

***Alma Mater Studiorum – Università di Bologna***

DOTTORATO DI RICERCA IN  
Scienze dello sviluppo e del movimento umano

Ciclo XXIII

**Settore/i scientifico-disciplinare/i di afferenza: BIO/13**

Caratterizzazione strutturale e funzionale di nuovi geni del cromosoma 21 umano con approccio integrato: dallo studio del *locus CYXR1* alla meta-analisi di dati di espressione.

**Presentata da: Pelleri Maria Chiara**

**Coordinatore Dottorato**

**Prof.ssa Caramelli Elisabetta**

**Relatore**

**Prof.ssa Frabetti Flavia**

**Esame finale anno 2011**

<b>INTRODUZIONE</b>	<b>1</b>
<b>1. La Sindrome di Down</b>	<b>2</b>
1.1. Caratteristiche fenotipiche e patologie associate	2
1.2. Cause dell'aneuploidia del cromosoma 21 umano	6
<b>2. Espressione genica</b>	<b>9</b>
2.1. Il cromosoma 21 umano (HSA21)	9
2.2. Analisi dell'espressione genica	12
2.3. Variazione del numero di copie ( <i>Copy Number Variation</i> ) e HSA21	14
2.4. Cambiamenti del proteoma e SD	15
<b>3. Regolazione genica</b>	<b>16</b>
3.1. Ruolo dei microRNA e della mutilazione nella regolazione di HSA21	16
3.2. Sequenze regolatorie non codificanti	17
<b>4. Analisi post-genomica</b>	<b>19</b>
4.1. Analisi funzionale	19
4.2. Modello animale del pesce zebra: vantaggi e limiti	20
4.3. Cenni sullo sviluppo embrionale del pesce zebra	23
<b>5. Il sistema multi-trascritto <i>CYYR1</i> come modello genico nello studio della variabilità di espressione dei geni del cromosoma 21 umano</b>	<b>29</b>
<b>SCOPO DELLA TESI</b>	<b>32</b>
<b>MATERIALI E METODI</b>	<b>37</b>
<b>1. CARATTERIZZAZIONE DEL <i>LOCUS</i> GENICO UMANO <i>CYYR1</i></b>	<b>38</b>
1.1. Analisi bioinformatica	38
1.2. Materiale biologico	38
1.3. Progettazione dei <i>primers</i> per il clonaggio genico	40
1.4. RT-PCR	41
1.5. Elettroforesi, purificazione e quantificazione prodotti di RT-PCR	42
1.6. Sequenziamento del cDNA	43
1.7. Clonaggio in vettore, trasformazione in <i>E.Coli</i> e recupero del plasmide	43

<b>2. IL MODELLO ANIMALE <i>ZEBRAFISH</i> PER LO STUDIO FUNZIONALE E DELLA REGOLAZIONE GENICA</b>	<b>45</b>
2.1. Pesce zebra ( <i>Danio rerio</i> )	45
2.1.1. Accoppiamento dei pesci e raccolta degli embrioni	46
2.2. Estrazione di acidi nucleici	48
2.2.1. Estrazione di RNA da pesce zebra	48
2.2.2. Estrazione di DNA genomico da <i>Ciona intestinalis</i>	49
2.3. RT-PCR	50
2.3.1. RT-PCR per l'identificazione di geni <i>housekeeping</i> in pesce zebra	50
2.3.2. RT-PCR per la caratterizzazione di CNEs	52
2.4. L'utilizzo della real-time PCR nell'analisi dell'espressione genica	53
2.5. Progettazione di costrutti da microiniettare in embrioni di pesce zebra: <i>CYYR1</i> e CNEs	55
2.5.1. mRNA del gene <i>cyrr1</i>	55
2.5.2. CNEs	59
2.6. Procedura di microiniezione in embrioni di <i>zebrafish</i>	59
<b>3. SVILUPPO DI UN <i>SOFTWARE</i> IN GRADO DI PRODURRE MAPPA CROMOSOMICHE DI ESPRESSIONE (TRAM)</b>	<b>63</b>
<b>RISULTATI</b>	<b>68</b>
<b>1. CARATTERIZZAZIONE DEL <i>LOCUS</i> GENICO UMANO <i>CYYR1</i></b>	<b>69</b>
1.1. Analisi bioinformatica	69
1.2. Materiale biologico ed estrazione di RNA	69
1.3. Clonaggio genico dei trascritti del <i>locus</i> umano <i>CYYR1</i>	71
1.4. Analisi bioinformatiche dei prodotti proteici e di espressione dei trascritti del <i>locus CYYR1</i>	75

<b>2. IL MODELLO ANIMALE <i>ZEBRAFISH</i> PER LO STUDIO FUNZIONALE E DELLA REGOLAZIONE GENICA</b>	<b>82</b>
<b>2.1. RT-PCR</b>	<b>82</b>
2.1.1. Identificazione di geni <i>housekeeping</i> in pesce zebra	83
2.1.2. Caratterizzazione di CNEs	86
<b>2.2. PCR <i>real time</i> per lo studio di espressione di <i>cyrr1</i></b>	<b>86</b>
<b>2.3. Studio funzionale di <i>cyrr1</i> in <i>zebrafish</i></b>	<b>89</b>
<b>2.4. Studio della regolazione genica di CNEs di <i>Ciona intestinalis</i> attraverso l'utilizzo di <i>zebrafish</i></b>	<b>93</b>
<b>3. SVILUPPO DI UN <i>SOFTWARE</i> IN GRADO DI PRODURRE MAPPA CROMOSOMICHE DI ESPRESSIONE (TRAM)</b>	<b>97</b>
<b>DISCUSSIONE e CONCLUSIONI</b>	<b>99</b>
<b>BIBLIOGRAFIA</b>	<b>114</b>
<b>TABELLE</b>	<b>128</b>

# **INTRODUZIONE**

## **1. La sindrome di Down**

La Sindrome di Down (SD) è uno dei più frequenti disordini cromosomici nell'uomo.

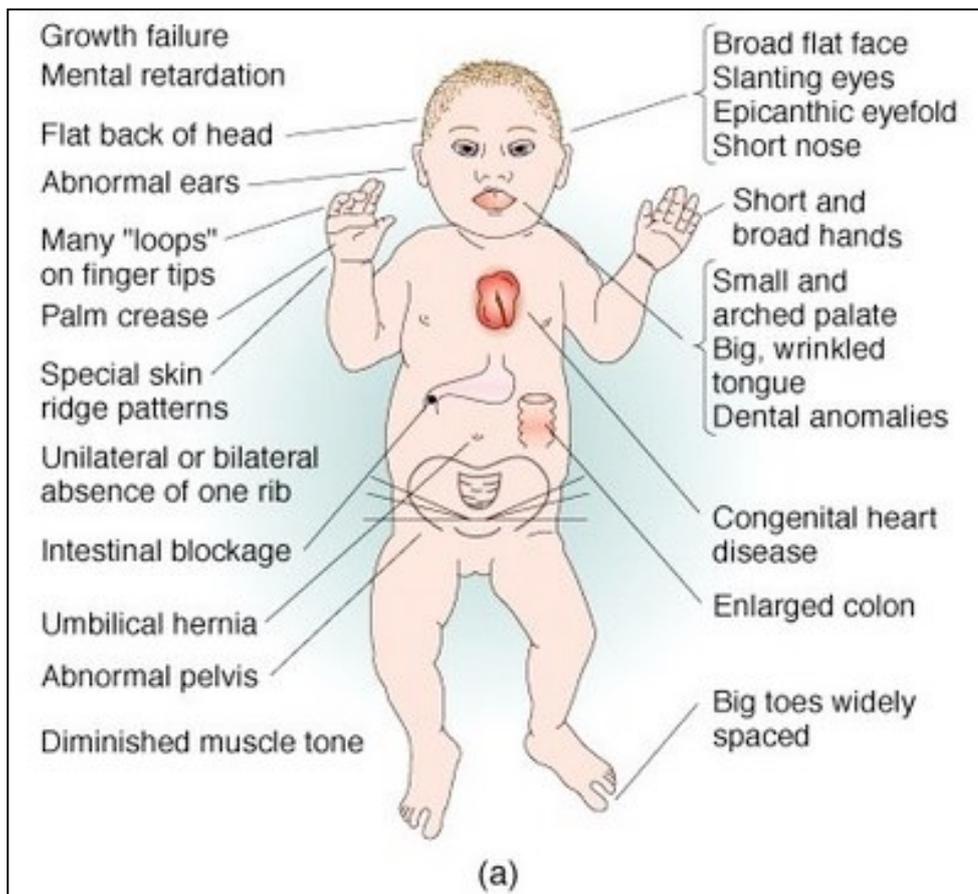
E' caratterizzata dalla presenza di una copia soprannumeraria (trisomia 21) o di una parte (dovuta a traslocazione) del materiale genetico presente sul cromosoma 21 umano (HSA21). Più raramente, la Sindrome di Down può essere causata da mosaicismo: in questo caso alcune, ma non tutte le cellule dell'individuo affetto saranno trisomiche per il cromosoma 21.

Circa lo 0,45% dei concepimenti è caratterizzato da trisomia 21 (Hassold *et al.*, 1996). L'incidenza della malattia è influenzata dall'età materna e varia a seconda delle popolazioni (tra 1/319 e 1/1000 nati vivi) (Wiseman *et al.*, 2009).

In particolare, l'età materna avanzata come fattore di rischio per le aneuploidie autosomiche è supportata da studi sulle origini parentali degli eventi di non-disgiunzione. Occasionalmente i cromosomi omologhi possono presentare piccole differenze di bandeggio. Attraverso l'analisi del bandeggio cromosomico di bambini con SD e dei loro genitori, si può determinare l'origine della non-disgiunzione che è, nel 94% dei casi, di origine materna e nel 6% di origine paterna (da "Eredità. Principi e problematiche della genetica umana", Cummings MR, Ed. EdiSES, 2009).

### **1.1. Caratteristiche fenotipiche e patologie associate**

Le caratteristiche fenotipiche dei pazienti affetti da Sindrome di Down variano da individuo a individuo. Si può descrivere un quadro clinico caratteristico (Figura 1) con "*facies*" larga ed appiattita, occhi con pieghe epicantiche, configurazione anormale delle linee palmari, bassa statura.



**Figura 1.** Quadro clinico dei pazienti affetti da Sindrome di Down

Il ritardo mentale è un disturbo che si presenta in tutti i casi, anche se il livello di disabilità può variare da moderato a severo. Con il progredire della malattia aumentano il ritardo psicomotorio e mentale e nel cervello si riscontrano, in seguito ad esame autoptico, le caratteristiche tipiche del morbo di Alzheimer. Questa neuropatia si può manifestare sin dall'età di trent'anni, nonostante sia una patologia tipica dell'età avanzata. Il morbo di Alzheimer è la causa più frequente di demenza senile ed è caratterizzata da un progressivo peggioramento delle capacità cognitive e funzionali. Nel tempo, il progredire della patologia porta al deterioramento delle reti sinaptiche, perdita neuronale e atrofia della massa cerebrale (Victor e Ropper, 2001). Le lesioni caratteristiche del morbo di Alzheimer sono: a) placche neuritiche, depositi extracellulari della proteina Beta-Amiloide nella corteccia cerebrale; b) degenerazione neurofibrillare, che risulta nell'aggregazione intraneurale di filamenti di proteine Tau. Alcuni fattori chiave della neurobiologia del morbo di Alzheimer sono alterati nei pazienti con SD, per questo esistono ad oggi diverse ipotesi che cercano di

correlare le due patologie. Innanzitutto, il gene *APP*, che codifica per la proteina precursore della *Beta*-amiloide si trova sul braccio lungo del Cromosoma 21 (21q21.3) e la sua sovraespressione è considerata un fattore critico per lo sviluppo precoce di amiloidosi, osservato nei pazienti con SD (Zigman *et al.*, 2008). Mentre l'amiloidosi precoce sembra possa essere correlata alla sovraespressione di *APP* in pazienti affetti da SD, il meccanismo che determina la degenerazione neurofibrillare non è ancora conosciuto. Studi recenti (Wegiel *et al.*, 2008; Ryoo *et al.*, 2007) suggeriscono che la sovraespressione del gene *minibrain-kinase* (*MNB/DYRK1A*, 21q22.2) possa influenzare la degenerazione neurofibrillare in soggetti affetti da SD. Questa protein-kinasi è coinvolta nei processi di fosforilazione che, a livello neuronale, controllano la crescita delle cellule nervose e il trasporto lungo gli assoni; una alterata fosforilazione potrebbe determinare la degenerazione neurofibrillare con l'accumulo di filamenti di proteine Tau iperfosforilate (Hardy e Allsop, 1991).

Circa il 40% dei bambini affetti da Sindrome di Down presenta alla nascita malattie cardiovascolari congenite, in particolare sono molto frequenti difetti del setto. Non sono stati identificati geni candidati per questo fenotipo patologico, ma numerosi geni che codificano per proteine correlate alla matrice extracellulare (*alfa-1* e *alfa-2* collagene di tipo VI, molecole di adesione, integrina *beta-2* e *alfa-2* Collagene di tipo XVIII) sono localizzati sul cromosoma 21 (Vis *et al.*, 2009). Ad esempio, la sovraespressione del Collagene di tipo VI potrebbe essere correlata alla patogenesi dei difetti del setto atrioventricolare (Gittenberger-de Groot *et al.*, 2003). Non tutti i pazienti con SD presentano difetti del setto atrioventricolare che invece possono insorgere in persone non affette da SD. Di conseguenza, non solo i geni del Cromosoma 21, ma tutto il contesto genomico e fattori ambientali sono determinanti nello sviluppo di difetti cardiaci.

Il rischio di sviluppare una leucemia durante l'infanzia è aumentato, mentre si può notare un diminuito rischio di sviluppare tumori solidi, come il cancro alla mammella (Patterson, 2009; Lott e Head, 2005; Patterson e Costa, 2005).

In particolare per quanto riguarda, le leucemie, l'incidenza della leucemia linfoblastica acuta (ALL) in bambini con SD è 20 volte più alta, mentre aumenta di 500 volte l'incidenza di leucemie megacarioblastiche acute (AMKL) (Lange, 2000). Inoltre, tra il 4 e il 10% dei pazienti pediatrici con SD presentano alla nascita una patologia mieloproliferativa transiente (*Transient Myeloproliferative Disorder*, TMD). Nonostante in molti casi si tratti di una patologia che va incontro a risoluzione spontanea, è considerata una sindrome preleucemica, infatti circa il 20% dei bambini a cui è diagnosticata la TMD, sviluppa AMKL entro 4 anni (Tabella 1 da Fonatsch, 2010).

Questo scenario sembra suggerire che la trisomia 21 contribuisca alla trasformazione in senso maligno delle cellule ematopoietiche, anche se la SD non può essere classificata come una classica sindrome da instabilità genomica, in quanto il rischio complessivo di sviluppare un tumore, in particolare un tumore solido, è più basso rispetto a quello della popolazione (Hasle, 2001).

Molte osservazioni portano a pensare a un contributo funzionale della trisomia 21 nelle patologie ematopietiche maligne: l'alta incidenza di leucemie in pazienti affetti da SD, i blasti di pazienti con TMD o DS-AMKL presentano sempre la trisomia 21 (anche in bambini non affetti da SD), infine, la trisomia 21 è spesso acquisita da cellule blastiche in diversi tipi di leucemia. Una questione aperta riguarda però quali siano i geni presenti su HSA21 che promuovono alterazioni nell'ematopoesi in pazienti con SD (Tabella 2 da Fonatsch, 2010). Negli ultimi anni vi sono stati interessanti studi che hanno cercato di rispondere a questa domanda.

Anche se tutte queste osservazioni portano a ritenere ragionevole l'ipotesi che la trisomia 21 sia una componente coinvolta nelle patologie ematopietiche maligne, non si conoscono, ad oggi, i geni specifici la cui disregolazione possa essere direttamente coinvolta nella tumorigenesi. Sono necessari quindi ulteriori studi che mirino ad approfondire i meccanismi molecolari che potrebbero spiegare l'insorgenza così frequente di leucemie in pazienti affetti da SD.

Si possono infine presentare altri effetti fenotipici dovuti all'alterato dosaggio genico, con una variabilità nella manifestazione

clinica, che sicuramente in parte riflette quella della popolazione. Questa variabilità, può dipendere inoltre da differenze riscontrabili a livello di espressione genica nei pazienti con Sindrome di Down, aspetto recentemente approfondito da numerosi studi (Sommer *et al.*, 2008; Lockstone *et al.*, 2007; Altug-Teber *et al.*, 2007; Conti *et al.*, 2007; Lotte *et al.*, 2006; Helguera *et al.*, 2005), allo scopo di definire come varia l'espressione di geni trisomici ed euploidi e di delineare una possibile correlazione con gli effetti fenotipici.

## **1.2. Cause dell'aneuploidia del cromosoma 21 umano**

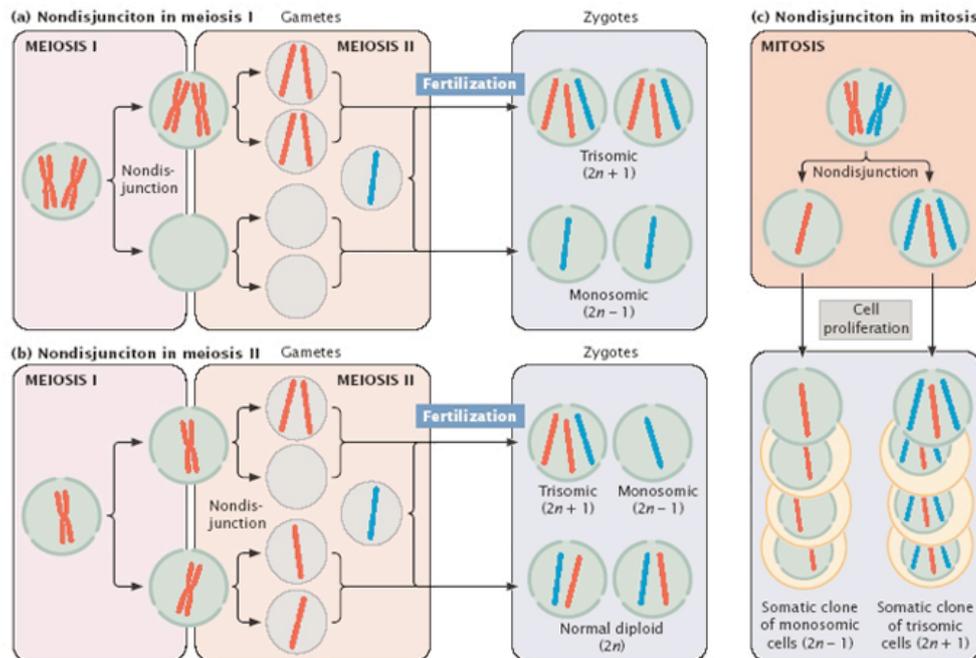
Il 95% dei casi di SD si verifica per una non-disgiunzione durante la prima divisione meiotica dell'ocita (trisomia libera) (Lejeune, Gautier e Turpin, 1959; Antonarakis *et al.*, 2004; Lyle *et al.*, 2001; Scriver *et al.*, 2001).

I meccanismi biologici che portano alla produzione e divisione degli oociti sono alla base dell'ipotesi ormai comprovata da tempo (Penrose 1933; Penrose, 1934) di una correlazione tra incidenza della patologia ed età materna. Sono state proposte diverse teorie che mirano a spiegare l'effetto dell'età materna (Tabella 3 da Hultèn *et al.*, 2010). L'ipotesi ad oggi più accreditata è che avvengano errori di segregazione durante la meiosi di oociti disomici normali, come dimostrano numerosi studi (Hunt e Hassold, 2008; Jones, 2008; Mailhes, 2008; Martin, 2008; Oliver *et al.*, 2008; Allen *et al.*, 2009; Ghosh *et al.*, 2009). In particolare, la non-disgiunzione meiotica potrebbe essere influenzata dalla durata della meiosi femminile: gli oociti primari si formano molto precocemente durante lo sviluppo ed entrano nella profase della prima divisione meiotica molto prima della nascita. La meiosi I si completa solo all'ovulazione, di conseguenza l'ovocita di una donna con più di 35 anni può avere accumulato, durante tutto questo periodo, diversi errori metabolici, oppure fattori ambientali possono danneggiare la cellula e favorire l'aneuploidia al momento del completamento della meiosi.

In casi più rari, il cromosoma soprannumerario è di origine paterna. L'età paterna non risulta, però, essere un fattore di rischio. Gli

spermatozoi, a differenza degli oociti, vengono prodotti continuamente a partire dalla pubertà, senza intervalli nelle divisioni meiotiche e questo potrebbe spiegare l'origine materna della maggior parte delle aneuploidie nell'uomo.

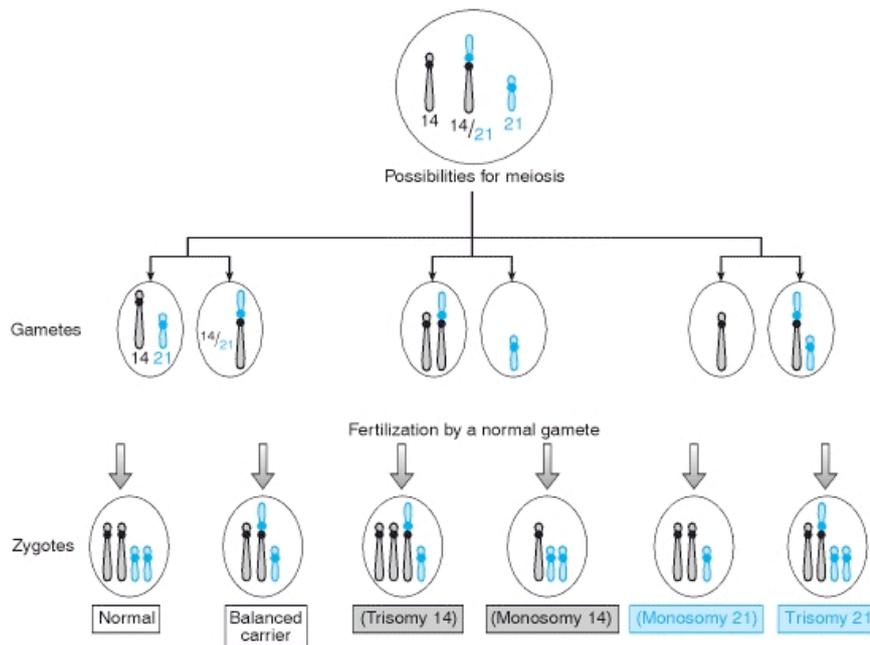
La non-disgiunzione meiotica porta alla formazione di un gamete con 24 cromosomi e può avvenire in I o II divisione meiotica (Figura 2a e b da Miko, 2008).



**Figura 2.** Non disgiunzione meiotica.

L'unione di questo gamete con uno caratterizzato da assetto cromosomico normale, darà origine a un embrione con 47 cromosomi (cariotipo 47,XX,+21 oppure 47 XY,+21). Nel 3-4% dei casi, il cromosoma 21 soprannumerario può derivare da una traslocazione robertsoniana (Figura 3 da Strachan *et al.*, 1999).

Il braccio lungo del Cromosoma 21 è fuso con un altro cromosoma, generalmente acrocentrico (Pueschel *et al.*, 1982; Hook *et al.*, 1983; Koremberg *et al.*, 1994; Lyle *et al.*, 2008).



**Figura 3.** Schema raffigurante il meccanismo della traslocazione Robertsoniana

La più comune traslocazione robertsoniana coinvolge i cromosomi 14 e 21 ( $t(14;21)$ ) ed è prodotta da scambi che avvengono in corrispondenza della porzione prossimale dei bracci corti. La fusione dei bracci corti porterà alla formazione di un cromosoma in cui sono presenti entrambi i centromeri, ma funzionerà come se ce ne fosse uno solo e risulterà stabile. Il piccolo frammento acentrico viene perso, ma senza conseguenze patologiche poiché è costituito soltanto da sequenze ripetute di rDNA, presenti anche su altri cromosomi acrocentrici.

Un individuo portatore, presenta quindi due copie del braccio lungo del cromosoma 21 e del cromosoma 14 (21q e 14q), ma una sola copia del braccio corto di entrambi i cromosomi (21p e 14p); nonostante posseggano soltanto 45 cromosomi, si tratta di persone sane, che però hanno maggiore probabilità di generare figli affetti da SD. Durante la meiosi, infatti, i portatori della traslocazione possono produrre gameti sbilanciati, che, se fecondati, daranno origine a zigoti monosomici (abortiti spontaneamente) oppure trisomici. La SD da traslocazione robertsoniana viene definita anche familiare, in quanto nel 50% dei casi la traslocazione è presente in

un genitore eterozigote. Nel restante 50% dei casi, si origina per una traslocazione *de novo*. La trisomia familiare non ha alcuna relazione con l'età materna e può derivare sia dal padre che dalla madre. Altre traslocazioni che possono dare origine alla trisomia 21 sono t(21;22) e t(21;21).

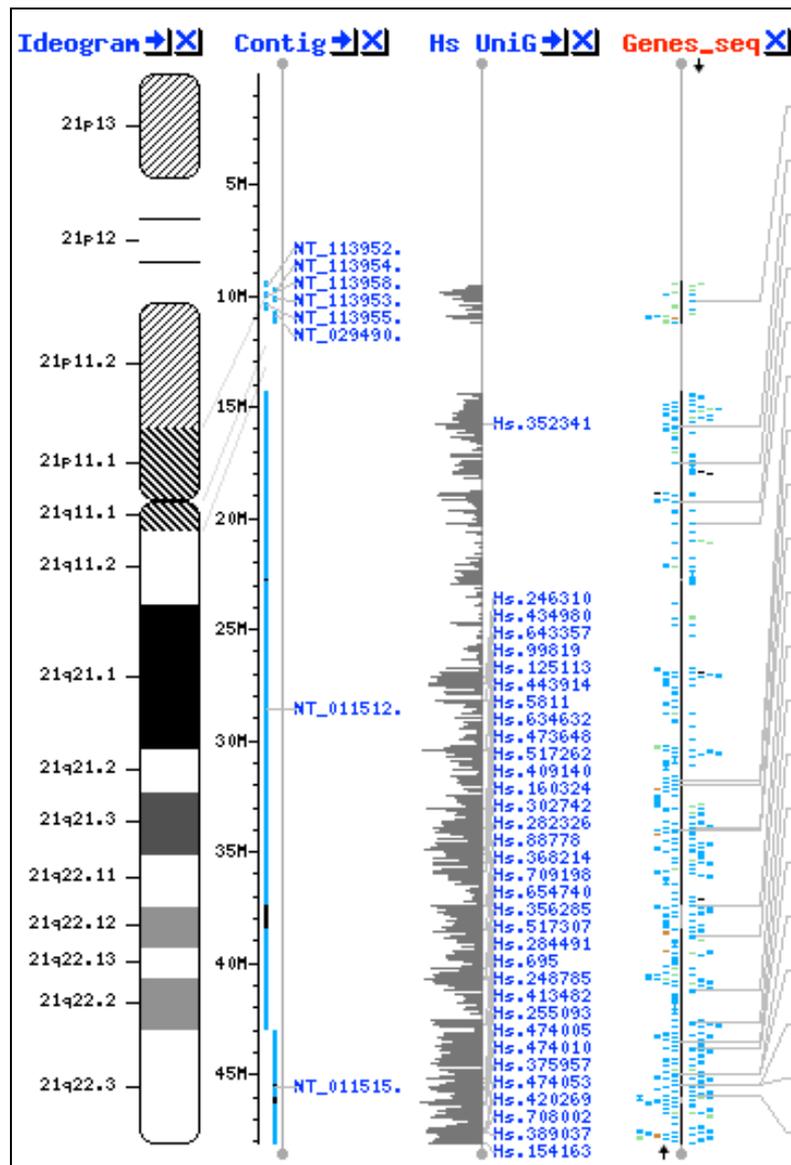
Nell'1-2% dei casi, la SD può essere dovuta ad un mosaicismo aneuploide (Figura 2c da Miko, 2008) causato da un errore nella separazione dei cromatidi (non-disgiunzione mitotica) durante la divisione cellulare dell'embrione in via di sviluppo. In questo caso, solo parte delle cellule dell'embrione avranno un assetto trisomico, mentre le altre presenteranno un cariotipo normale. La distribuzione tra cellule euploidi e trisomiche è altamente variabile, in funzione dello stadio dell'embrione nel momento in cui si è verificata la non-disgiunzione. Anche il quadro clinico, quindi, potrà comprendere segni più o meno gravi.

## **2. Espressione genica**

### **2.1. Il cromosoma 21 umano (HSA21)**

Il cromosoma 21 umano (HSA21) è il più piccolo autosoma e rappresenta circa l'1,5% del genoma (Figura 4).

L'interesse per lo studio di questo cromosoma è aumentato a partire dal 1959, anno della scoperta che la Sindrome di Down (Lejeune *et al.*, 1959). Da quel momento, e grazie alle comuni tecniche d'analisi citogenetica e di linkage, sono stati localizzati sul braccio lungo del cromosoma 21 (21q) i loci di venti malattie monogeniche, fra cui la Sclerosi Laterale Amiotrofica (gene responsabile *SOD1*), la Malattia Polighiandolare Autoimmune (*AIRE*), l'Omocistinuria (*CBS*), la Epilessia Mioclonica Progressiva (*CSTB*) e una forma di malattia di Alzheimer (*APP*). Ad alterazioni in 21q sono state anche associate la predisposizione



**Figura 4.** Mappa del cromosoma 21 umano

Fonte: <http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?org=hum&chr=21>

a leucemie (AML1) e due comuni patologie complesse, l'Iperlipidemia Familiare Combinata e il Disturbo Affettivo Bipolare. Inoltre nei tumori solidi di diversi organi, fra cui testa e collo, seno, pancreas, bocca, stomaco, esofago e polmoni, è stata osservata una perdita di eterozigosi (LOH, *loss of heterozygosity*) in regioni del cromosoma 21, che sembra indicare la probabile presenza in tali regioni di almeno un gene soppressore di tumore (TS, tumour suppressor) non ancora identificato.

Il cromosoma 21 è stato il secondo cromosoma umano, dopo il cromosoma 22, la cui sequenza sia stata accuratamente determinata

(Hattori *et al.*, 2000). Il sequenziamento, effettuato dal “Chromosome 21 mapping and sequencing consortium”, copre 33.546 Kbp (migliaia di paia di basi) del braccio lungo e 280 Kbp del braccio corto; complessivamente, questa sequenza corrisponde al 99.7% della porzione eucromatinica di HSA21.

Il numero totale dei geni (codificanti per proteine e non codificanti) presenti sul braccio lungo del cromosoma 21, non è ancora completamente conosciuto. Quando l’iniziale sequenza di 21q è stata pubblicata, il numero di geni stimati era di 225 (Hattori *et al.*, 2000). Studi successivi basati su analisi bioinformatica, studio delle EST (*expressed sequence tags*, etichette di sequenze espresse), verifiche in laboratorio e analisi comparative di sequenze genomiche, hanno portato a stimare il numero di geni codificanti per proteine localizzati su 21q tra 261 e 364 (Antonarakis *et al.*, 2004). Questa stima include anche potenziali trascritti supportati solo dalla determinazione di una sequenza parziale di RNA comprendente almeno due esoni (“spliced EST”, *spliced expressed sequence tags* - etichette di sequenze espresse comprendenti almeno una giunzione di *splicing*).

E’ stato stimato anche il numero di pseudogeni, attualmente valutato in 141 (Ensembl, <http://www.ensembl.org/index.html>).

La densità genica così determinata risulta relativamente bassa (circa 300 geni/33.800 Kbp) per il cromosoma 21 umano, in particolare rispetto al cromosoma 22 (circa 700 geni/33.400 Kbp). Tale densità è compatibile con la osservazione che la trisomia 21 è una delle poche (la più frequente) delle trisomie autosomiche umane compatibili con la vita.

L’identificazione e la caratterizzazione dei geni presenti su HSA21 può contribuire alla comprensione delle basi molecolari della patologia. La descrizione di alcuni soggetti con SD in assenza di Trisomia 21, ma in presenza di una duplicazione lungo lo stesso HSA21 di una specifica regione all’interno della sottobanda 21q22.2, fa ritenere che i geni presenti in questa regione possano essere particolarmente importanti nel determinare il fenotipo Down. Per questo tale regione, compresa tra lo pseudogene *LOC266693* e il gene *ETS2*, è stata denominata “regione critica per la sindrome di Down” (*Down Syndrome Critical Region*,

DSCR). Nonostante la ricerca si sia concentrata per anni sull'identificazione di una "regione critica", l'esistenza di loci responsabili delle caratteristiche fenotipiche della SD non è mai stata dimostrata. Inoltre, studi recenti (Olson *et al.*, 2007) hanno dimostrato come la trisomia della DSCR sia condizione necessaria, ma non sufficiente per determinare le caratteristiche fenotipiche del cervello in topi trisomici.

Anche se si ipotizza che molte caratteristiche fenotipiche dei soggetti affetti da SD siano una conseguenza della sovraespressione di alcuni geni presenti su HSA21, è molto difficile selezionare geni candidati per un fenotipo così complesso. Definire come una copia aggiuntiva di tutto o parte di HSA21 porti al fenotipo tipico della Sindrome di Down, è un caso specifico di un problema più generale che riguarda gli effetti di uno squilibrio cromosomico sulla morfologia e sulle funzioni biologiche. Sicuramente non esiste un meccanismo univoco in grado di spiegare gli effetti di una aneuploidia e occorre tener presente la correlazione tra geni diversi.

Inoltre, l'avanzamento nella conoscenza della complessità del genoma umano, ha messo in luce l'esistenza di numerosissimi meccanismi che influenzano l'espressione genica o che sono con essa embricati: la regolazione genica mediata da tantissimi attori tra i quali miRNA (microRNA), sequenze regolatrici conservate non codificanti, fattori di trascrizione, l'epigenetica e cambiamenti nel proteoma stesso.

## **2.2. Analisi dell'espressione genica**

Sono stati effettuati numerosi studi che mirano a definire se vi siano cambiamenti nei livelli di espressione genica in pazienti, modelli animali o cellulari con trisomia 21 (Tabella 4 da Patterson, 2009).

In alcuni studi, i livelli di espressione genica sono stati determinati in specifici tessuti di pazienti affetti da SD. Sommer *et al.* (2008), utilizzando l'analisi seriale dell'espressione genica (*Serial Analysis of Gene Expression*, SAGE), hanno verificato alterazioni dell'espressione genica in linfociti di pazienti pediatriche con SD. La conclusione generale dello studio fu che si può osservare una modesta disregolazione dei geni

euploidi e che, a causa della bassa sensibilità del metodo, non vi sono evidenze di un effetto dose-dipendente per i geni di HSA21. L'osservazione più interessante riguarda la bassa espressione di alcuni geni coinvolti nelle alterazioni neurologiche degli individui con SD.

Lockstone *et al.* (2007), attraverso la tecnica dei *microarrays* hanno analizzato l'espressione genica in campioni di cervello di pazienti affetti da SD, dimostrando la sovraespressione del 27% dei geni di HSA21 (percentuale molto più alta rispetto al 4,4% relativa ai geni presenti su altri cromosomi). In particolare, vi è una significativa sovraespressione di geni correlati al citoscheletro. La tecnica dei *microarrays* è stata utilizzata anche per studiare l'espressione genica in campioni di cuore ottenuti da feti con SD (Conti *et al.*, 2007), dimostrando una interessante sovraespressione di geni codificanti per proteine della matrice extracellulare, accompagnata da una diminuita espressione di geni nucleari codificanti per proteine mitocondriali. Si tratta di una osservazione che potrebbe correlare con l'ipotesi che vede lo stress ossidativo e la disfunzione mitocondriale tra le caratteristiche della SD (Lott *et al.*, 2006; Helguera *et al.*, 2005). Tutti questi studi hanno portato a numerose osservazioni interessanti, ma difficili da interpretare.

In particolare, una domanda rimasta aperta riguarda la variazione di espressione genica nella popolazione euploide: in generale, si pensa che un gene la cui espressione vari più di 1.5 volte da individuo a individuo nella popolazione normale, potrebbe non essere coinvolto nella SD (Deutsh *et al.*, 2005). Non è chiaro, poi se la presenza di un cromosoma 21 aggiuntivo possa incidere sulla variazione dell'espressione genica. Recentemente, sono stati condotti alcuni studi che hanno voluto confrontare linee cellulari linfoblastoidi e fibroblastiche ottenute da pazienti con SD con quelle ottenute da soggetti euploidi (Yaha-Graison *et al.*, 2007; Prandini *et al.*, 2007). Solo un limitato numero di geni di HSA21 sembra sovraespresso in SD e si ipotizza che geni soggetti a regolazione e che mostrano differenze di espressione tra linee cellulari trisomiche ed euploidi, potrebbero essere i maggiori responsabili della variabilità fenotipica della SD. Infine, risultati analoghi sono stati raggiunti da gruppi di ricerca che hanno utilizzato il modello murino Ts65Dn per la SD (Sultan *et al.*, 2007). Analizzando la

differenza di espressione dei geni in tre regioni distinte del cervello di topi, sono state identificate tre categorie: geni con livelli di espressione molto più alti nei tessuti derivati da topi trisomici, geni la cui espressione non varia tra tessuti trisomici ed euploidi e geni con livelli di espressione molto variabili, tanto da non poter essere discriminanti. Anche in questo caso l'ipotesi avanzata sostiene che i geni che mostrano una significativa e costante sovraespressione nei tessuti ottenuti dai topi Ts65Dn potrebbero avere un ruolo nel determinare le caratteristiche tipiche della SD, mentre quelli che mostrano un più alto grado di variabilità potrebbero contribuire alla variabilità del fenotipo. Come si può notare dalla Tabella 4 (Patterson, 2009), i tipi di campioni utilizzati per condurre tutte queste analisi di espressione genica sono numerosi e molto diversi fra loro, cosa che rende difficile paragonare studi differenti.

Proprio una meta-analisi di studi indipendenti, potrebbe invece fornire elementi utili che mettano in evidenza regioni cromosomiche caratterizzate da una consistente e stabile sovraespressione genica. Non esistono, ad oggi, strumenti bioinformatici in grado di effettuare una meta-analisi di questo tipo in modo semplice e facilmente fruibile. Per questo la messa a punto di *software* in grado di produrre mappe cromosomiche di espressione a partire da dati di analisi genica globale potrebbe risultare molto utile per l'interpretazione di dati così numerosi. Questo approccio bioinformatico potrebbe consentire di identificare in modo predittivo intere regioni cromosomiche i cui geni hanno lo stesso schema di espressione e profili stabili di sovraespressione in campioni patologici.

### **2.3. Variazione del numero di copie (*Copy Number Variation*) e HSA21**

Un ulteriore aspetto caratteristico del genoma umano che può influenzare variazioni nell'espressione genica è il fenomeno denominato variazione del numero di copie (*Copy Number Variations*, CNVs).

Esistono, infatti, regioni di DNA il cui numero di copie varia da individuo a individuo. CNVs potrebbero riguardare fino al 12% del genoma umano e potrebbero essere responsabili del 20% delle variazioni nell'espressione genica (Stranger *et al.*, 2007). Inoltre, recenti studi hanno

dimostrato come alcuni geni localizzati all'interno di regioni con CNVs mostrino un livello di espressione basso, ma più variabile rispetto a geni posizionati all'esterno di queste zone; infine, si ipotizza che CNVs possano influire non solo sull'espressione dei geni all'interno della regione coinvolta, ma anche sull'espressione di geni che si trovano a una determinata distanza da essa (Henrichsen *et al.*, 2009: Segmental copy number variation shapes tissue transcriptomes; Copy number variants, disease and gene expression). In particolare, sono stati condotti anche alcuni studi su CNVs riguardanti HSA21 e vi sono evidenze di una correlazione tra questo fenomeno e la SD (Sleeper *et al.*, 2006; Rovelet-Lecrux *et al.*, 2006; Strefford *et al.*, 2006).

Anche la presenza della terza copia di HSA21 potrebbe essere di per sé una condizione che influenza la variabilità di espressione di geni euploidi. Recentemente è stato dimostrato come l'espressione dei geni euploidi sembri variare in maniera più consistente in campioni DS rispetto a campioni di controllo non trisomici.

## **2.4 Cambiamenti del proteoma e SD**

Variazioni nell'espressione genica sono particolarmente rilevanti se determinano una diversa quantità di prodotto funzionale proteico nelle cellule.

Ad oggi, gli strumenti che permettono l'analisi del proteoma sono ancora poco efficienti. Alcuni studi hanno utilizzato la 2DGE per indagare la variazione di alcune proteine in tessuti umani trisomici ed euploidi e in tessuti ottenuti da modelli murini per la SD (Harris *et al.*, 2007; Lubec, 2001). Non si può ancora arrivare ad una ipotesi conclusiva poiché i risultati mostrano una variabilità nei livelli di proteine codificate da geni euploidi e da geni presenti su HSA21. Si tratta quindi di un campo ancora tutto da esplorare e che potrà portare a nuove informazioni che potranno fare da ponte tra gli studi di espressione genica e di tipo funzionali.

