



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

FACULTAD DE CIENCIAS

ANÁLISIS BIOINFORMÁTICO DEL  
PROMOTOR DE *SNO* DE *HOMO SAPIENS* Y  
*MUS MUSCULUS*

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGA

P R E S E N T A :

ELISA DOMÍNGUEZ HÜTTINGER

TUTORA:

DRA. MARINA MACÍAS SILVA

2008



FACULTAD DE CIENCIAS  
UNAM



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## Hoja de Datos del Jurado

<p>1. Datos del alumno Domínguez Hüttinger Elisa 52 07 63 72 Universidad Nacional Autónoma de México Facultad de Ciencias Biología 301540377</p>
<p>2. Datos del tutor Dra Marina Macías Silva</p>
<p>3. Datos del sinodal 1 Dr Emilio Joaquín Córdova Alarcón</p>
<p>4. Datos del sinodal 2 M en C José Antonio Vilchis Peluyera</p>
<p>5. Datos del sinodal 3 Dra Patricia Ramos Morales</p>
<p>6. Datos del sinodal 4 Dr Arturo Carlos II Becerra Bracho</p>
<p>7. Datos del trabajo escrito Análisis bioinformático del promotor de <i>sno</i> de <i>Homo sapiens</i> y <i>Mus musculus</i> 120 p 2008</p>

El presente trabajo se realizó bajo la dirección de la Dra. Marina Macías Silva, en el Laboratorio 225N del Departamento de Biología Celular en el Instituto de Fisiología Celular de la UNAM. El proyecto fue apoyado por los donativos del CONACYT 49493 y DGAPA-PAPIIT IN2142206.

# Agradecimientos

Un agradecimiento especial a

M. en C. Ángeles Concepción Tecalco Cruz

A los miembros del Jurado:

Dra. Marina Macías Silva

Dr. Emilio Joaquín Córdova Alarcón

M. en C. Alfonso José Vilchis Peluyera

Dra. Patricia Ramos Morales

Dr. Arturo Carlos II Becerra Bracho

A la

Universidad Nacional Autónoma de México

Particularmente a la Facultad de Ciencias

*A la familia-banda*



# Índice

Resumen	1
Lista de abreviaturas	3
Introducción	5
La transducción de señales	5
TGF $\beta$	7
Papel biológico	7
Papel del TGF $\beta$ en el cáncer	8
Vía de señalización del TGF $\beta$	9
Regulación de la transcripción mediada por el TGF $\beta$	13
Los factores de transcripción del TGF $\beta$ : Las Smad	13
Unión de las Smad al DNA	15
Genes blanco del TGF $\beta$	15
Factores de Transcripción asociados a las Smad	18
El promotor de un gen	21
Contexto celular	23
Sno	24
Mecanismos de acción de Sno y Ski	25
Papel biológico de Ski y Sno	26
Regulación bifásica de <i>sno</i> por TGF $\beta$	27
Antecedentes	29
Regulación de <i>sno</i> por TGF $\beta$	29
TGF $\beta$ induce la transcripción del gen <i>sno</i>	29
Zhu <i>et al.</i> encuentran SBEs y un SIE en el promotor de <i>sno</i> de <i>Mus musculus</i>	29
Ruoyun y colaboradores detectan un elemento de respuesta a HGF en el promotor del gen <i>sno</i> de <i>Homo sapiens</i>	30
Planteamiento del problema	32
Hipótesis	33
Objetivos	34
Objetivo general	34
Objetivos específicos	34
Metodología	35
Obtención de la secuencia promotora	36
Determinación de los Sitios de Inicio de la Transcripción	40
Búsqueda de cajas regulatorias	42
Predicción de Sitios de Unión a Factores de Transcripción	44
Validación del programa	44
Detección de TFBS putativos	45
Resultados	47
Obtención de la secuencia promotora de <i>sno</i>	47
Validación de la secuencia	47
Comparación entre las secuencias regulatorias	48

	Determinación del TSS	50
	Búsqueda de cajas regulatorias en la zona núcleo	56
	Detección de TFBS	59
	Validación de los programas	59
	Detección de TFBS putativos	61
	Factores de Trascrición candidatos	71
	Receptor a Glucocorticoides (GR)	71
	C/EBP $\alpha$ , $\beta$ y $\gamma$	72
	p53	73
	Receptor a ácido retinóico (RXR $\alpha$ y $\beta$ )	74
	Proteína específica (Sp1)	75
	Receptor a Andrógenos (AR)	76
	NF $\kappa$ B	77
	STAT4	79
	E2F	79
	YY1	81
	MyoD/Miogenina	82
Discusión		83
	Obtención de la secuencia promotora del gen <i>sno</i>	83
	Determinación del TSS	85
	Búsqueda de cajas consenso en la zona núcleo del promotor del gen de <i>sno</i>	86
	Predicción de TFBS	87
	Propuestas para la validación <i>in vitro</i> de las predicciones <i>in silico</i>	105
	Predicciones con base en análisis <i>in silico</i>	108
Conclusión		111
Hipergrafía		112
Bibliografía		113

# Resumen

El TGF $\beta$  es una familia de citocinas que a través de sus efectores, las Smad, regula la expresión de una variedad de genes blanco, entre los que destaca un correpresor de la vía denominado *sno*.

Hasta la fecha no se sabe qué factores regulan la expresión de *sno*. En concreto, no existe una caracterización detallada del promotor de *sno*, aunque se sabe desde los trabajos de Stroschein y colaboradores (1999) que *sno* es activado transcripcionalmente por TGF $\beta$ . De hecho, Zhu *et al.* (2005) encontraron un elemento de respuesta al TGF $\beta$  en el promotor de *sno*.

El TGF $\beta$  es capaz de ejercer una gran gama de respuestas biológicas, fundamentalmente a través de la regulación de genes blanco precisos. La acción que el TGF $\beta$  desencadene en determinado tiempo y espacio depende del contexto celular. En concreto, esta versatilidad se debe, en gran medida, a la capacidad que poseen las Smad de asociarse con un gran número de factores dentro del núcleo celular.

En cuanto a su papel fisiológico, se ha observado que los niveles de mRNA del correpresor *sno* aumentan durante la regeneración hepática (Macías-Silva *et al.* 2002). Ski y SnoN inducen, además, la diferenciación terminal de músculo esquelético (Wrighton *et al.* 2007). Durante la carcinogénesis, Sno juega un papel importante, pero ambiguo. En diversos tipos de cáncer, a este correpresor se le adjudica una actividad de oncogen, mientras que en otros, parece comportarse como supresor de tumores (Luo, 2004).

Por lo anterior, resulta interesante analizar los Sitios de Unión a Factores de Transcripción (TFBS) en el elemento de respuesta al TGF $\beta$  en el promotor de *sno*.

Para tales efectos, es necesario determinar primero la región promotora; predecir el o los sitios de inicio de la transcripción; encontrar las regiones conservadas en *sno*, y, finalmente predecir TFBS.

Este análisis se llevó a cabo con herramientas bioinformáticas, principalmente con los programas *Blast*, *ClustalX*, *DBTSS* y *Alggen*.

Se encontró un alto grado de identidad entre las secuencias regulatorias (71%) y los elementos de respuesta al TGF $\beta$  (79%) de las especies de estudio, *Homo sapiens* y *Mus musculus*. En la zona núcleo del promotor se detectó un sitio de inicio de la transcripción (2280 pb río abajo de SBE4 en *Mus musculus* y 2035 en *Homo sapiens*) dentro de una isla rica en CpGs; y no se encontraron cajas regulatorias. En la vecindad de cada SBE ( $\pm 10$ pb) se predijo la presencia de sitios de unión para doce factores de transcripción capaces de interactuar física y funcionalmente con las Smad y que responden a una gran variedad de estímulos (GR, STAT4, AR, C/EBP, E2F, NF $\kappa$ B, p53, RAR-RXR, MyoD, Miogenina, Sp1 y YY1). Estos TFBS putativos se encuentran 100% conservadas entre *Homo sapiens* y *Mus musculus*, lo que aumenta la probabilidad de que, efectivamente, se trate de secuencias regulatorias. Varios de los TF para los que se detectaron secuencias de unión putativas en el TRE de *sno* de las especies estudiadas protagonizan procesos fisiológicos y/o patológicos en los que se sabe que el TGF $\beta$  juega un papel importante y los niveles de expresión de *sno* se encuentran alterados (sistema inmune e inflamación, cáncer y ciclo celular, y miogénesis).

En conclusión, en el presente trabajo se realizó una caracterización *in silico* del promotor de *sno* de *Homo sapiens* y *Mus musculus*, en particular de su zona núcleo y el TRE. Debido a que se trata de predicciones, el trabajo invita a realizar una verificación experimental de los resultados obtenidos.

# Lista de abreviaturas

NOMBRE (ABREVIADO)	SIGNIFICADO DE LAS SIGLAS (EN INGLÉS)	SIGNIFICADO EN ESPAÑOL (TRADUCCIÓN)	FUNCIÓN
<i>Alggen</i>	<i>Algorithm and genetics group</i>	Grupo de algoritmos y genética	Programa de búsqueda de TFBS
AR	<i>Androgen receptor</i>	Receptor a andrógenos	TF
ARE	<i>Androgen Response Element</i>	Elemento de respuesta a andrógenos	TFBS para AR
BLAST	<i>Basic Local Alignment and Search Tool</i>	Herramienta básica de alineamientos locales y búsqueda	Programa en PubMed que sirve para realizar alineamientos entre secuencias
BMP	<i>Bone Morphogenic Protein</i>	Proteína Morfogénica de Hueso	Miembro de la superfamilia del TGF $\beta$
BR-LZ	<i>Basic Region – Leucine Zipper</i>	Región Básica - Cierre de Leucina	Dominio de unión a DNA
bZIP	<i>Basic Zipper</i>	Cierre básico	Familia de factores de transcripción con un <i>Zipper</i> de leucina.
C-Smad	<i>Common Smad</i>	Smad común	Smad que se asocia a R-Smad fosforiladas
C/EBP	<i>CCAAT/Enhancer Binding Protein</i>	Proteína de unión al potenciador CCAAT	TF
CBP/p300	Core Binding Factors, polynoma enhancer binding protein, PEBP	Factores de unión nucleares, proteína de unión al <i>enhancer</i> de polinoma	Cofactor
CDS	<i>Coding secuencia</i>	Secuencia codificante	Región del mRNA que se transcribe a proteína
CEBP	<i>cAMP responsive element binding protein</i>	Proteína de unión al elemento	TF
C-ter	/	De respuesta a AMP cíclico	Región extrema de un polipéptido
DBTSS	<i>Data Base of Transcriptional Start Sites</i>	Carboxilo Terminal	Herramienta bioinformática de búsqueda de TSS
DHT	<i>5<math>\alpha</math>-dihydrotestosterone</i>	Base de datos de sitios de inicio de la transcripción	Ligando de AR
E2F-1,2,4	<i>Adenovirus E2 promoter Factor</i>	5 $\alpha$ -dihydrotestosterona	Reguladores de ciclo celular cuya transcripción es activada por el TGF $\beta$ ; Factores de transcripción bHLH capaces de asociarse con las Smad
GC	<i>Glucocorticoid</i>	Factor celular descubierto por su asociación al promotor 2 de adenovirus	Ligando de GR
GR	<i>Glucocorticoid receptor</i>	Glucocorticoide	TF
GRE	<i>Glucocorticoid Response Element</i>	Receptor a glucocorticoides	TFBS para GR
HDAC	<i>Histone desacetylases</i>	Elemento de Respuesta a Glucocorticoide	Inhibición de la transcripción
MH	<i>Mad Homology</i>	Desacetilasas de Histona	Dominio característico de las Smad
N-coR	<i>Nuclear co Repressor</i>	Homología a Mad	Correpressor Nuclear, es reclutado por Ski y SnoN
NF $\kappa$ B	<i>B- cell specific nuclear factor binding to the intronic <math>\kappa</math> light chain enhancer</i>	Correpressor nuclear	
PAI-I	<i>Plasminogen activator inhibitor I</i>	Factor Nuclear específico de células B de unión a la cadena intrónica ligera $\kappa$	Gen blanco del TGF $\beta$ prototipo
R-Smad	<i>Receptor activated Smad</i>	Inhibidor del activador del plasminógeno I	/
SBE	<i>Smad binding DNA element</i>	Smad activada por receptor	TFBS para las Smad3, 4.
SIE	<i>Smad Inhibitory Site</i>	Elemento de unión a las Smad	TFBS para Smad3. Unión de Smad3 a SIE inhibe la transcripción del gen blanco (Zhu <i>et al.</i> , 2004)
SIM	<i>Smad Interaction Motiv</i>	Sitio de inhibición de las Smad	/
		Dominio de interacción con las Smad	

Sno	<i>c-ski novel related</i>	nuevo asociado a c-ski	Correpressor y gen blanco del TGF $\beta$ , Skil (Ski-like)
Sp1	<i>Specific protein 1</i>	Proteína específica 1	TF
STAT4	<i>Signal Transducer and Activator of Transcription</i>	Transductor de Señales y Activador de la Transcripción	Factor de transcripción activado por IL12
SXS	<i>Serine-Aminoacid-Serine</i>	Serina-Aminoácido-Serina	Motivo rico en serinas, presente en las Smad
TRE	<i>TGF<math>\beta</math> Reponse Element</i>	Elemento de respuesta al TGF $\beta$	Secuencia génica que presenta SBEs y SIEs
TBP	<i>Tata box binding protein</i>	Proteína de unión a caja TATA	Factor de transcripción basal que recluta la a la RNA polimerasa al TSS
TF	<i>Transcription Factor</i>	Factor de Transcripción	/
TFBS	<i>Transcription Factor Binding Site</i>	Sitio de Unión a Factor de Transcripción	Parte de la zona regulatoria de un gen a la que se une determinado TF
TGF $\beta$	<i>Transforming Grow Factor <math>\beta</math></i>	Factor de Crecimiento transformante $\beta$	Citocina antiproliferativa
TSS	<i>Transcriptional start site</i>	<i>Sitio de inicio de la transcripción</i>	/
VDR	<i>Vitamin D Receptor</i>	Receptor a vitamina D	Receptor hormonal nuclear
YY1	<i>Yin yang 1</i>	/	Factor de transcripción con dedos de Zinc

# Introducción

## La transducción de señales

Las células están constantemente interactuando con su entorno. Deben ser capaces de censar cambios en el ambiente y responder a éstos. En los organismos multicelulares, existe además una fina coordinación e integración entre todas las células del organismo.

Gomperts, Kramer y Tatham mencionan en su libro *Signal Transduction* que la transducción de señales es la disciplina que se ocupa de responder a la pregunta de cómo estímulos o influencias externas –representados por primeros mensajeros, como las hormonas o los neurotransmisores, pero también nutrientes e incluso luz ultravioleta- determinan un proceso dentro de su célula blanco.

La membrana celular es impermeable a la mayoría de estas moléculas. Por ello, los primeros mensajeros deben ser capaces de ejercer su efecto en la célula desde la parte externa de ésta.

Receptores en la superficie celular unen a estas moléculas –los ligandos- y *transducen* la señal hacia el interior de la célula por medio de activaciones en cadena de diversas proteínas denominados mensajeros intracelulares o segundos mensajeros.

Las cadenas de activación correspondientes a cada ligando/receptor se denominan *vía de señalización*. Las vías de señalización culminan en cambios protéicos y/o genéticos que tienen como consecuencia modificaciones estructurales en la célula.

Desde el punto de vista evolutivo, se conoce la existencia de estructuras químicas muy relacionadas a hormonas tiroideas en algas y esponjas; estradiol en helechos y coníferas; catecolaminas en protozoarios; efedrinas en la gimnosperma *Ephedra* e incluso péptidos relacionadas a la insulina y la endorfina en protozoarios, hongos y bacterias.

Esto indica que la transducción de señales apareció muy temprano en la evolución. Sin embargo, resulta interesante que no siempre existe una correlación entre la existencia de ligandos y sus receptores.

Algunos autores consideran como el antecesor de la endocrinología y por tanto de la transducción de señales a Plinio el Viejo. En épocas más recientes - finales del Siglo XIX- destacan los trabajos de Brown-Séquard, Claude Bernard, Bayliss y Starling. No obstante, la transducción de señales se trata como una disciplina *per se* sólo a partir del trabajo de Gilman y Rodbell, quienes fueron galardonados en 1994 con el premio Nóbel por su descubrimiento de las proteínas G (Magner, 2002).

En la actualidad se clasifica a los distintos tipos de transducción de señales en familias según su tipo de ligandos, de receptores y/o de efectores. A continuación se enumeran las principales parejas ligando/receptor que desencadenan determinados tipos de señalización, según Heldin (2002).

- ♦ Integrinas / Moléculas de la matriz extracelular.
- ♦ Contactos célula-célula directos, por ejemplo, por medio de las cadherinas.
- ♦ Ligando soluble / Receptor con actividad intrínseca de cinasa de tirosina.
- ♦ Interleucina (u otra citocina) / receptor a citocinas.
- ♦ Antígeno / receptor para antígenos de células T o B.
- ♦ Factores de Necrosis Tumoral (TNF) / receptor a TNF.
- ♦ Factor de crecimiento / receptor acoplado a proteínas G.
- ♦ Miembros de la familia de Wnt / Frizzled.
- ♦ Hedgehog / Patched.
- ♦ Ligando (como AMP) / Canales iónicos ligando-dependientes.
- ♦ Miembro de la familia de los Factores de Crecimiento Transformante (TFG $\beta$ ) / Receptor con actividad intrínseca de cinasa de serinas y treoninas.

El presente trabajo estudia algunos detalles de este último tipo de vía de señalización; la del TGF $\beta$ .

## **EI TGF $\beta$**

El Factor de Crecimiento Transformante  $\beta$  (TGF $\beta$ ) pertenece a una superfamilia de polipéptidos que comprende a las Proteínas Morfogenéticas de Hueso (BMP's); a Nodal, a las Activinas, a la Hormona Antimülleriana (AMH) y a otros factores estructuralmente relacionados en invertebrados, como Daf en el nemátodo *Caenorhabditis elegans* y Dpp (*Decapentaplegic*) en *Drosophila melanogaster* (Massagué y Wotton, 2000).

El genoma humano codifica 28 genes que pertenecen a la superfamilia del TGF $\beta$ , y sus ligandos, receptores y efectores intracelulares, están conservados en los eucariontes desde *C. elegans* y *Drosophila* hasta los mamíferos (Moustakas y Souchelnytski, 2001).

### ***Papel biológico***

Este conjunto de proteínas es un importante orquestador del programa genético de un organismo (Massagué y Wotton, 2000). Es decir, las respuestas celulares que desencadena el TGF $\beta$  se deben en gran medida a la diversidad de las respuestas transcripcionales que esta citocina, como son la regulación de genes de matriz extracelular y de ciclo celular (Macías-Silva *et al.*, 2002).

De esta manera, el TGF $\beta$  afecta a un gran número de procesos biológicos tales como la respuesta inmune, el crecimiento celular, la diferenciación, el desarrollo y la apoptosis de diversos tejidos y tipos celulares (Liu *et al.*, 2001). Además, controla procesos de migración, adhesión y, en general, procesos que involucran la remodelación de la matriz extracelular. Por lo tanto, esta proteína

está involucrada de manera importante en el desarrollo, en la regeneración de diversos tejidos y en el cáncer.

La diversidad de funciones que un miembro de la familia del TGF $\beta$  es capaz de desencadenar depende del contexto celular. Por ejemplo, en una misma herida, el TGF $\beta$  puede tanto estimular como inhibir la proliferación celular, dependiendo de si la célula blanco es un fibroblasto o un queratinocito. En células epiteliales de mamífero, el TGF $\beta$  puede ocasionar arresto celular o un comportamiento metastático, dependiendo del nivel de la actividad oncogénica de Ras presente en la célula (Oft *et al.*, 1996). Así mismo, las BMPs humanas y su homólogo en *Drosophila*, Dpp, pueden señalar dorsalización en la mosca y formación de hueso en un vertebrado (Massagué y Wotton, 2000).

En suma, el TGF $\beta$  puede estimular o inhibir el crecimiento y la diferenciación de prácticamente todas las células, dependiendo del tipo celular, del estado de diferenciación, de la fisiología y de las condiciones de cultivo (Feng y Derynck, 2005).

### ***Papel del TGF $\beta$ en el cáncer***

El TGF $\beta$  juega un papel dual en el desarrollo de cáncer.

En las primeras fases de la tumorigénesis, la actividad del TGF $\beta$  inhibe el crecimiento celular (Luo, 2004). Así, el TGF $\beta$  actúa como una citocina antimitótica en diversos tipos celulares, y a varios componentes de su ruta se les adjudica el título de supresor de tumores (Miyazono *et al.*, 2003).

De acuerdo con lo anterior, en diversos tipos de cáncer se han detectado frecuentemente mutaciones que ocasionan pérdida de la función de varios genes directamente relacionados con el TGF $\beta$ . Entre éstas destacan las mutaciones en el receptor al TGF $\beta$  tipo II en el síndrome de Lynch (*Hereditary Nonpolyposis Colon Cáncer*, HNPCC), así como en Smad4 en el carcinoma pancreático (de hecho, las siglas de su sinónimo, DPC, significan *Deleted in Pancreatic Carcinoma*) y en cáncer de colon avanzado (Miyazono, 2003).

Sin embargo, una vez que las células han perdido la capacidad de responder a las señales antimitóticas del TGF $\beta$  -por ejemplo, a través de la activación oncogénica de *ras*, que le confiera independencia del ligando-, esta molécula induce la transición epitelio-mesénquima, propiciando así un aumento en la tasa de metástasis e invasión de tejidos. Este evento puede suceder de manera dependiente o independiente de las Smad (Luo, 2004).

Al inducir la producción de matriz extracelular, el TGF $\beta$  estimula la angiogénesis. Además, esta citocina es inmunosupresora. Así, en fases tardías del cáncer, el TGF $\beta$  actúa como una citocina oncogénica (Miyazono *et al.*, 2003).

### ***Vía de señalización del TGF $\beta$***

A continuación se trazarán los rasgos generales de la vía del TGF $\beta$ , considerando como referencias principales las reseñas realizadas por Rahimi y Leof (2007) y por Moustakas (2002). Véase la figura 1 al final de la sección.

Ligandos diméricos de la superfamilia del TGF $\beta$  se unen con alta afinidad a los Receptores al TGF $\beta$  (T $\beta$ RI y T $\beta$ RII) con actividad de cinasas de serina y treonina en la superficie de la célula. Esta unión induce la formación de complejos de receptor heterotetraméricos.

En estos complejos, el T $\beta$ RII constitutivamente activo es capaz de transfosforilar el dominio rico en Glicina y Serina (GS) del T $\beta$ RI.

Los acarreadores anclados a la membrana, SARA (Smad Anchor Receptor Activator) y HRS/HGS (Hepatocyte growth factor-regulated tyrosine kinase substrate), que unen fosfolípidos a través de su dominio de dedos FYVE (Fab1p, YOTB, Vsp27p, EEA1; las primeras proteínas purificadas con este dominio), contribuyen a la presentación de los efectores principales del TGF $\beta$ , las Smad (palabra que se compone de la fusión de los nombres Sma de *C. elegans* y MAD de *D. melanogaster*), con el dominio con actividad de cinasa del receptor T $\beta$ RI activado.

Los T $\beta$ RI fosforilan a las de Smads activadas por el receptor (R-Smad), tales como Smad2 y 3 cuando la activación se debe al TGF $\beta$ , y Smad1, 5 y 8 en el caso de la señalización de BMP's. La activación de las R-Smads ocurre por medio de una región denominada horquilla L45 del receptor T $\beta$ RI. El dominio de las R-Smads que entra en contacto con la horquilla L45 se conoce como Dominio de Homología a Mad (MH2); MH2 se localiza en la zona carboxilo terminal (C-ter) y es capaz de unir fosfoserinas. Eventualmente, T $\beta$ RI fosforila a las R-Smad en su motivo rico en Serinas SXSS. Esta fosforilación provoca un cambio conformacional en la R-Smad, lo que le permite disociarse rápidamente del receptor y de SARA.

Las fosfoserinas C-ter de las R-Smad son ahora reconocidas por el MH2 de otra Smad, lo que permite la formación de complejos heterotriméricos, formados por dos R-Smad y una Smad Común, Co-Smad, esta última representada en vertebrados exclusivamente, por la Smad4.

Posteriormente, la Smad3 se asocia con la importina B1, lo que le permite a todo el complejo heterotrimérico translocarse al núcleo.

Una vez en el núcleo, el complejo R-Smad/Co-Smad regula la expresión genética. Smad3 y 4 se unen a un Sitio de Unión a Factor de Transcripción (TFBS) en la región regulatoria de un gen blanco del TGF $\beta$  denominado SBE (*Smad Binding Element*) a través de sus dominios MH1, en concreto por una horquilla  $\beta$  que interactúa con el Surco Mayor del DNA. En contraste, la Smad2 se une difícilmente a este TFBS, ya que carece de la horquilla  $\beta$ .

Tanto las R-Smads como las Co-Smad interactúan con Factores de Transcripción (TF) tejido-específicos vía su dominio MH1 o MH2. De hecho, los SBEs muchas veces están adyacentes a sitios de unión a otros TF.

El dominio MH2 recluta además Coactivadores (CoA) como p300 que conllevan a la acetilación de histonas nucleosomales y/o TF asociados, eventos que son cruciales para la inducción transcripcional.

Alternativamente, el dominio MH2 de las Smad nucleares reclutan Correprosores (CoR) que se asocian con Desacetilasas de Histona (HDAC), reprimiendo la transcripción del gen blanco.

Este modo de acción de las Smad, caracterizado por una versatilidad de asociaciones con otros factores, explica muchos efectos fisiológicos del TGF $\beta$ .

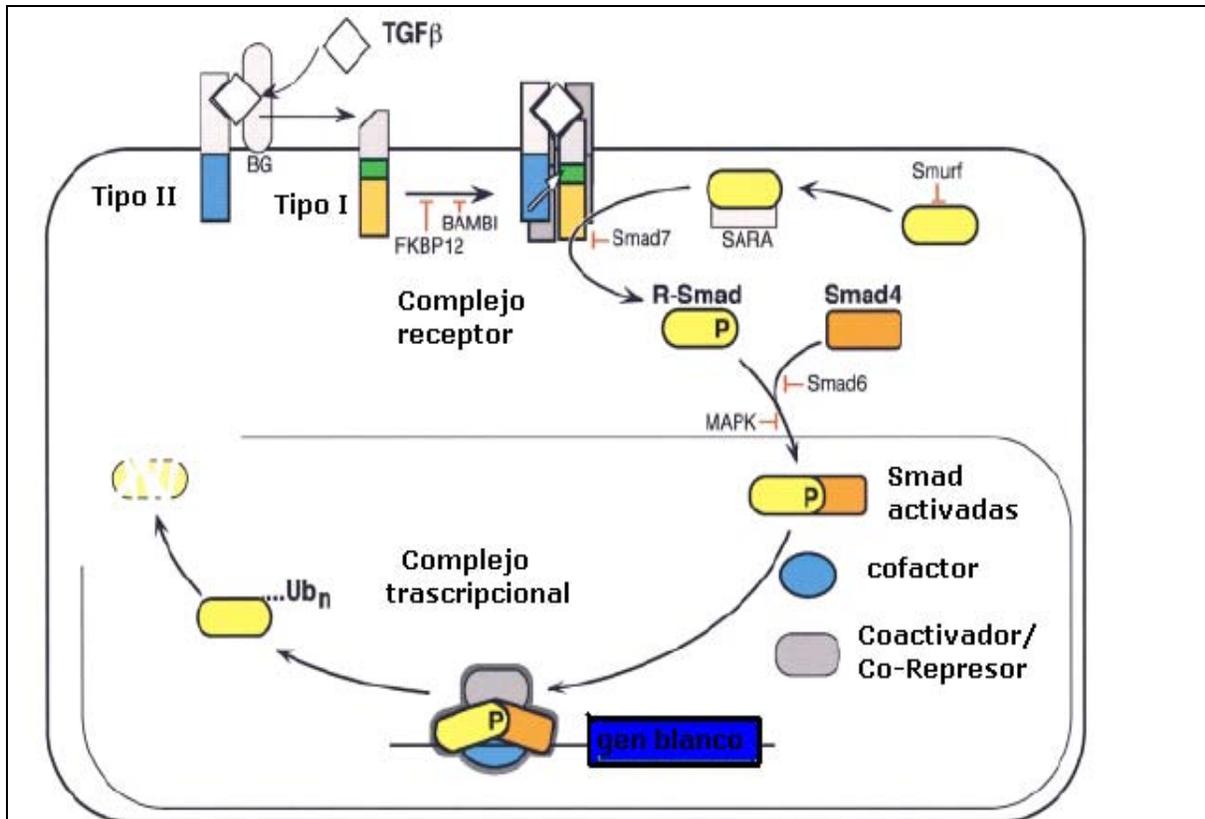
La regulación de esta vía es un proceso muy complejo que involucra un gran número de moléculas activadas por la misma vía del TGF $\beta$  (inhibición homóloga, asa de retroalimentación negativa) o por alguna otra vía (inhibición heteróloga o paróloga, comunicación cruzada).

Algunos de los procesos de regulación de la vía del TGF $\beta$  son:

- Ubiquitinación y degradación proteosomal de R-Smads en estado basal por las ligasas E3 Roc1/SCF y Smurf (*Smad ubiquitin regulatory factors*), entre otras.
- Smads Inhibitorias (I-Smad) Smad7 y 6. Smad7 se une con su dominio MH2 al receptor, lo que bloquea la unión de R-Smad al receptor (inhibición competitiva). Cuando Smad 7 une, además del receptor, a Smurf, induce la ubiquitinación y subsiguiente degradación lisosomal y proteólisis del receptor. La proteína adaptadora STRAP-1 aumenta la afinidad de Smad7 por T $\beta$ RI.
- El pseudoreceptor BAMBI juega también un papel de inhibición competitiva del receptor (Akiyoshi, 2001).
- FKBP12 se une al motivo GS en el T $\beta$ RI, lo que impide su fosforilación (Massagué y Wotton, 2000).
- La vía de las Cinasas de Proteínas Activadas por Mitógenos (MAPK) por medio de su efector Erk (*Extracellular Signal regulated kinase*) inhibe la vía del TGF $\beta$  al fosforilar a las R-Smad en la región linker, impidiendo su acumulación nuclear (Massagué y Wotton, 2000).
- Internalización de los receptores T $\beta$ RI y T $\beta$ RII.

La aparente simplicidad de la vía del TGF $\beta$  adquiere un carácter altamente complejo por la capacidad que poseen sus efectores de asociarse con una enorme variedad de proteínas.

El presente trabajo centra su atención en la dinámica de regulación de la transcripción por las Smads de un gen blanco del TGF $\beta$  particular, del cual se hablará más adelante con detalle.



**Figura 1.** Esquema de la vía de señalización del TGF $\beta$ . El ligando TGF $\beta$  se une al T $\beta$ RII. Posteriormente, se induce la fosforilación y activación del T $\beta$ RI. T $\beta$ RI fosforila a las Smad dependientes de Receptor (R-Smad). Dos R-Smad se asocian con una Smad común (Smad4 en vertebrados). Este complejo heterotrimérico se transloca al núcleo. El complejo heterotrimérico de Smad frecuentemente se asocia a otros TF y/o cofactores, y en la región reguladora de genes blanco del TGF $\beta$  se unen a un SBE, regulando la transcripción de dicho gen. Este proceso está finamente regulado por las I-Smad, la degradación proteosómica, etcétera.

Modificado de Massagué y Wotton, 2000

## Regulación de la transcripción mediada por TGF $\beta$

La regulación de la transcripción genética -el proceso de síntesis de RNA a partir de DNA- es el primer medio de control de la expresión génica. Este proceso culmina con la producción de diferentes proteínas en diferentes tipos celulares o en respuesta a diferentes estímulos (Lachman, 2001).

Como se mencionó, el TGF $\beta$  es un importante modulador de la transcripción de un gran número de genes blanco. A continuación se profundizará en la parte de la vía del TGF $\beta$  que se lleva a cabo en el núcleo celular.

### ***Los Factores de Transcripción de la vía del TGF $\beta$ : Las Smad***

Los principales efectores del TGF $\beta$ , las Smad, son codificados por genes altamente conservados. Se han encontrado en el humano, en la rata, en el ratón, en *Xenopus*, en el pez cebra, en el platelminto *Schistosoma mansoni*, en *Drosophila* y en *C. elegans* (Moustakas, A., Souchelnytski, 2001).

Estas proteínas se dividen en R-Smad, Co-Smad e I-Smad (Figura 2):

- Las R-Smad que responden a BMP están representadas en *Drosophila* por Mad (*Mothers against dpp*); y en vertebrados, por Smad1, 5 y 8. Las R-Smad reguladas por el TGF $\beta$  o por activina son dSmad2, Smad2 y Smad3, respectivamente.
- Medea, Smad4 y Smad4 $\beta$  son las Co-Smad en *Drosophila*, vertebrados y en *Xenopus*, respectivamente.
- Las I-Smad son Dad en *Drosophila* y Smad6 y 7 en vertebrados.

