

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN
Scienze Biochimiche e Biotecnologiche

Ciclo XXVIII

Settore Concorsuale di afferenza: 05/E1

Settore Scientifico disciplinare: BIO/10

**UN'INDAGINE SU LARGA SCALA DELL'INTERAZIONE
ENHANCER - PROMOTORE MEDIATA DA FATTORI DI TRASCRIZIONE**

Presentata da: Francesco Aggazio

Coordinatore Dottorato

Prof. Santi Mario Spampinato

Relatore

Prof.ssa Rita Casadio

Correlatori

Prof. Pier Luigi Martelli

Prof.ssa Sara C. Madeira

Esame finale anno 2017

INDICE

INTRODUZIONE

CAPITOLO I - L'attivazione dell'espressione genica: un sommario introduttivo

1.1 Proteine coinvolte nell'attivazione della trascrizione: I fattori

I fattori basali e il complesso di pre-inizio della trascrizione

I fattori di trascrizione

I cofattori

1.2 Sequenze nucleotidiche coinvolte nell'attivazione della trascrizione: Gli elementi di regolazione *cis*-agenti

Nucleo del promotore

Elementi prossimi al promotore

Enhancer

Silencer

Isolatori

1.3 Comunicazione enhancer-promotore: il looping della cromatina

CAPITOLO II - Siti di legame per i fattori di trascrizione: rappresentazione e predizione

2.1 Rappresentazione dei motivi

Rappresentazione testuale: le sequenze consenso

Rappresentazione matematica: le matrici di posizione

Rappresentazione grafica: sequence logos

2.2 Matrici di posizione e predizione di siti di legame di fattori di trascrizione

Position Weight Matrices

Predizione di siti di legame di fattori di trascrizione mediante PWM

Significatività statistica dei siti di legame predetti

CAPITOLO III - Identificazione di coppie e gruppi di fattori di trascrizione su coppie di enhancer e promotori di geni regolati

3.1 Identificazione delle coppie di FT

3.2 Identificazione dei gruppi di FT

CAPITOLO IV - Descrizione delle banche dati e delle principali risorse computazionali utilizzate

4.1 Sequenze degli enhancer: VISTA enhancer browser

4.2 Sequenze dei promotori: UCSC genome browser

4.3 Modelli dei siti di legame dei fattori di trascrizione: HOCOMOCO

4.4 Predittore di siti di legame di fattori di trascrizione: MOODS

4.5 Algoritmo di biclustering: BicNET

CAPITOLO V - Organizzazione ed elaborazione dei dati

5.1 Selezione delle sequenze di enhancer e promotori

5.2 Predizione dei SLFT su enhancer e promotori

5.3 Riduzione del set di coppie enhancer-promotore

5.4 Generazione di CFT e identificazione di CFT significative

5.5 Validazione delle CFT mediante bootstrap

5.6 Ricerca delle CFT all'interno dell'interattoma umano

5.7 Analisi dell'espressione delle CFT

5.8 Generazione dei gruppi di FT

5.9 Grafo di CFT

5.10 Analisi mediante ontologie

CAPITOLO VI - Risultati e discussione

6.1 Dataset di sequenze enhancer e promotore

6.2 Predizione dei SLFT

6.3 Set di coppie enhancer - promotore

6.4 Coppie di fattori di trascrizione significative

Espressione delle CFT significative

Comparazione con l'interattoma umano

Grafo delle CFT

Annotazioni dei geni delle CFT significative

CFT: Caso di studio

6.5 Gruppi di fattori di trascrizione significativi

Annotazioni dei geni dei bicluster significativi

FT dei bicluster significativi

Bicluster: Caso di studio

6.6 Formazione del loop: Confronto tra biclustering e test binomiale

CONCLUSIONE

APPENDICE

BIBLIOGRAFIA

INTRODUZIONE

La regolazione dell'espressione è un processo di cruciale importanza nella cellula eucariotica. Questo processo è finemente regolato; nel suo controllo sono coinvolte porzioni genomiche non codificanti che contengono l'informazione necessaria al legame di fattori di trascrizione attraverso i quali modulano l'espressione dei geni regolati. I fattori di trascrizione a loro volta possono interagire con un vasto numero di proteine, tra cui cofattori o coregolatori: il risultato è un complesso proteico che catalizza l'attivazione della trascrizione attraverso il reclutamento della RNA-Polimerasi II.

Esistono diversi tipi di sequenze non codificanti coinvolte in questo processo, le quali sono ampiamente discusse nel paragrafo 1.2, ma in particolare le sequenze non codificanti protagoniste dell'attivazione della trascrizione sono enhancer e promotori (Orphanides and Reinberg, 2002). Mentre i promotori sono posti in prossimità dei geni regolati, gli enhancer possono essere anche molto distanti e come questi agiscano da tali distanze è un attuale argomento di studio e discussione. Il modello correntemente accettato per spiegare la loro azione è il modello del looping, che vede enhancer e promotori collaborare mediante la formazione di un'ansa della cromatina che pone le due sequenze di DNA vicine nello spazio, nonostante la distanza genomica che le separa (Matharu and Ahituv, 2015). L'ipotesi, largamente accettata, è che i fattori di trascrizione e le altre proteine che si legano ad enhancer e promotori possano catalizzare la formazione e stabilizzare il loop. Sebbene esistano tecniche sperimentali capaci di esaminare la struttura tridimensionale della cromatina, come le tecniche 3C ("Chromosome Conformation Capture") e derivate (Wit and Laat, 2012), sono poche le interazioni di questo tipo caratterizzate nel dettaglio (si veda il paragrafo 1.3), vale a dire sono poche le interazioni di looping per le quali si conoscono i fattori di trascrizione coinvolti nello specifico.

In questo studio si esplora la possibilità di utilizzare metodiche statistiche e computazionali volte all'identificazione di fattori di trascrizione che potrebbero avere un ruolo nella formazione del loop, in modo da complementare i casi resi noti da altre tecniche sperimentali e contribuire alla comprensione del fenomeno. L'idea in termini generali è la seguente: se enhancer e promotori possono interagire tramite un loop della cromatina, la cui formazione e stabilizzazione è mediata dal legame di fattori di trascrizione legati a queste sequenze, almeno parte dell'informazione necessaria a questa interazione deve essere presente sulle sequenze stesse. Di particolare importanza sono i siti di legame di fattori di trascrizione, la cui specificità emerge se si considerano le sequenze accoppiate di enhancer e promotori dei geni regolati.

In questo lavoro di tesi si prendono in considerazione due ipotesi limite e complementari: l'ipotesi A secondo la quale la specificità dell'interazione dipende solo da coppie di fattori di trascrizione sull'enhancer e sul promotore; l'ipotesi B secondo la quale più fattori di trascrizione sull'enhancer e sul promotore determinano il loop.

Questa nostra analisi è stata effettuata utilizzando coppie di enhancer e promotori, generate partendo da enhancer validati sperimentalmente forniti da "VISTA enhancer browser" (Visel et al., 2007) la maggiore risorsa del suo tipo, che mette a disposizione un'ampia collezione di enhancer altamente conservati nei vertebrati, importanti per lo sviluppo embrionale e validati sperimentalmente in topo transgenico. Tali enhancer sono stati

accoppiati ai promotori dei geni fiancheggiati. Questo accoppiamento è ragionevole considerando le caratteristiche di questi particolari enhancer, i quali sono attivi durante lo sviluppo embrionale e si trovano in prossimità di geni coinvolti nello stesso processo. Inoltre esistono casi in cui evidenze sperimentali suggeriscono che tali enhancer siano regolatori dei geni fiancheggiati (si veda il paragrafo 4.1).

Sui set di enhancer e promotori così ottenuti è stata effettuata la predizione dei siti di legame per i fattori di trascrizione avvalendosi di modelli per i siti di legame stessi provenienti dal database HOCOMOCO (Kulakovskiy et al., 2013), scelto per la sua completezza, alta qualità e specificità verso l'organismo umano. Gli algoritmi utilizzati per la predizione provengono da "The Motif Occurrence Detection Suite" (MOODS) (Korhonen et al., 2009; Pizzi et al., 2011), una collezione di algoritmi allo stato dell'arte per effettuare questo tipo di predizioni.

Dopo l'identificazione delle coppie e dei gruppi, i primi identificati per mezzo di un test binomiale (paragrafo 3.1) e i secondi per mezzo di algoritmi di biclustering (paragrafo 3.2), i risultati ottenuti sono stati validati utilizzando dati di espressione, annotazioni dei geni di interesse in ontologie e l'analisi degli intermedi tra le coppie di fattori di trascrizione significative.

Nelle prossime pagine prima di tutto spiegheremo i concetti fondamentali necessari alla comprensione di questo studio. Dopo una panoramica sulla corrente interpretazione della regolazione della trascrizione (capitolo 1) mostreremo la teoria relativa alle tecniche da noi utilizzate, tra cui la predizione dei siti di legame dei fattori di trascrizione (capitolo 2) e i modelli statistici utilizzati per la scoperta delle coppie e dei gruppi di siti di legame dei fattori di trascrizione (capitolo 3). Successivamente descriveremo le banche dati e le altre risorse e computazionali da noi scelte (capitolo 4) e come queste sono state utilizzate (capitolo 5). Infine esporremo e discuteremo i risultati (capitolo 6).

CAPITOLO I

L'attivazione dell'espressione genica: un sommario introduttivo

Il genoma di un organismo ne rappresenta il progetto dettagliato. I geni in esso contenuti assolvono funzioni specifiche che contribuiscono al funzionamento complessivo dell'organismo stesso. Così come le funzioni dei geni sono specifiche, altrettanto specifiche sono il momento e la localizzazione della loro espressione. Lo studio della sola funzione dei geni, infatti, sarebbe insufficiente, se non si prendesse in considerazione anche la loro regolazione. Questa è particolarmente complessa quando si considerano organismi pluricellulari come i metazoi, i quali sono composti da centinaia di tipi cellulari diversi ognuno straordinariamente specifico per un certo ruolo. Ciascuna cellula contiene gli stessi 35000 - 50000 geni, ma le loro diverse caratteristiche sono ottenute attraverso lo stretto controllo dell'espressione di un limitato e preciso sottoinsieme di geni (Orphanides and Reinberg, 2002). L'importanza della regolazione dell'espressione non si limita alla formazione di un organismo, altrettanto cruciale è la modulazione delle risposte agli stimoli che esso deve essere capace di integrare durante il suo mantenimento.

Nelle cellule eucariotiche la trascrizione si divide in tre classi, ciascuna caratterizzata dall'utilizzo di un certo tipo di RNA-Polimerasi (RNAP). La RNAP I trascrive RNA ribosomale (rRNA), RNAP II trascrive RNA messaggero (mRNA) ed infine la RNAP III trascrive RNA transfer (tRNA) ed altri piccoli RNA (Lewin, 2006). In questo testo ci focalizzeremo sullo studio della trascrizione mediata dalla RNAP II, responsabile della trascrizione dei geni codificanti proteine, i quali sono caratterizzati da una regolazione più complessa rispetto alle altre due categorie di geni.

1.1 Proteine coinvolte nell'attivazione della trascrizione: I fattori

L'inizio della trascrizione dei geni codificanti in organismi eucariotici richiede la presenza di fattori molteplici, riconducibili a tre categorie principali (Maston et al., 2006):

- I fattori basali, vale a dire gli elementi necessari e sufficienti ad attivare la trascrizione *in vitro*. Sono qui inclusi RNAP II e vari componenti ausiliari.
- Fattori di Trascrizione (FT), o talvolta chiamati attivatori, sono caratterizzati da un dominio di legame al DNA sequenza specifico e da un dominio di attivazione capace di stimolare la trascrizione.
- Cofattori, anche noti come coattivatori, sono proteine capaci di stimolare la trascrizione contattando altre proteine piuttosto che il DNA.

Questi elementi per espletare le proprie funzioni, devono essere localizzati nei pressi di particolari regioni di DNA, che fungono da siti di regolazione *cis*-agenti e che contengono i siti di legame per i fattori basali o i FT. Tali regioni includono il promotore, posto in prossimità del Sito di Inizio della Trascrizione (SIT), e siti di regolazione distali (una descrizione più dettagliata di queste regioni sarà esposta di seguito).

I fattori basali e il complesso di pre-inizio della trascrizione

L'attivazione della trascrizione richiede la formazione del Complesso di Pre-Inizio della trascrizione (CPI), che è composto dai fattori basali e che è necessario al corretto posizionamento della RNAP II al SIT. La formazione del CPI è un processo che coinvolge una moltitudine di proteine, le quali sono reclutate progressivamente seguendo un ordine preciso (Gupta et al., 2016). Nell'organismo umano, i fattori coinvolti in questo processo includono TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, TFIIH. Insieme alla RNAP II, vanno a formare un complesso di circa 2 MDa (Mega Dalton) che si posiziona sul nucleo del promotore (He et al., 2013). TFIID è il primo a legarsi al DNA, mediante la sua subunità "TATA-box Binding Protein" che riconosce la regione del nucleo del promotore nota come TATA-box (Riethoven, 2010). Successivamente altri fattori si uniscono al complesso fino al completamento del CPI (Gupta et al., 2016). La presenza del CPI è di per sé sufficiente a catalizzare la trascrizione a livello basale, ma FT e cofattori possono grandemente stimolare la trascrizione e modularla a seconda delle necessità. Una delle funzioni dei FT e dei cofattori è appunto quella di favorire la formazione del CPI contattando direttamente o indirettamente componenti dell'apparato della trascrizione (Maston et al., 2006).

I fattori di trascrizione

I fattori di trascrizione sono proteine capaci di modulare la trascrizione legandosi al DNA in siti specifici. Sono una classe di proteine differente dai fattori basali, poiché non sono né necessari né sufficienti all'attivazione della trascrizione *in vitro* (Maston et al., 2006). Tipicamente possiedono almeno due domini: uno necessario al legame col DNA sito specifico, chiamato Dominio di Legame al DNA (DLD) e un altro che conferisce al FT la capacità di modulazione della trascrizione, spesso facilitando la formazione del CPI mediante il contatto diretto o indiretto di componenti dell'apparato della trascrizione. (Maston et al., 2006; Ptashne and Gann, 1997). Possono legarsi su promotori, enhancer o altri elementi di regolazione *in cis*. Possono agire da attivatori o inibitori della trascrizione, l'una o l'altra funzione può essere assunta a seconda del contesto pertanto un FT può agire da attivatore per alcuni geni e da repressore per altri (Orphanides and Reinberg, 2002). Possono influire sulla struttura della cromatina reclutando complessi di rimodellamento (Venters and Pugh, 2009). La loro azione è altamente controllata e spesso rappresentano l'ultima tappa di molte vie di propagazione del segnale che possono modulare il legame dei FT al DNA. Possono essere regolati anche da varie modificazioni post-traduzionali come

fosforilazione, acetilazione, metilazione, ubiquitinazione (Orphanides and Reinberg, 2002; Park et al., 2015; Zhang and Reinberg, 2001).

I FT sono una classe di geni ben rappresentata all'interno del genoma umano: già le prime analisi del genoma completo di *Homo sapiens* stimavano un numero di FT compreso tra 2000 e 3000 (Lander et al., 2001; Venter et al., 2001). Una successiva analisi curata manualmente ne contava 1391 (Vaquerizas et al., 2009), mentre una stima più recente, derivante da un lavoro di catalogazione sistematica dei FT da parte di TFClass (Wingender et al., 2013) ne aumenta il numero a 1558. È probabile che questo numero, non ancora noto in maniera assoluta, continuerà a oscillare. Ad ogni modo, volendo fare una stima approssimativa considerando il numero dei geni codificanti pari a 20441 (così come riportati dalla versione GRCh38.p7 del genoma presentato da Ensembl) (Yates et al., 2016), i FT rappresentano circa il 7% dei geni codificanti.

L'aspetto più caratterizzante di un FT è il tipo di contatto che può effettuare col DNA. TFClass (Wingender et al., 2013) raggruppa 1558 FT umani in insiemi gerarchici basandosi su similarità strutturali al livello del DLD, il livello gerarchico più elevato è chiamato "superclasse" e ogni superclasse rappresenta una topologia del DLD. TFClass conta 10 superclassi, non uniformemente distribuite. Le più recenti statistiche mostrano che in uomo i tipi più comuni sono "Zinc-coordinated domain" (52%) ed "Helix-turn-helix domain" (27%), seguiti da "Basic domain" (11%), "immunoglobulin fold" (4%), "Other all- α -helical DNA binding domain" (3%), "Undefined" (1%), " β -Hairpin exposed by an α/β -scaffold" (1%), " α -Helices exposed by β -structures" (1%), " β -sheet binding to DNA" (<1%), " β -barrel DNA binding domain" (<1%) (Wingender et al., 2015).

I cofattori

Sono proteine coinvolte nella regolazione della trascrizione che non hanno un contatto diretto con il DNA. Le loro funzioni includono l'interazione con FT, il rimodellamento della cromatina, la regolazione della formazione o dell'attività del CPI. Ponendosi tra i FT e il CPI rappresentano un livello di complessità aggiuntivo per la modulazione della trascrizione alla quale contribuiscono (Roeder, 2005).

1.2 Sequenze nucleotidiche coinvolte nell'attivazione della trascrizione: Gli elementi di regolazione *cis*-agenti

Col termine "Elementi di Regolazione *Cis*-agenti" (ERC) si intendono regioni di DNA contenenti i siti di legame per le proteine coinvolte nella regolazione della trascrizione, sia che si trovino in posizioni prossime o distali dal gene regolato (Maston et al., 2006). Questo termine descrive una classe eterogenea di sequenze i cui membri possono distinguersi in cinque tipi principali, corrispondenti a cinque modi di contribuzione alla complessiva regolazione dell'espressione: nucleo del promotore, elementi prossimi al promotore, enhancer, silencer ed isolatori (Riethoven, 2010). Le proprietà di ogni tipo sono esposte nei prossimi paragrafi.

Caratteristiche tipiche degli ERC sono: 1) contengono i Siti di Legame per i Fattori di Trascrizione (SLFT) o per i fattori basali (Maston et al., 2006); 2) si possono trovare in regioni cromatiniche rilassate che esibiscono sensibilità alla DNAsi (Boyle et al., 2008); 3) possono esporre marker biochimici caratteristici dovuti alla presenza di proteine o modificazioni istoniche (Calo and Wysocka, 2013; Hardison and Taylor, 2012); 4) possono essere conservate durante l'evoluzione e di conseguenza mantenere similarità di sequenza tra specie diverse (Visel et al., 2008).

Classicamente questi elementi sono stati studiati in regioni genomiche di dimensioni limitate, sfruttando alcune delle caratteristiche sopracitate (Pennacchio and Rubin, 2001), con tecniche come gel-shift (Garner and Revzin, 1981), o saggi di ipersensibilità alla DNAsi (Galas and Schmitz, 1978) o ancora con tecniche di immunoprecipitazione (Orlando, 2000). Successivamente e con una logica simile, sono state sviluppate tecniche su larga scala, come ChiP-Seq (Chromatin ImmunoPrecipitation sequencing) (Johnson et al., 2007) e DNase-Seq (DNase I hypersensitive sites sequencing) (Boyle et al., 2008). Queste stesse caratteristiche possono venire sfruttate anche avvalendosi di tecniche computazionali, recentemente sono stati sviluppati algoritmi per la predizione di ERC che affrontano il problema in maniera varia, mediante approcci come: genomica comparativa, identificazione di cluster di SLFT e predizioni basate su marker epigenetici. Si vedano le review (Hardison and Taylor, 2012; Suryamohan and Halfon, 2015) per una completa panoramica di queste metodologie.

Nucleo del promotore

È la regione immediatamente circostante il SIT (Sito di Inizio della Trascrizione) contenente gli ERC (Elementi Regolatori *cis*-agenti) necessari al posizionamento del CPI (Complesso di Pre-Inizio), talvolta indicati come gli elementi del nucleo del promotore. L'architettura del nucleo del promotore può variare ed è definita dagli ERC contenuti, tra questi i più comuni e meglio caratterizzati sono TATA-box (sito di legame per "TATA-binding protein"), Inr ("initiator element"), DPE ("Downstream Promoter Element"), BRE ("TFIIB-Recognition Element"), DCE ("Downstream Core Element"), MTE ("Motif Ten Element") (Riethoven, 2010). Un'altra caratteristica delle sequenze dei promotori eucariotici è la frequenza di dinucleotidi CpG, e integrando queste osservazioni i promotori dei mammiferi sono stati divisi in due classi: 1) promotori a basso contenuto CpG, che definiscono il sito di inizio della trascrizione in maniera relativamente precisa, generalmente hanno un singolo SIT o un gruppo di SIT compatto e spesso contengono una TATA-box; 2) promotori ad alto contenuto CpG, ricchi di isole CpG, hanno diversi siti di inizio dispersi in una regione più vasta (Lenhard et al., 2012; Riethoven, 2010).

Questa non è una classificazione precisa, e divide i promotori in maniera sommaria, ciononostante viene usata anche oggi. Più recentemente, alla luce di nuove osservazioni, sono state distinte tre classi di promotori: 1) promotori di tipo 1, spesso contenenti TATA-box, con basso contenuto CpG, dal sito di inizio preciso e regolanti prevalentemente geni tessuto specifici; promotori di tipo 2, non associati a TATA-box, mostrano generalmente un'unica isola CpG che si sovrappone al SIT, dal sito di inizio non preciso e tipici dei geni ubiquitari; promotori di tipo 3, contenenti diverse isole CpG, più larghe, dal sito di inizio più preciso rispetto ai promotori di tipo 2 e spesso regolanti geni coinvolti nello sviluppo embrionale (Lenhard et al., 2012).

Elementi prossimi al promotore

Consistono in ERC localizzate a monte rispetto al nucleo del promotore. Contengono SLFT e si estendono per qualche centinaio di bp (Maston et al., 2006). Possono essere importanti per la collaborazione tra enhancer e promotore (Calhoun et al., 2002; Riethoven, 2010).

Enhancer

Il termine enhancer è stato coniato nel 1981 per descrivere la capacità di amplificazione della trascrizione di una sequenza di 72bp scoperta nel genoma del virus SV40 (Banerji et al., 1981). La capacità di questa sequenza di catalizzare la trascrizione indipendentemente dalla distanza o dall'orientamento rispetto a un gene reporter è successivamente diventata la definizione operativa di enhancer (Cho, 2012). La distanza tra un enhancer e un promotore regolato può essere considerevole, e nei genomi dei metazoi può misurare anche 2 - 3 Mbp (Krivega and Dean, 2012). Gli enhancer contengono vari SLFT, si quali si legano FT che a loro volta possono reclutare cofattori ed è l'attività combinata delle proteine posizionate sugli enhancer a determinarne l'attività (Shlyueva et al., 2014). La composizione dei motivi, vale a dire i SLFT presenti su un enhancer, è importante per la sua specificità, dato che determinati FT possono essere responsabili della regolazione della trascrizione specifica per determinati momenti o tipi cellulari. Il posizionamento dei motivi, invece, può essere importante per la sua dinamica di funzionamento (Spitz and Furlong, 2012). Il controllato posizionamento dei FT su un enhancer può infatti essere necessario alla formazione di un complesso proteico, o enhanceosoma, che andrà poi ad espletare la funzione dell'enhancer stesso (Merika and Thanos, 2001; Merika et al., 1998; Panne, 2008). Tra le proteine reclutate da un enhancer si nota la presenza di rimodellatori della cromatina, cosa consistente con un'altra delle proprietà degli enhancer, vale a dire la loro localizzazione in regioni cromatiniche rilassate caratterizzate da varianti istoniche come H3.3 e H2A.Z o modificazioni istoniche come H3K4me1 e H3K27ac (Calo and Wysocka, 2013).

Mentre la funzione degli enhancer è definita chiaramente, vale a dire l'amplificazione del tasso di trascrizione indipendentemente da posizione o orientamento rispetto al gene bersaglio, non altrettanto chiaro è il loro meccanismo di azione. Tra i modelli proposti per spiegarne il funzionamento due si contrappongono, pur non essendo mutualmente esclusivi, il "looping" e il "tracking". Il modello del looping pone l'enhancer in diretto contatto con il promotore del gene regolato e la distanza tra i due elementi lungo la molecola di DNA viene coperta attraverso la formazione di un'ansa o "loop" della cromatina. Il modello del tracking invece vede l'enhancer come un sito di posizionamento di proteine attivatrici capaci di spostarsi sul DNA, queste quindi si allontanerebbero dall'enhancer fino a raggiungere il promotore bersaglio (Bondarenko et al., 2003; Bulger and Groudine, 2011). Il modello del looping è stato recentemente confermato da esperimenti di cattura della conformazione della cromatina ("Chromosome Conformation Capture" o semplicemente 3C) (Dekker, 2002) e da altre tecniche derivate (Wit and Laat, 2012) come 4C ("Chromosome Conformation Capture-on-Chip") (Simonis et al., 2006), 5C ("Chromosome Conformation Capture Carbon Copy") (Dostie et al., 2006), Hi-C (Lieberman-

Aiden et al., 2009) e ChIA-PET ("Chromatin Interaction Analysis by Paired-End Tag sequencing") (Fullwood et al., 2009; Li et al., 2012).

Silencer

I silencer sono ERC che limitano o bloccano l'espressione dei geni regolati. Sono una classe di elementi regolatori relativamente poco compresa rispetto agli enhancer, di cui rappresentano la controparte (Riethoven, 2010). Dalla loro scoperta negli anni '80, diversi silencer sono stati identificati e descritti, e nel 1998 lo studio di questi elementi ha portato alla loro classificazione in due categorie (Ogbourne and Antalis, 1998): gli elementi regolatori negativi, la cui funzione di repressione è dipendente dalla posizione e dal contesto; e i silencer classici, che agiscono indipendentemente dalla loro posizione. Mentre i primi agiscono reclutando proteine che disturbano e inibiscono la formazione di complessi di attivazione della trascrizione ai promotori, i secondi inducono modificazioni della cromatina di carattere repressivo (Dean, 2011; Qi et al., 2015).

Isolatori

Sono elementi che bloccano l'influenza sulla regolazione della trascrizione da parte di altri elementi circostanti (Maston et al., 2006), hanno due proprietà fondamentali: possono bloccare l'azione di enhancer, impedendo la sua comunicazione con il promotore, e possono interferire con la diffusione di eterocromatina catalizzata da un silencer (Riethoven, 2010). Queste due proprietà hanno portato alla distinzione di altrettante classi di isolatori, gli isolatori "enhancer-blocking" e gli isolatori "barrier" rispettivamente, ma la distinzione non è netta (Dean, 2011). Gli isolatori limitano l'azione degli altri ERC in domini circoscritti (Maston et al., 2006), e un crescente numero di evidenze sperimentali suggerisce che lo possano fare definendo strutture cromatiniche tridimensionali che delimitano regioni genomiche tra le quali le interazioni nelle tre dimensioni sono frequenti (Phillips-Cremins and Corces, 2013), denominate "Topologically Associating Domains" (TAD) (Dixon et al., 2012; Nora et al., 2012) (i TAD saranno ulteriormente discussi nel paragrafo 1.3). Come le altre ERC, gli isolatori legano proteine che contribuiscono all'espletamento della loro funzione, e nei mammiferi la proteina isolatore principale è CCCTC-binding factor (CTCF) (Phillips-Cremins and Corces, 2013). Proteina definita "magistrato tessitore del genoma" (Phillips and Corces, 2009), mostra un'espressione ubiquitaria (Wendt et al., 2008) e i suoi siti di legame sono molto comuni all'interno del genoma dei mammiferi (Phillips-Cremins and Corces, 2013). È capace di mediare interazioni cromatiniche 3D (Handoko et al., 2011); questa funzione può assolverla anche in concerto con il complesso della coesina (Gause et al., 2008; Wendt et al., 2008). Per questi motivi si suppone che CTCF abbia un ruolo fondamentale per la configurazione dei domini genomici (Dean, 2011; Phillips-Cremins and Corces, 2013).

1.3 Comunicazione enhancer-promotore: il looping della cromatina

Correntemente, il modello più studiato e confermato per spiegare l'interazione tra enhancer e promotori è quello del "looping" della cromatina (Matharu and Ahituv, 2015). In quest'ottica un enhancer si posiziona vicino al promotore regolato attraverso la formazione di un'ansa o "loop" della cromatina. La prima evidenza concreta a supporto di questo modello è stata mostrata nello studio della "Locus Control Region" (LCR) della β -globina (Dean, 2011; Tolhuis et al., 2002). Successivamente questo modello è stato confermato dallo sviluppo di tecniche di cattura della conformazione cromosomica, anche dette 3C (Dekker, 2002), e derivate (Wit and Laats, 2012), le quali rappresentano gli strumenti più utilizzati per lo studio delle interazioni cromatiniche. Queste tecniche sperimentali in un primo momento fissano la struttura tridimensionale della cromatina utilizzando un fissante, che tipicamente è formaldeide, poi la cromatina viene esposta a enzimi di restrizione ed infine le estremità di DNA tagliate sono nuovamente legate in maniera da favorire il legame tra porzioni di DNA che sono poste in prossimità nello spazio tridimensionale (Wit and Laats, 2012). L'evoluzione di queste tecniche ha portato alla possibilità di investigare le interazioni cromatiniche su scala genomica mediante l'uso di tecniche Hi-C (Lieberman-Aiden et al., 2009) o sue derivazioni come ChIA-PET (Fullwood et al., 2009; Li et al., 2012), confermando l'esistenza del fenomeno del looping su larga scala.

È ipotesi correntemente accettata che le proteine che si legano a enhancer e promotori siano coinvolte nella formazione e stabilizzazione del loop, ma non sono molte le interazioni ben caratterizzate di questo tipo. Un esempio molto studiato che convalida questa ipotesi è il loop tra il gene della β -globina e la rispettiva LCR. È stato mostrato che il loop tra queste due regioni richiede la presenza di FT specifici degli eritrociti come EKLF ("Krueppel-like factor 1") e GATA1 ("GATA-binding factor 1") ed è mediato da cofattori come LDB1 ("LIM domain-binding protein 1") (Dean, 2011). Questi studi hanno portato a modelli per la formazione del loop in questa regione che prevedono la formazione di un complesso proteico formato da FT e cofattori che congiunge gli elementi regolatori distali (Dean, 2011; Maksimenko and Georgiev, 2014). Le osservazioni supportano l'idea che la formazione di un loop tra enhancer e promotore può dipendere dalla compatibilità biochimica dei FT legati alle due estremità dell'ansa (Matharu and Ahituv, 2015).

Esistono altri esempi di FT coinvolti in interazioni simili a queste; sempre nel locus della β -globina ad esempio è stata osservata la presenza di CTCF in qualità di mediatore di loop cromatinici (Splinter et al., 2006). CTCF è un FT la cui attività di formazione di loop è particolarmente studiata, e la cui attività sembra molto importante per la formazione di domini genomici come si è parlato nella sezione dedicata agli isolatori del paragrafo 1.2.

Un altro esempio è SP1 ("Transcription Factor SP1"), un FT ubiquitario i cui siti di legame sono posizionati su numerosi promotori sia di geni ubiquitari che tessuto specifici (Zhao and Meng, 2005), può modulare la trascrizione legandosi a enhancer distali, può congiungere segmenti di DNA mediante la formazione di omopolimeri (Mastrangelo et al., 1991; Su et al., 1991) e la sua attività di mediatore tra enhancer e promotore è stata osservata nella regolazione del gene HO-1 ("Heme Oxygenase-1") dell'organismo umano (Deshane et al., 2010). SP1 può anche reclutare il coattivatore EP300 ("Histone acetyltransferase p300") per attivare la trascrizione (Li and Davie, 2010; Suzuki et al., 2000).

EP300, e il suo paralogo CREBBP ("CREB-binding protein"), sono altre due proteine che potrebbero essere coinvolte in interazioni di looping; entrambi sono considerati adattatori della trascrizione capaci di agire da ponte di intermediazione tra FT, fattori basali ed RNAP II (Holmqvist and Mannervik, 2013). Questa idea è stata supportata dalla risoluzione di una struttura tridimensionale che mostra EP300 legata a tre fattori di trascrizione di tipo MEF2 ("Myocyte-specific enhancer factor") a loro volta legati al DNA, che ha portato a postulare il suo possibile ruolo di intermedio di interazioni tra segmenti di DNA distali (He et al., 2011). Sia EP300 che CREBBP sono tra i cofattori della trascrizione più comuni nei metazoi (Holmqvist and Mannervik, 2013) e in particolare EP300 è così frequentemente localizzato in regioni regolatrici *cis*-agenti che la sua presenza è stata utilizzata per identificare enhancer attivi (Visel et al., 2009).

Il complesso noto come "Mediator" è un'altro esempio di "ponte" che può mettere in comunicazione enhancer e promotori. Questo complesso è composto da 26 subunità nei mammiferi e può prendere contatto con vari FT (anche nello stesso momento) ed altre proteine coinvolte nella regolazione della trascrizione tra cui fattori basali ed RNAP II (Allen and Taatjes, 2015). Inizialmente il Mediator è stato descritto proprio in base a questa sua capacità di mediare tra FT, fattori basali ed RNAP II (Carlsten et al., 2013). Come questo regoli l'attività della RNAP II non è ancora chiaro (Allen and Taatjes, 2015); ciononostante è considerato un componente essenziale per la trasduzione dei segnali biologici comunicati dai FT nella risposta fisiologica (Poss et al., 2013). È stata osservata la presenza del Mediator su enhancer di cellule staminali embrionali (Plank and Dean, 2014) ed un numero crescente di evidenze suggeriscono che sia coinvolto nel looping tra enhancer e promotori (Allen and Taatjes, 2015; Carlsten et al., 2013).

Gli esempi e le evidenze fino ad ora citate mostrano che sono noti diversi esempi di proteine che possono mettere in comunicazione enhancer e promotori attraverso un loop. Rimane però aperta la questione su come enhancer e promotori distanti possano congiungersi correttamente all'interno dello spazio nucleare. I più recenti studi ad alta resa, effettuati con le tecniche di cattura della conformazione cromosomica già citate sopra, pongono l'interazione enhancer-promotore all'interno del contesto genomico e l'immagine che sta emergendo è quella di enhancer e promotori di geni regolati collocati all'interno degli stessi domini genomici o "Topological Associated Domains" (TAD) (Gibcus and Dekker, 2013; Plank and Dean, 2014). Questi domini, riconosciuti utilizzando in particolare tecniche Hi-C e 5C in genomi di uomo, topo e *Drosophila* (Dixon et al., 2012; Gibcus and Dekker, 2013; Nora et al., 2012; Sexton et al., 2012), definiscono zone dentro le quali sono osservabili un alto numero interazioni a lungo raggio tra i loci che ne fanno parte. Gli enhancer e i promotori dei geni pertinenti appaiono raggruppati all'interno dei TAD (Plank and Dean, 2014). Sebbene cosa definisca i confini di un TAD non sia del tutto chiaro, su tali confini è particolarmente presente la proteina CTCF, il che è conforme al ruolo strutturale riconosciuto in questa proteina (Gibcus and Dekker, 2013). Anche coesina e Mediator sono stati proposti come possibili elementi rilevanti per l'organizzazione dei TAD (Plank and Dean, 2014), sebbene contribuiscano in maniera diversa alla loro formazione. Evidenze sperimentali suggeriscono che CTCF sia implicato nella definizione dei confini dei TAD mentre il complesso della coesina sia implicato nella definizione di interazioni all'interno di tali domini. Infatti una comparazione degli effetti sulle interazioni cromatiniche a lungo raggio (effettuata nella linea cellulare HEK293T) dovuti alla privazione di RAD21 ("Double-strand-break repair protein rad21 homolog" una subunità del complesso della coesina) e di CTCF mostrano nel primo caso riduzioni delle

interazioni all'interno dei TAD, senza però alterarne l'architettura generale; nel secondo caso invece si sono osservate sia riduzioni di interazioni all'interno dei domini, sia un aumento di interazioni tra domini vicini (Zuin et al., 2014). Anche il Mediator sembra essere coinvolto nelle interazioni all'interno dei domini come osservato in uno studio effettuato su una popolazione di cellule NPC ("multipotent Neural Progenitor Cells") derivate da cellule staminali embrionali ("ES cell") dove è stato riconosciuto il coinvolgimento del complesso Mediator e di quello della coesina nella definizione di interazioni a corto raggio tra enhancer e promotori specifici delle cellule ES, nelle quali potrebbero essere coinvolti anche FT pionieri (Phillips-Cremins et al., 2013).

Recenti sforzi di integrazione di evidenze derivanti da tecniche di microscopia e sequenziamento del DNA tentano di mettere a fuoco questi ed altri fenomeni in un contesto ancora più ampio e mirano ad una definizione della struttura nucleare sia nello spazio che nel tempo in quello che è definito come "4D nucleome" (Cremer et al., 2015). Questa visione ancora in via di caratterizzazione rappresenta un tentativo di integrare i domini cromatinici e i componenti del nucleo in un unico modello integrato. Il nucleo è visto come un'entità dinamica la cui funzione ed attività è fortemente determinata dalla sua struttura e vengono distinti due compartimenti fondamentali: uno attivo detto ANC ("Active Nuclear Compartment") composto da cromatina nello stato rilassato, ed uno inattivo chiamato INC ("Inactive Nuclear Compartment") composto da cromatina condensata. ANC è a sua volta composto da IC ("Interchromatin Compartment") e PR ("Perichromatin Region"). Il primo costituisce una rete tridimensionale che si ramifica all'interno di INC e sfocia ai complessi del poro nucleare. Il secondo costituisce i confini del primo ed è il luogo dove è localizzata la cromatina attiva, incluso i loop attivi nella trascrizione.

Nonostante i meccanismi qui elencati non siano ancora stati descritti nel dettaglio, e i modelli ancora in fase di definizione, è sempre più evidente che il looping della cromatina rappresenta un fenomeno verificato, di estrema importanza, che non si limita alle interazioni enhancer-promotore. Le strutture genomiche tridimensionali che può mediare danno luogo a regioni nucleari chiamate "transcription factories" (Iborra et al., 1996; Li et al., 2012; Weipoltshammer and Schöfer, 2016) all'interno delle quali avviene la trascrizione dei geni co-espressi e che sono state comparate agli operoni (Allen and Taatjes, 2015; Li et al., 2012).

Nel corso degli ultimi anni sono stati fatti enormi passi avanti nelle tecniche dedicate alla comprensione della struttura della cromatina e assistiamo al loro continuo sviluppo. Molto recente è lo sviluppo della nuova tecnica di mappatura dell'architettura genomica chiamata GAM (Genome Architecture Mapping). Qui le cellule sono fissate, congelate, criosezionate e i profili nucleari delle criosezioni sono isolati, amplificati e sequenziati. Questa nuova tecnica ha permesso l'identificazione di numerose interazioni tra enhancer e geni attivi in cellule staminali embrionali di topo (Beagrie et al., 2017) confermando ancora una volta la presenza e l'importanza su larga scala dei loop enhancer-promotore.

CAPITOLO II

Siti di legame per i fattori di trascrizione: rappresentazione e predizione

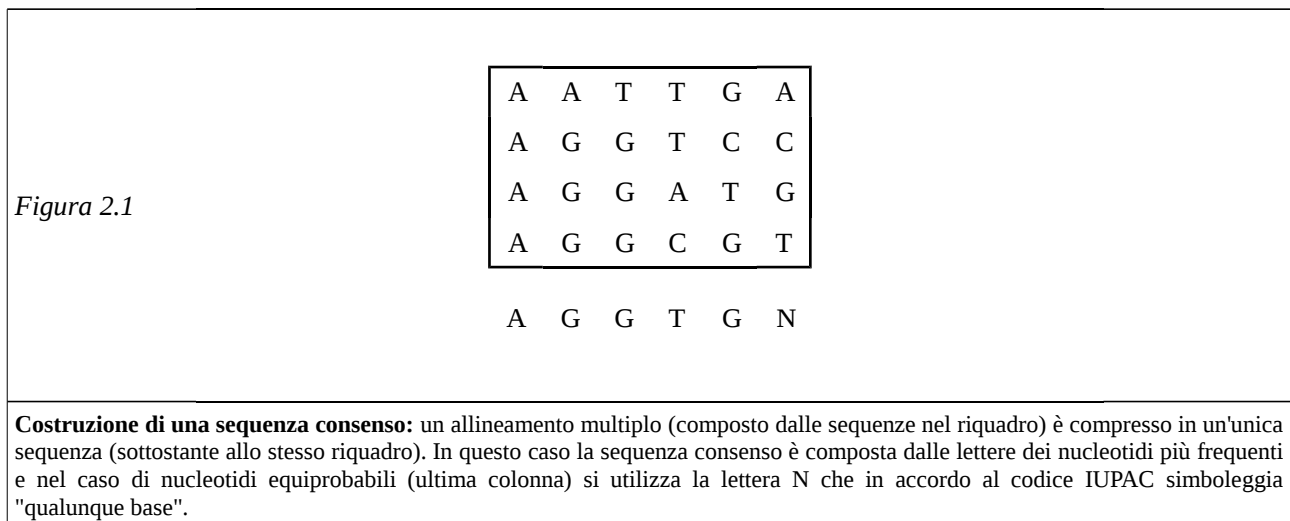
I "Siti di Legame per i Fattori di Trascrizione" (SLFT), chiamati anche semplicemente "motivi", sono le unità funzionali degli elementi di regolazione *cis*-agenti, vale a dire i siti sul DNA dove si legano i Fattori di Trascrizione (FT). Sono tipicamente lunghi dalle 6 alle 12 bp ma le sequenze sono generalmente variabili e la specificità del legame è definita principalmente da 4-6 bp (Maston et al., 2006). Questa variabilità si evidenzia quando si allineano diverse sequenze di siti di legame per lo stesso FT. Dato un insieme di SLFT noti è possibile sviluppare un modello capace di descrivere le proprietà delle sequenze di legame di un particolare FT e questo modello può essere successivamente usato per predire SLFT presenti su un genoma (Wasserman and Sandelin, 2004). Le sequenze di partenza provengono tipicamente da evidenze sperimentali o computazionali di vario tipo. Numerose tecniche sono state sviluppate per generare dati relativi all'interazione DNA-proteine *in vitro*: gel-shift (Garner and Revzin, 1981), saggi di ipersensibilità alla DNAsi (Galas and Schmitz, 1978), immunoprecipitazione (Orlando, 2000). Quest'ultima tecnica è stata anche adattata per esperimenti su larga scala combinandola con microarray (Ren et al., 2000) e sequencing (ChIP-Seq) (Johnson et al., 2007). Tra le tecniche che fanno uso di microarray menzioniamo anche "Protein Binding Microarray" (Mukherjee et al., 2004). Infine la tecnologia del sequencing è stata adottata anche a saggi di ipersensibilità alla DNAsi (Boyle et al., 2008) e alla tecnica SELEX ("Systematic Evolution of Ligands by Exponential Enrichment") (Pollock and Treisman, 1990) diventando SELEX-seq (Jolma et al., 2010).

2.1 Rappresentazione dei motivi

Le metodologie sviluppate per rappresentare e modellare correttamente i motivi in maniera da descrivere la loro variabilità intrinseca sono molteplici e possono essere divise in testuali, matematiche e grafiche (Suryamohan and Halfon, 2015; Wasserman and Sandelin, 2004). Nel primo caso si parla di sequenze consenso, nel secondo si utilizzano matrici per costruire un modello matematico volto alla corretta descrizione dei siti di legame, nel terzo si utilizzano delle rappresentazioni grafiche dette "logos".

Rappresentazione testuale: le sequenze consenso

Il metodo più semplice è quello testuale, che consiste nella produzione di una sequenza consenso. Esistono diversi modi per generare una sequenza consenso ma in generale la procedura genera una stringa capace di riassumere la distribuzione dei simboli in un allineamento di sequenze nucleotidiche (Day and McMorris, 1992). L'alfabeto utilizzato per scrivere una sequenza consenso è scelto in base al tipo di sequenza in esame, nel caso di sequenze di DNA ogni nucleotide è rappresentato dalla lettera iniziale del nome della sua base azotata, di conseguenza l'alfabeto conta quattro lettere: A per adenina, C per citosina, G per guanina e T per timina. Dopo aver allineato le sequenze di interesse, le frequenze di ogni nucleotide in ogni posizione vengono calcolate, e il nucleotide più rappresentativo di quella posizione viene scritto nella stessa posizione della sequenza consenso. Nel caso in cui due o più nucleotidi abbiano lo stesso peso in una data posizione, eventualità probabile nelle posizioni meno specifiche dei SLFT, si possono anche indicare le posizioni variabili o elencando tutti i nucleotidi possibili, o utilizzando il codice nucleotidico IUPAC ("International Union of Pure and Applied Chemistry") (Cornish-Bowden, 1985; Stormo, 2000) dove le combinazioni di diversi nucleotidi sono rappresentate da una singola lettera. Questa metodologia è stata utilizzata nelle prime raccolte di SLFT (Faisst and Meyer, 1992). L'effetto collaterale della semplicità di questo metodo è l'imprecisione: una sequenza consenso non descrive precisamente ed univocamente le preferenze relative dei nucleotidi nelle posizioni variabili, con conseguente perdita di informazione (Day and McMorris, 1992; Wasserman and Sandelin, 2004). D'altro canto, una sequenza consenso rappresenta un SLFT meglio della sequenza di un singolo sito di legame, e permette una veloce e comoda visualizzazione delle caratteristiche generali di un SLFT (Wasserman and Sandelin, 2004). Un semplice esempio di sequenza consenso è mostrato in *Figura 2.1*.

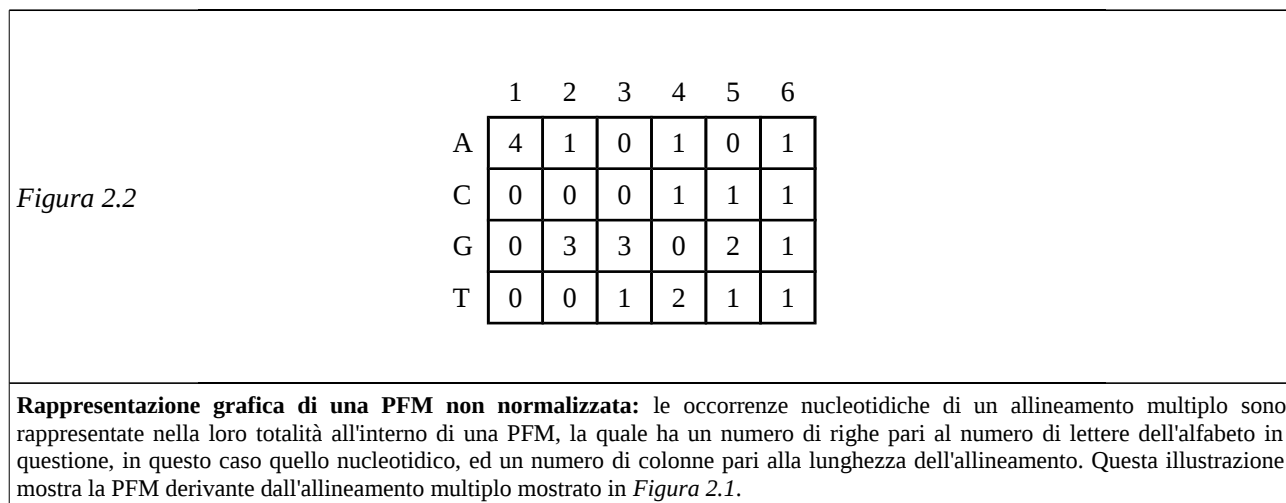


Rappresentazione matematica: le matrici di posizione

Per rappresentare matematicamente un SLFT tipicamente si utilizzano matrici posizione-specifiche le quali possono differenziarsi a seconda dell'algoritmo utilizzato per computarle. Queste generalmente hanno un numero

di righe pari al numero di lettere dell'alfabeto utilizzato e un numero di colonne pari alla lunghezza dell'allineamento. Permettono una rappresentazione quantitativa di un SLFT e a differenza delle sequenze consenso descrivono in maniera esaustiva le preferenze di tutti i nucleotidi, anche nel caso di posizioni degenerate (Wasserman and Sandelin, 2004). Le matrici di posizione più frequentemente utilizzate per rappresentare i SLFT sono le "Position Frequency Matrices" (PFM), le quali possono essere normalizzate o non normalizzate. Molto comuni sono anche le "Position Weight Matrices" (PWM), che derivano dalle PFM ma il cui utilizzo è particolarmente idoneo alla predizione di SLFT, e per tale ragione verranno discusse dettagliatamente nel paragrafo 2.2, dedicato appunto alla predizione.

Le PFM non normalizzate sono il tipo più semplice di matrice di posizione. Per computarle si parte da un allineamento multiplo di sequenze di SLFT noti e si calcolano le occorrenze di ogni nucleotide in ogni posizione dell'allineamento. Sono queste occorrenze che vanno a costituire la PFM, la quale nel caso di sequenze nucleotidiche avrà quattro righe, e come già anticipato un numero di colonne pari alla lunghezza dell'allineamento. Queste matrici sono talvolta chiamate matrici di occorrenza per differenziarle dalle PFM normalizzate contenenti invece le frequenze normalizzate dei nucleotidi (Wasserman and Sandelin, 2004). Una rappresentazione grafica di PFM non normalizzata è illustrata in *Figura 2.2*, calcolata partendo dall'allineamento di *Figura 2.1*.



Una PFM normalizzata deriva da una PFM non normalizzata, ma in questo caso gli elementi della matrice sono le frequenze nucleotidiche normalizzate per il numero di sequenze dell'allineamento multiplo di origine, che è pari alla somma delle lettere presenti in ogni colonna della PFM non normalizzata. Le frequenze sono quindi calcolate come illustrato in *Formula 2.1*, dove $f_{i,j}$ è la frequenza del nucleotide i in posizione j , $n_{i,j}$ è il numero di occorrenze del nucleotide i in posizione j , infine A è il numero di lettere dell'alfabeto, in caso di sequenze nucleotidiche $A = 4$.

Formula 2.1

$$f_{i,j} = \frac{n_{i,j}}{\sum_{i=1}^A n_{i,j}}$$

Rappresentazione grafica: sequence logos

Si tratta di loghi sviluppati per rappresentare graficamente insiemi di sequenze allineate. Per generare un logo le sequenze vengono prima allineate, poi vengono calcolate le frequenze di ogni nucleotide in ogni posizione, generando una tabella delle frequenze. Le frequenze in ogni posizione sono esaminate e ordinate portando le lettere più frequenti in cima. Infine viene calcolata la quantità di informazione presente in ogni posizione dell'allineamento (Schneider and Stephens, 1990). Tutti questi dati sono rappresentati nel logo, che mostra contemporaneamente: 1) l'ordine di predominanza di ogni elemento in ogni posizione, dato che tutte le lettere in ogni posizione dell'allineamento sono esposte nel logo dall'alto verso il basso in ordine decrescente per la frequenza; 2) il consenso delle sequenze, che è composto dalle lettere posizionate più in alto; 3) la frequenza relativa di ogni elemento in ogni posizione, che nel logo è rappresentata dall'altezza relativa di ogni lettera; 4) il contenuto in termini di informazione presente in ogni posizione espresso in bit, che nella rappresentazione grafica equivale all'altezza assoluta di tutte le lettere presenti nella stessa colonna (Schneider and Stephens, 1990). *Figura 2.3* mostra un esempio di logo generato dalle stesse sequenze che compongono l'allineamento multiplo mostrato in *Figura 2.1*.



L'informazione di ogni posizione dell'allineamento, è calcolata utilizzando *Formula 2.2*, dove R_j è il contenuto in termini di informazione in posizione j , $f_{i,j}$ è la frequenza del nucleotide i in posizione j , A il numero di lettere nell'alfabeto (Schneider and Stephens, 1990; Wasserman and Sandelin, 2004).

Formula 2.2

$$R_j = 2 + \sum_{i=1}^A f_{i,j} \log_2 f_{i,j}$$

Questa misura dell'informazione deriva dalla teoria della comunicazione di Shannon (Shannon, 1948), ed un guadagno in informazione è definito come una perdita di incertezza, definita come mostrato in *Formula 2.3*, dove H_j è l'incertezza in posizione j , $f_{i,j}$ è la frequenza del nucleotide i in posizione j , A il numero di lettere nell'alfabeto (Schneider et al., 1986).

Formula 2.3

$$H_j = - \sum_{i=1}^A f_{i,j} \log_2 f_{i,j}$$

Nel caso di una posizione in un allineamento contenente tutti i nucleotidi con la stessa probabilità, H_j assume il suo valore massimo pari a 2. Questa incertezza massima, si sottrae poi a quella calcolata usando la stessa *Formula 2.3* per la posizione del multi-allineamento in esame. La risultante perdita di incertezza è così interpretata un guadagno in informazione (Schneider et al., 1986), la quale è perciò calcolata come mostrato in *Formula 2.2*.

2.2 Matrici di posizione e predizione di siti di legame di fattori di trascrizione

Una matrice di posizione, essendo una rappresentazione quantitativa calcolata integrando una moltitudine di siti di legame, è effettivamente un modello del sito di legame che rappresenta e può essere utilizzata per identificare nuovi potenziali siti di legame diversi da quelli utilizzati per la generazione della matrice stessa (Wasserman and Sandelin, 2004). Nella pratica per effettuare questo tipo di predizioni si utilizzano delle matrici di posizione adatte allo scopo chiamate "Position Weight Matrices" o "PWM" per brevità (Hertz and Stormo, 1999).

Position Weight Matrices

Le PWM sono simili alle PFM dalle quali derivano, ma in questo caso gli elementi della matrice non sono frequenze bensì pesi. Un peso è il logaritmo del rapporto tra la frequenza osservata e la frequenza attesa dello stesso nucleotide. Questo concetto è mostrato in *Formula 2.4*, dove $w_{i,j}$ è il peso del nucleotide i in posizione j , $f_{i,j}$ è la frequenza del nucleotide i in posizione j e p_i è la frequenza attesa del nucleotide i (anche chiamata probabilità a priori o "background"). La probabilità a priori di un nucleotide può essere la sua frequenza nell'intero genoma dell'organismo studiato, o la sua frequenza all'interno di un sottoinsieme di sequenze genomiche (Hertz and Stormo, 1999).

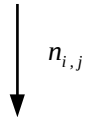
Formula 2.4

$$w_{i,j} = \ln \frac{f_{i,j}}{p_i}$$

Figura 2.4

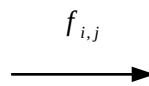
1) Allineamento multiplo

A	A	T	T	G	A
A	G	G	T	C	C
A	G	G	A	T	G
A	G	G	C	G	T



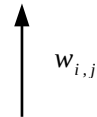
2) PFM (Occorrenze)

	1	2	3	4	5	6
A	4	1	0	1	0	1
C	0	0	0	1	1	1
G	0	3	3	0	2	1
T	0	0	1	2	1	1



4) PWM

	1	2	3	4	5	6
A	1,2	0	-1,6	0	-1,6	0
C	-1,6	-1,6	-1,6	0	0	0
G	-1,6	0,96	0,96	-1,6	0,59	0
T	-1,6	-1,6	0	0,59	0	0



3) PFM (Frequenze)

	1	2	3	4	5	6
A	0,85	0,25	0,05	0,25	0,05	0,25
C	0,05	0,05	0,05	0,25	0,25	0,25
G	0,05	0,65	0,65	0,05	0,45	0,25
T	0,05	0,05	0,25	0,45	0,25	0,25

Computazione di matrici di posizione: partendo da un allineamento multiplo (1) una matrice di occorrenza (2) viene generata inserendo in ogni elemento della matrice le occorrenze di ogni nucleotide ($n_{i,j}$). La matrice di occorrenza è la forma più semplice di "Position Frequency Matrix" o PFM. Una PFM può in alternativa contenere le frequenze normalizzate ($f_{i,j}$) (3). Le frequenze sono calcolate con pseudo-count come descritto in *Formula 2.5*, ponendo $k = 1$ e $p_i = 0.25$ per tutti gli i . Queste frequenze sono poi utilizzate per generare una "Position Weight Matrix" o PWM (4), dove in ogni elemento della matrice sono inseriti i pesi ($w_{i,j}$) calcolati applicando *Formula 2.4*. Nelle rappresentazioni grafiche delle matrici (sezioni 2 - 4), le righe indicano le lettere dell'alfabeto utilizzato (A, C, G, T), le colonne la posizione nell'allineamento multiplo (1 - 6).

La definizione di $f_{i,j}$ così come mostrata in *Formula 2.1* è però inappropriata per la risoluzione di *Formula 2.4* poiché nel caso in cui un nucleotide fosse completamente assente in una certa posizione dell'allineamento di partenza il rispettivo elemento della PWM sarebbe pari al logaritmo di zero (eventualità non remota data la scarsità di sequenze note per alcuni di SLFT). Per questa ragione nel calcolo di una PWM, la definizione di $f_{i,j}$ viene modificata aggiungendo uno "pseudo-count" il cui scopo è quello di prevenire l'esistenza di elementi pari a zero (Hertz and Stormo, 1999; Wasserman and Sandelin, 2004). Un esempio è mostrato in *Formula 2.5*, dove $n_{i,j}$ è il numero di occorrenze del residuo i in posizione j , p_i è la frequenza attesa del nucleotide i , k è lo pseudo-count e A è il numero di lettere dell'alfabeto in considerazione.

Formula 2.5

$$f_{i,j} = \frac{n_{i,j} + p_i k}{\sum_{i=1}^A n_{i,j} + k}$$

Sebbene le PWM siano state per la prima volta sviluppate per studiare i siti di inizio della traduzione su sequenze di mRNA in *E. Coli* (Stormo et al., 1982), sono oggi molto utilizzate per studiare i siti di legame sul DNA e per la predizione di SLFT in genomi eucariotici. Non è raro infatti che collezioni curate di SLFT noti siano distribuite sotto forma di PFM (Kulakovskiy et al., 2013; Sandelin et al., 2004), in modo da permettere una semplice conversione in PWM utilizzando il background più idoneo al caso.

La procedura di generazione di una PWM partendo da un allineamento multiplo, completa di stadi intermedi di computazione delle PFM è illustrata schematicamente in *Figura 2.4*.

Predizione di siti di legame di fattori di trascrizione mediante PWM

Le PWM sono matrici particolarmente convenienti per la predizione di SLFT, e il problema dell'individuazione di SLFT su una sequenza nucleotidica può essere descritto nel modo seguente: data una PWM chiamata M e una sequenza nucleotidica S (generalmente molto più lunga di M), si vogliono identificare tutte le regioni su S che ben corrispondono alle frequenze nucleotidiche date dal modello M . Per farlo si può allineare M contro tutti i segmenti di S di lunghezza uguale alla lunghezza di M , che chiameremo s , assegnando a ogni segmento s un punteggio che descrive la qualità del "match". I segmenti s con il punteggio migliore sono le regioni di S che ben corrispondono al modello M , vale a dire potenziali SLFT (Pizzi et al., 2011). Il punteggio è uguale alla somma degli elementi di M corrispondenti ai nucleotidi osservati nella sequenza s in esame. Questo calcolo è mostrato in *Formula 2.6*, dove W è il punteggio che descrive la qualità del match, $w_{i,j}$ è il peso del nucleotide i in posizione j , ed m è la lunghezza della matrice, che è anche pari alla lunghezza del segmento s .

Formula 2.6

$$W = \sum_{j=1}^m w_{i,j}$$

Utilizzare un punteggio così definito ha diversi vantaggi. In primo luogo il risultato W viene calcolato tenendo conto sia delle frequenze nucleotidiche del modello M , sia delle frequenze nucleotidiche attese, infatti la definizione di un peso w è equivalente ad un log odd ratio tra due ipotesi: 1) la sequenza osservata s mostra la presenza di un segnale; 2) la stessa sequenza s è una successione casuale di nucleotidi (Claverie and Audic, 1996). Questa equivalenza è mostrata molto chiaramente in *Formula 2.7*, dove il punteggio W equivale al log odd ratio tra la probabilità della sequenza s dato il modello M e la probabilità della stessa sequenza s dato il modello nullo N . La frequenza $f_{i,j}$ del nucleotide i in posizione j è la stessa frequenza definita in *Formula 2.5*. p_i e $w_{i,j}$ sono rispettivamente frequenza attesa del nucleotide i e il peso del nucleotide i in posizione j entrambi analoghi agli stessi valori già mostrati in *Formula 2.4*.