L'insieme di queste osservazioni rivela un quadro molto complesso in cui sono coinvolti molteplici meccanismi molecolari in grado di

influenzare l'espressione genica e quindi determinare le caratteristiche della SD, meccanismi che interagiscono tra loro e che sono sicuramente oggetto degli studi più attuali sulla SD.

Occorre, però, aggiungere importanti osservazioni riguardanti la regolazione genica: esistono, infatti, numerosi meccanismi che determinano variazioni dipendenti da funzione cellulare, condizione sana o patologica, contesto genomico ed altro.

### **3. Regolazione genica**

Vi sono numerosi e differenti meccanismi molecolari che possono essere coinvolti nella regolazione genica. Per quanto riguarda la Sindrome di Down, si tratta di un argomento poco esplorato, ma nondimeno interessante in quanto potrebbe portare ad una più completa comprensione di cascate di segnali o *pathway* molecolari coinvolti nella patologia.

Inoltre, lo studio di meccanismi coinvolti nella regolazione spaziale e temporale dell'espressione dei geni di interesse è un argomento di ricerca molto interessante che potrebbe portare a una migliore comprensione della funzione di geni ignoti.

#### **3.1. Ruolo dei microRNA e della metilazione nella regolazione di HSA21**

Sono stati recentemente descritti e studiati 5 microRNA (miRNA) localizzati su HSA21: miR-99a, let-7c, miR-125b-2, miR-155 e miR-802 (Kuhn *et al.* 2008). Tutte queste piccole molecole regolatorie di RNA sono sovraespresse in campioni di cervello e cuore di feti affetti da SD, cosa che suggerisce un possibile ruolo nel determinare l'insorgenza di difetti cognitivi e cardiaci in pazienti con SD.

Nuovi metodi sperimentali e computazionali hanno permesso di individuare un numero sempre maggiore di miRNA, cosa che favorisce l'ipotesi che possano esistere altri miRNA localizzati anche sul HSA21. Infine, recenti studi hanno dimostrato che la trascrizione stessa dei miRNA

può essere regolata da cambiamenti epigenetici, come la metilazione (Guil e Esteller, 2009).

Anche l'epigenetica gioca infatti un ruolo importante nella regolazione dell'espressione genica e nella formazione di determinati prodotti proteici, non sempre deducibili dalla sola analisi di sequenza nucleotidica.

In particolare, il Cromosoma 21 è coinvolto nel processo di metilazione sotto diversi aspetti:

- sono stati identificati alcuni geni localizzati su HSA21 coinvolti nei meccanismi di metilazione di DNA e proteine (OOi *et al.*, 2007; Hu *et al.*, 2008; Takashima *et al.*, 2009; El-Maarii *et al.*, 2009; Figaro *et al.*, 2008)
- la metilazione potrebbe alterare l'espressione di geni di HSA21. Sono state messe a punto nuove metodiche per studiare la metilazione del DNA ed identificare siti bersaglio su HSA21 (Carson *et al.*, 2008), seguendo l'ipotesi, ancora tutta da verificare, che vi possano essere siti di metilazione differenti tra SD e euploidi.

### **3.2. Sequenze regolatorie non codificanti**

Fra gli elementi regolatori, occorre nominare un insieme molto vasto e variegato di sequenze i cui nomi ricordano aspetti funzionali (*enhancer, repressor, silencer, promoter, insulator*), ma i cui meccanismi d'azione non sono ben conosciuti.

Si tratta di elementi funzionali non codificanti (*noncoding functional elements*, ncFEs) per i quali è difficile innanzitutto determinare un metodo di identificazione e caratterizzazione all'interno del genoma, inoltre occorre integrare la definizione della sequenza con il suo aspetto funzionale, infine è necessario prendere in considerazione la storia evolutiva per meglio comprendere il linguaggio molecolare che sottostà ad una determinata funzione regolatoria.

Sequenze che rimangono altamente conservate durante l'evoluzione sono, molto probabilmente, elementi funzionali.

Recentemente, numerosi studi comparativi su vasta scala hanno permesso l'identificazione di elementi non codificanti con funzione regolatoria,

conservati evolutivamente tra i vertebrati. Gli approcci che si possono utilizzare, però, sono differenti: un primo metodo si basa sull'utilizzo di parametri molto stringenti che permettono di selezionare sequenze ultraconservate (*ultraconserved sequence elements*, UCEs), caratterizzate dal 100% di identità per almeno 200 paia di basi tra i genomi di ratto, topo e uomo (Bejerano *et al.*, 2004). Il secondo approccio, invece, prevede la comparazione tra genoma umano e di pesce, permettendo l'identificazione di sequenze conservate da almeno 450 milioni di anni. Due studi indipendenti hanno caratterizzato sequenze non codificanti conservate paragonando il genoma umano con quello di *Fugu rubripes*: il primo ha definito queste sequenze (conservate nel topo) regioni ultraconservate (*ultraconserved regione*, UCRs) (Sandelin *et al.*, 2004). Il secondo, ha invece identificato sequenze conservate tra uomo e *Fugu* (*conserved non-coding elements*, CNEs) (Tabella 5 da Elgar e Vavouri, 2008).

L'alto livello di conservazione delle sequenze riflette molto probabilmente la presenza di meccanismi molecolari ugualmente conservati tra i vertebrati e coinvolti in funzioni chiave per l'organismo. In particolare, queste sequenze sono caratterizzate da una forte associazione spaziale con geni coinvolti nella regolazione trascrizionale e/o nello sviluppo embrionale (*genes involved in transcriptional regulation and/or development; trans-dev genes*).

In questo progetto di ricerca, sono stati condotti esperimenti utili alla maggiore comprensione della storia evolutiva di CNEs, grazie alla collaborazione col gruppo di ricerca del Prof. Elgar (MRC National Institute for Medical Research, Londra).

Nonostante siano ancora elementi poco caratterizzati, non si può escludere un loro coinvolgimento nella SD. Col progredire della definizione dei *pathway* molecolari coinvolti nella patologia potrà essere interessante verificare se all'interno di queste cascate di segnale, sono coinvolti anche elementi regolatori di questo tipo, soprattutto considerando il fatto che si tratta di sequenze molto importanti proprio durante lo sviluppo embrionale.

## 4. Analisi post-genomica

### 4.1. Analisi funzionale

La genomica funzionale è un settore di studio nato con la realizzazione dei progetti di sequenziamento genomico che si propone di studiare la funzione dei geni su larga scala, ovvero in modo globale. Una volta determinata la sequenza di un genoma, grazie ai metodi di approccio bioinformatici, essa viene scansionata da varie tipologie di *software* per annotare la presenza di geni noti o mai riconosciuti in precedenza. Gli studi funzionali si incaricano a questo punto di risolvere il problema legato al ruolo di tali geni assegnando una probabile funzione alle proteine codificate.

Un approccio per una analisi preliminare su una nuova sequenza codificante è sicuramente il metodo bioinformatico che ben si inserisce nella dimensione genomica, ma ha dei limiti che sono intrinseci nella procedura di analisi.

Infatti l'analisi bioinformatica, al fine di assegnare una funzione al gene in esame, si basa su un confronto di sequenze che richiede, quindi, la presenza di una o più sequenze a funzione nota significativamente simili a quella di interesse; inoltre viene dato come base di partenza il fatto che sequenze significativamente simili abbiano la stessa funzione in forza delle sole regioni di similarità, trascurando i possibili effetti differenziativi indotti dalle regioni dissimili. Venendo meno uno o entrambi questi presupposti, o per confermare sperimentalmente i risultati di un'analisi di questo tipo, bisogna selezionare nuove strategie di indagine.

I metodi per studiare la struttura, la funzione e l'espressione dei geni sono molteplici e possono basarsi su estratti di cellule, colture cellulari o modelli animali. Un esempio di strategia su larga scala è rappresentata dai *microarray*, che consentono di studiare sistematicamente l'espressione genica; altre metodiche usano l'ibridazione di sonde come nel Northern blot o nell'ibridazione *in situ*, altre ancora usano tecniche che mettono in evidenza le interazioni proteina-proteina come il test di interazione tra proteine chimeriche in lievito (test dei due ibridi) e il

riconoscimento di proteine tramite anticorpi (*Western blot*, immunistochemica, immunocitochemica, ecc.).

Particolarmente interessante sarebbe caratterizzare i trascritti e le proteine di interesse in linee cellulari, che potrebbero fungere da modello per lo studio dell'espressione e della funzione genica.

Tutte queste strategie operano, però, *ex vivo* e forniscono risposte che possono rivelarsi errate, in tutto o in parte, quando l'RNA o la proteina in esame vengano analizzati nel loro ambiente fisiologico. Per questo, la sperimentazione su modelli animali *in vivo* non ha mai definitivamente smesso di suscitare l'interesse all'interno della comunità scientifica, comunque attenta e sempre più sensibile alle problematiche etiche legate a questo tipo di esperimenti.

In questo lavoro, è stato utilizzato un approccio *in vivo* servendosi del modello animale del pesce zebra che permette di studiare geni di interesse, nel presente studio il gene *cyrr1*, nel contesto dell'organismo *in toto*.

#### **4.2. Modello animale del pesce zebra: vantaggi e limiti**

Il modello animale utilizzato per lo studio funzionale di *CYYRI* è quello del pesce zebra (*Danio rerio*), un pesce tropicale di piccole dimensioni comunemente chiamato anche *zebrafish* (pesce zebra) in lingua inglese.

Tale animale è un ottimo modello per lo studio dello sviluppo dei vertebrati sia per le caratteristiche fisiologiche e morfologiche di vertebrato che lo rendono un modello più vicino all'uomo rispetto agli invertebrati, sia per la sua semplicità di allevamento e di utilizzo sperimentale.

Sebbene il genoma del pesce zebra consista di circa  $1,7 \times 10^9$  bp (circa la metà del genoma umano) e *zebrafish* e uomo siano separati evolutivamente da 420 milioni di anni, ci sono comunque delle regioni conservate di sintenia (ossia regioni cromosomiche dove si co-localizzano due o più loci) tra le due specie (CGOW, 1996) che permettono al pesce zebra di fare da ponte tra modelli di invertebrati (come ad esempio

*Drosophila melanogaster* e *Caenorhabditis elegans*) e vertebrati complessi come topo e uomo.

Il pesce zebra è stato ampiamente utilizzato come modello di studio nell'ambito della biologia dello sviluppo ed utilmente impiegato in studi di embriogenesi e morfogenesi; successivamente, tuttavia, il campo di applicazione si è allargato alla genomica funzionale e alle potenzialità applicative dei mutanti di *zebrafish* per lo studio di malattie umane. A questo proposito è bene sottolineare che nel pesce zebra l'introduzione di mutazioni può avvenire utilizzando metodi diversi come la mutagenesi indotta chimicamente, tramite irradiazione o come conseguenza di trattamenti con vettori virali (Ingham e Whitfield, 2004).

In seguito alla introduzione di queste mutazioni, *screening* genetici hanno evidenziato le caratteristiche genotipiche di fenotipi mutati, molti dei quali corrispondono a malattie umane. L'estrema versatilità di questo vertebrato in studi mutazionali di genetica classica ha avvalorato l'utilizzo di mutanti di *zebrafish* come modello per lo studio di specifiche patologie umane che riguardano, ad esempio, il sistema emopoietico e gli apparati cardiovascolare e renale (Barut e Zon, 2000). L'utilizzo di questo modello animale sia nella biologia dello sviluppo che nella genetica è vantaggioso sotto molti punti di vista. Lo sviluppo embrionale del pesce zebra e il suo piano corporeo, infatti, sono conformi a quelli umani, con la struttura degli organi interni molto simile a quella dell'uomo, anche se, in confronto ad altri vertebrati, gli organi del pesce, possono essere definiti "in miniatura" in quanto è utilizzato un numero molto minore di cellule per assolvere alla medesima funzione.

Le uova, deposte dalle femmine in grande quantità, sono fecondate esternamente dal maschio. Almeno durante le prime 24 ore dalla fecondazione (hpf, *hours post fertilization*) gli embrioni rimangono trasparenti e possono essere osservati attraverso il corion.

Questo significa che, in particolare negli stadi precoci dello sviluppo, la completa trasparenza consente l'osservazione dettagliata delle strutture interne rendendo possibile l'osservazione *in vivo* della presenza di difetti morfologici. La trasparenza degli embrioni, inoltre, rende immediata l'osservazione di *pattern* di espressione genica sia in tessuti fissati e

colorati, sia in cellule marcate con sonde fluorescenti in campioni vivi (Ingham e Whitfield, 2004).

Lo sviluppo embrionale è piuttosto rapido e già a 24 ore dalla fecondazione è visibile il piano corporeo tipico dei vertebrati con la presenza del tubo neurale, dei muscoli, del sistema cardiovascolare e del cuore pulsante; dopo tre giorni dalla fertilizzazione gli embrioni, usciti dal corion, sono capaci di nuotare e di nutrirsi. La maturità sessuale si raggiunge nel giro di circa tre mesi e l'animale diventa capace di riprodursi.

L'analisi della espressione e della funzione genica è facilmente applicabile a questo vertebrato, in quanto trasparenza e dimensione relativamente grande delle uova (circa 0,8 mm) le rendono facilmente manipolabili. In particolare si possono mettere a punto procedure di microiniezione di costrutti marcati, per studiare la localizzazione del prodotto proteico o realizzarne la sovraespressione, o anche, utilizzare morfolino-oligonucleotidi antisense (MO) utili allo studio della funzione del gene in esame attraverso la sua transitoria inibizione (*knock-down*). La velocità dello sviluppo degli embrioni e la loro facile osservabilità rendono poi possibile osservare in tempi brevi l'effetto di mutazioni indotte con tale manipolazione.

I limiti nell'utilizzo di questo modello sono, invece, essenzialmente due. Il primo riguarda l'attuale incompletezza del sequenziamento del genoma del pesce zebra, atteso inizialmente per la fine del 2003. Al 14/10/2010 il progetto di sequenziamento del genoma del pesce zebra del "The Wellcome Trust Sanger Institute" ([http://www.sanger.ac.uk/Projects/D\\_rerio/](http://www.sanger.ac.uk/Projects/D_rerio/)) dichiara che l'assemblaggio più recente (Zv9) si basa su 11.099 cloni localizzati all'interno di una mappa fisica. Questi cloni coprono oltre l'87% del genoma totale (1.335 Gb). Questo porta ad avere regioni delle quali ancora non si conosce la sequenza completa che possono costituire un limite nell'analisi genomica; tale svantaggio, però, è destinato ad esaurirsi appena il sequenziamento sarà completo.

Il secondo limite risiede nel modello stesso; il pesce zebra, infatti, è un vertebrato inferiore e, nonostante la conservazione delle caratteristiche

di sviluppo e morfologiche che appartengono ai vertebrati, è innegabile che anche a livello macroscopico uomo e *zebrafish* siano diversi. Ciò nonostante, anche se l'organizzazione dell'organismo *in toto* porta a diversità morfologiche e fisiologiche importanti, i meccanismi molecolari di base che contraddistinguono e differenziano i vertebrati dagli invertebrati sono pressoché identici.

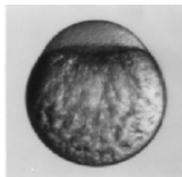
#### 4.3. Cenni sullo sviluppo embrionale del pesce zebra

Lo sviluppo embrionale del pesce zebra è molto veloce e abbraccia il periodo che va dal momento della fecondazione fino al terzo giorno di sviluppo.

Le fasi di seguito schematizzate, in cui questo periodo si suddivide, si riferiscono ad uno sviluppo alla temperatura di 28.5°C, tenendo conto che questa variabile influisce sulla cinetica dello sviluppo.

La fonte di riferimento per il testo e le immagini è il sito [http://zfin.org/zf\\_info/zfbook/stages/index.html](http://zfin.org/zf_info/zfbook/stages/index.html).

Fase di zigote (Figura 5): questa fase inizia nel momento della fecondazione e si conclude con la prima divisione che avviene circa dopo 3/4 d'ora. In questo breve lasso di tempo le contrazioni dell'actina citoscheletrica determinano lo spostamento del citoplasma non occupato dal tuorlo verso il polo animale della cellula. Quest'ultima quindi subisce una modificazione nella forma che da sferica diventa piriforme con il blastodisco localizzato in posizione apicale.

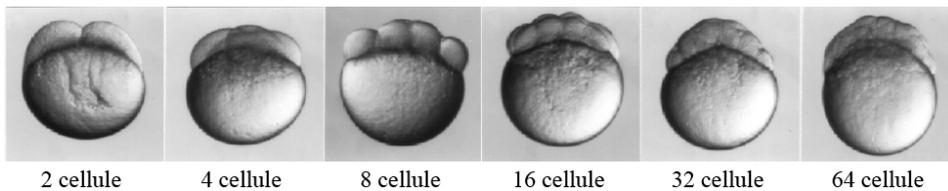


zigote

**Figura 5.** Zigote di pesce zebra

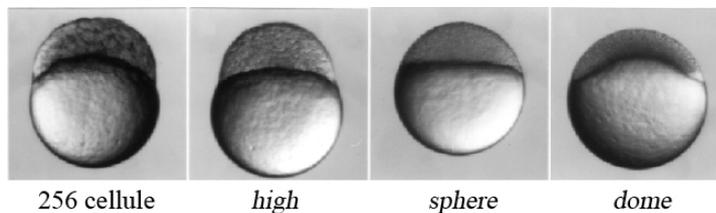
Fase di segmentazione (Figura 6): questo processo si compie in circa 2 ore a partire dalla prima divisione cellulare fino al momento in cui l'embrione

raggiunge le 64 cellule. Dopo la prima divisione, i blastomeri si dividono in modo sincrono a intervalli regolari di circa 1/4 d'ora; tali divisioni sono incomplete e perciò lasciano le cellule tra loro connesse attraverso ponti citoplasmatici che permettono il passaggio di molecole di peso inferiore ai 17 kDa circa.



**Figura 6.** Stadi principali della fase di segmentazione di pesce zebra

Fase di blastula (Figura 7): questa fase comincia da quando l'embrione è allo stadio di 128 cellule (e assume una forma *ball-like*) e finisce con l'inizio della gastrulazione. Durante questo periodo i piani lungo i quali avvengono le divisioni cellulari sono meno precisi rispetto a quelli che si hanno durante la segmentazione; per questo motivo la disposizione spaziale dei blastomeri è meno precisa. La caratteristica che permette di distinguere tra loro vari stadi di sviluppo (128 cellule, 256 cellule, 1000 cellule, *high*, *sphere* e *dome*) è il polo animale che con l'avanzare del tempo assume forma diversa, grazie all'aumento di dimensione delle cellule, e colore diverso a causa del numero sempre crescente di cellule.

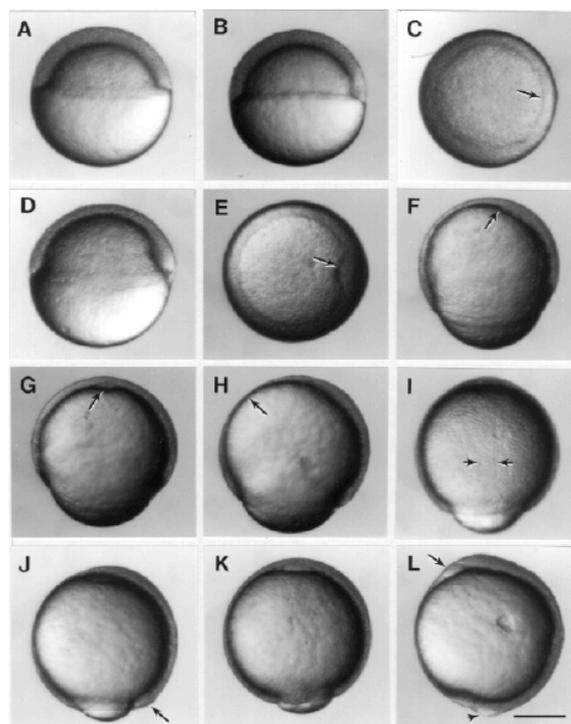


**Figura 7.** Stadi principali della fase di blastula di pesce zebra

Durante il periodo di blastula avvengono importanti processi come l'inizio dell'epibolia e l'entrata nella "*mid-blastula transition*", quest'ultima caratterizzata da una serie di cambiamenti come l'inizio della trascrizione del genoma zigotico, il rallentamento delle divisioni cellulari e l'inizio dei movimenti cellulari.

Il termine epibolia definisce, invece, un movimento uniforme di un foglietto di cellule di natura prevalentemente epiteliale sopra ad una superficie che si appresta a racchiudere, in questo caso, il tuorlo (Boncinelli, 2001).

Fase di gastrula (Figura 8): parallelamente all'epibolia, gli embrioni procedono nello sviluppo attraverso la gastrulazione, processo caratterizzato da movimenti cellulari morfogenetici di involuzione, convergenza ed evoluzione nonché dalla formazione degli assi di sviluppo embrionali. Alcuni degli stadi fondamentali della gastrulazione sono: 50% epibolia, *shield*, 80% epibolia e *tail bud* che sopraggiunge circa 10 ore dopo la fecondazione.



**Figura 8.** Stadi principali della fase di gastrula di pesce zebra  
**A.** 50% epibolia; **B.** *Germ ring*; **C.** Visione da polo animale del *Germ Ring* (indicato dalla freccia); **D, E** *Shield*; **F, G.** 70% epibolia; **H.** 75% epibolia; **I.** 80% epibolia; **J, K.** 90% epibolia; **L.** *tail bud*

Come accennato, la fase epibolica coincide con la fine della fase di blastula e prosegue parallelamente allo svolgersi della gastrulazione.

In questa fase l'epibolia è caratterizzata dalla migrazione e dall'assottigliamento del blastodisco sul nucleo. Il blastodisco cambia passando da un ammasso di cellule impilate (*high-piled*) a un multistrato cellulare a forma di tazza, di spessore pressoché uniforme.

Questo movimento è un *marker* molto utile per stadiare lo sviluppo che può essere classificato a seconda della percentuale di tuorlo coperto dal blastoderma (percentuale di epibolia). Quando il blastoderma ha coperto circa metà della superficie del tuorlo, si ha un ispessimento lungo tutto il margine del blastoderma che si sta muovendo. Questo ispessimento è detto *germ ring*; una volta formato quest'ultimo, le cellule dei due strati cellulari che lo formano si intercalano a formare un ispessimento localizzato, lo *shield*. Questa struttura è fondamentale nello stabilire ed organizzare l'asse dorso-ventrale nel pesce. Durante la formazione dello *shield* l'epibolia si arresta temporaneamente.

A livello molecolare ci sono dei segnali molto complessi che regolano lo sviluppo del piano corporeo lungo questo asse e le variazioni nell'equilibrio tra fattori ventralizzanti e dorsalizzanti sono fondamentali per un corretto sviluppo. Un esempio di proteine che agiscono in modo opposto sono *bmp2b* (*bone morphogenetic protein 2b*) che favorisce la ventralizzazione e *chd* (*chordin*) che lega *bmp2b* e lo inattiva, regolandone quindi la funzione. La mutazione del gene o la inibizione della funzione di *chordin* inducono quindi lo sviluppo di un fenotipo ventralizzato (Gilbert, 2000).

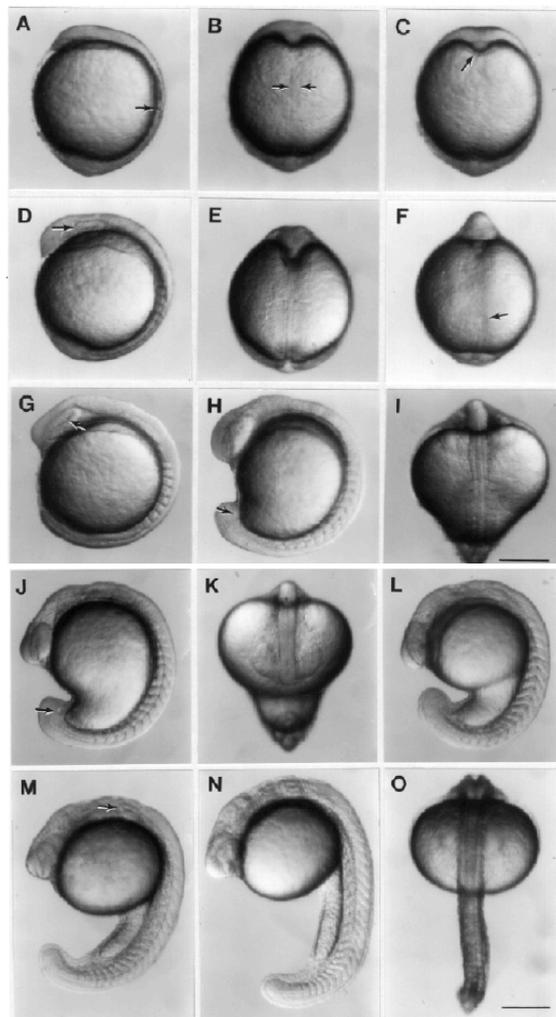
L'epibolia continua per tutta la gastrulazione fino a quando il tuorlo viene completamente ricoperto (*tail bud*).

Durante la gastrulazione vengono gettate le basi per la formazione del sistema nervoso in quanto i tre foglietti embrionali, ectoderma, endoderma e mesoderma, attraverso specifici movimenti cellulari, vengono correttamente posizionati uno rispetto all'altro, instaurando così nuove relazioni tessutali.

Una volta conclusosi questo processo, l'embrione va incontro ad una nuova fase di sviluppo, nota come neurulazione, che porta alla formazione del tubo neurale. Tale struttura è di origine ectodermica e successivamente si organizzerà a formare l'encefalo e il midollo spinale.

Fase della somitogenesi (Figura 9): questo periodo è caratterizzato dalla formazione dei somiti; durante questa fase l'embrione si accresce lungo l'asse antero-posteriore e cominciano ad essere visibili gli organi anche se ancora rudimentali. A livello cellulare comincia il differenziamento morfologico e, ultimato questo, i primi movimenti embrionali.

I somiti sono segmenti mesodermici che si formano in coppie, bilateralmente all'asse antero-posteriore ogni 30 minuti circa durante l'allungamento dell'embrione e sono formati dal riarrangiamento di *cluster* di cellule del mesoderma presomitico mesenchimale.



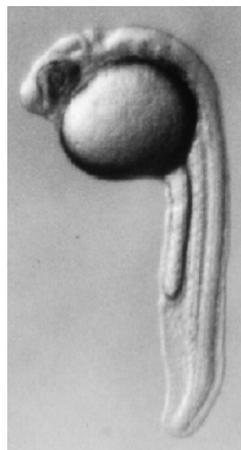
**Figura 9.** Stadi principali della fase di somitogenesi di pesce zebra

**A.** 2 somiti; **B.** visione dorsale della fase di sviluppo 2 somiti. Tra le frecce è mostrata la notocorda; **C.** Visione ventrale della fase di sviluppo 2 somiti; **D, E.** 4 somiti; **F.** 5 somiti; **G.** 8 somiti; **H.** 13 somiti; **J, K.** 15 somiti; **L.** 17 somiti; **M.** 20 somiti; **N, O.** 25 somiti

Durante la segmentazione, il numero dei somiti viene utilmente impiegato per stadiare lo sviluppo embrionale. I somiti si sviluppano in direzione antero- posteriore e la velocità di somitogenesi è relativamente costante; per questo è possibile distinguere le fasi di questo periodo e dividerle in: 1 somite, 5 somiti, 15 somiti e 20 somiti (circa 19 hpf).

Fase di faringula e *hatching* (schiusa): nel periodo successivo alla somitogenesi l'embrione continua ad accrescersi velocemente e ad allinearsi; l'angolo tra la testa e il tronco, infatti, diminuisce passando da 120° (24 hpf, figura 10) a 25° (72 hpf). In questo lasso di tempo avvengono ulteriori modificazioni nell'embrione che determinano una progressione nell'organogenesi. Per esempio, si ha lo sviluppo dell'intestino e della branchie, la comparsa di cellule pigmentate, la definizione delle cavità cardiache; è possibile osservare lo scorrimento del sangue nell'apparato circolatorio e contrazioni ritmiche coordinate che mimano i movimenti per il nuoto. Il completamento dello sviluppo embrionale porta alla liberazione dal corion e all'assorbimento del tuorlo rendendo l'animale progressivamente autonomo nel movimento e nell'alimentazione.

Dopo il terzo giorno di sviluppo comincia la fase larvale che, concludendosi, determina il passaggio allo stadio adulto (circa 90 giorni dopo la fecondazione).



**Figura 10.** Embrione di pesce zebra a 24 ore dalla fecondazione

## 5. Il sistema multi-trascritto *CYYR1*: modello genico nello studio della variabilità di espressione dei geni di HSA21

Come evidenziato dalla prima parte dell'introduzione a questa tesi, la ricerca di base si rivolge oggi ad una valutazione più precisa della variazione complessiva dei livelli di espressione genica, in cellule di individui affetti da DS o utilizzando modelli animali prevalentemente di topo trisomico. Lo scopo finale di questo approccio è comprendere come varino i livelli di espressione genica sia per quanto concerne la espressione di geni presenti in triplice copia sia per quelli euploidi. L'indicazione chiara di questi ultimi anni è che molti *loci* genici, ma non tutti, risentano almeno a livello dell'mRNA, di una reale sovraespressione (Sommer *et al.*, 2008; Lockstone *et al.*, 2007; Altug-Teber *et al.*, 2007; Conti *et al.*, 2007; Lott *et al.*, 2005; Helguera *et al.*, 2005). Questi studi hanno prodotto un insieme di dati la cui interpretazione è veramente molto complessa, soprattutto alla luce della inattesa complessità nella struttura, espressione e regolazione, anche epigenetica, della informazione contenuta nel genoma umano. La ricerca dunque è ancora molto aperta in questo campo e chiaramente indica come futura direzione la necessità di definire la funzione di tutti i geni del cromosoma 21 e comprendere il ruolo della variabilità di espressione e i meccanismi molecolari che sottendono a tali variazioni e che possono contribuire alla generazione del fenotipo caratteristico della DS.

Nel presente lavoro, si è voluto prendere in considerazione i geni identificati come sovraespressi in modo significativo e stabile nei soggetti DS. Da un recente studio di analisi ad ampio spettro sui geni del HC21 (Prandini *et al.*, 2007), in particolare, tra quelli a minore fluttuazione nella espressione genica è stato descritto il gene *cysteine/tyrosine-rich 1* (*CYYR1*).

Il gene *CYYR1* (*cysteine and tyrosine-rich protein 1*) umano è stato identificato e clonato nel Laboratorio di Genomica del Dipartimento di Istologia, Embriologia e Biologia Applicata dell'Università di Bologna (Reymond *et al.*, 2001; Vitale *et al.*, 2002). Questo gene è localizzato sul cromosoma 21 umano (21q21.2), è formato da 4 corti esoni e 3 lunghi

introni (di cui il secondo si estende per più di 85 Kb, una dimensione molto elevata in confronto alla media degli introni del genoma umano), ha una ORF (*open reading frame*) di 462 nucleotidi e codifica per una proteina di 154 amminoacidi. E' da notare che nell'ambito di progetti di clonaggio massivo (NEDO *human sequencing project*) è stato schedato un mRNA (GenBank #AK054581), più esteso in 5' che identificherebbe un prodotto proteico putativo di 262 amminoacidi.

La recente caratterizzazione del *locus CYYRI* ha determinato la presenza di numerose isoforme con conseguente necessità di un precisa nomenclatura concordata con HUGO (*Human Genome Organization*) secondo le linee guida per la nomenclatura dei geni umani (HGNC Guidelines, <http://www.genenames.org/>). Tale nomenclatura, nella presente tesi, verrà dettagliata per la prima volta.

La forma di riferimento (GenBank #NM\_052954), per prima identificata ed ufficialmente denominata *CYYRI* si riferisce alla isoforma che comprende i 4 esoni sopra detti.

Tramite l'analisi bioinformatica della proteina predetta codificata da *CYYRI* è stato possibile osservare, grazie al programma ClustalW di allineamento di proteine, che la proteina umana mostra identità e somiglianza con proteine di numerosi esseri viventi a partire da *Mus musculus* (81% di identità e 86% di somiglianza) fino a *Syngasteria tropicalis* (69% di identità e 78% di somiglianza).

Tramite metodi di predizione di strutture proteiche è stata identificata la presenza in *CYYRI* di un dominio transmembrana e, caratteristica strutturale fondamentale, la presenza di un dominio centrale ricco in tirosina e cisteina (CCSYAY) fortemente conservato dai vertebrati inferiori all'uomo. Questo dominio è stato proposto come segnale che contraddistingue questa nuova famiglia di prodotti proteici. *CYYRI*, infatti, non mostra somiglianze con altre proteine conosciute e il suo ruolo all'interno della cellula è a tutt'oggi sconosciuto. *CYYRI* appartiene, quindi, ad una famiglia genica fortemente conservata nell'evoluzione che nell'uomo ha un unico rappresentante, ma che è presente nei pesci, negli anfibi e nei mammiferi. All'estremità carbossi-terminale predetta è stata riscontrata la presenza di tre motivi PPPY

conservati in tutte le proteine della famiglia genica di *CYYRI* e studiati per le possibili implicazioni tra elementi proteici del citoscheletro e matrice extracellulare (Pereboev *et al.*, 2001; Ilsley *et al.*, 2002).

L'espressione di questo gene nell'uomo è stata analizzata con Northern blot e RT-PCR (*reverse transcription-polymerase chain reaction*, ovvero reazione a catena della polimerasi in seguito a retrotrascrizione) in un' ampia gamma di tessuti e organi (polmone, placenta, intestino tenue e colon, fegato, rene, milza, timo, muscolo scheletrico, cuore e cervello); l'unica popolazione cellulare che sembra non esprimere *CYYRI* sono i leucociti del sangue periferico.

In questo laboratorio, inoltre, è stato clonato e mappato l'ortologo murino di *CYYRI* che nel topo è localizzato sul cromosoma 16 e codifica per una proteina di 165 amminoacidi.

Infine, sono state fatte delle ricerche in banche dati di EST che hanno permesso l'identificazione di segmenti del trascritto di *cyrr1* nel pesce zebra adatti alla progettazione di *primers* per il clonaggio del gene in *zebrafish*. L'allineamento del prodotto proteico di 158 amminoacidi con quello di topo e uomo ha mostrato rispettivamente 57% di identità (73% somiglianza) e 58% di identità (74% somiglianza). In accordo con gli studi precedenti, anche nel pesce zebra è stato trovato il dominio conservato CCSYYAY e un dominio transmembrana nella stessa posizione identificata nell'uomo e nel topo.

Questi risultati costituiscono il punto di partenza per intraprendere una analisi dettagliata del gene *CYYRI*, che può essere considerato un modello genico nello studio della variabilità di espressione dei geni del cromosoma 21 umano, trattandosi di un sistema multi-trascrizione che occupa uno spazio fisico cospicuo pari allo 0,3% dell'intero cromosoma 21.

## **SCOPO DELLA TESI**

L'ambito della presente ricerca è stata la caratterizzazione dal punto di vista strutturale e funzionale di nuovi geni umani, con particolare attenzione rivolta ai geni del cromosoma 21, attraverso un approccio metodologico integrato.

La sequenza nucleotidica del cromosoma 21 umano (HC21) è stata determinata con alta accuratezza e pubblicata nel 2000 (Hattori *et al.*). Il numero totale dei geni (codificanti per proteine e non codificanti) presenti sul braccio lungo del cromosoma 21, non è ancora completamente conosciuto e per molti di questi *loci* genici non risulta completa la caratterizzazione funzionale

Il presente progetto di ricerca sottolinea la necessità di proseguire l'analisi genomica della sequenza del cromosoma 21 umano per una completa determinazione del suo contenuto genico. Inoltre si pone a più ampio raggio come studio avanzato del genoma umano, con un particolare riferimento ai geni del cromosoma 21, nel tentativo di portare contributi conoscitivi eventualmente utili a chiarire la complessa, ed ancora ampiamente sconosciuta, patogenesi molecolare della Sindrome di Down (DS) correlata con la trisomia 21.

In particolare, il lavoro si è articolato in tre parti: la caratterizzazione del *locus* genico umano *CYYRI* (*cysteine and tyrosine-rich protein 1 gene*), gene identificato e clonato nel 2002 nel laboratorio di Genomica del Dipartimento di Istologia, Embriologia e Biologia Applicata (Vitale *et al.*, 2002) e localizzato sul cromosoma 21; l'utilizzo del modello animale di *zebrafish* per lo studio funzionale dell'omologo di *CYYRI* e lo studio di sequenze non codificanti altamente conservate, coinvolte nella regolazione genica ed infine lo sviluppo di un *software* in grado di produrre mappe cromosomiche di espressione, al fine di inserire lo studio del gene *CYYRI* in un'analisi a più ampio raggio sui geni del Cromosoma 21.

## **1. CARATTERIZZAZIONE DEL *LOCUS* GENICO UMANO *CYYR1***

La prima parte del presente studio, si proponeva di effettuare la caratterizzazione del *locus* genico umano *CYYR1*, uno dei geni più estesi sul Cromosoma 21.

Il *locus* risulta molto complesso in quanto esistono numerose isoforme, codificanti per proteine putative diverse e che, probabilmente, mediano funzioni biologiche differenti; inoltre sono presenti trascritti in antisenso che arricchiscono ulteriormente la complessità del tratto genomico oggetto di studio.

In primo luogo, è stato quindi necessario effettuare esperimenti di clonaggio dei trascritti, identificati grazie a strumenti bioinformatici e *database* disponibili in rete. Successivamente l'analisi dell'espressione genica ha richiesto la messa a punto di tecniche quali *Real-time* PCR e RT-PCR quantitativa relativa. Tramite queste tecniche, infine, gli studi di espressione dei trascritti del *locus CYYR1* sono stati condotti in diverse linee cellulari, allo scopo di ottenere modelli cellulari che esprimessero in maniera significativa il gene di interesse e potessero essere impiegati per indagare aspetti funzionali di *CYYR1*.

## **2. IL MODELLO ANIMALE DI *ZEBRAFISH* PER LO STUDIO FUNZIONALE E DELLE REGOLAZIONE GENICA**

Nella seconda parte dello studio è stato utilizzato il modello animale del pesce zebra, che permette di studiare geni di interesse (nel presente lavoro, il gene *cyrr1*) nel contesto dell'organismo *in toto* e di verificarne aspetti funzionali.

In primo luogo, è emersa la necessità di effettuare una analisi sistematica per l'identificazione di un gene *housekeeping* "ideale" da utilizzare come gene di riferimento in studi di espressione sul pesce zebra. Il secondo obiettivo è stato quello di mettere a punto metodiche per studi di modulazione genica, quali la sovraespressione genica e la realizzazione di organismi *knockdown*. A queste metodiche utili per indagare la funzione

del gene *cyrr1*, è stata affiancata la *Real-time* PCR per verificare l'espressione del gene e una sua possibile modulazione, in particolare durante lo sviluppo embrionale.

Il modello animale di *zebrafish* è stato utilizzato anche per la caratterizzazione di sequenze conservate evolutivamente e correlate a funzioni regolatorie (CNE, *conserved non coding element*). Le CNEs, identificate in primo luogo con metodi bioinformatici in *Ciona intestinalis*, e correlate a geni coinvolti nello sviluppo embrionale, sono state utilizzate per la progettazione di *primers* specifici. Questi *primers* sono quindi stati utilizzati in PCR ai fini di clonare *in vitro* le sequenze CNEs di *C. intestinalis*, e gli ampliconi ottenuti sono stati utilizzati in esperimenti di microiniezione in embrioni di pesce zebra, per verificare se tali sequenze mantessero una funzione regolatoria anche nei vertebrati.

Esiste un quadro molto complesso in cui sono coinvolti molteplici meccanismi molecolari in grado di influenzare l'espressione genica e quindi determinare lo sviluppo embrionale, la funzione di cellule, tessuti ed organi o provocare patologie.

Con il progredire della definizione dei *pathway* molecolari considerati potrà essere interessante verificare se all'interno di queste cascate di segnale, sono coinvolti anche elementi regolatori; in particolare le indagini sul gene *cyrr1* fanno ipotizzare un suo coinvolgimento a livello di sviluppo embrionale e moltissimi geni chiave a questo livello sembrano essere regolati proprio da CNEs.

### **3. SVILUPPO DI UN *SOFTWARE* IN GRADO DI PRODURRE MAPPE CROMOSOMICHE DI ESPRESSIONE (TRAM)**

Lo studio sul gene *CYYRI* rientra in un'indagine a più ampio raggio sui geni del Cromosoma 21, coinvolti nella patogenesi della Sindrome di Down, ma anche nella tumorigenesi e nello sviluppo embrionale.

L'analisi dell'espressione genica fornisce numerose informazioni che possono contribuire ad associare un gene o un insieme di geni a un particolare meccanismo biologico normale o patologico

La meta-analisi di studi indipendenti potrebbe fornire elementi utili che mettano in evidenza regioni cromosomiche caratterizzate da una consistente e stabile sovraespressione o sottoespressione genica. Non esistono, ad oggi, strumenti bioinformatici in grado di effettuare una meta-analisi di questo tipo in modo semplice e facilmente fruibile. Per questo la messa a punto di un *software* in grado di produrre mappe cromosomiche di espressione a partire da dati di analisi genica globale potrebbe risultare molto utile per l'interpretazione di dati così numerosi, quali l'insieme delle analisi di espressione prodotte dal lavoro della comunità scientifica e spesso liberamente consultabili in rete.