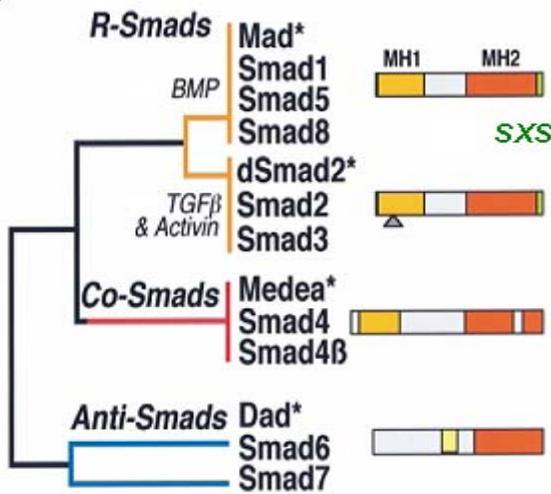
Esta familia de proteínas contiene en su estructura tres regiones altamente conservadas. En la región amino terminal se encuentra el dominio MH1, que es responsable de la unión al TFBS en la región regulatoria del DNA. En específico, esta interacción DNA-Smad se lleva a cabo por medio de una horquilla  $\beta$  (figura 3) que está ausente en Smad2 (triángulo gris, figura 2), lo que explica la incapacidad de Smad2 de unirse al DNA aún presentando el MH1. En las I-Smads, este

dominio se encuentra poco conservado, por lo que difícilmente se asocian a un SBE.

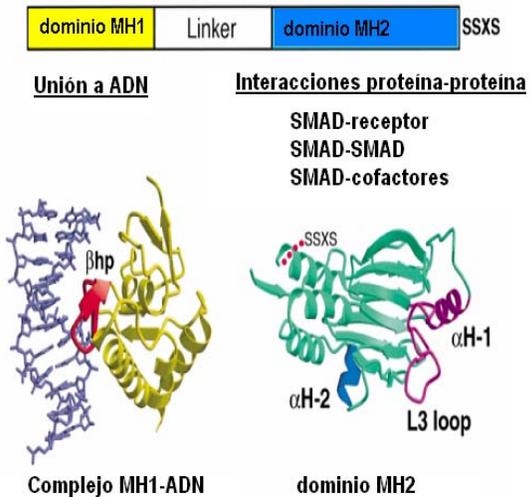
En medio se encuentra la región *linker*, cuya función principal es regulatoria al presentar numerosos sitios de fosforilación. Este dominio está pobremente conservado en comparación a MH1 y MH2.

Como se muestra en la figura 3, el dominio MH2 en la región carboxilo terminal se encarga de las interacciones proteína-proteína. En el caso de las R-Smad, este dominio presenta el motivo SXS que es fosforilado por los receptores tipo I. Además, MH2 contiene dos hélices  $\alpha$ , así como una asa L3 que interactúa también con el receptor; de hecho se propone que le confiere selectividad (Feng y Derynck, 2005).

El bajo grado de conservación del dominio de unión a MH1, así como la ausencia del motivo SXS en el dominio MH2, explican la actividad de inhibidores competitivos de los receptores tipo I por las I-Smad.



**Figura 2.** La familia de las R-Smad. Se muestra el dominio MH1 (N-ter) y el dominio MH2 (C-ter). La región intermedia corresponde al *linker*. El asterisco\* denota las Smads características de *Drosophila*, Smad4 $\beta$  está presente en *Xenopus*. Los demás ejemplares son de vertebrados.



**Figura 3.** Las interacciones de los dominios MH de las Smad. MH1 se une a un TFBS en el DNA por medio de una hélice  $\beta$ . MH2 es responsable de las uniones protéicas. Destaca el motivo de fosforilación SXS, dos hélices  $\alpha$  y un giro L3.

Modificado de Massagué y Wotton, 2002.

## **Unión de las Smad al DNA**

De lo anterior, es importante destacar que las Smad son factores de transcripción activados por miembros de la superfamilia del TGF $\beta$ .

Shi *et al.* (1998) describieron la estructura cristalográfica de la asociación entre el dominio MH1 de las Smad y el DNA. Detectaron que Smad3 y 4 se unen al SBE con la secuencia 5'GTCT 3' por medio de una horquilla  $\beta$  que interactúa con el surco mayor del DNA.

Johnson y colaboradores descubrieron en 1999 que la asociación Smad-SBE induce un cambio en la tasa de transcripción. Reportaron además la presencia de uno o más SBEs en genes de respuesta al TGF $\beta$ .

Así, la capacidad de responder a señales del TGF $\beta$  de los genes tempranos requiere de la presencia de uno o más SBEs (Massagué y Wotton, 2000).

## **Genes blanco del TGF $\beta$**

La amplia gama de funciones biológicas reguladas por el TGF $\beta$  está directamente asociada con la activación o inhibición de genes blanco precisos.

Los primeros genes que se asociaron con una regulación por el TGF $\beta$  fueron el gen del Inhibidor del Activador de Plasminógeno (*PAI1*), de la colagenasa 7 (*COL7A1*), *Mix2* (protagoniza la especificación de mesodermo en *Xenopus*), *gooseoid* (encargado de la inducción del mesodermo anterior en ratón), *JunB*, *Nodal*, *lefty2*, *IgA*, *c-Fos*, *Vent2*, *Tinman* y *Ultrabithorax*, los dos últimos en *Drosophila* (Massagué y Wotton, 2000, Akiyoshi *et al.*, 2001).

Akiyoshi y colaboradores realizaron en 2001 un microarreglo de células HaCaT tratadas con el TGF $\beta$  (400pM), mediante el que detectaron un gran número de genes regulados positiva y negativamente por esta citocina. A continuación se presenta una tabla en la que se enlistan algunos de los genes regulados por el TGF $\beta$  que se consideran más relevantes (Tabla 1).

**Tabla 1.** Genes blanco del TGF $\beta$ , encontrados por Akiyoshi *et al.* (2001) por medio de un microarreglo de células derivadas de queratinocitos humanos (HaCat) estimuladas con 400pM del TGF $\beta$

	<i>Regulados positivamente</i>	<i>Regulados negativamente</i>
<i>TF</i>	Fra-1, Ets-2, HEF1	NF-IL6-beta, SREBP-2, C2H2-150, HBZ17, DP-2, IRI1B, NF90, RAR1 y ESE-1b
<i>TF de interacción con las Smad</i>	c-Jun, JunB, Fra1, 2F-2, Dlx-2, E2F-1,2,4*	ESE, CtBP, ATF,GCN1, 2F-4, Tob, Fra, p53*
<i>Inhibidores del TGF<math>\beta</math></i>	Nma, Smad7 y LABP.	IL-1 $\beta$ , receptor a IL-1 $\beta$
<i>Matriz extracelular</i>	Proteína A similar a la hormona paratiroide, PAI-1 y proteína similar a plasminógeno	
<i>Proteína unida a membrana</i>	Involucrina	
<i>Transportador ATP-asa transportadora de aniones</i>		SLC6A10, GLVR1, RING4
<i>Cinasas</i>	Cinasa parecida a CDC, cinasa de queratina B	ATPasa humana estimulada por arsenito
<i>Cinasas de serinas y treoninas</i>	Pim-1	Cinasa relacionada a Cdc-7, cinasas de residuos de tirosina
<i>Fosfatasa de tirosina</i>	SHP-2	Calcineurina A1, fosfatasa de tirosina
<i>Proteína de andamiaje</i>	HEF1	
<i>Inhibidor de proteína</i>	TFPI-2, cistatína M	
<i>Inhibidor de CDK</i>	p21	
<i>Inhibidor de proteasa</i>		ALP
<i>Enzima biosintética</i>	Homólogo humano de la isomerasa IPP de levadura y FLAP	Fosfoetanolamina citidiltransferasa, CAD, GlcNAc-T1, PIG-C, Subunidad beta1 de la K-ATPasa, fosfatasa de la polifosfatasa de inosítidos
<i>Proteína mitocondrial</i>	hidroxilasa 1,25- dihidroxiviramina D3 24- mitocondrial	MIPEP
<i>Receptor a citocinas</i>		TNFR, IL-2 R
<i>Receptor a hormonas</i>		receptor membranal a esteroides, proteína putativa de unión a progesterona
<i>CPGR</i>	RDC1	Receptor 3 a B2-bradiquinina

<i>Glicoproteína</i>	BTF2	BTF1
<i>Ligando</i>	Jagged1	Hsema-I
<i>Moléculas de adhesión</i>		Ninjurina 1
<i>Citocromo</i>	CyP450	citocromo P450 inducido por dioxina
<i>Pirofosfatasa de nucleótidos</i>		Diadenosina tetrafosfatasa
<i>Enzima procesadora de nucleótidos</i>		RPP38
<i>Proteína cromosomal</i>		RbAp48, NHC
<i>Peptidil propil cis-trans isomerasa</i>		FKBP-12
<i>Antígeno de linfocitos</i>		HLA-DNA
<i>Gatekeeper</i>		Cdc46
<i>Oxidoreductasa</i>		PDIR
<i>Fosfoproteína asociada Src</i>		SKAP55
<i>Enzima metabólica</i>		cinasa de fosfomelvanonato CDC2,6,7,16,25A,25B; PITSLARE;
<i>Reguladores de ciclo celular</i>	CDC25C,42; Ciclina D1,I; p15/Ink4b; p16/Ink4a; p107,130; Rb y E2F-1,2,4*	CiclinaA,B1,C,D2,D3,F,G1,H; CDK2,4,6; p19/Ink4d; p27/Kip1; subunidad regulatoria de la cinasa cdk5, p35; p57/kip2; p53*; E2F-3,5; DP- 1,2; PCNA; c-Myc; PTEN; Wee

Los genes enlistados en la Tabla 1 codifican para una gran diversidad de proteínas, tales como receptores, proteínas cromosomales, transportadores, etcétera. Destaca la presencia de factores de transcripción, de componentes de matriz extracelular y de reguladores de ciclo celular.

Como se detallará más adelante, estas proteínas están íntimamente relacionadas con el papel que desempeña el TGFβ en los diversos tejidos.

En la Tabla 1 se distinguen dos tipos de genes activados por el TGFβ:

- Genes tempranos: Aquéllos activados directamente por los factores de transcripción de la vía del TGFβ, las Smad.

- Genes tardíos: Se trata de los genes que son activados por los productos de los genes tempranos regulados por el TGF $\beta$ ; por ejemplo, por un factor de transcripción o por algún integrante de otra vía de señalización. En este caso, el TGF $\beta$  actúa por medio de una *red genética*.

Además de los genes enlistados, recientemente se ha encontrado que *Sno*, un CoR transcripcional de las Smads, es activado transcripcionalmente por el TGF $\beta$ . La regulación transcripcional de *sno* es, por lo tanto, un ejemplo de un asa de retroalimentación negativa.

La regulación genética analizada en este trabajo corresponde precisamente al gen *sno*, por lo que se abundará al respecto más adelante.

### ***Factores de Transcripción asociados a las Smad***

Como mencionan Massagué y Wotton (2000), se ha calculado que la secuencia del SBE está presente, en promedio, una vez cada 1024 pares de bases en el genoma, es decir, una vez en cada región regulatoria de cada gen de tamaño promedio. Así, si la unión de un SBE fuera suficiente para la activación transcripcional dependiente de las Smads, una proteína Smad activada llevaría a la activación no selectiva de un gran número de genes. Sin embargo, esto no ocurre. La presencia exclusiva de un SBE no induce la activación de genes por las siguientes razones:

1. La afinidad del dominio MH1 de las R-Smad y de las Co-Smad por el SBE es baja ( $K_{disociación} = 10^{-7}M$ ), por lo que se requieren contactos adicionales con el DNA (Shi *et al.*, 1998).
2. La unión de las Smad a los SBE es poco selectiva ya que Smad1, 3 y 4 tienen una afinidad semejante por el SBE. Esto se debe a que la secuencia de horquilla  $\beta$  es idéntica entre todas las R-Smad y está altamente conservada en la Smad4.

Por lo tanto, una unión estable con el DNA aunada a una alta especificidad en la secuencia de unión se logra por la interacción con otros factores de unión a DNA específicos de una secuencia determinada (Akiyoshi *et al.*, 2001).

Massagué y Wotton (2000) proponen incluso que los factores de transcripción determinan la secuencia de DNA a la que el complejo Smad se une, así como los cofactores (correpresores y coactivadores) reclutados, los otros factores de transcripción que cooperan y el tiempo de duración de la regulación del gen.

Mientras que el dominio MH1 de las Smad se une al SBE, el dominio MH2, responsable de las interacciones proteína-proteína, puede unirse a otros Factores de Transcripción y/o a Cofactores. La asociación de las Smad con otros factores forma complejos de una composición y geometría determinados. Los sitios de unión a DNA de las Smad y los factores asociados en el mismo complejo son capaces de actuar sinérgicamente si sus secuencias de unión están presentes a la distancia exacta entre la una y la otra en un promotor blanco (Massagué y Wotton, 2000).

Así, una gran variedad de actividades biológicas del TGF $\beta$  resulta de la capacidad de las Smads de interactuar con varios factores de transcripción. Se han reportado más de 30 factores de transcripción, como FAST1, TFE3, Mixer, c-Jun, Sp1 y el receptor a vitamina D3, que interactúan con las Smad. Por lo tanto, al asociarse con un factor de transcripción particular, las Smad pueden exhibir efectos biológicos específicos en la célula en cuestión (Miyazono *et al.*, 2003).

Feng y Derynck realizaron en 2005 una extensa revisión de la complejidad de las interacciones del TGF $\beta$  en la que incluyeron una lista de los factores de transcripción y cofactores potencialmente capaces de asociarse a las Smad (Tabla 2).

**Tabla 2.** Cofactores y Factores de Transcripción capaces de asociarse a las Smad. *Reseñados por Feng y Derynck (2005)*

Coactivadores	Correpresores	Factores de transcripción			
ARC105	c-Myc	E2F4/5	FoxH1/FAST	VDR	HIF-1 $\alpha$
CBP/p300	c-Ski, SnoN	Max	FoxO	CBFA1/Runx2/AML	IRF-7
GCN5	Evi-1 (ZF)	MyoD	Dlx1	GATA3	Lef1/TCF
MSG1	SNIP1 (FHA)	TFE3	Hoxc-8	GATA4,5,6	MEF2 (caja MADS)
PCAF	TGIF (HD)	ATF2	Milk-Mixer	Gli $\Delta$ C-ter	Menin
SKIP	Tob	ATF3	AR	OAZ	NF $\kappa$ B p52
SMIF	YB-1	c-Fos	ER	Sp1	NICD
Swift	ZEB22/SIP1	c-Jun, JunB, JunD	GR	YY1	p53
ZEB1		CEBP $\alpha,\beta,\delta$	HNF4 RXR	ZNF198 $\beta$ -catenina	Pax8 SRF

Moustakas *et al.* (2001) proponen que las Smad son capaces de unirse a factores en el núcleo también a través de su dominio MH1. Mencionan en este sentido al coactivador pXHBC capaz de unirse a Smad4; a los represores HDAC y Hoxc-8, capaz de unirse a Smad3 y Smad1, respectivamente y que el factor de transcripción Sp3 se puede unir a Smad2/4.

Estos investigadores agregan además a la lista de Feng y Derynck, los factores de transcripción BF-1, E1A y Evi-1 capaces de unirse al dominio MH2 de las Smad1-4, Smad1-3 y Smad3, respectivamente.

La realización de esta interacción, depende de dos circunstancias:

- Contexto celular, es decir, la disponibilidad de estas proteínas en el núcleo celular en un determinado momento y espacio.
- Estructura de la región regulatoria del gen blanco del TGF $\beta$ , es decir, la disposición de los sitios de unión a factores de transcripción

Esta investigación se enfoca en el análisis de la estructura de la región regulatoria del gen blanco del TGF $\beta$ , SnoN.

### ***El promotor de un gen***

El promotor es la región regulatoria de un gen que controla la transcripción en respuesta a estímulos ambientales (Zhang, 2007). En términos más concretos, es la región genética que está involucrada de manera más íntima en el proceso de inicio de la transcripción. En esta región se encuentra una colección de secuencias regulatorias, cada una de una longitud de 6 a 10 pares de bases, a las que se unen proteínas de unión a DNA que regulan la transcripción (Smale, 2005). Estas proteínas son los factores y cofactores de transcripción.

Los **factores de transcripción** son proteínas que, en respuesta a un estímulo fisiológico ya sea intra o extracelular, se unen a su **Sitio de Unión para Factor de Transcripción (TFBS)** en el DNA, lo que promueve o inhibe la transcripción del gen blanco.

Además de los dominios de unión a DNA, los factores de transcripción pueden contener también regiones conocidas como dominio de activación, que son capaces de activar la transcripción. Estas regiones interactúan con componentes del complejo basal de la transcripción, ya sea de manera directa o indirecta. Cuando la acción se ejerce de manera indirecta, los factores de transcripción se unen a cofactores que funcionan como “puentes” entre el factor de transcripción y el aparato basal de la transcripción.

Los **cofactores** son moléculas que interactúan con el DNA indirectamente, es decir, a partir de la unión a factores de transcripción. Estas moléculas establecen un puente entre el factor de transcripción y la maquinaria basal de la transcripción.

Los **coactivadores** (como p300, CBP y P/CAF para el TGF $\beta$ ), contienen dominios de Acetilación de Histonas (*Histone Acetil Transferase*, HAT) que inducen la transcripción de genes blanco al permitir una estructura más abierta de la cromatina (Miyazono *et al.*, 2003). Dicho de otra manera, estas moléculas

neutralizan por acetilación la carga positiva de las histonas y las estructuras de la cromatina sueltas, lo que se traduce en una regulación positiva de la transcripción (Akiyoshi *et al.*, 2001).

Los **correpresores** reclutan desacetilasas de histona que fortalecen (“compactan”) la estructura de la cromatina, lo que reprime la transcripción del gen blanco (Akiyoshi *et al.*, 2001).

Como apunta Smale, 2005, el promotor puede dividirse en dos zonas regulatorias: la zona núcleo y la zona distal.

La zona núcleo se encarga de la tasa de transcripción basal de un gen. Se encuentra aproximadamente entre 80 y 100 pares de bases alrededor del Sitio de Inicio de la Transcripción (TSS) en los promotores de genes que codifican para proteínas (Zhang, 2007) y es el sitio al que se une la maquinaria basal de la transcripción, es decir, la RNA polimerasa y los Factores de Transcripción Basales (IIA, B, D, E, F y H). Estos factores son responsables de reclutar la RNA polimerasa al promotor, establecer el TSS preciso y facilitar el proceso de iniciación de la transcripción.

La zona núcleo presenta sitios de unión a estos factores de transcripción. Destaca la caja TATA, que se encuentra de -25 a -30 pb río arriba del TSS (Zhang, 2007).

La caja TATA es reconocida por la Proteína de Unión a la caja TATA (*TATA box binding protein*, TBP) que se asocia con un determinado número de factores asociados a TBP (*TBP associated Factors*, TAF) que en conjunto forman un complejo denominado TFIID.

El TFIID recluta a un gran número de otros factores de transcripción, la RNA polimerasa II y otras proteínas –que en conjunto forman la holoenzima transcripcional – al TSS, lo que permite la iniciación de la transcripción.

Aproximadamente el 60% de las zonas núcleo de los promotores de *Homo sapiens* y *Mus musculus* contienen una región rica en islas CpG. Los genes asociados a este tipo de zonas núcleo comprenden a todos aquellos genes que se expresan en la totalidad de los tejidos (genes *housekeeping*) y

aproximadamente la mitad de los genes tejido-específicos (Antequera, 2003). Los promotores que presentan islas CpG en su zona núcleo generalmente carecen de cajas regulatorias, como la caja TATA (Juven-Gershon *et al.*, 2008).

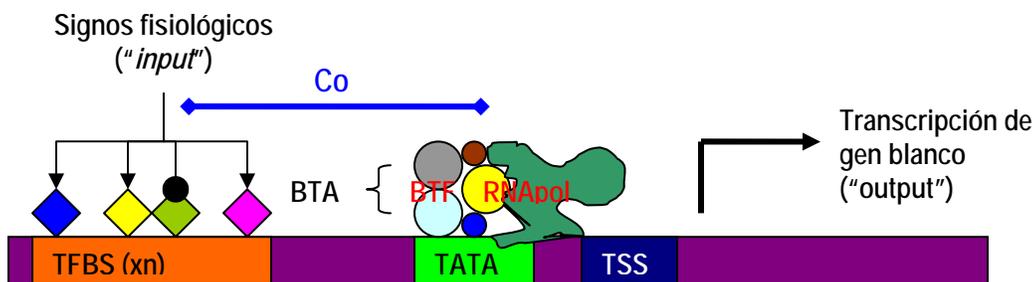
La zona regulatoria distal une a factores de transcripción que no son esenciales para el ensamblaje de la maquinaria transcripcional básica, pero que afectan de manera importante la tasa de transcripción basal.

El nivel de transcripción dictado por los Factores Basales de la Transcripción es muy bajo. Sin embargo, este nivel se aumenta considerablemente por la unión de TF que se activan únicamente en determinados tipos celulares o en respuesta a un estímulo particular. La activación de un factor de transcripción se da a nivel de su síntesis y/o de su actividad proteica (Lachmann, 2001).

Los coactivadores y las proteínas que modifican la estructura de la cromatina actúan como puentes entre la región regulatoria distal y la zona núcleo del promotor.

Las regiones regulatorias en el promotor se encuentran frecuentemente divididas en sub-regiones conservadas y/o regiones en donde elementos regulatorios forman *módulos* (Zhang, 2007).

Los componentes típicos de un promotor se esquematizan en la figura 4.



**Figura 4.** Regulación de la transcripción en respuesta a estímulos fisiológicos.

## **Contexto celular**

La heterogeneidad de respuestas biológicas que los miembros de la familia del TGF $\beta$  son capaces de desencadenar depende de manera importante del contexto celular, es decir, de las proteínas y otros factores que se encuentren en determinado momento del desarrollo y tipo celular y que sean capaces de unirse al promotor del gen blanco del TGF $\beta$  (Davidson y Coffman, 2007).

En el caso del TGF $\beta$ , el contexto celular se puede definir, en términos moleculares, como la mezcla de “factores acompañantes” en una célula determinada en el momento de estimulación con el TGF $\beta$ . Esta mezcla determina el resultado o tipo de respuesta genética (Massagué y Wotton, 2000).

Para fines de este trabajo, se resaltarán el papel de factores de transcripción que se encuentren en el núcleo celular en un tiempo y tejido específicos.

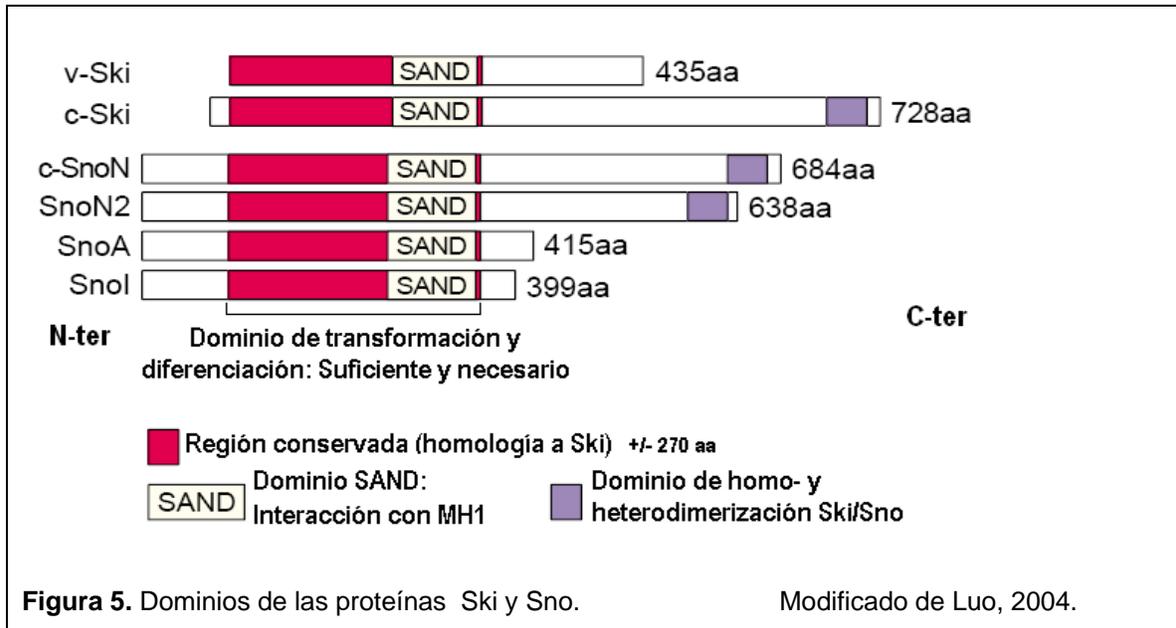
## **Sno**

Ski (Sloan-Kettering Institute protooncogene) y Sno (Ski-related novel gene) pertenecen a la familia de CoR de las Smad (Ruoyun *et al.*, 2007). Estas dos proteínas se caracterizan por la presencia de un dominio altamente conservado de diferenciación y transformación suficiente y necesario en el que se encuentra un motivo de interacción con las Smad denominado SAND (Sp100, AIRE-1, NucP41/75, DEAF-1), así como un dominio de homó y heteromerización hacia su extremo carboxilo terminal (Figura 5).

En el genoma humano, *sno* está representado por un solo gen, pero por medio de procesos de *splicing* alternativo se generan cuatro transcritos distintos (Takesu *et al.*, 2006).

En el caso de Ski, se conoce el producto de un oncogen retroviral (v-Ski) y su contraparte celular, identificada originalmente en pollo (c-Ski).

En cuanto a *sno*, se han identificado en *Homo sapiens* cuatro isoformas para las que este gen codifica: SnoN (*Ski-related novel gene Non alu containing*), SnoN2 (carece de 138 nucleótidos en el tercer exón), SnoI (*Insertion*) y SnoA (*Alu containing*) (Figura 5).



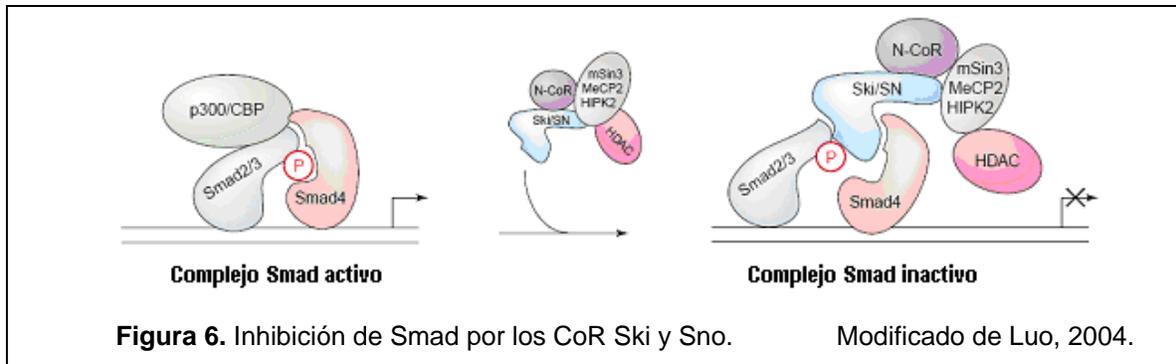
### ***Mecanismos de acción de Sno y Ski***

Las proteínas Ski y Sno *no* tienen una actividad catalítica específica, y carecen de dominios de unión a DNA. Su efecto represor parece deberse a la interacción con moléculas, como las Smad (Luo, 2004).

Las formas de interacción inhibitoria Ski/SnoN-Smad son las siguientes (Figura 6):

1. Dominio de unión a Ski en Smad4, MH1, es el mismo que se requiere para la unión a las R-Smad Fosforiladas: obstrucción del complejo heterotromérico Co-Smad (Smad4)- R-Smad
2. Interacción con correpresores (Nuclear Hormone Co-Repressor, N-CoR; mSin3A, HIPK2 y MeCP2) y reclutamiento del complejo de desacetilación de las histonas (HDAC).

3. Ski y SnoN impiden la unión del coactivador p300/CPB a las Smad.



Cuando ligandos de la familia del TGF $\beta$  están presentes, SnoN es rápidamente ubiquitinado y degradado, lo que permite a las Smad la activación de la expresión de genes blanco, incluyendo la transcripción del propio SnoN (Takesu *et al.*, 2004).

### ***Papel biológico de Ski y Sno***

Macías-Silva y colaboradores detectaron en 2002 una sobreexpresión de SnoN y Ski durante la regeneración hepática. En este trabajo se propone que el aumento en los niveles de estos correpresores durante la fase proliferativa fomenta la resistencia a la citocina antimitótica TGF $\beta$ , permitiendo la restauración de los niveles de masa hepática.

También se ha establecido un importante papel de Ski y SnoN Durante la miogénesis. De hecho, en células de codorniz se ha observado que la sobreexpresión de Ski y SnoN induce la diferenciación terminal de músculo esquelético (Wrighton *et al.* 2007).

Además, la sobreexpresión de estas proteínas es común en muchas líneas tumorales humanas derivadas de melanoma, cáncer de mama, de esófago, de pulmón, de epidermis y de próstata, que podría conferirles un papel de oncogen (Luo, 2004).

Sin embargo, se ha sugerido que Ski y SnoN pueden actuar también como supresores de tumores en algunas líneas celulares.

Por lo tanto, una regulación precisa de Ski y SnoN es esencial para el mantenimiento de la homeostasis tisular.

La función de SnoN durante el desarrollo es ambigua, como lo muestran dos estudios con ratones *knockout* para *sno*. Shinagwa y colaboradores descubrieron en el 2000 una incapacidad de implantación de ejemplares murinos *sno*<sup>-/-</sup>; es decir, el homocigoto es letal. Contrariamente, otro estudio sostiene que el homocigoto no es letal, pero tiene defectos en la activación de células T (Takesu *et al.* 2006).

El papel dual de estos correpresores está probablemente relacionado con el papel aparentemente contradictorio del TGFβ (protooncogen y supresor de tumores).

### ***Regulación bifásica de Sno por TGFβ***

La activación de la vía del TGFβ induce la degradación proteosomal de los correpresores Ski y Sno. Por otro lado, *sno*, pero no *ski*, está regulado transcripcionalmente por el TGFβ (Stroschein *et al.* 1999). Así, La regulación de Sno por esta citocina constituye un caso paradigmático de retroalimentación negativa. Por ello, este trabajo se centra en la regulación de la expresión de *sno* por TGFβ.

Poco tiempo después de un estímulo del TGFβ, Smad2 y 3 reclutan las ligasas de ubiquitina APC (*Anaphase Promoting Complex*) y Smurf2, entre otras (Moustakas, 2002). Estas moléculas median la degradación proteosomal de SnoN dependiente de ubiquitina al propiciar la ubiquitinación de este correpresor en sus tres residuos de lisina (Lys440, Lys446 y Lys449) y su subsiguiente degradación.

Desde el punto de vista de la transcripción, sucede el fenómeno contrario. Después de aproximadamente dos horas de estimulación con el TGFβ, aumenta

considerablemente la cantidad de mRNA y de proteína de Sno, al activarse la transcripción del gen *sno*. Este efecto positivo del TGF $\beta$  sobre la expresión de *sno* constituye un mecanismo de retroalimentación negativa, lo que permite una coordinación y regulación adecuada de la duración en intensidad de la señalización por TGF $\beta$  (Luo, 2004).

El mecanismo de retroalimentación negativa radica en que el reestablecimiento de los niveles basales del correpresor Sno inducido por la regulación transcripcional de *sno* por TGF $\beta$  inhibe el control de la expresión de genes blanco de TGF $\beta$ .

# Antecedentes

## ***Regulación de sno por TGF $\beta$***

### **TGF $\beta$ induce la transcripción del gen *sno***

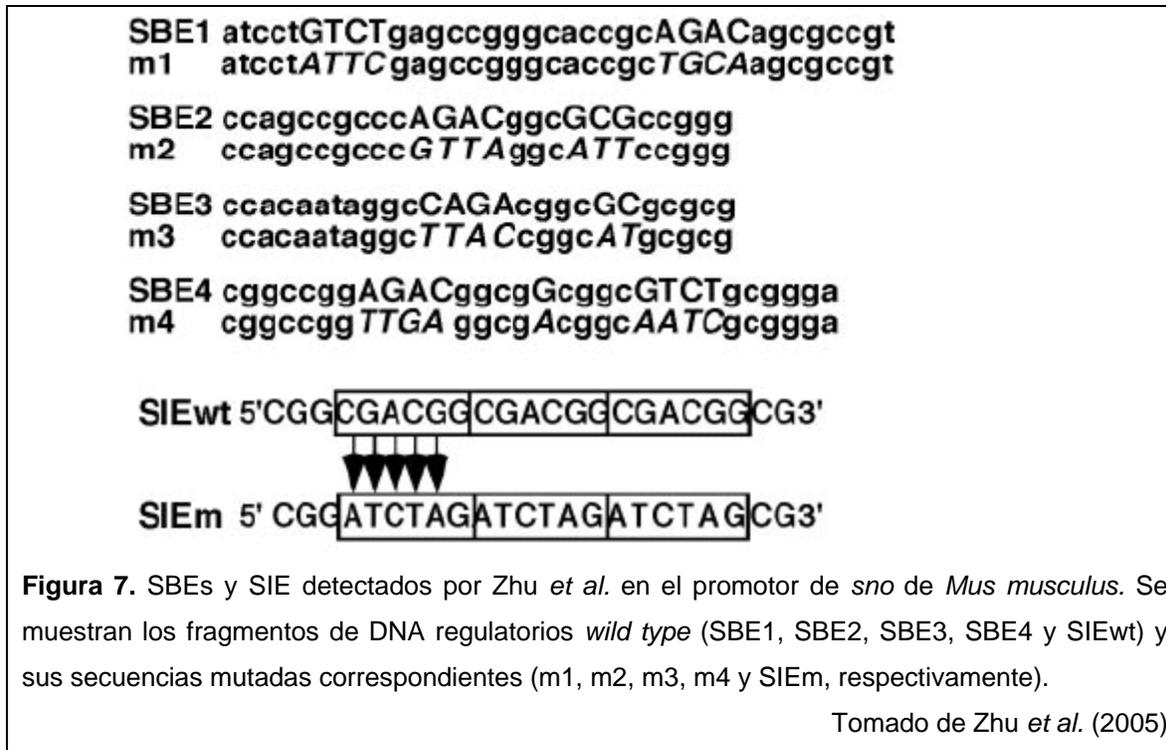
Stroschein *et al.* (1999) reportaron por primera vez que a las dos horas de estimulación, el TGF $\beta$  induce un aumento marcado en los niveles de mRNA de SnoN. El aumento en la expresión de *sno* inhibe a las Smad, interfiriendo sobre la regulación de genes blanco del TGF $\beta$ . Por lo tanto, la inducción en la expresión del gen *sno* por el TGF $\beta$  constituye un mecanismo de retroalimentación negativa.

