RNAhybrid

L'algoritmo di predizione RNAhybrid (Krüger et al., 2006), senza dubbio uno dei più utilizzati, ha un nucleo centrale che consiste nella predizione della struttura secondaria più favorevole, dal punto di vista energetico, di molecole di RNA a doppio filamento. In estrema sintesi, il programma cerca quali siano gli appaiamenti energeticamente più probabili tra la sequenza del miRNA e quella del trascritto target, e successivamente filtra i risultati ottenuti sulla base di una soglia energetica fornita dall'utente. Il software permette di effettuare modifiche su alcuni parametri, come ad esempio impostare la *seed sequence* desiderata (ossia la regione del miRNA

che dovrebbe avere maggiore complementarità con il target), acconsentire o meno gli appaiamenti G:U oppure i *loop* tra le due sequenze, e risulta essere più rapido rispetto a MiRanda.

1.5.3. Pita

Pita (Kertesz et al., 2007) è il più recente dei software presentati in questa tesi, e rispetto ai due illustrati precedentemente, presenta un'innovazione importante: esso, infatti, è in grado di considerare l'accessibilità del sito di target per il legame con il miRNA. L'ipotesi è che la struttura secondaria del mRNA partecipi attivamente al riconoscimento del target, in quanto esiste un costo energetico necessario per rompere la struttura secondaria del target e 'liberare' il sito di legame per il miRNA, rendendolo in questo modo accessibile. Questa particolare caratteristica rende Pita un algoritmo unico in questo campo perché, considerando anche la diversità della sequenza che circonda il sito specifico di target di un miRNA, è in grado di spiegare la variabilità di efficienza osservata per il riconoscimento, da parte di uno stesso miRNA, di target con siti di legame identici. Per il resto, Pita è simile a RNAhybrid, perché anch'esso cerca gli appaiamenti energeticamente più favorevoli tra le due molecole di RNA.

1.6. Scopo della tesi

Lo scopo di questo lavoro di tesi è studiare i profili di espressione dei miRNA nella progressione della malattia epatica ad eziologia prevalentemente virale che, a partire dall'epatite cronica, può portare allo sviluppo di fibrosi, cirrosi ed epatocarcinoma. In secondo luogo, una volta identificati dei miRNA la cui espressione risulti essere correlata con l'avanzamento della malattia, l'obiettivo è investigarne la funzione in relazione ai geni che sono bersaglio della loro azione. Per questa seconda parte è previsto lo sviluppo di metodi bioinformatici.

2. MATERIALI E METODI

2.1. Pazienti

Lo studio è stato condotto su due coorti distinte di pazienti: il primo gruppo è costituito da 40 biopsie ottenute da pazienti affetti da epatite virale cronica, mentre il secondo consiste di 29 coppie di campioni di HCC e di tessuto epatico non tumorale adiacente, prelevati da pazienti affetti da HCC.

2.1.1. Pazienti affetti da epatite cronica

Le biopsie sono state effettuate tra il 2001 e il 2008 dall'equipe del Prof. F. Farinati; i campioni di biopsia sono stati rapidamente congelati in azoto liquido. La casistica di pazienti affetti da epatite virale cronica comprende soggetti con stadi diversi di patologia, a partire da fibrosi lieve fino alla cirrosi avanzata. I parametri clinici e istologici di ciascun campione sono riassunti in tabella 2.

Tabella 2: Dati clinici ed istopatologici dei casi di epatite virale cronica studiati.

CODICE	SESSO	ETA ^a	PRELIEV	GRADING	STAGING	EZIOLOGIA ^b	STEATOSI
BIO1	F	59	2002	2	2	HCV	Si
BIO2	F	58	2006	2	2	HCV 2 (a-c)	Si
BIO3	F	63	2006	3	2	HCV 1b	Si
BIO4	F	43	2007	1	1	HCV 1b	No
BIO5	F	41	2004	2	1	HCV 3a	No
BIO6	M	56	2007	1	1	HCV 2	No
BIO7	M	54	2007	1	1	HCV 1b	No
BIO8	M	33	2007	1	2	HCV 1b	Si
BIO9	M	60	2008	2	2	HCV 1	Si
BIO10	M	30	2007	1	1	HCV 1b	No
BIO11	M	58	2008	3	2	HCV 1b	Si
BIO12	M	55	2007	3	2	HCV 3a	Si
BIO13	F	60	2002	-	CIRROSI	HCV 1b	Si
BIO14	F	62	2006	2	CIRROSI	HCV 2 (a-c)	No
BIO15	M	46	2003	2	CIRROSI	HCV	Si
BIO16	M	48	2006	2	3	HCV 1a	No
BIO17	M	40	2006	2	3	HBV	Si
BIO18	F	58	2007	2	3	HCV 1b - alcool	Si
BIO19	M	48	2006	2	3	HCV - alcool	Si
BIO20	M	49	2004	3	CIRROSI	HCV 3a	Si
BIO21	M	60	2007	2	CIRROSI	HCV	Si
BIO22	M	51	2007	2	CIRROSI	HCV 1b	Si
BIO23	M	39	2001	1	2	HBV	No
BIO24	F	56	2005	2	1	HBV	No
BIO25	M	60	2006	3	2	HBV	Si
BIO26	F	65	2006	2	CIRROSI	HCV 1b	Si
BIO27	F	61	2007	3	3	HCV	Si
BIO28	M	54	2006	1	3	HBV	Si
BIO29	F	52	2008	2	1	HBV	Si
BIO30	M	66	2007	3	2	HCV 1b	Si
BIO31	M	31	2008	1	2	HBV	Si
BIO32	M	56	2006	2	3	HCV 1b	Si
BIO33	M	39	2008	1	2	HBV	No

BIO34	F	57	2007	2	1	HCV 2 (a-c)	Si
BIO35	M	29	2004	1	2	HBV	No
BIO36	M	47	2008	0	CIRROSI	HBV	No
BIO37	F	65	2002	2	2	HBV	Si
BIO38	M	47	2007	2	2	HBV	No
BIO39	M	31	2007	2	2	HBV	Si
BIO40	M	30	2007	1	1	HCV 1a	No

Note: a: età al momento del prelievo; b: eziologia riportata dalle note cliniche disponibili (per i pazienti HCV-positivi è riportato anche il genotipo virale).

2.1.2. Pazienti affetti da HCC

I campioni di HCC e tessuto epatico adiacente al tumore derivano da 29 pazienti ricoverati presso la Clinica Chirurgica I dell'Azienda Ospedaliera di Padova nel periodo 2006 - 2008. I tessuti sono stati prelevati durante l'intervento di resezione dell'HCC in condizioni sterili e preservati dalla degradazione tramite congelamento rapido in azoto liquido. Per la maggior parte dei pazienti è stato possibile risalire ai dati clinici di archivio, e, qualora disponibili, alle risultanze diagnostiche di laboratorio di possibile utilità per il nostro studio.

Tab. 3: Dati clinici ed istopatologici dei casi di HCC studiati.

CODICE	SESSO	ETA ^a	PRELIEVO	PREGRESSO ^b	EZIOLOGIA ^c
HCC0	F	62	2006	CIRROSI	HCV
HCC1	M	67	2006	CIRROSI	ALCOOL
HCC2	F	70	2006	EPATITE CR	HCV
HCC3	F	NA	2006	NA	NA
HCC4	F	50	2006	SANO	NA
HCC6	M	62	2006	NA	HCV
HCC7	M	51	2006	CIRROSI	ALCOOL
HCC8	F	35	2006	SANO	NA
HCC9	M	67	2006	CIRROSI	NA
HCC10	M	78	2006	EPATITE CR	HBV
HCC12	M	73	2006	CIRROSI	ALCOOL
HCC13	M	60	2006	CIRROSI	HCV
HCC14	M	61	2006	SANO	NA
HCC15	F	69	2006	STEATOSI	HBV
HCC16	M	58	2006	CIRROSI	HBV
HCC17	F	51	2006	SANO	NA
HCC18	M	27	2006	CIRROSI	HBV
HCC19	M	61	2006	CIRROSI	ALCOOL
HCC20	M	52	2006	CIRROSI	HBV
HCC21	M	71	2006	CIRROSI	HCV
HCC22	F	78	2006	EPATITE CR	HCV
HCC23	F	72	2007	CIRROSI	HCV
HCC31	M	77	2008	NA	HCV
HCC32	M	53	2008	NA	HCV/HBV
HCC33	M	NA	2008	CIRROSI	ALCOOL
HCC34	M	69	2008	NA	NA

HCC35	M	81	2008	NA	NA
HCC36	M	67	2008	CIRROSI	HCV
HCC37	M	60	2007	CIRROSI	ALCOOL

Note: a : età al momento del prelievo; b: condizioni istopatologiche del fegato pregresse all'insorgenza del HCC; c: eziologia riportata dalle note cliniche disponibili.

2.2. Purificazione dell'RNA da tessuti

I tessuti, congelati in azoto liquido subito dopo il prelievo, sono stati conservati a -80°C in attesa di essere processati per l'estrazione dell'RNA.

Il protocollo impiegato nel presente studio, ovvero il mirVana miRNA isolation kit (Ambion, USA), è una combinazione dei metodi più frequentemente impiegati per l'estrazione dell'RNA totale: la cosiddetta estrazione chimica (o metodo del fenolo-cloroformio) e l'estrazione su fase solida (o metodo della filtrazione) ed è stato appositamente sviluppato per l'estrazione di miRNA e RNA totale arricchito in miRNA. I metodi di estrazione chimica e di estrazione su fase solida, usati singolarmente, sono ottimali per il recupero dell'RNA totale, ma non consentono un ottimale recupero del RNA a basso peso molecolare, fra cui i miRNA.

Seguendo le indicazioni del produttore, il metodo di purificazione dei miRNA con il mirVana miRNA isolation kit prevede la disgregazione meccanica dei frammenti di tessuto (circa 0,1 g per frammento), eseguita tramite omogenizzatore meccanico Mixer Mill 300 (Qiagen, Germania), accompagnata dalla lisi chimica ottenuta tramite una soluzione detergente-denaturante (contenente guanididio tiocianato). Il lisato ottenuto viene successivamente incubato in una soluzione acida di fenolo-cloroformio e centrifugato per 5 min a 10000 x g per permettere la separazione dei componenti cellulari nelle varie fasi liquide risultanti. Recuperata la fase acquosa contenente l'RNA, essa viene solubilizzata in etanolo assoluto e quindi sottoposta a lavaggi su filtro (fibra di vetro) per eliminare i detergenti e i solventi impiegati nelle fasi precedenti, e infine eluita in soluzione di acqua-DEPC (dietilpirocarbonato, agente alchilante inibitore delle RNAsi, comunemente impiegato in soluzioni destinate all'RNA). L'RNA totale (arricchito in miRNA) così ottenuto viene successivamente conservato a -80°C fino a successivo utilizzo.

2.3. Quantificazione e analisi di qualità del RNA

L'RNA estratto col protocollo precedentemente illustrato è stato analizzato con spettrofotometro (NanoDrop-1000, Thermo Scientific, USA) per verificarne la qualità e la quantità. Un successiva verifica qualitativa e quantitativa dell'RNA è stata eseguita mediante elettroforesi capillare su gel denaturante utilizzando microchip Agilent su apparecchiatura Agilent 2100 Bioanalyzer (Agilent Technologies, USA). I risultati dell'elettroforesi capillare sono visualizzati in forma digitale che ne consente la quantificazione e l'analisi di parametri, quali per esempio l'*RNA integrity number* (RIN). In particolare, l'elettroforesi permette la visualizzazione dei due picchi di rRNA 18S e

28S, che insieme contano per circa l'80% dell'RNA totale, oltre che di un picco relativo al complesso degli RNA a basso peso molecolare. L'integrazione dell'area sottesa ai due picchi di rRNA permette di ricavare una stima della concentrazione dell'RNA totale (meno accurata di quella ottenuta allo spettrofotometro) e il calcolo del citato RIN, il quale è un parametro di notevole importanza in particolare in vista di successivi utilizzi che richiedano un alto livello di qualità dell'RNA (es. microarrays). Recentemente sono state introdotte nell'uso versioni dei reagenti e del software da utilizzare nell'Agilent 2100 Bioanalyzer specifici per l'analisi dei miRNA.

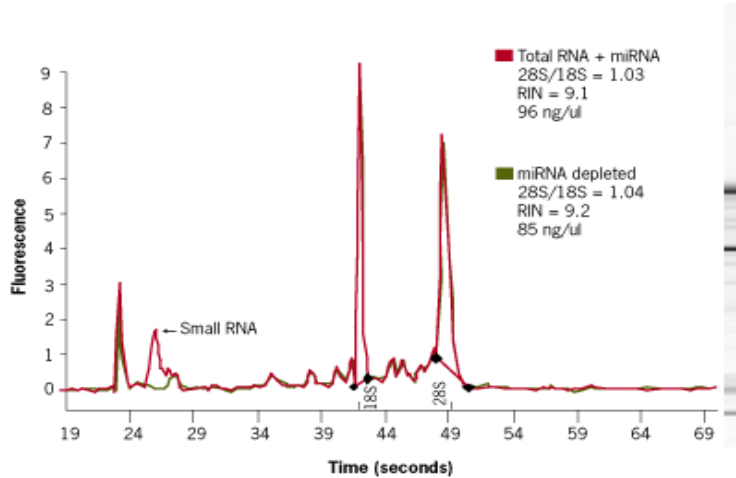


Fig. 4: esempio di analisi di qualità dell'RNA eseguita con Agilent Bioanalyzer. In questo caso, elettroferogramma di un campione di RNA totale (rosso) e di un campione di RNA privato dei miRNA (verde).

2.4. Analisi dell'espressione dei miRNA mediante real-time PCR

L'analisi dell'espressione dei miRNA nei campioni di tessuto è stata eseguita mediante RT-realtime-PCR quantitativa che permette di identificare gli specifici miRNA presenti nel campione e di valutarne l'espressione in rapporto a quella di un normalizzatore appositamente scelto. L'intero procedimento è basato sull'impiego di kit TaqMan microRNA assay (Applied Biosystems, USA). Brevemente, l'RNA è stato sottoposto a trascrizione inversa utilizzando primer (looped RT primer) specifici per ogni singolo miRNA e con una coda a forcina complementare ad uno dei primer impiegati per la real-time PCR (figura 5).

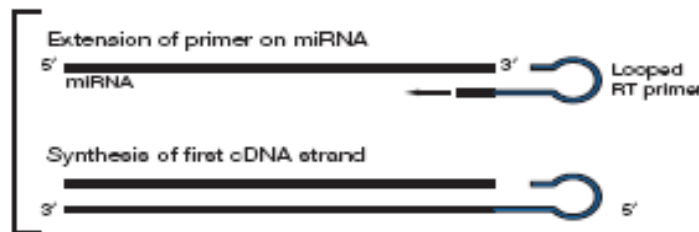


Fig. 5: Primers per retrotrascrizione utilizzati dal kit TaqMan microRNA assay (di kit TaqMan microRNA assay).

Tab. 4: composizione della miscela di reazione e protocollo di retrotrascrizione per miRNA.

Miscela di reazione	
<i>Componente</i>	<i>Quantità (µl)</i>
RNA totale	5
Buffer RT 10x	1.5
dNTPS 1mM	0.15
inibitore Rnasi 4U	0.19
Multiscribe Reverse Transcriptase 50U	1
RT primer 1x	3
acqua-DEPC	4.16
Protocollo di reazione	
<i>Durata (min)</i>	<i>Temperatura (°C)</i>
30	16
30	42
5	85

Dopo aver eseguito la RT seguendo il protocollo indicato in tabella 4 che richiede un quantitativo peraltro molto piccolo di RNA totale (1-10ng) , si allestisce la reazione di real-time PCR nell'apposita apparecchiatura (ABI-PRISM 7900HT, Applied Biosystems, USA) utilizzando il protocollo indicato dal produttore (Tabella 5).

Tab. 5: Composizione della miscela di reazione e protocollo di amplificazione real-time-PCR per miRNA.

Miscela di reazione	
<i>Componente</i>	<i>Quantità (µl)</i>
cDNA	1.33
TaqmanUniversal Master Mix 1x	10
Taqman miRNA assay 1x	1
acqua-mQ	7.67
Protocollo di reazione	
<i>Durata</i>	<i>Temperatura (°C)</i>
10 min	95°C
15 sec	95°C } 60°C } 40x
60 sec	

L'analisi mediante real-time RT-PCR dell'espressione dei miRNA in ciascun campione è stata eseguita in triplicato. In ogni reazione sono stati inseriti gli opportuni controlli (bianco di reazione e bianco di retrotrascrizione).

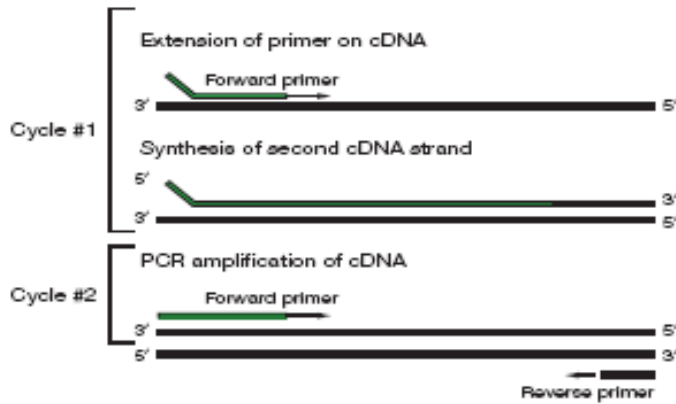


Fig. 6: meccanismo di funzionamento dei primers per real-time-PCR utilizzati nel nostro studio. Il primer Fw appaia un'ulteriore allungamento della sequenza da amplificare, mentre il primer Rw appaia alla sequenza artificialmente aggiunta nella fase di retrotrascrizione.

La reazione di PCR produce fluorescenza grazie all'impiego di sonde Taqman®, sinteticamente illustrata nella figura 7. Nel caso specifico, alla sonda è legato anche un elemento MGB (Minor Groove Binder), ovvero una modificazione chimica che, aggiungendo un secondo punto di legame (aspecifico) con la molecola di cDNA, permette di conferire alla sonda una maggiore temperatura di melting (T_m), caratteristica molto importante per mantenere alta la specificità di legame in una sonda di così breve lunghezza.

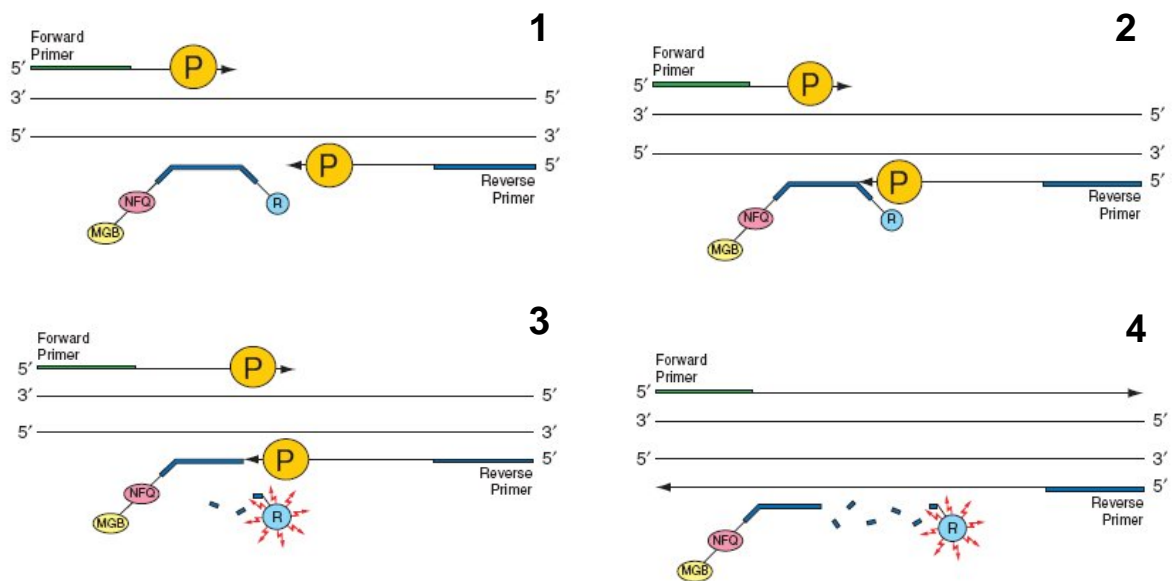


Fig. 7: meccanismo di funzionamento delle sonde TaqMan MGB per real-time-PCR. La sonda Taqman appaia il cDNA con legame specifico (1), successivamente a partire dai primers la polimerasi replica la sequenza fino ad incontrare (2) ed idrolizzare (3) la sonda ad esso appaiata. La degradazione della sonda permette al fluorocromo reporter R di allontanarsi dal fluorocromo quencher NFQ ed emettere fluorescenza (4).

2.5. Analisi dei risultati di real-time RT-PCR

Le curve di amplificazione (v. esempio del risultato grafico dell'analisi della fluorescenza) di ciascun campione possono essere utilizzate per effettuare analisi quantitative assoluta, in presenza di una curva standard con cui confrontarle, o relativa, utilizzando il sistema del $\Delta\Delta Ct$.

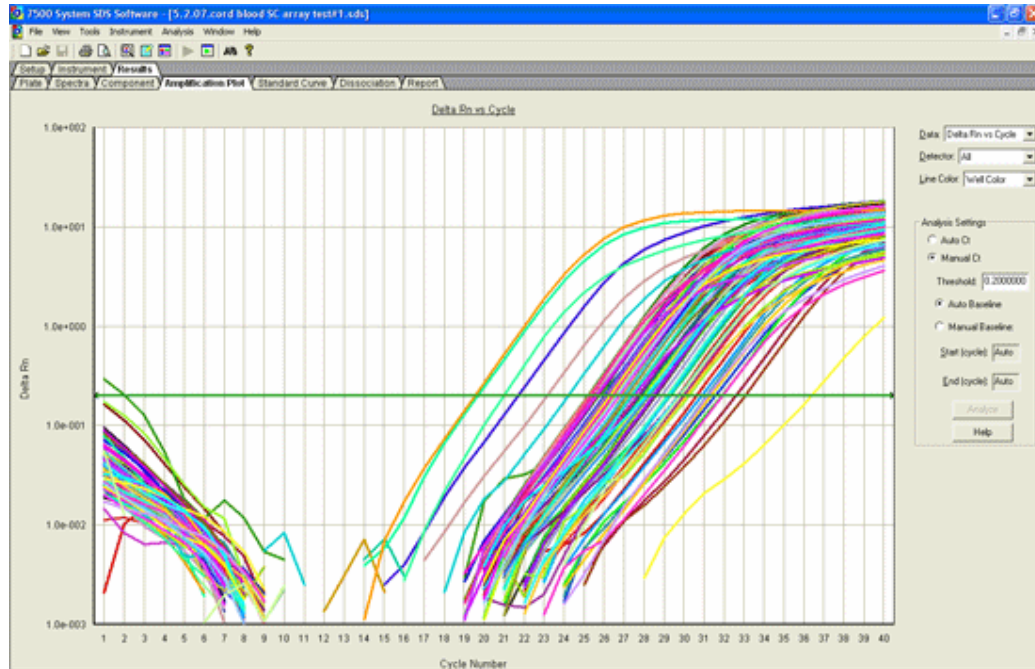


Fig. 8: esempio di risultato grafico dell'analisi della fluorescenza emessa durante l'amplificazione e registrata dall'apparecchiatura. Ogni curva rappresenta un campione diverso, la linea orizzontale è la soglia scelta dall'operatore. L'intersezione fra la soglia e la curva rappresenta il ciclo soglia Ct.

