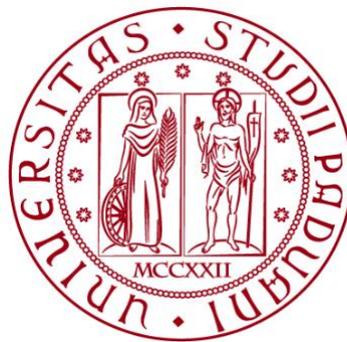


**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**DIPARTIMENTO DI BIOLOGIA**

**Corso di Laurea in Biotecnologie**



**ELABORATO DI LAUREA**

**Un RNA non codificante controlla il  
differenziamento muscolare interagendo con  
specifici microRNA nel citoplasma**

**Tutor: Prof. Gerolamo Lanfranchi  
Dipartimento di Biologia**

**Laureanda: Francesca Pertile  
ANNO ACCADEMICO 2022/2023**

## SOMMARIO

|  |           |
|--|-----------|
| <b>SOMMARIO</b> .....  | <b>2</b>  |
| <b>ABSTRACT</b> .....  | <b>3</b>  |
| <b>INTRODUZIONE</b> .....  | <b>3</b>  |
| <i>Long non coding RNA</i> .....   | <b>4</b>  |
| <i>Biogenesi dei lncRNA</i> .....  | <b>5</b>  |
| <i>Funzioni dei lncRNA</i> .....   | <b>6</b>  |
| <i>Competing endogenous RNA</i> .....  | <b>8</b>  |
| <b>SVILUPPO MUSCOLARE E MICRORNA</b> .....   | <b>9</b>  |
| <b>METODI</b> .....  | <b>12</b> |
| <i>Tecniche</i> .....  | <b>12</b> |
| <i>Chromatin immunoprecipitation (ChIP)</i> .....  | <b>12</b> |
| <i>Analisi 3C</i> .....  | <b>13</b> |
| <b>RISULTATI</b> .....   | <b>13</b> |
| <i>linc-MD1 è espresso durante il differenziamento muscolare</i> .....   | <b>13</b> |
| <i>Identificazione degli elementi regolatori che dirigono l'espressione di linc-MD1 e miR-206/133b: esperimenti di fusione del promotore e saggi in vivo</i> ..... | <b>15</b> |
| <i>Regolazione di linc-MD1 ha un ruolo nella differenziazione miogenica</i> .....  | <b>17</b> |
| <i>linc-MD1 è un target di miR-133 e miR-135</i> .....   | <b>18</b> |
| <i>linc-MD1 e l'espressione dei target di miR-133 e miR-135: in topo e nell'uomo</i> .....   | <b>19</b> |
| <b>CONCLUSIONI</b> .....   | <b>21</b> |
| <b>BIBLIOGRAFIA</b> .....  | <b>22</b> |

**ABSTRACT.** Grazie alle nuove tecnologie di analisi del trascrittoma è stato visto come il genoma dei mammiferi viene trascritto in diverse famiglie di RNA tra cui i long non coding RNA (lncRNA). Sebbene siano poco espressi nei diversi compartimenti del corpo grazie a tecnologie di rilevazione sensibili è stato possibile studiarli per identificarne il ruolo e la funzione. Uno di questi linc-MD1 ha un ruolo nella differenziazione dei muscoli, dove è in grado di competere con microRNA (miRNA) che interagiscono con fattori di trascrizione miogenici: infatti viene definito *competing endogenous RNA*. Questo sistema è stato scoperto grazie allo studio del processo di differenziamento muscolare in cui linc-MD1 agisce come spugna molecolare per miR-133 e miR-135 per regolare l'espressione dei geni codificanti 2 fattori di trascrizione MAML1 e MEF2C, che hanno un ruolo nell'espressione muscolo specifica. I livelli di espressione del ceRNA sono legati alle fasi differenziamento dei mioblasti, ma anche la comparsa di distrofia di Duchenne, una miopatia. Gli studiosi hanno deciso di utilizzare per lo studio delle linee cellulari murine estratte da mioblasti 2C. Grazie all'alto livello di omologia, i risultati sono stati poi confrontati con degli studi fatti in linee cellulari umane.

## INTRODUZIONE

Il concetto di genoma venne coniato per la prima volta da un botanico Hans Winkler nel 1920 che lo definì come "l'insieme dei cromosomi delle cellule aploidi degli Eucarioti localizzato esclusivamente nel nucleo". Per anni questa definizione venne usata solo in botanica e associata al numero di cromosomi.

In seguito, poi, grazie ai grandi ricercatori, pionieri della Biologia molecolare, *in primis* Watson e Crick, le ricerche non si limitarono solo alla struttura del genoma, ma anche alla sua funzione. Il genoma venne quindi definito come un insieme di geni, unità fondamentali alla base della trascrizione del DNA in RNA e della traduzione di RNA in proteine.

Definita la struttura primaria dei genomi, si fecero numerosi studi qualitativi e quantitativi che diventarono man mano maggiormente approfonditi con l'evoluzione delle tecnologie della genomica funzionale che riguardarono genomi di diversi organismi, compreso l'uomo. Con il completamento dei primi progetti di sequenziamento sistematico pubblicati nel 2001 dalle riviste scientifiche *Nature* e *Science* si poté stabilire che il genoma umano è costituito da 3,1 miliardi di paia di basi.

Già negli anni 70 e successivamente codificata da Ohno venne teorizzata la presenza del "*junk DNA*" o DNA spazzatura, cioè parti del genoma non codificanti proteine, che costituiscono il 98% del materiale genetico. Lo studio dello scienziato giapponese costituì uno spartiacque che ha portato e sta portando ogni anno a nuove scoperte sul ruolo dei diversi elementi funzionali che invece si sono rivelati presenti in questo "Junk DNA". All'interno di questi elementi, troviamo geni che codificano per due categorie di RNA non codificanti proteine: i long non coding RNA (lncRNA) e i microRNA (miRNA).

