

**UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II**



**DOTTORATO DI RICERCA IN  
FISIOPATOLOGIA CLINICA E MEDICINA  
SPERIMENTALE**

*INDIRIZZO IN SCIENZE*

*GERONTOLOGICHE*

**XIX Ciclo**

**Coordinatore: Prof. Gianni Marone**

**TESI DI DOTTORATO**

**RELAZIONE TRA I PRINCIPALI MRNA DELLE  
CELLULE STAMINALI PLURIPOTENTI ED I  
MECCANISMI DI REGOLAZIONE GENICA**

TUTOR

**Chiar.mo**

Prof. Franco Rengo

CANDIDATO/A

**Dott. Gianluca Testa**

## INDICE

<b>Introduzione</b>	<b>pag. 3</b>
<b>Materiali e Metodi</b>	<b>pag. 8</b>
<b>Risultati</b>	<b>pag. 13</b>
<b>Discussione</b>	<b>pag. 32</b>
<b>Bibliografia</b>	<b>pag. 34</b>
<b>Tabelle</b>	<b>pag. 48</b>
<b>Figure</b>	<b>pag. 49</b>

## INTRODUZIONE

Le cellule staminali costituiscono le unità di base dello sviluppo e della rigenerazione e sono caratterizzate dalla possibilità pressoché illimitata di self-renewal e differenziamento. Tuttavia, la maggiore capacità di capacità di sviluppo è esclusivo appannaggio delle cellule staminali embrionali (ESc) che sono isolate dall'inner cell mass (ICM) di embrioni allo stato pre-impianto (Wobus & Boheler, 2005).

Le ESc possono essere coltivate in vitro indefinitamente pur mantenendo un cariotipo normale e conservando la capacità di dare origine a tutte le cellule dell'organismo (pluripotenza).

La generazione di ESc umane (Thomson et al., 1998) ha dato luogo ad approfondite considerazioni sulle loro applicazioni nello studio della biologia dello sviluppo umano e per la medicina rigenerativa (Wobus & Boheler, 2005) ma, di fatto, la conoscenza dei complessi meccanismi che ne regolano la pluripotenza ed il potenziale di differenziamento sono tutt'ora limitati ad alcuni signaling pathways (LIF, BMP, Wnt, FGF4) e ad alcuni fattori di trascrizione (Oct-3/4, Nanog, Sox2, Bcl2, STAT3, c-Myc,

Id)(Cartwright et al., 2005; Keller, 2005; Wobus & Boheler, 2005; Yamane et al., 2005; Ying et al., 2003) che potrebbero essere diversi tra topo ed uomo. Sebbene la pluripotenza delle ESc sia indiscutibile, il loro utilizzo come modello di sviluppo evolutivo e come sorgente di definite popolazioni cellulari per scopi farmaceutici o rigenerativi è, quindi, limitato dal fatto che i meccanismi, che controllano lo stato pluripotente e la conseguente capacità rigenerativa, rimangono ancora perlopiù sconosciuti.

È oramai generalmente accettato il fatto che la biologia delle ESc sia gestita fondamentalmente attraverso meccanismi di controllo trascrizionale. La recente scoperta che Oct4, Sox2 e Nanog cooperano per formare un circuito regolatorio implicato nel self-renewal e nella pluripotenza supporta l'idea che le caratteristiche delle ESc siano realmente controllate tramite complessi meccanismi trascrizionali (Boyer *et al.*, 2005; Wang *et al.*, 2006).

Teoricamente, l'identificazione di trascritti specifici delle ESc dovrebbe facilitare l'identificazione di altri fattori di trascrizione che regolano le peculiari capacità delle ESc (Wobus & Boheler, 2005). Una possibilità di rintracciare networks trascrizionali co-regolati è

quella di esaminare i promotori dei geni i cui prodotti sono regolati in modo coordinato nel tempo e nello spazio. Dal punto di vista funzionale, i fattori di trascrizione si legano ai promotori od enhancers di specifici geni per regolarne la trascrizione. Analogamente, ogni promotore, nei mammiferi, deve essere controllato da un'insieme di frameworks conservate, costituite da vari cis-elements, che gestiscono le diverse risposte per assicurare la corretta gestione spaziale e temporale dell'espressione genica.

Numerose pubblicazioni hanno dimostrato che le frameworks regolatorie possono essere rappresentate da modelli computazionali (Werner, 2000; Werner et al., 2003), e che, una volta identificati, questi modelli possono essere utilizzati per screening su interi databases di acidi nucleici per identificare altri geni che potrebbero contenere una simile organizzazione della regione regolatoria (Dohr et al., 2005). La previsione di modelli di framework si basa su tre criteri fondamentali (Cartharius et al., 2005; Seifert et al., 2005). Il primo criterio consiste nel fatto che siti di legame di fattori di trascrizione conosciuti possono essere utilizzati per identificare potenziali cis-elements nei promotori. Secondo, la presenza di due o più cis-elements conservati

localizzati in un ordine specifico ed in uno spazio definito tra promotori genici “biologicamente correlati” si possono considerare un’unità di base di regolazione genica. Il terzo criterio si basa sulla descrizione numerica di un probabile modello di framework che consiste in siti di legame per fattori di trascrizione (Transcription Factor Binding Sites, TFBS), le somiglianze tra i diversi TFBS, la distanza tra gli ipotetici TFBS e l’orientamento di ciascun TFBS. I modelli di frameworks vengono classificati con un punteggio di qualità che è, essenzialmente, il rapporto tra selectivity e recall, laddove selectivity è la percentuale di promotori identificati in un dato genoma riconosciuto dal modello e recall è la percentuale di promotori dei geni inseriti riconosciuti dal modello.

Per sfruttare questi modelli computazionali, abbiamo ipotizzato che i geni con un profilo espressivo limitato o predominante nelle cellule pluripotenti potessero contenere cis-elements conservati che potessero predire i fattori di trascrizione implicati nella biologia delle cellule staminali. Per verificare questa ipotesi, abbiamo unito e confrontato profili quantitativi dei prodotti trascrizionali di tre linee di cellule staminali pluripotenti murine (ES cells: R1, D3; ed embryonic germ (EG) cells: EG-1) per identificare un gruppo di trascritti abbondanti nelle cellule

staminali pluripotenti. I corrispondenti geni espressi con maggiore abbondanza sono stati inseriti in un procedimento reiterativo per identificare frameworks conservate di cis-elements presenti nel sottogruppo di geni espressi nelle ESc (Dohr et al., 2005; Seifert et al., 2005). Questo ci portato all'identificazione di cinque frameworks che rispondevano a tali criteri, incluse alcune con siti di legami per octamer e sox. Il nostro modello ha predetto il ruolo di due nuovi fattori di trascrizione coinvolti nel mantenimento e nel differenziamento delle cellule staminali pluripotenti.

## MATERIALI E METODI

### ***Cellule staminali embrionali murine***

Per i nostri esperimenti abbiamo utilizzato tre linee di cellule staminali embrionali murine: D3 (Doetschman *et al.*, 1985), R1 (Nagy *et al.*, 1993), ESc ed EG-1 embryonic germ cell (Stewart *et al.*, 1994). Tali cellule sono state mantenute in coltura in presenza od in assenza di feeder layers di fibroblasti embrionali murini (MEFs) come descritto in letteratura (Wobus *et al.*, 2002).

### ***Generazione ed analisi delle librerie SAGE***

Sono state costruite librerie SAGE (Serial Analysis of Gene Expression) di cellule staminali puripotenti utilizzando D3, R1 ed EG-1 come in precedenza descritto da Anisimov *et al.* (Anisimov *et al.* 2002), partendo da RNA poliadenilato (PoliA+). In seguito tali librerie sono state combinate generando così un'unica lista di trascritti designati dall'acronimo RED (R1, EG-1 and D3 cell lines – *RED*). L'abbondanza dei trascritti è stata verificata mediante il confronto con 25 librerie SAGE, presenti nel laboratorio o disponibili in rete, di tessuti, cellule adulte o precursori. Tali librerie sono state contrassegnate dall'acronimo APC (Adult

tissues/Precursor Cells/Cell Lines). È stato poi determinato il rapporto di ogni elemento presente nel gruppo RED con quello presente nel gruppo APC. I trascritti presenti in rapporto RED/APC > di 3 sono stati considerati come predominanti nelle cellule pluripotenti.

### ***RT-PCR Quantitative***

Le reazioni di RT-PCR quantitativa sono state effettuate con l'apparato ABI PRISM 7900HT Sequence Detector System (PE Applied Biosystems) su RNA ottenuto mediante purificazione con metodica al CsCl<sub>2</sub> utilizzando il SYBR Green PCR Master Mix (Applied Biosystems) o il Platinum® SYBR® Green qPCR SuperMix (Invitrogen) in piastre da 384 pozzetti (Tarasova et al., 2006). Le analisi dei risultati sono state effettuate con il software SDS 2.1 usando il metodo della curva standard, Standard Curve Method (Applied Biosystems, User Bulletin #2). Per confermare i trascritti genici predominanti nelle ESc sono stati utilizzati primers specifici.

### ***Immunoistochimica***

L'immunofluorescenza di cellule ESc è stata effettuata come descritto in letteratura (Tarasova et al., 2006; Wobus et al., 2002) usando i seguenti anticorpi Maz/Pur1 di pollo (Abcam, USA; 1:20), B-myb di coniglio (Santa Cruz, 1:20), Sox-2 di coniglio (1:100, Abcam), Creb1 di coniglio (1:100, Chemicon) ed IgG murine, (Molecular Probes, 1:20). Le cellule sono state incubate con anticorpi secondari coniugati al fluorocromo [FITC-conjugated anti-chicken, anti-rabbit or anti-mouse IgG (1:100) or Alexa Fluor 488 or 568 conjugated goat anti-rabbit IgG(H+L)].