Questo metodo si basa sul presupposto che il confronto di curve di amplificazione di campioni diversi, se la reazione ha efficienza comparabile, può fornire la misura del rapporto fra le concentrazioni iniziali dei trascritti che vengono analizzati.

In sintesi, per ciascun campione, si analizzano sempre in parallelo il miRNA di interesse ed un secondo trascritto (in questo caso l'RNA nucleare RNU6B, espresso in modo ubiquitario e spesso impiegato in studi di espressione dei miRNA), e si normalizza il Ct (ciclo soglia) del miRNA con quello di RNU6B. Il dato ottenuto non è una quantità assoluta ma un dato di espressione relativo che può essere utilizzato per confrontarlo con altri dati dello stesso tipo, come vedremo in seguito. L'applicabilità del metodo del $\Delta\Delta Ct$ è subordinata al fatto che l'efficienza di amplificazione sia comparabile fra i vari trascritti analizzati, ed è stato pertanto necessario verificare preliminarmente anche questo aspetto conducendo apposite prove su diluizioni note di RNA di controllo.

2.6. Analisi dell'espressione dei miRNA mediante microarray

Lo studio del profilo di espressione dei miRNA nelle biopsie epatiche è stato eseguito mediante piattaforma microarray della ditta Agilent Technologies. In particolare, la versione di microarray utilizzata è la Human miRNA Microarray Kit (V2), contenente 723 sonde per miRNA umani e 76 per miRNA virali tratti dal Sanger Database del 2008, mentre la marcatura dei campioni è stata eseguita con il sistema monocolor miRNA Complete Labeling and Hybridization Kit, contenente la cianina 3-pCp che emette nella lunghezza d'onda del verde.

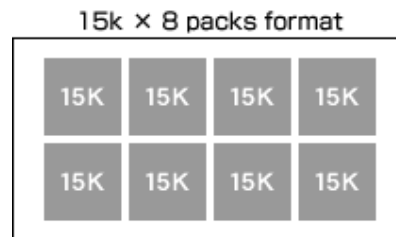


Fig. 9: Rappresentazione schematica di un vetrino per miRNA microarray (Agilent). Sul vetrino sono presenti 8 griglie identiche, sulle quali è possibile ibridare 8 campioni diversi. Le griglie contengono ciascuna 15000 spots costituiti da sonde relative a miRNA umani e virali presenti in più repliche.

Il principio su cui si basa la tecnica dei microarrays, anche nel caso di microarrays di miRNA, è quello dell'ibridazione tra molecole complementari di acido nucleico a singolo filamento: la prima, chiamata *probe*, è attaccata covalentemente alla superficie del vetrino ed ha una sequenza nota appositamente disegnata per potersi legare alla molecola complementare che si vuole studiare nel campione, chiamata *target*. Per amplificare il segnale, e renderlo quindi più facilmente rilevabile in fase di analisi, molte copie della stessa probe vengono posizionate nello stesso punto del vetrino, in modo da formare uno *spot*. Nel caso di avvenuta ibridazione, essa sarà rilevata in fase di scansione del vetrino grazie alla fluorescenza emessa dalla molecola target marcata con un opportuno fluoroforo.

Nel caso di microarrays di miRNA, per aumentare la scarsa specificità di ibridazione dovuta alla brevità delle sequenze in questione, il sistema sviluppato da Agilent prevede l'aggiunta di una Citosina al 3' delle molecole target in fase di marcatura e di una sequenza in grado di formare una forcina adiacente alla regione target-probe (figura 10).

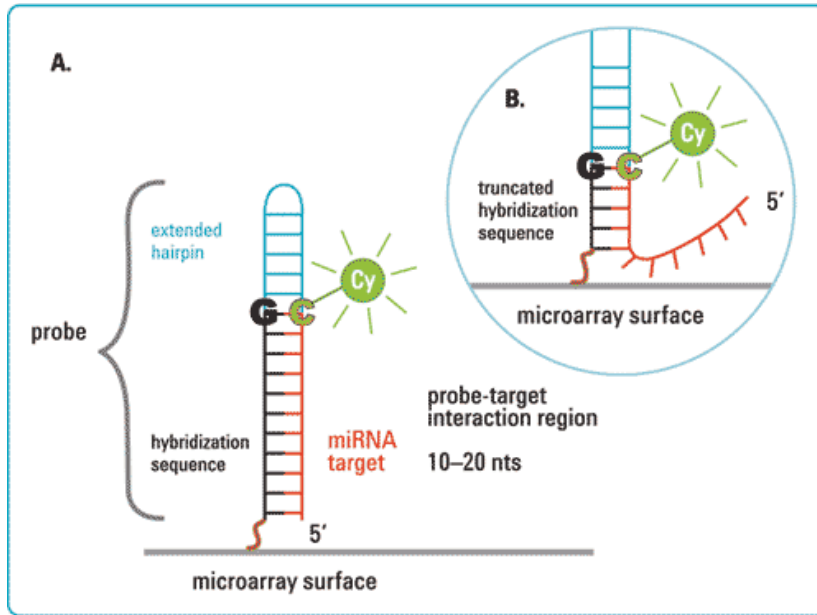


Fig. 10: Sistema di ibridazione target-probe nei microarrays di miRNA Agilent. A) Alla molecola probe (in nero), specifica per il riconoscimento di una molecola target (in rosso), viene aggiunta prima una guanina (G), in grado di appaiarsi alla citosina (C) inserita al 3' della molecola target in fase di marcatura, quindi una sequenza hairpin (in blu) in grado di formare una forcina per rendere più stabile l'ibridazione. **B)** Per destabilizzare sonde troppo stabili, che potrebbero creare un bias nel vetrino, viene ridotta la regione di appaiamento sonda-target.

Il protocollo di analisi dei microarray prevede l'impiego di RNA totale arricchito di miRNA. La quantità di RNA di partenza è di 100 ng, e il metodo di marcatura prevede la ligazione della cianina 3-pCp al 3' delle molecole di RNA del campione: la prima fase è la defosforilazione dell'RNA, che avviene per incubazione con l'enzima CIP (Calf Intestinal Alkaline Phosphatase) a 37°C per 30 minuti; quindi si procede con la denaturazione termica del campione (5-10 minuti a 100°C) seguita dalla fase di ligazione della cianina per mezzo dell'enzima T4 RNA ligasi (2 ore a 16°C). A questo punto, dopo una fase di purificazione su colonna, il campione viene seccato in una centrifuga a vuoto, fino ad ottenere un pellet sul fondo della provetta. Il campione deve essere risospeso in acqua RNase-free e buffer di ibridazione, e viene aggiunto inoltre il Blocking Agent, una soluzione che consente di saturare tutti i siti di legame aspecifico. Dopo un'incubazione a 100°C per 5 minuti il campione è pronto per l'ibridazione: questa fase è molto delicata, in quanto su un singolo vetrino vengono caricati 8 campioni diversi, separati uno dall'altro da una griglia chiamata *gasket*; una volta assemblata la camera di ibridazione, essa viene posta in un forno per 20 ore a 55°C, in rotazione a 20 rpm.

Al termine della fase di ibridazione sono necessari diversi passaggi di lavaggio del vetrino, per rimuovere la fluorescenza aspecifica: per questo si utilizzano delle apposite soluzioni (Gene Expression Wash Buffer Kit) contenenti Triton, un tensioattivo, e sali a diverse concentrazioni per

rendere il lavaggio sempre più stringente (per l'ultimo passaggio si usa inoltre una soluzione preriscaldata a 37°C).

A questo punto il vetrino è pronto per la scansione, che viene effettuata sullo Scanner Agilent G2565 (Agilent Technologies): il sistema è in grado di rilevare la fluorescenza emessa dal vetrino dopo l'eccitazione da parte di un laser a determinate lunghezze d'onda (550 nm quella del laser di eccitazione, 570 nm quella di emissione della cianina impiegata in questo studio). Con la scansione, che può essere eseguita a risoluzioni diverse e può arrivare fino a 2 µm, si ottiene un'immagine relativa alla fluorescenza emessa su tutto il vetrino, la quale dovrà essere successivamente analizzata con opportuni software per estrarre l'intensità di emissione di ciascuno spot. A questo scopo si utilizza il Feature Extraction, un programma che permette di estrarre i dati di intensità grezza di ciascuno spot i quali, unitamente ai valori di background, saranno raccolti in un file prodotto come output dal software e pronto per l'analisi statistica.

2.7. Analisi statistica dei microarray

Come spesso accade con le tecnologie *high throughput*, la fase di analisi *post-processing* dei dati sperimentali risulta essere di fondamentale importanza. Proprio per questo motivo, unitamente all'esistenza di una collaborazione già in atto, per le analisi dei microarray abbiamo chiesto il supporto del gruppo di bioingegneri del Prof. Cobelli, che da anni lavora su tematiche di questo tipo e possiede quindi competenze notevoli in campo biomedico e biostatistico. Di seguito saranno presentati brevemente i metodi utilizzati per tale analisi, i quali possono essere suddivisi in 2 sezioni principali:

- 1) Il *data processing*, che consiste principalmente nella valutazione della qualità dei dati, nella sottrazione del background, nella normalizzazione e nella ricerca dei valori estremi, detti *outlier*.
- 2) Il *data analysis*, che consente l'individuazione di espressione genica differenziale e, se richieste, l'analisi di *clustering* e l'analisi funzionale grazie ai nuovi approcci della *gene ontology*.

Nel nostro caso il *processing* dei dati è in parte automatizzato grazie al software Feature Extraction, il quale, per estrarre i dati grezzi di intensità del segnale e del background, effettua automaticamente il *gridding* (ossia l'identificazione della posizione di ogni singolo spot), la *segmentation* (definisce i pixel appartenenti al foreground e quelli appartenenti al background) e, per l'appunto, l'estrazione dell'intensità del segnale (calcolata come media o mediana dei singoli pixel), oltre a fornire anche un resoconto sulla qualità del vetrino.

Dall'analisi di tali resoconti, chiamati *QC report*, è stata valutata tutta una serie di caratteristiche che vengono adoperate per determinare se il vetrino in questione possa essere considerato di buona qualità, e quindi adatto per il proseguimento delle analisi, oppure di scarsa qualità, nel qual

caso viene scartato. In particolare si valuta la distribuzione spaziale degli *outlier* (in un vetrino buono essa dovrebbe essere uniforme e non presentare addensamenti), la presenza di eventuali gradienti di intensità del segnale (grazie all’algoritmo di *detrending* RMS_Resid), lo stato dei controlli negativi (ampie variazioni del segnale potrebbero indicare basse prestazioni dell’algoritmo di *gridding* o una forte componente di rumore), la qualità delle repliche intra-vetrino (si valuta il coefficiente di variazione mediano del segnale processato, ossia ripulito dal background) e la riproducibilità sulle eventuali repliche.

Ulteriori analisi di controllo della qualità sono state effettuate indipendentemente dai *QC report* forniti da Feature Extraction, e si riferiscono direttamente ai dati di intensità estratti dal vetrino: esistono diversi tipi di segnale che vengono attribuiti ad ogni spot, ma per le analisi che seguono sono stati considerati esclusivamente i seguenti:

- Il *Mean Signal*: è il segnale “grezzo” dello spot ricavato dalle intensità dei pixel “*inlier*”;
- Il *Processed Signal*: è il segnale ottenuto dall’algoritmo di Feature Extraction ripulito dal segnale di background (considerando, qualora selezionato, il *detrending*);
- Il *Total Probe Signal*: è la media dei processed signal di ogni probe replicata;
- Il *Total Gene Signal*: è la somma dei total probe signal per ogni gene.

Dato che questi segnali possono ottenere anche valori negativi è stato loro aggiunto un valore costante, chiamato offset, e sono stati trasformati su scala logaritmica applicando il logaritmo in base 2. Con tali dati sono stati costruiti dei *boxplot* e dei *density plot*, per valutare che la distribuzione delle intensità fosse omogenea in tutti gli array analizzati e presentasse medesima forma e range di valori (fig. 11).

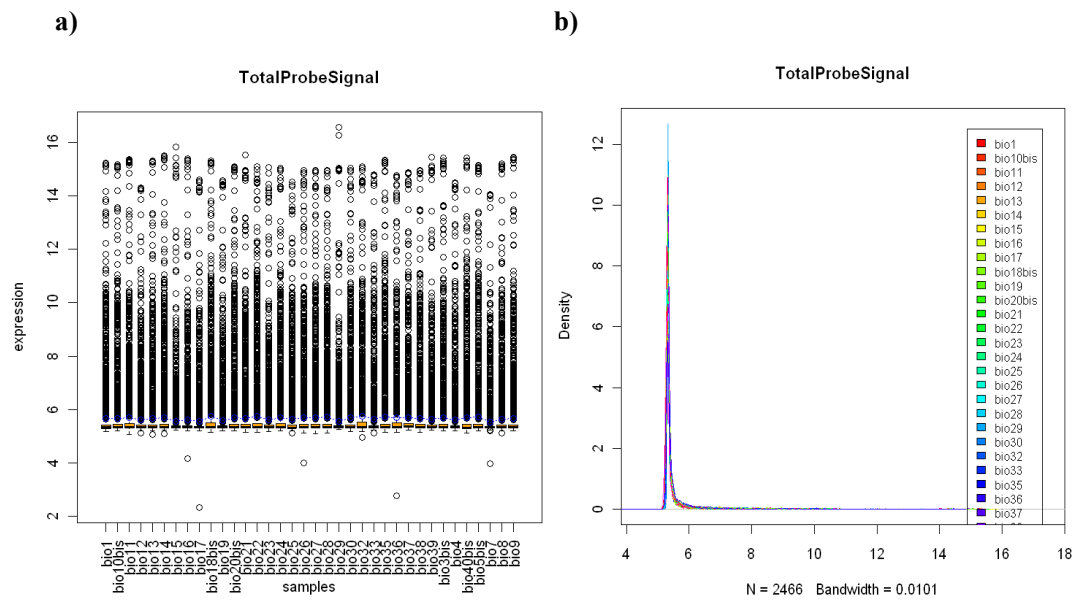


Fig. 11: Rappresentazione di un boxplot (a) e di un density plot (b) della distribuzione delle intensità tra gli array.

È stata determinata, inoltre, anche l'intensità relativa di ogni spot dell'array rispetto alla mediana dei valori del medesimo spot su tutti gli array analizzati (*Relative Log Expression*): se si assume che ci sia un'alta percentuale di geni non differenzialmente espressi, la maggior parte di questi valori dovrebbe essere prossima allo zero (fig. 12). Infine, per i segnali considerati sono stati creati gli MvA plot, i quali consentono di visualizzare, per ogni array del dataset, la differenza tra i valori di intensità degli spot nell'array osservato e il corrispondente valore mediano calcolato su tutti gli array (in ordinata) contro la media tra i valori dell'array e il valore mediano (in ascissa). Se si assume che la maggior parte degli spot corrispondano a geni non differenzialmente espressi, non dovrebbero risultare variazioni evidenti tra l'array osservato e l'array "mediano", e quindi la nuvola di punti dovrebbe essere attorno a $M=0$. (fig. 12).

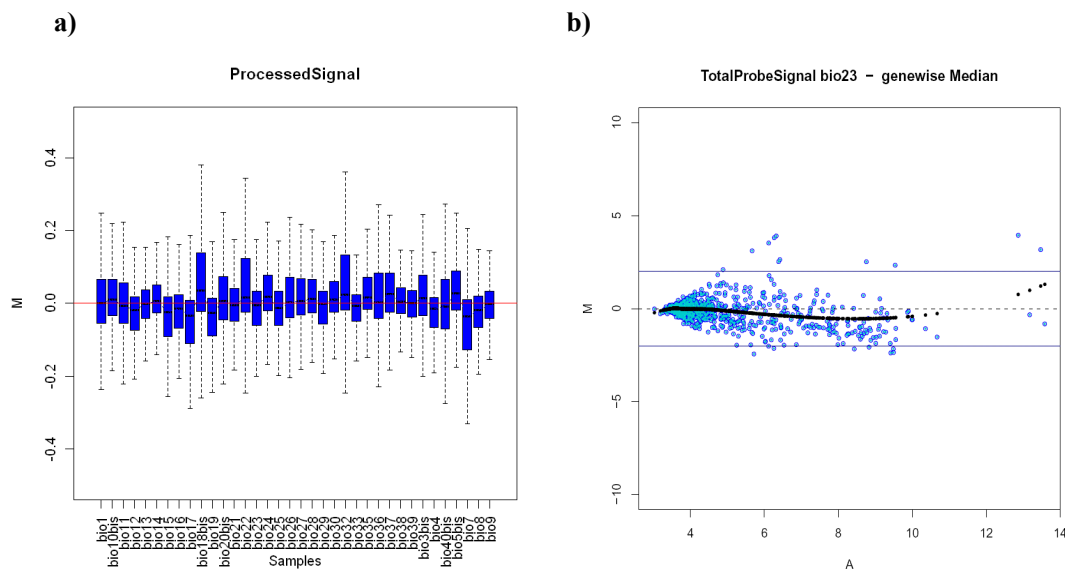


Fig. 12: Esempi di plot di Relative Log Expression (a) e MvA plot (b), utilizzati rispettivamente per valutare l'omogeneità di spot identici tra array diversi e la riproducibilità degli stessi rispetto ad un comportamento ideale.

Per la normalizzazione dei dati, necessaria in quanto si dovevano confrontare vetrini diversi ibridati in momenti successivi, è stata utilizzata la funzione loess. Questo metodo di normalizzazione si basa sui valori calcolati per gli MvA plot (media e differenza), e permette di ottenere una curva che approssimi al meglio la non linearità legata ad eventuali trend presenti nel plot. Applicando l'inversa di tale curva ai valori di intensità si può così compensare l'effetto della distorsione. Tale funzione ha lo scopo di rendere simili le distribuzioni dell'intensità degli spot tra array diversi e si basa sull'ipotesi che la maggior parte dei geni non dovrebbe avere intensità diverse in un array rispetto ad un altro.

In base all'esito di questa fase di pre-processing dei dati alcuni array sono stati classificati di bassa qualità, e sono stati quindi scartati dalle analisi successive. In particolare, i campioni che non hanno avuto buon esito sono bio2, bio6, bio11, bio15, bio29, bio31 e bio34, per i quali sarà probabilmente possibile effettuare degli ulteriori tentativi affinché in futuro possano essere inseriti anch'essi nei risultati di questo studio.

Per i rimanenti, invece, si è proceduto con la fase di *data analysis*, per la quale sono stati considerati soltanto gli spot che presentavano un segnale confrontabile per tutti i differenti gruppi patologici studiati. Per identificare i miRNA con espressione differenziale nei vari gruppi è stato impiegato il metodo di analisi della varianza ANOVA (per il confronto di due o più gruppi di campioni). L'analisi della varianza è un insieme di tecniche statistiche facenti parte della statistica inferenziale che permettono di confrontare due o più gruppi di dati confrontando la variabilità interna a questi gruppi con la variabilità tra i gruppi. L'ipotesi nulla, solitamente, prevede che i dati di tutti i gruppi abbiano la stessa origine, ovvero la stessa distribuzione stocastica, e che le differenze osservate tra i gruppi siano dovute solo al caso. L'analisi della varianza si basa sul modello additivo, il quale ipotizza che ogni dato sia composto da una somma di fattori. Se è valida l'ipotesi nulla, cioè se non c'è alcuna differenza reale tra i gruppi considerati, e le differenze nelle loro medie sono dovute solo al caso, dovremmo trovare la stessa variabilità sia all'interno dei gruppi che tra un gruppo e l'altro. Per i dati analizzati si è notato che i gruppi non sono indipendenti tra di loro e che alcuni geni possono essere maggiormente significativi considerando la combinazione di due o più gruppi.

In un primo momento, per la selezione dei miRNA statisticamente significativi si è adottato un P-value < 0.05, il quale è stato successivamente modificato applicando il metodo di correzione della *False Discovery Rate* (FDR) (Storey, 2002), grazie al quale è stata incrementata la stringenza della selezione. Per la FDR si è applicato un valore soglia di 0.05, corrispondente al valore di significatività $\alpha=0.0152$.

