

INSTITUTO DE QUÍMICA
Universidade de São Paulo
20.653

"Caracterização de novos genes humanos envolvidos no processo de regulação da expressão de genes homeóticos"

Tese de Doutorado

Diana Noronha Nunes

Orientação: Profa. Dra. Mari Cleide Sogayari
Departamento de Bioquímica
Instituto de Química da Universidade de São Paulo

SÃO PAULO
2004

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA

**“Caracterização de novos genes humanos
envolvidos no processo de regulação da expressão
de genes homeóticos”**

Diana Noronha Nunes

Tese de Doutorado

Orientadora: Prof. Dra. Mari Cleide Sogayar

SÃO PAULO

Data do Depósito do Trabalho na SPG: 28/07/2004

(Vinte e oito de Julho de dois mil e quatro)

Ficha Catalográfica
Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

Nunes, Diana Noronha
N972c Caracterização de novos genes humanos envolvidos no
processo de regulação da expressão de genes homeóticos / Diana
Noronha Nunes. - São Paulo, 2004.
143p

Tese (doutorado) - Instituto de Química da Universidade de
São Paulo. Departamento de Bioquímica.

Orientador : Sogayar, Mari Cleide

1. Genoma : Biologia molecular 2. Expressão gênica
3. Regulação gênica I. T. II. Sogayar, Mari Cleide,
orientador.

574.88 CDD

“Caracterização de novos genes humanos envolvidos no processo de regulação da expressão de genes homeóticos”

DIANA NORONHA NUNES

Tese de Doutorado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários à obtenção do grau de Doutor em Ciências – Área: Bioquímica.

Aprovado(a) por:

Profa. Dra. MARI CLEIDE SOGAYAR
IQ – USP
(Orientadora e Presidente)

Profa. Dra. GLÁUCIA MENDES SOUZA
IQ – USP

Prof. Dr. EDUARDO MORAES REGO REIS
IQ – USP

Profa. Dra. MARIA APARECIDA NAGAI
FM – USP

Prof. Dr. ROGER CHAMMAS
FM – USP

SÃO PAULO
03 DE SETEMBRO 2004

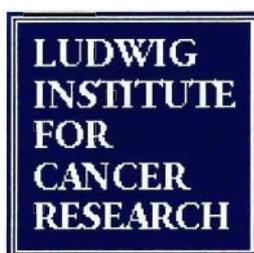
"This suspense is terrible. I hope it will last"

Oscar Wilde em seu conto "The importance of being Earnest".

“Ao meu querido Emmanuel, fundamental em todos os momentos. Por seu carinho, paciência e apoio. Por seu exemplo de integridade como pessoa e pesquisador”

À minha família por sua presença, carinho e apoio

Este trabalho contou com o apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (**FAPESP**) (Processo: 99/12508-0), da “International Union Against Cancer” (UICC), do Ludwig Institute for Cancer Research, Uppsala/Suécia, e foi desenvolvido pela Pós-Graduação do Departamento de Bioquímica, do Instituto de Química da USP.



AGRADECIMENTOS

À Dra. Mari Cleide Sogayar, por sua generosidade ao me acolher no seu laboratório e permitir a realização deste trabalho. Por seu entusiasmo com a Ciência e seu exemplo de imensa perseverança e empenho;

À Dra. Anamaria Aranha Camargo pelo pontapé inicial dado neste trabalho. Por sua disponibilidade e ajuda;

Ao Dr. Roger Chammas, por ter aberto as portas de seu laboratório e pelas agradáveis “divagações” sobre este projeto;

Aos meus colegas do Laboratório de Biologia Celular e Molecular do Instituto de Química da USP: Ana Paula, André, Antero, Áurea, Carlos, Christian, Cléber, ElizaBeth, Fernanda Festa, Fernanda Ortis, Fernando, Gabriela, Jae, Juan, Juliane, Karin, Katlin, Leonardo, Letícia, Lúcia, Luciana Cruz, Luciana Gomes, Marcos, Mariana, Marina, Mário, Patrícia, Renato, Rita, Roberto, Sheila, Sílvia, Tatiana, Thery e Wagner;