### ***Analisi dei Promotori e Modelli di Framework***

Per identificare le potenziali frameworks caratteristiche dei trascritti predominanti nelle ESc identificati dal SAGE, abbiamo utilizzato il pacchetto di software offerto da Genomatix (Dohr et al., 2005). Inizialmente, la nostra analisi ha portato all'identificazione di un unico sito di legame per fattori di trascrizione (TFBS) per ogni modello ipotizzato e, quindi, non abbiamo potuto considerare questo risultato significativo. Questa situazione ci ha indotto ad elaborare una Problem Specific List (PSL) costituita esclusivamente dai trascritti più abbondanti nelle

ESc: Fgf4, Pou5f1, Utf1, Zfp42, Dnmt3b, Lin28, Nanog, Tdgf1, Esrrb, Etv4, B-Myb, e Sox2.

### ***Chromosomal Immunoprecipitation (ChIP)***

Gli esperimenti di Chromosomal Immunoprecipitation sono stati effettuati come indicato nel kit EZ-Chip (Catalog #17-371, Upstate). Il DNA iniziale e quello immunoprecipitato sono stati analizzati mediante PCR quantitativa per verificare il legame di MYBL2 ai promotori dei geni *sox2*, *nanog* and *myc*. MYC è stato utilizzato come controllo positivo.

### ***Cloni overesprimenti B-Myb***

Per generare cloni di ESc overesprimenti stabilmente B-Myb, il cDNA di B-Myb delle dimensioni di 2.5 kb ottenuto mediante RT-PCR è stato clonato nel vettore di espressione pIRES-EGFP (BD Bioscience Clontech), sotto il controllo sia del promotore CMV che, dopo sostituzione, del promotore PGK. Il costrutto ottenuto è stato introdotto nelle ESc R1 mediante elettroporazione ed i cloni ottenuti selezionati mediante resistenza a G418 conferita da una cassetta contenuta nello stesso costrutto come descritto in letteratura (Tarasova et al., 2006). Ogni clone

così ottenuto è stato testato valutando i markers di pluripotenza mediante PCR, immunofluorescenza e valutazione dell'attività della fosfatasi alcalina (Vector Laboratories).

### ***Valutazione della Proliferazione Cellulare***

Per valutare la proliferazione cellulare è stato utilizzato il kit che prevede l'utilizzo del WST-1 (Roche) secondo le istruzioni della casa produttrice, aggiungendo 10 $\mu$ l di reagente WST-1 a 100 $\mu$ l di mezzo in ogni pozzetto per 1h. L'assorbanza è stata letta a 600 nm.

### ***Analisi Statistica***

I risultati sono presentati come media  $\pm$  deviazione standard. Inoltre è stata effettuata l'analisi della varianza (ANOVA) seguita test di Student-Newman-Keul per confronti multipli della media. Sono stati considerati significativi valori di P inferiori a 0.05.

## RISULTATI

### Analisi del Trascrittoma di cellule pluripotenti

#### ***Serial Analysis of Gene Expression (SAGE) di Linee Cellulari di ES ed EG***

I microarrays sono stati ampiamente utilizzati per identificare i trascritti predominanti nelle cellule staminali; comunque la variabilità emersa tra i diversi studi condotti ci ha spinti ad utilizzare il SAGE (Serial Analysis of Gene Expression), una tecnica imparziale e basata sulle sequenze che quantifica accuratamente l'abbondanza dell'mRNA (conosciuti e sconosciuti) (Velculescu et al., 1995) (Figure 1A).

Le librerie SAGE di ESc D3, R1 ed EG-1 sono state costruite utilizzando NlaIII come anchoring enzyme e BsmFI come tagging enzyme. La libreria finale di D3 conteneva 13.144 tags unici, da un totale di 37.652 valutabili, e la libreria di EG-1 conteneva 15.825 tags unici da un totale di 47.544 tags valutabili.

I dettagli completi sulle librerie generate (GSM 11348 e GSM 11349) sono disponibili all'indirizzo web <http://www.ncbi.nlm.nih.gov/sage/>.

I 50 tags più abbondanti costituivano più del 21% del totale dei trascritti e tutti i tags di entrambe le librerie erano presenti tra i 90 trascritti più abbondanti presenti in entrambe le librerie. La maggior parte dei tags singoli ed altamente rappresentati potevano essere raggruppate in tre categorie funzionali: 1) metabolismo, 2) crescita cellulare e 3) mantenimento e comunicazione cellulare. Entrambe le librerie contenevano quantitativamente più geni correlati alla sintesi ed al metabolismo (geni coinvolti in processi anabolici) che ogni altra libreria SAGE di tessuti fetali o adulti.

Dal momento che il SAGE è essenzialmente una tecnica di campionamento, circa il 5.5% dei tags poco abbondanti erano esclusivi per ogni libreria; comunque per i tags più abbondanti (>5 copie per libreria), i profili di espressione nelle due librerie erano statisticamente quasi sovrapponibili, a dimostrazione del fatto che il fenotipo molecolare delle ESc e delle EG è altamente conservato. A seguito della correzione di Bonferroni dell'abbondanza normalizzata dei tags (tags per milione), solo un tag (GTGGCTCACA) per l'interleukin-1 receptor-associated kinase 1 (Irak1) presentava un'abbondanza significativamente maggiore ( $p < 0.05$ ) nelle EG che nelle D3 alla conferma della PCR quantitativa.

### ***Trascritti Predominanti nelle Cellule ES***

Al fine di identificare in maniera inequivocabile i trascritti predominanti delle cellule staminali pluripotenti che ne potessero regolare il fenotipo molecolare, abbiamo unificato le librerie SAGE delle cellule EG-1, D3 ed R1 per confrontarle con le 25 librerie SAGE di diversi tessuti adulti e fetali, anch'esse riunite in un unico data base (Figura 1B).

Per facilitare il confronto, abbiamo eliminato tutti i tags non informativi, tutti quelli presenti in singola copia (per escludere possibili errori di sequenziamento) e tutti i tags che non fossero presenti in tutte le 3 librerie di cellule pluripotenti. Sulla base della sola abbondanza, questo procedimento ha consentito di ridurre il nostro database ad un totale di 2423 tags unici che potevano essere significativamente più abbondanti nelle cellule staminali pluripotenti rispetto ad ogni altro tipo cellulare (Figura 1C). Il confronto diretto delle sequenze dei tags tra tutte le nostre librerie ha, inoltre, indicato che solo 80 tags erano “esclusivamente” presenti nelle linee cellulari di ES ed EG e che ulteriori 50 tags erano quantitativamente più abbondanti di almeno 5 volte nelle 3 librerie di cellule ES/EG che in ogni altro tipo cellulare.

Dal momento che numerosi tags presenti nelle cellule staminali embrionali (ES e D3) e nelle cellule germinali (EG) potevano corrispondere a numerosi possibili trascritti, al fine di caratterizzare definitivamente i possibili trascritti abbiamo utilizzato la RT-PCR quantitativa sugli RNA più rappresentati nelle cellule staminali pluripotenti rispetto ai tessuti fetali ed adulti (Figura 1D).

Nonostante il fatto che tre tags non potevano essere definitivamente associati ad alcun gene o prodotto genico, e che per tale motivo non era possibile valutarli mediante RT-PCR quantitativa, questo approccio ha confermato che, tra altri, i trascritti predominanti nelle cellule staminali embrionali Tdgf1, Pou5f1 (Oct4), Nanog, Fgf4, Utf1, Zfp42, Sox2 e Dppa5.

Il confronto effettuato sulle librerie SAGE e l'analisi mediante RT-PCR quantitativa ha inoltre indicato che i trascritti più abbondanti erano associabili, fondamentalmente, sia alle cellule ES ed EG allo stato indifferenziato (EmS – Embryonic Stemness, Figura 1E) che, più in generale, alle cellule pluripotenti insieme ai tessuti in via di sviluppo (p.es. cornea neonatale) o derivati immaturi delle cellule staminali (p.es., cellule dendritiche

derivate da ES, cellule dendritiche derivate da midollo osseo) (SnD – Stemness and Differentiation, Figura 1E).

Basandoci sulla nostra ipotesi originale e sulla presenza di tali trascritti nei tessuti in via di sviluppo o nelle cellule in differenziamento, abbiamo eliminato, dalle nostre successive analisi, molti trascritti che in precedenza erano già stati associati alla staminalità. I restanti trascritti preponderanti nelle cellule staminali pluripotenti non erano riferibili ad alcun singolo locus cromosomico. Inoltre, i trascritti codificanti per i geni come Pou5f1, Nanog, Utf1, Sox2, B-Myb, DNA metiltransferasi 3a (Dnmt3a), 3b (Dnmt3b), and Tdgf1 erano presenti tra i trascritti correlati alla staminalità delle cellule staminali embrionali umane, come in precedenza già descritti dalla letteratura (Bhattacharya et al., 2004; Ginis et al., 2004; Richards et al., 2004; Sato et al., 2003; Sperger et al., 2003); in ogni caso esistevano differenze importanti.

## **Identificazione di *cis*-Elements Conservati e di Ipotetici *trans*-Regulators**

### ***Frameworks di cis-Elements specifiche delle cellule ES/EG***

Il profilo caratteristico dei trascritti predominanti nelle cellule ES/EG ci ha fornito una lista di possibili geni che noi avevamo postulato potessero essere utili nell'identificazione degli elementi genici conservati (i.e., *cis*- elements and *trans*-regulatory factors) e che potessero essere responsabili dei pattern di espressione esclusivi e caratteristici delle cellule staminali pluripotenti.

Cercando gli elementi regolatori conservati (*cis*-elements), abbiamo dato per assunto che i geni attivamente trascritti nelle cellule staminali ES ed EG contenessero una prevalenza maggiore di siti di legame per fattori di trascrizione (TFBSs) unici rispetto ad una previsione basata sul sequenziamento random del genoma, e che almeno una framework comune, costituita da vari *cis*-elements conservati, fosse condivisa da alcuni dei geni dei geni identificati. Per identificare questi ipotetici *cis*-elements, abbiamo utilizzato Genomatix, uno strumento bioinformatico che facilita l'identificazione di potenziali *cis*-elements implicati nel

controllo trascrizionale di un gruppo conservato di geni (nel nostro caso quelli della Embryonic Stemness – Ems)

La nostra analisi iniziale è stata eseguita sulle regioni prossimali del promotore (da -500 bp a +100 bp rispetto al punto d'inizio della trascrizione) dei geni identificati a partire dai 50 più presenti, nelle cellule ES/EG, utilizzando Gene2Promoter e MatInspector della suite Genomatix. Queste analisi, in particolare se estese a - 1000 ed a + 500, hanno indicato la presenza di molteplici siti di legame per Octamer, Ebox and Nkxh, Mybl, Sory, Stat e Smad, presenti ad una frequenza superiore a quella predetta a partire dall'analisi del genoma in toto.