Questo approccio bioinformatico potrebbe consentire di identificare in modo predittivo intere regioni cromosomiche i cui geni hanno lo stesso schema di espressione e profili stabili di sovra/sottoespressione in campioni biologici.

Su queste basi, nel presente studio è stato messo a punto un *software* in grado di produrre mappe cromosomiche di espressione a partire da dati di analisi genica globale (TRAM, *Transcriptome Mapper*). Questo programma può essere utilmente impiegato per approfondire le conoscenze riguardanti l'espressione genica durante lo sviluppo embrionale di *zebrafish* e trovare ulteriori dati da affiancare a quelli ottenuti mediante la *Real-time* PCR, consentendo una meta-analisi di dati di espressione genica disponibili in rete.

Allo scopo di ampliare la ricerca sull'espressione e la caratterizzazione dei geni del Cromosoma 21, potrebbe infine essere utilizzato per la meta-analisi di *microarray* per confrontare l'espressione genica di campioni trisomici ed euploidi.

## **MATERIALI E METODI**

## 1. CARATTERIZZAZIONE DEL *LOCUS* GENICO UMANO *CYYR1*

### 1.1. Analisi bioinformatica

La sequenza di riferimento utilizzata in questo studio è stata #NM\_052954, corrispondente alla sequenza del gene *CYYR1* (*CYYR1*-1,2,3,4) identificato e clonato presso il Laboratorio di Genomica del Dipartimento di Istologia, Embriologia e Biologia Applicata dell'Università di Bologna (Reymond *et al.*, 2001; Vitale *et al.*, 2002).

Il database di EST umane (Boguski *et al.*, 1993) è stato interrogato mediante BLASTN 2 (Altschul *et al.*, 1997) utilizzando la sequenza #NM\_052954 ed è stato possibile individuare numerosi splicing alternativi. Molti di essi sono riportati come EST anche nel database *UCSC Genome Browser*.

Inoltre, sono riportate anche EST che si riferiscono a trascritti in antisenso rispetto al gene. In particolare, è stato indagato il trascritto #AK125338, un RNA relativamente lungo (3.412 bp), embricato col gene *CYYR1* in antisenso; per questo è possibile definire il *locus* genico un *locus* multitrascritto.

### 1.2. Materiale biologico

Sono stati utilizzati pool di RNA umani commerciali della BD Clontech (Palo Alto, CA, USA), provenienti da diversi tipi tissutali (cuore, cervello, testicolo, sangue, intestino). Alcuni saggi sono stati anche condotti su RNA estratto da vari tessuti di origine neuroendocrina, tra cui pancreas, fegato, ghiandola mammaria e collo.

Inoltre, è stato impiegato RNA estratto a partire da linee cellulari tumorali, quali SJNKP (neuroblastoma), SKNMC (neuroblastoma), HT29 (adenocarcinoma del colon), MCF7 (adenocarcinoma della mammella), SK-MEL28 (melanoma), G361 (melanoma) ed infine linee tumorali di osteosarcoma (SaOS e U2OS). Sono state anche utilizzate due linee primarie fibroblastiche, una trisomica per il cromosoma 21 (CCL-54) e una euploide (CCL-110).

Le linee cellulari tumorali sono state gentilmente fornite dalla Prof.ssa Rocchi e dal Prof. Lollini (Dipartimento di Patologia Sperimentale, Università di Bologna).

Le linee cellulari primarie di fibroblasti (CCL-54 e CCL-110) sono state acquistate dalla ATCC e conservate in azoto liquido. Si è quindi proceduto allo scongelamento e successiva semina in fiasche T70 in terreno MEM (Minimum Essential Medium Eagle, SIGMA) addizionato al 10% con FBS scomplementato (Micro-tech), aminoacidi non essenziali (MEM-Non essential Amino Acid Solution, 100X, SIGMA), vitamine (MEM Eagle Vitamin Mixture, 100X Micro-Tech), L-glutammina (100x), Sodio piruvato 100mM (SIGMA).

Le cellule sono state mantenute in incubatore fino al raggiungimento della confluenza. In seguito, si è proceduto all'estrazione dell'RNA seguendo il protocollo TriReagent (Ambion): le cellule sono state staccate dalle fische utilizzando la tripsina, successivamente si è aggiunto 1 ml della soluzione TriReagent, contenente guanidina isotiocianato e fenolo; tale solvente permette la formazione di complessi tra l'RNA, il guanidinio e le molecole d'acqua, ed esclude il DNA e le proteine, ostacolando la formazione di interazioni idrofiliche del DNA e delle proteine stesse. Le cellule con TriReagent sono state lasciate incubare a temperatura ambiente per 5'; sono state poi trasferite in microprovette da 2 mL. Sono stati aggiunti 200  $\mu$ L di cloroformio a cui è seguita un'altra incubazione a temperatura ambiente per 10 minuti e una centrifugazione a 12000xg per 15' a 4°C. In seguito a questa centrifugazione sono state ottenute tre fasi:

- Rossa organica contenente le proteine (in basso)
- Interfase contenente il DNA (al centro)
- Fase acquosa incolore contenente l'RNA (in alto)

Dopo aver recuperato la fase acquosa e averla posta in una nuova provetta, sono stati aggiunti 500  $\mu$ L di isopropanolo e si è centrifugato a 12000xg per 8' a 4°C per far precipitare l'RNA e, rimosso il surnatante, il pellet è stato lavato con 1 ml di etanolo al 75% procedendo infine a centrifugazione a 7500xg per 5' a 4°C. Il pellet è stato asciugato all'aria e

poi risospeso in 20  $\mu\text{L}$  di Acqua DEPC (Sigma). L'RNA così estratto è stato conservato a  $-80^{\circ}\text{C}$  fino al momento dell'utilizzo.

L'RNA di una seconda linea cellulare di osteosarcoma (U2OS-1) ci è stato gentilmente fornito dalla Prof.ssa Valvassori (Dipartimento di Istologia, Embriologia e Biologia Applicata, Università di Bologna). Purezza e concentrazione dei campioni di RNA sono state determinate mediante lettura con Nanodrop® e corsa elettroforetica in un gel d'agarosio all'1% in presenza di etidio bromuro (0.5  $\mu\text{g}/\text{ml}$ ) che ha permesso di valutare la presenza di bande integre e di peso molecolare corrispondente a quello degli RNA ribosomiali.

### **1.3. Progettazione dei *primers* per il clonaggio genico**

I *primers* utilizzati per l'amplificazione sono stati progettati sulla base delle sequenze di riferimento pubblicate nelle banche dati, relative all'mRNA di diverse isoforme di *CYYRI* (numeri di accesso GenBank: #CN\_402794; #CN\_402791; #DB\_268533; #DB\_120436; #AK304124) e di trascritti in antisenso (#AK125338). Confrontando queste sequenze con la sequenza di DNA genomico #AP001696 (Cromosoma 21q, sezione 40/105), è stato possibile individuare la suddivisione esoni-introni. Le sequenze nucleotidiche e aminoacidiche sono state poi confrontate con quelle della prima isoforma dichiarata e utilizzata in questo studio come forma di riferimento *CYYRI-1,2,3,4* (#NM\_052954).

La progettazione dei *primers* è stata effettuata utilizzando il software Amplify 3X (Engels, 1993), disponibile liberamente sulla rete Internet e utilizzato in accordo con le istruzioni del programmatore. I *primers* sono stati progettati in modo che si appaiassero su esoni diversi e non sul medesimo esone per evitare di includere porzioni genomiche.

Mediante il software BLASTN ([www.ncbi.nlm.nih.gov/blast](http://www.ncbi.nlm.nih.gov/blast)) è stata valutata la validità biologica, per escludere eventuali somiglianze con altre zone del genoma oltre a quella attesa.

Gli oligonucleotidi corrispondenti sono stati fatti sintetizzare presso un servizio commerciale (Sigma-Aldrich, UK) e il preparato liofilizzato è stato risospeso in acqua bidistillata sterile alla concentrazione di 1  $\mu\text{g}/\mu\text{L}$ .

In Tabella 6 sono riportati i *primers* utilizzati.

#### 1.4. RT-PCR

La reazione a catena della polimerasi dopo retrotrascrizione (RT-PCR) è una metodica che permette di analizzare campioni di RNA che vengono in un primo momento retrotrascritti, cioè copiati in cDNA, e poi amplificati.

Tutte le retrotrascrizioni sono state effettuate secondo metodiche standard (Davis *et al.*, 1994). Due µg di RNA sono stati retrotrascritti in un volume finale di 50 µL in presenza di 500 µM di ciascun tipo di desossinucleotide trifosfato (dNTP), 400 U di trascrittasi inversa (Murine Moloney Leukemia Virus) e relativo tampone di retrotrascrizione 1× (Promega, Madison, WI). Per l'innesco della reazione sono state usate sequenze di *random* esameri con concentrazione finale di 2 µM e di oligonucleotidi dT (2.5 µM finale). L'incubazione per far avvenire la reazione è stata effettuata a 37°C per il periodo di 1 ora, di seguito è stato denaturato l'enzima ponendo la miscela a 70°C per 10 minuti per porre termine alla reazione.

Tutti i reagenti per la RT-PCR sono stati acquistati dalle ditte Promega (Madison, WI, USA) e TaKaRa (Shiga, Giappone).

La tappa successiva è stata l'amplificazione del segmento precedentemente retrotrascritto attraverso la reazione a catena della polimerasi.

Per amplificare il segmento nucleotidico di interesse si utilizza la PCR (*polymerase chain reaction*), reazione che si basa sull'impiego della DNA polimerasi del batterio *Thermus aquaticus* (*Taq*), che ha la proprietà di resistere ad alte temperature senza essere degradata.

La maggior parte degli esperimenti di PCR sono stati condotti in volumi finali di 50 µL, contenenti 5 µL del prodotto della retrotrascrizione, 1 U di Taq Polimerasi (TaKaRa, Shiga, Japan) con i relativi reagenti (0,2 mM di ciascun dNTP, 2 mM MgCl<sub>2</sub>, 5 µL di buffer di PCR 5×), e 0,3 µM di ciascun primer (senso e antisenso) che serve da innesco alla polimerasi.

La preparazione completa viene sottoposta a cicli successivi in cui si alternano diverse temperature. Nelle PCR fatte per questo studio è stato usato un programma che prevede un primo intervallo di 2 min. a 94°C per la denaturazione dei filamenti e poi, ciclicamente, 30 sec. a 94°C, sempre per la denaturazione, 30 sec. a 58-66°C per permettere l'appaiamento dei *primers* e l'allungamento della catena, 30 sec. a 72°C, temperatura alla quale l'efficienza della *Taq* polimerasi è massima e quindi continua l'allungamento. Questo si ripete fino a 20-45 cicli a cui seguono 7 min. a 72°C per terminare l'estensione di eventuali filamenti incompleti (Tabella 6).

Lo strumento utilizzato è un “*thermal cycler*”, una macchina che permette di impostare una variazione programmata della temperatura nel tempo. Sono stati allestiti anche controlli positivi (con i *primers* per l'mRNA di *B2M*) per verificare la presenza di RNA amplificabile e controlli negativi con acqua.

### **1.5. Elettroforesi, purificazione e quantificazione prodotti di RT-PCR**

I risultati dell'amplificazione sono stati osservati tramite elettroforesi su gel di agarosio all'1,5 % in Tris-Acetato-EDTA (TAE) 1×. Per evidenziare i prodotti è stato aggiunto Etidio Bromuro (EtBr) alla concentrazione finale di 0,02 µg/mL.

Come marcatore è stato utilizzato “GeneRuler” (Fermentas, Hannover, Maryland, USA), che contiene frammenti di DNA di lunghezza definita che si separano durante la corsa; questo permette di determinare la lunghezza del campione in base al confronto tra la posizione della relativa banda e quelle del marcatore.

La corsa è stata eseguita a voltaggio compreso tra 80 e 120V e a fine corsa il gel è stato osservato al transilluminatore UV e fotografato con una camera IBI QSP.

Nel caso delle amplificazioni effettuate per isolare l'isoforma *CYYRI-1,2,4*, è stato necessario effettuare il carotaggio delle bande ad altezza attesa, ma presenti insieme ad altri prodotti di amplificazione, in quanto i *primers* utilizzati non erano isoforma-specifici.

Le bande di dimensione attesa, sono state purificate utilizzando il kit della ditta Sigma (“Gene Elute” PCR Clean-up kit).

Al fine di quantificare il contenuto in DNA, i campioni purificati sono stati fatti migrare su gel di agarosio 1,5% in presenza di “Marker 5” (Fermentas, Hannover, Maryland, USA), marcatore che contiene frammenti di DNA di lunghezza definita e di cui è stata calcolata la quantità.

### **1.6. Sequenziamento del cDNA**

Ottenuta una banda singola di DNA dalla corsa elettroforetica (indicatore della presenza di un solo trascritto), dopo la purificazione su colonnine Sigma, si può procedere alla preparazione del campione per il sequenziamento.

Come innesco della reazione di Sanger sono stati utilizzati gli stessi *primers* con cui i cDNA erano stati originariamente ottenuti nella reazione di PCR e i “BigDye chain-terminator”. I campioni da sequenziare contengono tipicamente Con i “BigDye” bisogna usare 30-40 ng di DNA in massimo 15  $\mu\text{L}$  di acqua bidistillata sterile. Le provette inviate al sequenziamento contengono, in un volume finale di 16  $\mu\text{L}$ , 30-40 ng di cDNA e 4 pmol di *primer*. A ciascuna provetta a questo punto verranno aggiunti 2  $\mu\text{L}$  di “BigDye Terminator Mix”, che contiene enzima, dideossinucleotidi e 2  $\mu\text{L}$  di buffer 5 $\times$ .

La reazione di *cycle sequencing* consta di 25 cicli formati da: 96°C per 30 sec., 15 sec. alla temperatura di *annealing* della PCR, 60°C per 4 min.. Terminati i 25 cicli il campione viene lasciato a 4°C al Una volta effettuata la reazione di Sanger i campioni sono stati sequenziati con il sequenziatore automatico ABI310 (Perkin-Elmer, Foster City, CA).

Le sequenze ottenute sono state confrontate col programma BLAST con le sequenze degli mRNA registrati nelle banche dati informatiche.

### **1.7. Clonaggio in vettore, trasformazione in *E.Coli* e recupero del plasmide**

Nella impossibilità di ottenere un prodotto specifico di amplificazione per l'isoforma *CYYRI-1,2,2bis,3,4* (#AK304124) si è dovuto procedere con la metodica alternativa del clonaggio *in vivo*.

I prodotti di amplificazione sono stati clonati nel vettore pCR2.1 fornito dalla ditta (Kit TA Cloning, Invitrogen). Il vettore pCR2.1 si estende per circa 3.929 bp ed include: il promotore del gene *lac*, che permette l'espressione da parte della cellula batterica del gene *lacZ*-alfa; i primi 146 aminoacidi della Beta-galattosidasi, utile per la selezione delle colonie bianche (vettore con inserto) rispetto a quelle blu (vettore senza inserto); i geni per la resistenza alla kanamicina ed alla ampicillina; l'origine di replicazione pUC ori per la propagazione in *E. coli*; un promotore T7 per la trascrizione e la traduzione dell'RNA e un MCS (Multiple Cloning Site), sito di clonaggio multiplo che rende semplice il subclonaggio del prodotto dell'RT-PCR all'interno del vettore; siti di attacco per i *primers* M13 (forniti dal Kit per l'amplificazione dell'inserto di interesse).

Il vettore e i prodotti di amplificazione purificati sono stati utilizzati per effettuare una reazione di ligasi. Per ogni reazione, è stata stimata la quantità di prodotto di PCR da utilizzare mantenendo un rapporto molare inserto:vettore 1:1, secondo la seguente formula:

$$\text{Xng prodotto PCR} = (\text{Ybp prodotto PCR}) (50 \text{ ng vettore pCR2.1}) / (\text{bp vettore pCR2.1} \sim 3900)$$

La reazione di ligasi è stata effettuata in un volume finale di 10  $\mu\text{l}$  contenente x  $\mu\text{l}$  prodotto di PCR, 1  $\mu\text{l}$  *Ligation Buffer* 10X, 2  $\mu\text{l}$  vettore pCR2.1 (25ng/ $\mu\text{l}$ ), acqua, 1  $\mu\text{l}$  T4 DNA ligasi (4.0 U) ed incubata a 14°C *overnight*.

I prodotti ottenuti sono stati inseriti in cellule batteriche di *E. coli* chimicamente competenti (KIT One Shot TOP10 competent cells, Invitrogen) tramite la fase di *heat shock* (30 secondi a 42° e raffreddamento immediato in ghiaccio). Le cellule sono state risospese in 250 mL di terreno liquido SOC (Tryptone 2%, Estratto di Lievito 0,5%, NaCl 10 mM, KCl 2,5 mM, MgCl<sub>2</sub>·6H<sub>2</sub>O 10 mM, Glucosio 20 mM) e poste in incubatore con agitatore a 37° per 1 ora. Dopo l'incubazione le

cellule sono state piastrate su terreno LB-Agar (Luria-Bertani) in presenza di kanamicina e X-Gal e poste in incubatore 37° *overnight*.

Le colonie cresciute sono state analizzate mediante “*colony PCR*”. In breve, le colonie ben isolate sono state prelevate con un puntale sterile e trasferite ognuna in una microprovetta da 0,5 mL, contenente 50 µL di acqua sterile; successivamente sono state dissolte per agitazione con vortex. Per il salvataggio della colonia sono stati prelevati 10 µL della sospensione cellulare a cui sono stati aggiunti 200 µL di terreno liquido LB con antibiotico. Per effettuare l’analisi tramite PCR, le colonie sono state risospese e tenute per 5 minuti a 99°C. Dopo averle centrifugate a 14.000 rpm per 1 minuto a temperatura ambiente, sono stati trasferiti 10 µL del sopranatante nella reazione di PCR (volume finale 50 µL) per 25 cicli, usando i *primers* senso ed antisense specifici per il vettore pCR2.1, forniti nel kit TA Cloning (M13f e M13r), ed è stato utilizzato il programma: 2 min a 94°C, 25 cicli di amplificazione (30 sec a 94°C, 30 sec a 58°C, 30 sec a 72°C) e tappa finale di estensione per 7 min a 72°C. I prodotti di amplificazione ottenuti sono stati analizzati su gel di agarosio per identificare quelli di dimensioni compatibili alle isoforme indagate. Di seguito questi sono stati purificati e sequenziati, come già descritto, per confermare la sequenza nucleotidica.

## **2. IL MODELLO ANIMALE *ZEBRAFISH* PER LO STUDIO FUNZIONALE E DELLA REGOLAZIONE GENICA**

### **2.1. Pesce zebra (*Danio rerio*)**

Il modello animale utilizzato è il pesce zebra (*Danio rerio*), un pesce tropicale di piccole dimensioni, dal corpo affusolato e livrea a strie orizzontali gialle ben visibili sul fondo blu-argenteo. Per l'allevamento di questi pesci, provenienti dall'acquario di Bologna, sono state seguite le procedure standard (Westerfield, 1995). L’ambiente adibito alla stabulazione degli animali è stato opportunamente attrezzato con 2 armadi-acquari a 8 piani con vasche alimentate in continuo con un sistema di acqua ad opportuna composizione e osmolarità con ricircolo e filtri

meccanici e biologici (Figura 11). Il sistema garantisce oltre ad un costante ricambio d'acqua, il mantenimento costante della temperatura e la sterilizzazione dell'acqua mediante lampada UV. Sono stati eseguiti controlli settimanali dei parametri chimico-fisici che assicurano il benessere degli animali, quali il monitoraggio della temperatura, della osmolarità dell'acqua, del pH (tra 7 e 8), del KH, di nitriti e nitrati nonché le condizioni di ossigenazione delle vasche.

Gli esemplari adulti da riproduzione sono stati allevati in vasche da 5 l, divisi per generazione e ceppo biologico, mentre vasche da 3 l venivano utilizzate per fare crescere i giovani avannotti.

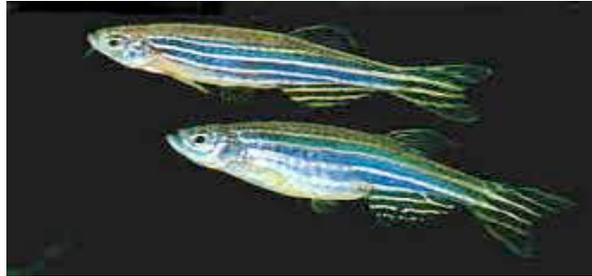


**Figura 11.** L'allestimento degli armadi-acquari presso il Dipartimento di Istologia, Embriologia e Biologia Applicata, Università di Bologna

### **2.1.1. Accoppiamento dei pesci e raccolta degli embrioni**

La temperatura ideale per l'accoppiamento e la produzione di uova è di circa 28°C. *Danio rerio* ha un fotoperiodo ottimale di 13 ore di luce e 11 ore di buio ed inizia il corteggiamento alle prime luci del giorno; queste condizioni sono state ricreate mediante illuminazione indipendente degli armadi controllata da un meccanismo di accensione e spegnimento automatico.

La femmina è leggermente più grande del maschio da cui si distingue facilmente solo quando accumula le uova e presenta quindi un ventre più prominente (Figura 12).



**Figura 12.** Esemplari maschile e femminile di *Danio rerio*

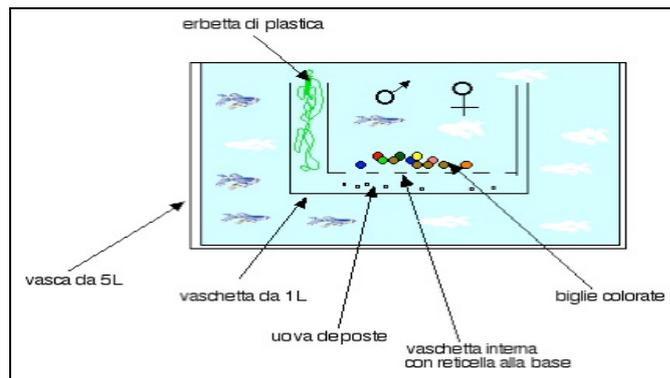
La fecondazione è esterna e avviene non appena la femmina, rincorsa dal maschio durante il rituale di corteggiamento, depone le uova sul fondo.

Poiché i pesci, una volta deposte le uova, tendono spontaneamente a cibarsi di esse, è necessario trasferire gli animali in un sistema costituito da un contenitore con fondo a rete, a sua volta collocato in una vaschetta che possa raccogliere le uova deposte ed eventualmente fecondate; tale sistema a doppia vasca viene riempito di biglie colorate per stimolare l'entrata dei pesci nella vaschetta più piccola (Figura 13A).

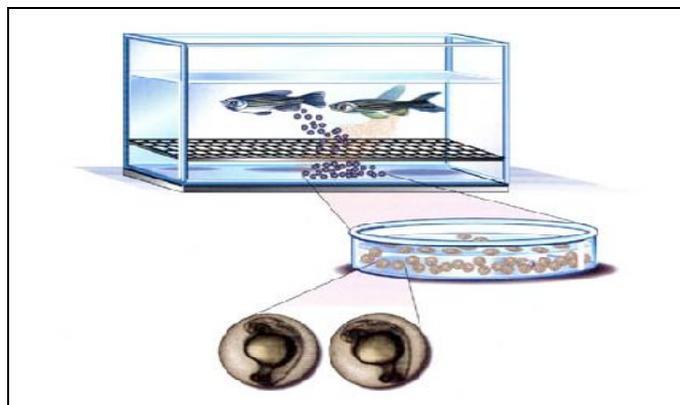
Le uova fecondate si presentano trasparenti e con un diametro di circa 1 mm. Dopo l'accoppiamento, con una pipetta Pasteur esse vengono raccolte, lavate con un tampone isosmotico e trasferite su piastre Petri (Figura 13B) con acqua pulita e possono essere poste in incubatore a circa 28°C con l'aggiunta di blu di metilene (2 ppm), un potente antibatterico e antifungino.

Una volta raggiunto il periodo di sviluppo desiderato, gli embrioni vengono osservati allo stereomicroscopio e selezionati gli stadi embrionali scelti per le fasi sperimentali successive.

**A.**



**B.**



**Figura 13.** **A.** Allestimento dei contenitori da riproduzione **B.** Raccolta embrioni in piastre Petri

## 2.2. Estrazione di acidi nucleici

### 2.2.1. Estrazione di RNA da pesce zebra

L'RNA di pesce zebra utilizzato per diversi studi, quali l'analisi dei geni *housekeeping* e l'analisi di espressione del gene *cyr1*, è stato ottenuto a partire da embrioni a diverso stadio di sviluppo e da tessuti di pesce adulto.

L'allestimento dei campioni ha previsto sia una fase di raccolta degli embrioni ottenuti dall'accoppiamento dei pesci zebra, sia una fase di prelievo in sede biotica di diversi tessuti adulti, seguita da estrazione, purificazione e quantificazione dell'RNA totale.

In particolare, per lo studio dei geni *housekeeping*, l'RNA è stato estratto da *pools* di 23-24 embrioni a diversi stadi di sviluppo: sfera (4 hpf), cupola (4 1/3 hpf), gastrula (*shield*, 6 hpf), 1-5 somiti (10 1/3 hpf), 48 hpf. Inoltre l'RNA è stato estratto da tessuti quali cervello, muscolo, epidermide, ossa, occhio e ovaio ottenuti dalla dissezione di singoli pesci adulti; per ogni tessuto, l'RNA è stato estratto da cinque pesci adulti, tranne che per l'RNA di ossa che è stato ottenuto da quattro pesci.

Per l'analisi di espressione di *cyrr1*, l'RNA è stato ottenuto a partire da circa 20 embrioni a diversi stadi di sviluppo: 16 cellule (1 1/2 hpf), 1000 cellule (3 hpf), sfera (4 hpf), cupola (4 1/3 hpf), epibolia al 50% (5 1/4 hpf), 1-5 somiti, 24 hpf, 48 hpf e larva di 5 giorni.

L'estrazione dell'RNA è avvenuta seguendo il protocollo TriReagent, già descritto in precedenza.

Purezza e concentrazione dei campioni sono state determinate mediante lettura con Nanodrop® e corsa elettroforetica in un gel d'agarosio all'1% in presenza di etidio bromuro (0.5 µg/ml) che ha permesso di valutare la presenza di bande integre e di peso molecolare corrispondente a quello degli RNA ribosomiali.

### **2.2.2. Estrazione di DNA genomico da *Ciona intestinalis***

Per l'analisi degli elementi regolatori non codificanti conservati (CNEs) è stato utilizzato DNA genomico estratto da *Ciona intestinalis*, seguendo il protocollo per l'estrazione di DNA da coda di roditori del Kit DNeasy Tissue (Qiagen).

In breve, circa 0,5 cm di organismo sono stati inseriti in una microprovetta da 1,5 ml. Si aggiungono 180 µl di tampone di lisi (*Buffer ATL*) e 20 µl di Proteinasi K. Dopo un periodo di incubazione a 55°C per lisare completamente il tessuto, si aggiungono 400 µl di *Buffer AL* ed etanolo per favorire l'adesione del DNA alla membrana della colonnina. Si trasferisce il campione nella colonnina posizionata all'interno di una microprovetta da 2 ml. Si effettuano lavaggi della colonnina con passaggi successivi di centrifugazione a 8000 rpm e aggiunta di tampone per lavare

la colonnina (*Buffer* AW1 e AW2). Il campione di DNA viene infine eluito in 200  $\mu$ l di soluzione di eluizione (*Buffer* AE).

Il DNA genomico così estratto è stato caricato su gel di agarosio 1% TBE e quantificato prima di procedere all'amplificazione delle CNEs da microiniettare in embrioni di pesce zebra.

## 2.3. RT-PCR

### 2.3.1. RT-PCR per l'identificazione di geni *housekeeping* in pesce zebra

Allo scopo di caratterizzare geni *housekeeping* per il modello animale di pesce zebra, è stata messa a punto una tecnica di RT-PCR quantitativa relativa (Vitale *et al.*, 2007; Facchin *et al.*, 2008).

Gli esperimenti di RT-PCR quantitativa relativa sono stati effettuati per 8 geni candidati, selezionati in base a diversi criteri: innanzitutto, è stata effettuata una analisi della letteratura per identificare i geni *housekeeping* maggiormente utilizzati in pesce zebra, in secondo luogo è stato interrogato il database UniGene per identificare i geni di pesce zebra col maggior numero di EST.

Infine, sono stati effettuati alcuni esperimenti di RT-PCR preliminari, allo scopo di escludere quei geni che mostrassero mancata espressione in uno o più tessuti/stadi embrionali utilizzati.

La progettazione dei *primers* è stata effettuata utilizzando il software Amplify 3X (Engels, 1993) secondo i criteri standard già descritti (Sharrocks, 1994). Le sequenze dei *primers* sono indicate in Tabella 7.

In breve, 2  $\mu$ g di RNA totale sono stati retrotrascritti in un volume finale di 50  $\mu$ l, utilizzando 500  $\mu$ M di ciascun tipo di desossinucleotide trifosfato (dNTP), 400 U di trascrittasi inversa (Murine Moloney Leukemia Virus) e relativo tampone di retrotrascrizione 1 $\times$  (Promega, Madison, WI). Per l'innesco della reazione sono state usate sequenze di random esameri con concentrazione finale di 2  $\mu$ M e di oligonucleotidi dT (2  $\mu$ M finale). L'incubazione per far avvenire la reazione è stata effettuata

a 37°C per il periodo di 1 ora, di seguito è stato denaturato l'enzima ponendo la miscela a 70°C per 10 minuti per porre termine alla reazione.

Tutti i reagenti per la RT-PCR sono stati acquistati dalle ditte Promega (Madison, WI, USA) e TaKaRa (Shiga, Giappone).

La tappa successiva è stata l'amplificazione del segmento precedentemente retroscritto attraverso la reazione a catena della polimerasi.

Gli esperimenti di PCR sono stati condotti in volumi finali di 25 µL, contenenti 2,5 µL del prodotto della retroscrittura, 0,5 U di Taq Polimerasi (TaKaRa, Shiga, Japan) con i relativi reagenti (0,2 mM di ciascun dNTP, 2 mM MgCl<sub>2</sub>, 2,5 µL di buffer di PCR 5×), e 0,3 µM di ciascun primer (senso e antisenso) che serve da innesco alla polimerasi.

La preparazione completa viene sottoposta a cicli successivi in cui si alternano diverse temperature. Nelle PCR fatte per questo studio è stato usato un programma che prevede un primo intervallo di 2 min. a 94°C per la denaturazione dei filamenti e poi, ciclicamente, 30 sec. a 94°C, sempre per la denaturazione, 30 sec. a 64°C per permettere l'appaiamento dei *primers* e l'allungamento della catena, 30 sec. a 72°C, temperatura alla quale l'efficienza della Taq polimerasi è massima e quindi continua l'allungamento. In esperimenti di PCR preliminari, sono stati valutati i prodotti di PCR dopo 20, 25, 30 e 35 cicli per determinare le condizioni che permettono la quantificazione dei prodotti di PCR per ogni gene candidato alla massima distanza dal ciclo corrispondente al plateau della reazione. Il numero di cicli specifico utilizzato per ogni gene candidato è indicato in Tabella 7. Lo strumento utilizzato è un “*thermal cycler*”, una macchina che permette di impostare una variazione programmata della temperatura nel tempo.

I risultati dell'amplificazione sono stati osservati tramite elettroforesi su gel di agarosio all'1,5 %, senza l'aggiunta di EtBr. Sono stati caricati 10 µL di prodotto di PCR. Come marcatore è stato utilizzato “Marker 5” (Fermentas, Hannover, Maryland, USA) a due diverse concentrazioni. A fine corsa, il gel è stato colorato immergendolo in una vaschetta contenente il tampone TAE e EtBr (0.5 µg/mL) e di seguito analizzato utilizzando Gel Doc 2000 Imaging System. Le immagini sono

state analizzate e quantificate utilizzando Quantity One software (Bio-Rad). I valori di intensità delle bande corrispondenti ai prodotti di PCR sono stati calcolati confrontandoli con una curva standard costruita a partire dai valori di concentrazione attribuiti ad almeno quattro bande del marcatore “Marker 5”.

Infine, i dati così ottenuti sono stati analizzati utilizzando diversi strumenti statistici.

Sono state calcolate media e deviazione standard di tutti i valori ottenuti per ogni gene e la deviazione standard è stata espressa anche come percentuale della media, allo scopo di identificare i geni più stabili.

Utilizzando il software JMP 5.1.2 (SAS Institute Inc., Cary, USA), sono state effettuate analisi statistiche quali analisi della varianza (ANOVA) e Tukey’s test.

Infine il t-test è stato utilizzato per calcolare le differenze tra i valori dei geni negli stadi embrionali e i valori dei tessuti, considerando  $p < 0.05$  come significativo e  $p < 0.01$  come altamente significativo.

### **2.3.2. RT-PCR per la caratterizzazione di CNEs**

Allo scopo di amplificare CNEs di *Ciona intestinalis*, da microiniettare in embrioni di pesce zebra, sono state messe a punto reazioni di PCR effettuate come segue.

La progettazione dei *primers* è avvenuta utilizzando il *software* Amplify 3X (Engels, 1993); i *primers* sono stati progettati secondo i criteri già descritti e cercando di farli cadere il più vicino possibile agli estremi 5’ e 3’ delle CNEs, in modo da evitare di amplificare zone fiancheggianti che avrebbero potuto falsare il risultato del saggio funzionale *in vivo*.

Le sequenze dei *primers* e la loro localizzazione sono mostrate in Tabella 8 e Tabella 9.

Gli esperimenti di PCR sono stati condotti in volumi finali di 100  $\mu$ L, contenenti 2  $\mu$ L di DNA genomico, 2,5 U di Taq Polimerasi (KAPA Biosystems) con i relativi reagenti (2 mM di dNTPs mix, 16 mM MgCl, 10

$\mu\text{L}$  di buffer di PCR 10 $\times$ ), e 0,5  $\mu\text{M}$  di ciascun primer (senso e antisenso) che serve da innesco alla polimerasi.

Nelle PCR fatte per questo studio è stato usato un programma che prevede un primo intervallo di 2 min. a 96°C per la denaturazione dei filamenti e poi, ciclicamente, 30 sec. a 96°C, sempre per la denaturazione, 30 sec. da 65°C a 50°C ( con decremento di un grado ad ogni ciclo) per permettere l'appaiamento dei *primers* e l'allungamento della catena, 60 sec. a 72°C, temperatura alla quale l'efficienza della *Taq* polimerasi è massima e quindi continua l'allungamento. Seguono 5 min. a 72°C per terminare l'estensione di eventuali filamenti incompleti.

I risultati dell'amplificazione sono stati osservati tramite elettroforesi su gel di agarosio all'1,5 %. Sono stati caricati 5  $\mu\text{L}$  di prodotto di PCR. Come marcatore è stato utilizzato "Marker 5" (Fermentas, Hannover, Maryland, USA). Per ogni CNEs sono state effettuate 12 reazioni di PCR, unite insieme prima di procedere alla purificazione.

La purificazione dei prodotti di PCR è stata effettuata secondo il seguente protocollo: si aggiungono 1/10 del volume di Sodio acetato e il doppio di Etanolo 100% e si mantiene la soluzione a -20°C *overnight*. Dopo aver centrifugato alla massima velocità per 30 secondi e aver eliminato il surnatante, si risospende il precipitato in 200  $\mu\text{L}$  di Etanolo 100%. Dopo aver centrifugato alla massima velocità per 2 minuti e aver eliminato il surnatante, si risospende il precipitato in 100  $\mu\text{L}$  di acqua.

Infine, è stato seguito il protocollo del Kit HeathCare per la purificazione da PCR, eluendo il precipitato finale in 30  $\mu\text{L}$  di acqua.

Il prodotto di purificazione è stato visualizzato e quantificato su gel di agarosio 1,5%, utilizzando come marcatore "Marker 5" (Fermentas, Hannover, Maryland, USA).

#### **2.4. L'utilizzo della *Real-time* PCR nell'analisi dell'espressione genica**

E' stata scelta la *Real-time* PCR quantitativa relativa come metodo per l'analisi del gene *cyrr1*. Tale tecnica è stata effettuata ponendo a confronto i vari campioni oggetto di studio con l'amplificazione di uno o

più geni *housekeeping*, ovvero geni ubiquitariamente espressi nell'organismo di interesse e che vengono quindi presi come riferimento. Sono state dunque progettate coppie di *primers* per *cyvr1* e per i geni di riferimento *housekeeping* (*ef1a*, *slc25a5* e *bactin2*), utilizzando il software Amplify 3X (Engels, 1993). Per le sequenze dei *primers* e dettagli aggiuntivi si veda la Tabella 10.

Per verificare che i *primers* progettati amplificassero una banda singola ed altezza attesa, sono stati effettuati una serie di esperimenti preliminari, attraverso la tecnica di RT-PCR tradizionale con condizioni standard precedentemente descritte (paragrafo 1.4 di Materiali e Metodi). I risultati dell'amplificazione sono stati osservati tramite elettroforesi su gel di agarosio all'1,5 %. Sono stati caricati 5  $\mu$ L di prodotto di PCR. Come marcatore è stato utilizzato "GeneRuler" (Fermentas, Hannover, Maryland, USA).

I geni utilizzati in questa analisi come geni di riferimento *bactin2*, *ef1a* e *slc25a5*, sono stati scelti in quanto considerati geni *housekeeping* "ideali" in base ad una recente analisi sistematica condotta nel nostro laboratorio (Casadei *et al.*, 2011).

Le reazioni di amplificazioni sono state eseguite in volume finale di 20  $\mu$ L utilizzando piastre a 96 pozzetti. Nella miscela di reazione sono stati impiegati i seguenti reagenti: 300 nM per ciascun primer forward e reverse, 10  $\mu$ L di SsoFast EvaGreen® Supermix 1X (Bio-Rad, Hercules, CA, USA) contenente la Sso7d, una proteina che lega il DNA a doppio filamento e che stabilizza la polimerasi aumentandone inoltre la processività, acqua a volume e cDNA. Ogni campione di retrotrascrizione è stato verificato in duplicato (doppio tecnico). Un controllo negativo, contenente tutti i reagenti citati e acqua come stampo, e un campione positivo di reazione, contenente in questo caso reagenti e un cDNA noto, sono stati previsti per ogni coppia di *primers*. Per ogni gene è stata quindi preparata una miscela per un totale di 24 campioni contenente la Supermix, la coppia di *primers*, acqua PCR a volume; successivamente è stata distribuita nei pozzetti della piastra a 96 (18  $\mu$ L) e infine aggiunti 2  $\mu$ L di cDNA alla concentrazione di 2 ng/ $\mu$ L (diluizione 1:20 a partire dalla madre a concentrazione 40 ng/ $\mu$ L).

Tutte le reazioni di *Real-time* PCR sono state effettuate con la macchina CFX-96™ (Bio-Rad, Hercules, CA, USA ) con il seguente ciclo di reazione:

95°C per 30"

95°C per 5"

60°C per 5"

seguiti dalla curva di *melting* per l'identificazione di eventuali prodotti aspecifici o dimeri di primers. Terminata la reazione, i risultati sono stati esaminati utilizzando il programma interno allo strumento "CFX Manager™ v1.1" (Bio-Rad, Hercules, CA, USA).

Un esperimento preliminare è stato condotto per generare una curva standard, utile per determinare l'efficienza di reazione per ogni coppia di primer. E' stata quindi realizzata una piastra di amplificazione con le stesse condizioni previste per gli esperimenti di quantificazione (ed appena descritte), utilizzando come stampo due campioni di cDNA scelti dal nostro pannello sperimentale, diluiti a concentrazioni seriali dell'ordine di 10 (5 diluizioni da 40 ng/ml a 0,004 ng/ml).

## **2.5. Progettazione di costrutti da microiniettare in embrioni di pesce zebra: *CYYR1* e CNEs**

### **2.5.1. mRNA del gene *cyrr1***

Le sonde per gli esperimenti di microiniezione sono state realizzate utilizzando opportune condizioni di RT-PCR, clonaggio in vettore e trascrizione *in vitro*. Gli mRNA specifici ottenuti corrispondono a:

- CDS completa di *cyrr1*, per indagare gli effetti dell'espressione ectopica del gene completo;
- CDS di *cyrr1* priva del codone di stop, per realizzare un costrutto di fusione con la GFP (*green fluorescent protein*) utile per determinare la localizzazione intracellulare della proteina;
- CDS priva del sito putativamente codificante una regione transmembrana ( $\Delta$ TM) oltre al codone di stop, per realizzare un costrutto di fusione con la GFP;

Innanzitutto è stato identificato il vettore per l'inserimento della sonda al fine di realizzare il costrutto.