### **Zhu *et al.* encuentran SBEs y un SIE en el promotor de *sno* de *Mus musculus***

Zhu *et al.* detectaron en 2005 que TGF $\beta$  induce la expresión de *sno* a través de la unión del complejo Smad2/4 al elemento de respuesta al TGF $\beta$  (TRE) en el promotor de *sno* murino. Por otro lado, el complejo Smad3/4 inhibe la transcripción de *sno* al unirse a un sitio de inhibición de las Smad (SIE) en este mismo promotor.

Por medio de ensayos de retardo (Electrophoretic Mobility Shift Assay, EMSA) determinaron cuatro SBEs en el TRE y en el SIE. Ensayos de luciferasa usando genes reporteros con el TRE y el SIE alterados por mutagénesis dirigida mostraron en cada SBE y el SIE las bases indispensables para la regulación de *sno* por el TGF $\beta$ . Además, determinaron que todos los SBEs tiene el mismo efecto relativo sobre la regulación transcripcional, y que sólo la deshabilitación simultánea de los cuatro SBE abate la activación de la transcripción de *sno* mediada por el TGF $\beta$ .

En el artículo, únicamente se muestran las secuencias correspondientes a los SBEs y al SIE (Figura 7), sin reportar la secuencia promotora de *sno*.



### Ruoyun y colaboradores detectan un elemento de respuesta a HGF en el promotor del gen *sno* de *Homo sapiens*

Rouyun y colaboradores encontraron en 2007 que el Factor de Crecimiento de Hepatocitos (HGF) inhibe la acción del TGF $\beta$  a través de la inducción de la expresión de *sno* en las células proximales tubulares epiteliales HKC-8. La inducción de la transcripción de *sno* por HGF se lleva a cabo a través de la ruta de las MAPK que involucra a Erk1,2. Erk 1,2 activa a CREB (*cAMP responsive element binding protein*), un factor de transcripción que se une a un Elemento de Respuesta a AMPc (CRE) y únicamente en cooperación con el TF Sp1 -que se une a los sitios de unión a Sp1 localizados cerca de CRE- induce la expresión de *sno*.

Estos datos moleculares sugieren una explicación del papel fisiológico del HGF con relación al TGF $\beta$ . La citocina HGF es un potente antifibrótico -es decir, inhibe la inducción de componentes de matriz extracelular por el TGF $\beta$ . Además, bloquea la transición epitelio–mesénquima, proceso que es también favorecido por TGF $\beta$ . Estos efectos fisiológicos se pueden deber, al menos en parte, al antagonismo que ejerce HGF sobre TGF $\beta$  al inducir la expresión del correpresor de los efectores del TGF $\beta$ , SnoN.

Aunque es importante destacar el papel antagónico que ejerce HGF sobre TGF $\beta$  a nivel fisiológico, para el presente trabajo resulta de suma importancia exclusivamente que Rouyun *et al.* (2007) hayan reportado sitios de unión a factores de transcripción determinados en el promotor de *sno* humano, ya que es el único dato experimental que se tiene sobre la secuencia del promotor de *sno* de *Homo sapiens*.

Ruoyun *et al.* presentan la siguiente secuencia como parte del promotor de *sno* humano (los números al inicio y al final de la misma son arbitrarios, ya que no se conoce con precisión el TSS).

En la figura 8 se muestran los TFBS de respuesta a HGF en el promotor de *sno* de *Homo sapiens* detectados por Ruoyun *et al.* (2007).

1838----

ggtcggggccagagggggtcgctagggcgggtgttccacggctgatcacgcttcggtggccccagggcgggtggggc  
 cggcagggcagggcgctgtagtgacgggagccgcggggaccccggagaccgggacggccgtcaccgcccctcc  
 ccagtcagcagcgctccgcggggatgccctggtctcgaccgggtccg – 1639

Sitios de unión a Sp1	CRE
-----------------------	-----

**Figura 8.** Elemento de respuesta a HGF en el promotor de SnoN humano. Ruoyun *et al.* 2007

# Planteamiento del problema

El TGF $\beta$  y Sno juegan un papel importante en la homeostasis de un organismo. Esto se evidencia por su papel dual y aparentemente contradictorio en el desarrollo de diversos tipos de cáncer, padecimiento en el que pueden jugar un papel tanto de supresores de tumores como de oncogenes.

El TGF $\beta$  y Sno se regulan mutuamente; es decir, Sno es un correpresor transcripcional de la vía del TGF $\beta$  mientras que el TGF $\beta$  ejerce sobre Sno una regulación bifásica (a tiempos cortos, un estímulo de TGF $\beta$  disminuye los niveles de la proteína Sno; aproximadamente a las dos horas, induce la expresión del gen *sno*).

Se sabe que el TGF $\beta$  induce la expresión de *sno*; de hecho, se conocen los SBEs en el promotor de este gen. Sin embargo, no existe hasta la fecha una caracterización detallada del promotor de *sno*. Se desconoce la locación genética correspondiente a esta región regulatoria, así como la existencia de otros reguladores de la expresión de este gen.

Además, aunque es conocido que TGF $\beta$  induce la expresión de *sno*, se sabe también que los efectores del TGF $\beta$ , las Smad, se necesitan asociar con otros factores de transcripción y/o cofactores para unirse a las SBE y regular así, conjuntamente, la expresión del gen blanco del TGF $\beta$ .

Por lo anterior, resulta interesante averiguar qué otros factores son capaces de asociarse al promotor de *sno*, en concreto a su región regulatoria.

Así, en el presente trabajo se busca, con herramientas bioinformáticas, delimitar y ubicar la zona de respuesta a TGF $\beta$  en el promotor de *sno* de *Homo sapiens* y *Mus musculus* y predecir en ésta Sitios de Unión a Factores de Transcripción.

# Hipótesis

Existen en el elemento de respuesta al TGF $\beta$  del promotor de *sno*, secuencias de unión a factores de transcripción capaces de interactuar con las Smad. Estas secuencias están conservadas entre *Homo sapiens* y *Mus musculus*.

# Objetivos

## Objetivo general

Caracterización *in silico* del promotor de *sno* en *Homo sapiens* y en *Mus musculus*.

## Objetivos específicos

- Delimitar la secuencia del promotor de *sno* humano y murino
- Establecer el sitio de inicio de la transcripción.
- Determinar el grado de conservación entre las secuencias regulatorias del promotor de *sno* en *Homo sapiens* y *Mus musculus*.
- Predecir los sitios de unión a factores de transcripción en el promotor de *sno*, en concreto en la región regulada por TGF $\beta$ .
- Seleccionar sitios de unión para factores de transcripción capaces de interactuar con las Smad.
- Analizar, por medio de una revisión bibliográfica, mecanismos de acción y la situación fisiológica en la que están presentes estos factores de transcripción.
- Proponer un modelo de regulación de la expresión de *sno*.
- Diseñar una estrategia experimental para comprobar el modelo.

# Métodología

El presente trabajo tiene como objetivo un análisis teórico de la secuencia promotora del gen de *sno*.

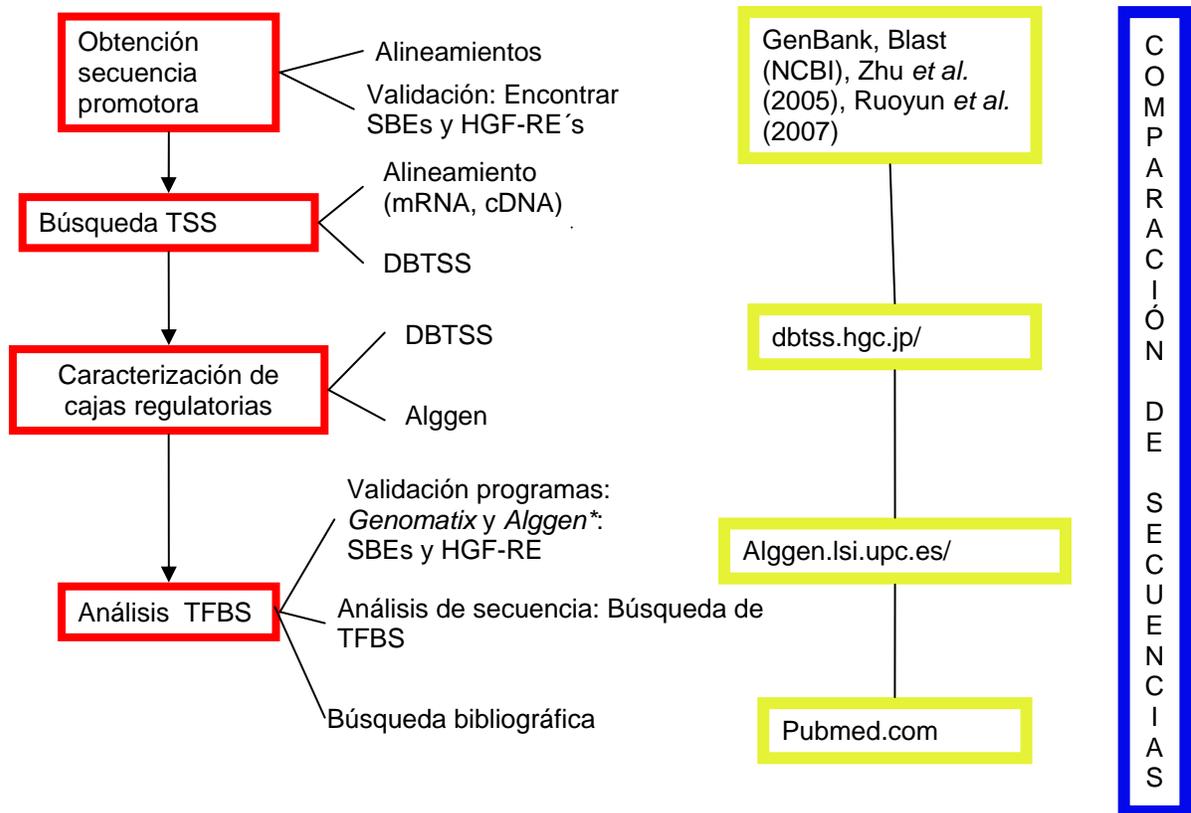
La herramienta que se utilizó a lo largo de esta investigación fue la bioinformática. En particular, se utilizaron los programas *Blast* y *ClustalX* para los alineamientos; *DBTSS* para establecer el TSS; *Alggen* para la determinación de los TFBS putativos; y, finalmente, PubMed para el análisis bibliográfico de los resultados obtenidos (Figura 9).

Debido a que la secuencia del promotor de *sno* no se había reportado anteriormente, fue necesario obtener, por medio de alineamientos, dicha secuencia.

Posteriormente, se determinó el Sitios de Inicio de la Transcripción (TSS) y se buscaron las cajas regulatorias que frecuentemente se encuentran asociadas a éste. Se utilizó el programa de predicción de TSS *DBTSS* y se realizaron alineamientos.

Finalmente, se realizó un análisis de los Sitios de Unión a Factores de Transcripción (TFBS) putativos en la secuencia del promotor de *sno* que contiene al elemento de respuesta al TGF $\beta$ , es decir, a los SBEs y al SIE. Este análisis se efectuó con el programa *Alggen*, y se robustecieron los resultados por medio de comparaciones entre las secuencias.

A lo largo de todo el proceso se compararon las regiones conservadas; el o los TSS; y la presencia y ubicación de los TFBS putativos entre *Homo sapiens* y *Mus musculus*.



**Figura 9:** Esquema de la metodología seguida en este trabajo. Se muestra a la izquierda el proceso; al centro, el modo de realización y la dirección en Internet de los programas que se utilizaron; y el panel de la derecha se refiere a las comparaciones entre las secuencias realizadas a lo largo de todo el trabajo.

### Obtención de la secuencia promotora

La técnica empleada fue el alineamiento.

Para obtener la secuencia promotora de *sno* en *Mus musculus* se alineó la secuencia de mRNA de SnoN2 con número de acceso en el GenBank U14655 con la región del cromosoma 3 del ratón AC117590. La elección de la secuencia AC117590 se debe a reportes anteriores (Luo *et al.*, 2004) que localizan al gen *sno* en el cromosoma 3, en particular en la región 3q26, correspondiente a esta secuencia.

ncbi.nlm.nih.gov → All Databases → BLAST (*B*asic *L*ocal *A*lignment and *S*earch *T*ool) → Align two sequences (bl2seq)

Se seleccionaron 5000 bases en la secuencia del cromosoma 3 río arriba del alineamiento con el mRNA de SnoN2.

El mRNA se alinea con la secuencia complementaria del cromosoma 3 a partir de la base 84,121. Por lo tanto, la región promotora del gen debe encontrarse 5 000 bases río arriba, es decir, en el intervalo

**(5´) 79,121 - 84,121 (3´)**

de la secuencia complementaria del cromosoma 3 de *Mus musculus*.

La elección de la secuencia de mRNA de SnoN2 fue arbitraria, ya que el primer exón, que contiene al TSS, está presente en todas las isoformas de mRNA codificadas por el gen (Lewin, 2008).

Para constatar que la secuencia de 5000 bases seleccionada efectivamente corresponde a la región promotora de *sno*, se buscaron en ésta los elementos de respuesta al TGFβ -los cuatro SBEs y el SIE- reportados por Zhu y colaboradores (2005) en el promotor de *sno* de ratón, esperando encontrarlos con una similitud del 100%.

Adicionalmente se buscó el elemento de respuesta a HGF –los dos CRE y los dos elementos de unión a Sp1- reportado por Rouyun *et al.* (2007) en el promotor de *sno* humano.

En el caso de *Homo sapiens*, se alineó la secuencia de respuesta a HGF del promotor de *sno* reportada por Ruoyun y colaboradores en 2007 con todo el genoma humano.

BLAST → BLAST human genome

El programa BLAST detecta que la secuencia de respuesta a HGF se encuentra 54,811 pb río abajo del gen de la PKC $\phi$  y a 2,195 pb río arriba del gen de Sno (=SkiL).

Esta región del genoma humano corresponde al intervalo

**(5´) 76,570,876 – 76,571,075 (3´)**

de la secuencia NT\_005612 del cromosoma 3 humano.

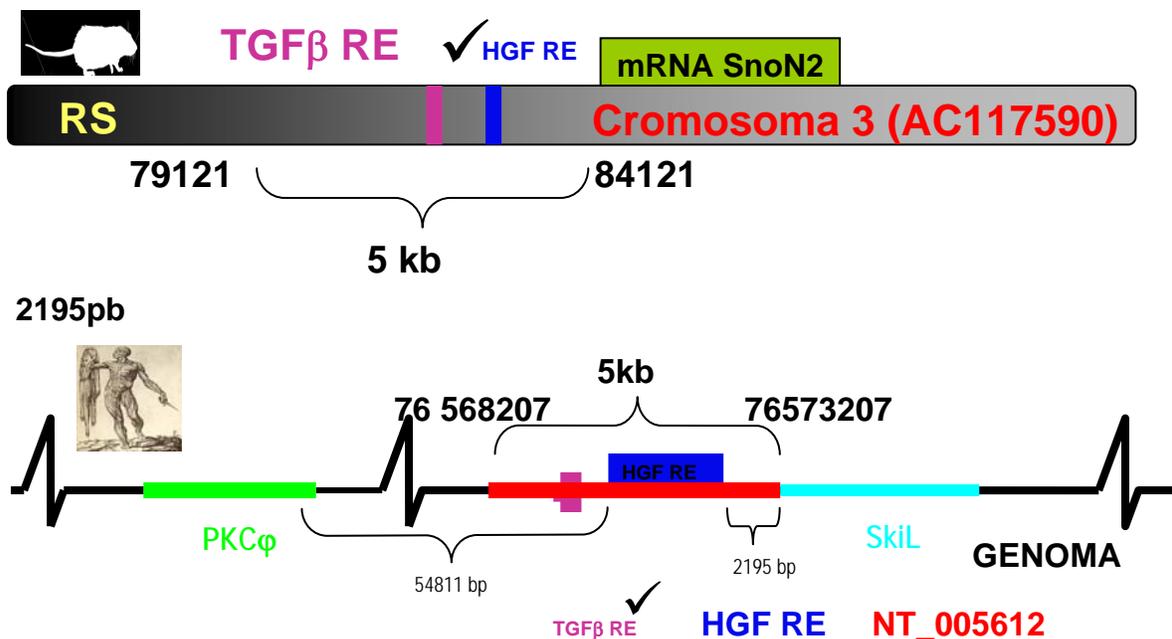
Por la distancia del alineamiento al gen *sno* (2,195pb), se calcula que los 5000 pares de bases correspondientes al promotor de *sno* se encuentran de

**76,571,075 + 2,195 = 76,573,270 río abajo (cota en 3´) y  
76,573,270 - 5,000 = 76,568,270 río arriba (cota en 5´).**

Así, la región del Cromosoma 3 de *Homo sapiens* que corresponde a la región promotora de *sno* está conformada por las bases:

**(5´) 76,568,270 a 76,573,270 (3´)**

En la figura 10 se esquematiza el proceso de determinación de la región promotora de *sno* en *Mus musculus* y *Homo sapiens*.





## Determinación de los Sitios de Inicio de la Transcripción

El Sitio de Inicio de la Transcripción (TSS) es la región del promotor en la que comienza la transcripción, es decir, la síntesis de RNA a partir de DNA. La localización del TSS es un paso básico en el análisis bioinformático de la regulación genética (Zhang, 2007). Para el presente trabajo, determinar la posición del TSS en el promotor de *sno* es fundamental para relativizar las posiciones de los TFBS que se predigan. Para determinar el TSS en el promotor de *sno* de *Homo sapiens* y de *Mus musculus* se utilizaron tres técnicas.

1. Alineamientos entre una isoforma de mRNAs de Sno y el promotor putativo. Se eligió la isoforma de mRNA de Sno de manera arbitraria, únicamente asegurando la presencia de la Secuencia Codificante Completa (CDS) y de la Región No Traducible 5' (5'UTR).

Se alineó el mRNA con la secuencia del promotor putativo de *sno* con el programa Blast2Sequences; las secuencias se obtuvieron de la base de datos de GenBank en formato FASTA.

Para *Mus musculus*, se seleccionó la secuencia de mRNA de snoN2 U14655. Como se detalla en su descripción (GenBank), esta secuencia de 2,018pb, está dividida en la Región No Traducible 5' (UTR) de las bases 1 a 64; la CDS, que se encuentra entre las bases 65 y 1,951 y que presenta entre las bases 1,341 y 1,342 un sitio de inserción de 138 nucleótidos para la isoforma SnoN (ausentes en SnoN2 por la presencia de un donador críptico); y finalmente la región 3'UTR, se encuentra entre las bases 1,852 y 2,018.

Debido a que en la selección de la zona promotora se consideraron exactamente 5kb río arriba del TSS putativo, para realizar el alineamiento fue necesario aumentar 100pb a estos 5kb.

En el caso de *Homo sapiens* se seleccionó el mRNA U70730 de SnoN2.

Este mRNA mide 2702pb; el sitio 5'UTR corresponde a las bases 1 a 709; la CDS, a las bases 710 a 2626. Entre las bases 2000 y 2001 se encuentra un sitio de inserción de 138 nucleótidos adicionales que codifican para 46 aminoácidos. Cuando esto sucede, la proteína codificada es SnoN. El sitio 3'UTR corresponde a las bases 2627 a 2702.

2. Determinación del TSS con el programa DBTSS (*Data base Transcriptional Start Sites*).

Este programa identifica TSS basándose en una amplia base de datos de cDNA. La utilidad de este método para detectar zonas núcleo de promotores ha sido resaltada por Zhang en 2007.

A continuación se presenta la ruta de acceso al buscador de TSS en DBTSS.

<http://dbtss.hgc.jp/> → Database search → Species: Homo sapiens (*Mus musculus*)Solexa:-  
- → Category: Product → Keyword: SKI-like\*(SKI-like isoform a\*)

Para estandarizar los resultados obtenidos, se despliega la secuencia promotora que se alinea con SBE4 y se comparan las distancias  $TSS_{DBTSS} - SBE4$  y  $TSS_{BLAST\ mRNA} - SBE4$ .

3. Alienamientos “manuales” entre el promotor putativo y las diferentes isoformas de cDNA que se encuentran en GenBank. Esta técnica ha sido utilizada previamente por Takaesu *et al.* en 2006.

Se alinearon los mRNA correspondientes a las diferentes isoformas de sno en *Homo sapiens* y en *Mus musculus* con los promotores putativos de estas dos especies (Tabla 4) con el programa Blast2sequences. Como secuencias de cDNA se consideraron las diferentes isoformas de mRNA, ya que las isoformas de cDNA de *sno* no se encuentran representadas en su totalidad en la base de datos consultada (PubMed) y las secuencias de cDNA y mRNA son equivalentes.

Los mRNA se obtuvieron de la base de datos de PubMed como sigue:

www.ncbi.nlm.nih.gov→Nucleotide→sno mRNA→ Core nucleotide records

**Tabla 4.** Isoformas de mRNA de Sno desplegadas en la página de PubMed.

Isoforma	Mus musculus	Homo sapiens
SnoN	<a href="#">U36203</a>	<a href="#">X15219</a>
SnoN2	<a href="#">U14655</a>	<a href="#">U70730</a>
SnoI		<a href="#">Z19588</a>
SnoA		<a href="#">X15217</a>

Para alinear el mRNA de SnoI, fue necesario aumentar 100 pb al promotor putativo de *Homo sapiens*, porque la secuencia correspondiente a esta isoforma (Z19588) no está completa.

### Búsqueda de cajas regulatorias

Las cajas regulatorias son sitios de unión a factores basales de la transcripción. Cuando están presentes en un promotor, se localizan rodeando el TSS, es decir, en la zona núcleo del promotor.

La búsqueda de TSS con el programa DBTSS muestra que *sno* en *Homo sapiens* y en *Mus musculus* presenta una zona rica en CpG cerca de la zona núcleo del promotor; es decir, en aquella zona en la que supuestamente se encuentran las cajas regulatorias.

La zona rica en CpG está asociada a los TSS detectados por DBTSS que *no* corresponden al TSS que el presente trabajo pondera.

Como comenta Smale (2005), las zonas núcleo ricas en CpGs frecuentemente están asociadas a una ausencia de las cajas regulatorias TATA, Inr y DPE.

Para determinar si efectivamente la abundancia de CpGs en la zona núcleo de los promotores de *sno* apunta hacia una ausencia de cajas regulatorias, se

analizó con el programa *Alggen* la zona núcleo del promotor con el objetivo de encontrar TFBS para factores basales de la transcripción.

Como zona núcleo se consideraron 100 pb río arriba y 100 pb río abajo del TSS, tal como lo plantea Zhang en 2007. Se enfatizó la búsqueda de la caja TATA en la región 25 – 30 pb río arriba del TSS.

El programa que se utilizó para predecir los TFBS fue *Alggen* (*Algorithmics Genetics Group*).

Como TSS se consideró aquel que se obtuvo a partir del alineamiento con el mRNA.

Así, las secuencias que se consideraron como zona núcleo son:

### ***Homo sapiens***

**4191 –**

```
GGGAGGCGGAGGTTGCAGTATGCCGAGATCGCTGCACTTCAGCC
TGGGCGACAGAGCGAGACTCCGTAAAAAAAAGAAAAAAAAAAAAA
TAAAAGAAACTCGTTTCAAATTGGCCCTTTGGCCTCTGGAGCAAA
TTCAAATGTAACCTTCCCAACCCCTTCTCTTCTTCCAGATTA
ATTAAGAAGAATGAATAT
```

**- 4392**

**TSS: 4292**

### ***Mus musculus***

**4900-**

```
AACAGAGTTGTACAATCTATTGAGTAAGAAAATTCATCTTTGGA
AAGTAAACATTTCTAAGTGAGTCAATAAGTATTTGTATACATTCC
CCTTTTCTCTTCAGATTAGGAAGTGAACAACATGCCAGTTG
CAGACTTAATCGTTTAAAGCACCAAGTGTGGCCATGGAAAACCT
CCAGTCTAAGTTCTCCTTAGTTCAGG
```

**- 5100**

**TSS: 5000**

El análisis con el programa *Alggen* se realizó como se detallará más adelante.

## Predicción de Sitos de Unión a Factores de Transcripción

### Validación del programa

La detección de Sitios de Unión a Factores de Transcripción (TFBS) putativos en las secuencias regulatorias de los promotores de *sno* de *Homo sapiens* y *Mus musculus* constituye la parte medular de este trabajo.

Por lo tanto, primero se seleccionó el programa, exigiéndole que encontrara los TFBS detectados experimentalmente por Zhu *et al.* (2005) y por Rouyun *et al.* (2007).

En este sentido, se probaron los programas Genomatix y *Alggen* como sigue:

[www.genmatix.de](http://www.genmatix.de) → Login → Search Transcription Factor Binding Sites/MatInspector → Secuencia formato Fasta → Putative Transcription Factor Binding Sites

<http://Alggen.lsi.upc.es/> → Research → Promo 3.0 → TRANSFAC Versión 8.3 → Step1: Select Species → Factors/Sites of: *Animals* → Step 2: Search Sites → Maximum Matrix Disimilarity Rate: 25 → Copiar y pegar secuencia a analizar.

El programa que cumplió con los requerimientos del trabajo al detectar los TFBS de los elementos de respuesta al TGF $\beta$  y a HGF fue *Alggen*. Esto se debe a que entre los programas examinados, es el único al que se le puede ajustar el valor de disimilitud máximo. Este valor indica el porcentaje máximo por el que la subsecuencia a analizar puede diferir en relación con la secuencia consenso (establecida en base de datos de TRANSFAC), de tal manera que *Alggen* considere a la subsecuencia a analizar como un TFBS putativo.

Por lo tanto, el análisis de las secuencias se realizó con el programa *Alggen*.





# Resultados

## Obtención de la secuencia promotora de *sno*

La región promotora de *sno* se encuentra en el intervalo (5´) 79,121 - 84,121 (3´) de la secuencia complementaria del cromosoma 3 AC117590 de *Mus musculus*.

La región del cromosoma 3 de *Homo sapiens* (NT\_005612) que corresponde a la región promotora de *sno* está comprendida entre las bases (5´) 76,568,270 y 76,573,270 (3´).

## Validación de la secuencia

Los únicos reportes de fragmentos del promotor de *sno* en *Homo sapiens* y *Mus musculus* son el elemento de respuesta a HGF y al TGF $\beta$ , respectivamente. Por lo tanto, para corroborar que las secuencias que se obtuvieron efectivamente corresponden a estos promotores, se buscó en éstas los SBEs, el SIE, los CRE y los elementos de unión a Sp1.

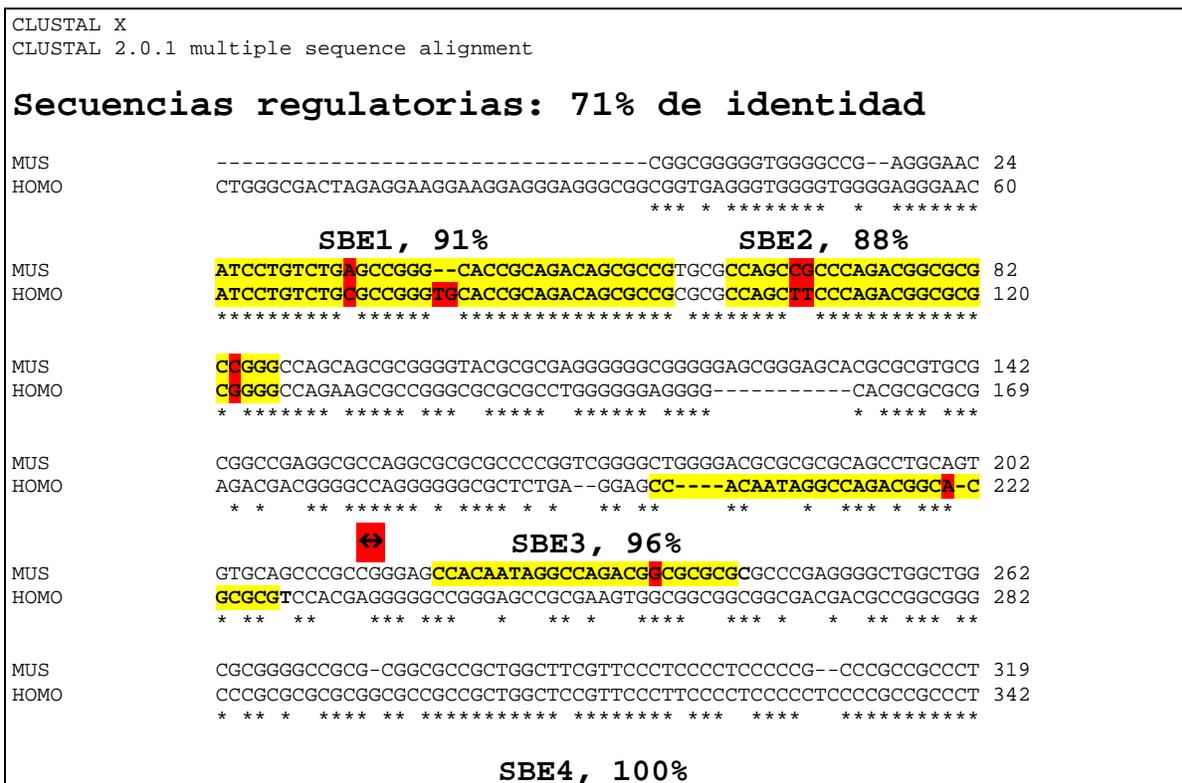
En la figura 11 se presenta el alineamiento entre las secuencias regulatorias de *sno* de *Homo sapiens* y *Mus musculus*. Se resaltan los SBEs reportadas por Zhu y colaboradores en 2005; y los CRE y Sitios de Unión a Sp1, reportados por Ruoyun *et al.* (2007). Como se esperaba, se encontró una concordancia del 100% entre los SBEs y el SIE reportados por Zhu *et al.* (2005) y la región promotora de *sno* en *Mus musculus*, ya que los SBEs y el SIE se reportaron originalmente en esta especie. De manera análoga, se vislumbró que hay una identidad del 100% entre el elemento de respuesta a HGF –los CRE y los sitios de unión a Sp1- y la región promotora de *sno* en *Homo sapiens*, debido a que el HGF se detectó en el genoma humano.

Ya que el TRE y el HGFRE son regiones regulatorias, y las regiones regulatorias están frecuentemente conservadas, se esperó encontrar una identidad muy cercana al 100% entre los SBEs y el SIE reportados por Zhu *et al.*

(2005) y la región promotora de *sno* en *Homo sapiens*; así como entre el elemento de respuesta a HGF y la región promotora de *sno* en *Mus musculus*. Efectivamente, se detectaron los componentes del TRE en *Homo sapiens* (91%, 88%, 96%, 100% y 82.6% de identidad en relación con los SBE y el SIE de *Mus musculus*) y los TFBS que conforman el HGFRE en *Mus musculus* (con una identidad del 100%, 75%, 83% y 75% con relación al Sp1ABE, CRE1, Sp1BBE y CRE2 de *Homo sapiens*), como se muestra en la Figura 11.

### Comparación entre las secuencias regulatorias

Un alineamiento entre las secuencias regulatorias de *sno* de *Homo sapiens* y *Mus musculus* con ClustalX muestra un grado de conservación del 71% entre las mismas. Los TFBS detectados experimentalmente por Ruoyun *et al.* (2007) y Zhu *et al.* (2005) en *Homo sapiens* y *Mus musculus*, respectivamente, se compararon con el programa Blast (Align two sequences). Este resultado se muestra en la figura 11. Se enfatizan los TFBS detectados experimentalmente y se señala el grado de similitud entre los TFBS de las dos especies analizadas.





## Determinación del TSS

A continuación se detallan los resultados obtenidos con las tres técnicas de detección de TSS empleadas. Al final de la sección se compararán dichos resultados.

### 1. Alineamiento de mRNA con la secuencia promotora.

#### a. *Mus musculus*

La secuencia [U14655](#) de mRNA SnoN2 se alineó con los 5101 pares de bases (ver método) de la región promotora de la base 5000 a 5101 (Figura 12).

Por tanto, según este método el TSS corresponde a la base 5000 de la secuencia promotora.

#### b. *Homo sapiens*

La secuencia del promotor de *sno* y la del mRNA [U70730](#) se alinearon entre las bases 4292 y 5000 del promotor y las primeras 709 bases del mRNA. Estas últimas corresponden justamente a la región 5'UTR (Figura 12).

Por lo anterior, este alineamiento predice que el TSS corresponde a la base 4292 de la secuencia promotora.