## Long non coding RNA

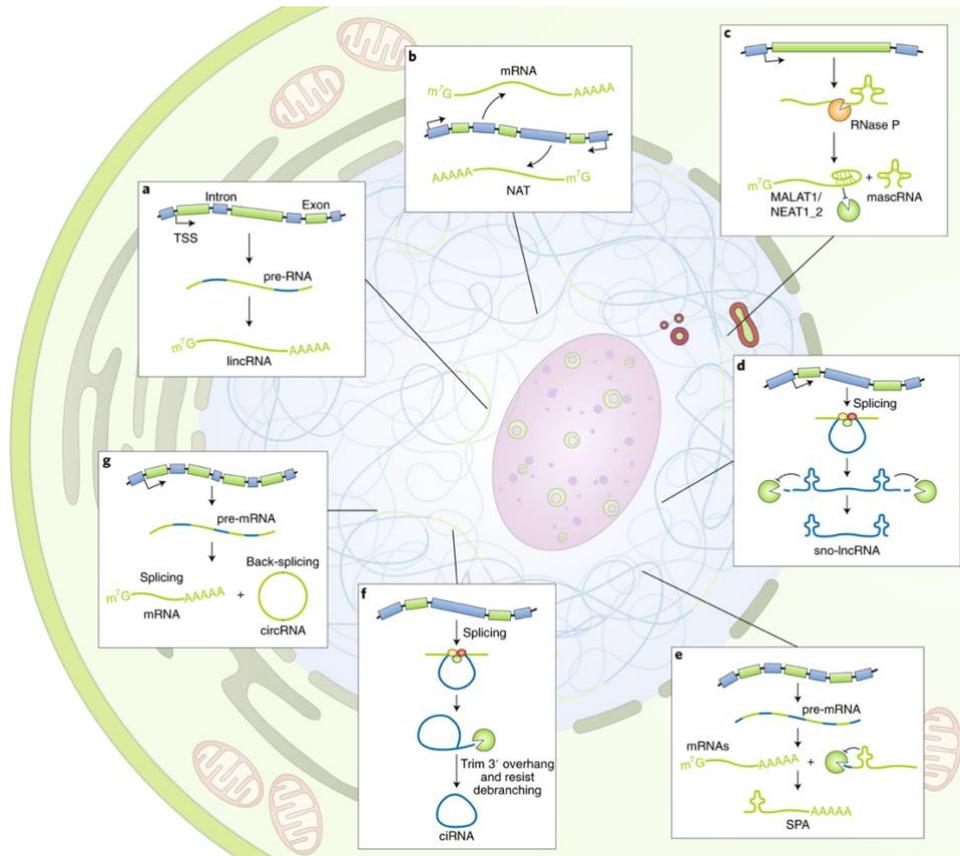
Una parte consistente dei geni contenuti nel genoma dei mammiferi non è codifica proteine, ma porta alla produzione di RNA non codificanti. Tra questi, un numero crescente codifica per RNA non codificanti “lunghi”, i cosiddetti lncRNA. Per lncRNA si intendono frammenti di RNA lunghi più di 200 nucleotidi che non contengono evidenti ORF (*open reading frame*) tradotte in proteine.

Per quanto riguarda la trascrizione di questi geni per lncRNA, anche se le RNA polimerasi trascrivono preferenzialmente un solo filamento del DNA iniziando di solito a livello di uno specifico TSS (*transcription start site*), si è visto che la loro trascrizione può iniziare anche a una certa distanza dal promotore ed eventualmente in direzione opposta, oppure può iniziare da siti che pur non essendo promotori funzionali, presentano una qualche omologia di sequenza con essi e quindi con una certa propensione alla trascrizione (Amaldi et al., 2018).

Nel genoma umano, sono stati identificati finora circa 16.000 geni (fonte Human GENCODE) per lncRNA che hanno una struttura generalmente simile ai geni strutturali che codificano per RNA messaggeri. I trascritti prodotti da questi geni sono strutturalmente simili ad mRNA canonici quando i geni che li producono sono collocati come unità indipendenti in spazi intergenici. lncRNA possono essere trascritti in antisenso a geni codificanti proteine a partire da promotori alternativi con orientamento opposto. Esistono poi lncRNA non convenzionali trascritti dall'RNA polimerasi II, ma la cui stabilizzazione avviene a seguito del taglio nella regione 3'-terminale da parte RNAsi P che produce una tripla elica UAU o formando strutture circolari, chiuse covalentemente per evitare che vengano degradati una volta esportati al di fuori del nucleo.

Alcuni elementi regolatori della trascrizione come *enhancer* e promotori possono favorire la trascrizione bidirezionale da parte dell'RNA pol II producendo *enhancer* RNA (eRNA) oppure trascritti a monte del promotore (PROMPT). Gli eRNA partecipano alle modificazioni della cromatina che sono indotte dalle sequenze enhancer, mentre i PROMPT non hanno ancora una funzione definita. La degradazione di PROMPT è associata con la scelta della direzionalità del promotore. Talvolta, la trascrizione dei geni per lncRNA può avvenire da parte di ~~altre~~ RNA polimerasi diverse dall'RNAPol II (Statello et al., 2022).

I trascritti prodotti dai geni per lncRNA possiedono un'estremità modificata al terminale 5' con 7-metilguanosa, una coda poliadenilica all'estremità opposta 3' e sono sottoposti a splicing in modo simile ai canonici mRNA.



**Figura 1** Diverse tipologie di lncRNA nelle cellule dei mammiferi: (a) RNA intergenici lunghi non codificanti (lincRNA) solitamente che sono sottoposti a capping, poliadenilazione e splicing; (b) trascritti antisense rispetto al gene che codifica proteine; (c) esempio di trascritto MALAT1 e NEAT1\_2 stabilizzato in 3' da tripla elica UAU; (d) ed (e) lncRNA processati dal complesso snoRNA; (f) RNA circolari intronici; (g) RNA circolari (Yao et al., 2019).

Si è visto che almeno un ristretto numero di lncRNA può contenere brevi ORF che vengono tradotte in piccoli polipeptidi. Recentemente è stato dimostrato che molti lncRNA influenzano lo stato della cromatina agendo sull'espressione di geni vicini, ma lontani rispetto al *locus* che permette l'espressione del lncRNA.

È stato visto anche come questi elementi possono avere anche una funzione strutturale andando a stabilizzare i complessi ribonucleoproteici nucleari o citoplasmatici.

Recenti evidenze sperimentali hanno dimostrato che la loro espressione può essere aumentata o diminuita in diverse patologie come ad esempio nei tumori: per questo motivo sono stati usati come biomarcatori o target per lo sviluppo di terapie antitumorali (Statello et al., 2022. Yao et al., 2019).

## Biogenesi dei lncRNA

I geni codificanti lncRNA sono molto meno conservati rispetto a quelli codificanti proteine, presentano meno esoni e sono generalmente espressi a livelli bassi

nelle cellule ~~espressi~~. Questa bassa espressione sembra dovuta alla presenza di modificazioni repressive degli istoni cromatinici presenti nei nucleosomi dei loro promotori. Le modalità di trascrizione dei lncRNA presentano delle peculiarità. Il dominio carbossi-terminale della RNA polimerasi II (RNAPII) presenta diversi stati di fosforilazione che sono legati ai diversi momenti della trascrizione in cui l'enzima è coinvolto, anche se un numero cospicuo di questi RNA è tradotto da una polimerasi in cui lo stato di fosforilazione della polimerasi non è ben regolato. I lncRNA sembrano non avere un processo di splicing contemporaneo alla trascrizione e il segnale di poliadenilazione è indipendente dal processo, portando all'accumulo di trascritti immaturi a livello della cromatina che possono essere successivamente degradati. Questo spiega la predominante localizzazione nucleare dei lncRNA, che spesso hanno sviluppato delle strategie per evadere la sorveglianza nucleare. Alcuni di questi trascritti possono anche legarsi direttamente alla cromatina con un meccanismo che coinvolge il piccolo RNA nucleare (snoRNA) U1.