I vettori scelti sono stati pCS2<sup>+</sup> e pCS2+GFP che sono stati linearizzati tramite reazione standard di digestione con enzima di restrizione appropriato. In particolare il vettore pCS2+GFP è stato digerito con l'enzima di restrizione *BamHI* mentre il vettore pCS2<sup>+</sup> è stato digerito con l'enzima *EcoRI*. Brevemente, in entrambi i casi, 2 µg di plasmide sono stati trattati con 20 U di enzima e relativo tampone (Stratagene, La Jolla, CA); l'incubazione è stata condotta a 30°C *overnight*; di seguito, la miscela è stata posta a 75°C per 10 minuti per porre termine alla reazione. Dopo la digestione il plasmide è stato trattato con la fosfatasi alcalina (CALF, Calf Intestinal Alkaline Phosphatase, TaKaRa), aggiungendo direttamente al campione digerito 1 mL di CALF (circa 20 U/mL) e relativo tampone. L'incubazione è stata effettuata a 37°C per 2 ore, di seguito la miscela è stata posta a 75°C per 15 minuti per porre termine alla reazione.

I cDNA dei messaggeri da inserire nei vettori sono stati amplificati mediante reazioni di PCR con *primers* progettati *ad hoc* per il clonaggio in vettore, in cui vengono aggiunte in 5' le basi del sito di restrizione: *BamHI* (GGATCC) e *EcoRI* (GAATTC). In questo modo è stato possibile ottenere dei prodotti di PCR compatibili con l'enzima di restrizione utilizzato anche per il taglio dei vettori (rispettivamente pCS2+GFP e pCS2<sup>+</sup>).

Gli esperimenti di PCR sono stati condotti in condizioni standard, già descritte.

Per le coppie di *primers* utilizzate nelle reazioni di PCR e dettagli aggiuntivi si veda la Tabella 11. Nello specifico, da reazioni di amplificazione con i *primers* #71 e #72 si sono ottenuti prodotti senza codoni di stop e deleti di un tratto della regione 3'-UTR (3'-*untranslated region*), corrispondenti agli ultimi 27 amminoacidi della proteina della di riferimento #NP\_998047 (costrutto *cyrr1-Δ3'end- GFP*); con le coppie di *primers* #73 e #74, #75 e #76 sono stati amplificati dei prodotti deleti anche del tratto transmembrana (ΔTM). L'amplificazione con i *primers* #77 e #78 ha fornito la CDS completa di *cyrr1*.

Gli ampliconi ottenuti sono stati a loro volta sottoposti a reazione di

digestione con enzima di restrizione appropriato. In particolare 30 ng di prodotto di amplificazione sono stati trattati con 20 U di enzima e relativo tampone (Stratagene, La Jolla, CA). L'incubazione è stata effettuata a 30°C *overnight*, di seguito la miscela è stata posta a 75°C per 10 minuti per porre termine alla reazione.

Successivamente ampliconi e vettori digeriti sono stati purificati rispettivamente con il kit a colonnine "GenElute™ PCR Clean-up Kit" (Sigma, Germany) e con il kit di purificazione "Plasmid miniKit" (Qiagen, Hilden, Germania). I campioni purificati sono stati analizzati su gel di agarosio 1,5%, al fine di quantificare il DNA utilizzando marcatori quali GeneRuler e Marker 5 (Fermentas, Hannover, Maryland, USA).

Alla quantificazione dei prodotti è seguita la reazione di ligasi. Plasmidi ed inserti sono stati trattati con 4 U di T4 ligasi, relativo tampone e rATP 10 mM (Stratagene); per ottenere una resa maggiore è stato mantenuto un rapporto tra plasmide ed inserto di 1:3. L'incubazione è stata effettuata a 23°C per 2 ore.

I prodotti di ligasi ottenuti sono stati trasformazione in batteri E. coli, mediante clonaggio *in vivo* con il kit "One Shot TOP10 competent cells" (Invitrogen) e le colonie cresciute sono state prelevate ed analizzate mediante "colony PCR" (procedura già descritta nel paragrafo 1.7 di MATERIALI E METODI).

Mediante coppie di *primers* opportune sono state verificate la presenza dell'inserto nel vettore (primer #79 e #80 specifici per il vettore, tabella 11) e l'orientamento (un *primer* specifico per il vettore e uno specifico per l'inserto). Successivamente le colonie prescelte sono state espanse a 37° *overnight* e trattate per il recupero del plasmide mediante kit di purificazione "Plasmid miniKit" (Qiagen) secondo il protocollo fornito dalla ditta.

Infine la preparazione dei costrutti da microiniettare in embrioni di pesci zebra ha previsto la trascrizione *in vitro*.

Le reazioni di trascrizione sono state effettuate a partire da prodotti di amplificazione o da DNA plamidico linearizzato, entrambi purificati e quantificati su gel di agarosio, le cui sequenze sono state verificate mediante sequenziamento automatico, usando i medesimi *primers*

utilizzati nelle rispettive reazioni di PCR.

La trascrizione *in vitro* è stata effettuata utilizzando il kit “mMESSAGE mMACHINE SP6” (Ambion, Austin, USA), partendo da 1 µg di plasmide linearizzato o da 0,1-0,2 µg di prodotto di PCR. Questo kit di trascrizione è adatto per la sintesi di mRNA “capped” che, cioè, mimano la maggior parte degli mRNA presenti *in vivo* grazie alla presenza della struttura, all'estremità 5', del cap con la 7- metil guanosina.

Per la reazione di trascrizione *in vitro* è stato utilizzato un tampone di trascrizione 10 X adeguato all'enzima, DNA lineare nella quantità indicata precedentemente, RNase-Inhibitor (35 U/µL) 1 µL, RNA polimerasi 2 µL, mix dNTPs 2 µL, H<sub>2</sub>O RNase free fino ad arrivare ad un volume finale di 20 µL.

La miscela di trascrizione è stata incubata a 37°C per 2 ore, a cui è seguito un trattamento di 15 minuti con *DNase RNase free*. Per purificare la sonda trascritta dai nucleotidi non incorporati è stata effettuata una precipitazione in alcool (etanolo, EtOH) aggiungendo 2,5 µL di litio cloruro 4 M e 75 µL di EtOH assoluto. Dopo aver mescolato, e precipitato il tutto a -20° C per almeno 4 ore, è stata fatta una centrifugazione a 12000 rpm a 4°C, per 20 minuti; poi, al sedimento sono stati aggiunti 500 µL di EtOH 70% freddo e tutto è stato centrifugato a 12000 rpm a 4°C per 5-10 min. Il pellet, dopo evaporazione dell'etanolo residuo, è stato risospeso in 20 µL di H<sub>2</sub>O RNase free + 20 µL di formammide.

Gli mRNA sintetizzati sono stati purificati con kit Sigma e quantificati su gel di agarosio e gel denaturante.

Oltre ai costrutti così ottenuti, oligonucleotidi morfolinici antisense (Gene Tools, LLC, OR, USA) sono stati progettati sulla sequenza complementare alla regione 5'-UTR a monte del codone di inizio AUG di *cyrr1* (sequenza #81, Tabella 11).

Insieme al morfolino specifico per il gene *cyrr1*, sono stati ordinati un morfolino di controllo positivo complementare al trascritto del gene *chordin* del pesce zebra (#82, Tabella 11) e un controllo negativo standard (#83, Tabella 11), utili per verificare la specificità degli effetti dati dal morfolino diretto contro il messaggero di *cyrr1*. Tutti gli oligonucleotidi sono stati marcati con un marcatore fluorescente all'estremità 3'. Gli

oligonucleotidi antisense acquistati sono stati dissolti nella “soluzione Danieau” (58 mM NaCl, 0,7 mM KCl, 0,4 mM MgSO<sub>4</sub>, 0,6 mM Ca(NO<sub>3</sub>)<sub>2</sub>, 5,0 mM HEPES, pH 7,6) e iniettati con una concentrazione nell'ago che varia tra 2 e 6 µg/µL, corrispondenti a 1-3 ng di oligonucleotide per ogni embrione. Il volume della soluzione e la quantità di mRNA microiniettati sono stati determinati dal diametro della goccia iniettata pari a 1/10 di quello dell'embrione.

### 2.5.2. CNEs

Gli elementi non codificanti conservati (CNEs) sono sequenze di DNA ottenute tramite amplificazione a partire da DNA genomico di *Ciona intestinalis*, come descritto in precedenza.

I prodotti di amplificazione, dopo essere stati purificati, sono stati microiniettati in embrioni di pesce zebra in 5 µL di soluzione finale composta da 25 ng/µL di costrutto formato da cDNA di gene *reporter* (Enhanced Green Fluorescent Protein, EGFP) sotto il controllo del promotore del gene umano della Beta-globina, rosso fenolo (0,1%) e 250-300 ng/µL di CNEs.

## 2.6. Procedura di microiniezione in embrioni di *zebrafish*

La microiniezione è una tecnica per introdurre acidi nucleici in cellule od organismi; questa metodica applicata a uova appena fecondate è una procedura relativamente veloce che, con un po' di esperienza, consente di manipolare centinaia di cellule uovo fecondate in un'ora; inoltre è una metodica che permette di avere un alto numero di sopravvivenza degli embrioni e risultati riproducibili.

L'equipaggiamento richiesto per la microiniezione prevede tre componenti fondamentali:

- a) microscopio;
- b) microiniettore collegato ad un compressore;
- c) micromanipolatore con base magnetica, *joystick*, *holder* nel quale posizionare l'ago e braccio trainante.

Oltre a ciò è fondamentale la disponibilità di aghi di vetro finemente tirati con uno strumento opportuno (*puller*) o di aghi commerciali già pronti per eseguire la microiniezione.

Lo strumento per la microiniezione permette di ottenere iniezioni di pressione in modo alternativo alla elettroionoforesi. Nel sistema vengono applicate due pressioni differenti: una alta che permette di espellere sostanze da un sistema di iniezione connesso alla pompa, una più bassa che previene il reflusso nell'ago di tali sostanze. Ai fini della microiniezione, il controllo quantitativo sul volume di fluido espulso dall'ago è ottenuto modulando la pressione o la durata dell'immissione di pressione attraverso una micropompa.

Il micromanipolatore viene fissato ad una pesante base magnetica saldamente posizionata su un tavolo al fine di garantire un sistema immobile e il più possibile libero da vibrazioni. Alla base magnetica è connesso un braccio metallico al quale viene collegato il braccio trainante orizzontale della micropipetta mediante il *joystick*. Quest'ultimo ha il compito di controllare il micromanipolatore grazie alla possibilità di guidare i movimenti submillimetrici antero-posteriori e laterali dell'*holder* nel quale è inserito l'ago.

La soluzione da microiniettare deve essere ben risospesa (per non creare intasamenti meccanici nel sistema), alla concentrazione giusta (in funzione delle necessità sperimentali) e microiniettata ad un volume opportuno (non superiore a quello corrispondente a 1/10 del raggio delle uova).

Si può opportunamente aggiungere rosso fenolo che, oltre ad essere un tracciante che mostra l'entrata della goccia microiniettata nell'embrione, è anche un indicatore di pH.

Le uova da microiniettare vengono prelevate dalla vaschetta di accoppiamento con una pipetta pasteur di plastica prestando attenzione a selezionare solo le uova fertilizzate (che appaiono traslucide) e a non mescolare progenie diverse.

In una piastra Petri si posiziona un vetrino che viene fatto aderire alla base della piastra grazie ad una goccia di *fish water* (0,5 mM  $\text{NaH}_2\text{PO}_4$ , 0,5 mM  $\text{Na}_2\text{HPO}_4$ , 1,5 g *Tropic Marine*, pH 7,2); lungo un lato

del vetrino vengono appoggiate e allineate le uova secondo una singola fila con una pellicola di liquido. Questo serve per bloccare le uova durante la penetrazione dell'ago da microiniezione che solitamente viene inserito nel tuorlo che offre una minore resistenza.

Dal tuorlo, grazie ai ponti citoplasmatici che si mantengono almeno nelle fasi iniziali dello sviluppo embrionale tra blastomero e blastomero e tra tuorlo e blastomeri, il costrutto si diffonde anche nelle cellule. La microiniezione può essere fatta anche a livello del blastomero la cui membrana offre però una maggiore elasticità e resistenza nei confronti dell'ago.

Una volta preparate le cellule uovo fecondate bisogna allestire l'apparato di microiniezione caricando l'ago con la soluzione da microiniettare. Questo viene fatto caricando un piccolo volume di soluzione (tipicamente 3  $\mu\text{L}$ ) in Gilson da 10  $\mu\text{L}$  con puntali *Geloader tips* utili a depositare il piccolo volume nell'ago di vetro, il quale viene subito montato sull' *holder*. Viene poi posizionata sotto al microscopio la piastra Petri con le uova disposte lungo il vetrino e si cerca di mettere a fuoco sia le uova che la punta dell'ago che, essendo chiusa ad una estremità, deve essere delicatamente "sbeccata" per permettere la fuoriuscita della soluzione.

La soluzione viene fatta uscire dall'ago in seguito alla pressione esercitata sul pedale collegato al compressore, ciò determina l'instaurarsi di una pressione di eiezione.

A questo punto, aumentando l'ingrandimento del microscopio, si procede alla microiniezione mediante il controllo, con una mano, della piastra Petri che viene spostata man mano e, con l'altra mano, del micromanipolatore in modo da forare il corion con l'ago e contemporaneamente premere il pedale collegato alla pompa così da iniettare la quantità opportuna di soluzione.

Osservazioni reiterate degli embrioni in via di sviluppo sono state fatte ad intervalli regolari, mantenendo sempre, negli intervalli, gli embrioni in incubatore a 28°C. In particolare sia nei casi della microiniezione di costrutti di *cyyr1* fluorescinati, sia nel caso degli

oligonucleotidi morfolinati, le osservazioni sono state fatte sugli embrioni realmente microiniettati e quindi fluorescenti.

Annotazioni numeriche e fenotipiche precise hanno permesso di seguire i tempi di sviluppo e i relativi dati ottenuti sono descritti puntualmente nei risultati.

Per lo studio sugli elementi conservati non codificanti sono state osservate le procedure di seguito descritte.

Dopo 2-3 ore dalla microiniezione delle CNEs, gli embrioni che si sono sviluppati in modo anormale sono stati eliminati. Gli embrioni normali sono fatti crescere in *medium* contenente 0,003% di feniltiocarbamide.

Durante il secondo giorno dopo la microiniezione (26-33 hpf) gli embrioni sono stati decorionati utilizzando Pronasi E, anestetizzati in Tricaina e analizzati per verificare l'espressione di GFP direttamente sotto l'esposizione di luce UV, utilizzando il microscopio motorizzato invertito IX81 (Olympus, Tokyo, Japan). L'osservazione è stata ripetuta durante il terzo giorno dopo la microiniezione (48 hpf).

Poiché il DNA microiniettato non si integra nel genoma di *zebrafish*, si ottiene una espressione a mosaico; per ottenere, quindi, un profilo di espressione complessivo, è stato effettuato lo screening di un minimo di 25 embrioni per ogni elemento microiniettato.

Il profilo di espressione della GFP così ottenuto è stato schematizzato individuando differenti siti anatomici di espressione (prosencefalo, mesencefalo, rombencefalo, midollo spinale, notocorda, occhio, orecchio, tessuto muscolare, sangue, cuore/zona pericardica, epidermide, pinne). La localizzazione e il sito anatomico di espressione di GFP per ogni embrione è stato schematizzato utilizzando Adobe Photoshop software (Adobe Systems, San Jose, CA, USA)

Come controllo, circa 200 embrioni sono stati microiniettati il costruito contenete il gene *reporter* GFP; una sequenza conservata codificante e una non conservata non codificante, co-microiniettate col gene reporter: nessuno di questi casi ha mostrato espressione di GFP.

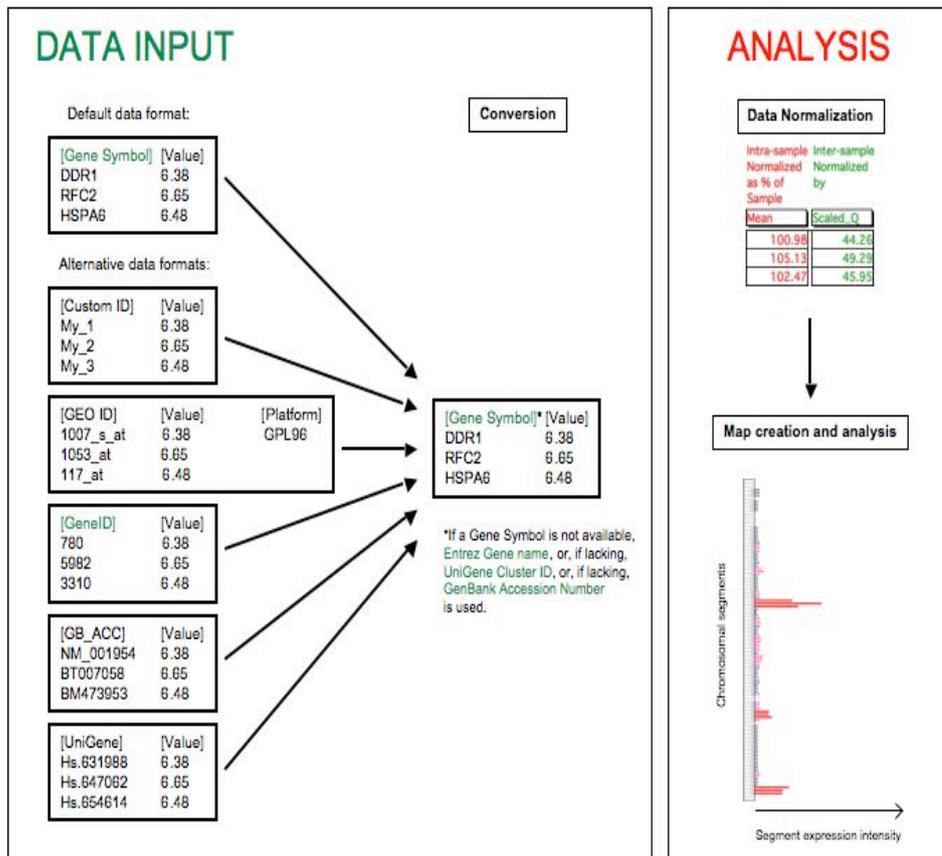
### 3. SVILUPPO DI UN *SOFTWARE* IN GRADO DI PRODURRE MAPPE CROMOSOMICHE DI ESPRESSIONE (TRAM)

TRAM (Transcriptome Mapper) è un *software* in grado di produrre mappe cromosomiche di espressione a partire da dati di analisi genica globale ed è stato sviluppato utilizzando FileMakerPro. Una delle caratteristiche principali e interessanti di questo *software* riguarda la possibilità di utilizzare ed analizzare dati appartenenti a tabelle di database messe in relazione tra loro; di conseguenza, ogni insieme di dati in ogni tabella (per esempio, valori di espressione genica, nomi dei geni, valori di espressione di un segmento cromosomico o di un *cluster* genico) è strutturato come una serie di *records* che può essere interrogata facilmente a partire dal criterio di interesse.

Al fine di collegare il codice identificativo di un gene (*gene identifier*) con il corrispondente nome o simbolo genico, è possibile importare in TRAM file di testo contenenti una serie di dati quali il codice identificativo di ogni sonda utilizzata nell'esperimento di partenza (per esempio in un *microarray*) e il corrispettivo simbolo genico (o numero di accesso alla sequenza GeneBank) (Figura 14).

Ad esempio, è possibile caricare una o più piattaforme GEO (Barrett *et al.*, 2009) ed effettuare analisi sui dati contenuti nella piattaforma. E' stata effettuata la messa a punto dei dati e delle piattaforme GEO più utilizzate per uomo, topo e pesce zebra.

La versione TRAM 1.0 è stata messa a punto utilizzando i dati disponibili al Gennaio 2011: sono stati scaricati i dati relativi alla localizzazione genica da Entrez Gene e sono stati correlati alle tabelle UniGene al fine di ottenere la conversione di ogni numero di accesso GenBank delle sequenze di RNA o di EST (*expression sequence tags*) nel corrispettivo simbolo genico. Inoltre, la localizzazione di cluster di EST (sequenze non ancora caratterizzate come geni ufficiali, ma comunque presenti nel trascrittoma) è stata ottenuta a partire dai dati presenti in UCSC Genome Browser.



**Figura 14.** Struttura generale del software TRAM.

L'utente è guidato nell'importazione e analisi dei profili di espressione genica. Il numero identificativo di ogni gene (*gene identifier*) è convertito nel simbolo genico ufficiale; viene poi eseguita la normalizzazione dei valori *intra* e *inter-sample*. I dati di espressione sono visualizzati graficamente lungo i cromosomi. La sovra o sottoespressione genica è determinata in seguito ad analisi statistica.

Ogni serie di dati (TRAM "Sample") è definita come "campione biologico distinto": ad esempio, un campione potrebbe corrispondere a un singolo canale nel caso di esperimenti di *microarray* condotti su due canali.

Ogni file contenente dati di espressione deve essere un file di testo contenente il codice identificativo del gene (*gene identifier*) e un valore numerico di espressione. Una funzione interna al programma assiste l'utente nella creazione dei file di testo.

Ogni campione o insieme di campioni può essere importato in uno dei due "pool" "a" o "B", in modo da poter distinguere e confrontare due

condizioni biologiche differenti. TRAM è in grado di effettuare una serie di normalizzazioni dei dati in modo da ottenere un confronto dei valori di espressione ottenuti da campioni biologici differenti o piattaforme sperimentali diverse: la normalizzazione *intra-sample* (per esempio, *intra-array*) avviene all'interno di ogni singolo campione e può portare a visualizzare i valori di espressione come media o mediana; la normalizzazione *inter-sample* consiste in una rielaborazione del metodo dei quantili. In particolare, ad ogni valore normalizzato è assegnata una posizione in ordine ascendente; successivamente viene calcolata la media dei valori che si trovano nella stessa posizione e il valore così calcolato è assegnato ad ogni gene che si trova nella stessa posizione in ogni campione. Questo metodo permette di effettuare confronti *inter-sample*.

Vi sono due possibili modalità di analisi: “Map” mode e “Cluster” mode. Con la prima modalità, è possibile generare una mappa grafica del trascrittoma che mostra linee verticali che rappresentano i cromosomi (Figura 15).

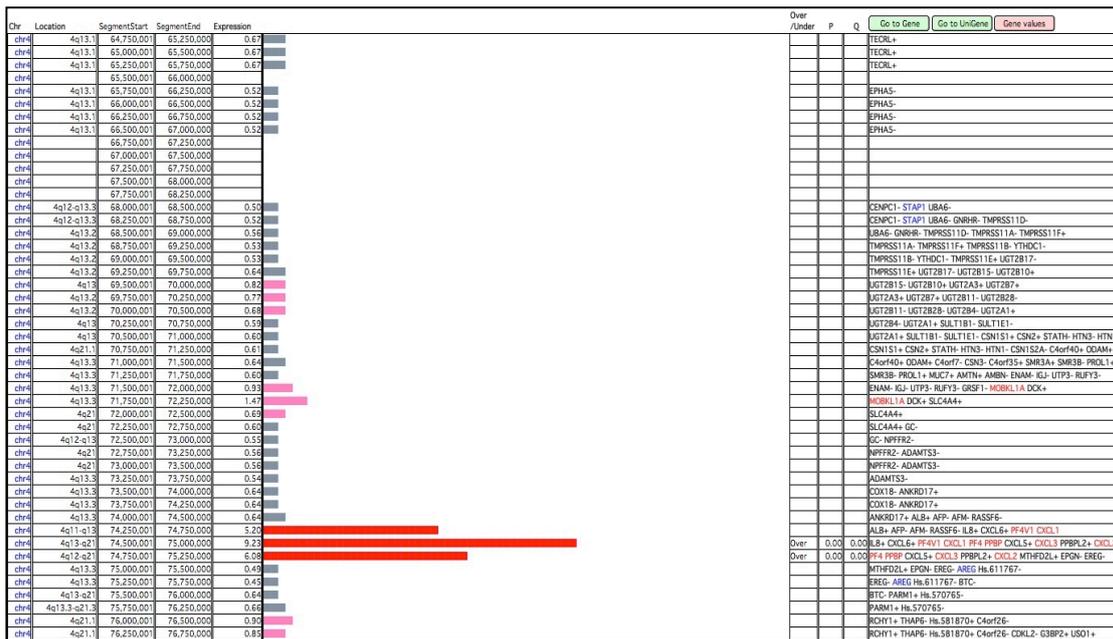


Figura 15. Visualizzazione grafica della modalità di analisi “Map”.

La lunghezza delle barre orizzontali è proporzionale alla media dell’espressione genica all’interno di un segmento di 0.5 Mb. La linea verticale rappresenta il cromosoma 4 umano. Sono indicati i valori di espressione di ogni segmento e la presenza di sovra o sottoespressione. I valori statistici *p* e *q* sono calcolati per ogni regione analizzata.

Un valore di espressione è associato a ogni segmento all'interno della linea, segmento la cui dimensione può essere impostata dall'utente. Per ogni segmento di cromosoma verrà visualizzata una barra orizzontale la cui lunghezza sarà proporzionale al valore di espressione assegnato. Questo valore corrisponderà alla media di tutti i dati di espressione disponibili associati ai geni presenti nel segmento analizzato.

Attraverso la modalità "Cluster", invece, TRAM effettua una analisi andando a ricercare un insieme di almeno due geni contigui, espressi entro una certa soglia fissata dall'utente (Figura 16). In questo modo, i risultati mostrano le differenze di espressione tra loci senza il limite fissato dalla lunghezza dei segmenti cromosomici analizzati.



**Figura 16.** Visualizzazione grafica della modalità di analisi "Cluster".

La lunghezza delle barre orizzontali è proporzionale alla media dell'espressione genica calcolata come rapporto tra due campioni o insiemi di campioni (A/B). Le barre rosse indicano la sovraespressione genica ed è visualizzato il valore medio di espressione genica di ogni cluster. I valori statistici  $p$  e  $q$  sono calcolati per ogni regione analizzata.

Infine, per ottenere risultati statisticamente significativi, TRAM utilizza la distribuzione ipergeometrica per calcolare la probabilità  $p$  che la colocalizzazione o la sovra/sottoespressione dei geni analizzati sia dovuta al caso. Il valore  $p$  deve poi essere corretto per tener conto della *False*

*Discovery Rate (FDR)*, in quanto si tratta di analisi di un numero molto elevato di dati. Il valore di  $p$  così corretto è espresso come  $q$ . I risultati sono ritenuti statisticamente significativi se  $q < 0.05$ .

L'utente può decidere se effettuare i calcoli statistici tenendo conto di tutti i geni nel genoma oppure solo dei geni che si trovano all'interno del cromosoma a cui appartiene il segmento o il cluster analizzato.

## **RISULTATI**

## 1. CARATTERIZZAZIONE DEL *LOCUS* GENICO UMANO *CYYR1*

### 1.1. Analisi bioinformatica

Il database di EST umane (Boguski *et al.*, 1993) è stato interrogato mediante BLASTN 2 (Altschul *et al.*, 1997) utilizzando la sequenza #NM\_052954 relativa alla forma di riferimento *CYYR1-1,2,3,4* ed è stato possibile individuare numerosi *splicing* alternativi.

Le isoforme caratterizzate in questo studio si riferiscono alle seguenti EST: #CN\_402794, #CN\_402791, #DB\_268533, #DB\_120436, #AK\_304124.

In particolare l'analisi della EST #CN\_402794, dimostrava la presenza di un esone 4 alternativo (*CYYR1-1,2,3,4b*); la EST #CN\_402791, dimostrava la presenza di un esone 3 alternativo e la completa assenza dell'esone 4 (*CYYR1-1,2,3b*); la EST #DB\_268533, dimostrava la presenza di un esone 1 alternativo (*CYYR1-1b,2,3,4*); la EST #DB\_120436 dimostrava la mancanza dell'esone 3 (*CYYR1-1,2,4*) ed infine la EST #AK\_304124 dimostrava la presenza di un esone aggiuntivo tra l'esone 2 e l'esone 3 (*CYYR1-1,2,2bis,3,4*).

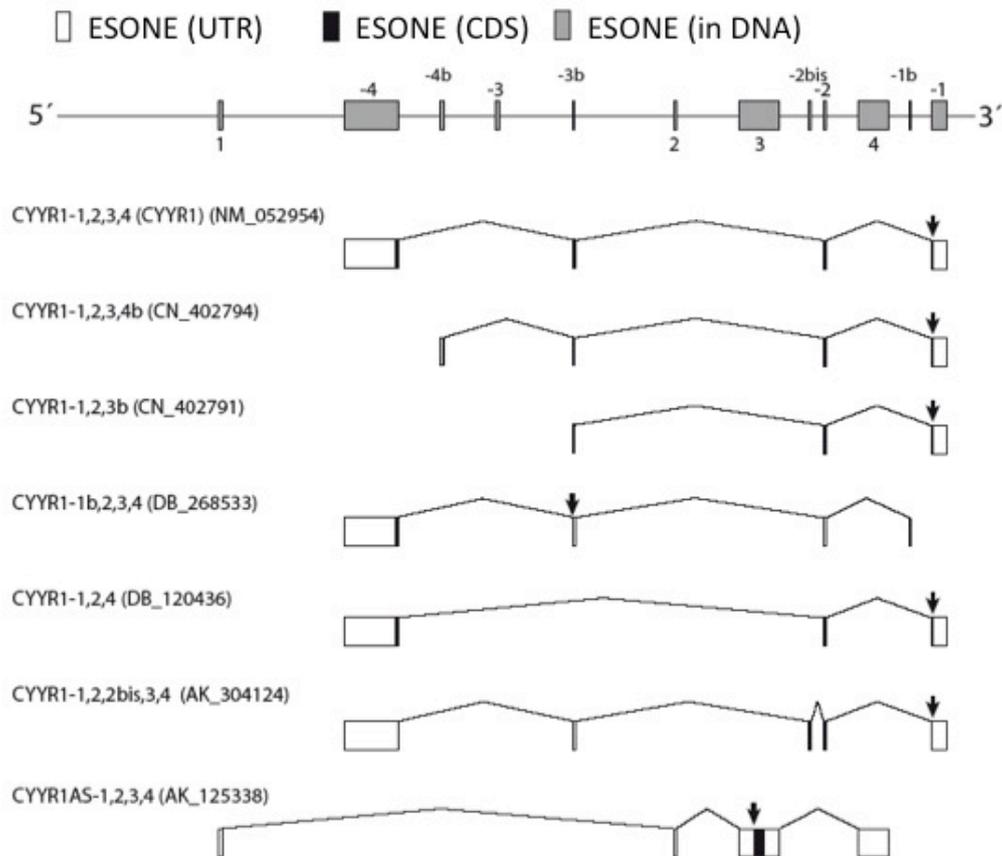
Inoltre, sono riportate anche EST che si riferiscono a trascritti in antisenso rispetto al gene. In particolare, è stato ritrovato un trascritto #AK\_125338 (schedato in Genbank come *Homo sapiens cDNA FLJ43348 fis, clone NT2RI3008974, moderately similar to Probable transposase*) e da noi denominato *CYYR1-ASI* in accordo con HUGO (*Human Genome Organization*). La sequenza corrisponde infatti ad un RNA molto lungo (3.412 bp) organizzato in 4 esoni sul filamento opposto rispetto a *CYYR1* ed in parte imbricato con esso. Per questo è possibile definire il *locus* genico un *locus* multitrascritto.

Per dettagli sul *locus* si veda Figura 17.

### 1.2. Materiale biologico ed estrazione di RNA

In seguito ad estrazione di RNA dalle diverse fonti di materiale biologico descritte, sono state valutate le rese nonché la purezza dei campioni. I valori del rapporto dell'assorbanza alla lunghezza d'onda di

260 e 280 nm, misurata al Nanodrop, erano tutti approssimati al valore di ~2.0, generalmente considerato corrispondente a campioni di RNA “puro” (privo di contaminazioni da DNA). La conferma della integrità dell’RNA è stata ottenuta in seguito a corsa elettroforetica in un gel d’agarosio, che ha permesso anche di verificare ulteriormente l’assenza di tracce di DNA genomico.



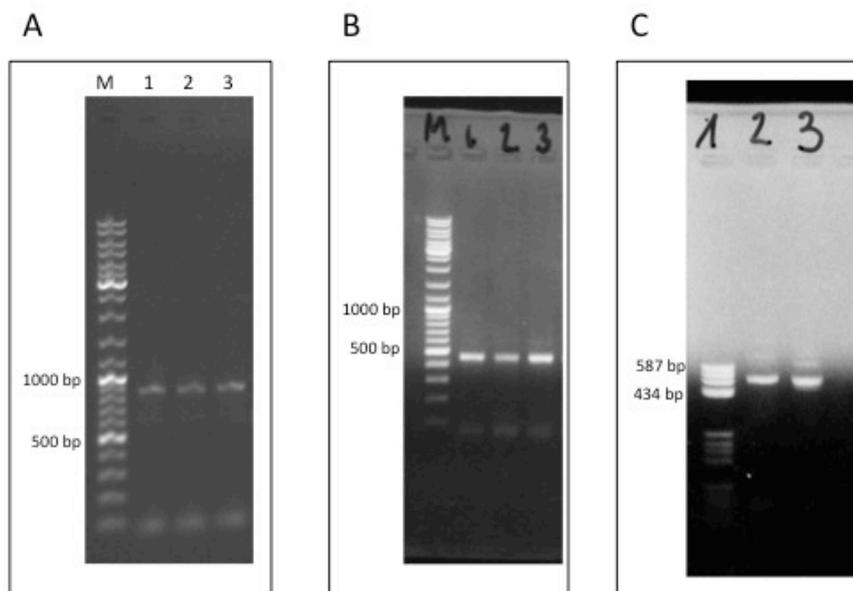
**Figura 17.** Rappresentazione schematica con organizzazione esoni-introni dei vari trascritti del *locus CYR1*.

Nello schema del DNA gli introni non sono rappresentati in scala. Gli esoni 1, 2, 3 e 4 sono da intendere sul filamento senso.

I diagrammi rappresentano gli mRNA delle diverse isoforme del *locus*. Nei trascritti la direzione della traduzione è da destra a sinistra e la freccia indica l’inizio della CDS, nel trascritto *CYR1-ASI* la traduzione è da sinistra a destra.

### 1.3. Clonaggio genico dei trascritti del locus umano *CYYR1*

Esperimenti di RT-PCR sono stati condotti, secondo la metodica descritta (paragrafo 1.4. MATERIALI E METODI), a partire da cDNA di tessuti umani (cuore, cervello e tessuto neuroendocrino di fegato) utilizzando le coppie di *primers* #1 e #4 (EST #CN\_402794), #1 e #6 (EST #CN\_402791) e #8 e #2 (EST #DB\_268533) indicate in Tabella 6. Si sono così ottenuti prodotti di amplificazione di altezza compatibile con quella attesa (Figura 18), la cui analisi di sequenza ha consentito l'identificazione di due nuove isoforme di *CYYR1*: *CYYR1-1,2,3,4b* con una porzione alternativa a partire dalla posizione 656 della sequenza di riferimento #NM\_052954 (fine dell'esone 3); *CYYR1-1,2,3b* con una porzione alternativa a partire dalla posizione 498 della sequenza di riferimento #NM\_052954 (fine dell'esone 2).



**Figura 18.** A. Elettroforesi su gel di agarosio esemplificativo per i prodotti ottenuti in RT-PCR (5  $\mu$ L) con coppie di *primer* #1 e #4 (Tabella 6) per la caratterizzazione della isoforma *CYYR1-1,2,3,4b*.

Corsia 1-2-3: rispettivamente cuore, cervello e tumore neuroendocrino di fegato umani. Corsia M: 1  $\mu$ L (500 ng) di marcatore GeneRuler DNA Ladder Mix (spettro: 100-10000 paia di basi).

**B.** Elettroforesi su gel di agarosio esemplificativo per i prodotti ottenuti in RT-PCR (5  $\mu$ L) con coppie di *primer* #1 e #6 (Tabella 6) per la caratterizzazione della isoforma *CYYRI-1,2,3b*.

Corsia 1-2-3: rispettivamente cuore, cervello e tumore neuroendocrino di fegato umani. Corsia M: 1  $\mu$ L (500 ng) di marcatore GeneRuler DNA Ladder Mix (spettro: 100-10000 paia di basi).

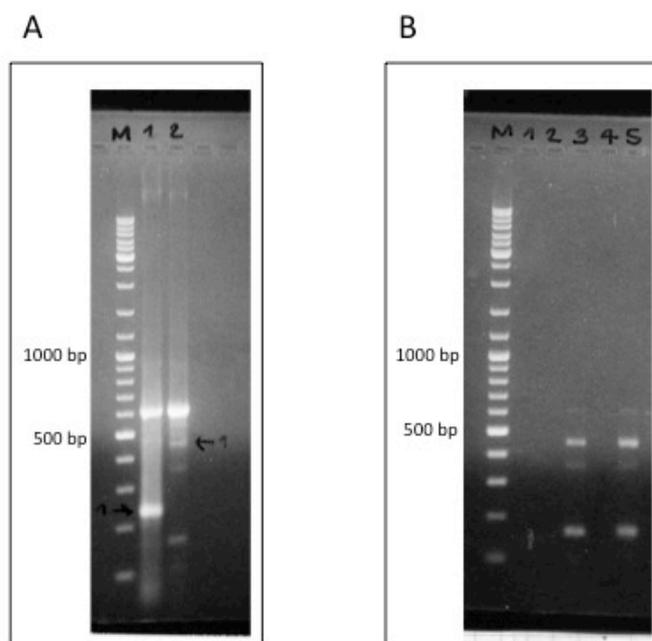
**C.** Elettroforesi su gel di agarosio esemplificativo per i prodotti ottenuti in RT-PCR (5  $\mu$ L) con coppie di *primer* #8 e #2 (Tabella 6) per la caratterizzazione della isoforma *CYYRI-1b,2,3,4*.

Corsia 2: cervello umano. Corsia 1: 1  $\mu$ L (250 ng) di marcatore Marker 5 (spettro: 80-587 paia di basi).

Le nuove isoforme sono state confermate in seguito a sequenziamento di prodotti ottenuti da almeno 2 clonaggi indipendenti e schedate nella banca dati di sequenze nucleotidiche GenBank.

L'analisi su gel di agarosio effettuata dopo RT-PCR condotta con le coppie di *primers* #1 e #2 (Tabella 6) mostrava la presenza di diversi amplificati insieme al prodotto ad altezza attesa, alcuni dei quali compatibili con le dimensioni delle isoforme relative alle EST #DB\_120436 (Figura 19A) e #AK304124 (Figura 20A). Per ottenere un clonaggio specifico per queste isoforme si è deciso quindi di procedere con altre tecniche di clonaggio.

In particolare, per isolare la banda attesa corrispondente alla isoforma mancante dell'esone 3 (#DB\_268533) è stato eseguito il carotaggio di una banda ottenuta in seguito a PCR a partire da cDNA di cervello umano (vedi paragrafo 1.5, Materiali e Metodi) e la riamplicazione di una aliquota del prodotto di eluizione della banda in *buffer* di PCR 1 $\times$ . L'analisi di sequenza dell'amplicone ottenuto ha mostrato la mancanza di 158 basi corrispondenti all'intervallo 499-656 della sequenza di riferimento #NM\_052954 (quindi mancanza completa dell'esone 3, *CYYRI-1,2,4*) e la presenza di una forma di "splicing sottile" CAG+, precedentemente descritta (Vitale *et al.*, 2007).



**Figura 19. A.** Elettroforesi su gel di agarosio dei prodotti ottenuti in RT-PCR con coppie di primer #1 e #2 (Tabella 6) per l'amplificazione della isoforma *CYYRI-1,2,4*.

Corsia 2: cervello umano; la banda 1 indicata dalla fraccia è stata caratterizzata mediante carotaggio (paragrafo 1.5 di Materiali e Metodi).

Corsia M: 1  $\mu$ L (500 ng) di marcatore GeneRuler DNA Ladder Mix.

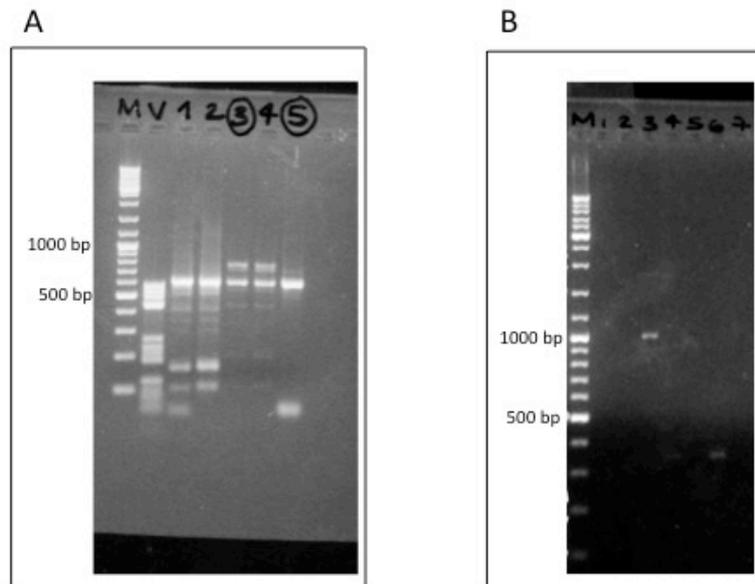
**B.** Elettroforesi su gel di agarosio dei prodotti di riamplicazione (5  $\mu$ L) da banda carotata, con primer #1 e #5 (Tabella 6).