<i>Mus musculus</i>	
Score = 194 bits (101), Expect = 2e-45	
Identities = 101/101 (100%), Gaps = 0/101 (0%)	
Strand=Plus/Plus	
Promotor	5001 CAGATTAGGAACTGAGAACAACATGCCAGTTGCAGACTTAATCGTTTAAAGCACCAGTGT 5060
mRNA SnoN2	1 CAGATTAGGAACTGAGAACAACATGCCAGTTGCAGACTTAATCGTTTAAAGCACCAGTGT 60
(...)	
Promotor	5061 GGCCATGGAAAACCTCCAGTCTAAGTTCTCCTTAGTTCAGG 5101
mRNA SnoN2	61 GGCCATGGAAAACCTCCAGTCTAAGTTCTCCTTAGTTCAGG 101
<i>Homo sapiens</i>	
Score = 1358 bits (706), Expect = 0.0	

```

Identities = 708/709 (99%), Gaps = 0/709 (0%)
Strand=Plus/Plus

Promotor 4292 GGTTCAAATTGGCCCTTTGGCCTCTGGAGCAAATTCAAATGTAACCTTCCCCAACCCC 4351
|||||
mRNA SnoN2 1 GGTTCAAATTGGCCCTTTGGCCTCTGGAGCAAATTCAAATGTAACCTTCCCCAATCCC 60

( . . . )

Promotor 4952 AAATAGAAATGCTAATCTCAGACTTAATTATTTAACAGAAGAGTGTTACC 5000
|||||
mRNA SnoN2 661 AAATAGAAATGCTAATCTCAGACTTAATTATTTAACAGAAGAGTGTTACC 709

```

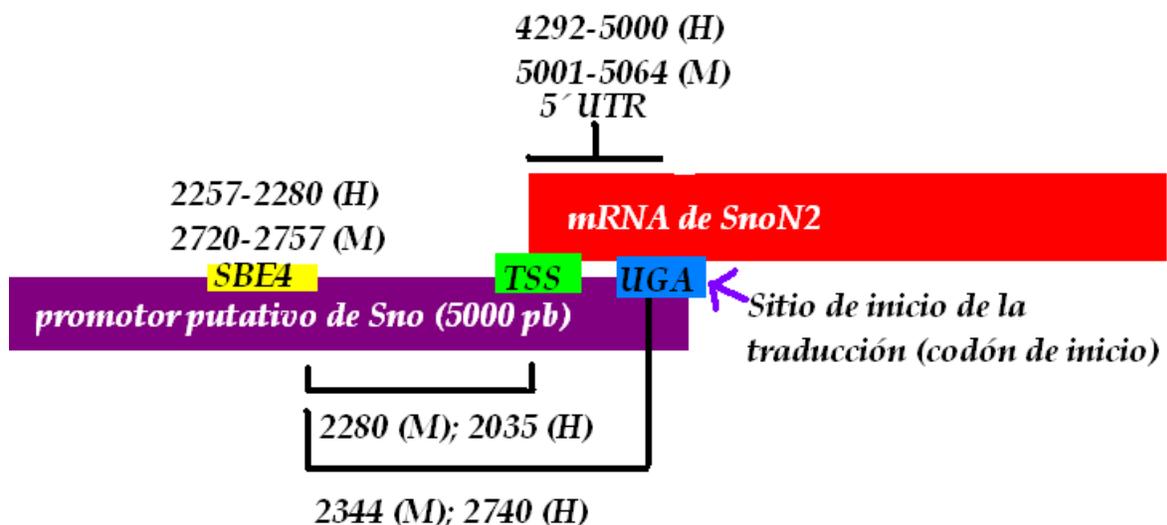
**Figura 12.** Resultado de los alineamientos entre las región promotora y los mRNA de SnoN2 de *Homo sapiens* y *Mus musculus*, realizados con el programa Blast2Sequences.

El SBE4 se encuentra en los intervalos 2,720–2,757 y 2,257–2,280 de las secuencias promotoras de 5kb de *Mus musculus* y *Homo sapiens*, respectivamente.

Por lo tanto, la distancia entre la SBE4 y el TSS es de 2280 pb en *Mus musculus*, y de 2035 en *Homo sapiens*.

Si se agrega a estas distancias la longitud de la región 5' UTR, la distancia entre el SBE4 y el codón de inicio (primer triplete del CDS) es de 2344 en *Homo sapiens*, y de 2740 pb en *Mus musculus*. Este sitio corresponde al sitio de inicio de la traducción (mRNA → proteína).

En la figura 13 se representan estos alineamientos y las distancias al SBE4 calculadas.



**Figura 13.** Esquema de los alineamientos entre los mRNA de SnoN y la secuencia promotora.

## 2. Detección de TSS con DBTSS

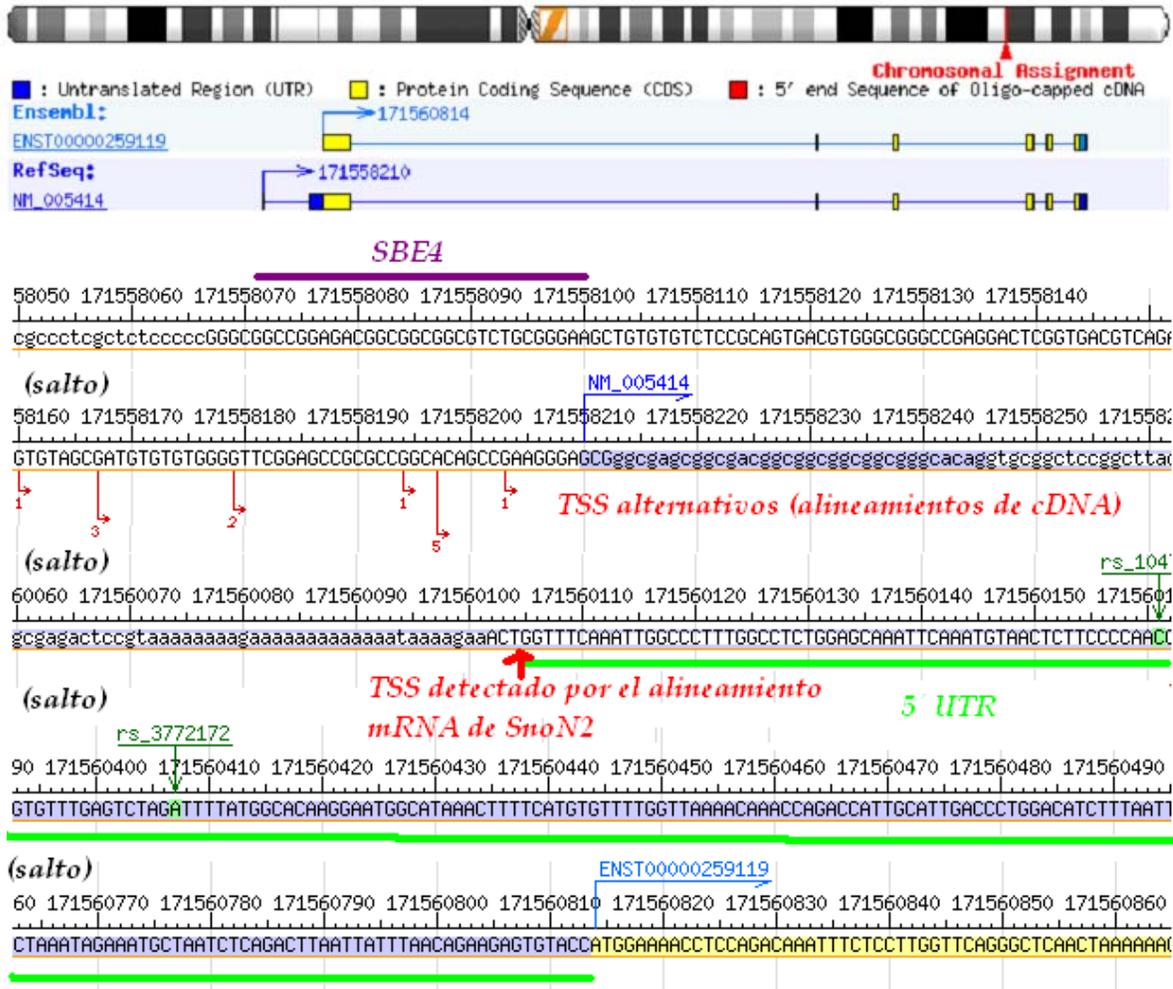
DBTSS (*Data Base of Transcriptional Start Sites*) es una base de datos de TSS que contiene una colección de más de 19 millones de secuencias 5'-terminal de cDNA determinadas experimentalmente (Wakaguri *et al.*, 2007). DBTSS contiene herramientas analíticas entre las que se encuentra un buscador de TSS.

Como se muestra en la figura 14, DBTSS detectó un gran número de TSS putativos en los promotores de *sno* humano (NM\_005414) y murino (NM\_011386 y NM\_001039090). En particular, detecta un TSS en la región correspondiente al final de la región no traducible (UTR) 5' de SnoN2 en ambas especies.

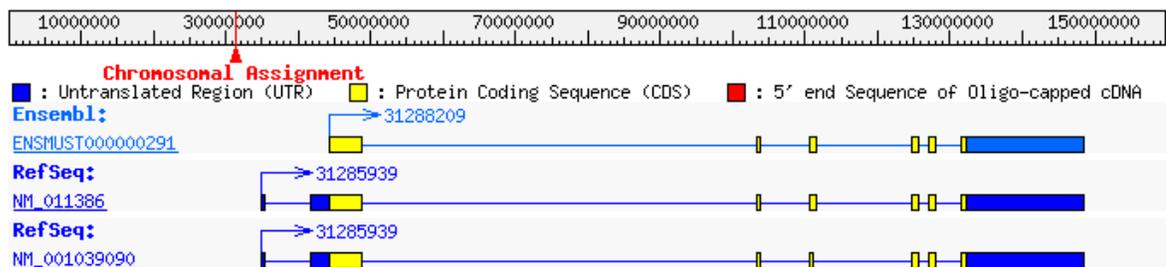
Se muestra, para cada especie de estudio, la región cromosómica en la que se encuentra *sno*. Más abajo, se encuentra un esquema del gen *sno*. En amarillo se marca la región traducible (CDS), es decir, los exones. En azul se marcan las regiones no traducibles (UTR); se trata tanto de los intrones como de las regiones correspondientes al sitio de poliadenilación (3' UTR) y de *Capping* (5' UTR). La flecha en la primera línea (en la posición 171560814 del cromosoma 3 de *Homo sapiens*, y 31288209 de *Mus musculus*) indica en sitio de inicio de la traducción. Las secuencias ENST00000259119 en *Homo sapiens* y ENSMUST000000291 en *Mus musculus* forman parte del CDS del gen *sno*. A continuación se observa el TSS putativo que DBTSS predice como principal. En el caso de *Homo sapiens*, se trata del sitio 17155821 del cromosoma 3, y en *Mus musculus*, al 31285939. Las flechas rojas indican TSS alternativos putativos, que DBTSS detecta por patrones de extremos de cDNA 5'.

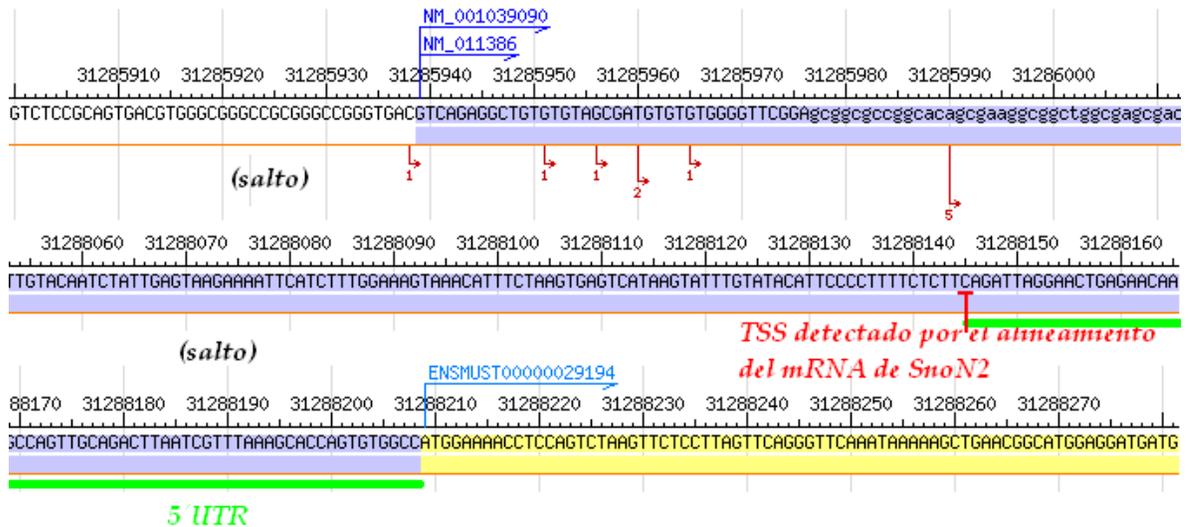
Ninguno de los TSS que DBTSS predice, corresponde exactamente con el TSS que se encontró por medio del alineamiento con mRNA de SnoN2. Sin embargo, este sitio se localiza entre los TSS y el sitio de inicio de la traducción que DBTSS predice. En verde se marca la región 5'UTR, y en rojo se señala el TSS que se encontró mediante el primer método.

**Homo sapiens**



**Mus musculus**





**Fig 14.** Detección de TSS en los promotores de *sno* de *Homo sapiens* y *Mus musculus* con el programa DBTSS. Las flechas rojas muestran que DBTSS detecta un gran número de TSS putativos en los promotores de *sno* de ambas especies de estudio. La flecha azul (ENST00000259119 en *Mus musculus* y ENSMUST00000029194 en *Homo sapiens*) corresponde al final del 5'UTR de mRNA de SnoN2 de ambas especies.

### 3. Alineamiento “manual” de cDNAs con las secuencias promotoras.

Existen dos isoformas de cDNA para *sno* en *Mus musculus* (SnoN2, SnoN) y cuatro en *Homo sapiens* (SnoN, SnoN2, Sno1 y SnoA). Estas isoformas son producto de *splicing alternativo* (Macías-Silva y Briones-Orta, 2008). Como última aproximación para predecir el TSS, se alinearon estos cDNA con el promotor de *sno* de las especies de estudio.

En el caso de *Mus musculus*, las secuencias de mRNA U36203 (SnoN) y U14655 (SnoN2) se alinearon con una identidad del 100% con las bases 5001 – 5101. Este resultado era de esperarse, ya que SnoN2 y SnoN se diferencian sólo por una inserción de 138 nucleótidos entre las bases 1343 y 1344 de SnoN que no presenta SnoN2 gracias a la presencia de un donador de *splice* críptico. Por lo tanto, las primeras 1342 bases son idénticas entre ambas isoformas, y se alinean así con la misma región del promotor. En ambos casos, la secuencia 5'UTR corresponde a las bases 1-64. El TSS corresponde así a la base 5000, mientras que el Sitio de Inicio de la Traducción corresponde a la

base 5064. Estos resultados son idénticos a los que se obtuvieron con el primer procedimiento (Figura 13). Así mismo, el final de la región 5'UTR corresponde al TSS ENSMUST000000291 31288202 detectado por DBTSS (Figura 14).

El promotor de *Homo sapiens* se alinea con las secuencias X15219 (SnoN, versión oncogénica; presenta dos mutaciones puntuales en los nucleótidos 269 y 302), U70730 (SnoN2) y X15217 (SnoA, versión oncogénica; el nucleótido 302 exhibe una mutación puntual de G a A) en sus últimas 709 bases, es decir, de la base 4292 a la 5000. En estos casos, la secuencia 5'UTR corresponde a las bases 1-709 de las secuencias de mRNA. De hecho, las isoformas SnoN, SnoN2 y SnoA difieren entre sí por la presencia o ausencia de determinadas secuencias. SnoN es en este sentido la secuencia de mRNA mínima; SnoN2 presenta una inserción de 138 pb entre las bases 2000 y 2001 de SnoN y SnoA, una secuencia *Alu* a partir de la base 1807. Así, estas diferencias entre las isoformas no afectan el primer exón y por lo tanto la determinación del TSS por medio de alineamientos. El TSS detectado está a 2035 del SBE4; el final de la región 5'UTR, a 2344 pb. Estos datos corresponden al TSS detectado por medio del primer método (Figura 13). El codón de inicio (final de la región 5'UTR) corresponde al TSS ENTS00000259119 que se encuentra en la base 171560814 del cromosoma 3 de *Homo sapiens* y que DBTSS detecta (Figura 14). La isoforma Snol no está representada en la base de datos de GenBank por una secuencia completa. La secuencia seleccionada, Z19588, no presenta la región 5'UTR. El alineamiento entre Z19588 con X15219 (SnoN) se da sólo a partir de la base 747 de SnoN, lo que sugiere que tampoco la región CDS –a partir de la base 709- está completa en Snol. Lo anterior explica que Z19588 se alinea con la secuencia promotora de *Homo sapiens* únicamente a partir de la base 5038 de la última. Sin embargo, debido a que la secuencia Z19588 de Snol no presenta la región 5'UTR, no se considera este alineamiento para determinar un nuevo TSS.

## Comparación de los resultados obtenidos.

El común denominador entre los tres métodos es el TSS que se encuentra a 2280 pb en *Mus musculus* y de 2035 en *Homo sapiens* del SBE4. Esta ubicación corresponde al inicio del alineamiento de la región 5'UTR de mRNA de SnoN y de las diferentes isoformas de cDNA con los promotores de *sno*.

El programa DBTSS consideró como TSS a la posición que se encuentra al *final* de esta región 5'UTR. Sin embargo, ya que esta base forma parte del codón de inicio, corresponde más bien al sitio de inicio de la traducción. Así, en el presente trabajo se considera como TSS putativo al nucleótido que se encuentra 2280pb y 2035pb río abajo del SBE4 en *Mus musculus* y en *Homo sapiens*, respectivamente.

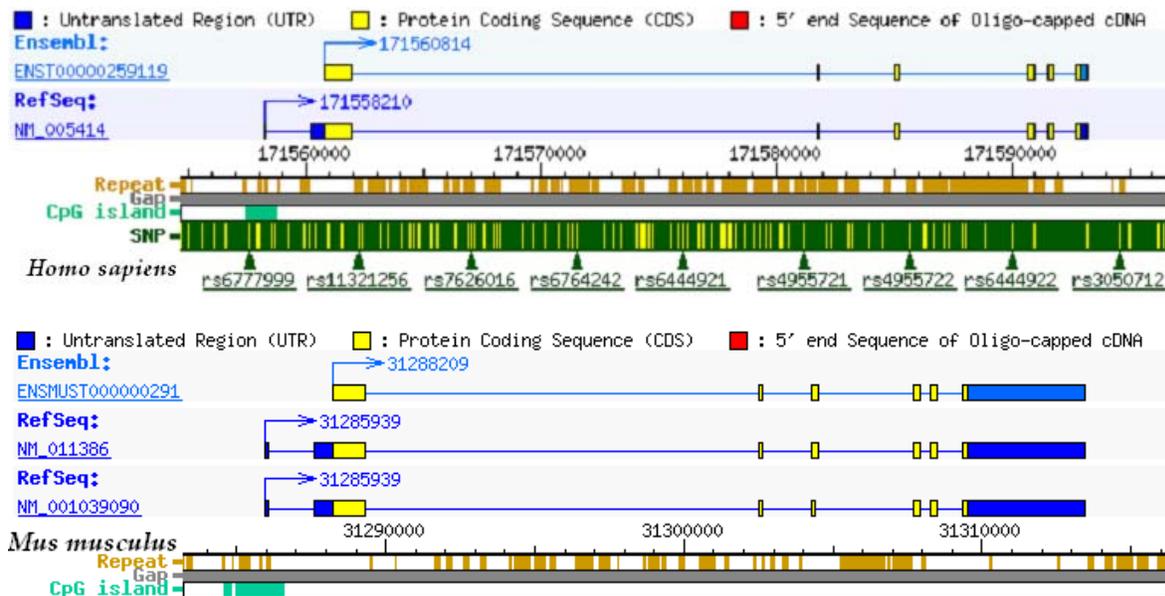
## **Búsqueda de cajas regulatorias en la zona núcleo**

La zona núcleo es la región del promotor que se encuentra rodeando el TSS y a la que se une el aparato basal de la transcripción (Sandelin *et al.*, 2007). La unión entre el aparato basal de la transcripción y la zona núcleo del promotor sucede por medio de motivos de unión a factores basales de la transcripción. Estos motivos frecuentemente están representados por cajas regulatorias, que son TFBS para factores basales de la transcripción. En este sentido, el TFBS más común, conocida como caja TATA, está constituida preferencialmente por las bases TATAAAA. Sin embargo, existe un gran número de secuencias ricas en AT de seis a siete pb que pueden fungir como caja TATA siempre y cuando se encuentren en la posición 25 – 30 río arriba del TSS (Smale, 2005). La caja TATA une al Factor de Transcripción IID (TFIID) y a la TBP (Sandelin *et al.*, 2007).

Por otro lado, se ha propuesto que existe una alta correlación entre una alta frecuencia de CpGs en la zona núcleo del promotor y la ausencia de las cajas regulatorias, en particular a la caja TATA (Bajic, *et al.*, 2003).

En este sentido, el análisis de la zona núcleo del promotor de *sno* de *Homo sapiens* y *Mus musculus* ha reflejado dos resultados aparentemente contradictorios.

Por un lado, un examen preliminar de esta secuencia con el programa DBTSS muestra que la región en la que se encuentran inmersos algunos de los TSS detectados por este programa, es rica en CpGs tanto en *Homo sapiens* como en *Mus musculus* (Figura 15).



**Figura 15.** Inmersión de los TSS 171558210 (*Homo sapiens*) y 31285939 (*Mus musculus*) en una isla rica en CpGs. Resultados obtenidos con el programa DBTSS.

Por otro lado, la exploración más detallada de esta zona con el programa *Algen* muestra que sí existen TFBS putativos para TBP y TFIID.

En el promotor de *Homo sapiens*, las bases 13 y 6 río arriba del TSS corresponden a la secuencia AATAAAA, que es 95.89% idéntica a la secuencia consenso para TBP (caja TATA) ATATAAAA. Así mismo, las bases 12 a 5 presentan sólo un 2.66% de disimilitud con respecto a la secuencia consenso TATAAAA que une TFIID (Tabla 5).

**Tabla 5.** Búsqueda de cajas regulatorias en la zona núcleo del promotor de *sno* de *Homo sapiens* y *Mus musculus* con el programa *Alggen*.

	<i>Homo sapiens</i>	<i>Mus musculus</i>
<b>TBP</b>		
Alta similitud	AAATAAAA (13-6, 4.11%)	
	CCGTAAA (36- 29, 22.31%)	
	GTAAAAA (34-27, 19.02%)	
	TAAAAAAA (33-74, 23.13%)	
Locación cercana a las bases 25 a 30 río arriba del TSS	AAAAAAA (32-25, 22.85%)	TCATAAGT (34-27, 24.08%)
	AAAAAAG (31-24, 24.17%)	ATTTGTAT (26-19, 20.51%)
	AAAGAAAA (27-20, 22.85%)	TTGTATAC (24-17, 19.02%)
	AGAAAAA (25-18, 21.20%)	GTATACAT (22-15, 20.79%)
	GAAAAAA (24-17, 23.13%)	
	AAAAAAA (23-16, 22.85%)	
	AAAAAAA (22-15, 22.85%)	
<b>TFIID</b>		
Alta similitud	AATAAAG (12- 5, 2.66%)	
	CGTAAAA (35-28, 20.17%)	
	TAAAAAAA (33-26, 18.99%)	
	AAAAAAA (32-25, 21.36%)	CATAAGTA (33-26, 21.38%)
Locación cercana a las bases 25 a 30 río arriba del TSS	AAAAAAG (31-24, 21.64%)	TATTTGTA (27-20, 20.19%)
	AAAAAAGA (30- 23, 23.77%)	TTTGTATA (25-18, 18.99%)
	AAGAAAA (26-19, 21.36%)	TATACATT (21-14, 20.48%)
	GAAAAAAA (24-17, 20.17%)	
	AAAAAAA (23-16, 21.36%)	
	AAAAAAA (22-15, 21.36%)	
	AAAAAAA (21-14, 21.36%)	

En el caso de *Mus musculus*, no hay una secuencia que *Alggen* detecte como caja TATA (TFBP para TBP y o para TFIID) con un índice de disimilitud menor al 19% (Tabla 5).

Este hecho concuerda con los resultados preliminares obtenidos con el programa DBTSS, que muestran para esta especie una mayor cantidad de CpGs en la zona núcleo de su promotor de *sno* que en la secuencia análoga de *Homo sapiens* (Figura 15).

El resto de la zona núcleo no presenta TFBS putativos con una disimilitud menor al 20% para factores basales de la transcripción tales como TBP o el TFIID. Sin embargo, se sabe que es más importante la ubicación de la caja TATA (entre

las bases 25 y 30 río arriba del TSS) que el grado de conservación de su secuencia (ATATAAAA) (Smale, 2005).

Las secuencias que *Alggen* detecta en esta región están efectivamente poco conservadas; la disimilitud oscila entre el 19 – 25%. En la Tabla 5 se muestra este resultado anotando la secuencia detectada como caja TATA y entre paréntesis la posición de la misma con relación al TSS, así como el grado de disimilitud respecto a la secuencia consenso ATATAAAA.

La probabilidad de que un TFBS putativo sea funcional aumenta considerablemente si esta región se encuentra conservada (Loots y Ovcharenko, 2004). Los TFBS putativos para TBP y para TFIID se detectaron exclusivamente en *Homo sapiens*, por lo que no están conservados entre las dos especies de estudio. Por lo tanto, es probable que ningún promotor de *sno* de las especies de estudio presente cajas regulatorias. Esta visión se apoya además en el alto contenido de CpGs en la zona núcleo del promotor y en que DBTSS no detecta cajas regulatorias (Figura 14). Sin embargo, no se puede descartar *a priori* una diferencia interespecífica, por lo que la ausencia de cajas regulatorias en los promotores de *sno* de *Homo sapiens* y *Mus musculus* debe ser comprobada experimentalmente.

## **Detección de TFBS**

### ***Validación de los programas***

Existe un gran número de programas de búsqueda de TFBS putativos. En el presente trabajo, se eligió aquel que fue capaz de encontrar todos los TFBS que previamente se determinaron experimentalmente, es decir, los SBE y el SIE determinados por Zhu *et al.* (2005) y el Elemento de Respuesta a HGF que encontraron Ruoyun *et al.* (2007). En este sentido, se probaron los programas *Alggen* ([Alggen.lsi.upc.es/](http://Alggen.lsi.upc.es/)) y *Genomatix* ([www.genomatix.de/](http://www.genomatix.de/)).

Entre todos los TFBS que los programas debían encontrar, Genomatix detectó únicamente SBE2 y SBE3 en *sno* de *Mus musculus* y SBE3, CRE1, CRE2 y Sp1B en *sno* de *Homo sapiens* (Figura 16).

### *Mus musculus*

	Opt.	Position	Str.	Core sim.	Matrix sim.	Sequence (red: ci-value > 60 capital: core sequence)
		from - to				
Sbe2	0.99	68 - 76	(-)	1.000	0.993	GTCTgggcg
Sbe3	0.99	227 - 235	(-)	1.000	0.991	GTCTggcct

### *Homo sapiens*

	Opt.	Position	Str.	Core sim.	Matrix sim.	Sequence (red: ci-value > 60 capital: core sequence)
		From - to				
Sbe3	0.99	209 - 217	(-)	1.000	0.991	GTCTggcct
CRE1	0.89	795 - 815	(+)	1.000	0.905	cgctgtagtTGACgggagccg
CRE2	0.90	831 - 851	(-)	1.000	0.919	gggcccTGACgggcgtccggg
	0.89	834 - 854	(-)	1.000	0.932	ggggggcggTGACgggcgtcc
Sp1A	0.88	841 - 855	(-)	1.000	0.911	agggGGGCggtgacg
Sp1B	0.88	848 - 862	(-)	0.807	0.911	ctggGGGAggggggc

### En resumen:

	<i>Mus musculus</i>	<i>Homo sapiens</i>
Sbe1	No	No
Sbe2	Sí GTCTgggcg, 99%	No
Sbe3	Sí GTCTggcct, 99%	Sí GTCTggcct, 99%
Sbe4	No	No
Sp1A	No	No
CRE1	No	Sí cgctgtagtTGACgggagccg, 90%
CRE2	No	Sí gggcccTGACgggcgtccggg, 91%
Sp1B	No	Sí agggGGGCggtgacg, 88%

**Figura 16.** Búsqueda fallida de los TFBS detectados experimentalmente en los promotores de *sno* de *Mus musculus* y *Homo sapiens* con el programa Genomatix.

Por otro lado, el programa *Algen* fue capaz de detectar todos los SBEs, el SIE, los CRE y los elementos de unión a Sp1 en *sno* tanto de *Mus musculus* como de *Homo sapiens* (Tabla 5).

**Tabla 6.** Detección eficaz de los TFBS previamente determinados experimentalmente con el programa *Alggen*. Se muestra la secuencia del TFBS determinado experimentalmente, la posición del mismo en la secuencia regulatoria, así como los motivos que *Alggen* detectó. Los porcentajes muestran el índice de disimilitud con relación a la secuencia consenso.

TFBS	Posición*	Secuencia ( <i>TFBS putativo</i> )**	%***
<b><i>Mus musculus</i></b>			
SBE1	24-58	atcc <b>tgctgagcc</b> gggc <b>accgcagaca</b> gcgccc	94, 95
SBE2	62-87	ccagc <b>cgcccagacg</b> gcgcccggg	79
SBE3	219-245	ccacaat <b>aggccagacg</b> gcgcccgcg	81
SBE4	333-365	Cggccggagacggcggcg <b>cgctgcggg</b> aag	79
CRE1	758-766	<b>tgacccga</b>	79
CRE2	792-800	<b>cctggtca</b>	88
Sp1B	801-806	<b>tcgccc</b>	91
<b><i>Homo sapiens</i></b>			
SBE1	61-96	Atcc <b>tgctgcgcc</b> ggggtg <b>caccg</b> cagacagcgcg	95, 93
SBE2	101-125	Ccagc <b>ttcccagacggcg</b> cgccggg	79
SBE3	202-227	Ccacaat <b>aggccagacg</b> gca cgcccgcg	82
SBE4	357-386	Cggccggagacggcggcg <b>tctgcggg</b> a	84
Sp1A	729-736	<b>aggcgg</b> g	91
CRE1	803-811	<b>Tgacggga</b>	93
CRE2	837-845	<b>gcccgta</b>	90
Sp1B	846-852	<b>ccgcc</b>	100

\*Posición relativa a la secuencia regulatoria.

\*\* *Alggen* predice la unión de Smad3.

\*\*\*identidad entre secuencia y TFBS prototipo (TRANSFAC)

En conclusión, el programa que se empleó para el presente trabajo es *Alggen*, ya que predijo satisfactoriamente los TFBS determinados experimentalmente en *sno* de *Homo sapiens* y *Mus musculus*.

### ***Detección de TFBS putativos***

La secuencia regulatoria se encuentra en la región comprendida entre las bases 2709 y 1791 río arriba del TSS en *Mus musculus* y 3090 a 2190 río arriba del TSS en *Homo sapiens*. Recuérdese que la secuencia regulatoria se refiere a las 918 y 900 pares de bases que contienen tanto al Elemento de Respuesta al TGF $\beta$  (TRE) –es decir, los SBE y el SIE- como al HGF-RE de *Mus musculus* y *Homo sapiens*, respectivamente.

Los análisis de TFBS putativos en las secuencias regulatorias del promotor de *sno* de *Homo sapiens* y *Mus musculus*, realizados con el programa *Alggen*, muestran que existen TFBS potenciales para prácticamente todos los factores de transcripción capaces de asociarse a las Smad y reseñados por Feng y Derynck en 2005.

En la tabla 7 se muestran los TFBS putativos que el programa *Alggen* detectó en la secuencia regulatoria del promotor de *sno* de las dos especies estudiadas.

En la parte superior (A) de la Tabla 7 se muestran las predicciones de los sitios de unión para TF capaces de interactuar con las Smad. Se menciona la Smad y el dominio con el que cada TF interactúa, así como las características y el mecanismo de interacción, tal como lo reseñan Feng y Derynck en 2005. La última columna muestra el nombre y la clave de la base de datos TRANSFAC del TF para el cual *Alggen* detecta un TFBS putativo con menos de 25% de disimilitud.

En la parte inferior (B) de la Tabla 7 se presentan los TF para los cuales *Alggen* detectó un TFBS putativo con menos de 25% de disimilitud con relación a la secuencia consenso. Los TF que ahí se presentan no necesariamente son capaces de interactuar con las Smad, por lo que no se comentan mayormente.

**Tabla 7:** TF para los que *Alggen* detecta un TFBS putativo en la secuencia regulatoria de *sno* de *Homo sapiens* y de *Mus musculus* con al menos 85% de identidad con relación a la secuencia consenso.