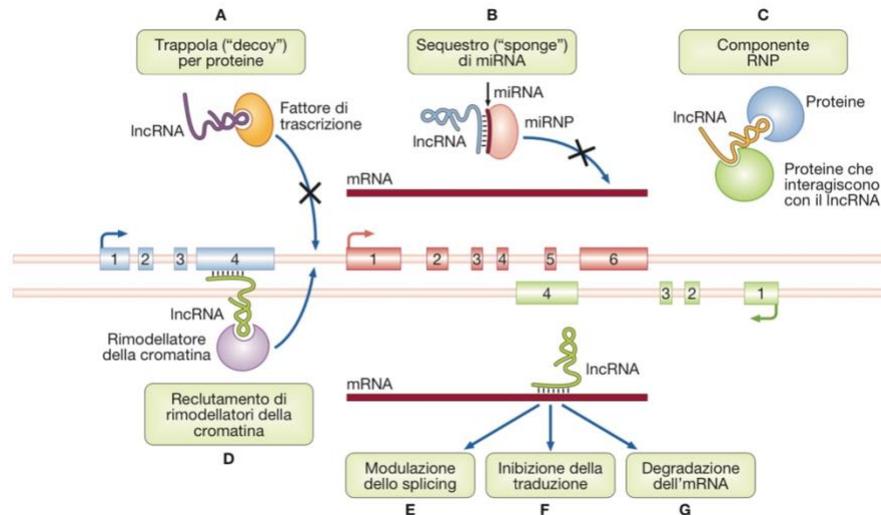
Nel complesso, i lncRNA subiscono lo *splicing* in maniera meno efficiente degli mRNA a causa dei segnali di splicing ~~poche~~ deboli nella loro sequenza e alla presenza di una distanza maggiore tra il sito di *splicing* in 3' e il punto di ramificazione. Questa struttura contribuisce anche alla ritenzione nucleare che è anche legata ad altri fattori come la presenza di diversi regolatori dello *splicing*. Accanto a queste caratteristiche, alcuni lncRNA contengono nella loro ~~stessa~~ sequenza motivi che possono reclutare fattori nucleari che favoriscono ulteriormente la loro localizzazione nucleare e la loro funzione. In sintesi, la localizzazione nucleare e il destino di questi RNA non codificanti è regolato a livello trascrizionale, grazie alla presenza di sequenze regolatrici in *cis* ed ~~elementi~~ in *trans*. Le ragioni per le quali i lncRNA siano legati solamente a specifici domini nucleari sono in gran parte ancora da decifrare ~~ancora sconosciute~~.

Una parte consistente di lncRNA è esportata nel citosol dove probabilmente condividono gli stessi processi a cui sono sottoposti gli mRNA. Una volta giunti nel citoplasma, infatti, sono distribuiti tra i vari organelli cellulari grazie anche alla presenza di proteine che li legano e li indirizzano. I lncRNA possono anche essere dirottati verso i mitocondri.

La comprensione della localizzazione dei diversi lncRNA nei diversi compartimenti della cellula è ancora molto limitata (Statello et al., 2022. Yao et al., 2019, Amaldi et al., 2018 )

## Funzioni dei lncRNA

I lncRNA possono svolgere un ruolo nella regolazione dell'espressione genica a diversi livelli sia nel nucleo che nel citoplasma.



**Figura 2 Funzioni lncRNA:** (A) sequestro di fattori di trascrizione; (B) competizione e sequestro come "spugne" di microRNA; (C) formazione di complessi RNA-proteine (RNP); (D) reclutamento di rimodellatori e modificatori della cromatina; (E) modulazione dello splicing; (F) blocco del mRNA ed inibizione della traduzione; (G) avvio del mRNA verso la degradazione. (Amaldi et al., 2018).

Essi possono interagire con elementi che vanno a modificare la cromatina promuovendoli o sequestrandoli. In questo caso, i lncRNA agiscono come "esche" (*decoy*) sequestrando proteine regolative e fattori di trascrizione impedendo la loro interazione con il DNA (Figura 2-A).

Un classico esempio di lncRNA che agisce sulla conformazione della cromatina è costituito da Xist (*X inactive specific transcript*), responsabile dell'inattivazione del cromosoma X a livello nucleare. Una volta trascritto, Xist innesca una cascata di segnali che culmina con il rimodellamento della cromatina di una copia del cromosoma X il cui risultato è la repressione trascrizionale della maggior parte dei geni ospitati da quel cromosoma. Il cromosoma X così condensato forma il corpuscolo di Barr che solitamente è localizzato a livello della membrana perinucleare e della lamina nucleare. È stato visto come a livello della lamina si trovi un recettore associato a Xist. Quando questa interazione non avviene nella fase embrionale precoce, l'inattivazione del cromosoma viene compromessa, dimostrando il ruolo del trascritto Xist nel cambiamento dello stato della cromatina. Questo trascritto è antagonizzato da un altro lncRNA: Tsix che viene trascritto in direzione opposta a Xist. Tsix agisce legandosi in maniera complementare a Xist inibendo la sua azione di silenziamento.

I lncRNA possono fungere da adattatori (*scaffold*) per la formazione di complessi ribonucleoproteici che agiscono come modificatori della cromatina alterandone lo stato trascrizionale (Figura 2-B e 2-C). Un esempio di questo meccanismo d'azione è il complesso multiproteico Polycomb che silenzia i geni bersaglio portando la cromatina ad uno stato condensato. Polycomb si lega infatti a lncRNA che favoriscono la sua azione repressiva a livello della cromatina. Un altro esempio di questo meccanismo riguarda gli RNA trascritti dalle sequenze ripetute della famiglia Alu che sopprimono l'azione della RNA polimerasi II durante la trascrizione degli mRNA.

Molti lncRNA agiscono da guida (*guide*) indirizzando la localizzazione di enzimi modificatori della cromatina e del DNA come le DNA metiltransferasi sui loro siti bersaglio. Questi enzimi metilando il DNA, contribuiscono al silenziamento trascrizionale della cromatina. I lncRNA guida possono agire sia in *cis* che in *trans*. I primi agiscono con un controllo nella regione che si trova nelle vicinanze del sito in cui l'RNA viene trascritto, mentre i secondi attivano o reprimono l'espressione di geni in *loci* indipendenti.

Alcuni lncRNA possono interagire con delle sequenze *enhancer*, attivando in *trans* la trascrizione di un gene. Possono anche legarsi ad un trascritto primario interferendo con il meccanismo dello *splicing*, oppure inibendone la traduzione o avviando la sua degradazione.

Alcuni lncRNA possono agire come ceRNA (*competing endogenous RNA*) entrando in competizione per il legame con specifici miRNA (azione "spugna"). I lncRNA sono capaci di interagire con il DNA genomico formando strutture ibride (solitamente triple eliche o R-loops). Le triple eliche RNA-DNA-DNA sono delle strutture che solitamente sono associate al silenziamento o all'attivazione di un gene.