Corsie 3 e 5: cervello umano. Corsia "M": 1  $\mu$ L (500 ng) di marcatore GeneRuler DNA Ladder Mix (spettro: 100-10000 paia di basi).

Per quanto riguarda, invece, l'isoforma con esone 2 aggiuntivo (#AK304124), è stato necessario effettuare un clonaggio *in vivo* come descritto nel paragrafo 1.7 di MATERIALI E METODI. L'analisi di sequenza ha confermato il clonaggio di una sequenza con 167 basi aggiuntive tra gli esoni 2 e 3 della sequenza di riferimento #NM\_052954, denominata *CYYRI-1,2,2bis,3,4*.

Le sequenze ottenute provengono da esperimenti di clonaggio effettuati a partire da cDNA di cuore umano.

Per le ultime due nuove isoforme servirà la conferma di almeno un ulteriore clonaggio indipendente prima della registrazione nella banca dati di sequenze nucleotidiche GenBank, in quanto contengono piccoli *mismatch* rispetto alla sequenza del messaggero schedato in banca dati (EST) e del DNA genomico.



**Figura 20. A.** Elettroforesi su gel di agarosio dei prodotti ottenuti in RT-PCR (5  $\mu$ L) con coppia di primer #1 e #2 (Tabella 6), per l'amplificazione della isoforma *CYYRI-1,2,2bis,3,4*.

Corsie 3 e 5: prodotti utilizzati in reazione di ligasi con vettore pCR2.1 (paragrafo 1.7 di MATERIALI E METODI). Corsia M: 1  $\mu$ L (500 ng) di marcatore GeneRuler DNA Ladder Mix (spettro: 100-10000 paia di basi).

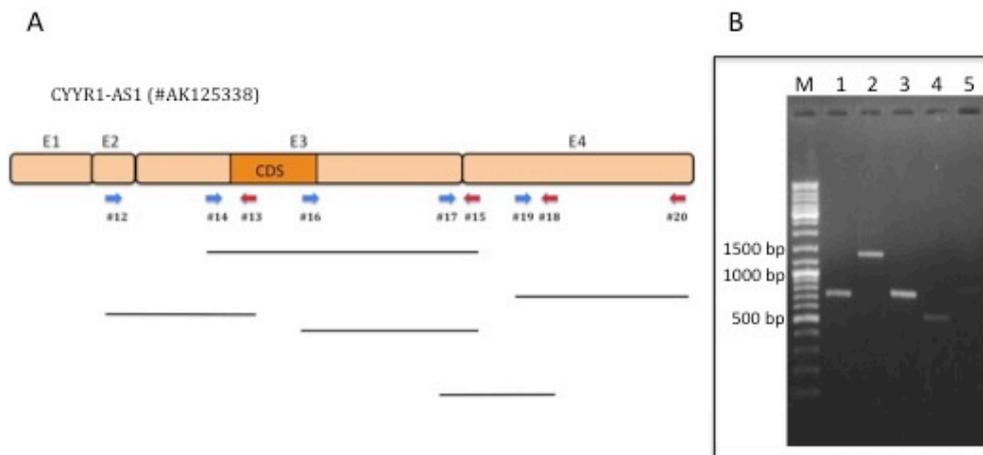
**B.** Elettroforesi su gel di agarosio dei prodotti ottenuti in *colony* PCR con coppia di primer M13f e M13r (Kit TA Cloning, Invitrogen) per la ricerca di isoforme di *CYYRI*.

Corsia 3: banda ad altezza attesa compatibile con l'amplificazione della isoforma *CYYRI-1,2,2bis,3,4* più vettore. Corsia M: 1  $\mu$ L (500 ng) di marcatore GeneRuler DNA Ladder Mix (spettro: 100-10000 paia di basi).

Esperimenti di RT-PCR sono stati infine effettuati con le diverse coppie di *primers* relative al trascritto #AK\_125338 (Tabella 6). Le successive analisi di sequenza, di prodotti di amplificazione in parte

sovrapposti, hanno permesso di ottenere il clonaggio completo di questo lungo messaggero (Figura 21) che si presenta sul filamento opposto rispetto a *CYYR1* ed in parte embricato con esso (Figura 17).

La sequenza è stata registrata nella banca dati di sequenze nucleotidiche GenBank.



**Figura 21. A.** Schema del mRNA di *CYR1-AS1*. Le frecce blu e rosse rappresentano i *primers* senso ed antisenso utilizzati per il clonaggio dell'isoforma in RT-PCR (Tabella 6). Le barre sovrapposte rappresentano gli amplificati ottenuti (per le altezze si veda Tabella 6).

**B.** Elettroforesi su gel di agarosio dei prodotti ottenuti in RT-PCR (5  $\mu$ L) con le diverse coppia di *primer* per il clonaggio di *CYR1-AS1* (Tabella 6). Corsia 1: testicolo umano, *primers* #12 e #13; corsia 2: cervello umano, *primers* #14 e #15; corsia: 3: cervello umano, *primers* #16 e #15; corsia 4: cuore umano, *primers* #17 e #18; corsia 5: cuore umano, *primers* #19 e #20. Corsia M: 1  $\mu$ L (500 ng) di marcatore GeneRuler DNA Ladder Mix (spettro: 100-10000 paia di basi).

#### 1.4. Analisi bioinformatica dei prodotti proteici predetti e di espressione dei trascritti del locus umano *CYR1*

Analisi bioinformatiche del prodotto proteico sono state condotte per tutte le isoforme clonate. Programmi quali BlastP (parametri di *default*, no filtro) e SMART sono stati usati per la ricerca di eventuali domini proteici noti (Figura 22B). Inoltre è stato utilizzato il programma ClustalW

1.83 (<http://www.ebi.ac.uk/clustalw>) per ottenere l'allineamento multiplo tra *CYYR1* e le varie isoforme proteiche (Figura 22A).

### A

```

CYYR1                      -----MDAPRLPVRPGVLLPKLVLLFVYADDCLAQCGK 33
CYYR1-1_2_3_4b           -----PFANMDAPRLPVRPGVLLPKLVLLFVYADDCLAQCGK 37
CYYR1-1_2_3b             -----FPSPPFPVPGSPFAPSLRPPANMDAPRLPVRPGVLLPKLVLLFVYADDCLAQCGK 56
CYYR1-1b_2_3_4          -----QDDCLAQCGK 10
CYYR1-1_2_4             PSDRPPSPFPVPGSPFAPSLRPPANMDAPRLPVRPGVLLPKLVLLFVYADDCLAQCGK 60
CYYR1-1_2_2bis_3_4      -----PGSPFAPSLRPPANMDPRLPVRPGVLLPKLVLLFVYADDCLAQCGK 47
                          *****

CYYR1                      DCKSYCCDGTTPYCCSYAYIGNILSGTAIAGIVFGIVFIMGVIAGIAICICMCKNHRA 93
CYYR1-1_2_3_4b           DCKSYCCDGTTPYCCSYAYIGNILSGTAIAGIVFGIVFIMGVIAGIAICICMCKNHRA 97
CYYR1-1_2_3b             DCKSYCCDGTTPYCCSYAYIGNILSR-----EPKDHQT 90
CYYR1-1b_2_3_4          DCKSYCCDGTTPYCCSYAYIGNILSGTAIAGIVFGIVFIMGVIAGIAICICMCKNHRA 70
CYYR1-1_2_4             DCKSYCCDGTTPYCCSYAYIGNILSRITTL-----RSRPRDGLCLASSILPHT 112
CYYR1-1_2_2bis_3_4      DCKSYCCDGTTPYCCSYAYIGNILSC-----CCPR-- 78
                          *****

CYYR1                      TRVGILRTTHINTVSSYPGPPPYGHDEMEYCADLPPYSPTPQGPAQRSPPPYPGNAR 153
CYYR1-1_2_3_4b           TRVGILRTTHINTVSSYP---EY-ELHEAEILVYFVQHSIPSM----- 136
CYYR1-1_2_3b             -----FEEN-----INYE----- 98
CYYR1-1b_2_3_4          TRVGILRTTHINTVSSYPGPPPYGHDEMEYCADLPPYSPTPQGPAQRS----- 120
CYYR1-1_2_4             GSSTAFSTPSL---SWKRKEIILYLPQNM----- 139
CYYR1-1_2_2bis_3_4      -----

CYYR1                      K 154
CYYR1-1_2_3_4b           -
CYYR1-1_2_3b             -
CYYR1-1b_2_3_4          -
CYYR1-1_2_4             -
CYYR1-1_2_2bis_3_4      -

```

### B



**Figura 22.** **A.** Allineamento multiplo tra *CYYR1* e le varie isoforme proteiche effettuato tramite il programma ClustalW. **B.** Rappresentazione schematica delle predizioni proteiche delle isoforme di splicing del locus. La regione tratteggiata si riferisce alla sequenza più estesa in 5' (NEDO human sequencing project); in rosso è indicato il dominio conservato ricco in tirosina e cisteina (CCSYAY); in verde è indicato il dominio transmembrana; in blu, sono indicati i motivi PPPY.

La sequenza della isoforma denominata *CYYR1-1,2,3,4b* è caratterizzata da un esone 4 alternativo le cui basi si mantengono *in frame* con le basi dell'esone 3. Dalla analisi di allineamento con il DNA genomico si evidenzia la riduzione dell'introne 3 da 11.639 bp a 2.135 bp. Le basi alternative contengono un codone di STOP *in frame* e portano alla formazione di una proteina ipotetica con una porzione C terminale diversa rispetto alla proteina di riferimento (CYYR1 #NP\_443186), con conseguente perdita dei tre motivi PPPY conservati.

La sequenza della isoforma denominata *CYYR1-1,2,3b* è caratterizzata da un esone 3 alternativo contenente un codone di STOP *in frame*. La proteina ipotetica risultante sarà molto più corta rispetto alla proteina di riferimento (CYYR1 #NP\_443186), con perdita degli aminoacidi che identificano il dominio caratteristico transmembrana e dei motivi PPPY conservati. Dalla analisi di allineamento con il DNA genomico si evidenzia una leggera riduzione del lungo introne 2 da 85.836 bp a 82.857 bp.

La sequenza della isoforma denominata *CYYR1-1b,2,3,4* è caratterizzata da un esone 1 alternativo che, dagli esperimenti di clonaggio finora effettuati e dalla analisi di allineamento con il DNA genomico, si presenta a valle dell'esone 1 utilizzato dalla forma di riferimento e molto più piccolo. La proteina ipotetica risultante avrà una porzione N terminale diversa rispetto alla proteina di riferimento (CYYR1 #NP\_443186), con spostamento del codone di inizio della CDS.

La sequenza della isoforma denominata *CYYR1-1,2,4* manca dell'intero esone 3 rispetto alla sequenza di riferimento NM\_052954. La ricerca nelle banche dati di DNA genomico evidenzia un introne formato dalla somma dell'introne 2 (85.836 bp), dell'esone 3 (158 bp) e dell'introne 3 (11.640 bp) relativamente alla forma di riferimento. Viene rimosso un introne di lunghezza complessiva di 97.631 bp in quanto si osserva l'incorporazione di tre basi aggiuntive (CAG) al confine tra gli esoni 2 e 4 che dovrebbero risultare contigui secondo la forma di riferimento. A livello della traduzione si ha dunque una perdita del normale schema di lettura determinando nella proteina predetta il

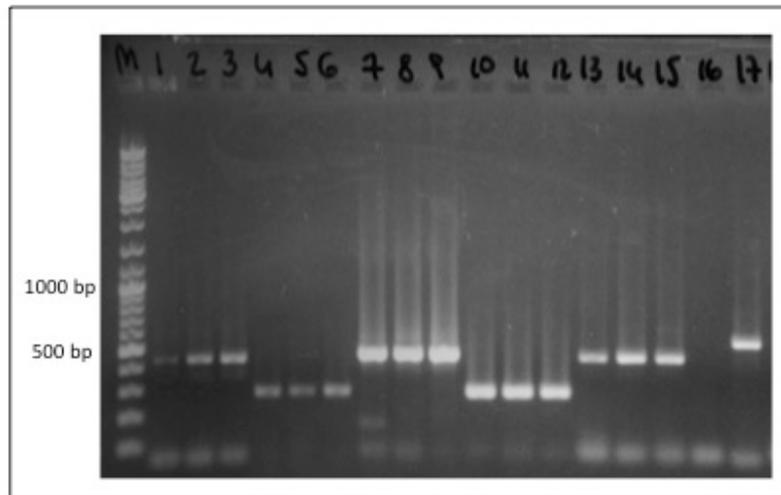
cambiamento della catena polipeptidica a partire dall'aminoacido in posizione 61.

Infine la sequenza della isoforma denominata *CYYR1-1,2,2bis,3,4* presenta 167 basi aggiuntive tra gli esoni 2 e 3 della sequenza di riferimento #NM\_052954. La ricerca nelle banche dati di DNA genomico evidenzia la formazione di un nuovo esone e di due nuovi introni rispettivamente di 1.439 bp (tra esone 2 e esone 2bis) e 84.228 (tra esone 2bis e esone 3). L'esone 2bis introduce uno STOP *in frame* precoce e la proteina predetta risulta dunque più corta rispetto alla proteina di riferimento (CYYR1 #NP\_443186), con perdita degli aminoacidi che identificano il dominio caratteristico transmembrana e dei motivi PPPY conservati.

L'analisi delle sequenze proteiche delle isoforme di splicing indagate non ha mostrato alcuna somiglianza con domini proteici noti ad eccezione del dominio centrale ricco in tirosina e cisteina (CCSYAY) fortemente conservato dai vertebrati inferiori all'uomo che contraddistingue questa nuova famiglia di prodotti proteici (DUF2668 nella banca di domini conservati CDD, <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>).

Analisi di espressione dei trascritti del *locus CYYR1* sono tuttora in corso con metodi di RT-PCR quantitativa relativa e RT-PCR *Real-time* in diversi pannelli di RNA di tessuti umani, in linee tumorali continue e in linee primarie normali vs trisomiche.

Indagini preliminari condotte in diversi tessuti umani normali (quali cervello, cuore, testicolo, sangue, intestino) e tessuti tumorali (tumori neuroendocrini di fegato, pancreas, ghiandola mammaria) non hanno mostrato particolari tessuto-specificità nè per le isoforme di splicing, nè per il trascritto in antisenso (Figura 23).



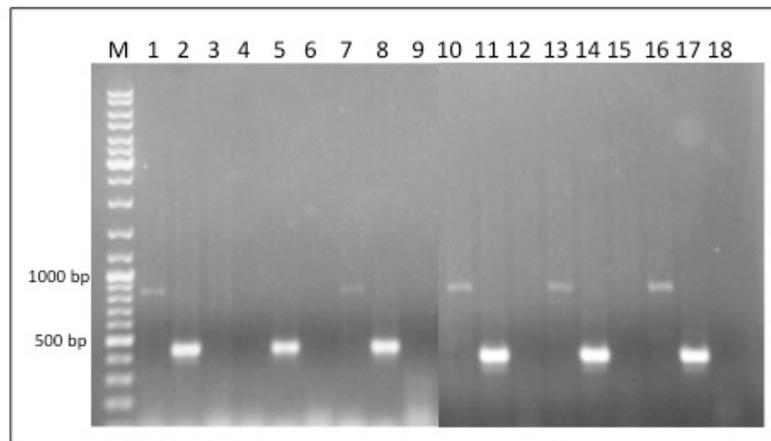
**Figura 23.** Elettroforesi su gel di agarosio dei prodotti ottenuti in RT-PCR (5  $\mu$ l) con coppie di primer per le diverse isoforme di splicing di *CYR1* (Tabella 6).

Corsie 1-2-3: *CYR1-1,2,3,4b* (primers #1 e #5) rispettivamente in tessuto di cervello, testicolo e neuroendocrino di fegato umani. Corsie 4-5-6: *CYR1-1,2,3b* (primers #1 e #7) rispettivamente in tessuto di cervello, testicolo e neuroendocrino di fegato umani. Corsie 7-8-9: *CYR1-1b,2,3,4* (primers #9 e #2) rispettivamente in tessuto di cervello, testicolo e neuroendocrino di fegato umani. Corsie 10-11-12: *CYR1-1,2,2bis,3,4* (primers #1 e #11) rispettivamente in tessuto di cervello, testicolo e neuroendocrino di fegato umani. Corsia 16: controllo negativo. Corsia 17: controllo positivo. Corsia M: 1 $\mu$ l (500 ng) di marcatore GeneRuler DNA Ladder Mix.

Per quanto riguarda le linee cellulari tumorali, è stata indagata la presenza delle isoforme del *locus* in numerose linee continue: SJNKP e SKNMC di neuroblastoma; HT29 di adenocarcinoma del colon; MCF7 di adenocarcinoma della mammella; SK-MEL28 e G361 di melanoma ed infine linee tumorali di osteosarcoma (SaOS, U2OS e U2OS-1). In particolare è da notare che la forma di riferimento *CYR1* non è risultata espressa nelle linee continue sopra dette ad esclusione delle linee tumorali di osteosarcoma (dato non mostrato). Per quanto riguarda le isoforme di splicing indagate, interessante è il dato che mostra l'espressione costante di

*CYYRI-1,2,3,4b* e *CYYRI-1,2,3b* e la assenza costante dell'isoforma *CYYRI-1b,2,3,4* (Figure 24 e 25).

In Figura 25 si nota anche la presenza in linee di osteosarcoma della isoforma di riferimento *CYYRI* e di *CYYRI-AS1*.



**Figura 24.** Elettroforesi su gel di agarosio dei prodotti ottenuti in RT-PCR (5 $\mu$ l) con coppie di primer per le diverse isoforme del *locus* di *CYYRI* (Tabella 6) in cellule tumorali continue.

Corsie 1-2-3: rispettivamente *CYYRI-1,2,3,4b* (primers #1 e #4), *CYYRI-1,2,3b* (primers #1 e #6), *CYYRI-1b,2,3,4* (primers #8 e #2) in HT29.

Corsie 4-5-6: rispettivamente *CYYRI-1,2,3,4b* (primers #1 e #4), *CYYRI-1,2,3b* (primers #1 e #6), *CYYRI-1b,2,3,4* (primers #8 e #2) in SKMEL-28.

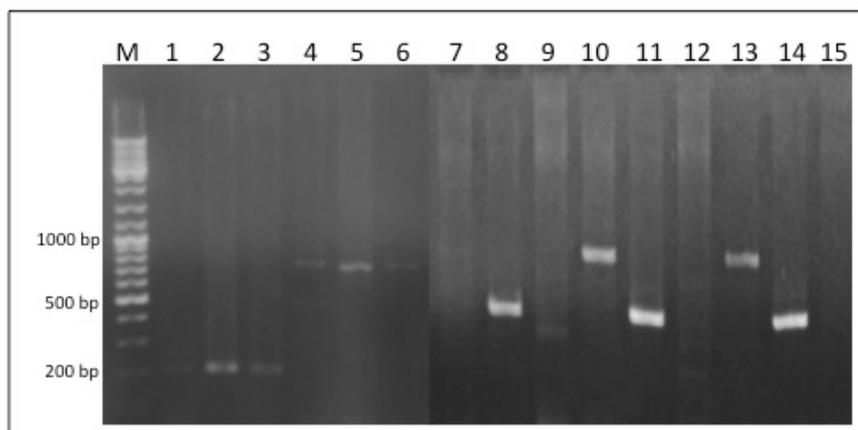
Corsie 7-8-9: rispettivamente *CYYRI-1,2,3,4b* (primers #1 e #4), *CYYRI-1,2,3b* (primers #1 e #6), *CYYRI-1b,2,3,4* (primers #8 e #2) in G361.

Corsie 10-11-12: rispettivamente *CYYRI-1,2,3,4b* (primers #1 e #4), *CYYRI-1,2,3b* (primers #1 e #6), *CYYRI-1b,2,3,4* (primers #8 e #2) in SJNKB.

Corsie 13-14-15: rispettivamente *CYYRI-1,2,3,4b* (primers #1 e #4), *CYYRI-1,2,3b* (primers #1 e #6), *CYYRI-1b,2,3,4* (primers #8 e #2) in SKNMC.

Corsie 16-17-18: rispettivamente *CYYRI-1,2,3,4b* (primers #1 e #4), *CYYRI-1,2,3b* (primers #1 e #6), *CYYRI-1b,2,3,4* (primers #8 e #2) in MCF7.

Corsia M: 1 $\mu$ l (500 ng) di marcatore Gene Ruler DNA Ladder Mix.



**Figura 25.** Elettroforesi su gel di agarosio dei prodotti ottenuti in RT-PCR (5 $\mu$ l) con coppie di primer per le diverse isoforme del *locus* di CYYR1 (Tabella 6) in cellule di osteosarcomi.

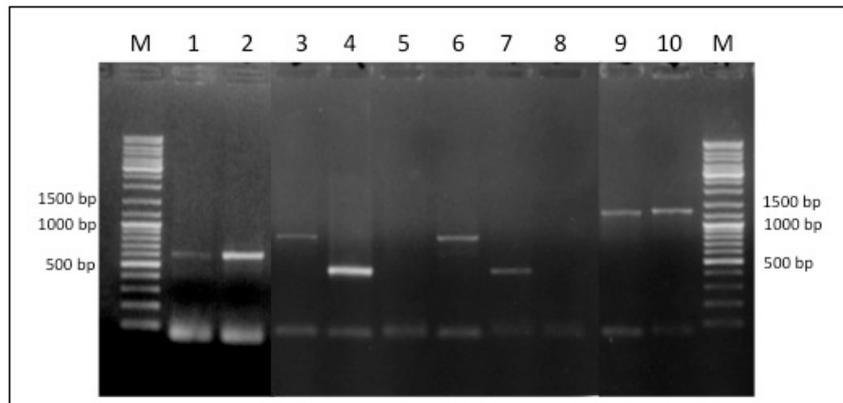
Corsie 1-2-3: *CYYR1* (primers #3 e #2) rispettivamente in SaOS, U2OS e U2OS-1. Corsie 4-5-6: *CYYR1-ASI* (primers #16 e #15) rispettivamente in SaOS, U2OS e U2OS-1. Corsie 7, 10 e 13: *CYYR1-1,2,3,4b* (primers #1 e #4) rispettivamente in SaOS, U2OS e U2OS-1. Corsia 8, 11 e 14: *CYYR1-1,2,3b* (primers #1 e #6) rispettivamente in SaOS, U2OS e U2OS-1. Corsia 9, 12 e 15: *CYYR1-1b,2,3,4* (primers #8 e #2) rispettivamente in SaOS, U2OS e U2OS-1. Corsia M: 1 $\mu$ l (500 ng) di marcatore Gene Ruler DNA Ladder Mix.

Ulteriori studi saranno necessari per verificare l'espressione delle restanti isoforme (*CYYR1-1,2,4* e *CYYR1-1,2,2bis,3,4*), così come per meglio definire il *pattern* di espressione del gene in antisense denominato *CYYR1-ASI*. Dati preliminari condotti su *CYYR1-ASI* indicano che si ha una elevata variabilità tra le linee, in particolare l'amplicone corrispondente alla zona centrale di *CYYR1-ASI* è presente in maniera significativa solo nelle linee di osteosarcoma e in una linea di melanoma (G362), mentre sembra poco espresso in una linea di neuroblastoma (SJNKP) e in MCF7 (adenocarcinoma della mammella). Non risulta affatto espresso in SKNMC (neuroblastoma), HT29 (adenocarcinoma del colon) e SKMEL28 (melanoma), (dati non mostrati).

Per quanto riguarda invece le linee cellulari primarie, sono state utilizzate due linee primarie fibroblastiche, una trisomica per il cromosoma 21 (CCL-54) e una euploide (CCL-110). In questo caso sono presenti tutti i

trascritti verificati in RT-PCR (*CYYRI-1,2,3,4*; *CYYRI-1,2,3,4b*; *CYYRI-1,2,3b*; *CYYRI-ASI*), tranne l'isoforma *CYYRI-1,2,3b* (Figura 26).

I dati ottenuti verranno ulteriormente approfonditi mediante una analisi quantitativa assoluta di PCR *Real-time*, i cui primi esperimenti hanno permesso di mettere a punto la preparazione degli standard molecolari da utilizzare come riferimento, di calcolare il numero esatto di copie del gene di interesse e verificarne l'andamento dell'espressione.



**Figura 26.** Elettroforesi su gel di agarosio dei prodotti ottenuti in RT-PCR (5 $\mu$ l) con coppie di primer per le diverse isoforme del *locus* di *CYYRI* (Tabella 6) in linee primarie di fibroblasti euploidi (CCL-110) e trisomiche per il cromosoma 21 (CCL-54).

Corsie 1-2: *CYYRI* (primers #1 e #2) rispettivamente in CCL-110 e CCL-54. Corsie 3 e 6: *CYYRI-1,2,3,4b* (primers #1 e #4) rispettivamente in CCL-110 e CCL-54. Corsie 4 e 7: *CYYRI-1,2,3b* (primers #1 e #6) rispettivamente in CCL-110 e CCL-54. Corsie 5 e 8: *CYYRI-1b,2,3,4* (primers #8 e #2) rispettivamente in CCL-110 e CCL-54. Corsie 9-10: *CYYRI-ASI* (primers #14 e #15) rispettivamente in CCL-110 e CCL-54. Corsia M: 1 $\mu$ l (500 ng) di marcatore Gene Ruler DNA Ladder Mix.

## 2. IL MODELLO ANIMALE *ZEBRAFISH* PER LO STUDIO FUNZIONALE E DELLA REGOLAZIONE GENICA

### 2.1. RT-PCR

### 2.1.1. Identificazione di geni *housekeeping* in pesce zebra

Lo studio dell'espressione genica necessita la caratterizzazione di geni di riferimento per le valutazioni quantitative relative. Per quanto riguarda il modello di zebrafish sono stati ritrovati in letteratura pochi dati significativi. In questo senso si è intrapreso una analisi sistematica al fine della identificazione di *housekeeping* "ideali", utili nella caratterizzazione della espressione genica sia in stadi embrionali che in tessuti adulti del pesce.

Seguendo i criteri descritti nella sezione 2.3.1. dei Materiali e metodi, sono stati selezionati otto geni candidati descritti in breve in Tabella 7. In particolare, *ef1a*, *slc25a5*, *zp2.4*, *ybx1* e *bactin2* sono stati scelti sulla base dei risultati ottenuti in seguito all'analisi del database di EST; *bactin2* e *ef1a* sono anche i geni maggiormente utilizzati come *housekeeping* secondo l'analisi della letteratura. Inoltre, sono stati selezionati *gapdhs* e *b2m* in quanto si tratta di geni ampiamente utilizzati come riferimento nei vertebrati; infine è stato deciso di considerare gene candidato anche *cyyr1*, in quanto ubiquitariamente espresso nell'uomo e con una espressione che si presenta in numerosi tessuti/stadi embrionali in *zebrafish*.

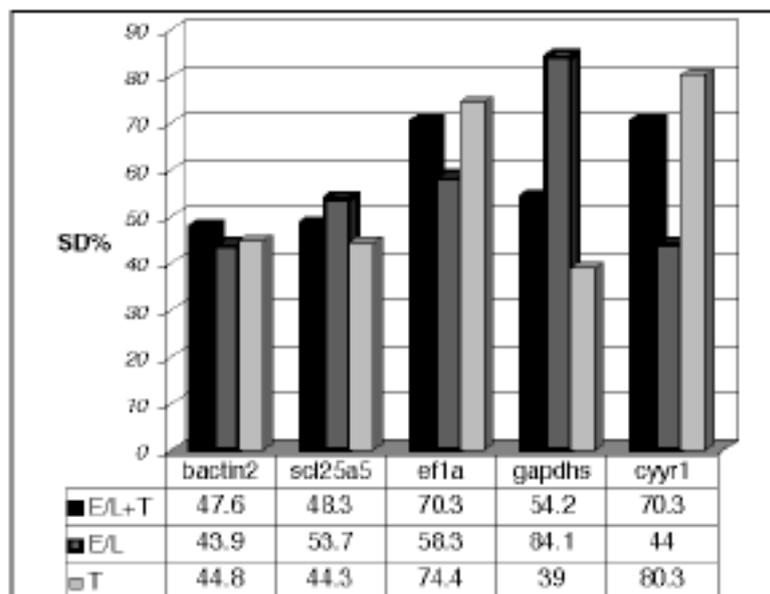
Sono stati ottenuti i prodotti di amplificazione attesi degli mRNA *ef1a*, *slc25a5*, *cyyr1*, *gapdhs* e *bactin2*: in tutti i casi, l'analisi su gel in seguito a corsa elettroforetica ha mostrato singole bande ad altezza attesa. Per tre geni (*ybx1*, *b2m* e *zp2.4*), l'mRNA è risultato poco o per nulla espresso in alcuni tessuti analizzati, per questo sono stati esclusi dalle analisi successive di RT-PCR quantitativa relativa.

Il valore di espressione di ogni gene negli stadi embrionali e nei tessuti analizzati è mostrato nelle Tabelle 12 e 13. Tutti i valori sono espressi in ng.

Ogni reazione di RT-PCR è stata condotta in doppio per verificare l'affidabilità del metodo: la media (espressa in percentuale) di differenza tra le misurazioni dei duplicati tecnici è del 4% con una deviazione standard di  $\pm 6.7\%$ .

Analisi di tipo statistico sono state condotte sui dati ottenuti, in particolare, l'espressione genica nell'intero pannello di RNA (stadi embrionali e tessuti) è stata espressa come percentuale della media dei valori di deviazione standard e il gene *bactin2* è risultato essere il più stabile tra tutti i campioni studiati ( $71.8 \pm 34.2$  ng, 47.6%, Figura 27).

Per verificare le differenze di espressione tra i diversi campioni biologici analizzati, è stato condotto il test statistico ANOVA, che ha permesso di rilevare che non vi sono differenze significative ( $p < 0.0001$ ). Una rappresentazione grafica dei valori di espressione per ogni gene è stata ottenuta attraverso il software di analisi statistica JMP (Figura 28); in figura sono mostrati anche i risultati ottenuti dall'analisi *post hoc* fatta attraverso il test di Tukey, (in particolare, i "livelli" denominati con lettere diverse mostrano una differenza significativa tra loro).



**Figura 27.** Stabilità dell'espressione genica dei candidati.

Il valore della deviazione standard è espresso come percentuale della media (SD%). E/L: embrioni/larva; T: tessuti adulti

A partire da queste analisi, è possibile effettuare una serie di osservazioni: i dati riguardanti il gene *bactin2* sono suddivisi in un minore numero di "livelli" (A, B e C), confermando la maggiore stabilità del gene nelle diverse condizioni analizzate.

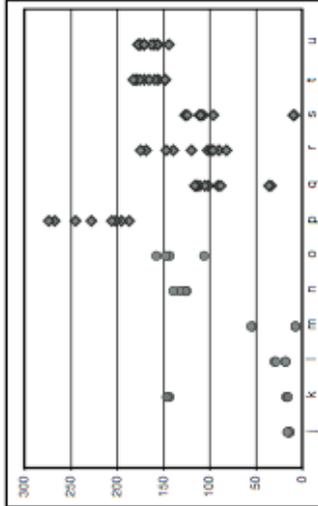
*bactin2*

Sample	Level	Mean
Ovary	A	144.0
5 days	B	79.4
Skin	B	78.6
48hpf	B	76.3
Eye	B	74.8
Bone	B	60.9
Muscle	B	57.3
Brain	B	54.6
1-5	B	51.9
Somites	B	48.4
Shield	B	45.4
Dome	B	45.4
Sphere	C	25.1



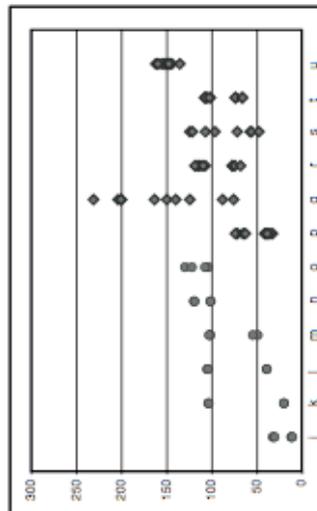
*gapdh*

Sample	Level	Mean
Brain	A	225.0
Eye	B	166.2
Ovary	B	164.8
5 days	B	133.1
48hpf	B	131.0
Skin	B	121.5
Muscle	C	90.9
Bone	C	86.4
Dome	D	81.1
1-5	D	31.3
Somites	D	31.3
Shield	E	24.1
Sphere	F	15.1



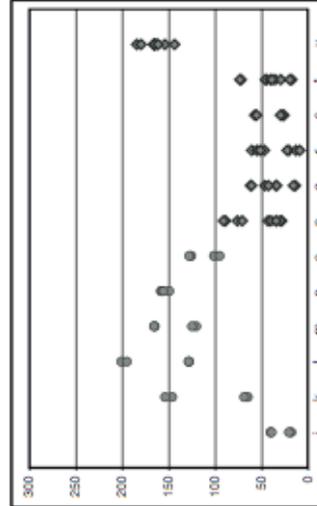
*sic25a5*

Sample	Level	Mean
Muscle	A	157.4
Ovary	A	150.3
5 days	A	116.1
48hpf	A	110.7
Skin	B	97.5
Eye	B	90.8
Bone	B	84.9
1-5	B	77.3
Somites	B	72.2
Shield	B	62.1
Dome	B	62.1
Brain	C	49.7
Sphere	D	21.5



*cypr1*

Sample	Level	Mean
Ovary	A	164.8
Shield	A	163.6
48hpf	A	155.5
1-5	A	144.4
Somites	B	112.9
5 days	B	108.7
Dome	C	53.4
Brain	D	41.5
Eyes	D	38.4
Muscle	D	36.8
Bone	D	34.7
Skin	D	34.7
Sphere	D	29.8



*efla*

Sample	Level	Mean
Ovary	A	131.9
5 days	A	106.7
48hpf	A	106.6
1-5	B	97.6
Somites	B	88.7
Shield	C	54.7
Eye	C	53.2
Skin	C	36.4
Muscle	D	24.2
Bone	D	19.9
Brain	D	16.8
Dome	D	16.8
Sphere	D	11.1



**Figura 28.** Rappresentazione grafica dei dati di espressione genica per ogni gene candidato.

● Stadi di sviluppo embrionale (j: sfera; k: dome; l: shield; m: 1-5 somiti; n: 48 hpf; o: 5 giorni).

◆ Tessuti adulti (p: cervello; q: muscolo; r: pelle; s: ossa; t: occhi; u: ovaio).

I dati sono espressi in nanogrammi. Le tabelle sono state ottenute utilizzando il software JMP per l'analisi di Tukey. "Livelli" non connessi fra loro dalla stessa lettera mostrano differenze statisticamente significative ( $p < 0.05$ ).

Inoltre, analizzando l'espressione genica separatamente per i due gruppi di campioni (stadi embrionali e tessuti), si può osservare come *bactin2* rimanga il gene più stabile tra gli stadi embrionali ( $54.4 \pm 23.9$  ng, 43.9%), mentre fra i tessuti sembra essere più stabile il gene *gapdhs* ( $141.5 \pm 55.1$  ng, 39.0%, Figura 27)

### **2.1.2. Caratterizzazione di CNEs**

Allo scopo di amplificare CNEs di *Ciona intestinalis*, da microiniettare in embrioni di pesce zebra, sono state messe a punto reazioni di PCR come descritto nel paragrafo 2.3.2 dei MATERIALI E METODI.

Sono state ottenute bande singole e ad altezza attesa per 19 su 21 CNEs (non sono state ottenute bande attese per le CNEs Mesi\_8 e Zfhx1B\_2).

### **2.2. PCR *Real-time* per lo studio di espressione di *cyrr1***

Lo studio della espressione di *cyrr1* in diversi stadi dello sviluppo embrionale è stato realizzato attraverso un approccio di tipo quantitativo relativo in PCR *Real-time* e sulla base di dati pregressi, di tipo descrittivo, ottenuti con tecniche di ibridazione *in situ whole mount* su embrioni di *zebrafish*.

*Primers* specifici sono stati progettati per lo studio di espressione di *cyrr1* in reazioni di PCR *Real-time* (vedi MATERIALI E METODI) e verificati in RT-PCR tradizionale per l'amplificazione di bande singole e ad altezza attesa per tutti i geni, sia *cyrr1* che i geni *housekeeping* scelti come geni di riferimento: *bactin2*, *ef1a* e *slc25a5* (per la scelta si veda il paragrafo 2.1.1. dei Risultati).

Piastre preliminari di *Real-time* PCR sono state realizzate per l'ottimizzazione delle varie condizioni sperimentali, in particolare sono state verificate le coppie di *primers* a diverse concentrazioni finali, 300 e 500 nM.

Questa fase iniziale ha richiesto anche la messa a punto di una piastra per determinare l'efficienza di reazione, misura della velocità alla quale la polimerasi converte i reagenti (dNTPs, oligonucleotidi e stampo di cDNA) in amplicone.

Per ogni coppia di *primers*, quindi per ogni gene considerato nella reazione, è stata determinata una curva standard, intesa come *log* della quantità di partenza dello stampo contro i valori di  $C_q$  ottenuti. Il termine *standard* "ciclo di quantificazione" (*quantification cycle*,  $C_q$ ) è stato recentemente introdotto per uniformare le diverse terminologie (tra cui "ciclo soglia", *threshold cycle*,  $C_t$ ); la proposta fa parte infatti di una serie di linee guida, utili a fornire le informazioni di base a chi pubblica e a chi revisiona analisi di PCR quantitativa *Real-time*, per assicurare la più corretta interpretazione, accuratezza e ripetibilità degli esperimenti (Bustin *et al.*, 2009; Bustin *et al.*, 2010).

Due campioni di cDNA scelti dal nostro pannello sperimentale, sono stati diluiti a concentrazioni seriali dell'ordine di 10 (in particolare, 5 diluizioni da 40 ng/ml a 0,004 ng/ml) ed amplificati in duplicato in condizioni standard.

La pendenza della retta (*slope*) della curva standard è un indice della efficienza della reazione in base alla formula:

$$E = 10^{-1/\text{slope}}$$

Con una efficienza del 100% *slope* = -3,32, ossia i valori di  $C_q$  dovrebbero essere separati da 3,32 cicli. I valori di E accettabile variano tra 90-110%. Per quanto riguarda i geni considerati i valori sono risultati ottimali e sono

stati inseriti nel programma di analisi della espressione genica “CFX Manager™ v1.1” (Bio-Rad, Hercules, CA, USA), che tiene conto appunto delle singole efficienze di reazione (Pfaffl, 2001).

Un altro parametro che si ottiene dalla curva standard è  $r^2$  (coefficiente di determinazione) che rappresenta quanto i dati sperimentali si adattano alla linea di regressione, cioè quanto lineari sono i dati. I valori di  $r^2$  per i geni considerati sono risultati tutti accettabili, ossia maggiori di 0.985.

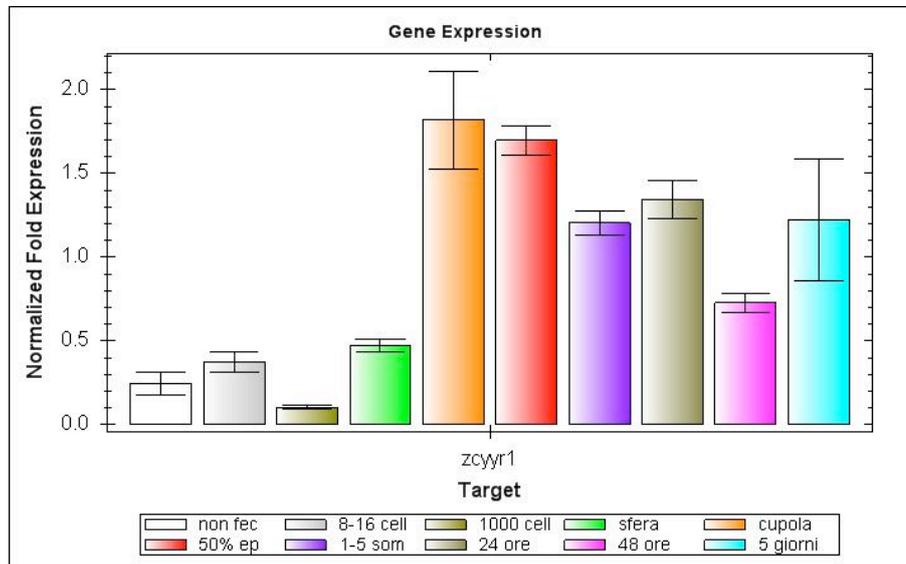
La stabilità dei geni di riferimento, utilizzati per la normalizzazione dei dati di espressione, è stata calcolata dallo strumento stesso, che utilizza il metodo geNorm (Vandesompele *et al.*, 2002) e fornisce il valore  $M$ , indicativo appunto della stabilità di espressione di un gene in differenti condizioni. La  $M$  dei nostri geni *housekeeping* è risultata compresa tra 0.5 e 1, definiti valori ottimali per un buon gene di riferimento.