Tra le regioni prossimali dei promotori di geni relativi alla staminalità si sono così individuate numerose frameworks, che potenzialmente definivano unità trascrizionali regolatorie attive, contenenti a) VMyb, TBPF, MAZ, CREB, b) Nkx/Dlx, CREB ed Ets e c) Ets e CREB; comunque non è stato possibile identificare un singolo modello che avesse un qualità score accettabile (alta specificità e bassa prevalenza nell'intero genoma).

Allora abbiamo focalizzato la nostra analisi sui più abbondanti geni riguardanti la staminalità includendo i ben noti regolatori delle ES Oct4, nanog, fgf4, and zfp42. Di nuovo,

impostando la regione prossimale del promotore da -500 bp a +100 bp rispetto al punto d'inizio della trascrizione, solo nove promotori contenevano il sito di legame per octamer e nessuno conteneva un sito di legame per Sry-related HMG (Sory).

Sette dei dodici geni caratteristici delle ES, contenevano siti di legame per octamer nella regione prossimale del promotore ma nessuno conteneva siti di legame per Sory. Comunque, analizzando le regioni distali (da -1000 a + 400), tutti questi geni contenevano siti per octamer e solo due contenevano siti per Sory, suggerendo che fosse necessario considerare regioni di promotore più lunghe di 600 bp. Tra tali geni associati alla stemness, tuttavia, nessuno conteneva una prevalenza di siti di legame per octamer e Sory nella regione prossimale del promotore rispetto al genoma in toto.

Dal momento che nessuno di questi modelli di framework raggiungeva un quality score soddisfacente, basandoci sui criteri precedentemente definiti, abbiamo esteso la ricerca aumentando le dimensioni delle regioni di promotore su cui lavorare (da - 2000 a + 500 bp) ed abbiamo considerato un sottogruppo di 12 promotori genici con la più alta prevalenza nelle cellule ES/EG

(Problem Specific List, escluso RIKEN cDNA 2410146L05 gene, Tabella 1).

Abbiamo analizzato differenti gruppi e sottogruppi di questi 12 geni mediante il software e, dopo ottimizzazione e classificazione, sono emersi 6 modelli che definivano un framework comune di siti di legame per fattori di trascrizione (cis-elements) che tenessero in considerazione distanze, matrici ed caratteristiche del filamento di DNA su cui si trovavano (orientamento).

Quindi, abbiamo screenato tutto il genoma con questi sei modelli ed abbiamo trovato che il promotore di c-myc, il cui prodotto genico è stato recentemente dimostrato come critico per la regolazione da Lif nelle ESc di topo (Cartwright et al., 2005), conteneva 2 dei nostri modelli. Abbiamo quindi utilizzato ModellInspector e Frameworker del pacchetto Genomatix, utilizzando c-Myc, ed il modello con il miglior quality score (0.00126), basato sulla lunghezza di promotore automatica definita dal software (650 bp contro le 2500 bp usate nella nostra analisi finale), era costituito da quattro siti di legame conservati presenti in cinque dei 14 promotori esaminati (Figura 2A).

Questo modello consisteva nei siti di legame per fattori di trascrizione (TFBSs) per l'omologo dell'oncogene della mieloblastosi virale acuta (B-Myb), siti di legame per Sox (SORY), Myc-associated Zinc Finger (MAZ) ed un cAMP responsive element binding protein (CREB). Il modello con il secondo miglior punteggio conteneva gli stessi elementi eccetto CREB. Sono emersi inoltre quattro altri modelli con un buon quality score. Tutti i modelli sono elencati nella figura 2B. Questi modelli contenevano varie combinazioni dei siti di legame per fattori di trascrizione per SORY (cinque modelli), MAZ (quattro modelli), CMYB (tre modelli) ed ETS (due modelli) e numerosi altri modelli contenevano singoli siti di legame per Octamer, EGRF, ZBPF, o MZF1.

### **Modello di Framework ed Identificazione di Ipotetici Regolatori dei Fattori di Trascrizione**

Fino a questo punto, l'identificazione dei cis-elements attraverso l'analisi delle frameworks può solo essere considerata

una predizione e non stabilisce se qualcuno di questi modelli o di questi cis-elements sia realmente funzionalmente importante.

In pratica, più fattori di trascrizione della stessa famiglia possono effettivamente legarsi ad un qualsiasi sito di legame tra quelli identificati nei nostri modelli, e se i modelli sono effettivamente funzionanti le dinamiche di legame dei fattori di trascrizione ai molteplici siti di una framework possono essere funzionali, da soli od in concerto, per attivare o reprimere l'espressione genica e regolare il fenotipo indifferenziato o i primi steps di differenziamento.

Abbiamo allora definito che, per essere funzionalmente significativo, il nostro modello doveva soddisfare tre criteri fondamentali, anche per dimostrare la rilevanza trascrizione di uno o più di questi fattori di trascrizione nelle ESc. Ma, a causa della complessità della valutazione di ognuno dei modelli per ognuno dei potenziali fattori di trascrizione coinvolti abbiamo scelto di limitare la nostra scelta ad un solo candidato presente nella nostra lista EmS. Quindi i criteri da soddisfare sono stati: primo, il candidato doveva essere presente nelle cellule staminali pluripotenti. Secondo, almeno alcuni dei candidati dovevano effettivamente legare il sito per il legame sul promotore predetto

dal nostro modello e, terzo, il fattore di trascrizione candidato doveva essere in grado di regolare aspetti critici del fenotipo delle cellule staminali pluripotenti, come proliferazione, differenziamento od apoptosi.

### ***Potenziali trans-Binding Factors Presenti nelle Cellule ES***

A tutt'oggi nel topo sono stati identificati almeno 25 membri della famiglia Sry (Hmgb3, Hmga1, Sox1-Sox19, Sox21, Sox30 Sry, Hbp1), 13 membri della famiglia Creb (Atf1-Atf7, Creb1, Creb3, Creb5, Crebbp, Crem, Nfil3), 3 membri della famiglia Myb (Mybl, Mybl1, B-Myb) ed un singolo membro della famiglia Maz. Dai risultati della RT-PCR quantitativa e dalle valutazioni emerse dal SAGE nelle ESc, abbiamo identificato due Sox (Sox2 and Sox4), ma il più abbondante era di gran lunga Sox2. Abbiamo inoltre osservato SAGE tags per 2 proteine Creb (Atf4 and Atf6), ma dalla RT'PCR quantitativa abbiamo anche rilevato Creb 1 e Creb 2 così come trascritti per B-Myb e Maz.

All'analisi mediante immuofluorescenza, i fattori di trascrizione B-MYB, SOX2 e Creb1,3 erano tutti presenti nelle ES ed EG indifferenziate, sebbene il segnale per MAZ fosse solo debolmente rilevabile.

Dopo sottrazione di Lif dal mezzo di coltura, per indurre il differenziamento, l'RNA per Creb 1,3 risultava incrementato nei passaggi iniziali del differenziamento, ma in maniera non significativa (dati non mostrati).

Gli altri tre fattori di trascrizione, comunque, dimostravano una risposta dinamica a livello sia di RNA che di proteine, infatti Sox2, proteine e trascritto, si riduceva con il differenziamento, MAZ mostrava un rapido incremento nella sua distribuzione intracellulare (Figura 3B), associato ad un modesto incremento del contenuto di RNA.

I trascritti per B-Myb aumentavano in modo transiente, mostrando un picco alla quarta ora per poi decrescere durante il differenziamento, ma la proteina mostrava, in generale, una riduzione nelle EG-1 durante il differenziamento (Figura 3C).

Questi ultimi dati ci hanno suggerito che non solo Sox2 ma anche Maz e B-Myb erano regolati in maniera dinamica durante il differenziamento precoce delle cellule pluripotenti. Dal momento che né MAZ né CREB erano tra i trascritti identificati nella nostra lista di più abbondanti nelle cellule ES/EG e siccome Sox2 è risaputo essere critico per le ESc, abbiamo focalizzato la nostra

attenzione su B-Myb che risultava tra i più abbondanti trascritti nelle ESc.

### ***Chromatin Immunoprecipitation (ChIP)***

Abbiamo utilizzato il DNA estratto dal ES-R1, per determinare se B-Myb potesse legare direttamente le regioni del promotore dei geni identificate dalla analisi delle frameworks.

Per l'analisi ChIP le cellule sono state tripsinizzate ed i nuclei sono stati trattati con formaldeide per legare i fattori di trascrizione al DNA in maniera irreversibile. I lisati cellulari sono stati immunoprecipitati con anticorpi per B-Myb ed in seguito sono state effettuate RT-PCR per amplificare i frammenti dei promotori candidati per legame con B-Myb (nanog, sox2).

La porzione prossimale di ognuno dei promotori di questi geni conteneva numerosi siti di legame per B-Myb, e dal momento che è stato in precedenza dimostrato il legame di B-Myb al promotore di myc (Nakagoshi et al., 1992; Tashiro et al., 2004), tali siti sono stati utilizzati come ipotetici controlli positivi, prima di verificare il legame di B-Myb ai siti predetti del promotore. Come controlli negativi per l'immunoprecipitazione sono stati utilizzati anticorpi anti IgG di coniglio.