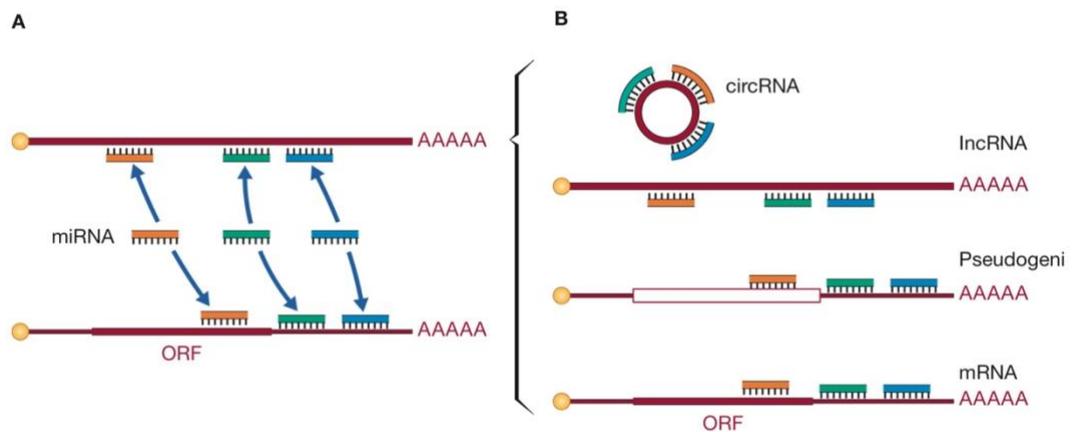
Alcuni lncRNA trascritti in antisenso possono invece regolare la trascrizione formando R-loop. Gli R-loop sono delle strutture composte da un ibrido DNA-RNA associato a un DNA a singolo filamento. Questa struttura può sequestrare cofattori che si vanno a legare alle regioni promotoriali vicine. È stato visto che lncRNA della categoria dei eRNA abbiano un ruolo nell'attivazione di SPHK1, un protooncogene che localizza nelle vicinanze di un *enhancer* coinvolto nella sua regolazione trascrizionale.

Le recenti osservazioni sul ruolo dei lncRNA e, in generale, degli RNA non codificanti nella regolazione fine dell'espressione del genoma suggeriscono le molteplici e complicate interazioni tra RNA codificanti e non codificanti in numerose e complesse reti regolative del trascrittoma. (Statello et al., 2022. Yao et al., 2019, Amaldi et al., 2018).

## Competing endogenous RNA

I *competing endogenous RNA* (ceRNA) sono RNA spugna che sequestrano specifici miRNA o trascritti senza open reading frame (Figura 3 A).

I ceRNA hanno un ruolo nel *cross-talk* in malattie complesse per il legame dei miRNA. Agiscono come spugne molecolari attraverso le sequenze complementari ai miRNA (MRE), andando a regolare indirettamente altri geni bersaglio di questi specifici miRNA. È stato dimostrato il ruolo di questo *cross-talk* nello sviluppo del cancro e in malattie complesse dell'uomo.



**Figura 3 Come funzionano ceRNA (A)** interazione tra miRNA e mRNA e ncRNA; **(B)** lncRNA, circRNA, pseudogeni e mRNA possiedono sequenze complementari a miRNA e quindi hanno un ruolo nella loro regolazione (Amaldi et al., 2018).

Un numero crescente di studi sul ruolo dei ceRNA in specifici tumori sta costruendo una visione complessiva del ruolo dei ceRNA nella genesi e progressione di queste patologie. Nel lavoro di *Statello et al.* sono confrontati campioni di diversi tumori con controlli normali per analizzare la regolazione del rapporto tra mRNA e ceRNA. I risultati si sono rivelati molto utili non solo perché hanno permesso di identificare il ruolo dei ceRNA, ma anche implicitamente il ruolo dei miRNA nella regolazione dell'oncogenesi.

Anche alcuni lncRNA possono svolgere una funzione simile a quella dei ceRNA. Ad esempio, linc-MD1 sembra svolgere un ruolo ceRNA-simile nella differenziazione muscolare, come illustrerò più avanti.

Recentemente nelle cellule neuronali è stata identificata una nuova classe di RNA non codificanti circolari che, come i ceRNA e i lncRNA possono agire come spugne per i miRNA. Al contrario di questi, che, avendo emi-vita breve permettono un controllo rapido e variabile della traduzione, i circRNA mostrano invece una maggiore stabilità e loro turn-over è regolato dall'appaiamento perfetto con un miRNA che causa il taglio endonucleotidico della molecola (Figura 3B) (Statello et al., 2022. Yao et al., 2019, Amaldi et al., 2018).

## SVILUPPO MUSCOLARE E MICRORNA

Lo sviluppo muscolare determina fisiologicamente il numero di fibre, la massa muscolare, e il tipo di fibre muscolari ma anche alcuni cambiamenti patologici che possono colpire la muscolatura scheletrica. Il processo può essere diviso in 4 stadi:

1. Allo stadio embrionale le cellule muscolari si originano dal mesoderma e sono poi divise dopo la formazione in masse a forma di blocchi. Queste strutture sono dette mesomeri che portano alla formazione dei dermatomi da cui si originano i miotomi.
2. A partire dai miotomi per differenziazione si formano i mioblasti che poi migrano verso le regioni che porteranno allo sviluppo degli arti dove formano le masse ventrali e dorsali.



I myomiR possono anche avere un ruolo nella regolazione post-trascrizionale: questo è ad esempio il caso di miR-135 che va ad inibire MEF2C arrestando lo sviluppo dei miociti (Wen Luo et al., 2013, Chal J et al.,2017).

## METODI

Lo studio descritto in questa tesi ha lo scopo di identificare il ruolo di linc-MD1 nella regolazione del differenziamento muscolare: infatti questo lncRNA interviene anticipando o ritardando il differenziamento cellulare interagendo con altri RNA e, in particolare, con miRNA. A questo proposito linc-MD1 può essere definito come *competing endogenous RNA* perché entra in competizione con alcuni miRNA agendo come spugna e bloccando il loro legame ai trascritti di due geni che codificano per fattori di trascrizione muscolari. Infine, è stato visto che il *cross talk* tra i vari RNA ha un ruolo non solo nella tempistica del differenziamento, ma è legato anche alla comparsa di miopatie come la distrofia di Duchenne. Modelli animali e cellulari.

Nel progetto sono stati utilizzati mioblasti primari di topo C2, derivati dal muscolo scheletrico di topi sani di 2 mesi prelevati dopo 3 giorni dall'induzione di un danno. Queste cellule sono state trasfettate con i rispettivi reagenti con un plasmide che esprime uno *small interfering RNA* (siRNA) che è in grado di legare gli esoni 2 e 3 di linc-MD1, per cui miR-133a/b e miR-135a/b non sono espressi. Tutte le sequenze sono state progettate attraverso uno studio approfondito sul ruolo dei trascritti in esame.

Sono stati utilizzati anche mioblasti Duchenne con delezione dell'esone 44 e di controllo che sono stati fatti crescere con vettori lentivirali contenenti la sequenza di controllo e la sequenza MD1.