2.8. Predizione bioinformatica dei geni target dei miRNA

Al fine di conferire un significato biologico al comportamento alterato di alcuni dei miRNA studiati, si è proceduto con una ricerca bioinformatica dei geni target ad essi associati, in un primo momento utilizzando dei software già esistenti, tra cui Miranda, RNAhybrid e Pita (i cui principi di funzionamento sono già stati trattati nell'introduzione), e in un secondo tempo implementando direttamente degli script in Perl, un linguaggio di programmazione ad alto livello, ottimizzato per la manipolazione di file di testo. Lo scopo degli *script* implementati è quello di analizzare i risultati dei 3 software precedentemente citati, confrontandoli tra loro e andando alla ricerca di un *consensus* che possa incrementarne la validità, dato che nessuno di essi preso singolarmente è in grado di fornire delle predizioni sufficientemente affidabili. Inoltre è stato sviluppato un sistema


```

LOCUS       XM_939316                336 bp    mRNA    linear    PRI 29-FEB-2008
DEFINITION PREDICTED: Homo sapiens hCG1820764 (LOC645100), mRNA.
ACCESSION  XM_939316
VERSION    XM_939316.1  GI:89036235
KEYWORDS
SOURCE     *
           Homo sapiens (human)
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
           Catarrhini; Hominidae; Homo.
COMMENT    MODEL REFSEQ; This record is predicted by automated computational
           analysis. This record is derived from a genomic sequence
           (NW_925395) annotated using gene prediction method: GNOMON.
           Also see:
           Documentation of NCBI's Annotation Process

FEATURES             Location/Qualifiers
     source           1..336
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
     gene            1..336
                     /gene="LOC645100"
                     /note="Derived by automated computational analysis using
                     gene prediction method: GNOMON. Supporting evidence
                     includes similarity to: 2 Proteins"
                     /db_xref="GeneID:645100"
     CDS             1..336
                     /gene="LOC645100"
                     /codon_start=1
                     /product="similar to hCG1820764"
                     /protein_id="XP_944409.1"
                     /db_xref="GI:89036236"
                     /db_xref="GeneID:645100"
                     /translation="METAGSGTGQPASLLEAPGSTDDRLFLVKGIGFLGTVAAGTLA
                     GFITTLSLAKKKSPWFNKGSMVTTALPESGSSLALRALGCGSLYAMCGVGVISFSVM
                     KALGVHSDM"
     misc_feature    1..327
                     /gene="LOC645100"
                     /note="Protein of unknown function (DUF1358); Region:
                     DUF1358; pfam07096"
                     /db_xref="CDD:115731"
     variation       24
                     /gene="LOC645100"
                     /replace="g"
                     /replace="t"
                     /db_xref="dbSNP:12426017"
     variation       27
                     /gene="LOC645100"
                     /replace="g"
                     /replace="t"
                     /db_xref="dbSNP:11174744"
     variation       295
                     /gene="LOC645100"
                     /replace="g"
                     /replace="t"
                     /db_xref="dbSNP:11174745"
ORIGIN
1 atggagacag cgggctctgg tactgggcag ccggcctctc tgctggaggc tcccgggtcc
61 acggatgacc ggcttttctt ggttaaaggt ggaattttcc ttggtactgt tgctgcagca
121 ggaacgctag ctggatttat tacaacatta tcattggcta aaaagaaaag ccttgaatgg
181 ttcaataaag gaagtatggt cactgactgca ttaccggaaa gcgggtcttc ccttgctctg
241 cgagctctgg gctgcggctc actgtatgca tgggtgtggg ttggtgtgat tagcttctca
301 gtctgaaaag ctttaggggt tcacagtgac atgtaa
//

```

Fig. 14: Esempio di file genbank per un trascritto.

Una volta ottenuto il *dataset* su cui lavorare, è necessario effettuare le predizioni dei target con Miranda, RNAhybrid e Pita per ogni miRNA che si vuole studiare: i 3 software sono stati

installati in locale sui server a nostra disposizione, in modo da rendere le operazioni più rapide rispetto all'utilizzo dei web server che sono messi a disposizione per ciascuno di essi. In questa fase le predizioni vengono lanciate con parametri poco stringenti, in quanto un massiccio filtraggio dei dati sarà effettuato nei passaggi successivi.

```

=====
Performing Scan: miR-UL70-5p vs NM_007294
=====

Forward:   Score: 127,000000  Q:2 to 19  R:830 to 850 Align Len (17) (64,71%) (82,35%)

Query:    3' aGACCUUGCUCGGCUCUGcgu 5'
          |||||::||  :|||||
Ref:      5' cCTGGGTGACAGTGAGAGctgt 3'

Energy:   -22,500000 kCal/Mol

Scores for this hit:
>miR-UL70-5p  NM_007294      127,00 -22,50 0,00  2 19  830 850 17      64,71% 82,35%

Forward:   Score: 108,000000  Q:3 to 19  R:525 to 549 Align Len (21) (57,14%) (71,43%)

Query:    3' agACCUGCUC---GGCUCUGcgu 5'
          |||: ||||  |: |||:|
Ref:      5' atTGGG-GAGGAAATTCGAGGCagg 3'

Energy:   -21,290001 kCal/Mol

```

Fig. 15: Esempio di un tipico output di Miranda di una predizione eseguita per il miRNA hcmv-miR-UL70-5p sulla banca dati di trascritti umani (porzione del file).

```

target: NM_001136041(HDAC11)
length: 999
miRNA : hcmv-miR-UL22A
length: 20

mfe: -25,7 kcal/mol
p-value: 0,347538

position 66
target 5' C      GU  GG  G  3'
          CUCAUGGG  GGU  AG
          GAGUGCCC  CCG  UC
miRNA  3' A      UU  A  AAU 5'

target: NM_002819(polypyrimidine)
length: 999
miRNA : hcmv-miR-UL22A
length: 20

mfe: -22,9 kcal/mol
p-value: 0,815819

position 377
target 5' G      UG  CCCAA  G  GG  G  3'
          UC  C  GGG  GGG  UGG  G
          AG  G  CCCU  UCC  AUC  U
miRNA  3'  A  UG      G  AA  5'

```

Fig 16: Esempio di un tipico output di RNAhybrid di una predizione eseguita per il miRNA hcmv-miR-UL22A sulla banca dati di trascritti umani (porzione del file).

RefSeq	microRNA	Sites	Score	
NM_001103150	hcmv-miR-UL148D	13	-32,75	
NM_001103151	hcmv-miR-UL148D	13	-32,75	
NM_052926	hcmv-miR-UL148D	13	-32,75	
NM_172239	hcmv-miR-UL70-3p	11	-32,29	
NM_001136154	hcmv-miR-UL148D	5	-31,44	
NM_001136155	hcmv-miR-UL148D	5	-31,44	
NM_004449	hcmv-miR-UL148D	5	-31,44	
NM_182918	hcmv-miR-UL148D	5	-31,44	
NM_001159286	hcmv-miR-UL70-3p	8	-30,41	
NM_014971	hcmv-miR-US4	4	-30,66	
NM_020709	hcmv-miR-UL148D	11	-30,41	
NM_032019	hcmv-miR-UL70-3p	8	-30,41	
NM_000774	hcmv-miR-UL70-3p	3	-29,51	
NM_001005476	hcmv-miR-UL70-3p	1	-29,94	
NM_001112741	hcmv-miR-UL148D	3	-29,51	
NM_003628	hcmv-miR-UL70-3p	1	-29,94	
NM_032189	hcmv-miR-UL70-3p	11	-29,07	
NM_000425	hcmv-miR-UL70-3p	17	-28,20	
NM_001001418	hcmv-miR-UL70-3p	4	-28,00	
NM_001024845	hcmv-miR-UL70-3p	9	-28,88	
NM_001099338	hcmv-miR-UL148D	4	-28,38	
NM_001143963	hcmv-miR-UL70-3p	17	-28,20	
NM_005650	hcmv-miR-UL70-3p	6	-28,93	
NM_006934	hcmv-miR-UL70-3p	9	-28,88	
NM_024003	hcmv-miR-UL70-3p	17	-28,20	
NM_032258	hcmv-miR-UL70-3p	4	-28,00	

Fig 17: Esempio di un tipico output di Pita di una predizione eseguita per miRNA diversi sulla banca dati di trascritti umani (porzione del file).

VariableThresholdPredictors.pl (A.2) è lo *script* implementato per la generazione delle soglie di energia di legame, in base alle quali filtra i risultati prodotti in precedenza: per quest'ultima funzione il programma necessita di 3 *script* accessori, chiamati rispettivamente *mirandaOutputFilterForVariableThreshold.pl*, *RNAhybridOutputFilterForVariableThreshold.pl* e *pitaOutputFilterForVariableThreshold.pl* (A.3, A.4 e A.5). Brevemente, nella *pipeline* di esecuzione di questo programma viene generata una tabella per ogni miRNA, nella quale ad ogni possibile gene target viene assegnato un valore, maggiore o uguale a zero, associato a ciascuno dei programmi Miranda, RNAhybrid e Pita, a seconda che tale gene sia stato predetto o meno come target e che abbia superato la soglia di cut-off. Le tabelle vengono poi analizzate alla ricerca dei geni che abbiano ottenuto un punteggio positivo da almeno 2 dei 3 programmi, e tale punteggio tiene conto anche di quanti siti di legame un miRNA può avere sul 3'-UTR di uno stesso gene. Il risultato finale sarà 4 liste di geni: la prima conterrà tutti quelli che sono stati predetti come target da tutti e tre i programmi di predizione utilizzati, e a seguire verranno mostrati i target predetti da almeno 2 programmi su 3. Inoltre, per ogni predizione, sarà possibile valutare la presenza di siti multipli di legame di un miRNA sullo stesso gene, fattore che sembra incrementare la validità delle predizioni.

3. RISULTATI

3.1. Parte I: Espressione dei microRNA nell'epatocarcinoma e nelle epatiti virali croniche

L'analisi dell'espressione dei miRNA è stata eseguita su una casistica comprendente 29 campioni di HCC associato ad infezione da virus dell'epatite B e/o C e del relativo tessuto epatico adiacente al tumore e 40 campioni di biopsia epatica prelevati da pazienti con diverso grado di epatite virale cronica B e C. Lo scopo della ricerca era studiare le modificazioni dell'espressione dei miRNA nelle varie fasi di sviluppo e progressione dell'HCC, al fine di ottenere una maggiore comprensione del ruolo dei miRNA in questo processo ancora ampiamente sconosciuto. L'analisi dell'espressione dei miRNA è stata eseguita sia mediante real-time RT-PCR su un pannello di miRNA, elencati in tabella 6, i quali, dai dati riportati nella letteratura scientifica, risultavano potenzialmente coinvolti nella patogenesi dell'HCC, sia su tutti i miRNA umani noti, utilizzando la metodica di analisi dei miRNA microarray.

Tab. 6: Elenco dei miRNA studiati con metodica di realtime PCR.

miRNA	SEQUENZA
hsa-miR-18a	uaaggugcaucuagugcagauag
hsa-miR-92-a	uauugcacuuguccggccugu
hsa-miR-122a	uggagugugacaauagguguuug
hsa-miR-145	guccaguuuuccaggaaucuccu
hsa-miR-195	uagcagcacagaaaauuuggc
hsa-miR-199a-5p	cccaguguucagacuaccuguuc
hsa-miR-199b-5p	cccaguuuuagacuauucuguuc
hsa-miR-222	agcuacaucuggcuacugggu
hsa-miR-302c	uaagugcuuccauguuucagugg

3.1.1. Espressione dei miRNA nell'HCC rispetto al tessuto non tumorale adiacente

L'analisi dell'espressione dei miRNA miR-18a, miR-92-a, miR-122a, miR-145, miR-195, miR-199a-5p, miR-199b-5p, miR-222, e miR-302c nei campioni di HCC è stata valutata mediante real-time RT-PCR. In questi campioni non è stato possibile eseguire l'analisi di espressione dei miRNA mediante microarray poiché la qualità dell'RNA purificato dai campioni non era adeguata. I valori di espressione dei miRNA nel campione di HCC rispetto al tessuto non tumorale adiacente sono stati calcolati con il metodo $2^{-\Delta\Delta Ct}$, normalizzando i valori di espressione rispetto al miRNA *housekeeping* RNU6B.

Nella tabella 7 sono riportati i valori medi dei valori di espressione dei miRNA ottenuti da 3 prove eseguite in duplicato. E' stato considerato, convenzionalmente, come valore soglia per

indicare un significativo aumento o diminuzione dell'espressione dei miRNA nel campione di HCC rispetto al tessuto adiacente il valore 2. In tabella 7 sono stati evidenziati con diversi colori i valori di espressione dei miRNA il cui rapporto HCC/tessuto adiacente (N) era >2 o $<0,5$.

Dall'esame della tabella 7 emerge come alcuni dei miRNA esaminati tendano, nei limiti della variabilità intercampionaria, ad avere una orientamento comune all'interno della casistica. In particolare, i miRNA miR-122, miR-195, miR-199a-5p, miR-199b-5p risultavano sottoespressi (ovvero rapporto HCC/N $<0,5$) nella maggior parte dei campioni di HCC, mentre il miR-222 era sovraespresso (ovvero rapporto HCC/N > 2) nel 50% dei campioni di HCC (figura 18). L'espressione del miR-302c non era rilevabile in molti campioni, per cui non è stato possibile calcolare l'espressione relativa.

campione	mir-122	mir-145	mir-18a	mir-195	mir-199a-5p	mir-199b-5p	mir-222	mir-92a	mir-302c
HCC0	0,417	0,588	12,891	0,526	0,762	0,371	2,186	1,041	ND
HCC1	3,728	1,845	3,135	1,013	0,091	0,155	2,033	2,141	ND
HCC2	0,567	0,790	1,925	0,651	3,019	0,671	3,758	1,440	192,672
HCC3	1,146	1,707	0,563	0,741	2,184	2,297	1,178	0,523	ND
HCC4	0,029	0,576	0,849	0,229	0,232	0,834	0,574	0,512	ND
HCC6	0,524	0,065	0,438	0,237	0,003	0,003	0,826	0,290	0,683
HCC7	0,031	0,169	0,455	0,024	0,023	0,025	0,407	0,139	19,472
HCC8	0,007	3,706	1,469	0,257	0,557	0,737	2,173	0,509	ND
HCC9	0,288	2,878	1,735	3,352	3,470	4,942	2,871	0,705	ND
HCC10	0,155	0,019	15,437	0,168	1,505	0,011	0,489	0,491	ND
HCC12	0,056	0,393	0,796	0,051	0,053	0,512	4,332	0,381	ND
HCC13	0,037	0,012	0,028	0,003	0,001	0,007	0,023	0,026	0,045
HCC14	2,321	0,801	1,079	0,801	0,557	0,727	0,763	0,620	2,471
HCC15	0,003	0,753	0,983	0,082	0,354	0,296	1,682	0,107	ND
HCC16	1,406	0,525	2,517	0,761	0,285	0,051	1,948	3,045	10,642
HCC17	0,104	0,566	1,628	0,170	0,106	1,741	8,959	0,731	ND
HCC18	0,124	1,161	3,375	0,230	0,290	0,131	5,716	4,228	11,353
HCC19	1,560	0,548	2,019	0,268	0,074	0,033	4,037	0,924	ND
HCC20	0,717	0,462	0,334	0,092	0,629	1,253	0,727	0,485	0,514
HCC21	0,152	0,486	0,834	0,103	0,024	0,768	0,958	0,792	19,607
HCC22	0,189	0,245	1,584	0,131	0,436	0,310	0,916	0,789	1,398
HCC23	1,847	3,719	3,745	3,204	6,964	8,846	6,255	1,759	ND
HCC31	1,939	0,527	0,676	ND	0,173	0,200	1,042	0,644	ND
HCC32	2,639	0,104	0,382	0,076	0,030	0,047	0,189	0,443	ND
HCC33	0,551	1,014	0,859	0,336	0,191	0,146	1,699	0,747	ND
HCC34	0,524	0,308	1,028	0,853	0,142	0,107	0,644	1,390	ND
HCC35	0,247	4,532	1,705	1,025	0,061	3,655	5,134	3,106	ND
HCC36	2,751	0,096	0,655	0,457	0,298	0,014	1,218	0,419	ND
HCC37	0,090	1,979	0,838	0,350	0,254	0,374	2,378	0,646	ND

HCC/N ≤ 0,5

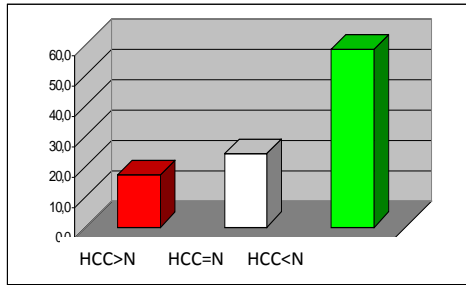
HCC/N ≥ 2

0,5 < HCC/N < 2

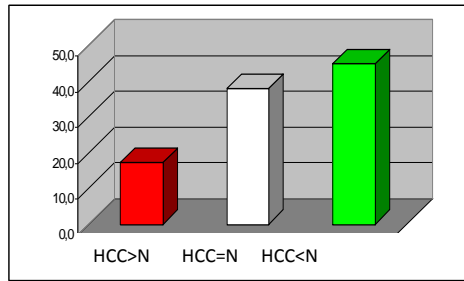
Non rilevabile

Tab. 7: Quadro riassuntivo dei risultati dell'analisi di espressione relativa dei miRNA studiati nei campioni raccolti. Ogni dato rappresenta il rapporto di espressione HCC/N calcolato con la metodica del $\Delta\Delta Ct$ ($K/N = 2^{-\Delta\Delta Ct}$) sulla media delle ripetizioni di ciascun campione. L'inclusione nelle categorie di espressione ha per soglia il rapporto 2 (o 0,5) con margine di confidenza del 5%.

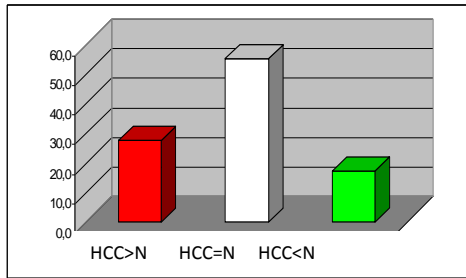
mir-122



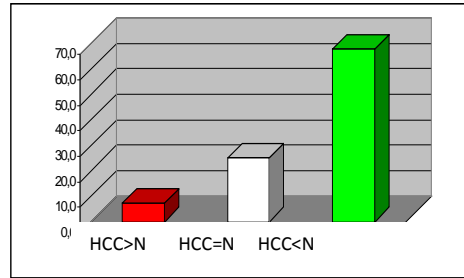
mir-145



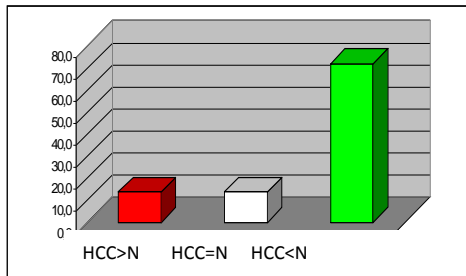
mir-18a



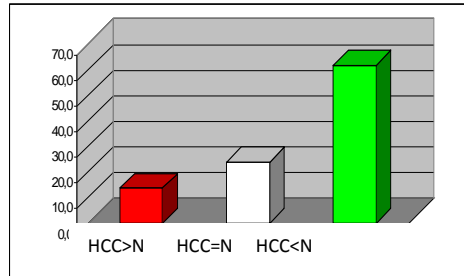
mir-195



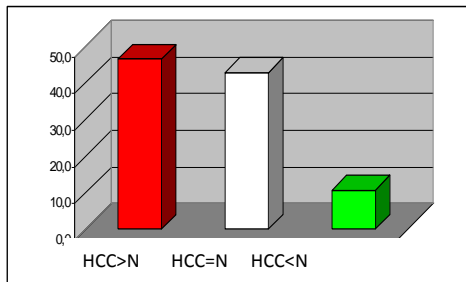
mir-199a-5p



mir-199b-5p



mir-222



mir-92-a

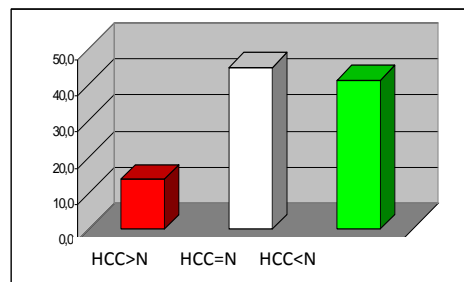


Fig 18 : Rappresentazione grafica della distribuzione dei campioni fra le tre categorie di espressione relativa per ciascuno dei miRNA studiati. Tutti i dati sono espressi come distribuzione percentuale.

Suddividendo i campioni di HCC sulla base dell'eziologia virale (Tabella 8), i valori di espressione del miR-122 sono risultati significativamente più bassi nei campioni HCV-negativi rispetto ai campioni HCV-positivi.

Tab. 8: Quadro riassuntivo dei risultati dell'analisi di espressione relativa dei miRNA nei campioni di HCC HCV-positivi e HCV-negativi.

HCC	miR-122	miR-145	miR-18a	miR-195	miR-199a	miR-199b	miR-222	miR-92a	miR-302c
<i>Campioni HCV-positivi</i>									
HCC0	0.42	0.59	12.89	0.53	0.76	0.37	2.19	1.04	ND
HCC1	3.73	1.85	3.14	1.01	0.09	0.16	2.03	2.14	ND
HCC2	0.57	0.79	1.93	0.65	3.02	0.67	3.76	1.44	192.67
HCC4	0.03	0.58	0.85	0.23	0.23	0.83	0.57	0.51	ND
HCC6	0.52	0.07	0.44	0.24	0.00	0.00	0.83	0.29	0.68
HCC7	0.03	0.17	0.46	0.02	0.02	0.03	0.41	0.14	19.47
HCC9	0.29	2.88	1.74	3.35	3.47	4.94	2.87	0.71	ND
HCC10	0.16	0.02	15.44	0.17	1.51	0.01	0.49	0.49	ND
HCC12	0.06	0.39	0.80	0.05	0.05	0.51	4.33	0.38	ND
HCC13	0.04	0.01	0.03	0.00	0.00	0.01	0.02	0.03	0.05
HCC14	2.32	0.80	1.08	0.80	0.56	0.73	0.76	0.62	2.47
HCC16	1.41	0.53	2.52	0.76	0.29	0.05	1.95	3.05	10.64
HCC18	0.12	1.16	3.38	0.23	0.29	0.13	5.72	4.23	11.35
HCC19	1.56	0.55	2.02	0.27	0.07	0.03	4.04	0.92	ND
HCC21	0.15	0.49	0.83	0.10	0.02	0.77	0.96	0.79	19.61
HCC22	0.19	0.25	1.58	0.13	0.44	0.31	0.92	0.79	1.40
HCC23	1.85	3.72	3.75	3.20	6.96	8.85	6.26	1.76	ND
HCC31	1.94	0.53	0.68	ND	0.17	0.20	1.04	0.64	ND
HCC32	2.64	0.10	0.38	0.08	0.03	0.05	0.19	0.44	ND
HCC36	2.75	0.10	0.66	0.46	0.30	0.01	1.22	0.42	ND
Media	1.04	0.78	2.73	0.65	0.91	0.93	2.03	1.04	ND
DS	1.14	0.98	4.07	0.97	1.73	2.15	1.87	1.05	ND
<i>Campioni HCV-negativi</i>									
HCC3	1.15	1.71	0.56	0.74	2.18	2.30	1.18	0.52	ND
HCC8	0.01	3.71	1.47	0.26	0.56	0.74	2.17	0.51	ND
HCC15	0.00	0.75	0.98	0.08	0.35	0.30	1.68	0.11	ND
HCC17	0.10	0.57	1.63	0.17	0.11	1.74	8.96	0.73	ND
HCC20	0.72	0.46	0.33	0.09	0.63	1.25	0.73	0.49	0.51
HCC33	0.55	1.01	0.86	0.34	0.19	0.15	1.70	0.75	ND
HCC34	0.52	0.31	1.03	0.85	0.14	0.11	0.64	1.39	ND
HCC35	0.25	4.53	1.71	1.03	0.06	3.66	5.13	3.11	ND
HCC37	0.09	1.98	0.84	0.35	0.25	0.37	2.38	0.65	ND
Media	0.38	1.67	1.05	0.43	0.50	1.18	2.73	0.92	ND
DS	0.39	1.51	0.47	0.35	0.66	1.20	2.69	0.89	ND
Test T*	0.03	0.13	0.08	0.41	0.36	0.70	0.49	0.74	ND

*Test *t* di Student a due code; HCV-positivi vs. HCV-negativi.