Formula 2.7

$$W = \ln \left(\frac{P(s|M)}{P(s|N)} \right) = \ln \left(\frac{\prod_{j=1}^m f_{i,j}}{\prod_{j=1}^m p_i} \right) = \sum_{j=1}^m \ln \left(\frac{f_{i,j}}{p_i} \right) = \sum_{j=1}^m w_{i,j}$$

Un altro vantaggio è che trattandosi di valori logaritmici, il punteggio W relativo a una intera sequenza s viene calcolato mediante le semplici somme dei pesi $w_{i,j}$ associati ad ogni singolo nucleotide di s . Se così non fosse, le frequenze $f_{i,j}$ e p_i di ogni nucleotide dovrebbero essere moltiplicate tra loro, rendendo il calcolo meno vantaggioso da un punto di vista computazionale. Si tenga infine in considerazione che nel calcolo del punteggio W qui esposto si accetta l'assunzione che le occorrenze di nucleotidi consecutivi siano eventi indipendenti. Sebbene questa assunzione possa essere contestabile da un punto di vista strettamente biologico esperimenti utilizzando modelli probabilistici più complessi non si sono mostrati particolarmente vantaggiosi da un punto di vista matematico, e per questo quella descritta rimane la procedura canonica più utilizzata (Benos et al., 2002; Wasserman and Sandelin, 2004).

Significatività statistica dei siti di legame predetti

Una PWM di buona qualità è capace di associare punteggi (o "score") alti a SLFT reali e punteggi bassi a sequenze prive di tale funzionalità. Di conseguenza le predizioni dei siti di legame possono essere accettate o meno definendo una punteggio di soglia: se l'allineamento del segmento s con la matrice M ritorna un punteggio superiore o uguale alla soglia la predizione viene accettata, in caso contrario viene scartata. Questa semplice procedura ha però i suoi limiti: 1) il punteggio W dipende dalla lunghezza e dalla composizione della PWM allineata (Medina-Rivera et al., 2011; Pan and Phan, 2009) e di conseguenza i punteggi ottenuti da PWM diverse non sono comparabili; 2) la scelta di un livello di soglia è arbitraria e deve essere basata sul problema biologico.

Il primo limite sopra citato è particolarmente rilevante quando si vogliono predire contemporaneamente SLFT per FT diversi e si desidera accettare solo previsioni che abbiano la stessa qualità, in maniera ininfluenza dal FT considerato. Questo problema è stato risolto assegnando una significatività statistica ai punteggi (Staden, 1989). La distribuzione dei punteggi che possono risultare da una data PWM viene calcolata, in modo da rendere possibile la definizione di una soglia in termini di p-value e non di punteggio. In questa maniera si possono accettare o rifiutare previsioni di SLFT mantenendo per tutti i FT lo stesso livello di significatività statistica. Per un dato p-value, il punteggio soglia corrispondente t è un valore tale per cui la probabilità di avere punteggi superiori o uguali a t è pari al p-value stesso (Pizzi et al., 2011). In altri termini, il p-value è la probabilità di ottenere un punteggio superiore o uguale a t allineando la matrice a un segmento nucleotidico casuale (Wu et al., 2000).

Per scegliere una soglia adeguata bisogna considerare che le predizioni di SLFT basate su PWM sono caratterizzate da un alto numero di falsi positivi, vale a dire predizioni per siti di legame che non sono effettivamente funzionali *in vivo* (Wasserman and Sandelin, 2004). Questo fenomeno però non è dovuto ad una

inadeguatezza del modello del SLFT così come definito dalla PWM, piuttosto è dovuto alla presenza nei genomi di numerosi segmenti di DNA composti da sequenze nucleotidiche che potrebbero essere potenziali SLFT ma che non espletano tale funzione per motivi diversi dalla loro composizione nucleotidica. Infatti è stato mostrato sia che sequenze predette come SLFT possono legare i FT in vitro (Tronche et al., 1997), sia che esiste una relazione diretta tra il punteggio generato dalle PWM e l'affinità DNA-proteina (Berg and von Hippel, 1987; Stormo, 2000). Per ovviare a tali inconvenienti è buona norma utilizzare livelli di soglia particolarmente stringenti. Inoltre limitare la ricerca di SLFT su sequenze nucleotidiche di cui l'attività di regolazione della trascrizione è nota aiuta a mitigare il numero di falsi positivi (Wasserman and Sandelin, 2004).

CAPITOLO III

Identificazione di coppie e gruppi di fattori di trascrizione su coppie di enhancer e promotori di geni regolati

Precedentemente sono state descritte le caratteristiche delle sequenze di DNA responsabili della regolazione della trascrizione e la loro classificazione. Tra queste abbiamo menzionato gli enhancer e i promotori, che sono le sequenze regolatrici sulle quali questo lavoro si è concentrato e delle quali discuteremo successivamente. Abbiamo descritto la presenza di siti di legame di fattori di trascrizione sia sugli enhancer che sui promotori, e di come tali sedi di ancoraggio siano di primaria importanza per la funzione di tali sequenze nucleotidiche, le quali agiscono in concerto durante la regolazione dell'espressione dei geni regolati. Abbiamo anche mostrato quali sono le ipotesi e i modelli proposti e correntemente accettati per comprendere le dinamiche della loro cooperazione, in particolare il modello del looping, il quale pone enhancer e promotori in una stretta relazione spaziale mediante la formazione di un'ansa, o loop, della cromatina, e come la formazione dell'ansa possa essere mediata e stabilizzata dai Fattori di Trascrizione (FT) e da altre proteine o complessi proteici. Sia nei casi in cui si ipotizza che i FT possano interagire direttamente tra loro per la formazione del loop, (come ad esempio nei casi già citati di CTCF o SP1) o nei casi in cui si suppone la presenza di proteine intermedie (come nel caso di EP300 o Mediator), i FT presenti su enhancer e su promotori sono necessari alla formazione di queste interazioni, qualunque sia il numero o la natura delle proteine intermedie. Per questi motivi è ragionevole supporre che almeno parte dell'informazione necessaria alla corretta interazione enhancer-promotore sia localizzata sulle stesse sequenze di DNA regolatrici sotto forma di Siti di Legame per Fattori di Trascrizione (SLFT). In accordo con questa ipotesi, è possibile correlare l'interazione enhancer-promotore ai fattori di trascrizione che si legano su queste sequenze; di conseguenza lo studio dei SLFT (la cui presenza può essere predetta come esposto nel capitolo precedente) sia sugli enhancer che sui promotori può dare indicazioni sugli elementi necessari alla corretta interazione tra le due sequenze regolatrici.

Per quanto è a nostra conoscenza non sono numerosi i lavori di carattere computazionale che hanno esplorato questa possibilità. Nel 2012, in uno studio delle interazioni cromatiniche derivate da esperimenti Hi-C effettuati in linee cellulari K562 (Lan et al., 2012), una particolare enfasi è stata posta nell'identificazione dei SLFT aventi un ruolo nella generazione e/o mantenimento di contatti tra sequenze nucleotidiche utilizzando un algoritmo di data-mining chiamato Apriori (Agrawal and Srikant, 1994). Un grafo di interazioni tra FT coinvolti nel looping è stato generato e gli autori di questo studio hanno ipotizzato che un gran numero di FT potrebbero avere un ruolo nel looping della cromatina, e che questa attività possa essere mediata sia da FT identici posti alle due estremità dell'ansa, sia da fattori diversi, sempre posti alle stesse estremità. In uno studio del progetto FANTOM5 del 2014 (Andersson et al., 2014) invece, delle sequenze la cui attività di enhancer è stata proposta utilizzando esperimenti di tipo CAGE ("Cap Analysis of Gene Expression") (Kodzius et al., 2006) sono state accoppiate a dei

promotori tenendo in considerazione sia il loro pattern di espressione che la loro prossimità. Successivamente sono stati identificati i SLFT co-occorrenti per mezzo di un test binomiale allo scopo di avvalorare l'ipotesi che tali enhancer e promotori potessero essere in cooperazione. Questi esempi mostrano come le analisi statistiche e computazionali siano utilizzate e possano contribuire a definire i SLFT/FT necessari alla formazione di loop, complementando i casi resi noti da altre tecniche sperimentali.

Il nostro studio si differenzia da questi sopra citati per i seguenti motivi: 1) il set di enhancer e promotori esaminato è composto a partire da enhancer umani validati sperimentalmente *in-vivo*; 2) su queste sequenze regolatrici analizziamo sia le coppie di FT che i gruppi di FT corrispondenti ai SLFT predetti; 3) non ci limitiamo alla ricerca di co-occorrenze come nel caso dello studio del progetto FANTOM, ma prendiamo in considerazione anche coppie e gruppi di FT composti da FT diversi. In prima istanza abbiamo preso in considerazione la presenza di coppie di FT su coppie di enhancer e promotori e sono state studiate le caratteristiche delle interazioni individuate utilizzando questo modello minimo di interazione. Successivamente sono stati invece individuati ed analizzati gruppi di FT presenti in gruppi di enhancer e promotori regolati, che rappresentano un modello di cooperazione enhancer-promotore più esteso. In questa maniera abbiamo studiato il problema della cooperazione enhancer-promotore mediata da fattori di trascrizione tenendo in considerazione i due casi estremi: quello in cui pochi fattori di trascrizione sono responsabili dell'interazione (le coppie) e quello in cui sono invece coinvolti molti FT (i gruppi). Questi due approcci sono stati espletati utilizzando gli stessi dati di partenza in modo da poterli comparare.

3.1 Identificazione delle coppie di FT

Una Coppia di Fattori di Trascrizione (CFT) è composta da due Fattori di Trascrizione (FT), uno all'enhancer e l'altro al promotore. Ogni FT può far parte di una coppia se un suo sito di legame specifico è presente almeno una volta sull'enhancer o sul promotore in considerazione. Un semplice metodo per identificare CFT significativamente presenti sia su enhancer che su promotori, è quello di considerare la presenza delle CFT all'interno di un set di Coppie Enhancer-Promotore (CEP). Se una CFT è capace di mediare la formazione di un loop e il suo ruolo è abbastanza generico da poter essere funzionale in diverse interazioni di questo tipo, i Siti di Legame per Fattori di Trascrizione (SLFT) corrispondenti dovrebbero essere conservati sia sugli enhancer che sui promotori che compongono le CEP; di conseguenza, dovrebbe essere possibile osservare l'occorrenza di tale CFT nel set di CEP più frequentemente di quanto atteso. In questo caso i due eventi necessari alla composizione di una CFT (presenza di almeno un SLFT all'enhancer e di almeno un SLFT al promotore) non sono eventi indipendenti e la frequenza di occorrenza di tale CFT deve essere superiore al prodotto delle frequenze di occorrenza dei singoli FT nel caso in cui questi siano eventi indipendenti.

In altre parole si vuole rifiutare l'ipotesi nulla mostrata in *Formula 3.1*, dove H_0 indica l'ipotesi nulla, t_E indica il FT all'enhancer e t_P il FT al promotore della CFT in considerazione.

Formula 3.1
$$H_0: P(t_E \cap t_P) \leq P(t_E) \cdot P(t_P)$$

Per testare questa ipotesi, abbiamo sviluppato un test binomiale dove ogni CEP è considerata come una variabile dicotomica che può avere due stati mutualmente esclusivi: 1) il successo, che si verifica se una CFT è presente sulla CEP; 2) il fallimento, che si verifica in caso contrario al primo. In questa maniera è stato possibile calcolare un p-value relativo ad ogni CFT che corrisponde alla probabilità di rinvenire un numero di CFT pari o superiore a quello osservato data la probabilità di un accoppiamento casuale dei due membri della coppia. Questo calcolo è mostrato in *Formula 3.2* dove: n è pari al numero totale di CEP, k è uguale al numero di CEP contenenti la CFT in considerazione e p è pari alla probabilità di rinvenimento casuale della stessa CFT. La definizione di p è mostrata in *Formula 3.3* dove: data una CFT composta da un FT all'enhancer (FT_E) e un FT al promotore (FT_P), t_E è il numero di CEP che contengono FT_E , t_P è il numero di CEP che contengono FT_P ed n è di nuovo il numero totale di CEP.

Formula 3.2

$$P(X \geq x) = \sum_{k=x}^n \binom{n}{k} p^k (1-p)^{n-k}$$

Formula 3.3

$$p = \frac{t_E}{n} \cdot \frac{t_P}{n}$$

3.2 Identificazione dei gruppi di FT

L'identificazione dei gruppi di Fattori di Trascrizione (FT) all'interno del set di Coppie Enhancer-Promotore (CEP) è stata effettuata avvalendosi di tecniche di biclustering (Hartigan, 1972). Per biclustering si intende un insieme eterogeneo di analisi ed algoritmi il cui scopo è l'identificazione di gruppi di dati dalle caratteristiche simili (Pontes et al., 2015a). Gli algoritmi di biclustering possono essere considerati un'evoluzione degli algoritmi di semplice clustering (Pontes et al., 2015b). Per comprendere le differenze tra i due approcci e il problema che entrambi affrontano, si prenda in considerazione un set di dati sotto forma di una matrice; questo è il tipico caso di dati di espressione genica che useremo come esempio per esporre i concetti fondamentali. In una matrice di espressione le righe rappresentano geni, le colonne rappresentano condizioni, e gli elementi della matrice rappresentano il livello di espressione del gene nella condizione corrispondente. Se si volessero raggruppare i geni con pattern di espressione simile si potrebbe utilizzare un algoritmo di clustering, il quale raggrupperebbe i geni (le righe) tenendo in considerazione tutte le condizioni (le colonne) (Pontes et al., 2015b). L'espressione di geni diversi però, non è necessariamente correlata in tutte le condizioni; gruppi di geni potrebbero essere coregolati solo in determinate condizioni e gli algoritmi di biclustering sono più indicati in quanto mirano all'identificazione di pattern locali raggruppando gruppi di geni (le righe) e gruppi di condizioni (le colonne) simultaneamente (Madeira and Oliveira, 2004).

Gli algoritmi di clustering possono produrre modelli globali dove l'inserimento di un gene in un cluster è determinato prendendo in considerazione tutte le condizioni, oppure possono effettuare l'operazione contraria,

vale a dire possono raggruppare le condizioni tenendo in considerazione tutti i geni, ma non possono effettuare queste due operazioni simultaneamente. Gli algoritmi di biclustering invece producono modelli locali dove righe e colonne sono raggruppate simultaneamente. Per l'inclusione di un gene (o una condizione) in un gruppo, un sotto-insieme delle condizioni (o dei geni) viene considerato (Madeira and Oliveira, 2004).

Formalmente il problema del biclustering può essere così definito: dato un set di dati sotto forma di una matrice A composta da n righe ed m colonne, e posto che tale matrice sia composta da un insieme di righe $X = \{x_1, \dots, x_n\}$ ed un insieme di colonne $Y = \{y_1, \dots, y_m\}$, e che ogni elemento di tale matrice sia identificato con la dicitura a_{ij} , che indica l'elemento alla riga i ed alla colonna j , si vogliono identificare tutte le sotto-matrici A_{IJ} composte da un sotto-insieme di righe $I = \{i_1, \dots, i_k\}$ ($I \subseteq X$ e $k \leq n$) che esibiscono un comportamento simile su un sotto-insieme di colonne $J = \{j_1, \dots, j_s\}$ ($J \subseteq Y$ e $s \leq m$) e viceversa. Il problema del biclustering è NP-completo e per questo motivo la maggior parte degli algoritmi utilizza metodi euristici per risolverlo (Madeira and Oliveira, 2004), i quali ricercano i bicluster ottimali utilizzando misure che ne determinano la qualità (Pontes et al., 2015a). Esistono anche approcci al problema di tipo diverso, e l'utilizzo o meno di tali misure è una prima forma di classificazione degli algoritmi di biclustering (Pontes et al., 2015b).

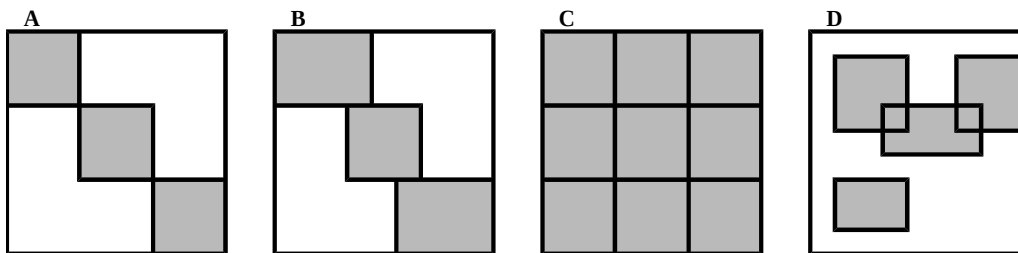
Una delle caratteristiche più importanti di un algoritmo di biclustering è il tipo di bicluster che è in grado di produrre. Ne esistono vari tipi, i quali sono definiti in base a come gli elementi all'interno dei bicluster sono raggruppati. Nella letteratura si trovano generalmente distinti i seguenti tipi (Madeira and Oliveira, 2004; Pontes et al., 2015b):

- 1) Bicluster contenenti valori costanti, dove tutti i valori del bicluster sono simili. Nel caso più semplice, quello di un bicluster perfetto, vale la relazione $a_{ij} = \mu$, dove μ è il valore tipico del bicluster. Nella realtà i valori possono non essere esattamente identici ma può venire accettata una certa quantità di rumore.
- 2) Bicluster contenenti valori costanti su righe o colonne, dove i valori non sono costanti su tutti gli elementi presi singolarmente, ma mostrano variazioni coerenti sulle righe o sulle colonne. Questa coerenza può essere di tipo additivo o moltiplicativo. Nel caso di un bicluster perfetto con righe costanti vale la relazione $a_{ij} = \mu + \alpha_i$ nel caso additivo e $a_{ij} = \mu \times \alpha_i$ nel caso moltiplicativo. In entrambi i casi α_i è un aggiustamento per la riga. Similmente nel caso di bicluster perfetto con colonne costanti valgono le relazioni $a_{ij} = \mu + \beta_j$ e $a_{ij} = \mu \times \beta_j$ dove β_j è un aggiustamento per la colonna.
- 3) Bicluster contenenti valori coerenti sia sulle righe che sulle colonne, qui la coerenza tra gli elementi è più complessa e prende in considerazione righe e colonne contemporaneamente. Anche in questo caso valgono le definizioni dei tipi additivo e moltiplicativo. Nel caso additivo per i valori di un bicluster perfetto vale la relazione $a_{ij} = \mu + \alpha_i + \beta_j$ mentre nel caso moltiplicativo $a_{ij} = \mu \times \alpha_i \times \beta_j$.
- 4) Bicluster con evoluzioni coerenti, in questo caso i valori numerici degli elementi della matrice non sono utilizzati per la generazione dei bicluster e gli elementi sono considerati come simboli, che rappresentano i dati e dai quali gli algoritmi identificano sotto-insiemi di elementi che mostrano comportamenti coerenti. Nel caso dell'espressione genica, ad esempio, questo approccio può essere utile se si vogliono identificare sotto-insiemi di geni regolati positivamente o negativamente nelle diverse condizioni senza considerare le quantità dei livelli di espressione.

Un altro tipo di classificazione dei bicluster è quello riguardante la loro struttura, vale a dire il modo in cui le colonne e le righe sono incorporate all'interno del set di bicluster. Su questa base i bicluster, possono essere (Pontes et al., 2015b):

- 1) Esaustivi, se ogni riga e/o colonna deve essere inclusa in almeno un bicluster.
- 2) Non esaustivi, se tutte le righe e/o colonne non devono necessariamente fare parte di un bicluster.
- 3) Esclusivi, se ogni riga e/o colonna può far parte al massimo di un solo bicluster.
- 4) Non esclusivi, nel caso righe e/o colonne siano ammesse in più di un bicluster, condizione che può portare alla formazione di bicluster sovrapposti.

Figura 3.1



Esempi di strutture di bicluster: Ogni quadrato rappresenta la matrice di partenza, le regioni delimitate con bordo nero e colorate in grigio rappresentano i bicluster identificati. (A) bicluster esaustivi ed esclusivi sia per le righe che per le colonne; (B) bicluster esaustivi per righe e colonne ed esclusivi solo per le righe; (C) bicluster esaustivi e non esclusivi sia per righe che per colonne senza sovrapposizioni; (D) bicluster non esaustivi e non esclusivi sia per righe che per colonne e con sovrapposizioni.

La procedura di biclustering può essere utilizzata per identificare gruppi di FT presenti su gruppi di CEP previa preparazione di una matrice che possa rappresentare i dati in una maniera compatibile con questi algoritmi. Questa matrice può essere preparata facendo corrispondere alle righe le CEP; alle colonne si fanno corrispondere tutti i FT i cui siti di legame corrispondenti siano stati predetti almeno una volta su enhancer e/o su promotori, specificandone la loro localizzazione sull'una o sull'altra sequenza nucleotidica. Gli elementi della matrice si fanno corrispondere in maniera binaria alla presenza o meno di tale FT sulle CEP. La matrice così composta è equivalente alla matrice delle adiacenze di un grafo bipartito dove le CEP e i FT sono i due tipi di nodi. I bicluster che ne derivano sono del tipo 1 (con valori costanti) e rappresentano gruppi di CEP che mostrano gli stessi gruppi di FT su enhancer e promotori. In questo caso non ci si aspetta una ripartizione esclusiva degli elementi, dato che diverse sequenze possono essere regolate dagli stessi FT, e non sono richiesti bicluster esaustivi, dato che non ci si aspetta che tutti i FT siano responsabili delle interazioni enhancer-promotore.

CAPITOLO IV

Descrizione delle banche dati e delle principali risorse computazionali utilizzate

4.1 Sequenze degli enhancer: VISTA enhancer browser

In questo studio una particolare attenzione è stata dedicata all'acquisizione di sequenze nucleotidiche di cui l'attività di enhancer *in vivo* è stata validata sperimentalmente. Per questa ragione sono stati utilizzati i dati provenienti dal "VISTA enhancer browser" (o semplicemente VISTA) (Visel et al., 2007) (<https://enhancer.lbl.gov/>), la principale risorsa di elementi di regolazione *cis*-agenti di vertebrati (Suryamohan and Halfon, 2015) la cui attività di enhancer è validata sperimentalmente in topo transgenico (Visel et al., 2007) e il cui obiettivo è l'identificazione e catalogazione di enhancer nei genomi di uomo e topo. Gli elementi inclusi in questo database sono identificati sia sfruttando la conservazione di sequenze non codificanti tra il genoma umano e quello di altri vertebrati (Nobrega et al., 2003; Pennacchio and Rubin, 2001; Pennacchio et al., 2006; Visel et al., 2008) che utilizzando esperimenti ChIP-Seq mirati alla proteina EP300 che ha la caratteristica di essere abbondantemente localizzata su enhancer attivi *in vivo* (Visel et al., 2009).

Due tipi di conservazione sono presi in considerazione per la selezione delle sequenze (Pennacchio et al., 2006; Visel et al., 2007):

- 1) La conservazione "antica", vale a dire la conservazione di sequenze non codificanti che emerge dalla comparazione di sequenze di organismi evolutivamente distanti. In questo caso il genoma umano o di altri vertebrati è comparato ai genomi di organismi posti a diverse distanze evolutive come pollo, rana, pesce palla o pesce zebra.
- 2) L' ultra-conservazione (Bejerano et al., 2004; Visel et al., 2008), vale a dire un'estrema similarità di sequenza, pari al 100% su segmenti di DNA lunghi almeno 200bp, osservata comparando genomi di mammiferi come uomo, topo e ratto. Questi elementi sono conservati anche in altri genomi di vertebrati e sembra siano mantenuti da un'estrema selezione negativa (Bejerano et al., 2004).

Successivamente viene verificato che tutte le sequenze così selezionate possano svolgere una funzione da enhancer mediante un test in topo transgenico che prevede l'inserimento in oociti fecondati di costrutti genetici composti da il potenziale enhancer proveniente dal genoma umano, posto a monte di un promotore Hsp68 ("Heat shock protein 68") che a sua volta precede un gene reporter LacZ ("Beta-galactosidase"). Tale promotore è inattivo in tessuti embrionali di topo (Kothary et al., 1989) ma può catalizzare l'espressione del gene reporter se accoppiato ad enhancer tessuto-specifici (Visel et al., 2007). Gli embrioni risultanti da questa procedura sono poi ritirati nel giorno embrionale E11.5. Infine viene osservato il pattern di espressione del gene reporter, se questo viene riscontrato in almeno tre diversi embrioni, la sequenza nucleotidica sotto esame è annotata come enhancer

positivo. Se nessun pattern di espressione è stato osservato in maniera consistente in almeno 5 embrioni, la sequenza è annotata come negativa.

VISTA enhancer browser fornisce tutte le sequenze così analizzate, ognuna annotata come enhancer "positivo" o "negativo" e completata dalle annotazioni riguardanti i tessuti in cui l'espressione del gene reporter è stata osservata, le coordinate genomiche dell'elemento regolatore e i geni più vicini all'enhancer.

Dato il tipo di validazione sperimentale a cui gli enhancer provenienti da VISTA enhancer browser sono soggetti, è noto che tali sequenze agiscono da regolatori dell'espressione durante lo sviluppo embrionale. Il loro alto grado di conservazione suggerisce che regolino processi particolarmente conservati nei mammiferi e nei vertebrati, il che è consistente con il loro ruolo di regolatori dello sviluppo. Sia gli enhancer identificati per ultra-conservazione che quelli identificati per conservazione tra vertebrati evolutivamente molto distanti si trovano adiacenti a geni le cui funzioni sono arricchite per regolazione dello sviluppo embrionale e regolazione della trascrizione (Bejerano et al., 2004; Visel et al., 2008).

Gli enhancer identificati con tecniche di genomica comparativa e fiancheggiati i geni DACH e SALL1, ricapitolano diversi aspetti dell'espressione di tali geni fiancheggiati nei test in topo transgenico (Nobrega et al., 2003; Pennacchio et al., 2006). Per quanto riguarda gli enhancer identificati utilizzando ChIP-Seq contro la proteina EP300, i picchi di questa proteina identificati in prosencefalo nel giorno E11.5 sono significativamente distribuiti in prossimità di geni espressi nello stesso tessuto (Visel et al., 2009). Tutte queste evidenze suggeriscono che gli enhancer individuati in questi studi siano capaci di regolare l'espressione dei geni adiacenti (Bejerano et al., 2004; Visel et al., 2009).

4.2 Sequenze dei promotori: UCSC genome browser

Le sequenze dei promotori dei geni fiancheggiati gli enhancer segnalati dalle annotazioni di VISTA sono rinvenute avvalendosi di UCSC genome browser (Kent et al., 2002) (<http://genome.ucsc.edu/>) un servizio che mette a disposizione diversi genomi incluso quello umano. Una delle caratteristiche che lo contraddistinguono è la sua natura estendibile che lo rende pratico per l'integrazione di annotazioni provenienti da terze parti, e anche VISTA enhancer browser lo utilizza per annotare i propri segmenti genomici.

Ogni promotore è definito come le 2000bp a monte del SIT di ogni relativo gene. Questa lunghezza permette di prendere in considerazione sia il nucleo del promotore che gli elementi prossimi al promotore, in modo da poter considerare un ampio spettro dei FT capaci di regolare i promotori selezionati, e di includere quegli elementi a monte del nucleo del promotore che potrebbero essere coinvolti nelle interazioni tra enhancer e promotori (Calhoun et al., 2002; Riethoven, 2010).

Da UCSC genome browser, i promotori dei geni sono stati rinvenuti avvalendosi del "Table Browser data retrieval tool" (Karolchik et al., 2004), selezionando lo stesso assembly del genoma umano utilizzato da VISTA vale a dire GRCh37/hg19 (Release: Feb. 2009) ed utilizzando le tabelle "knownCanonical" e "knownGene" (Release: 2011-12-21) (Hsu et al., 2006).

4.3 Modelli dei siti di legame dei fattori di trascrizione: HOCOMOCO

La risorsa chiamata "*HOmo sapiens* COmprehensive MOdel COllection" o semplicemente "HOCOMOCO" è una collezione di modelli per siti di legame di fattori di trascrizione specifica per l'organismo umano, completa, non ridondante e curata manualmente (Kulakovskiy et al., 2013) (<http://www.cbrc.kaust.edu.sa/hocomoco/>). A differenza di altre risorse simili, come JASPAR (Sandelin et al., 2004) o TRANSFAC (Matys et al., 2003), diversi tipi di evidenze sperimentali sono integrate nella generazione dei modelli, e ogni modello, specifico per l'organismo umano, è mappato al database UniProt (The UniProt Consortium, 2017) dai curatori, evitando la presenza di più modelli per la stessa proteina. Questa risorsa evita quindi sia la presenza di più modelli relativi allo stesso FT, tipicamente derivanti da esperimenti diversi e che rende difficile la scelta del modello più idoneo da utilizzare per la predizione dei SLFT, sia la presenza di modelli generati da poche sequenze grazie all'integrazione di dati di diversa origine (Kulakovskiy et al., 2013).

I dati utilizzati da HOCOMOCO per la generazione dei modelli includono: evidenze sperimentali di tipo ChIP-Seq dal progetto ENCODE Yale/HudsonAlpha (ENCODE Project Consortium, 2011), evidenze sperimentali di tipo SELEX (Jolma et al., 2010), dati relativi ai vertebrati provenienti da TRANSFAC 2011.2 SITE (Matys et al., 2006) e dati provenienti da JASPAR CORE, sempre relativi ai vertebrati (Portales-Casamar et al., 2010).

La qualità dei dati così collezionati viene valutata dai curatori di HOCOMOCO, i quali nella creazione dei modelli assegnano un peso diverso alle sequenze usate a seconda dell'affidabilità dei dati. I modelli risultanti sono classificati utilizzando un codice alfabetico che va dalla lettera "A" alla "E" che indica un ordine decrescente di affidabilità. Sebbene tutti i modelli siano resi disponibili dalla risorsa, inclusi quelli a bassa affidabilità, solo quelli classificati nei gruppi che vanno dalla lettera "A" alla lettera "D" sono stati da noi utilizzati per predire i SLFT.

La non ridondanza, la mappatura dei modelli verso le voci in UniProt delle proteine relative, la specificità del progetto riguardo l'organismo umano e l'alta qualità dei modelli (i quali sono curati manualmente e includono un gran numero di dati diversi, inclusi quelli provenienti dalle risorse simili e alternative), sono tutti ottimi elementi a favore di questa banca dati, che è anche più recente rispetto alle altre. Per questi motivi HOCOMOCO è stato selezionato come risorsa di modelli di SLFT in questo nostro studio.

4.4 Predittore di siti di legame di fattori di trascrizione: MOODS

"MOtif Occurrence Detection Suite" o semplicemente "MOODS" (Korhonen et al., 2009; Pizzi et al., 2011) (<https://www.cs.helsinki.fi/group/pssmfind/>) è una collezione di algoritmi allo stato dell'arte per effettuare predizione di Siti di Legame per Fattori di Trascrizione (SLFT) *in-silico*. Permette di effettuare predizioni del tipo "*knowledge based*" vale a dire basate sulla conoscenza a priori (sotto forma di PFM) dei SLFT che vengono predetti. Questi predittori si differenziano da quelli "*ab-initio*" i quali ricercano SLFT senza una precedente conoscenza dei siti di legame, basandosi sull'identificazione di pattern ricorrenti all'interno del set di sequenze analizzate (Abnizova and Gilks, 2006). Nel caso di predizioni "*knowledge based*", poiché ogni predizione è effettuata utilizzando un modello del sito di legame di un Fattore di Trascrizione (FT) specifico e ben noto è facile

correlare ogni SLFT predetto al relativo FT. Utilizzando algoritmi "*ab-initio*" questa associazione tra SLFT e FT non è altrettanto immediata; pertanto sono spesso utilizzati per identificare SLFT nuovi, non ancora caratterizzati e per i quali un modello del sito di legame non è disponibile. In questo studio è nostra intenzione associare ogni SLFT predetto al corrispettivo FT senza ambiguità, e per questo è stato scelto un algoritmo di predizione appartenente alla categoria "*knowledge based*".

Gli algoritmi che utilizzano PFM per la predizione di SLFT possono a loro volta distinguersi in "*on-line*" ed "*index-based*". Mentre gli algoritmi "*on-line*" ricercano i siti di legame direttamente sulle sequenze grezze, quelli "*index-based*" prevedono una fase di pre-processamento delle sequenze che tipicamente consiste nella generazione di una struttura dati che fa da indice e che ha lo scopo di velocizzare la successiva identificazione dei SLFT sullo stesso set di sequenze. Se da una parte la fase di predizione è effettivamente velocizzata, la fase di generazione dell'indice richiede molte risorse sia in termini di spazio che di tempo e questo rende gli algoritmi "*index-based*" meno pratici di quelli "*on-line*" (Korhonen et al., 2009).

Per mantenere la praticità degli algoritmi "*on-line*" senza sacrificare la velocità di computazione delle previsioni, MOODS mette a disposizione una collezione di algoritmi altamente ottimizzati, che fanno uso di scorciatoie per minimizzare i tempi di ricerca ed implementati utilizzando il linguaggio di programmazione C++. Sono messe a disposizione anche delle librerie per utilizzare tali algoritmi mediante linguaggi di programmazione più pratici come Python e Perl senza dover rinunciare alla velocità dell'implementazione in C++.

Oltre a effettuare la ricerca dei SLFT velocemente, MOODS fornisce anche gli algoritmi necessari per generare PWM ("Position Weight Matrix") partendo da PFM ("Position Frequency Matrix"), dando così all'utente la possibilità di specificare le frequenze attese dei nucleotidi relative al proprio set di sequenze. Inoltre implementa gli algoritmi per la conversione da p-value a score in modo da poter associare una significatività statistica alle previsioni.

4.5 Algoritmo di biclustering: BicNET

BicNET (Henriques and Madeira, 2015, 2016) (<https://web.ist.utl.pt/rmch/bicnet/>) è un algoritmo allo stato dell'arte ed ad alta efficienza per la ricerca esaustiva di bicluster massimi in matrici binarie. Se la matrice binaria rappresenta la matrice delle adiacenze di un grafo BicNET può essere utilizzato per identificare moduli della rete sotto forma di bicluster coerenti. Gli algoritmi di questo tipo generalmente si focalizzano sull'identificazione di moduli densi, composti da nodi altamente interconnessi, il che spesso porta: 1) all'identificazione di associazioni ben note; 2) all'incapacità di scoprire interazioni tra geni/proteine meno studiati ed annotati, le cui interazioni sono tipicamente associate a bassi valori di confidenza (Henriques and Madeira, 2016).

BicNET invece è capace di identificare moduli non densi ma coerenti (che possono quindi essere biologicamente rilevanti), ed è capace di identificare coerenze non banali. Questo è ottenuto mediante (Henriques and Madeira, 2015): 1) la riduzione della complessità della matrice di adiacenza attraverso la generazione di un grafo eterogeneo bipartito minimo, vale a dire un grafo composto da due set di nodi diversi dove ogni nodo di un set (es. geni) può avere interazioni solo con i nodi dell'altro set (es. proteine); 2) una ricerca esaustiva dei moduli

in questo grafo eseguendo il biclustering sia sulla matrice di adiacenza che sulla sua trasposta; 3) L'utilizzo di algoritmi di biclustering del tipo "pattern-based", i quali sono gli unici algoritmi capaci di garantire una ricerca esaustiva ed efficiente dei bicluster e che al tempo stesso danno la possibilità di definire parametri relativi alla coerenza e alla qualità dei bicluster identificati (Henriques and Madeira, 2015). I bicluster trovati da questo algoritmo rappresentano gruppi di nodi di un set associati a gruppi di nodi dell'altro set che possono esibire i seguenti tipi di coerenza (Henriques and Madeira, 2015, 2016): 1) coerenze costanti, dove i nodi di un set mostrano un solo tipo di interazione con i nodi dell'altro set. Ciò garantisce che un gene associato con un gruppo di proteine esibisca lo stesso tipo di interazione con tutte le proteine del gruppo, evidenziando interazioni tra gruppi omogenei di geni e proteine; 2) coerenze simmetriche, dove i nodi di un set esibiscono valori assoluti simili ma con segno differente e che possono esibire pattern di attivazione e repressione; 3) modello plaid (Lazzeroni and Owen, 2002), che permette la scoperta di bicluster sovrapposti.

CAPITOLO V

Organizzazione ed elaborazione dei dati

5.1 Selezione delle sequenze di enhancer e promotori

Da VISTA enhancer browser (VISTA), sono state prelevate tutte le sequenze del genoma umano annotate come enhancer positivi assieme alle rispettive annotazioni riguardanti i geni fiancheggiati. VISTA distingue gli enhancer in intragenici (quando l'enhancer si trova in un introne) ed intergenici, nel primo caso un solo gene è segnalato come fiancheggiante mentre nel secondo caso sono segnalati i due geni ai lati dell'enhancer.

Sono stati estratti 807 enhancer sotto forma di file in formato FASTA. È stata verificata la presenza di sovrapposizioni all'interno dell'insieme di enhancer onde evitare la presenza sequenze diverse localizzate nelle stesse regioni genomiche ed stata osservata la sovrapposizione di 13 enhancer. Di questi, 10 erano sovrapposti in coppie e i rimanenti erano tutti e tre sovrapposti tra di loro. Per eliminare questa ridondanza senza perdere informazione le sequenze sovrapposte sono state unite generando sequenze senza sovrapposizioni. VISTA contraddistingue ogni enhancer umano mediante un codice che inizia con le lettere "hs" seguite da un valore numerico (ci riferiremo a questo codice con il nome VISTA ID), nel caso delle sequenze sovrapposte uno dei VISTA ID delle sequenze di origine è stato utilizzato per indicare la corrispondente sequenza da noi generata. Le denominazioni risultanti sono mostrate nella *Lista A1* dell'appendice. Una volta eliminate le sovrapposizioni il set di enhancer aveva una dimensione di 800 elementi, di cui 268 annotati come intragenici e 532 come intergenici. A questi 800 enhancer erano annotati in maniera non univoca un totale di 786 geni.

Le sequenze dei promotori dei geni sono state ottenute estraendo le 2000bp a monte del SIT di ogni gene. Questa lunghezza ha permesso di includere gli elementi prossimi al promotore i quali potrebbero essere responsabili della corretta interazione enhancer-promotore (Calhoun et al., 2002; Riethoven, 2010). Utilizzando UCSC genome browser è stato possibile rinvenire le sequenze dei promotori di 740 geni, i quali a loro volta erano accoppiabili a un totale di 790 enhancer.

I 790 enhancer e i 740 promotori dei geni associati sono stati organizzati in Coppie Enhancer-Promotore (CEP), e dai due set di sequenze è stato generato un totale di 1247 coppie. Una volta generate le CEP, è stata verificata la presenza di sovrapposizioni tra sequenze enhancer e promotori accoppiati: 13 CEP si sono rivelate sovrapposte e sono state eliminate dal set ottenendo un totale di 1234 CEP, formate da 786 enhancer e 730 promotori.

Il set completo di 786 enhancer utilizzato per le 1234 CEP è mostrato in appendice in *Lista A1*, dove ogni enhancer è identificato dal rispettivo VISTA ID. Sono anche riportate le coordinate genomiche relative all'assembly del genoma umano GRCh37/hg19. Similmente, il set completo di 730 promotori è mostrato in *Lista A2* dell'appendice, dove per ogni promotore sono elencati il gene symbol del gene regolato, il codice identificativo di UCSC dello stesso gene e le coordinate genomiche del promotore sempre relative all'assembly GRCh37/hg19

del genoma umano. La lista completa di 1234 CEP è mostrata in *Lista A3* dell'appendice, qui per ogni CEP è mostrato un codice numerico identificativo della coppia, il VISTA ID dell'enhancer e il gene symbol del genere regolato. Vi sono anche elencate informazioni riguardo la predizione dei SLFT e il significato di tali dati sarà spiegato nel paragrafo 5.2.

5.2 Predizione dei SLFT su enhancer e promotori

Da HOCOMOCO v9 sono state acquisite tutte le PFM ("Position Frequency Matrix") di alta qualità (vale a dire appartenenti ai gruppi A, B, C e D) e specifiche per l'organismo umano, per un totale di 426 PFM relative a 400 FT. Tali PFM sono state utilizzate per la predizione dei SLFT, effettuata utilizzando software da noi sviluppato avvalendosi degli algoritmi e delle librerie messe a disposizione da MOODS v1.0. La predizione di SLFT è stata effettuata indipendentemente su enhancer e promotori specificando per ogni set di sequenze i relativi valori delle frequenze attese (o background) dei nucleotidi. Le frequenze attese sono state calcolate per ogni set assumendo indipendenza tra nucleotidi consecutivi, e sono mostrate in *Tabella 5.1*, arrotondate alla quarta cifra decimale.

	A	C	G	T
Enhancer	0.2876	0.2112	0.2119	0.2892
Promotori	0.2564	0.2434	0.2447	0.2556

Frequenze attese: Ogni cella contiene la frequenza attesa di un nucleotide (colonna) nel relativo set di sequenze (righe). le frequenze attese sono utilizzate per la predizione dei SLFT in ogni set di sequenze, e sono calcolate assumendo indipendenza tra i nucleotidi consecutivi.

Prima della predizione dei SLFT, ogni PFM viene convertita in una PWM ("Position Weight Matrix") tenendo in considerazione le frequenze di background (si veda paragrafo 2.2). Questi valori possono influire sui risultati della predizione ed è preferibile specificare un background adatto al set di sequenze in esame. Utilizzando frequenze di background specifiche per ogni set, si compensano eventuali sbilanciamenti delle distribuzioni dei nucleotidi nei set stessi. Esaminando le frequenze attese di enhancer e promotori si nota una leggera discrepanza: mentre per i promotori le frequenze dei nucleotidi sono quasi uniformi, lo stesso non vale per gli enhancer, i quali mostrano una leggera preferenza composizionale per A e T.

La predizione dei SLFT è stata effettuata utilizzando una soglia stringente, pari a un p-value di $1 \cdot 10^{-7}$, in modo da limitare l'alto numero di falsi positivi tipico di questo tipo di predizioni. Sono stati predetti un totale di 5917 SLFT, di cui 3343 su enhancer e 2574 su promotori, corrispondenti a 231 FT. Dei 786 enhancer e 730 promotori esaminati 738 enhancer e 639 promotori esibivano almeno un SLFT per i FT ricercati. In appendice, la *Lista A3* mostra il numero di SLFT predetti e il numero di FT ad essi corrispondenti per ogni sequenza del set di CEP.

5.3 Riduzione del set di coppie enhancer-promotore

Prima di procedere alle analisi successive, dal set di Coppie Enhancer-Promotore (CEP) di 1234 elementi sono state eliminate le CEP composte da almeno una sequenza priva di Siti di Legame per Fattori di Trascrizione (SLFT) predetti. In questa maniera sono state selezionate 1011 CEP, composte da 702 enhancer e 621 promotori (si veda *Lista A3*). Su questi enhancer erano presenti SLFT per 159 Fattori di Trascrizione (FT), mentre sui promotori un numero pari a 169. Il totale dei FT per i due set di sequenze conta 229 elementi, di cui 99 condivisi tra enhancer e promotori (si veda *Lista A4*).

5.4 Generazione di CFT e identificazione di CFT significative

I Fattori di Trascrizione (FT) corrispondenti ai Siti di Legame per Fattori di Trascrizione (SLFT) predetti sulle sequenze di ogni Coppia Enhancer-Promotore (CEP) sono utilizzati per generare Coppie di Fattori di trascrizione (CFT). Queste rappresentano FT la cui presenza, uno all'enhancer e l'altro al promotore, potrebbe avere una funzione nella formazione e stabilizzazione del loop tra le due sequenze. L'accoppiamento è effettuato sull'intero set di CEP, abbinando ogni FT il cui SLFT fosse presente almeno una volta su un enhancer con ogni FT il cui SLFT fosse presente almeno una volta sul rispettivo promotore. Questa procedura, una volta applicata al set di 1011 CEP descritto precedentemente, porta alla generazione di 2837 coppie di FT per un numero di FT pari a 229.

Allo scopo di determinare quali CFT fossero presenti nel set di CEP in maniera statisticamente significativa (il set di coppie enhancer-promotore è mostrato in *Lista A3* dell'appendice), per ogni CFT è stata prima calcolata la probabilità di accoppiamento casuale dei suoi FT; poi è stata calcolata la significatività statistica del numero di occorrenze della coppia con un test binomiale.

Il test binomiale è stato effettuato utilizzando la *Formula 3.2* dove n è pari al numero totale di CEP, k è uguale al numero di CEP contenenti la CFT in considerazione e p è pari alla probabilità di rinvenimento casuale della CFT. Il calcolo di p è mostrato in *Formula 3.3* dove, data una CFT composta dai FT FT_E ed FT_P (il primo sull'enhancer e il secondo sul promotore), t_E è il numero di CEP che contengono FT_E , t_P è il numero di CEP che contengono FT_P ed n è il numero totale di CEP.

Questi calcoli sono stati implementati in Python usando la libreria SciPy (Jones et al., 2001). La lista completa di tutte le CFT significative è mostrata in *Lista A7* dell'appendice.

5.5 Validazione delle CFT mediante bootstrap

Per validare statisticamente le Coppie di Fattori di Trascrizione (CFT) significative trovate usando il test binomiale è stata utilizzata una procedura di bootstrap. Gli enhancer e i promotori del set di Coppie Enhancer-Promotore (CEP) sono stati divisi e nuovamente accoppiati tra di loro in maniera casuale generando un set di CEP

Randomizzate (CEP-R). Questa procedura è stata ripetuta fino ad ottenere 100000 set di CEP-R. Successivamente, come mostrato in *Formula 5.1*, il p-value di ogni CFT è stato calcolato dividendo il numero di set di CEP-R contenenti tale CFT un numero di volte maggiore o uguale a quello osservato nel set di CEP di partenza (r) per il numero totale di set di CEP-R generati (R). Questa procedura conferma la significatività statistica di tutte le coppie del binomiale. In *Lista A7* dell'appendice sono elencati tutti i p-value così calcolati.

Formula 5.1

$$p\text{-value} = \frac{r}{R}$$

5.6 Ricerca delle CFT all'interno dell'interattoma umano

Secondo la nostra ipotesi le Coppie di Fattori di Trascrizione (CFT) sono proteine interagenti; per verificare la significatività di questa assunzione, ogni CFT significativa è stata ricercata all'interno dell'interattoma umano. La risorsa utilizzata a questo scopo è stata il database STRING v9.1 (Franceschini et al., 2013) (<http://string-db.org/>). STRING integra informazioni relative ad interazioni ed associazioni tra proteine provenienti da risorse di vario tipo, tra cui: evidenze sperimentali di interazioni proteina-proteina; associazioni funzionali di gruppi di proteine rispetto a pathway metabolici, di trascrizione o di segnale; predizioni di associazioni proteina-proteina di vario tipo. Queste informazioni si complementano tra loro permettendo la generazione di una risorsa completa e contenente molte annotazioni relative a un vasto numero di proteine (von Mering et al., 2005).

STRING rappresenta le interazioni tra proteine sotto forma di grafo formato da 20770 nodi e 4850628 archi. Ad ogni associazione è assegnato un punteggio chiamato "combined score" il quale esprime il livello di confidenza di quella associazione tenendo in considerazione tutte le varie risorse (von Mering et al., 2005). Per comparare le CFT significative con STRING solo le interazioni con un combined score superiore o uguale a 400 (equivalente a un livello medio di confidenza) sono state considerate, ottenendo un grafo composto da 19059 nodi e 938370 archi.

All'interno di questo grafo è stato possibile reperire 193 dei 196 FT costituenti 356 delle 364 coppie significative. Per ognuna di queste 356 CFT, attraverso l'utilizzo dell'interattoma umano come riportato in STRING, si è calcolato il cammino minimo tra i due FT accoppiati. In questo modo ogni CFT recupera un numero minimo di proteine intermedie necessario per collegarle. Questa computazione è stata effettuata avvalendosi della libreria NetworkX (Hagberg et al., 2008). Lo stesso si è fatto per mettere in relazione RNAP II con ogni FT.

Per verificare se le CFT fossero significativamente vicine in termini di intermedi proteici, la distribuzione dei cammini minimi di STRING è stata comparata alla distribuzione dei cammini minimi delle CFT per mezzo di un test binomiale dalla formula identica a quella esposta in *Formula 3.2*. Dato un numero n di CFT, si calcola la probabilità di rinvenire in STRING, k o più coppie di proteine collegate da i o meno proteine intermedie. In questo calcolo n è pari al numero di CFT connesse da qualunque numero di proteine intermedie in STRING (356), k è il numero di CFT collegate da i o meno proteine intermedie, p è pari alla probabilità di trovare in STRING un numero di intermedi minore o uguale ad i . Tale probabilità è stata calcolata a partire dalla distribuzione dei

cammini minimi di STRING, mostrata in *Tabella 5.2*, dalla quale è stata calcolata la relativa distribuzione cumulativa (p è quindi equivalente alla probabilità cumulativa su i , dove i è il numero di proteine intermedie per ogni coppia). La stessa procedura è stata effettuata in maniera analoga considerando le coppie composte dai FT e la RNAP II. Oltre alle lunghezze dei cammini minimi tra i FT accoppiati, vengono identificate le proteine intermedie in modo stabilire quali proteine siano più frequentemente presenti nella rete di FT.

Tabella 5.2

Lunghezza (# nodi)	Cammini minimi (#)	Cammini minimi (%)
0	938370	0,25%
1	36202384	10%
2	180882752	50%
3	115367154	32%
4	24815998	6,8%
5	3293030	0,91%
6	327322	$9,0 \cdot 10^{-2}\%$
7	26954	$7,5 \cdot 10^{-3}\%$
8	1520	$4,2 \cdot 10^{-4}\%$
9	70	$1,9 \cdot 10^{-5}\%$
10	2	$5,5 \cdot 10^{-7}\%$

Distribuzione dei cammini minimi: Dopo aver selezionato da STRING le interazioni annotate con un "combined score" superiore a 400 il grafo risultante mostra questa distribuzione dei cammini minimi. I cammini sono qui espressi in numeri di nodi e rappresentano il numero di proteine intermedie minimo necessario a mettere in comunicazione tutte le coppie di proteine del grafo. Le percentuali sono calcolate rispetto a tutti i cammini minimi presenti tra le proteine di STRING selezionate.

5.7 Analisi dell'espressione delle CFT

Per verificare che le Coppie di Fattori di Trascrizione (CFT) fossero composte da FT espressi in tessuti comuni sono state utilizzate due risorse: EBI Expression Atlas (Petryszak et al., 2016) e Human Proteome Map (Kim et al., 2014). EBI Expression Atlas è una selezione di studi su scala genomica caratterizzati da una buona qualità dei dati grezzi e una vasta copertura in termini di tessuti provenienti dall'archivio ArrayExpress (Kolesnikov et al., 2015). Nella sezione "Baseline Atlas" da noi utilizzata sono resi disponibili dati di tipo RNA-seq (Nagalakshmi et al., 2008) misurati in FPKM ("Fragments Per Kilobase of exon model per Million mapped reads"). Human Proteome Map invece fornisce dati di tipo proteomico, derivanti da studi condotti con tecniche di spettrometria di massa e misurati in termini di "spectral counts".

Da EBI Expression Atlas sono state selezionati due studi: 1) RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues (E-MTAB-2836), (Djureinovic et al.; Habuka et al., 2015; Uhlén et al., 2015), o più brevemente "Tissues 32", che fornisce profili per 32 tessuti umani; 2) RNA-Seq of human individual tissues and mixture of 16 tissues (E-MTAB-513) (Asmann et al., 2012; Barbosa-Morais et al., 2012; Derrien et al., 2012; Kechavarzi and Janga, 2014; Kim et al., 2013; Zhao, 2014), anche noto come "Illumina

Body Map", che fornisce profili per 16 tessuti umani. Human Proteome Map invece fornisce profili riguardanti 17 tessuti di organismi adulti, 7 tessuti fetali e 6 tipi cellulari ematopoietici.

Per ogni studio sono stati ricercati i FT di ogni CFT ed è stato verificato se entrambi i FT fossero presenti in almeno un tessuto in comune. Basandosi sui livelli di espressione riportati da EBI Expression Atlas è stato utilizzato un livello di soglia pari a 0.4. Per Human Proteome Map invece si è utilizzato un livello di soglia pari a 1.1. Dai dati provenienti da EBI Expression Atlas è stato possibile rinvenire i pattern di espressione per tutti i 196 FT che compongono le 364 CFT significative. Da Human Proteome Map invece sono stati trovati 185 FT, corrispondenti a 335 CFT.

5.8 Generazione dei gruppi di FT

Tutte le Coppie Enhancer-Promotore (CEP) e tutti i Fattori di Trascrizione (FT) corrispondenti ai Siti di Legame per Fattori di trascrizione (SLFT) predetti, sono stati utilizzati per generare una matrice binaria. Qui ogni riga è una CEP, ogni colonna è un FT (di cui è stata specificata la presenza su enhancer e/o su promotore), e gli elementi della matrice rappresentano la presenza o meno dei FT sulle CEP corrispondenti. È stata così ottenuta una matrice composta da 1011 righe (il numero di CEP) e 328 colonne (corrispondenti a 159 più 169 FT su enhancer e promotori rispettivamente). La matrice così composta rappresenta una matrice delle adiacenze di un grafo minimo (ogni nodo è specificato una sola volta senza ripetizioni su righe e/o colonne), bipartito (i nodi delle righe possono formare archi solo con i nodi delle colonne) ed eterogeneo (i nodi sulle righe e sulle colonne rappresentano entità biologiche di diverso tipo). Le caratteristiche del grafo precedentemente elencate sono i requisiti per il corretto utilizzo di BicNET (Henriques and Madeira, 2015, 2016), l'algoritmo di biclustering tramite il quale sono stati identificati in maniera esaustiva i moduli coerenti del grafo, sotto forma di bicluster, rappresentanti gruppi di FT coerentemente presenti su gruppi di CEP.

È stato utilizzato "BicNET Source v2.0.0 programmatic version" mantenendo le opzioni predefinite. Per calcolare la significatività statistica dei bicluster i p-value sono stati calcolati come descritto in (Madeira et al., 2010) con alcune modifiche. In sostanza, dato un bicluster composto da un sotto-insieme di righe I ed un sotto-insieme di colonne J : 1) la probabilità a priori p del pattern è calcolata come probabilità congiunta di avere un valore 1 in ogni colonna; 2) il p-value è calcolato con un test binomiale, come la probabilità che il pattern occorra un numero di volte almeno pari a $|I|$. Per esempio, per il bicluster $A_{I,J} = \{y2, y4, y5\}$, la probabilità di occorrenza del pattern è semplicemente $p = P(y2=1) \cdot P(y4=1) \cdot P(y5=1)$, dove la probabilità per ognuna delle colonne ($P(y2=1)$, $P(y4=1)$ e $P(y5=1)$) è calcolata con la distribuzione di Bernoulli sulla colonna stessa. Il valore di ciascuna colonna è differente perché molto differente è la probabilità di avere il valore 1 (vale a dire, la probabilità di avere un dato FT all'enhancer o al promotore è molto differente per ogni FT). Data la probabilità del pattern (p) il p-value è calcolato come la somma per $n \geq |I|$ della distribuzione Binomiale (n, p).

Per la selezione dei bicluster rilevanti sono stati applicati filtri ad hoc. Nello specifico sono stati accettati solo i bicluster comprendenti più di una CEP, composti da almeno un FT su enhancer ed un FT su promotore. Inoltre sono stati selezionati solo bicluster privi di rumore, vale a dire che sono ammessi solo bicluster i cui elementi sono pari ad uno, e nel caso di bicluster con una sovrapposizione superiore al 75% è stato mantenuto solo

il bicluster col minor p-value. Infine i bicluster sono stati filtrati in accordo alla loro significatività statistica utilizzando una soglia in termini di p-value pari a $1 \cdot 10^{-3}$. È stato osservato che nella maggior parte dei casi tra i bicluster sovrapposti quelli con la maggiore significatività statistica erano anche i bicluster dalle dimensioni maggiori. Utilizzando questi filtri abbiamo quindi selezionato i bicluster più estesi per i quali i FT fossero presenti sia su enhancer che su promotori e che coinvolgessero il maggior numero di FT e CEP. Tale procedura ha portato all'identificazione di 48 bicluster comprendenti da 222 CEP e 24 FT.

5.9 Grafo di CFT

Tutte le Coppie di Fattori di Trascrizione significative (CFT) risultate significative dal test binomiale esposto nel paragrafo 5.4 sono state utilizzate per generare un grafo dove ogni nodo rappresenta un Fattore di trascrizione (FT) di una CFT significativa e gli archi sono tracciati solo se tali FT fanno parte della stessa CFT significativa. Questa rete rappresenta interazioni tra FT posti ad enhancer e promotori che potrebbero essere coinvolte nel processo del looping della cromatina. Il grafo è stato generato utilizzando Cytoscape v2.8.2 (Smoot et al., 2011) e le relative statistiche sono state calcolate utilizzando il plugin NetworkAnalyzer version 2.7 (Assenov et al., 2008).

5.10 Analisi mediante ontologie

Funzioni ed annotazioni riguardanti geni relativi alle CFT significative e ai bicluster significativi sono state studiate utilizzando tecniche analitiche di arricchimento (enrichment analysis). Per queste analisi è stato selezionato lo strumento NET-GE (Bovo et al., 2016; Di Lena et al., 2015) (<http://net-ge.biocomp.unibo.it/enrich>). Rispetto agli altri strumenti paragonabili NET-GE ha il vantaggio di poter ricercare termini significativamente arricchiti utilizzando liste di input che possono essere anche molto brevi, con un minimo di due geni (Di Lena et al., 2015), cosa resa possibile dalla sua particolare implementazione. Gli strumenti di enrichment tradizionali si basano sulle annotazioni del set di geni di input e trattano ogni gene come un elemento isolato; l'analisi effettuata da NET-GE invece si avvale anche di informazione proveniente dal network di interazione delle proteine/geni in esame. Il network di riferimento è STRING e sulla base delle annotazioni dei geni in input viene generata una sotto-rete di proteine che condivide le stesse annotazioni minimizzando la grandezza della sotto-rete e massimizzandone la compattezza. L'analisi di enrichment verifica che le annotazioni dei geni di input siano significativamente condivise con le annotazioni della sotto-rete generata e poi viene effettuato un test esatto di Fisher seguito da una correzione per le comparazioni multiple. La procedura permette di identificare termini arricchiti che possono anche non far parte delle annotazioni proprie dei geni di partenza, ma che possono aiutare a delineare la funzione del gene/proteina all'interno del suo contesto funzionale. I database di annotazione ed ontologie supportate da NET-GE sono Gene Ontology (Ashburner et al., 2000; The Gene Ontology Consortium, 2015), Reactome (Fabregat et al., 2016) e KEGG (Kanehisa et al., 2016). Data la necessità di investigare gruppi di geni/proteine di dimensioni anche ridotte questa risorsa è stata scelta per il nostro studio. Lo strumento consente anche di effettuare enrichment tradizionali.

Le analisi di enrichment, salvo dove specificato diversamente, sono state effettuate utilizzando enrichment sia standard che network-based, applicando la correzione di Bonferroni ed impostando un livello di significatività di soglia pari a un p-value di $5 \cdot 10^{-2}$.