Queste linee sono state affiancate da cellule muscolari satelliti umane e murine per comprendere il ruolo di questi RNA non codificanti in questa patologia neuromuscolare. Infine, sono stati utilizzati anche topi mdx in cui è presente una mutazione puntiforme nell'esone 23 del gene distrofina che rende questi animali dei modelli consolidati della distrofia di Duchenne. (Cesana et al., *Nature*)

## Tecniche

### Chromatin immunoprecipitation (ChIP)

Le analisi ChIP sono state effettuate su estratti di mioblasti (GM) e miotubi (DM). L'immunoprecipitazione della cromatina, o ChIP, è una tecnologia basata su anticorpi utilizzata per arricchire selettivamente specifiche proteine leganti il DNA insieme alle loro sequenze bersaglio. La ChIP viene utilizzata per studiare una singola interazione proteina-DNA, diverse interazioni proteina-DNA con un'analisi allargata all'intero genoma oppure ristretta ad un sottoinsieme di geni o a singoli loci. Nella tecnologia ChIP adottata in questo lavoro, sono stati utilizzati anticorpi specifici per l'RNA polimerasi II, per il fattore MyoD e anti-acetyl-HistoneH3 (Lys9), anti-acetyl-HistoneH3 (Lys4), e anti-trimethyl Histone H3 (Lys27).

I frammenti di DNA che sono stati ottenuti con la precipitazione CHIP, sono stati poi analizzati tramite PCR quantitativa con l'utilizzo di una coppia di primer disegnati appositamente e poi sono stati confrontati con uno standard (Cesana et al.).

## **Analisi 3C**

La *chromosome conformation capture* (3C) è una tecnica che permette l'analisi dell'organizzazione della cromatina nelle cellule. Questo metodo permette di evidenziare e quantificare l'interazione tra *loci* diversi del genoma. Sono stati usati 2 metodi di controllo: il primo per vedere se l'efficienza dei primer alterasse i risultati dell'esperimento; il secondo per monitorare la quantità di DNA che era stata usata. È stato confermato che ogni primer era in grado di amplificare una sequenza di controllo artificiale, ma non la cromatina digerita o legata (Cesana et al.)

## **RISULTATI**

### **linc-MD1 è espresso durante il differenziamento muscolare**

Le regioni che codificano i miRNA si trovano per il 50% in porzioni esoniche o introniche del genoma codificante proteine, mentre l'altra metà in geni che ospitano non coding RNA come nel caso dei pre-miR-206 e pre-miR-133. Questi 2 miRNA muscolari sono stati analizzati con la tecnica 5'RACE (*rapid amplification of cDNA ends*). Questa tecnica utilizza 2 primer: uno basato sulla sequenza di DNA nota e un secondo che è basato sugli oligonucleotidi che sono aggiunti in 5' alla sequenza da analizzare. Il frammento viene amplificato per PCR e poi segue il sequenziamento RACE-seq. Questa analisi ha dimostrato la presenza in un sito di inizio di trascrizione (TSS) 600 bp a monte rispetto alla sequenza pre-miR-206 con delle E-box, una regione legante proteine che ha un ruolo nella espressione genica. Per miR-133 è stato identificato solo un TSS 133 Kb a monte della regione codificante l'elemento (Figura 5B).

L'analisi della regione genomica ha individuato 2 introni e 3 esoni: nel secondo introne si trova miR-206, mentre miR-133 nel terzo esone. Nessuna sequenza AUG nelle vicinanze presenta la sequenza consensus Kozak o presenta delle regioni che sono conservate confermando la tesi per cui non è una sequenza codificante. Il trascritto viene quindi definito RNA intergenico non codificante MD1 (linc-MD1).

Grazie ad analisi filogenetiche di linc-MD1 è stato visto come siano conservati gli esoni 1 e 2, mentre l'omologia è limitata nella sequenza pre-miR-133b nell'esone 3. Le sequenze introniche, invece, che sono state sottoposte a splicing sono conservate. L'analisi *in silico* ha evidenziato la presenza di sequenze E-box

conservate nella regione prossimale e distale come nella regione dell'esone 2 dove si trovano meno siti di trascrizione alternativi (Cesana et al.).

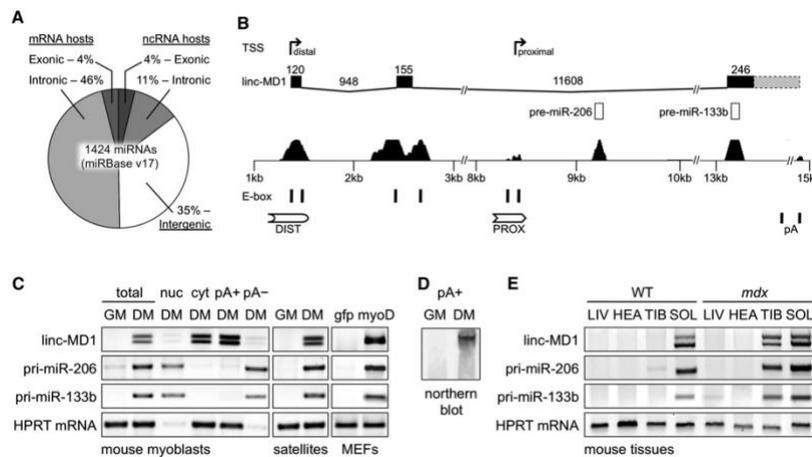
Grazie a RT-PCR è stato visto come linc-MD1 si trovi nel citoplasma e sia poliadenilato. Risulta assente in condizione di crescita. Quando i mioblasti murini entrano in stadio di differenziamento si alzano i livelli di linc-MD1, ma è presente anche in cellule satellite e in fibroblasti transdifferenziati da MyoD, una proteina che fa parte dei fattori di differenziazione miogenica. L'espressione di linc-MD1 avviene durante la differenziazione parallelamente a miR-133 ed è distinta da miR-206 già presente in linee C2 di mioblasti proliferanti. Nell'esone 2 grazie alla RT-PCR è stata individuata una variante di splicing di 70 nucleotidi.

Con il Northern blot è stato visto come linc-MD1 poliadenilato sia prodotto maggiormente in mioblasti transdifferenziati sebbene le 2 forme alternative non siano distinguibili su gel.

L'analisi *in situ* ha evidenziato come linc-MD1 non è espresso in condizioni proliferanti, ma la sua espressione è indotta durante il differenziamento (Figura 5D) (Cesana et al.).

L'analisi RT-PCR ha evidenziato la presenza di linc-MD1 tessuti muri di esemplari distrofici *mdx* contemporaneamente alla sintesi di miR-206 e miR-133. In animali *wild type* i livelli di linc-MD1 sono bassi solo nel soleo forse a causa della rigenerazione delle fibre grazie alla presenza di cellule satellite, mentre risulta assente in tutti gli altri muscoli scheletrici. L'analisi *in situ* ha evidenziato come linc-MD1 sia espresso solo in fibre con alto tasso di rigenerazione, mentre non sia espresso in fibre terminalmente differenziate. Tuttavia, la sua espressione nel soleo suggerisce che possa essere espresso nel caso in cui avvenga una rigenerazione del muscolo (figura 5E).

Questi dati suggeriscono che linc.MD1 sia muscolo specifico e la sua espressione è attivata con la differenziazione dei mioblasti (Cesana et al.)