A) TFBS putativos para TF capaces de interactuar con las SMAD (Feng y Derynck 2005)			
TF	Smad /dominio	Características/mecanismos de acción	Nombre y clave de TF (TRANSFAC)
<u>Familia bHLH</u>			
E2F4/5	Smad3 (MH2)	Recluta p107 a Smad3 para reprimir al gen <i>c-myc</i>	<u>E2F [T01547]</u>
MyoD	Smad3 (MH1, linker)	Smad3 interfiere en la formación del complejo MyoD/proteína E/DNA	<u>MyoD [T00526]</u>
<u>Familia bZIP</u>			
ATF2	Smad3/4 (MH1)	Smad3/4 estimula la transactivación de ATF2	<u>ATF-2 [T00167]</u>
ATF3	Smad3 (MH2)	Reprimen transcripción de Id1	<u>ATF3 [T01313]</u>
c-Fos	Smad3 (MH2)	Sinergia en la regulación de genes blanco del TGFβ dependientes de AP-1	<u>c-Fos [T00122], [T00123], [T00124], JunB [T00436], JunD [T00437, cJun [T00132], [T00133], [T00131]</u>
c-Jun, JunB, JunD	Smad3 (MH1), Smad4	Regulan positiva y negativamente a las Smad	
<u>Familia de Receptores nucleares</u>			
AR	Smad3 (MH2)	Inhibición recíproca de la actividad de unión a DNA de Smad3 y de AR	<u>AR [T00042], AR [T00040]</u>
ER	Smad1/3/4	Represión de genes blanco del TGFβ	<u>ER-beta [T04651],</u>

	(MH2)		<a href="#">ER-alpha [T00261]</a> <a href="#">GR-beta [T01920]</a> , <a href="#">GR [T00335]</a> , <a href="#">GR-alpha [T00337]</a> , GR [T00333] <a href="#">HNF-4alpha1 [T00372]</a> <a href="#">RXR-alpha [T01345]</a> , <a href="#">RXR-beta [T01349]</a> , RXR:RXR-alpha [T05670] <a href="#">VDR [T00885]</a>
GR	Smad3 (MH2)	Inhibición de la actividad de transactivación de Smad3	
HNF4	Smad3/4	Activación Cooperativa	
RXR	Smad3 (MH2)	Feng y Derynck (2005) no reportan mecanismos de interacción entre estos dos TF	
VDR	Smad3 (MH1)	Smad3 sinergiza la transactivación de VDR inducida por ligando <i>Familia Runx</i>	
CBFA1/ Runx2/A ML	Smad1/2/3/5 (MH2)	Este complejo coopera con BMP en la regulación de la respuesta inmune <i>Familia de proteínas con dedos de Zinc</i>	<a href="#">AML1 [T01067]</a>
GATA3	Smad3	GATA3 recluta Smad3 a sitios GATA para activar la transcripción conjuntamente	<a href="#">GATA-3 [T00311]</a>
Sp1	Smad2 (MH1) Smad4 (MH2)	Sinergia en la inducción de la transcripción de genes blanco del TGF- $\beta$ , como p15Ink4B, p21Cip1, Smad7, PAI-1 y colágena	<a href="#">Sp1 [T00754]</a> , <a href="#">[T00759]</a> , <a href="#">[T00755]</a> , <a href="#">[T00753]</a> , <a href="#">[T00752]</a> <a href="#">YY1 [T00865]</a> , <a href="#">[T00915]</a> , <a href="#">[T00278]</a> , <a href="#">[T04970]</a>
YY1	Smad1, Smad4 (MH1)	Forma complejo con Smads y GATA	
HIF-1 $\alpha$	Smad3 (MH1, MH2)	<i>Otros</i> Cooperación del TGF- $\beta$ con vía de hipoxia y angiogénesis	<a href="#">HIF-1 [T01609]</a>
IRF-7 (IRFs)	Smad3 (MH2)	Activación de Smad3 de la función de transactivación de IRF-7	<a href="#">IRF-2 [T01491]</a> , <a href="#">[T00425]</a> , <a href="#">IRF-1 [T00423]</a> <a href="#">LEF-1 [T02905]</a> , <a href="#">TCF-2 [T01110]</a> , <a href="#">TCF-4E [T02878]</a> , <a href="#">TCF-4 [T02918]</a> , <a href="#">TCF-1(P) [T01109]</a> <a href="#">NF-kappaB [T00588]</a> , <a href="#">[T00590]</a> <a href="#">[T00671]</a> , <a href="#">[T01806]</a> <a href="#">SRF [T00763]</a>
Lef1/ TCF	Smad1/2/3/4 (MH1, MH2)	Smad sinergiza a LEF1	
NFkB p52 p53 SRF	Smad3 Smad2/4 Smad3	Coactivación de sitio kB Sinergia y antagonismo Media transcripción de SM22 $\alpha$ inducida por TGF- $\beta$	
c-Myc	Smad2/3 (MH2)	Inhibición del complejo activador Smad-Sp1	<a href="#">c-Myc [T00140]</a>

B) TFBS putativos para TF que no necesariamente son capaces de interactuar con las SMAD

Adf	CRE-BP2	HIF-1	Nrf2:MafK	T3R-a,b,b1
AhR	CREMtau1,2,a	HNF1,1C,3,3a, B,4a	Oct-B1	TBP
AML	Crx	HOXA4,5	p53 (supresor de tumores) Pax1,2.2,2a,5,6, 9a,9b (desarrollo)	TCF-1P,2,4,4E
AP-1,2,3,4	DBP	HOXD8		TFIID
AR	DEC2	HSF1 (long, short)	PEA3	TGGCA-Binding Protein
AREB6	DI	INSAF	PEBP2	TII
ARP1	DSXF	IPF1	POU(1- F1,2F2(Oct-	TSAP

			2.6),2FB	
ATF2,3	DSXM	IRF-1,2	PR A,B	Ttk 69K
Bcd	DTF-1	Kr	Prd	unc-86
BTEB3,4	E2F-1	LEF-1	PTF-1b	USF1,2
C/EBPa,b,g,d	E47	LF-A1	PU.1	VDR
CAG-binding protein	Elk.-1	LyF1	PURa,b	WT1,KTS
CArG-box binding protein	ENKTF-1	Mad	RARa1,b,g	XBP-1
CBF	ER-a,b	MAZ	RelA	XPF-1
c-Ets-1,2	F(a)-f(e)	MF3	RXRa,b,SXR:R XR-a	Yi
c-Fos	FOXP3	MyoD	SF-1	YY1 (Miogénesis)
c-Jun, JunB, JunD	Fra-1	Myogenin	Smad3 (TGFβ)	Zeste
c-Myc	Ftz	NF-1,1L,AT1,AT2,E4,X3,Y	Sp1	
COE	GAGA factor	NFI/CTF	Sp3	
COUP-TF1	GATA1,2,3	NFκB (sistema inmune)	SREBP-1c	
CP2	GCMa,b	NHP-1	SRF	
CREB	GR(a,b)	Nkx2-1,6-2	STAT1b,4 (vía Jak/STAT)	
CREBb	HES-1			

Con el fin de inspeccionar más detalladamente el papel de los factores de transcripción que se enlistan en la Tabla 7a sobre la regulación de *sno* en el contexto del TGFβ, se analizó posteriormente el TRE. El TRE forma parte de la secuencia regulatoria. Se encuentra entre 2461 y 2127 bases río arriba del TSS identificado para el gen de *Mus musculus* y entre la base 2381 y 1882 río arriba del TSS del gen de *Homo sapiens* y contiene los SBEs y el SIE detectados por Zhu y colaboradores en 2005. Como este mismo grupo mostró a través de mutagénesis dirigida, estos “SBE” presentan más de un TFBS para las Smad. Los sitios de unión a las Smad, detectados experimentalmente de manera puntual por mutagénesis dirigida y que se encuentran inmersos en cada “SBE” se enfatizan por medio de un marco y se nombran como SBE1A, SBE1B, SBE2A, etcétera.

El TRE se analizó nuevamente con el programa *Algen*, pero se disminuyó el grado de similitud máximo de 25% (porcentaje que permitió detectar la totalidad de TFBS encontrados experimentalmente) a 15%.

Con estas restricciones (tamaño de la secuencia, grado de disimilitud), el número de TFBS putativos disminuyó significativamente con relación al primer

análisis. Como se muestra en la tabla 7, con el análisis de la secuencia regulatoria entera se detectaron 120 TF para los que *Alggen* predice TFBS con una disimilitud menor al 25% con relación a la secuencia consenso. De estos 120 TF, 25 son capaces de asociarse con las Smad, como reseñan Feng y Derynck (2005).

Los factores de transcripción que se descartaron con este método fueron AP1,2 y 3; ARP1; BDC; Cbf; CoupTF1; Crx; dec2; DI; dtf1; Elk-1 (MAPK); ftz; GCM $\alpha$  y  $\beta$ ; HSF1 *long* y *short*; IPF1 y 2; Kr, Lyf1, NF1L; NFAT2; NFY; NFICTF; PEBP2; POU2F2 y 2F2 (Oct2.6); Prd; Ptf-1b; Pu1, SREBP1c, STAT1b, TFIID, TGGCA-BP; TII; TSAP; Ttk69K, XBO-1, XBF1 y Zeste. Se descartaron también los TF que potencialmente son capaces de interactuar con las Smad que a continuación se mencionan: ER- $\alpha$  y  $\beta$ ; c-EBP $\gamma$ ; FoxP3, GATA1,2 y 3; HIF1, HoxA4, HoxD8, HNF1, 1c, 3 y 4a, Factor Regulado por Interferón (IRF)-1 y 2; Lef 1; Nkx2-1,6-2; OctB1, RAR-A1, SRF, TCF-1p, 2, 2 y 4E y TBP.

En la Tabla 8 se muestran los 66 TF detectados por *Alggen* después de esta depuración. En un recuadro se resaltan los 17 grupos de TF capaces de interactuar con las Smad.

En la figura 16 se muestran los TFBS putativos para los factores de transcripción capaces de asociarse a las Smad que *Alggen* detecta en los TRE del gen de *sno* de *Homo sapiens* y *Mus musculus* con un índice de disimilitud máximo del 15%.

**Tabla 8.** Factores de Transcripción para los cuales se detectaron TFBS putativos en el TRE de *Homo sapiens* y *Mus musculus*, según el programa *Alggen* (grado de disimilitud del 15%). En amarillo se resaltan aquellos que son capaces de interactuar con las Smads.

Adf Ahr	Coe c-Ets-1,2	DSXF DSXM	HOXA5 NF $\kappa$ B	RARb,g RelA	USF1,2 VDR
AML	c-Fos	E2F-1	NHP-1	RXRa,b, SXR:RXR-a	Yi
AR	c-Jun, JunB, JunD	E47	Nrf2:MafK	SF-1	YY1
AREB6	CP2	ENKTF-1	p53	Smad3	
ATF2,3	CREB	f( $\alpha$ )-f( $\epsilon$ )	Pax1,2,5,6,9	Sp1	
BTEB3,4	CREBb	Fra-1	PEA3	Sp3	
C/EBPa,b,d	CRE-BP2	GAGA factor	POU(1F1,2FB)	STAT4	
CAC-BP	CREMtau1,2,a	GR(a,b)	PR a,b	T3R-a,b	
CaRG-box BP	DBP	HES-1	PURa,b	Unc-85	



Con el fin de determinar el grado de conservación entre los TFBS predichos, se realizó un alineamiento de los TRE del gen de *sno* de *Homo sapiens* y *Mus musculus* con los programas Blast (Blast 2 Sequences, Expect: 8.0, Filter: No) y ClustalW2.

El programa Blast detectó las siguientes tres secuencias conservadas entre los TRE de ambas especies con una identidad mayor al 85%:

- ♦ 86% de la base 13 de *Mus musculus* y *Homo sapiens* a la base 71 de *Mus musculus* y 73 de *Homo sapiens*.
- ♦ 96% de la base 212 de *Mus musculus* y 158 de *Homo sapiens* a la base 241 de *Mus musculus* y 187 de *Homo sapiens*.
- ♦ 90% de la base 270 de *Mus musculus* y 255 de *Homo sapiens* a la base 508 de *Mus musculus* y 499 de *Homo sapiens*.

El programa ClustalW2 detectó una identidad global del 79% entre las secuencias analizadas (resultado corroborado con el programa LALIGN, datos no mostrados). Este resultado se muestra a continuación (Figura 17); en azul se marcan las secuencias que Blast2Sequences detecta como semejantes.

SeqA Name	Len(nt)	SeqB Name	Len(nt)	Score
1 Mus	514	2 Homo	508	79

CLUSTAL 2.0.5 multiple sequence alignment



**Figura 17.** Alineamiento de los Elementos de Respuesta al TGFβ (TRE) de *sno* de *Homo sapiens* y *Mus musculus*. La identidad global entre estas dos secuencias (79%) se determinó por medio de ClustalW2. Blast2Sequences determinó tres subsecuencias conservadas, que se marcan en azul, con un porcentaje de identidad de 86, 96 y 90. En estas subsecuencias se encuentran los SBE y el SIE (cuadro) con sus respectivas vecindades (±10pb).

Entre los sitios de unión para factores de transcripción capaces de asociarse a las Smad que se detectaron en el TRE de las especies estudiadas (*Algen*, grado máximo de disimilitud: 15%), se seleccionaron aquellos TFBS putativos que se encuentran a menos de 10 pares de bases río arriba o río abajo de las SBEs (SBE1A, SBE1B, etcétera) y del SIE -es decir, en la vecindad de estos TFBS determinados experimentalmente-, y que están presentes en las dos especies de estudio (Tabla 9).

**Tabla 9.** Presencia (Sí) o ausencia (-) de TFBS putativos para Factores de Transcripción capaces de asociarse a las Smad y que se encuentran a menos de 10pb de los SBEs y el SIE en el TRE de *sno* de *Homo sapiens* y *Mus musculus*. En amarillo se marcan las diferencias (presencia/ausencia) entre las especies (*Homo sapiens*, **H** y *Mus musculus*, **M**). Se resaltan los TF con los que después de esta depuración se continúa trabajando.

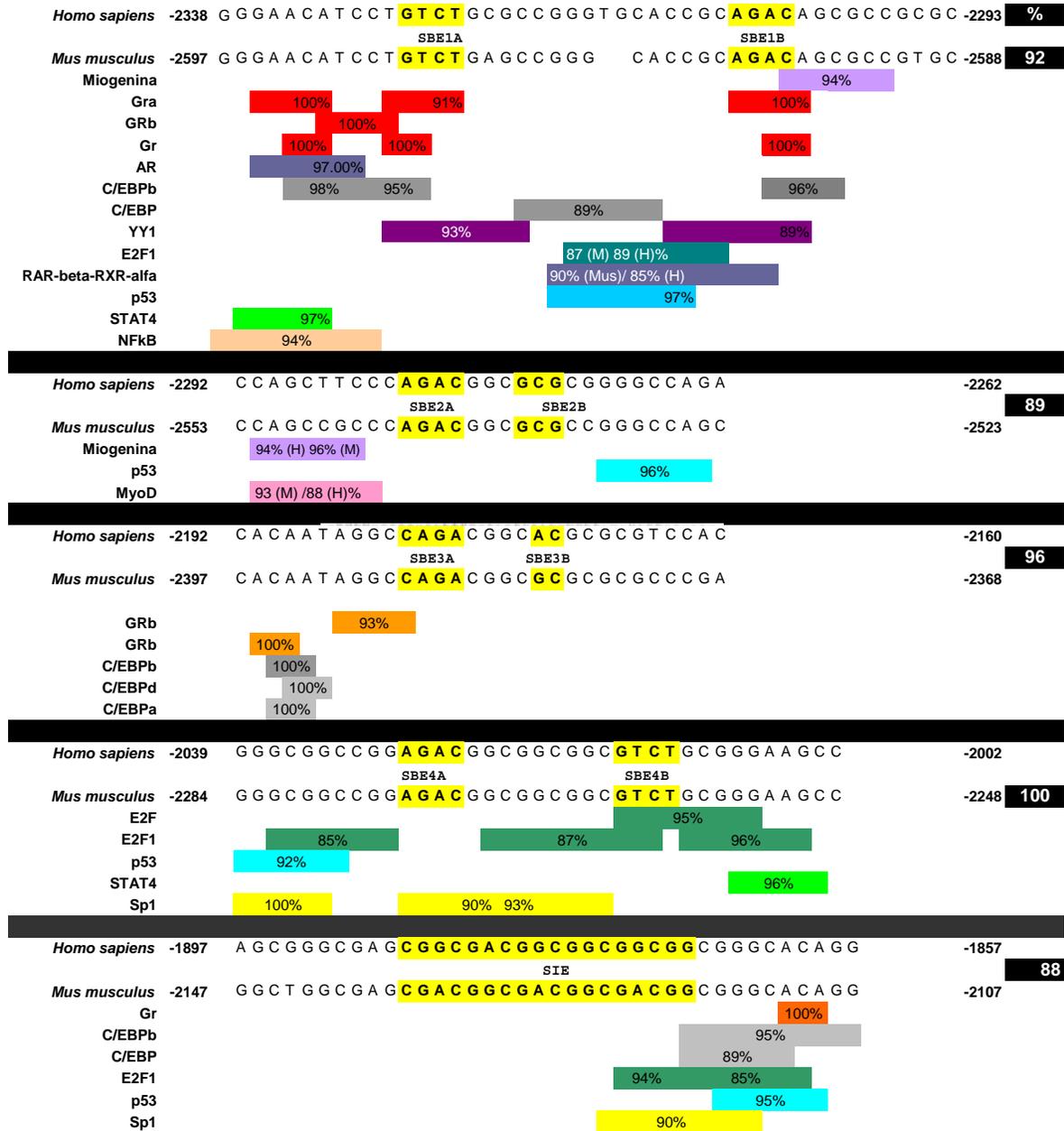
SBE	SBE1A		SBE1B		SBE2A		SBE2B		SBE3A		SBE3B		SBE4A		SBE4B		SIE	
Organismo	H	M	H	M	H	M	H	M	H	M	H	M	H	M	H	M	H	M
GR $\alpha,\beta$	Sí	Sí	Sí	Sí	-	-	-	-	Sí	Sí	-	-	-	-	Sí	■	Sí	Sí
HNF3 $\beta$	Sí	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AML1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AR	Sí	Sí	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ATF1,2,3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C/EBP $\alpha,\beta,\delta$	Sí	Sí	Sí	Sí	■	Sí	-	-	Sí	Sí	-	-	Sí	Sí	Sí	Sí	Sí	Sí
c-Fos	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
c-Jun	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E2F (-1, -4)	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	-	-	-	-	Sí	Sí	Sí	Sí	Sí	Sí
IRF-2	-	/	-	/	-	/	-	/	-	/	-	/	-	/	-	/	-	/
JunB	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
JunD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LF-A1	-	-	-	-	-	-	-	-	Sí	■	-	-	-	-	-	-	Sí	■
NF $\kappa$ B	Sí	Sí	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P53	■	Sí	■	Sí	Sí	Sí	Sí	Sí	■	Sí	Sí	Sí	Sí	Sí	-	-	Sí	Sí
RXR $\alpha,\beta$	■	Sí	Sí	Sí	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sp1	-	-	-	-	■	Sí	■	Sí	-	-	■	Sí	Sí	Sí	■	Sí	■	Sí
VDR	Sí	-	Sí	■	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YY1	Sí	Sí	Sí	Sí	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MyoD	-	-	-	-	Sí	Sí	-	-	-	-	-	-	-	-	Sí	■	■	Sí
Hoxa5	-	-	-	-	-	-	-	-	■	Sí	-	-	-	-	-	-	-	-
STAT4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Miogenina	-	-	Sí	Sí	Sí	Sí	Sí	Sí	-	-	-	-	-	-	-	-	-	-

Este procedimiento permitió descartar a los TF HNF3 $\beta$ ; AML1; ATF (1,2,3); c-fos; c-jun; IRF-2; JunB y JunD; LF-A1; VDR y Hoxa5; ya que no presentan un TFBS putativo en la vecindad ( $\pm 10$ pb) de algún SBE o SIE determinados experimentalmente y/o no se detectaron en las dos especies de estudio cerca del mismo SBE o SIE (Tabla 9).

Por lo anterior, el papel que se explorará más detalladamente en la regulación de *sno* en el contexto del TGF $\beta$  es la de los doce TF capaces de interactuar con las Smad **GR (GR, GR $\alpha$  y GR $\beta$ ), C/EBP $\alpha,\beta$  y  $\gamma$ , p53, RXR $\alpha$  y  $\beta$ , Sp1, AR, NF $\kappa$ B, E2F (-1, -4), YY1, STAT4, Miogenina y MyoD.**

En la figura 18 se muestran los TFBS putativos que *Algen* detecta en las vecindades de los SBE y el SIE tanto de *Homo sapiens* como de *Mus musculus*. En la columna derecha se muestra el porcentaje de identidad entre las secuencias de cada SBE o SIE y su vecindad ( $\pm 10$ pb) correspondientes a las dos especies de estudio. Este porcentaje se obtiene por medio de un alineamiento con el

programa Blast2Sequences. Bajo cada SBE o SIE y su vecindad se muestra la ubicación del TFBS putativo, así como el porcentaje de identidad con relación a la secuencia consenso, que se encuentra en la base de datos de TRANSFAC. Las posiciones relativas se calcularon en relación al TSS putativo.



**Figura 18.** TFBS putativos para TF capaces de interactuar con las Smad en las vecindades de los SBE y el SIE del TRE de *sno* en *Homo sapiens* y en *Mus musculus*. Se muestra el porcentaje de identidad de los TFBS putativos con relación a la secuencia consenso de la base de datos de TRANSFAC, como se despliega en el programa *Aliggen*. En negro se destaca el porcentaje de identidad entre las secuencias de las dos especies de estudio, determinado por medio de Blast2Sequences.

## **Factores de Transcripción candidatos**

En la sección anterior se presentaron a los doce TF capaces de interactuar con las Smad para los que se detectaron TFBS putativos en la vecindad de al menos un SBE o SIE del TRE del gen de *sno* tanto de *Homo sapiens* como de *Mus musculus*.

Con el objetivo de esclarecer los posibles estímulos que interfieren positiva o negativamente en la regulación de la transcripción de *sno* mediada por el TGF $\beta$ , presentes en determinados escenarios fisiopatológicos, se presentará a continuación una breve descripción de estos TF. Se enfatiza la atención en los estímulos o vías de señalización que los regula, las condiciones fisiopatológicas en las que juegan un papel protagónico, y finalmente el mecanismo de interacción con el TGF $\beta$ .

Estas breves monografías son producto de una revisión bibliográfica, efectuada con ayuda de la base de datos de PubMed.

### **♦ Receptor a Glucocorticoides (GR)**

Como mencionan Kassel y Herrlich (2007), los glucocorticoides (GC), representados en humanos por el cortisol, ejercen una amplia gama de efectos sobre el organismo. Por ejemplo, influyen sobre el metabolismo de glucosa y de lípidos, la homeostasis ósea, la respuesta al estrés, la diferenciación hematopoiética y la respuesta inmune e inflamatoria y otros efectos fisiológicos e incluso conductuales.

Los GC ejercen sus efectos al alterar la expresión génica a través del Receptor a Glucocorticoides (GR). Cuando el GR se encuentra acoplado a su ligando, se transloca del citosol al núcleo, donde activa la transcripción de genes que presentan un TFBS denominado Elemento de Respuesta a Glucocorticoides (GRE).

GR puede también modular, positiva o negativamente, directa o indirectamente, la actividad de otros factores de transcripción a través de la comunicación cruzada. En este sentido, GR interfiere ya sea con vías de

transducción que se encuentran río arriba de otros TF, o bien directamente con los TF, modulando su función de transactivación.

Existen numerosos reportes sobre eventos de comunicación cruzada directa entre el GC y el TGF $\beta$ . Estos eventos involucran TFBS para Smad3/4 y/o para GR (Li *et al.*, 2003).

En el TRE del promotor de *sno* de *Homo sapiens* y *Mus musculus* se detectaron numerosos TFBS putativos para GR, GR $\alpha$  y GR $\beta$ . En particular, a  $\pm 10$  pb de SBE1A, SBE1B, SBE3A, SBE3B y SIE se encontraron GRE putativos con un porcentaje de identidad mínimo de 91% (Figura 18). En la vecindad de SBE1A, SBE1B, SBE3A y SIE se encontraron GRE putativos con un porcentaje de identidad del 100%.

Li *et al.* (2003) demostraron que el dominio de unión a ligando (LBD) de GR interactúa físicamente con el MH2 de Smad3 en respuesta a la estimulación con TGF $\beta$ , y que esta interacción se traduce en una represión por los GC de la inducción del gen blanco del TGF $\beta$ , *PAI-1*. Estos mismos autores comentan que también existen evidencias de efectos de sinergia sobre la regulación de la expresión de genes de estos dos ligandos.

#### ♦ C/EBP $\alpha, \beta$ y $\gamma$

C/EBP (C/CAAT/Enhancer Binding Protein) es una familia de factores de transcripción integrada por seis isoformas (C/EBP $\alpha$ ,  $-\beta$ ,  $-\chi$ ,  $-\delta$ ,  $-\varepsilon$ ,  $-\phi$  y  $-\gamma$ ). Las isoformas más estudiadas y con el más amplio espectro de expresión son el C/EBP $\alpha$  y el C/EBP $\beta$ . Estos TF presentan un dominio de unión a DNA básico con cierre de leucina (Nerlov, 2007).

Nerlov (2007) comenta que C/EBP controla de manera importante la diferenciación de un amplio rango de tipos celulares y que tiene un papel fundamental en la regulación de la proliferación al interactuar con proteínas del ciclo celular. La posición estratégica de este TF, entre la proliferación y la diferenciación, hace que C/EBP sea un importante regulador de la carcinogénesis.

Diferentes miembros de la familia de C/EBP se han catalogado tanto como protooncogenes como supresores de tumores.

Aunque los C/EBP pueden controlar tanto la proliferación como la diferenciación celular, en general se catalogan como importantes antimitóticos en varios contextos celulares. Esto concuerda con que *C/EBP* es blanco de mutaciones en varios tipos de cáncer, como la Leucemia Mieloide Aguda. La decisión entre inducir proliferación o diferenciación depende de la presencia de factores de transcripción colaborantes, tales como las Smad. En cáncer de mama, los C/EBP y el TGF $\beta$  regulan conjuntamente el ciclo celular al reprimir la expresión de *c-myc* e inducir la transcripción de los inhibidores de cinasas dependientes de ciclinas *p21* y *p15*. Además, se sabe que Smad3/4 y C/EBP sinergizan la inhibición de la diferenciación de adipocitos y la transcripción de *HIF* (Feng y Derynck, 2005).

Por otro lado, Chen *et al.* (2002) determinaron que la subunidad CBF-Cb de C/EBP es capaz de asociarse con el dominio MH2 de Smad2 y de Smad3 en ausencia o presencia del TGF $\beta$ . Esta interacción se traduce en una inhibición de la inducción de la transcripción mediada por el TGF $\beta$ .

El análisis efectuado en este trabajo revela que existen TFBS putativos con más de 89% de identidad con relación a la secuencia consenso para C/EBP en las vecindades de SBE1, SBE3 y SIE de gen *sno* (Figura 18).

#### ♦ p53

El gen que con más frecuencia se encuentra mutado en el cáncer humano es el de p53; aproximadamente el 50% de los cánceres humanos han perdido a p53 o expresan una forma mutante inactiva. Así, se considera a p53 como uno de los supresores de tumores más importantes. p53 es un factor de transcripción maestro que bajo condiciones normales se encuentra funcionalmente inactivo, ya que rápidamente es degradado por la ligasa de ubiquitina MDM2. Cuando la célula se encuentra en condiciones de estrés, se bloquea la degradación mediada por MDM2. Entonces, p53 se acumula y echa a andar su capacidad de regulación de

la transcripción. En concreto, p53 regula genes característicos del ciclo celular y genes que codifican para proteínas antiapoptóticas (Efeyan y Serrano, 2007).

Un análisis *in silico* reveló que p53 se encuentra en el nodo de una compleja red de interacciones con otros factores de transcripción (Rodríguez-Caso *et al.*, 2005). Entre las interacciones que p53 puede formar, resultan interesantes para el presente trabajo aquellas que p53 forma con los efectores de la vía del TGF $\beta$ .

Wilkinson *et al.* estudiaron en 2008 la interacción entre p53 y Smad2/3 en el promotor de la alfafetoproteína (APF). El gen de la APF se mantiene reprimido durante prácticamente toda la vida postparto de un organismo. Sin embargo, su nivel de expresión aumenta durante la regeneración hepática y en numerosos tipos de cáncer. De hecho, el nivel de expresión del *APF* es usado como un marcador de hepatomas. La expresión de este marcador es regulada negativamente por la acción conjunta entre el TGF $\beta$  y p53. Presenta en su promotor un elemento de respuesta a p53 y a las Smad (SBE/p53RE) al que se une el complejo Smad4/p-Smad2/p53. Es muy interesante que en esta represión participan también SnoN y el correpresor mSin3A. De hecho, los autores incluso proponen que SnoN es *indispensable* para la represión de APF debido a que influye sobre mSin3A, probablemente manteniendo su estabilidad.

En el promotor de *sno* se detectaron TFBS putativos para p53 en las vecindades de SBE1A, SBE1B, SBE2A, SBE2B, SBE4A y SIE. La identidad de estos TFBS putativos con relación a la secuencia consenso para p53RE oscila entre el 92% y el 98%. En ningún caso, p53RE sobrelapa al SBE (Figura 18).

#### ♦ **Receptor a ácido retinoico (RXR $\alpha$ y $\beta$ )**

El ácido retinoico (RA) es una hormona importante en la regulación de la proliferación y diferenciación de una gran variedad de células humanas y murinas. RA ejerce sus efectos biológicos a través de dos receptores diferentes, RAR y

RXR. Ambos receptores se unen a su TFBS en el DNA, regulando la transcripción de los genes blanco de RA. La expresión de  $RXR\alpha$ , es regulada por el  $TGF\beta$  (Chen *et al.* 1996).

Pedaries *et al.* (2003) demostraron por medio de un promotor artificial que responde exclusivamente al  $TGF\beta$  ([CAGA]<sub>9</sub>) que el  $TGF\beta$  y el RA cooperan positivamente en la expresión de genes blanco por medio de la interacción del dominio MH2 de Smad3 con RAR-RXR.

#### ♦ **Proteína específica (Sp1)**

Sp1 es un TF con dedos de Zinc, involucrado en una gran variedad de procesos fisiológicos. Por ejemplo, Sp1 regula la progresión del ciclo celular, la activación hormonal, la apoptosis, la angiogénesis, entre otros. Se han encontrado una gran variedad de señales de naturaleza divergente que inducen la fosforilación (regulación) de Sp1. Entre estas señales se encuentran infecciones virales, factores de crecimiento, algunos fármacos, citocinas, estrés mecánico, etcétera (Chu y Ferro, 2005).

Sp1 induce la transcripción de los genes blanco del  $TGF\beta$   $p15^{Ink4b}$ ,  $p21^{Cip1}$ , *colágena  $\alpha 2(I)$* , *integrina  $\beta 5$* , *smad7* y *PAI-1* (Traylor *et al.* 2007). Sp1 se asocia con el heterotrímero Smad2/3/4 a través de un dominio rico en glutamina (Feng y Derynck, 2005).

Traylor y colaboradores detectaron en 2007 un elemento de respuesta al  $TGF\beta$  en el promotor de la Hemooxigenasa Humana (*HO-1*) que contiene un SBE putativo a 13 pares de bases de un TFBS para Sp1. El  $TGF\beta$  sinergiza la inducción de la expresión de *HO-1* mediada por Sp1. A través de las Smad, el  $TGF\beta$  induce la expresión de *HO-1*. Este proceso sucede exclusivamente cuando Sp1 se encuentra anclado a su sitio de unión, como lo demuestra la falta de inducción de *HO-1* en presencia de mitramicina A, un antibiótico que une específicamente secuencias ricas en CpGs impidiendo así la unión de Sp1 a su TFBS; y el uso de un siRNA dirigido contra Sp1. Además, los autores demostraron, por mutagénesis dirigida, que para la expresión de *HO-1* mediada

por el TGF $\beta$  y la Sp1 es indispensable la presencia del SBE y de un sitio de unión a Sp1.

En el TRE de *sno* de *Homo sapiens* y *Mus musculus* se detectaron TFBS putativos en las vecindades del SBE3 y del SIE. El porcentaje de identidad entre estos TFBS putativos y las secuencias consenso se encuentra entre el 90% y el 100% (Figura 18).

#### ♦ **Receptor a Andrógenos (AR)**

Los andrógenos controlan el desarrollo y la función correcta del sistema reproductor masculino, incluyendo la próstata y el epidídimo, y de varios sistemas somáticos, tales como los músculos, la piel, los folículos pilosos y el cerebro. Además, la señalización por andrógenos media la progresión del cáncer de próstata (Kang *et al.* 2002).

El Receptor a Andrógenos (AR) es un regulador transcripcional dependiente de andrógeno. Después de la unión con su ligando, por ejemplo, 5 $\alpha$ -dihidrotestosterona (DHT), el AR activado reconoce secuencias palindrómicas en el promotor blanco, denominadas elementos de respuesta a andrógenos (ARE). En este TFBS, AR interactúa con proteínas asociadas a AR e induce la expresión de genes blanco de AR (Kang *et al.*, 2001).

Es ampliamente conocido que el AR es un efector crítico del desarrollo y la progresión del cáncer prostático. De hecho, se ha sugerido que el cáncer de próstata depende de la señalización de los andrógenos (Balk y Knudsen, 2007).

Se ha propuesto la participación del TGF $\beta$  en el desarrollo del cáncer prostático. De hecho, los niveles del TGF $\beta$  plasmáticos se encuentran significativamente elevados en pacientes con metástasis y se correlacionan con un aumento en los niveles séricos del Antígeno Específico de Próstata (PSA) (Kang *et al.*, 2001).

La correlación entre el TGF $\beta$  y los andrógenos no sólo es funcional. Kang *et al.* (2002) demostraron que Smad3 y Smad4 son capaces de interactuar con AR a través de su dominio de unión a DNA y el dominio de unión al ligando. Esta interacción, comentan, se traduce en una regulación diferencial de la expresión de genes blanco tanto del TGF $\beta$  como de la DHT. Que la comunicación cruzada entre el TGF $\beta$  y la DHT genere una inhibición o una activación de la transcripción depende del contexto celular (es decir, de la presencia de determinados correguladores) y de la secuencia.