Considerando i valori soglia di espressione relativa dei miRNA nell'HCC rispetto al tessuto circostante (tabella 9), il miR-145 risultava più frequentemente sottoespresso negli HCC HCV-positivi (55% dei casi) rispetto ai campioni di HCC HCV-negativi (22% dei casi), mentre il miR-18a era sovraespresso nel 40% degli HCC HCV-positivo ma in nessun HCC HCV-negativo.

Tab. 9 : Tabella riassuntiva dei risultati dell'analisi di espressione relativa dei miRNA.

miRNA	all samples	HCV+	HCV-
miR-199a-5p	70%	70%	67%
miR-199b-5p	62%	70%	45%
miR-195	67%	68%	67%
miR-122a	58%	55%	67%
miR-145	45%	55%	22%
miR-92a	41%	40%	45%
miR-18a	27%	40%	0%
miR-222	45%	45%	44%

Ogni dato rappresenta la percentuale di campioni appartenenti alla categoria di espressione relativa contraddistinta dal colore (verde, $HCC/N \leq 0,5$; rosso, $HCC/N \geq 2$) nei gruppi di HCC HCV-positivi (HCV+) e HCV-negativi (HCV-).

Di seguito è riportata un'analisi dettagliata dei valori di espressione relativa (cancro/tessuto circostante) di ciascun miRNA.

- miR-122 : questo miRNA è stato studiato molto approfonditamente dalla comunità scientifica, sia per la sua alterata espressione in alcune tipologie di cancro (in particolare HCC) (Murakami et al., 2006; Jiang et al., 2008; Li et al., 2008), sia per il suo possibile coinvolgimento nell'infezione di HCV (Jopling et al., 2005; Niepmann M, 2009). In effetti, nel nostro studio, mostra dati interessanti, in quanto risulta sottoespresso (espressione relativa $\leq 0,5$) in circa il 58% dei campioni analizzati. Questo dato è in accordo con quanto riportato in alcuni altri studi pubblicati in letteratura (Bai et al., 2009, Henke et al., 2008), e mostra un possibile coinvolgimento di tale miRNA nella patogenesi di HCC. Se possibile, ancora più interessante appare la differenza sostanziale e statisticamente significativa fra i due sottogruppi relativi all'infezione da HCV, che mostra come la sottoespressione sia molto più frequente fra i campioni HCV-negativi che fra i positivi. Tale dato sembra supportare l'ipotesi di una regolazione positiva di miR-122 sulla replicazione e traduzione di HCV dovuta alla diretta interazione del miRNA con il genoma virale, suggerita da diversi studi (Jopling et al., 2005; Jopling et al., 2006; Henke et al., 2008).

- miR-18a: questo miRNA, così come il prossimo, appartiene al cluster di miRNA con attività oncogena mir-17-92a. E' risultato essere sovraespresso in circa il 25% dei campioni, come

è spesso risultato anche in altri studi sul cancro (Liu et al., 2009; Takakura et al., 2008). Altro dato singolare è che tutti i casi di sovraespressione appartengono al gruppo HCV-positivo, lasciando ipotizzare un qualche nesso tra il virus e l'espressione di questo miRNA.

- miR-92a: altro miRNA spesso oggetto di studio per la sua attività di oncogene (insieme agli altri membri del cluster di mir-17) (van Haafte et al., 2010; Rinaldi et al., 2007; Diosdado et al., 2009). Nel nostro studio risulta però sottoespresso in circa il 40% degli HCC, in modo indipendente dalla presenza di HCV.

- miR-145: questo miRNA è stato indicato da diversi studi avere un ruolo di oncosoppressore, come nel caso del cancro colon-rettale (Sachdeva et al., 2010; Arndt et al., 2009; Slaby et al., 2007). I dati ottenuti nel nostro studio sono concordi con la letteratura, in quanto esso risulta sottoespresso in poco più del 40% dei casi; i casi di sottoespressione sono circa due volte più frequenti fra gli HCV-positivi che fra i negativi.

- miR-222: questo miRNA è spesso sovraespresso nel cancro (insieme all'omologo mir-221), comportandosi pertanto come un cosiddetto 'onco-mir' (Mercatelli et al., 2008; Visone et al., 2007). Effettivamente, anche nel nostro studio, miR-222 risulta sovraespresso in più del 45% dei casi, senza particolari differenze tra i due gruppi di pazienti con e senza infezione da HCV.

- miR-199a: risulta frequentemente sottoespresso nella nostra casistica, in quasi il 70% dei casi, suggerendo quindi che questo miRNA abbia un ruolo nella patogenesi dell'HCC. Non risultano invece differenze fra campioni HCV positivi e negativi.

- miR-199b: anch'esso risulta notevolmente de-regolato nel nostro studio, in quasi due terzi dei campioni. La sottoespressione è molto più frequente nei campioni HCV positivi rispetto a quelli negativi.

- miR-195: questo miRNA è risultato sottoespresso in circa il 65% dei campioni, senza differenze significative fra campioni HCV-positivi e negativi. Il dato è consistente con la letteratura e, recentemente, è stato dimostrato un coinvolgimento di questo miRNA nella regolazione del ciclo cellulare (in particolare nella transizione G1/S) in linee cellulari di HCC, proponendolo come un possibile target terapeutico (Xu et al., 2009).

3.1.2. Espressione del profilo globale di espressione dei miRNA nell'epatite virale cronica

Dagli esperimenti di ibridazione dell'RNA, estratto dai campioni bioptici, su microarray di miRNA è stato possibile valutare l'espressione di tutti i miRNA ad oggi noti. Dei 40 campioni esaminati, l'analisi di *pre-processing* dei risultati dei miRNA microarray ha dimostrato che 33 campioni erano di qualità adeguata per essere sottoposti alla successiva analisi statistica.

L'analisi dei dati è stata mirata all'identificazione di quei miRNA che presentassero una correlazione della loro espressione con i parametri clinici ed istopatologici, ovvero lo *staging*, il *grading*, l'eziologia virale della malattia, e la presenza di cirrosi, come riportato in Tabella 10.

Tab. 10: Definizione dei gruppi di campioni in base ai parametri clinico-patologici.

Parametri clinici	1° gruppo	Numerosità	2° gruppo	Numerosità
Staging	Stage 0-1-2	19	Stage 3-cirrosi	14
Grading	Grade 1-2	26	Grade 3-4	6
Eziologia virale	HCV	22	HBV	11
Fibrosi/cirrosi	Fibrosi	26	Cirrosi	7

Il metodo ANOVA di analisi della varianza ha permesso di identificare i miRNA con una espressione significativamente diversa nei campioni appartenenti ai vari gruppi, e le analisi sono state effettuate sia separatamente per ciascun parametro, sia confrontando più parametri contemporaneamente.

Si noti come per ciascun miRNA potevano essere presenti sul microarray probe diverse, ciascuna delle quali era replicata nel microarray fino ad un massimo di 16 volte. Tra le repliche è stata effettuata la media, mentre probe diverse relative allo stesso miRNA sono state considerate separatamente per aumentare la consistenza del risultato. Nelle tabelle che saranno presentate in seguito sono riportate le medie dei valori di espressione di ciascuna probe nei gruppi di campioni specificati in tabella 10.

3.1.2.1. Analisi in base all'eziologia virale

Dal confronto tra i campioni HBV-positivi e HCV-positivi sono emersi numerosi miRNA aventi espressione differenziale nei due gruppi: in particolare 34 miRNA sono risultati avere un'espressione più elevata nel gruppo di epatiti da HCV, mentre 17 hanno mostrato una maggiore espressione nelle epatiti da HBV.

Tra di essi, hsa-miR-211 ha mostrato un comportamento concorde (ossia una maggiore espressione nei campioni HBV-positivi) con quello riportato nel lavoro di Ura et al., 2009, relativo a campioni di HCC. Da questo dato sembra plausibile una sua implicazione in una delle vie patogenetiche seguite dai due virus, anche se le predizioni sui possibili geni target non hanno fornito indicazioni in merito. Presentano invece una maggiore espressione nei campioni HCV-positivi i miRNA hsa-miR-17, hsa-miR-17* e hsa-miR-18a, tutti appartenenti al cluster miR-17-92, e hsa-miR-21. È interessante notare come questi ultimi siano stati trovati sovraespressi in molti tipi di cancro (Volinia et al., 2006), tra cui anche l'HCC associato all'infezione da HBV

(Connolly et al., 2008): questo dato lascia ipotizzare, a nostro parere, un maggiore coinvolgimento dei miRNA del cluster miR-17-92 nell'epatite cronica indotta da HCV.

Un altro dato interessante è la maggiore espressione di miR-548b-3p nei campioni con eziologia da HBV, in quanto tra i suoi potenziali target figura un gene implicato specificamente nella risposta cellulare all'infezione da HBV, come avremo modo di vedere nella sezione riguardante i risultati bioinformatici.

Nella tabella 11 è riportato l'elenco completo dei miRNA la cui espressione è risultata significativamente diversa nei campioni HCV-positivi rispetto a quelli HBV positivi.

Tab. 11: Elenco dei miRNA con espressione significativamente differente nei due gruppi di campioni, divisi in base all'eziologia virale (HBV vs HCV). In grigio è evidenziato il gruppo che presenta un'espressione più elevata.

miRNA	Probe ID	p-value	HCV* (n=22)	HBV* (n=11)
hsa-let-7i	A_25_P00012145	0.0041	7.5368	7.0501
hsa-let-7i	A_25_P00012146	0.0107	7.8163	7.4095
hsa-miR-106b	A_25_P00010433	5.07E-06	7.4114	6.9037
hsa-miR-106b	A_25_P00010434	0.0048	6.5876	6.2381
hsa-miR-107	A_25_P00011069	0.0005	8.9756	8.6265
hsa-miR-125b	A_25_P00010980	0.0136	8.2801	8.6400
hsa-miR-140-3p	A_25_P00012177	0.0060	5.7237	5.2869
hsa-miR-140-5p	A_25_P00010607	0.0079	5.3503	5.0625
hsa-miR-142-3p	A_25_P00011016	0.0056	9.9416	8.9613
hsa-miR-142-3p	A_25_P00013937	0.0082	9.1297	8.1965
hsa-miR-150	A_25_P00014846	0.0030	6.4118	5.3937
hsa-miR-150	A_25_P00014847	0.0046	7.2926	6.1778
hsa-miR-150	A_25_P00010490	0.0051	5.3379	4.4900
hsa-miR-150*	A_25_P00013452	0.0145	2.8046	3.1650
hsa-miR-17	A_25_P00013841	0.0015	6.5646	6.3225
hsa-miR-17	A_25_P00014819	0.0059	7.2129	7.0222
hsa-miR-17*	A_25_P00013151	0.0019	5.0765	4.6151
hsa-miR-185	A_25_P00012239	0.0017	5.0821	4.6677
hsa-miR-186	A_25_P00012243	0.0017	5.0433	4.7708
hsa-miR-18a	A_25_P00011994	0.0007	3.9849	3.3530
hsa-miR-18a	A_25_P00011993	0.0115	3.2718	2.6903
hsa-miR-202*	A_25_P00013588	4.41E-05	0.8427	2.0020
hsa-miR-21	A_25_P00010975	0.0040	11.7909	11.2439
hsa-miR-211	A_25_P00013898	0.0065	1.0444	1.6728
hsa-miR-221*	A_25_P00013345	0.0045	1.0075	0.3787
hsa-miR-24-1*	A_25_P00014494	0.0064	3.4078	2.9947
hsa-miR-29b	A_25_P00010053	0.0013	9.4754	8.8786
hsa-miR-301a	A_25_P00010839	0.0018	3.2128	2.4299
hsa-miR-302a	A_25_P00010982	0.0120	0.8785	1.4351
hsa-miR-30c	A_25_P00010815	0.0148	6.6770	7.1762
hsa-miR-328	A_25_P00010263	0.0116	1.2145	1.9088

hsa-miR-331-3p	A_25_P00011117	0.0004	6.3348	5.8390
hsa-miR-331-3p	A_25_P00011118	0.0005	5.7809	5.3778
hsa-miR-361-3p	A_25_P00012305	0.0012	4.7569	4.0852
hsa-miR-378*	A_25_P00010152	0.0110	2.0974	2.8102
hsa-miR-500*	A_25_P00010502	0.0014	2.6109	2.1134
hsa-miR-517b	A_25_P00011941	0.0085	0.5621	0.9802
hsa-miR-532-3p	A_25_P00012691	0.0094	0.8199	1.2241
hsa-miR-548b-3p	A_25_P00010124	1.25E-06	0.3233	1.7546
hsa-miR-548b-3p	A_25_P00010122	0.0054	0.1566	0.4391
hsa-miR-548d-5p	A_25_P00012838	0.0151	0.5324	0.3695
hsa-miR-556-3p	A_25_P00012717	0.0116	0.3737	0.9157
hsa-miR-590-5p	A_25_P00010656	5.40E-05	3.6149	3.0597
hsa-miR-590-5p	A_25_P00010655	0.0002	3.9513	3.3504
hsa-miR-622	A_25_P00010807	0.0006	1.8063	2.6945
hsa-miR-624*	A_25_P00011054	0.0138	0.7597	1.2926
hsa-miR-640	A_25_P00010372	0.0052	0.5496	1.0765
hsa-miR-7	A_25_P00012078	7.65E-05	2.3711	1.6076
hsa-miR-7	A_25_P00012077	0.0003	1.9117	0.8060
hsa-miR-767-3p	A_25_P00011299	0.0115	0.8444	0.5506
hsa-miR-99a	A_25_P00010472	0.0056	7.0620	7.5690

*: valori medi di espressione; HCV: campioni di epatite HCV-positivi; HBV: campioni di epatite HBV-positivi.

3.1.2.2. Confronto tra i gruppi di campioni con fibrosi o cirrosi

Dividendo i campioni in fibrotici e cirrotici abbiamo osservato come 37 miRNA presentino un'espressione significativamente maggiore nelle cirrosi, mentre 19 sono più espressi nelle fibrosi. Un andamento di questo tipo, in cui la maggior parte dei miRNA differenzialmente espressi tendono ad incrementare il livello di espressione col progredire della patologia epatica, sembra essere in contrasto con quanto riportato in letteratura in relazione all'HCC. Nel cancro infatti, come osservato anche in questo studio, molti dei miRNA deregolati tendono ad una sottoespressione rispetto al tessuto su cui insorge il tumore, che spesso presenta uno stato di cirrosi avanzata; un simile andamento risulta essere di difficile interpretazione, ma lascia presupporre che il probabile drastico cambiamento degli effettori cellulari in gioco prima e dopo l'insorgenza della neoplasia possa portare, come causa o come conseguenza, ad una sostanziale modificazione dei pattern di espressione di molti miRNA. Un'altra spiegazione potrebbe essere la presenza nei campioni con epatite cronica di infiltrato infiammatorio: l'espressione dei miRNA in tali tessuti potrebbe essere quindi condizionata dalla presenza di un numero elevato di cellule del sistema immunitario, che contribuirebbero anch'esse alla variazione dei livelli di espressione. Tra i miRNA che presentano un simile andamento citiamo hsa-miR-199a, hsa-miR-199b e hsa-miR-145, tutti sottoespressi nei nostri campioni tumorali, ma presenti in quantità più elevata nella

cirrosi rispetto alle fibrosi. Al contrario, hsa-miR-92a mostra un comportamento lineare, ossia una costante diminuzione della sua espressione al progredire della malattia. Tra gli altri segnaliamo hsa-miR-221, che è stato spesso trovato sovraespresso in vari tipi di cancro tra cui anche l'HCC (Garofalo et al., 2009; Fornari et al., 2008) e presenta un andamento crescente con l'avanzare della patologia, e hsa-miR-15a, che al contrario tende a diminuire e spesso è sottoespresso in altri tipi di cancro come la leucemia linfatica cronica (Calin et al., 2002).

Tab. 12: Elenco dei miRNA con espressione significativamente differente nei due gruppi di campioni, divisi tra fibrotici e cirrotici (fibrosi vs. cirrosi). In grigio è evidenziato il gruppo che presenta un'espressione più elevata.

miRNA	probe ID	p_value	Fibrosi (n=26)	Cirrosi (n=7)
hsa-miR-10a*	A_25_P00013301	0.0026	0.2632	0.6815
hsa-miR-10b	A_25_P00012082	0.0090	2.6595	3.3459
hsa-miR-1228	A_25_P00014938	0.0130	3.3077	2.8164
hsa-miR-125a-5p	A_25_P00012209	0.0004	5.3065	5.7952
hsa-miR-125a-5p	A_25_P00012208	0.0051	3.6249	4.2246
hsa-miR-139-5p	A_25_P00012071	0.0005	4.8158	4.3185
hsa-miR-142-5p	A_25_P00014844	0.0114	6.4208	7.2879
hsa-miR-142-5p	A_25_P00012181	0.0148	6.4974	7.3409
hsa-miR-143	A_25_P00012183	0.0031	2.9384	4.1228
hsa-miR-143	A_25_P00012184	0.0089	6.5117	7.0134
hsa-miR-143*	A_25_P00013418	0.0018	0.9525	1.5686
hsa-miR-145	A_25_P00012192	0.0035	6.5858	7.1655
hsa-miR-150	A_25_P00010490	0.0069	4.8970	5.6431
hsa-miR-155*	A_25_P00013471	0.0133	0.5444	0.0742
hsa-miR-15a	A_25_P00014817	0.0108	8.5260	8.2166
hsa-miR-17	A_25_P00013841	0.0022	6.5566	6.2139
hsa-miR-17*	A_25_P00013151	0.0093	5.0144	4.5823
hsa-miR-181c	A_25_P00010288	0.0064	2.1791	3.0589
hsa-miR-181c	A_25_P00010287	0.0126	1.5103	2.2674
hsa-miR-192	A_25_P00010869	0.0026	9.4947	9.0294
hsa-miR-192	A_25_P00010868	0.0028	10.0193	9.4019
hsa-miR-193b	A_25_P00012512	0.0086	7.3258	6.8539
hsa-miR-199a-3p	A_25_P00014828	0.0003	8.4210	9.1197
hsa-miR-199a-3p	A_25_P00014829	0.0115	9.0609	9.6160
hsa-miR-199a-5p	A_25_P00010701	3.54E-05	6.9024	7.8245
hsa-miR-199a-5p	A_25_P00010700	0.0003	8.3876	9.0892
hsa-miR-199b-5p	A_25_P00010698	0.0016	2.9995	3.7360
hsa-miR-19b-2*	A_25_P00013166	0.0150	0.2271	0.6073
hsa-miR-200a	A_25_P00010208	0.0027	4.5194	5.9633
hsa-miR-200a	A_25_P00010209	0.0095	5.5566	6.7697
hsa-miR-200b	A_25_P00012138	0.0122	5.1865	6.2023
hsa-miR-200b	A_25_P00012139	0.0145	6.0850	7.1298
hsa-miR-206	A_25_P00010528	0.0140	3.9920	2.9311

hsa-miR-214*	A_25_P00013335	0.0003	1.8099	2.6101
hsa-miR-214*	A_25_P00014557	0.0027	2.5779	3.1785
hsa-miR-215	A_25_P00010927	0.0060	9.1950	8.6826
hsa-miR-215	A_25_P00010926	0.0064	9.1266	8.6076
hsa-miR-221	A_25_P00010689	0.0017	2.6919	3.6752
hsa-miR-221	A_25_P00010690	0.0056	4.3324	5.0307
hsa-miR-23a	A_25_P00010843	0.0047	8.1974	8.6525
hsa-miR-339-5p	A_25_P00012400	0.0118	0.3951	0.9337
hsa-miR-374a*	A_25_P00013525	0.0130	0.5658	1.0363
hsa-miR-429	A_25_P00014860	0.0014	3.2029	4.4782
hsa-miR-429	A_25_P00010273	0.0036	3.2581	4.5856
hsa-miR-483-3p	A_25_P00012462	0.0136	4.5995	4.2915
hsa-miR-497	A_25_P00010130	0.0059	5.9324	6.2975
hsa-miR-500*	A_25_P00010501	0.0085	0.5189	1.1726
hsa-miR-544	A_25_P00012701	0.0005	0.3505	0.7097
hsa-miR-548b-3p	A_25_P00010124	0.0064	0.9285	0.1597
hsa-miR-625*	A_25_P00013642	0.0061	1.3675	0.9727
hsa-miR-656	A_25_P00014346	0.0079	1.0623	0.4778
hsa-miR-875-3p	A_25_P00012958	0.0049	0.9292	1.7699
hsa-miR-875-5p	A_25_P00012954	0.0126	0.5864	1.1278
hsa-miR-877*	A_25_P00015000	0.0026	1.8587	1.5659
hsa-miR-877*	A_25_P00014999	0.0079	2.3140	2.0378
hsa-miR-92a	A_25_P00012031	0.0065	7.9753	7.4893

3.1.2.3. Analisi in base alla classificazione per *grading*

Il confronto tra il gruppo di campioni caratterizzati da un *grading* assente o lieve (grade 0-1-2) rispetto a quelli con *grading* moderato o severo (grade 3-4) ha portato all'identificazione di 23 miRNA differenzialmente espressi, dei quali 8 sono risultati essere espressi maggiormente a *grade* più bassi e 15 a *grade* più elevati.