Dopo l'immunoprecipitazione con anticorpi contro B-Myb, si sono ottenuti i frammenti di PCR predetti per myc (Figura 4A). L'analisi mediante ChIP ha inoltre dimostrato che B-Myb legava i promotori di nanog (Figure 4A e 4B) e sox2, e che in ognuno di tali casi non si osservava amplificazione dopo immunoprecipitazione con gli anticorpi anti IgG di coniglio.

Questi dati ci hanno indicato che B-Myb aveva la capacità di legare i promotori ad i siti predetti dal nostro modello di analisi delle Frameworks. Comunque, solo il legame di B-Myb al promotore del gene sox2 è apparso regolato dinamicamente al momento della sospensione del Lif e con l'induzione del differenziamento (sospensione di Lif ed FBS) (Figura 4C). Questa è stata la prima evidenza concreta a suggerirci la possibile implicazione di B-Myb nella regolazione dell'espressione genica nelle cellule staminali embrionali.

### **B-Myb è un Regolatore della Proliferazione e del Differenziamento nelle cellule ES**

Per meglio chiarire, a questo punto, il ruolo di B-Myb nel self-renewal e nel differenziamento delle cellule Staminali

Embionali, abbiamo generato un clone di ESc overesprimenti B-Myb murino. In queste linee clonali, il fattore di trascrizione B-Myb era sotto il controllo del promotore della fosfoglicerato chinasi (PGK) e l'espressione era visualizzabile mediante l'espressione bicistrionica di una proteina di fluorescenza verde (Figura 5A). Tutte le linee clonali crescevano compatte in colonie e presentavano positività alla colorazione per la fosfatasi alcalina (Figura 5B). Tre delle linee clonali generate presentavano un incremento significativo del trascritto per B-Myb alla PCR quantitativa.

A seconda della linea clonale analizzata, infatti, l'incremento del trascritto variava da circa il 50 al 90% ed il maggiore incremento era riscontrabile nelle linee PGK-10 e PGK-11; comunque l'incremento della proteina B-Myb risultava modesto e non significativo (Figura 5C).

In tali linee clonali né il trascritto di Oct4 né quello di Nanog risultava incrementato in maniera significativa, invece il trascritto di Sox2 incrementava in modo significativo (Figura 5D). Inoltre, non abbiamo osservato incremento significativo nell'abbondanza dei trascritti di markers di differenziamento come CoupTF, Fgf5,

Sox17, TdGF1 (Cripto)(Figura 5E), Gata4, Brachyury or Nkx2.5 (non mostrati).

In contrasto, il trascritto del marker associato al neuroectoderma Nestin, ma non Sox1, mostrava un incremento significativo nelle nostre linee clonali rispetto alle linee parentali di R1. Questo incremento è stato confermato da esperimenti di immunofluorescenza (Figure 5E). Al fine di valutare la proliferazione cellulare in vitro, abbiamo utilizzato il WST-1. L'incremento di conversione del substrato nelle linee clonali PGK-10 e PGK-11, rispetto alle linee parentali di controllo, ci ha suggerito che nei nostri cloni si potesse osservare sia un'incremento di proliferazione che una possibile riduzione del tasso di apoptosi.

Per stabilire se l'overespressione di B-Myb fosse sufficiente a mantenere le Esc allo stato indifferenziato, abbiamo eliminato il LIF dal mezzo di coltura. Tale sospensione, per 48 ore, ha portato ad una riduzione, ma non alla scomparsa dell'attività della fosfatasi alcalina (Figura 5G). Era presente, inoltre, anche una riduzione del rapporto nucleo citoplasmatico in tutte le linee cellulari ed erano evidenti delle differenze morfologiche tra le linee clonali e le R1 di controllo. In particolare, le linee clonali hanno

cominciato a crescere in monolayer, nel quale era possibile distinguere chiaramente le singole cellule. Tale comportamento in coltura faceva rassomigliare, in particolare, la morfologia dei nostri cloni a quella delle cellule di carcinoma embrionali P19.

In tutte le linee cellulari osservate, l'abbondanza dei trascritti per Oct4, Sox2 e Nanog era ridotta: nelle linee cellulari overesprimenti, tuttavia, la riduzione di Sox2 era meno pronunciata che nelle linee parentali di controllo R1, confermando la nostra precedente osservazione. Abbiamo inoltre osservato altre due importanti differenze al 48 ore dalla sospensione del Lif.

Primo, il marker di differenziamento CoupTF era aumentato nelle R1 parentali, ma non nelle linee clonali; secondo, il marker FGF5 era aumentato in tutte le linee cellulari, ma l'entità dell'incremento era di circa 10 volte maggiore nelle linee clonali rispetto alle R1 parentali ( $p < 0.01$ ).

La presenza di attività della fosfatasi alcalina e di Oct4, il marcato incremento di FGF5 e la morfologia a monolayer delle cellule assunta durante la crescita in assenza di Lif ci hanno fatto ipotizzare la presenza in coltura di early primitive ectoderm-like (EPL) cells (Rathjen et al., 1999). Abbiamo inoltre osservato un modesto incremento una modesta riduzione di Zfp42 (Rex1) a 48

ore dalla riduzione del Lif, anche questa caratteristica tipica delle EPL. È inoltre da evidenziare, che continuando a mantenere in coltivazione in assenza di Lif le nostre linee clonali, si sarebbero potuti osservare altri markers di differenziamento. Noi abbiamo ipotizzato che l'overespressione di B-Myb da sola non è sufficiente a mantenere le ESc allo stato indifferenziato (o come EPL), ma è sufficiente per modificare i meccanismi che controllano il differenziamento di tali cellule in assenza di Lif.

## DISCUSSIONE

Il presente studio è stato ideato allo scopo di identificare potenziali fattori di trascrizione e frameworks di cis-elements di geni altamente espressi e necessari per il mantenimento di cellule staminali embrionali pluripotenti. Per tale scopo, abbiamo utilizzato un nuovo approccio per identificare nuovi fattori di trascrizione implicati nel mantenimento delle cellule ES allo stato indifferenziato.

Il razionale è stato simile a quello utilizzato in due recenti studi, nei quali i siti di legame per fattori di trascrizione sono stati identificati mediante un processo ripetitivo per l'analisi dei promotori genici (Dohr et al., 2005; Seifert et al., 2005) e nei quali la previsione di specifiche frameworks si è basata su tre criteri fondamentali (Cartharius et al., 2005; Seifert et al., 2005). Più specificamente, presentiamo una strategia che definisce potenziali frameworks derivanti dall'analisi quantitative del trascrittoma e forniamo alcune evidenze a supporto di un ruolo funzionale per due dei fattori di trascrizione identificati nelle cellule ES.

Questa strategia si è basata, per prima cosa, sulla valutazione quantitativa del prodotto genico in vari tipi di cellule staminali pluripotenti. I risultati dell'analisi del SAGE sono stati poi paragonati con altre librerie SAGE derivanti da linee cellulari adulte e tessuti, ed il tutto seguito dall'analisi mediante PCR quantitativa. Ciò ci ha portati alla elaborazione di una Problem Specific List costituita dai geni di maggior interesse. Abbiamo, in seguito, analizzato tali geni per identificare eventuali frameworks comuni ed abbiamo, infine, eseguito uno screening dell'intero genoma murino per identificare altri geni che contenessero geni candidati che presentassero le frameworks da noi identificate (p.es. myc).

Da tale analisi sono emersi sei modelli. Il modello "migliore" era quello con il quality score più basso ed era costituito dalla distribuzione spaziale dei siti di legame per Maz, Mybl, Sory e Creb nei promotori di myc, B-Myb, pou5f1, nanog e sox2 (Figura 2). Un secondo modello emerso dalla nostra analisi conteneva anch'esso siti di legame per Maz, Mybl e Sory, ma aveva un quality score leggermente più basso.

Questa strategia, inoltre, ci ha indicato che i siti di legame per *octamer* e *sory* sono prevalenti nel set di geni EmS ma non

nel set SnD. È rilevante il fatto che i membri di ognuna di queste famiglie di fattori di trascrizione risultavano prevalentemente espressi nelle linee cellulari di ES ed EG e mostravano profili di espressione dinamici, in seguito all'induzione del differenziamento, suggerendoci un loro ruolo funzionale. Noi non abbiamo potuto dare dimostrazione diretta a suffragio di un ruolo critico svolto dalle frameworks predetta dal nostro modello sperimentale ed non siamo riusciti a mostrare alcun potenziale sito di legame per Nanog in tale studio a causa delle limitazioni esistenti nel software al tempo in cui l'analisi è stata condotta. Tali riscontri unitamente alle limitazioni suddette ci hanno indotti a focalizzarci su due fattori di trascrizione: Maz e B-Myb.

Il fattore di trascrizione Maz contiene regioni C2H2-type zinc-finger capaci di legare cis-elements ricchi in GC distribuiti nei promotori di numerosi geni nei mammiferi (Bossone et al., 1992; Komatsu et al., 1997; Song et al., 2003). Sebbene esso non sia particolarmente abbondante nelle cellule staminali pluripotenti indifferenziate, noi abbiamo dimostrato che la proteina MAZ presentava notevoli cambiamenti nella distribuzione intracellulare ed una aumentata espressione in seguito all'induzione del differenziamento, sebbene i livelli di RNA fossero solo

marginalmente aumentati. È stato infatti descritto che le deacetilasi istoniche sono coinvolte nell'autorepressione del gene di MAZ, mentre l'espressione della metiltransferasi I si associa ad una repressione dell'espressione di MAZ ad opera di Sp1 (Song et al., 2003).

Tale risultato sarebbe coerente con le evidenze sperimentali che dimostrano che le cellule staminali presentano zone di forte repressione della trascrizione dovute alla deacetilazione degli istoni (Gan et al., 2007; Lee et al., 2004). L'attivazione di MAZ, in seguito all'induzione del differenziamento, potrebbe, a questo punto, essere utilizzata come marker per le variazioni dello stato di acetilazione e verosimilmente di metilazione degli istoni che potrebbero essere segnali di controllo trascrizionale di singoli geni ad opera di Maz in seguito all'induzione del differenziamento.