À Débora, Irenice, Sandra e Zizi pelo trabalho técnico e por estarem sempre disponíveis a ajudar;

Aos amigos Cristiano, Mário e Katlin. Vocês fazem muita falta no Laboratório;

A querida amiga Maki, por sua ajuda, seu exemplo de humildade e competência;

À querida Helena, por ser sempre tão prestativa e amiga;

A Ana Carla pelas conversas e discussões;

À Dahabada, pelos bons momentos dentro e fora do Laboratório e pela grande ajuda com a purificação de proteína;

Ao querido Chris, por seu apoio, carinho e por nossas discussões “acaloradas”. Por sua disponibilidade e ajuda imprescindíveis. Por dividir sua inteligência, entusiasmo e alegria com a Ciência.

À Ana Paula, por sua amizade, seu bom humor e seu exemplo de dedicação e empenho;

Ao LIM 27: Camila, Carolina, Cíntia, Daniel, Elida, Fábio, Sheila, e Dr. Wagner Gattaz, pela amizade e por estarem sempre disponíveis a ajudar;

À Elida, por sua amizade e por ser sempre tão solícita profissional e pessoalmente;

Ao Dr. Paulo Oliveira, pela ajuda com as análises de bioinformática;

Ao Dr. Carl-Henrik Heldin, pela oportunidade de colaboração com o “Ludwig Institute for Cancer Research”, em Uppsala, Suécia e pelas discussões científicas;

Ao Dr. Aristidis Moustakas pela inesquecível oportunidade de minha estada em Uppsala. Por sua competência, amizade, disponibilidade, paciência, humildade e por ter me possibilitado um grande aprendizado científico e humano.

Aos grupos “Signaling in TGF- β ”: Aristidis, Anita, Katerina, Marcin, Ulrich e “Integration Signaling in TGF- β ”: Hiroyuki, Ivan, Ihor, Maria, Marta, Serhiy, Taras, Yevhen. Pelos ótimos momentos de convivência no laboratório;

Aos amigos Federica, Jeannette, Maite, Marta, Sami, Trias e Ulrich. Pelos ótimos momentos no Ludwig Institute e nossas inenarráveis “Festas de arromba”;

Ao Airton, Cibele, Emiliano, Marcelo e Milton da Secretária de Pós-Graduação do Instituto de Química da USP;

À FAPESP pelo financiamento de minha bolsa de doutorado;

À UICC, pelo financiamento de minha bolsa sanduíche em Uppsala.

PREFÁCIO

Esta tese de doutorado foi organizada da seguinte maneira:

Parte I – Introdução, Objetivos, Resultados, Discussão: Nesta parte é apresentada uma introdução ao conceito dos genes do grupo dos Polycomb (PcG) e uma descrição de como estas proteínas atuam na repressão da transcrição gênica. Também é mostrada a relação existente entre os genes PcG, o controle de proliferação celular e o câncer. Da mesma maneira, apresentamos o papel das proteínas do grupo dos Trithorax (trxG) e como estas se opõem aos PcG, ativando a transcrição gênica. Um terceiro grupo, o das proteínas “Enhancers of Polycomb and trithorax” (ETP), que associa ambas as características dos PcG e dos trxG, também é descrito. A relação entre o tema das proteínas PcG com essa tese de doutorado se deu durante o Projeto Genoma Humano do Câncer (HCGP), uma iniciativa do Instituto Ludwig para o estudo do Câncer (LICR) e da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Foram geradas mais de 1 milhão de ESTs derivadas de tecidos tumorais e normais, e utilizando o banco de dados das ESTs geradas no HCGP, foi possível identificar 2 ESTs que mostraram similaridade com o gene *de Drosophila melanogaster* da família dos ETP, o gene “Enhancer of Polycomb” (*Epc*). A partir destas seqüências, nasceram os objetivos desta tese, que consistiram inicialmente na utilização destas ESTs como ponto de partida para a caracterização do ortólogo humano deste gene. Este gene foi escolhido, devido ao interesse particular do grupo em estudar genes/proteínas que fossem relacionados com regulação de transcrição epigenética, e pudessem estar relacionados com a tumorigênese e a progressão tumoral. Os resultados de clonagem dos dois ortólogos humanos do gene *EPC*, *hEPC1* e *hEPC2*, são mostrados, assim como sua análise de expressão em tecidos normais e tumorais. Diante da descrição do gene *hEPC1* pelo grupo de Shimono *et al*, 2001, optamos por aprofundar nossos estudos no gene de *hEPC2*. São apresentados resultados de análise de domínios protéicos, similaridade do gene entre diferentes espécies e identificação e caracterização de variantes de “splicing” de *hEPC2*. A parte inicial deste