Essendo il *grading* un parametro assegnato in base al livello di attività necro-infiammatoria presente nel campione, è probabile che almeno alcuni di questi miRNA siano implicati a diversi livelli in *pathway* coinvolte nella regolazione dell'infiammazione, o che possano avere un ruolo nel controllo del sistema immunitario innato e adattativo coinvolto nei meccanismi patogenetici delle epatiti. Infatti, nella risposta immunitaria innata, specifici miRNA possono essere regolati da stimoli infiammatori oppure possono agire come mediatori degli stimoli stessi (Schetter et al., 2009). Tra i miRNA da noi identificati in base a questo criterio di selezione, hsa-miR-223 e hsa-miR-224 sono già stati associati in letteratura all'insorgenza di processi infiammatori in altri organi (Moschos et al., 2007) e hanno mostrato una sovraespressione in seguito all'induzione di infiammazione con lipopolisaccaridi. Anche nei nostri risultati è evidente un andamento simile,

essendo essi espressi maggiormente nelle epatiti di grado moderato o severo. Altri miRNA che ci sembrano importanti da segnalare e che presentano un comportamento dello stesso tipo sono hsa-miR-202* e hsa-miR-548b-3p, in quanto dalle predizioni bioinformatiche sui possibili geni target sono emersi dei candidati interessanti come la fibrosina e il recettore dell'interleuchina-18, che saranno discussi più approfonditamente nella sezione dei risultati bioinformatici.

Tab. 13: Elenco dei miRNA con espressione significativamente differente nei due gruppi di campioni, divisi in base al *grading*. In grigio è evidenziato il gruppo che presenta un'espressione più elevata.

miRNA	probe ID	p value	grade 1-2 (n=26)	grade 3-4 (n=6)
hsa-miR-1227	A_25_P00014934	0,0135	1,6711	1,3063
hsa-miR-1227	A_25_P00014933	0,0136	1,9506	1,5606
hsa-miR-138-2*	A_25_P00013409	0,0099	1,4417	3,1686
hsa-miR-142-3p	A_25_P00011016	0,0094	9,4526	10,1318
hsa-miR-142-3p	A_25_P00013937	0,0094	8,6604	9,3411
hsa-miR-197	A_25_P00010836	0,0109	4,4141	3,5750
hsa-miR-202*	A_25_P00013588	0,0001	1,0262	2,3468
hsa-miR-223	A_25_P00012130	0,0130	6,5931	7,0647
hsa-miR-224	A_25_P00012133	0,0014	2,6358	3,9915
hsa-miR-224	A_25_P00012134	0,0109	3,0018	4,3016
hsa-miR-302a*	A_25_P00013493	0,0047	0,3274	1,7322
hsa-miR-302b*	A_25_P00013513	0,0018	0,3013	0,6916
hsa-miR-373	A_25_P00010156	0,0123	3,2023	2,1754
hsa-miR-383	A_25_P00010384	0,0036	2,9064	1,4251
hsa-miR-383	A_25_P00010383	0,0142	2,7524	1,6237
hsa-miR-450b-5p	A_25_P00012910	0,0002	0,5270	1,3125
hsa-miR-452	A_25_P00012451	0,0061	1,5604	2,2015
hsa-miR-520e	A_25_P00010561	0,0122	0,5429	1,1287
hsa-miR-548b-3p	A_25_P00010124	0,0000	0,3924	2,9788
hsa-miR-557	A_25_P00010345	0,0066	3,7440	2,5398
hsa-miR-647	A_25_P00011371	0,0066	0,2081	0,4712
hsa-miR-769-3p	A_25_P00011230	0,0080	1,6697	0,6155
hsa-miR-877*	A_25_P00015000	0,0001	1,7137	2,1626

3.1.2.4. Analisi in base alla classificazione per *staging*

Un ulteriore confronto è stato effettuato suddividendo i campioni in base alla stadiazione: anche in questo caso abbiamo individuato dei miRNA con espressione significativamente diversa nei due gruppi considerati, in particolare 16 maggiormente espressi allo *stage* 1-2 e 47 allo *stage* 3/cirrosi. Questo confronto è simile a quello illustrato precedentemente tra il gruppo di fibrosi e quello di cirrosi, ma mentre in precedenza si voleva caratterizzare una possibile netta distinzione tra i due stati patologici, ora si vuole valutare più specificamente il processo di progressione della malattia. I risultati sono in parte diversi. È il caso, ad esempio, di hsa-miR-21, il quale tende ad

aumentare la sua espressione con l'aggravarsi dello stage della fibrosi, ed è frequentemente riportato anche in letteratura come uno dei miRNA più spesso sovraespressi in vari tipi di tumore, incluso l'HCC (Meng et al., 2007). Un altro miRNA significativo è hsa-miR-34a, che è stato dimostrato essere indotto direttamente da p53: di conseguenza, la sua espressione risulta, per la grande maggioranza dei casi, fortemente diminuita in molte neoplasie. Dai nostri risultati emerge un suo incremento al progredire della fibrosi, che può essere interpretato come una risposta protettiva al danno cellulare dovuto al continuo susseguirsi di fasi di necrosi e rigenerazione; tale risposta viene però persa di pari passo alla funzionalità di p53, motivo per cui nei tumori lo si trova poco espresso.

Tab. 14: Analisi in base alla classificazione per staging. In grigio è evidenziato il gruppo che presenta un'espressione più elevata.

miRNA	probe ID	p_value	stage 1-2 (n=19)	stage 3-c (n=14)
hsa-miR-10a*	A_25_P00013301	0,0031	0,1564	0,6687
hsa-miR-1227	A_25_P00014934	0,0027	1,7809	1,3602
hsa-miR-1229	A_25_P00014940	0,0039	1,9728	1,4674
hsa-miR-1236	A_25_P00014958	0,0059	1,1955	0,8710
hsa-miR-125a-5p	A_25_P00012209	0,0115	5,2151	5,6748
hsa-miR-130a	A_25_P00010440	0,0108	7,5262	7,8005
hsa-miR-135a	A_25_P00010981	0,0056	0,9067	1,6332
hsa-miR-139-5p	A_25_P00012071	0,0010	4,9378	4,4015
hsa-miR-142-3p	A_25_P00011016	0,0022	9,2218	10,1483
hsa-miR-142-3p	A_25_P00013937	0,0034	8,4433	9,3281
hsa-miR-142-5p	A_25_P00014844	0,0071	6,1790	7,1825
hsa-miR-142-5p	A_25_P00012181	0,0080	6,2551	7,2480
hsa-miR-143*	A_25_P00013419	0,0108	1,1144	1,7640
hsa-miR-150	A_25_P00010490	0,0051	4,7113	5,5221
hsa-miR-18b*	A_25_P00013568	0,0011	1,1734	0,7762
hsa-miR-199a-3p	A_25_P00014828	0,0015	8,2556	8,9948
hsa-miR-199a-3p	A_25_P00014829	0,0105	8,9195	9,5303
hsa-miR-199a-5p	A_25_P00010700	0,0006	8,2271	8,9564
hsa-miR-199a-5p	A_25_P00010701	0,0021	6,7260	7,6029
hsa-miR-199b-5p	A_25_P00010698	0,0144	2,8713	3,5418
hsa-miR-200a	A_25_P00010208	0,0007	4,0889	5,8256
hsa-miR-200a	A_25_P00010209	0,0007	5,1609	6,7001
hsa-miR-200b	A_25_P00012138	0,0008	4,8449	6,1580
hsa-miR-200b	A_25_P00012139	0,0011	5,7567	7,0530
hsa-miR-21	A_25_P00010976	0,0019	10,8605	11,2845
hsa-miR-21	A_25_P00010975	0,0033	11,4247	11,8582
hsa-miR-211	A_25_P00010405	0,0055	1,2027	0,6459
hsa-miR-214*	A_25_P00013335	0,0014	1,6246	2,4974
hsa-miR-214*	A_25_P00014557	0,0014	2,4022	3,1167
hsa-miR-218	A_25_P00010211	0,0017	1,9654	2,7266
hsa-miR-218	A_25_P00010210	0,0018	2,2481	2,9758

hsa-miR-221	A_25_P00010689	0,0063	2,4552	3,5049
hsa-miR-221*	A_25_P00013345	0,0029	0,6645	0,9266
hsa-miR-222	A_25_P00012126	0,0019	3,6215	4,4774
hsa-miR-224	A_25_P00012133	0,0024	2,4160	3,6452
hsa-miR-26a-2*	A_25_P00013508	0,0028	0,5121	0,8066
hsa-miR-27a	A_25_P00010797	0,0116	7,2833	7,7068
hsa-miR-301a	A_25_P00010840	0,0070	1,7432	2,4455
hsa-miR-302a*	A_25_P00013493	0,0060	0,9615	0,0512
hsa-miR-32*	A_25_P00013219	0,0096	0,7512	1,2702
hsa-miR-328	A_25_P00010262	0,0022	2,5843	2,1508
hsa-miR-33b	A_25_P00012827	0,0011	1,0292	0,5765
hsa-miR-34a	A_25_P00012085	0,0092	7,4773	7,9775
hsa-miR-375	A_25_P00010986	0,0147	3,8787	3,4469
hsa-miR-378*	A_25_P00010152	0,0114	2,6244	1,9303
hsa-miR-423-3p	A_25_P00012422	0,0099	0,2160	0,4932
hsa-miR-429	A_25_P00014860	0,0002	2,8189	4,3617
hsa-miR-429	A_25_P00010273	0,0003	2,8338	4,4978
hsa-miR-448	A_25_P00010441	0,0132	0,3663	0,7294
hsa-miR-452	A_25_P00012451	0,0080	1,2901	2,0688
hsa-miR-483-3p	A_25_P00012462	0,0070	4,7393	4,2557
hsa-miR-489	A_25_P00010830	0,0099	1,1201	0,9278
hsa-miR-505	A_25_P00012653	0,0053	3,9786	3,5096
hsa-miR-544	A_25_P00012701	0,0096	0,4024	0,4521
hsa-miR-548a-3p	A_25_P00010747	0,0005	0,2195	0,5610
hsa-miR-548b-3p	A_25_P00010124	0,0010	1,1477	0,3663
hsa-miR-601	A_25_P00010641	0,0071	1,6406	2,1417
hsa-miR-622	A_25_P00010805	0,0103	2,3390	2,6168
hsa-miR-655	A_25_P00010021	0,0031	0,2450	0,7092
hsa-miR-875-3p	A_25_P00012958	0,0037	0,6796	1,8718
hsa-miR-921	A_25_P00013041	0,0061	0,9374	0,8024
hsa-miR-96	A_25_P00012035	0,0072	1,6541	2,6122
hsa-miR-96	A_25_P00012034	0,0076	1,3315	2,3654

3.1.2.5. L'espressione di alcuni miRNA è associata a più parametri di progressione dell'epatite virale cronica

Abbiamo riscontrato, inoltre, come alcuni miRNA correlassero allo stesso tempo con più parametri clinico-patologici, sia relativi all'eziologia virale che alla progressione dell'epatite cronica. Una simile associazione multipla tra variabili considerate in modo indipendente rafforza in modo significativo i risultati ottenuti da questa analisi, e i miRNA presentati in tabella 15 potrebbero rivelarsi i più interessanti sui quali svolgere ulteriori indagini e da utilizzare come marcatori del rischio di progressione della malattia. Tra di essi sono presenti hsa-miR-202* e hsa-miR-548b-3p, di cui abbiamo già parlato in precedenza, ma anche hsa-miR-186, già associato in letteratura alla regolazione della crescita cellulare e dell'apoptosi (Zhou et al., 2008).

Tab. 15: Elenco dei miRNA la cui espressione risulta significativamente associata a più parametri clinico-patologici.

miRNA	p value	parametro	fibrosi	cirrosi	s12	s3c	g12	g3c	HCV	HBV
miR-154*	0.0101	fibrcirr*grade	1.10	0.92	1.17	0.92	1.06	1.11	1.06	1.06
	0.0129	stage*grade	0.93	0.51	0.92	0.73	0.82	1.02	0.86	0.80
miR-186	0.0097	fibrcirr*etiology	4.13	4.24	4.11	4.21	4.11	4.35	4.22	4.02
	0.0092	grade*etiology	4.13	4.24	4.11	4.21	4.11	4.35	4.22	4.02
miR-200b*	0.0014	fibrcirr*etiology	0.79	0.97	0.76	0.89	0.93	0.57	0.80	1.15
	0.0041	grade*etiology	1.16	1.47	1.06	1.44	1.06	1.72	1.20	1.35
miR-202*	0.0002	fibrcirr*grade	1.36	0.65	1.46	0.90	1.03	2.35	0.84	2.00
	0.0000	grade*etiology	1.36	0.65	1.46	0.90	1.03	2.35	0.84	2.00
miR-206	0.0004	fibrcirr*etiology	3.99	2.93	4.12	3.22	3.94	2.91	3.61	4.00
	0.0061	fibrcirr*grade	3.99	2.93	4.12	3.22	3.94	2.91	3.61	4.00
miR-373	0.0014	fibrcirr*etiology	3.11	2.21	3.16	2.49	3.20	2.18	2.59	3.37
	0.0100	fibrcirr*grade	3.11	2.21	3.16	2.49	3.20	2.18	2.59	3.37
miR-382	0.0031	fibrcirr*etiology	3.77	2.37	3.98	2.60	3.62	2.70	3.34	3.44
	0.0100	fibrcirr*grade	3.77	2.37	3.98	2.60	3.62	2.70	3.34	3.44
miR-450b-5p	0.0000	grade*etiology	0.74	0.53	0.78	0.60	0.53	1.31	0.62	0.86
	0.0081	stage*grade	0.74	0.53	0.78	0.60	0.53	1.31	0.62	0.86
miR-548b-3p	0.0000	grade*etiology	0.93	0.16	1.15	0.37	0.39	2.98	0.32	1.75
	0.0000	stage*grade	0.93	0.16	1.15	0.37	0.39	2.98	0.32	1.75
miR-648	0.0046	fibrcirr*etiology	1.40	1.30	1.31	1.48	1.45	1.04	1.28	1.55
	0.0038	stage*grade	1.40	1.30	1.31	1.48	1.45	1.04	1.28	1.55
miR-877*	0.0107	fibrcirr*etiology	1.86	1.57	1.81	1.78	1.71	2.16	1.84	1.70
	0.0001	stage*grade	1.86	1.57	1.81	1.78	1.71	2.16	1.84	1.70

3.2. Parte II: Predizione dei target dei miRNA mediante analisi bioinformatica

La parte bioinformatica del lavoro di tesi ha portato al conseguimento di un duplice risultato: da un lato c'è lo sviluppo di un metodo per la predizione dei geni target dei miRNA il quale, pur basandosi su algoritmi già esistenti, introduce una componente di novità per quanto riguarda il calcolo dell'energia di legame tra miRNA e target da utilizzare come soglia nelle predizioni; dall'altro, invece, ci sono i risultati ottenuti grazie all'impiego del metodo sviluppato, ossia le predizioni sui possibili target dei miRNA la cui espressione era stata trovata significativamente modulata dalle analisi dei microarray sulle biopsie delle epatiti virali croniche.

Il metodo bioinformatico prodotto è composto da una serie di script implementati nel linguaggio di programmazione Perl (tutti gli algoritmi creati sono riportati in Appendice A). Per il calcolo delle soglie di energia specifiche per ogni miRNA il programma si basa sull'osservazione di una correlazione inversa esistente tra la percentuale di basi guanina e citosina (GC) presenti nella sequenza del miRNA e l'energia di legame tra miRNA e target: quanto più elevata risulta essere la percentuale di GC di una sequenza, tanto più bassa sarà l'energia di legame con la sua sequenza

complementare e, di conseguenza, tanto più stabile sarà l'appaiamento tra le due sequenze (fig.19).

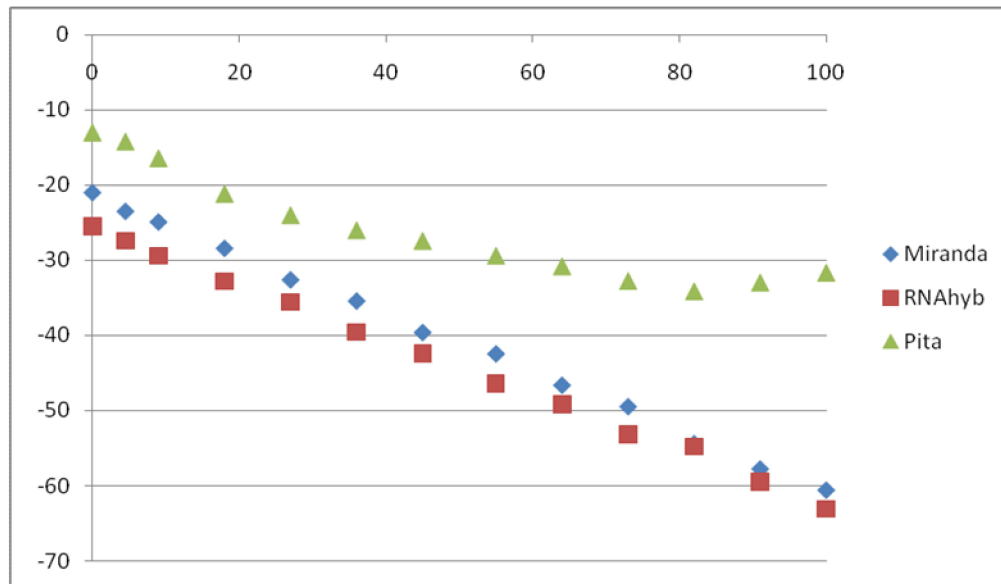


Fig.19: Correlazione lineare tra contenuto in GC ed energia minima di legame tra due sequenze nucleotidiche complementari. Pita sembra comportarsi in modo diverso per valori di GC molto elevati, ma raramente un miRNA ha un contenuto in GC che supera il 60%.

Per ogni miRNA viene quindi calcolata una soglia specifica per ognuno dei tre algoritmi di predizione su cui si basa il nostro software (Miranda, RNAhybrid e Pita) e proporzionale al contenuto in GC di ciascun miRNA. La pipeline del software, già trattata in dettaglio nella sezione “Materiali e metodi”, procede con la fase di predizione dei target, effettuata da ciascuno dei tre algoritmi indipendentemente, e da una successiva fase di analisi dei risultati che va alla ricerca dei target che siano stati predetti da almeno due dei tre programmi. L’output per ognuno dei miRNA analizzati sarà composto, quindi, solamente dai geni target che abbiano superato questa fase di filtraggio.

Come detto, quindi, tale metodo è stato applicato su alcuni dei miRNA la cui espressione era risultata significativamente correlata alla progressione o all’eziologia delle epatiti, oppure presentava sostanziali differenze nei tessuti tumorali rispetto ai tessuti non tumorali adiacenti. In particolare sono stati analizzati sette miRNA, e una selezione dei risultati delle predizioni è riportata in tabella 16.

Tab 16: Risultati delle predizioni dei geni target dei miRNA effettuate con il metodo bioinformatico prodotto appositamente per questo lavoro. I target già validati sperimentalmente da studi riportati in letteratura sono stati sottolineati.

miRNA	Geni target predetti
miR-7	NM_000345: synuclein alpha (SNCA), transcript variant 1, mRNA NM_001040056: mitogen-activated protein kinase 3 (MAPK3) NM_001102396: suppressor of IKK epsilon (SIKE), transcript variant 1 <u>NM_002880: v-raf-1 murine leukemia viral oncogene homolog 1 (RAF1)</u> <u>NM_005228: epidermal growth factor receptor (EGFR), transcript variant 1</u>
miR-145	NM_005419: signal transducer and activator of transcription 2, (STAT2) NM_000076: cyclin-dependent kinase inhibitor 1C (CDKN1C) <u>NM_002467: v-myc myelocytomatosis viral oncogene homolog (MYC)</u> NM_000633: B-cell CLL/lymphoma 2 (BCL2), transcript variant alpha NM_000875: insulin-like growth factor 1 receptor (IGF1R) NM_001258: cyclin-dependent kinase 3 (CDK3) <u>NM_024615: poly (ADP-ribose) polymerase family, member 8 (PARP8)</u> NM_001261: cyclin-dependent kinase 9 (CDK9)
miR-202*	NM_001105079: fibrosin (FBR3) NM_003336: ubiquitin-conjugating enzyme E2A (UBE2A), transcript variant 1 NM_015894: stathmin-like 3 (STMN3) NM_021205: ras homolog gene family, member U (RHOU)
miR-222	NM_005707: programmed cell death 7 (PDCD7) NM_198207: LAG1 homolog, ceramide synthase 1 (LASS1) <u>NM_004064: cyclin-dependent kinase inhibitor 1B (p27, Kip1)(CDKN1B)</u> <u>NM_001122742: estrogen receptor 1 (ESR1)</u>
miR-548b-3p	NM_005562: laminin, gamma 2 (LAMC2), transcript variant 1 NM_018414: ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1, 3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1 (ST6GALNAC1) NM_003855: interleukin 18 receptor 1 (IL18R1) NM_005385: natural killer-tumor recognition sequence (NKTR) NM_138927: SON DNA binding protein (SON), transcript variant f
miR-622	NM_005914: minichromosome maintenance complex component 4 (MCM4) NM_004368: calponin 2 (CNN2), transcript variant 1 polymerase (DNA directed), lambda (POLL) NM_022094 : Homo sapiens cell death-inducing DFFA-like effector c (CIDEC)
miR-625*	NM_006981: Homo sapiens nuclear receptor subfamily 4, group A, member 3 (NR4A3), transcript variant 1 NM_001001852: Homo sapiens pim-3 oncogene (PIM3) NM_001134477: Homo sapiens poly(A)-specific ribonuclease (deadenylation nuclease) (PARN) NM_138709: Homo sapiens DAB2 interacting protein (DAB2IP), transcript variant 2

Per quanto riguarda hsa-miR-7, il candidato target più interessante sembra essere senza dubbio SIKE, una proteina che interagisce con altri effettori cellulari come IKK-epsilon e TBK1 e opera una soppressione di TLR3 e delle vie di attivazione della risposta interferonica scatenata dall'infezione virale (Huang et al., 2005). Questo miRNA ha mostrato un'espressione maggiore nelle epatiti correlate all'infezione da HCV, e potrebbe essere un meccanismo di reazione

cellulare per aumentare la risposta antivirale: infatti esso andrebbe ad inibire SIKE, col risultato di una maggiore attivazione della via dell'interferone.

Il miRNA hsa-miR-145, che risultava più espresso nei campioni cirrotici rispetto ai fibrotici ma che va incontro ad una forte diminuzione di espressione nel HCC, presenta tra i suoi possibili target due distinte chinasi ciclina-dipendenti. Tra i target validati per questo miRNA è presente il gene MYC, il quale ha tra le sue funzioni quella di promuovere la proliferazione e la progressione del ciclo cellulare, pathway nelle quali rivestono un ruolo importante anche le cicline: una diretta modulazione dell'espressione di tali geni da parte di hsa-miR-145 sembra quindi interessante da valutare con ulteriori studi.

Per hsa-miR-202* non esistono al momento dei target validati, e tra quelli proposti dal nostro software il gene che codifica per la fibrosina è senza dubbio il più intrigante da studiare. Si tratta infatti di una molecola ad attività infiammatoria, che viene prodotta dai linfociti attivati e stimola la proliferazione dei fibroblasti.