L’espressione del gene *cyyr1* in un pannello di diversi tessuti embrionali di *zebrafish* è stata dunque valutata come detto come la quantità relativa del gene di interesse normalizzata con le quantità dei geni di riferimento, *bactin2*, *ef1a* e *slc25a5*. Il valore di espressione normalizzato, o  $\Delta\Delta C_t$  ( $DDC_t$ ), calcolato automaticamente dal programma CFX Manager™ Bio-Rad, è inteso come quantità relativa (RQ) del gene di interesse (*target*), in un certo campione (*sample*), normalizzata con le quantità dei geni di riferimento (*Ref1*, *Ref2...Refn*) negli stessi campioni, ovvero:

$$\Delta\Delta C_{t \text{ sample (target)}} = \frac{RQ_{\text{ sample (target)}}}{(RQ_{\text{ sample (Ref1)}} \times RQ_{\text{ sample (Ref2)}} \times \dots \times RQ_{\text{ sample (Refn)}})^{1/n}}$$

In Figura 29 è riportato il grafico che mostra l’andamento dell’espressione del gene *cyyr1* nei diversi stadi embrionali di *zebrafish* considerati, ottenuto dall’analisi dei dati di uno dei due set di campioni generati da due esperimenti indipendenti eseguiti. Lo stesso andamento è stato confermato

della analisi dei dati prodotti dal secondo esperimento, a partire degli stessi stadi embrionali, ma da campioni diversi di RNA.



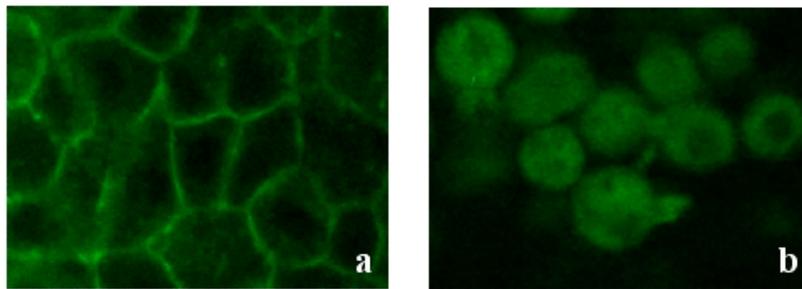
**Figura 29.** Analisi della espressione del gene *zcyyr1* di *zebrafish* in diversi stadi embrionali (l'errore è rappresentato come Errore Standard della media).

### 2.3. Studio funzionale di *zcyyr1* in *zebrafish*

Dall'analisi bioinformatica è emerso, come citato nell'Introduzione, che la proteina putativa codificata da *zcyyr1* presenta un dominio transmembrana, caratteristico della famiglia genica. Per verificare questa predizione e dimostrare sperimentalmente che la proteina di *zcyyr1* si localizza in membrana è stato progettato un costrutto di fusione tra la sequenza codificante di *zcyyr1* privata del segnale di stop e il gene per la GFP (*Green Fluorescent Protein*). Come controprova è stato anche realizzato un costrutto di fusione tra la sequenza codificante di *zcyyr1*, deleva della porzione transmembrana e del segnale di stop, e la GFP. Tali costrutti sono stati poi microiniettati in embrioni di 1-2 cellule e rilevati all'interno delle cellule seguendo l'espressione della GFP facilmente visibile al microscopio per la sua proprietà di autofluorescenza.

Gli embrioni dopo la microiniezione sono stati osservati ad intervalli regolari per 10 ore. L'analisi microscopica ha permesso di

verificare che la proteina di fusione Cyvr1-GFP si localizza sulla membrana plasmatica, principalmente ai siti di contatto cellula-cellula (Figura 30A). L'osservazione degli embrioni microiniettati con il costrutto di fusione in cui la regione codificante di *cyvr1* è stata deleta della sequenza corrispondente alla regione transmembrana, permette di constatare agli stessi tempi di osservazione, la presenza di una diffusa fluorescenza a livello citoplasmatico, confermando l'importanza e la funzione della regione transmembrana deleta (Figura 30B).



**Figura 30.** Localizzazione subcellulare della proteina di fusione Cyvr1-GFP. (a) Localizzazione della proteina di fusione tradotta dall'mRNA *cyvr1*- $\Delta$ 3'end- GFP a 10 ore dopo la microiniezione. (b) Localizzazione della proteina di fusione tradotta dall'mRNA *cyvr1*-  $\Delta$ TM-GFP a 10 ore dopo la microiniezione.

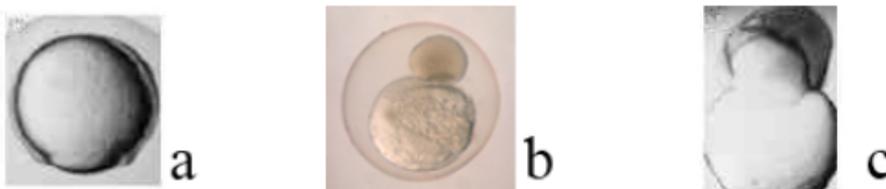
In parallelo a questo approccio sperimentale, si sono valutati gli effetti della sovraespressione del gene *cyvr1* durante lo sviluppo, microiniettando gli embrioni con il mRNA relativo alla sequenza codificante completa di *cyvr1*, ad una concentrazione di 100 ng/ $\mu$ L, per un totale di 50 pg per iniezione.

Conseguenza della espressione ectopica realizzata è stata la alterazione evidente del processo epibolico durante la fase di gastrulazione (5 1/4 - 10 1/3 hpf).

L'osservazione è stata reiterata a intervalli regolari fino a 24 hpf. Gli embrioni iniettati hanno mostrato un ritardo nell'ingresso e nella progressione attraverso le fasi dell'epibolia. Circa il 78% degli iniettati hanno presentato un restringimento del blastoderma tale da estrarre il

turlo, fino a conferire all'embrione una tipica morfologia ad "8", in cui la sfera più piccola è costituita dalle cellule ammassate del blastoderma, quella più grande invece dal turlo (Figura 31b).

Questo fenotipo è del tutto analogo al fenotipo B descritto da Strähle e Jesuthasan (1993) a proposito degli effetti delle radiazioni UV sugli embrioni di *zebrafish*, e risulta letale prima di 24 ore per il 64% degli iniettati (Figura 30c). Gli embrioni che sono riusciti a superare la fase di epibolia, all'osservazione dopo 24 ore, hanno presentato anomalie nello sviluppo del piano corporeo, con difetti di varia tipologia e grado (non mostrati).



**Figura 31.** Effetti della sovraespressione di *cyrr1*.

(a) Embrione normale allo stadio di 80% di epibolia. (b) Embrione anomalo in seguito alla microiniezione dell'mRNA di *cyrr1*. (c) Embrione anomalo in seguito all'esposizione a raggi UV. Pannelli (a) e (c) da Strähle *et al.*, 1993

Contemporaneamente, embrioni di controllo non microiniettati e osservati nelle stesse fasi di sviluppo analizzate nei campioni microiniettati, hanno mostrato un fenotipo normale e il 100% di vitalità.

Come noto anche la inibizione transiente della espressione genica può essere fortemente informativa della funzione genica. A questo fine si sono ottenuti organismi *knock-down* mediante la microiniezione di oligonucleotidi antisenso morfolinati.

In particolare, gli oligonucleotidi morfolino antisenso (MO) sono stati utilizzati nel campo della biologia dello sviluppo dal 2000 e la loro introduzione è stata accolta con entusiasmo, poiché essi rappresentano una nuova tecnologia per l'ottenimento di "*loss of function*" (Heasman, 2002) e

risultano quindi essere ottimali per lo studio della funzione genica *in vivo* (Corey e Abrams, 2001).

Queste molecole hanno numerose caratteristiche che ne fanno degli ottimi strumenti di analisi; essi, infatti, non hanno effetti tossici, si appaiano in modo specifico alla sequenza bersaglio impedendone la traduzione e hanno una buona stabilità.

Per ottenere indicazioni sul ruolo di *cyrr1*, durante l'embriogenesi, sono stati microiniettati oligonucleotidi morfolinici antisenso in embrioni di *zebrafish* a 1-2 cellule.

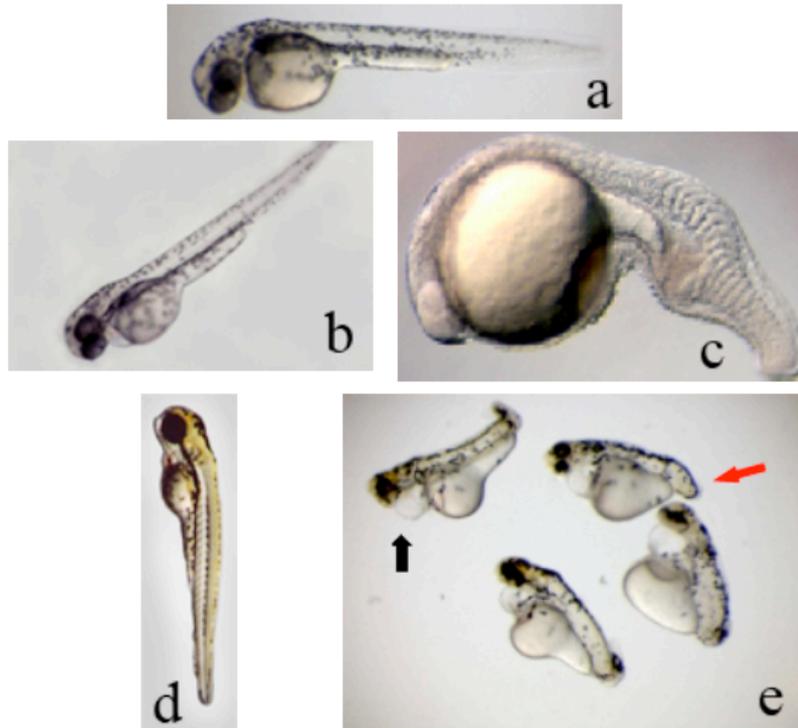
Ogni embrione è stato iniettato con 1 nL di soluzione circa a concentrazione compresa fra 2 e 6 mg/mL. Gli embrioni sono stati quindi osservati al microscopio, allo scopo di rilevare alterazioni morfologiche indotte dagli oligonucleotidi morfolinici.

Lo sviluppo embrionale negli stadi precedenti alla gastrulazione è apparso normale, tranne che per un generalizzato e lieve ritardo temporale nella progressione dello sviluppo.

Dopo la gastrulazione, gli embrioni microiniettati hanno mostrato una varietà di alterazioni fenotipiche. Sono stati, infatti, riscontrati difetti (Figura 32e):

- dell'estremità caudale, con modificazioni della curvatura naturale da minori a gravi;
- degli occhi e della testa, di dimensioni significativamente minori rispetto agli embrioni di controllo;
- del pericardio, con edema rilevante;
- della coordinazione motoria, con notevoli contrazioni parossistiche ritmiche;
- della velocità di riassorbimento del tuorlo, inferiore al normale.

Gli effetti della inibizione genica sono stati osservati fino a 72 ore dopo l'iniezione, in cui una percentuale dal 93% al 95% degli embrioni sopravvissuti risultava anomala.



**Figura 32.** Effetti del *cyrr1*-MO sul normale sviluppo embrionale del pesce zebra.

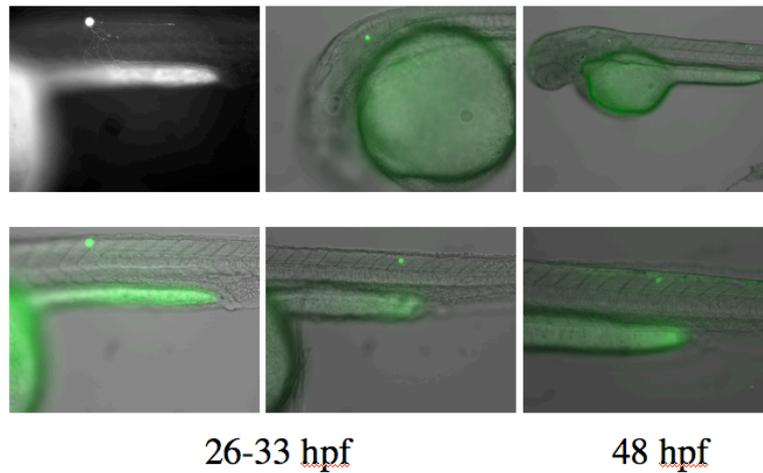
(a) Embrione normale a 27 hpf. (b) Embrione microiniettato con morfolino di controllo a 48 hpf. (c) Embrioni microiniettati con *chd*-MO a 24 hpf. (d) Embrione normale a 72 hpf. (e) Embrioni a 72 hpf microiniettati con *cyrr1*-MO mostrano una dimensione anomala della testa e degli occhi con un evidente edema pericardiale (indicate dalla freccia nera). Inoltre sono presenti anomalie a livello della coda (indicate dalla freccia rossa).

#### **2.4. Studio della regolazione genica di CNEs di *Ciona intestinalis* attraverso l'utilizzo di *zebrafish***

In seguito a microiniezione della soluzione contenente la CNE di *Ciona intestinalis* studiata, sono state effettuate osservazioni per verificare l'espressione di GFP negli embrioni di pesce zebra microiniettati (Paragrafo 3.8.B Materiali e Metodi).

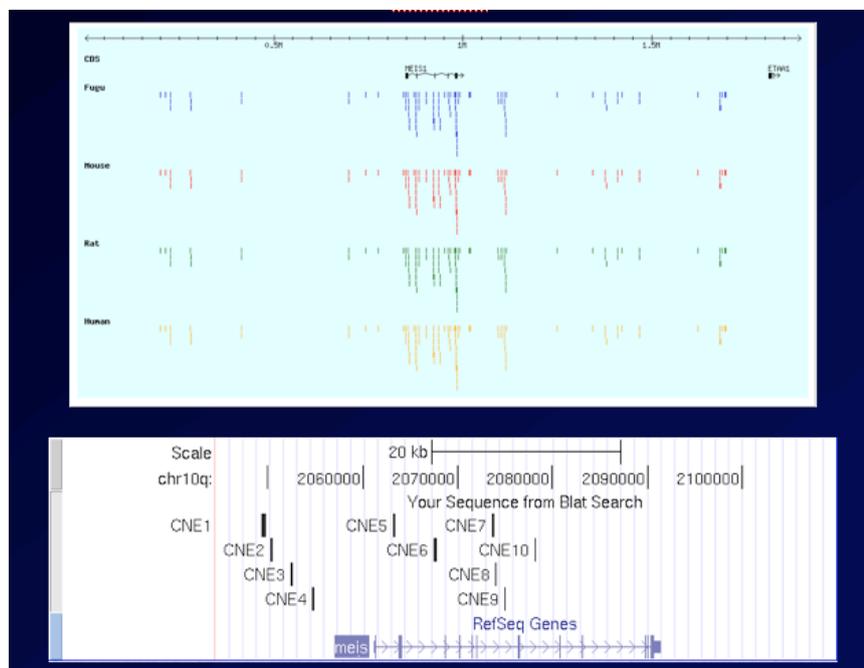
Sono stati ottenuti risultati positivi per due CNEs microiniettate: Pax6\_2 e Meis\_1.





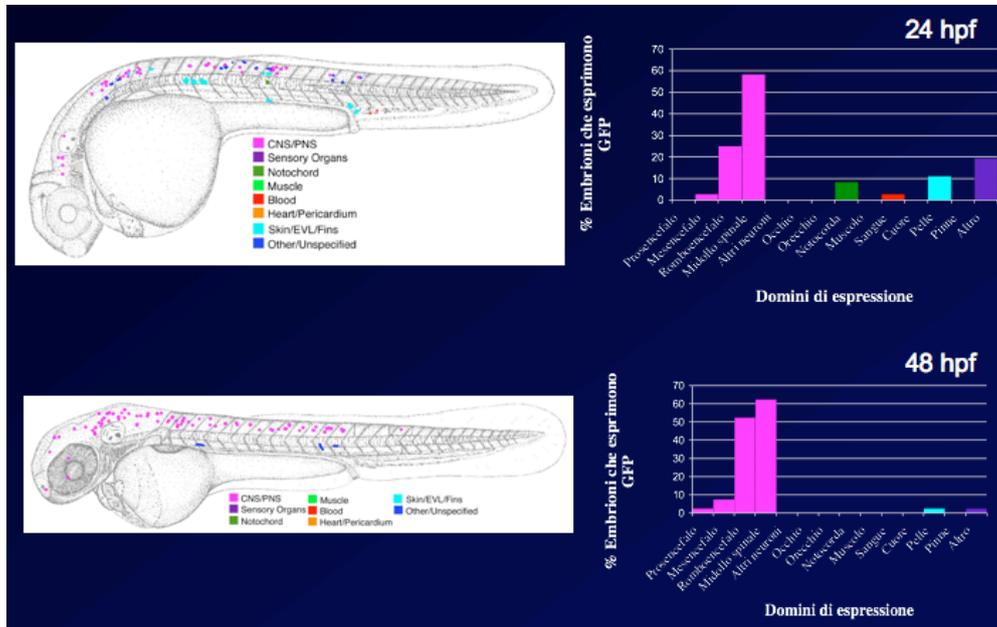
**Figura 35.** Espressione del gene reportere GFP dopo microiniezione della CNE Pax6\_2. Le osservazioni sono state effettuate a 24 hpf e 48 hpf.

Per quanto riguarda il gene *meis*, sono state ad esso associate 9 CNEs in *Ciona intestinalis*, un numero molto elevato che rispecchia anche la situazione che si osserva nei vertebrati dove i geni della famiglia Meis sono associati al numero più elevato di CNEs (Figura 36).

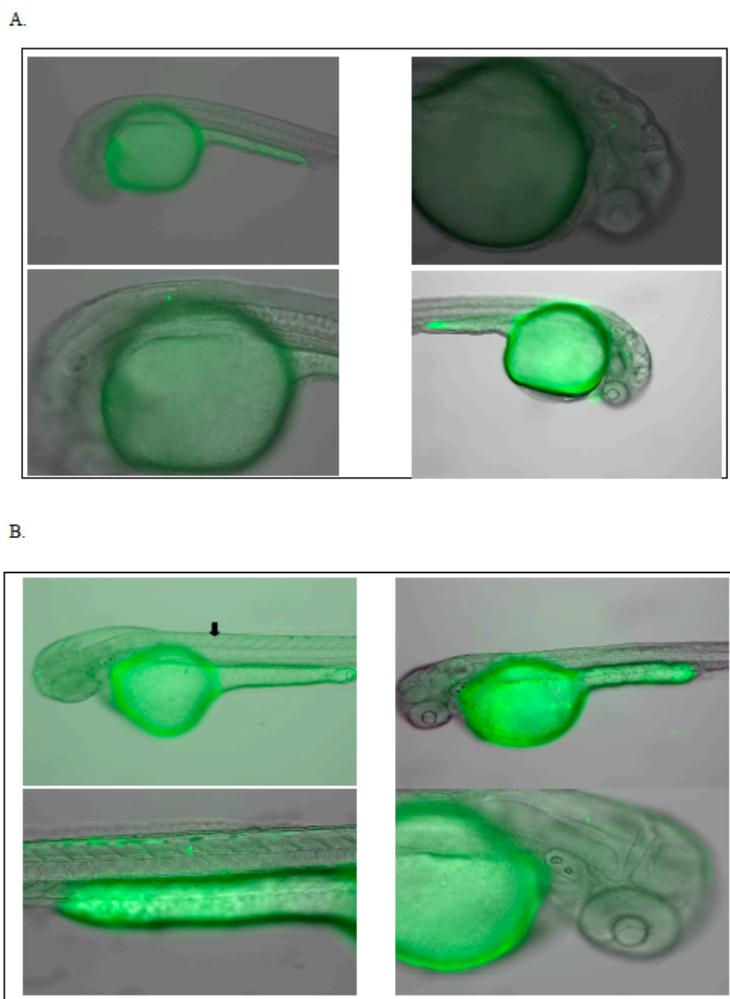


**Figura 36. A.** Localizzazione delle CNE associate al famiglia genica Meis in diverse specie di vertebrati; **B.** Localizzazione delle CNE associate al famiglia genica Meis in *Ciona intestinalis*.

L'espressione della GFP in seguito a microiniezione in embrioni di pesce zebra dell'elemento Meis\_1 è stata osservata nel 18,2% dei casi, in particolare il gene *reporter* si accende in diverse zone del sistema nervoso come schematizzato in Figura 37 e Figura 38.



**Figura 37.** Schematizzazione grafica dei risultati ottenuti in seguito a microiniezione della CNE Meis\_1. Le osservazioni sono state effettuate a 24 hpf e 48 hpf.



**Figura 38.** Espressione del gene reporte GFP dopo microiniezione della CNE Meis\_1. Le osservazioni sono state effettuate a 24 hpf (A) e 48 hpf (B).

### 3. SVILUPPO DI UN SOFTWARE IN GRADO DI PRODURRE MAPPE CROMOSOMICHE DI ESPRESSIONE (TRAM)

TRAM è un nuovo *software* che permette di creare ed analizzare mappe quantitative del trascrittoma, a partire da valori di espressione genica (per esempio, valori di espressione ottenuti tramite *microarrays*).

Il programma è costituito da 37 tabelle di database correlate fra loro da 118 relazioni; alcune tabelle sono costruite in modo da permettere la conversione del codice identificativo di un gene (*gene identifier*) nel simbolo genico ufficiale. Le tabelle di conversione possono essere caricate

dall'utente oppure, per quanto riguarda uomo, topo e pesce zebra, si trovano già preconfigurate all'interno del programma.

Un campione (TRAM *sample*) è definito da una serie omogenea di codici identificativi di geni e dal corrispondente valore di espressione. Il programma è in grado di effettuare confronti tra un campione (o un insieme di campioni) definito "Sample A" e un campione (o insieme di campioni) distinto nominato "Sample B". In questo caso, verrà analizzato il rapporto A/B tra i valori di espressione per ogni locus. Inoltre, TRAM è in grado di effettuare la normalizzazione dei dati sia *intra-sample* (media o mediana) sia *inter-sample* (metodo dei quantili scalati), cosa che permette di ottenere valori confrontabili fra loro, anche quando si tratta di confrontare piattaforme con una quantità molto differente di dati. In particolare, l'utilizzo del metodo dei quantili scalati permette sempre di ottenere un risultato analogo se non migliore rispetto a quello che si otterrebbe utilizzando il metodo standard dei quantili.

E' possibile effettuare analisi in due modalità: "*Map*" *mode* e "*Cluster*" *mode*. Attraverso la prima funzione, il software genera una rappresentazione quantitativa del trascrittoma di un campione (o di un insieme di campioni) e determina se segmenti cromosomici di lunghezza definita si possano definire sovra- o sottoespressi rispetto a una certa soglia fissata dall'utente. Quest'ultimo può facilmente navigare tra le mappe di geni e le tabelle di valori di espressione; i segmenti cromosomici possono essere esplorati ed indagati in base ai criteri fissati dall'utente. La modalità "*Cluster*", invece, permette di ricercare geni contigui sovra- o sottoespressi.

La significatività statistica dei segmenti o cluster sovra- o sottoespressi può essere calcolata rispetto all'insieme di tutti i geni nel genoma oppure solo rispetto ai geni che si trovano all'interno del cromosoma a cui appartiene il segmento o il cluster analizzato. Il calcolo viene eseguito come descritto nella sezione "Materiali e metodi" e mostrato al termine dell'analisi.

## **DISCUSSIONE E CONCLUSIONI**

La Sindrome di Down (trisomia 21) è la più frequente anomalia cromosomica dell'uomo, con frequenza di 1 su 400 concepiti e 1 su 700 nati vivi.

I sintomi sono attribuiti a effetti di dosaggio genico per geni localizzati sul cromosoma 21, presente in tre copie nei soggetti con la sindrome. Sebbene la Sindrome di Down sia la prima alterazione genetica dimostrata nell'uomo, e la forma più frequente di ritardo mentale causato da una aberrazione cromosomica microscopicamente dimostrabile, la patogenesi molecolare è ancora ampiamente sconosciuta.

Il processo di annotazione genica è ancora in corso per il cromosoma 21 umano: attualmente circa 420 geni e modelli genici esistono sul braccio lungo del cromosoma 21 e 4 sul braccio corto, di questi ci sono informazioni su solo 145 (Patterson *et al.*, 2009). L'identificazione e l'analisi di tutti i geni del cromosoma 21 umano e l'analisi della espressione genica è attualmente al centro dell'interesse negli studi biomedici volti a comprendere la patogenesi della Sindrome di Down.

Nonostante si presuma che le caratteristiche fenotipiche della SD siano una diretta conseguenza della sovraespressione di un certo numero di geni contenuti all'interno del cromosoma, è difficile selezionare i geni candidati per questo fenotipo complesso, anche se alcuni prodotti genici possono essere più sensibili agli squilibri di dosaggio genico di altri. Mentre alcuni *loci* possono avere un maggiore effetto fenotipico, è l'effetto cumulativo di squilibrio di più geni che determina il fenotipo globale. La ricerca di base si rivolge oggi ad una valutazione più precisa della variazione complessiva dei livelli di espressione genica, in cellule di individui affetti da SD o utilizzando modelli animali principalmente di topo trisomico. Lo scopo finale di questo approccio è comprendere come varino i livelli di espressione genica sia per quanto concerne la espressione di geni presenti in triplice copia, sia per quelli euploidi. L'indicazione chiara di questi ultimi anni è che molti *loci* genici, ma non tutti risentano, almeno a livello dell'mRNA, di una reale sovraespressione (Sommer *et al.*, 2008; Lockstone *et al.*, 2007; Altug-Teber *et al.*, 2007; Conti *et al.*, 2007;

Lott *et al.*, 2005; Helguera *et al.*, 2005). Questi studi hanno prodotto una pletera di dati la cui interpretazione è veramente molto complessa, soprattutto alla luce della inattesa complessità nella struttura, espressione e regolazione, anche epigenetica, della informazione contenuta nel genoma umano. La ricerca dunque è ancora molto aperta in questo campo e chiaramente indica come futura direzione la necessità di definire la funzione di tutti i geni del cromosoma 21 e comprendere il ruolo della variabilità di espressione e i meccanismi molecolari che sottendono a tali variazioni e che possono contribuire alla generazione del fenotipo caratteristico della SD.

Il presente lavoro si propone di effettuare studi su HSA21 in relazione alla Sindrome di Down cercando di rispondere alle due domande che la ricerca scientifica lascia ancora aperte: da una parte, la necessità di caratterizzare dal punto di vista strutturale e funzionale i geni del Cromosoma 21, dall'altra la necessità di effettuare studi che permettano di analizzare la variazione complessiva dei livelli di espressione genica, le interazione tra geni diversi e i meccanismi di regolazione dei *pathway* molecolari che sottendono le diverse funzioni biologiche.

Vi sono numerosi geni sovraespressi in modo significativo e stabile nei soggetti SD: da un recente studio di analisi ad ampio spettro sui geni del HSA21 (Prandini *et al.*, 2007), tra quelli a minor fluttuazione nella espressione genica è stato riscontrato il gene *cysteine/tyrosine-rich 1* (*CYYR1*). La prima parte dello studio è stata dunque dedicata alla caratterizzazione strutturale e di espressione del *locus* genico umano e allo studio funzionale, utilizzando come modello animale il pesce zebra.

*CYYR1*, descritto per la prima volta nel 2002 (Vitale *et al.*, 2002), è localizzato sul Cromosoma 21 umano (21q21.2), è espresso in una ampia gamma di tessuti umani normali e codifica per una proteina che sembra essere altamente conservata nei vertebrati e assente nei procarioti e negli invertebrati. L'analisi molecolare della sequenza del *locus* *CYYR1* ha consentito di dimostrare che si tratta di un sistema multi-trascrizione (che occupa uno spazio fisico cospicuo pari allo 0,3% dell'intero cromosoma 21) con trascritti alternativi, posti anche in antisenso alla forma di riferimento inizialmente clonata. Quest'ultima è formata da 4 esoni e 3

lunghi introni, ha una ORF di 462 nucleotidi e codifica per una proteina di 154 amminoacidi. *CYYRI* non mostra somiglianze con altre proteine conosciute e il suo ruolo all'interno della cellula è a tutt'oggi sconosciuto. La proteina predetta è caratterizzata da tre domini conservati: uno centrale ricco in cisterna e tirosina (CCSYAY), tre motivi PPPY all'estremità carbossi-terminale, studiati per le possibili implicazioni tra elementi proteici del citoscheletro e matrice cellulare (Pereboev *et al.*, 2001; Ilsley *et al.*, 2002) e un dominio transmembrana.

La combinazione di analisi *in silico* e *in vitro* ha permesso di caratterizzare il *locus* del gene *CYYRI*, in particolare di identificare 5 nuove isoforme e un trascritto in antisenso rispetto a *CYYRI*: *CYYRI-1,2,3,4b* con una porzione alternativa a partire dalla posizione 656 della sequenza di riferimento #NM\_052954 (fine dell'esone 3); *CYYRI-1,2,3b* con una porzione alternativa a partire dalla posizione 498 della sequenza di riferimento #NM\_052954 (fine dell'esone 2); *CYYRI-1,2,4* che mostra la mancanza di 158 basi corrispondenti all'intervallo 499-656 della sequenza di riferimento #NM\_052954 (quindi mancanza completa dell'esone 3) e la presenza di una forma di "splicing sottile" CAG+, precedentemente descritta (Vitale *et al.*, 2007); *CYYRI-1,2,2bis,3,4* che consiste di una sequenza con 167 basi aggiuntive tra gli esoni 2 e 3 della sequenza di riferimento #NM\_052954; *CYYRI-ASI* un lungo messaggero che si presenta sul filamento opposto rispetto a *CYYRI* antisenso ed è in parte embricato con esso.

Tutte le isoforme di splicing di *CYYRI* codificano per proteine putative che mantengono il dominio centrale ricco in tirosina e cisteina (CCSYAY) fortemente conservato dai vertebrati inferiori all'uomo che contraddistingue questa nuova famiglia di prodotti proteici. Esistono però delle differenze rilevanti tra di esse che potrebbero far ipotizzare funzioni differenti mediate dalle varie forme di splicing: in particolare, *CYYRI-1,2,3,4b* potrebbe codificare per una proteina ipotetica con una porzione C terminale diversa rispetto alla proteina di riferimento (*CYYRI* #NP\_443186), con conseguente perdita dei tre motivi PPPY conservati; la proteina ipotetica risultante della isoforma *CYYRI-1,2,3b* sarà molto più corta rispetto alla proteina di riferimento con perdita degli amminoacidi che

identificano il dominio caratteristico transmembrana e dei motivi PPPY conservati; la proteina ipotetica codificata dalla forma di *splicing* denominata *CYYRI-1b,2,3,4* avrà una porzione N-terminale diversa rispetto alla proteina di riferimento, con spostamento del codone di inizio della CDS; *CYYRI-1,2,4* manca dell'intero esone 3 e questo determina la perdita del normale schema di lettura, quindi, nella proteina predetta, il cambiamento della catena polipeptidica; infine, l'isoforma denominata *CYYRI-1,2,2bis,3,4* è caratterizzata dalla presenza dell'esone 2bis che introduce uno STOP in *frame* precoce e la proteina predetta risulta dunque più corta rispetto alla proteina di riferimento, con perdita degli aminoacidi che identificano il dominio caratteristico transmembrana e dei motivi PPPY conservati (Figura 22B).

Potrebbe essere molto interessante sviluppare queste osservazioni andando a verificare la localizzazione delle diverse isoforme a livello cellulare, la possibile interazione dei domini conservati con altre proteine o con specifiche sequenze ed infine le funzioni mediate da geni codificanti prodotti con differenze così rilevanti, in quanto si tratta di perdita o mantenimento di domini funzionali.

La complessità del *locus* genico *CYYRI*, inoltre, è arricchita dalla presenza di trascritti in antisenso rispetto al gene; in particolare, è stato clonato il trascritto denominato *CYYRI-ASI*. La comprensione dell'organizzazione genomica di questo messaggero in antisenso sarà il punto di partenza per effettuare studi funzionali che mirino a verificare un suo possibile ruolo di tipo regolativo, per esempio come RNA non codificante (ncRNA).

La caratterizzazione di un *locus* così complesso, avvalorata la tesi secondo cui la regolazione dell'espressione genica negli eucarioti superiori è legata a numerosi fattori, tra cui la presenza di isoforme derivanti da *splicing* alternativo, sequenze geniche sovrapposte ("overlapping genes"), RNA non codificanti e *natural antisense* RNA (NAT), la cui presenza permette di definire un *locus* "multi-trascritto" (Carninci *et al.*, 2008). Si stima che il 40-60% di tutti i geni umani e il 74% dei geni umani multiesonici vadano incontro a *splicing* alternativo (Kim *et al.*, 2004). Esistono, poi, numerosi meccanismi che possono mediare lo *splicing*

alternativo dei geni umani, dalla perdita di tre basi di un esone (*splicing* sottile) (Vitale *et al.*, 2006; Vitale *et al.* 2007) fino alla perdita di uno o più esoni (Blencowe *et al.*, 2006). La ricchezza di meccanismi molecolari potrebbe spiegare la complessità funzionale dei vertebrati rispetto agli organismi invertebrati (Blencowe *et al.*, 2006).

Dati in letteratura mostrano come *CYYRI* sia un gene espresso in una vasta gamma di tessuti (Vitale *et al.*, 2002); alla luce della caratterizzazione di nuove isoforme, sono stati effettuati esperimenti di RT-PCR per verificare la loro espressione in diversi tessuti e linee cellulari.

Per quanto riguarda le indagini preliminari condotte in diversi tessuti umani normali (quali cervello, cuore, testicolo, sangue, intestino) e tessuti tumorali (tumori neuroendocrini di fegato, pancreas, ghiandola mammaria), non hanno mostrato particolari tessuto-specificità nè per le isoforme di *splicing*, nè per il trascritto in antisenso (Figura 23).

E' stata verificata la presenza delle isoforme del *locus* in numerose linee cellulari tumorali continue tra cui: SJNKP e SKNMC di neuroblastoma; HT29 di adenocarcinoma del colon; MCF7 di adenocarcinoma della mammella; SK-MEL28 e G361 di melanoma ed infine linee tumorali di osteosarcoma (SaOS, U2OS e U2OS-1). La isoforma di riferimento di *CYYRI* ha mostrato livelli significativi di espressione esclusivamente nelle linee di osteosarcoma considerate; le isoforme denominate *CYYRI-1,2,3,4b* e *CYYRI-1,2,3b* sono invece sempre presenti, mentre non risulta essere espressa l'isoforma *CYYRI-1b,2,3,4*. A questo proposito è interessante notare che quest'ultima isoforma presenta un esone 1 alternativo non codificante. La presenza di esoni 1 alternativi contribuisce ad incrementare la complessità della struttura del *locus* genico. Infatti, regioni con funzioni regolatorie nel 5'-UTR diverse per isoforme differenti, potrebbero essere correlate all'esistenza di promotori alternativi. Questo potrebbe determinare una regolazione genica tessuto specifico o legata a particolari stadi dello sviluppo (Baek *et al.* 2007). Recenti studi suggeriscono che più del 50% dei geni codificanti proteine contengono promotori alternativi (Davuluri *et al.*, 2008) Sarà necessaria una analisi approfondita per capire se esistono promotori alternativi legati

a esoni 1 differenti. Ulteriori studi, inoltre, saranno necessari per verificare l'espressione delle restanti isoforme (*CYYRI-1,2,4* e *CYYRI-1,2,2bis,3,4*) nelle linee cellulari tumorali, così come per definire il *pattern* di espressione del gene in antisenso denominato *CYYRI-ASI*. Dati preliminari indicano che si ha una elevata variabilità tra le linee, in particolare l'amplicone corrispondente alla zona centrale di *CYYRI-ASI* è presente in maniera significativa solo nelle linee di osteosarcoma e in una linea di melanoma (G362), sembra poco espresso in una linea di neuroblastoma (SJNKP) e in MCF7 (adenocarcinoma della mammella). Non risulta espresso in SKNMC (neuroblastoma), HT29 (adenocarcinoma del colon) e SKMEL28 (melanoma). Sarà interessante effettuare indagini più approfondite anche attraverso l'utilizzo di tecniche quantitative quali *Real-time* PCR.

Per quanto riguarda invece le linee cellulari primarie, sono state utilizzate due linee primarie fibroblastiche, una trisomica per il cromosoma 21 (CCL-54) e una euploide (CCL-110). In questo caso sono presenti tutti i trascritti verificati in RT-PCR (l'isoforma di riferimento *CYYRI*; *CYYRI-1,2,3,4b*; *CYYRI-1,2,3b*; *CYYRI-ASI*), tranne l'isoforma *CYYRI-1,2,3b*.

In questa prima parte della tesi, è stata quindi dimostrata l'esistenza del *locus* multi-trascritto *CYYRI*, caratterizzato da 6 isoforme alternative e un lungo messaggero in antisenso rispetto a *CYYRI*. L'analisi dei diversi trascritti, delle proteine putative e del *pattern* di espressione ha permesso di caratterizzare l'organizzazione genomica del *locus* e può fungere da base per studi funzionali.

In particolare, l'identificazione di linee cellulari in cui si può osservare un apprezzabile livello di espressione dei diversi trascritti, è molto utile per mettere a punto modelli cellulari che potrebbero permettere di indagare la funzione del gene. Le linee cellulari CCL-54 e CCL-110 sono linee primarie che mostrano livelli di espressione differenti per le diverse componenti del *locus* e che potrebbero essere utilizzate per verificare se vi siano differenze significative tra una situazione di euploidia e una di triploidia del cromosoma 21 non solo a livello di espressione, ma anche a livello funzionale. Le linee cellulari di osteosarcoma, invece, dai risultati ottenuti sembrano le uniche linee non primarie che mostrano

livelli significativi di espressione dei trascritti della isoforma di riferimento *CYYR1* e potrebbero essere utilizzate come modello cellulare per indagare la funzione del gene nella tumorigenesi. Il fatto che si tratti proprio di linee di osteosarcoma, inoltre, è molto interessante in quanto dati in letteratura mostrano come in alcuni di questi tumori vi sia una sovraespressione del *pathway* molecolare di *Sonic Hedgehog* (Shh) (Hirotsu *et al.*, 2010). Da osservazioni preliminari condotte nel vertebrato *zebrafish*, *cyrr1* presenterebbe un motivo di legame comune a tutti i geni sovraespressi in risposta alla regolazione da parte di Shh e potrebbe dunque costituire un possibile bersaglio di alterata regolazione durante la tumorigenesi (Xu *et al.*, 2006; Bergeron *et al.*, 2007).

La seconda parte del presente studio, è stata dedicata alla messa a punto e all'utilizzo di approcci basati sull'impiego di un modello animale quale *zebrafish*. Si tratta di un approccio *in vivo* che permette di studiare geni di interesse, nel presente studio il gene *cyrr1*, nel contesto dell'organismo *in toto*.

In primo luogo, è emersa la necessità di effettuare una analisi sistematica per l'identificazione di un gene *housekeeping* "ideale" da utilizzare come gene di riferimento in studi di espressione sul pesce zebra. Le caratteristiche ideali di un gene perchè possa essere utilizzato come riferimento in studi di espressione genica sono un ampio *range* di espressione e un livello di espressione stabile ed elevato.

Il modello animale di *zebrafish* è largamente utilizzato per studi durante lo sviluppo embrionale, ma anche per indagare specifiche caratteristiche geniche in tessuti adulti ed è possibile che geni definiti *housekeeping* e spesso utilizzati come riferimento, possano in realtà subire variazioni di espressione in stadi embrionali o tessuti differenti.

Attraverso la tecnica di RT-PCR semi quantitativa relativa, è stato possibile dimostrare come numerosi geni ritenuti stabili come livello di espressione, subiscano variazioni notevoli in tessuti diversi, mentre altri mostrano livelli di espressione molto bassi. *Bactin2* è risultato essere il gene che rispondeva maggiormente ai criteri indicati per definire un gene *housekeeping* "ideale", dimostrando di possedere un ampio *range* di espressione in un pannello costituito da 6 stadi embrionali e 6 tessuti adulti

ed un livello di espressione relativamente stabile ed elevato. E' da notare che esiste una differenza statisticamente significativa tra stadi embrionali e tessuti adulti. Questa osservazione suggerisce che utilizzare *bactin2* all'interno di una sola categoria potrebbe aumentare la sua validità come gene di riferimento.

La procedura più solida e che meglio risponde al problema della variabilità intrinseca dei geni consiste nella validazione degli esperimenti con più geni di riferimento. Tra i geni analizzati, i più stabili dopo *bactin2* sono risultati essere *efla* e *slc25a5* (Casadei *et al.*, 2011).

La definizione di un insieme di geni di riferimento utile in studi di espressione condotti su stadi embrionali o tessuti di pesce zebra è stata molto utile per le successive analisi effettuate allo scopo di ottenere informazioni sull'ortologo del gene *CYYR1* e definirne gli aspetti funzionali.