**Figura 5** (A) Localizzazione miRNA nei geni umani; (B) Rappresentazione del *locus* miR-206/133b murino: sono indicati i siti di inizio trascrizione con le frecce (TSS) ottenuti con analisi 5' RACE, sono indicate le regioni di linc-MD1 e nella porzione in basso sono indicati i livelli di conservazione nei vertebrati; (C) Risultati dell'analisi PCR in mioblasti di topo in

condizioni di crescita (GM) e in condizione di differenziamento (DM). Sono mostrate le frazioni nucleari (nuc), citoplasmatiche (cyt), poliadenilate e non (pA+ e pA-). La stessa analisi è stata fatta per cellule satelliti e in fibroblasti murini embrionali. HRPT è stato usato come controllo; (D) analisi di Northern blot per linc-MD1 nella frazione poliadenilata in conduzioni di DM e GM; (E) RT-PCR per i frammenti nel fegato (LIV), cuore (HEA), nel muscolo tibiale anteriore (TIB) e nel soleo (SOL) in topi con *mdx* e in topi *wild type*. (Cesana et al.)

## **Identificazione degli elementi regolatori che dirigono l'espressione di linc-MD1 e miR-206/133b: esperimenti di fusione del promotore e saggi *in vivo***

Sono stati fatti degli esperimenti di fusione del promotore per capire il ruolo nella trascrizione della regione prossimale (PROX) e distale (DIST) di linc-MD1 e miR-206/133b. A monte di pre-miR-233 o della regione della luciferasi, che codifica per un enzima che catalizza una reazione che porta alla produzione di energia luminosa, sono state clonate le regioni dei promotori distali e prossimali. La loro attività è stata testata in mioblasti C2 in stato di proliferazione (GM) (Cesana et al.).

Il promotore PROX è attivo in GM, come visto per l'espressione basale di miR-206 (Figura 6A). In condizioni di differenziamento (DM) la regione prossimale porta all'espressione di miR-233 e della luciferasi, usate come reporter. La regione DIST invece è attiva solo in DM. Se gli elementi DIST e PROX sono nello stesso costrutto avviene la massima attivazione (Cesana et al.).

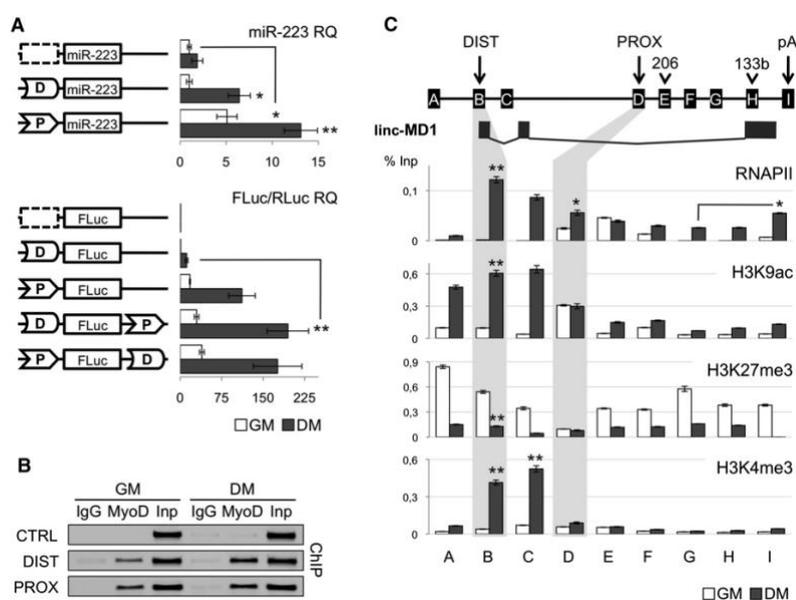
*In vivo* le regioni E-box contenute in linc-MD1 e pre-miR-206/133b legano MyoD come visto nella figura 6B di nell'analisi immunoprecipitazione. MyoD viene acetilato e quindi durante in differenziamento muscolare e quindi lega la regione del promotore distale (Cesana et al.).

9 diverse regioni del *locus* sono state analizzate per modifiche istoniche in fase GM e DM (Figura 6C). È stato visto come parallelamente all'esperimento di fusione del promotore l'RNA polimerasi II (RNAPII) aumenti in condizioni di crescita. Da notare come in queste condizioni l'RNAPII non è legata a miR-133b per cui il promotore PROX non dirige la lettura durante la trascrizione: questi dati sono confermano il fatto che l'espressione di miR-133b e mir-206 sia disaccoppiata.

Quando viene indotta la differenziazione RNAPII immunoprecipita nel promotore DIST: l'espressione di RNAPII diminuisce nel cluster genico e aumenta in 3' come avviene per molte altre unità trascrizionali.

Le modificazioni istoniche sono in accordo con l'attività trascrizionale dei promotori: bassi livelli H3K27me3 (associato al silenziamento) e alti livelli H3K9ac (legato all'espressione del gene) trovati in GM sono mantenuti in DM per l'elemento PROX. Al contrario per DIST i livelli dei 2 istoni modificati sono alti in GM, mentre i pattern sono bassi in DM. Il marker H3K4me3 ha confermato la presenza di TSS nella regione DIST riconosciuti dalla polimerasi. La stessa modificazione è stata mappata nella regione C suggerendo la presenza di ulteriori TSS minori (Figura 6C) (Cesana et al.).

I dati raccolti confermano che il promotore PROX è responsabile per l'espressione di miR-206 in GM, mentre in DM sia PROX che DIST hanno un ruolo nella trascrizione del *locus*.



**Figura 6 (A)** Saggio di attività del promotore in GM e DM dove le regioni distali D e prossimali P sono state clonate a monte di miR-133b e della regione che codifica per la luciferasi (FLuc); **(B)** ChIP analisi con arricchimento di myoD nelle regioni DIST, PROX e nel controllo (CTRL), sono state confrontate con immunoprecipitazione delle IgG di controllo e del 10% delle cromatina di input; **(C)** in altro si trova una rappresentazione schematica del trascritto che è stato analizzato negli esperimenti di ChIP con le regioni rappresentate da lettere maiuscole, a seguire le diverse analisi dei frammenti in base alla presenza di RNAPII o modificazione istoniche. (Cesana et al.)