Un miRNA molto studiato è invece hsa-miR-222, per il quale sono già stati validati geni target come l'inibitore di una chinasi ciclina-dipendente (CDKN1B) e il recettore degli estrogeni ESR1. Nel nostro studio correla direttamente con l'avanzare dello stage, e anche nel tumore presenta un'espressione maggiore rispetto al tessuto non tumorale circostante. Esso, quindi sembra promuovere la progressione del ciclo cellulare e la proliferazione: PDCD7, un gene che promuove l'apoptosi, sembra essere un ulteriore buon candidato dell'azione di hsa-miR-222.

Un altro miRNA che presenta dei target interessanti è hsa-miR-548b-3p, per il quale in letteratura non esistono al momento studi mirati. Ciò che emerge dall'analisi di espressione è una sua correlazione con l'eziologia virale, in particolare una sovraespressione nei campioni HBV-positivi. Tra i target predetti è presente il gene SON, il quale codifica per una proteina in grado di legare una specifica sequenza di DNA a monte della sequenza regolatoria del promotore del gene Core di HBV, sopprimendo l'espressione dei geni virali e inibendo la produzione di virioni. Non è chiaro il motivo per cui la cellula dovrebbe inibire l'espressione di questo gene attraverso l'attivazione di hsa-miR-548b-3p, favorendo in questo modo la replicazione del virus, ma non è da escludere che il virus stesso abbia sviluppato dei meccanismi volti all'attivazione di miRNA cellulari per favorire la propria sopravvivenza.

Anche hsa-miR-622 mostra una correlazione con l'eziologia virale delle epatiti e anch'esso risulta più espresso in quelle HBV-positive, ma al contrario del precedente non ha target predetti che possano essere associati ad uno dei due virus in particolare. Tuttavia, potrebbe avere un ruolo nella proliferazione, e in particolare nel processo di replicazione del DNA, in quanto tra i suoi target predetti ci sono il gene POLL (polimerasi DNA-dipendente) e MCM4, che codifica per una proteina essenziale per l'inizio della replicazione del genoma eucariotico.

Infine, hsa-miR-625*: nel nostro studio questo miRNA risulta espresso maggiormente nei campioni fibrotici, mentre diminuisce nelle cirrosi. Al momento non ci sono dati in letteratura relativi ad un suo eventuale comportamento alterato nel HCC, ma tra i suoi potenziali target segnaliamo l'oncogene PIM3, che appartiene ad una famiglia di proto oncogeni che codificano per chinasi serin/treoniniche, e DAB2IP, un gene che agisce nella pathway che coinvolge Ras. Per completezza va ribadito che i risultati di predizione presentati in questa sezione necessitano tutti di una validazione sperimentale prima di poter essere accettati come target, ma un simile lavoro bioinformatico risulta necessario per indirizzare la ricerca verso quelli che sono i candidati più probabili all'interno di una vastità di molecole tutte potenzialmente valide.

4. DISCUSSIONE

La scoperta, relativamente recente, che variazioni dell'espressione dei miRNA si associano all'insorgenza di molti tipi di tumori, compresi i più comuni come il cancro al polmone, alla mammella, alla prostata e le neoplasie gastrointestinali, ha portato in breve tempo alla definizione di profili di espressione di miRNA specificamente associati alle diverse patologie neoplastiche.

In questo lavoro abbiamo studiato l'andamento dell'espressione dei miRNA in relazione allo sviluppo di epatite cronica/HCC, a partire dallo stadio di fibrosi lieve fino allo sviluppo di cirrosi e, infine, all'insorgenza dell'epatocarcinoma. Abbiamo identificato un pannello di miRNA che vanno incontro ad una decisa diminuzione della loro espressione nel tessuto tumorale (hsa-miR-122, hsa-miR-145, hsa-miR-195, hsa-miR-199a, hsa-miR-199b), mentre alcuni altri hanno mostrato un comportamento opposto, risultando più espressi nel tumore (hsa-miR-222, hsa-miR-18a). In entrambi i casi questi risultati sono concordi con la letteratura a riguardo (Wong et al., 2010; Xu et al., 2009; Bai et al., 2009; Li et al., 2008; Murakami et al., 2006). Per alcuni di questi miRNA sono state definite anche funzioni specifiche nella patogenesi: hsa-miR-145 inibisce la formazione di metastasi e limita l'invasività della massa tumorale (Sachdeva et al., 2010); hsa-miR-222, al contrario, conferisce capacità migratorie alle cellule agendo sulla via di Akt (Wong et al., 2010); hsa-miR-195 regola la transizione G1/S del ciclo del ciclo cellulare (Xu et al., 2009), mentre hsa-miR-122 è stato implicato con l'infezione da HCV ed è stato dimostrato interagire direttamente col genoma virale (Jopling et al., 2006; Henke et al., 2008).

Abbiamo inoltre individuato altri miRNA che presentano variazioni significative dell'espressione nel fegato affetto da epatite, in associazione a livelli differenti di *staging* (hsa-miR-21, hsa-miR-34a) e *grading* (hsa-miR-223, hsa-miR-224, hsa-miR-202*, hsa-miR-548b-3p), oppure che risultano correlati all'eziologia virale da HBV o da HCV (hsa-miR-211, hsa-miR-17, hsa-miR-17*, hsa-miR-18a, miR-548b-3p).

Infine, per attribuire un significato biologico a questi dati, abbiamo sviluppato un metodo bioinformatico per la predizione dei geni target dei miRNA, che è stato poi applicato per alcuni di essi e ha fornito dei candidati interessanti da studiare nel prossimo futuro.

La valutazione dell'espressione dei miRNA nel tessuto di epatocarcinoma è stata effettuata mediante l'utilizzo della metodica di realtime PCR: questa scelta è dovuta alla tipologia dei campioni in nostro possesso, per i quali la qualità dell'RNA estratto non presentava i requisiti richiesti all'applicazione di altre metodiche a più ampio spettro, come ad esempio i microarray, evento che crediamo possa essere spiegato con la maggiore difficoltà di preservazione del campione durante le fasi di raccolta. La scelta dei miRNA da studiare con questa metodica è stata un punto di criticità del protocollo ed è stata guidata dalla letteratura: tra i criteri di selezione

citiamo principalmente una loro documentata presenza nel tessuto epatico e le indicazioni bioinformatiche su una possibile funzione che potesse essere correlata con la patologia. Lo studio dell'espressione dei miRNA è stato effettuato ricorrendo al confronto diretto di due diversi campioni ottenuti contestualmente da ciascun paziente, ovvero al rapporto fra l'espressione di ciascun miRNA nel tessuto tumorale e nel tessuto circostante di ciascun soggetto. Questa scelta, adottata comunemente in studi di questo tipo, può presentare dei margini di imprecisione dovuti essenzialmente alla natura, eterogenea e complessa, dei cosiddetti tessuti 'circostanti' il tumore.

L'analisi dei profili di espressione dei miRNA nei campioni bioptici relativi alle epatiti ad eziologia virale sono state invece effettuate impiegando i microarray, tecnologia che ha permesso di monitorare contemporaneamente l'espressione di tutti i miRNA noti; inoltre, la coorte di campioni a nostra disposizione ci ha consentito di ottenere dei quadri specifici per le diverse fasi della patologia epatica e per le due distinte eziologie virali. Lo studio è stato condotto confrontando i valori di espressione dei miRNA all'interno di gruppi di campioni definiti in base all'eziologia e a parametri clinico-patologici quali il *grading* e lo *staging*, ed ha portato alla definizione di insiemi di miRNA significativamente correlati alla variazione di tali parametri. Una simile strategia è stata adottata recentemente anche in altri lavori riguardanti studi di espressione genica mirati a identificare dei marcatori epatici specifici che riflettessero i cambiamenti biologici nella progressione della fibrosi epatica (Takahara et al., 2008) ma, al momento, non sono presenti in letteratura dati inerenti ai miRNA, se non relativi esclusivamente all'epatocarcinoma.

Il fatto che molti miRNA siano deregolati nel cancro o nelle diverse fasi di lesioni pre-neoplastiche suggerisce che molte delle pathway implicate nella tumorigenesi possano avere dei miRNA come target o effettori a valle delle loro vie di segnalazione. Per questo motivo, nelle condizioni patologiche in cui sia dimostrata la perdita o la sovraespressione di tali molecole, si potrebbe considerare l'ipotesi di impiegarli come farmaci, sia come miRNA che come anti-miRNA. La loro introduzione all'interno dei tessuti specifici, come ad esempio nel caso di una massa tumorale, può essere ottenuta in modi differenti, ad esempio per mezzo di vettori virali, quali adenovirus, virus adeno-associati o retrovirus, che sono in grado di trasferire i miRNA nei tessuti di interesse, oppure possono essere impiegati vettori non virali, come i liposomi. In alternativa, miRNA o anti-miRNA possono essere usati direttamente senza l'utilizzo di un vettore di espressione, sia modificati per aumentare la loro emivita *in vivo*, sia in forma non modificata. Esperimenti di trasferimento sono già stati effettuati in modelli animali, come nel caso di miR-15a e miR-16-1, i quali sono stati usati nel topo per inibire la crescita di cellule leucemiche in cui l'espressione di questi miRNA era ridotta (Calin et al., 2008), o nel caso dei componenti della famiglia di miR-29, usati come oncosoppressori nel cancro al polmone (Fabbri et al., 2007). Per quanto riguarda l'epatocarcinoma, inoltre, un lavoro recente realizzato su modello murino ha dimostrato la possibilità di utilizzare vettori virali adeno-associati per il trasferimento *in vivo* di

miRNA (in questo caso hsa-miR-26a), con il risultato di inibire la crescita e la proliferazione delle cellule tumorali (Kota et al., 2009). Esperimenti di questo tipo fanno supporre che simili strategie possano essere impiegate anche sull'uomo, soprattutto per la scarsità di effetti collaterali osservati.

Lo studio dei profili di espressione dei miRNA può essere mirato anche all'individuazione di marcatori dei differenti stati patologici, utili sia per la classificazione dei pazienti che per la scelta dei trattamenti terapeutici, in quanto alcuni di essi sono già stati associati a forme di resistenza a farmaci (Pogribny et al., 2010). Inoltre, è interessante notare come alcune delle modificazioni epigenetiche osservate nel cancro possano essere dovute alla deregolazione dei miRNA, a causa della loro azione come modulatori di geni che codificano per proteine coinvolte nella struttura e nella funzione della cromatina, come metilasi o acetilasi (Noonan et al., 2009).

La parte del lavoro di tesi fin qui discussa, relativa ad attività di laboratorio, ha portato al conseguimento di risultati che, se da un lato sono molto significativi, dall'altro suscitano tutta una serie di punti interrogativi su quello che potrebbe essere il significato biologico del comportamento alterato di alcuni dei miRNA studiati nei vari stadi di epatite cronica/HCC. La risposta a tali quesiti risiede inevitabilmente nella comprensione della funzione che ciascuno di essi svolge all'interno della cellula, e la chiave per accedere a questa conoscenza è l'identificazione dei geni che direttamente subiscono gli effetti dell'azione regolatoria di queste molecole. Purtroppo, però, la struttura intrinseca dei miRNA e il loro peculiare meccanismo di riconoscimento (peraltro non ancora del tutto chiarito) generano delle problematiche che non facilitano questo processo: la lunghezza esigua delle sequenze, e ancor di più la possibilità della formazione di *mismatch* nella fase di appaiamento con il trascritto target, accrescono enormemente il numero di potenziali siti di legame e impongono lo sviluppo di metodi rapidi di selezione dei candidati più probabili. A tal fine, la scelta della comunità scientifica è ricaduta sulla bioinformatica: in breve tempo sono stati infatti sviluppati e pubblicati numerosi algoritmi in grado di predire, in modo più o meno accurato, quali siano i geni che potenzialmente possano rappresentare il bersaglio d'azione di uno specifico miRNA (John et al., 2004; Krüger et al., 2006; Kertesz et al., 2007; Rusinov et al., 2005; Lewis et al., 2003; Kiriakidou et al., 2004). Tutti gli algoritmi proposti finora, però, presentano delle limitazioni piuttosto importanti, dovute principalmente all'incompleta comprensione dei meccanismi che regolano l'interazione miRNA-target: se, infatti, sembra ormai appurato che un perfetto appaiamento della porzione di sequenza nella regione 2-8 al 5' del miRNA (la cosiddetta *seed sequence*) sia fondamentale in tal senso (Lai, 2002), per molti altri fattori esistono soltanto delle supposizioni che spesso incontrano pareri discordanti in letteratura. Si parla, infatti, della possibile presenza di siti di legame anche nella regione 5' dei geni (Tsai et al., 2009) o nella regione codificante (Zhou et al., 2009), mentre è ancora di difficile interpretazione la variabilità del contesto in cui il sito di legame è contenuto e

l'eventuale presenza di altri fattori che possano determinarne una maggiore o minore disponibilità. Queste e probabilmente altre variabili che intervengono nel meccanismo di riconoscimento miRNA-trascritto target sono difficilmente valutabili dagli algoritmi di ricerca automatici i quali, invece, basano le loro predizioni quasi esclusivamente sull'appaiamento tra le sequenze, assegnando particolare importanza alla *seed sequence*. Di conseguenza, la probabilità di ottenere falsi positivi risulta alquanto consistente e comporta la generazione di un numero spesso molto elevato di predizioni.

Per cercare di ovviare a questo problema si è deciso di sviluppare un nuovo metodo che, pur basandosi sull'utilizzo di algoritmi già esistenti (Miranda, RNAhybrid e Pita), ci consentisse di migliorare l'attendibilità delle predizioni. Il nuovo software si fonda sulla strategia del *consensus*, ossia della concordanza tra programmi diversi: infatti, se un risultato viene conseguito sempre uguale a sé stesso utilizzando software diversi, basati su algoritmi differenti, la sua consistenza è in genere molto superiore rispetto allo stesso dato generato da un singolo programma.

Entrando nel dettaglio del metodo bioinformatico prodotto, la sezione più importante è senza dubbio quella del calcolo delle soglie di energia di legame specifiche per ogni miRNA. Infatti, la determinazione di tali soglie è cruciale nella fase di filtraggio dei risultati da parte di tutti gli algoritmi utilizzati in questo lavoro, e da esse dipende la presenza o meno di un determinato target nella predizione finale. L'utilizzo di una soglia non appropriata può portare alla generazione di risultati anche molto lontani dalla realtà: una soglia troppo stringente causa un notevole abbassamento della sensibilità del metodo, con la perdita di molti risultati potenzialmente validi, mentre una soglia troppo permissiva causa, in genere, un aumento spropositato del numero di risultati e, di conseguenza, porta alla produzione di un numero elevato di falsi positivi. Nel caso si stia lavorando con uno o pochi miRNA, la calibrazione di tale parametro è un processo che può essere effettuato manualmente, ad esempio facendo diversi tentativi con valori di energia di legame differenti; diversamente, però, nel caso si stia lavorando con un numero maggiore di miRNA e si stia impiegando più di un software di predizione (come nel nostro caso), il processo di attribuzione delle soglie può risultare difficile e oneroso in termini di tempo. Proprio per questo motivo abbiamo deciso di implementare un algoritmo automatico in grado di compiere questa operazione in base a criteri stabiliti a priori. Il principio su cui si basa questa particolare sezione dell'algoritmo è dovuto all'osservazione della diretta correlazione che intercorre tra la percentuale in GC di un miRNA e il numero di target che uno qualsiasi degli algoritmi di predizione utilizzati fornisce ad una determinata soglia. In altre parole, abbiamo osservato come miRNA con elevato contenuto di GC avessero, a parità di soglia, un numero molto maggiore di target predetti rispetto a miRNA con basso contenuto di GC. Questo fenomeno, già riportato in letteratura in relazione ad un altro software di predizione non incluso nel nostro algoritmo (Davis et al., 2008), è dovuto al fatto che, durante l'appaiamento tra due molecole di acido nucleico (come ad esempio avviene tra

miRNA e molecola target), il legame che si forma tra le basi guanina e citosina, costituito da tre legami ponte-idrogeno, è più stabile di quello tra adenina e timina, costituito da due ponti-idrogeno solamente, e conferisce di conseguenza una maggiore solidità all'appaiamento.

Grazie a questo software abbiamo identificato dei candidati target per alcuni miRNA potenzialmente molto interessanti: tra di essi citiamo in particolare il gene SIKE, predetto per hsa-miR-7 e implicato nella via di risposta anti-virale mediata dall'interferone, e SON, predetto per hsa-miR-548b-3p e associato alla regolazione del ciclo replicativo di HBV. Una fase successiva del progetto, già avviata, prevederà la verifica sperimentale di questi dati: in un primo momento si effettueranno i cosiddetti saggi della luciferasi, nei quali si valuta la potenziale interazione tra miRNA e target mediante l'osservazione di un gene reporter che produce una proteina fluorescente; quindi, si andrà a valutare direttamente la presenza del trascritto e della proteina target in relazione a differenti livelli di espressione del miRNA, mediante esperimenti di realtime PCR e saggi di Western-Blot.

In conclusione, questo lavoro di tesi ha portato all'individuazione di alcuni miRNA diversamente espressi nell'epatocarcinoma, nei diversi stadi di epatite virale cronica, e in relazione all'infezione da HBV e HCV. Per la prima volta è stata inoltre stabilita una correlazione tra il pattern di espressione dei miRNA e i parametri clinico-patologici che definiscono la progressione dell'epatite virale cronica dalla fibrosi lieve alla cirrosi. Nel tentativo di approfondire la nostra conoscenza sul ruolo biologico di queste molecole è stato quindi sviluppato un software di predizione dei geni target, applicato poi ad alcuni dei miRNA proponendo un modello di possibile applicazione. I promettenti risultati ottenuti suscitano nuovi spunti di riflessione relativi alla delucidazione dei geni target e pongono le basi per futuri studi sull'utilizzo dei miRNA come nuovi strumenti di indagine diagnostica, definizione prognostica e utilizzo terapeutico.

APPENDICE A:

IMPLEMENTAZIONE DEGLI ALGORITMI IN PERL

A.1: Human3primeUTR.pl

```
#!/usr/bin/perl -w

use Bio::SeqIO;
use Getopt::Long;
my ($script) = ($0 =~ m|([^\s]*)$|);
my $Uso = "use of $script:
-rnadb      human.rna.fna
-gbff      human.rna.gbff
-dir       output directory";

my ($rnadb,$gbff,$dir);
my $seq = '';
my $result = GetOptions ("rnadb=s"=>\$rnadb,
                        "gbff=s"=>\$gbff,
                        "dir=s"=>\$dir);

unless ($result)
{
    print STDERR "Some options are not well formatted\n";
    exit (1);
}
if ((!$rnadb)||(!$gbff)||(!$dir))
{
    print "$Uso";
    exit;
}

#load all RNA sequences in a hash
my ($header,@fields,%hash,%cds);
open (FILE,"$rnadb");
while (my $line = <FILE>)
{
    chomp $line;
    if ($line =~ /^>/)
    {
        if ($seq)
        {
            $hash{$header} = $seq;
            $seq = '';
        }
        @fields = split (/\\|/, $line);
        ($header) = ($fields[3] =~ /(\w+_d+)/);
    }
    else
    {
        $seq .= $line;
    }
}
if ($seq)
{
    $hash{$header} = $seq;
    $seq = '';
}
close FILE;

#split gbff file into pieces, one for each mRNA

my $split = '';
my $i =1;
qx{mkdir $dir/singlesGenbank};

#create a directory with a file for each mRNA
```

```

my $line;
open (GBFF,"$gbff");
while ($line = <GBFF>)
{
  chomp $line;
  if ($line eq ("\\"))
  {
    if ($split)
    {
      $split .= "$line\n";
      open (OUT,">$dir/singlesGenbank/gb$i");
      print OUT "$split";
      $i++;
    }
    $split = '';
  }
  else
  {
    $split .= "$line\n";
  }
}
if ($split)
{
  $split .= "$line\n";
  open (OUT,">$dir/singlesGenbank/gb$i");
  print OUT "$split";
  $i++;
}
close GBFF;

#now extract CDS end from each of these files and print them in a new file

my $endCDS = '';
open (FIN,">$dir/singlesGenbank/human3UTR");
opendir (DIR,"$dir/singlesGenbank");
while (my $file = readdir (DIR))
{
  next if ($file =~ /\^\.\/);
  my $seqio_object = Bio::SeqIO->new(-file => "$dir/singlesGenbank/$file" );
  my $seq_object = $seqio_object->next_seq;
  $id = $seq_object->display_id();
  $seq = $seq_object->seq();
  print FIN ">$id\n";
  foreach $feat ( $seq_object->get_SeqFeatures() )
  {
    if ($feat->primary_tag() eq 'gene')
    {
      $endGene = $feat->end;
    }
    elsif ($feat->primary_tag() eq 'CDS')
    {
      $endCDS = $feat->end;
    }
  }
  if ($endCDS)
  {
    if ($endCDS ne $endGene)
    {
      if (($endGene-$endCDS) <= 1000)
      {
        print FIN $seq_object->subseq($endCDS,$endGene),"\n";
      }
      else
      {
        print FIN $seq_object->subseq($endCDS,($endCDS+999)), "\n";
      }
    }
    else
    {
      print FIN "no 3'-utr\n";
    }
  }
  else
  {

```

```

        print FIN "no 3'-utr\n";
    }
    $endCDS = '';
}
close FIN;
close DIR;