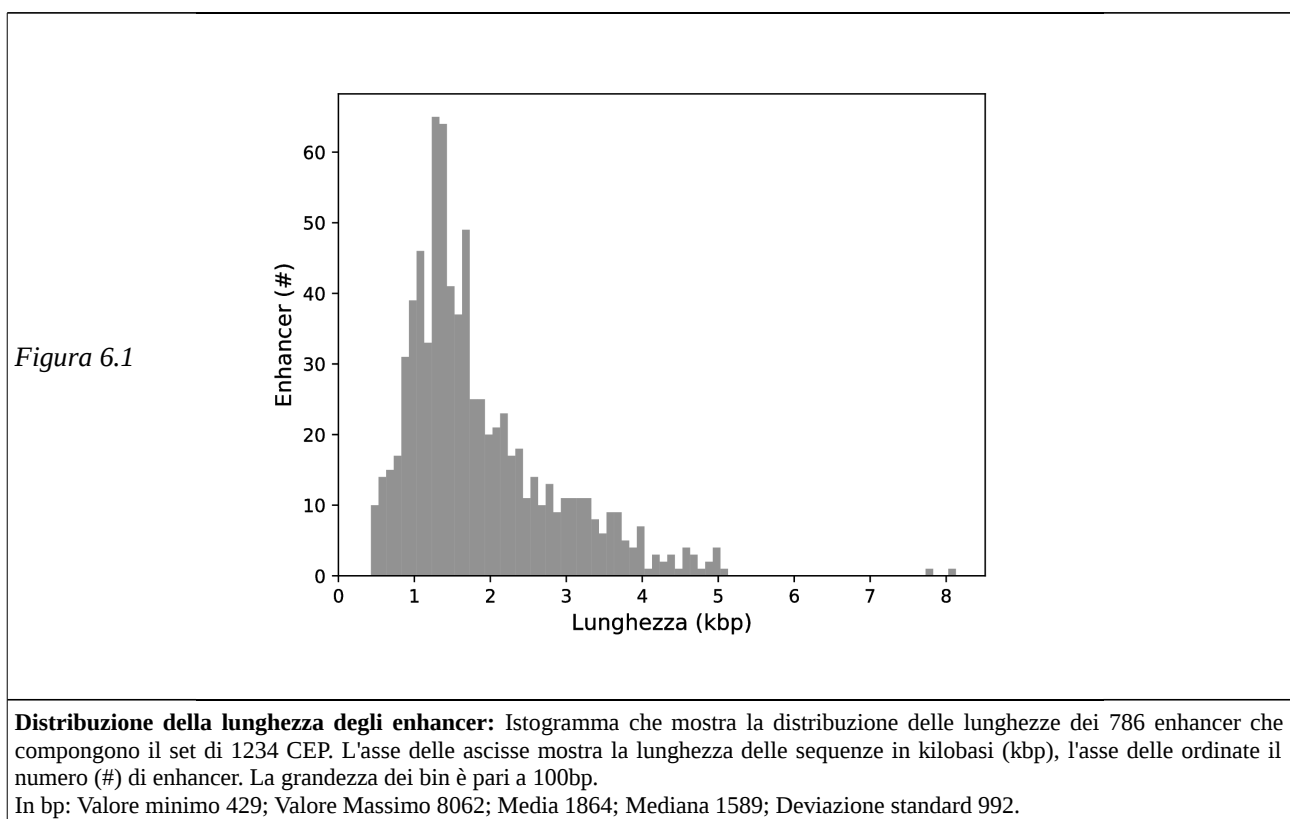
CAPITOLO VI

Risultati e discussione

6.1 Dataset di sequenze enhancer e promotore

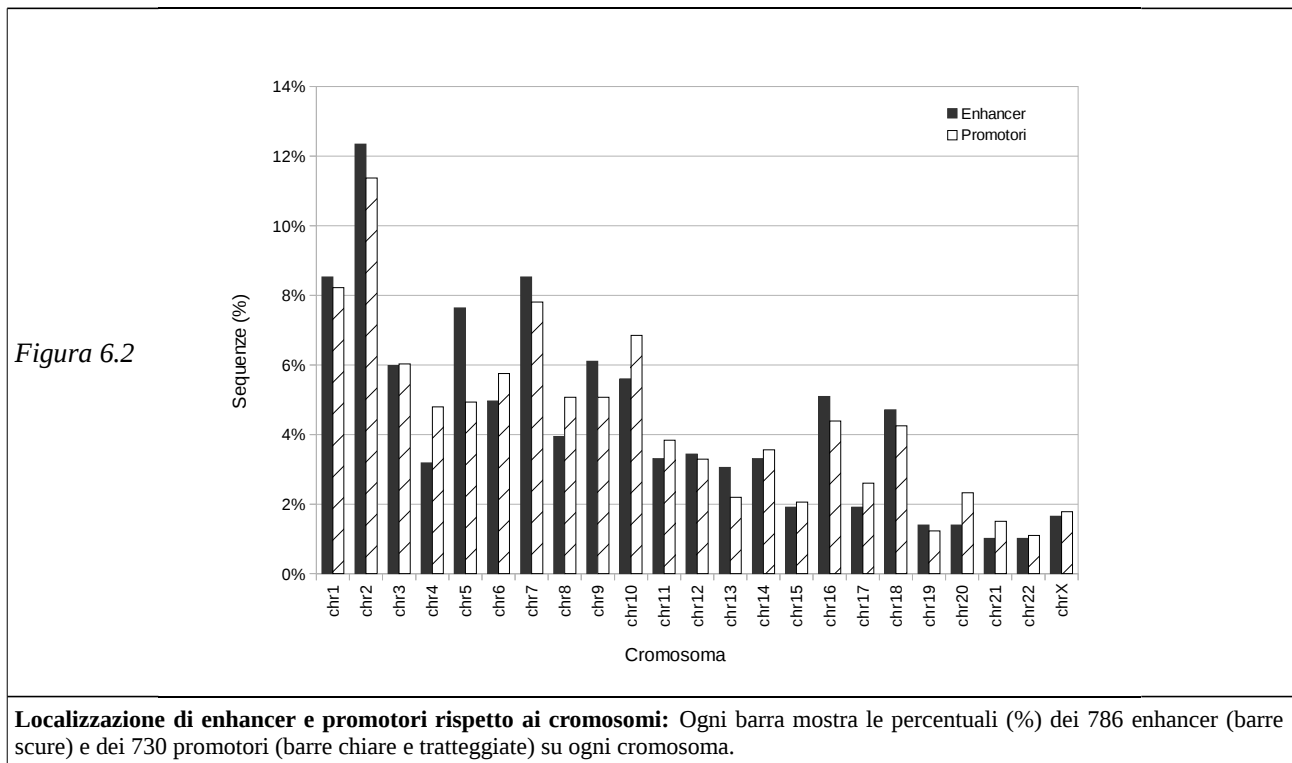
La procedura di selezione degli enhancer da VISTA enhancer browser, l'accoppiamento di tali sequenze ai promotori dei geni fiancheggianti e l'eliminazione della ridondanza sia all'interno delle sequenze enhancer che tra sequenze enhancer e promotore accoppiate ha portato alla generazione di due dataset di sequenze nucleotidiche, il primo composto da 786 enhancer e il secondo da 730 promotori.

Mentre i promotori hanno una lunghezza costante di 2000bp (si veda paragrafo 4.2 e 5.1), gli enhancer hanno una lunghezza variabile e la distribuzione di queste lunghezze è mostrata in *Figura 6.1*. Le lunghezze degli enhancer vanno da un minimo di 429bp a un massimo di 8062bp, e la lunghezza media è di circa 1864bp.



Sia i promotori che gli enhancer sono presenti in proporzioni variabili sui cromosomi dell'organismo umano, come mostrato nel grafico a barre in *Figura 6.2*, dove per ogni cromosoma le percentuali delle sequenze enhancer e promotore sono mostrate da barre fiancheggianti di diverso colore. Le percentuali tra i due tipi di sequenze sono simili, dato che ogni enhancer è accoppiato ai promotori dei geni fiancheggianti. Sebbene la distribuzione di

enhancer e promotori rispetto ai cromosomi non sia uniforme e si osservino cromosomi contenenti più sequenze (es. chr2) rispetto ad altri (es. chr22), questa distribuzione si riferisce solo al set di enhancer e promotori da noi selezionato e ovviamente non implica l'esistenza di una distribuzione non uniforme tra tutti gli enhancer all'interno del genoma umano. Non siamo a conoscenza di preferenze in termini di cromosomi studiati da parte di VISTA enhancer browser ma sappiamo che seleziona tipi di enhancer specifici, la maggior parte dei quali coinvolti nello sviluppo embrionale e altamente conservati, che possono rappresentare solo una frazione degli enhancer funzionali di un organismo.



6.2 Predizione dei SLFT

La predizione di Siti di Legame per Fattori di Trascrizione (SLFT), effettuata indipendentemente sui due dataset di sequenze enhancer e promotore, ha portato all'individuazione di un totale di 5917 SLFT, di cui 3343 su enhancer e 2574 su promotori (*Tabella 6.1*). I siti di legame predetti corrispondono a 162 e 169 Fattori di Trascrizione (FT) su enhancer e promotori rispettivamente (*Tabella 6.2*); di questi 100 FT sono condivisi tra i due set di sequenze. Collettivamente i SLFT corrispondevano a 231 FT. Il modo in cui i FT sono ripartiti tra i set di sequenze enhancer e promotore è riassunto in *Tabella 6.3* ed esposto per esteso in *Lista A4* dell'appendice dove sono riportati tutti gli identificativi dei FT di cui sono stati predetti i SLFT.

Il livello di soglia impostato per accettare o rifiutare le predizioni è equivalente a un p-value di $1 \cdot 10^{-7}$. Questa è una soglia stringente e garantisce che alle predizioni corrispondano segnali significativi di legame presenti sulle sequenze nucleotidiche. Inoltre la predizione è avvenuta su regioni genomiche di cui l'attività

regolatrice è nota o altamente probabile, e questo minimizza ulteriormente la probabilità di ottenere false predizioni. Per quanto riguarda gli enhancer, questi corrispondono a sequenze molto conservate tra genomi di mammiferi e/o vertebrati di cui l'attività di attivatori della trascrizione è validata sperimentalmente. Per quanto riguarda i promotori, non possiamo escludere la possibilità che nell'ambito delle 2000bp di ogni promotore ci siano segmenti che non legano FT in vivo, ma è ragionevole considerare le regioni a monte del sito di inizio della trascrizione come potenziali sequenze regolatrici e l'alto livello di significatività richiesto minimizza il rischio di falsi positivi.

Il costo dell'utilizzo di una soglia stringente è l'aumento di falsi negativi, vale a dire l'incapacità di predire SLFT funzionali in vivo caratterizzati da un segnale relativamente meno forte. A questo effetto è probabilmente dovuta la mancanza di predizione di SLFT per 48 enhancer e 91 promotori. In appendice la *Lista A3* mostra il numero di SLFT predetti e il numero di FT ad essi corrispondenti per ogni sequenza del set di CEP.

<i>Tabella 6.1</i>		Enhancer	Promotori
	Num. SLFT predetti	3343	2574
	Num. Minimo di SLFT per sequenza	0	0
	Num. Massimo di SLFT per sequenza	77	37
	Num. Medio di SLFT per sequenza	4,25	3,53

Siti di legame predetti: La prima riga della tabella mostra il numero totale di Siti di Legame dei Fattori di Trascrizione (SLFT) predetti sull'intero set di enhancer e promotori. In questo conteggio sono considerati tutti i SLFT significativi. Nelle righe successive sono esposte delle statistiche descrittive riguardanti il numero di SLFT predetti per ogni sequenza di ogni dataset.

<i>Tabella 6.2</i>		Enhancer	Promotori
	Num. FT	162	169
	Num. Minimo di FT per sequenza	0	0
	Num. Massimo di FT per sequenza	18	13
	Num. Medio di FT per sequenza	2,88	2,6

FT corrispondenti ai siti di legame predetti: La prima riga mostra il numero totale di Fattori di Trascrizione (FT) corrispondenti ai Siti di Legame dei Fattori di Trascrizione (SLFT) predetti per ogni dataset. In questo conteggio è considerata la presenza di ogni FT di cui almeno un SLFT è stato predetto. Nelle righe successive sono elencate statistiche descrittive riguardanti il numero di FT di cui almeno un SLFT è stato identificato per ogni sequenza.

<i>Tabella 6.3</i>	Num. FT totale	231
	Num. FT solo su enhancer	62
	Num. FT solo su promotori	69
	Num. FT condivisi tra enhancer e promotori	100

Ripartizione FT tra enhancer e promotori: Il totale di 231 FT è ripartito tra enhancer e promotori come qui mostrato. Le identità dei singoli FT sono elencate in *LISTA A4* dell'appendice.

6.3 Set di coppie enhancer - promotore

Il set finale di CEP ottenuto dopo l'eliminazione degli elementi composti da almeno una sequenza priva di predizioni per SLFT conta 1011 CEP, composte da 702 enhancer e 621 promotori. In questo set ogni enhancer può essere associato a uno o due promotori, come mostrato in *Tabella 6.4* il che è consistente con la procedura di generazione delle CEP, dove ogni coppia è formata associando ogni enhancer ai suoi geni fiancheggiati, che possono essere uno o due.

<i>Tabella 6.4</i>	Promotori accoppiati (#)	Enhancer (#)	Enhancer (%)
	1	393	56%
	2	309	44%

Numero di enhancer rispetto al numero di promotori accoppiati: Nel set di coppie enhancer-promotore ogni enhancer è accoppiato a uno o due promotori. La percentuale è calcolata rispetto al numero totale di enhancer pari a 702.

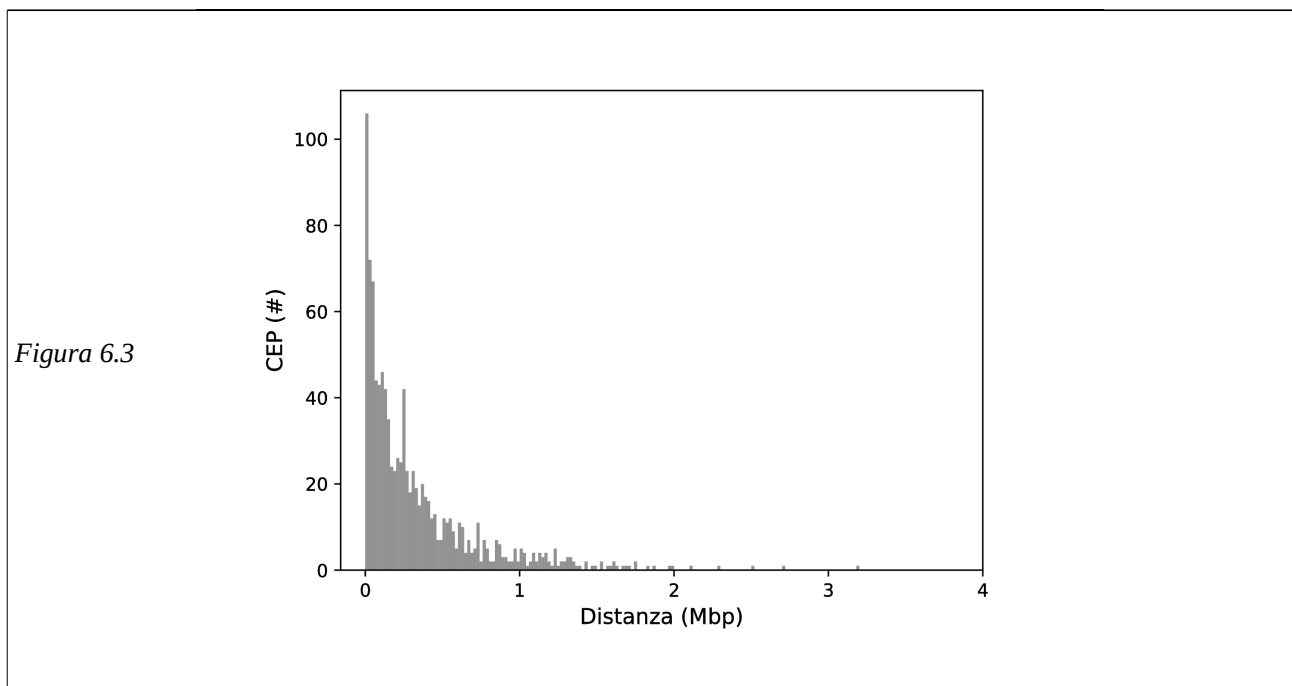
<i>Tabella 6.5</i>	Enhancer accoppiati (#)	Promotori (#)	Promotori (%)
	1	457	74%
	2	81	13%
	3	41	6.6%
	4	11	1.8%
	5	9	1.4%
	6	9	1.4%
	7	4	0.64%
	8	2	0.32%
	9	3	0.48%
	11	1	0.16%
	14	2	0.32%
	16	1	0.16%

Numero di promotori rispetto al numero di enhancer accoppiati: Nel set di coppie enhancer-promotore ogni promotore è accoppiato a un numero variabile di enhancer che dipende dalla loro posizione genomica rispetto agli enhancer circostanti. Le percentuali sono calcolate rispetto al numero totali di promotori pari a 621.

Per quanto riguarda i promotori, come mostrato in *Tabella 6.5*, il numero di enhancer associati ad ogni promotore può andare da un minimo di 1, che è il caso più comune, a un massimo di 16, che è un caso molto raro e che riguarda un solo promotore. Questa variabilità è dovuta al fatto che per diversi enhancer i geni fiancheggiati più vicini possono essere gli stessi. VISTA enhancer browser include tra i suoi enhancer sequenze localizzate in "deserti genici", vale a dire zone caratterizzate da una scarsa presenza di sequenze geniche che coprono all'incirca il 25% del genoma (Nobrega et al., 2003; Visel et al., 2007). In queste regioni sono stati individuati enhancer capaci di ricapitolare aspetti dell'espressione di geni fiancheggiati nonostante le grandi

distanze, tra cui DACH1 (Nobrega et al., 2003) e SALL1 (Pennacchio et al., 2006). I promotori di questi stessi geni risultano nel nostro set di CEP accoppiati a 9 enhancer, il che spiega l'alto numero di enhancer accoppiati per alcuni promotori. I promotori accoppiati a un numero maggiore di enhancer (11, 14 o 16), sono per la maggior parte relativi a loci poco annotati, ad eccezione del promotore di TLE4 che risulta associato a 14 enhancer. La lista completa degli enhancer e la lista dei promotori ordinati tenendo in considerazione il numero di enhancer o promotori accoppiati sono mostrate in appendice in *Lista A5* e *Lista A6*, rispettivamente.

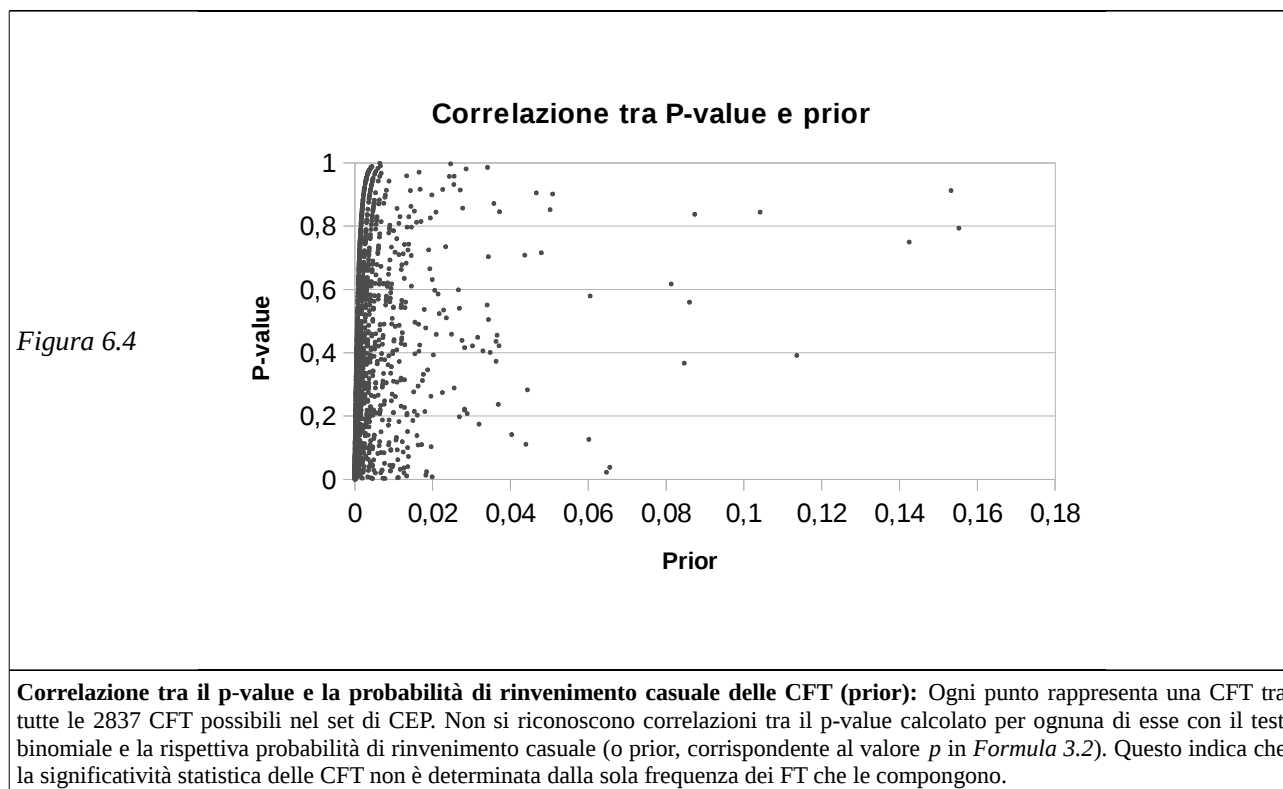
Data la procedura di associazione dei promotori agli enhancer, basata sulla prossimità, ci si aspetterebbe che tutte le distanze tra enhancer e promotori del set di 1011 CEP siano relativamente brevi. Questo è vero nella maggior parte dei casi, ma esistono casi estremi in cui le distanze tra enhancer e promotori possono superare le 3Mbp. La distribuzione delle distanze enhancer-promotore è mostrata nell'istogramma in *Figura 6.3*, dove ogni distanza è calcolata contando le coppie di basi poste tra le estremità più vicine di ogni enhancer ed il rispettivo promotore. Per mostrare chiaramente le distanze tra gli elementi di tutte le CEP è stata scelta una dimensione dei bin pari a 20kbp e l'asse delle ascisse mostra le distanze in Mbp. La distribuzione è asimmetrica, spostata verso sinistra a indicare che la maggior parte delle CEP sono relativamente vicine. La distanza minima osservata è pari a 345bp, la massima 3198122bp. La presenza di enhancer particolarmente distanti dai rispettivi promotori può essere ricondotta ai sopra menzionati deserti genici.



Distribuzione delle distanze tra enhancer e promotori: Istogramma che mostra la distribuzione delle distanze tra ogni enhancer ed ogni corrispettivo promotore di ogni CEP. Le distanze sono misurate in bp, la grandezza dei bin è pari a 20000bp. I valori dell'asse delle ascisse rappresentano intervalli di 1Mbp. L'asse delle ordinate rappresenta il numero (#) di CEP. Il numero totale di CEP è pari a 1011. La maggior parte degli enhancer e dei promotori delle coppie sono vicini tra di loro (distanza minima 345bp), ma si nota la presenza di alcuni casi estremi in cui le distanze sono molto grandi, superando le 3Mb (distanza massima 3198122bp), il che è riconducibile alla presenza nel set di enhancer di sequenze poste in "deserti genici" e quindi lontani da sequenze geniche. In bp: Valore minimo 345; Valore massimo 3198122; Media 334084,7; Mediana: 206023, deviazione standard: 394514,9.

6.4 Coppie di fattori di trascrizione significative

Mediante il test binomiale, tra le 2837 CFT che possono essere costituite accoppiando tutti i FT rispettivi ai SLFT predetti sulle sequenze del set di CEP, sono state identificate 364 CFT significative a un livello di significatività pari a $5 \cdot 10^{-2}$, tutte validate con la procedura di bootstrap. Queste coppie sono composte da 196 FT. L'elenco di tutte le CFT significative completo dei p-value e dei dati utilizzati per calcolarli è mostrato in appendice nella *Lista A7*. I p-value calcolati variano da un minimo di $6,9 \cdot 10^{-5}$ a un massimo di $4,9 \cdot 10^{-2}$. *Figura 6.4* mostra la correlazione tra il p-value calcolato con il test binomiale e la probabilità di rinvenimento casuale di ogni possibile CFT. Non si notano correlazioni, a indicare che la significatività statistica così calcolata non è dipendente dalla sola frequenza dei FT accoppiati. Con questo test è stato possibile riconoscere CFT significative costituite sia da FT frequenti all'interno del set che da FT meno frequenti. Infatti il numero di CEP associate alle CFT varia da un minimo di 1 a un massimo di 82 con un valore medio pari a 3. Le CFT significative riguardano un totale di 377 CEP composte da 315 enhancer e 293 promotori. La lista completa di CEP per ogni CFT significativa è mostrata in *Lista A8* dell'appendice.



Espressione delle CFT significative

In accordo ai dati di espressione provenienti da EBI Expression Atlas e dai profili proteomici provenienti da Human Proteome Map, la gran parte delle CFT significative risulta composta da FT che sono espressi in almeno un tessuto comune. È stato possibile rinvenire i pattern di espressione per tutti i FT delle CFT significative

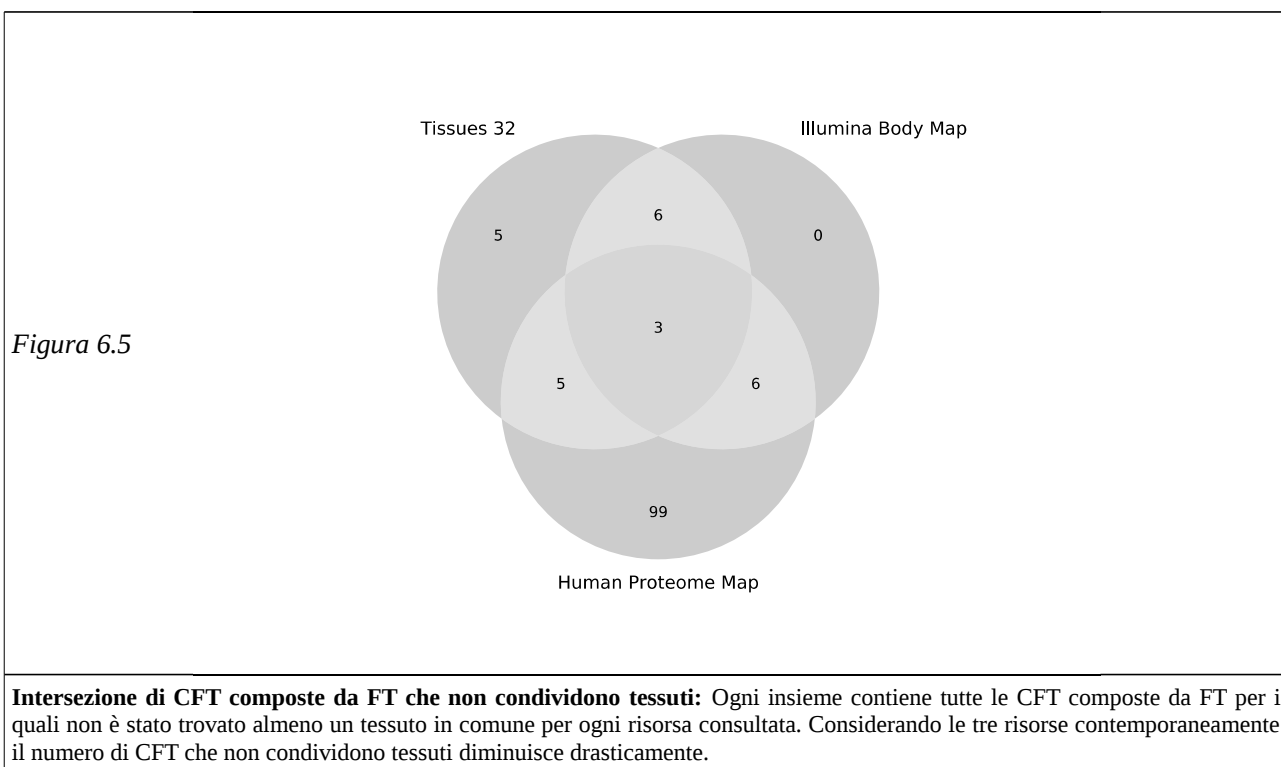
considerando i dati RNA-Seq selezionati (Tissues 32 ed Illumina Body Map). Human Proteome Map invece presentava profili per un numero più limitato di FT. Ciononostante questa risorsa è stata inclusa per integrare nell'analisi informazioni riguardanti l'effettiva presenza della proteina tradotta.

Considerando le risorse singolarmente in tutti e tre i casi, la maggior parte delle CFT risulta composta da FT espressi in almeno un tessuto in comune (Tabella 6.6). Il numero di CFT composte da FT che non condividono tessuti diminuisce drasticamente se si considerano i dati di tutte e tre le risorse. Questo è mostrato in Figura 6.5 dove ogni insieme contiene tutte le CFT che non condividono tessuti per ogni risorsa. La loro intersezione mostra la presenza di sole 3 CFT per i quali i FT non risultano espressi in almeno un tessuto in comune in tutti e tre i casi. Ad ogni modo, se si prendessero in considerazione livelli di espressione più bassi di quelli da noi selezionati anche queste 3 CFT risulterebbero composte da FT espressi negli stessi tessuti.

Tabella 6.6

	Tissues 32	Illumina body map	Human Proteome Map
Tessuti	32	16	30
FT inclusi nella risorsa consultata	196/196 (100%)	196/196 (100%)	185/196 (94%)
CFT composte da FT inclusi	364/364 (100%)	364/364 (100%)	335/364 (92%)
CFT composte da FT espressi in tessuti comuni	345/364 (95%)	349/364 (96%)	251/335 (75%)

CFT espresse in almeno un tessuto comune: In tutti e tre le risorse selezionate la maggior parte delle CFT è composta da FT espressi in almeno un tessuto in comune. "Tissues 32" ed "Illumina Body Map" mettono a disposizione dati di tipo RNA-Seq, "Human Proteome Map" dati di tipo proteomico. A differenza degli altri quest'ultimo non forniva informazioni riguardanti tutti i FT componenti le CFT significative.



Comparazione con l'interattoma umano

I FT delle CFT significative sono più vicine di quanto atteso all'interno dell'interattoma umano. In *Tabella 6.7* è mostrato il numero di proteine intermedie necessario a mettere in comunicazione le 356 CFT composte dai 193 FT che è stato possibile identificare in STRING. Nella stessa tabella è anche mostrato il numero di proteine intermedie necessario a mettere in comunicazione i FT e la RNAP II. Attraverso un test binomiale è stato accertato che queste distanze sono significativamente ridotte se paragonate alla distribuzione dei cammini minimi di STRING. I risultati dei due test binomiali, uno per le CFT e l'altro per le coppie FT-RNAP II sono mostrati in *Tabella 6.8*.

Tabella 6.7

Proteine intermedie (<i>b</i>)	CFT	FT-RNAPII
0	25 (7%)	4 (2%)
1	205 (58%)	158 (82%)
2	118 (33%)	31 (16%)
3	8 (2%)	0 (0%)

Proteine intermedie: Lunghezza dei cammini minimi osservati tra i FT di ogni CFT e tra gli stessi FT e la RNAPII.

Tabella 6.8

CFT

Proteine intermedie (<i>b</i>)	CFT collegate da <i>b</i> o meno intermedi (<i>k</i>)	Coppie attese collegate da <i>b</i> o meno intermedi	Prior (<i>p</i>)	P-value
0	25	0,9	$2,6 \cdot 10^{-3}$	$1,6 \cdot 10^{-27}$
1	230	37	$1,0 \cdot 10^{-1}$	$6,6 \cdot 10^{-135}$
2	348	215	$6,0 \cdot 10^{-1}$	$1,0 \cdot 10^{-64}$
3	356	328	$9,2 \cdot 10^{-1}$	$2,2 \cdot 10^{-13}$

FT-RNAP II

Proteine intermedie (<i>b</i>)	FT collegati a RNAP II da <i>b</i> o meno intermedi (<i>k</i>)	Coppie attese collegate da <i>b</i> o meno intermedi	Prior (<i>p</i>)	P-value
0	4	0,5	$2,6 \cdot 10^{-3}$	$1,7 \cdot 10^{-3}$
1	162	20	$1,0 \cdot 10^{-1}$	$1,6 \cdot 10^{-126}$
2	193	116	$6,0 \cdot 10^{-1}$	$3,4 \cdot 10^{-43}$

Prossimità delle CFT: Risultati del test binomiale volto a determinare la probabilità di rinvenire in STRING un numero di coppie di proteine superiore o uguale a quello osservato (*k*) collegato da un numero inferiore o uguale di proteine intermedie (*b*). Il prior *p* è ricavato dalla distribuzione cumulativa delle lunghezze dei cammini minimi di STRING. Sia le CFT che le coppie FT-RNAPII risultano significativamente vicine in termini di proteine intermedie. Il numero di coppie attese è ottenuto moltiplicando la probabilità *p* per il numero di coppie presenti in STRING (356 per CFT, 193 per FT-RNAPII).

Di particolare interesse è l'analisi delle proteine intermedie più frequentemente coinvolte nella composizione di un cammino minimo. Come si vede in *Lista A9* dell'appendice, dove sono riportate le proteine che più frequentemente fanno da intermedio alle CFT, quelle più frequenti sono CREBBP ed EP300. Inoltre in diversi casi è possibile riconoscere subunità del complesso Mediator tra cui: MED1, MED6, MED12, MED14, MED15, MED17, MED21, MED23, MED24 (Questi identificativi indicano le subunità 1, 6, 12, 14, 15, 17, 21, 23, 24 di "Mediator of RNA polymerase II transcription"), CDK8 (Cyclin-dependent kinase 8). Come discusso nel paragrafo 1.3 queste proteine agiscono da ponte dei FT e sono possibili mediatori del loop enhancer-promotore.

Gli enrichment dei termini Gene Ontology (GO-term) "cellular component" associati alle proteine intermedie che possono collegare i FT delle CFT tra loro (si veda *Lista A10* in appendice) e gli stessi FT con RNAP II (*Lista A12*) mostrano termini arricchiti riguardanti il Mediator (GO:0016592; "mediator complex") e la localizzazione nel nucleo (GO:0005634; "nucleus"). Le stesse proteine nella sezione "biological process" (*Lista A11* e *Lista A13*) mostrano anche termini collegati ai processi dello sviluppo (es. GO:0048856; "anatomical structure development").

La presenza degli intermedi elencati nei cammini minimi, le annotazioni associate agli intermedi, e la prossimità che caratterizza questi cammini minimi suggerisce che i componenti delle CFT potrebbero facilmente collaborare e possibilmente partecipare alla formazione di complessi proteici coinvolti nella formazione/stabilizzazione dei loop.

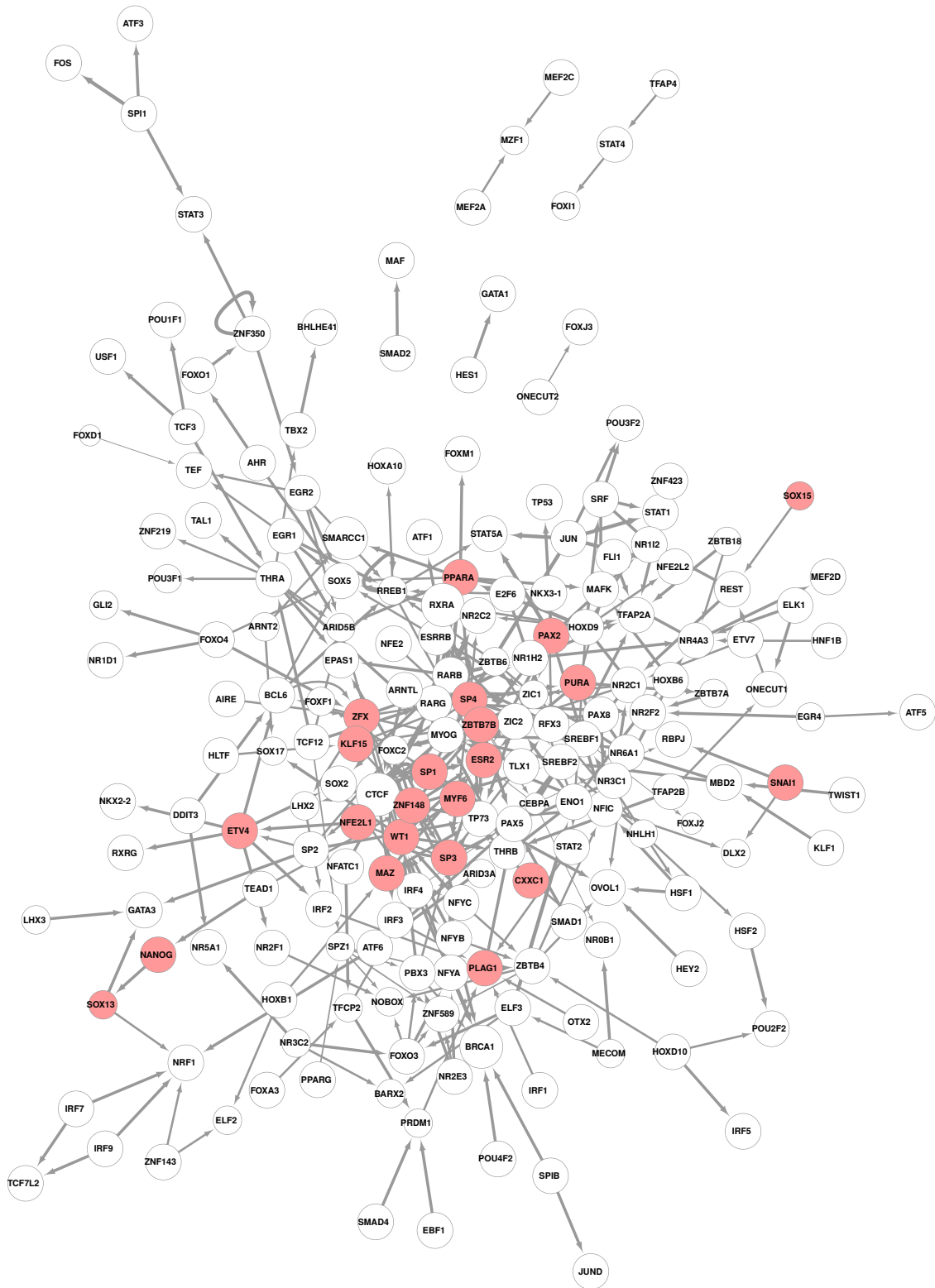
Grafo delle CFT

Le CFT significative possono essere rappresentate sotto forma di grafo, dove i nodi corrispondono ai FT e gli archi indicano la formazione di una CFT significativa. Il grafo così ottenuto è composto da 196 nodi (FT) e 364 archi (CFT) ed è mostrato nella sua interezza in *Figura 6.6*. Il grafo è caratterizzato da un coefficiente di clustering pari a 0,018, un cammino minimo medio pari a 4,3, un diametro di 11 e il numero medio di vicini è 3,6. È privo di nodi isolati ed è composto da 6 componenti connesse, di queste 3 sono composte da 2 nodi ciascuna, 2 da 3 nodi ciascuna e la rimanente componente principale è formata da 184 nodi e 357 archi.

La distribuzione dei gradi è mostrata in *Figura 6.7*, nonostante le dimensioni ridotte della rete si nota la presenza di un numero non nullo, seppur limitato, di nodi caratterizzati da un alto grado che può essere ricondotta alla presenza di hub. Questo suggerisce che la rete così ottenuta non è casuale. La *Lista A14* in appendice mostra l'elenco completo di tutti i nodi in ordine decrescente di grado.

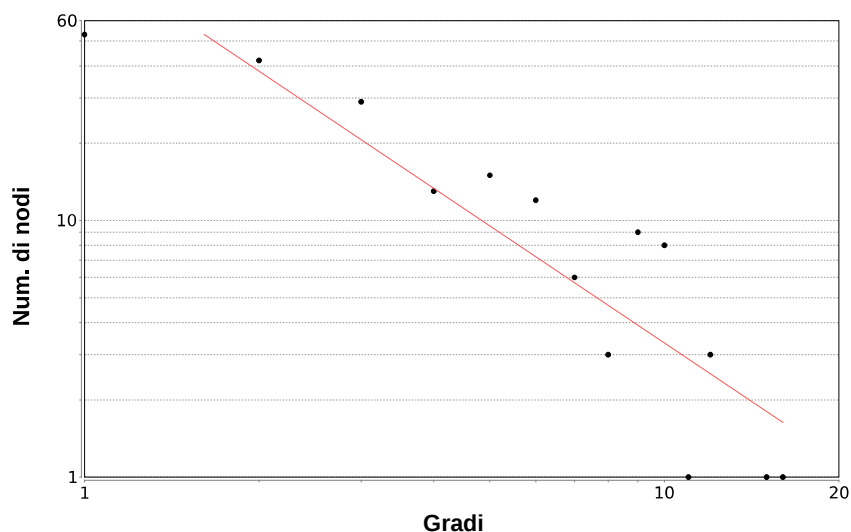
Il grafo così composto contiene tutte le CFT significativamente presenti nel set di CEP e rappresenta interazioni tra FT putativamente coinvolti nella formazione di loop tra enhancer e promotori, le cui interazioni potrebbero essere importanti per lo sviluppo embrionale dell'organismo umano. Questa possibilità è avvalorata, come scritto sopra, dalla composizione e lunghezza dei cammini minimi che congiungono le CFT e la RNAP II all'interno dell'interattoma umano, dalla presenza degli elementi delle CFT negli stessi tessuti e, come vedremo successivamente, anche dall'analisi delle CFT particolarmente significative e delle annotazioni associate ai geni del grafo.

Figura 6.6



Grafo delle CFT: Ogni nodo rappresenta un FT, ogni arco rappresenta un'interazione significativa identificata con il test binomiale. Ogni coppia di nodi connessi rappresenta una CFT, la direzione degli archi indica la localizzazione dei FT (ogni arco parte dal FT all'enhancer e termina al FT al promotore). Lo spessore degli archi è proporzionale alla prossimità dei FT in termini di numero di proteine intermedie. Similmente la grandezza dei nodi è proporzionale alla prossimità dei FT alla RNAP II. Il colore rosso di alcuni nodi indica che tali FT sono presenti anche all'interno dei gruppi di FT individuati con il biclustering.

Figura 6.7



Grado	Nodi (#)	Nodi (%)
16	1	1%
15	1	1%
14	0	0%
13	0	0%
12	3	2%
11	1	1%
10	8	4%
9	9	5%
8	3	2%
7	6	3%
6	12	6%
5	15	8%
4	13	7%
3	29	15%
2	42	21%
1	53	27%

Distribuzione dei gradi: Il grado di un nodo (o "degree") rappresenta il numero di nodi a cui è connesso tramite archi. La maggior parte dei nodi ha un grado basso. La linea rossa rappresenta la legge di potenza nella forma $y=ax^b$ calcolata con il metodo dei minimi quadrati. $a=109,01$; $b=-1,515$; correlazione= $0,893$; $R\text{-squared}=0,808$. Nonostante le dimensioni ridotte della rete (196 nodi, 364 archi) si nota la presenza di un numero non nullo seppur limitato di nodi con un alto grado. Questo indica la presenza di hub all'interno della rete e suggerisce che la rete non è casuale. Nella tabella a destra le percentuali sono calcolate in considerazione del numero di nodi (196).

Annotazioni dei geni delle CFT significative

I FT che compongono le CFT significative e i geni delle CEP associati sono arricchiti per termini Gene Ontology (GO-term) simili. Ogni CFT può essere localizzata su un numero variabile di CEP, ed ogni CEP può essere ricondotta al gene del rispettivo promotore. L'intero set di CEP che presentano CFT significative conta 377 elementi. I promotori all'interno di questo set sono 293. Dei geni di questi promotori è stato possibile ottenere l'identificativo Uniprot per 257 elementi. Un'analisi di enrichment dei GO-term "Biological Process" di questi geni mostra che questi sono in gran parte implicati in funzioni relative alla regolazione della trascrizione e allo sviluppo embrionale, il che è in sintonia con quanto atteso data la localizzazione dei geni in prossimità di enhancer importanti per lo sviluppo. Tra i GO-term significativi si nota la presenza di termini specifici per lo sviluppo di una varietà di tessuti tra cui tessuto nervoso, cardiaco e renale. (si veda *Lista A15* in appendice per i dettagli). La significativa presenza di termini simili si osserva anche quando si effettua un enrichment per i 196 FT componenti le CFT significative (risultati esposti in *Lista A16* dell'appendice). La presenza di questi stessi termini sia tra i FT che tra i geni conferma la validità dei SLFT predetti e la selezione delle CFT significative.

CFT: Caso di studio

Tra le CFT più significative si nota la presenza di buoni candidati per la regolazione della trascrizione mediata dal looping della cromatina. Le CFT 4, 5 e 6 sono tutte caratterizzate da un p-value pari a $1,1 \cdot 10^{-4}$ e sono composte dal FT RARG (Retinoic acid receptor gamma) all'enhancer e dai FT NR2C2 (Nuclear receptor subfamily 2 group C member 2), PPARA (Peroxisome proliferator-activated receptor alpha) ed RXRA (Retinoic acid receptor RXR-alpha) al promotore. Tutte e tre sono localizzate sulle stesse due CEP: hs1931,CCDC85C (la coppia è qui rappresentata scrivendo il VISTA ID dell'enhancer e il gene symbol del gene del promotore corrispondente separati da una virgola; una CEP può anche essere rappresentata dal suo codice identificativo, in questo caso CEP 110) ed hs1971,RBM38 (CEP 759). VISTA enhancer browser annota entrambi gli enhancer come regolatori dell'espressione nel tessuto cardiaco. RBM38 (RNA-binding protein 38) è una proteina capace di legarsi ad RNA preferenzialmente espressa nei tessuti muscolare e cardiaco coinvolta nel differenziamento dei miociti (Miyamoto et al., 2009). CCDC85C (Coiled-coil domain-containing protein 85C) è una proteina relativamente poco annotata, Gene Ontology riporta termini riguardanti lo sviluppo della corteccia cerebrale, ciononostante in accordo ai dati RNA-Seq dello studio "Tissues 32" precedentemente citato è espressa in una varietà di tessuti incluso cuore, tessuto muscolare liscio e scheletrico.

RARG è un recettore dell'acido retinoico (RA). RA gioca un ruolo nella regolazione dello sviluppo embrionale e la sua funzione è mediata da recettori (indicati con la sigla RAR) i quali formano eterodimeri con i recettori X dell'acido retinoico (indicati con la sigla RXR) (Zuchegna et al., 2015). È stato osservato che RA è responsabile di cambiamenti conformazionali della cromatina che danno luogo alla formazione di loop (Zuchegna et al., 2015) ed è stato anche postulato un modello secondo il quale proteine RXR potrebbero essere direttamente coinvolte nella formazione di loop della cromatina (Yasmin et al., 2004). Nella CFT 6, RARG è associato ad RXRA. Le due proteine sono direttamente connesse in STRING, possono dimerizzare ed entrambe giocano un ruolo nello sviluppo cardiaco (Dollé, 2009; Mark et al., 2009). Queste evidenze suggeriscono che la CFT 6 potrebbe avere un ruolo per la generazione e/o stabilizzazione di loop della cromatina.

Le altre due CFT mostrano PPARA (CFT 5) e NR2C2 (CFT 4) associati a RARG. PPARA è un membro della superfamiglia dei recettori nucleari, espleta la sua funzione eterodimerizzando con proteine RXR come RXRA (Gorla-Bajszczak et al., 1999) e nel database Ensembl (Yates et al., 2016) è annotata con il GO-term "heart development". NR2C2 invece agisce da repressore del pathway dell'acido retinoico e può competere con proteine RXR legandosi agli stessi SLFT (Hirose et al., 1995). Il rinvenimento sulle stesse CEP di diverse CFT composte da proteine coinvolte negli stessi pathway conferma la rilevanza biologica delle CFT significative.

6.5 Gruppi di fattori di trascrizione significativi

Oltre all'analisi di Coppie di Fattori di Trascrizione (CFT) su Coppie Enhancer-Promotore (CEP), con la tecnica del biclustering si procede all'analisi di un livello di complessità maggiore, questo presuppone che il looping enhancer-promotore dipenda dall'interazione di più FT su enhancer e di più FT al promotore. In questo senso, la procedura di buclustering ha portato all'identificazione di 48 bicluster significativi, con un livello di significatività pari a $1 \cdot 10^{-3}$, composti da 222 CEP (a loro volta composte da 200 enhancer e 171 promotori) e 24 FT, di cui 7 condivisi tra enhancer e promotori, 12 specifici per gli enhancer e 5 specifici per i promotori.

Ogni bicluster può contenere da un minimo di 2 a un massimo di 6 FT (circa 3 in media), e da un minimo di 7 a un massimo di 32 CEP (circa 13 in media). I FT agli enhancer variano da 1 a 4 (circa 1 in media), quelli ai promotori da 1 a 5 (circa 2 in media). Le sequenze enhancer vanno da un minimo di 6 a un massimo di 32 (media circa 12), le sequenze promotore da 4 a 27 (media circa 12). Infine i p-value dei bicluster variano da un minimo di $5 \cdot 10^{-29}$ a un massimo di $5.4 \cdot 10^{-4}$. I valori specifici per ogni singolo bicluster sono elencati in *Lista A17* dell'appendice; in *Lista A18* invece sono elencati tutti i FT e tutte le CEP per ogni bicluster significativo.

Annotazioni dei geni dei bicluster significativi

I geni dei bicluster significativi sono arricchiti per GO-term riguardanti lo sviluppo embrionale, in particolare lo sviluppo del sistema nervoso. In totale i 48 bicluster contengono 171 promotori, associabili a un eguale numero di geni. Di questi è stato possibile rinvenire il codice identificativo Uniprot per 163 geni, i quali sono stati utilizzati per le analisi di enrichment. I geni di ogni bicluster sono stati analizzati impostando un livello di soglia pari a un p-value di $5 \cdot 10^{-2}$. Dei 48 bicluster significativi 43 sono risultati arricchiti per almeno un termine appartenente alla sezione "biological process". Il 51% di questi risulta arricchito per il termine GO:0030182 (neuron differentiation) che risulta il termine più comune. Il termine GO:0006357 (regulation of transcription from RNA polymerase II promoter) risulta arricchito per il 44% dei bicluster. Termini riguardanti lo sviluppo embrionale specifici di altri tessuti sono anche presenti ma per un numero limitato di bicluster e riguardano il 16% dei bicluster o meno. In *Lista A19* dell'appendice sono mostrati questi risultati per esteso. Questi risultati indicano che la maggior parte dei bicluster significativi potrebbero avere un ruolo nella regolazione dello sviluppo.

FT dei bicluster significativi

I FT inclusi nei bicluster sono per la maggior parte presenti anche nel set di FT delle CFT significative. Dei 24 FT totali solo due non sono condivisi con le CFT (PGR e MECP2). La maggior parte dei rimanenti 22 coincide con i nodi di grado maggiore del grafo di CFT, come mostrato sia in *Figura 6.6* che in *Lista A14*. Il rinvenimento dei FT più connessi del grafo anche con la procedura di biclustering conferma un ruolo centrale di questi elementi ed indica la presenza di un segnale considerevole che è stato identificato con entrambe le procedure.

Tra i FT condivisi e caratterizzati da un alto grado nel grafo si nota la presenza di: SP1 (Transcription factor Sp1), SP3 (Transcription factor Sp3), SP4 (Transcription factor Sp4), ZNF148 (Zinc finger protein 148). Di SP1 è stato già discusso nel paragrafo 1.3 e sono state mostrate le sue caratteristiche di FT ubiquitario e di importante regolatore della trascrizione che può esercitare la sua funzione anche mediando la formazione di loop enhancer-promotore. Tale FT è anche importante durante lo sviluppo embrionale e può essere coinvolto nel differenziamento del tessuto nervoso (Okamoto et al., 2002; Zhao and Meng, 2005). SP3 è un FT bifunzionale che può sia assistere che ostacolare la modulazione della trascrizione effettuata da SP1, possiede un dominio di legame al DNA altamente simile a quello di SP1 (Li and Davie, 2010) e anch'esso è espresso in maniera ubiquitaria. Inoltre è coinvolto in diversi processi dello sviluppo embrionale dei vertebrati (Zhao and Meng, 2005). Rispetto ad SP1 ed SP3, SP4 mostra pattern di espressione più specifici e risulta altamente espresso nel tessuto nervoso durante lo sviluppo embrionale (Supp et al., 1996; Zhao and Meng, 2005). Infine ZNF148 è un FT espresso in maniera aspecifica che può opporsi o cooperare all'attività regolatrice di SP1 e che potrebbe collaborare nel reclutamento di EP300 per la formazione di un complesso di attivazione della trascrizione comprendente tutte e tre queste proteine (Bai and Merchant, 2000).

La tecnica di biclustering permette di evidenziare FT che sono frequenti anche tra le coppie derivanti dal test binomiale e sono comunque coinvolti in meccanismi di regolazione comuni. Ciò suggerisce che i FT individuati abbiano un'elevata probabilità di assolvere un ruolo fondamentale per la formazione dei loop. Ad esempio sono state identificate 3 CFT contenenti SP1 e ZNF148 ad enhancer e/o promotore (CFT numero 67, 113, 326 di *Lista A8* dell'appendice), e gli stessi FT sono presenti all'interno di 5 Biclustri (Bic numero 1, 4, 5, 8, 12 di *Lista A18* dell'appendice). Nei biclustri questi FT sono accompagnati da altri (tra cui SP3 ed SP4) e disposti su enhancer e promotori in varie combinazioni. Queste combinazioni, specifiche per gruppi di CEP e di conseguenza gruppi di geni, potrebbero rappresentare un livello superiore di regolazione, non visualizzabile con le più semplici coppie risultanti dal test binomiale.

Biclustri: Caso di studio

Il biclustri numero 5 è tra i biclustri più significativi con p-value pari a $1.8 \cdot 10^{-21}$. È composto da 11 CEP a loro volta comprendenti 9 enhancer e 11 promotori ed inoltre da 3 FT agli enhancer e 1 FT ai promotori. Il FT ai promotori è SP1, il quale si trova anche tra i FT agli enhancer, insieme a SP3 e ZNF148. Come esposto precedentemente sia SP3 che ZNF148 sono FT capaci di cooperare con SP1 per la modulazione dell'espressione; inoltre SP1 è anche noto per essere capace di mediare la formazione di loop della cromatina mediante la formazione omodimeri (Mastrangelo et al., 1991; Su et al., 1991).

Un'analisi di enrichment dei GO-term relativi alla sezione "Biological Process" (effettuata con NET-GE applicando la correzione Benjamini-Hockberg e mostrata in *Lista A20* dell'appendice), indica che degli 11 geni associati alle CEP, 8 sono coinvolti a loro volta nella regolazione dell'espressione (GO:0010468; "regulation of gene expression") e 5 sono coinvolti nella regolazione della neurogenesi (GO:0050767; "regulation of neurogenesis"). Queste funzioni sono consistenti con l'attività di SP1 di regolatore dello sviluppo del tessuto

nervoso citata precedentemente. Questo bicluster individua quindi un gruppo di FT coinvolti nella regolazione dell'espressione di queste sequenze mediate da looping della cromatina.

6.6 Formazione del loop: Confronto tra biclustering e test binomiale

Gli stessi Fattori di Trascrizione (FT) ad enhancer e promotore di 19 Coppie di Fattori di Trascrizione (CFT) significative (secondo il test binomiale) sono presenti all'interno di 18 bicluster significativi. Queste coppie con i relativi bicluster sono tutte mostrate in *Tabella 6.9*. In due casi la stessa CFT si trova all'interno di due bicluster (CFT 113 e CFT 171) ed in un solo caso la stessa CFT si trova in sei bicluster (CFT 282). Esistono anche casi in cui lo stesso bicluster può contenere più di una CFT (BIC: 2, 3, 10, 12, 14, 24).

Tabella 6.9

Coppie di FT (test binomiale)			Bicluster		
ID	FT enhancer	FT promotore	ID	FT enhancer	FT promotore
CFT20	KLF15	ZBTB7B	BIC24	KLF15	SP4,ZBTB7B
CFT21	SP4	ZNF148	BIC3	SP4,ZBTB7B,SP3	ZNF148
CFT39	PAX2	ZBTB7B	BIC16	PAX2	SP1,SP4,ZBTB7B
CFT67	ZNF148	ZNF148	BIC48	ZNF148,MYF6	ZNF148
CFT79	ZNF148	SP3	BIC12	ZNF148	SP1,SP3
CFT99	ZNF148	ZBTB7B	BIC18	ZNF148	SP4,ZBTB7B
CFT113	ZNF148	SP1	BIC5	ZNF148,SP1,SP3	SP1
			BIC12	ZNF148	SP1,SP3
CFT167	KLF15	SP4	BIC24	KLF15	SP4,ZBTB7B
CFT171	MYF6	SP4	BIC6	MYF6	SP4,ZBTB7B,MAZ
CFT188	MYF6	WT1	BIC22	MYF6,SP3	SP4,ZBTB7B
			BIC14	MYF6	WT1,SP3
CFT197	ZBTB7B	ESR2	BIC2	SP1,SP4,ZBTB7B,SP3	ESR2
CFT198	SP3	ZNF148	BIC3	SP4,ZBTB7B,SP3	ZNF148
CFT229	SP1	ZBTB7B	BIC36	SP1	ZBTB7B,SP3
CFT265	SP4	SP1	BIC10	MYF6,SP4	SP1,SP3
			BIC1	MYF6	ZNF148,SP1,SP3
			BIC8	MYF6,PPARA	ZNF148,SP1,ZBTB7B,SP3
			BIC10	MYF6,SP4	SP1,SP3
			BIC14	MYF6	WT1,SP3
			BIC21	MYF6	ZNF148,KLF15,SP3
CFT304	SP1	ESR2	BIC40	MYF6	ZFX,SP3
			BIC2	SP1,SP4,ZBTB7B,SP3	ESR2
CFT305	SP4	SP4	BIC7	ZNF148,SP4,MAZ	SP4
CFT327	SP3	ESR2	BIC2	SP1,SP4,ZBTB7B,SP3	ESR2
CFT329	CXXC1	SP1	BIC32	CXXC1	SP1,SP3

Coppie di Fattori di Trascrizione presenti nei bicluster: Elenco delle Coppie di Fattori di Trascrizione (CFT) i cui Fattori di Trascrizione (FT) sono presenti all'interno dei bicluster. 19 CFT sono contenute in 18 bicluster. Le CFT sono mostrate sul lato sinistro nel campo "Coppie di FT (test binomiale)", i rispettivi bicluster sul lato destro nel campo "Bicluster". I campi "ID" mostrano i codici identificativi di CFT e bicluster, nei campi "FT enhancer" ed "FT promotore" sono elencati i FT presenti sulle sequenze corrispondenti.

I bicluster tendono a contenere le CFT caratterizzate dal più alto numero di Coppie Enhancer-Promotore (CEP), vale a dire le CFT più frequentemente presenti su enhancer e promotori (si veda paragrafo 6.4). Questo è mostrato in *Tabella 6.10*, dove sono elencate le 27 CFT significative con il più alto numero di CEP assieme ai bicluster che le contengono. In tutti i casi le CEP comprese nei bicluster sono un sotto-insieme delle CEP associate alla CFT corrispondente.

Tabella 6.10

CEP (#)	CFT (ID, FT)	Bic contenenti CFT (ID)
82	171 (MYF6,SP4)	6, 22
81	282 (MYF6,SP3)	1, 8, 10, 14, 21, 40
32	79 (ZNF148,SP3)	12
29	113 (ZNF148,SP1)	5, 12
28	188 (MYF6,WT1)	14
23	99 (ZNF148,ZBTB7B)	18
21	67 (ZNF148,ZNF148)	48
21	305 (SP4,SP4)	7
21	45 (ZNF148,ESR2)	
21	166 (MAZ,ESR2)	
20	265 (SP4,SP1)	10
19	229 (SP1,ZBTB7B)	36
17	21 (SP4,ZNF148)	3
16	304 (SP1,ESR2)	2
16	326 (SP1,ZNF148)	
16	197 (ZBTB7B,ESR2)	2
16	198 (SP3,ZNF148)	3
16	28 (PPARA,SOX5)	
15	327 (SP3,ESR2)	2
13	224 (NFE2L1,ETV4)	
13	167 (KLF15,SP4)	24
13	208 (SNAI1,DLX2)	
12	20 (KLF15,ZBTB7B)	24
11	39 (PAX2,ZBTB7B)	16
11	164 (NANOG,SOX13)	
10	223 (PURA,ESR2)	
9	329 (CXXC1,SP1)	32

I bicluster tendono a contenere le coppie di FT col maggior numero di coppie enhancer-promotore: La tabella mostra le 27 Coppie di Fattori di Trascrizione (CFT) significative con il maggior numero di Coppie Enhancer-Promotore (CEP) associate. Il numero di CEP è mostrato nel primo campo "CEP (#)", segue il codice numerico identificativo della CFT con i FT che la compongono tra parentesi. L'ultima colonna mostra i codici numerici identificativi dei Bicluster (Bic) che contengono ogni CFT. Si nota che i bicluster tendono a contenere le CFT con il più alto numero di CEP, vale a dire le CFT più frequentemente presenti su enhancer e promotori.

Nei casi in cui la stessa CFT, individuata con il test binomiale, sia presente in diversi bicluster, è interessante analizzare quali geni siano regolati dalla coppia, ma anche dalla coppia in associazione agli altri FT. Ad esempio, la CFT 171, è la coppia di fattori di trascrizione con il più alto numero di Coppie Enhancer Promotore (CEP) ed è caratterizzata dall'interazione dei FT MYF6 ed SP4. Le sue 82 CEP corrispondono a 79 enhancer e 65 promotori, che regolano 65 geni, come mostrato in Figura 6.8. La coppia è comunque presente in associazione con altri FT in Bicluster (Bic) 6 e Bic 22. La coppia, in Bic 6, che contiene 18 CEP corrispondenti a 18 enhancer e 14 promotori, controlla insieme a MAZ e ZBTB7B 14 geni. In Bic 22, che contiene 9 CEP corrispondenti a 8 enhancer e 9 promotori, la stessa coppia insieme a SP3 e ZBTB7B controlla 9 geni.