Il secondo obiettivo, infatti, è stato quello di mettere a punto metodiche per studi di modulazione genica, quali la sovraespressione genica e la realizzazione di organismi *knockdown*. A queste metodiche utili per indagare la funzione del gene *cyrr1*, è stata affiancata la *Real-time* PCR per verificare l'espressione del gene e una sua possibile modulazione, in particolare durante lo sviluppo embrionale.

Il gene *cyrr1* è stato identificato e clonato attraverso analisi *in silico* e RT-PCR a partire da *pool* di embrioni di pesce zebra. E' localizzato in una regione del genoma che mostra sintenia tra il genoma umano e quello di *zebrafish*. La proteina putativa mostra il 58% di identità con quella umana e il 57% con quella di topo, confermando che si tratta di una famiglia proteica conservata. In particolare, vi sono tre domini proteici conservati: il tratto transmembrana, tre motivi PPPY nella regione C-terminale e un dominio ricco in cisteina e tirosina (CCSYAY). Gli esperimenti di localizzazione cellulare hanno confermato che la proteina codificata da *cyrr1* si localizza a livello della membrana plasmatica, inoltre i motivi PPPY sono particolarmente interessanti in quanto sembrano essere coinvolti nelle interazioni tra membrana plasmatica e citoscheletro (Pereboev *et al.*, 2001; Ilsley *et al.*, 2002). Infine è da notare che la perdita di questo tipo di interazioni determina l'alterazione di meccanismi

fondamentali durante lo sviluppo quali adesione cellulare, migrazione, proliferazione e differenziamento (James *et al.*, 2000).

Al fine di comprendere il ruolo di *cyrr1* durante lo sviluppo embrionale di *zebrafish*, sono stati effettuati esperimenti di inibizione e sovraespressione della proteina. In entrambi i casi sono state osservate alterazioni morfologiche che hanno dimostrato come *cyrr1* possa essere coinvolto nei meccanismi che permettono uno sviluppo normale e completo dell'asse corporeo del pesce. In particolare, *cyrr1* sembra avere un ruolo principalmente durante i primi stadi di sviluppo embrionale, contribuendo ai movimenti cellulari durante l'epibolia.

La sovraespressione del gene *cyrr1* ha determinato l'alterazione dei movimenti epibolici, con i margini della gastrula che si avvicinano a "strozzare" il tuorlo, portandolo all'esterno del lume della gastrula.

L'inibizione temporanea dell'espressione genica è stata realizzata con la microiniezione di oligonucleotidi morfolinici antisenso. Questa tipologia di antisenso di nuova generazione ha superato, almeno in buona parte, i problemi degli oligo di prima generazione. I nuovi oligo, infatti, hanno una specificità maggiore, non hanno effetti tossici, ma soprattutto agiscono grazie ad un meccanismo d'azione diverso che consiste nel blocco della traduzione e non più nella degradazione dell'ibrido DNA-RNA a mezzo dell'RNasiH. Queste caratteristiche ne fanno uno strumento molto utile ed oggi ampiamente impiegato nelle analisi funzionali *in vivo*.

Anche le alterazioni derivanti dalla microiniezione di *cyrr1* – MO osservate nei morfanti a 24-72 ore sono probabilmente da ascrivere all'alterazione della fase dell'epibolia, con movimenti anomali dei foglietti embrionali, che si manifestano in maniera visibile solo più tardi con difetti riscontrabili a livello dello sviluppo ed organizzazione del piano corporeo, nonché delle estremità cefalica e caudale.

Saranno necessari ulteriori studi per comprendere il meccanismo attraverso cui agisce la proteina codificata dal gene *cyrr1* e se vi siano interazioni specifiche con altri prodotti proteici o sequenze nucleotidiche. Questi risultati, comunque, permettono di ipotizzare ragionevolmente un coinvolgimento di *cyrr1* nei movimenti cellulari fondamentali durante l'epibolia, movimenti che collaborano al raggiungimento di un corretto

sviluppo del piano e degli assi corporei.

Dati precedenti ottenuti tramite la tecnica dell'ibridazione *in situ*, avevano dimostrato una espressione ubiquitaria del gene *cypr1*, ma i dati emersi dagli esperimenti funzionali hanno suggerito la necessità di indagarne l'andamento di espressione durante lo sviluppo embrionale e in particolare durante la gastrulazione, momento caratteristico dello sviluppo dove avvengono i movimenti epibolici.

Tale caratterizzazione è stata condotta tramite reazione a catena della polimerasi in tempo reale dopo retrotrascrizione (*Real-time* RT-PCR), in accordo con le linee guida recentemente pubblicate per uniformare ed indirizzare i ricercatori ad un utilizzo proprio di questa tecnica (Bustin *et al.*, 2009; Bustin *et al.*, 2010).

Innanzitutto, è stato necessario affrontare la messa a punto delle condizioni migliori per valutare una espressione quantitativa relativa del trascritto genico di interesse, quindi elementi critici quali estrazione e quantificazione dell'RNA degli stadi prescelti; verifica e scelta della concentrazione più opportuna dello stampo (cDNA); progettazione dei primer e verifica della loro specificità (curva di *melting*); scelta del metodo di normalizzazione (ovvero scelta dei geni di riferimento); curva standard e calcolo della efficienza per ogni coppia di primer.

In particolare, la scelta dei geni di riferimento deve essere fatta in modo opportuno ed essere basata su dati empirici il più possibile aderenti alle condizioni sperimentali studiate considerando che l'uso di un singolo gene di riferimento può portare ad errori nella stima quantitativa relativa di ordini che vanno da 3 a 6 volte (Vandesompele *et al.*, 2002). In base ai risultati dello studio precedentemente descritto sulla valutazione di geni *housekeeping* ideali per studi di espressione in *zebrafish*, è stato deciso di utilizzare come geni di riferimento *bactin2*, *ef1a* ed *slc25a5*.

I risultati ottenuti indicano un chiaro andamento bimodale del prodotto di trascrizione genica attraverso gli stadi e i tempi di sviluppo da noi osservati. Questi dati (vedi Figura 29), sono stati ottenuti scegliendo la modalità di analisi "Espressione genica" del programma CFX Manager<sup>TM</sup> Bio-Rad, che calcola direttamente le quantità relative (RQ) del gene di interesse riportandole su un grafico. Ulteriori analisi statistiche potranno

essere condotte sui dati grezzi dopo ulteriori verifiche sperimentali indipendenti per ridurre gli effetti della variabilità biologica e tecnica. I risultati sono stati comunque già confermati da un secondo esperimento indipendente (dati non mostrati). Il dato di espressione a 5 giorni risulta suscettibile di maggiore variabilità e dovrà essere vagliato con maggiore attenzione in future caratterizzazioni queste potranno anche essere utili per conferme quantitative più precisione rispetto alla espressione nei tessuti adulti.

Questi dati sono in linea con quelli ottenuti dagli studi funzionali sul ruolo di questo gene in quanto si può notare la presenza di una quantità maggiore di trascritto in stadi direttamente connessi ai movimenti morfogenetici epibolici, che si innescano proprio a partire dallo stadio di cupola (*dome*) e che sono responsabili della organizzazione spaziale delle cellule che si differenzieranno durante la fase di gastrulazione a definire gli assi corporei. Durante l'intera gastrulazione e somitogenesi, come dimostrato dai risultati ottenuti, la quantità relativa del trascritto si mantiene alta, con una leggera flessione a 48 hpf.

Il coinvolgimento di *cyrr1* durante lo sviluppo embrionale può essere ulteriormente indagato andando a verificare possibili *pathway* di regolazione genica. Dati in letteratura dimostrano che *cyrr1* presenta un motivo di legame comune a tutti i geni sovraespressi in risposta alla regolazione da parte di Shh (Xu *et al.*, 2006). Sarà quindi interessante mettere a punto esperimenti volti a verificare se *cyrr1* possa far realmente parte della cascata di segnale di Shh.

Inoltre, il modello animale di *zebrafish* può essere utilmente utilizzato per la caratterizzazione di sequenze conservate evolutivamente e correlate a funzioni regolatorie (CNE, *conserved non coding element*). Le CNEs ad oggi caratterizzate sono sequenze correlate a geni coinvolti durante lo sviluppo embrionale e formano una famiglia di elementi regolatori che trasmettono le informazioni ereditarie essenziali per lo sviluppo embrionale; tale famiglia è conservata nei vertebrati (Elgar *et al.*, 2008).

I dati ottenuti in questo studio hanno dimostrato, per la prima volta, una conservazione di tipo funzionale di questi elementi dai vertebrati agli

invertebrati.

Le CNEs, identificate in primo luogo con metodi bioinformatici in *Ciona intestinalis*, sono state clonate *in vitro* tramite RT-PCR e gli ampliconi ottenuti sono stati utilizzati in esperimenti di microiniezione in embrioni di pesce zebra, appunto per verificare se tali sequenze mantengono una funzione regolatoria anche nei vertebrati.

Col progredire della definizione dei *pathway* molecolari in esame potrà essere interessante verificare se all'interno di queste cascate di segnale, sono coinvolti anche elementi regolatori di questo tipo, altamente conservati e correlati a una funzione biologica essenziale quale l'insieme dei meccanismi dello sviluppo embrionale; in particolare, le indagini sul gene *cyrr1* fanno ipotizzare un suo coinvolgimento a livello di sviluppo embrionale, quindi potrà essere interessante verificare se *cyrr1* stesso o geni coinvolti nel suo *pathway* molecolare siano regolati da CNEs.

Lo studio sul gene *CYYRI* e sul suo ortologo in *zebrafish* rientra in un'indagine a più ampio raggio sui geni del Cromosoma 21, coinvolti nella patogenesi della Sindrome di Down, ma anche nella tumorigenesi e nello sviluppo embrionale.

L'analisi dell'espressione genica fornisce numerose informazioni che possono contribuire ad associare un gene o un insieme di geni ad un particolare meccanismo biologico normale o patologico.

La meta-analisi di studi indipendenti potrebbe fornire elementi utili che mettano in evidenza regioni cromosomiche caratterizzate da una consistente e stabile sovraespressione o sottoespressione genica.

Sulla base di queste considerazioni, in questo studio è stato messo a punto un *software* in grado di produrre mappe cromosomiche di espressione a partire da dati di analisi genica globale (TRAM, Transcriptome Mapper) (Lenzi *et al.*, 2011). Si tratta di un programma con caratteristiche innovative quali la possibilità di sfruttare dati organizzati in database interconnessi tra loro, due modalità di analisi (Map e Cluster) per ricercare regioni genomiche sovra/sottoesprese, a partire da qualsiasi fonte di profili di espressione; inoltre l'utente è sempre in grado di visualizzare i valori numerici di espressione visualizzabili al fianco dello schema grafico, infine, non è necessario fornire alcuna coordinata

genomica in quanto la localizzazione dei geni è già caricata nel database.

Si tratta di un modello di analisi “esplorativo” che potrebbe essere utilmente impiegato per approfondire le conoscenze riguardanti l’espressione genica durante lo sviluppo embrionale di *zebrafish* e trovare ulteriori dati da affiancare a quelli ottenuti mediante la *Real-time* PCR.

Allo scopo di ampliare la ricerca sull’espressione e la caratterizzazione dei geni del Cromosoma 21, potrebbe infine essere utilizzato per meta-analisi di *microarray* per confrontare l’espressione genica di campioni trisomici ed euploidi.

Concludendo, il presente progetto di ricerca sottolinea la necessità di proseguire l’analisi genomica della sequenza del cromosoma 21 umano per una completa determinazione del suo contenuto genico. Inoltre si pone a più ampio raggio come studio avanzato del genoma umano, con un particolare riferimento ai geni del cromosoma 21, nel tentativo di portare contributi conoscitivi eventualmente utili a chiarire la complessa, ed ancora ampiamente sconosciuta, patogenesi molecolare della Sindrome di Down correlata con la trisomia 21.

In particolare, il lavoro si è articolato in tre parti: la caratterizzazione del *locus* genico umano di *CYYR1* (*cysteine and tyrosine-rich protein 1 gene*) localizzato sul cromosoma 21; l’utilizzo del modello animale di *zebrafish* per lo studio funzionale dell’omologo di *CYYR1* e parallelamente per lo studio di sequenze non codificanti altamente conservate coinvolte nella regolazione genica ed infine lo sviluppo di un *software* in grado di produrre mappe cromosomiche di espressione, al fine di inserire lo studio del gene *CYYR1* in un’analisi a più ampio raggio sui geni del Cromosoma 21.

I dati ottenuti aprono alla possibilità di proseguire nella ricerca approfondendo diversi aspetti di ognuna delle parti affrontate in questa tesi.

Per quanto riguarda la caratterizzazione del *locus* genico umano di *CYYR1*, è stata messa a punto la tecnica di *Real-time* PCR per una analisi quantitativa assoluta dei diversi trascritti, allo scopo di verificare

l'andamento dell'espressione in diversi tessuti e modelli cellulari. Sarà necessario approfondire la possibilità di utilizzare le linee cellulari CCL 54 e 110 e le linee di osteosarcoma U2OS e SaOS come modello per lo studio delle funzioni di *CYYRI* e delle sue possibili interazioni con altre proteine o sequenze. In particolare, le CCL potrebbero essere considerate un buon modello cellulare per paragonare la condizione euploide con quella di trisomia 21, approfondendo l'analisi quantitativa reciproca delle diverse isoforme di senso ed antiseno trascritte dal *locus CYYRI* alla ricerca di possibili correlazioni utili a comprendere la regolazione genica del sito alla luce del diverso dosaggio genico.

Tra le diverse linee studiate, invece, le linee di osteosarcoma si sono finora dimostrate le uniche con livelli significativi dei diversi componenti del *locus CYYRI*. Questo potrebbe costituire un ottimo modello cellulare di partenza per la realizzazione di esperimenti tesi ad indagare possibili *pathway* regolativi potenzialmente coinvolti nella tumorigenesi, nonché per la messa a punto di esperimenti di *gain e loss of function* utili per esplorare la funzione cellulare svolta dal gene *CYYRI*.

Il modello animale di *zebrafish*, invece, potrà essere utilizzato per effettuare ulteriori esperimenti che permettano di verificare la specifica funzione di *cyyr1* durante lo sviluppo embrionale, per esempio indagando quali siano le componenti biochimiche del *pathway* molecolare di *cyyr1*.

L'aspetto della regolazione genica potrà essere approfondito sia andando a verificare il coinvolgimento di *Shh* nella regolazione di *cyyr1*, sia effettuando indagini sulla funzione delle CNEs durante lo sviluppo embrionale.

Infine, si intende utilizzare il *software* TRAM per effettuare meta-analisi di dati di espressione genica disponibili in rete ed integrare i risultati sperimentali con una analisi a più ampio raggio che permette di ottenere una visione dell'andamento complessivo dell'espressione genica in determinate condizioni di interesse, quali per esempio la presenza della trisomia 21.

## **BIBLIOGRAFIA**

Aït Yahya-Graison E, Aubert J, Dauphinot L, Rivals I, Prieur M, Golfier G, Rossier J, Personnaz L, Creau N, Bléhaut H, Robin S, Delabar JM, Potier MC. (2007) Classification of human chromosome 21 gene-expression variations in Down syndrome: impact on disease phenotypes. *Am J Hum Genet.*; 81(3):475-91.

Allen EG, Freeman SB, Druschel C, Hobbs CA, O’Leary LA, Romitti PA, Royle MH, Torfs CP & Sherman SL (2009) Maternal age and risk for trisomy 21 assessed by the origin of chromosome nondisjunction: a report from the Atlanta and National Down Syndrome Projects. [Hum Genet.](#); 125(1):41-52.

Altschul SF, Madden TL, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein data base search programs. *Nucleic Acids Research.*; 25(17): 3389-3402.

Altug-Teber O, Bonin M, Walter M, Mau-Holzmann UA, Dufke A, Stappert H, Tekesin I, Heilbronner H, Nieselt K, Riess O (2007) Specific transcriptional changes in human fetuses with autosomal trisomies. *Cytogenet Genome Res* 119:171–184

Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S. (2004) Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nat Rev Genet.*, 5(10):725-38. Review.

Baek D, Davis C, Ewing B, Gordon D, Green P (2007) Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. *Genome Res* 17: 145-155.

Barrett T, Troup DB, et al. (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res*, 37: D885-890.

Barut BA, Zon LI. (2000) Realizing the potential of zebrafish as a model for human disease. *Physiol Genomics.* ; 2(2): 49-51

Bejerano G. et al. , (2004). Ultraconserved elements in the human genome. *Science* 304, 1321-1325

Bergeron SA, Milla LA, Villegas R, Shen MC, Burgess SM, Allende ML, Karlstrom RO, Palma V. (2008) Expression profiling identifies novel Hh/Gli-regulated genes in developing zebrafish embryos. *Genomics.*; 91(2):165-77.

BLASTN ([www.ncbi.nlm.nih.gov/blast](http://www.ncbi.nlm.nih.gov/blast))

Blencowe BJ (2006) Alternative splicing: new insights from global analyses. *Cell* 126: 37-47

Boguski MS, Lowe TM, Tolstoshev CM. (1993) dbEST--data base for "expressed sequence tags". *Nature Genetics.*; 4(4): 332-333. PMID: 8401577

Bustin S.A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* 25: 169-193.

Bustin S.A., Benes V., Garson J.A., Hellemans J., Huggett J., Kubista M., Mueller R., Nolan T., Pfaffl M.W., Shipley G.L., et al (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 55: 611-622.

Carninci P, Yasuda J, Hayashizaki Y (2008) Multifaceted mammalian transcriptome. *Curr Opin Cell Biol* 20: 274-280.

Carson MB, Langlois R, Lu H (2008) Mining knowledge for the methylationstatus of CpG islands using alternating decision trees. Conf Proc IEEE Eng Med Biol Soc 2008:3787–3790

Casadei R, Pelleri MC, Vitale L, Facchin F, Lenzi L, Canaider S, Strippoli P, Frabetti F. (2011) Identification of housekeeping genes suitable for gene expression analysis in the zebrafish. *Gene Expr Patterns*.

CGOW [Comparative Genome Organization of Vertebrates: First International Workshop] (1996) *Mamm Genome* 7:717-734.

ClustalW 1.83 (<http://www.ebi.ac.uk/clustalw>)

Conti A, Fabbrini F, D’Agostino P, Negri R, Greco D, Genesio R, D’Armiento M, Olla C, Paladini D, Zannini M, Nitsch L (2007) Altered expression of mitochondrial and extracellular matrix genes in the heart of human fetuses with chromosome 21 trisomy. *BMC Genomics* 8:268

Corey DR, Abrams JM. (2001) Morpholino antisense oligonucleotides: tools for investigating vertebrate development. *Genome Biol*; 2(5):REVIEWS1015.

Cummings M.R. (2002) *Human Heredity: Principles and Issues*, EdiSES, Napoli

Davis LG, Kuehl WM, and Battey JF. (1994) *Basic methods in molecular Biology*. II ed. Appleton & Lange, Norwalk.

Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH (2008) The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet* 24: 167-177.

Deutsch S, Lyle R, Dermitzakis ET, Attar H, Subrahmanyam L, Gehrig C, Parand L, Gagnebin M, Rougemont J, Jongeneel CV, Antonarakis SE.

(2005) Gene expression variation and expression quantitative trait mapping of human chromosome 21 genes. *Hum Mol Genet.*;14(23):3741-9.

Elgar G, Vavouri T. (2008) Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet.*; 24(7):344-52.

El-Maarri O, Kareta MS, Mikeska T, Becker T, Diaz-Lacava A, Junen J, Nü sgen N, Behne F, Wienker T, Waha A, Oldenburg J, Chédin F (2009) A systematic search for DNA methyltransferase polymorphisms reveals a rare DNMT3L variant associated with subtelomeric hypomethylation. *Hum Mol Genet* 18(10): 1755–1768

Engels WR. (1993) Contributing software to the internet: the Amplify program. *Trends in Biochemical Science*; 18(11): 448-450. PMID: 8291093

Facchin F, Canaider S, Vitale L, Frabetti F, Griffoni C, Lenzi L, Casadei R, Strippoli P. (2008) Identification and analysis of human RCAN3 (DSCR1L2) mRNA and protein isoforms. *Gene.*; 407(1-2):159-68.

Figaro S, Scrima N, Buckingham RH, Heurgue-Hamard V (2008) HemK2 protein, encoded on human chromosome 21, methylates translation termination factor eRF1. *FEBS Lett* 582: 2352–2356

Fonatsch C. (2010) The role of chromosome 21 in hematology and oncology. *Genes Chromosomes Cancer*, 49(6): 497-508.

Ghosh S, Feingold E & Dey SK (2009) Etiology of Down syndrome: evidence for consistent association among altered meiotic recombination, nondisjunction, and maternal age across populations. *American Journal of Medical Genetics. Part A* 149A 1415–1420.

Gilbert SF., (2000) *Developmental Biology*. 6th edition. Sunderland (MA): Sinauer Associates

Gittenberger-de Groot AC, Bartram U, Oosthoek PW, Bartelings MM, Hogers B, Poelmann RE, Jongewaard IN, Klewer SE. (2003) Collagen type VI expression during cardiac development and in human fetuses with trisomy 21. *Anat Rec A Discov Mol Cell Evol Biol.*, 275(2):1109-16.

Guil S, Esteller M (2009) DNA methylomes, histone codes and miRNAs: tying it all together. *Int J Biochem Cell Biol* 41:87–95

Hardy J, Allsop D. (1991) Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends in Pharmacological Sciences*, 12: 383–388.

Harris LJW, Swatton JE, Wengenroth M, Wayland M, Lockstone H, Holland A, Faull RLM, Lilley KS, Bahn S (2007) Differences in protein profiles in schizophrenia prefrontal cortex compared to other major brain disorders. *Clin Schizophrenia Relat Psychoses* 1: 73–91

Hasle H. (2001) Pattern of malignant disorders in individuals with Down's Syndrome *Lancet Oncol.*, 2(7):429-36.

Hassold T., Abruazzo M., Adkins K., Griffin D., Merrill M., Millie E., Saker D., Shen J. and Zaragoza M. (1996) Human aneuploidy: incidence, origin, and etiology. *Environ. Mol. Mutagen.*, 28, 167–175.

Hattori M, Fujiyama A, et al. (2000) The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium. *Nature.*; 405(6784): 311-319.

Heasman J, Wessely O, Langland R, Craig EJ, Kessler DS. (2001) Vegetal localization of maternal mRNAs is disrupted by VegT depletion. *Dev Biol.*; 240(2):377-86.

Helguera P, Pelsman A, Pigino G, Wolvetang E, Head E, Busciglio J (2005) *ets-2* promotes the activation of a mitochondrial death pathway in Down's syndrome neurons. *J Neurosci* 25:2295–2303.

Henrichsen CN, Chaignat E, Reymond A (2009b) Copy number variants, diseases and gene expression. *Hum Mol Genet* 18(R1): R1–R8.

Henrichsen CN, Vinckenbosch N, Zöllner S, Chaignat E, Pradervand S, Schütz F, Ruedi M, Kaessmann H, Reymond A (2009a) Segmental copy number variation shapes tissue transcriptomes. *Nat Genet* 41: 424–429.

Hirotsu M, Setoguchi T, Sasaki H, Matsunoshita Y, Gao H, Nagao H, Kunigou O, Komiya S. (2010) Smoothed as a new therapeutic target for human osteosarcoma. *Mol Cancer*.; 9:5.

Hook EB, Cross PK, Schreinemachers DM (1983) Chromosomal abnormality rates at amniocentesis and in live-born infants. *JAMA* 249: 2034-2038.

Hu YG, Hirasawa R, Hu JL, Hata K, Li CL, Jin Y, Chen T, Li E, Rigolet M, Viegas-Péquignot E, Sasaki H, Xu GL (2008) Regulation of DNA methylation activity through *Dnmt3L* promoter methylation by *Dnmt3* enzymes in embryonic development. *Hum Mol Genet* 17: 2654–2664.

Hultén MA, Patel S, Jonasson J, Iwarsson E (2010) On the origin of the maternal age effect in trisomy 21 Down syndrome: the Oocyte Mosaicism Selection model. *Reproduction*.139(1): 1-9.

Hunt PA & Hassold TJ (2008) Human female meiosis: what makes a good egg go bad? *Trends in Genetics* 24 86–93.

Ilseley JL, Sudol M, Winder SJ. (2002) The WW domain: linking cell signalling to the membrane cytoskeleton. *Cell Signal.* ;14(3): 183-9. Review.

James, M., Nuttall, A., Ilsey, J.L., Ottersbach, K., Tinsley, J.M., Sudol, M., Winder, S.J., (2000) Adhesion-dependent tyrosine phosphorylation of beta-dystroglycan regulates its interaction with utrophin. *J. Cell Science* 113, 1717-1726.

Jones KT (2008) Meiosis in oocytes: predisposition to aneuploidy and its increased incidence with age. *Human Reproduction Update* 14:143–158.

Kim H, Klein R, Majewski J, Ott J (2004) Estimating rates of alternative splicing in mammals and invertebrates. *Nat Genet* 36: 915-916; author reply 916-917.

Koremborg JR, Chen XN, Shipper R, Sun Z, Gonsky R, et al. (1994) DS phenotypes: the consequences of chromosomal imbalance. *Proc Natl Acad Sci USA* 91: 4997-5001.

Kuhn DE, Nuovo GJ, Martin MM, Malana GE, Pleister AP, Jiang J, Schmittgen TD, Terry AV, Gardiner K, Head E, Feldman DS, Elton TS (2008) Human chromosome 21-derived miRNAs are overexpressed in Down syndrome brains and hearts. *Biochem Biophys Res Commun* 370: 473–477.

Lange B. (2000) The management of neoplastic disorders of haematopoiesis in children with Down's syndrome. *Br J Haematol.*, 110(3):512-24.

Lejeune J, Gautier e M, Turpin, R (1959;) Etude des chromosomes somatique de neuf enfants mongoliens. *C.R. Acad Sci* 248: 1721-1722.

Lejeune J, Gautier M, Turpin R. (1959) Etude des chromosomes somatique des neufs enfants mongoliens. CR Academical Science. Paris.; 248, pp: 1721-1722. Paris, France.

Lenzi L, Facchin F, Piva F, Giulietti M, Pelleri MC, Frabetti F, Vitale L, Casadei R, Canaider S, Bortoluzzi S, Coppe A, Danieli GA, Principato G, Ferrari S, Strippoli P. (2011) TRAM (Transcriptome Mapper): database-driven creation and analysis of transcriptome maps from multiple sources. BMC Genomics.; 12:121.

Lockstone HE, Harris LW, Swatton JE, Wayland MT, Holland AJ, Bahn S (2007) Gene expression proWling in the adult Down syndrome brain. Genomics 90: 647–660.

Lott IT, Head E (2005) Alzheimer disease and Down syndrome: factors in pathogenesis. Neurobiol Aging 26: 383–389.

Lott IT, Head E, Doran E, Busciglio J (2006) Beta-amyloid, oxidative stress and down syndrome. Curr Alzheimer Res 3: 521–528.

Lubec G (ed) (2001) Protein expression in Down syndrome brain. Springer, Austria.

Lyle R, Bena F, Gagos S, Gehrig C, Lopez G et al. (2008) Genotype-phenotype correlations in DS identified by array CGH in 30 cases of partila trisomy and partila monosomy chromosome 21. Eur J of Hum Gen, pp1-13.

Lyle R, Dermitzakis E, Reymond A, Deutsh S (2004) Chromosome 21 and DS:from genomics to pathophysiology. Nar Rev Genet 5:725-738.

Mailhes JB (2008) Faulty spindle checkpoint and cohesion protein activities predispose oocytes to premature chromosome separation and aneuploidy. Environmental and Molecular Mutagenesis 49 642–658.

Martin RH (2008) Meiotic errors in human oogenesis and spermatogenesis. *Reproductive Biomedicine Online* 16 523–531.

Miko, I. (2008) Mitosis, meiosis, and inheritance. *Nature Education*.

Oliver TR, Feingold E, Yu K, Cheung V, Tinker S, Yadav-Shah M, Masse N & Sherman SL (2008) New insights into human nondisjunction of chromosome 21 in oocytes. *PLoS Genetics* 4 e1000033.

Olson LE, Roper RJ, Sengstaken CL, Peterson EA, Aquino V, Galdzicki Z, Siarey R, Pletnikov M, Moran TH, Reeves RH. (2007) Trisomy for the Down syndrome 'critical region' is necessary but not sufficient for brain phenotypes of trisomic mice. *Hum Mol Genet.*; 16(7):774-82.

Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin S-P, Allis CD, Cheng X, Bestor TH (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448: 714–717.

Patterson D, Costa ACS (2005) Down syndrome and genetics—a case of linked histories. *Nat Rev Genet* 6: 137–147.

Patterson D., (2009) Molecular genetic analysis of Down syndrome *Hum Genet*, 126:195–214.

Penrose LS (1933) The relative effect of paternal and maternal age in mongolism. *Journal of Genetics* 27 219–224.

Penrose LS (1934) The relative aetiological importance of birth order and maternal age in mongolism. *Proceedings of the Royal Society (Biology)* 115 431–450.

Pereboev AV, Ahmed N, thi Man N, Morris GE. (2001) Epitopes in the interacting regions of beta-dystroglycan (PPxY motif) and dystrophin (WW domain). *Biochim Biophys Acta*;1527(1-2):54-60.

Pfaffl M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29: e45.

Prandini P, Deutsch S, Lyle R, Gagnebin M, Delucinge Vivier C, Delorenzi M, Gehrig C, Descombes P, Sherman S, Dagna Bricarelli F, Baldo C, Novelli A, Dallapiccola B, Antonarakis SE. (2007) Natural gene-expression variation in Down syndrome modulates the outcome of gene-dosage imbalance. *Am J Hum Genet.*; 81(2):252-63.

Pueschel SM, Rynders JE. (1982). *DS: Advances in Biomedicine and the Behavioral Sciences*. Cambridge: Ware Press. 133 p.

Reymond A, Camargo AA, Deutsch S, Stevenson BJ, Parmigiani RB, Ucla C, Bettoni F, Rossier C, Lyle R, Guipponi M, de Souza S, Iseli C, Jongeneel CV, Bucher P, Simpson AJ, Antonarakis SE. (2002) Nineteen additional unpredicted transcripts from human chromosome 21. *Genomics*; 79(6): 824-32.

Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerrière A, Vital A, Dumanchin C, Feuillette S, Brice A, Vercelletto M, Dubas F, Frebourg T, Campion D (2006) APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet* 38: 24–26.

Ryoo SR, Jeong HK, Radnaabazar C, Yoo JJ, Cho HJ, Lee HW, et al. (2007) DYRK1A-mediated hyperphosphorylation of Tau. A functional link between Down syndrome and Alzheimer disease. *Journal of Biological Chemistry*, 282: 34850–34857.

Sandelin et al., (2004) Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics* 5, 99.

Scriver CR, Beaudet AL, Sly WS, Valle D, (2001) *The metabolic and molecular bases of inherited disease* 8th ed New York: McGraw Hill. Pp1223-1256.

Sharrocks AD. (1994) in *The design of primer for PCR. PCR Technology - Current Innovations*. Chapter 2, pp: 5-11. CRC Press, Boca Raton.

Slegers K, Brouwers N, Gijssels I, Theuns J, Goossens D, Wauters J, Del-Favero J, Cruts M, van Duijn CM, Van Broeckhoven C (2006) APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain* 129:2977–2983.

Sommer CA, Pavarino-Bertelli EC, Goloni-Bertollo EM, Henrique-Silva F (2008) Identification of dysregulated genes in lymphocytes from children with Down syndrome. *Genome* 51:19–29.

Strachan T, Read AP. (1999) *Human Molecular Genetics*. 2nd edition. New York: Wiley-Liss.

Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavaré S, Deloukas P, Hurles ME, Dermitzakis ET (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315: 848–853.

Strefford JC, van Delft FW, Robinson HM, Worley H, Yiannikouris O, Selzer R, Richmond T, Hann I, Bellotti T, Raghavan M, Young BD, Saha V, Harrison CJ. (2006) Complex genomic alterations and gene expression in acute lymphoblastic leukemia with intrachromosomal amplification of chromosome 21. *Proc Natl Acad Sci U S A.*; 103(21):8167-72.

Sultan M, Piccini I, Balzereit D, Herwig R, Saran NG, Lehrach H, Reeves RH, Yaspo ML. (2007) Gene expression variation in Down's syndrome mice allows prioritization of candidate genes. *Genome Biol.*; 8(5):R91.

Takashima S, Takehashi M, Lee J, Chuma S, Okano M, Hata K, Suetake I, Nakatsuji N, Miyoshi H, Tajima S, Tanaka Y, Toyokuni S, Sasaki H, Kanatsu-Shinohara M, Shinohara T (2009) Abnormal DNA methyltransferase expression in mouse germline stem cells results in spermatogenic defects. *Biol Reprod.*; 81(1):155-64.

Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A., Speleman F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*; 3(7):RESEARCH0034

Victor M., Ropper AH. (2001) *Adams and Victor's Principles of Neurology*. Vol. 7. New York: McGraw-Hill.

Vis JC, Duffels MG, Winter MM, Weijerman ME, Cobben JM, Huisman SA, Mulder BJ. (2009) Down syndrome: a cardiovascular perspective. *J Intellect Disabil Res.*, 53(5):419-25.

Vitale L, Casadei R, Canaider S, Lenzi L, Strippoli P, D'Addabbo P, Giannone S, Carinci P, Zannotti M. (2002) Cysteine and tyrosine-rich 1 (CYYR1), a novel unpredicted gene on human chromosome 21 (21q21.2), encodes a cysteine and tyrosine-rich protein and defines a new family of highly conserved vertebrate-specific genes. *Gene*; 290(1-2): 141-51.

Vitale L, Frabetti F, Huntsman SA, Canaider S, Casadei R, Lenzi L, Facchin F, Carinci P, Zannotti M, Coppola D, Strippoli P. (2007) Sequence, "subtle" alternative splicing and expression of the CYYR1 (cysteine/tyrosine-rich 1) mRNA in human neuroendocrine tumors. *BMC Cancer.*; 7:66.

Vitale L, Lenzi L, Huntsman SA, Canaider S, Frabetti F, et al. (2006) Differential expression of alternatively spliced mRNA forms of the insulin-like growth factor 1 receptor in human neuroendocrine tumors. *Oncol Rep* 15: 1249-1256.

Wegiel J, Dowjat K, Kaczmarek W, Kuchna I, Nowicki K, Frackowiak J, Mazur K, Kolecka B, Wegiel J, Silverman WP, Reisberg B, DeLeon M, Wisniewski T, Gong CX, Liu F, Adayev T, Chen-Hwang MC, Hwang YW. (2008) The role of overexpressed DYRK1A protein in the early onset of neurofibrillary degeneration in Down syndrome. *Acta Neuropathol.*, 116(4):391-407.

Westerfield M. (1995) in *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio)*. 4th ed. Univ. of Oregon Press, Eugene.

Wiseman FK, Alford KA, Tybulewicz VL, Fisher EM. (2009) Down syndrome—recent progress and future prospects. *Hum Mol Genet.*, 18:R75–83.

Xu J, Srinivas BP, Tay SY, Mak A, Yu X, Lee SG, Yang H, Govindarajan KR, Leong B, Bourque G, Mathavan S, Roy S. (2006) Genomewide expression profiling in the zebrafish embryo identifies target genes regulated by Hedgehog signaling during vertebrate development. *Genetics.*; 174(2):735-52.

Zigman WB, Devenny DA, Krinsky-McHale SJ, Jenkins EC, Urv TK, Wegiel J, Schupf N, Silverman W. (2008) Alzheimer's Disease in Adults with Down Syndrome. *Int Rev Res Ment Retard.*, 1;36:103-145.