```

A.2: VariableThresholdsPredictor.pl

```

#!/usr/bin/perl -w

use strict;
use Getopt::Long;
use FileHandle;
use IO::File;
use Bio::SeqIO;

my ($script) = ($0 =~ m|([^\s]*)$|);
my $Uso = "use of $script:

-mirna                mirna's multiple fasta file
-mirandaOut           miranda output slimmed file (not stringent)
-RNAhybridOut         RNAhybrid output file (not stringent)
-pitaOut              pita output file (_targets.tab file, not filtered)
-utr                  utr fasta file
-gbff                 genbank file of the database used (each sequence block divided by
'//');

my ($mirna,$mirandaOut,$RNAhybridOut,$pitaOut,$utr,$gbff);
my $result = GetOptions ("mirna=s"=>\$mirna,
                        "mirandaOut=s"=>\$mirandaOut,
                        "RNAhybridOut=s"=>\$RNAhybridOut,
                        "pitaOut=s"=>\$pitaOut,
                        "utr=s"=>\$utr,
                        "gbff=s"=>\$gbff);

unless ($result)
{
    print STDERR "Some options are not well formatted\n";
    exit (1);
}
if ((!$mirna)||(!$mirandaOut)||(!$RNAhybridOut)||(!$pitaOut)||(!$utr)||(!$gbff))
{
    print "$Uso";
    exit;
}

#set absolute paths for auxiliary programs

#path of mirandaOutputFilterForVariableThreshold.pl
my $binMiranda = "/home/enrico/Desktop/cytomegalovirus/bin/downloaded_softwares/miRanda-1.9";

#path of RNAhybridOutputFilterForVariableThreshold.pl
my $binRNAhyb = "/home/enrico/Desktop/cytomegalovirus/bin/downloaded_softwares/RNAhybrid-2.1";

#path of pitaOutputFilterForVariableThreshold.pl
my $binPita = "/home/enrico/Desktop/cytomegalovirus/bin/downloaded_softwares/pita";

#path of outputCrossAnalyser.pl
my $binOutputCrossAnalyser = "/home/enrico/Desktop/cytomegalovirus/bin";

#calculate GC% for each miRNA and store in a hash

print "Calculating GC% for each miRNA...\n";
my ($header,$gpcPercentage,$newKey,$outMiranda,$decimals,$outRNAhyb,$outPita);
my $seq = '';
my (@nucleotides,%gcContent,%thresholds);
my $tot=0;
my $gc=0;

open (MIRNA,"$mirna");

```

```

while (my $line = <MIRNA>)
{
    chomp $line;
    if ($line =~ /^>/)
    {
        if ($seq)
        {
            @nucleotides = split(//,$seq);
            foreach my $nuc (@nucleotides)
            {
                $tot++;
                if (($nuc =~ /c/) || ($nuc =~ /g/))
                {
                    $gc++;
                }
            }
            $gcPercentage = ($gc/$tot);
            ($gcPercentage) = ($gcPercentage =~ /(\d+)\$/);
            $gcPercentage = "$gcPercentage".'0';
            ($gcPercentage) = ($gcPercentage =~ /(^\\d\\d)/);
            $gcContent{$header} = $gcPercentage;          #hash containing gc contents
        }
        for each mirna
        {
            $tot =0;
            $gc =0;
        }
        ($header) = ($line =~ /(miR.+)/);
        $seq = '';
    }
    else
    {
        $seq .= $line;
    }
}
if ($seq)
{
    @nucleotides = split(//,$seq);
    foreach my $nuc (@nucleotides)
    {
        $tot++;
        if (($nuc =~ /c/) || ($nuc =~ /g/))
        {
            $gc++;
        }
    }
    $gcPercentage = ($gc/$tot);
    ($gcPercentage) = ($gcPercentage =~ /(\d+)\$/);
    $gcPercentage = "$gcPercentage".'0';
    ($gcPercentage) = ($gcPercentage =~ /(^\\d\\d)/);
    $gcContent{$header} = $gcPercentage;
    $tot =0;
    $gc =0;
}
close MIRNA;

#for each mirna, filter output results from different softwares with the specific
calculated threshold

print "For each prediction software calculate each mirna's specific threshold and filter
output files...\n";

#calculate deltaGCmax

my $max;
my $min;
foreach my $key (keys (%gcContent))
{
    if ($max)
    {
        if ($gcContent{$key} > $max)
        {
            $max = $gcContent{$key};
        }
    }
    else
    {

```



```

    $max = $gcContent{$key};
}
if ($min)
{
    if ($gcContent{$key} < $min)
    {
        $min = $gcContent{$key};
    }
}
else
{
    $min = $gcContent{$key};
}
}

my $deltaGCmax = ($max - $min);    #max gc% difference between mirnas gc content

#ENERGY PARAMETERS SECTION

my $deltaGC;                                #max gc% (85) minus current gc%
my $deltaMaxMiranda = 14;                  #miranda delta between max and min thresholds (-37
and -25)
my $deltaMaxRNAhyb = 13;                   #RNAhybrid delta between max and min thresholds (-39 and -
26)
my $deltaMaxPita = 12;                     #pita delta between max and min thresholds (-25 and
-9)
my $mirandaMinThreshold = (-34);          #miranda most stringent min thresholds used in
predictions
my $RNAhybMinThreshold = (-35);          #RNAhybrid most stringent min thresholds used in
predictions
my $pitaMinThreshold = (-17);             #pita most stringent min thresholds used in
predictions

foreach my $key (keys(%gcContent))
{
    if ($key =~ /\*$/)
    {
        $newKey = $key;
        $newKey =~ s/\*/asteriscato/;    #this is mirna name to use for filtering files
    }
    else
    {
        $newKey = $key;                  #this is mirna name to use for filtering files
    }
    $deltaGC = $max-($gcContent{$key});

    #calculate miranda threshold differential

    my $mirandaDifferential = (($deltaGC*$deltaMaxMiranda)/$deltaGCmax);
    ($decimals) = ($mirandaDifferential =~ /\d+$/);
    $decimals .= 00;
    ($decimals) = ($decimals =~ /^(^d\d)/);
    if ($decimals >= 50)
    {
        ($mirandaDifferential) = ($mirandaDifferential =~ /^(^d+));
        $mirandaDifferential++;          #rounds off to the superior integer if decimals are
above 0.50
    }
    else
    {
        ($mirandaDifferential) = ($mirandaDifferential =~ /^(^d+));
    }

    #filter miranda output for this mirna with its specific threshold

    my $filterMiranda = ($mirandaMinThreshold + $mirandaDifferential);
    $outMiranda .= qx{$binMiranda/mirandaOutputFilterForVariableThreshold.pl -
mirandaOutSlimmed $mirandaOut -filter $filterMiranda -mirna $newKey};
    $thresholds{$newKey}->{'miranda'} = $filterMiranda;    #record the threshold for this
mirna

    #calculate RNAhybrid threshold differential

    my $RNAhybDifferential = (($deltaGC*$deltaMaxRNAhyb)/$deltaGCmax);
    ($decimals) = ($RNAhybDifferential =~ /\d+$/);

```

```

$decimals .= 00;
($decimals) = ($decimals =~ /^(^d\d)/);
if ($decimals >= 50)
{
    ($RNAhybDifferential) = ($RNAhybDifferential =~ /^(^d+)/);
    $RNAhybDifferential++; #rounds off to the superior integer if decimals are
above 0.50
}
else
{
    ($RNAhybDifferential) = ($RNAhybDifferential =~ /^(^d+)/);
}

#filter RNAhybrid output for this mirna with its specific threshold

my $filterRNAhyb = ($RNAhybMinThreshold + $RNAhybDifferential);
$outRNAhyb .= qx{$binRNAhyb/RNAhybridOutputFilterForVariableThreshold.pl -RNAhybridOut
$RNAhybridOut -filter $filterRNAhyb -mirna $newKey};
$thresholds{$newKey}->{'RNAhyb'} = $filterRNAhyb; #record the threshold for this
mirna

#calculate pita threshold differential

my $pitaDifferential = (($deltaGC*$deltaMaxPita)/$deltaGCmax);
($decimals) = ($pitaDifferential =~ /(\d+)/);
$decimals .= 00;
($decimals) = ($decimals =~ /^(^d\d)/);
if ($decimals >= 50)
{
    ($pitaDifferential) = ($pitaDifferential =~ /^(^d+)/);
    $pitaDifferential++; #rounds off to the superior integer if decimals are above
0.50
}
else
{
    ($pitaDifferential) = ($pitaDifferential =~ /^(^d+)/);
}

#filter pita output for this mirna with its specific threshold

my $filterPita = ($pitaMinThreshold + $pitaDifferential);
$outPita .= qx{$binPita/pitaOutputFilterForVariableThreshold.pl -pitaOut $pitaOut -
threshold $filterPita -mirna $newKey};
$thresholds{$newKey}->{'pita'} = $filterPita; #record the threshold for this mirna
}

#create an array with miRNA names

my @mirna;
open (MIRNA,"$mirna");
while (my $line = <MIRNA>)
{
    chomp $line;
    if ($line =~ /^>/)
    {
        ($line) = ($line =~ /(miR.+)/);
        push (@mirna,$line);
    }
}
close MIRNA;

my (%hash);
open (UTR,"$utr");
while (my $line = <UTR>)
{
    chomp $line;
    if ($line =~ /^>/)
    {
        (my $id) = ($line =~ /(\w+)/);
        foreach my $value (@mirna)
        {
            $hash{$id}->{'miranda'}->{$value}=0;
        }
        foreach my $value (@mirna)
        {

```

```

        $hash{$id}->{'RNAhybrid'}->{$value}=0;
    }
    foreach my $value (@mirna)
    {
        $hash{$id}->{'pita'}->{$value}=0;
    }
}
close UTR;

#now start parsing output files from prediction programs
#1: miranda

open (MIRANDA, '<', \SoutMiranda); #opens a file in memory to read
while (my $line = <MIRANDA>)
{
    chomp $line;
    my @fields = split (/\\t/, $line);
    (my $mir) = ($fields[0] =~ /([^\>].+)/);
    if ($mir =~ /\*$/)
    {
        $mir =~ s/\\*/asteriscato/;
    }
    $hash{$fields[1]}->{'miranda'}->{$mir}=1;
}
close MIRANDA;

#2: RNAhybrid

open (HYBRID, '<', \SoutRNAhyb); #opens a file in memory to read
while (my $line = <HYBRID>)
{
    next if ($line =~ /^\\-/); #skip separation lines
    next if ($line =~ /^(\\s)*$/); #skip blank lines
    next if ($line =~ /^Micro/); #skip header lines
    chomp $line;
    my @fields = split (/\\:/, $line);
    my $mir = $fields[2];
    if ($mir =~ /\*$/)
    {
        $mir =~ s/\\*/asteriscato/;
    }
    $hash{$fields[0]}->{'RNAhybrid'}->{$mir}=1;
}
close HYBRID;

#3: pita

open (PITA, '<', \SoutPita); #opens a file in memory to read
while (my $line = <PITA>)
{
    chomp $line;
    next if ($line =~ /^RefSeq/);
    my @fields = split (/\\t/, $line);
    (my $mir) = ($fields[1] =~ /(miR\\-.+)/);
    if ($mir =~ /\*$/)
    {
        $mir =~ s/\\*/asteriscato/;
    }
    $hash{$fields[0]}->{'pita'}->{$mir}=1;
}
close PITA;

#finally create output: 1 table for each miRNA with all values (0 or 1) indicating a
negative or positive prediction for each ID (utr) by each predictive program

print "Determining the targets shared by 2 or more softwares...\\n";
qx{mkdir ./mirnaTables};
foreach my $value (@mirna)
{
    my $out = '';
    $out .= "ID\\tMIRANDA\\tRNAhybrid\\tPITA\\n";
    foreach my $ID (keys (%hash))
    {

```

```

        my $progNameRef = $hash{$ID};
        $out .= "$ID\t$$progNameRef{'miranda'}->{$value}\t$$progNameRef{'RNAhybrid'}-
>{$value}\t$$progNameRef{'pita'}->{$value}\n";
    }
    open (OUT,">./mirnaTables/$value");
    print OUT $out;
    close OUT;
}

qx{mkdir ./mirnaCommonTargets};
opendir (DIR,"./mirnaTables");
while (my $file = readdir (DIR))
{
    next if ($file =~ /^\.\/); #skip hidden files
    #first filter targets predicted by all the programs
    my $out = '';
    $out .= "Targets predicted by Miranda, RNAhybrid and Pita:\n";
    $out .= qx{$binOutputCrossAnalyser/outputCrossAnalyser.pl -results ./mirnaTables/$file
-filter 111};
    #then those common to 2 out of 3 programs
    $out .= "Targets predicted by Miranda and RNAhybrid:\n";
    $out .= qx{$binOutputCrossAnalyser/outputCrossAnalyser.pl -results ./mirnaTables/$file
-filter 110};
    $out .= "Targets predicted by Miranda and Pita:\n";
    $out .= qx{$binOutputCrossAnalyser/outputCrossAnalyser.pl -results ./mirnaTables/$file
-filter 101};
    $out .= "Targets predicted by RNAhybrid and Pita:\n";
    $out .= qx{$binOutputCrossAnalyser/outputCrossAnalyser.pl -results ./mirnaTables/$file
-filter 011};
    open (OUT,">./mirnaCommonTargets/$file");
    print OUT $out;
    close OUT;
}
close DIR;

#now search for multiple binding sites

print "Verifying the presence of multiple mirna sites in previously predicted
targets...\n";
my (%tmp,%targets,%final);
open (RNAHYB,"$RNAhybridOut");
while (my $line = <RNAHYB>)
{
    next if ($line =~ /^-\/); #skip separation lines
    next if ($line =~ /^(\/s)*$/); #skip blank lines
    next if ($line =~ /^Micro\/); #skip header lines
    chomp $line;
    my @fields = split (\/:\/,$line);
    if ($fields[2] =~ /\s*/)
    {
        $fields[2] =~ s\/\*/asteriscato\/;
    }
    if ($tmp{$fields[2]}->{$fields[0]})
    {
        $tmp{$fields[2]}->{$fields[0]}++;
    }
    else
    {
        $tmp{$fields[2]}->{$fields[0]} = 1;
    }
}
close RNAHYB;
opendir (COMMON,"./mirnaCommonTargets");
while (my $file = readdir (COMMON))
{
    next if ($file =~ /^\.\/); #skip hidden files
    open (FILE,"./mirnaCommonTargets/$file");
    while (my $line = <FILE>)
    {
        next if ($line =~ /^(\/s)*$/); #skip blank lines
        next if ($line =~ /^Targets\/);
        next if ($line =~ /^miR\/);
        chomp $line;
        $targets{$line} = $line;
    }
}

```

```

close FILE;
my $fileFormatted = $file;
if ($fileFormatted =~ /\*$/)
{
    $fileFormatted =~ s/\*/asteriscato/;
}
open (MIRANDA,"$mirandaOut");
while (my $line = <MIRANDA>)
{
    my @fields = split (/\\t/, $line);
    if ($fields[0] =~ /\*$/)
    {
        $fields[0] =~ s/\*$*/asteriscato/;
    }
    if (($fields[0] =~ /$fileFormatted/) && ($targets{$fields[1]}))
    {
        my @sites = split (/\\s/, $fields[9]);
        $final{$fileFormatted}->{$fields[1]}->{'miranda'} = ($#sites);
    }
}
close MIRANDA;

#search multiple sites in RNAhybrid output
foreach my $key (keys (%targets))
{
    if ($tmp{$fileFormatted}->{$key})
    {
        $final{$fileFormatted}->{$key}->{'RNAhyb'} = $tmp{$fileFormatted}->{$key};
    }
    else
    {
        $final{$fileFormatted}->{$key}->{'RNAhyb'} = 0;
    }
}

#search multiple sites in pita output
open (PITA,"$pitaOut");
while (my $line = <PITA>)
{
    next if ($line =~ /^Ref/);          #skip header line
    chomp $line;
    my @fields = split (/\\t/, $line);
    if ($fields[1] =~ /\*$/)
    {
        $fields[1] =~ s/\*$*/asteriscato/;
    }
    if (($fields[1] =~ /$fileFormatted/) && ($targets{$fields[0]}))
    {
        $final{$fileFormatted}->{$fields[0]}->{'pita'} = $fields[2];
    }
}
close PITA;
%targets = ();
}
close COMMON;
$tmp = ();          #I don't need it anymore
print "Loading target info from gbff file (this step may take a little bit of time)...\\n";
my $split = '';          #variable used to split the gbff file in singles
genbank files
my %descriptions;
open (GBFF,"$gbff");
while (my $line = <GBFF>)
{
    chomp $line;
    if ($line eq ("\\\\\\"))
    {
        if ($split)
        {
            my $var = '';
            $split .= "$line\\n";
            open(MEMORY, '>', \\$var)          #opens a file in memory to
        }
        or die "Can't open memory file: !$\\n";
        print MEMORY "$split";          #output will end up in $var
        close MEMORY;
    }
}

```

```

open (MEMORY, '<', \svar); #opens a file in memory to
read
    my $seqio_object = Bio::SeqIO->new(-fh => \*MEMORY);
    my $seq_object = $seqio_object->next_seq;
    my $id = $seq_object->display_id();
    my $desc = $seq_object->desc();
    $descriptions{$id} = $desc;
}
$split = '';
}
else
{
    $split .= "$line\n";
}
}
close GBFF;

#FINAL STEP: create output

print "Creating output...\n";
qx{mkdir ./PredictionFinalResults/};
my $final = '';
opendir (COMMON, "./mirnaCommonTargets");
while (my $file = readdir (COMMON))
{
    next if ($file =~ /^\.\/); #skip hidden files
    open (FILE, "./mirnaCommonTargets/$file");
    my $mirandaThreshold = $thresholds{$file}->{'miranda'};
    my $RNAhybThreshold = $thresholds{$file}->{'RNAhyb'};
    my $pitaThreshold = $thresholds{$file}->{'pita'};
    $final .= "ID\tMiranda Sites ($mirandaThreshold)\tRNAhybrid sites
($RNAhybThreshold)\tPita sites ($pitaThreshold)\tTarget description\n";
    while (my $line = <FILE>)
    {
        next if ($line =~ /^s$/); #skip blank lines
        next if ($line =~ /^miR/);
        chomp $line;
        if ($line =~ /^Targets/)
        {
            if (($line =~ /Miranda/) && ($line =~ /RNAhybrid/) && ($line =~ /Pita/))
            {
                $final .= "$file targets with at least 1 site above the threshold predicted
by Miranda, RNAhybrid and Pita:\n";
            }
            elsif (($line =~ /Miranda/) && ($line =~ /RNAhybrid/))
            {
                $final .= "$file targets with at least 1 site above the threshold predicted
by Miranda and RNAhybrid:\n";
            }
            elsif (($line =~ /Miranda/) && ($line =~ /Pita/))
            {
                $final .= "$file targets with at least 1 site above the threshold predicted
by Miranda and Pita:\n";
            }
            elsif (($line =~ /RNAhybrid/) && ($line =~ /Pita/))
            {
                $final .= "$file targets with at least 1 site above the threshold predicted
by RNAhybrid and Pita:\n";
            }
        }
        else
        {
            my $mirandaSites = $final{$file}->{$line}->{'miranda'} if ($final{$file}-
>{$line}->{'miranda'});
            $mirandaSites = 0 if (!$final{$file}->{$line}->{'miranda'});
            my $RNAhybSites = $final{$file}->{$line}->{'RNAhyb'} if ($final{$file}-
>{$line}->{'RNAhyb'});
            $RNAhybSites = 0 if (!$final{$file}->{$line}->{'RNAhyb'});
            my $pitaSites = $final{$file}->{$line}->{'pita'} if ($final{$file}->{$line}-
>{'pita'});
            $pitaSites = 0 if (!$final{$file}->{$line}->{'pita'});
            $final .=
"$line\t$mirandaSites\t$RNAhybSites\t$pitaSites\t$descriptions{$line}\n";
        }
    }
}
}

```

```

    open (FINAL, ">./PredictionFinalResults/$file");
    print FINAL "$final";
    $final = '';
}
close COMMON;
print "Remove unnecessary output files and directories...\n";
qx{rm -r ./mirnaTables};
qx{rm -r ./mirnaCommonTargets};

```

A.3: mirandaOutputFilterForVariableThreshold.pl

```

#!/usr/bin/perl -w

$|=1;
use strict;
use Getopt::Long;
my ($script) = ($0 =~ m|([^\s]*)$|);
my $Uso = "use of $script";

-mirandaOutSlimmed      miranda output file to be filtered (slimmed with
'mirandaOutputSlimmer.pl')
-filter                  threshold for filtering (binding energy)
-mirna                  mirna to select for filtering";

my ($mirandaOutSlimmed,$filter,$mirna);
my $result = GetOptions ("mirandaOutSlimmed=s"=>\$mirandaOutSlimmed,
                        "filter=i"=>\$filter,
                        "mirna=s"=>\$mirna);
unless ($result)
{
    print STDERR "Some options are not well formatted\n";
    exit (1);
}
if ((!$mirandaOutSlimmed)||(!$filter)||(!$mirna))
{
    print "$Uso";
    exit;
}

my @fields;
my $out = '';
open (FILE,"$mirandaOutSlimmed");
while (my $line = <FILE>)
{
    chomp $line;
    my $mirnaFormatted = $mirna;
    if ($mirnaFormatted =~ /\*$/)
    {
        $mirnaFormatted =~ s/\*/asteriscato/;
    }
    @fields = split (/\\t/, $line);
    ($fields[0]) = ($fields[0] =~ /(miR.+)/);
    if ($fields[0] =~ /\*$/)
    {
        $fields[0] =~ s/\*/asteriscato/;
    }
    if ($fields[0] eq $mirnaFormatted)
    {
        if ($fields[5] <= $filter)
        {
            $out .= "$line\n";
        }
    }
}
close FILE;
print "$out";

```

A.4: RNAhybridOutputFilterForVariableThreshold.pl

```
#!/usr/bin/perl -w

use strict;
use Getopt::Long;
my ($script) = ($0 =~ m|([^\s]*)$|);
my $Uso = "use of $script:

-RNAhybridOut          RNAhybrid output file (-c option)
-filter                energy cutoff (-#)
-mirna                 mirna to select for filtering";

my ($RNAhybridOut,$filter,$mirna);
my $result = GetOptions ("RNAhybridOut=s"=>\$RNAhybridOut,
                        "filter=i"=>\$filter,
                        "mirna=s"=>\$mirna);

unless ($result)
{
    print STDERR "Some options are not well formatted\n";
    exit (1);
}
if ((!$RNAhybridOut)||(!$filter)||(!$mirna))
{
    print "$Uso";
    exit;
}
#filter the output with a given energy threshold

my @fields;
my $out;
if ($mirna =~ /\*$/)
{
    $mirna =~ s/\*/asteriscato/; #sostituisco * con 'asteriscato', in modo che non venga
interpretato nel pattern matching
}
open (FILE,"$RNAhybridOut");
while (my $line = <FILE>)
{
    chomp $line;
    if ($line =~ /\:\/)
    {
        @fields = split (/\:\/,$line);
        if ($fields[2] =~ /\*$/)
        {
            $fields[2] =~ s/\*/asteriscato/;
        }
        if ($fields[2] eq $mirna)
        {
            if ($fields[4] <= $filter)
            {
                $out .= "$line\n";
            }
        }
    }
}
close FILE;
print $out;
```