Diversamente da Maz, il fattore di trascrizione B-Myb è interessante come regolatore nelle cellule staminali pluripotenti indifferenziate. Altri due membri della famiglia *myb* sono già stati associati alla regolazione delle cellule staminali, A-Myb e C-Myb. Il primo è particolarmente abbondante nelle cellule germinali maschili ed è presente in quantità inferiori nelle ovaie, nel cervello

e nelle cellule B del centro germinale. La perdita di tale proteina conduce a difetti nella spermatogenesi (Toscani et al., 1997).

C-Myb è generalmente presente nelle cellule ematopoietiche immature ed è assolutamente essenziale al controllo della proliferazione e del differenziamento delle cellule staminali ematopoietiche (Sandberg et al., 2005). Al contrario, B-Myb è ubiquitariamente espresso e risulta essere particolarmente abbondante nelle prime fasi dello sviluppo, nelle quali è anche molto abbondante (Tanaka et al., 1999). Esso costituisce il gene ancestrale di una famiglia di fattori di trascrizione altamente conservata implicate nella regolare crescita cellulare, nello sviluppo e nel differenziamento (Joaquin & Watson, 2003; Sala, 2005; Sala & Watson, 1999).

Numerose evidenze sperimentali, inoltre, supportano il ruolo essenziale di B-MYB nel controllo trascrizionale dei trascritti predominanti nelle ESc. Primo, B-Myb è abbondante nelle cellule staminali embrionali murine ed è espresso dinamicamente a nel differenziamento (Figura 3). Secondo, B-Myb lega dinamicamente al promotore di *sox2* endogeno nelle ESc murine, suggerendo un ruolo potenziale nel controllo trascrizionale (Figura 4).

Terzo, i trascritti di Sox2 sono elevati nei cloni che overesprimono B-Myb e tali trascritti non diminuiscono così rapidamente in seguito alla sottrazione del siero e del, rispetto alle cellule R1 parentali di controllo. La capacità di B-Myb di legarsi a promotori di geni endogeni critici nelle cellule ES, ne suggerisce una funzione di regolatore della trascrizione.

Inoltre, un ruolo cruciale nel mantenimento dello stato indifferenziato delle cellule ES è supportato da altri due studi. Iwai e coll. hanno dimostrato che la perdita della funzione di B-Myb, attraverso l'induzione condizionale di una forma interferente di dominante negativo di B-Myb, ha causato, nelle cellule ES, cambiamenti morfologici caratterizzati dalla dissociazione di colonie di cellule, strettamente adese tra loro, in singole cellule disperse e ne ha inoltre causato l'arresto del ciclo cellulare in fase G1-S.

Quest'ultima evidenza si accompagnava ad una ridotta espressione di  $\alpha$ 1-integrina ed al distacco delle cellule (Iwai et al., 2001). Questi esperimenti rimangono, tuttavia, aperti ad altre interpretazioni dal momento che le modificazioni alla molecola B-MYB potrebbero averne alterata la funzione intrinseca (Johnson et al., 2002; Tashiro et al., 1995).

Tanaka e coll. Hanno, inoltre, dimostrato che la perdita di Mybl2, mediante gene targeting ha portato a letalità embrionale precoce allo stadio embrionale E4.5 - E6.5 (Tanaka et al., 1999). Sebbene rimanga non spiegato se la morte sia stata il risultato diretto degli effetti sulla proliferazione cellulare, la coltivazione in vitro delle blastocisti di questi topi knockout, suggerisce che B-Myb sia assolutamente necessario per lo sviluppo e l'espansione della ICM (Inner Cell Mass) negli embrioni post-impianto ma che non lo sia per la formazione del TE (Trophoectoderm). Inoltre, nessuno è stato in grado, fino ad ora, di generare linee di ES che mancassero di B-Myb, avvalorando il ruolo cruciale di questo fattore di trascrizione per la loro generazione.

Riassumendo, la capacità delle cellule staminali embrionali di differenziarsi in ogni tipo cellulare dell'organismo rimane indiscussa e l'utilizzo terapeutico di tali cellule è, a tutt'oggi, limitato dalla limitata conoscenza che si ha dei meccanismi che ne controllano il differenziamento. In questo lavoro abbiamo utilizzato informazioni ottenute dai vari progetti genoma per identificare due nuovi fattori di trascrizione, B-Myb e Maz, coinvolti nel mantenimento o nel differenziamento di cellule staminali pluripotenti. Noi prevediamo che l'uso di un'approccio analogo ad

altri ampi set di dati possa facilitare ulteriormente l'identificazione di fattori critici nella biologia delle cellule staminali e che nel momento in cui ciò si realizzasse la nostra comprensione dei meccanismi che regolano il destino delle cellule ES in vitro accrescerebbe notevolmente.

## BIBLIOGRAFIA

- Anisimov, S.V., Tarasov, K.V., Tweedie, D., Stern, M.D., Wobus, A.M. & Boheler, K.R. (2002). SAGE identification of gene transcripts with profiles unique to pluripotent mouse R1 embryonic stem cells. *Genomics* 79: 169-176.
- Bhattacharya, B., Miura, T., Brandenberger, R., Mejido, J., Luo, Y., Yang, A.X., Joshi, B.H., Ginis, I., Thies, R.S., Amit, M., Lyons, I., Condie, B.G., Itskovitz-Eldor, J., Rao, M.S. & Puri, R.K. (2004). Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood* 103, 2956-64.
- Bossone, S.A., Asselin, C., Patel, A.J. & Marcu, K.B. (1992). MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination. *Proc Natl Acad Sci U S A* 89: 7452-6.
- Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R. & Young, R.A. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122: 947-56.

- Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M. & Werner, T. (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 21: 2933-42.
- Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K. & Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* 132: 885-96.
- Doetschman, T.C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. (1985). The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* 87: 27-45.
- Dohr, S., Klingenhoff, A., Maier, H., Hrabe de Angelis, M., Werner, T. & Schneider, R. (2005). Linking disease-associated genes to regulatory networks via promoter organization. *Nucleic Acids Res* 33: 864-72.
- Gan, Q., Yoshida, T., McDonald, O.G. & Owens, G.K. (2007). Concise review: epigenetic mechanisms contribute to

pluripotency and cell lineage determination of embryonic stem cells. *Stem Cells* 25: 2-9.

Ginis, I., Luo, Y., Miura, T., Thies, S., Brandenberger, R., Gerecht-Nir, S., Amit, M., Hoke, A., Carpenter, M.K., Itskovitz-Eldor, J. & Rao, M.S. (2004). Differences between human and mouse embryonic stem cells. *Dev Biol* 269: 360-80.

Iwai, N., Kitajima, K., Sakai, K., Kimura, T. & Nakano, T. (2001). Alteration of cell adhesion and cell cycle properties of ES cells by an inducible dominant interfering Myb mutant. *Oncogene* 20: 1425-34.

Joaquin, M. & Watson, R.J. (2003). Cell cycle regulation by the B-Myb transcription factor. *Cell Mol Life Sci* 60: 2389-401.

Johnson, L.R., Johnson, T.K., Desler, M., Luster, T.A., Nowling, T., Lewis, R.E. & Rizzino, A. (2002). Effects of B-Myb on gene transcription: phosphorylation-dependent activity and acetylation by p300. *J Biol Chem* 277: 4088-97.

Keller, G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* 19: 1129-55.

Komatsu, M., Li, H.O., Tsutsui, H., Itakura, K., Matsumura, M. & Yokoyama, K.K. (1997). MAZ, a Myc-associated zinc finger

protein, is essential for the ME1a1-mediated expression of the c-myc gene during neuroectodermal differentiation of P19 cells. *Oncogene* 15: 1123-31.

Lee, J.H., Hart, S.R. & Skalnik, D.G. (2004). Histone deacetylase activity is required for embryonic stem cell differentiation. *Genesis* 38: 32-8.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J.C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci U S A* 90: 8424-8.

Nakagoshi, H., Kanei-Ishii, C., Sawazaki, T., Mizuguchi, G. & Ishii, S. (1992). Transcriptional activation of the c-myc gene by the c-myb and B-myb gene products. *Oncogene* 7: 1233-40.

Rathjen, J., Lake, J.A., Bettess, M.D., Washington, J.M., Chapman, G. & Rathjen, P.D. (1999). Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors. *J Cell Sci* 112: 601-12.

Richards, M., Tan, S.P., Tan, J.H., Chan, W.K. & Bongso, A. (2004). The transcriptome profile of human embryonic stem cells as defined by SAGE. *Stem Cells* 22: 51-64.

- Sala, A. (2005). B-MYB, a transcription factor implicated in regulating cell cycle, apoptosis and cancer. *Eur J Cancer* 41: 2479-84.
- Sala, A. & Watson, R. (1999). B-Myb protein in cellular proliferation, transcription control, and cancer: latest developments. *J Cell Physiol* 179: 245-50.
- Sandberg, M.L., Sutton, S.E., Pletcher, M.T., Wiltshire, T., Tarantino, L.M., Hogenesch, J.B. & Cooke, M.P. (2005). c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation. *Dev Cell* 8: 153-66.
- Sato, N., Sanjuan, I.M., Heke, M., Uchida, M., Naef, F. & Brivanlou, A.H. (2003). Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol* 260: 404-13.
- Seifert, M., Scherf, M., Epple, A. & Werner, T. (2005). Multievidence microarray mining. *Trends Genet* 21: 553-8.
- Song, J., Ugai, H., Nakata-Tsutsui, H., Kishikawa, S., Suzuki, E., Murata, T. & Yokoyama, K.K. (2003). Transcriptional regulation by zinc-finger proteins Sp1 and MAZ involves interactions with the same cis-elements. *Int J Mol Med* 11: 547-53.