trabalho está incluída em uma publicação do HCGP, onde *hEPC1* e *hEPC2* são apresentados como exemplos de genes humanos que foram clonados utilizando ESTs geradas neste projeto. Um segundo manuscrito, contendo uma análise mais detalhada de isoformas de splicing, evolução dos genes EPC humanos e correlação entre câncer algumas isoformas transcricionais de *hEPC2* se encontra submetido para a publicação na revista Gene, e anexado a esta tese.

Parte II - Introdução, Resultados e Discussão da Interação de hEPC2 com as proteínas SMADs, e sua modulação por TGF- β : Nesta parte do trabalho é apresentada a hipótese da possível interação entre as proteínas hEPC2 e as SMADs. Durante a clonagem do gene *hEPC2*, duas ESTs descritas como da patente WO 9855512, apresentaram similaridade com *hEPC2*. Estas seqüências foram obtidas através de ensaios de duplo híbrido, e são descritas como polipeptídeos parciais, que interagem com proteínas da família SMAD. As SMADs são uma família de fatores de transcrição, que são receptoras e propagadoras da transdução de sinal da citocina TGF- β (fator de crescimento tumoral). A partir deste dado foi montada uma colaboração com o grupo do Dr. Aristidis Moustakas, chefe do grupo de sinalização de TGF- β (TSL) do Ludwig Institute for Cancer Research (Uppsala, Suécia), com o objetivo de confirmar a interação entre hEPC2 e as SMADs. Desta forma, são mostrados os experimentos de validação das construções de vetores de expressão constitutiva em mamífero e dos anticorpos policlonais anti-EPC2, reagentes fundamentais para a investigação em questão. Em seguida são mostrados os resultados dos experimentos de co-imunoprecipitação de hEPC2 e das SMADs 1,2, 3, 4, 5, 6, 7 e 8 e sua modulação por TGF- β 1. Também são mostrados os resultados da localização celular de hEPC2 por imunofluorescência indireta, assim como de investigação da co-localização de hEPC2 e SMAD2 e sua modulação por TGF- β 1. Por fim são apresentados os resultados de ensaio de gene repórter realizado com o objetivo de investigar se a transcrição das SMADs seria dependente de hEPC2. Com o objetivo de confirmar a interação entre hEPC2 e as SMADs através de outra metodologia, a seqüência codificante de *hEPC2* foi clonada no vetor pGEX de expressão indutível em

bactérias para sua utilização em ensaios de “pull-down”. São mostrados os resultados das várias tentativas de super-expressão de 4T1-pGEX-hEPC2 induzidas por IPTG. Diante das falhas seguidas na expressão deste construção de *hEPC2*, foi feita uma sub-clonagem de *hEPC2* para o vetor de expressão indutível em bactérias, pET-32a. Os resultados da expressão de pET-32a-hEPC2 são mostrados, assim como a purificação da proteína recombinante em condições desnaturantes. Em seguida todos os resultados são discutidos. Para a confirmação da interação entre hEPC2 e as SMADs 2, 3, 4, 7 e 8 ainda são necessários vários experimentos, que ainda deverão ser feitos em colaboração com o grupo sueco.