A.5: pitaOutputFilterForVariableThreshold.pl

```
#!/usr/bin/perl -w

$|=1;
use strict;
use Getopt::Long;
my ($script) = ($0 =~ m|([^\s]*)$|);
my $Uso = "use of $script:
-pitaOut          pita output file to be reduced(*_targets.tab)
-threshold        maximum score allowed
-mirna            mirna name for filtering";

my ($pitaOut,$threshold,$mirna);
my $result = GetOptions ("pitaOut=s"=>\$pitaOut,
                        "threshold=i"=>\$threshold,
                        "mirna=s"=>\$mirna);

unless ($result)
{
    print STDERR "Some options are not well formatted\n";
    exit (1);
}
if ((!$pitaOut)||(!$threshold)||(!$mirna))
{
    print "$Uso";
    exit;
}
my (@fields);
my $ seq = '';
if ($mirna =~ /\*$/)
{
    $mirna =~ s/\*/asteriscato/;
}
open (FILE,"$pitaOut");
while (my $line = <FILE>)
{
    chomp $line;
    if ($line =~ /^RefSeq/)
    {
        $seq .= "$line\n";
        next;
    }
    @fields = split(/\t/, $line);
    next if ($fields[3] > $threshold);
    ($fields[1]) = ($fields[1] =~ /(miR.*));
    if ($fields[1] =~ /\*$/)
    {
        $fields[1] =~ s/\*/asteriscato/;
    }
    next if ($fields[1] ne $mirna);
    $seq .= "$line\n";
}
close FILE;
print $seq;
```


BIBLIOGRAFIA

1. Arndt GM, Dossey L, Cullen LM, Lai A, Druker R, Eisbacher M, Zhang C, Tran N, Fan H, Retzlaff K, Bittner A, Raponi M. Characterization of global microRNA expression reveals oncogenic potential of miR-145 in metastatic colorectal cancer. *BMC Cancer*. 2009 Oct 20;9:374.
2. Aronsohn A, Reau N: Long-term outcomes after treatment with interferon and ribavirin in HCV patients. *J Clin Gastroenterol*. 2009 Aug;43(7):661-71.
3. Bai S, Nasser MW, Wang B, Hsu SH, Datta J, Kutay H, Yadav A, Nuovo G, Kumar P, Ghoshal K. MicroRNA-122 inhibits tumorigenic properties of hepatocellular carcinoma cells and sensitizes these cells to sorafenib. *J Biol Chem*. 2009 Nov 13;284(46):32015-27. Epub 2009 Sep 2.
4. Bartenschlager R, Kaul A, Sparacio S. 2003. Replication of the hepatitis C virus in cell culture. *Antiviral Res*. 60:91–102.
5. Blenkiron C, Miska EA: miRNAs in cancer: approaches, aetiology, diagnostics and therapy. *Hum Mol Genet*. 2007 Apr 15;16 Spec No 1:R106-13.
6. Bouchard MJ, Schneider RJ. The enigmatic X gene of hepatitis B virus. *J Virol* 2004;78:12725-12734.
7. Brechot C. 2004. Pathogenesis of hepatitis B virus-related hepatocellular carcinoma: old and new paradigms. *Gastroenterology* 127:S56–61.
8. Buckwold VE, Beer BE, Donis RO. 2003. Bovine viral diarrhea virus as a surrogate model of hepatitis C virus for the evaluation of antiviral agents. *Antiviral Res*. 60:1–15.
9. Caldas C, Brenton JD: Sizing up miRNAs as cancer genes. *Nat Med*. 2005 Jul;11(7):712-4.
10. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti L, Kipps T, Negrini M, Bullrich F, Croce CM. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. 2002 Nov 26;99(24):15524-9. Epub 2002 Nov 14.
11. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M, Croce CM. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A*. 2004;101:2999-3004.

12. Chang J, Guo JT, Jiang D, Guo H, Taylor JM, Block TM: Liver-specific microRNA miR-122 enhances the replication of hepatitis C virus in nonhepatic cells. *J Virol*. 2008 Aug;82(16):8215-23.
13. Chisari FV, Ferrari C. 1995. Hepatitis B virus immunopathogenesis. *Annu. Rev. Immunol.* 13:29–60.
14. Cho WC: OncomiRs: the discovery and progress of microRNAs in cancers. *Mol Cancer*. 2007 Sep 25;6(1):60.
15. Connolly E, Melegari M, Landgraf P, Tchaikovskaya T, Tennant BC, Slagle BL, Rogler LE, Zavolan M, Tuschl T, Rogler CE. Elevated expression of the miR-17-92 polycistron and miR-21 in hepatitis B virus-associated hepatocellular carcinoma contributes to the malignant phenotype. *Am J Pathol*. 2008 Sep;173(3):856-64. Epub 2008 Aug 7.
16. Cullen, B. R.. Viruses and microRNAs. *Nat. Genet.* 2006; 38:S25-S30.
17. Davis N, Biddlecom N, Hecht D, Fogel GB. On the relationship between GC content and the number of predicted microRNA binding sites by MicroInspector. *Comput Biol Chem*. 2008 Jun;32(3):222-6. Epub 2008 Mar 10.
18. Diosdado B, van de Wiel MA, Terhaar Sive Droste JS, Mongera S, Postma C, Meijerink WJ, Carvalho B, Meijer GA. MiR-17-92 cluster is associated with 13q gain and c-myc expression during colorectal adenoma to adenocarcinoma progression. *Br J Cancer*. 2009 Aug 18;101(4):707-14.
19. Disson O, Haouzi D, Desagher S, Loesch K, Hahne M, et al. 2004. Impaired clearance of virus-infected hepatocytes in transgenic mice expressing the hepatitis C virus polyprotein. *Gastroenterology* 126:859–72.
20. Elmén J, Lindow M, Silahatoglu A, Bak M, Christensen M, Lind-Thomsen A, Hedtjærn M, Hansen JB, Hansen HF, Straarup EM, McCullagh K, Kearney P, Kauppinen S: Antagonism of microRNA-122 in mice by systemically administered LNA-antimiR leads to up-regulation of a large set of predicted target mRNAs in the liver. *Nucleic Acids Res*. 2008 Mar;36(4):1153-62. Epub 2007 Dec 23.
21. Fausto N. 2004. Liver regeneration and repair: hepatocytes, progenitor cells, and stem cells. *Hepatology* 39:1477–87.
22. Fornari F, Gramantieri L, Ferracin M, Veronese A, Sabbioni S, Calin GA, Grazi GL, Giovannini C, Croce CM, Bolondi L, Negrini M. MiR-221 controls CDKN1C/p57 and CDKN1B/p27 expression in human hepatocellular carcinoma. *Oncogene*. 2008 Sep 25;27(43):5651-61. Epub 2008 Jun 2.

23. Fukutomi T, Zhou Y, Kawai S, Eguchi H, Wands JR, Li J: Hepatitis C virus core protein stimulates hepatocyte growth: correlation with upregulation of wnt-1 expression. *Hepatology*. 2005 May;41(5):1096-105.
24. Ganem D, Prince AM. 2004. Hepatitis B virus infection—natural history and clinical consequences. *N. Engl. J. Med.* 350:1118–29.
25. Garofalo M, Di Leva G, Romano G, Nuovo G, Suh SS, Ngankeu A, Taccioli C, Pichiorri F, Alder H, Secchiero P, Gasparini P, Gonelli A, Costinean S, Acunzo M, Condorelli G, Croce CM. miR-221&222 regulate TRAIL resistance and enhance tumorigenicity through PTEN and TIMP3 downregulation. *Cancer Cell*. 2009 Dec 8;16(6):498-509.
26. Gerolami R, Uch R, Brechot C, Cannoni P, Bagnis C: Gene therapy of hepatocarcinoma: a long way from the concept to the therapeutical impact. *Cancer Gene Therapy* 2003; 10:640-660.
27. Girard M, Jacquemin E, Munnich A, Lyonnet S, Henrion-Caude A: miR-122, a paradigm for the role of microRNAs in the liver. *J Hepatol*. 2008 Apr;48(4):648-56.
28. Gramantieri L, Ferracin M, Fornari F, Veronese A, Sabbioni S, Liu CG, Calin GA, Giovannini C, Ferrazzi E, Grazi GL, Croce CM, Bolondi L, Negrini M: Cyclin G1 is a target of miR-122a, a microRNA frequently down-regulated in human hepatocellular carcinoma. *Cancer Res*. 2007 Jul 1;67(13):6092-9.
29. Guidotti LG, Chisari FV. 2001. Noncytolytic control of viral infections by the innate and adaptive immune response. *Annu. Rev. Immunol.* 19:65–91.
30. Guidotti LG, Chisari FV. Immunobiology and Pathogenesis of Viral Hepatitis. *Annu. Rev. Pathol. Mech. Dis.* 2006. 1:23–61.
31. Gusev Y: Computational methods for analysis of cellular functions and pathways collectively targeted by differentially expressed microRNA. *Methods* 2008 Jan; 44(1):61-72.
32. Habib NA, Mitry RR, Sarraf CE, Jiao LR, Havlik R, Nicholls J, Jensen SL: Assessment of growth inhibition and morphological changes in in vitro and in vivo hepatocellular carcinoma models post treatment with dl1520 adenovirus. *Cancer Gene Therapy* 2002; 9:414-420.
33. Henke JI, Goergen D, Zheng J, Song Y, Schüttler CG, Fehr C, Jünemann C, Niepmann M. microRNA-122 stimulates translation of hepatitis C virus RNA. *EMBO J*. 2008 Dec 17;27(24):3300-10. Epub 2008 Nov 20.

34. Herzer K, Weyer S, Krammer PH, Galle PR, Hofmann TG. Hepatitis C virus core protein inhibits tumor suppressor protein promyelocytic leukemia function in human hepatoma cells. *Cancer Res.* 2005 Dec 1;65(23):10830-7.
35. Hoofnagle JH. 2002. Course and outcome of hepatitis C. *Hepatology* 36:S21–29.
36. Huang J, Liu T, Xu LG, Chen D, Zhai Z, Shu HB. SIKE is an IKK epsilon/TBK1-associated suppressor of TLR3- and virus-triggered IRF-3 activation pathways. *EMBO J.* 2005 Dec 7;24(23):4018-28. Epub 2005 Nov 10.
37. Ishak K, Baptista A, Bianchi L, Callea F, De Groote J, Gudat F, Denk H, Desmet V, Korb G, MacSween RN, et al.: Histological grading and staging of chronic hepatitis. *J Hepatol.* 1995 Jun;22(6):696-9.
38. Jiang J, Gusev Y, Aderca I, Mettler TA, Nagorney DM, Brackett DJ, Roberts LR, Schmittgen TD: Association of MicroRNA expression in hepatocellular carcinomas with hepatitis infection, cirrhosis, and patient survival. *Clin Cancer Res.* 2008 Jan 15;14(2):419-27.
39. John B., Enright A.J., Aravin A., Tuschl T., Sander C., Marks D.S. Human MicroRNA targets. *PLoS Biol.* 2004;2:e363.
40. Jopling CL, Norman KL, Sarnow P. Positive and negative modulation of viral and cellular mRNAs by liver-specific microRNA miR-122. *Cold Spring Harb Symp Quant Biol.* 2006;71:369-76.
41. Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P: Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science.* 2005 Sep 2;309(5740):1577-81.
42. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E.: The role of site accessibility in microRNA target recognition. *Nat Genet.* 2007 Oct;39(10):1278-84. Epub 2007 Sep 23.
43. Kiriakidou M, Nelson PT, Kouranov A, Fitziev P, Bouyioukos C, Mourelatos Z, Hatzigeorgiou A. A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.* 2004 May 15;18(10):1165-78. Epub 2004 May 6.
44. Kota J, Chivukula RR, O'Donnell KA, Wentzel EA, Montgomery CL, Hwang HW, Chang TC, Vivekanandan P, Torbenson M, Clark KR, Mendell JR, Mendell JT. Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. *Cell.* 2009 Jun 12;137(6):1005-17.
45. Krüger J, Rehmsmeier M.: RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* 2006 Jul 1;34(Web Server issue):W451-4.

46. Lai EC. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet.* 2002 Apr;30(4):363-4. Epub 2002 Mar 18.
47. Lanford RE, Bigger C, Bassett S, Klimpel G. 2001. The chimpanzee model of hepatitis C virus infections. *Ilar. J.* 42:117–26.
48. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854.
49. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. *Cell.* 2003 Dec 26;115(7):787-98.
50. Li W, Xie L, He X, Li J, Tu K, Wei L, Wu J, Guo Y, Ma X, Zhang P, Pan Z, Hu X, Zhao Y, Xie H, Jiang G, Chen T, Wang J, Zheng S, Cheng J, Wan D, Yang S, Li Y, Gu J: Diagnostic and prognostic implications of microRNAs in human hepatocellular carcinoma. *Int J Cancer.* 2008 Oct 1;123(7):1616-22.
51. Liang TJ, Heller T. 2004. Pathogenesis of hepatitis C-associated hepatocellular carcinoma. *Gastroenterology* 127:S62–71.
52. Liu WH, Yeh SH, Lu CC, Yu SL, Chen HY, Lin CY, Chen DS, Chen PJ. MicroRNA-18a prevents estrogen receptor- α expression, promoting proliferation of hepatocellular carcinoma cells. *Gastroenterology.* 2009 Feb;136(2):683-93. Epub 2008 Nov 1.
53. Llovet JM, Fuster J, Bruix J: Intention-to-treat analysis of surgical treatment for early hepatocellular carcinoma: resection versus transplantation. *Hepatology* 1999; 30:1434-1440.
54. Meng F, Henson R, Wehbe-Janek H, Ghoshal K, Jacob ST, Patel T: MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology.* 2007 Aug;133(2):647-58. Epub 2007 May 21.
55. Mercatelli N, Coppola V, Bonci D, Miele F, Costantini A, Guadagnoli M, Bonanno E, Muto G, Frajese GV, De Maria R, Spagnoli LG, Farace MG, Ciafrè SA. The inhibition of the highly expressed miR-221 and miR-222 impairs the growth of prostate carcinoma xenografts in mice. *PLoS One.* 2008;3(12):e4029. Epub 2008 Dec 24.
56. Moschos SA, Williams AE, Perry MM, Birrell MA, Belvisi MG, Lindsay MA. Expression profiling in vivo demonstrates rapid changes in lung microRNA levels following lipopolysaccharide-induced inflammation but not in the anti-inflammatory action of glucocorticoids. *BMC Genomics.* 2007 Jul 17;8:240.

57. Murakami Y, Yasuda T, Saigo K, Urashima T, Toyoda H, Okanou T, Shimotohno K: Comprehensive analysis of microRNA expression patterns in hepatocellular carcinoma and non-tumorous tissues. *Oncogene*. 2006 Apr 20;25(17):2537-45.
58. Niepmann M. Activation of hepatitis C virus translation by a liver-specific microRNA. *Cell Cycle*. 2009 May 15;8(10):1473-7. Epub 2009 May 4.
59. Noonan EJ, Place RF, Pookot D, Basak S, Whitson JM, Hirata H, Giardina C, Dahiya R. miR-449a targets HDAC-1 and induces growth arrest in prostate cancer. *Oncogene*. 2009 Apr 9;28(14):1714-24. Epub 2009 Mar 2.
60. Pawlotsky JM. 2004. Pathophysiology of hepatitis C virus infection and related liver disease. *Trends Microbiol*. 12:96–102.
61. Pfeffer S, Sewer A, Lagos-Quintana M, Sheridan R, Sander C, Grasser FA, van Dyk LF, Ho CK, Shuman S, Chien M, Russo JJ, Ju J, Randall G, Lindenbach BD, Rice CM, Simon V, Ho DD, Zavolan M, Tuschl T. Identification of microRNAs of the herpesvirus family. *Nat Methods*. 2005; 2:269-76.
62. Pogribny IP, Filkowski JN, Tryndyak VP, Golubov A, Shpyleva SI, Kovalchuk O. Alterations of microRNAs and their targets are associated with acquired resistance of MCF-7 breast cancer cells to cisplatin. *Int J Cancer*. 2010 Jan 22. [Epub ahead of print]
63. Ramadori G, Saile B. 2004. Inflammation, damage repair, immune cells, and liver fibrosis: specific or nonspecific, this is the question. *Gastroenterology* 127:997–1000.
64. Rinaldi A, Poretti G, Kwee I, Zucca E, Catapano CV, Tibiletti MG, Bertoni F. Concomitant MYC and microRNA cluster miR-17-92 (C13orf25) amplification in human mantle cell lymphoma. *Leuk Lymphoma*. 2007 Feb;48(2):410-2.
65. Rusinov V, Baev V, Minkov IN, Tabler M. MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence. *Nucleic Acids Res*. 2005 Jul 1;33(Web Server issue):W696-700.
66. Sachdeva M, Mo YY. MicroRNA-145 suppresses cell invasion and metastasis by directly targeting mucin 1. *Cancer Res*. 2010 Jan 1;70(1):378-87. Epub 2009 Dec 8.
67. Schetter AJ, Heegaard NH, Harris CC. Inflammation and cancer: interweaving microRNA, free radical, cytokine and p53 pathways. *Carcinogenesis*. 2010 Jan;31(1):37-49. Epub 2009 Dec 2.
68. Scheuer PJ: Classification of chronic viral hepatitis: a need for reassessment. *J Hepatol*. 1991 Nov;13(3):372-4.

69. Seeger C, Mason WS. 2000. Hepatitis B virus biology. *Microbiol. Mol. Biol. Rev.* 64:51–68.
70. Shah YM, Morimura K, Yang Q, Tanabe T, Takagi M, Gonzalez FJ: Peroxisome proliferator-activated receptor alpha regulates a microRNA-mediated signalling cascade responsible for hepatocellular proliferation. *Mol Cell Biol.* 2007 Jun;27(12):4238-47. Epub 2007 Apr 16.
71. Shoukry NH, Cawthon AG, Walker CM. 2004. Cell-mediated immunity and the outcome of hepatitis C virus infection. *Annu. Rev. Microbiol.* 58:391–424.
72. Slaby O, Svoboda M, Fabian P, Smerdova T, Knoflickova D, Bednarikova M, Nenutil R, Vyzula R. Altered expression of miR-21, miR-31, miR-143 and miR-145 is related to clinicopathologic features of colorectal cancer. *Oncology.* 2007;72(5-6):397-402. Epub 2008 Jan 15.
73. Stahlhut Espinosa CE, Slack FJ : The Role of MicroRNAs in Cancer. *Yale J Biol Med.* 2006 Dec;79(3-4):131-40.
74. Storey JD. A direct approach to false discovery rates. *J R Stat Soc* 2002, 3:479-498.
75. Takahara Y, Takahashi M, Zhang QW, Wagatsuma H, Mori M, Tamori A, Shiomi S, Nishiguchi S. Serial changes in expression of functionally clustered genes in progression of liver fibrosis in hepatitis C patients. *World J Gastroenterol.* 2008 Apr 7;14(13):2010-22.
76. Takakura S, Mitsutake N, Nakashima M, Namba H, Saenko VA, Rogounovitch TI, Nakazawa Y, Hayashi T, Ohtsuru A, Yamashita S. Oncogenic role of miR-17-92 cluster in anaplastic thyroid cancer cells. *Cancer Sci.* 2008 Jun;99(6):1147-54. Epub 2008 Apr 21.
77. Thimme R, Wieland S, Steiger C, Ghayeb J, Reimann KA, et al. 2003. CD8(+) T cells mediate viral clearance and disease pathogenesis during acute hepatitis B virus infection. *J. Virol.* 77:68–76.
78. Tong AW and J Nemunaitis. Modulation of miRNA activity in human cancer: a new paradigm for cancer gene therapy? *Cancer Gene Therapy* 2008; 15:341-355.
79. Tsai NP, Lin YL, Wei LN. MicroRNA mir-346 targets the 5'-untranslated region of receptor-interacting protein 140 (RIP140) mRNA and up-regulates its protein expression. *Biochem J.* 2009 Dec 10;424(3):411-8.
80. Ura S, Honda M, Yamashita T, Ueda T, Takatori H, Nishino R, Sunakozaka H, Sakai Y, Horimoto K, Kaneko S. Differential microRNA expression between hepatitis B and

- hepatitis C leading disease progression to hepatocellular carcinoma. *Hepatology*. 2009 Apr;49(4):1098-112.
81. van Haaften G, Agami R. Tumorigenicity of the miR-17-92 cluster distilled. *Genes Dev*. 2010 Jan 1;24(1):1-4.
 82. Varnholt H, Drebber U, Schulze F, Wedemeyer I, Schirmacher P, Dienes HP, Odenthal M. MicroRNA gene expression profile of hepatitis C virus-associated hepatocellular carcinoma. *Hepatology*. 2008 Apr;47(4):1223-32.
 83. Venugopal SK, Jiang J, Kim TH, Li Y, Wang SS, Torok NJ, Wu J, Zern MA. Liver fibrosis causes down-regulation of miRNA-150 and miRNA-194 in hepatic stellate cells and their over-expression causes decreased stellate cell activation. *Am J Physiol Gastrointest Liver Physiol*. 2009 Nov 5.
 84. Visone R, Russo L, Pallante P, De Martino I, Ferraro A, Leone V, Borbone E, Petrocca F, Alder H, Croce CM, Fusco A. MicroRNAs (miR)-221 and miR-222, both overexpressed in human thyroid papillary carcinomas, regulate p27Kip1 protein levels and cell cycle. *Endocr Relat Cancer*. 2007 Sep;14(3):791-8.
 85. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, Harris CC, Croce CM. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A*. 2006 Feb 14;103(7):2257-61. Epub 2006 Feb 3.
 86. Wong QW, Ching AK, Chan AW, Choy KW, To KF, Lai PB, Wong N. MiR-222 Overexpression Confers Cell Migratory Advantages in Hepatocellular Carcinoma through Enhancing AKT Signaling. *Clin Cancer Res*. 2010 Jan 26. [Epub ahead of print]
 87. Xu T, Zhu Y, Xiong Y, Ge YY, Yun JP, Zhuang SM. MicroRNA-195 suppresses tumorigenicity and regulates G1/S transition of human hepatocellular carcinoma cells. *Hepatology*. 2009 Jul;50(1):113-21.
 88. Zhou L, Qi X, Potashkin JA, Abdul-Karim FW, Gorodeski GI. MicroRNAs miR-186 and miR-150 down-regulate expression of the pro-apoptotic purinergic P2X7 receptor by activation of instability sites at the 3'-untranslated region of the gene that decrease steady-state levels of the transcript. *J Biol Chem*. 2008 Oct 17;283(42):28274-86. Epub 2008 Aug 5.
 89. Zhou X, Duan X, Qian J, Li F. Abundant conserved microRNA target sites in the 5'-untranslated region and coding sequence. *Genetica*. 2009 Nov;137(2):159-64. Epub 2009 Jul 4.