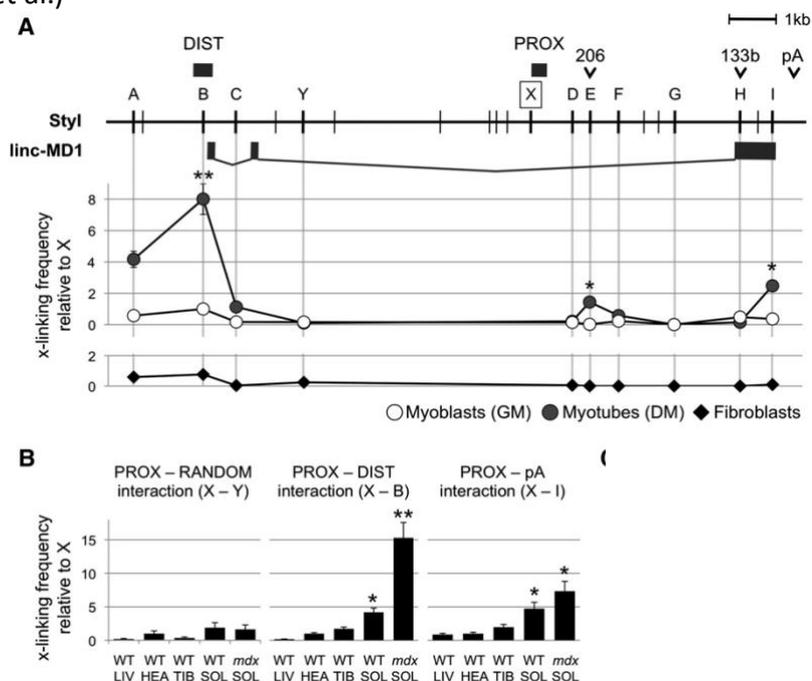
Dimostrata la cooperazione tra i promotori per l'espressione dei miRNA *in vitro*, gli studiosi hanno voluto vedere se l'interazione avvenisse allo stesso modo *in vivo*. Questo è stato fatto con la *Chromosome Conformation Capture* (3C) permette di identificare la conformazione della cromatina e le interazioni tra le regioni di interesse. Il *locus* miR-206/133b è stato testato in mioblasti in condizioni di crescita e di differenziamento e anche nei fibroblasti dove i due miRNA non sono espressi. Un *reverse primer* comune per la regione PROX con altri primer che mappano nel *locus* genomico sono stati usati per una qPCR. I risultati ottenuti concordano con quanto ottenuto tramite immunoprecipitazione della cromatina: infatti durante il DM è stata osservata un'interazione tra PROX e

DIST. Nessuna interazione, invece, è osservata fibroblasti dove il sito è silente (Figura 7A).

L'analisi 3C è stata fatta anche in diversi tipi di tessuto murino *wild type* e *mdx* ed è stato visto come l'interazione avviene solo nei muscoli scheletrici con alto tasso di rigenerazione. L'interazione tra i 2 promotori è particolarmente evidente nei muscoli *mdx*. La stessa specificità si ha tra la regione PROX e la coda poliadenilata, ma è assente tra PROX e regione di controllo (Figura 7B).

Questo ha dimostrato che le interazioni tra DIST e PROX sono funzionali per l'espressione di miRNA e linc-MD1 anche *in vivo*.

(Cesana et al.)



**Figura 7 (A)** Con le lettere maiuscole sono indicati i siti per le regioni studiate indicate nella fascia in alto. Un primer comune X è usato insieme a dei primer antisenso. Il grafico indica i risultati che sono stati ottenuti da analisi 3C della cromatina estratta da 3 diversi tipi cellulari indicati con colori diversi. **(B)** La stessa analisi 3C è stata fatta in diversi tessuti: fegato (LIV), cuore (HEA), tibiale anteriore (TIB) e sole (SOL) *wild type* e SOL *mdx*. (Cesana et al.)

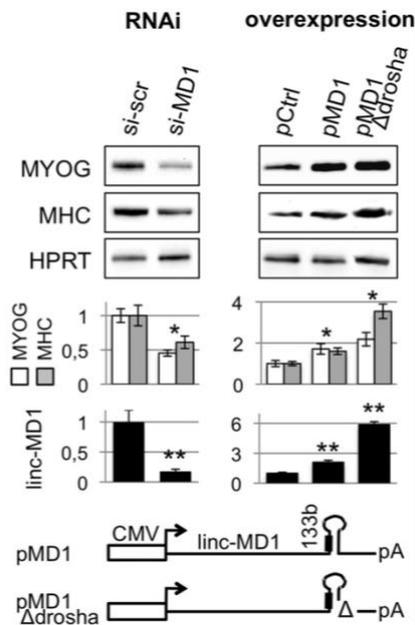
## Regolazione di linc-MD1 ha un ruolo nella differenziazione miogenica

Sono state studiate *in vitro* le interazioni tra miosina a catena pesante (MHC) e miogenina (MYOG), due proteine miogeniche, linc-MD1, miR-1, miR-206 e miR-133, tre miRNA specifici per il muscolo durante il differenziamento muscolare in mioblasti C2. L'analisi ha evidenziato come miR-206 sia espressa in GM con attività basale del promotore PROX; l'espressione di miR-1 e miR-133 è ritardata rispetto a miR-206; l'espressione di linc-MD1 inizia al terzo giorno dall'inizio del differenziamento.

Per capire meglio il ruolo di linc-MD1 nel muscolo scheletrico sono stati fatti esperimenti con RNA *interference* per spegnerne o aumentarne l'espressione genica. I risultati sono stati confrontati con un pannello di controllo. Dopo 5

giorni, l'espressione di MHC e MYOG in mioblasti C2 diminuisce con l'uso del siRNA corrispondente rispetto al controllo (Figura 8 sinistra).

Per l'espressione ectopica di linc-MD1 vengono usati 2 costrutti: pMD1 con la porzione conservata di linc-MD1 e pMD1ΔDrosha con una mutazione fiancheggiante miR-133b, che permette di evitare il taglio e il rilascio di quest'ultimo. pMD1 permette l'espressione di alti livelli di miR-133 al contrario di pMD1ΔDrosha. Entrambi i costrutti (Figura 8 destra) causano un aumento nel numero di MYOG e MHC rispetto al controllo (pCtrl). Sembra che pMD1ΔDrosha ha un'attività maggiore di pMD1 forse legata al sovradosaggio di linc-MD1 e quindi non legata a miR-133b. Nella figura 8 in basso si vedono i livelli di linc-MD1 rispetto ai controlli. È presente una disproporzione tra l'abbondanza di linc-MD1 e gli effetti nelle sintesi dei target miogenici, indicando la presenza di un livello massimo oltre il quale il sistema non può più essere influenzato (Cesana et al.).



**Figura 8** A sinistra vediamo l'analisi di siRNA per linc-MD1 in mioblasti C2 e sono mantenuti in stato DM per 5 giorni. A destra invece si trova cosa si ottiene con sovraespressione di linc-MD1 ottenuta con trasfezione di pMD1 e che pMD1ΔDrosha insieme con un GFP cDNA di controllo. In questo caso invece i campioni sono stati raccolti 4 giorni dopo l'espressione. (Cesana et al.)

## **linc-MD1 è un target di miR-133 e miR-135**

Le analisi di bioinformatica delle sequenze per il riconoscimento dei miRNA in linc-MD1 hanno evidenziato la presenza di 36 siti miRNA. Se escludiamo target che non sono espressi o non hanno un ruolo nella fisiologia muscolare, ne rimangono: uno è miR-135 con 2 siti possibili sul trascritto linc-MD1, l'altro è miR-133 con un solo possibile sito in linc-MD1.

cDNA di linc-MD1 e mutati dove non sono espressi le sequenze di riconoscimento di miR-133 e miR-135 sono clonati a monte del gene della luciferasi. I trascritti sono poi trasfettati con i plasmidi per l'espressione dei geni miRNA in mioblasti C2. Con i plasmidi di controllo l'espressione della luciferasi è ridotta del 50% e del 20% quando rispettivamente miR-135 e miR-133 sono

espressi. Sono invece aboliti invece quando i mutanti sono espressi. Dal momento che la sovraespressione di entrambi i miRNA non altera la stabilità dell'mRNA della luciferasi significa che linc-MD1 può legare sia miR-135 sia miR-133.