ÍNDICE

ABREVIATURAS

LISTA DE FIGURAS

LISTA DAS TABELAS

RESUMO

ABSTRACT

PARTE I – CLONAGEM E CARACTERIZAÇÃO DOS GENES *hEPC1* E

<i>hEPC2</i>	1
INTRODUÇÃO	1
O Projeto genoma humano do câncer como ferramenta de descoberta de novos genes humanos	1
Mecanismos da memória celular transcricional	2
Repressão por complexos de PcG	4
Os genes <i>Trithorax</i> e a ativação de transcrição	5
O grupo dos genes “Enhancers of trithorax and Polycomb”	6
Repressão pelas proteínas PcG: da memória celular ao controle de proliferação e o câncer	7
<i>Hematopoese</i>	9
Alvos que conectam a repressão pelos PcGs ao controle de proliferação celular	12
Desregulação de PcG no câncer humano	13
O Gene “Enhancer of Polycomb” de <i>Drosophila</i>	15
OBJETIVOS DO TRABALHO	17
MATERIAIS	18
Reagentes	18

Soluções	18
MÉTODOS	19
Clonagem do cDNA de <i>hEPC1</i> e da isoforma canônica de <i>hEPC2</i> por RT-PCR	19
Caracterização das variantes Iso-B e Iso-C e análise quantitativa por RT-PCR	20
Ensaio de Northern-blot	21
<i>Preparo de sonda de DNA radiotiva</i>	21
Análises de expressão gênica por Dot-blot	22
Sequenciamento de DNA	22
Análises de bioinformática e bancos de dados	23
Western-blot	24
<i>Extração e dosagem de proteínas totais</i>	24
<i>Transferência úmidadas proteínas para membrana de nitrocelulose</i>	25
<i>Imunoreação</i>	25
Imunoprecitação	26
Condições de cultura e manutenção de linhagens celulares	26
Clonagem em vetores plasmidiais	27
Transfecção celular	29
<i>Por cloreto de cálcio</i>	29
<i>Transfecção utilizando o reagente Lipofectamina</i>	29
Infecção celular utilizando adenovírus	30
Ensaio de gene repóter	32
Imunofluorescência	33
Expressão da construção 4T1 pGEX-hEPC2	33

Expressão e purificação da proteína de fusão His-hEPC2	34
RESULTADOS	38
<i>Estratégia para a obtenção da sequência completa dos cDNAs de hEPC1 e hEPC2</i>	38
Estrutura genômica de <i>hEPC1</i> e <i>hEPC2</i>	44
Análise dos domínios protéicos de <i>hEPC2</i>	48
Análise da similaridade da proteína <i>hEPC2</i> entre diferentes espécies	49
Análise da expressão de <i>hEPC1</i> e <i>hEPC2</i>	50
Identificação de variantes de “splicing” de <i>hEPC2</i>	52
Análise de bioinformática e identificação da isoforma D de <i>hEPC2</i>	56
DISCUSSÃO	60
Comparação da expressão de <i>hEPC1</i> e <i>hEPC2</i> em diferentes tecidos humanos	64
O ortólogo humano de <i>hEPC1</i>	69
<i>hEPC1</i> e <i>hEPC2</i> como membros do grupo “Enhancer of Polycomb and Trithorax Proteins” (ETP)	71
O uso de um sítio alternativo de poli-adenilação em Iso-D	75

ANEXO 1: Artigo publicado no PNAS: “**Contribution of 700,000 ORF sequence tags to the definition of the human transcriptome**”. Camargo *et al*, 2001

ANEXO 2: MANUSCRITO SUBMETIDO À PUBLICAÇÃO: “**Cloning and Characterization of the human *EPC2* gene, an ortholog of *Drosophila EPC* (Enhancer of Polycomb)**”