- Sperger, J.M., Chen, X., Draper, J.S., Antosiewicz, J.E., Chon, C.H., Jones, S.B., Brooks, J.D., Andrews, P.W., Brown, P.O. & Thomson, J.A. (2003). Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc Natl Acad Sci U S A* 100: 13350-5.
- Stewart, C.L., Gadi, I. & Bhatt, H. (1994). Stem cells from primordial germ cells can reenter the germ line. *Dev Biol* 161: 626-8.
- Tanaka, Y., Patestos, N.P., Maekawa, T. & Ishii, S. (1999). B-myb is required for inner cell mass formation at an early stage of development. *J Biol Chem* 274: 28067-70.
- Tarasova, Y.S., Riordon, D.R., Tarasov, K.V. & Boheler, K.R. (2006). In vitro differentiation of mouse ES cells to muscle cells. In *Embryonic Stem Cells*, Notarianni, E. & Evans, M.J. (eds) A Practical Approach. Oxford University Press: New York. pp. 130-168.
- Tashiro, S., Sumi, T., Uozumi, N., Shimizu, T. & Nakamura, T. (2004). B-Myb-dependent regulation of c-Myc expression by cytosolic phospholipase A2. *J Biol Chem* 279: 17715-22.
- Tashiro, S., Takemoto, Y., Handa, H. & Ishii, S. (1995). Cell type-specific trans-activation by the B-myb gene product:

requirement of the putative cofactor binding to the C-terminal conserved domain. *Oncogene* 10: 1699-707.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S. & Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science*, 282: 1145-1147.

Toscani, A., Mettus, R.V., Coupland, R., Simpkins, H., Litvin, J., Orth, J., Hatton, K.S. & Reddy, E.P. (1997). Arrest of spermatogenesis and defective breast development in mice lacking A-myb. *Nature* 386: 713-7.

Velculescu, V.E., Zhang, L., Vogelstein, B. & Kinzler, K.W. (1995). Serial analysis of gene expression. *Science* 270: 484-487.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W. & Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. *Nature* 444: 364-8.

Werner, T. (2000). Identification and functional modelling of DNA sequence elements of transcription. *Brief Bioinform* 1: 372-80.

- Werner, T., Fessele, S., Maier, H. & Nelson, P.J. (2003).  
Computer modeling of promoter organization as a tool to  
study transcriptional coregulation. *Faseb J*, 17: 1228-37.
- Wobus, A.M. & Boheler, K.R. (2005). Embryonic stem cells:  
prospects for developmental biology and cell therapy.  
*Physiol Rev* 85: 635-78.
- Wobus, A.M., Guan, K., Yang, H.-T. & Boheler, K.R. (2002).  
Embryonic stem cells as a model to study cardiac, skeletal  
muscle and vascular smooth muscle cell differentiation. In  
*Methods Mol Biol*, Turksen, K. (ed), Vol. 185. Humana  
Press: Totowa, New Jersey. pp. 127-156.
- Yamane, T., Dylla, S.J., Muijtjens, M. & Weissman, I.L. (2005).  
Enforced Bcl-2 expression overrides serum and feeder cell  
requirements for mouse embryonic stem cell self-renewal.  
*Proc Natl Acad Sci U S A* 102: 3312-7.
- Ying, Q.L., Nichols, J., Chambers, I. & Smith, A. (2003). BMP  
induction of Id proteins suppresses differentiation and  
sustains embryonic stem cell self-renewal in collaboration  
with STAT3. *Cell* 115: 281-92.

**Tabella 1: Problem Specific List generata dall'analisi SAGE**

Mm.5090	Tdgf1, teratocarcinoma-derived growth factor
Mm.6047	Nanog, Nanog homeobox
Mm.4594	Mybl2, myeloblastosis oncogene-like 2
Mm.17031	Pou5f1, POU domain, class 5, transcription factor 1
Mm.4956	Fgf4, fibroblast growth factor 4
Mm.5025	Etv4, ets variant gene 4 (E1A enhancer binding protein, E1AF)
Mm.4541	Sox 2, SRY-box containing gene 2
Mm.46461	Tdh, L-threonine dehydrogenase
Mm.10205	Utf1, undifferentiated embryonic cell transcription factor 1
Mm.235550	Esrrb, estrogen related receptor, beta
Mm.26883	2410146L05Rik RIKEN cDNA 2410146L05 gene
Mm.285848	Zfp42 zinc finger protein 42
Mm.28369	Fbxo15, F-box only protein 15

I numeri Mm fanno riferimento ai numeri di accessi in Unigene per *Mus musculus*.

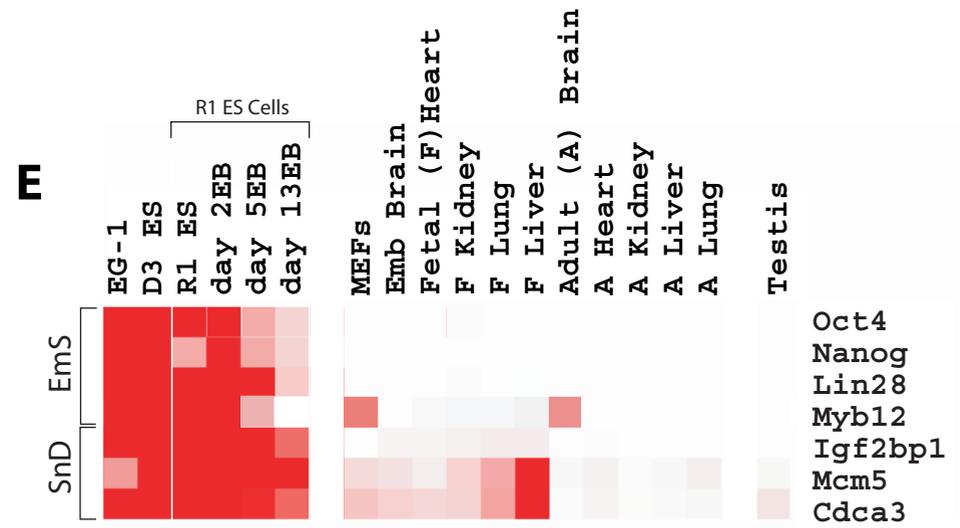
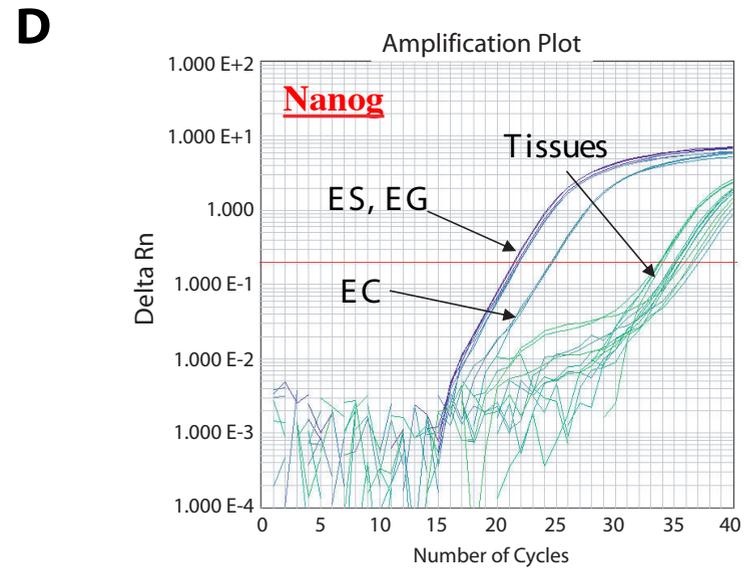
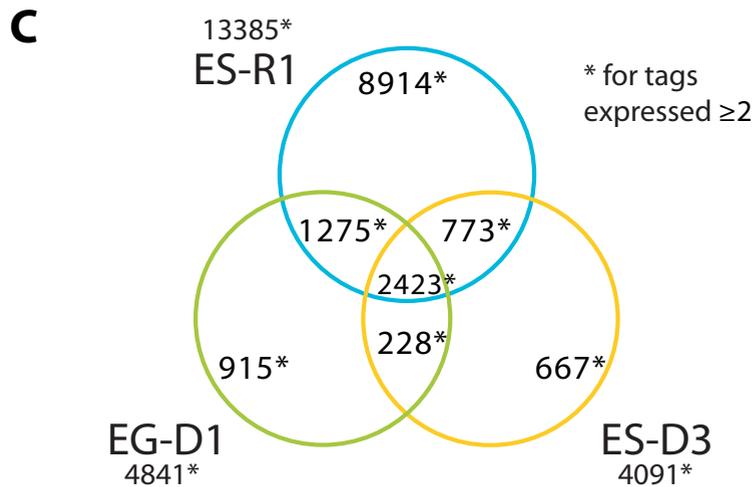
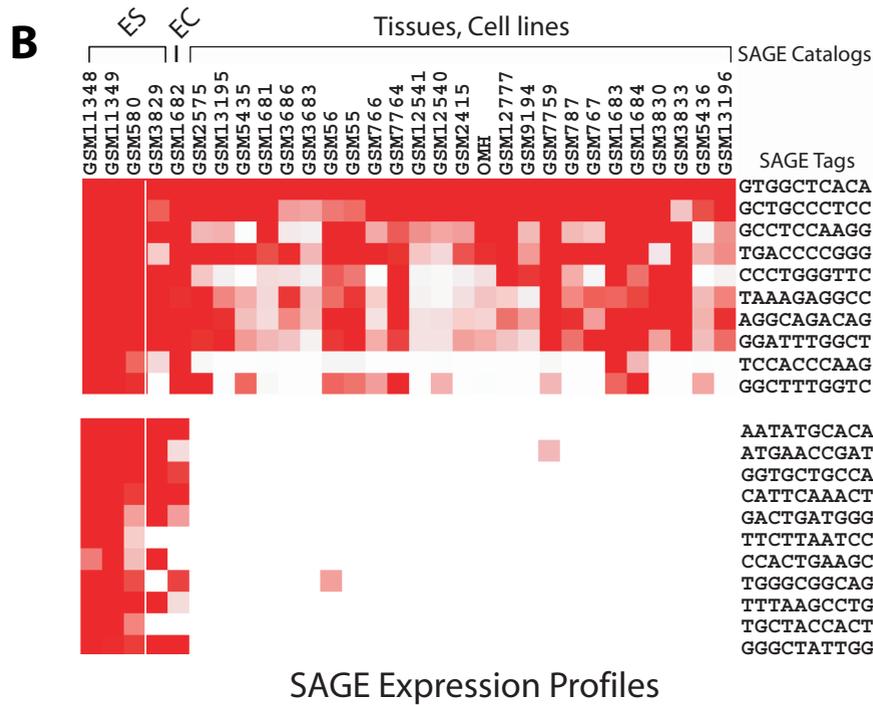
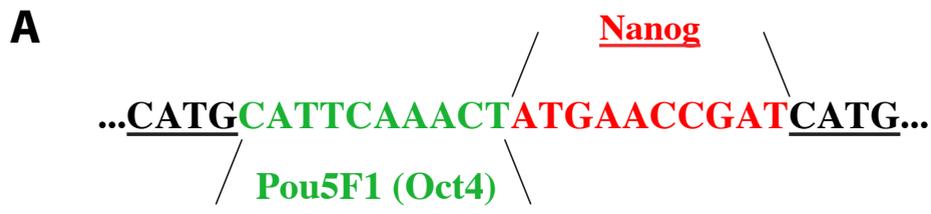
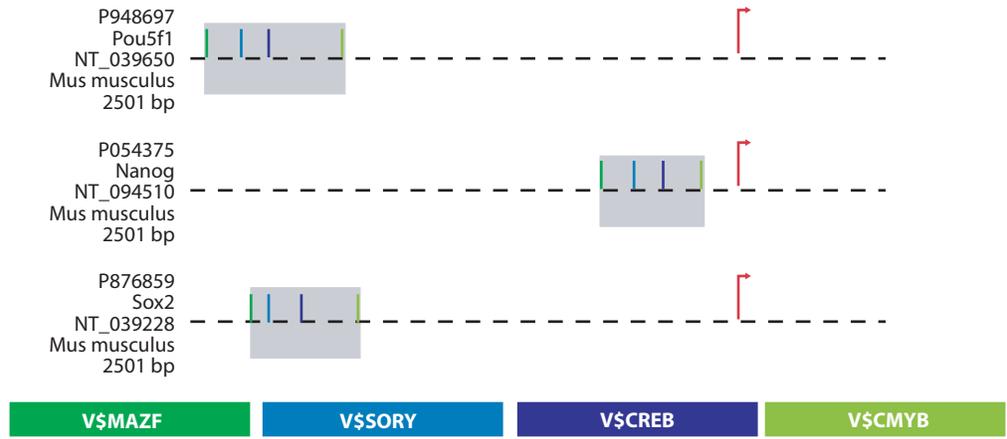