Así, el AR y el TGF $\beta$  sinergizan la inducción de la expresión del virus tumoral mamario murino, a través de la unión del dominio MH2 de Smad3 al AR y del AR a un ARE. Además, el TGF $\beta$  coopera con el AR en la inducción de la expresión del Antígeno Específico de Próstata (PSA) (Kang *et al.*, 2001). Por otro lado, Hayes *et al.* establecieron en 2001 que el TGF $\beta$  antagoniza la inducción de la expresión de genes blanco del AR por medio de la interacción física entre el MH2 de Smad3 y el N-ter del AR. Esta interacción previene la unión del AR al ARE. Entonces, los efectores del TGF $\beta$  y el AR actúan en sinergia o antagonismo en la regulación de genes blanco del uno o del otro, dependiendo del contexto celular y del TFBS en el promotor del gen blanco.

En el TRE del gen *sno* se detectó un ARE putativo (97% de similitud) 2pb río arriba de SBE1A (Figura 18).

#### ♦ **NF $\kappa$ B**

El factor nuclear  $\kappa$ B (NF $\kappa$ B) es un factor de transcripción que responde a estímulos del sistema inmune innato. NF $\kappa$ B es activado, por ejemplo, por la unión del ácido lisofosfatídico (LPS) al receptor tipo *Toll* (TLR4), o en respuesta a citocinas proinflamatorias como el factor de necrosis tumoral (TNF $\alpha$ ) y la interleucina  $\beta$  (IL $\beta$ ) (Akira y Takeda, 2004).

Smad3 y NF $\kappa$ B interactúan directamente; además, tanto NF $\kappa$ B como las C-Smad son capaces de asociarse con CBP/p300, lo que puede llevar a una cooperación en el reclutamiento de este cofactor (Feng y Derynck, 2005).

Kon *et al.* describieron en 1999 un efecto aditivo entre el TGF $\beta$  y las citocinas proinflamatorias TNF $\alpha$  e IL $\beta$  sobre la inducción de la transcripción del gen de la colagenasa humana tipo VII (*col7*). La sinergia entre estas citocinas sucede por una interacción funcional, pero no física, entre Smads y NF $\kappa$ B. Los autores comentan que los efectos aditivos del TNF $\alpha$  y el TGF $\beta$  dependen de la presencia de un SBE y un sitio de unión a NF $\kappa$ B en el promotor de *col7*. Sin embargo, estos TFBS están separados por 200pb, lo que hace poco probable su interacción física.

López-Rovira *et al.* (2000) describieron un efecto sinérgico entre los efectores del TGF $\beta$  y de NF $\kappa$ B sobre la inducción de la transcripción de *junB*. En este trabajo, los autores demuestran que Smad3 interactúa directamente con el componente de NF $\kappa$ B p52 *in vivo* en respuesta a la unión del TGF $\beta$  a su receptor. Además, demuestran que en el promotor de *junB* existe un TRE que contiene un SBE y un elemento de respuesta a NF $\kappa$ B (NF $\kappa$ B RE). Por medio de mutaciones puntuales, establecen que el NF $\kappa$ B RE, pero no el SBE, es indispensable para la inducción de la transcripción de *junB* por Smad3 y NF $\kappa$ B.

Kon y colaboradores (1999) mencionan que, a diferencia del gen *col7*, la transcripción del gen de la colagenasa tipo I es inducida por TGF $\beta$  y reprimida por TNF $\alpha$ . Es decir, que los efectores de estas dos citocinas no siempre tienen un efecto aditivo sobre la regulación de un gen, sino que en muchos casos, actúan de manera antagónica.

El análisis bioinformático realizado en el presente trabajo, muestra que existe un TFBS putativo para NF $\kappa$ B en la vecindad río arriba de SBE1A. Esta secuencia comparte un 97% de identidad con relación a la secuencia consenso para NF $\kappa$ B (Figura 18).

#### ♦ **STAT4**

La Interleucina 12 (IL12) es una citocina proinflamatoria que activa a STAT4 (*Signal Transducer and Activator of Transcription*), por medio de la activación de cinasas de la familia de Janus (JAK). IL12 es interferida negativamente por TGF $\beta$  (Bright y Sriram, 1998).

Pardoux *et al.* (1999) demostraron que el TGF $\beta$  disminuye la capacidad de unión al DNA de STAT4 en linfocitos humanos y murinos.

En la figura 18 se muestra que se detectaron TFBS putativos para STAT4 en las vecindades de SBE1A y de SBE4B con un 97% y 96% de identidad con relación a la secuencia consenso de unión a STAT4, respectivamente.

#### ♦ **E2F**

E2F es una familia de factores de transcripción tipo bHLH que interfieren sobre la progresión del ciclo celular. Hasta la fecha se han detectado siete miembros de la familia de E2F. En particular, E2F1 funge como activador, mientras que a E2F4 se le ha adjudicado un papel de represor transcripcional (Tang-Hu, 2008).

E2F ejerce sus acciones sobre la progresión del ciclo celular al formar parte de una red de señalización que incluye Ciclinas (cic), Cinasas dependientes de ciclinas (Cdk), inhibidores de cinasas dependientes de ciclinas (Ink) y al Retinoblastoma (Rb). En células quiescentes, E2F forma complejos con Rb. Estos complejos inhiben la transcripción de genes necesarios para la replicación del DNA. La estimulación mitogénica induce la fosforilación de Rb por CiclinaD-Cdk4/6. pRb se disocia de E2F. Entonces, E2F funge como un activador transcripcional de factores relacionados a la progresión del ciclo celular y la replicación del DNA. La importancia biológica de esta vía se evidencia por el hecho de que la mayoría de los tumores humanos exhiben alteraciones que

alteran la capacidad de pRb de inhibir E2F, lo que conlleva a una activación permanente del último (Ivanova *et al.*, 2005).

La generación de modelos murinos mutantes para E2F ha hecho evidente la relación entre E2F y el TGF $\beta$ . La inactivación de E2F1, por ejemplo, resulta en una alteración de la proliferación, la tasa de adhesión, la migración y la quimiotaxis en respuesta al TGF $\beta$  en queratinocitos (Ivanova *et al.*, 2005). Además de la interacción funcional entre TGF $\beta$  y E2F, Lacerete y colaboradores descubrieron en 2008 que Smad3 (y Smad2/3) y E2F actúan en sinergia en la inhibición del gen del componente protéico de la Telomerasa Humana (*hTERT*). De hecho, el TGF $\beta$  depende de la presencia de E2F1, como lo muestra la falta de inhibición de *hTERT* por TGF $\beta$  en un modelo de *knockdown* para E2F. Proponen que este efecto regulatorio depende además de MAPK, lo que demuestran por medio del uso de inhibidores de Erk1/2 y p38. Además, la inhibición de la expresión de *hTERT* mediada por el TGF $\beta$  depende de HDAC, como lo demuestra el uso de Tricostatina (TSA), un inhibidor de esta enzima. La sinergia entre E2F1 y Smad2/3 en la inhibición de *hTERT* depende de la presencia de TFBS para E2F1 en una región de 250pb. Aunque un análisis de la secuencia regulada por el TGF $\beta$  y E2F en *hTERT* revela que contiene un SBE, los autores no caracterizaron su papel en la cooperación E2F1-TGF $\beta$ .

Por otro lado, Chen *et al.* demostraron en 2002 que E2F4/5 y el TGF $\beta$  reprimen conjuntamente la expresión del protooncogen *c-myc*. El promotor de *c-myc* presenta un elemento inhibitorio por el TGF $\beta$  (TIE) que contiene un sitio de unión a E2F flanqueado por un SBE imperfecto (GGCT) pero capaz de unir Smad3/4, así como un sitio de unión a E2F. Mostraron por ensayos de *pull down* que E2F4/5 y Smad3 interactúan directamente por el dominio de unión a DNA y el dominio MH2, respectivamente.

En el promotor del gen *sno* se predicen TFBS para E2F en las vecindades de SBE1, de SBE4 y de SIE. El grado de identidad con relación a una secuencia consenso de unión a E2F oscila entre el 85% y el 96% (Figura 18).

## ♦ YY1

Ying Yang 1 (YY1) es un factor de transcripción con dedos de Zinc expresado ubicuamente y cuyos niveles protéicos se modifican durante la diferenciación de mioblastos y durante el envejecimiento (Yao *et al.*, 1998). YY1 juega un papel importante en la embriogénesis, la diferenciación, la replicación y la proliferación (Gordon *et al.*, 2006). YY1 activa y reprime la transcripción al unirse a su TFBS 5´-[C/A]CAT-3´ por medio de su dominio de dedos de Zinc (Kurisaki *et al.* 2003). La naturaleza multifuncional de YY1 se debe en parte a la gran diversidad de factores celulares que interactúan físicamente con YY1 (Yao *et al.*, 1998).

Kurisaki y colaboradores determinaron en 2003 que YY1 reprime la expresión de genes blanco del TGF $\beta$  tales como *PAI1*. Los autores establecieron que la inhibición de las Smad por YY1 se debe a la interacción física entre YY1 y el dominio MH1 de Smad4 y en menor medida con Smad3. Esta interacción impide la unión de Smad3,4 al SBE en el promotor del gen blanco del TGF $\beta$ . Concluyen que la inhibición del TGF $\beta$  por YY1 depende, al menos en parte, de la presencia de múltiples SBEs en el promotor del gen blanco del TGF $\beta$ .

La regulación transcripcional mediada por YY1 se lleva a cabo por múltiples mecanismos. Gordon *et al.* (2006) mencionan que de manera frecuente, YY1 compite con factores de transcripción que presenten un TFBS sobrelapado con el TFBS de YY1.

El promotor de *sno* de las especies estudiadas presenta TFBS putativos en las vecindades de SBE1A y SBE1B. Estas secuencias son 93% y 89% idénticas a la secuencia consenso, respectivamente (Figura 18).

#### ♦ MyoD/Miogenina

MyoD y Miogenina pertenecen a la familia de factores de transcripción bHLH. Protagonizan la miogénesis al regular la expresión de genes característicos de músculo, como el gen de la Cadena Pesada de la Miosina (MHC). Esta regulación sucede a través de la heterodimerización de MyoD o Miogenina con un integrante de la familia de las proteínas E, E12 o E47 seguida de la unión de este heterodímero a un TFBS en el promotor del gen blanco denominada caja E (Liu *et al.*, 2001).

Liu y colaboradores encontraron en 2001 los mecanismos moleculares a través de los que sucede la inhibición de la miogénesis por el TGF $\beta$ . Encontraron que Smad3/4, pero no Smad2, inhiben el aumento en la expresión de genes característicos de músculo (como MHC) mediados por los genes maestros de la diferenciación MyoD y Miogenina. Esta inhibición, proponen, se debe a que Smad3 se une al dominio bHLH de MyoD y Miogenina, interfiriendo así con la formación del complejo heterodimérico MyoD/Miogenina-E12/E47. Al ocupar el dominio bHLH e impedir la formación del complejo heterotrimérico, Smad3 impide la unión de MyoD/Miogenina a su TFBS en el promotor blanco, la caja E, y, con ello, la expresión del gen.

Se detectaron TFBS putativos para MyoD y Miogenina en el TRE de *sno* de *Homo sapiens* y *Mus musculus*. Los TFBS putativos para Miogenina se encuentran en las vecindades de SBE1B (94% de identidad con relación a la secuencia consenso) y SBE2A (94% en *Homo sapiens*, 96% en *Mus musculus*). El sitio de unión a MyoD putativo traslapa al segundo TFBS putativo para Miogenina. El porcentaje de identidad de esta secuencia con relación a la secuencia consenso de unión a MyoD es de 93% en el caso de *Mus musculus*, y 88% en el caso de *Homo sapiens*.

# Discusión

En el presente trabajo se realizó un análisis bioinformático del promotor de *sno* de *Homo sapiens* y *Mus musculus*. Se enfatizó en la búsqueda de TFBS putativos en la región responsiva al TGF $\beta$ . Además, se predijo un TSS, se realizaron comparaciones entre las secuencias analizadas correspondientes a las dos especies de estudio, y se efectuó una búsqueda bibliográfica en torno a los resultados de la predicción de TFBS.

En la presente sección se discute cada punto mencionado. Posteriormente, se propone una estrategia de validación experimental de las predicciones efectuadas y, finalmente, se analiza brevemente el alcance de la investigación *in silico*, en particular de las predicciones.

## Obtención de la secuencia promotora del gen *sno*

No existe hasta la fecha en las bases de datos una secuencia catalogada como promotor de *sno*. En particular, no se ha reportado una secuencia promotora de *sno* de ni de *Mus musculus* ni de *Homo sapiens*. Por lo tanto, el primer paso en este trabajo fue obtener dicha secuencia.

Para ello, se consideró como secuencia promotora a las 5000 bases río arriba del sitio de inicio de la transcripción del gen de estudio. Este sitio se determinó en el caso del gen de *sno* de *Mus musculus* alineando el mRNA de SnoN2 U14655 con el fragmento del cromosoma 3 AC117590; y, en el caso de *Homo sapiens*, rastreando el elemento de respuesta a HGF del gen de *sno* en todo el genoma humano y sumando la distancia en pares de bases que restaban de este alineamiento al inicio del gen de *sno* (SkiL) determinado por el programa utilizado (Blast the Human Genome). Es decir, se sumaron 2195pb correspondientes a la distancia entre el sitio de inicio del alineamiento de HGFRE con el genoma humano y la región codificante de *sno* para obtener el extremo 3' del promotor. A este sitio se restaron 5kb para obtener la cota 5' de la región regulatoria. Se validaron estas secuencias como promotores de *sno* de *Homo sapiens* y *Mus musculus*, respectivamente, al encontrar satisfactoriamente los

SBEs, el SIE y el HGFRE determinados de manera experimental en las secuencias. Un esquema de este procedimiento se encuentra en la figura 10, y la detección satisfactoria de los TFBS determinados experimentalmente se muestra en la tabla 3.

La elección del tamaño del promotor putativo -5kb- fue arbitraria. Existen numerosos reportes que proponen diferentes longitudes de partida de promotores putativos para su inspección *in silico*. Por ejemplo, según Smale (2005), un promotor eucarionte típico contiene el TSS y entre 100 y 400 pb alrededor de éste. Por otro lado, Zhang (2007) considera que el promotor se encuentra entre ~250 a 1000 pb río arriba del promotor núcleo, y Rico *et al.* (2007) sugieren realizar el análisis considerando 10kb río arriba y 1kb río abajo del TSS.

De cualquier manera, las 5kb elegidas definitivamente contienen la secuencia regulatoria de *sno*, como lo muestra la validación de la secuencia (Figura 10), y no interfieren con el gen que se encuentra río arriba de *sno*. En el caso de *Homo sapiens*, el alineamiento a nivel genómico del HGFRE mostró que PKC $\phi$  se encuentra a **54811 + 200<sub>HGF-RE</sub> + 2195<sub>Dist.</sub>** al TSS de *sno* = **57206pb** río arriba del TSS de *sno*. Análogamente, el TSS de *sno* de *Mus musculus* se encuentra a 46137 pares de bases de la PKC $\phi$ . Este dato se obtiene alineando la secuencia regulatoria con el genoma murino. El programa PubMed detecta que el alineamiento se encuentra flanqueado por el gen de la PKC $\phi$  a 44376 pares de bases río arriba (5') y por SkiLb a 1761 pares de bases río abajo (3').

En conclusión, las 5kb consideradas en el presente trabajo como promotor putativo de *sno* son suficientes, porque presentan al TRE y al HGFRE, y no demasiado abundantes, ya que no transgreden la región genética que se encuentra río arriba de *sno*.

## Determinación del TSS

Un paso clave en el análisis *in silico* de un promotor es la determinación del TSS putativo. En el presente trabajo, se emplearon tres estrategias diferentes para una predicción acertada de este TSS:

1. Alineamiento de mRNA de SnoN.
2. Detección de TSS con DBTSS.
3. Alineamiento “manual” de cDNAs con las secuencias promotoras.

Las estrategias 1 y 3 produjeron resultados idénticos; tanto el comienzo de la región 5'UTR del mRNA de SnoN (U14655, en *Mus musculus*, U70730 en *Homo sapiens*) como los cDNA de las diferentes isoformas de *sno* (SnoN y SnoN2 en *Mus musculus*, adicionalmente Sno1 y SnoA en *Homo sapiens*; para detalles consultar la tabla 4) se alinean a 2280 pb de 2035 del SBE4 en *Mus musculus* y *Homo sapiens*, respectivamente (Figura 12).

El programa DBTSS detectó un gran número de TSS putativos (Figura 14). Uno de estos sitios concuerda con aquel que se detectó con las otras dos estrategias.

Por lo tanto, en el presente trabajo se consideró como TSS a la base que se encuentra a 2280 pb en *Mus musculus* y de 2035 en *Homo sapiens* del SBE4, posición que concuerda con los resultados obtenidos con las tres aproximaciones mencionadas.

Sin embargo, no se descarta la presencia de múltiples TSS, como predice DBTSS. Análisis *in silico* de los extremos 5' de cDNAs revelaron que en muchos genes humanos y murinos, los TSS se encuentran distribuidos a lo largo de una región relativamente amplia y no corresponden a un sitio discreto, como se sugería anteriormente (Tsuritani *et al.*, 2007). Experimentos *in vitro* han mostrado que este fenómeno ocurre especialmente en promotores que carecen de la caja TATA. Esto sugiere que el TSS es una región, en vez de una posición estática, y que, de hecho, la distribución de los TSS debe reflejar la naturaleza dinámica de los eventos de inicio de la transcripción *in vivo*. No obstante esta posibilidad,

debido a que el presente trabajo se enfoca en la predicción de TFBS, más que en la determinación del número preciso de TSS, no se abunda más en la exploración de los posibles TSS alternativos en *sno* de *Homo sapiens* y *Mus musculus*. Así, se considera al TSS que se encuentra a una distancia de 2344 pb en el promotor de *Mus musculus* y a 2740 pb en *Homo sapiens*.

Finalmente, como una perspectiva del presente trabajo, se sugiere realizar una corroboración experimental del TSS determinado. Hay numerosas técnicas para corroborar un TSS, tales como la expansión de *primer* o la determinación de TSS por la endonucleasa S1 (Clark, 2005).

### **Búsqueda de cajas regulatorias en la zona núcleo del promotor del gen de *sno***

Para determinar la presencia o ausencia de cajas regulatorias en el promotor del gen de *sno*, particularmente de la caja TATA, se consideró primero el resultado desplegado por DBTSS. El programa DBTSS no sólo detecta TSS, sino que también exhibe la presencia de TFBS putativos. Como se observa en la figura 14, DBTSS no encontró ningún TFBS. Por otro lado, DBTSS apunta sobre una zona rica en CpGs en la región núcleo del promotor (Figura 15). La presencia de una isla rica en CpGs está correlacionada estadísticamente con la ausencia de cajas regulatorias tales como la caja TATA (Bajic *et al.*, 2004). Así, los resultados obtenidos con DBTSS sugieren que el promotor de *sno* de *Homo sapiens* y *Mus musculus* no presenta cajas regulatorias, particularmente caja TATA.

Para corroborar este resultado, se analizó la zona núcleo del promotor de *sno* de las especies estudiadas con el programa *Alggen* (25% de disimilitud máxima). Como zona núcleo se consideraron 100pb río arriba y 100pb río abajo del TSS, tal como lo plantea Zhang (2007). Se enfatizó el análisis entre la base 25 y la 30 río arriba del TSS, ya que, como comenta Smale (2005), es ahí donde, preferentemente, se encuentran las cajas TATA.

La tabla 5 muestra los resultados obtenidos con *Alggen*. Es importante resaltar que se detectaron TFBS putativos para TBP y para TFIID entre 25 y 30 pb con un alto grado de identidad con relación a la caja TATA consenso en *Homo*

*sapiens*, pero no en *Mus musculus*. Aunque esta falta de conservación entre los TFBS putativos se puede deber a una divergencia interespecífica, es probable que ninguna de las dos especies de estudio presente caja TATA, ya que, como mencionan Levy y Hannenhalli (2002), una predicción de TFBS es mucho más acertada cuando los TFBS putativos se encuentran conservados entre especies.

Así, en el presente análisis se concluye provisionalmente que los promotores de *sno* de *Homo sapiens* y *Mus musculus* no presentan cajas regulatorias, por lo que *sno* forma parte del 50% de los genes que carecen de caja TATA (Zhang, 2007). Esta afirmación se debe comprobar experimentalmente, por ejemplo por medio de un ensayo de retardo (EMSA) en el que se trate a la zona núcleo del promotor de *sno* de las especies de estudio con la TBP.

### **Predicción de TFBS**

El tema central del presente trabajo es la predicción de TFBS. A partir de los TFBS putativos en el TRE de *sno* de *Homo sapiens* y *Mus musculus*, se infieren los TF que pueden estar involucrados en la regulación de *sno* mediada por el TGF $\beta$ . Estos TF son, en algunos casos, característicos de determinadas condiciones fisiológicas y/o patológicas, por lo que su predicción permite inferir el escenario en el que la posible interacción TFX-Smad se lleva a cabo.

Por la importancia de esta sección, en primer lugar fue necesario validar el programa de predicción de TFBS. Una vez que se encontró el programa que satisface los requerimientos, se analizaron las secuencias regulatorias –es decir, el TRE y el HGFRE- de las especies de estudio. El número de TFBS putativos detectados en las secuencias regulatorias (n=70) se restringió por cuatro vías. En primer lugar, se consideró en un segundo análisis sólo el TRE, es decir, los SBEs y el SIE  $\pm 10$ pb, equivalentes a las vecindades. En este segundo análisis, el porcentaje de disimilitud máximo de *Algggen* se disminuyó de 25% a 15%. Finalmente, se consideraron sólo aquellos TFBS putativos capaces de interactuar con las Smad (revisado por Feng y Derynck, 2005) y que están conservados en *Homo sapiens* y *Mus musculus*. Después de estos filtros, el número de TF para los que se detectó al menos un TFBS putativo en las vecindades de los SBE y/o el

SIE se redujo de 70 a doce: GR, AR, RXR, C/EBP, E2F, p53, MyoD, Miogenina, YY1, Sp1, NFκB y STAT4.

Finalmente, se rastreó el expediente de estos TF, poniendo particular énfasis en su relación con el TGFβ. Con esta información se elaboraron modelos de regulación de *sno* en el contexto del TGFβ, y se propuso un diseño experimental para la validación del modelo.

El programa Genomatix no fue capaz de detectar los TFBS determinados experimentalmente; por su parte, *Alggen*, sí (Figura 16, Tabla 5). Esta diferencia entre los programas se debe a la posibilidad de modificar el porcentaje de disimilitud máxima en *Alggen*, lo que no es posible en el programa Genomatix (15% de *fault*). Por lo tanto, la predicción de TFBS se realizó con *Alggen*.

En primer lugar, se analizó la secuencia regulatoria –el HGFRE y el TRE, que se muestran en la tabla 3 y en la figura 11- con *Alggen*. Se consideró un porcentaje de disimilitud máximo del 25%, ya que sólo con este porcentaje se detectaron todos los TFBS determinados experimentalmente de forma satisfactoria. Se detectaron así TFBS putativos para 120 TF; de éstos TF, 25 son capaces de interactuar con las Smad (Tabla 7).

Con el objetivo de disminuir el número de falsos positivos, y para restringir el análisis de TFBS putativos, se analizó a continuación exclusivamente el TRE, y se redujo el porcentaje de disimilitud máximo al 15%. Se detectaron ahora TFBS putativos para 70 TF, de los cuales 19 son capaces de interactuar con las Smad (Tabla 6). La posición de los sitios de unión para estos 19 TF se presenta en la figura 16.

La identidad entre los TRE en el gen de *sno* de las dos especies es mayor (79%, Figura 17) que entre las secuencias regulatorias (71%, Figura 11). Esto sugiere que la probabilidad de que en el TRE se encuentren falsos positivos es menor a esta probabilidad en las secuencias regulatorias, ya que, como mencionan Loots y Ovcharenko (2004), la conservación entre secuencias no codificantes es indicio de que estas secuencias presentan una función regulatoria.

Como siguiente paso, se consideraron exclusivamente aquellos TF para los que se detectaron TFBS putativos en ambas especies y que se encuentran a una

distancia no mayor a 10pb de algún SBE y/o SIE (Tabla 9). Estos criterios obedecen, por un lado, a la importancia de la conservación entre TFBS putativos en regiones no codificantes mencionada anteriormente; y por el otro, a que si los probables sitios de unión para los TF capaces de interactuar con las Smad se encuentran cerca ( $\pm 10$ pb) de los SBE y/o el SIE, la posibilidad de que estos TF interactúen físicamente con los efectores del TGF $\beta$  aumenta considerablemente (Massagué y Wotton, 2000).

Después de esta depuración, el número de TF capaces de interactuar con las Smad, presentes en las vecindades de ambas especies de estudio, es de doce; se trata de los siguientes TF candidatos: Receptor a Glucocorticoides (GR), Receptor a Andrógenos (AR), Receptor a Vitamina D (RXR), Miogenina, Factor Maestro de la Diferenciación Muscular (MyoD), Ying Yang 1 (YY1), Proteína Específica 1 (Sp1), E2F, p53, Factor Nuclear  $\kappa$ B (NF $\kappa$ B), Transductor de Señales y Regulador de la Transcripción 4 (STAT4) y Proteína de Unión al Acrecentador CAAT (C/EBP). El grado de conservación entre las vecindades es de 92%, 89%, 96%, 100% y 88% para SBE1, SBE2, SBE3, SBE4 y SIE, respectivamente (Figura 18).

La selección de los TFBS putativos consideró así varios aspectos. El aspecto evolutivo, en el sentido de que un alto grado de conservación entre dos secuencias no codificantes es indicio de una huella filogenética (Levy y Hannenhali, 2002); el aspecto funcional, representado por la distancia entre un SBE y/o SIE y el TFBS para el TF capaz de interactuar con las Smad influye sobre la posible geometría de la interacción entre este TF y los efectores de la vía del TGF $\beta$  (Massagué y Wotton, 2000; y Hua *et al.* 1999) y, finalmente, el aspecto fenológico; el programa *Alggen* detecta patrones en las secuencias por medio de un algoritmo que le asigna a cada subsecuencia un valor o puntaje determinado (Long *et al.*, 2004).

Estos tres criterios permiten reducir el número de falsos positivos de manera considerable, lo que hace que las afirmaciones realizadas en el presente análisis sean plausibles.

Como apuntan Massagué y Wotton en 2000, la regulación genética por la vía del TGF $\beta$  en una determinada célula depende en gran medida de los factores que se encuentren en el núcleo celular al momento de estimulación con el TGF $\beta$ . Para la unión de estos factores al promotor y su asociación con las Smad se requieren TFBS.

En el presente análisis se parte de TFBS putativos para inferir los TF que pueden colaborar con las Smad en determinados escenarios. Combinaciones específicas entre TF permiten una especificidad espacio-temporal de la expresión por medio de un reducido número de factores de transcripción bajo una gran variedad de condiciones fisiológicas y patológicas (Chu y Ferro, 2005).

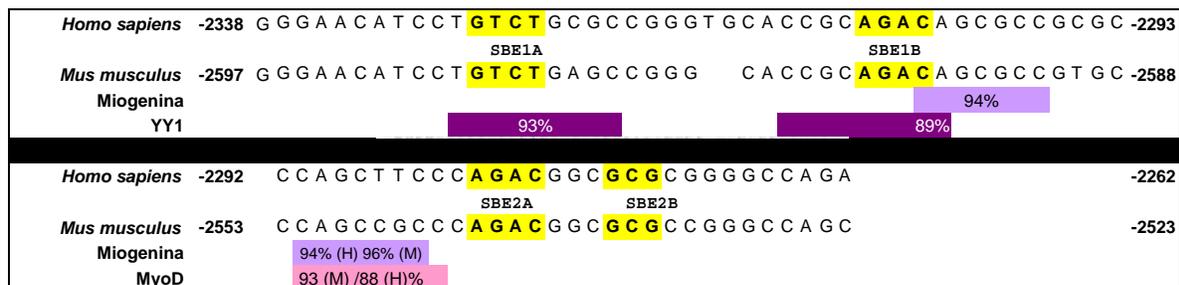
A continuación se presentan los modelos de regulación de *sno* en el contexto del TGF $\beta$  que se dedujeron a partir de la búsqueda de TFBS putativos y el análisis bibliográfico de los TF candidatos resultantes de la búsqueda.

El primer TF que se analizó fue GR. Como se muestra en la figura 18, se detectaron TFBS putativos para GR con una identidad del 100% (94% en la vecindad de SBE3A) en las vecindades de SBE1A, SBE1B, SBE3A y SIE.

No obstante, Liu *et al.* (2003) comentan que los GCs y el TGF $\beta$  cooperan o se reprimen mutuamente a nivel de sus efectores, GR y Smads. Que la relación entre GR y las Smads sea sinérgica o antagónica, comentan, depende del contexto celular, es decir, de la presencia de cofactores. Por otro lado, tanto los GC como el TGF $\beta$  son dos ligandos que protagonizan una amplia gama de procesos fisiológicos diferentes (Kassel y Herrlich, 2007; Moustakas *et al.*, 2002).

La ambigüedad del efecto de los GC sobre el TGF $\beta$ , aunada al pleiotropismo de estas dos moléculas, dificulta conjeturar sobre un modelo de regulación de *sno* por GC en el contexto del TGF $\beta$ . Así, aunque se detectó una gran cantidad de GRE putativos con una identidad muy alta con relación al GRE prototipo en las vecindades de los SBE y el SIE en *sno* de *Homo sapiens* y *Mus musculus* (Figura 18), el presente trabajo no se aventura a predecir un escenario fisiopatológico en el que *sno* se encuentre sub- o sobreexpresado por la acción conjunta del TGF $\beta$  y los GC.

El siguiente grupo de TF a analizar es MyoD, Miogenina y YY1. Estos TF están profundamente involucrados con el proceso de la diferenciación muscular. Se detectaron TFBS putativos para éstos en las vecindades de SBE1 y SBE2 (Figura 19).



**Figura 19.** TFBS putativos para los TF típicos de la miogénesis en las vecindades de SBE1B y SBE2A.

Smad3, pero no Smad2, es capaz de mediar la inhibición de la miogénesis inducida por el TGF $\beta$  (Liu *et al.* 2001). Aunque Smad3 y Smad2 son muy similares, desde el punto de vista funcional difieren en la capacidad de unión al DNA de Smad3, pero no (o con mucho menor medida) de Smad2. Físicamente hablando, esta incapacidad parcial de Smad2 de unirse al DNA se debe a una inserción en el exón 2 en la región que codifica para una horquilla  $\beta$ , la estructura responsable de unión a DNA del dominio MH1 (Massagué y Wotton, 2002, Figuras 2 y 3).

Liu y colaboradores detectaron en su trabajo de 2001 que una mutante de Smad3 que carece de su dominio MH2 es capaz de reprimir la expresión de MHC, mientras que la mutante de Smad3 carente de MH1 y la de región linker reprime la expresión de MHC mediada por TGF $\beta$  en un porcentaje significativamente menor. Aunque estos autores no analizaron si la represión de la miogénesis mediada por Smad3 depende de la presencia de un SBE en la vecindad de la caja E, los resultados obtenidos con las mutantes de MH1, así, como la incapacidad de Smad2 de ejercer este efecto inhibitorio, sugieren que la cercanía entre un SBE y un TFBS para MyoD/Miogenina pueda jugar un papel determinante en la inhibición de la miogénesis mediada por TGF $\beta$ .