Figura 6.8

	FT enhancer	FT promotore	AGAP1	ARID3B	C1orf198	C5orf38	CBLN1	CDYL	CELF4	CISH	DLX1	ECI2	EMX1	ERBB4	ERGIC1	FAM78A	FIGN	FOXP4	GDPD5	HEMK1	IRX2	LINC00466	LOC100134229	LOC283050
CFT 171	MYF6	SP4																						
Bic 6	MYF6	MAZ,SP4,ZBTB7B																						
Bic 22	MYF6,SP3	SP4,ZBTB7B																						

	MECOM	MED13L	MGMT	MRPL23	NBEA	OLA1	PBX3	PDK4	POU3F2	QKI	RANBP17	SMARCA2	SOX21	TBC1D8	TEAD1	TLE4	TUBB6	UNC5D	VIPR2	ZFPM2	ZMIZ1	ZNF521	ZNF800	CLMN	FGF8	IGF1R	JAZF1	MEIS2	PLEKHA7	RREB1	SALL3	SETBP1
CFT 171																																
Bic 6																																
Bic 22																																

	SLC22A23	SOX11	IRF2BPL	OLIG2	SALL1	CDC85C	CCND1	CYLD	GPC1	INPP5A	TEAD3
CFT 171											
Bic 6											
Bic 22											

Ripartizione dei geni della coppia di fattori di trascrizione 171 tra i bicluster 6 e 22: Le righe mostrano i Fattori di Trascrizione (FT) e i geni sia delle Coppie di FT (CFT) che dei Bicluster (Bic). I campi "FT enhancer" ed "FT promotore" mostrano i FT che compongono la CFT e i bicluster. Successivamente ogni colonna rappresenta un gene e gli elementi in nero rappresentano la presenza del gene alla CFT/bicluster corrispondente. La tabella così composta ha un numero di colonne (geni) elevato, e per poterla visualizzare nella sua interezza è stata partizionata in 3 sotto-tabelle più piccole. Di conseguenza ogni sotto-tabella rappresenta la continuazione della precedente. I nomi delle righe sono comunque riportati per ogni sotto-tabella. A CFT 171 corrispondono 65 geni, a Bic 6 quattordici, a Bic 22 nove. Bic 6 e Bic 22 condividono 3 geni.

Un'analisi di enrichment dei termini Gene Ontology (o più brevemente "termini GO" o "GO-term") "biological process" dei 62 su 65 geni associabili alla CFT 171 di cui è stato possibile rinvenire il corrispondente codice identificativo Uniprot mostra la presenza di termini significativamente presenti riguardanti la regolazione dell'espressione (GO:0010468; regulation of gene expression), la differenziazione del tessuto nervoso (GO:0030182; "neuron differentiation"), lo sviluppo dello scheletro (GO:0001501; "skeletal system development") e meno significativamente lo sviluppo di tessuto muscolare (GO:0060537; "muscle tissue development") e rene (GO:0001822; "kidney development"). Si veda *Lista A21* in appendice per i dettagli. La stessa analisi per i 14 geni del Bic 6 (*Lista A22*) mostra di nuovo la presenza di termini simili, inclusi termini relativi allo sviluppo nervoso, scheletrico e renale, ma si nota anche l'emergenza di termini riguardanti la formazione dell'arto (GO:0035108; "limb morphogenesis") che precedentemente non risultavano significativamente presenti. Infine i 9 geni del Bic 22 (*Lista A23*) mostrano termini relativi al differenziamento del tessuto nervoso.

I FT presenti in Bic 6 e Bic 22 sono simili ed includono FT coinvolti in processi dello sviluppo embrionale noti. SP3 ed SP4 sono già stati discussi precedentemente, MAZ similmente ad SP1 è un FT ubiquitario che

potrebbe essere coinvolto nella regolazione di geni specifici dei neuroni (Okamoto et al., 2002). ZBTB7B è un FT coinvolto nello sviluppo embrionale con un'attività nella regolazione del collagene (Widom et al., 1997) e può interagire con SP1 ed SP3 (Kyriotou et al., 2007). MYF6 è un FT noto per il suo ruolo nella differenziazione del muscolo scheletrico (Moncaut et al., 2013). Sebbene non siamo a conoscenza di meccanismi di regolazione noti riguardanti lo sviluppo del sistema nervoso per tutti questi FT, tutti risultano espressi in diverse regioni sistema nervoso durante lo sviluppo embrionale umano (Lindsay et al., 2016), il che potrebbe indicare un loro ruolo in questi processi non ancora approfondito.

Da questi dati è possibile corroborare la seguente ipotesi: l'analisi del test binomiale evidenzia coppie in grado di controllare un numero elevato di geni. Il biclustering alla stessa coppia, aggiungendo altri FT sia all'enhancer che al promotore evidenzia una diminuzione del numero di geni controllati. In concomitanza da termini GO che descrivono processi generali si passa a termini GO più specifici. Questi dati supportano quindi l'ipotesi che il biclustering da un lato confermi i dati ottenuti con il test binomiale, dall'altro consenta di rendere più specifici i gruppi che corrispondono alle transcription factories recentemente suggerite altre con altri approcci (Cremer et al., 2015; Iborra et al., 1996; Li et al., 2012; Weipoltshammer and Schöfer, 2016).

CONCLUSIONE

Questa tesi ha analizzato l'ipotesi che l'interazione tra fattori di trascrizione sia alla base del meccanismo del looping per la regolazione della trascrizione genica. Il fenomeno è estremamente complesso, poco dettagliato sperimentalmente e in linea con i modelli attuali di cromatina, secondo cui super-complessi proteici noti come "transcription factories" (Cremer et al., 2015; Iborra et al., 1996; Li et al., 2012; Weipoltshammer and Schöfer, 2016) sarebbero responsabili dell'attività della regolazione dell'espressione. Il fenomeno è stato affrontato mediante tecniche computazionali, volte a identificare con validazione statistica coppie di o più fattori di trascrizione interagenti nel loop enhancer-promotore.

La strategia individuata parte da una banca dati che contiene set di enhancer validati sperimentalmente, coinvolti nella regolazione dello sviluppo embrionale dei vertebrati e posizionati nei pressi di geni anch'essi coinvolti nello stesso processo. Dal set, viene poi generato un set di coppie enhancer-promotore accoppiando ogni enhancer ai promotori dei geni più vicini. Partendo da questi dati in parte sperimentali, sulle coppie enhancer-promotore si sono predetti i siti di interazione del DNA con i fattori di trascrizione. Questo consente di predire anche i fattori di trascrizione che vengono accoppiati sulla base delle coppie enhancer-promotore. L'ipotesi sottesa a questa procedura è che il loop sia stabilizzato dalla interazione tra fattori di trascrizione specifici per l'enhancer e il promotore appaiato.

Una prima validazione statistica della procedura avviene mediante test binomiale, che identifica 364 coppie di fattori di trascrizione significative (validate mediante una procedura di bootstrap) per un totale di 196 fattori. Le coppie sono state rappresentate sotto forma di grafo ed annotate utilizzando informazioni provenienti da varie risorse.

La validazione della procedura avviene mediante approcci diversi. Prima di tutto è evidente che alla base della formazione delle coppie interagenti debba riscontrarsi un fenomeno di co-espressione. Utilizzando le risorse dati più recenti, almeno il 99% delle coppie è associato a dati di co-espressione. La co-espressione dei FT che compongono le coppie conferma quindi la possibilità di una loro interazione.

Una ulteriore validazione sperimentale, anche se indiretta, ha avuto origine da una indagine dell'interattoma umano, cioè dalla mappa di tutte le interazioni tra proteine umane contenute nella banca dati STRING. In questa banca dati le proteine interagenti sono nodi di un grafo altamente connesso, dove la distanza tra due proteine è il numero di altre proteine che le connettono secondo un cammino minimo (zero per proteine direttamente interagenti). Qui la mappatura delle coppie di fattori di trascrizione indica che queste in media sono significativamente più vicine di quanto atteso dalla distribuzione dei cammini minimi nell'interattoma umano. Una simile prossimità significativa si osserva anche tra i fattori di trascrizione individuati e la RNAP II, la quale è presente nei complessi di attivazione della trascrizione. Oltre a essere brevi, i cammini minimi mostrano la frequente presenza di noti adattatori della trascrizione (EP300, CREBBP, Mediator), che la letteratura elenca tra i possibili mediatori del looping della cromatina.

Questa prima analisi individua quindi le coppie di fattori di trascrizione statisticamente significative che potrebbero essere alla base delle interazioni enhancer-promotore a lungo raggio e che quindi potrebbero essere un utile contributo alla comprensione della regolazione dei processi coinvolti nello sviluppo embrionale umano.

L'analisi accurata della letteratura sull'argomento conferma questa ipotesi ed in particolare tra gli esempi noti va citato che per una delle coppie più significative nella nostra analisi (CFT 6: RARG, RXRA; si veda paragrafo 6.4) è già stata descritta la possibilità di interazione (Yasmin et al., 2004; Zuchegna et al., 2015).

Nella fattorie di trascrizione, l'ipotesi è che più fattori di trascrizione diano origine alla specificità della regolazione della trascrizione. Per analizzare questa possibilità a livello della formazione dei loop, abbiamo proceduto con una analisi mediante l'uso di algoritmi di biclustering, utilizzando da un lato i fattori di trascrizione individuati mediante l'analisi precedente e dall'altro le coppie di enhancer-promotore a loro corrispondenti. L'analisi permette di individuare su basi statistiche non più coppie, ma gruppi di fattori di trascrizione coinvolti nello stabilizzare la coppia enhancer-promotore. La procedura ha identificato 48 bicluster significativi, composti in media da 3 fattori di trascrizione coinvolti in media in 13 coppie enhancer-promotore. I bicluster includono spesso fattori di trascrizione che giocano un ruolo nella regolazione dello sviluppo embrionale; inoltre in un caso (Bic 5; si veda paragrafo 6.5) sono state identificate proteine ad enhancer e promotori la cui attività di mediatori del loop della cromatina è nota (Mastrangelo et al., 1991; Su et al., 1991).

Dalla comparazione dei risultati del test binomiale con i risultati del biclustering, si può concludere che esistono coppie di fattori di trascrizione associate, che quando associate ad altre fattori di trascrizione sulla stessa coppia enhancer-promotore, rendono la regolazione dell'espressione genica più specifica. Ciò è corroborato dal fatto che un'analisi sulle funzioni dei geni regolati mette in luce che all'aumentare del numero di fattori di trascrizione presenti sulla stessa coppia enhancer-promotore, i processi biologici coinvolti diventano più specifici (*Figura 6.8*). In quest'ottica rispetto al test binomiale, il biclustering può essere utilizzato per investigare un livello superiore di organizzazione della regolazione della trascrizione.

Sebbene studi dedicati alla ricerca di coppie di FT siano già stati effettuati, per quanto è a nostra conoscenza non era ancora stata esplorata la possibilità di considerare sia le coppie che i gruppi di FT a partire da un set di enhancer specifici dello sviluppo embrionale e validati sperimentalmente.

Nonostante le assunzioni che sono state accettate per poter effettuare questo studio (accoppiamento enhancer-promotore basato sulla prossimità delle sequenze, definizione di coppie e gruppi di FT basata sulla predizione dei siti di binding), i risultati ottenuti confermano la nozione secondo cui almeno parte della informazione necessaria alla formazione e stabilizzazione del loop tra enhancer e promotore risiede in porzioni ben specifiche delle sequenze nucleotidiche, dove i fattori di trascrizione hanno una elevata affinità di interazione anche se mediata da altre proteine.

ABBREVIAZIONI

3C	Chromosome Conformation Capture	TATA-box	sito di legame per TATA-binding protein
ANC	Active Nuclear Compartment	SELEX	Systematic Evolution of Ligands by Exponential Enrichment
A	Adenina	GO-term	Termini Gene Ontology
bp	Base-Pair	BRE	TFIIB-Recognition Element
LacZ	Beta-galattosidase	T	Timina
CTCF	CCCTC-binding factor	TAD	Topologically Associating Domains
ES cell	Cellule Staminali Embrionali	SP1	Transcription Factor SP1
ChIP-seq	Chromatin ImmunoPrecipitation sequencing	SP1	Transcription factor Sp1
ChIA-PET	Chromatin Interaction Analysis by Paired-End Tag sequencing	SP3	Transcription factor Sp3
5C	Chromosome Conformation Capture Carbon Copy	SP4	Transcription factor Sp4
4C	Chromosome Conformation Capture-on-Chip	UCSC	University of California Santa Cruz genome browser
C	Citosina	VISTA	VISTA Enhancer Browser
CCDC85C	Coiled-coil domain-containing protein 85C	ZBTB7B	Zinc finger and BTB domain-containing protein 7B
CPI	Complesso di Pre-Inizio della trascrizione	ZNF148	Zinc finger protein 148
CST	Coppia di Siti di legame per fattori di Trascrizione		
CEP	Coppia Enhancer-Promotore		
CREBBP	CREB-binding protein		
CDK8	Cyclin-dependent kinase 8		
Da	Dalton		
DNase-seq	DNase I hypersensitive sites sequencing		
DLD	Dominio di Legame al DNA		
DCE	Downstream Core Element		
DPE	Downstream Promoter Element		
ERC	Elementi di Regolazione Cis-agenti		
FT	Fattore di Trascrizione		
FPKM	Fragments Per Kilobase of exon model per Million mapped reads		
GATA1	GATA-binding factor 1		
GAM	Genome Architecture Mapping		
G	Guanina		
Hsp68	Heat shock protein 68		
HO-1	Heme Oxygenase-1		
EP300	Histone acetyltransferase p300		
HOCOMOCO	HOmo sapiens COmprehensive MOdel COllection		
INC	Inactive Nuclear Compartment		
Inr	Initiator element		
IC	Interchromatin Compartment		
IUPAC	International Union of Pure and Applied Chemistry		
EKLF	Kruppel-like factor 1		
LDB1	LIM domain-binding protein 1		
LCR	Locus Control Region		
MFP	Matrici di Frequenza Posizione-specifiche		
MED1	Mediator of RNA polymerase II transcription subunit 1		
MED12	Mediator of RNA polymerase II transcription subunit 12		
MED14	Mediator of RNA polymerase II transcription subunit 14		
MED15	Mediator of RNA polymerase II transcription subunit 15		
MED17	Mediator of RNA polymerase II transcription subunit 17		
MED21	Mediator of RNA polymerase II transcription subunit 21		
MED23	Mediator of RNA polymerase II transcription subunit 23		
MED24	Mediator of RNA polymerase II transcription subunit 24		
MED6	Mediator of RNA polymerase II transcription subunit 6		
MECP2	Methyl-CpG-binding protein 2		
MOODS	MOtif Occurrence Detection Suite		
MTE	Motif Ten Element		
NPC	multipotent Neural Progenitor Cells		
MAZ	Myc-associated zinc finger protein		
MEF2	Myocyte-specific enhancer factor		
MYF6	Myogenic factor 6		
NR2C2	Nuclear receptor subfamily 2 group C member 2		
PR	Perichromatin Region		
PPARA	Peroxisome proliferator-activated receptor alpha		
PFM	Position Frequency Matrix		
PWM	Position Weight Matrix		
PGR	Progesterone receptor		
RARG	Retinoic acid receptor gamma		
RXRA	Retinoic acid receptor RXR-alpha		
mRNA	RNA messaggero		
RNAP	RNA polimerasi		
rRNA	RNA ribosomale		
RNA-Seq	RNA sequencing		
tRNA	RNA transfer		
RBM38	RNA-binding protein 38		
SIT	Sito di Inizio della Trascrizione		
SLFT	Sito di Legame per Fattore di Trascrizione		

APPENDICE

Lista A1: Set delle sequenze enhancer

Elenco completo degli enhancer. La lista contiene i 786 enhancer utilizzati per la generazione del set di 1234 coppie EP. Gli enhancer non presenti anche nel set di 1011 coppie EP sono contraddistinti da un asterisco ed elencati in fondo. Il campo "VISTA ID" indica il codice dell'enhancer associato da VISTA enhancer browser.

Si tenga presente che per evitare sovrapposizioni tra le sequenze 5 coppie di enhancer così come rilasciate da VISTA sono state unite tra di loro e lo stesso è stato fatto per un gruppo di 3 enhancer. Per denominare le sequenze risultanti è stato selezionato uno dei codici di identificazione delle sequenze di origine, come mostrato nella tabella qui sotto.

VISTA ID selezionato	VISTA ID di origine
hs341	hs341,hs189
hs1523	hs1523,hs342
hs1032	hs1032,hs935
hs174	hs174,hs322
hs1717	hs1717,hs1597
hs1540	hs1540,hs1354,hs998

VISTA ID	Coordinate genomiche
hs1	chr16:86430087-86430726
hs4	chr16:80372593-80373755
hs12	chr16:78510608-78511944
hs16	chr16:72980900-72982545
hs20	chr16:72738568-72740149
hs23	chr16:55424416-55425738
hs26	chr16:55223185-55224283
hs37	chr16:54650598-54651882
hs43	chr16:54575878-54577925
hs51	chr16:54323420-54324336
hs52	chr16:54143854-54144867
hs59	chr16:53726316-53727556
hs65	chr16:51958288-51959688
hs71	chr16:51671181-51672039
hs73	chr16:51576725-51578101
hs74	chr16:51560196-51560925
hs76	chr16:51231121-51232362
hs79	chr16:51056311-51057592
hs108	chr16:72404533-72407591
hs110	chr7:21003280-21004750
hs111	chr7:42191728-42193638
hs112	chr9:973435-975288
hs113	chr11:32197490-32198734
hs118	chrX:24894335-24896084
hs119	chrX:24915382-24918272
hs121	chrX:25007879-25009581
hs122	chrX:25017067-25018756
hs123	chrX:25400224-25402334
hs124	chr16:49094466-49095242
hs129	chr13:71774848-71775368
hs131	chr13:72771271-72773876
hs135	chr13:73149054-73150028
hs137	chr13:72300849-72302934
hs138	chr13:72425787-72428335
hs141	chr13:73001083-73003053
hs142	chr13:73081447-73083754
hs145	chrX:25018871-25020532
hs149	chr2:45106927-45107653
hs151	chr16:55329289-55330760
hs152	chr16:51147012-51148642
hs155	chr16:53948201-53949846
hs156	chr16:53994058-53995943
hs157	chr16:53986630-53988193
hs161	chr16:52446050-52447237
hs162	chr16:50869692-50871098
hs168	chr1:63443534-63444884
hs169	chr1:63464282-63465717
hs170	chr2:164450144-164451758
hs174	chr1:87821621-87823082
hs181	chr15:37240805-37242498
hs186	chr9:129198400-129200739
hs187	chr3:71290418-71292584
hs188	chr13:72345980-72348230
hs191	chr5:91036888-91038899
hs192	chr3:180773639-180775802
hs194	chr1:51034546-51036289
hs195	chr8:106333530-106334859
hs200	chr1:51165195-51166786

VISTA ID	Coordinate genomiche
hs204	chr1:213597964-213599524
hs205	chr2:66297527-66299214
hs208	chr3:136982707-136984175
hs209	chr3:137047544-137049271
hs210	chr3:137067622-137068925
hs213	chr3:147049536-147051100
hs215	chr5:50335050-50336643
hs217	chr6:51148668-51149710
hs218	chr7:114056847-114058647
hs222	chr7:114295109-114296373
hs229	chr10:102546590-102548095
hs230	chr5:158340962-158342611
hs232	chr10:131691086-131692848
hs234	chr11:31685357-31686884
hs236	chr11:16316207-16317562
hs238	chr1:38560194-38561606
hs240	chr9:83727123-83728378
hs241	chr1:10965129-10966144
hs243	chr2:174985544-174987464
hs244	chr2:174988737-174990363
hs246	chr2:176940070-176941410
hs248	chr2:177503157-177503671
hs249	chr2:237071364-237071848
hs253	chr3:114433184-114435415
hs258	chr3:180436427-180437958
hs259	chr4:13009318-13010327
hs260	chr4:105345575-105346895
hs262	chr5:76940836-76941396
hs266	chr5:87168414-87169433
hs267	chr5:87239942-87241645
hs269	chr5:90928612-90929226
hs271	chr5:93226985-93228322
hs277	chr1:44715420-44716129
hs278	chr5:170627990-170630018
hs280	chr1:44989824-44991149
hs281	chr6:41523224-41523677
hs282	chr6:98116085-98116943
hs284	chr6:98719759-98720510
hs289	chr1:10732070-10733118
hs293	chr7:1265154-1266318
hs294	chr7:21811370-21812002
hs296	chr7:26728697-26729802
hs297	chr7:31401914-31403016
hs298	chr7:96633582-96634303
hs302	chr8:37249594-37251182
hs304	chr9:8095553-8096166
hs307	chr9:16710536-16711184
hs309	chr9:23496285-23497394
hs311	chr1:63369349-63370894
hs312	chr9:81471747-81473114
hs313	chr9:81870622-81872224
hs314	chr9:126537718-126539929
hs326	chr10:103266649-103267972
hs327	chr1:88926796-88928508
hs329	chr10:124851813-124853071
hs330	chr10:126905322-126906003
hs341	chr13:95618464-95619850
hs348	chr14:36020024-36020998

VISTA ID	Coordinate genomiche
hs355	chr15:36819021-36820695
hs358	chr15:67665164-67665685
hs359	chr15:70391576-70392613
hs362	chr17:35011848-35012557
hs369	chr18:22864753-22866457
hs371	chr18:35063482-35064528
hs376	chr18:53089625-53091079
hs381	chr19:30494614-30495503
hs382	chr19:30565950-30567016
hs385	chr19:31587058-31588070
hs388	chr2:7774393-7775070
hs393	chr2:59198905-59200529
hs394	chr2:59746377-59746992
hs399	chr2:60441495-60442515
hs401	chr2:104736518-104737365
hs407	chr2:145188099-145189812
hs411	chr2:156726581-156727605
hs413	chr2:157551014-157551952
hs415	chr2:157899792-157900721
hs416	chr2:162094895-162095451
hs421	chr2:164843797-164844783
hs422	chr2:172955879-172957052
hs426	chrX:81788884-81790571
hs427	chrX:139169379-139171545
hs428	chr14:26377372-26378803
hs430	chr19:30840299-30843536
hs433	chr14:30741750-30743626
hs434	chr3:62350726-62351718
hs435	chr3:62359866-62360569
hs442	chr1:83345366-83346271
hs446	chr7:35454800-35456218
hs463	chr7:35505184-35506342
hs466	chr7:35268514-35269380
hs472	chr7:35458551-35459522
hs483	chr12:115124394-115125300
hs484	chr10:77164564-77165639
hs488	chr13:95358263-95360017
hs491	chr22:38429240-38430515
hs495	chr10:92690021-92690674
hs502	chr18:19773983-19774532
hs503	chr5:88548400-88550500
hs511	chr10:103527193-103528393
hs513	chr10:103540180-103541030
hs515	chr22:19749356-19750055
hs516	chr5:44389166-44390467
hs521	chr2:236773458-236774696
hs522	chr2:121832418-121833317
hs526	chr4:1613479-1614106
hs529	chr9:17322200-17324371
hs533	chr5:3186439-3187926
hs536	chr19:32452701-32453894
hs541	chr2:45030569-45032739
hs543	chr5:50960132-50961567
hs546	chr5:92811465-92813217
hs550	chr7:13506207-13507276
hs553	chr2:172936519-172938249
hs556	chr9:84299210-84300117
hs559	chr4:112421802-112422905

VISTA ID	Coordinate genomice
hs2071	chr7:115993663-115997109
hs2073	chr11:69054197-69056378
hs2077	chr7:135742628-135745716
hs2078	chr12:54412238-54415445
hs2081	chr7:107587162-107588892
hs2082	chr7:139910588-139914247
hs2084	chr21:35915964-35919662
hs2088	chr1:38728185-38732516
hs2089	chr1:212704746-212707293
hs2090	chr11:44289927-44293764
hs2094	chr1:10795106-10799241
hs2096	chr11:16913136-16917745
hs2263	chr1:109816874-109818905
hs2306	chr7:18868716-18870405
hs2307	chr7:18874272-18875369
hs2308	chr7:95705089-95705696
hs2312	chr7:96075405-96075991
hs2313	chr7:95765880-95766473
* hs22	chr16:72254566-72255825
* hs27	chr16:55079237-55080429
* hs72	chr16:51657810-51658716
* hs101	chr16:48912816-48914144
* hs126	chr13:71343609-71344868
* hs242	chr2:174114371-174115933
* hs261	chr5:3511978-3513399
* hs320	chr10:77726861-77727618
* hs335	chr11:131867949-131868377
* hs336	chr12:16715315-16716360
* hs357	chr15:57425244-57425958
* hs363	chr17:35329349-35329944
* hs367	chr18:22692416-22693640
* hs378	chr18:72357475-72358333
* hs383	chr19:30767058-30768273
* hs384	chr19:31002941-31004262
* hs389	chr1:10781239-10781744
* hs408	chr1:10851570-10852173
* hs486	chr22:38394345-38395199
* hs492	chr22:38442140-38442879
* hs508	chr8:11604182-11604695
* hs532	chr13:28395961-28397536
* hs540	chr13:71358093-71359507
* hs545	chr1:243876467-243877893
* hs568	chr2:146692288-146693283
* hs593	chr14:37726340-37727348
* hs595	chr10:76177765-76179000
* hs623	chr15:57426028-57426952
* hs628	chr9:159657-160780
* hs645	chr1:32510030-32510913
* hs646	chr2:172820365-172821314
* hs653	chr3:137185964-137186866
* hs662	chr2:157720628-157721586
* hs679	chr18:45087290-45088074
* hs742	chr8:78042571-78044201
* hs759	chr13:106015598-106016927
* hs778	chr17:59893119-59894141
* hs816	chr7:14379627-14380740
* hs843	chr3:137638911-137640108
* hs847	chr4:42150091-42151064
* hs863	chr11:31502035-31503157
* hs865	chr6:50685244-50686237
* hs868	chr2:105697211-105698158
* hs882	chr13:71533037-71534195
* hs906	chr4:145276813-145278719
* hs924	chr17:2081998-2084289
* hs926	chr4:112668302-112669915
* hs967	chr12:103484342-103485519
* hs1007	chr7:20997668-20999102
* hs1052	chr6:18536628-18538267
* hs1086	chr20:39334182-39335059
* hs1112	chr2:103909358-103911221
* hs1128	chr6:98829860-98831049
* hs1144	chr1:164668592-164669823
* hs1148	chr7:21019551-21021200
* hs1169	chr18:22235986-22237388
* hs1185	chr1:163507965-163509139
* hs1227	chr5:91271776-91272886
* hs1246	chr13:112268067-112269088
* hs1257	chr1:218207432-218208498
* hs1268	chr7:155547616-155548788
* hs1305	chr8:80874361-80876746
* hs1325	chr7:25791903-25794282
* hs1339	chr9:92292484-92293889
* hs1368	chr1:204423958-204424935
* hs1385	chr2:3268196-3270849
* hs1392	chr14:68634843-68636060

VISTA ID	Coordinate genomice
* hs1434	chr10:78282592-78284988
* hs1445	chr17:2079772-2081236
* hs1449	chr5:131600830-131602104
* hs1475	chr3:79042278-79045053
* hs1480	chr2:37643870-37645460
* hs1529	chr2:104578156-104580488
* hs1563	chr3:193489359-193491333
* hs1677	chr7:25533607-25536880
* hs1683	chr10:78156517-78158752
* hs1752	chr5:148801693-148805076
* hs1759	chr11:65254704-65258394
* hs1856	chr17:61751592-61755090
* hs2027	chr2:159885988-159889012
* hs2041	chr3:141579463-141580810
* hs2065	chr10:46052882-46055670
* hs2309	chr7:95726279-95727680
* hs2311	chr7:96124918-96125415

Gene symbol	UCSC ID	Strand e coordinate genomice
* LOC100506172	uc002fcn.3	(+) chr16:73418704-73420703
* LOC150568	uc002tch.3	(+) chr2:105048805-105050804
* LOC285501	uc010iru.3	(+) chr4:178647911-178649910
* LOC285577	uc003jdd.3	(-) chr5:3536209-3538208
* LOC729950	uc002kvj.1	(+) chr18:22206146-22208145
* MAFB	uc002xji.3	(-) chr20:39317877-39319876
* MAGEB18	uc004dbq.2	(+) chrX:26154460-26156459
* MALAT1	uc010roh.2	(+) chr11:65263233-65265232
* MAP3K3	uc002jbg.3	(+) chr17:61697801-61699800
* MARCH8	uc001jch.2	(-) chr10:46030820-46032819
* MGST1	uc001rdg.3	(+) chr12:16498579-16500578
* MIPOL1	uc001wuc.3	(+) chr14:37665118-37667117
* MIR143HG	uc003lpq.2	(+) chr5:148784440-148786439
* MIR2054	uc021xrx.1	(+) chr4:126426414-126428413
* MIR208B	uc021rqy.1	(-) chr14:23887273-23889272
* MIR2113	uc021zdb.1	(+) chr6:98470407-98472406
* MIR4251	uc021oex.1	(+) chr1:3042539-3044538
* MIR548A1	uc021yme.1	(+) chr6:18570015-18572014
* MIR548A3	uc021xbz.1	(-) chr3:103946106-103948105
* MIR601	uc022bnd.1	(-) chr9:126164883-126166882
* MIR875	uc022aze.1	(-) chr8:100549090-100551089
* MLK7-AS1	uc002uib.3	(-) chr2:174146765-174148764
* MRPS28	uc003ybp.3	(-) chr8:80942507-80944506
* MRPS9	uc002tcn.4	(+) chr2:105652483-105654482
* MYEOV	uc001oov.3	(+) chr11:69059622-69061621
* NES	uc001fpq.3	(-) chr1:156647190-156649189
* NPM3	uc001ktt.3	(-) chr10:103543171-103545170
* NPVF	uc003sxo.3	(-) chr7:25268106-25270105
* NTM	uc001qgq.3	(+) chr11:131778712-131780711
* NUF2	uc001gcr.1	(+) chr1:163289723-163291722
* OTOS	uc002vyv.3	(-) chr2:241080074-241082073
* PDLIM4	uc003kwn.3	(+) chr5:131591351-131593350
* PDP2	uc002eqk.2	(+) chr16:66912383-66914382
* PDX1	uc001urt.2	(+) chr13:28492168-28494167
* PIK3C2B	uc001haw.3	(-) chr1:204459475-204461474
* PLSCR5	uc010hvc.3	(-) chr3:146324004-146326003
* PSMG1	uc002yxi.3	(-) chr21:40555441-40557440
* PVRL1	uc001pww.3	(-) chr11:119599436-119601435
* QPCT	uc002rqg.3	(+) chr2:37569753-37571752
* RAD51B	uc001xkf.2	(+) chr14:68284496-68286495
* RBM33	uc010lqk.1	(+) chr7:155435203-155437202
* RNF144B	uc003ncs.3	(+) chr6:18385581-18387580
* ROBO1	uc003dqe.2	(-) chr3:79817060-79819059
* RPS24	uc001jzs.3	(+) chr10:79791518-79793517
* RRAGC	uc001ccq.2	(-) chr1:39325341-39327340
* RRP15	uc001hlj.3	(+) chr1:218456629-218458628
* SKP1P2	uc021qvt.1	(+) chr12:17139681-17141680
* SMAD2	uc002lcz.3	(-) chr18:45457516-45459515
* SMG6	uc002fub.1	(-) chr17:2207070-2209069
* SOX14	uc003erm.2	(+) chr3:137481134-137483133
* SOX1	uc001vsb.1	(+) chr13:112719913-112721912
* STXBP6	uc001wpu.3	(-) chr14:25519096-25521095
* TANK	uc002ubr.2	(+) chr2:161991466-161993465
* TBX3	uc001tvt.1	(-) chr12:115121970-115123969
* TCF12	uc002aea.3	(+) chr15:57208833-57210832
* TERC	uc003ffr.1	(-) chr3:169482849-169484848
* TFAP2D	uc003paf.3	(+) chr6:50679257-50681256
* TMCOS5A	uc001zjw.3	(+) chr15:38225458-38227457
* TMEM39B	uc010ogv.2	(+) chr1:32536503-32538502
* TOP1	uc002xjl.3	(+) chr20:39655462-39657461
* TSSC1	uc002qxj.2	(-) chr2:3381654-3383653
* UNQ6494	uc004aqs.2	(+) chr9:92252698-92254697
* USP6NL	uc001iks.1	(-) chr10:11574275-11576274
* YTHDF3	uc003xuy.3	(+) chr8:64079121-64081120
* ZBTB20-AS1	uc003ebo.2	(+) chr3:114068658-114070657
* ZNF407	uc002llw.2	(+) chr18:72340923-72342922
* ZNF536	uc002nsu.1	(+) chr19:30861328-30863327
* ZNF644	uc001dnw.3	(-) chr1:91487813-91489812

LISTA A3: Set delle coppie enhancer - promotore

Elenco completo di tutte le CEP. La lista include tutte le 1234 coppie, quelle non presenti anche nel set di 1011 CEP sono contraddistinte da un asterisco ed elencate in fondo. Queste ultime hanno la caratteristica di essere composte da almeno una sequenza per la quale non è stata predetta la presenza di SLFT (si vedano i paragrafi 5.2 e 6.2). Per ogni riga della lista il primo campo contiene un codice numerico identificativo della CEP, i campi da 2 a 4 contengono informazioni riguardo all'enhancer della coppia, mentre i restanti 3 campi contengono informazioni riguardo il promotore accoppiato. Il campo "VISTA ID" indica il codice dell'enhancer, "Gene symbol" indica il nome del gene del promotore associato. I campi "SLFT" e "FT" indicano il numero di SLFT e il numero FT corrispondenti predetti sulle rispettive sequenze. I valori del campo "SLFT" sono ridondanti, vale a dire che il numero corrisponde al conteggio di tutti i SLFT predetti contando anche occorrenze multiple di SLFT per lo stesso FT. I valori del campo "FT" invece corrispondono al numero di FT predetti almeno una volta sulle sequenze.

CEP	VISTA ID	SLFT	FT	Gene symbol	SLFT	FT	CEP	VISTA ID	SLFT	FT	Gene symbol	SLFT	FT
1	hs1319	4	1	ACADSB	3	3	70	hs355	2	2	C15orf41	3	3
2	hs329	1	1	ACADSB	3	3	71	hs828	2	1	C15orf41	3	3
3	hs1487	18	11	ACOT11	2	2	72	hs1948	3	3	C19orf69	1	1
4	hs1854	28	9	ACSF3	1	1	73	hs1766	6	5	C1orf198	9	5
5	hs249	1	1	AGAP1	13	6	74	hs1346	5	2	C21orf54	1	1
6	hs521	2	2	AGAP1	13	6	75	hs1188	2	2	C21orf62	1	1
7	hs921	6	4	AGAP1	13	6	76	hs1548	12	8	C21orf62	1	1
8	hs680	7	6	AGPAT9	1	1	77	hs1886	14	13	C2orf57	3	3
9	hs712	2	2	AGPAT9	1	1	78	hs434	2	2	C3orf14	2	2
10	hs1115	1	1	AGTR1	1	1	79	hs1387	1	1	C3orf49	3	3
11	hs1173	2	2	AGTR1	1	1	80	hs1251	4	2	C3orf55	1	1
12	hs1203	5	4	AGTR1	1	1	81	hs638	1	1	C3orf55	1	1
13	hs1224	3	3	AGTR1	1	1	82	hs981	2	2	C4orf21	4	4
14	hs1573	3	1	AGTR1	1	1	83	hs559	2	1	C4orf32	4	4
15	hs654	2	2	AGTR1	1	1	84	hs930	6	4	C4orf32	4	4
16	hs1351	3	3	AH11	1	1	85	hs1473	4	3	C5orf20	2	2
17	hs1672	2	2	AKR7A2P1	3	3	86	hs533	9	6	C5orf38	8	5
18	hs2090	12	8	ALX4	2	2	87	hs603	1	1	C5orf38	8	5
19	hs495	1	1	ANKRD1	1	1	88	hs754	3	2	C5orf38	8	5
20	hs1334	6	3	ANKRD30A	2	2	89	hs1955	7	6	C6orf118	2	2
21	hs1567	5	2	ANKRD30A	2	2	90	hs304	2	2	C9orf123	4	3
22	hs1589	6	4	ANKRD30A	2	2	91	hs1717	9	6	C9orf156	1	1
23	hs1783	8	7	AP3B1	2	1	92	hs1316	5	5	CADPS	2	2
24	hs675	3	3	ARHGAP15	3	3	93	hs435	1	1	CADPS	2	2
25	hs1912	25	18	ARHGEF16	6	6	94	hs1809	6	6	CADPS2	1	1
26	hs705	1	1	ARHGEF16	6	6	95	hs1498	4	3	CAND1	1	1
27	hs569	2	2	ARID1A	14	8	96	hs1858	10	7	CAPRIN1	2	2
28	hs1510	4	4	ARID3B	20	7	97	hs1881	2	2	CASC4	4	4
29	hs550	5	3	ARL4A	2	2	98	hs2094	12	8	CASZ1	4	4
30	hs749	1	1	ARL4A	2	2	99	hs241	22	7	CASZ1	4	4
31	hs118	7	4	ARX	1	1	100	hs289	35	12	CASZ1	4	4
32	hs119	5	4	ARX	1	1	101	hs2071	6	4	CAV2	10	6
33	hs121	3	3	ARX	1	1	102	hs1854	28	9	CBFA2T3	2	2
34	hs122	7	5	ARX	1	1	103	hs124	1	1	CBLN1	10	6
35	hs123	3	2	ARX	1	1	104	hs1448	5	5	CBLN4	2	2
36	hs145	3	3	ARX	1	1	105	hs297	1	1	CCDC129	1	1
37	hs1738	1	1	ASCC3	3	3	106	hs258	1	1	CCDC39	1	1
38	hs1114	2	1	ASCL1	6	6	107	hs655	2	2	CCDC39	1	1
39	hs1122	5	5	ASCL1	6	6	108	hs1013	1	1	CCDC68	3	3
40	hs1540	10	8	ASCL1	6	6	109	hs1574	3	1	CCDC77	2	2
41	hs1575	5	3	ASCL1	6	6	110	hs1931	12	10	CCDC85C	19	8
42	hs1723	1	1	ASCL1	6	6	111	hs1725	1	1	CCL2	1	1
43	hs1328	5	4	ASS1	3	3	112	hs1920	37	11	CCND1	12	5
44	hs2089	4	4	ATF3	5	5	113	hs266	1	1	CCNH	1	1
45	hs1242	3	3	ATF7IP2	2	1	114	hs267	4	2	CCNH	1	1
46	hs1432	5	5	ATG10	6	6	115	hs853	6	4	CCNH	1	1
47	hs168	5	3	ATG4C	3	3	116	hs807	2	2	CDCA7L	3	3
48	hs169	2	2	ATG4C	3	3	117	hs1959	1	1	CDH13	1	1
49	hs311	2	2	ATG4C	3	3	118	hs1760	9	9	CDH16	4	4
50	hs1652	12	7	ATOX1	2	2	119	hs1634	4	2	CDH2	8	5
51	hs775	2	2	ATP9B	4	4	120	hs1636	1	1	CDH2	8	5
52	hs1425	3	2	AUTS2	23	6	121	hs947	1	1	CDH7	1	1
53	hs1660	1	1	AUTS2	23	6	122	hs1383	1	1	CDH8	2	2
54	hs658	6	2	AUTS2	23	6	123	hs1633	3	3	CDK14	1	1
55	hs925	2	2	AUTS2	23	6	124	hs1340	10	6	CDKAL1	2	1
56	hs612	1	1	BARHL2	3	2	125	hs2060	5	5	CDYL	6	5
57	hs1891	17	14	BCAN	6	5	126	hs371	4	4	CELF4	4	2
58	hs1142	2	2	BCL11A	6	5	127	hs2263	5	5	CELSR2	3	3
59	hs957	3	2	BCL11A	6	5	128	hs840	1	1	CENPC1	1	1
60	hs622	2	2	BCL11B	4	3	129	hs1651	3	3	CEP135	4	4
61	hs2050	2	2	BCL2L1	4	4	130	hs678	1	1	CHIC2	2	1
62	hs1336	3	3	BMPER	5	4	131	hs687	1	1	CHIC2	2	1
63	hs1438	1	1	BMPR1B	1	1	132	hs1636	1	1	CHST9	12	6
64	hs307	2	2	BNC2	11	8	133	hs1751	4	2	CIDEA	2	1
65	hs1366	2	2	BTBD9	4	4	134	hs953	5	2	CIR1	2	2
66	hs326	3	2	BTRC	2	2	135	hs1747	10	8	CISH	3	3
67	hs578	4	3	BUB3	1	1	136	hs1130	5	3	CLINT1	3	2
68	hs1699	7	4	C10orf35	3	2	137	hs1208	3	3	CLINT1	3	2
69	hs1466	16	9	C14orf166B	3	3	138	hs1486	4	4	CLMN	37	9

CEP	VISTA ID	SLFT	FT	Gene symbol	SLFT	FT
1159	* hs742	0	0	PEX2	1	1
1160	* hs486	0	0	PICK1	4	3
1161	* hs492	0	0	PICK1	4	3
1162	* hs1368	6	4	PIK3C2B	0	0
1163	* hs926	0	0	PITX2	1	1
1164	* hs742	0	0	PKIA	3	3
1165	* hs213	5	3	PLSCR5	0	0
1166	* hs22	0	0	PMFBP1	3	3
1167	* hs1128	0	0	POU3F2	4	3
1168	* hs1516	1	1	PSMG1	0	0
1169	* hs1915	6	5	PVRL1	0	0
1170	* hs1480	0	0	QPCT	1	1
1171	* hs1392	0	0	RAD51B	1	1
1172	* hs1268	0	0	RBM33	0	0
1173	* hs1052	2	2	RNF144B	0	0
1174	* hs1475	6	4	ROBO1	0	0
1175	* hs1995	11	5	ROBO1	0	0
1176	* hs1007	0	0	RPL23P8	1	1
1177	* hs1148	0	0	RPL23P8	1	1
1178	* hs1974	9	5	RPS24	0	0
1179	* hs1001	8	8	RRAGC	0	0
1180	* hs1031	4	1	RRAGC	0	0
1181	* hs1139	6	6	RRAGC	0	0
1182	* hs1546	2	2	RRAGC	0	0
1183	* hs2088	7	5	RRAGC	0	0
1184	* hs1257	0	0	RRP15	3	3
1185	* hs72	0	0	SALL1	19	9
1186	* hs1532	7	4	SKP1P2	0	0
1187	* hs579	3	1	SKP1P2	0	0
1188	* hs605	2	1	SKP1P2	0	0
1189	* hs661	8	2	SKP1P2	0	0
1190	* hs993	3	3	SKP1P2	0	0
1191	* hs679	0	0	SMAD2	3	3
1192	* hs1445	2	1	SMG6	0	0
1193	* hs924	5	3	SMG6	0	0
1194	* hs1246	0	0	SOX1	3	2
1195	* hs208	4	2	SOX14	0	0
1196	* hs209	2	2	SOX14	0	0
1197	* hs210	3	2	SOX14	0	0
1198	* hs653	0	0	SOX14	0	0
1199	* hs843	3	2	SOX14	0	0
1200	* hs1007	0	0	SP4	5	5
1201	* hs1148	0	0	SP4	5	5
1202	* hs428	1	1	STXBP6	0	0
1203	* hs2027	0	0	TANC1	5	3
1204	* hs416	2	2	TANK	0	0
1205	* hs483	1	1	TBX3	0	0
1206	* hs357	0	0	TCF12	0	0
1207	* hs623	1	1	TCF12	0	0
1208	* hs1433	7	3	TERC	0	0
1209	* hs865	1	1	TFAP2D	0	0
1210	* hs572	2	1	TMC05A	0	0
1211	* hs599	4	2	TMC05A	0	0
1212	* hs830	4	4	TMC05A	0	0
1213	* hs1112	0	0	TMEM182	1	1
1214	* hs1529	0	0	TMEM182	1	1
1215	* hs645	0	0	TMEM39B	3	3
1216	* hs1086	0	0	TOP1	1	1
1217	* hs1385	3	3	TSSC1	0	0
1218	* hs1339	7	5	UNQ6494	0	0
1219	* hs383	0	0	URI1	3	3
1220	* hs1382	2	2	USP6NL	0	0
1221	* hs1239	5	3	YTHDF3	0	0
1222	* hs253	5	4	ZBTB20-AS1	0	0
1223	* hs22	0	0	ZFH3	3	3
1224	* hs378	4	3	ZNF407	0	0
1225	* hs824	4	3	ZNF407	0	0
1226	* hs367	0	0	ZNF521	7	4
1227	* hs382	4	2	ZNF536	0	0
1228	* hs383	0	0	ZNF536	0	0
1229	* hs384	1	1	ZNF536	0	0
1230	* hs385	3	2	ZNF536	0	0
1231	* hs430	4	4	ZNF536	0	0
1232	* hs721	3	3	ZNF536	0	0
1233	* hs858	2	2	ZNF536	0	0
1234	* hs612	1	1	ZNF644	0	0

LISTA A4: Fattori di trascrizione corrispondenti ai siti di legame predetti

Elenco completo di tutti i 231 FT di cui siti di legame sono stati predetti su enhancer e promotori. Ogni FT è identificato dal suo gene symbol. I SLFT corrispondenti a 62 e 69 FT sono stati predetti solo sugli enhancer e solo sui promotori rispettivamente, i primi sono elencati sul lato sinistro, i secondi sul lato destro. I SLFT corrispondenti a 100 FT sono stati invece predetti su entrambi i tipi di sequenze nucleotidiche e sono elencati al centro.

Si tenga presente che una volta effettuata la riduzione del set di coppie enhancer-promotore a 1011 elementi (si veda paragrafo 5.3) i FT contrassegnati da un asterisco sono andati persi e non sono stati considerati per le analisi successive. Inoltre il fattore di trascrizione POU5F1 (contrassegnato da un tratto) non risultava più condiviso tra enhancer e promotori, bensì risultava presente solo su promotori.

FT solo su enhancer

AHR
AIRE
* ALX1
ARNT2
ATF6
CEBPG
CXXC1
DDIT3
EBF1
EGR1
ELK1
ESRRB
ETV7
FOXO4
GFI1B
HES1
HEY2
HOXB1
HOXB6
HOXD10
HOXD9
INSM1
IRF1
IRF7
JUN
KLF3
LHX3
MECOM
MECP2
MEF2A
MYF6
NFAT5
NFE2
NR2E3
NR2F6
NR6A1
ONECUT2
OTX2
PAX2
PAX8
POU4F2
* PPAR
PPARG
RARB
RORA
RUNX1
SMAD1
SMAD2
SMAD4
SNAI1

FT sia su enhancer che su promotori

ARID3A
ARNTL
ATF5
BCL6
CEBPA
CTCF
E2F6
EGR2
EGR4
ELF3
ENO1
EPAS1
ETV4
FLI1
FOXA3
FOXC2
FOXD1
FOXO1
FOXO3
FUBP1
GATA1
HLTF
HNF1B
HSF1
HSF2
IRF2
IRF3
IRF4
IRF9
KLF1
KLF15
LHX2
MAZ
MBD2
MEF2C
MYOG
NANOG
NFATC1
NFE2L1
NFIC
NFYA
NFYB
NFYC
NHLH1
NKX3-1
NOBOX
NR1H2
NR1H2
NR2C1
NR2C2

NR2F1
NR2F2
NR3C1
NR3C2
NR4A3
NR5A1
ONECUT1
PBX3
PLAG1
- POU5F1
PPARA
PRDM1
PTF1A
PURA
RARG
REST
RFX3
RREB1
RXRA
SOX13
SP1
SP2
SP3
SP4
SPI1
SPIB
SPZ1
SREBF1
SREBF2
SRF
STAT2
STAT4
TBX2
TCF7L2
TFAP4
TFCP2
THRA
THRB
TP53
TP73
WT1
ZBTB4
ZBTB7A
ZBTB7B
ZFX
ZIC2
ZNF143
ZNF148
ZNF219
ZNF350

FT solo su promotori

ARID5B
ATF1
ATF3
BARX2
BHLHE41
BRCA1
DLX2
EHF
ELF1
ELF2
ELK3
ESR1
ESR2
FOS
FOXA1
FOXA2
FOXC1
FOXF1
FOXI1
FOXJ2
FOXJ3
FOXM1
FOXQ1
GATA3
GLI2
HNF1A
HOXA10
HOXA9
HOXB8
IRF5
IRF8
JUND
MAF
MAFK
MEF2D
MZF1
NFE2L2
NKX2-2
NR0B1
NR1D1
NRF1
OTX1
OVOL1
PAX5
PGR
POU1F1
POU2F2
POU3F1
POU3F2
PROP1

RARA
RBPJ
RFX1
RXRG
SMARCC1
SOX17
SOX2
SOX5
STAT1
STAT3
STAT5A
TAL1
TEF
TFAP2A
USF1
YY1
ZIC1
ZNF384
ZNF589

LISTA A5: Enhancer divisi per numero di promotori associati

Elenco completo di tutti gli enhancer suddivisi per il numero di promotori associati. Gli enhancer possono essere associati a uno o due promotori. I numeri in evidenza in cima ad ogni lista indicano il numero di promotori associati agli enhancer sottostanti.

1

hs1
hs16
hs52
hs59
hs111
hs112
hs123
hs129
hs137
hs138
hs155
hs156
hs157
hs174
hs181
hs187
hs188
hs191
hs194
hs195
hs200
hs204
hs208
hs209
hs210
hs213
hs215
hs218
hs229
hs230
hs234
hs236
hs241
hs243
hs244
hs248
hs253
hs269
hs277
hs278
hs280
hs281
hs282
hs284
hs289
hs294
hs296
hs298
hs307
hs314
hs326
hs341
hs348
hs358
hs369
hs371
hs376
hs381
hs382
hs385
hs393
hs407
hs416
hs427
hs428
hs430
hs433
hs435
hs442
hs461
hs463
hs466
hs472
hs483
hs484
hs495
hs502

hs513
hs515
hs516
hs533
hs546
hs553
hs556
hs563
hs564
hs565
hs566
hs567
hs569
hs572
hs579
hs590
hs598
hs599
hs603
hs605
hs607
hs608
hs611
hs612
hs618
hs619
hs622
hs625
hs630
hs632
hs634
hs636
hs640
hs641
hs656
hs658
hs661
hs667
hs672
hs675
hs676
hs678
hs682
hs687
hs688
hs690
hs693
hs697
hs698
hs702
hs705
hs717
hs720
hs721
hs741
hs748
hs752
hs754
hs762
hs764
hs769
hs774
hs775
hs782
hs798
hs800
hs809
hs813
hs818
hs824
hs828
hs830
hs840
hs851
hs852
hs855
hs858
hs864
hs872
hs876

hs883
hs886
hs891
hs895
hs901
hs909
hs914
hs919
hs925
hs952
hs957
hs965
hs969
hs971
hs978
hs987
hs990
hs993
hs999
hs1001
hs1011
hs1015
hs1017
hs1020
hs1022
hs1024
hs1027
hs1030
hs1031
hs1035
hs1039
hs1042
hs1043
hs1049
hs1056
hs1060
hs1064
hs1067
hs1071
hs1078
hs1079
hs1080
hs1081
hs1082
hs1085
hs1092
hs1100
hs1102
hs1109
hs1114
hs1116
hs1117
hs1123
hs1131
hs1139
hs1149
hs1150
hs1151
hs1152
hs1153
hs1166
hs1168
hs1170
hs1172
hs1174
hs1177
hs1180
hs1181
hs1186
hs1187
hs1199
hs1202
hs1210
hs1213
hs1218
hs1220
hs1223
hs1235
hs1239
hs1304

hs1309
hs1311
hs1314
hs1316
hs1318
hs1321
hs1322
hs1324
hs1326
hs1328
hs1330
hs1331
hs1333
hs1334
hs1335
hs1336
hs1340
hs1344
hs1345
hs1348
hs1350
hs1358
hs1359
hs1360
hs1361
hs1362
hs1363
hs1366
hs1369
hs1374
hs1382
hs1389
hs1391
hs1393
hs1425
hs1430
hs1432
hs1433
hs1435
hs1437
hs1440
hs1442
hs1450
hs1453
hs1460
hs1461
hs1463
hs1464
hs1465
hs1469
hs1477
hs1484
hs1486
hs1491
hs1497
hs1500
hs1510
hs1516
hs1517
hs1523
hs1525
hs1532
hs1534
hs1537
hs1539
hs1540
hs1546
hs1553
hs1564
hs1567
hs1574
hs1575
hs1577
hs1578
hs1586
hs1588
hs1589
hs1633
hs1634
hs1638

hs1650
hs1652
hs1653
hs1658
hs1659
hs1660
hs1665
hs1670
hs1673
hs1675
hs1702
hs1720
hs1723
hs1724
hs1725
hs1734
hs1741
hs1746
hs1750
hs1753
hs1760
hs1763
hs1764
hs1766
hs1767
hs1769
hs1791
hs1800
hs1802
hs1809
hs1811
hs1828
hs1859
hs1862
hs1867
hs1881
hs1887
hs1891
hs1912
hs1915
hs1920
hs1931
hs1932
hs1933
hs1937
hs1948
hs1958
hs1959
hs1962
hs1967
hs1974
hs1991
hs1995
hs2007
hs2026
hs2033
hs2040
hs2048
hs2050
hs2059
hs2061
hs2062
hs2063
hs2064
hs2070
hs2073
hs2077
hs2078
hs2081
hs2084
hs2088
hs2090
hs2094
hs2096
hs2308
hs2313

2

hs4
hs12
hs20
hs23
hs26
hs37
hs43
hs51
hs65
hs71
hs73
hs74
hs76
hs79
hs108
hs110
hs113
hs118
hs119
hs121
hs122
hs124
hs131
hs135
hs141
hs142
hs145
hs149
hs151
hs152
hs161
hs162
hs168
hs169
hs170
hs186
hs192
hs205
hs217
hs222
hs232
hs238
hs240
hs246
hs249
hs258
hs259
hs260
hs262
hs266
hs267
hs271
hs293
hs297
hs302
hs304
hs309
hs311
hs312
hs313
hs327
hs329
hs330
hs355
hs359
hs362
hs388
hs394
hs399
hs401
hs411
hs413
hs415
hs421
hs422
hs426
hs434

hs488
hs491
hs503
hs511
hs521
hs522
hs526
hs529
hs536
hs541
hs543
hs550
hs559
hs562
hs573
hs576
hs578
hs580
hs582
hs592
hs609
hs627
hs629
hs631
hs633
hs638
hs643
hs644
hs649
hs654
hs655
hs660
hs669
hs680
hs692
hs699
hs701
hs704
hs712
hs722
hs737
hs749
hs755
hs765
hs779
hs781
hs789
hs793
hs796
hs799
hs807
hs831
hs834
hs836
hs841
hs842
hs844
hs848
hs853
hs860
hs878
hs881
hs908
hs921
hs930
hs932
hs947
hs953
hs956
hs966
hs974
hs975
hs980
hs981
hs1002
hs1004
hs1006
hs1013
hs1019
hs1025

hs1032
hs1041
hs1093
hs1107
hs1115
hs1118
hs1119
hs1122
hs1130
hs1142
hs1147
hs1161
hs1163
hs1173
hs1175
hs1188
hs1191
hs1192
hs1198
hs1203
hs1205
hs1208
hs1212
hs1217
hs1224
hs1226
hs1236
hs1242
hs1244
hs1251
hs1258
hs1271
hs1273
hs1278
hs1300
hs1302
hs1303
hs1308
hs1315
hs1319
hs1320
hs1327
hs1329
hs1332
hs1341
hs1346
hs1351
hs1352
hs1364
hs1371
hs1375
hs1383
hs1384
hs1387
hs1388
hs1394
hs1397
hs1418
hs1419
hs1422
hs1427
hs1428
hs1436
hs1438
hs1441
hs1447
hs1448
hs1452
hs1462
hs1466
hs1467
hs1468
hs1472
hs1473
hs1482
hs1485
hs1487
hs1488
hs1496
hs1498

hs1507
hs1526
hs1527
hs1530
hs1535
hs1538
hs1543
hs1544
hs1548
hs1551
hs1554
hs1555
hs1557
hs1559
hs1566
hs1571
hs1573
hs1581
hs1582
hs1585
hs1635
hs1636
hs1640
hs1642
hs1643
hs1644
hs1651
hs1657
hs1667
hs1672
hs1699
hs1707
hs1709
hs1715
hs1717
hs1726
hs1738
hs1743
hs1745
hs1747
hs1748
hs1751
hs1754
hs1783
hs1798
hs1831
hs1833
hs1837
hs1854
hs1857
hs1858
hs1860
hs1866
hs1875
hs1882
hs1886
hs1909
hs1945
hs1950
hs1951
hs1955
hs1963
hs1971
hs1989
hs2060
hs2071
hs2082
hs2089
hs2263
hs2306
hs2307
hs2312

LISTA A7: Risultati del test binomiale e bootstrap

Elenco completo delle CFT significative identificate per mezzo del test binomiale e validate mediante bootstrap. Ogni riga rappresenta una CFT significativa formata da i FT indicati nei campi "FT-enh" (FT all'enhancer) e "FT-pro" (FT al promotore). Le coppie sono disposte in ordine di p-value (calcolato con il test binomiale) crescente. Nei campi "k", "te", "tp", "p" sono riportati i valori utilizzati per il calcolo del test binomiale. I nomi di questi campi hanno un significato identico ai valori mostrati nella *Formula 3.2*, vale a dire: "k" è il numero di CEP contenenti la CFT, "te" indica il numero di CEP che mostrano la presenza di FT-enh all'enhancer, similmente "tp" il numero di CEP che mostrano la presenza di FT-pro al promotore. "te" e "tp" sono utilizzati per calcolare la probabilità di rinvenimento casuale della CFT riportata nel campo "p". Si ricorda che il valore di *n* della *Formula 3.2* è sempre pari a 1011 e non è riportato delle righe dell'elenco per brevità. Il campo "r" corrisponde al valore dell'omonima variabile in *Formula 5.1* usata per la procedura di bootstrap. Nella stessa formula il valore "R" è sempre pari a 100000 e non è ripetuto nelle righe dell'elenco. "r" ed "R" sono usati per calcolare il p-value mostrato nel campo "P-value (bootstrap)". Il campo "CFT" infine è un codice numerico identificativo della coppia.