PARTE II - INTERAÇÃO ENTRE hEPC2 E AS SMADS E SUA	
MODULAÇÃO POR TGF-β	81
INTRODUÇÃO	81
As proteínas SMAD como sinalizadoras da via de TGF- β	81
RESULTADOS	84
Interação entre hEPC2 e as SMADs	84
co-Imunoprecipitação (IP) de hEPC2	85
Super-expressão de pCXN2-hEPC2 e 6X myc pCDNA3-hEPC2	89
Investigação da interação de hEPC2 e SMADs por IP.....	90
Investigação da localização celular de hEPC2	98
Investigação se hEPC2 afeta a transcrição dependente de SMAD	101
Super expressão das construções pGEX- hEPC2	103
Super expressão e purificação da proteína recombinante pETA- hEPC2	110
DISCUSSÃO	113
Interação de hEPC2 com as proteínas SMADs e modulação do complexo	
por TGF-β	113
Transcrição dependente de SMADs	116
Localização nuclear de hEPC2	119
CONCLUSÕES	120
BIBLIOGRAFIA	124
<i>CURRÍCULUM VITAE</i>	141

ABREVIATURAS

36B4	Fosfoproteína PO ribossomal ácida
AMV	"Avian Myeloblastosis Virus"
Asx	gene "Additional sex combs" de <i>Drosophila</i>
ATP	Trifosfato de adenosina
BBS	"BES-buffered saline"
bp	"base pair"; pares de base
BSA	"Bovine Serum Albumin"; soro albumina bovina
Bmi1	gene de camundongo ortólogo ao gene <i>Psc</i> de <i>Drosophila</i>
BMI1	gene humano ortólogo ao gene <i>Psc</i> de <i>Drosophila</i>
BMP	"bone morphogenetic protein"
B-NHLs	linfomas de células B não-Hodgkin
<i>Brm</i>	gene "Brahma" de <i>Drosophila</i>
cDNA	DNA complementar
Da	Dalton
dCTP	Trifosfato de desoxicitosina
ddNTP	didesoxinucleotídeo
DMEM	"Dulbecco's Modified Eagle's Medium"
DMSO	Dimetilsulfóxido
DNA	Ácido desoxirribonucléico
DNase	Desoxiribonuclease
dNTP	desoxinucleotídeo
DTT	Ditiotreitol
EDTA	Ácido etilenodinitrilotetracético sal dissódico
<i>Eed</i>	gene "embryonic ectoderm development" de camundongo
<i>EED</i>	gene ortólogo humano de <i>Eed</i> de camundongo
<i>Enx1</i>	gene de camundongo ortólogo ao gene <i>E(z)</i> de <i>Drosophila</i>
<i>Enx2</i>	gene de camundongo ortólogo ao gene <i>E(z)</i> de <i>Drosophila</i>
<i>EPC</i>	gene "Enhancer of Polycomb" de <i>Drosophila</i>
<i>EPL1</i>	gene de levedura ortólogo ao gene <i>EPC</i> de <i>Drosophila</i>

<i>Esc</i>	gene "extra sex combs" de <i>Drosophila</i>
EST	"Expressed Sequence Tag"
ETP	grupo dos genes "Enhancer of Trithorax and Polycomb"
<i>E(z)</i>	gene "Enhancer of Zeste" de <i>Drosophila</i>
<i>EZH1</i>	gene humano ortólogo ao gene <i>Enx1</i> de camundongo
<i>EZH2</i>	gene humano ortólogo ao gene <i>Enx2</i> de camundongo
FBS	"Fetal bovine serum"; soro fetal bovino
FITC	isoticianato de fluoresceína
GAPDH	Gliceraldeído 3-fosfato desidrogenase
GST	glutathiona-S-transferase
HAT	"histone acetyltransferase"; histona acetil transferase
HCGP	Projeto Genoma do Câncer Humano
HDAC	"histone deacetylase"; histona deacetilase
<i>hEPC1</i>	gene humano ortólogo ao gene "Enhancer of Polycomb" de <i>Drosophila</i>
<i>hEPC2</i>	gene humano ortólogo ao gene "Enhancer of Polycomb" de <i>Drosophila</i>
Hox	genes homeobox
<i>HPC1</i>	gene humano ortólogo ao gene <i>Pc</i> de <i>Drosophila</i>
<i>HPC2</i>	gene humano ortólogo ao gene <i>Pc</i> de <i>Drosophila</i>
<i>HPC3</i>	gene humano ortólogo ao gene <i>Pc</i> de <i>Drosophila</i>
IF	ensaio de imunofluorescência indireta
IgG	imunoglobulina G
IP	ensaio de imunoprecipitação
IPTG	β -D-tiogalactopiranosídeo de isopropila
Iso-A	Isoforma A, ou transcrito canônico do gene <i>hEPC2</i>
Iso-B	Isoforma B do gene <i>hEPC2</i>
Iso-C	Isoforma C do gene <i>hEPC2</i>
Iso-D	Isoforma D do gene <i>hEPC2</i>
Kbp	Kilopares de base
kDa	Kilo Daltons
LB	"Luria-Bertani Medium"