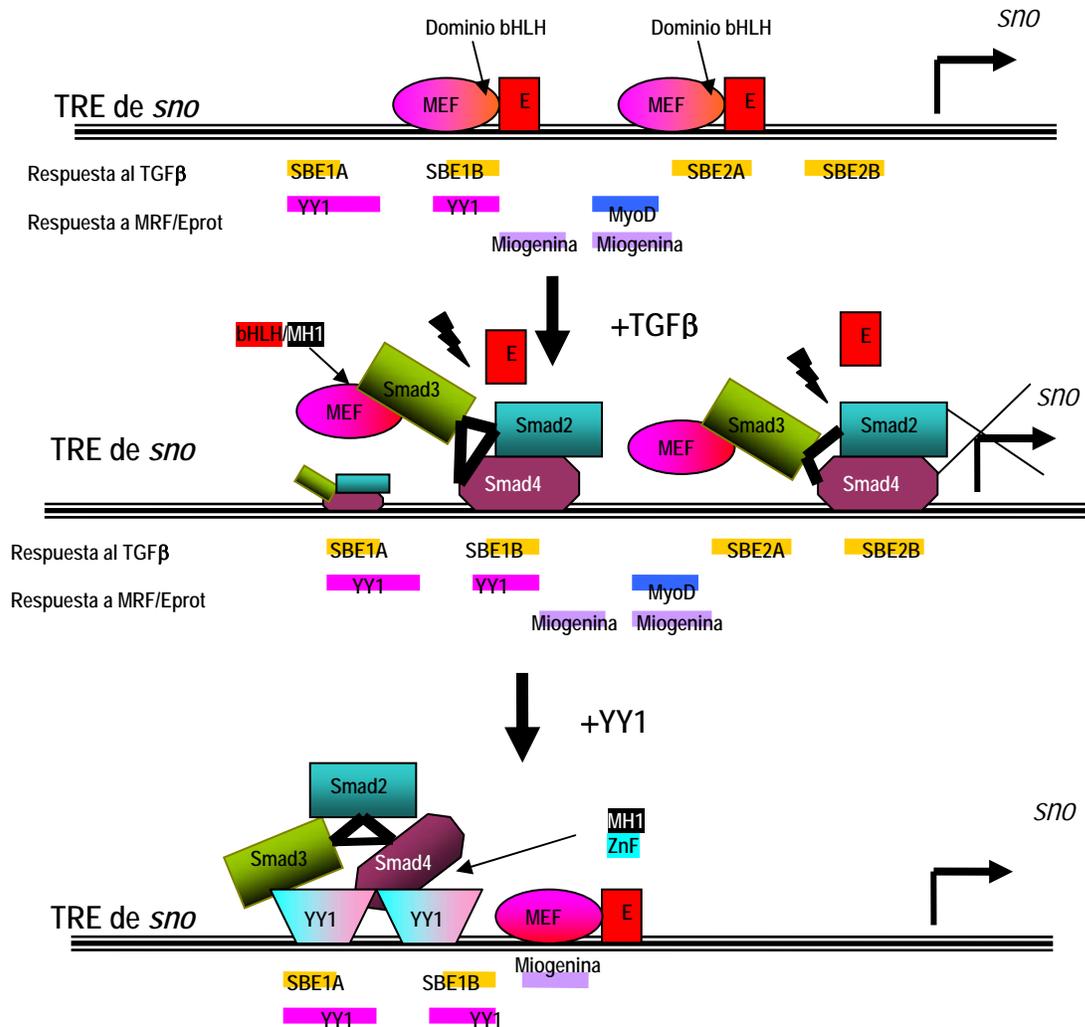
Altos niveles de *sno* inducen la diferenciación muscular en fibroblastos embrionarios de codorniz (Shinagawa *et al.*, 2000). Wrighton *et al.* comentan en su artículo de 2007 sobre el incremento en la expresión de *sno* durante la miogénesis. MyoD y Miogenina protagonizan este proceso de diferenciación, mientras que TGF $\beta$ , lo inhibe (Liu *et al.*, 2001). De esta manera, la presencia de TFBS putativos para MyoD en la vecindad de SBE2 y para Miogenina cerca de SBE1 y SBE2 (Figura 19) sugiere que MyoD y Miogenina son capaces de inducir la expresión de *sno* en *Homo sapiens* y *Mus musculus*, y que esta inducción es reprimida en la presencia del TGF $\beta$ .

Por otro lado, YY1 juega también un papel importante durante la diferenciación muscular. Kurisaki y colaboradores determinaron en 2003 que este factor de transcripción inhibe la regulación de genes blanco del TGF $\beta$  al asociarse con el dominio MH1 de Smad3 y Smad4 e impedir así la unión de estos factores a su TFBS.

En el promotor de *sno* se encuentran secuencias de unión a Miogenina, MyoD y YY1 en las vecindades de los SBE1 y 2 (Figura 19). Así, es posible que TGF $\beta$  antagonice los efectos de MyoD y Miogenina (Liu *et al.*, 2001) sobre la expresión de *sno* durante la miogénesis al inducir una interacción directa entre el dominio MH1 de Smad3 y el dominio bHLH de MyoD/Miogenina, mecanismo que ha sido reportado anteriormente por Liu *et al.* (2001) en otro contexto genómico. Por otro lado, Kurisaki y colaboradores mencionan en 2003 que YY1 inhibe la regulación genética mediada por TGF $\beta$  justamente por la asociación directa entre el dominio de dedos de Zinc de este factor de transcripción con el dominio MH1 de Smad3 y 4. De esta manera, es posible que YY1 inhiba la represión de la expresión de *sno* durante la miogénesis por el TGF $\beta$  al asociarse con Smad3,4 e impedir así que estos efectores se asocien de manera represiva con MyoD y Miogenina. Este modelo de acción se esquematiza en la figura 20.

El efecto global del TGF  $\beta$  sobre la expresión de *sno* es positivo (Zhu *et al.*, 2005). Así, el antagonismo entre el TGF $\beta$ , MyoD y Miogenina; así como entre el TGF $\beta$  y YY1, se puede reflejar sobre la expresión de *sno* inducida por el TGF $\beta$  de alguna manera no contemplada en el modelo propuesto. Evidentemente, este

modelo es sólo una propuesta y debe someterse a una verificación o refutación experimental. El diseño experimental para probar este modelo se menciona en otra sección.



**Figura 20.** Modelo de regulación de la expresión de *sno* durante la miogénesis. En ausencia del TGFβ, MyoD y Miogenina (MEF:E) inducen la expresión de *sno*. La estimulación con TGFβ inhibe esta inducción, ya que Smad3, por medio de su dominio MH1, se une al dominio bHLH de MEF, impidiendo su heterodimerización con la proteína E y por lo tanto su unión al DNA. Finalmente, la presencia de YY1 contraresta la inhibición de la expresión de *sno* durante la miogénesis mediada por TGFβ al asociarse con el dominio MH1 de Smad3 y 4, impidiendo así que estos factores se asocien de manera inhibitoria al dominio bHLH de MEF.

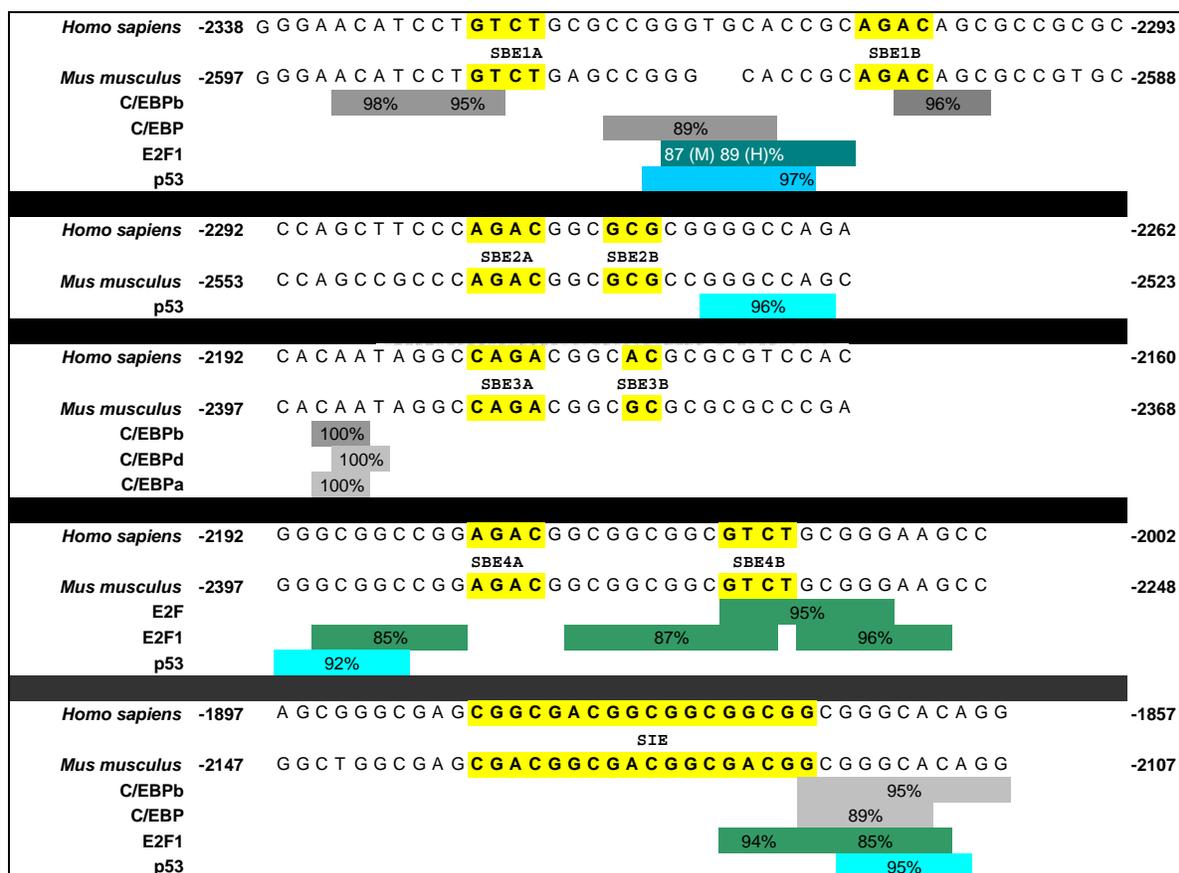
Los siguientes escenarios examinados son el ciclo celular y el cáncer.

El antagonismo del TGF $\beta$  y Sno en la oncogénesis de diversos tipos de cáncer ha sido ampliamente reportado (Macías-Silva y Briones-Orta, 2008). La regulación del TGF $\beta$  y Sno está frecuentemente alterada en un gran número de tipos de cáncer. La malignidad está además asociada a una incapacidad de las células de responder a los efectos antiproliferativos del TGF $\beta$ , en parte por una sobreexpresión de *sno* (Liu *et al.*, 2001). El alto índice de expresión de *sno* en células malignas se debe en gran medida al aumento en la tasa de transcripción (Krakowski *et al.*, 2005).

El TGF $\beta$  inhibe la progresión del ciclo celular por diversas vías. Por ejemplo, el TGF $\beta$  inhibe la expresión de *c-myc*, y por el otro, induce la transcripción de p21 (Liu *et al.*, 2001). En el desarrollo de cáncer se le ha adjudicado al TGF $\beta$  un papel dual. En las primeras fases de la carcinogénesis, el TGF $\beta$  actúa como supresor de tumores, ya que es un potente antimitótico. Empero, en fases avanzadas del cáncer, el TGF $\beta$  induce la metástasis y suprime el sistema inmune, por lo que se califica como protooncogen (Miyazono *et al.*, 2003).

Desde 1995, Pearson-White *et al.* observaron que los niveles de expresión de *sno* fluctúan con relación al ciclo celular. Estas observaciones fueron corroboradas por Wan *et al.* (2001), quienes establecieron que SnoN se acumula durante la fase G2, probablemente inhibiendo las señales antiproliferativas del TGF $\beta$ . Después de la mitosis, cuando la célula entra a G1, SnoN es degradada proteolíticamente, evento mediado por el complejo promotor de la anafase (APC).

En el TRE del gen *sno* de *Homo sapiens* y *Mus musculus* se encontraron TFBS putativos para TF íntimamente ligados a estos procesos fisiopatológicos. Se trata de E2F, p53 y C/EBP (Figura 21). Aunque RXR y AR son también importantes mediadores de la progresión del ciclo celular y frecuentemente están alterados en cáncer, por el papel unívoco de AR en el cáncer prostático, así como de RXR en el control de la proliferación y la diferenciación, estos TF se analizan separadamente.



**Figura 21.** TFBS putativos para TF que juegan un importante papel en el cáncer y la progresión del ciclo celular en las vecindades de SBE1-4 y del SIE de *sno* de *Homo sapiens* y *Mus musculus*.

La desregulación de la función de E2F1 sucede en una gran cantidad de tipos de cáncer humano. De manera interesante, tanto E2F1, como el TGF $\beta$ , juega un papel dual en el cáncer, actuando como supresor de tumores y como oncogen (Lacerte *et al.*, 2008).

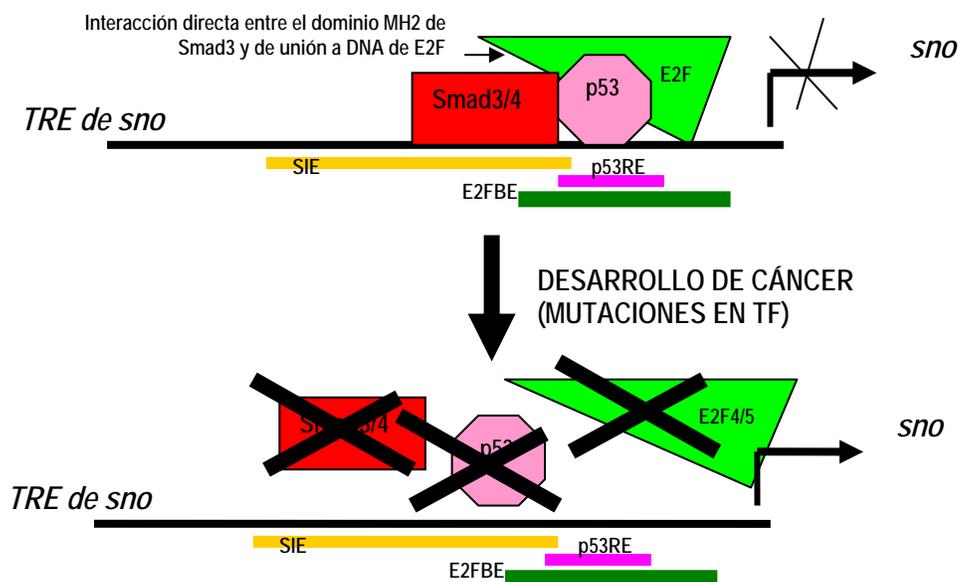
El “guardián del genoma” p53 es un supresor de tumores que previene el desarrollo de cáncer. Además de inducir la apoptosis y la reparación de DNA dañado, ejerce su acción al inhibir la expresión de protooncogenes (Macleod, 2000).

El TGF $\beta$  actúa en los primeros estadios de la carcinogénesis como supresor de tumores (Miyazono *et al.*, 2003). De hecho, se ha reportado que interactúa con p53, E2F1 y E2F4/5 en la inhibición de los protooncogenes *APF*, *hTERT* y *c-myc*, respectivamente (Wilkinson *et al.*, 2008; Lacerete *et al.*, 2008 y Chen *et al.*, 2002). Los niveles de SnoN se encuentran aumentados en diversos

tipos de cáncer (Luo, 2004). Así mismo, la condición oncológica está frecuentemente asociada con una pérdida en la función del TGF $\beta$  (Gold, 1999), p53 (Weisz *et al.* 2007) y E2F (Chen *et al.*, 2002).

El TRE del gen *sno* presenta dos zonas. Como demostraron Zhu *et al.* (2005), cuando Smad2/4 se asocia a los SBEs, el TGF $\beta$  induce la expresión de *sno*. Por otro lado, la unión de Smad3/4 al SIE inhibe la expresión de *sno*. SBE1, SBE2 y SBE4, pertenecientes a la región activadora de la transcripción de *sno*, presentan en sus vecindades sitios de unión para E2F y p53 (Figura 21). También la vecindad del SIE presenta elementos de unión a estos dos factores de transcripción. Así, se pueden distinguir dos secciones regulatorias propensas a un cambio en el contexto celular (presencia/actividad de determinados factores transcripcionales) durante el desarrollo del cáncer; SBE1,2 y 3 y SIE.

Debido a la evidencia de la acción sinérgica entre p53/E2F y el TGF $\beta$  en la represión de protooncogenes, se propone que p53/E2F y las Smad se encuentran reprimiendo la transcripción de *sno*. La aparición de mutaciones en los genes de estos factores transcripcionales, que en este caso están actuando como supresores de tumores, disminuye la capacidad represiva de *sno*, lo que contribuye a explicar el aumento en los índices protéicos y de mRNA de SnoN en el cáncer (Figura 22).



**Figura 22:** Modelo de regulación de la expresión de *sno* por supresores de tumores.

Hay varias analogías entre el SIE del gen *sno* y el TIE del gen *c-myc*. Por ejemplo, Chen y colaboradores (2008) reportan que al TIE del gen *c-myc* se une preferentemente el complejo Smad3/4, de la misma manera que Luo *et al.* (2005) demostraron que el efecto inhibitorio del TGF $\beta$  sobre la expresión de *sno* se da por medio de la unión de Smad3/4 al SIE. Además, el TIE del gen *c-myc* y el SIE de *sno* no presentan un elemento de unión a Smad perfecto; mientras que en *c-myc* se trata de un GGTC, el SIE está conformado por la secuencia **5'CGACGGCGACGGCGACGG 3'**.

Aunque el SIE de *sno* se asemeje al TIE de *c-myc*, el TRE de *sno* presenta además, ocho SBEs que bajo influencia del TGF $\beta$  inducen la expresión de *sno*, como lo reportan Zhu *et al.* (2005). Además, el efecto global del TGF $\beta$  sobre la expresión de *sno* es de inductor (Macías-Silva *et al.*, 2002).

Así, es indispensable considerar también la presencia de TFBS putativos para p53 y E2F en la vecindad de SBE1,2 y 4. Aunque numerosas observaciones sustentan el hecho de que p53 y E2F sinergizan la acción del TGF $\beta$  (Wilkinson *et al.*, 2008; Lacerete *et al.*, 2008; Chen *et al.*, 2002), no es posible descartar *a priori* un efecto antagónico entre p53/E2F y los efectores del TGF $\beta$  en el contexto genómico de *sno*. De hecho, el TGF $\beta$  induce una transformación oncogénica en algunas líneas celulares, como AKR-2B y NRK, a través de la inducción de la expresión de *sno* (Zhu *et al.* 2005). De esta manera, es posible que el TGF $\beta$  se encuentre desempeñando su papel de protooncogen, más que de supresor de tumores. Este papel se ha reportado ampliamente, por ejemplo en Miyazono *et al.* (2003).

En conclusión, la presencia de TFBS putativos para p53 y E2F en la vecindad de prácticamente todos los SBE y el SIE de *sno* con un grado de identidad muy alto con relación a los elementos de unión de estos TF prototipo (figura 21), los múltiples reportes de interacciones físicas y funcionales entre estos TF y las Smad (Wilkinson *et al.*, 2008; Lacerete *et al.*, 2008; Chen *et al.*, 2002), el papel del TGF $\beta$  en el cáncer, así como la alteración en los niveles de expresión de *sno* en el cáncer y la progresión del ciclo celular, sugieren ampliamente que E2F y p53 son capaces de interferir en la regulación de *sno* mediada por el TGF $\beta$ . El tipo de interferencia –positiva o negativa- debe ser comprobado experimentalmente.

En cuanto a C/EBP, se ha demostrado que este TF y el TGF $\beta$  actúan en sinergia sobre la regulación de la expresión de *c-myc* y *p21* en cáncer de mama (Nerlov, 2007). Como reseñan Macías-Silva y Briones-Orta en 2008, SnoN tiene un efecto importante en este padecimiento. Líneas celulares de cáncer de mama con niveles altos de expresión de *sno* son capaces de promover el crecimiento de tumores *in vivo*; sin embargo, el silenciamiento de *sno* promueve la metástasis. De hecho, según comentan Macías-Silva y Briones-Orta, recientemente se ha propuesto a SnoN como un marcador pronóstico para cáncer de mama positivo a receptores de estrógeno.

Coyle-Rink y colaboradores (2002) demostraron que Smad3 y Smad4 son capaces de interactuar físicamente con C/EBP $\beta$ . En ese mismo año, Chen *et al.* detectaron que C/EBP se une particularmente al dominio MH2 de Smad2 y Smad3. Estos autores proponen que esta unión tiene un efecto inhibitorio sobre la regulación de genes blanco del TGF $\beta$ , ya que –según comentan- C/EBP impide la unión de coactivadores como p300 a las Smad e incluso la unión al SBE.

Nerlov (2007) y Feng y Derynck (2005) mencionan que C/EBP y el TGF $\beta$  actúan en sinergia en la regulación de genes blanco, tales como *c-myc*, *p21*, *p15* y *HIF*. Además, estas dos moléculas juegan un importante papel en la proliferación y la diferenciación celular (Nerlov, 2007; Liu *et al.*, 2001).

Por lo tanto, se puede predecir que el TGF $\beta$ , a través de las Smad, y C/EBP colaboran en la regulación de la transcripción de *sno*. Esta cooperación se da, probablemente, en situaciones fisiológicas en las que la proliferación y la diferenciación juegan un papel importante, como en el cáncer y la regeneración. En estos escenarios, los niveles de *sno* se encuentran frecuentemente alterados. Luo (2004) sostiene que los niveles de SnoN se encuentran elevados en diversas líneas celulares humanas derivadas de melanoma, de cáncer de mama, de carcinoma del esófago, del pulmón, del estómago, de la epidermis y de la próstata. En particular, en cáncer de mama, padecimiento en el que se ha comprobado que hay sinergia entre el TGF $\beta$  y los C/EBP (Nerlov, 2007), *sno* se encuentra tan sobreexpresado que funciona incluso como marcador (Macías-Silva y Briones-Orta, 2008).

En conclusión, la presencia de TFBS putativos para C/EBP en el gen *sno* de *Homo sapiens* y *Mus musculus*, en particular en las vecindades de los SBE1, SBE3 y SIE, así como reportes previos de cooperación entre las Smads y los C/EBP, sugieren que en determinados escenarios fisiopatológicos, los C/EBP podrían sinergizar con el TGF $\beta$  en la regulación de la transcripción de *sno*.

Por lo anterior, es probable que los C/EBP y el TGF $\beta$  colaboren o antagonicen la regulación transcripcional de *sno*. El tipo de interacción funcional que estas dos moléculas tengan sobre la expresión de *sno* debe ser corroborada experimentalmente.

El siguiente escenario fisiológico a analizar es la respuesta inmune, en el que el TGF $\beta$  juega un papel importante. Como citocina inmunosupresora, el TGF $\beta$  inhibe el crecimiento y la diferenciación de muchas líneas celulares, incluyendo células B y T. Además, el TGF $\beta$  regula la expresión de moléculas de adhesión y de matriz extracelular, y es un importante quimioatrayente de monocitos y macrófagos e inhibe la activación de células del sistema inmune por presentación de antígeno o interleucina (Moustakas *et al.*, 2002). De hecho, Pardoux *et al.* (1999) comentan que ratones *knockout* para el TGF $\beta$  nacen con inflamación crónica, así como con alteraciones en la hematopoyesis. Lin *et al.* (2005) mencionan además que el TGF- $\beta$ 1 suprime el desarrollo y la intensidad de enfermedades autoinmunes.

Pearson-White y McDuffie (2003) demostraron que *sno* se encuentra constitutivamente expresado en las células del timo y del bazo. Además, mostraron que células T carentes de *sno* son incapaces de proliferar. Este efecto se debe a que *sno* regula negativamente los efectos antiproliferativos del TGF $\beta$ .

En el TRE del promotor del gen *sno* se detectaron TFBS putativos para dos factores de transcripción característicos del sistema inmune, NF $\kappa$ B y STAT4. Estos dos TFBS putativos están en las vecindades de SBE1A y SBE4B (Figura 18).

NF $\kappa$ B y Smad3 son capaces de interactuar física y funcionalmente en la regulación transcripcional de genes blanco de citocinas proinflamatorias y/o del TGF $\beta$  (Feng y Derynck, 2005; Kon *et al.*, 1999). Sin embargo, el TGF $\beta$  y las

citocinas proinflamatorias pueden tener un efecto antagónico o sinérgico (Kon *et al.*, 1999, López-Rovira *et al.*, 2000). 10pb río arriba de SBE1A hay un sitio de unión a NFκB putativo (Figura 18). Por lo tanto, es posible que el TGFβ y las citocinas proinflamatorias como IL6 o TNFα actúen de manera sinérgica o antagónica en la regulación transcripcional de *sno*.

Así mismo, se detectó una secuencia consenso de unión a STAT4 en la vecindad tanto de SBE1A (97% de identidad con relación a la secuencia consenso) como de SBE4B (95%). Reportes como los de Pardoux *et al.* (1999) sugieren que el TGFβ antagoniza a STAT4 al impedir la unión de este factor de transcripción a su TFBS. Lin *et al.* (2005) confirmaron este resultado al demostrar que el TGFβ1 inhibe la inducción de la expresión de IFN-γ mediada por IL-12.

Resulta difícil realizar una predicción sobre la influencia de NFκB y STAT4 sobre la transcripción de *sno* mediada por el TGFβ. Por un lado, el TGFβ es un potente inmunosupresor (Pardoux *et al.*, 1999; Moustakas *et al.*, 2002; Lin *et al.*, 2005), mientras que los estímulos que activan a los factores de transcripción NFκB y STAT4 son frecuentemente proinflamatorios (LPS, IL, TNFα, etcétera; e IL12, respectivamente).

Este hecho fisiológico sugiere un antagonismo entre NFκB/STAT4 y Smad2/3/4 en la inducción de la transcripción de *sno*. Además, el trabajo de Pearson-White y McDuffie (2003) muestra que los niveles de mRNA de *sno* se encuentran constitutivamente elevados en células del bazo y del timo, y que SnoN es un importante regulador negativo del TGFβ en el sistema inmune, como lo muestra la incapacidad de proliferación de células T en ausencia de *sno* y en presencia del TGFβ.

Sin embargo, la regulación genética mediada por el TGFβ pocas veces es tan lineal. La inducción de la transcripción de *sno* por el TGFβ es un ejemplo de que esta citocina inmunosupresora induce a sus propios antagonistas.

Por lo tanto, bajo esta perspectiva, es posible que NFκB se encuentre cooperando con el TGFβ en la inducción de la expresión de *sno*. Se ha observado que NFκB y Smad3 son capaces de cooperar física y funcionalmente en la

inducción de la transcripción de genes blanco del TGF $\beta$  (López-Rovira, *et al.*, 2000). Por lo tanto, es posible que estos dos factores de transcripción se encuentren sinergizando la regulación de la transcripción del gen blanco del TGF $\beta$  *sno*.

Existen pocos reportes en torno a la interacción directa entre STAT4 y los efectores del TGF $\beta$ . Por lo tanto, realizar una predicción en torno a la forma de cooperación (si es que se da) entre STAT4 y Smad2/3/4 en el gen de *sno* humano y murino es aventurada, por lo que entre las perspectivas del presente trabajo se encuentra la realización de una inspección experimental.

Las predicciones *in silico* de TFBS necesariamente contienen un gran número de falsos positivos (Lässig, 2007). En el presente trabajo se intentó disminuir drásticamente el número de falsos positivos considerando únicamente aquellos TFBS putativos conservados entre las dos especies de estudio y que correspondan a TF típicos de algún escenario fisiológico o patológico en el que se sepa que el TGF $\beta$  juega un papel importante y que los niveles de expresión de *sno* se encuentren alterados.

Se encontró en el TRE del gen *sno* en *Homo sapiens* y *Mus musculus* una gran cantidad de TFBS putativos que incluso se sobrelapan frecuentemente (Figura 18). Desde el punto de vista fisiológico, el hecho de que los TFBS se sobrelapan sugiere que la regulación de la transcripción es compleja. Es decir, que bajo determinadas circunstancias o estímulos, y en la presencia simultánea de factores de transcripción capaces de asociarse a un TFBS que se sobrelapa, probablemente suceda un proceso de competencia o cooperación entre estos factores.

Por ejemplo, se ha reportado que YY1 ejerce su efecto inhibitorio sobre la regulación transcripcional de un gen al desplazar por competencia a un factor de transcripción activador cuya secuencia de unión en el DNA esté traslapada con la de YY1 (Gordon *et al.*, 2006). En el presente trabajo se propone que este mecanismo subyace la regulación de la transcripción del gen *sno* durante la

diferenciación muscular (figura 20). No obstante, no se efectuarán más conjeturas con relación al traslape de algunos TFBS putativos hasta no comprobar experimentalmente la funcionalidad de estas secuencias.

Antes de realizar un diseño experimental para la comprobación de las predicciones efectuadas en el presente trabajo, se centrará brevemente la atención sobre el gran número de secuencias de unión a las Smads en el promotor del gen *sno* con relación a otros genes blanco del TGF $\beta$ . Como contraste se considera a *smad7*, ya que, como *sno*, es un regulador negativo del TGF $\beta$  que es activado transcripcionalmente por esta citocina.

El promotor de *sno* de las especies estudiadas presenta ocho SBEs y un SIE, es decir, en total nueve elementos de unión a Smads (Zhu *et al.*, 2005). Por otro lado, el promotor blanco del TGF $\beta$ , *smad7*, presenta únicamente un SBE. Este SBE es indispensable para la respuesta transcripcional ejercida por el TGF $\beta$  sobre *smad7* (Stopa *et al.*, 2000). Estudios de mutagénesis dirigida sobre los sitios de unión a Smad en el gen *sno* muestran que la eliminación de un SBE no es suficiente para inhibir la respuesta transcripcional al TGF $\beta$  sobre este gen (Zhu *et al.*, 2005).

El TRE en el gen *smad7* detectado por Stopa *et al.* (2000) es mucho más pequeño (141pb) que en *sno* (514pb en *Mus musculus* y 508pb en *Homo sapiens*). Desde un punto de vista estadístico, un mayor número de pares de bases en el TRE de *sno* permite la presencia de un mayor número de TFBS aledaños y capaces de interactuar con las Smad.

En suma, la regulación de genes blanco del TGF $\beta$ , en particular de *sno*, sucede por la unión de Smads a un SBE. Sin embargo, la unión de Smad al DNA no siempre es suficiente para inducir la expresión o represión de un gen blanco del TGF $\beta$ . Hay ahora un creciente número de ejemplos en los que Smad coopera con otros factores de transcripción para regular la transcripción (Subramaniam *et al.*, 2001).

Se han demostrado este tipo de interacciones entre las Smad y otros factores de transcripción –tales como GR, AR, C/EBP, YY1, etcétera- en una gran cantidad de genes blanco del TGF $\beta$  (Feng y Derynck, 2005).

Desde el artículo de Stroschein *et al.* (1999) se sabe que *sno* es un gen blanco del TGF $\beta$ . Sin embargo, fue sólo hasta el trabajo de Ruoyun y colaboradores en 2005 que se demostró que la inducción de la expresión de *sno* depende de la presencia de TFBS para Smads en el promotor de *sno*.

En el presente trabajo, se detectaron sitios de unión putativos para factores de transcripción capaces de asociarse a las Smad (revisado en Feng y Derynck, 2005). Para ello, se utilizó el programa de predicción de TFBS, *Alggen*. Se eligió este programa porque fue el único capaz de detectar los SBE determinados experimentalmente por Ruoyun y colaboradores en 2005. Además, se seleccionaron sólo aquellos TFBS putativos que se encuentran conservados entre ambas especies de estudio (*Homo sapiens* y *Mus musculus*), ya que estudios, como el de Levy y Hannenhalli (2002) que proponen que las zonas no codificantes que se encuentran conservadas (similitud mayor o igual a 80%) muy probablemente constituyen elementos regulatorios.

Se detectaron TFBS putativos para GR, AR, C/EBP, YY1, E2F, RXR, p53, STAT4, MyoD, Miogenina, NF $\kappa$ B, STATy Sp1 en la vecindad ( $\pm$ 10pb) de los SBE del gen *sno* determinados por Ruoyun y colaboradores en 2005 (Figura 18).

Los factores de transcripción mencionados responden a una gran diversidad de estímulos, presentes en distintas condiciones fisiológicas e incluso patológicas. Por ejemplo, MyoD, Miogenina y YY1 son proteínas características de la miogénesis; E2F, RXR, C/EBP y p53 juegan un papel importante en la progresión del ciclo celular; los receptores intercelulares GR y AR responden a cambios en la concentración hormonal, y NF $\kappa$ B y STAT4 son importantes moduladores de la respuesta inmune innata.

Esto concuerda con los datos experimentales sobre las condiciones fisiológicas y patológicas en las que se encuentra sobreexpresado *sno*. Wrighton y colaboradores mostraron que *sno* está sobreexpresado durante la miogénesis, Luo revisó en 2004 su papel dual en el desarrollo de cáncer, y Marina-Macías *et al.* determinaron en 2002 que la expresión de *sno* aumenta durante la regeneración

hepática, evento en el que el índice de proliferación aumenta, y por tanto hay una progresión del ciclo celular más acelerada que en condiciones normales.

Paralelamente, se sabe que el TGF $\beta$  protagoniza estos procesos de miogénesis, cáncer y regeneración hepática a través de sus efectores, las Smad. Durante la miogénesis, el TGF $\beta$  juega un papel inhibitorio (Liu *et al.*, 2001). En la regeneración hepática, el TGF $\beta$  orchestra la síntesis de matriz extracelular (Macías-Silva *et al.*, 2002). Además, se sabe que el TGF $\beta$  juega un papel importante en la respuesta inmune (Liu *et al.*, 2001). De manera análoga a SnoN, se le ha adjudicado al TGF $\beta$  un papel bifásico en el desarrollo de cáncer (Miyazono, 2003 y Luo, 2004).

De esta manera, es probable que bajo las circunstancias fisiológicas en las que aumenta la expresión de *sno*, estén interactuando sobre el promotor de este gen los efectores de la vía del TGF $\beta$ , las Smad, y otros factores de transcripción que responden a determinados estímulos presentes en las mencionadas circunstancias fisiológicas y/o patológicas. Esta interacción tiene como consecuencia un cambio en la expresión del gen *sno*, cambio que previamente se había detectado experimentalmente.

En el presente trabajo, se dedujeron estos TF a partir de un análisis *in silico* del TRE de *sno* de *Homo sapiens* y *Mus musculus*.

## Propuestas para la validación *in vitro* de las predicciones *in silico*

Esta sección se centra en proponer una estrategia experimental para la validación de las predicciones en torno a los TFBS. También el TSS putativo se debe comprobar experimentalmente; sin embargo, no se abunda sobre técnicas que permitan corroborar la posición precisa del TSS en el gen *sno* de *Mus musculus* y *Homo sapiens* porque se mencionó anteriormente.

La presencia de TFBS putativos para TF capaces de asociarse con las Smad en un promotor determinado no implica que estos factores modulen la regulación de la expresión del gen. Por ejemplo, la secuencia de respuesta a E2F1 y al TGF $\beta$  de *hTERT* presenta dos sitios de unión para la caja E, cinco cajas CpGs de unión a Sp1, cuatro sitios de unión a E2F1 y sólo un SBE. Sin embargo, Lacerete *et al.* (2008) demostraron por medio de mutaciones puntuales de los elementos de unión a Sp1 y a la proteína E, que interferir sobre la unión de los TF no siempre modifican el efecto del TGF $\beta$  sobre la expresión de este gen blanco, ya que sólo la abolición de todos los sitios de unión a E2F1 suprimió el efecto del TGF $\beta$  sobre la expresión de *hTERT*. Cabe destacar que los TFBS mencionados no se encuentran en la vecindad del SBE de *hTERT*.