CFT	FT-enh	FT-pro	k	te	tp	p	P-value (binomiale)	r	P-value (bootstrap)
1	SP2	FOXC2	2	4	3	1.17e-05	6.98e-05	3	3.00e-05
2	SP2	TP73	2	4	3	1.17e-05	6.98e-05	3	3.00e-05
3	NHLH1	SREBF1	2	3	5	1.47e-05	1.09e-04	9	9.00e-05
4	RARG	NR2C2	2	3	5	1.47e-05	1.09e-04	2	2.00e-05
5	RARG	PPARA	2	3	5	1.47e-05	1.09e-04	2	2.00e-05
6	RARG	RXRA	2	3	5	1.47e-05	1.09e-04	2	2.00e-05
7	RARG	NR2F2	2	3	7	2.05e-05	2.13e-04	11	1.10e-04
8	ZFX	NR1H2	2	15	2	2.94e-05	4.31e-04	23	2.30e-04
9	TP73	NFIC	2	3	12	3.52e-05	6.19e-04	49	4.90e-04
10	NFYA	WT1	3	3	65	1.91e-04	1.03e-03	34	3.40e-04
11	NFYB	WT1	3	3	65	1.91e-04	1.03e-03	34	3.40e-04
12	PBX3	WT1	3	3	65	1.91e-04	1.03e-03	34	3.40e-04
13	JUN	POU3F2	2	9	6	5.28e-05	1.38e-03	96	9.60e-04
14	EBF1	PRDM1	1	2	1	1.96e-06	1.98e-03	200	2.00e-03
15	DDIT3	BCL6	1	2	1	1.96e-06	1.98e-03	209	2.09e-03
16	FOXC2	EPA51	1	1	2	1.96e-06	1.98e-03	204	2.04e-03
17	NR3C2	FOXO3	1	2	1	1.96e-06	1.98e-03	216	2.16e-03
18	TCF3	POU1F1	1	1	2	1.96e-06	1.98e-03	197	1.97e-03
19	NFYC	WT1	3	4	65	2.54e-04	2.33e-03	99	9.90e-04
20	KLF15	ZBTB7B	12	30	155	4.55e-03	2.79e-03	82	8.20e-04
21	SP4	ZNF148	17	61	130	7.76e-03	2.95e-03	112	1.12e-03
22	EGR4	ATF5	1	3	1	2.94e-06	2.96e-03	286	2.86e-03
23	NFYA	ZBTB4	1	3	1	2.94e-06	2.96e-03	275	2.75e-03
24	NFYB	ZBTB4	1	3	1	2.94e-06	2.96e-03	275	2.75e-03
25	PBX3	ZBTB4	1	3	1	2.94e-06	2.96e-03	275	2.75e-03
26	SPI1	FOS	1	3	1	2.94e-06	2.96e-03	273	2.73e-03
27	ZNF350	EGR2	1	3	1	2.94e-06	2.96e-03	267	2.67e-03
28	PPARA	SOX5	16	164	45	7.22e-03	3.46e-03	94	9.40e-04
29	ZFX	ZNF148	7	15	130	1.91e-03	3.70e-03	123	1.23e-03
30	ARNT2	SOX17	1	2	2	3.91e-06	3.95e-03	360	3.60e-03
31	ELF3	FOXO3	1	4	1	3.91e-06	3.95e-03	443	4.43e-03
32	FOXO1	TEF	1	2	2	3.91e-06	3.95e-03	405	4.05e-03
33	FOXO1	ZNF350	1	2	2	3.91e-06	3.95e-03	377	3.77e-03
34	LHX2	IRF2	1	4	1	3.91e-06	3.95e-03	391	3.91e-03
35	NFYC	ZBTB4	1	4	1	3.91e-06	3.95e-03	383	3.83e-03
36	NR2C1	PAX5	1	1	4	3.91e-06	3.95e-03	425	4.25e-03
37	SMAD1	PAX5	1	1	4	3.91e-06	3.95e-03	411	4.11e-03
38	TCF3	USF1	1	1	4	3.91e-06	3.95e-03	411	4.11e-03
39	PAX2	ZBTB7B	11	28	155	4.25e-03	4.67e-03	150	1.50e-03
40	MECOM	ELF3	1	5	1	4.89e-06	4.93e-03	484	4.84e-03

CFT	FT-enh	FT-pro	k	te	tp	p	P-value (binomiale)	r	P-value (bootstrap)
41	SRF	NFE2L2	1	1	5	4.89e-06	4.93e-03	483	4.83e-03
42	SRF	STAT1	1	1	5	4.89e-06	4.93e-03	483	4.83e-03
43	TFCP2	PRDM1	1	5	1	4.89e-06	4.93e-03	482	4.82e-03
44	ZBTB6	SREBF1	1	1	5	4.89e-06	4.93e-03	507	5.07e-03
45	ZNF148	ESR2	21	88	128	1.10e-02	5.11e-03	185	1.85e-03
46	ATF6	THR3	1	3	2	5.87e-06	5.92e-03	573	5.73e-03
47	ONECUT1	REST	1	1	6	5.87e-06	5.92e-03	588	5.88e-03
48	IRF1	PLAG1	1	2	3	5.87e-06	5.92e-03	555	5.55e-03
49	IRF2	PLAG1	1	2	3	5.87e-06	5.92e-03	555	5.55e-03
50	IRF7	TCF7L2	1	2	3	5.87e-06	5.92e-03	600	6.00e-03
51	IRF9	TCF7L2	1	2	3	5.87e-06	5.92e-03	600	6.00e-03
52	KLF1	MBD2	1	6	1	5.87e-06	5.92e-03	590	5.90e-03
53	NFATC1	RARG	1	3	2	5.87e-06	5.92e-03	603	6.03e-03
54	PPARG	SPZ1	1	3	2	5.87e-06	5.92e-03	572	5.72e-03
55	PRDM1	PLAG1	1	2	3	5.87e-06	5.92e-03	555	5.55e-03
56	RARG	NR1H2	1	3	2	5.87e-06	5.92e-03	593	5.93e-03
57	SPI1	ATF3	1	3	2	5.87e-06	5.92e-03	566	5.66e-03
58	SPI1	STAT3	1	3	2	5.87e-06	5.92e-03	566	5.66e-03
59	SPIB	JUND	1	3	2	5.87e-06	5.92e-03	567	5.67e-03
60	SRF	POU3F2	1	1	6	5.87e-06	5.92e-03	575	5.75e-03
61	STAT2	PLAG1	1	2	3	5.87e-06	5.92e-03	555	5.55e-03
62	STAT4	FOXO1	1	3	2	5.87e-06	5.92e-03	606	6.06e-03
63	TLX1	MYOG	1	3	2	5.87e-06	5.92e-03	583	5.83e-03
64	ZNF143	ELF2	1	3	2	5.87e-06	5.92e-03	602	6.02e-03
65	ZNF350	STAT3	1	3	2	5.87e-06	5.92e-03	634	6.34e-03
66	ZNF350	ZNF350	1	3	2	5.87e-06	5.92e-03	623	6.23e-03
67	ZNF148	ZNF148	21	88	130	1.12e-02	6.04e-03	199	1.99e-03
68	NR2C1	ZIC1	1	1	7	6.85e-06	6.90e-03	714	7.14e-03
69	NR2C1	ZIC2	1	1	7	6.85e-06	6.90e-03	714	7.14e-03
70	SRF	MAFK	1	1	7	6.85e-06	6.90e-03	690	6.90e-03
71	PLAG1	PAX5	2	32	4	1.25e-04	7.36e-03	541	5.41e-03
72	TFAP2B	THR3	1	4	2	7.83e-06	7.88e-03	806	8.06e-03
73	ENO1	STAT2	1	2	4	7.83e-06	7.88e-03	763	7.63e-03
74	NR3C1	PAX5	1	2	4	7.83e-06	7.88e-03	780	7.80e-03
75	RFX3	NHLH1	1	8	1	7.83e-06	7.88e-03	801	8.01e-03
76	THR3	MYOG	1	4	2	7.83e-06	7.88e-03	739	7.39e-03
77	TWIST1	MBD2	1	8	1	7.83e-06	7.88e-03	815	8.15e-03
78	ZFX	SP4	9	15	228	3.35e-03	7.93e-03	181	1.81e-03
79	ZNF148	SP3	32	88	231	1.99e-02	8.07e-03	181	1.81e-03
80	NR2C2	SOX5	2	3	45	1.32e-04	8.15e-03	566	5.66e-03

CFT	FT-enh	FT-pro	k	te	tp	p	P-value (binomiale)	r	P-value (bootstrap)
297	SPZ1	PBX3	1	2	20	3.91e-05	3.88e-02	3853	3.85e-02
298	TCF3	THRA	1	1	40	3.91e-05	3.88e-02	3925	3.93e-02
299	ZBTB6	RREB1	1	1	40	3.91e-05	3.88e-02	3937	3.94e-02
300	EGR2	SOX5	2	7	45	3.08e-04	3.95e-02	3512	3.51e-02
301	SOX13	GATA3	8	96	40	3.76e-03	3.97e-02	2883	2.88e-02
302	HLTF	KLF15	1	1	41	4.01e-05	3.97e-02	3930	3.93e-02
303	SNAI1	RBPJ	2	79	4	3.09e-04	3.97e-02	3213	3.21e-02
304	SP1	ESR2	16	77	128	9.64e-03	3.98e-02	2508	2.51e-02
305	SP4	SP4	21	61	228	1.36e-02	4.02e-02	1967	1.97e-02
306	NFE2L1	ARID3A	6	109	23	2.45e-03	4.05e-02	3088	3.09e-02
307	ATF6	NRF1	1	3	14	4.11e-05	4.07e-02	4041	4.04e-02
308	NR4A3	MAFK	1	6	7	4.11e-05	4.07e-02	4073	4.07e-02
309	SMAD2	MAF	1	1	42	4.11e-05	4.07e-02	4173	4.17e-02
310	ZBTB4	ZNF589	1	3	14	4.11e-05	4.07e-02	4005	4.01e-02
311	ZNF143	NRF1	1	3	14	4.11e-05	4.07e-02	4069	4.07e-02
312	PPARA	ARID5B	2	164	2	3.21e-04	4.25e-02	2611	2.61e-02
313	CXXC1	NR0B1	1	22	2	4.30e-05	4.26e-02	4318	4.32e-02
314	RREB1	ESR2	9	36	128	4.51e-03	4.26e-02	2886	2.89e-02
315	KLF15	ZNF148	8	30	130	3.82e-03	4.28e-02	3058	3.06e-02
316	IRF4	WT1	3	12	65	7.63e-04	4.34e-02	3590	3.59e-02
317	AHR	SOX5	1	1	45	4.40e-05	4.35e-02	4363	4.36e-02
318	BCL6	SOX5	1	1	45	4.40e-05	4.35e-02	4498	4.50e-02
319	JUN	NFE2L2	1	9	5	4.40e-05	4.35e-02	4352	4.35e-02
320	JUN	STAT1	1	9	5	4.40e-05	4.35e-02	4352	4.35e-02
321	PAX2	NFIC	2	28	12	3.29e-04	4.44e-02	4110	4.11e-02
322	ENO1	ARID3A	1	2	23	4.50e-05	4.45e-02	4404	4.40e-02
323	FOXO3	NFYA	1	2	23	4.50e-05	4.45e-02	4513	4.51e-02
324	NR2E3	NFYA	1	2	23	4.50e-05	4.45e-02	4582	4.58e-02
325	SPZ1	NFYA	1	2	23	4.50e-05	4.45e-02	4513	4.51e-02
326	SP1	ZNF148	16	77	130	9.79e-03	4.46e-02	2885	2.89e-02
327	SP3	ESR2	15	72	128	9.02e-03	4.46e-02	2973	2.97e-02
328	SP4	PURA	3	61	13	7.76e-04	4.51e-02	3864	3.86e-02
329	CXXC1	SP1	9	22	212	4.56e-03	4.53e-02	2597	2.60e-02
330	AHR	FOXO1	1	1	47	4.60e-05	4.54e-02	4548	4.55e-02
331	HOXD10	POU2F2	2	9	38	3.35e-04	4.58e-02	4253	4.25e-02
332	SOX13	NRF1	4	96	14	1.31e-03	4.60e-02	3607	3.61e-02
333	TCF12	THRA	3	20	40	7.83e-04	4.61e-02	4141	4.14e-02
334	CTCF	SP2	1	6	8	4.70e-05	4.64e-02	4525	4.52e-02
335	ETV4	BCL6	1	48	1	4.70e-05	4.64e-02	4779	4.78e-02
336	ETV4	NR2F1	1	48	1	4.70e-05	4.64e-02	4803	4.80e-02
337	ETV4	IRF2	1	48	1	4.70e-05	4.64e-02	4809	4.81e-02
338	ETV4	NKX2-2	1	48	1	4.70e-05	4.64e-02	4753	4.75e-02
339	ETV4	RXRG	1	48	1	4.70e-05	4.64e-02	4803	4.80e-02
340	ETV4	SOX2	1	48	1	4.70e-05	4.64e-02	4719	4.72e-02
341	HOXB6	NFIC	1	4	12	4.70e-05	4.64e-02	4597	4.60e-02
342	SREBF2	PAX5	1	12	4	4.70e-05	4.64e-02	4735	4.74e-02
343	SOX15	REST	4	225	6	1.32e-03	4.66e-02	2488	2.49e-02
344	MEF2A	MZF1	1	1	49	4.79e-05	4.73e-02	4723	4.72e-02
345	MEF2C	MZF1	1	1	49	4.79e-05	4.73e-02	4723	4.72e-02
346	HOXD9	NFE2L2	1	10	5	4.89e-05	4.83e-02	4759	4.76e-02
347	HOXD9	NR2C2	1	10	5	4.89e-05	4.83e-02	4738	4.74e-02
348	HOXD9	PPARA	1	10	5	4.89e-05	4.83e-02	4738	4.74e-02
349	HOXD9	RXRA	1	10	5	4.89e-05	4.83e-02	4738	4.74e-02
350	HOXD9	STAT1	1	10	5	4.89e-05	4.83e-02	4759	4.76e-02

CFT	FT-enh	FT-pro	k	te	tp	p	P-value (binomiale)	r	P-value (bootstrap)
351	NR1H2	TFAP2A	1	2	25	4.89e-05	4.83e-02	4818	4.82e-02
352	RARB	NR2C2	1	10	5	4.89e-05	4.83e-02	4874	4.87e-02
353	RARB	PPARA	1	10	5	4.89e-05	4.83e-02	4874	4.87e-02
354	RARB	RXRA	1	10	5	4.89e-05	4.83e-02	4874	4.87e-02
355	RARB	SREBF1	1	10	5	4.89e-05	4.83e-02	4955	4.95e-02
356	ZBTB18	TFAP2A	1	2	25	4.89e-05	4.83e-02	4818	4.82e-02
357	PPARA	SMARCC1	3	164	5	8.02e-04	4.90e-02	3242	3.24e-02
358	HSF1	OVOL1	1	3	17	4.99e-05	4.92e-02	4946	4.95e-02
359	NFYA	BRCA1	1	3	17	4.99e-05	4.92e-02	4976	4.98e-02
360	NFYB	BRCA1	1	3	17	4.99e-05	4.92e-02	4976	4.98e-02
361	TP73	OVOL1	1	3	17	4.99e-05	4.92e-02	4967	4.97e-02
362	PBX3	BRCA1	1	3	17	4.99e-05	4.92e-02	4976	4.98e-02
363	SPIB	BRCA1	1	3	17	4.99e-05	4.92e-02	4928	4.93e-02
364	ZBTB4	OVOL1	1	3	17	4.99e-05	4.92e-02	4989	4.99e-02

LISTA A8: CEP per ogni CFT

Elenco completo di tutte le CEP sulle quali è presente ogni CFT. Le righe in grassetto mostrano il codice identificativo della CFT e i FT che la compongono (primo FT all'enhancer e secondo al promotore). A seguire le CEP, una per riga, comprese del proprio codice di identificazione seguito dall'identificativo dell'enhancer (equivalente al VISTA ID) e dal "gene symbol" del gene del promotore associato.

CFT1 - SP2,FOXC2
CEP99: hs241,CASZ1
CEP100: hs289,CASZ1

CFT2 - SP2,TP73
CEP99: hs241,CASZ1
CEP100: hs289,CASZ1

CFT3 - NHLH1,SREBF1
CEP24: hs675,ARHGAP15
CEP57: hs1891,BCAN

CFT4 - RARG,NR2C2
CEP110: hs1931,CCDC85C
CEP759: hs1971,RBM38

CFT5 - RARG,PPARA
CEP110: hs1931,CCDC85C
CEP759: hs1971,RBM38

CFT6 - RARG,RXRA
CEP110: hs1931,CCDC85C
CEP759: hs1971,RBM38

CFT7 - RARG,NR2F2
CEP110: hs1931,CCDC85C
CEP759: hs1971,RBM38

CFT8 - ZFX,NR1H2
CEP331: hs1665,GPR124
CEP759: hs1971,RBM38

CFT9 - TP73,NFIC
CEP638: hs1484,NFIA
CEP785: hs152,SALL1

CFT10 - NFYA,WT1
CEP332: hs488,GPR180
CEP848: hs488,SOX21
CEP941: hs2007,TSPAN17

CFT11 - NFYB,WT1
CEP332: hs488,GPR180
CEP848: hs488,SOX21
CEP941: hs2007,TSPAN17

CFT12 - PBX3,WT1
CEP332: hs488,GPR180
CEP848: hs488,SOX21
CEP941: hs2007,TSPAN17

CFT13 - JUN,POU3F2
CEP211: hs1418,EN2
CEP595: hs566,MIR548AI

CFT14 - EBF1,PRDM1
CEP806: hs1748,SERPINH1

CFT15 - DDIT3,BCL6
CEP135: hs1747,CISH

CFT16 - FOXC2,EPAS1
CEP767: hs1517,RGS9

CFT17 - NR3C2,FOXO3
CEP371: hs1527,ID2

CFT18 - TCF3,POU1F1
CEP231: hs200,FAF1

CFT19 - NFYC,WT1
CEP332: hs488,GPR180
CEP848: hs488,SOX21

CEP941: hs2007,TSPAN17

CFT20 - KLF15,ZBTB7B

CEP154: hs152,CYLD
CEP155: hs162,CYLD
CEP243: hs782,FGF17
CEP326: hs1937,GPC1
CEP378: hs1763,INPP5A
CEP401: hs1500,KAT6B
CEP641: hs1205,NKX2-2
CEP695: hs1958,PFKP
CEP742: hs1955,PKI
CEP785: hs152,SALL1
CEP786: hs162,SALL1
CEP807: hs1362,SETBP1

CFT21 - SP4,ZNF148

CEP25: hs1912,ARHGEF16
CEP43: hs1328,ASS1
CEP57: hs1891,BCAN
CEP150: hs1971,CTCF
CEP317: hs921,GBX2
CEP325: hs1750,GPC1
CEP326: hs1937,GPC1
CEP334: hs1882,GPR20
CEP632: hs627,NEUROD2
CEP635: hs1699,NEUROG3
CEP641: hs1205,NKX2-2
CEP660: hs1548,OLIG2
CEP678: hs1202,PBX1
CEP759: hs1971,RBM38
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3
CEP978: hs1909,ZFP36L2

CFT22 - EGR4,ATF5
CEP218: hs277,ERI3

CFT23 - NFYA,ZBTB4
CEP941: hs2007,TSPAN17

CFT24 - NFYB,ZBTB4
CEP941: hs2007,TSPAN17

CFT25 - PBX3,ZBTB4
CEP941: hs2007,TSPAN17

CFT26 - SPI1,FOS
CEP357: hs629,HIBADH

CFT27 - ZNF350,EGR2
CEP616: hs1488,MRPL23

CFT28 - PPARA,SOX5
CEP46: hs1432,ATG10
CEP89: hs1955,C6orf118
CEP202: hs309,ELAVL2
CEP232: hs1364,FAM110B
CEP480: hs1525,LOC285419
CEP481: hs1582,LOC285419
CEP512: hs1496,LOC727982
CEP588: hs205,MIR4778
CEP713: hs108,PMFBP1
CEP714: hs20,PMFBP1
CEP823: hs296,SKAP2
CEP826: hs1440,SLC14A2
CEP827: hs1464,SLC14A2
CEP871: hs1582,SPRY1
CEP1007: hs631,ZNF703
CEP1008: hs834,ZNF703

CFT29 - ZFX,ZNF148
CEP150: hs1971,CTCF
CEP331: hs1665,GPR124

CEP334: hs1882,GPR20
CEP384: hs1327,IRX2
CEP739: hs1886,PTMA
CEP759: hs1971,RBM38
CEP785: hs152,SALL1

CFT30 - ARNT2,SOX17
CEP69: hs1466,C14orf166B

CFT31 - ELF3,FOXO3
CEP371: hs1527,ID2

CFT32 - FOXD1,TEF
CEP460: hs161,LOC100505619

CFT33 - FOXO1,ZNF350
CEP104: hs1448,CBLN4

CFT34 - LHX2,IRF2
CEP629: hs1341,NEDD1

CFT35 - NFYC,ZBTB4
CEP941: hs2007,TSPAN17

CFT36 - NR2C1,PAX5
CEP118: hs1760,CDH16

CFT37 - SMAD1,PAX5
CEP25: hs1912,ARHGEF16

CFT38 - TCF3,USF1
CEP231: hs200,FAF1

CFT39 - PAX2,ZBTB7B
CEP125: hs2060,CDYL
CEP154: hs152,CYLD
CEP198: hs2060,ECI2
CEP211: hs1418,EN2
CEP355: hs1543,HES1
CEP394: hs23,IRX6
CEP605: hs23,MMP2
CEP785: hs152,SALL1
CEP789: hs73,SALL1
CEP816: hs2062,SHB
CEP865: hs953,SP9

CFT40 - MECOM,ELF3
CEP142: hs1951,COL6A3

CFT41 - SRE,NFE2L2
CEP594: hs1539,MIR548AI

CFT42 - SRE,STAT1
CEP594: hs1539,MIR548AI

CFT43 - TFCEP2,PRDM1
CEP806: hs1748,SERPINH1

CFT44 - ZBTB6,SREBF1
CEP57: hs1891,BCAN

CFT45 - ZNF148,ESR2
CEP25: hs1912,ARHGEF16
CEP64: hs307,BNC2
CEP69: hs1466,C14orf166B
CEP140: hs1418,CNRY1
CEP155: hs162,CYLD
CEP321: hs522,GLI2
CEP332: hs488,GPR180
CEP336: hs1371,GRAMD3
CEP470: hs388,LOC100506274
CEP483: hs1001,LOC339442
CEP487: hs2088,LOC339442
CEP491: hs388,LOC339788

CEP553: hs699,MGMT
CEP618: hs516,MRPS30
CEP664: hs262,OTP
CEP759: hs1971,RBM38
CEP799: hs118,SCARNA23
CEP835: hs1507,SLC4A3
CEP868: hs1754,SPC24
CEP883: hs262,TBCA
CEP978: hs1909,ZFP36L2

CFT46 - ATF6,THRB
CEP882: hs1783,TBCA

CFT47 - ONECUT1,REST
CEP407: hs748,KCNMA1

CFT48 - IRF1,PLAG1
CEP648: hs580,NR4A2

CFT49 - IRF2,PLAG1
CEP648: hs580,NR4A2

CFT50 - IRF7,TCF7L2
CEP115: hs853,CCNH

CFT51 - IRF9,TCF7L2
CEP115: hs853,CCNH

CFT52 - KLF1,MBD2
CEP102: hs1854,CBFA2T3

CFT53 - NFATC1,RARG
CEP235: hs1860,FAM65C

CFT54 - PPARG,SPZ1
CEP558: hs1388,MIR3605

CFT55 - PRDM1,PLAG1
CEP648: hs580,NR4A2

CFT56 - RARG,NR1H2
CEP759: hs1971,RBM38

CFT57 - SPI1,ATF3
CEP357: hs629,HIBADH

CFT58 - SPI1,STAT3
CEP357: hs629,HIBADH

CFT59 - SPIB,JUND
CEP380: hs1389,INTU

CFT60 - SRE,POU3F2
CEP594: hs1539,MIR548AI

CFT61 - STAT2,PLAG1
CEP648: hs580,NR4A2

CFT62 - STAT4,FOXJ1
CEP290: hs1064,FOXG1

CFT63 - TLX1,MYOG
CEP760: hs2084,RCAN1

CFT64 - ZNF143,ELF2
CEP509: hs831,LOC644838

CFT65 - ZNF350,STAT3
CEP617: hs1488,MRPL23-AS1

CFT66 - ZNF350,ZNF350
CEP617: hs1488,MRPL23-AS1

CFT67 - ZNF148,ZNF148

CEP25: hs1912,ARHGEF16
CEP57: hs1891,BCAN
CEP64: hs307,BNC2
CEP140: hs1418,CNRY1
CEP150: hs1971,CTCF
CEP326: hs1937,GPC1
CEP332: hs488,GPR180
CEP334: hs1882,GPR20
CEP382: hs1466,IRF2BPL
CEP401: hs1500,KAT6B
CEP603: hs699,MKI67
CEP635: hs1699,NEUROG3
CEP641: hs1205,NKX2-2
CEP678: hs1202,PBX1
CEP759: hs1971,RBM38
CEP786: hs162,SALL1
CEP845: hs1496,SOX11
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3
CEP905: hs974,TLE1
CEP978: hs1909,ZFP36L2

CFT68 - NR2C1,ZIC1
CEP118: hs1760,CDH16

CFT69 - NR2C1,ZIC2
CEP118: hs1760,CDH16

CFT70 - SRE,MAFK
CEP594: hs1539,MIR548AI

CFT71 - PLAG1,PAX5
CEP25: hs1912,ARHGEF16
CEP118: hs1760,CDH16

CFT72 - TFAP2B,THRB
CEP883: hs262,TBCA

CFT73 - ENO1,STAT2
CEP693: hs799,PER4

CFT74 - NR3C1,PAX5
CEP25: hs1912,ARHGEF16

CFT75 - RFX3,NHLH1
CEP845: hs1496,SOX11

CFT76 - THRB,MYOG
CEP77: hs1886,C2orf57

CFT77 - TWIST1,MBD2
CEP102: hs1854,CBFA2T3

CFT78 - ZFX,SP4
CEP135: hs1747,CISH
CEP154: hs152,CYLD
CEP331: hs1665,GPR124
CEP354: hs1747,HEMK1
CEP384: hs1327,IRX2
CEP739: hs1886,PTMA
CEP759: hs1971,RBM38
CEP785: hs152,SALL1
CEP942: hs1967,TTC7A

CFT79 - ZNF148,SP3
CEP8: hs680,AGPAT9
CEP25: hs1912,ARHGEF16
CEP40: hs1540,ASCL1
CEP64: hs307,BNC2
CEP110: hs1931,CCDC85C
CEP112: hs1920,CCND1
CEP153: hs260,CXXC4
CEP193: hs1208,EBF1

CEP464:
hs1198,LOC100505768
CEP658: hs1346,OLIG1
CEP692: hs793,PER4

CFT307 - ATF6,NRF1
CEP23: hs1783,AP3B1

CFT308 - NR4A3,MAFK
CEP588: hs205,MIR4778

CFT309 - SMAD2,MAF
CEP278: hs1170,FLJ42709

CFT310 - ZBTB4,ZNF589
CEP916: hs313,TLE4

CFT311 - ZNF143,NRF1
CEP758: hs1673,RBM19

CFT312 - PPARA,ARID5B
CEP571: hs1507,MIR4268
CEP606: hs1374,MMRN1

CFT313 - CXXC1,NR0B1
CEP641: hs1205,NKX2-2

CFT314 - RREB1,ESR2
CEP92: hs1316,CADPS
CEP127: hs2263,CELSR2
CEP140: hs1418,CNPY1
CEP485: hs1139,LOC339442
CEP800: hs119,SCARNA23
CEP803: hs145,SCARNA23
CEP835: hs1507,SLC4A3
CEP868: hs1754,SPC24
CEP978: hs1909,ZFP36L2

CFT315 - KLF15,ZNF148
CEP326: hs1937,GPC1
CEP401: hs1500,KAT6B
CEP632: hs627,NEUROD2
CEP641: hs1205,NKX2-2
CEP678: hs1202,PBX1
CEP742: hs1955,QKI
CEP785: hs152,SALL1
CEP786: hs162,SALL1

CFT316 - IRF4,WT1
CEP228: hs1866,EXOC6
CEP333: hs796,GPR180
CEP849: hs796,SOX21

CFT317 - AHR,SOX5
CEP175:
hs1802,DKFZp686O1327

CFT318 - BCL6,SOX5
CEP378: hs1763,INPP5A

CFT319 - JUN,NFE2L2
CEP595: hs566,MIR548AI

CFT320 - JUN,STAT1
CEP595: hs566,MIR548AI

CFT321 - PAX2,NFIC
CEP785: hs152,SALL1
CEP789: hs73,SALL1

CFT322 - ENO1,ARID3A
CEP693: hs799,PER4

CFT323 - FOXO3,NFYA
CEP918: hs765,TLE4

CFT324 - NR2E3,NFYA
CEP909: hs1118,TLE4

CFT325 - SPZ1,NFYA
CEP918: hs765,TLE4

CFT326 - SP1,ZNF148
CEP25: hs1912,ARHGEF16

CEP57: hs1891,BCAN
CEP150: hs1971,CTCF
CEP302: hs1100,FOXP4
CEP334: hs1882,GPR20
CEP382: hs1466,IRF2BPL
CEP384: hs1327,IRX2
CEP385: hs633,IRX2
CEP641: hs1205,NKX2-2
CEP739: hs1886,PTMA
CEP759: hs1971,RBM38
CEP786: hs162,SALL1
CEP867: hs1320,SPATA8
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3
CEP978: hs1909,ZFP36L2

CFT327 - SP3,ESR2
CEP25: hs1912,ARHGEF16
CEP69: hs1466,C14orf166B
CEP118: hs1760,CDH16
CEP154: hs152,CYLD
CEP321: hs522,GLI2
CEP470: hs388,LOC100506274
CEP483: hs1001,LOC339442
CEP491: hs388,LOC339788
CEP664: hs262,OTP
CEP759: hs1971,RBM38
CEP802: hs122,SCARNA23
CEP835: hs1507,SLC4A3
CEP868: hs1754,SPC24
CEP883: hs262,TBCA
CEP978: hs1909,ZFP36L2

CFT328 - SP4,PURA
CEP18: hs2090,ALX4
CEP43: hs1328,ASS1
CEP374: hs1932,IGF1R

CFT329 - CXXC1,SP1
CEP25: hs1912,ARHGEF16
CEP55: hs925,AUTS2
CEP237: hs1345,FBXL17
CEP378: hs1763,INPP5A
CEP676: hs1082,PAX6
CEP680: hs762,PBX1
CEP861: hs701,SP4
CEP875: hs1487,SSBP3
CEP1005: hs641,ZNF618

CFT330 - AHR,FOXO1
CEP175:
hs1802,DKFZp686O1327

CFT331 - HOXD10,POU2F2
CEP666: hs649,PABPC1P2
CEP979: hs1800,ZFPM2

CFT332 - SOX13,NRF1
CEP22: hs1589,ANKRD30A
CEP23: hs1783,AP3B1
CEP757: hs1571,RBM19
CEP758: hs1673,RBM19

CFT333 - TCF12,THRA
CEP82: hs981,C4orf21
CEP152: hs427,CXorf66
CEP458:
hs1726,LOC100287225

CFT334 - CTCE,SP2
CEP712: hs914,PLK1S1

CFT335 - ETV4,BCL6
CEP135: hs1747,CISH

CFT336 - ETV4,NR2F1
CEP779: hs2061,RREB1

CFT337 - ETV4,IRF2
CEP629: hs1341,NEDD1

CFT338 - ETV4,NKX2-2
CEP489: hs1042,LOC339568

CFT339 - ETV4,RXRG
CEP779: hs2061,RREB1

CFT340 - ETV4,SOX2
CEP228: hs1866,EXOC6

CFT341 - HOXB6,NFIC
CEP791: hs76,SALL1

CFT342 - SREBF2,PAX5
CEP118: hs1760,CDH16

CFT343 - SOX15,REST
CEP92: hs1316,CADPS
CEP342: hs1242,GRIN2A
CEP407: hs748,KCNMA1
CEP831: hs2313,SLC25A13

CFT344 - MEF2A,MZF1
CEP237: hs1345,FBXL17

CFT345 - MEF2C,MZF1
CEP237: hs1345,FBXL17

CFT346 - HOXD9,NFE2L2
CEP593: hs1523,MIR548AI

CFT347 - HOXD9,NR2C2
CEP989: hs213,ZIC4

CFT348 - HOXD9,PPARA
CEP989: hs213,ZIC4

CFT349 - HOXD9,RXRA
CEP989: hs213,ZIC4

CFT350 - HOXD9,STAT1
CEP593: hs1523,MIR548AI

CFT351 - NR1H2,TFAP2A
CEP975: hs20,ZFH3

CFT352 - RARB,NR2C2
CEP110: hs1931,CCDC85C

CFT353 - RARB,PPARA
CEP110: hs1931,CCDC85C

CFT354 - RARB,RXRA
CEP110: hs1931,CCDC85C

CFT355 - RARB,SREBF1
CEP57: hs1891,BCAN

CFT356 - ZBTB18,TFAP2A
CEP975: hs20,ZFH3

CFT357 - PPARA,SMARCC1
CEP69: hs1466,C14orf166B
CEP246: hs1707,FGG
CEP939: hs1915,TRIM29

CFT358 - HSF1,OVOL1
CEP637: hs1450,NFIA

CFT359 - NFYA,BRCA1
CEP332: hs488,GPR180

CFT360 - NFYB,BRCA1
CEP332: hs488,GPR180

CFT361 - TP73,OVOL1
CEP638: hs1484,NFIA

CFT362 - PBX3,BRCA1
CEP332: hs488,GPR180

CFT363 - SPIB,BRCA1
CEP777: hs1324,RPS6KC1

CFT364 - ZBTB4,OVOL1
CEP637: hs1450,NFIA

LISTA A9: Proteine intermedie più comuni

Elenco delle proteine intermedie più comuni capaci di collegare i FT delle CFT significative. Il valore numerico a sinistra indica il numero di CFT collegate da tale proteina intermedia. L'identità della proteina intermedia è mostrata sul lato destro così come fornita dal database STRING. Solo le proteine che fanno da intermedio ad almeno 3 CFT sono qui riportate.

CFT collegate	Intermedio	CFT collegate	Intermedio	CFT collegate	Intermedio	CFT collegate	Intermedio
98	CREBBP	13	SMARCD3	7	PTEN	5	TWIST1
93	EP300	13	NR4A2	7	PRL	5	TRIM24
77	UBC	13	MAPK14	7	PIAS3	5	TRAF2
71	SP1	13	BMP4	7	PCNA	5	TGS1
71	HDAC1	13	BCL6	7	MTA2	5	TCF12
60	TP53	12	YY1	7	MSX1	5	TAT
52	FOS	12	XRCC6	7	MMP2	5	SUV39H1
46	ESR1	12	TBL1XR1	7	MLL	5	STAT6
45	JUN	12	PAX6	7	MED6	5	SREBF2
43	HDAC2	12	MMP9	7	MED12	5	SPAG6
42	INS	12	ESRRB	7	KAT2A	5	SOX6
38	MYC	12	EGR1	7	ID3	5	SDC1
35	SUMO2	11	VIM	7	HRAS	5	RUNX1T1
34	RXRA	11	PIAS1	7	HMGA2	5	RPS4X
32	KAT2B	11	PBX1	7	GTF2B	5	RAN
32	HDAC3	11	NFKB1	7	FLT3	5	PRKCA
32	CCND1	11	NCOA6	7	EZH2	5	PPP1CA
31	SMARCA4	11	MYB	7	ERBB2	5	PPARD
31	SHH	11	MAX	7	EGF	5	POLR2A
31	PPARG	11	HIST2H3PS2	7	EED	5	PGR
29	TBP	11	HDAC4	7	DNMT3A	5	NUMB
29	SMAD3	11	GATA2	7	CSF2	5	NR0B1
29	NCOR1	11	CEBPB	7	CD44	5	NME1
29	AKT1	11	ALB	7	CASP3	5	MTA1
28	VEGFA	10	TRIM28	6	ZFPM2	5	MMP1
28	SUMO1	10	THRA	6	YAP1	5	MITF
28	SIN3A	10	SRY	6	XRCC5	5	MED24
28	NCOA2	10	PPARGC1A	6	UHRF2	5	MECP2
28	MAPK1	10	PAX2	6	UBA1	5	MEA1
27	CDKN1A	10	LEP	6	TRRAP	5	MAF
26	RELA	10	JAK1	6	SPP1	5	KPNA1
24	SMAD4	10	IGF1	6	SP11	5	KNTC1
24	RB1	10	HMGB1	6	SP7	5	KLK3
24	NOTCH1	10	ETS1	6	SOX9	5	KDM5C
24	AR	9	SIRT1	6	SOX10	5	JUNB
23	SRC	9	RUNX2	6	SOD2	5	ING2
23	PPARA	9	POU5F1	6	SMARCC2	5	IHH
23	NR0B2	9	POU2F1	6	RORA	5	IGFBP3
23	NCOR2	9	PIAS2	6	RBL2	5	ID2
23	MYOD1	9	PAX3	6	PTGS2	5	ICAM1
22	SMAD2	9	NRIP1	6	PRKDC	5	HOXD9
21	STAT3	9	NCOA3	6	PAX5	5	HOXA9
21	NR4A1	9	KPNA2	6	NEUROD1	5	HMGA1
21	NANOG	9	IGF2	6	MLL4	5	HIST3H3
21	HDAC9	9	H3F3B	6	MGMT	5	HCFC1
20	MAPK8	9	H3F3A	6	MEF2A	5	GPATCH2
19	VDR	9	GLI3	6	MED1	5	GATA1
19	RUNX1	9	GATA6	6	MDM2	5	GADD45A
19	HNF4A	9	DNMT1	6	MCL1	5	FOXL2
19	CTNNB1	8	UTY	6	MAPK11	5	FOSB
18	E2F1	8	TCF3	6	MAPK10	5	FNTB
18	CD4	8	SIN3B	6	LMO2	5	EPOR
17	UBE2I	8	RBBP4	6	JUND	5	DDX3X
17	NR1H2	8	OCT4	6	JAK2	5	DAXX
17	GATA4	8	NR2E1	6	IKBKAP	5	CYLD
17	GAPDH	8	MNX1	6	IFNG	5	CXCR4
16	HSPA4	8	MEIS1	6	HSP90AA1	5	CTNND1
16	ESR2	8	KDM6A	6	HOXD10	5	CSNK1E
16	BRCA1	8	KDM1A	6	HOXB1	5	COMP
15	TSPO	8	IGF1R	6	HIF1A	5	COL2A1
15	SRF	8	IFNB1	6	GLI1	5	CHD9
15	SOX2	8	GLI2	6	GATA5	5	CDK4
15	RARA	8	FOXO1	6	GATA3	5	CDK2
15	PARP1	8	DIF	6	FTH1	5	CDC25A
15	NR5A1	8	DDIT3	6	FOXA2	5	CCND3
15	NR3C1	8	CTGF	6	FASN	5	CBX8
15	NCOA1	8	CSNK2A1	6	ENSG00000230336	5	CASP7
15	MAPK3	8	CIITA	6	EIF2S3	5	AVPR2
15	EGFR	8	CHD4	6	EIF2AK3	5	ATF3
15	CARM1	8	CDK8	6	CTNNA1	5	APOE
14	TGFB1	8	CDH1	6	CTBP1	5	AATF
14	PML	8	CD34	6	CEBPD	4	ZFYVE9
14	IL6	7	TP73	6	CDKN2A	4	WDR16
14	FOXM1	7	TP63	6	CDK6	4	VCAM1
14	DNMT3B	7	THRB	6	CCND2	4	TEAD1
14	CREB1	7	TCEA1	5	ZFPM1	4	TCL1A
13	UBA52	7	SP3	5	YBX1	4	TAF2E
13	STAT1	7	REL	5	WNT3A	4	STAT5A

CFT collegate	Intermedio	CFT collegate	Intermedio	CFT collegate	Intermedio
4	SREBF1	4	CALCA	3	MPO
4	SOCS1	4	BTRC	3	MGP
4	SH3GL1	4	BRCA2	3	MED23
4	SERPINE1	4	BCR	3	MED21
4	SALL4	4	ATF7IP	3	MED17
4	RHOB	4	ATF6	3	MED15
4	REV3L	4	ATF2	3	MCM8
4	RET	4	ASXL1	3	MAPK7
4	PTPN1	4	ACTA1	3	MAPK13
4	PRPS1	4	ABL1	3	MAFK
4	POU5F1B	4		3	LPL
4	POMC	3	ZRSR2	3	LDB1
4	POLE	3	ZHX2	3	KPNA6
4	POLD1	3	ZBTB7A	3	IRF3
4	POLA1	3	YWHAQ	3	IRF1
4	PHKA1	3	XPO1	3	INTU
4	PDGFRB	3	XBP1	3	IL4
4	NUP98	3	WWTR1	3	IL1B
4	NT5E	3	VSX2	3	IL10
4	NR2C1	3	USP7	3	HTT
4	NPM1	3	UIMC1	3	HSPA5
4	NKX3-1	3	TYR	3	HOXA5
4	NFYA	3	TYK2	3	HOPX
4	NES	3	TRAF6	3	HMOX1
4	NDN	3	TNNT2	3	HMGB2
4	NCAPH	3	TLE2	3	HIST2H2AC
4	NCAM1	3	TGFBI	3	HIST2H2AA4
4	MYST3	3	TFAP2A	3	HIST2H2AA3
4	MYOG	3	TES	3	HIST1H4L
4	MYCN	3	TECR	3	HIST1H3A
4	MORF4	3	TEAD4	3	HDAC7
4	MLLT1	3	TCF7L2	3	HAND2
4	MLL2	3	TBL1X	3	GMPS
4	MEN1	3	TAF4	3	FUS
4	MEIS2	3	STAT5B	3	FSH
4	MED14	3	STAT2	3	FRZB
4	MBD3	3	SON	3	FOXP3
4	LMO4	3	SOCS3	3	FOXA1
4	LIFR	3	SMO	3	FN1
4	KLF5	3	SMARCC1	3	FKBP8
4	KLF4	3	SMARCB1	3	FHL2
4	KDM5B	3	SMARCA2	3	FGFR2
4	JAK3	3	SMAD7	3	FGF10
4	ISL1	3	SLC11A1	3	FCGR2B
4	ISG15	3	SKP1	3	FASLG
4	IRS1	3	SIX3	3	ENSG00000235307
4	IRF8	3	SIX1	3	ENSG00000204301
4	IL8	3	SIM1	3	ENSG00000168477
4	IL2	3	SFPQ	3	ELP2
4	IKBKB	3	RUNX3	3	ELN
4	ICK	3	ROS1	3	ELL
4	HOXB7	3	RORB	3	EDA
4	HNF1A	3	REF1	3	DMRT1
4	HDAC5	3	REN	3	DHX15
4	GNB2L1	3	RBBP7	3	DHCR7
4	GMNN	3	RARS	3	CYBB
4	GFAP	3	RARB	3	CRX
4	GCK	3	RAG2	3	CPS1
4	GCG	3	PTPN11	3	COPS2
4	GBP1	3	PRPS2	3	COLQ
4	FGF4	3	PRKCD	3	COL14A1
4	FGF2	3	PRDM1	3	CHUK
4	FGF1	3	POU3F2	3	CHKA
4	ETV6	3	PIN1	3	CDKN2B
4	ESRRG	3	PHF6	3	CDK5
4	ESRRA	3	PAX4	3	CD5
4	EPO	3	OTX2	3	CD40LG
4	ENSG00000204531	3	OPN3	3	CCT4
4	ELF3	3	NTSR1	3	CBX5
4	EIF2AK2	3	NRL	3	CAT
4	EBF1	3	NRD1	3	CALM1
4	DMAP1	3	NR2F1	3	BHMT
4	DAD1	3	NR1H4	3	BCL3
4	CYP7A1	3	NOTCH4	3	BCL2
4	CYP2B6	3	NOS3	3	ATRX
4	CUL1	3	NIN	3	ARID3A
4	CSNK1D	3	NGF	3	ARID1A
4	CSN1S1	3	NFIC	3	AIMP1
4	COL1A1	3	NFATC2	3	AHR
4	CENPH	3	NAGPA	3	AES
4	CEBPE	3	MTTP		
4	CDX2	3	MTOR		
4	CD36	3	MSX2		
4	CD19	3	MRPS7		
4	CAMK4	3	MRPL49		

LISTA A10: Enrichment delle proteine intermedie 1

Enrichment dei GO-term (cellular component) delle 1322 su 1547 proteine intermedie che possono collegare i FT delle CFT significative e di cui è stato possibile rinvenire un identificativo Uniprot (reviewed). 1309 proteine risultavano annotate nell'ontologia. I p-value sono corretti con il metodo Bonferroni, sono mostrati solo i termini con un p-value inferiore a $1 \cdot 10^{-5}$. A sinistra sono elencati i GO-term, seguiti da una loro descrizione. Il campo "Geni" indica il numero di geni annotati per il termine corrispondente. "Classe" indica il numero totale di geni annotati per lo stesso termine nell'intero database di riferimento (NET-GE, standard enrichment). I termini sono ordinati in ordine di p-value crescente.

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0005654	nucleoplasm	355	1088	3,72E-153	GO:0070161	anchoring junction	70	454	8,51E-08
GO:0005634	nucleus	813	5811	3,39E-131	GO:0016592	mediator complex	17	39	1,30E-07
GO:0044428	nuclear part	517	2616	1,02E-127	GO:0071565	nBAF complex	10	13	3,93E-07
GO:0043231	intracellular membrane-bounded organelle	966	8574	5,33E-105	GO:0000123	histone acetyltransferase complex	23	79	1,20E-06
GO:0043227	membrane-bounded organelle	1059	10081	5,78E-105	GO:0090575	RNA polymerase II transcription factor complex	19	55	1,29E-06
GO:0043229	intracellular organelle	997	9610	1,17E-85	GO:0031248	protein acetyltransferase complex	24	86	1,37E-06
GO:0043226	organelle	1082	11068	6,11E-85	GO:1902493	acetyltransferase complex	24	86	1,37E-06
GO:0044424	intracellular part	1192	13748	1,31E-67	GO:0009986	cell surface	72	505	1,59E-06
GO:0044446	intracellular organelle part	764	6941	6,29E-61	GO:0098552	side of membrane	43	238	3,04E-06
GO:0044422	organelle part	775	7145	2,49E-59	GO:0031988	membrane-bounded vesicle	292	3140	4,00E-06
GO:0005667	transcription factor complex	113	270	1,60E-57	GO:0031982	vesicle	299	3245	6,35E-06
GO:0032991	macromolecular complex	591	4806	2,36E-57	GO:0005730	nucleolus	92	733	7,12E-06
GO:0000785	chromatin	112	283	6,13E-54	GO:0032993	protein-DNA complex	33	161	8,08E-06
GO:0043234	protein complex	509	4050	2,43E-49	GO:0009897	external side of plasma membrane	37	195	9,52E-06
GO:0044427	chromosomal part	162	637	2,49E-48					
GO:0005829	cytosol	373	2631	4,70E-45					
GO:0044451	nucleoplasm part	157	636	6,67E-45					
GO:0000790	nuclear chromatin	79	172	3,25E-43					
GO:0044454	nuclear chromosome part	106	317	1,14E-42					
GO:0005737	cytoplasm	534	4656	1,08E-39					
GO:0044464	cell part	1264	16438	7,43E-38					
GO:1990234	transferase complex	114	514	3,00E-27					
GO:1902494	catalytic complex	138	784	1,15E-22					
GO:0017053	transcriptional repressor complex	35	70	8,91E-20					
GO:0070603	SWI/SNF superfamily-type complex	32	70	1,91E-16					
GO:0035097	histone methyltransferase complex	29	60	1,34E-15					
GO:0000792	heterochromatin	30	68	8,77E-15					
GO:0005615	extracellular space	158	1158	1,04E-14					
GO:0016605	PML body	32	84	1,34E-13					
GO:0034708	methyltransferase complex	30	77	5,69E-13					
GO:0044421	extracellular region part	365	3669	9,54E-13					
GO:0016604	nuclear body	64	314	1,39E-12					
GO:0044444	cytoplasmic part	647	7542	1,16E-11					
GO:0071564	npBAF complex	11	11	1,13E-10					
GO:0000118	histone deacetylase complex	20	42	3,99E-10					
GO:0043228	non-membrane-bounded organelle	251	2423	1,82E-09					
GO:0043232	intracellular non-membrane-bounded organelle	251	2423	1,82E-09					
GO:0016514	SWI/SNF complex	12	15	2,87E-09					
GO:0090544	BAF-type complex	13	19	9,53E-09					
GO:0005912	adherens junction	70	437	1,46E-08					
GO:0005925	focal adhesion	63	373	1,54E-08					
GO:0045121	membrane raft	43	205	2,24E-08					
GO:0005720	nuclear heterochromatin	17	36	2,68E-08					
GO:0005924	cell-substrate adherens junction	63	378	2,75E-08					
GO:0044798	nuclear transcription factor complex	25	79	3,15E-08					
GO:0030055	cell-substrate junction	63	382	4,34E-08					

LISTA A11: Enrichment delle proteine intermedie 2

Enrichment dei GO-term (biological process) delle 1322 su 1547 proteine intermedie che possono collegare i FT delle CFT significative e di cui è stato possibile rinvenire un identificativo Uniprot (reviewed). 1305 proteine risultavano annotate nell'ontologia. I p-value sono corretti con il metodo Bonferroni, sono mostrati solo i termini con un p-value inferiore a $1 \cdot 10^{-112}$. A sinistra sono elencati i GO-term, seguiti da una loro descrizione. Il campo "Geni" indica il numero di geni annotati per il termine corrispondente. "Classe" indica il numero totale di geni annotati per lo stesso termine nell'intero database di riferimento (NET-GE, standard enrichment). I termini sono ordinati in ordine di p-value crescente.

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0010604	positive regulation of macromolecule metabolic process	615	2219	9,32E-240	GO:0031327	negative regulation of cellular biosynthetic process	359	1161	4,32E-142
GO:0006357	regulation of transcription from RNA polymerase II promoter	513	1526	1,20E-235	GO:0051239	regulation of multicellular organismal process	488	2148	1,07E-139
GO:0009893	positive regulation of metabolic process	637	2463	1,34E-231	GO:0051172	negative regulation of nitrogen compound metabolic process	347	1112	4,36E-138
GO:0031325	positive regulation of cellular metabolic process	614	2331	1,30E-225	GO:0045934	negative regulation of nucleobase-containing compound metabolic process	343	1088	7,81E-138
GO:0010628	positive regulation of gene expression	461	1291	1,12E-220	GO:0050793	regulation of developmental process	442	1798	1,43E-137
GO:0009891	positive regulation of biosynthetic process	495	1506	1,25E-220	GO:0070887	cellular response to chemical stimulus	466	1991	1,89E-137
GO:0031328	positive regulation of cellular biosynthetic process	491	1484	5,26E-220	GO:0071310	cellular response to organic substance	426	1695	1,94E-135
GO:0060255	regulation of macromolecule metabolic process	909	5351	5,47E-219	GO:0048869	cellular developmental process	507	2370	1,46E-134
GO:0010557	positive regulation of macromolecule biosynthetic process	471	1386	2,84E-215	GO:0032774	RNA biosynthetic process	531	2621	1,28E-131
GO:0051173	positive regulation of nitrogen compound metabolic process	478	1433	6,54E-215	GO:0009059	macromolecule biosynthetic process	589	3211	9,62E-129
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	471	1405	2,76E-212	GO:0044260	cellular macromolecule metabolic process	883	6635	8,42E-128
GO:0048522	positive regulation of cellular process	766	3867	6,03E-211	GO:0050794	regulation of cellular process	1114	10220	1,25E-126
GO:1902680	positive regulation of RNA biosynthetic process	440	1224	2,03E-210	GO:0050789	regulation of biological process	1141	10755	4,31E-125
GO:0051254	positive regulation of RNA metabolic process	443	1245	6,25E-210	GO:1901362	organic cyclic compound biosynthetic process	572	3105	1,21E-124
GO:0010468	regulation of gene expression	798	4217	1,04E-209	GO:0019438	aromatic compound biosynthetic process	559	2982	1,96E-124
GO:0045893	positive regulation of transcription, DNA-templated	432	1184	4,58E-209	GO:0044271	cellular nitrogen compound biosynthetic process	564	3036	4,83E-124
GO:0009889	regulation of biosynthetic process	790	4148	1,70E-208	GO:0034654	nucleobase-containing compound biosynthetic process	549	2898	1,29E-123
GO:0031326	regulation of cellular biosynthetic process	783	4103	2,69E-206	GO:0000122	negative regulation of transcription from RNA polymerase II promoter	249	625	2,82E-123
GO:2001141	regulation of RNA biosynthetic process	725	3575	2,51E-201	GO:0065007	biological regulation	1163	11240	1,08E-122
GO:0010556	regulation of macromolecule biosynthetic process	762	3950	3,08E-201	GO:0018130	heterocycle biosynthetic process	556	2984	2,60E-122
GO:0051252	regulation of RNA metabolic process	729	3661	1,32E-197	GO:0009719	response to endogenous stimulus	349	1271	8,67E-120
GO:0048518	positive regulation of biological process	808	4492	1,48E-197	GO:0006351	transcription, DNA-templated	494	2444	1,42E-119
GO:0006355	regulation of transcription, DNA-templated	712	3531	1,58E-194	GO:0090304	nucleic acid metabolic process	646	3950	2,61E-119
GO:0048519	negative regulation of biological process	732	3732	2,16E-194	GO:0043170	macromolecule metabolic process	940	7621	3,14E-119
GO:2000112	regulation of cellular macromolecule biosynthetic process	742	3835	2,77E-194	GO:0042221	response to chemical	579	3276	5,47E-118
GO:0080090	regulation of primary metabolic process	911	5785	1,04E-192	GO:0045595	regulation of cell differentiation	347	1299	4,98E-115
GO:0048523	negative regulation of cellular process	697	3422	1,72E-191	GO:0034645	cellular macromolecule biosynthetic process	549	3029	9,13E-115
GO:0019222	regulation of metabolic process	956	6406	9,98E-191	GO:2000026	regulation of multicellular organismal development	353	1355	1,09E-113
GO:0031323	regulation of cellular metabolic process	905	5753	4,79E-190	GO:0009058	biosynthetic process	668	4302	3,92E-113
GO:0019219	regulation of nucleobase-containing compound metabolic process	787	4495	2,53E-181					
GO:0051171	regulation of nitrogen compound metabolic process	795	4591	7,33E-181					
GO:0032502	developmental process	786	4537	8,99E-178					
GO:0044767	single-organism developmental process	748	4137	8,16E-177					
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	344	873	4,21E-174					
GO:0048856	anatomical structure development	574	2687	1,58E-157					
GO:0010033	response to organic substance	530	2307	2,58E-157					
GO:0010629	negative regulation of gene expression	355	1038	2,03E-156					
GO:0045892	negative regulation of transcription, DNA-templated	331	939	3,55E-149					
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	349	1055	4,04E-148					
GO:1902679	negative regulation of RNA biosynthetic process	332	959	6,69E-147					
GO:0009892	negative regulation of metabolic process	452	1786	1,03E-146					
GO:0010605	negative regulation of macromolecule metabolic process	431	1630	2,21E-146					
GO:0010558	negative regulation of macromolecule biosynthetic process	354	1104	1,50E-145					
GO:0031324	negative regulation of cellular metabolic process	431	1655	1,19E-143					
GO:0051253	negative regulation of RNA metabolic process	333	991	7,68E-143					
GO:0009890	negative regulation of biosynthetic process	362	1177	1,36E-142					

LISTA A12: Enrichment delle proteine intermedie 3

Enrichment dei GO-term (cellular component) delle 612 su 684 proteine intermedie che possono collegare i FT delle CFT significative con RNAP II e di cui è stato possibile rinvenire un identificativo Uniprot (reviewed). 608 proteine risultavano annotate nell'ontologia. I p-value sono corretti con il metodo Bonferroni, sono mostrati solo i termini con un p-value inferiore a $1 \cdot 10^{-4}$. A sinistra sono elencati i GO-term, seguiti da una loro descrizione. Il campo "Geni" indica il numero di geni annotati per il termine corrispondente. "Classe" indica il numero totale di geni annotati per lo stesso termine nell'intero database di riferimento (NET-GE, standard enrichment). I termini sono ordinati in ordine di p-value crescente.

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0005654	nucleoplasm	252	1088	3,21E-154	GO:0005689	U12-type spliceosomal complex	10	24	7,35E-07
GO:0044428	nuclear part	339	2616	4,90E-135	GO:0005675	holo TFIIF complex	7	10	2,01E-06
GO:0044451	nucleoplasm part	149	636	7,93E-86	GO:0016592	mediator complex	11	39	1,31E-05
GO:0005634	nucleus	419	5811	2,00E-85	GO:0044452	nucleolar part	11	40	1,75E-05
GO:0032991	macromolecular complex	337	4806	2,54E-56	GO:0017053	transcriptional repressor complex	14	70	1,98E-05
GO:0044446	intracellular organelle part	408	6941	9,71E-53	GO:0035097	histone methyltransferase complex	13	60	2,23E-05
GO:0043231	intracellular membrane-bounded organelle	457	8574	2,74E-51	GO:0005736	DNA-directed RNA polymerase I complex	7	13	2,64E-05
GO:0044422	organelle part	410	7145	4,92E-50	GO:0016363	nuclear matrix	16	95	2,74E-05
GO:0043229	intracellular organelle	480	9610	3,09E-47	GO:1902554	serine/threonine protein kinase complex	12	51	2,79E-05
GO:0005667	transcription factor complex	71	270	2,81E-42	GO:0000791	euchromatin	9	26	3,45E-05
GO:0043227	membrane-bounded organelle	480	10081	7,90E-40					
GO:0043234	protein complex	271	4050	7,25E-37					
GO:0043226	organelle	501	11068	8,76E-37					
GO:0044424	intracellular part	562	13748	2,22E-35					
GO:1990234	transferase complex	86	514	3,82E-35					
GO:0000785	chromatin	59	283	1,05E-28					
GO:1902494	catalytic complex	96	784	4,17E-28					
GO:0016604	nuclear body	59	314	3,93E-26					
GO:0044427	chromosomal part	80	637	1,10E-23					
GO:0071013	catalytic step 2 spliceosome	30	80	6,48E-22					
GO:0005681	spliceosomal complex	39	153	8,56E-22					
GO:0044464	cell part	593	16438	7,18E-21					
GO:0000428	DNA-directed RNA polymerase complex	23	47	1,05E-19					
GO:0055029	nuclear DNA-directed RNA polymerase complex	23	47	1,05E-19					
GO:0005737	cytoplasm	255	4656	1,15E-19					
GO:0030880	RNA polymerase complex	23	48	1,96E-19					
GO:0090575	RNA polymerase II transcription factor complex	24	55	4,04E-19					
GO:0000790	nuclear chromatin	38	172	8,38E-19					
GO:0044798	nuclear transcription factor complex	27	79	2,51E-18					
GO:0044454	nuclear chromosome part	49	317	1,49E-17					
GO:0030529	ribonucleoprotein complex	73	688	4,62E-17					
GO:0043228	non-membrane-bounded organelle	149	2423	2,14E-13					
GO:0043232	intracellular non-membrane-bounded organelle	149	2423	2,14E-13					
GO:0008023	transcription elongation factor complex	17	40	7,07E-13					
GO:0005666	DNA-directed RNA polymerase III complex	12	17	2,92E-12					
GO:0016605	PML body	22	84	6,46E-12					
GO:0030532	small nuclear ribonucleoprotein complex	16	40	1,57E-11					
GO:0016607	nuclear speck	29	174	1,29E-10					
GO:0032806	carboxy-terminal domain protein kinase complex	10	13	1,47E-10					
GO:0034708	methyltransferase complex	19	77	1,36E-09					
GO:0005665	DNA-directed RNA polymerase II, core complex	10	16	3,78E-09					
GO:0005730	nucleolus	60	733	1,04E-08					
GO:1902911	protein kinase complex	17	69	2,02E-08					
GO:0005669	transcription factor TFIID complex	10	21	1,44E-07					
GO:0005829	cytosol	139	2631	2,48E-07					
GO:0005682	U5 snRNP	6	6	5,79E-07					

LISTA A13: Enrichment delle proteine intermedie 4

Enrichment dei GO-term (biological process) delle 612 su 684 proteine intermedie che possono collegare i FT delle CFT significative con RNAP II e di cui è stato possibile rinvenire un identificativo Uniprot (reviewed). 604 proteine risultavano annotate nell'ontologia. I p-value sono corretti con il metodo Bonferroni, sono mostrati solo i termini con un p-value inferiore a $1 \cdot 10^{-38}$. A sinistra sono elencati i GO-term, seguiti da una loro descrizione. Il campo "Geni" indica il numero di geni annotati per il termine corrispondente. "Classe" indica il numero totale di geni annotati per lo stesso termine nell'intero database di riferimento (NET-GE, standard enrichment). I termini sono ordinati in ordine di p-value crescente.