LBCM	Laboratório de Biologia Celular e Molecular, Instituto de Química da USP
LLA	Leucemia linfóide aguda
LMC	Leucemia mielóide crônica
LML	"mixed lineage leukemia"; Leucemia de linhagem mixta ou leucemia bifenotípica
LOH	"loss of heterozigosity"; perda de heterozigose
M33	o mesmo que HPC1
<i>Mel 18</i>	gene de camundongo ortólogo ao gene <i>Psc</i> de <i>Drosophila</i>
<i>MEL 18</i>	gene humano ortólogo ao gene <i>Psc</i> de <i>Drosophila</i>
<i>Mph1/rae28</i>	gene de camundongo ortólogo ao gene <i>Ph</i> de <i>Drosophila</i>
<i>Mph2</i>	gene de camundongo ortólogo ao gene <i>Ph</i> de <i>Drosophila</i>
MOPS	"3-[N-Morpholino] propanosulfonic acid"
<i>Mor</i>	gene "Moira" de <i>Drosophila</i>
mRNA	RNA mensageiro
NSCLC	"non-small cell lung cancer"; câncer de pulmão de células grandes
OD	"Optic density"
ORESTES	"Open Reading Frame Expressed Sequence Tags"
PAGE	"Polyacrilamide electrophoresis gel", eletroforese em gel de poliacrilamida
pb	pares de base
PBSA	Solução salina de PBS ("Phosphate Buffered Saline") sem cálcio e magnésio
<i>Pc</i>	gene " Polycomb" de <i>Drosophila</i>
<i>Pcl</i>	gene " Polycomblike" de <i>Drosophila</i>
PCR	"Polymerase Chain Reaction"
PcG	grupo de genes da família dos Polycomb
PcGi	complexo Polycomb de iniciação
PcGm	complexo Polycomb de manutenção
PEV	efeito posicional de variação

pfu	unidade formadora de partícula viral
<i>Ph</i>	gene "Polyhomeotic" de <i>Drosophila</i>
<i>Pho</i>	gene "Phehomeotic" de <i>Drosophila</i>
PM	Peso Molecular
PMSF	"Fenil Metil Sulfonil Fluoride"
PRC-1	o mesmo que PcGm
PRE	"Polycomb responsive elements"; elementos responsivos à Polycomb
<i>Psc</i>	gene "Posterior sex combs" de <i>Drosophila</i>
rDNA	DNA ribossomal
RNA	Ácido ribonucléico
Rnase	Ribonuclease
RPM	rotações por minuto
RT-PCR	"Reverse Transcription Polimerase Chain Reaction"
SBE	"SMAD binding elements"; elementos responsivos a SMAD
<i>Sce</i>	gene " <i>Sex combs extra</i> " de <i>Drosophila</i>
<i>Scm</i>	gene "sex comb on midleg" de <i>Drosophila</i>
<i>SCMH1</i>	gene humano ortólogo do gene "sex comb on midleg" de <i>Drosophila</i>
SDS	"Sodium Dodecyl Sulfate"
SMAD	"Small Mothers Against Dodecaplegic"
SSC	Solução salina contendo Citrato de sódio, pH 7,0
Su(var)	genes supressores de variegação
<i>Su(z)2</i>	gene "supressor 2 of zeste" de <i>Drosophila</i>
<i>sxc</i>	gene "super sex combs" de <i>Drosophila</i>
TBS	Tampão Tris-salina, pH 7,5
TBST	Tampão Tris-salina-Tween, pH 7,5
TGF- β	"Transforming Growth Factor Beta"
TP	"Terminal Protein"; proteína terminal
trxG	grupo de genes da família dos trithorax
<i>Trl</i>	gene "Trithoraxlike" de <i>Drosophila</i>
TRITC	isotiocianato B de tetrametilrodamida