De esta manera, es necesario realizar un análisis experimental para determinar la importancia relativa de cada uno de los TFBS putativos que se detectaron en la vecindad de cada SBE de *sno*.

Hay una amplia gama de técnicas para la determinación experimental de TFBS y de su efecto sobre la expresión del gen blanco. La caracterización de TRE de *smad7*, por ejemplo (Stopa *et al.*, 2005), se realizó por medio de EMSA (Electrophoretic Mobility Shift Assay) utilizando oligonucleótidos con TFBS alterados, de tal suerte que es posible establecer la especificidad de unión a DNA de los complejos TF-Smad con relación a la presencia de TFBS. La influencia de la asociación de TF a sus sitios de unión sobre la tasa de expresión del gen blanco se estableció en este mismo trabajo por medio de ensayos de luciferasa con TFBS mutados.

Para la comprobación de las predicciones realizadas en el presente trabajo se propone una estrategia similar.

En primer lugar, es necesario determinar si el TF se une al TFBS que se predijo en el presente trabajo. Para tales efectos se propone utilizar la técnica EMSA. Para cada TF para el cual se detectó un TFBS putativo en las vecindades de los SBE y/o el SIE en *sno* (TFX) se sugiere realizar el experimento con el TRE de *sno*, añadiendo las proteínas en la combinación que en la Tabla 10 se señalan. Por el alto grado de identidad entre los TRE (79%, figura 17), y porque los TFBS putativos a probar están presentes en ambas especies de estudio, debe ser indistinto si se utiliza el TRE de *Homo sapiens* o de *Mus musculus*.

**Tabla 9.** Propuesta para comprobar la funcionalidad de los TFBS putativos por medio de un EMSA. El oligonucleótido que se debe someter al ensayo es el TRE de *Homo sapiens* o de *Mus musculus*. TFX: TF para el cual se detectó un TFBS putativo en las vecindades de los SBE y/o el SIE en *sno*.

	Control	Se espera unión, por Zhu <i>et al.</i> 2005	Mide independiente del TGFβ	unión del TGFβ	Unión dependiente del TGFβ
Smad	-	+	-	+	
TFX	-	-	+	+	

Una vez que se ha detectado cuáles son los TF que efectivamente se unen a su TFBS en el TRE, se analiza el efecto de la eliminación de los TFBS sobre la expresión del gen blanco.

Para ello es necesario, en primer lugar, seleccionar una línea celular que exprese los factores de transcripción que se someterán a prueba.

Los efectos sobre la expresión de *sno* al alterar (“eliminar”) los TFBS se pueden analizar por medio de un ensayo de luciferasa. Para ello, se propone utilizar una población de células sensibles al TGFβ transfectadas con un plásmido de expresión con el TRE de *sno* intacto y otra población con los SBEs y el SIE mutados, siguiendo las modificaciones a estos TFBS realizadas por Zhu *et al.* (2005).

La “eliminación” de los sitios de unión para los factores de transcripción que se desean probar se puede realizar por medio de la introducción de oligonucleótidos *decoy*, que unen a los TF, impidiendo que se unan a su TFBS.

Los oligonucleótidos *decoy* son fragmentos de aproximadamente 20 pb que contienen secuencias de unión para el factor de transcripción en estudio. En

los últimos años, los oligonucleótidos (ODN) han recibido mucha atención, ya que constituyen una manera de regular de manera selectiva la expresión genética. Una vez dentro de la célula, los oligonucleótidos sintéticos que contienen el TFBS son reconocidos específicamente por el TF que se desea inhibir. Así, el factor de transcripción unido al oligonucleótido, no es capaz de unirse a su TFBS en el promotor del gen blanco (Crinelli *et al.* 2002).

Así, la introducción de oligonucleótidos *decoy* para un factor de transcripción determinado al núcleo celular inhibe la unión del factor de transcripción a su TFBS. Por ejemplo, la transcripción de p21 mediada por p53 se bloquea completamente en respuesta a la introducción del oligonucleótido *decoy* (gancho/seductor) para p53 (Sakaguchi *et al.*, 2005).

Para verificar que las mutaciones en el TRE y el uso de oligonucleótidos *decoy* efectivamente previenen la unión de las Smad y los TFX a su TFBS, se debe realizar un EMSA (Morishita *et al.*, 1998).

Para determinar si los TF se unen a su TFBS putativo, se puede realizar, además o en vez del EMSA, una inmunoprecipitación de cromatina (ChIP). Un Re-ChIP permitiría visualizar además los TF que se unen al TRE y a alguna Smad. Además, esta técnica permite evaluar la regulación de la expresión del gen de *sno* de manera endógena.

## Predicciones con base en análisis *in silico*

En los últimos años se ha generado un gran número de datos en las ciencias de la vida. Con esta inmensa cantidad de información, se antoja necesario realizar una síntesis de la misma con el objetivo de empezar a predecir fenómenos. Los estudios *in silico* pretenden contribuir a este objetivo. No obstante, la limitada capacidad de los algoritmos, el importante componente azaroso e histórico en la biología, así como la inmensa complejidad de los seres vivos dificulta esta empresa. Hasta la fecha, las predicciones *in silico* representan invitaciones para los investigadores experimentales a seguir determinado camino. Es decir, se pretenden reducir los callejones sin salida de la investigación.

En el caso de las secuencias génicas regulatorias, la predicción de TFBS reduce considerablemente el número de TF candidatos en un proceso determinado de modulación de la transcripción. Por ejemplo, en el presente trabajo se descartaron 44 de 56 factores de transcripción y cofactores candidatos capaces de participar en la regulación genética de *sno* mediada por las Smad. Las predicciones que se realizan en torno a los doce TF evidentemente deben ser corroboradas experimentalmente. Las predicciones *in silico* normalmente crean falsos positivos, pero los falsos negativos prácticamente no aparecen (Halfon y Michelson, 2002). Sin embargo, cabe mencionar que en el presente trabajo, los falsos negativos pueden estar representados por los TF capaces de modular la acción del TGF $\beta$  independientemente de la presencia de un TFBS.

Dado que un factor de transcripción es capaz de encontrar su TFBS funcional, actualmente se busca que los investigadores también seamos capaces de predecir esta secuencia. Si esto es posible, entonces podemos predecir las interacciones genéticas específicas construyendo redes regulatorias a partir de datos de secuencias. El análisis de DNA regulatorio es uno de los mayores tópicos en la investigación en el área de la bioinformática. Sin embargo, la determinación de interacciones génicas a partir de la predicción de TFBS no es tarea fácil. El análisis de secuencias regulatorias requiere de criterios estadísticos que permitan distinguir entre TFBS funcionales y falsos positivos (Lässig, 2007).

El presente trabajo predice TFBS en el promotor del gen *sno* de *Homo sapiens* y *Mus musculus* con el fin de apuntar sobre las señales intra- y extracelulares que interfieren con la regulación de la expresión de *sno* mediada por el TGFβ, ya sea positiva- o negativamente, en sinergia o antagonismo.

El método que se siguió fue un análisis *in silico* del promotor de *sno*, en particular de la secuencia regulada por el TGFβ (TRE). Se combinó el método de predicción de TFBS cuyo punto de partida es la base de datos TRANSFAC con la selección de las posibles secuencias regulatorias conservadas evolutivamente. La búsqueda de TFBS putativos se realizó con el programa *Alggen* que detecta subsecuencias por medio de una matriz posicional dependiente del peso (*Positional Wight Matrix*, PWM).

Esta matriz contiene una descripción cuantitativa de las variaciones de pares de nucleótidos en cada posición del TFBS. Las PWM para cada TFBS se han obtenido a partir de datos experimentales, en particular por medio del estudio de la unión de un factor de transcripción a su TFBS y la subsiguiente determinación de la secuencia consenso. En la figura 23 se muestra un ejemplo.

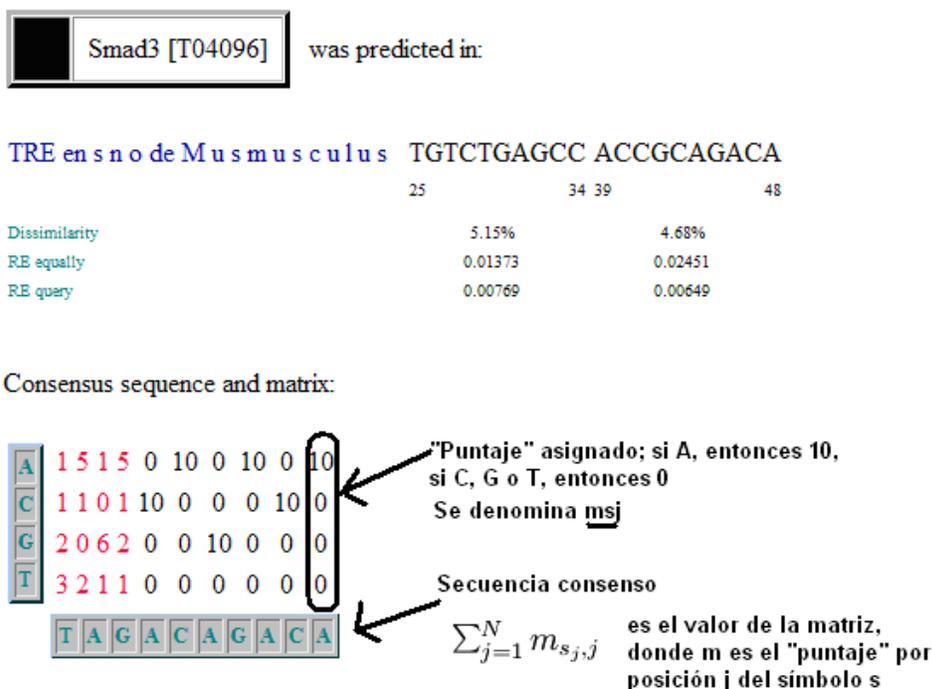


Figura 23. Ejemplo del *modus operandi* del programa *Alggen*.

Además de la detección de subsecuencias semejantes a determinados TFBS, la presencia de sitios de unión para factores de transcripción capaces de interactuar con las Smad (reseñado en Feng y Derynck, 2005) se propuso exclusivamente para los TFBS detectados por *Alggen* para las dos especies de estudio. Es decir, que la probabilidad de la funcionalidad de elementos regulatorios aumenta considerablemente si estas secuencias están conservadas (Levy y Hannenhalli, 2002; Loots y Ovcharenko, 2004).

Este método de detección de TFBS putativos es análogo al que utiliza el programa rVista 2.0. Loots y Ovcharenko (2004) probaron el poder predictivo de este programa intentando detectar SBEs en el promotor de *NKX2.5*, un gen blanco del TGF $\beta$ . En una secuencia de 7kb, el programa detectó cuatro SBEs dentro de un elemento conservado de 2 kb que previamente se había catalogado como fundamental para la regulación de *NKX2.5*. Además, un SBE coincidió con una base que anteriormente se había determinado experimentalmente como fundamental para la inducción de la expresión de este gen.

Los resultados de Loots y Ovcharenko (2004) muestran que la predicción *in silico* de secuencias regulatorias puede servir como un primer esbozo de la regulación genética. Evidentemente, los resultados que por esta vía se obtengan, deben ser siempre corroborados experimentalmente.

## Conclusión

En el presente trabajo se realizó un análisis bioinformático de los promotores de *sno* de *Homo sapiens* y de *Mus musculus*. En primer lugar, se detectó satisfactoriamente una secuencia de 5kb para cada especie de estudio, que llamamos promotor del gen *sno*, que presenta los dos módulos regulatorios que habían sido detectados experimentalmente con anterioridad. Tres métodos distintos permitieron establecer un TSS putativo que, aunque debe ser corroborado experimentalmente, da un panorama de las distancias entre los módulos regulatorios y el inicio de la transcripción. Provisionalmente se descartó la presencia de TFBS para factores basales de la transcripción en la zona núcleo del promotor de *sno* de ambas especies. La ausencia de estos TFBS putativos concuerda con el hallazgo de una zona rica en CpGs en esta zona. A lo largo de la investigación se compararon las secuencias analizadas, obteniendo un alto grado de conservación entre las regiones regulatorias. Un gran número de criterios permitió establecer doce TF candidatos que pueden influir sobre la tasa de transcripción de *sno* en el contexto del TGF $\beta$ . La mayoría de estos TF son típicos de determinados escenarios fisiopatológicos que concuerdan, en muchos casos, con una alteración en los niveles de *sno*, así como algún papel sobresaliente del TGF $\beta$ .

Como perspectivas del trabajo, se propone corroborar las predicciones *in silico*, por ejemplo a través de técnicas experimentales como el EMSA, el ChIP y Re-ChIP, y ensayos con genes reporteros.

Finalmente, el presente trabajo representa una invitación a investigadores experimentales a considerar para la elaboración de sus hipótesis las predicciones realizadas *in silico*, ya que pueden disminuir -de manera considerable- los senderos sin salida en la investigación.

# Hipergrafía

- ♦ Programas:
  - Traducción de secuencia en formato GenBank a FASTA  
<http://www.ebioinfogen.com/biotools/genbank-fasta.htm>
  - Alineación de dos secuencias  
<http://www.ncbi.nlm.nih.gov/BLAST/bl2seq/wblast2.cgi>  
[http://www.ch.embnet.org/software/LALIGN\\_form.html](http://www.ch.embnet.org/software/LALIGN_form.html)  
<http://www.ebi.ac.uk/Tools/clustalw2/index.html>
  - Comparación de una secuencia con el genoma total de un organismo:  
  
*H. sapiens:*  
<http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9606>  
  
*M. musculus:*  
<http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=10090>
  - Análisis de sitios de unión a factores transcripcionales  
<http://Alqgen.lsi.upc.es/>

# Bibliografía

Akira, S. y K. Takeda. 2004. Toll-like receptor signalling. Review *En: Nature Reviews Immunology* **4**: 499-510

Akiyoshi S., M. Ishii, N. Nemoto, M. Kawabata, H. Aburatani y K. Miyazono. 2001. Targets of Transcriptional Regulation by Transforming Growth Factor- $\beta$ : Expression Profile Analysis Using Oligonucleotide Arrays *En: Japanese Journal of Cancer Research* **92**: 257–268

Antequera, F. 2003. Structure, function and evolution of CpG island promoters. Review *En: Cellular and Molecular Life Sciences* **60**: 1647-1658

Balk, S. y K. Knudson. 2007. AR, the cell cycle, and prostate cancer. Review *En: Nuclear Receptor Signaling* **6**: e001- e012

Bajic, V., V. Choudhary y C. Hock. 2003. Content analysis of the core promoter region of human genes *En: In Silico Biology* **4** (2): 109-25

Bright, J. y S. Sriram. 1998. TGF- $\beta$  Inhibits IL-12-Induced Activation of Jak-STAT Pathway in T Lymphocytes *En: The Journal of Immunology* **161**: 1772-1777

Chen, C., Y. Kang, P. Siegel y J. Massagué. 2002. E2F4/5 and p107 as Smad Cofactors Linking the TGF $\beta$  Receptor to *c-myc* Repression *En: Cell* **110**: 19-32.

Chen, F., K. Ogawa, X. Liu, T. Stringfield y Y. Chen. 2002. Repression of Smad2 and Smad3 transactivating activity by association with a novel splice variant of CCAAT-binding factor C subunit *En: Biochemical Journal* **364**: 571-577.

Chen, Y., A. Takeshita, K. Ozaki, S. Kitano y S. Hanazawa. 1996. Transcriptional Regulation by Transforming Growth Factor  $\beta$  of the Expression of Retinoic Acid and Retinoid X Receptor Genes in Osteoblastic Cells Is mediated through AP-1 *En: The Journal of Biological Chemistry* **271** (49): 31602-31606.

Chu, S. y T. J. Ferro. 2005. Sp1: Regulation of gene expression by phosphorylation (Review) *En: Gene* **348**: 1- 11

Clark, D. 2005. *Molecular Biology. Understanding Genetic Revolution*. Ed. Elsevier, Reino Unido, pp. 694-713

Coyle-Rink, J., T. Sweet, S. Abraham, B. Sawaya, O. Batuman, K. Khalli y S. Amini. 2002. Interaction between TGF $\beta$  Signaling Proteins and C/EBP Controls Basal and Tat-Mediated Transcription of HIV-1 LTR in Astrocytes *En: Virology* **299**: 240-247

Crinelli, R., M. Bianchi, L. Gentilini y M. Magnani. 2002. Design and characterization of decoy oligonucleotides containing locked nucleic acids *En: Nucleic Acids Research* **30** (11): 2425-2443

Davidson E. y J. Coffman. 2007. Genetic Networks *En: els.net*

Efeyan, A. y M. Serrano. 2007. p53: guardian of the genome and policeman of the oncogenes *En: Cell Cycle* **6** (9): 1006-1010

Feng, X. y R. Derynck. 2005. Specificity and Versatility in TGF- $\beta$  Signaling Through Smads *En: Annual Review of Cell and Developmental Biology* **21**: 659–93

Gold, L. 1999. The role for transforming growth factor-beta (TGF-beta) in human cancer. Review *En: Critical Reviews in Oncogenesis* **10** (4): 303 - 360.

Gomperts, B., Kramer, I y P. Tatham. 2002. *Signal Transduction*, Ed. Elsevier, E.U.A. 423 pp.

Gordon, S., G. Akopyan, H. Garban y B. Bonavida. 2006. Transcription Factor YY1: structure, function, and thertapeutic implications in cancer biology. Review *En: Oncogene* **25**: 1125-1142.

Halfon, M. y A. Michelson. 2002. Exploring genetic regulatory networks in metazoan development: methods and models. Review *En: Physiological Genomics* **10**: 131–143

Hayes S., M. Zarnegar, M. Sharma, F. Yang, D. Peehl, P. Dijke y Z. Sun. 2001. SMAD3 Represses Androgen Receptor-mediated Transcription *En: Cancer Research* **61**: 2112-2118

Heldin, CH. 2002. Signal Transduction: Overview *En: els.net*

Hsu, Y-H., K. Sarker, I. Pot, A. Chan, S. Netheton y S. Bonni. 2006. Sumoylated SnoN Repress Transcription in a Promoter-specific Manner *En: Journal of Biological Chemistry* **281** (44): 33008-33018

Hua, X., Z. Miller, G. Wu, Y. Shi y H. Lodish. 1999. Specificity in transforming growth factor  $\beta$ -induced transcription of the plasminogen activator inhibitor-1 gene: Interactions of promoter DNA, transcription factor mE3, and Smad proteins *En: PNAS* **96** (23): 13130–13135

Ivanova I., S. D'Souza y L. Dagnino. 2005. Signalling In The Epidermis: The E2f Cell Cycle Regulatory Pathway In Epidermal Morphogenesis, Regeneration And Transformation *En: International Journal of Biological Sciences* **1**: 87-95

Johnson, K., H. Kirkpatrick, A. Comer, F. M. Hoffmann y A. Laughon. 1999. Interaction of Smad Complexes with Tripartite DNA-binding Sites *En: The Journal of Biological Chemistry* **274** (29): 20709–20716

Juven-Gershon, T., J. Hsu, J. Theisen y J. Kadonaga. 2008. The RNA polymerase II core promoter — the gateway to transcription *En: Current Opinion in Cell Biology* **20**:253–259

Kang, H., K. Huang, S. Chang, W. Ma, W. Lin y C. Chang. 2002. Differential Modulation of Androgen Receptor-mediated Transactivation by Smad3 and Tumor Suppressor Smad4 *En: The Journal of Biological Chemistry* **277** (46): 43749–43756

Kang, H., H. Lin, Y. Hu, S. Yeh, K. Huang y C. Chang. 2001. From transforming growth factor- $\beta$  signaling to androgen action: Identification of Smad3 as an androgen receptor coregulator in prostate cancer cells *En: PNAS* **98** (6): 3018-3023

Kassel, O y P. Herrlich. 2007. Crosstalk between the glucocorticoid receptor and other transcription factors: Molecular aspects. Review *En: Molecular and Cellular Endocrinology* **275**: 13–29

Kon, A., L. Vindevoghel, D. J. Kouba, Y. Fujimura, J. Uitto y A. Mauviel. 1999. Cooperation between SMAD and NF- $\kappa$ B in growth factor regulated type VII collagen gene expression *En: Oncogene* **18**: 1837-1844

Krakowski, A., J. Laboreau, A. Mauviel, M. Bissel y K. Luo. 2005. Cytoplasmic SnoN in normal tissues and nonmalignant cells antagonizes TGF- $\beta$  signaling by sequestration of the Smad proteins *En: Proceedings of the National Academy of Sciences* **102** (35): 12437-12442

Kurisaki, K., A. Kurisaki, U. Valcourt, A. Terentiev, K. Pardali, P. Dijke, C. Heldin, J. Ericsson y A. Moustakas. 2003. Nuclear Factor YY1 Inhibits Transforming Growth Factor  $\beta$ - and Bone Morphogenic Protein-Induced Cell Differentiation *En: Molecular and Cellular Biology* **23** (13): 4494-4510

Lacerete, A., J. Korah, M. Roy, X. Yang, S. Lemay y J. Lebrun. 2008. Transforming growth factor- $\beta$  inhibits telomerase through SMAD2 and E2F transcription factors *En: Cellular Signaling* **20**: 50-59

Lachman, David. 2001. Transcriptional gene regulation in Eukaryotes. *En els.net*

Lässig, M. 2007. From biophysics to evolutionary genetics: statistical aspects of gene regulation. Review *En: BioMedical Center Bioinformatics* **8**: S7

Levy, S. y S. Hannenhalli. 2002. Identification of transcription factor binding sites in the human genome sequence *En: Mammalian Genome* **13**: 510-514

Lewin, B. 2008. *Genes IX*. Ed. Jones and Bartlett Publishers, 9a ed., EUA, 892 pp.

- Li, G., S. Wang y T. Gelehrter. 2003. Identification of Glucocorticoid Receptor Domains Involved in Transrepression of Transforming Growth Factor- $\beta$  Action *En: The Journal of Biological Chemistry* **278** (43): 41779-41778
- Lin, J., S. Martin, L. Xia y J. Gorham. 2005. TGF- $\beta$ 1 Uses distinct Mechanisms to inhibit IFN- $\gamma$  Expression in CD4<sup>+</sup> T Cells at Priming and at Recall: Differential Involvement of Stat4 and T-bet *En: The Journal of Immunology* **174**: 5950-5958
- Liu, D., B. Black y R. Derynck. 2001. TGF- $\beta$  inhibits muscle differentiation through functional repression of myogenic transcription factors by Smad3 *En: Genes and Development* **15**: 2950-2966
- Liu, X, Y. Sun, R. Weinberg y H. Lodish. 2001. Ski/Sno and TGF- $\beta$  signalling. Mini Review *En: Cytokine and Growth Factor Reviews* **12** (2001) 1–8.
- Long, F., H. Liu, C. Hahn, P. Sumazin, M. Zhang y A. Zilberstein. 2004. Genome-wide prediction and analysis of function-specific transcription factor binding sites *En: In Silico Biology* **4**: e0033
- Loots, G. e I. Ovcharenko. 2004. rVista 2.0: evolutionary analysis of transcription factor binding sites *En: Nucleic Acids Research* **32**: W217-W221.
- López-Rovira, T., E. Chalaux, J. Rosa, R. Bartons y F. Ventura. 2000. Interaction and Functional Cooperation of NF- $\kappa$ B with Smads *En: The Journal of Biological Chemistry* **272** (37): 28937-28946
- Luo, K. 2004. Ski and SnoN: Negative regulators of TGF- $\beta$  signaling *En: Current Opinion in Genetics and Development* **14**: 65-70
- Macías-Silva, M., W. Li, J. Leu, M. A. Crissey y R. Taub. 2002. Up-regulated Transcriptional repressors SnoN and Ski Bind Smad Proteins to antagonize Transforming Growth Factor beta Signaling during Liver Regeneration *En: The Journal of Biological Chemistry* **277** (32): 28483-28490
- Macías-Silva, M. y M. Briones-Orta. 2008. SNO (SKIL) *En: Targeted Protein Database (TPdb), Current Biodata Ltd. En prensa.*
- Macleod, K. 2000. Tumor suppressor genes *En: Current opinion in Genetics and Development* **10**: 81-93.
- Magner, L. 2002. *A History of the Life Sciences*, 3a ed., Ed. Marcel Dekker, NYC.
- Massagué J., D. Wotton. 2000. Transcriptional control by the TGF $\beta$ /Smad signaling system *En: The European Molecular Biology Organization Journal* **19** (8): 1745-1754
- Miyazono, K., H. Suzuki y T. Imamura. 2003. Regulation of TGF- $\beta$  signaling and its role in progression in tumors. Review. *En: Cancer Science* **94** (3): 230-234.

Morishita, R. J. Higaki, N. Tomita y T. Ogihara. 1998. Application of Transcription Factor "Decoy" Strategy as Jeans of Gene Therapy and Study of Gene Expression in Cardiovascular Disease *En: Circulation Research* **82**: 1023-1028

Moustakas, A. 2002. Smad signaling Network *En: Journal of Cell Science* **115**: 3355-3356

Moustakas, A., K. Pardali, A. Gaal y C. H. Heldin. 2002. Mechanisms of TGF $\beta$  Signaling in regulation of cell grow and differentiation *En: Immunology Letters* **82**: 82-91

Moustakas, A., Souchelnytski, S. y C.H. Heldin. 2001. Smad regulation in TGF $\beta$  signal transduction *En: Journal of Cell Science* **114**: 4359-4369

Nerlov, C. 2007. The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control. Review *En: TRENDS in Cell Biology* **17** (7): 318-324

Oft, M., J. Pelf, C. Rudaz, H. Schwarz, H. Beug y E. Reichmann. TGF- $\beta$ 1 and Ha-Ras collaborate in modulating the phenotypic plasticity and invasiveness of epithelial tumor cells *En: Genes and Development* **10**: 2462-247

Pardoux, C., X. Ma, S. Gubert, S. Pellegrini, P. Mayeux, F. Gay, G. Trinchiteri y S. Chouaib. 1999. Downregulation of Interleukin-12 (IL-12) Responsiveness in Human T Cells by Transforming Growth Factor- $\beta$ : Relationship With IL-12 Signaling *En: Blood* **93** (5): 1448-1445

Pearson-White, S., D. Deacon, R. Crittenden, G. Brady, N. Iscove y P. Quesenberry. 1995. The *ski/sno* Protooncogene Family in Hematopoietic Development *En: Blood* **86**: 2146-2155

Pearson-White, S. y M. McDuffie. 2003 Defective T-Cell Activation Is associated with Augmented Transforming Growth Factor $\beta$  Sensitivity in Mice with Mutations in the *Sno* Gene *En: Molecular and Cellular Biology* **23** (15): 5446–5459

Pedaries, V., F. Verrecchia, S. Michel y A. Mauviel 2003. Retinoic acid receptors interfere with the TGF- $\beta$ /Smad signaling pathway in a ligand-specific manner *En: Oncogene* **22**: 8212-8220

Qingwei Zhu, Q., S. Pearson-White y K. Luo. 2005. Requirement for the SnoN Oncoprotein in Transforming Growth Factor  $\beta$ -Induced Oncogenic Transformation of Fibroblast Cells *En: Molecular and Cellular Biology* **25** (24): 10731–10744.

Rahimi, R. y E. Leof. 2007. TGF- $\beta$  Signaling: A Tale of Two Responses *En: Journal of Cellular Biochemistry* **102**: 593–608

Rico, D., Vaquerizas, J., Dopazo, J. y Boscá, L. 2007. Identification of conserved domains in the promoter regions of nitric oxide synthetase 2: implications for the species – specific transcription and evolutionary differences *En: BMC Genomics* **8**: 271.

Rodríguez-Caso C., M. A. Medina y R. V. Solé. 2005. Topology, tinkering and evolution of the human transcription factor network *En: Federation of European Biochemical Societies* **272**: 6423-6432

Rouyun Tn., X. Zhang, J. Yang, Y. Liu y Y. Li. 2007. Molecular Basis for the Cell Type-Specific Induction of SnoN Expresión by Hepatocyte Growth Factor *En: Journal of the American Society of Nephrology* **18**: 2340 -2349

Sakaguchi, M., T. Nukui, H. Sonegawa, H. Murata, J. Futami, H. Yamada y N. Huh. 2005. Targeted disruption of transcriptional regulatory function of p53 by a novel efficient method for introducing a decoy oligonucleotide into nuclei *En: Nucleic Acids Research* **33** (9): e88

Sandelin, A., P. Carninci, B. Lenhard, J. Ponjavic, Y. Hayashizaki y D. Hume. 2007. Mammalian RNA polymerase II core promoters: insights from genome-wide studies *En: Nature Reviews Genetics* **8**: 436-424

Shinagwa, T., H-D. Dong, M. Xu, T. Maekawa y S. Ishii. 2000. The *sno* gene, which encodes a component of the histone deacetylase complex, acts as a tumor suppressor in mice *En: The EMBO Journal* **19** (10): 2280 – 2291.

Smale, S. 2005. Promoters *En: els.net*

Shi, Y., Y. Wang, L. Jayaraman, H. Yang, J. Massague, y N. Pavletich, N. 1998. Cristal Structure of a Smad MH1 domain Bound to DNA: Insights of DNA binding in TGF $\beta$  Signaling *En: Cell* **94**: 585–594

Stopa, M., D. Anhuf, L. Terstegen, P. Gatsios, Ax. Gressner y S. Dooley. 2000. Participation of Smad2, Smad3, and Smad4 in Transforming Growth Factor  $\beta$  (TGF- $\beta$ )-induced Activation of Smad7 *En: The Journal of Biological Chemistry* **275** (38): 29308-29317.

Stroschein S., W. Wang, S. Zhou, Q. Zhou y K. Luo. 1999. Negative Feedback Regulation of TGF- $\beta$  Signalling by the SnoN Oncoprotein *En: Science* **286** (5440): 771-774

Subramaniam, N., G. Leong, T. Cock, J. Flanagan, C. Fong, J. Eisman y A. Kouzmenko. 2001. Cross-talk between 1,25Dihydroxyvitamin D<sub>3</sub> and Transforming Growth Factor- $\beta$  Signaling Requires Binding of VDR and Smad3 Proteins to Their Cognate DNA Recognition Elements *En: The Journal of Biological Chemistry* **276** (19): 15741-15746.

Takaesu, N., C. Hyman-Walsh, Y. Ye, R. Wisotzkey, M. Stinchfield, M. O'Connor, D. Wotton y S. Newfield. 2006. dSno Facilitates Baboon signaling in the Drosophila Brain by Switching the Affinity of Medea away From Mad and Toward dSmad2 *En: Genetics* **174**: 1299-1313

Tang-Hu, X. 2008. TGF $\beta$ - mediated formation of pRb-E2F complexes in human myeloid leukemia cells. Review *En: Biochemical and Biophysical Research communications* **369**: 227-280

Traylor, A., T. Hock y N. Hill-Kapurczak. 2007. Specific protein 1 and Smad-dependent regulation of human heme oxygenase-1 gene by transforming growth factor- $\beta$ 1 in renal epithelial cells *En: American Journal of Physiology – Renal Physiology* **293**: F885-F894

Tsuritani K, T. Irie, R. Yamashita, Y. Sakakibara, H. Wakaguri, A. Kanai, J. Mizushima-Sugano, S. Sugano, K. Nakai y Y. Suzuki Y. 2007. Distinct class of putative "non-conserved" promoters in humans: comparative studies of alternative promoters of human and mouse genes *En: Genome Research* **17** (7):1005-14.

Zhang, M. 2007 Computational analyses of eukaryotic promoters. Review *En: Bio Med Central Bioinformatics*, **8** (6):S3

Zhu, Q., S. Pearson-White y K. Luo. 2005. Requirement for the SnoN Oncoprotein in Transforming Growth Factor  $\beta$ -induced Oncogenic Transformation of Fibroblast Cells *En: Molecular and Cellular Biology* **25** (24): 10731-10744

Wakaguri H., R. Yamashita, Y. Suzuki, S. Sugano y K. Nakai. 2008. DBTSS: database of transcription start sites, progress report 2008 *En: Nucleic Acids Research* **16** (Database issue)

Wan, Y., X. Liu y M. Kirschner. 2001. The Anaphase-Promoting Complex Mediates TGF- $\beta$  Signaling by Targeting SnoN for Destruction *En: Molecular Cell* **8**: 1027-1039.

Weisz, L., M. Oren y V. Rotter.T 2007. Transcription regulation by mutant p53. Review *En: Oncogene* **26** (15): 2202- 2211.

Wilkinson D. S., W. W. Tsai, M. A. Schumacher, y M. C. Barton. 2008. Chromatin-Bound p53 Anchors Activated Smads and the mSin3A Corepressor To Confer Transforming Growth Factor  $\beta$ -Mediated Transcription Repression *En: Molecular and Cellular Biology* **28** (6): 1988-1998.

Wrighton K. H., M. Liang, B. Bryan, K. Luo, M. Liu, X. H. Feng y X. Lin. 2007. Transforming growth factor-beta-independent regulation of myogenesis by SnoN sumoylation *En: Journal of Biological Chemistry* **282** (9): 6517-24

Ya-Li Yao, B. Dupont, S. Ghosh, Y. Fang, R. Leach y E. Seto. 1998. Cloning, chromosomal localization and promoter analysis of the human transcription factor YY1 *En: Nucleic Acid Research* **26** (16): 3776–3783