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0090304	nucleic acid metabolic process	391	3950	1,48E-118	GO:0006366	transcription from RNA polymerase II promoter	130	479	1,44E-80
GO:0006139	nucleobase-containing compound metabolic process	427	4834	1,51E-117	GO:0019222	regulation of metabolic process	437	6406	1,35E-79
GO:0046483	heterocycle metabolic process	431	5027	1,85E-114	GO:0045893	positive regulation of transcription, DNA-templated	190	1184	4,95E-79
GO:0006725	cellular aromatic compound metabolic process	431	5040	5,01E-114	GO:0080090	regulation of primary metabolic process	413	5785	4,31E-78
GO:1901360	organic cyclic compound metabolic process	439	5271	2,50E-113	GO:0044237	cellular metabolic process	516	9122	2,41E-77
GO:0016070	RNA metabolic process	357	3377	3,43E-112	GO:0048522	positive regulation of cellular process	333	3867	7,64E-76
GO:0034654	nucleobase-containing compound biosynthetic process	330	2898	1,13E-109	GO:0044238	primary metabolic process	521	9443	1,23E-74
GO:0034641	cellular nitrogen compound metabolic process	433	5240	3,25E-109	GO:0045944	positive regulation of transcription from RNA polymerase II promoter	161	873	1,93E-74
GO:0006357	regulation of transcription from RNA polymerase II promoter	247	1526	4,56E-109	GO:0048518	positive regulation of biological process	354	4492	6,31E-72
GO:0019438	aromatic compound biosynthetic process	331	2982	8,15E-107	GO:0071704	organic substance metabolic process	524	9717	1,95E-71
GO:0018130	heterocycle biosynthetic process	331	2984	9,95E-107	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	74	180	1,03E-58
GO:0044271	cellular nitrogen compound biosynthetic process	333	3036	3,56E-106	GO:0000398	mRNA splicing, via spliceosome	74	180	1,03E-58
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	235	1405	7,20E-106	GO:0000375	RNA splicing, via transesterification reactions	74	187	3,09E-57
GO:0051173	positive regulation of nitrogen compound metabolic process	237	1433	7,68E-106	GO:0006354	DNA-templated transcription, elongation	56	93	7,36E-56
GO:1901362	organic cyclic compound biosynthetic process	335	3105	6,31E-105	GO:0051172	negative regulation of nitrogen compound metabolic process	155	1112	8,98E-54
GO:0006807	nitrogen compound metabolic process	444	5690	1,61E-104	GO:0045934	negative regulation of nucleobase-containing compound metabolic process	153	1088	1,72E-53
GO:0009891	positive regulation of biosynthetic process	239	1506	8,11E-103	GO:0051253	negative regulation of RNA metabolic process	146	991	2,58E-53
GO:0031328	positive regulation of cellular biosynthetic process	237	1484	2,17E-102	GO:0010629	negative regulation of gene expression	149	1038	4,61E-53
GO:1902680	positive regulation of RNA biosynthetic process	217	1224	1,58E-101	GO:0008152	metabolic process	531	11004	1,12E-52
GO:0010468	regulation of gene expression	382	4217	3,21E-101	GO:0010558	negative regulation of macromolecule biosynthetic process	153	1104	1,26E-52
GO:0051254	positive regulation of RNA metabolic process	218	1245	6,14E-101	GO:0010605	negative regulation of macromolecule metabolic process	186	1630	1,76E-52
GO:0044260	cellular macromolecule metabolic process	472	6635	6,99E-101	GO:1902679	negative regulation of RNA biosynthetic process	142	959	6,42E-52
GO:0032774	RNA biosynthetic process	304	2621	4,18E-100	GO:0045892	negative regulation of transcription, DNA-templated	140	939	1,85E-51
GO:0051252	regulation of RNA metabolic process	355	3661	2,10E-99	GO:2000113	negative regulation of cellular macromolecule biosynthetic process	148	1055	2,49E-51
GO:0010604	positive regulation of macromolecule metabolic process	280	2219	6,58E-99	GO:0009890	negative regulation of biosynthetic process	156	1177	3,71E-51
GO:0009889	regulation of biosynthetic process	375	4148	5,22E-98	GO:0009892	negative regulation of metabolic process	192	1786	2,22E-50
GO:0031326	regulation of cellular biosynthetic process	373	4103	5,69E-98	GO:0006368	transcription elongation from RNA polymerase II promoter	47	69	3,08E-50
GO:0009893	positive regulation of metabolic process	292	2463	2,57E-97	GO:0031327	negative regulation of cellular biosynthetic process	153	1161	1,13E-49
GO:0031325	positive regulation of cellular metabolic process	284	2331	7,92E-97	GO:0048519	negative regulation of biological process	286	3732	1,76E-49
GO:0010557	positive regulation of macromolecule biosynthetic process	224	1386	9,64E-97	GO:0048523	negative regulation of cellular process	272	3422	2,62E-49
GO:2001141	regulation of RNA biosynthetic process	347	3575	3,32E-96	GO:0006397	mRNA processing	93	405	6,33E-49
GO:0010556	regulation of macromolecule biosynthetic process	362	3950	1,31E-94	GO:0031324	negative regulation of cellular metabolic process	182	1655	1,15E-48
GO:0009059	macromolecule biosynthetic process	324	3211	9,13E-92	GO:0000122	negative regulation of transcription from RNA polymerase II promoter	111	625	2,89E-47
GO:0010467	gene expression	160	672	1,05E-91	GO:0008380	RNA splicing	81	320	1,46E-45
GO:2000112	regulation of cellular macromolecule biosynthetic process	353	3835	1,31E-91	GO:0016071	mRNA metabolic process	107	613	1,32E-44
GO:0060255	regulation of macromolecule metabolic process	415	5351	2,21E-91	GO:0006396	RNA processing	112	706	1,24E-42
GO:0019219	regulation of nucleobase-containing compound metabolic process	379	4495	9,66E-90	GO:0044767	single-organism developmental process	289	4137	1,34E-41
GO:0006355	regulation of transcription, DNA-templated	336	3531	3,08E-89	GO:0006352	DNA-templated transcription, initiation	66	221	1,83E-41
GO:0051171	regulation of nitrogen compound metabolic process	382	4591	5,09E-89	GO:0050793	regulation of developmental process	179	1798	2,16E-41
GO:0043170	macromolecule metabolic process	488	7621	9,40E-89	GO:0050434	positive regulation of viral transcription	38	54	6,21E-41
GO:0006351	transcription, DNA-templated	280	2444	2,37E-88	GO:0006367	transcription initiation from RNA polymerase II promoter	61	186	8,41E-41
GO:0031323	regulation of cellular metabolic process	422	5753	1,87E-85	GO:0048856	anatomical structure development	221	2687	1,58E-39
GO:0034645	cellular macromolecule biosynthetic process	306	3029	8,55E-85	GO:0045595	regulation of cell differentiation	147	1299	3,84E-39
GO:1901576	organic substance biosynthetic process	359	4218	1,54E-83	GO:0048513	organ development	138	1153	4,07E-39
GO:0044249	cellular biosynthetic process	351	4094	1,35E-81	GO:0006353	DNA-templated transcription, termination	44	86	6,83E-39
GO:0010628	positive regulation of gene expression	200	1291	4,47E-81					
GO:0009058	biosynthetic process	359	4302	5,93E-81					

LISTA A14: Nodi in ordine di grado

Elenco di tutti i nodi del grafo composto dalle CFT significative. Ogni nodo rappresenta un FT e sono disposti in ordine di grado decrescente. L'asterisco che precede alcuni nomi indica che tali FT oltre a essere presenti tra le CFT significative sono presenti anche all'interno dei bicluster significativi.

Nodo / FT	Grado	Nodo / FT	Grado	Nodo / FT	Grado
* ZNF148	16	GATA3	3	MAF	1
* SP4	15	TCF3	3	TWIST1	1
MYOG	12	ONECUT1	3	BHLHE41	1
* ZBTB7B	12	HSF1	3	FOXF1	1
* SP1	12	IRF2	3	FOXA3	1
* SP3	11	STAT2	3	ATF3	1
PAX5	10	ZBTB6	3	POU4F2	1
* ESR2	10	ATF6	3	USF1	1
* PPARA	10	E2F6	3	GATA1	1
* ZFX	10	STAT1	3	POU1F1	1
FOXC2	10	NR3C2	3	FOS	1
* WT1	10	STAT5A	3	MEF2D	1
TP73	10	HOXD10	3	ZNF423	1
RREB1	10	NR3C1	3	ATF5	1
ZIC2	9	* NFE2L1	3	MEF2C	1
THRA	9	NFE2L2	3	SMAD1	1
* PAX2	9	* MYF6	3	HLTF	1
SOX5	9	* SOX13	3	TLX1	1
RARB	9	TEF	3	ZNF219	1
RARG	9	ESRRB	3	ETV7	1
ZBTB4	9	RBPJ	3	NKX2-2	1
HOXD9	9	EPAS1	3	PPARG	1
NFIC	9	ARNTL	3	OTX2	1
* ETV4	8	POU2F2	2	ONECUT2	1
SREBF1	8	DLX2	2	FOXD1	1
SREBF2	8	NFATC1	2	LHX3	1
PBX3	7	NKX3-1	2	JUND	1
NFYA	7	IRF7	2	IRF1	1
* PLAG1	7	HOXB6	2	FOXJ3	1
* PURA	7	MECOM	2	HES1	1
NR2C1	7	ELF2	2	TAL1	1
NR4A3	7	POU3F2	2	EBF1	1
ZIC1	6	ZBTB18	2	KLF1	1
TFAP2A	6	NR5A1	2	TFAP4	1
ARID5B	6	DDIT3	2	SMAD4	1
BRCA1	6	FLI1	2	ATF1	1
IRF4	6	TBX2	2	IRF5	1
BCL6	6	TCF7L2	2	FOXI1	1
ENO1	6	STAT3	2	SMAD2	1
FOXO3	6	NR1I2	2	FOXM1	1
CEBPA	6	STAT4	2		
* MAZ	6	NR2F1	2		
SPZ1	6	ELK1	2		
CTCF	6	* CXXC1	2		
ZNF350	5	ARNT2	2		
JUN	5	ARID3A	2		
RXRA	5	TEAD1	2		
NRF1	5	HSF2	2		
THRB	5	EGR4	2		
NHLH1	5	HOXB1	2		
SP2	5	SMARCC1	2		
NR6A1	5	MZF1	2		
OVOL1	5	ZNF143	2		
NR2C2	5	SOX17	2		
NFYB	5	* SNAI1	2		
SRF	5	ZBTB7A	2		
EGR2	5	PAX8	2		
NR1H2	5	* NANOG	2		
RFX3	5	BARX2	2		
ELF3	4	SPIB	2		
NOBOX	4	AHR	2		
EGR1	4	FOXO1	2		
REST	4	IRF9	2		
PRDM1	4	IRF3	2		
TFCP2	4	TFAP2B	2		
MBD2	4	RXRG	1		
TCF12	4	POU3F1	1		
ZNF589	4	FOXJ2	1		
MAFK	4	NFE2	1		
FOXO4	4	HEY2	1		
NFYC	4	MEF2A	1		
* KLF15	4	TP53	1		
SPI1	3	NR1D1	1		
NR2E3	3	HNF1B	1		
LHX2	3	GLI2	1		
SOX2	3	HOXA10	1		
NR2F2	3	AIRE	1		
NR0B1	3	* SOX15	1		

LISTA A15: Enrichment dei geni delle CFT significative

Enrichment dei GO-term (biological process) dei geni dei promotori inclusi nel set di CEP associati a CFT significative. Sono stati utilizzati come input tutti i geni per i quali è stato possibile rinvenire un identificativo Uniprot, vale a dire 257 elementi. I p-value sono corretti con il metodo Bonferroni, il livello di significatività di soglia usato è stato pari a un p-value di $5 \cdot 10^{-2}$. Sono stati selezionati solo i termini riguardanti un numero di geni superiore o uguale a 20. A sinistra sono elencati i GO-term, seguiti da una loro descrizione. Poi il campo "Geni" indica il numero di geni annotati per il termine corrispondente. "Classe" indica il numero totale di geni annotati per lo stesso termine nell'intero database di riferimento (NET-GE). I termini evidenziati in grassetto sottolineano il coinvolgimento di questi geni nello sviluppo embrionale di vari tessuti e nella regolazione della trascrizione. Tutti i termini sono ordinati in ordine di numero di geni arricchiti decrescente.

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0009987	cellular process	207	26673	3,64E-10	GO:0019438	aromatic compound biosynthetic process	94	3919	1,82E-30
GO:0065007	biological regulation	182	19296	6,43E-16	GO:0044271	cellular nitrogen compound biosynthetic process	94	3991	7,75E-30
GO:0044699	single-organism process	180	20881	2,00E-10	GO:1901362	organic cyclic compound biosynthetic process	94	4099	6,42E-29
GO:0050789	regulation of biological process	175	18486	2,01E-14	GO:0018130	heterocycle biosynthetic process	93	3931	1,42E-29
GO:0050794	regulation of cellular process	170	17465	8,82E-15	GO:0051172	negative regulation of nitrogen compound metabolic process	93	3798	1,42E-27
GO:0044763	single-organism cellular process	153	16559	3,59E-09	GO:0034654	nucleobase-containing compound biosynthetic process	92	3760	2,68E-30
GO:0071704	organic substance metabolic process	147	17115	1,38E-05	GO:0034645	cellular macromolecule biosynthetic process	92	4013	4,22E-28
GO:0044237	cellular metabolic process	144	15758	1,72E-07	GO:0032774	RNA biosynthetic process	91	3250	1,83E-34
GO:0044238	primary metabolic process	144	16587	1,20E-05	GO:0045934	negative regulation of nucleobase-containing compound metabolic process	91	3691	5,21E-27
GO:0019222	regulation of metabolic process	141	10848	7,51E-22	GO:0006351	transcription, DNA-templated	89	2949	4,72E-36
GO:0080090	regulation of primary metabolic process	137	9808	4,98E-24	GO:0010558	negative regulation of macromolecule biosynthetic process	89	3771	7,94E-25
GO:0031323	regulation of cellular metabolic process	135	9728	3,88E-23	GO:0051094	positive regulation of developmental process	87	3200	1,48E-28
GO:0043170	macromolecule metabolic process	132	12986	1,02E-09	GO:0010629	negative regulation of gene expression	87	3561	3,66E-25
GO:0060255	regulation of macromolecule metabolic process	131	8942	2,30E-24	GO:2000113	negative regulation of cellular macromolecule biosynthetic process	87	3633	1,55E-24
GO:0051171	regulation of nitrogen compound metabolic process	126	7927	2,32E-26	GO:0048585	negative regulation of response to stimulus	85	3759	4,58E-22
GO:0044260	cellular macromolecule metabolic process	126	11069	7,71E-13	GO:0045944	positive regulation of transcription from RNA polymerase II promoter	84	2989	2,57E-28
GO:0019219	regulation of nucleobase-containing compound metabolic process	124	7797	9,88E-26	GO:1902679	negative regulation of RNA biosynthetic process	84	3295	2,59E-25
GO:0010468	regulation of gene expression	121	7158	2,54E-27	GO:0048869	cellular developmental process	84	3694	1,04E-24
GO:0032502	developmental process	121	7299	1,68E-26	GO:0051253	negative regulation of RNA metabolic process	84	3369	1,22E-24
GO:0031326	regulation of cellular biosynthetic process	120	6987	1,20E-27	GO:0048856	anatomical structure development	84	4289	2,73E-20
GO:0009889	regulation of biosynthetic process	120	7051	2,90E-27	GO:0045892	negative regulation of transcription, DNA-templated	83	3247	5,30E-25
GO:2000112	regulation of cellular macromolecule biosynthetic process	117	6579	4,80E-28	GO:0048731	system development	81	2612	5,20E-30
GO:0010556	regulation of macromolecule biosynthetic process	117	6784	8,71E-27	GO:0048646	anatomical structure formation involved in morphogenesis	79	2968	1,31E-24
GO:0051252	regulation of RNA metabolic process	116	6360	1,00E-28	GO:0009893	positive regulation of metabolic process	78	3630	6,91E-21
GO:0044767	single-organism developmental process	116	6740	2,31E-26	GO:0010648	negative regulation of cell communication	78	3347	1,68E-20
GO:0006355	regulation of transcription, DNA-templated	114	6146	1,10E-28	GO:0023057	negative regulation of signaling	77	3347	8,30E-20
GO:2001141	regulation of RNA biosynthetic process	114	6249	5,15E-28	GO:0060284	regulation of cell development	75	2979	1,61E-21
GO:0006807	nitrogen compound metabolic process	113	8714	6,96E-15	GO:0009968	negative regulation of signal transduction	75	3204	1,38E-19
GO:0048518	positive regulation of biological process	110	6624	4,95E-23	GO:0045597	positive regulation of cell differentiation	74	2514	2,32E-25
GO:1901360	organic cyclic compound metabolic process	109	8010	1,19E-15	GO:0051960	regulation of nervous system development	74	2584	1,31E-24
GO:0006725	cellular aromatic compound metabolic process	108	7662	1,30E-16	GO:0010604	positive regulation of macromolecule metabolic process	74	3285	9,74E-21
GO:0034641	cellular nitrogen compound metabolic process	108	7912	1,58E-15	GO:0032501	multicellular organismal process	74	4447	3,64E-13
GO:1901576	organic substance biosynthetic process	106	6037	7,96E-24	GO:0010647	positive regulation of cell communication	73	3801	5,09E-14
GO:0009058	biosynthetic process	106	6217	9,27E-23	GO:0044707	single-multicellular organism process	73	4361	4,66E-13
GO:0046483	heterocycle metabolic process	106	7642	1,33E-15	GO:0048468	cell development	72	2280	2,16E-26
GO:0048522	positive regulation of cellular process	104	5768	3,92E-24	GO:0000122	negative regulation of transcription from RNA polymerase II promoter	72	2314	5,41E-26
GO:0006139	nucleobase-containing compound metabolic process	104	7311	5,72E-16	GO:0008284	positive regulation of cell proliferation	72	2808	6,91E-21
GO:0009653	anatomical structure morphogenesis	103	4591	2,25E-28	GO:0031325	positive regulation of cellular metabolic process	72	3418	2,49E-18
GO:0044249	cellular biosynthetic process	103	5837	4,85E-23	GO:0009967	positive regulation of signal transduction	71	3621	5,65E-14
GO:0090304	nucleic acid metabolic process	100	5696	6,43E-22	GO:0006357	regulation of transcription from RNA polymerase II promoter	70	2379	3,91E-26
GO:0048519	negative regulation of biological process	99	5756	6,41E-21	GO:0044702	single organism reproductive process	69	3664	1,55E-12
GO:0009059	macromolecule biosynthetic process	98	4344	5,66E-30	GO:0050767	regulation of neurogenesis	68	2377	4,20E-22
GO:0016070	RNA metabolic process	95	4467	9,82E-27	GO:0040011	locomotion	68	3142	2,09E-15
GO:0048523	negative regulation of cellular process	95	5279	3,48E-21	GO:0009888	tissue development	67	2570	1,91E-19

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0051173	positive regulation of nitrogen compound metabolic process	65	2108	5,12E-25	GO:0010646	regulation of cell communication	50	3714	1,54E-04
GO:0030154	cell differentiation	65	2446	2,02E-21	GO:0040007	growth	49	1496	4,16E-17
GO:0051093	negative regulation of developmental process	65	2690	5,72E-17	GO:0006629	lipid metabolic process	49	3598	3,17E-03
GO:0014070	response to organic cyclic compound	65	2783	3,34E-16	GO:0048729	tissue morphogenesis	48	1359	4,92E-18
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	64	2060	9,84E-25	GO:0040008	regulation of growth	48	2308	4,58E-09
GO:0008283	cell proliferation	64	2526	1,07E-17	GO:0009966	regulation of signal transduction	48	3261	1,93E-05
GO:0016477	cell migration	64	2574	2,84E-17	GO:0080134	regulation of response to stress	48	3072	6,97E-05
GO:0048870	cell motility	64	2808	2,43E-15	GO:0001932	regulation of protein phosphorylation	48	3361	1,07E-03
GO:0022603	regulation of anatomical structure morphogenesis	64	2832	3,73E-15	GO:0042592	homeostatic process	48	3571	6,13E-03
GO:0051239	regulation of multicellular organismal process	64	3432	4,08E-13	GO:0030182	neuron differentiation	47	911	1,63E-24
GO:0010628	positive regulation of gene expression	63	1919	1,38E-25	GO:0007275	multicellular organismal development	47	1673	1,24E-13
GO:0009891	positive regulation of biosynthetic process	63	2208	2,93E-22	GO:0045595	regulation of cell differentiation	47	2111	1,96E-11
GO:0060548	negative regulation of cell death	63	2642	5,22E-16	GO:0023052	signaling	47	2287	1,26E-08
GO:0007389	pattern specification process	62	1481	1,23E-28	GO:0044700	single organism signaling	47	2287	1,26E-08
GO:0031328	positive regulation of cellular biosynthetic process	62	2177	8,72E-22	GO:0016265	death	47	2848	1,87E-05
GO:0009628	response to abiotic stimulus	62	3444	6,26E-10	GO:0045184	establishment of protein localization	47	3728	4,51E-02
GO:0051254	positive regulation of RNA metabolic process	61	1838	7,22E-25	GO:0007267	cell-cell signaling	46	2185	9,81E-09
GO:0010557	positive regulation of macromolecule biosynthetic process	61	2042	1,93E-22	GO:1901698	response to nitrogen compound	46	2693	9,68E-06
GO:0009887	organ morphogenesis	60	1719	2,45E-23	GO:0042327	positive regulation of phosphorylation	46	2706	1,13E-05
GO:0045664	regulation of neuron differentiation	60	1894	3,74E-21	GO:0008219	cell death	46	2827	4,36E-05
GO:0048598	embryonic morphogenesis	59	1566	1,37E-24	GO:0032270	positive regulation of cellular protein metabolic process	46	3279	3,31E-03
GO:0045893	positive regulation of transcription, DNA-templated	59	1762	4,27E-24	GO:0065009	regulation of molecular function	46	3941	3,00E-02
GO:1902680	positive regulation of RNA biosynthetic process	59	1811	1,77E-23	GO:0022604	regulation of cell morphogenesis	45	1615	8,92E-13
GO:0045596	negative regulation of cell differentiation	59	2226	6,80E-17	GO:0042127	regulation of cell proliferation	45	2008	6,55E-11
GO:0008285	negative regulation of cell proliferation	59	2354	9,94E-16	GO:0070848	response to growth factor	45	1987	1,47E-09
GO:0071495	cellular response to endogenous stimulus	59	3038	1,14E-10	GO:0007610	behavior	45	1993	1,64E-09
GO:0010605	negative regulation of macromolecule metabolic process	58	2452	3,42E-16	GO:0010033	response to organic substance	45	3487	2,74E-03
GO:0003006	developmental process involved in reproduction	58	2335	3,31E-15	GO:0032989	cellular component morphogenesis	44	1879	9,10E-10
GO:0009892	negative regulation of metabolic process	58	2679	2,10E-14	GO:0010243	response to organonitrogen compound	44	2509	1,05E-05
GO:0050793	regulation of developmental process	58	2884	5,95E-13	GO:0031401	positive regulation of protein modification process	44	2947	1,12E-03
GO:0050877	neurological system process	58	2735	4,38E-12	GO:0048513	organ development	43	1910	2,36E-10
GO:0051726	regulation of cell cycle	58	2760	6,55E-12	GO:0007167	enzyme linked receptor protein signaling pathway	43	2504	3,06E-05
GO:0009725	response to hormone	58	2824	1,79E-11	GO:0030030	cell projection organization	43	2507	3,17E-05
GO:0065003	macromolecular complex assembly	58	3681	1,02E-06	GO:0030855	epithelial cell differentiation	42	1135	4,45E-16
GO:0043069	negative regulation of programmed cell death	57	2511	4,17E-13	GO:0002009	morphogenesis of an epithelium	42	1159	9,55E-16
GO:0033993	response to lipid	57	2604	2,08E-12	GO:0031344	regulation of cell projection organization	42	1640	1,83E-10
GO:0051130	positive regulation of cellular component organization	57	2782	3,67E-11	GO:0071363	cellular response to growth factor stimulus	42	1903	2,36E-08
GO:0031324	negative regulation of cellular metabolic process	56	2479	1,36E-14	GO:0040012	regulation of locomotion	42	2224	2,93E-06
GO:0003002	regionalization	55	988	4,21E-31	GO:0033043	regulation of organelle organization	42	2759	1,32E-03
GO:0043066	negative regulation of apoptotic process	55	2487	5,05E-12	GO:0007155	cell adhesion	42	2817	2,29E-03
GO:0007154	cell communication	55	2752	3,43E-10	GO:0022610	biological adhesion	42	2824	2,44E-03
GO:0044092	negative regulation of molecular function	54	3258	8,16E-07	GO:0007411	axon guidance	41	1242	7,23E-14
GO:0060341	regulation of cellular localization	54	3341	2,04E-06	GO:0097485	neuron projection guidance	41	1243	7,43E-14
GO:0048583	regulation of response to stimulus	54	4515	1,87E-03	GO:0051050	positive regulation of transport	41	2507	2,79E-04
GO:0006366	transcription from RNA polymerase II promoter	53	1448	4,83E-21	GO:0007399	nervous system development	40	1018	4,54E-16
GO:0031327	negative regulation of cellular biosynthetic process	52	1764	3,06E-18	GO:0010769	regulation of cell morphogenesis involved in differentiation	40	1115	1,09E-14
GO:0009890	negative regulation of biosynthetic process	52	1786	5,30E-18	GO:0048878	chemical homeostasis	40	2410	2,80E-04
GO:0051240	positive regulation of multicellular organismal process	52	2185	1,88E-12	GO:1902533	positive regulation of intracellular signal transduction	40	2418	3,06E-04
GO:1901701	cellular response to oxygen-containing compound	52	2540	7,66E-10	GO:0048545	response to steroid hormone	39	1516	1,47E-09
GO:2000026	regulation of multicellular organismal development	51	2229	3,53E-13	GO:0044087	regulation of cellular component biogenesis	39	1812	3,18E-07
GO:0010562	positive regulation of phosphorus metabolic process	51	3001	1,25E-06	GO:0051046	regulation of secretion	39	2094	1,93E-05
GO:0045937	positive regulation of phosphate metabolic process	51	3001	1,25E-06	GO:0051270	regulation of cellular component movement	39	2182	5,92E-05
GO:0006461	protein complex assembly	51	3196	1,09E-05	GO:0043086	negative regulation of catalytic activity	39	2637	7,57E-03
GO:0051247	positive regulation of protein metabolic process	51	3557	3,60E-04	GO:0010720	positive regulation of cell development	38	977	5,83E-15
GO:0023051	regulation of signaling	50	3708	1,47E-04	GO:0010975	regulation of neuron projection development	38	1281	3,80E-11

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0048609	multicellular organismal reproductive process	38	2109	7,55E-05	GO:0001525	angiogenesis	28	949	1,93E-07
GO:0012501	programmed cell death	38	2247	3,90E-04	GO:0010035	response to inorganic substance	28	1420	1,16E-03
GO:0001934	positive regulation of protein phosphorylation	38	2371	1,51E-03	GO:0055082	cellular chemical homeostasis	28	1433	1,39E-03
GO:0048732	gland development	37	1014	1,36E-13	GO:0051241	negative regulation of multicellular organismal process	28	1513	4,05E-03
GO:0006915	apoptotic process	37	2154	4,00E-04	GO:0006325	chromatin organization	28	1572	8,47E-03
GO:0045165	cell fate commitment	36	501	7,57E-23	GO:0010564	regulation of cell cycle process	28	1606	1,27E-02
GO:0060429	epithelium development	36	1006	7,00E-13	GO:2001233	regulation of apoptotic signaling pathway	27	1313	8,46E-04
GO:0032870	cellular response to hormone stimulus	36	1636	1,03E-06	GO:0009953	dorsal/ventral pattern formation	26	264	3,45E-19
GO:0030334	regulation of cell migration	36	1926	7,50E-05	GO:0048562	embryonic organ morphogenesis	26	484	1,23E-12
GO:2000145	regulation of cell motility	36	2039	3,12E-04	GO:0048754	branching morphogenesis of an epithelial tube	26	542	1,80E-11
GO:0032880	regulation of protein localization	36	2200	1,96E-03	GO:0048638	regulation of developmental growth	26	605	2,34E-10
GO:0002684	positive regulation of immune system process	36	2287	4,85E-03	GO:0061138	morphogenesis of a branching epithelium	26	624	4,76E-10
GO:1900542	regulation of purine nucleotide metabolic process	36	2485	3,15E-02	GO:0000902	cell morphogenesis	26	800	1,24E-07
GO:0006140	regulation of nucleotide metabolic process	36	2500	3,59E-02	GO:0030198	extracellular matrix organization	26	1111	1,18E-04
GO:0051129	negative regulation of cellular component organization	35	1913	2,05E-04	GO:0043062	extracellular structure organization	26	1116	1,29E-04
GO:0051248	negative regulation of protein metabolic process	35	2189	4,89E-03	GO:0051249	regulation of lymphocyte activation	26	1264	1,43E-03
GO:0035295	tube development	34	664	1,03E-16	GO:0042493	response to drug	26	1473	2,35E-02
GO:0071396	cellular response to lipid	34	1198	3,90E-09	GO:0001822	kidney development	25	489	1,47E-11
GO:0010942	positive regulation of cell death	34	1542	3,28E-06	GO:0001764	neuron migration	25	578	6,34E-10
GO:0007600	sensory perception	34	1675	2,64E-05	GO:0051098	regulation of binding	25	826	1,33E-06
GO:0001775	cell activation	34	2070	3,90E-03	GO:0071407	cellular response to organic cyclic compound	25	1148	8,56E-04
GO:0009991	response to extracellular stimulus	33	1313	2,16E-07	GO:0046649	lymphocyte activation	25	1196	1,82E-03
GO:0043065	positive regulation of apoptotic process	33	1406	1,25E-06	GO:0010638	positive regulation of organelle organization	25	1260	4,68E-03
GO:0043068	positive regulation of programmed cell death	33	1416	1,50E-06	GO:0016568	chromatin modification	25	1339	1,37E-02
GO:0010941	regulation of cell death	33	2079	1,26E-03	GO:0032386	regulation of intracellular transport	25	1443	4,92E-02
GO:0034097	response to cytokine	33	2017	6,24E-03	GO:0021953	central nervous system neuron differentiation	24	365	1,96E-13
GO:0032101	regulation of response to external stimulus	33	2173	3,01E-02	GO:0045787	positive regulation of cell cycle	24	401	1,64E-12
GO:0001501	skeletal system development	32	620	1,16E-15	GO:0009792	embryo development ending in birth or egg hatching	24	937	7,72E-05
GO:0048608	reproductive structure development	32	1062	4,05E-09	GO:0050867	positive regulation of cell activation	24	996	2,41E-04
GO:0032269	negative regulation of cellular protein metabolic process	32	1804	1,62E-03	GO:0032990	cell part morphogenesis	24	1056	7,04E-04
GO:0070201	regulation of establishment of protein localization	32	1849	2,77E-03	GO:0097305	response to alcohol	24	1084	1,13E-03
GO:0009952	anterior/posterior pattern specification	31	585	2,07E-15	GO:0001101	response to acid chemical	24	1170	4,35E-03
GO:0050678	regulation of epithelial cell proliferation	31	1020	7,60E-09	GO:0030155	regulation of cell adhesion	24	1217	8,59E-03
GO:0009790	embryo development	31	1155	1,75E-07	GO:0051493	regulation of cytoskeleton organization	24	1336	4,11E-02
GO:0031667	response to nutrient levels	31	1219	6,60E-07	GO:0035107	appendage morphogenesis	23	348	8,10E-13
GO:0050865	regulation of cell activation	31	1568	2,37E-04	GO:0035108	limb morphogenesis	23	348	8,10E-13
GO:0006928	cellular component movement	31	1973	3,64E-03	GO:0007423	sensory organ development	23	423	5,34E-11
GO:0019725	cellular homeostasis	31	1855	8,67E-03	GO:0016055	Wnt signaling pathway	23	814	2,56E-05
GO:0046903	secretion	31	1910	1,57E-02	GO:0001701	in utero embryonic development	23	893	1,40E-04
GO:0048666	neuron development	30	539	2,05E-15	GO:0043009	chordate embryonic development	23	909	1,93E-04
GO:0051090	regulation of sequence-specific DNA binding transcription factor activity	30	1091	2,07E-07	GO:0002696	positive regulation of leukocyte activation	23	941	3,58E-04
GO:0001558	regulation of cell growth	30	1390	5,80E-05	GO:0048858	cell projection morphogenesis	23	963	5,40E-04
GO:0002694	regulation of leukocyte activation	30	1443	1,33E-04	GO:0048568	embryonic organ development	22	476	5,29E-09
GO:0044708	single-organism behavior	30	1465	1,85E-04	GO:0010721	negative regulation of cell development	22	609	6,16E-07
GO:0051223	regulation of protein transport	30	1645	2,18E-03	GO:0021700	developmental maturation	22	666	3,27E-06
GO:0009719	response to endogenous stimulus	30	2012	1,56E-02	GO:0048534	hematopoietic or lymphoid organ development	22	682	5,07E-06
GO:0050769	positive regulation of neurogenesis	29	660	4,21E-12	GO:0002521	leukocyte differentiation	22	929	1,18E-03
GO:0007346	regulation of mitotic cell cycle	29	1240	1,78E-05	GO:0022412	cellular process involved in reproduction in multicellular organism	22	998	3,88E-03
GO:0044057	regulation of system process	29	1316	6,57E-05	GO:0001708	cell fate specification	21	264	4,17E-13
GO:0045321	leukocyte activation	29	1486	8,64E-04	GO:0007417	central nervous system development	21	521	2,23E-07
GO:0043067	regulation of programmed cell death	29	1982	3,26E-02	GO:0000904	cell morphogenesis involved in differentiation	21	542	4,61E-07
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	29	1794	3,57E-02	GO:0090092	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	21	586	1,89E-06
GO:0001763	morphogenesis of a branching structure	28	690	9,76E-11	GO:0007507	heart development	21	638	8,58E-06
GO:0048589	developmental growth	28	714	2,26E-10	GO:0051091	positive regulation of sequence-specific DNA binding transcription factor activity	21	677	2,42E-05
GO:0030111	regulation of Wnt signaling pathway	28	814	5,34E-09					

GO-term	Descrizione	Geni	Classe	P-value
GO:0007584	response to nutrient	21	690	3,37E-05
GO:0007626	locomotory behavior	21	829	7,43E-04
GO:0051251	positive regulation of lymphocyte activation	21	874	1,76E-03
GO:0031346	positive regulation of cell projection organization	21	897	2,67E-03
GO:0051052	regulation of DNA metabolic process	21	905	3,07E-03
GO:0048511	rhythmic process	21	930	4,74E-03
GO:0045666	positive regulation of neuron differentiation	20	351	1,27E-09
GO:0060537	muscle tissue development	20	387	7,64E-09
GO:0048705	skeletal system morphogenesis	20	410	2,18E-08
GO:0016202	regulation of striated muscle tissue development	20	469	2,42E-07
GO:0030856	regulation of epithelial cell differentiation	20	470	2,51E-07
GO:1901861	regulation of muscle tissue development	20	471	2,61E-07
GO:0048634	regulation of muscle organ development	20	482	3,92E-07
GO:0001667	ameboidal cell migration	20	542	3,02E-06
GO:0050953	sensory perception of light stimulus	20	769	9,72E-04
GO:1901342	regulation of vasculature development	20	774	1,08E-03
GO:1901214	regulation of neuron death	20	808	2,11E-03
GO:0010817	regulation of hormone levels	20	868	6,32E-03

LISTA A16: Enrichment dei FT delle CFT significative

Enrichment dei GO-term (biological process) annotati ai FT delle CFT significative. Sono stati utilizzati come input tutti e 196 i FT. I p-value sono corretti con il metodo Bonferroni, il livello di significatività di soglia usato è stato pari a un p-value di $5 \cdot 10^{-2}$. Sono stati selezionati solo i termini riguardanti un numero di geni superiore o uguale a 40. A sinistra sono elencati i GO-term, seguiti da una loro descrizione. Il campo "Geni" indica il numero di geni arricchiti per l'annotazione corrispondente, "Classe" indica il numero totale di geni annotati per lo stesso termine nell'intero database di riferimento (NET-GE). Tutti i termini sono ordinati in ordine di numero di geni annotati decrescente. I termini evidenziati in grassetto sottolineano il coinvolgimento di questi geni nello sviluppo embrionale di vari tessuti.

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0006355	regulation of transcription, DNA-templated	196	6146	8,02E-153	GO:0006357	regulation of transcription from RNA polymerase II promoter	166	2379	2,57E-164
GO:2001141	regulation of RNA biosynthetic process	196	6249	2,20E-151	GO:0044699	single-organism process	166	20881	5,36E-15
GO:0051252	regulation of RNA metabolic process	196	6360	7,31E-150	GO:0010629	negative regulation of gene expression	163	3561	4,53E-123
GO:2000112	regulation of cellular macromolecule biosynthetic process	196	6579	6,17E-147	GO:0048522	positive regulation of cellular process	162	5768	2,08E-94
GO:0010556	regulation of macromolecule biosynthetic process	196	6784	2,76E-144	GO:0048518	positive regulation of biological process	162	6624	6,09E-85
GO:0031326	regulation of cellular biosynthetic process	196	6987	9,70E-142	GO:0045944	positive regulation of transcription from RNA polymerase II promoter	161	2989	4,72E-132
GO:0009889	regulation of biosynthetic process	196	7051	5,94E-141	GO:0045934	negative regulation of nucleobase-containing compound metabolic process	161	3691	2,94E-117
GO:0010468	regulation of gene expression	196	7158	1,19E-139	GO:0051172	negative regulation of nitrogen compound metabolic process	161	3798	2,89E-115
GO:0019219	regulation of nucleobase-containing compound metabolic process	196	7797	2,81E-132	GO:0051253	negative regulation of RNA metabolic process	160	3369	5,55E-122
GO:0051171	regulation of nitrogen compound metabolic process	196	7927	7,49E-131	GO:0045892	negative regulation of transcription, DNA-templated	159	3247	6,86E-123
GO:0060255	regulation of macromolecule metabolic process	196	8942	1,78E-120	GO:1902679	negative regulation of RNA biosynthetic process	159	3295	7,11E-122
GO:0031323	regulation of cellular metabolic process	196	9728	3,15E-113	GO:2000113	negative regulation of cellular macromolecule biosynthetic process	159	3633	3,87E-115
GO:0080090	regulation of primary metabolic process	196	9808	1,60E-112	GO:0010558	negative regulation of macromolecule biosynthetic process	159	3771	1,41E-112
GO:0019222	regulation of metabolic process	196	10848	7,31E-104	GO:0010628	positive regulation of gene expression	154	1919	7,89E-157
GO:0050794	regulation of cellular process	196	17465	4,93E-63	GO:0010604	positive regulation of macromolecule metabolic process	154	3285	2,01E-120
GO:0050789	regulation of biological process	196	18486	3,59E-58	GO:0009893	positive regulation of metabolic process	154	3630	8,97E-114
GO:0065007	biological regulation	196	19296	1,68E-54	GO:0045893	positive regulation of transcription, DNA-templated	153	1762	8,31E-161
GO:0009987	cellular process	195	26673	6,64E-25	GO:1902680	positive regulation of RNA biosynthetic process	153	1811	6,30E-159
GO:0090304	nucleic acid metabolic process	190	5696	6,00E-144	GO:0051254	positive regulation of RNA metabolic process	153	1838	6,48E-158
GO:0044249	cellular biosynthetic process	190	5837	6,58E-142	GO:0010557	positive regulation of macromolecule biosynthetic process	153	2042	9,76E-151
GO:1901576	organic substance biosynthetic process	190	6037	4,24E-139	GO:0045935	positive regulation of nucleobase-containing compound metabolic process	153	2060	3,86E-150
GO:0009058	biosynthetic process	190	6217	1,19E-136	GO:0051173	positive regulation of nitrogen compound metabolic process	153	2108	1,41E-148
GO:0006139	nucleobase-containing compound metabolic process	190	7311	3,55E-123	GO:0031328	positive regulation of cellular biosynthetic process	153	2177	2,16E-146
GO:0046483	heterocycle metabolic process	190	7642	1,67E-119	GO:0009891	positive regulation of biosynthetic process	153	2208	1,96E-145
GO:0006725	cellular aromatic compound metabolic process	190	7662	2,75E-119	GO:0031325	positive regulation of cellular metabolic process	153	3418	3,36E-116
GO:0034641	cellular nitrogen compound metabolic process	190	7912	1,26E-116	GO:0032502	developmental process	148	7299	1,77E-61
GO:1901360	organic cyclic compound metabolic process	190	8010	1,32E-115	GO:0044767	single-organism developmental process	145	6740	7,41E-63
GO:0006807	nitrogen compound metabolic process	190	8714	1,23E-108	GO:0044763	single-organism cellular process	145	16559	3,44E-14
GO:0044260	cellular macromolecule metabolic process	190	11069	6,33E-89	GO:0009653	anatomical structure morphogenesis	142	4591	4,90E-76
GO:0043170	macromolecule metabolic process	190	12986	7,83E-76	GO:0000122	negative regulation of transcription from RNA polymerase II promoter	132	2314	3,56E-102
GO:0044237	cellular metabolic process	190	15758	4,55E-60	GO:0050896	response to stimulus	132	11721	4,93E-22
GO:0044238	primary metabolic process	190	16587	6,53E-56	GO:0048519	negative regulation of biological process	130	5756	2,41E-55
GO:0071704	organic substance metabolic process	190	17115	2,24E-53	GO:0048523	negative regulation of cellular process	127	5279	1,24E-56
GO:0008152	metabolic process	190	20927	3,16E-37	GO:0006366	transcription from RNA polymerase II promoter	122	1448	1,06E-112
GO:0032774	RNA biosynthetic process	189	3250	1,23E-188	GO:0044702	single organism reproductive process	121	3664	3,60E-63
GO:0034654	nucleobase-containing compound biosynthetic process	189	3760	2,21E-176	GO:0048869	cellular developmental process	118	3694	2,12E-64
GO:0019438	aromatic compound biosynthetic process	189	3919	6,55E-173	GO:0048856	anatomical structure development	118	4289	3,19E-57
GO:0018130	heterocycle biosynthetic process	189	3931	1,18E-172	GO:0009888	tissue development	116	2570	1,44E-74
GO:0044271	cellular nitrogen compound biosynthetic process	189	3991	2,19E-171	GO:0051094	positive regulation of developmental process	116	3200	5,55E-64
GO:1901362	organic cyclic compound biosynthetic process	189	4099	3,76E-169	GO:0048585	negative regulation of response to stimulus	115	3759	2,84E-55
GO:0009059	macromolecule biosynthetic process	189	4344	2,68E-164	GO:0051093	negative regulation of developmental process	113	2690	1,24E-68
GO:0016070	RNA metabolic process	189	4467	5,74E-162	GO:0071495	cellular response to endogenous stimulus	111	3038	1,25E-60
GO:0034645	cellular macromolecule biosynthetic process	168	4013	1,43E-129	GO:0033993	response to lipid	110	2604	1,72E-66
GO:0006351	transcription, DNA-templated	167	2949	2,22E-150	GO:0010648	negative regulation of cell communication	110	3347	3,80E-55

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0023057	negative regulation of signaling	110	3347	3,80E-55	GO:0070887	cellular response to chemical stimulus	85	2904	4,69E-39
GO:0048731	system development	109	2612	3,81E-65	GO:0042327	positive regulation of phosphorylation	85	2706	8,66E-38
GO:0009968	negative regulation of signal transduction	109	3204	5,41E-56	GO:0007154	cell communication	85	2752	3,16E-37
GO:0051716	cellular response to stimulus	109	9450	2,23E-16	GO:0071310	cellular response to organic substance	81	2482	3,08E-40
GO:0010605	negative regulation of macromolecule metabolic process	108	2452	1,66E-71	GO:1901698	response to nitrogen compound	81	2693	3,68E-34
GO:0009892	negative regulation of metabolic process	108	2679	1,66E-67	GO:0060341	regulation of cellular localization	81	3341	1,69E-27
GO:0031324	negative regulation of cellular metabolic process	107	2479	9,40E-70	GO:0048545	response to steroid hormone	79	1516	1,37E-50
GO:0014070	response to organic cyclic compound	107	2783	5,80E-60	GO:0045664	regulation of neuron differentiation	79	1894	2,36E-43
GO:0009725	response to hormone	107	2824	2,55E-59	GO:0048729	tissue morphogenesis	78	1359	5,68E-53
GO:0031327	negative regulation of cellular biosynthetic process	106	1764	8,38E-84	GO:0051726	regulation of cell cycle	78	2760	1,11E-30
GO:0009890	negative regulation of biosynthetic process	106	1786	3,08E-83	GO:0016265	death	78	2848	9,59E-30
GO:0048468	cell development	106	2280	1,40E-67	GO:0006955	immune response	78	3063	1,35E-27
GO:0060284	regulation of cell development	106	2979	7,25E-56	GO:0016310	phosphorylation	78	4277	3,54E-18
GO:0060548	negative regulation of cell death	102	2642	1,72E-56	GO:0003002	regionalization	77	988	3,33E-62
GO:0045597	positive regulation of cell differentiation	101	2514	2,02E-57	GO:0048609	multicellular organismal reproductive process	77	2109	6,85E-38
GO:0008284	positive regulation of cell proliferation	101	2808	7,44E-53	GO:0040008	regulation of growth	77	2308	3,86E-35
GO:0040011	locomotion	101	3142	2,82E-48	GO:0007167	enzyme linked receptor protein signaling pathway	77	2504	1,10E-32
GO:0045596	negative regulation of cell differentiation	100	2226	2,64E-61	GO:0048513	organ development	76	1910	2,99E-43
GO:0008285	negative regulation of cell proliferation	100	2354	5,48E-59	GO:0010243	response to organonitrogen compound	76	2509	1,05E-31
GO:0043069	negative regulation of programmed cell death	99	2511	3,39E-55	GO:0008219	cell death	75	2827	2,24E-27
GO:0010647	positive regulation of cell communication	99	3801	8,29E-39	GO:0045595	regulation of cell differentiation	74	2111	4,20E-38
GO:0043066	negative regulation of apoptotic process	98	2487	1,86E-54	GO:0001934	positive regulation of protein phosphorylation	74	2371	1,64E-31
GO:0008283	cell proliferation	98	2526	7,80E-54	GO:0050776	regulation of immune response	74	2377	1,94E-31
GO:0051960	regulation of nervous system development	98	2584	6,32E-53	GO:0050877	neurological system process	74	2735	1,86E-27
GO:0022603	regulation of anatomical structure morphogenesis	98	2832	2,77E-49	GO:1902533	positive regulation of intracellular signal transduction	73	2418	4,87E-30
GO:0080134	regulation of response to stress	98	3072	4,36E-46	GO:0048732	gland development	72	1014	1,14E-54
GO:0009628	response to abiotic stimulus	98	3444	1,18E-41	GO:0001775	cell activation	72	2070	1,70E-33
GO:0009967	positive regulation of signal transduction	98	3621	9,85E-40	GO:0070848	response to growth factor	71	1987	1,12E-33
GO:0042221	response to chemical	98	4712	7,10E-34	GO:0051338	regulation of transferase activity	71	2819	3,91E-24
GO:0003006	developmental process involved in reproduction	97	2335	7,58E-56	GO:0065003	macromolecular complex assembly	71	3681	2,61E-17
GO:0051247	positive regulation of protein metabolic process	97	3557	1,80E-39	GO:0048583	regulation of response to stimulus	71	4515	4,44E-15
GO:0048646	anatomical structure formation involved in morphogenesis	96	2968	2,22E-45	GO:0030855	epithelial cell differentiation	70	1135	1,01E-48
GO:0030154	cell differentiation	95	2446	6,51E-56	GO:0002009	morphogenesis of an epithelium	70	1159	4,16E-48
GO:0050767	regulation of neurogenesis	94	2377	9,08E-52	GO:0040007	growth	70	1496	1,08E-40
GO:0048870	cell motility	94	2808	1,95E-45	GO:2000026	regulation of multicellular organismal development	70	2229	1,42E-32
GO:0032501	multicellular organismal process	94	4447	1,45E-32	GO:0032101	regulation of response to external stimulus	70	2173	2,99E-30
GO:0007389	pattern specification process	93	1481	4,90E-69	GO:0006629	lipid metabolic process	70	3598	3,46E-17
GO:0010033	response to organic substance	93	3487	2,11E-40	GO:0071407	cellular response to organic cyclic compound	69	1148	3,66E-47
GO:0044707	single-multicellular organism process	92	4361	1,47E-31	GO:0007610	behavior	69	1993	1,17E-31
GO:0044092	negative regulation of molecular function	91	3258	4,42E-37	GO:0051270	regulation of cellular component movement	69	2182	3,22E-29
GO:0032270	positive regulation of cellular protein metabolic process	91	3279	7,45E-37	GO:0002684	positive regulation of immune system process	69	2287	5,74E-28
GO:0001932	regulation of protein phosphorylation	91	3361	5,52E-36	GO:0051130	positive regulation of cellular component organization	68	2782	4,49E-22
GO:0016477	cell migration	90	2574	1,61E-44	GO:0007275	multicellular organismal development	67	1673	2,11E-34
GO:0042592	homeostatic process	90	3571	5,48E-33	GO:0009719	response to endogenous stimulus	67	2012	2,01E-32
GO:0009887	organ morphogenesis	89	1719	3,30E-58	GO:0071363	cellular response to growth factor stimulus	67	1903	5,64E-31
GO:0050793	regulation of developmental process	89	2884	2,81E-43	GO:0040012	regulation of locomotion	67	2224	6,43E-27
GO:0010562	positive regulation of phosphorus metabolic process	89	3001	4,37E-38	GO:0012501	programmed cell death	67	2247	1,18E-26
GO:0045937	positive regulation of phosphate metabolic process	89	3001	4,37E-38	GO:0043549	regulation of kinase activity	67	2576	3,39E-23
GO:0051239	regulation of multicellular organismal process	89	3432	3,48E-37	GO:0071396	cellular response to lipid	66	1198	2,24E-42
GO:0007165	signal transduction	89	7592	3,86E-12	GO:0023052	signaling	66	2287	2,46E-25
GO:0048598	embryonic morphogenesis	88	1566	1,95E-60	GO:0044700	single organism signaling	66	2287	2,46E-25
GO:0051240	positive regulation of multicellular organismal process	88	2185	2,78E-48	GO:0009790	embryo development	65	1155	3,36E-42
GO:0031401	positive regulation of protein modification process	88	2947	9,15E-38	GO:0006325	chromatin organization	65	1572	5,34E-34
GO:1901701	cellular response to oxygen-containing compound	87	2540	6,40E-42	GO:0006915	apoptotic process	65	2154	5,83E-26
GO:0032870	cellular response to hormone stimulus	85	1636	3,70E-55	GO:0048878	chemical homeostasis	65	2410	3,39E-23

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0022402	cell cycle process	65	2771	7,56E-20	GO:0071345	cellular response to cytokine stimulus	55	1640	2,53E-23
GO:0055114	oxidation-reduction process	65	3568	4,43E-14	GO:0030030	cell projection organization	55	2507	1,12E-14
GO:0010942	positive regulation of cell death	64	1542	1,81E-33	GO:0007155	cell adhesion	55	2817	1,86E-12
GO:0032989	cellular component morphogenesis	64	1879	1,85E-28	GO:0022610	biological adhesion	55	2824	2,07E-12
GO:2000145	regulation of cell motility	64	2039	1,98E-26	GO:0030182	neuron differentiation	54	911	3,53E-35
GO:0007267	cell-cell signaling	64	2185	9,76E-25	GO:0031667	response to nutrient levels	54	1219	9,78E-29
GO:0051248	negative regulation of protein metabolic process	64	2189	1,08E-24	GO:0009991	response to extracellular stimulus	54	1313	3,95E-27
GO:0045859	regulation of protein kinase activity	64	2438	4,24E-22	GO:0080135	regulation of cellular response to stress	54	1388	6,12E-26
GO:0051050	positive regulation of transport	64	2507	1,95E-21	GO:0002694	regulation of leukocyte activation	54	1443	4,10E-25
GO:0044711	single-organism biosynthetic process	64	3675	8,11E-13	GO:0042493	response to drug	54	1473	1,12E-24
GO:0006950	response to stress	64	4134	9,91E-13	GO:0032880	regulation of protein localization	54	2200	1,71E-16
GO:0031347	regulation of defense response	63	1612	2,50E-31	GO:0035295	tube development	53	664	4,25E-41
GO:0043086	negative regulation of catalytic activity	63	2637	1,77E-19	GO:0010467	gene expression	53	1348	1,38E-25
GO:0060429	epithelium development	62	1006	2,74E-42	GO:0071417	cellular response to organonitrogen compound	53	1481	1,25E-23
GO:0045321	leukocyte activation	62	1486	2,38E-32	GO:0007276	gamete generation	53	1582	2,84E-22
GO:0001817	regulation of cytokine production	62	1774	5,37E-28	GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	53	1794	9,86E-20
GO:0042127	regulation of cell proliferation	62	2008	1,07E-27	GO:0044087	regulation of cellular component biogenesis	53	1812	1,56E-19
GO:0030334	regulation of cell migration	62	1926	5,13E-26	GO:0051345	positive regulation of hydrolase activity	53	2953	2,35E-10
GO:0009607	response to biotic stimulus	62	2197	6,67E-23	GO:0006508	proteolysis	53	3357	3,58E-08
GO:0023051	regulation of signaling	62	3708	8,99E-14	GO:0019637	organophosphate metabolic process	53	4312	2,76E-04
GO:0010646	regulation of cell communication	62	3714	9,71E-14	GO:0007411	axon guidance	52	1242	2,62E-26
GO:0050878	regulation of body fluid levels	61	1837	3,12E-26	GO:0097485	neuron projection guidance	52	1243	2,72E-26
GO:0034097	response to cytokine	61	2017	4,86E-24	GO:0010564	regulation of cell cycle process	52	1606	4,41E-21
GO:0048608	reproductive structure development	60	1062	1,55E-38	GO:0045860	positive regulation of protein kinase activity	52	1636	1,03E-20
GO:0043065	positive regulation of apoptotic process	60	1406	1,19E-31	GO:0033674	positive regulation of kinase activity	52	1706	6,93E-20
GO:0043068	positive regulation of programmed cell death	60	1416	1,77E-31	GO:2001233	regulation of apoptotic signaling pathway	51	1313	3,41E-24
GO:1901699	cellular response to nitrogen compound	60	1626	3,53E-28	GO:0031344	regulation of cell projection organization	51	1640	8,40E-20
GO:0030162	regulation of proteolysis	60	1768	3,26E-26	GO:0032269	negative regulation of cellular protein metabolic process	51	1804	5,67E-18
GO:0045087	innate immune response	60	1935	4,00E-24	GO:0043408	regulation of MAPK cascade	51	1837	1,25E-17
GO:0051046	regulation of secretion	60	2094	2,53E-22	GO:0006259	DNA metabolic process	51	2443	2,07E-12
GO:0006468	protein phosphorylation	60	3039	3,04E-14	GO:0005975	carbohydrate metabolic process	51	2526	7,97E-12
GO:0006461	protein complex assembly	60	3196	3,26E-13	GO:0031329	regulation of cellular catabolic process	51	3094	2,07E-08
GO:0010035	response to inorganic substance	59	1420	2,16E-30	GO:0043436	oxoacid metabolic process	51	3244	1,18E-07
GO:0051129	negative regulation of cellular component organization	59	1913	1,66E-23	GO:0006082	organic acid metabolic process	51	3284	1,85E-07
GO:0043207	response to external biotic stimulus	59	2065	8,51E-22	GO:0006352	DNA-templated transcription, initiation	50	564	9,69E-41
GO:0033043	regulation of organelle organization	59	2759	1,46E-15	GO:0001101	response to acid chemical	50	1170	1,60E-25
GO:0044255	cellular lipid metabolic process	59	2817	3,93E-15	GO:0051249	regulation of lymphocyte activation	50	1264	5,49E-24
GO:0016568	chromatin modification	58	1339	9,55E-31	GO:1901652	response to peptide	50	1373	2,31E-22
GO:0051347	positive regulation of transferase activity	58	1933	2,08E-22	GO:0016032	viral process	50	1665	1,15E-18
GO:0018193	peptidyl-amino acid modification	58	2224	2,31E-19	GO:0044403	symbiosis, encompassing mutualism through parasitism	50	1665	1,15E-18
GO:0048511	rhythmic process	57	930	3,05E-38	GO:0044764	multi-organism cellular process	50	1682	1,79E-18
GO:0009792	embryo development ending in birth or egg hatching	57	937	4,61E-38	GO:0010720	positive regulation of cell development	49	977	4,41E-28
GO:0046649	lymphocyte activation	57	1196	2,54E-32	GO:0032844	regulation of homeostatic process	49	1155	8,93E-25
GO:0051241	negative regulation of multicellular organismal process	57	1513	5,96E-27	GO:1901615	organic hydroxy compound metabolic process	49	1475	4,38E-20
GO:0050865	regulation of cell activation	57	1568	3,76E-26	GO:0022604	regulation of cell morphogenesis	49	1615	2,15E-18
GO:0044419	interspecies interaction between organisms	57	1908	7,75E-22	GO:0070201	regulation of establishment of protein localization	49	1849	6,30E-16
GO:0009894	regulation of catabolic process	57	3408	3,82E-10	GO:0045184	establishment of protein localization	49	3728	1,30E-04
GO:0001701	in utero embryonic development	56	893	5,25E-38	GO:0006811	ion transport	49	3852	3,65E-04
GO:0043009	chordate embryonic development	56	909	1,37E-37	GO:0007166	cell surface receptor signaling pathway	49	4368	4,51E-04
GO:0051090	regulation of sequence-specific DNA binding transcription factor activity	56	1091	2,37E-33	GO:0006367	transcription initiation from RNA polymerase II promoter	48	419	2,60E-44
GO:0022414	reproductive process	56	1738	2,04E-25	GO:0009952	anterior/posterior pattern specification	48	585	2,60E-37
GO:0007600	sensory perception	56	1675	8,95E-24	GO:0048589	developmental growth	48	714	3,05E-33
GO:0009966	regulation of signal transduction	56	3261	1,67E-12	GO:0051098	regulation of binding	48	826	2,55E-30
GO:0045165	cell fate commitment	55	501	1,65E-50	GO:0070482	response to oxygen levels	48	1076	3,84E-25
GO:0002521	leukocyte differentiation	55	929	6,65E-36	GO:1903034	regulation of response to wounding	48	1274	6,29E-22

GO-term	Descrizione	Geni	Classe	P-value
GO:0007599	hemostasis	48	1382	2,08E-20
GO:0065008	regulation of biological quality	48	3888	3,27E-05
GO:0046907	intracellular transport	48	3768	4,96E-04
GO:0048562	embryonic organ morphogenesis	47	484	7,23E-40
GO:0048666	neuron development	47	539	1,14E-37
GO:0030522	intracellular receptor signaling pathway	47	544	1,76E-37
GO:2000027	regulation of organ morphogenesis	47	574	2,14E-36
GO:0042110	T cell activation	47	787	3,98E-30
GO:1902532	negative regulation of intracellular signal transduction	47	1289	8,83E-21
GO:0051051	negative regulation of transport	47	1375	1,32E-19
GO:0007283	spermatogenesis	47	1413	4,10E-19
GO:0048232	male gamete generation	47	1420	5,04E-19
GO:0032386	regulation of intracellular transport	47	1443	9,79E-19
GO:0051223	regulation of protein transport	47	1645	2,05E-16
GO:0010941	regulation of cell death	47	2079	2,30E-14
GO:0015031	protein transport	47	3506	1,51E-04
GO:0030098	lymphocyte differentiation	46	654	1,52E-32
GO:0052548	regulation of endopeptidase activity	46	1238	1,38E-20
GO:0010975	regulation of neuron projection development	46	1281	5,63E-20
GO:0052547	regulation of peptidase activity	46	1292	8,00E-20
GO:0043434	response to peptide hormone	46	1303	1,13E-19
GO:0044057	regulation of system process	46	1316	1,70E-19
GO:0007596	blood coagulation	46	1367	8,01E-19
GO:0050817	coagulation	46	1367	8,01E-19
GO:0044708	single-organism behavior	46	1465	1,31E-17
GO:0009314	response to radiation	46	1557	1,49E-16
GO:0009057	macromolecule catabolic process	46	2318	5,12E-10
GO:0019752	carboxylic acid metabolic process	46	2978	2,65E-06
GO:0048534	hematopoietic or lymphoid organ development	45	682	1,57E-30
GO:1902105	regulation of leukocyte differentiation	45	834	9,01E-27
GO:0010769	regulation of cell morphogenesis involved in differentiation	45	1115	1,65E-21
GO:0050778	positive regulation of immune response	45	1484	1,51E-16
GO:1901700	response to oxygen-containing compound	45	1851	8,75E-15
GO:0019725	cellular homeostasis	45	1855	7,32E-13
GO:1901361	organic cyclic compound catabolic process	45	3169	5,82E-05
GO:0030111	regulation of Wnt signaling pathway	44	814	4,08E-26
GO:0001819	positive regulation of cytokine production	44	963	4,10E-23
GO:0050678	regulation of epithelial cell proliferation	44	1020	4,21E-22
GO:0043067	regulation of programmed cell death	44	1982	6,29E-13
GO:1903047	mitotic cell cycle process	44	1769	6,91E-13
GO:0050863	regulation of T cell activation	43	912	4,85E-23
GO:0001666	response to hypoxia	43	987	1,12E-21
GO:0006954	inflammatory response	43	1314	7,03E-17
GO:0001558	regulation of cell growth	43	1390	5,79E-16
GO:0042981	regulation of apoptotic process	43	1970	2,70E-12
GO:0046903	secretion	43	1910	5,42E-11
GO:0071383	cellular response to steroid hormone stimulus	42	394	7,08E-37
GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	42	618	9,73E-29
GO:0001763	morphogenesis of a branching structure	42	690	8,38E-27
GO:0007568	aging	42	848	3,01E-23
GO:0036293	response to decreased oxygen levels	42	999	1,73E-20
GO:0007346	regulation of mitotic cell cycle	42	1240	5,96E-17
GO:0043410	positive regulation of MAPK cascade	42	1269	1,40E-16
GO:0002252	immune effector process	42	1396	4,57E-15
GO:0007010	cytoskeleton organization	42	2220	3,59E-08

GO-term	Descrizione	Geni	Classe	P-value
GO:1900542	regulation of purine nucleotide metabolic process	42	2485	1,22E-06
GO:0006140	regulation of nucleotide metabolic process	42	2500	1,47E-06
GO:0051246	regulation of protein metabolic process	42	2954	6,19E-06
GO:0055086	nucleobase-containing small molecule metabolic process	42	3391	8,10E-03
GO:0001822	kidney development	41	489	1,28E-31
GO:0001764	neuron migration	41	578	1,07E-28
GO:0021700	developmental maturation	41	666	2,90E-26
GO:1901342	regulation of vasculature development	41	774	1,01E-23
GO:0000902	cell morphogenesis	41	800	3,59E-23
GO:0019216	regulation of lipid metabolic process	41	845	2,91E-22
GO:1901653	cellular response to peptide	41	892	2,28E-21
GO:0016569	covalent chromatin modification	41	935	1,35E-20
GO:0097305	response to alcohol	41	1084	3,31E-18
GO:0051346	negative regulation of hydrolase activity	41	1348	8,67E-15
GO:0006974	cellular response to DNA damage stimulus	41	1933	1,79E-09
GO:0046700	heterocycle catabolic process	41	3029	1,12E-03
GO:0016192	vesicle-mediated transport	41	3055	1,40E-03
GO:0019439	aromatic compound catabolic process	41	3055	1,40E-03
GO:0061138	morphogenesis of a branching epithelium	40	624	3,33E-26
GO:0007507	heart development	40	638	7,82E-26
GO:0051052	regulation of DNA metabolic process	40	905	3,95E-20
GO:0002237	response to molecule of bacterial origin	40	918	6,67E-20
GO:0016570	histone modification	40	929	1,03E-19
GO:0002764	immune response-regulating signaling pathway	40	1248	3,95E-15
GO:0010563	negative regulation of phosphorus metabolic process	40	1353	6,45E-14
GO:0045936	negative regulation of phosphate metabolic process	40	1353	6,45E-14
GO:0031400	negative regulation of protein modification process	40	1368	9,41E-14
GO:0044270	cellular nitrogen compound catabolic process	40	3027	3,01E-03
GO:0009117	nucleotide metabolic process	40	3194	1,16E-02
GO:0006753	nucleoside phosphate metabolic process	40	3223	1,45E-02

LISTA A17: Bicluster significativi

Elenco dei bicluster. "Bic": codice identificativo del bicluster. "FT (#)": numero totale di FT del bicluster. "FT-enh (#)", "FT-pro (#)": numero di FT su enhancer e su promotori rispettivamente. "CEP (#)": Numero CEP del bicluster. "Enhancer (#)", "Promotori (#)": numero di enhancer e promotori che compongono le CEP del bicluster. "P-value": significatività statistica del bicluster.