Figure 1

**A****B**

Model Name	# of Promoters with Framework in PSL	# of Promoters in the mouse genome containing this framework (69108 promoters)	Quality Score
<b>MAZF, SORY, CREB, CMYB</b>	5	31	0.00126
<b>MAZF, SORY, CMYB</b>	7	211	0.00611
SORY, ETSF, MAZF, ETSF	5	325	0.0131
SORY, MZF1, CMYB	5	360	0.146
SORY, ETSF, OCT1	6	838	0.0284
MAZF, ZBPF, EGRF	5	>2000	>0.08

Number of recognized promoters in the mouse genome: 69,108

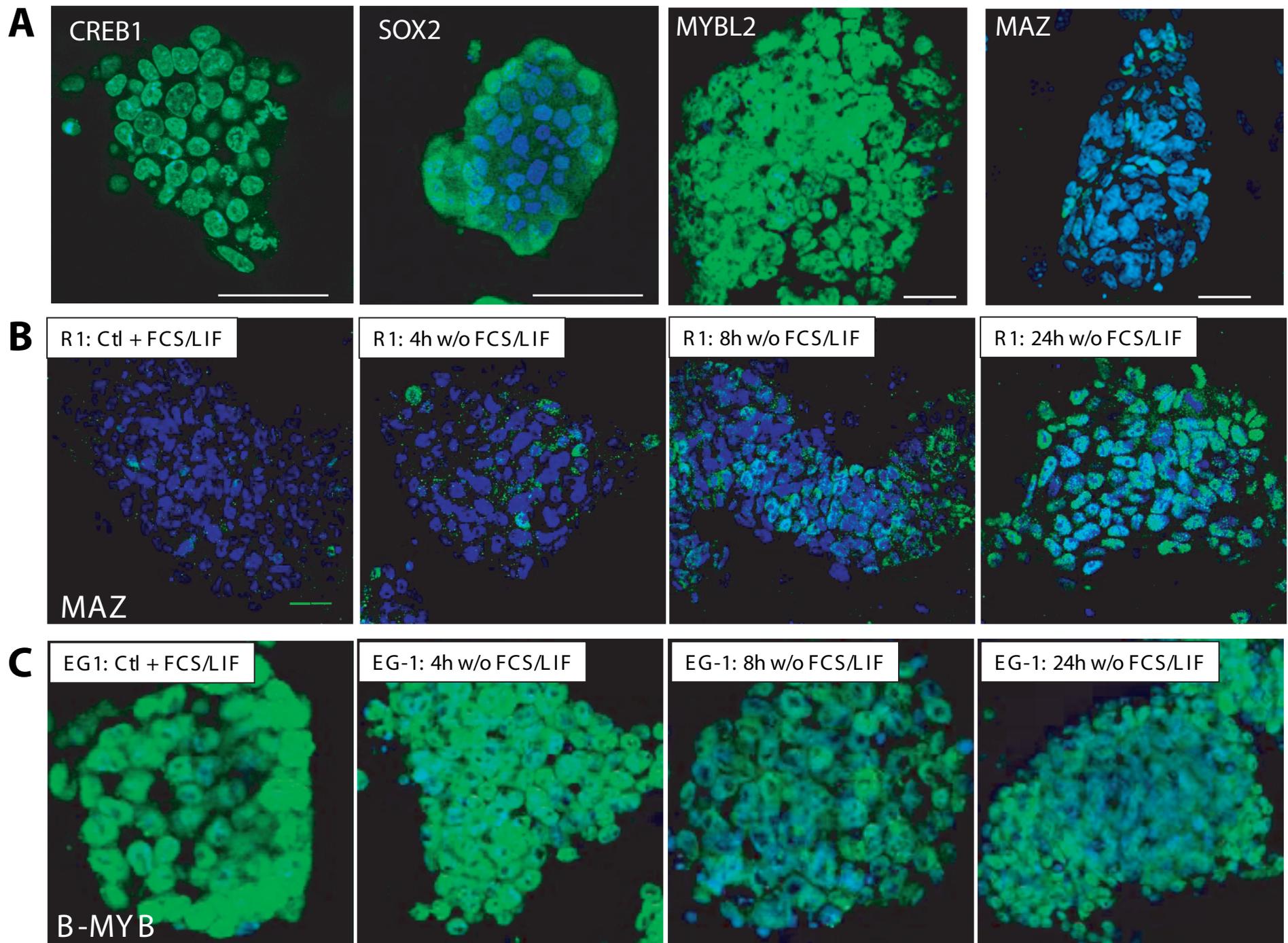


Figure 3

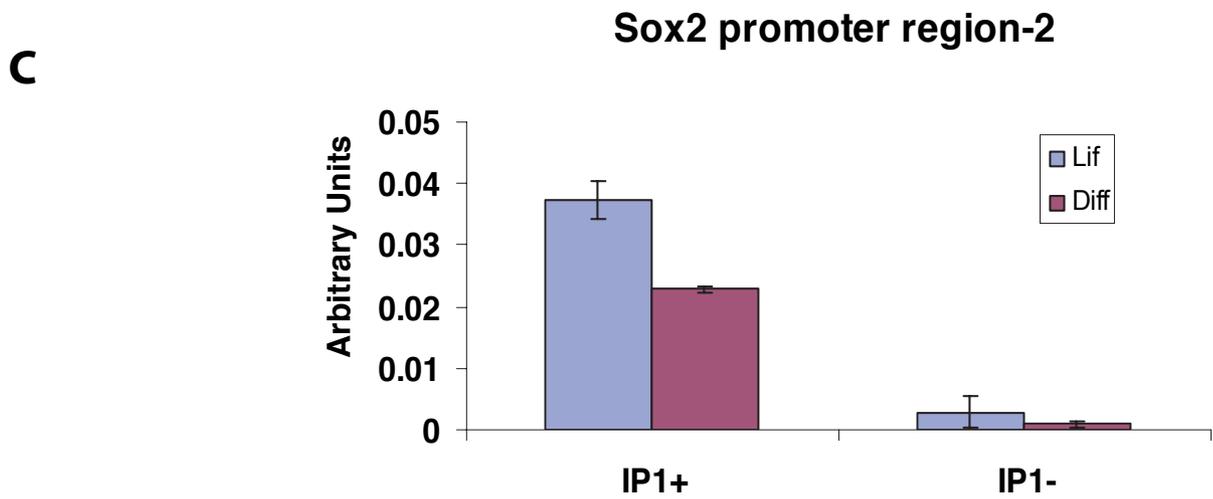
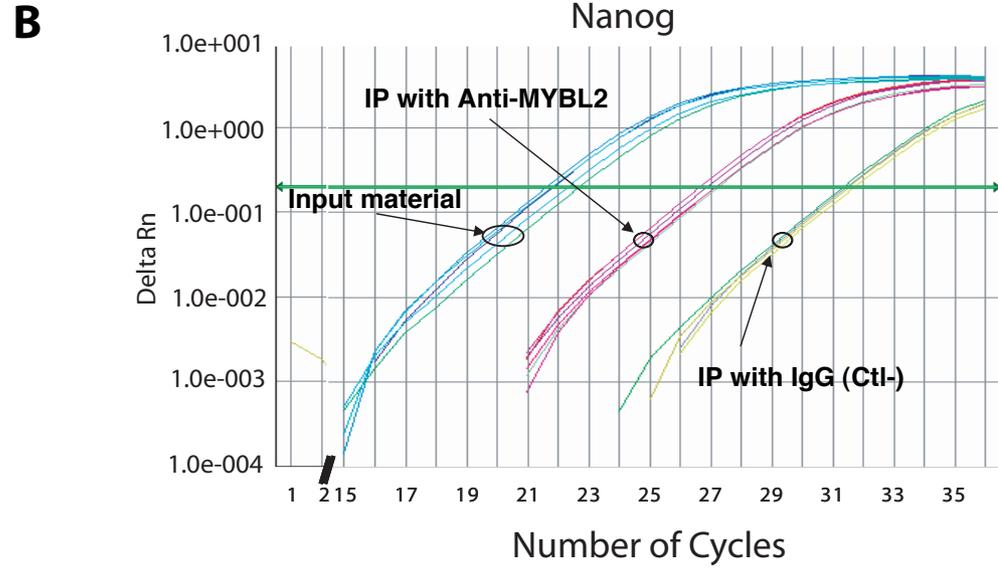
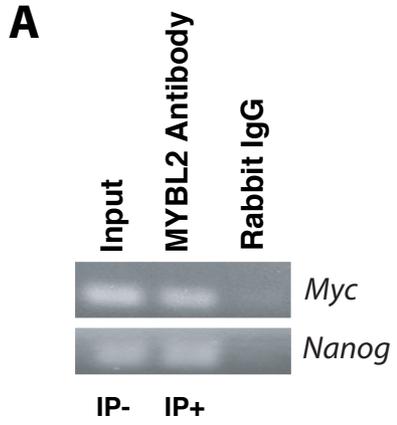