I diversi livelli di espressione dei 2 miRNA possono essere legati al fatto che linc-MD1 ha un 2 siti per miR-135 e uno solo per miR-133(Cesana et al.).

## **linc-MD1 e l'espressione dei target di miR-133 e miR-135: in topo e nell'uomo**

Come target sono stati studiati gli mRNA MEFC2 con 1 sito per miR-135 e MAML1 con 2 siti per miR-133 che svolgono un ruolo importante nella differenziazione miogenica. Tramite un'analisi comparativa è stato visto come le regioni 3' UTR dei 2 mRNA son conservate. Le regioni 3'UTR di MAML1 e MEFC2 sono state fuse con la regione codificante per la luciferasi e poi trasfettate in mioblasti C2 con plasmidi codificanti miR-133 e miR-135 accanto al plasmide di controllo. Grazie ai saggi della luciferasi è stato evidenziato come MEFC2 e MAML1 sono target rispettivamente miR-133 e miR-135. Un'ulteriore conferma sono stati plasmidi mutati nel sito riconosciuto dai miRNA per cui è stata confermato il ruolo di miR-133 e miR-206 come repressori (Cesana et al.).

Inoltre, i trascritti contenenti la luciferasi e i 3'UTR dei target sono stati trasfettati in mioblasti C2 con pMD1ΔDrosha o mutanti pMD1-Δ135 and pMD1-δ133 . Con pMD1ΔDrosha entrambi i reporter 3'UTR sono sovra regolati e questo indica come i miRNA espressi a partire da linc-MD1 agiscono come "spugne" abolendo l'attività di MAML1 e MEFC2 3'UTR. Al contrario, nel caso venga usato un mutante di pMD1ΔDrosha MAML1 e MEF2C svolgono la loro funzione. Nel caso però in cui sono trasfettati entrambi i trascritti mutanti di MAML1 e MEF2C la loro attività viene abolita (Cesana et al.).

L'effetto di linc-MD1 su MAML1 e MEF2C è stata monitorato in mioblasti C2 differenziati murini con LNA che bloccano l'espressione dei miRNA sono capaci di ripristinare la sintesi di entrambe le proteine codificate dai due mRNA. Evidentemente livelli troppo alti di linc-MD1 nella forma *wild type* o nel suo mutante Drosha portano ad un aumento delle 2 proteine endogene codificate da MAML1 e MEF2C. Questo indica un *crosstalk* tra i miRNA di linc-MD1, MAML1 e MEF2C che entrano in competizione con il legame dei miRNA (Cesana et al.).

È stata studiata una chinasi attiva durante la differenziazione del muscolo chiamata MCK controllata da MEF2C e MAML1. Per studiare la sua espressione in relazione a linc-MD1 è stato fatto un test sull'effetto di un knockdown o una sovraespressione di linc-MD1 in questo target a valle. L'analisi ha dimostrato che la quantità di MCK è collegata ai suoi attivatori trascrizionali, dimostrando che il circuito linc-MD1-miRNA ha un ruolo nell'espressione genica del muscolo.



## CONCLUSIONI

Grazie a questo studio è stato possibile comprendere qual è il ruolo di linc-MD1 nel differenziamento muscolare, come viene regolata la sua espressione in base allo stato della cellula. Grazie ai numerosi confronti tra linee cellulari murine sane e affette da Duchenne è stato possibile vedere come linc-MD1 non abbia solo un ruolo fisiologico, ma anche patologico. Queste osservazioni sono poi state traslate anche in cellule umane, confermando quello che è stato osservato *in vitro* in cellule murine. linc-MD1 rappresenta un long non coding RNA che, pur non portando direttamente alla produzione di proteine, ne influenza l'espressione interagendo anche con altri RNA che hanno diversi ruoli nella cellula in questo caso miRNA e mRNA.

Questo è stato possibile grazie allo studio approfondito di un processo fondamentale, il differenziamento muscolare, confermando l'importanza della ricerca di base per la scoperta di nuovi elementi biologici. Una maggiore conoscenza dei network creati dai diversi RNA nella cellula potrebbe infatti essere di spunto per la creazione di nuovi farmaci usando l'RNA come target o sfruttandolo come molecola alla base di nuovi farmaci biotecnologici.

Gli RNA attualmente sono molto studiati non solo perché intervengono in diversi ruoli nella cellula come ad esempio nella traduzione, ma anche perché rappresentano talvolta la causa di patologie- in questo caso legato a una malattia genetica la distrofia di Duchenne.

Sarebbe infatti interessante studiare linc-MD1 come target per la cura di miopatie, rappresentando una nuova prospettiva nella cura di questa categoria di malattie per cui molto spesso non si trova una cura andando a migliorare così l'aspettativa di vita del malato (Weskamp et al., Cesana et al.)

## BIBLIOGRAFIA

1. Ohno, S. So Much “Junk” DNA in Our Genome. *Brookhaven Symp. Biol.* 1972, 23, 366–370. [[Google Scholar](#)] [[PubMed](#)]
2. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921 (2001). <https://doi.org/10.1038/35057062>
3. Yao, RW., Wang, Y. & Chen, LL. Cellular functions of long noncoding RNAs. *Nat Cell Biol* 21, 542–551 (2019). <https://doi.org/10.1038/s41556-019-0311-8>
4. Statello, L., Guo, CJ., Chen, LL. *et al.* Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* 22, 96–118 (2021). <https://doi.org/10.1038/s41580-020-00315-9>
5. Amaldi F., Benedetti P., Pesole G., Plevani P. (2018) *Biologia molecolare*. Casa Editrice Ambrosiana, Milano, pp. 468-470
6. Wen Luo, Qinghua Nie, Xiquan Zhang, MicroRNAs Involved in Skeletal Muscle Differentiation, *Journal of Genetics and Genomics*, Volume 40, Issue 3, 2013, Pages 107-116, ISSN 1673-8527, <https://doi.org/10.1016/j.jgg.2013.02.002>
7. Chal J, Pourquié O. Making muscle: skeletal myogenesis *in vivo* and *in vitro*. *Development*. 2017 Jun 15;144(12):2104-2122. doi: [10.1242/dev.151035](https://doi.org/10.1242/dev.151035) PMID: 28634270
8. Weskamp K, Olwin BB, Parker R. Post-Transcriptional Regulation in Skeletal Muscle Development, Repair, and Disease. *Trends Mol Med*. 2021 May;27(5):469-481. doi: [10.1016/j.molmed.2020.12.002](https://doi.org/10.1016/j.molmed.2020.12.002) Epub 2020 Dec 29. PMID: 33384234.