Bic	FT (#)	FT-enh (#)	FT-pro (#)	CEP (#)	Enhancer (#)	Promotori (#)	P-value
1	4	1	3	31	31	25	5.02e-29
2	5	4	1	7	6	7	1.69e-25
3	4	3	1	10	9	10	1.05e-23
4	5	1	5	8	8	8	1.39e-22
5	3	3	1	11	9	11	1.80e-21
6	4	1	3	18	18	14	7.25e-17
7	3	3	1	9	9	9	1.42e-16
8	6	2	4	7	7	7	3.52e-16
9	4	3	1	7	7	7	8.38e-15
10	4	2	2	14	13	13	2.26e-14
11	3	3	1	7	7	7	3.63e-14
12	3	1	2	26	25	26	9.26e-14
13	3	1	2	32	32	27	2.92e-12
14	3	1	2	24	23	22	4.81e-12
15	4	1	3	13	13	13	1.31e-11
16	4	1	3	7	6	6	5.82e-11
17	3	2	1	13	13	12	6.59e-11
18	3	1	2	19	18	19	9.24e-11
19	3	1	2	30	29	27	2.72e-10
20	4	1	3	7	7	7	2.80e-10
21	4	1	3	7	7	4	3.62e-09
22	4	2	2	9	8	9	4.78e-09
23	3	1	2	16	16	16	1.07e-08
24	3	1	2	10	8	8	1.39e-08
25	3	2	1	7	6	7	1.83e-07
26	3	1	2	16	16	16	4.99e-07
27	3	1	2	30	28	25	1.42e-06
28	3	2	1	7	7	7	1.45e-06
29	3	1	2	19	19	19	2.07e-06
30	3	1	2	18	18	16	3.13e-06
31	3	1	2	11	10	11	8.42e-06
32	3	1	2	7	7	7	1.46e-05
33	3	1	2	7	6	7	2.05e-05
34	3	1	2	10	9	9	2.46e-05
35	3	1	2	15	14	14	2.47e-05
36	3	1	2	11	11	11	2.52e-05
37	3	1	2	9	9	9	2.57e-05
38	3	1	2	11	11	10	2.84e-05
39	3	1	2	10	10	10	2.85e-05
40	3	1	2	7	7	6	3.22e-05
41	3	2	1	10	10	10	7.27e-05
42	2	1	2	7	7	7	7.74e-05
43	3	1	2	7	7	7	1.42e-04
44	3	2	1	7	7	7	1.45e-04
45	3	1	2	11	10	11	1.74e-04
46	3	1	2	10	10	10	2.32e-04
47	3	2	1	9	8	9	4.40e-04
48	2	2	1	10	10	10	5.43e-04

LISTA A18: FT e CEP per ogni bicluster significativo

Elenco completo di FT e CEP che compongono ogni bicluster. Le sigle "BIC#" in grassetto indicano il codice identificativo del bicluster. Successivamente per ogni bicluster sono riportati i FT su enhancer (gene symbol) e i FT sui promotori (gene symbol) separati da virgole. Il campo riservato alle CEP comincia con la sigla "CEP" dopo la quale sono scritte tutte le CEP del bicluster, una per riga. Ognuna è completa del proprio codice identificativo ("CEP#"), del codice identificativo dell'enhancer (VISTA ID) e quello del gene del promotore (gene symbol).

BIC1

FT enhancer:
MYF6

FT promotore:
SP1,SP3,ZNF148

CEP:
CEP25: hs1912,ARHGEF16
CEP28: hs1510,ARID3B
CEP86: hs533,C5orf38
CEP88: hs754,C5orf38
CEP217: hs1828,ERIG1
CEP245: hs513,FGF8
CEP302: hs1100,FOXP4
CEP317: hs921,GBX2
CEP325: hs1750,GPC1
CEP326: hs1937,GPC1
CEP334: hs1882,GPR20
CEP382: hs1466,IRF2BPL
CEP400: hs1430,JAZF1
CEP540: hs483,MED13L
CEP603: hs699,MKI67
CEP659: hs1188,OLIG2
CEP660: hs1548,OLIG2
CEP663: hs1273,OSR1
CEP711: hs2096,PLEKHA7
CEP742: hs1955,QKI
CEP785: hs152,SALL1
CEP789: hs73,SALL1
CEP830: hs1391,SLC22A23
CEP843: hs1387,SNTN
CEP845: hs1496,SOX11
CEP851: hs667,SOX3
CEP891: hs1887,TEAD3
CEP900: hs1163,TLE1
CEP901: hs1359,TLE1
CEP905: hs974,TLE1
CEP943: hs1751,TUBB6

BIC2

FT enhancer:
SP1,SP3,SP4,ZBTB7B

FT promotore:
ESR2

CEP:
CEP25: hs1912,ARHGEF16
CEP118: hs1760,CDH16
CEP321: hs522,GLI2
CEP664: hs262,OTP
CEP759: hs1971,RBM38
CEP835: hs1507,SLC4A3
CEP883: hs262,TBCA

BIC3

FT enhancer:
SP3,SP4,ZBTB7B

FT promotore:
ZNF148

CEP:
CEP25: hs1912,ARHGEF16
CEP57: hs1891,BCAN
CEP150: hs1971,CTCF
CEP326: hs1937,GPC1
CEP334: hs1882,GPR20
CEP641: hs1205,NKX2-2

CEP660: hs1548,OLIG2
CEP759: hs1971,RBM38
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3

BIC4

FT enhancer:
ZBTB7B

FT promotore:
SP1,SP3,SP4,ZBTB7B,ZNF148

CEP:
CEP64: hs307,BNC2
CEP244: hs511,FGF8
CEP326: hs1937,GPC1
CEP382: hs1466,IRF2BPL
CEP660: hs1548,OLIG2
CEP759: hs1971,RBM38
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3

BIC5

FT enhancer:
SP1,SP3,ZNF148

FT promotore:
SP1

CEP:
CEP25: hs1912,ARHGEF16
CEP112: hs1920,CCND1
CEP150: hs1971,CTCF
CEP334: hs1882,GPR20
CEP382: hs1466,IRF2BPL
CEP403: hs1467,KCNJ2
CEP643: hs680,NKX6-1
CEP759: hs1971,RBM38
CEP858: hs1467,SOX9
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3

BIC6

FT enhancer:
MYF6

FT promotore:
MAZ,SP4,ZBTB7B

CEP:
CEP138: hs1486,CLMN
CEP245: hs513,FGF8
CEP374: hs1932,IGF1R
CEP382: hs1466,IRF2BPL
CEP400: hs1430,JAZF1
CEP548: hs599,MEIS2
CEP549: hs830,MEIS2
CEP659: hs1188,OLIG2
CEP660: hs1548,OLIG2
CEP711: hs2096,PLEKHA7
CEP779: hs2061,RREB1
CEP785: hs152,SALL1
CEP789: hs73,SALL1
CEP795: hs1020,SALL3
CEP798: hs634,SALL3
CEP807: hs1362,SETBP1
CEP830: hs1391,SLC22A23
CEP845: hs1496,SOX11

BIC7

FT enhancer:
MAZ,SP4,ZNF148

FT promotore:
SP4

CEP:
CEP110: hs1931,CCDC85C
CEP112: hs1920,CCND1
CEP641: hs1205,NKX2-2
CEP678: hs1202,PBX1
CEP835: hs1507,SLC4A3
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3
CEP936: hs1300,TOX3
CEP992: hs1437,ZNF503-AS1

BIC8

FT enhancer:
MYF6,PPARA

FT promotore:
SP1,SP3,ZBTB7B,ZNF148

CEP:
CEP302: hs1100,FOXP4
CEP382: hs1466,IRF2BPL
CEP603: hs699,MKI67
CEP711: hs2096,PLEKHA7
CEP742: hs1955,QKI
CEP845: hs1496,SOX11
CEP891: hs1887,TEAD3

BIC9

FT enhancer:
SP1,SP3,ZBTB7B

FT promotore:
PGR

CEP:
CEP3: hs1487,ACOT11
CEP642: hs1205,NKX2-4
CEP646: hs1320,NR2F2
CEP728: hs876,PRDM12
CEP759: hs1971,RBM38
CEP844: hs872,SORL1
CEP895: hs522,TFCP2L1

BIC10

FT enhancer:
MYF6,SP4

FT promotore:
SP1,SP3

CEP:
CEP7: hs921,AGAP1
CEP25: hs1912,ARHGEF16
CEP110: hs1931,CCDC85C
CEP112: hs1920,CCND1
CEP317: hs921,GBX2
CEP325: hs1750,GPC1
CEP326: hs1937,GPC1
CEP334: hs1882,GPR20
CEP378: hs1763,INPP5A
CEP660: hs1548,OLIG2

CEP671: hs1860,PARD6B
CEP716: hs1081,POU3F2
CEP881: hs1933,TBC1D8
CEP891: hs1887,TEAD3

BIC11

FT enhancer:
MAZ,SP3,ZBTB7B

FT promotore:
ZBTB7B

CEP:
CEP112: hs1920,CCND1
CEP242: hs781,FGF17
CEP382: hs1466,IRF2BPL
CEP641: hs1205,NKX2-2
CEP660: hs1548,OLIG2
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3

BIC12

FT enhancer:
ZNF148

FT promotore:
SP1,SP3

CEP:
CEP25: hs1912,ARHGEF16
CEP40: hs1540,ASCL1
CEP64: hs307,BNC2
CEP110: hs1931,CCDC85C
CEP112: hs1920,CCND1
CEP153: hs260,CXXC4
CEP193: hs1208,EBF1
CEP326: hs1937,GPC1
CEP334: hs1882,GPR20
CEP382: hs1466,IRF2BPL
CEP403: hs1467,KCNJ2
CEP553: hs699,MGMT
CEP603: hs699,MKI67
CEP643: hs680,NKX6-1
CEP670: hs2064,PAQR8
CEP678: hs1202,PBX1
CEP750: hs608,RANBP17
CEP759: hs1971,RBM38
CEP786: hs162,SALL1
CEP845: hs1496,SOX11
CEP848: hs488,SOX21
CEP875: hs1487,SSBP3
CEP890: hs1314,TEAD1
CEP891: hs1887,TEAD3
CEP905: hs974,TLE1
CEP990: hs1767,ZMIZ1

BIC13

FT enhancer:
NANOG

FT promotore:
SP1,SP3

CEP:
CEP41: hs1575,ASCL1
CEP51: hs775,ATP9B
CEP54: hs658,AUTS2
CEP55: hs925,AUTS2
CEP86: hs533,C5orf38

CEP195: hs813,EBF1
CEP234: hs526,FAM53A
CEP289: hs1717,FOXO1
CEP325: hs1750,GPC1
CEP355: hs1543,HES1
CEP359: hs329,HMX3
CEP371: hs1527,ID2
CEP400: hs1430,JAZF1
CEP557: hs1465,MIR148A
CEP650: hs1352,NRSN1
CEP656: hs243,OLA1
CEP657: hs244,OLA1
CEP659: hs1188,OLIG2
CEP671: hs1860,PARD6B
CEP720: hs563,POU3F2
CEP727: hs2040,PPP3R1
CEP742: hs1955,QKI
CEP749: hs278,RANBP17
CEP750: hs608,RANBP17
CEP787: hs65,SALL1
CEP790: hs74,SALL1
CEP848: hs488,SOX21
CEP862: hs1019,SP8
CEP875: hs1487,SSBP3
CEP955: hs430,URI1
CEP956: hs721,URI1
CEP990: hs1767,ZMIZ1

BIC14

FT enhancer:
MYF6

FT promotore:
SP3,WT1

CEP:
CEP7: hs921,AGAP1
CEP28: hs1510,ARID3B
CEP51: hs775,ATP9B
CEP103: hs124,CBLN1
CEP138: hs1486,CLMN
CEP198: hs2060,ECI2
CEP245: hs513,FGF8
CEP302: hs1100,FOXP4
CEP333: hs796,GPR180
CEP400: hs1430,JAZF1
CEP423: hs314,LHX2
CEP553: hs699,MGMT
CEP659: hs1188,OLIG2
CEP660: hs1548,OLIG2
CEP663: hs1273,OSR1
CEP711: hs2096,PLEKHA7
CEP847: hs341,SOX21
CEP849: hs796,SOX21
CEP863: hs844,SP8
CEP881: hs1933,TBC1D8
CEP891: hs1887,TEAD3
CEP941: hs2007,TSPAN17
CEP943: hs1751,TUBB6
CEP1009: hs1308,ZNF800

BIC15

FT enhancer:
MAZ

FT promotore:
SP1,SP3,SP4

CEP:
CEP73: hs1766,C1orf198
CEP110: hs1931,CCDC85C

CEP112: hs1920,CCND1
CEP153: hs260,CXXC4
CEP196: hs632,EBF4
CEP382: hs1466,IRF2BPL
CEP552: hs656,MGMT
CEP660: hs1548,OLIG2
CEP678: hs1202,PBX1
CEP752: hs807,RAPGEF5
CEP785: hs152,SALL1
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3

BIC16

FT enhancer:
PAX2

FT promotore:
SP1,SP4,ZBTB7B

CEP:
CEP125: hs2060,CDYL
CEP198: hs2060,ECI2
CEP355: hs1543,HES1
CEP605: hs23,MMMP2
CEP785: hs152,SALL1
CEP789: hs73,SALL1
CEP816: hs2062,SHB

BIC17

FT enhancer:
SP1,SP3

FT promotore:
SP4

CEP:
CEP112: hs1920,CCND1
CEP378: hs1763,INPP5A
CEP382: hs1466,IRF2BPL
CEP641: hs1205,NKX2-2
CEP646: hs1320,NR2F2
CEP686: hs987,PBX3
CEP759: hs1971,RBM38
CEP835: hs1507,SLC4A3
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3
CEP911: hs1360,TLE4
CEP913: hs1585,TLE4
CEP992: hs1437,ZNF503-AS1

BIC18

FT enhancer:
ZNF148

FT promotore:
SP4,ZBTB7B

CEP:
CEP64: hs307,BNC2
CEP110: hs1931,CCDC85C
CEP112: hs1920,CCND1
CEP153: hs260,CXXC4
CEP155: hs162,CYLD
CEP326: hs1937,GPC1
CEP374: hs1932,IGF1R
CEP382: hs1466,IRF2BPL
CEP553: hs699,MGMT
CEP641: hs1205,NKX2-2
CEP695: hs1958,PFKP
CEP759: hs1971,RBM38
CEP786: hs162,SALL1
CEP845: hs1496,SOX11
CEP875: hs1487,SSBP3
CEP890: hs1314,TEAD1
CEP891: hs1887,TEAD3
CEP936: hs1300,TOX3
CEP990: hs1767,ZMIZ1

BIC19

FT enhancer:
PPARA

FT promotore:
SP1,SP3

CEP:
CEP25: hs1912,ARHGFEF16
CEP41: hs1575,ASCL1
CEP73: hs1766,C1orf198
CEP110: hs1931,CCDC85C
CEP138: hs1486,CLMN
CEP189: hs1022,EBF1
CEP200: hs1734,EFNA5
CEP215: hs1578,ERBB4
CEP302: hs1100,FOXP4
CEP334: hs1882,GPR20
CEP335: hs578,GPR26
CEP382: hs1466,IRF2BPL
CEP443: hs2082,LOC100134229
CEP516: hs1447,LPAR1
CEP527: hs1397,MALT1
CEP553: hs699,MGMT
CEP603: hs699,MKI67
CEP644: hs712,NKX6-1
CEP645: hs428,NOVA1
CEP711: hs2096,PLEKHA7
CEP718: hs1233,POU3F2
CEP742: hs1955,QKI
CEP845: hs1496,SOX11
CEP848: hs488,SOX21
CEP891: hs1887,TEAD3
CEP953: hs381,UR11
CEP954: hs382,UR11
CEP955: hs430,UR11
CEP956: hs721,UR11
CEP1009: hs1308,ZNF800

BIC20

FT enhancer:
MECP2

FT promotore:
MAZ,SP4,ZBTB7B

CEP:
CEP138: hs1486,CLMN
CEP288: hs1657,FOS
CEP355: hs1543,HES1
CEP375: hs1886,PTMA
CEP759: hs1971,RBM38
CEP793: hs848,SALL1
CEP952: hs293,UNCX

BIC21

FT enhancer:
MYF6

FT promotore:
KLF15,SP3,ZNF148

CEP:
CEP603: hs699,MKI67
CEP779: hs2061,RREB1
CEP785: hs152,SALL1
CEP789: hs73,SALL1
CEP900: hs1163,TLE1
CEP901: hs1359,TLE1
CEP905: hs974,TLE1

BIC22

FT enhancer:
MYF6,SP3

FT promotore:
SP4,ZBTB7B

CEP:
CEP110: hs1931,CCDC85C
CEP112: hs1920,CCND1
CEP154: hs152,CYLD
CEP326: hs1937,GPC1
CEP378: hs1763,INPP5A
CEP382: hs1466,IRF2BPL
CEP660: hs1548,OLIG2
CEP785: hs152,SALL1
CEP891: hs1887,TEAD3

BIC23

FT enhancer:
MECP2

FT promotore:
SP1,SP3

CEP:
CEP25: hs1912,ARHGFEF16
CEP101: hs2071,CAV2
CEP138: hs1486,CLMN
CEP219: hs1833,ERRFI1
CEP326: hs1937,GPC1
CEP355: hs1543,HES1
CEP388: hs51,IRX3
CEP616: hs1488,MRPL23
CEP679: hs1235,PBX1
CEP738: hs2263,PSRC1
CEP759: hs1971,RBM38
CEP793: hs848,SALL1
CEP816: hs2062,SHB
CEP872: hs1640,SPRY4
CEP952: hs293,UNCX
CEP960: hs1753,VIPR2

BIC24

FT enhancer:
KLF15

FT promotore:
SP4,ZBTB7B

CEP:
CEP154: hs152,CYLD
CEP155: hs162,CYLD
CEP326: hs1937,GPC1
CEP378: hs1763,INPP5A
CEP641: hs1205,NKX2-2
CEP695: hs1958,PFKP
CEP742: hs1955,QKI
CEP785: hs152,SALL1
CEP786: hs162,SALL1
CEP807: hs1362,SETBP1

BIC25

FT enhancer:
SP1,SP4

FT promotore:
NFE2L1

CEP:
CEP346: hs1909,HAAO
CEP459: hs1300,LOC100505619
CEP600: hs359,MIR629
CEP883: hs262,TBCA
CEP891: hs1887,TEAD3
CEP895: hs522,TFCP2L1
CEP947: hs359,UACA

BIC26

FT enhancer:
SNAI1

FT promotore:
SP1,SP3

CEP:
CEP25: hs1912,ARHGFEF16
CEP28: hs1510,ARID3B
CEP73: hs1766,C1orf198
CEP101: hs2071,CAV2
CEP112: hs1920,CCND1
CEP289: hs1717,FOXE1
CEP325: hs1750,GPC1
CEP334: hs1882,GPR20
CEP378: hs1763,INPP5A
CEP403: hs1467,KCNJ2
CEP443: hs2082,LOC100134229
CEP523: hs1945,LRRFIP1
CEP527: hs1397,MALT1
CEP616: hs1488,MRPL23
CEP671: hs1860,PARD6B
CEP943: hs1751,TUBB6

BIC27

FT enhancer:
SOX15

FT promotore:
SP3,SP4

CEP:
CEP58: hs1142,BCL11A
CEP90: hs304,C9orf123
CEP119: hs1634,CDH2
CEP120: hs1636,CDH2
CEP132: hs1636,CHST9
CEP207: hs1025,EMX1
CEP216: hs1472,ERC2
CEP335: hs578,GPR26
CEP353: hs1950,HECTD2
CEP355: hs1543,HES1
CEP407: hs748,KCNMA1
CEP554: hs737,MGMT
CEP626: hs1330,NBEA
CEP660: hs1548,OLIG2
CEP678: hs1202,PBX1
CEP679: hs1235,PBX1
CEP711: hs2096,PLEKHA7
CEP716: hs1081,POU3F2
CEP717: hs1085,POU3F2
CEP719: hs284,POU3F2
CEP741: hs304,PTPRD
CEP786: hs162,SALL1
CEP790: hs74,SALL1
CEP795: hs1020,SALL3
CEP824: hs1394,SLAIN1
CEP847: hs341,SOX21
CEP860: hs1226,SP4
CEP922: hs267,TMEM161B
CEP997: hs1544,ZNF521
CEP1009: hs1308,ZNF800

BIC28

FT enhancer:
PURA,ZNF148

FT promotore:
SP4

CEP:
CEP110: hs1931,CCDC85C
CEP835: hs1507,SLC4A3
CEP890: hs1314,TEAD1
CEP891: hs1887,TEAD3
CEP911: hs1360,TLE4
CEP936: hs1300,TOX3
CEP990: hs1767,ZMIZ1

BIC29

FT enhancer:
PPARA

FT promotore:
SP4,ZBTB7B

CEP:
CEP73: hs1766,C1orf198
CEP110: hs1931,CCDC85C
CEP138: hs1486,CLMN
CEP200: hs1734,EFNA5
CEP302: hs1100,FOXP4
CEP382: hs1466,IRF2BPL
CEP429: hs311,LINC00466
CEP516: hs1447,LPAR1
CEP553: hs699,MGMT
CEP627: hs1333,NBEA
CEP645: hs428,NOVA1
CEP711: hs2096,PLEKHA7
CEP742: hs1955,QKI
CEP807: hs1362,SETBP1
CEP845: hs1496,SOX11
CEP891: hs1887,TEAD3
CEP937: hs161,TOX3
CEP979: hs1800,ZFPM2
CEP1009: hs1308,ZNF800

BIC30

FT enhancer:
NANOG

FT promotore:
SP4,ZBTB7B

CEP:
CEP54: hs658,AUTS2
CEP55: hs925,AUTS2
CEP86: hs533,C5orf38
CEP234: hs526,FAM53A
CEP289: hs1717,FOXE1
CEP325: hs1750,GPC1
CEP355: hs1543,HES1
CEP371: hs1527,ID2
CEP400: hs1430,JAZF1
CEP557: hs1465,MIR148A
CEP371: hs1527,ID2
CEP400: hs1430,JAZF1
CEP557: hs1465,MIR148A
CEP660: hs1548,OLIG2
CEP742: hs1955,QKI
CEP787: hs65,SALL1
CEP790: hs74,SALL1
CEP798: hs634,SALL3
CEP875: hs1487,SSBP3
CEP980: hs195,ZFPM2
CEP990: hs1767,ZMIZ1

BIC31

FT enhancer:
PPARA

FT promotore:
SP3,WT1

CEP:
CEP90: hs304,C9orf123
CEP138: hs1486,CLMN
CEP302: hs1100,FOXP4
CEP332: hs488,GPR180
CEP353: hs1950,HECTD2
CEP553: hs699,MGMT
CEP645: hs428,NOVA1
CEP711: hs2096,PLEKHA7
CEP848: hs488,SOX21
CEP891: hs1887,TEAD3
CEP1009: hs1308,ZNF800

BIC32

FT enhancer:
CXXC1

FT promotore:
SP1,SP3

CEP:

CEP25: hs1912,ARHGEF16
CEP55: hs925,AUTS2
CEP378: hs1763,INPP5A
CEP676: hs1082,PAX6
CEP680: hs762,PBX1
CEP861: hs701,SP4
CEP875: hs1487,SSBP3

BIC33

FT enhancer:
PLAG1

FT promotore:
SP4,ZBTB7B

CEP:
CEP154: hs152,CYLD
CEP236: hs1963,FAM78A
CEP245: hs513,FGF8
CEP785: hs152,SALL1
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3
CEP952: hs293,UNCX

BIC34

FT enhancer:
NANOG

FT promotore:
SP3,WT1

CEP:
CEP51: hs775,ATP9B
CEP54: hs658,AUTS2
CEP55: hs925,AUTS2
CEP332: hs488,GPR180
CEP400: hs1430,JAZF1
CEP557: hs1465,MIR148A
CEP659: hs1188,OLIG2
CEP727: hs2040,PPP3R1
CEP848: hs488,SOX21
CEP862: hs1019,SP8

BIC35

FT enhancer:
SOX13

FT promotore:
SP1,SP3

CEP:
CEP6: hs521,AGAP1
CEP58: hs1142,BCL11A
CEP87: hs603,C5orf38
CEP112: hs1920,CCND1
CEP290: hs1064,FOXG1
CEP316: hs521,GBX2
CEP438: hs186,LMX1B
CEP670: hs2064,PAQR8
CEP676: hs1082,PAX6
CEP716: hs1081,POU3F2
CEP717: hs1085,POU3F2
CEP786: hs162,SALL1
CEP851: hs667,SOX3
CEP862: hs1019,SP8
CEP901: hs1359,TLE1

BIC36

FT enhancer:
SP1

FT promotore:
SP3,ZBTB7B

CEP:
CEP112: hs1920,CCND1
CEP302: hs1100,FOXP4

CEP378: hs1763,INPP5A
CEP382: hs1466,IRF2BPL
CEP643: hs680,NKX6-1
CEP759: hs1971,RBM38
CEP786: hs162,SALL1
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3
CEP936: hs1300,TOX3
CEP990: hs1767,ZMIZ1

BIC37

FT enhancer:
PURA

FT promotore:
SP1,SP3

CEP:
CEP25: hs1912,ARHGEF16
CEP110: hs1931,CCDC85C
CEP334: hs1882,GPR20
CEP615: hs362,MRM1
CEP676: hs1082,PAX6
CEP839: hs660,SMAD6
CEP890: hs1314,TEAD1
CEP891: hs1887,TEAD3
CEP990: hs1767,ZMIZ1

BIC38

FT enhancer:
SNAI1

FT promotore:
SP4,ZBTB7B

CEP:
CEP28: hs1510,ARID3B
CEP73: hs1766,C1orf198
CEP112: hs1920,CCND1
CEP289: hs1717,FOXE1
CEP325: hs1750,GPC1
CEP378: hs1763,INPP5A
CEP523: hs1945,LRRFIP1
CEP549: hs830,MEIS2
CEP739: hs1886,PTMA
CEP796: hs1056,SALL3
CEP797: hs1186,SALL3

BIC39

FT enhancer:
ETV4

FT promotore:
SP1,SP3

CEP:
CEP171: hs1151,DDHD1
CEP189: hs1022,EBF1
CEP228: hs1866,EXOC6
CEP516: hs1447,LPAR1
CEP670: hs2064,PAQR8
CEP679: hs1235,PBX1
CEP717: hs1085,POU3F2
CEP859: hs110,SP4
CEP875: hs1487,SSBP3
CEP904: hs722,TLE1

BIC40

FT enhancer:
MYF6

FT promotore:
SP3,ZFX

CEP:
CEP103: hs124,CBLN1
CEP378: hs1763,INPP5A

CEP423: hs314,LHX2
CEP443:
hs2082,LOC100134229
CEP847: hs341,SOX21
CEP849: hs796,SOX21
CEP890: hs1314,TEAD1

BIC41

FT enhancer:
SOX13,SOX15

FT promotore:
MAZ

CEP:
CEP50: hs1652,ATOH8
CEP84: hs930,C4orf32
CEP239: hs1361,FBXW7
CEP328: hs415,GPD2
CEP424: hs1571,LHX5
CEP508: hs1244,LOC644838
CEP786: hs162,SALL1
CEP809: hs1538,SFTA3
CEP938: hs2073,TPCN2
CEP963: hs1472,WNT5A

BIC42

FT enhancer:
ZBTB7B

FT promotore:
MAZ,ZBTB7B

CEP:
CEP244: hs511,FGF8
CEP382: hs1466,IRF2BPL
CEP424: hs1571,LHX5
CEP660: hs1548,OLIG2
CEP759: hs1971,RBM38
CEP798: hs634,SALL3
CEP971: hs407,ZEB2

BIC43

FT enhancer:
MAZ

FT promotore:
SP4,WT1

CEP:
CEP153: hs260,CXXC4
CEP196: hs632,EBF4
CEP552: hs656,MGMT
CEP641: hs1205,NKX2-2
CEP660: hs1548,OLIG2
CEP891: hs1887,TEAD3
CEP942: hs1967,TTC7A

BIC44

FT enhancer:
MAZ,ZNF148

FT promotore:
PGR

CEP:
CEP3: hs1487,ACOT11
CEP4: hs1854,ACSF3
CEP40: hs1540,ASCL1
CEP68: hs1699,C10orf35
CEP642: hs1205,NKX2-4
CEP643: hs680,NKX6-1
CEP895: hs522,TFCP2L1

BIC45

FT enhancer:
SOX13

FT promotore:
SP4,ZBTB7B

CEP:
CEP58: hs1142,BCL11A
CEP87: hs603,C5orf38
CEP112: hs1920,CCND1
CEP155: hs162,CYLD
CEP290: hs1064,FOXG1
CEP384: hs1327,IRX2
CEP427: hs168,LINC00466
CEP438: hs186,LMX1B
CEP627: hs1333,NBEA
CEP779: hs2061,RREB1
CEP786: hs162,SALL1

BIC46

FT enhancer:
MAZ

FT promotore:
SP4,ZNF148

CEP:
CEP331: hs1665,GPR124
CEP382: hs1466,IRF2BPL
CEP384: hs1327,IRX2
CEP641: hs1205,NKX2-2
CEP660: hs1548,OLIG2
CEP678: hs1202,PBX1
CEP739: hs1886,PTMA
CEP785: hs152,SALL1
CEP875: hs1487,SSBP3
CEP891: hs1887,TEAD3

BIC47

FT enhancer:
SOX13,SOX15

FT promotore:
NFE2L1

CEP:
CEP45: hs1242,ATF7IP2
CEP139: hs1557,CMC1
CEP144: hs909,COX6C
CEP299: hs218,FOXP2
CEP342: hs1242,GRIN2A
CEP406: hs1811,KCNJ6
CEP601: hs1441,MIRLET7C
CEP882: hs1783,TBCA
CEP1011: hs1388,ZSCAN20

BIC48

FT enhancer:
MYF6,ZNF148

FT promotore:
ZNF148

CEP:
CEP25: hs1912,ARHGEF16
CEP326: hs1937,GPC1
CEP334: hs1882,GPR20
CEP382: hs1466,IRF2BPL
CEP603: hs699,MKI67
CEP635: hs1699,NEUROG3
CEP845: hs1496,SOX11
CEP891: hs1887,TEAD3
CEP905: hs974,TLE1
CEP978: hs1909,ZFP36L2

LISTA A19: Enrichment dei geni dei bicluster significativi

È stata effettuata un'analisi di enrichment dei GO-term "biological process" per ogni set di geni associati a ognuno dei 48 bicluster significativi. Per ogni bicluster sono stati utilizzati in input tutti i geni ai quali è stato possibile associare un codice identificativo Uniprot. In totale sono stati analizzati 163 geni, i p-value sono stati corretti con il metodo Bonferroni, ed è stato utilizzato un livello di soglia è pari a un p-value di $5 \cdot 10^{-2}$. 43 bicluster su 48 sono risultati arricchiti per almeno un termine dell'ontologia. Poi sono state calcolate le occorrenze di tutti i termini associati ai bicluster. La lista di sotto mostra tutti i termini arricchiti, uno per riga. Il campo "Bic (#)" e il campo "Bic (%)" mostrano il numero e la percentuale di bicluster di cui i geni sono risultati arricchiti per il termine, "GO-term" e "Descrizione" descrivono il termine in questione, "P-value (min)" e "P-value (max)" mostrano il minimo e il massimo p-value calcolato per i rispettivi termini. Gli enrichment sono stati effettuati avvalendosi di NET-GE. I termini sono disposti in ordine di numero di bicluster decrescente.

Bic (#)	Bic (%)	GO-term	Descrizione	P-value (min)	P-value (max)	Bic (#)	Bic (%)	GO-term	Descrizione	P-value (min)	P-value (max)
22	51,00%	GO:0030182	neuron differentiation	6,23E-08	3,70E-02	12	28,00%	GO:0006366	transcription from RNA polymerase II promoter	2,36E-05	2,69E-02
20	47,00%	GO:0032774	RNA biosynthetic process	5,90E-08	3,44E-02	11	26,00%	GO:0080090	regulation of primary metabolic process	2,23E-04	3,83E-02
19	44,00%	GO:0006357	regulation of transcription from RNA polymerase II promoter	1,29E-07	4,06E-02	11	26,00%	GO:0051960	regulation of nervous system development	7,20E-08	3,63E-02
18	42,00%	GO:0051171	regulation of nitrogen compound metabolic process	4,83E-05	4,22E-02	11	26,00%	GO:0051254	positive regulation of RNA metabolic process	9,88E-05	4,96E-02
18	42,00%	GO:0019219	regulation of nucleobase-containing compound metabolic process	3,98E-05	3,82E-02	11	26,00%	GO:0051173	positive regulation of nitrogen compound metabolic process	1,46E-05	2,35E-02
17	40,00%	GO:0044271	cellular nitrogen compound biosynthetic process	7,32E-07	4,98E-02	11	26,00%	GO:0050767	regulation of neurogenesis	1,40E-06	2,69E-02
17	40,00%	GO:0034654	nucleobase-containing compound biosynthetic process	3,84E-07	3,54E-02	11	26,00%	GO:0045934	negative regulation of nucleobase-containing compound metabolic process	3,81E-07	3,94E-02
17	40,00%	GO:0019438	aromatic compound biosynthetic process	6,01E-07	4,49E-02	11	26,00%	GO:0031325	positive regulation of cellular metabolic process	1,10E-04	4,40E-02
17	40,00%	GO:0018130	heterocycle biosynthetic process	6,21E-07	4,56E-02	11	26,00%	GO:0031323	regulation of cellular metabolic process	1,94E-04	3,62E-02
17	40,00%	GO:0010468	regulation of gene expression	2,26E-05	2,87E-02	10	23,00%	GO:2000113	negative regulation of cellular macromolecule biosynthetic process	3,20E-07	3,43E-02
17	40,00%	GO:0007389	pattern specification process	7,19E-07	4,41E-02	10	23,00%	GO:1902680	positive regulation of RNA biosynthetic process	8,93E-05	4,63E-02
17	40,00%	GO:0006351	transcription, DNA-templated	1,47E-08	4,14E-02	10	23,00%	GO:1902679	negative regulation of RNA biosynthetic process	1,10E-07	3,46E-02
16	37,00%	GO:1901362	organic cyclic compound biosynthetic process	9,77E-07	4,30E-02	10	23,00%	GO:0060284	regulation of cell development	1,28E-05	4,58E-02
16	37,00%	GO:0051252	regulation of RNA metabolic process	6,37E-05	2,53E-02	10	23,00%	GO:0051253	negative regulation of RNA metabolic process	1,41E-07	4,01E-02
16	37,00%	GO:0010604	positive regulation of macromolecule metabolic process	2,80E-05	3,63E-02	10	23,00%	GO:0051172	negative regulation of nitrogen compound metabolic process	5,20E-07	3,63E-02
15	35,00%	GO:2001141	regulation of RNA biosynthetic process	9,00E-05	4,98E-02	10	23,00%	GO:0051094	positive regulation of developmental process	2,56E-05	4,63E-02
15	35,00%	GO:0048522	positive regulation of cellular process	1,76E-05	4,07E-02	10	23,00%	GO:0045893	positive regulation of transcription, DNA-templated	7,40E-05	4,06E-02
15	35,00%	GO:0045944	positive regulation of transcription from RNA polymerase II promoter	3,50E-07	4,04E-02	10	23,00%	GO:0045892	negative regulation of transcription, DNA-templated	9,41E-08	3,14E-02
15	35,00%	GO:0045665	negative regulation of neuron differentiation	1,88E-07	4,43E-02	10	23,00%	GO:0045597	positive regulation of cell differentiation	7,55E-05	4,21E-02
15	35,00%	GO:0045664	regulation of neuron differentiation	6,31E-06	4,52E-02	10	23,00%	GO:0031328	positive regulation of cellular biosynthetic process	3,13E-04	4,93E-02
15	35,00%	GO:0010628	positive regulation of gene expression	2,09E-05	4,53E-02	10	23,00%	GO:0010629	negative regulation of gene expression	2,58E-07	2,87E-02
15	35,00%	GO:0009059	macromolecule biosynthetic process	2,54E-07	4,95E-02	10	23,00%	GO:0010558	negative regulation of macromolecule biosynthetic process	4,81E-07	4,75E-02
15	35,00%	GO:0006355	regulation of transcription, DNA-templated	7,54E-05	4,52E-02	10	23,00%	GO:0010557	positive regulation of macromolecule biosynthetic process	1,92E-04	3,12E-02
15	35,00%	GO:0003002	regionalization	1,96E-08	4,52E-02	9	21,00%	GO:1901576	organic substance biosynthetic process	2,14E-06	3,26E-02
14	33,00%	GO:2000112	regulation of cellular macromolecule biosynthetic process	1,18E-04	4,28E-02	9	21,00%	GO:0048666	neuron development	9,54E-09	4,12E-02
14	33,00%	GO:0016070	RNA metabolic process	3,87E-07	2,62E-02	9	21,00%	GO:0048468	cell development	1,85E-08	3,43E-02
14	33,00%	GO:0009893	positive regulation of metabolic process	9,20E-05	4,91E-02	9	21,00%	GO:0021879	forebrain neuron differentiation	2,76E-03	4,66E-02
14	33,00%	GO:0000122	negative regulation of transcription from RNA polymerase II promoter	1,72E-07	4,70E-02	9	21,00%	GO:0009891	positive regulation of biosynthetic process	3,45E-04	3,05E-02
13	30,00%	GO:0090304	nucleic acid metabolic process	8,12E-06	4,40E-02	9	21,00%	GO:0009887	organ morphogenesis	2,35E-05	4,95E-02
13	30,00%	GO:0060255	regulation of macromolecule metabolic process	4,08E-04	4,91E-02	9	21,00%	GO:0009653	anatomical structure morphogenesis	2,09E-06	4,21E-02
13	30,00%	GO:0048518	positive regulation of biological process	8,29E-05	4,94E-02	9	21,00%	GO:0009058	biosynthetic process	3,03E-06	4,32E-02
13	30,00%	GO:0010556	regulation of macromolecule biosynthetic process	1,83E-04	4,91E-02	8	19,00%	GO:0048869	cellular developmental process	5,09E-06	2,35E-02
12	28,00%	GO:0048731	system development	6,08E-04	4,79E-02	8	19,00%	GO:0048598	embryonic morphogenesis	2,30E-08	2,90E-02
12	28,00%	GO:0045935	positive regulation of nucleobase-containing compound metabolic process	1,18E-05	4,48E-02	8	19,00%	GO:0048562	embryonic organ morphogenesis	1,32E-03	3,27E-02
12	28,00%	GO:0045165	cell fate commitment	4,73E-07	2,45E-02	8	19,00%	GO:0046483	heterocycle metabolic process	1,17E-04	4,92E-02
12	28,00%	GO:0032502	developmental process	3,44E-07	3,27E-02	8	19,00%	GO:0044767	single-organism developmental process	1,03E-06	1,19E-02
12	28,00%	GO:0031326	regulation of cellular biosynthetic process	2,79E-04	2,82E-02	8	19,00%	GO:0044249	cellular biosynthetic process	2,06E-05	2,29E-02
12	28,00%	GO:0030154	cell differentiation	4,07E-07	2,64E-02	8	19,00%	GO:0034645	cellular macromolecule biosynthetic process	7,77E-07	1,89E-02
12	28,00%	GO:0009889	regulation of biosynthetic process	3,18E-04	3,14E-02	8	19,00%	GO:0009888	tissue development	3,55E-05	3,47E-02
						8	19,00%	GO:0006807	nitrogen compound metabolic process	9,82E-05	3,56E-02
						8	19,00%	GO:0006139	nucleobase-containing compound metabolic process	5,90E-05	3,09E-02

Bic (#)	Bic (%)	GO-term	Descrizione	P-value (min)	P-value (max)	Bic (#)	Bic (%)	GO-term	Descrizione	P-value (min)	P-value (max)
7	16,00%	GO:1901360	organic cyclic compound metabolic process	2,40E-04	2,61E-02	4	9,00%	GO:0021884	forebrain neuron development	1,37E-05	4,47E-02
7	16,00%	GO:0045596	negative regulation of cell differentiation	6,65E-04	4,98E-02	4	9,00%	GO:0021675	nerve development	4,38E-03	2,54E-02
7	16,00%	GO:0045168	cell-cell signaling involved in cell fate commitment	5,11E-03	3,92E-02	4	9,00%	GO:0021521	ventral spinal cord interneuron specification	1,51E-03	2,62E-02
7	16,00%	GO:0034641	cellular nitrogen compound metabolic process	1,99E-04	2,23E-02	4	9,00%	GO:0021516	dorsal spinal cord development	4,14E-03	4,75E-02
7	16,00%	GO:0031327	negative regulation of cellular biosynthetic process	4,02E-04	4,48E-02	4	9,00%	GO:0014015	positive regulation of gliogenesis	1,12E-02	2,85E-02
7	16,00%	GO:0030855	epithelial cell differentiation	3,84E-05	4,27E-02	4	9,00%	GO:0014013	regulation of gliogenesis	1,52E-03	3,00E-02
7	16,00%	GO:0009890	negative regulation of biosynthetic process	4,46E-04	4,75E-02	4	9,00%	GO:0009968	negative regulation of signal transduction	4,08E-04	4,48E-02
7	16,00%	GO:0006725	cellular aromatic compound metabolic process	1,22E-04	1,49E-02	4	9,00%	GO:0009953	dorsal/ventral pattern formation	3,10E-11	2,13E-02
7	16,00%	GO:0001568	blood vessel development	1,20E-04	2,44E-02	4	9,00%	GO:0008285	negative regulation of cell proliferation	1,02E-03	5,01E-03
6	14,00%	GO:0060429	epithelium development	5,86E-05	2,74E-02	4	9,00%	GO:0008283	cell proliferation	2,42E-04	4,37E-02
6	14,00%	GO:0060323	head morphogenesis	5,38E-03	1,78E-02	4	9,00%	GO:0007498	mesoderm development	2,92E-04	5,23E-03
6	14,00%	GO:0051093	negative regulation of developmental process	2,84E-03	3,75E-02	4	9,00%	GO:0003281	ventricular septum development	6,81E-03	4,88E-02
6	14,00%	GO:0050793	regulation of developmental process	6,53E-03	3,62E-02	4	9,00%	GO:0003006	developmental process involved in reproduction	2,56E-04	1,73E-02
6	14,00%	GO:0048856	anatomical structure development	3,56E-04	3,57E-02	4	9,00%	GO:0001822	kidney development	7,45E-04	3,40E-02
6	14,00%	GO:0048663	neuron fate commitment	1,37E-03	1,99E-02	4	9,00%	GO:0001501	skeletal system development	5,95E-05	4,51E-02
6	14,00%	GO:0045687	positive regulation of glial cell differentiation	5,10E-03	4,70E-02	3	7,00%	GO:2000026	regulation of multicellular organismal development	1,96E-03	3,50E-02
6	14,00%	GO:0045595	regulation of cell differentiation	6,26E-04	4,06E-02	3	7,00%	GO:0051726	regulation of cell cycle	9,47E-05	4,45E-02
6	14,00%	GO:0021953	central nervous system neuron differentiation	1,46E-04	4,26E-02	3	7,00%	GO:0050794	regulation of cellular process	8,63E-03	2,80E-02
6	14,00%	GO:0021889	olfactory bulb interneuron differentiation	1,25E-03	3,84E-02	3	7,00%	GO:0048732	gland development	6,19E-05	4,43E-02
6	14,00%	GO:0019222	regulation of metabolic process	1,23E-03	3,24E-02	3	7,00%	GO:0045666	positive regulation of neuron differentiation	1,35E-02	2,30E-02
6	14,00%	GO:0010720	positive regulation of cell development	1,01E-03	4,05E-02	3	7,00%	GO:0042127	regulation of cell proliferation	6,88E-03	4,01E-02
6	14,00%	GO:0008284	positive regulation of cell proliferation	9,52E-05	1,88E-02	3	7,00%	GO:0035295	tube development	1,67E-04	1,96E-02
5	12,00%	GO:0060021	palate development	1,66E-05	2,78E-02	3	7,00%	GO:0035108	limb morphogenesis	1,39E-04	2,41E-02
5	12,00%	GO:0050769	positive regulation of neurogenesis	5,05E-05	6,07E-03	3	7,00%	GO:0035107	appendage morphogenesis	1,39E-04	2,41E-02
5	12,00%	GO:0048713	regulation of oligodendrocyte differentiation	1,33E-03	4,30E-02	3	7,00%	GO:0032501	multicellular organismal process	1,25E-03	3,19E-02
5	12,00%	GO:0048523	negative regulation of cellular process	1,51E-05	2,21E-02	3	7,00%	GO:0031324	negative regulation of cellular metabolic process	2,09E-04	6,73E-03
5	12,00%	GO:0048519	negative regulation of biological process	3,80E-05	3,67E-02	3	7,00%	GO:0031016	pancreas development	2,12E-03	4,98E-02
5	12,00%	GO:0045931	positive regulation of mitotic cell cycle	1,85E-03	4,33E-02	3	7,00%	GO:0030856	regulation of epithelial cell differentiation	1,42E-02	1,95E-02
5	12,00%	GO:0044702	single organism reproductive process	8,10E-03	2,93E-02	3	7,00%	GO:0030177	positive regulation of Wnt signaling pathway	9,47E-03	2,84E-02
5	12,00%	GO:0044260	cellular macromolecule metabolic process	2,65E-03	2,62E-02	3	7,00%	GO:0023057	negative regulation of signaling	6,57E-04	4,36E-02
5	12,00%	GO:0035137	hindlimb morphogenesis	1,82E-04	3,21E-02	3	7,00%	GO:0021984	adenohypophysis development	8,04E-03	2,20E-02
5	12,00%	GO:0031128	developmental induction	4,45E-03	3,42E-02	3	7,00%	GO:0021794	thalamus development	7,83E-03	4,11E-02
5	12,00%	GO:0021780	glial cell fate specification	1,78E-03	1,98E-02	3	7,00%	GO:0014070	response to organic cyclic compound	9,91E-04	4,82E-02
5	12,00%	GO:0021778	oligodendrocyte cell fate specification	1,78E-03	1,98E-02	3	7,00%	GO:0010648	negative regulation of cell communication	6,57E-04	4,43E-02
5	12,00%	GO:0021545	cranial nerve development	4,90E-03	3,51E-02	3	7,00%	GO:0010647	positive regulation of cell communication	2,50E-03	2,96E-02
5	12,00%	GO:0021530	spinal cord oligodendrocyte cell fate specification	1,96E-04	6,17E-03	3	7,00%	GO:0010605	negative regulation of macromolecule metabolic process	1,92E-04	6,16E-03
5	12,00%	GO:0021522	spinal cord motor neuron differentiation	5,83E-03	2,76E-02	3	7,00%	GO:0009967	positive regulation of signal transduction	1,49E-03	1,86E-02
5	12,00%	GO:0021515	cell differentiation in spinal cord	9,67E-04	4,43E-02	3	7,00%	GO:0009952	anterior/posterior pattern specification	7,88E-05	1,78E-02
5	12,00%	GO:0010975	regulation of neuron projection development	3,06E-04	1,48E-02	3	7,00%	GO:0009892	negative regulation of metabolic process	3,79E-04	1,26E-02
5	12,00%	GO:0009954	proximal/distal pattern formation	1,02E-03	2,65E-02	3	7,00%	GO:0009267	cellular response to starvation	3,30E-03	4,83E-02
5	12,00%	GO:0007399	nervous system development	3,27E-04	2,49E-02	3	7,00%	GO:0007275	multicellular organismal development	1,86E-02	3,59E-02
5	12,00%	GO:0002009	morphogenesis of an epithelium	8,15E-08	4,92E-02	3	7,00%	GO:0001763	morphogenesis of a branching structure	4,36E-06	3,40E-02
5	12,00%	GO:0001764	neuron migration	7,34E-05	2,78E-02	2	5,00%	GO:2000179	positive regulation of neural precursor cell proliferation	3,10E-04	6,81E-03
5	12,00%	GO:0001708	cell fate specification	4,65E-03	4,76E-02	2	5,00%	GO:2000177	regulation of neural precursor cell proliferation	1,01E-04	5,15E-03
4	9,00%	GO:0060573	cell fate specification involved in pattern specification	1,51E-03	2,62E-02	2	5,00%	GO:1902751	positive regulation of cell cycle G2/M phase transition	1,61E-02	4,45E-02
4	9,00%	GO:0048754	branching morphogenesis of an epithelial tube	8,17E-07	4,21E-02	2	5,00%	GO:0097150	neuronal stem cell maintenance	1,46E-02	2,64E-02
4	9,00%	GO:0048729	tissue morphogenesis	5,64E-09	2,23E-02	2	5,00%	GO:0072164	mesonephric tubule development	1,23E-02	4,60E-02
4	9,00%	GO:0048714	positive regulation of oligodendrocyte differentiation	6,26E-03	2,11E-02	2	5,00%	GO:0072148	epithelial cell fate commitment	3,31E-02	4,45E-02
4	9,00%	GO:0048608	reproductive structure development	2,62E-05	3,53E-02	2	5,00%	GO:0071495	cellular response to endogenous stimulus	3,91E-04	2,31E-03
4	9,00%	GO:0045787	positive regulation of cell cycle	1,30E-02	3,40E-02	2	5,00%	GO:0071363	cellular response to growth factor stimulus	1,99E-05	2,74E-02
4	9,00%	GO:0045685	regulation of glial cell differentiation	3,31E-03	3,92E-02	2	5,00%	GO:0070848	response to growth factor	3,11E-05	3,50E-02
4	9,00%	GO:0040007	growth	7,86E-07	2,15E-02	2	5,00%	GO:0061387	regulation of extent of cell growth	1,71E-03	3,66E-02
4	9,00%	GO:0035148	tube formation	4,44E-03	2,41E-02	2	5,00%	GO:0061138	morphogenesis of a branching epithelium	2,17E-06	2,29E-02
4	9,00%	GO:0031344	regulation of cell projection organization	1,63E-03	4,92E-02	2	5,00%	GO:0060425	lung morphogenesis	3,18E-02	4,69E-02
4	9,00%	GO:0030111	regulation of Wnt signaling pathway	1,65E-02	2,89E-02	2	5,00%	GO:0060412	ventricular septum morphogenesis	2,46E-02	3,15E-02

Bic (#)	Bic (%)	GO-term	Descrizione	P-value (min)	P-value (max)	Bic (#)	Bic (%)	GO-term	Descrizione	P-value (min)	P-value (max)
2	5,00%	GO:0060411	cardiac septum morphogenesis	4,38E-04	9,57E-03	1	2,00%	GO:1901342	regulation of vasculature development	1,31E-02	1,31E-02
2	5,00%	GO:0051090	regulation of sequence-specific DNA binding transcription factor activity	3,06E-03	3,66E-02	1	2,00%	GO:0097154	GABAergic neuron differentiation	2,22E-02	2,22E-02
2	5,00%	GO:0050789	regulation of biological process	2,77E-02	4,78E-02	1	2,00%	GO:0090103	cochlea morphogenesis	4,47E-03	4,47E-03
2	5,00%	GO:0050770	regulation of axonogenesis	1,73E-03	2,71E-02	1	2,00%	GO:0080134	regulation of response to stress	3,53E-02	3,53E-02
2	5,00%	GO:0050768	negative regulation of neurogenesis	4,04E-02	4,80E-02	1	2,00%	GO:0072395	signal transduction involved in cell cycle checkpoint	4,94E-02	4,94E-02
2	5,00%	GO:0050678	regulation of epithelial cell proliferation	2,81E-02	4,99E-02	1	2,00%	GO:0072163	mesonephric epithelium development	1,36E-02	1,36E-02
2	5,00%	GO:0048863	stem cell differentiation	2,02E-03	2,44E-02	1	2,00%	GO:0072095	regulation of branch elongation involved in ureteric bud branching	5,01E-03	5,01E-03
2	5,00%	GO:0048711	positive regulation of astrocyte differentiation	1,29E-02	2,10E-02	1	2,00%	GO:0072091	regulation of stem cell proliferation	1,08E-02	1,08E-02
2	5,00%	GO:0048646	anatomical structure formation involved in morphogenesis	5,99E-03	2,67E-02	1	2,00%	GO:0072089	stem cell proliferation	7,12E-03	7,12E-03
2	5,00%	GO:0048638	regulation of developmental growth	3,96E-03	4,01E-02	1	2,00%	GO:0072073	kidney epithelium development	3,87E-02	3,87E-02
2	5,00%	GO:0048585	negative regulation of response to stimulus	2,44E-04	2,45E-03	1	2,00%	GO:0071504	cellular response to heparin	1,39E-02	1,39E-02
2	5,00%	GO:0048568	embryonic organ development	4,93E-04	3,06E-02	1	2,00%	GO:0071503	response to heparin	2,14E-02	2,14E-02
2	5,00%	GO:0048538	thymus development	1,31E-02	4,87E-02	1	2,00%	GO:0071496	cellular response to external stimulus	3,50E-02	3,50E-02
2	5,00%	GO:0044707	single-multicellular organism process	1,06E-03	2,65E-02	1	2,00%	GO:0071345	cellular response to cytokine stimulus	3,91E-02	3,91E-02
2	5,00%	GO:0043433	negative regulation of sequence-specific DNA binding transcription factor activity	7,55E-04	5,36E-03	1	2,00%	GO:0071000	response to magnetism	1,85E-03	1,85E-03
2	5,00%	GO:0035329	hippo signaling	1,15E-02	1,30E-02	1	2,00%	GO:0061351	neural precursor cell proliferation	2,98E-03	2,98E-03
2	5,00%	GO:0035136	forelimb morphogenesis	9,61E-04	1,04E-02	1	2,00%	GO:0061004	pattern specification involved in kidney development	4,66E-02	4,66E-02
2	5,00%	GO:0033993	response to lipid	3,55E-02	4,33E-02	1	2,00%	GO:0060606	tube closure	6,83E-03	6,83E-03
2	5,00%	GO:0032989	cellular component morphogenesis	1,80E-04	2,51E-04	1	2,00%	GO:0060485	mesenchyme development	2,74E-02	2,74E-02
2	5,00%	GO:0031346	positive regulation of cell projection organization	9,76E-04	2,59E-03	1	2,00%	GO:0060415	muscle tissue morphogenesis	3,88E-02	3,88E-02
2	5,00%	GO:0031018	endocrine pancreas development	6,79E-03	4,59E-02	1	2,00%	GO:0060253	negative regulation of glial cell proliferation	1,46E-02	1,46E-02
2	5,00%	GO:0030857	negative regulation of epithelial cell differentiation	2,53E-02	3,67E-02	1	2,00%	GO:0060251	regulation of glial cell proliferation	4,56E-02	4,56E-02
2	5,00%	GO:0030516	regulation of axon extension	5,70E-04	1,61E-02	1	2,00%	GO:0060174	limb bud formation	4,01E-02	4,01E-02
2	5,00%	GO:0030325	adrenal gland development	7,98E-03	2,13E-02	1	2,00%	GO:0060164	regulation of timing of neuron differentiation	4,80E-02	4,80E-02
2	5,00%	GO:0022604	regulation of cell morphogenesis	1,55E-06	4,57E-02	1	2,00%	GO:0060023	soft palate development	7,32E-03	7,32E-03
2	5,00%	GO:0022414	reproductive process	3,12E-02	4,74E-02	1	2,00%	GO:0051593	response to folic acid	2,96E-03	2,96E-03
2	5,00%	GO:0021983	pituitary gland development	8,03E-03	3,30E-02	1	2,00%	GO:0051301	cell division	1,48E-02	1,48E-02
2	5,00%	GO:0021954	central nervous system neuron development	5,49E-07	1,52E-02	1	2,00%	GO:0051240	positive regulation of multicellular organismal process	5,76E-04	5,76E-04
2	5,00%	GO:0021798	forebrain dorsal/ventral pattern formation	3,00E-02	4,12E-02	1	2,00%	GO:0051239	regulation of multicellular organismal process	3,37E-02	3,37E-02
2	5,00%	GO:0021782	glial cell development	1,82E-03	1,04E-02	1	2,00%	GO:0050877	neurological system process	4,90E-02	4,90E-02
2	5,00%	GO:0021555	midbrain-hindbrain boundary morphogenesis	1,65E-02	3,62E-02	1	2,00%	GO:0050679	positive regulation of epithelial cell proliferation	3,24E-02	3,24E-02
2	5,00%	GO:0016477	cell migration	9,28E-05	3,95E-02	1	2,00%	GO:0048870	cell motility	1,98E-04	1,98E-04
2	5,00%	GO:0014003	oligodendrocyte development	9,99E-03	2,89E-02	1	2,00%	GO:0048793	pronephros development	3,60E-02	3,60E-02
2	5,00%	GO:0010971	positive regulation of G2/M transition of mitotic cell cycle	1,61E-02	4,45E-02	1	2,00%	GO:0048715	negative regulation of oligodendrocyte differentiation	2,09E-02	2,09E-02
2	5,00%	GO:0010721	negative regulation of cell development	1,20E-02	3,58E-02	1	2,00%	GO:0048710	regulation of astrocyte differentiation	3,99E-02	3,99E-02
2	5,00%	GO:0009880	embryonic pattern specification	5,83E-04	7,24E-04	1	2,00%	GO:0048709	oligodendrocyte differentiation	4,57E-02	4,57E-02
2	5,00%	GO:0009790	embryo development	4,26E-03	4,25E-02	1	2,00%	GO:0048706	embryonic skeletal system development	1,31E-02	1,31E-02
2	5,00%	GO:0008589	regulation of smoothened signaling pathway	2,19E-05	3,33E-02	1	2,00%	GO:0048705	skeletal system morphogenesis	2,85E-02	2,85E-02
2	5,00%	GO:0007507	heart development	4,45E-03	5,13E-03	1	2,00%	GO:0048665	neuron fate specification	7,48E-03	7,48E-03
2	5,00%	GO:0007423	sensory organ development	1,15E-05	1,60E-02	1	2,00%	GO:0048664	neuron fate determination	3,73E-02	3,73E-02
2	5,00%	GO:0007346	regulation of mitotic cell cycle	1,42E-03	3,84E-02	1	2,00%	GO:0048589	developmental growth	8,88E-03	8,88E-03
2	5,00%	GO:0003323	type B pancreatic cell development	2,07E-02	2,79E-02	1	2,00%	GO:0048565	digestive tract development	3,63E-02	3,63E-02
2	5,00%	GO:0003279	cardiac septum development	8,21E-03	1,40E-02	1	2,00%	GO:0048513	organ development	1,32E-02	1,32E-02
2	5,00%	GO:0001823	mesonephros development	3,48E-02	4,62E-02	1	2,00%	GO:0048483	autonomic nervous system development	2,63E-02	2,63E-02
2	5,00%	GO:0001658	branching involved in ureteric bud morphogenesis	2,98E-04	3,09E-02	1	2,00%	GO:0048469	cell maturation	2,00E-02	2,00E-02
2	5,00%	GO:0001657	ureteric bud development	8,25E-03	3,08E-02	1	2,00%	GO:0048286	lung alveolus development	4,14E-02	4,14E-02
2	5,00%	GO:0001570	vasculogenesis	2,35E-05	3,76E-02	1	2,00%	GO:0045937	positive regulation of phosphate metabolic process	2,92E-02	2,92E-02
1	2,00%	GO:2000648	positive regulation of stem cell proliferation	3,03E-03	3,03E-03	1	2,00%	GO:0045927	positive regulation of growth	1,18E-02	1,18E-02
1	2,00%	GO:2000027	regulation of organ morphogenesis	2,90E-03	2,90E-03	1	2,00%	GO:0045683	negative regulation of epidermis development	4,81E-02	4,81E-02
1	2,00%	GO:1902749	regulation of cell cycle G2/M phase transition	2,10E-02	2,10E-02	1	2,00%	GO:0045605	negative regulation of epidermal cell differentiation	2,35E-02	2,35E-02
1	2,00%	GO:1902692	regulation of neuroblast proliferation	2,38E-02	2,38E-02	1	2,00%	GO:0045598	regulation of fat cell differentiation	1,05E-02	1,05E-02
1	2,00%	GO:1901992	positive regulation of mitotic cell cycle phase transition	2,82E-02	2,82E-02	1	2,00%	GO:0044273	sulfur compound catabolic process	3,69E-02	3,69E-02
1	2,00%	GO:1901989	positive regulation of cell cycle phase transition	3,12E-02	3,12E-02	1	2,00%	GO:0044255	cellular lipid metabolic process	2,09E-03	2,09E-03
						1	2,00%	GO:0044092	negative regulation of molecular function	4,61E-03	4,61E-03