Figure 4

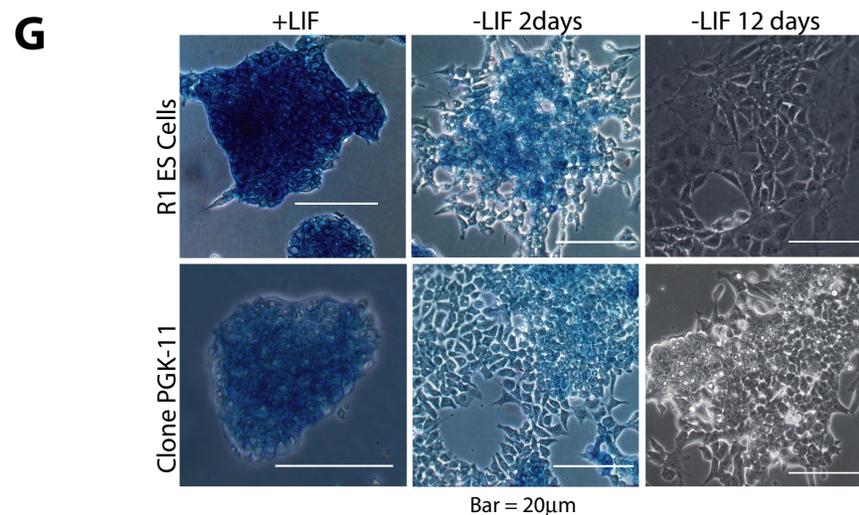
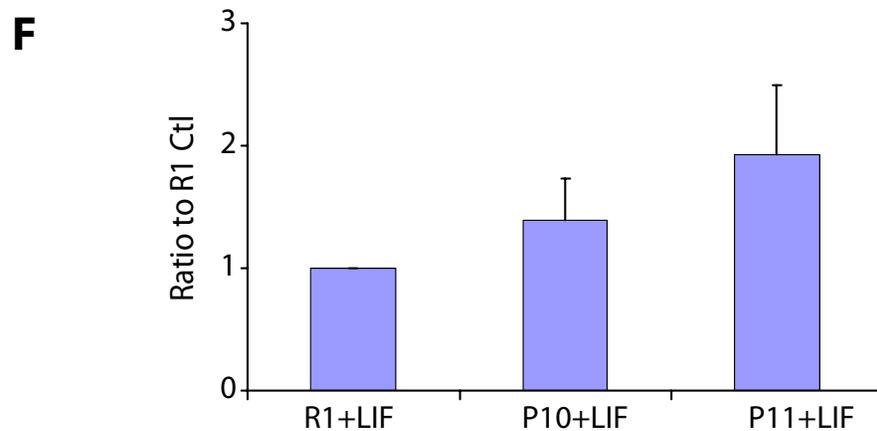
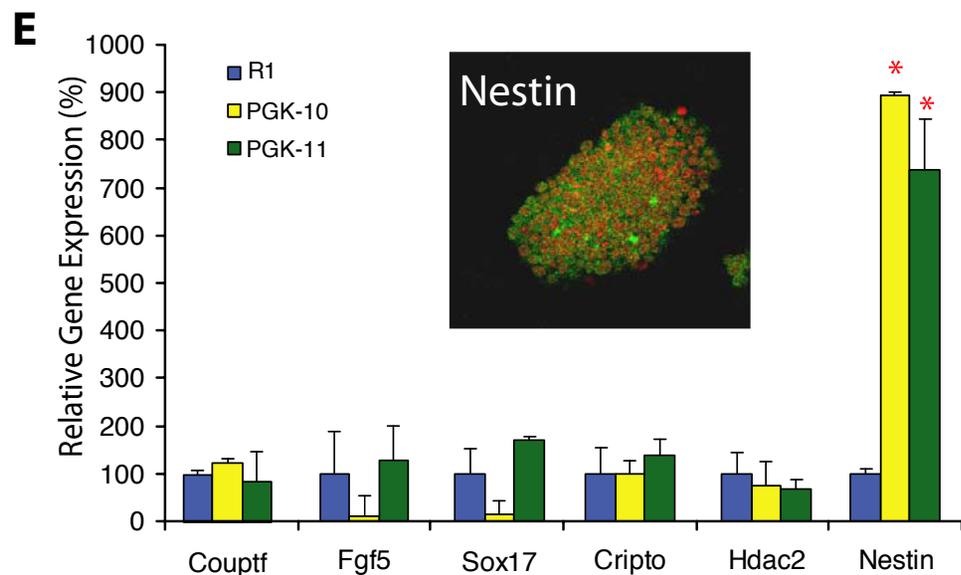
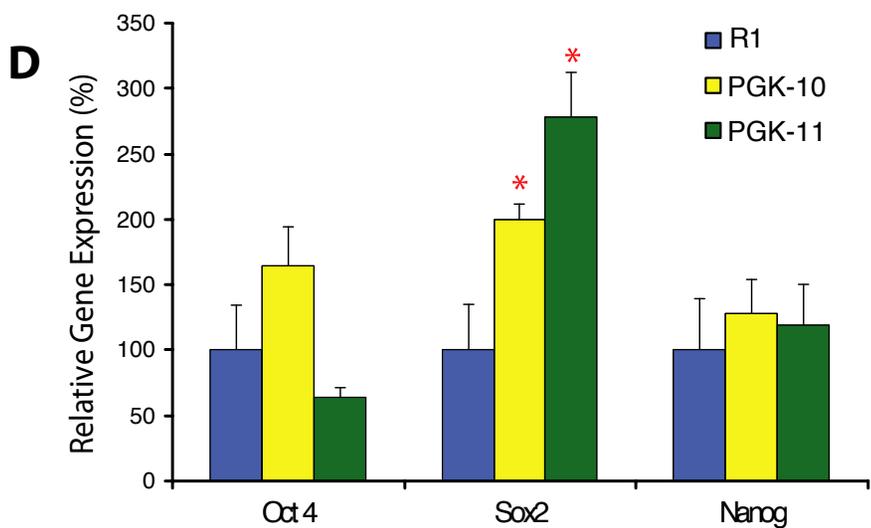
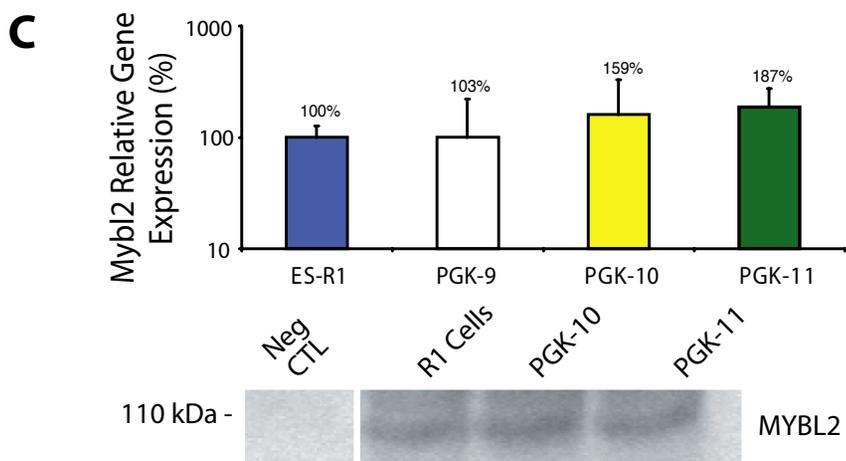
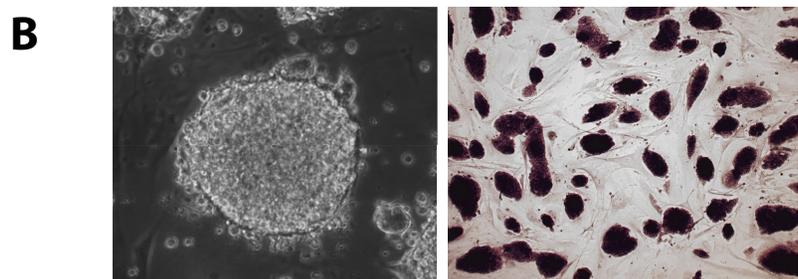
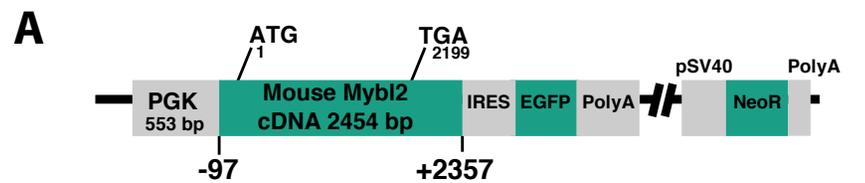


Figure 5