## **TABELLE**

**Tabella 1.** Patologia mieloproliferativa transiente (TMD) e Leucemie megacarioblastiche acute (AMKL) in pazienti con Sindrome di Down (SD)

---

- 1954: prima descrizione di DS-TMD
- TMD colpisce il 5-10% dei nuovi nati con SD
- Regressione entro i primi 3 mesi di vita nel 70-80% dei casi
- ~20% di DS-TMD evolve in DS-AMKL

Caratteristiche di DS-AMKL:

- Elevato numero di megacariociti, micromegacariociti, maturazione anomala dei megacarioblasti
  - Meno del 10% di DS-AMKL insorge entro il primo anno di vita, il 50% durante il secondo
  - Il 70% dei pazienti presenta una prefase mielodisplastica (anemia, trombocitopenia)
  - Rischio di sviluppare DS-AMKL per bambini con SD entro i 5 anni aumentato di 150 volte rispetto a bambini non affetti da SD
- 

DS- TMD: Patologia mieloproliferativa transiente (*Transient Myeloproloferative Disorder*) in pazienti con Sindrome di Down (DS);  
DS-AMKL: leucemie megacarioblastiche acute in pazienti con DS

Fonte: Fonatsch, *Genes Chromosomes Cancer*, 2010

**Tabella 2.** Geni candidati per la leucemogenesi in pazienti affetti da Sindrome di Down (DS)

---

- *BACH1* (21q22, repressore della differenziazione megacariocitica): aumentata espressione in blasti di DS-AMKL
  - *SON* (21q22, omologo della famiglia MYC): sovraespressione in blasti di DS-AMKL
  - 5 microRNA (21q): probabile ruolo nella leucemogenesi
  - *RUNX1* (21q22, fattore di trascrizione): no mutazioni, no sovraespressione in blasti di DS-AMKL
  - *ETS2*, *ERG*, (21q22, fattori di trascrizione): espressione analoga a quella nei blasti non DS
- 

Fonte: Fonatsch, *Genes Chromosomes Cancer*, 2010

**Tabella 3.** Ipotesi sull'effetto dell'età materna sull'insorgenza della Sindrome di Down

<b>Ipotesi</b>	<b>Bibliografia</b>
Alterata ricombinazione meiotica / Instabilità meccanica durante il passaggio MI-AI / Segregazione anomala dei cromosomi	Penrose (1954), Henderson & Echwards (1968), Warren <i>et al.</i> (1987), Sherman <i>et al.</i> (1991), Lamb <i>et al.</i> (2005), Hunt & Hassold (2008), Mailhes (2008), Oliver <i>et al.</i> (2008), Allen <i>et al.</i> (2009)
Alterazioni del DNA centromerico	Maratou <i>et al.</i> (2000)
Ridotta coesione dei cromatidi fratelli durante il passaggio MI-AI	Watanabe & Nurse (1999), Schon <i>et al.</i> (2000), de Bruijnet <i>et al.</i> (2004), Eichenlaub-Ritter (2004), Hodges <i>et al.</i> (2005), Steuerwaid <i>et al.</i> (2005), Mailhes <i>et al.</i> (2008), Subramanian & Nickel (2008), Vogt <i>et al.</i> (2008)
Ridotta condensazione della cromatina /Instabilità meccanica durante il passaggio MI-AI	Hulten <i>et al.</i> (1990)
Alterato fuso durante AI e AII	Hawley <i>et al.</i> (1994), Battaglia <i>et al.</i> (1996), Van Blerkom <i>et al.</i> (1997), LeMaire-Adkins <i>et al.</i> (1997), Eichenlaub-Ritter (1998), Vollarcik <i>et al.</i> (1998), Steuerwald <i>et al.</i> (2001), Homer <i>et al.</i> (2005), Ma <i>et al.</i> (2005), Mailhes (2008), Vogt <i>et al.</i> (2008)
Accumulo di effetti tossici/ Ambiente/ Virus/ Mutazioni mitocondriali/ Agenti citoplasmatici/ Deficit energetici/ Deficit di proteine degli oociti	Penrose (1954), Stoller & Collmann (1965), Evans (1967), Beermann & Hansmann (1986), Van Berkorn <i>et al.</i> (1998), Jansen & de Boer (1998), Schon <i>et al.</i> (2000), Sherman <i>et al.</i> (2007)
Trisomia 21 a mosaico, modello OM5	Vig (1984), Zheng & Byers (1992), Sensi & Ricci (1993), Zheng & Byers (1996), Hulten <i>et al.</i> (2008)
Limitato numero di oociti, menopausa prematura	Eichenlaub-Ritter & Boll (1989), Warburton (1989), Gauden (1992), Kline & Levin (1992), Freeman <i>et al.</i> (2000), Kline <i>et al.</i> (2000), van Montirans <i>et al.</i> (2002), Kline <i>et al.</i> (2004), Warburton (2005)
Difetti degli ovociti dovuti a squilibri ormonali	van Montirans <i>et al.</i> (1999), Nasser <i>et al.</i> (1999), Seifer <i>et al.</i> (2007), Sherman <i>et al.</i> (2007)
Diminuita selezione verso il concepimento di feti trisomici	Erckson (1978), Aymè & Lippmahand (1982), Neuhauser & Krackov (2007)

Fonte: Hultén *et al.*, *Reproduction*, 2010

**Tabella 4.** Analisi di espressione genica in campioni SD umani e murini

<b>Bibliografia</b>	<b>Metodo</b>	<b>Sorgente campione</b>	<b>Tessuto</b>
Sommer <i>et al.</i> (2008)	SAGE	SD <i>Homo sapiens</i>	Linfociti da bambini (1-4 anni)
Chrast <i>et al.</i> (2000)	SAGE	Ts65Dn	Cervello di topi di 30 giorni
Lockstone <i>et al.</i> (2007)	<i>Microarray</i>	SD <i>Homo sapiens</i>	Cervello da soggetti adulti
Altug-Teber <i>et al.</i> (2007)	<i>Microarray</i>	SD <i>Homo sapiens</i>	Feti di 13-16 settimane
Conti <i>et al.</i> (2007)	<i>Microarray</i>	SD <i>Homo sapiens</i>	Cuore da feti di 18-22 settimane
Prandini <i>et al.</i> (2007)	<i>Real-time PCR</i>	SD <i>Homo sapiens</i>	Linfoblasti, fibroblasti da feti e soggetti fino ai 25 anni
Ait-Yahya-Graison <i>et al.</i> (2007)	<i>Microarray, Real-time PCR</i>	SD <i>Homo sapiens</i>	Linfoblasti
Sultan <i>et al.</i> (2007)	<i>Real-time PCR</i>	Ts65Dn	Cervelletto, mesencefalo, corteccia cerebrale

Fonte: Patterson, *Hum. Genet.*, 2009

**Tabella 5.** Caratteristiche principali degli elementi non codificanti conservati dei vertebrati (Elgar *et al.*, 2008)

Acronimo	Breve descrizione	Metodo di identificazione e caratteristiche	Bibliografia
MCS	Multi-species conserved sequences	<p><b>Sequenze genomiche utilizzate:</b> 1.8 Mb del Crom.7 umano e regioni ortologhe di 12 specie di vertebrati</p> <p><b>Metodo di identificazione:</b> Viene assegnato un punteggio alla conservazione delle sequenze lungo l'evoluzione. Il livello soglia è fissato in modo che il 5% delle sequenze umane potrebbero rientrare nell'insieme delle MCS. Le MCSs possono sovrapporsi a sequenze geniche</p> <p><b>Caratteristiche principali:</b> 1194 MCSs identificate nel menoma umano con una lunghezza media di 58 bp. Il 68% non si sovrappone a equenze geniche. Solo due MCSs umane trovano un allineamento con sequenze genomiche di pesce.</p>	<p>Margulies, EH <i>et al</i> 2003 <i>Genome Res</i></p> <p>Thomas JW <i>et al</i> 2003 <i>Nature</i></p>
CNC	Conserved noncoding sequences	<p><b>Sequenze genomiche utilizzate:</b> Genomi umano e murino</p> <p><b>Metodo di identificazione:</b> Allineanti locali per identificare regioni omologhe, seguiti da allineamenti globali eseguiti a partire dalle regioni locali di similarità. Selezione delle regioni col 70% di identità lungo 100 bp Regioni conservate che non si sovrappongono a sequenze geniche.</p>	<p>Courinne, O <i>et al</i> 2003 <i>Genome Res</i></p>
UCE	Ultraconserved elements	<p><b>Sequenze genomiche utilizzate:</b> Genomi di uomo, ratto e topo.</p> <p><b>Metodo di identificazione:</b> Allineamento dei genomi di uomo, ratto e topo. Selezione di regioni con 100% di identità con una lunghezza <math>\geq 200</math> bp</p> <p><b>Caratteristiche principali:</b> Il menoma umano contiene 481 UCEs con una lunghezza media di 262 bp. Il 23% si sovrappone a sequenze geniche, il 53% non presenta evidenze di trascrizione. Le Uces non codificanti si trovano spesso in regioni povere di geni e vicino a geni coinvolti nello sviluppo. La maggior parte sono conservate nel menoma di pollo, numerose sono conservate in quello di Fugu.</p>	<p>Bejerano G <i>et al</i> 2004 <i>Science</i></p>

UCR	Ultraconserved regions	<p><b>Sequenze genomiche utilizzate:</b> Genomi di uomo, topo e Fugu</p> <p><b>Metodo di identificazione:</b> Allineamenti di menoma umano e murino, selezione delle regioni con &gt;95% di identità di sequenza per almeno 50 bp. Queste regioni sono confrontate col menoma di Fugu. Regioni che si sovrappongono a sequenze geniche vengono escluse dall'analisi.</p> <p><b>Caratteristiche principali:</b> Il menoma umano contiene 3583 UCrs con una lunghezza media di 125 bp. Sono sequenze raggruppate in cluster e si trovano vicino a geni coinvolti nello sviluppo</p>	Sandelin A <i>et al.</i> 2004 <i>BMC Genomics</i>
CNE	Conserved noncoding elements	<p><b>Sequenze genomiche utilizzate:</b> Genomi di uomo e Fugu</p> <p><b>Metodo di identificazione:</b> Similarità di sequenza locale. Lunghezza minima di allineamento di 100 bp. CNEs che si sovrappongono a sequenze codificanti e regioni ripetute sono escluse dall'analisi</p> <p><b>Caratteristiche principali:</b> Il menoma umano contiene 1373 CNEs con una lunghezza media di 199 bp. Le CNEs sono raggruppate in cluster e si trovano vicino a geni coinvolti nello sviluppo</p>	Woolfe A <i>et al.</i> 2007 <i>BMC Dev Biol</i>

**Tabella 6.** Primers utilizzati nelle reazioni di RT-PCR per amplificare i cDNA dei trascritti del locus del gene *CYYR1*

N. accesso GenBank	Primer n°	Sigla	Sequenza dei primers 5' → 3'	Regione	Prodotto (bp)	PCR
NM052954	#1	SYYA CDS s	gctgctctctccatctgatcgc	5' UTR	605 bp	PCR61 40Cy
	#2	SYYA CDS as	caatgggcgatcttgctggaat	3' UTR		
	#3	CYYR1 E3/E4s	gtctctctctatctctggaccac	E3/E4	196 bp	PCR61 40Cy
	#2	SYYA CDS as	caatgggcgatcttgctggaat	3' UTR		
CN402794	#1	SYYA CDS s	gctgctctctccatctgatcgc	5' UTR	855 bp	PCR61 40Cy
	#4	CYYR1 spl1 3' as	gaattgctttgagataacgttctag	3' UTR		
	#1	SYYA CDS s	gctgctctctccatctgatcgc	5' UTR	436 bp	PCR61 45Cy
	#5	CYYR1_SPL1 E3/E4b s	gagttcatattcaggataggaggag	E3/E4b		
CN402791	#1	SYYA CDSs	gctgctctctccatctgatcgc	5' UTR	449 bp	PCR58 40Cy
	#6	CYYR1 spl2 3' as2	gacgaggaaatacaaacagtagag	3' UTR		
	#1	SYYA CDS s	gctgctctctccatctgatcgc	5' UTR	276 bp	PCR61 45Cy
	#7	CYYR1_SPL2 E2/E3b as	gaattctctcgagaggatattcc	E2/E3b		
DB268533	#8	CYYR1 spl3 s	ctaagcaagtgcctcaagccag	5' UTR	508 bp	PCR61 45Cy
	#2	SYYA CDS as	caatgggcgatcttgctggaat	3' UTR		
	#9	CYYR1_SPL3 E1b/E2 s	aggcaggatgattgcttgctc	E1b/E2	448 bp	PCR61 45Cy
	#2	SYYA CDS as	caatgggcgatcttgctggaat	3' UTR		
DB120436	#1	SYYA CDS s	gctgctctctccatctgatcgc	5' UTR	447 bp	PCR61 20Cy
	#2	SYYA CDS as	caatgggcgatcttgctggaat	3' UTR		
	#1	SYYA CDS s	gctgctctctccatctgatcgc	5' UTR	274 bp	PCR61 45Cy
	#10	CYYR1_SPL4 E2/E4(CAG-) as	gggtgctcgagaggatattccaat	E2/E4		
AK304124	#1	SYYA CDS s	gctgctctctccatctgatcgc	5' UTR	775 bp	PCR61 20Cy
	#2	SYYA CDS as	caatgggcgatcttgctggaat	3' UTR		
	#1	SYYA CDS s	gctgctctctccatctgatcgc	5' UTR	448 bp	PCR61 45Cy
	#11	CYYR1_SPL5_E2bis/E3 as	cagtgccccagggtggaggtag	E2bis/E3		
AK125338	#12	AK125338_E2(277)s	gagctgacagaaatcatcaattgc	Esone 2	736 bp	PCR66 45Cy
	#13	AK125338_E3(1012)as	gtcagacactaatacagctgctgac	Esone 3		
	#14	AK125338_E3(827)s	gagctagctcagatcattgataaacgtg	Esone 3	1330 bp	PCR64 45Cy
	#15	AK125338_E4(2156)as	ccttgaggcttacctctggattgtg	Esone 4		
	#16	AK125338_E3(1425)s	gatggatccgggcaaagtaaattg	Esone 3	732 bp	PCR64 45Cy
	#15	AK125338_E4(2156)as	ccttgaggcttacctctggattgtg	Esone 4		
	#17	AK125338_E3(2060)s	acaacaactgctgaaggctcggatg	Esone 3	551 bp	PCR64 45Cy
	#18	AK125338_E4(2610)as	ctggactcctatctttaccacatac	Esone 4		
	#19	AK125338_E4(2551)s	ggataaccagttggccatacacag	Esone 4	809 bp	PCR61 45Cy
	#20	AK125338_E4(3359)as	gattcatccaagtgttcattaatag	Esone 4		

**Tabella 7.** Lista dei geni *housekeeping* candidati e *primers* relativi per l'analisi in pesce zebra.

<b>Simbolo genico</b>	<b>Nome del gene</b>	<b>N. accesso GenBank</b>	<b>Sequenza dei primers 5'→ 3'</b>	<b>Dimensione prodotto di RT-PCR</b>	<b>n. cicli PCR</b>
<i>ef1a</i>	elongation factor 1-alpha	NM_131263	ctgtacctgtgggtcgtgtggagact g cagccttctgtgcagactttgtgacc	594 bp	21
<i>slc25a5</i>	Solute carrier family 25 alpha, member 5	NM_173247	gtaacctggcttcaggtggtgctg cttctccagcagtcattgtgcc	437 bp	20
<i>zp2.4</i>	zona pellucida glycoprotein 2.4	NM_131829	ctgactaccagtcaccaaagtcc ctaaggcaccacagtaactgtcc	462 bp	
<i>ybx1</i>	Y box binding protein 1	NM_131620	gtagcgttggggacggagagactg ggccggatggtctggatgttaag	727 bp	
<i>bactin2</i>	bactin2	NM_181601	gcagaaggagatcacatcctggc cattgccgtcacctcaccgttc	322 bp	21
<i>gapdhs</i>	glyceraldehyde-3-phosphate dehydrogenase, spermatogenic	NM_213094	cagtcgctgatgtccgttgc gagacgaacgcagcgtctcgtag	416 bp	27
<i>b2m</i>	beta-2-microglobulin	NM_131163	ctccagatatctccattgaactgc gtccaaagacacaccaatccatc	246 bp	
<i>cyyr1</i>	cysteine and tyrosine-rich protein 1	NM_212882	catggctgaggtatc gatgcactg cagcaggtcagcgtctggcctc	610 bp	30

**Tabella 8.** Primers utilizzati in reazioni di RT-PCR per l'amplificazione di CNEs di *Ciona intestinalis*.

CNE	Primers n°	Sigla	sequenza dei primers 5' → 3'	Prodotto (bp)
Pax6_1	#21	Pax6 9q CNE1 fw2	gaacactgatctacaccgag	111
	#22	Pax6 9q CNE1rv	cgatgctactataacgtaacggg	
Pax6_2	#23	Pax6 9q CNE2 fw	ggtctctggtttccattagttg	421
	#24	Pax6 9q CNE2 rv	gctaacaactacttagctctgg	
Otx/Ptx_1	#25	OTX1/2 4q CNE3 fw	ttgattaacacacttcagacg	235
	#26	OTX1/2 4q CNE3 rv	atcacaagcgctggcaagcc	
Nkx2-4_1	#27	NKX2-4 10qCNE4 fw	catacactctggacgtatgc	317
	#28	NKX2-4 10qCNE4 rv	ttcagtgatcataaccacgg	
Nkx2-4_2	#29	NKX2-410q CNE5fw	gcttattttgcattaaaccgg	217
	#30	NKX2-410q CNE5rv2	aagtttgggccacaaagcgc	
Nkx2-4_3	#31	NKX2-4 10q CNE6 fw	ggaatttattgtgaatggatac	175
	#32	NKX2-4 10q CNE6 rv	caagttcaacgaaaaagttcg	
Otp_1	#33	OTP 14q CNE7 fw	atttctgcgattaatttctgtg	241
	#34	OTP 14q CNE7 rv	agaatatttattgtcataagttg	
Phox2Blike_1	#35	PHOX2Blike 5q CNE8 fw	gcaagaacgacaaattaccac	168
	#36	PHOX2Blike 5q CNE8 rv	atgaaacatacgaagaaaatcc	
Meis_1	#37	MEIS1 10q CNE9 fw	ccggctccagggtgttaaac	442
	#38	MEIS1 10q CNE9 rv	gccacaaattgattgaacgaaa	
Meis_2	#39	MEIS1 10q CNE10 fw	gtcgatttctgagttacgtcg	153
	#40	MEIS1 10q CNE10 rv	gtgctccccttctactggac	
Meis_3	#41	MEIS1 10q CNE11 fw	cagaactggcatttatcgacg	227
	#42	MEIS1 10q CNE11 rv	gctggccatattgtctcatc	
Meis_4	#43	MEIS1 10q CNE12 fw	gctgctgtcacagacataac	274
	#44	MEIS1 10q CNE12 rv	ttcaatttccgtctcggtcg	
Meis_5	#45	MEIS1 10q CNE13 fw	gggtttgttttgaggcgag	190
	#46	MEIS1 10q CNE13 rv	attaatctactgctctgcag	
Meis_6	#47	MEIS1 10q CNE14 fw	tgaattacgcagtacaatcacg	326
	#48	MEIS1 10q CNE14 rv	tatcgacctctgctgtttc	
Meis_7	#49	MEIS1 10q CNE15 fw	atgtctgggactgcgatattg	309
	#50	MEIS1 10q CNE15 rv	cagtctgtactataattatgtac	
Meis_8	#51	MEIS1 10q CNE16 fw	tcaaaggcctgtaaatagc	131
	#52	MEIS1 10q CNE16 rv	acagtcgatttggtagcattac	
Meis_9	#53	MEIS1 10q CNE17 fw	ccaggttactagacagggtg	122
	#54	MEIS1 10q CNE17 rv	ccatcttgatcttatagtttg	
Meis_10	#55	MEIS1 10q CNE18 fw	agagcgtctgtatagcgtttg	116
	#56	MEIS1 10q CNE18 rv	cagcccgatgcatttctattag	
Ebf3_1	#57	ciEBF3scaff186 CNE19 fw	gaatcgctccgagttgtccac	121
	#58	ciEBF3scaff186 CNE19 rv	tcctagtccaggaactagag	
Zfhx1B_1	#59	ZFHX1B 1q CNE20 fw	caaagcaggctcagttcggtg	289
	#60	ZFHX1B 1q CNE20 rv	gttctgaccatactggacc	
Zfhx1B_2	#61	ZFHX1B 1q CNE21 fw	attataaaccacctgaaccac	161
	#62	ZFHX1B 1q CNE21 rv	agatgatcaaacctaacgag	

**Tabella 9.** Localizzazione dei *primers* progettati per l'aplicazione delle CNEs di *Ciona intestinalis*.

<i>C. intestinalis</i> CNE	Sequenze delle CNEs estese di 50 basi sia al 5' che al 3'
Pax6_1	CNE1>9q dna: chromosome:JGI2:9q:3321806:3322008:1 ATCAAACCTGGGCCGTTACTTTTTCTCTACGGTCCGTCACCTTTTTCTCCGGGAACACTGATC TACACCGAGCAACAACAATAGCCACCGAACCTTGTTCCACGATGGCCGAGAAAAGGAAA AAAGCGGGGGATTCCCGATTACGTATATAGTAGCATCGACGAAACATTTCTTGATT TGGGAGCCGTCGTCATCCAAATT
Pax6_2	CNE2>9q dna: chromosome:JGI2:9q:3328137:3328649:1 TTGCCTCTCTAAAATAATTACGTTGAGCTTAAACCATCAACGGGTCTCTGGTTTCCATTA GTTGAAAAATAAAAACTATTTCCCGACCACTGACAGCCGTATTAATGATTAATTGGACG GGATTTCAAGACGATTTCCGAGCCGCTTCGAAAATAAATAAAAAGCTCTTTGGATCT CTCGCCCCGTCGCGTCACGTAAGATCCGTCGACCGCAACCGCCAGCCAGCCCCGCAT TTCCTATCTGACGCTGACGCTAATTCCAAGATAAACCGAGTTATTATGAATCAAGTTGCA CCGCGATCGATACCGGCGTGCATCGATCCGTCGGCGGGCTCTCGACGACCAAAAACCCC GCCGAAGTTTCGCTTAAAACCGGCCTACTCGATTATGAGCGATCGATTTTGGAGAGAAT GAAAGACATCCGTGACACGGCCAGAGCTAAGTAGTTTTGTTAGCGAATTCTGTTGACATTC CGTAGCTTGCAACATAACTCTATTTCAAGCCTG
Otx/Ptx_1	CNE3>4q dna: chromosome:JGI2:4q:4291066:4291387:1 CATTGAGTGCTGGGCAAAGTTTCCCTGTCGTGTTGAAGAGAAGTGGTTGGATTAAACAC ACTTCAGACGTGTGAAAATTCGAGTTGTTTTCAAAAACAAAGTGCGAACGGTAGCGCAA GAGCGTCGTTGTTGTGACACTGCGCTCGTTCCCAAGCCAGCTCGTATAATAGAAAAACAT AAACGCGAAACCTCAAGCAGTACTCCATCAGGTCCATTACATAAGCAGCTACTTGGAC ATTTTCGCACCTTCTATAATGGGTTGCCAGCGCTTTGTGATCGATCGGGGCCAACTGAA ATAGGCTCGTTAACCGTGAACG
Nkx2-4_1	CNE4>10q dna: chromosome:JGI2:10q:3237715:3238110:1 CCTTTTGCTACTGAGTTTAGTCAAACATACACACATACACTCTTGGACGTATGCATTAA AATTACCTTTGCCAATGCCAATATAGCTCGACTAATGGCGACCTAACAAAATTCGCTTC TCATATACCGACTGTAGTGCTTACTATAGCCCTCGTCTCATCTTAGGCCATCTTAAAG GGCGCTCTGGGATATGTTGTAAGTGCATTTGAGTCTAAAGTAGCCTGCCATCGGGCTC AAGTCCCCCGCTCAACAAGAAGAGCCTCCATAATTCAACATTAATTACTTTACTAGTGT GTCCATAGCTAAGGGTGCTTACCAGCATTTCCGTGGTTATGTACACTCGAAACGCGGCAT GTTGCTTTGCAAACAAGACACTTTCCCCGAATGATA
Nkx2-4_2	CNE5>10q dna: chromosome:JGI2:10q:3242400:3242710:1 ATATCTATCCGATTTACGTTACAGGTTTGTGAAAATTATTCTTGCTTTATTTGCATTA AACCGGGCAATTGTTACGTATCCGGAGAAACAATCACGACAACCAAAAATACACAGAAAGC AAATAAACTCTGACAGGCCGTGTGACTTCCACGCTAGAGTTGTCTTAAGCTCACACCAAT CAGTGATATCAAAGAGGGCTCTATGACATCATAATACGGAGTCTGGAACCTGTTGGCATA AGCGCTTTGTGGCCAAACTTTCAGCACCCGAATCCGTATAGCTCACAAAACGGCTATAC CATCCAATCAT

Nkx2-4_3	CNE6>10q dna: chromosome:JGI2:10q:3244024:3244297:1 AGTTATTTGGCAACGATAATGAAACGACAAGGTTCTCTCCTGGAATTAGGAATTTATTT GTGAATGGATACAATCTAGAACTCAGGTAACGGCATTGTCCGTGTTTCTTACGTCACAAT GGTCAGATCACCGGCTCAGCGGGGAAAACCCACGTCCTCAAAGCACCCCGCTGGGGA AGATTAGTTCACAAATTAAGTCA <b>CGAACTTTTTCGTTGAACTTG</b> AACTGTATTGGCACAA GGGAAAACATAATTAGAAATGGTAAAAATCGGG
Otp_1	CNE7>14q dna:chromosome chromosome:JGI2:14q:2219097:2219424:1 AGTCATAGCCGGTTAGATACGTCGACGTCGTTTACGGCAAATTTCTGCGATTAATTTTC GTGAATTTTATTTCCGTACGACGCCTCTGTTTCAAGCTAACTTTAGTCTTCCATTTGTC TGTTTATATCAAAGCAGTTTGACTTCGATCTTCGACAGGCCATTGACAGTCATACCCAAA TCAATACACCCTCGGCGAAGGAATATTTACACGCGAAATAATACGCAGAAAATTTTATA AATATTTAATTAATAAATT <b>CAACTTATGACAATAAATATTCT</b> AAAAAAGACGATTTTTT TTAACCGTAATTGCTCTAATAAAGAAAT
Phox2Blike_1	CNE8>5q dna:chromosome chromosome:JGI2:5q:870692:870956:1 AAACGAAATGCCAACGAGCAAAGAAAACCTATTTCCACGCCTACCGAC <b>GCAAGAACGAC</b> <b>AAATTACCAC</b> GTTTCAATTATAGTTTCCCCGAGCCAGCGTAAATAATAGCGAGCTTTGGC GAAATCAAATTTTGTGAGGGCTTACTGAACAAGAGAATAGATCAGAGAAGTGGTGACTT GACAAATCCTAGGCT <b>GGATTTTCTTCGTATGTTTCAT</b> CCTTACCAATACCGCACGCAAGC TGCTGGTCGTAATGATAGGCTTTCG
Meis_1	CNE9>10q dna:chromosome chromosome:JGI2:10q:2049300:2049836:1 TCCCATCTCTGCTTATTATTCTTGATCGGTAACGGCGACCGGTCCCAG <b>GC</b> CGGTCCAGG <b>GTTGTAAAAC</b> GTGACGAGTTCGGCGAGGGCCGTAAATCACTCGATGCGTTGGTTTGGG GGGAGGAGAGTTACAAGCGCTCTCGGCAGGGTCATCTAAAGGTGAAGCCCGGGTCAGGAT CAGGTGGCTCTCCCGAGACGTTCCGGAGGACCCCCCTCTCAGCGCGCATTGCACGATCC GCTTTCGTTTATTAAAGTTTATAGCTCGTTTATTACGGCGCGTTCGCCTACGCCAAAG GCCATCGATTAACGTCGCGAAGAGGTCTGGGGTCTTGATCTTGCTCCCCGGTCCAGTGAG CTCCTTTCCCCGAGATTAATTGCGAATGGCCCGTGTCTCGGCCTCAACAGATGCTGCCGA TTAAACTTGATTATTTTTCCGCCCTTAAGCCTAAATTGCGCGAGGATTA <b>TTTCGTTCAA</b> <b>TCAATTTGTGGC</b> GATGTCGGCACAGCAGAAGGCGGTGATCCGGGCGTAAGTGATCCG
Meis_2	CNE10>10q dna: chromosome:JGI2:10q:2050261:2050506:1 ACATTGGTTCTACTGTTTGGGTTGCTTATCAATCTGGCCTTCGT <b>GTCGATTTTCGAGTT</b> <b>TACGT</b> CGAAACTCGAGACTCGTTGCCTCTGCCACGTCATCCGTCGTAAGCAACCGGAAC TGACTGGCCAGGAGGCTTCGCGTTGACCGATCTGTGACATAATAATGAAGATGTAT <b>CGTC</b> <b>CAAGT</b> GAAAGGGGAG <b>CAC</b> GAAAAATGGGATTCGAATTTAGAGAGAGTTCGTATAAATTG CAGCCG
Meis_3	CNE11>10q dna: chromosome:JGI2:10q:2052358:2052674:1 GGCCGAAATTAGCTAGCGCCGCCAAGGAAATGTCACCGCG <b>CAGA</b> ACTGGCATTAT <b>TCGA</b> <b>CG</b> CGCCGGTCGCCATTTTGTACTCGCTGATGGCTGCTGTCAATATGACATCTTGACGTAG ACGGAGGCGATTTGTTTGCCTACTGCGGCTTGCAAGTAGGTGCAGCGCGCTGGTAGAT CATAATTAGCAAGTGGACGCAAAGGCGGCATTGGAAACAATATTCCAGGCGTGAAAAGC GAAATAT <b>GATGAGACAAATATGGCCAGC</b> GTGAAATCGCGCTGTGAGTGTCTCTCCAGAC GTCCAACCTGACGGAAT

Meis_4	CNE12>10q dna: chromosome:JGI2:10q:2054580:2054950:1 CGCGAGTTTCTGAGATGCTAGAGCGGCTGTAACAATCTATTCTTGCATAA <b>GCTGTCGTT</b> <b>ACAGACATAAC</b> TTTATTACCTTTGACGGGTAGGCGGTTTTAATGGTGCCTTGTGCATAT ATTCCAAAGCCAGGTCATTAATAACGGAATTTGACGTTAAACGGCGCCGCATATGCCCGA AGTTCTAAACGCTCTGGCAATTATAACTGCACGGAAATTGAATTAGAAATAAATCGAAGT TCGATGATCTGACGGGCGCTCTCAAACCTCAAATTCGATTTGGATGAATATTCAATACGCA ATCG <b>CGACCGAGACGGAAATTGAA</b> CGAAAAAGAATTGCGCGCTTTCCAGACCCGAGACTT TAATACGTGAC
Meis_5	CNE13>10q dna: chromosome:JGI2:10q:2063167:2063455:1 GGAGGCGAGCAAATTCACGAGCAAATAGAAAAACAGC <i>CGAGTTTCATCGGGGTTTGT</i> <i>TTGAGGCGAG</i> CGAGGGCCAGCAAGCAGTTTTATCACGCTTGCTAGAAAAGAAGAAGCC GTCGTAGTATTGGGCACAGTGACGCATACTCTGGCCTGTGTCAGTGTGATTCGGACAGCA GAGAAAAGGCTTATATGTAGCCTTCTAGTCTTGCAAC <b>CTGCAGAGCAGTGAGATTAAT</b> GTCGACCTGCGTGCTTTTCGTTTCAGATCGTTGATAGGTGCCTAACTCAG
Meis_6	CNE14>10q dna: chromosome:JGI2:10q:2067413:2067819:1 AAATTAGCCGTCATCGGCGCCGTCAAAGACGTCACGTTGACGTATGAATTACGCAGTACA <i>ATCACGAAATATTGGTAGGAATTTGGATCACTGGTCGCGGGGCCCTAGGGACTTGTGT</i> <i>ATTGTGTGAGGGGCTTTGTAGTGGCGGTGTCTCTGCTTCTAATCCCACCGCCTCGTACA</i> <i>AAGTGCTGCTTCCAAGGATAGGGATAAAAGGCGACTGGGCTTCGTGGGAAGGCACGACT</i> <i>CAGATCGATCCTTCGATCGTCGAATCTTGAAGAAGATAATATACAGAGATACTTTTACA</i> <i>TGGCGCACCGAGCAGACGAAAGCGCGCATTATCAGAGAAATATCA7TTTAGAAAGCAGCG</i> <i>AAGGTCGATA</i> CCGGGAAAATCAATGGATTTGTTGCAGGGCCTTGAA
Meis_7	CNE15>10q dna: chromosome:JGI2:10q:2073566:2073960:1 AGTCGTATTACGTCGCATAATTTACGGCATATACGGCGATAGTTTCGATGTCTGGGACTG <i>CGATATTGTTACACATGTTTCAATATTTGCGGAACTCGGCGCTCGGTGATTGTGCAA</i> <i>ACTCGGCCGGGCTGCGAGCCGCCACTCCCGCCTCGATAATCATTAAATGCCCGATGTTGAAA</i> <i>GAGCGGCCGACCCTTGTGATCGTCGTTTCTCGAGTGGGAGGCTTCGGCGCCTCAGAAGC</i> <i>GAATCTTGAGTGACGTCTGCTTCGAACGACCTACGTCGTAACGATGTTTGTTTTTTCTGA</i> <i>TCGTGAAAAATTAACAATATTGAAATGAAAGTACATAATTATAGTACAGACTGTTTCG</i> <i>AAACGGTATGTTTGGCGGCGTTACAGTTCTAACCC</i>
Meis_8	CNE16>10q dna: chromosome:JGI2:10q:2073947:2074169:1 TACAGTTCTAACCTTATCTCCAAAGTCCCGTACGAGTAATCATATGCTCAAAGGCCTG <i>GTAATAGCTTAAAAATTCAGTATTTAGGGACACCTTTTGTTTAATATTAATTCTGTTT</i> <i>ATGCCACGATCGTCTTGGTGTGGGAAAGCCTAAAATCGTAATGCTAACAAATCGACTGTC</i> <i>GGCGTAAGGTCGTTTCTTCAACCGCGCAAGTATAATAATAATT</i>
Meis_9	CNE17>10q dna: chromosome:JGI2:10q:2074920:2075118:1 CGTCAAATTTAACTTAACCACGGACGCAGTTTTTCGTGGAAGAA <b>CCAGGTTTACTAGAG</b> <b>CAGGTG</b> CGATTGTTTGTGACAGACTGATTGGCGACCACGAATGGCACGATCGCCGCACT AAGGCATGAGTTGTGTTTATTTA <b>CAACACTATAAGATCCAAGATGG</b> ACGAAACCCGCGT ATAAAGAAATCCTATGTAA

Meis_10	CNE18>10q dna: chromosome:JGI2:10q:2078121:2078327:1 TCTCCCCCTCCGAGAAATGGGTCGCTCAGCAGCACTCTTATTCTACAGAGCGTCTTGTATAGCGTTTGTAAATGGGATAAGCGGCCTTTAATGCGTCATAATGCAAACGCGCGCAGTATGTGATTAATATTTTCATAATGCACTAATAGAAATGCATCGGGCTGCTTCAGTGACCTAACGATACGTGTTCTACGTCATAACTGCTTT
Ebf3_1	CNE19>:scaffold:JGI2:scaffold_186:69251:69459:1 GGACCCGGCCCATGGGTCGTCCGGGACGAATCCACCAAGTTAGTGGCGAATCGTCCGAGTTGTCCACTGAGAGAAGTCTCTGTATTGATGTGTCCGTTGAGACACTGGCTTGCTGGTTTTAAACTCCACGACGCGTGGCCTTCTATACTCTAGTATTCCTGACTAGGACAGAAAGAAAAAGCGGCGCATGTCTTGGAGTAAGTTCT
Zfhx1B_1	CNE20>1q dna: chromosome:JGI2:1q:6452291:6452679:1 CAGGTCACGCAAATTTGCATACACAGCGCGAGTCAGCAAGCCGGACGGTCAAAGCAGGTCAGTTCCGGTGAACCTGTCACTACGTCATCATGCGTGACGTCACACCGCGGGTGGTGGACTTTCTCCATTACAGCGCTCGCCTAGATCAGGGCAGGATAAAGCTTAATTGTAATCGTAACAAAGAGGAAACGCTCTCCTTATTCTCTCACTTCTCACTCTGCTCTTTTTTTGTTTCTCTTTATTCCCCCTCGCCGTATGTGACAGTGTGGTGCGAATATTAAGCAAAGGAAGATGAGACTCGACAGCATTCCGGTGCTCGGTCCAGGTATGGTCAGGAACTCGAAGGGAATTGTACAAGATACACAGCAGACAGTACAGCAGTTTCGAA
Zfhx1B_2	CNE21>1q dna: chromosome:JGI2:1q:6481600:6481854:1 ATACAAAACCTTTAGGAAAAGAATTCAGTGAGGATACTAGGAAGAATTATAAACCCACCTTGAACCCACTTTGATATTTCTGTGCGAAAGCGTAACCTGCTCTTCGTACACGGACATTTTCCATCAAACCTAAATACCTCACTTGGTGGCGCACAAGATCACTGCTGAACACCGATGACCTTTTGACCTCGTTAGGTTTGTATGATCTGAGGATAATTAATCGAGTTTGGGTTATAAGATTTCTATATTAGGTG

**Tabella 10.** Lista dei geni utilizzati in *Real-time* PCR per l'analisi quantitativa relativa del gene *cyrr1* in *zebrafish*.

<b>Simbolo genico</b>	<b>Funzione del gene</b>	<b>N. accesso GenBank</b>	<b>Primer n°</b>	<b>Sequenza dei primers 5'→3'</b>	<b>Prodotto (bp)</b>
<i>cyrr1</i>	Sconosciuta	NM_212882	#63	gatgtgctgtcaggaacggctatc	594 bp
			#64	ctgaaaactcggactctggaccgc	
<i>efla</i>	Fattore di allungamento	NM_131263	#65	ctgtacctgtgggtcgtgtggagactg	437 bp
			#66	gttcaggatgatgacctgagcg	
<i>slc25a5</i>	Trasportatore	NM_173247	#67	ctgcccgatccaagaacac	462 bp
			#68	cttcctccagcagtcattgtgcc	
<i>bactin2</i>	Legami con ATP, nucleotidi e proteine	NM_181601	#69	gaaggagatcacatccctggc	727 bp
			#70	cattgccgtcaccttcaccgttc	

**Tabella 11.** Sequenze dei *primers* per la realizzazione di costrutti del gene *cyrr1* e degli oligonucleotidi morfolinati da microiniettare in embrioni di *zebrafish*.

In maiuscolo le basi del sito di restrizione specifico per il clonaggio in vettore: *BamHI* (GGATCC) e *EcoRI* (GAATTC).

<b>Primer n°</b>	<b>Sequenza del primer 5' → 3'</b>
#71	GCGGATCCcggattatggagaatcccagaagc
#72	GCGGATCCgagatcaggaggaaacatctc
#73	GCGGATCCcggattatggagaatcccagaagc
#74	cttcacacacattcctgacagcacatctcc
#75	ggagatgtgctgtcaatgtgtgtgaagaacagc
#76	GCGGATCCcctgctgaatccagggtatggtgg
#77	CGGAATTCcggattatggagaatcccagaagc
#78	CGGAATCgcctcaggtctggaaatggtc
#79	ggagcaagcttgatttaggtgac
#80	ctcactatagttctagaggctcg
#81	cgttctgggattctccataatc
#82	atccacagcagcagcccctccatcatcc
#83	cctcttacctcaGttacaattata

**Tabella 12.** Valori dei prodotti di RT-PCR (espressi in nanogrammi) dei geni *housekeeping* candidati per l'analisi di espressione in stadi embrionali di *zebrafish*.

	<i>bactin2</i> 21 cicli*	<i>slc25a5</i> 20 cicli	<i>ef1a</i> 21 cicli	<i>gapdhs</i> 27 cicli	<i>cyr1</i> 30 cicli
<b>Sfera</b>					
Pool A (n=24)	32.6	32.5	13.9	15.2	40.7
"	32.2	30.5	17.2	14.2	39.5
Pool B (n=23)	16.8	12.1	5.9	14.2	18.2
"	18.7	11.0	7.2	16.7	20.7
<b>Cupola</b>					
Pool A (n=24)	70.1	104.2	<i>n.d.</i>	143.8	146.4
"	68.0	103.5	<i>n.d.</i>	146.5	154.3
Pool B (n=23)	21.6	20.2	15.2	17.8	65.1
"	21.9	20.4	18.3	16.1	69.0
<b>Gastrula</b>					
Pool A (n=24)	39.8	39.5	61.1	18.5	128.3
"	35.0	39.2	60.2	18.6	129.1
Pool B (n=23)	62.8	105.3	117.7	30.4	201.3
"	56.0	104.8	115.9	28.7	195.6
<b>1-5 Somiti</b>					
Pool A (n=24)	32.6	49.3	70.4	7.8	120.8
"	31.5	54.8	65.6	7.8	125.2
Pool B (n=23)	71.2	102.0	119.8	53.9	165.5
"	72.4	103.2	134.7	55.6	166.1
<b>48hpf<sup>§</sup></b>					
Pool A (n=24)	77.5	120.0	105.3	125.2	158.6
"	80.3	119.6	103.4	126.6	157.7
Pool B (n=23)	74.3	101.3	111.2	132.6	149.7
"	73.2	101.8	106.4	139.5	156.1
<b>5 giorni</b>					
Pool A (n=24)	71.8	130.1	116.9	143.4	126.1
"	90.3	122.4	118.3	157.8	128.3
Pool B (n=23)	76.3	104.1	101.0	120.7	101.6
"	79.0	107.6	90.6	110.3	95.6
<b>Media</b>	<b>54.4</b>	<b>76.6</b>	<b>76.2</b>	<b>69.3</b>	<b>119.2</b>

\* numero specifico cicli corrispondente al plateau della reazione di amplificazione

§ hpf: *hours post fertilization*, ore dopo fertilizzazione

*n.d.*: non disponibile.

**Tabella 13.** Valori dei prodotti di RT-PCR (espressi in nanogrammi) dei geni housekeeping candidati per l'analisi di espressione in tessuti di zebrafish adulto.

*Differenti pesci adulti sono indicati con lettere diverse da A a G. I sessi sono descritti con /M (maschio) e /F (femmina).*

	<i>bactin2</i> 21 cicli*	<i>slc25a5</i> 20 cicli	<i>ef1a</i> 21 cicli	<i>gapdhs</i> 27 cicli	<i>cyyr1</i> 30 cicli
<b>Cervello (n=5)</b>					
A/M	45.6	39.8	10.3	186.6	28.1
"	51.2	36.7	10.3	195.1	33.8
B/F	46.1	32.4	10.2	200.9	43.0
"	45.2	35.6	11.6	205.2	39.7
D/F	51.6	41.2	15.5	227.5	29.3
"	27.4	38.4	14.9	244.6	34.2
E/M	61.3	71.4	33.9	266.9	88.5
"	68.2	73.5	32.0	273.5	90.9
F/F	72.5	65.1	30.8	<i>n.d.</i>	75.8
"	76.7	62.5	29.5	<i>n.d.</i>	70.5
<b>Muscolo (n=5)</b>					
A/M	65.5	139.6	39.6	101.3	34.2
"	68.4	163.1	43.4	110.9	33.8
B/F	21.5	87.6	6.4	33.5	<i>n.a.</i>
"	20.9	75.6	6.1	35.7	<i>n.a.</i>
C/F	77.1	200.3	72.8	114.4	46.3
"	82.0	230.8	60.6	115.7	42.4
D/F	48.6	149.6	19.4	90.9	13.2
"	46.1	124.1	20.1	87.9	15.5
E/M	71.8	199.9	47.9	105.1	61.5
"	70.7	203.6	48.0	113.5	60.5
<b>Pelle (n=5)</b>					
A/M	92.1	117.3	77.8	168.5	60.4
"	95.1	118.3	80.3	174.1	55.0
C/F	67.8	107.1	45.0	89.6	12.8
"	67.9	109.4	48.8	98.0	8.7
D/F	64.6	76.5	39.6	101.9	20.3
"	67.6	76.8	40.7	119.5	22.1
E/M	75.5	114.1	49.9	139.0	<i>n.a.</i>
"	79.2	113.0	52.3	146.5	<i>n.a.</i>
F/F	86.2	74.1	50.5	96.2	46.9
"	90.1	67.9	47.4	81.8	51.3
<b>Osso (n=4)</b>					
B/F	38.9	57.1	11.3	110.9	25.7
"	41.8	55.5	10.1	106.9	26.6
E/M	71.7	96.3	28.3	126.2	57.1
"	71.0	106.7	29.7	123.9	55.2
F/F	82.8	124.0	40.7	109.6	<i>n.a.</i>
"	83.6	121.1	39.9	95.8	<i>n.a.</i>
G/F	52.8	47.3	16.0	8.5	26.7
"	44.3	71.5	17.3	9.3	29.4
<b>Occhi (n=5)</b>					
A/M	99.3	101.3	72.2	179.7	35.2
"	86.6	106.0	69.0	179.0	40.4
B/F	60.5	73.1	46.1	147.9	44.8
"	66.4	73.7	49.5	155.2	45.4

C/F	59.3	66.2	33.2	148.0	17.2
"	58.3	65.5	34.4	159.1	19.3
D/F	71.4	107.1	52.6	175.4	29.1
"	76.0	102.4	48.4	182.2	38.7
E/M	92.8	105.1	70.3	170.3	71.7
"	77.6	107.5	71.2	165.0	73.2
<b>Ovaio (n=5)</b>					
B/F	142.7	150.2	129.5	155.7	165.0
"	149.0	149.0	133.2	156.6	165.6
C/F	136.7	153.4	140.1	174.4	184.2
"	160.3	161.7	133.8	176.0	179.5
F/F	156.6	159.4	131.6	177.0	165.7
"	164.7	147.4	130.7	170.5	161.3
G/F	165.3	147.0	138.9	162.9	165.6
"	99.7	156.0	138.4	170.9	163.2
H/F	129.2	143.9	119.3	143.7	143.6
"	135.3	135.1	123.9	160.0	153.9
<b>Media</b>	<b>79.0</b>	<b>105.8</b>	<b>54.4</b>	<b>141,5</b>	<b>65.4</b>

\* numero specifico cicli corrispondente al plateau della reazione di amplificazione

*n.d.*: non disponibile.