Bic (#)	Bic (%)	GO-term	Descrizione	P-value (min)	P-value (max)	Bic (#)	Bic (%)	GO-term	Descrizione	P-value (min)	P-value (max)
1	2,00%	GO:0044087	regulation of cellular component biogenesis	1,05E-02	1,05E-02	1	2,00%	GO:0010389	regulation of G2/M transition of mitotic cell cycle	1,94E-02	1,94E-02
1	2,00%	GO:0043436	oxoacid metabolic process	4,86E-02	4,86E-02	1	2,00%	GO:0009991	response to extracellular stimulus	9,20E-03	9,20E-03
1	2,00%	GO:0043170	macromolecule metabolic process	1,69E-02	1,69E-02	1	2,00%	GO:0009966	regulation of signal transduction	1,62E-02	1,62E-02
1	2,00%	GO:0043068	positive regulation of programmed cell death	1,05E-04	1,05E-04	1	2,00%	GO:0009950	dorsal/ventral axis specification	4,12E-02	4,12E-02
1	2,00%	GO:0043065	positive regulation of apoptotic process	9,94E-05	9,94E-05	1	2,00%	GO:0009798	axis specification	4,11E-03	4,11E-03
1	2,00%	GO:0043009	chordate embryonic development	1,06E-03	1,06E-03	1	2,00%	GO:0009792	embryo development ending in birth or egg hatching	1,26E-03	1,26E-03
1	2,00%	GO:0042594	response to starvation	3,47E-02	3,47E-02	1	2,00%	GO:0009725	response to hormone	1,04E-02	1,04E-02
1	2,00%	GO:0042476	odontogenesis	1,69E-02	1,69E-02	1	2,00%	GO:0008593	regulation of Notch signaling pathway	3,88E-03	3,88E-03
1	2,00%	GO:0042472	inner ear morphogenesis	1,77E-03	1,77E-03	1	2,00%	GO:0008406	gonad development	2,78E-02	2,78E-02
1	2,00%	GO:0042327	positive regulation of phosphorylation	1,47E-02	1,47E-02	1	2,00%	GO:0007565	female pregnancy	2,54E-02	2,54E-02
1	2,00%	GO:0040011	locomotion	5,24E-04	5,24E-04	1	2,00%	GO:0007420	brain development	4,67E-04	4,67E-04
1	2,00%	GO:0040008	regulation of growth	1,60E-02	1,60E-02	1	2,00%	GO:0007417	central nervous system development	3,97E-05	3,97E-05
1	2,00%	GO:0035883	enteroendocrine cell differentiation	2,28E-02	2,28E-02	1	2,00%	GO:0007154	cell communication	1,65E-02	1,65E-02
1	2,00%	GO:0035502	metanephric part of ureteric bud development	1,25E-02	1,25E-02	1	2,00%	GO:0006629	lipid metabolic process	8,89E-03	8,89E-03
1	2,00%	GO:0035113	embryonic appendage morphogenesis	4,29E-03	4,29E-03	1	2,00%	GO:0006461	protein complex assembly	2,16E-02	2,16E-02
1	2,00%	GO:0034504	protein localization to nucleus	6,95E-03	6,95E-03	1	2,00%	GO:0003156	regulation of organ formation	1,05E-02	1,05E-02
1	2,00%	GO:0034097	response to cytokine	1,40E-02	1,40E-02	1	2,00%	GO:0003151	outflow tract morphogenesis	2,81E-02	2,81E-02
1	2,00%	GO:0033273	response to vitamin	4,68E-02	4,68E-02	1	2,00%	GO:0002070	epithelial cell maturation	7,88E-03	7,88E-03
1	2,00%	GO:0032870	cellular response to hormone stimulus	3,16E-02	3,16E-02	1	2,00%	GO:0002068	glandular epithelial cell development	4,60E-02	4,60E-02
1	2,00%	GO:0032808	lacrimal gland development	3,40E-02	3,40E-02	1	2,00%	GO:0002064	epithelial cell development	3,51E-04	3,51E-04
1	2,00%	GO:0031669	cellular response to nutrient levels	2,53E-02	2,53E-02	1	2,00%	GO:0002052	positive regulation of neuroblast proliferation	3,19E-02	3,19E-02
1	2,00%	GO:0031667	response to nutrient levels	5,31E-03	5,31E-03	1	2,00%	GO:0001843	neural tube closure	6,53E-03	6,53E-03
1	2,00%	GO:0031401	positive regulation of protein modification process	2,59E-02	2,59E-02	1	2,00%	GO:0001709	cell fate determination	4,16E-02	4,16E-02
1	2,00%	GO:0030901	midbrain development	1,62E-02	1,62E-02	1	2,00%	GO:0001701	in utero embryonic development	9,51E-04	9,51E-04
1	2,00%	GO:0030539	male genitalia development	2,79E-02	2,79E-02	1	2,00%	GO:0001654	eye development	8,45E-03	8,45E-03
1	2,00%	GO:0030514	negative regulation of BMP signaling pathway	1,65E-02	1,65E-02	1	2,00%	GO:0001569	patterning of blood vessels	1,93E-04	1,93E-04
1	2,00%	GO:0030326	embryonic limb morphogenesis	4,29E-03	4,29E-03	1	2,00%	GO:0000902	cell morphogenesis	1,54E-02	1,54E-02
1	2,00%	GO:0030099	myeloid cell differentiation	1,04E-02	1,04E-02						
1	2,00%	GO:0023051	regulation of signaling	4,14E-02	4,14E-02						
1	2,00%	GO:0022603	regulation of anatomical structure morphogenesis	2,13E-04	2,13E-04						
1	2,00%	GO:0022412	cellular process involved in reproduction in multicellular organism	1,82E-03	1,82E-03						
1	2,00%	GO:0021930	cerebellar granule cell precursor proliferation	1,54E-02	1,54E-02						
1	2,00%	GO:0021924	cell proliferation in external granule layer	1,54E-02	1,54E-02						
1	2,00%	GO:0021846	cell proliferation in forebrain	4,39E-05	4,39E-05						
1	2,00%	GO:0021799	cerebral cortex radially oriented cell migration	2,41E-02	2,41E-02						
1	2,00%	GO:0021766	hippocampus development	3,30E-02	3,30E-02						
1	2,00%	GO:0021758	putamen development	4,64E-02	4,64E-02						
1	2,00%	GO:0021757	caudate nucleus development	4,64E-02	4,64E-02						
1	2,00%	GO:0021750	vestibular nucleus development	3,71E-03	3,71E-03						
1	2,00%	GO:0021700	developmental maturation	1,81E-03	1,81E-03						
1	2,00%	GO:0021549	cerebellum development	2,38E-02	2,38E-02						
1	2,00%	GO:0021546	rhombomere development	8,00E-05	8,00E-05						
1	2,00%	GO:0021534	cell proliferation in hindbrain	3,02E-02	3,02E-02						
1	2,00%	GO:0021510	spinal cord development	1,03E-02	1,03E-02						
1	2,00%	GO:0019752	carboxylic acid metabolic process	2,77E-02	2,77E-02						
1	2,00%	GO:0019230	proprioception	3,62E-02	3,62E-02						
1	2,00%	GO:0017015	regulation of transforming growth factor beta receptor signaling pathway	4,99E-02	4,99E-02						
1	2,00%	GO:0010942	positive regulation of cell death	2,02E-04	2,02E-04						
1	2,00%	GO:0010941	regulation of cell death	3,55E-02	3,55E-02						
1	2,00%	GO:0010830	regulation of myotube differentiation	3,71E-02	3,71E-02						
1	2,00%	GO:0010769	regulation of cell morphogenesis involved in differentiation	3,04E-06	3,04E-06						
1	2,00%	GO:0010646	regulation of cell communication	4,19E-02	4,19E-02						
1	2,00%	GO:0010562	positive regulation of phosphorus metabolic process	2,92E-02	2,92E-02						

LISTA A20: Enrichment dei geni del bicluster 5

Enrichment dei GO-term (biological process) degli 11 geni i cui promotori erano presenti nelle CEP del Bic 5. I p-value sono corretti con il metodo Benjamini-Hockberg, il livello di significatività di soglia usato è stato pari a un p-value di $5 \cdot 10^{-2}$. A sinistra sono elencati i GO-term, seguiti da una loro descrizione. Il campo "Geni" indica il numero di geni annotati per il termine corrispondente. "Classe" indica il numero totale di geni annotati per lo stesso termine nell'intero database di riferimento (NET-GE). I termini sono ordinati in ordine di numero di geni arricchiti decrescente. La lista non include tutti i termini significativi e sono stati selezionati quelli relativi al maggior numero di geni.

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0050794	regulation of cellular process	11	17465	5,76E-03	GO:0044271	cellular nitrogen compound biosynthetic process	6	3991	7,50E-03
GO:0050789	regulation of biological process	11	18486	7,61E-03	GO:0018130	heterocycle biosynthetic process	6	3931	7,60E-03
GO:0065007	biological regulation	11	19296	8,33E-03	GO:0019438	aromatic compound biosynthetic process	6	3919	7,67E-03
GO:0048522	positive regulation of cellular process	10	5768	5,25E-05	GO:0034654	nucleobase-containing compound biosynthetic process	6	3760	7,73E-03
GO:0048518	positive regulation of biological process	10	6624	1,02E-04	GO:1901362	organic cyclic compound biosynthetic process	6	4099	8,12E-03
GO:0019219	regulation of nucleobase-containing compound metabolic process	9	7797	1,11E-03	GO:0009059	macromolecule biosynthetic process	6	4344	8,41E-03
GO:0051171	regulation of nitrogen compound metabolic process	9	7927	1,21E-03	GO:2000113	negative regulation of cellular macromolecule biosynthetic process	6	3633	1,26E-02
GO:0031323	regulation of cellular metabolic process	9	9728	4,84E-03	GO:0045934	negative regulation of nucleobase-containing compound metabolic process	6	3691	1,30E-02
GO:0080090	regulation of primary metabolic process	9	9808	5,00E-03	GO:0051253	negative regulation of RNA metabolic process	6	3369	1,32E-02
GO:0019222	regulation of metabolic process	9	10848	7,38E-03	GO:0010629	negative regulation of gene expression	6	3561	1,35E-02
GO:0051252	regulation of RNA metabolic process	8	6360	2,40E-03	GO:0010558	negative regulation of macromolecule biosynthetic process	6	3771	1,35E-02
GO:0010468	regulation of gene expression	8	7158	4,90E-03	GO:0051172	negative regulation of nitrogen compound metabolic process	6	3798	1,37E-02
GO:0060255	regulation of macromolecule metabolic process	8	8942	9,81E-03	GO:0060341	regulation of cellular localization	6	3341	1,39E-02
GO:0051716	cellular response to stimulus	8	9450	9,83E-03	GO:1902679	negative regulation of RNA biosynthetic process	6	3295	1,47E-02
GO:0050896	response to stimulus	8	11721	2,04E-02	GO:0045892	negative regulation of transcription, DNA-templated	6	3247	1,54E-02
GO:0006357	regulation of transcription from RNA polymerase II promoter	7	2379	2,07E-04	GO:0044249	cellular biosynthetic process	6	5837	1,58E-02
GO:0031325	positive regulation of cellular metabolic process	7	3418	7,29E-04	GO:1901576	organic substance biosynthetic process	6	6037	1,71E-02
GO:0010604	positive regulation of macromolecule metabolic process	7	3285	8,08E-04	GO:0009058	biosynthetic process	6	6217	1,85E-02
GO:0009893	positive regulation of metabolic process	7	3630	8,84E-04	GO:0009653	anatomical structure morphogenesis	6	4591	1,93E-02
GO:0016070	RNA metabolic process	7	4467	2,31E-03	GO:0032502	developmental process	6	7299	2,96E-02
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	7	2989	4,00E-03	GO:0030154	cell differentiation	5	2446	7,59E-03
GO:0090304	nucleic acid metabolic process	7	5696	7,24E-03	GO:0006351	transcription, DNA-templated	5	2949	8,69E-03
GO:2001141	regulation of RNA biosynthetic process	7	6249	8,20E-03	GO:0006366	transcription from RNA polymerase II promoter	5	1448	9,38E-03
GO:0006355	regulation of transcription, DNA-templated	7	6146	9,26E-03	GO:0050793	regulation of developmental process	5	2884	9,71E-03
GO:0010556	regulation of macromolecule biosynthetic process	7	6784	9,43E-03	GO:0071345	cellular response to cytokine stimulus	5	1640	1,10E-02
GO:0031326	regulation of cellular biosynthetic process	7	6987	9,75E-03	GO:0000122	negative regulation of transcription from RNA polymerase II promoter	5	2314	1,29E-02
GO:2000112	regulation of cellular macromolecule biosynthetic process	7	6579	9,84E-03	GO:0050767	regulation of neurogenesis	5	2377	1,35E-02
GO:0009889	regulation of biosynthetic process	7	7051	1,00E-02	GO:0045664	regulation of neuron differentiation	5	1894	1,37E-02
GO:0006139	nucleobase-containing compound metabolic process	7	7311	1,16E-02	GO:0048869	cellular developmental process	5	3694	1,41E-02
GO:0046483	heterocycle metabolic process	7	7642	1,30E-02	GO:0034097	response to cytokine	5	2017	1,50E-02
GO:0006725	cellular aromatic compound metabolic process	7	7662	1,31E-02	GO:0051960	regulation of nervous system development	5	2584	1,63E-02
GO:0007165	signal transduction	7	7592	1,33E-02					
GO:0034641	cellular nitrogen compound metabolic process	7	7912	1,47E-02					
GO:1901360	organic cyclic compound metabolic process	7	8010	1,50E-02					
GO:0006807	nitrogen compound metabolic process	7	8714	1,88E-02					
GO:0044260	cellular macromolecule metabolic process	7	11069	4,37E-02					
GO:0051254	positive regulation of RNA metabolic process	6	1838	5,67E-04					
GO:1902680	positive regulation of RNA biosynthetic process	6	1811	6,07E-04					
GO:0045893	positive regulation of transcription, DNA-templated	6	1762	6,22E-04					
GO:0010628	positive regulation of gene expression	6	1919	6,37E-04					
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	6	2060	6,98E-04					
GO:0010557	positive regulation of macromolecule biosynthetic process	6	2042	7,30E-04					
GO:0051173	positive regulation of nitrogen compound metabolic process	6	2108	7,30E-04					
GO:0031328	positive regulation of cellular biosynthetic process	6	2177	7,53E-04					
GO:0009891	positive regulation of biosynthetic process	6	2208	7,63E-04					
GO:0032774	RNA biosynthetic process	6	3250	4,11E-03					

LISTA A21: Enrichment dei geni della CFT 171

Enrichment dei GO-term (biological process) dei 62 geni associabili alla CFT 171 e di cui è stato possibile rinvenire i codici identificativi Uniprot. I p-value sono corretti con il metodo Bonferroni, il livello di significatività di soglia usato è stato pari a un p-value di $5 \cdot 10^{-2}$. A sinistra sono elencati i GO-term, seguiti da una loro descrizione. Il campo "Geni" indica il numero di geni annotati per il termine corrispondente. "Classe" indica il numero totale di geni annotati per lo stesso termine nell'intero database di riferimento (NET-GE). I termini sono ordinati in ordine di p-value crescente.

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0030182	neuron differentiation	19	911	8,78E-13	GO:0044767	single-organism developmental process	30	6740	6,67E-06
GO:0031326	regulation of cellular biosynthetic process	36	6987	3,80E-10	GO:0051960	regulation of nervous system development	20	2584	9,33E-06
GO:0009889	regulation of biosynthetic process	36	7051	5,06E-10	GO:0032502	developmental process	31	7299	9,55E-06
GO:0010468	regulation of gene expression	36	7158	8,09E-10	GO:0045935	positive regulation of nucleobase-containing compound metabolic process	17	2060	9,67E-06
GO:0019219	regulation of nucleobase-containing compound metabolic process	37	7797	1,72E-09	GO:0006139	nucleobase-containing compound metabolic process	31	7311	9,95E-06
GO:0051171	regulation of nitrogen compound metabolic process	37	7927	2,91E-09	GO:0030855	epithelial cell differentiation	14	1135	1,13E-05
GO:0032774	RNA biosynthetic process	25	3250	2,97E-09	GO:1901576	organic substance biosynthetic process	28	6037	1,27E-05
GO:2000112	regulation of cellular macromolecule biosynthetic process	34	6579	3,03E-09	GO:0051173	positive regulation of nitrogen compound metabolic process	17	2108	1,36E-05
GO:0048869	cellular developmental process	26	3694	6,66E-09	GO:0048519	negative regulation of biological process	27	5756	2,19E-05
GO:0009887	organ morphogenesis	20	1719	6,83E-09	GO:0045664	regulation of neuron differentiation	17	1894	2,20E-05
GO:0010556	regulation of macromolecule biosynthetic process	34	6784	7,40E-09	GO:0009058	biosynthetic process	28	6217	2,45E-05
GO:0051252	regulation of RNA metabolic process	33	6360	7,65E-09	GO:0021953	central nervous system neuron differentiation	9	365	2,58E-05
GO:0006355	regulation of transcription, DNA-templated	32	6146	1,92E-08	GO:0045165	cell fate commitment	10	501	2,74E-05
GO:0006351	transcription, DNA-templated	23	2949	2,47E-08	GO:0007275	multicellular organismal development	16	1673	2,75E-05
GO:2001141	regulation of RNA biosynthetic process	32	6249	3,02E-08	GO:0044249	cellular biosynthetic process	27	5837	2,96E-05
GO:0009059	macromolecule biosynthetic process	27	4344	3,84E-08	GO:0046483	heterocycle metabolic process	31	7642	2,97E-05
GO:0030154	cell differentiation	21	2446	4,34E-08	GO:0006725	cellular aromatic compound metabolic process	31	7662	3,17E-05
GO:0031323	regulation of cellular metabolic process	39	9728	5,40E-08	GO:0009888	tissue development	19	2570	5,46E-05
GO:0080090	regulation of primary metabolic process	39	9808	7,06E-08	GO:0031327	negative regulation of cellular biosynthetic process	15	1764	6,11E-05
GO:0034654	nucleobase-containing compound biosynthetic process	25	3760	7,21E-08	GO:0000122	negative regulation of transcription from RNA polymerase II promoter	18	2314	6,60E-05
GO:0016070	RNA metabolic process	27	4467	7,34E-08	GO:0034641	cellular nitrogen compound metabolic process	31	7912	6,93E-05
GO:0048598	embryonic morphogenesis	18	1566	1,36E-07	GO:0009890	negative regulation of biosynthetic process	15	1786	7,19E-05
GO:0019438	aromatic compound biosynthetic process	25	3919	1,76E-07	GO:0003006	developmental process involved in reproduction	18	2335	7,58E-05
GO:0018130	heterocycle biosynthetic process	25	3931	1,88E-07	GO:1901360	organic cyclic compound metabolic process	31	8010	9,33E-05
GO:0044271	cellular nitrogen compound biosynthetic process	25	3991	2,60E-07	GO:0050767	regulation of neurogenesis	18	2377	9,96E-05
GO:0019222	regulation of metabolic process	40	10848	3,41E-07	GO:0060284	regulation of cell development	20	2979	1,03E-04
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	23	2989	4,34E-07	GO:0048729	tissue morphogenesis	14	1359	1,08E-04
GO:1901362	organic cyclic compound biosynthetic process	25	4099	4,59E-07	GO:0006807	nitrogen compound metabolic process	32	8714	1,71E-04
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	25	3633	5,30E-07	GO:0010628	positive regulation of gene expression	15	1919	1,83E-04
GO:0090304	nucleic acid metabolic process	29	5696	6,07E-07	GO:0048856	anatomical structure development	22	4289	1,94E-04
GO:0060255	regulation of macromolecule metabolic process	36	8942	6,97E-07	GO:0003002	regionalization	12	988	1,98E-04
GO:0045934	negative regulation of nucleobase-containing compound metabolic process	25	3691	7,42E-07	GO:0045597	positive regulation of cell differentiation	18	2514	2,33E-04
GO:0048468	cell development	20	2280	1,07E-06	GO:0060429	epithelium development	12	1006	2,40E-04
GO:0001501	skeletal system development	12	620	1,13E-06	GO:0009893	positive regulation of metabolic process	20	3630	2,86E-04
GO:0010558	negative regulation of macromolecule biosynthetic process	25	3771	1,17E-06	GO:0010604	positive regulation of macromolecule metabolic process	19	3285	3,03E-04
GO:0051172	negative regulation of nitrogen compound metabolic process	25	3798	1,36E-06	GO:0044260	cellular macromolecule metabolic process	36	11069	3,05E-04
GO:0048731	system development	21	2612	1,65E-06	GO:0007389	pattern specification process	14	1481	3,12E-04
GO:0006357	regulation of transcription from RNA polymerase II promoter	19	2379	1,67E-06	GO:0048522	positive regulation of cellular process	25	5768	4,88E-04
GO:0051094	positive regulation of developmental process	23	3200	1,67E-06	GO:0001763	morphogenesis of a branching structure	10	690	5,40E-04
GO:0034645	cellular macromolecule biosynthetic process	24	4013	1,84E-06	GO:0009954	proximal/distal pattern formation	5	79	5,53E-04
GO:0010629	negative regulation of gene expression	24	3561	2,22E-06	GO:0048666	neuron development	9	539	7,08E-04
GO:0045892	negative regulation of transcription, DNA-templated	23	3247	2,23E-06	GO:0051254	positive regulation of RNA metabolic process	14	1838	7,11E-04
GO:0009653	anatomical structure morphogenesis	27	4591	2,40E-06	GO:0010605	negative regulation of macromolecule metabolic process	16	2452	7,36E-04
GO:1902679	negative regulation of RNA biosynthetic process	23	3295	2,97E-06	GO:0031324	negative regulation of cellular metabolic process	16	2479	8,52E-04
GO:0048523	negative regulation of cellular process	27	5279	3,26E-06	GO:0031328	positive regulation of cellular biosynthetic process	15	2177	9,16E-04
GO:0006366	transcription from RNA polymerase II promoter	16	1448	3,53E-06	GO:0009891	positive regulation of biosynthetic process	15	2208	1,09E-03
GO:0051253	negative regulation of RNA metabolic process	23	3369	4,57E-06	GO:2000026	regulation of multicellular organismal development	15	2229	1,23E-03

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0045596	negative regulation of cell differentiation	16	2226	1,38E-03	GO:0010769	regulation of cell morphogenesis involved in differentiation	10	1115	3,86E-02
GO:0009952	anterior/posterior pattern specification	9	585	1,40E-03	GO:0001822	kidney development	7	489	4,35E-02
GO:0050794	regulation of cellular process	45	17465	1,45E-03	GO:0014706	striated muscle tissue development	6	333	4,97E-02
GO:0048518	positive regulation of biological process	26	6624	1,77E-03					
GO:0021954	central nervous system neuron development	5	102	1,98E-03					
GO:0009892	negative regulation of metabolic process	16	2679	2,37E-03					
GO:0040007	growth	13	1496	2,44E-03					
GO:0010557	positive regulation of macromolecule biosynthetic process	14	2042	2,48E-03					
GO:0050789	regulation of biological process	46	18486	2,59E-03					
GO:0031325	positive regulation of cellular metabolic process	18	3418	2,74E-03					
GO:0045893	positive regulation of transcription, DNA-templated	13	1762	2,78E-03					
GO:0051239	regulation of multicellular organismal process	18	3432	2,90E-03					
GO:0010648	negative regulation of cell communication	19	3347	3,37E-03					
GO:0023057	negative regulation of signaling	19	3347	3,37E-03					
GO:0048663	neuron fate commitment	5	115	3,60E-03					
GO:0048562	embryonic organ morphogenesis	8	484	3,66E-03					
GO:1902680	positive regulation of RNA biosynthetic process	13	1811	3,76E-03					
GO:0008284	positive regulation of cell proliferation	17	2808	6,06E-03					
GO:0048589	developmental growth	9	714	7,19E-03					
GO:0002009	morphogenesis of an epithelium	11	1159	8,08E-03					
GO:0009968	negative regulation of signal transduction	18	3204	8,15E-03					
GO:0048754	branching morphogenesis of an epithelial tube	8	542	8,44E-03					
GO:0060537	muscle tissue development	7	387	9,57E-03					
GO:0044237	cellular metabolic process	41	15758	1,02E-02					
GO:0065007	biological regulation	46	19296	1,15E-02					
GO:0060548	negative regulation of cell death	16	2642	1,29E-02					
GO:0044702	single organism reproductive process	19	3664	1,30E-02					
GO:0001708	cell fate specification	6	264	1,34E-02					
GO:0009953	dorsal/ventral pattern formation	6	264	1,34E-02					
GO:0048705	skeletal system morphogenesis	7	410	1,40E-02					
GO:0060021	palate development	6	269	1,49E-02					
GO:0001658	branching involved in ureteric bud morphogenesis	5	155	1,55E-02					
GO:0048706	embryonic skeletal system development	5	155	1,55E-02					
GO:0051093	negative regulation of developmental process	16	2690	1,63E-02					
GO:0007399	nervous system development	10	1018	1,76E-02					
GO:0007498	mesoderm development	5	161	1,87E-02					
GO:0048638	regulation of developmental growth	8	605	1,88E-02					
GO:0048585	negative regulation of response to stimulus	19	3759	1,89E-02					
GO:0043170	macromolecule metabolic process	36	12986	2,04E-02					
GO:0010975	regulation of neuron projection development	11	1281	2,08E-02					
GO:0051726	regulation of cell cycle	16	2760	2,25E-02					
GO:0061138	morphogenesis of a branching epithelium	8	624	2,35E-02					
GO:0048608	reproductive structure development	10	1062	2,54E-02					
GO:0040036	regulation of fibroblast growth factor receptor signaling pathway	4	82	2,69E-02					
GO:0002065	columnar/cuboidal epithelial cell differentiation	5	175	2,80E-02					
GO:0050793	regulation of developmental process	15	2884	2,84E-02					
GO:0051240	positive regulation of multicellular organismal process	14	2185	3,04E-02					
GO:0021544	subpallium development	2	5	3,05E-02					
GO:0007167	enzyme linked receptor protein signaling pathway	15	2504	3,12E-02					
GO:0048704	embryonic skeletal system morphogenesis	6	310	3,32E-02					
GO:0009880	embryonic pattern specification	5	185	3,66E-02					
GO:0048568	embryonic organ development	7	476	3,66E-02					
GO:0035295	tube development	8	664	3,68E-02					
GO:1901701	cellular response to oxygen-containing compound	15	2540	3,69E-02					
GO:0031344	regulation of cell projection organization	12	1640	3,86E-02					

LISTA A22: Enrichment dei geni del bicluster 6

Enrichment dei GO-term (biological process) dei 14 geni associabili al bicluster 6. I p-value sono corretti con il metodo Bonferroni, il livello di significatività di soglia usato è stato pari a un p-value di $5 \cdot 10^{-2}$. A sinistra sono elencati i GO-term, seguiti da una loro descrizione. Il campo "Geni" indica il numero di geni annotati per il termine corrispondente. "Classe" indica il numero totale di geni annotati per lo stesso termine nell'intero database di riferimento (NET-GE). I termini sono ordinati in ordine di p-value crescente.

GO-term	Descrizione	Geni	Classe	P-value	GO-term	Descrizione	Geni	Classe	P-value
GO:0048523	negative regulation of cellular process	11	5279	1,51E-05	GO:0019219	regulation of nucleobase-containing compound metabolic process	10	7797	1,38E-02
GO:0030182	neuron differentiation	7	911	1,59E-05	GO:0007423	sensory organ development	4	423	1,60E-02
GO:0048519	negative regulation of biological process	11	5756	3,80E-05	GO:0051171	regulation of nitrogen compound metabolic process	10	7927	1,61E-02
GO:0045934	negative regulation of nucleobase-containing compound metabolic process	10	3691	5,48E-05	GO:0009888	tissue development	7	2570	1,74E-02
GO:0001501	skeletal system development	6	620	5,95E-05	GO:0007275	multicellular organismal development	6	1673	1,92E-02
GO:0051172	negative regulation of nitrogen compound metabolic process	10	3798	7,23E-05	GO:0010557	positive regulation of macromolecule biosynthetic process	6	2042	1,96E-02
GO:0035107	appendage morphogenesis	5	348	1,39E-04	GO:0045935	positive regulation of nucleobase-containing compound metabolic process	6	2060	2,06E-02
GO:0035108	limb morphogenesis	5	348	1,39E-04	GO:0010720	positive regulation of cell development	5	977	2,17E-02
GO:0035137	hindlimb morphogenesis	4	137	1,82E-04	GO:0048729	tissue morphogenesis	4	578	2,23E-02
GO:0010605	negative regulation of macromolecule metabolic process	8	2452	1,92E-04	GO:0009887	organ morphogenesis	6	1719	2,24E-02
GO:0031324	negative regulation of cellular metabolic process	8	2479	2,09E-04	GO:0003002	regionalization	5	988	2,29E-02
GO:0001658	branching involved in ureteric bud morphogenesis	4	155	2,98E-04	GO:0051173	positive regulation of nitrogen compound metabolic process	6	2108	2,35E-02
GO:0045892	negative regulation of transcription, DNA-templated	9	3247	3,72E-04	GO:0032774	RNA biosynthetic process	7	3250	2,40E-02
GO:0009892	negative regulation of metabolic process	8	2679	3,79E-04	GO:0021884	forebrain neuron development	2	27	2,44E-02
GO:0031327	negative regulation of cellular biosynthetic process	7	1764	4,12E-04	GO:0010604	positive regulation of macromolecule metabolic process	7	3285	2,57E-02
GO:1902679	negative regulation of RNA biosynthetic process	9	3295	4,22E-04	GO:0031328	positive regulation of cellular biosynthetic process	6	2177	2,82E-02
GO:0009890	negative regulation of biosynthetic process	7	1786	4,48E-04	GO:0009891	positive regulation of biosynthetic process	6	2208	3,05E-02
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	8	2314	4,79E-04	GO:0060174	limb bud formation	2	25	4,01E-02
GO:0051253	negative regulation of RNA metabolic process	9	3369	5,11E-04	GO:0048666	neuron development	4	539	4,12E-02
GO:0001822	kidney development	5	489	7,45E-04	GO:0048754	branching morphogenesis of an epithelial tube	4	542	4,21E-02
GO:0010629	negative regulation of gene expression	9	3561	8,24E-04	GO:0060284	regulation of cell development	7	2979	4,58E-02
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	9	3633	9,78E-04	GO:0060425	lung morphogenesis	2	27	4,69E-02
GO:0060429	epithelium development	6	1006	1,02E-03	GO:0060255	regulation of macromolecule metabolic process	10	8942	4,91E-02
GO:0010558	negative regulation of macromolecule biosynthetic process	9	3771	1,35E-03	GO:0009893	positive regulation of metabolic process	7	3630	4,91E-02
GO:0006355	regulation of transcription, DNA-templated	10	6146	1,47E-03	GO:0002009	morphogenesis of an epithelium	5	1159	4,92E-02
GO:2001141	regulation of RNA biosynthetic process	10	6249	1,73E-03					
GO:0051252	regulation of RNA metabolic process	10	6360	2,04E-03					
GO:2000112	regulation of cellular macromolecule biosynthetic process	10	6579	2,81E-03					
GO:0006357	regulation of transcription from RNA polymerase II promoter	7	2379	3,06E-03					
GO:0009954	proximal/distal pattern formation	3	79	3,45E-03					
GO:0010556	regulation of macromolecule biosynthetic process	10	6784	3,75E-03					
GO:0030326	embryonic limb morphogenesis	4	303	4,29E-03					
GO:0035113	embryonic appendage morphogenesis	4	303	4,29E-03					
GO:0031128	developmental induction	3	86	4,45E-03					
GO:0031326	regulation of cellular biosynthetic process	10	6987	4,95E-03					
GO:0045168	cell-cell signaling involved in cell fate commitment	3	90	5,11E-03					
GO:0009889	regulation of biosynthetic process	10	7051	5,40E-03					
GO:0010468	regulation of gene expression	10	7158	6,21E-03					
GO:0048468	cell development	7	2280	7,90E-03					
GO:0006366	transcription from RNA polymerase II promoter	6	1448	8,40E-03					
GO:0021953	central nervous system neuron differentiation	4	365	8,95E-03					
GO:0007389	pattern specification process	6	1481	9,56E-03					
GO:0035136	forelimb morphogenesis	3	114	1,04E-02					
GO:0006351	transcription, DNA-templated	7	2949	1,27E-02					
GO:0045787	positive regulation of cell cycle	4	401	1,30E-02					
GO:0048598	embryonic morphogenesis	6	1566	1,32E-02					
GO:0010628	positive regulation of gene expression	6	1919	1,38E-02					

LISTA A23: Enrichment dei geni del bicluster 22

Enrichment dei GO-term (biological process) dei 9 geni associabili al bicluster 22. I p-value sono corretti con il metodo Bonferroni, il livello di significatività di soglia usato è stato pari a un p-value di $5 \cdot 10^{-2}$. A sinistra sono elencati i GO-term, seguiti da una loro descrizione. Il campo "Geni" indica il numero di geni annotati per il termine corrispondente. "Classe" indica il numero totale di geni annotati per lo stesso termine nell'intero database di riferimento (NET-GE). I termini sono ordinati in ordine di p-value crescente.

GO-term	Descrizione	Geni	Classe	P-value
GO:0030182	neuron differentiation	5	911	2,92E-03
GO:0045165	cell fate commitment	4	501	1,01E-02
GO:0030111	regulation of Wnt signaling pathway	3	325	2,84E-02
GO:0048522	positive regulation of cellular process	7	5768	2,92E-02
GO:0007389	pattern specification process	5	1481	3,14E-02
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	4	1008	3,18E-02
GO:0045665	negative regulation of neuron differentiation	3	213	3,53E-02

BIBLIOGRAFIA

- Abnizova, I., and Gilks, W.R. (2006). Studying statistical properties of regulatory DNA sequences, and their use in predicting regulatory regions in the eukaryotic genomes. *Brief. Bioinform.* 7, 48–54.
- Agrawal, R., and Srikant, R. (1994). Fast Algorithms for Mining Association Rules in Large Databases. In *Proceedings of the 20th International Conference on Very Large Data Bases*, (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.), pp. 487–499.
- Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* 16, 155–166.
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.
- Asmann, Y.W., Necela, B.M., Kalari, K.R., Hossain, A., Baker, T.R., Carr, J.M., Davis, C., Getz, J.E., Hostetter, G., Li, X., et al. (2012). Detection of redundant fusion transcripts as biomarkers or disease-specific therapeutic targets in breast cancer. *Cancer Res.* 72, 1921–1928.
- Assenov, Y., Ramírez, F., Schelhorn, S.-E., Lengauer, T., and Albrecht, M. (2008). Computing topological parameters of biological networks. *Bioinforma. Oxf. Engl.* 24, 282–284.
- Bai, L., and Merchant, J.L. (2000). Transcription Factor ZBP-89 Cooperates with Histone Acetyltransferase p300 during Butyrate Activation of p21 waf1 Transcription in Human Cells. *J. Biol. Chem.* 275, 30725–30733.
- Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299–308.
- Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587–1593.
- Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C.A., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.-M., Branco, M.R., et al. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature advance online publication*.
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004). Ultraconserved elements in the human genome. *Science* 304, 1321–1325.
- Benos, P.V., Bulyk, M.L., and Stormo, G.D. (2002). Additivity in protein–DNA interactions: how good an approximation is it? *Nucleic Acids Res.* 30, 4442–4451.
- Berg, O.G., and von Hippel, P.H. (1987). Selection of DNA binding sites by regulatory proteins. *J. Mol. Biol.* 193, 723–743.
- Bondarenko, V.A., Liu, Y.V., Jiang, Y.I., and Studitsky, V.M. (2003). Communication over a large distance: enhancers and insulators. *Biochem. Cell Biol.* 81, 241–251.
- Bovo, S., Di Lena, P., Martelli, P.L., Fariselli, P., and Casadio, R. (2016). NET-GE: a web-server for NETwork-based human gene enrichment. *Bioinforma. Oxf. Engl.* 32, 3489–3491.
- Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* 132, 311–322.
- Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell* 144, 327–339.
- Calhoun, V.C., Stathopoulos, A., and Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila* Antennapedia complex. *Proc. Natl. Acad. Sci. U. S. A.* 99, 9243–9247.
- Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how and why? *Mol. Cell* 49.
- Carlsten, J.O.P., Zhu, X., and Gustafsson, C.M. (2013). The multitasking Mediator complex. *Trends Biochem. Sci.* 38, 531–537.
- Cho, K.W.Y. (2012). Enhancers: WIRE Review. *Wiley Interdiscip. Rev. Dev. Biol.* 1, 469–478.
- Claverie, J.M., and Audic, S. (1996). The statistical significance of nucleotide position-weight matrix matches. *Comput. Appl. Biosci.* CABIOS 12, 431–439.
- Cornish-Bowden, A. (1985). Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Res.* 13, 3021–3030.

- Cremer, T., Cremer, M., Hübner, B., Strickfaden, H., Smeets, D., Popken, J., Sterr, M., Markaki, Y., Rippe, K., and Cremer, C. (2015). The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments. *FEBS Lett.* *589*, 2931–2943.
- Day, W.H., and McMorris, F.R. (1992). Critical comparison of consensus methods for molecular sequences. *Nucleic Acids Res.* *20*, 1093–1099.
- Dean, A. (2011). In the loop: long range chromatin interactions and gene regulation. *Brief. Funct. Genomics* elq033.
- Dekker, J. (2002). Capturing Chromosome Conformation. *Science* *295*, 1306–1311.
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* *22*, 1775–1789.
- Deshane, J., Kim, J., Bolisetty, S., Hock, T.D., Hill-Kapturczak, N., and Agarwal, A. (2010). Sp1 Regulates Chromatin Looping between an Intronic Enhancer and Distal Promoter of the Human Heme Oxygenase-1 Gene in Renal Cells. *J. Biol. Chem.* *285*, 16476–16486.
- Di Lena, P., Martelli, P.L., Fariselli, P., and Casadio, R. (2015). NET-GE: a novel NETWORK-based Gene Enrichment for detecting biological processes associated to Mendelian diseases. *BMC Genomics* *16*, S6.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* *485*, 376–380.
- Djureinovic, D., Hallström, B.M., Horie, M., Mattsson, J.S.M., La Fleur, L., Fagerberg, L., Brunnström, H., Lindskog, C., Madjar, K., Rahnenführer, J., et al. Profiling cancer testis antigens in non-small-cell lung cancer. *JCI Insight* *1*.
- Dollé, P. (2009). Developmental expression of retinoic acid receptors (RARs). *Nucl. Recept. Signal.* *7*, e006.
- Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res.* *16*, 1299–1309.
- ENCODE Project Consortium (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* *9*, e1001046.
- Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., et al. (2016). The Reactome pathway Knowledgebase. *Nucleic Acids Res.* *44*, D481–D487.
- Faisst, S., and Meyer, S. (1992). Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Res.* *20*, 3–26.
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* *41*, D808–815.
- Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor- α -bound human chromatin interactome. *Nature* *462*, 58–64.
- Galas, D.J., and Schmitz, A. (1978). DNase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* *5*, 3157–3170.
- Garner, M.M., and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res.* *9*, 3047–3060.
- Gause, M., Schaaf, C.A., and Dorsett, D. (2008). Cohesin and CTCF: cooperating to control chromosome conformation? *BioEssays News Rev. Mol. Cell. Dev. Biol.* *30*, 715–718.
- Gibcus, J.H., and Dekker, J. (2013). The Hierarchy of the 3D Genome. *Mol. Cell* *49*, 773–782.
- Gorla-Bajszczak, A., Juge-Aubry, C., Pernin, A., Burger, A.G., and Meier, C.A. (1999). Conserved amino acids in the ligand-binding and α domains of the peroxisome proliferator-activated receptor α are necessary for heterodimerization with RXR. *Mol. Cell. Endocrinol.* *147*, 37–47.
- Gupta, K., Sari-Ak, D., Haffke, M., Trowitzsch, S., and Berger, I. (2016). Zooming in on Transcription Preinitiation. *J. Mol. Biol.* *428*, 2581–2591.
- Habuka, M., Fagerberg, L., Hallström, B.M., Pontén, F., Yamamoto, T., and Uhlen, M. (2015). The Urinary Bladder Transcriptome and Proteome Defined by Transcriptomics and Antibody-Based Profiling. *PLOS ONE* *10*, e0145301.
- Hagberg, A.A., Schult, D.A., and Swart, P.J. (2008). Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, (Pasadena, CA USA), pp. 11–15.

- Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W.H., Ye, C., Ping, J.L.H., Mulawadi, F., et al. (2011). CTCF-Mediated Functional Chromatin Interactome in Pluripotent Cells. *Nat. Genet.* 43, 630–638.
- Hardison, R.C., and Taylor, J. (2012). Genomic approaches to finding cis-regulatory modules in animals. *Nat. Rev. Genet.* 13, 469–483.
- Hartigan, J.A. (1972). Direct Clustering of a Data Matrix. *J. Am. Stat. Assoc.* 67, 123–129.
- He, J., Ye, J., Cai, Y., Riquelme, C., Liu, J.O., Liu, X., Han, A., and Chen, L. (2011). Structure of p300 bound to MEF2 on DNA reveals a mechanism of enhanceosome assembly. *Nucleic Acids Res.* 39, 4464–4474.
- He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. *Nature* 495, 481–486.
- Henriques, R., and Madeira, S.C. (2015). BicNET: Efficient Biclustering of Biological Networks to Unravel Non-Trivial Modules. In *Algorithms in Bioinformatics*, (Springer, Berlin, Heidelberg), pp. 1–15.
- Henriques, R., and Madeira, S.C. (2016). BicNET: Flexible module discovery in large-scale biological networks using biclustering. *Algorithms Mol. Biol.* 11, 14.
- Hertz, G.Z., and Stormo, G.D. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinforma. Oxf. Engl.* 15, 563–577.
- Hirose, T., Apfel, R., Pfahl, M., and Jetten, A.M. (1995). The orphan receptor TAK1 acts as a repressor of RAR-, RXR- and T3R-mediated signaling pathways. *Biochem. Biophys. Res. Commun.* 211, 83–91.
- Holmqvist, P.-H., and Mannervik, M. (2013). Genomic occupancy of the transcriptional co-activators p300 and CBP. *Transcription* 4, 18–23.
- Hsu, F., Kent, W.J., Clawson, H., Kuhn, R.M., Diekhans, M., and Haussler, D. (2006). The UCSC Known Genes. *Bioinformatics* 22, 1036–1046.
- Iborra, F.J., Pombo, A., Jackson, D.A., and Cook, P.R. (1996). Active RNA polymerases are localized within discrete transcription “factories” in human nuclei. *J. Cell Sci.* 109, 1427–1436.
- Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316, 1497–1502.
- Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J.M., Yan, J., Sillanpää, M.J., et al. (2010). Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res.* 20, 861–873.
- Jones, E., Oliphant, T., Peterson, P., and others (2001). SciPy: Open source scientific tools for Python.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462.
- Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 32, D493–D496.
- Kechavarzi, B., and Janga, S.C. (2014). Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biol.* 15, R14.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, and D. (2002). The Human Genome Browser at UCSC. *Genome Res.* 12, 996–1006.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36.
- Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., et al. (2014). A draft map of the human proteome. *Nature* 509, 575–581.
- Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., et al. (2006). CAGE: cap analysis of gene expression. *Nat. Methods* 3, 211–222.
- Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T., et al. (2015). ArrayExpress update—simplifying data submissions. *Nucleic Acids Res.* 43, D1113–D1116.
- Korhonen, J., Martinmäki, P., Pizzi, C., Rastas, P., and Ukkonen, E. (2009). MOODS: fast search for position weight matrix matches in DNA sequences. *Bioinformatics* 25, 3181–3182.
- Kothary, R., Clapoff, S., Darling, S., Perry, M.D., Moran, L.A., and Rossant, J. (1989). Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* 105, 707–714.

- Krivega, I., and Dean, A. (2012). Enhancer and promoter interactions -- long distance calls. *Curr. Opin. Genet. Dev.* 22, 79–85.
- Kulakovskiy, I.V., Medvedeva, Y.A., Schaefer, U., Kasianov, A.S., Vorontsov, I.E., Bajic, V.B., and Makeev, V.J. (2013). HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res.* 41, D195–D202.
- Kypriotou, M., Beauchef, G., Chadjichristos, C., Widom, R., Renard, E., Jimenez, S.A., Korn, J., Maquart, F.-X., Oddos, T., Stetten, O.V., et al. (2007). Human Collagen Krox Up-regulates Type I Collagen Expression in Normal and Scleroderma Fibroblasts through Interaction with Sp1 and Sp3 Transcription Factors. *J. Biol. Chem.* 282, 32000–32014.
- Lan, X., Witt, H., Katsumura, K., Ye, Z., Wang, Q., Bresnick, E.H., Farnham, P.J., and Jin, V.X. (2012). Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic Acids Res.* 40, 7690–7704.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Lazzeroni, L., and Owen, A. (2002). Plaid models for gene expression data. *Stat. Sin.* 12, 61–86.
- Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* 13, 233–245.
- Lewin, B. (2006). *Il gene VIII* (Zanichelli editore S.p.A.).
- Li, L., and Davie, J.R. (2010). The role of Sp1 and Sp3 in normal and cancer cell biology. *Ann. Anat. Anat. Anz. Off. Organ Anat. Ges.* 192, 275–283.
- Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive Promoter-centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation. *Cell* 148, 84–98.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragozcy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science* 326, 289–293.
- Lindsay, S.J., Xu, Y., Lisgo, S.N., Harkin, L.F., Copp, A.J., Gerrelli, D., Clowry, G.J., Talbot, A., Keogh, M.J., Coxhead, J., et al. (2016). HDBR Expression: A Unique Resource for Global and Individual Gene Expression Studies during Early Human Brain Development. *Front. Neuroanat.* 10.
- Madeira, S.C., and Oliveira, A.L. (2004). Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 1, 24–45.
- Madeira, S.C., Teixeira, M.C., Sa-Correia, I., and Oliveira, A.L. (2010). Identification of Regulatory Modules in Time Series Gene Expression Data Using a Linear Time Biclustering Algorithm. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 7, 153–165.
- Maksimenko, O., and Georgiev, P. (2014). Mechanisms and proteins involved in long-distance interactions. *Front. Genet.* 5.
- Mark, M., Ghyselinck, N.B., and Chambon, P. (2009). Function of retinoic acid receptors during embryonic development. *Nucl. Recept. Signal.* 7, e002.
- Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* 7, 29–59.
- Mastrangelo, I.A., Courey, A.J., Wall, J.S., Jackson, S.P., and Hough, P.V. (1991). DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement. *Proc. Natl. Acad. Sci. U. S. A.* 88, 5670–5674.
- Matharu, N., and Ahituv, N. (2015). Minor Loops in Major Folds: Enhancer–Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. *PLoS Genet.* 11.
- Matys, V., Fricke, E., Geffers, R., Gößling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., et al. (2003). TRANSFAC®: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* 31, 374–378.
- Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., et al. (2006). TRANSFAC® and its module TRANSCompel®: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* 34, D108–D110.
- Medina-Rivera, A., Abreu-Goodger, C., Thomas-Chollier, M., Salgado, H., Collado-Vides, J., and van Helden, J. (2011). Theoretical and empirical quality assessment of transcription factor-binding motifs. *Nucleic Acids Res.* 39, 808–824.
- Merika, M., and Thanos, D. (2001). Enhanceosomes. *Curr. Opin. Genet. Dev.* 11, 205–208.
- Merika, M., Williams, A.J., Chen, G., Collins, T., and Thanos, D. (1998). Recruitment of CBP/p300 by the IFN β Enhanceosome Is Required for Synergistic Activation of Transcription. *Mol. Cell* 1, 277–287.

- von Mering, C., Jensen, L.J., Snel, B., Hooper, S.D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M.A., and Bork, P. (2005). STRING: known and predicted protein–protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 33, D433–D437.
- Miyamoto, S., Hidaka, K., Jin, D., and Morisaki, T. (2009). RNA-binding proteins Rbm38 and Rbm24 regulate myogenic differentiation via p21-dependent and -independent regulatory pathways. *Genes Cells* 14, 1241–1252.
- Moncaut, N., Rigby, P.W.J., and Carvajal, J.J. (2013). Dial M(RF) for myogenesis. *FEBS J.* 280, 3980–3990.
- Mukherjee, S., Berger, M.F., Jona, G., Wang, X.S., Muzzey, D., Snyder, M., Young, R.A., and Bulyk, M.L. (2004). Rapid Analysis of the DNA Binding Specificities of Transcription Factors with DNA Microarrays. *Nat. Genet.* 36, 1331–1339.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science* 320, 1344–1349.
- Nobrega, M.A., Ovcharenko, I., Afzal, V., and Rubin, E.M. (2003). Scanning human gene deserts for long-range enhancers. *Science* 302, 413.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation center. *Nature* 485, 381–385.
- Ogbourne, S., and Antalis, T.M. (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem. J.* 331, 1–14.
- Okamoto, S., Sherman, K., Bai, G., and Lipton, S.A. (2002). Effect of the ubiquitous transcription factors, SP1 and MAZ, on NMDA receptor subunit type 1 (NR1) expression during neuronal differentiation. *Mol. Brain Res.* 107, 89–96.
- Orlando, V. (2000). Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem. Sci.* 25, 99–104.
- Orphanides, G., and Reinberg, D. (2002). A Unified Theory of Gene Expression. *Cell* 108, 439–451.
- Pan, Y., and Phan, S. (2009). Threshold for positional weight matrix. *Eng. Lett.* 16, 498–504.
- Panne, D. (2008). The enhanceosome. *Curr. Opin. Struct. Biol.* 18, 236–242.
- Park, J.-M., Jo, S.-H., Kim, M.-Y., Kim, T.-H., and Ahn, Y.-H. (2015). Role of transcription factor acetylation in the regulation of metabolic homeostasis. *Protein Cell* 6, 804–813.
- Pennacchio, L.A., and Rubin, E.M. (2001). Genomic strategies to identify mammalian regulatory sequences. *Nat. Rev. Genet.* 2, 100–109.
- Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444, 499–502.
- Petryszak, R., Keays, M., Tang, Y.A., Fonseca, N.A., Barrera, E., Burdett, T., Füllgrabe, A., Fuentes, A.M.-P., Jupp, S., Koskinen, S., et al. (2016). Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res.* 44, D746–D752.
- Phillips, J.E., and Corces, V.G. (2009). CTCF: Master Weaver of the Genome. *Cell* 137, 1194–1211.
- Phillips-Cremins, J.E., and Corces, V.G. (2013). Chromatin Insulators: Linking genome organization to cellular function. *Mol. Cell* 50, 461–474.
- Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3-D organization of genomes during lineage commitment. *Cell* 153, 1281–1295.
- Pizzi, C., Rastas, P., and Ukkonen, E. (2011). Finding Significant Matches of Position Weight Matrices in Linear Time. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 8, 69–79.
- Plank, J.L., and Dean, A. (2014). Enhancer Function: Mechanistic and Genome-Wide Insights Come Together. *Mol. Cell* 55, 5–14.
- Pollock, R., and Treisman, R. (1990). A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res.* 18, 6197–6204.
- Pontes, B., Giraldez, R., and Aguilar-Ruiz, J.S. (2015a). Quality Measures for Gene Expression Biclusters. *PLoS ONE* 10.
- Pontes, B., Giraldez, R., and Aguilar-Ruiz, J.S. (2015b). Biclustering on expression data: A review. *J. Biomed. Inform.* 57, 163–180.
- Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W.W., and Sandelin, A. (2010). JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 38, D105–D110.

- Poss, Z.C., Ebmeier, C.C., and Taatjes, D.J. (2013). The Mediator complex and transcription regulation. *Crit. Rev. Biochem. Mol. Biol.* 48, 575–608.
- Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. *Nature* 386, 569–577.
- Qi, H., Liu, M., Emery, D.W., and Stamatoyannopoulos, G. (2015). Functional Validation of a Constitutive Autonomous Silencer Element. *PLoS ONE* 10.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309.
- Riethoven, J.-J.M. (2010). Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. *Methods Mol. Biol. Clifton NJ* 674, 33–42.
- Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett.* 579, 909–915.
- Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* 32, D91–D94.
- Schneider, T.D., and Stephens, R.M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18, 6097–6100.
- Schneider, T.D., Stormo, G.D., Gold, L., and Ehrenfeucht, A. (1986). Information content of binding sites on nucleotide sequences. *J. Mol. Biol.* 188, 415–431.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the *Drosophila* Genome. *Cell* 148, 458–472.
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423.
- Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* 15, 272–286.
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nat. Genet.* 38, 1348–1354.
- Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinforma. Oxf. Engl.* 27, 431–432.
- Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 13, 613–626.
- Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the β -globin locus. *Genes Dev.* 20, 2349–2354.
- Staden, R. (1989). Methods for calculating the probabilities of finding patterns in sequences. *Comput. Appl. Biosci. CABIOS* 5, 89–96.
- Stormo, G.D. (2000). DNA binding sites: representation and discovery. *Bioinforma. Oxf. Engl.* 16, 16–23.
- Stormo, G.D., Schneider, T.D., Gold, L., and Ehrenfeucht, A. (1982). Use of the “Perceptron” algorithm to distinguish translational initiation sites in *E. coli*. *Nucleic Acids Res.* 10, 2997–3011.
- Su, W., Jackson, S., Tjian, R., and Echols, H. (1991). DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. *Genes Dev.* 5, 820–826.
- Supp, D.M., Witte, D.P., Branford, W.W., Smith, E.P., and Potter, S.S. (1996). Sp4, a Member of the Sp1-Family of Zinc Finger Transcription Factors, Is Required for Normal Murine Growth, Viability, and Male Fertility. *Dev. Biol.* 176, 284–299.
- Suryamohan, K., and Halfon, M.S. (2015). Overview Article: Identifying transcriptional cis-regulatory modules in animal genomes. *Wiley Interdiscip. Rev. Dev. Biol.* 4, 59–84.
- Suzuki, T., Kimura, A., Nagai, R., and Horikoshi, M. (2000). Regulation of interaction of the acetyltransferase region of p300 and the DNA-binding domain of Sp1 on and through DNA binding. *Genes Cells* 5, 29–41.
- The Gene Ontology Consortium (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 43, D1049–D1056.
- The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45, D158–D169.
- Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F., and Laat, W. de (2002). Looping and Interaction between Hypersensitive Sites in the Active β -globin Locus. *Mol. Cell* 10, 1453–1465.

- Tronche, F., Ringeisen, F., Blumenfeld, M., Yaniv, M., and Pontoglio, M. (1997). Analysis of the distribution of binding sites for a tissue-specific transcription factor in the vertebrate genome. *J. Mol. Biol.* *266*, 231–245.
- Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Tissue-based map of the human proteome. *Science* *347*, 1260419.
- Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* *10*, 252–263.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The Sequence of the Human Genome. *Science* *291*, 1304–1351.
- Venters, B.J., and Pugh, B.F. (2009). How eukaryotic genes are transcribed. *Crit. Rev. Biochem. Mol. Biol.* *44*, 117–141.
- Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res.* *35*, D88-92.
- Visel, A., Prabhakar, S., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Afzal, V., Rubin, E.M., and Pennacchio, L.A. (2008). Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat. Genet.* *40*, 158–160.
- Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* *457*, 854–858.
- Wasserman, W.W., and Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.* *5*, 276–287.
- Weipoltshammer, K., and Schöfer, C. (2016). Morphology of nuclear transcription. *Histochem. Cell Biol.* *145*, 343–358.
- Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* *451*, 796–801.
- Widom, R.L., Culic, I., Lee, J.Y., and Korn, J.H. (1997). Cloning and characterization of hcKrox, a transcriptional regulator of extracellular matrix gene expression. *Gene* *198*, 407–420.
- Wingender, E., Schoeps, T., and Dönitz, J. (2013). TFClass: an expandable hierarchical classification of human transcription factors. *Nucleic Acids Res.* *41*, D165-170.
- Wingender, E., Schoeps, T., Haubrock, M., and Dönitz, J. (2015). TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic Acids Res.* *43*, D97–D102.
- Wit, E. de, and Laats, W. de (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* *26*, 11–24.
- Wu, T.D., Nevill-Manning, C.G., and Brutlag, D.L. (2000). Fast probabilistic analysis of sequence function using scoring matrices. *Bioinforma. Oxf. Engl.* *16*, 233–244.
- Yasmin, R., Yeung, K.T., Chung, R.H., Gaczynska, M.E., Osmulski, P.A., and Noy, N. (2004). DNA-looping by RXR Tetramers Permits Transcriptional Regulation “at a Distance.” *J. Mol. Biol.* *343*, 327–338.
- Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. *Nucleic Acids Res.* *44*, D710–D716.
- Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* *15*, 2343–2360.
- Zhao, S. (2014). Assessment of the Impact of Using a Reference Transcriptome in Mapping Short RNA-Seq Reads. *PLOS ONE* *9*, e101374.
- Zhao, C., and Meng, A. (2005). Sp1-like transcription factors are regulators of embryonic development in vertebrates. *Dev. Growth Differ.* *47*, 201–211.
- Zuchegna, C., Aceto, F., Bertoni, A., Romano, A., Perillo, B., Laccetti, P., Gottesman, M.E., Avvedimento, E.V., and Porcellini, A. (2015). Mechanism of retinoic acid-induced transcription: histone code, DNA oxidation and formation of chromatin loops. *Nucleic Acids Res.* *42*, 11040–11055.
- Zuin, J., Dixon, J.R., Reijden, M.I.J.A. van der, Ye, Z., Kolovos, P., Brouwer, R.W.W., Corput, M.P.C. van de, Werken, H.J.G. van de, Knoch, T.A., IJcken, W.F.J. van, et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci.* *111*, 996–1001.