TS	"TGF- β signaling group"; grupo de sinalização de TGF- β , Uppsala/Suécia
UTR	"Untranslated region"
UV	Luz ultravioleta
X-gal	5- bromo, 4-cloro, 2-indolil, β -D-galactopiranos
YY1	gene humano ortólogo ao gene "pleiohomeotic" de <i>Drosophila</i>

Os nomes dos genes se encontram em itálico no texto

LISTA DE FIGURAS

Figura 1 – Representação das proteínas componentes dos complexos PcGi e PcGm e sua atuação na repressão da transcrição gênica	8
Figura 2 - Representação das sequências nucleotídicas utilizadas na montagem do transcrito completo de <i>hEPC2</i>	41
Figura 3 - Representação das sequências nucleotídicas utilizadas na montagem do transcrito completo de <i>hEPC1</i>	43
Figura 4 – Alinhamento mostrando os domínios de <i>EPc</i> em diferentes espécies	46
Figura 5 – Domínios preditos pelo CDD para a seqüência de aminoácidos de <i>hEPC2</i>	48
Figura 6 – Filograma mostrando a similaridade entre a seqüência de aminoácidos dos homólogos de <i>EPc</i> em diferentes espécies	49
Figura 7 – Análise de expressão de <i>hEPC1</i> e <i>hEPC2</i>	51
Figura 8 - Northern-blot de tecidos normais ensaiados para a análise de Iso-B e Iso-C	54
Figura 9 - RT-PCR semi-quantitativo de Iso-B e Iso-C	55
Figura 10 - Northern-blots de tecidos normais e linhagens celulares tumorais sondadas para <i>hEPC2</i>	59
Figura 11 - Representação da cobertura de contigs de ORESTES em relação aos cDNAs dos genes <i>hEPC1</i> e <i>hEPC2</i>	60

Figure 12 – Estrutura genômica das isoformas A, B C and D do gene humano <i>hEPC2</i>	74
Figura 13 – Transdução do sinal de TGF- β pelas proteínas SMADs	83
Figura 14 – Mapa do vetor de expressão constitutiva pCDNA3 e amplificação do inserto de <i>hEPC2</i> clonado em pCDNA3-6Xmyc	86
Figura 15 –Mapa do vetor de expressão indutível em bactéria pGEX e clones de 4T1-pGEX-hEPC2 digeridos com <i>EcoRI</i> e <i>XhoI</i>	87
Figura 16 – Mapa do vetor pCXN2 e clones de pCXN2-hEPC2 digeridos com <i>Sall</i>	88
Figura 17 - Super-expressão das construções pCDNA3-6Xmyc-hEPC2 e pCXN2-hEPC2, co-transfectados transitoriamente em células 293-T.....	90
Figura 18 - Super-expressão das construções pCDNA3-6Xmyc-hEPC2 e SMADs 2, 3, 4 e 7 em células 293-T	92
Figura 19 - Super-expressão das construções pCDNA3-6Xmyc-hEPC2 e SMADs 1, 5, 6 e 8 em células 293-T	93
Figura 20 – Análise das proteínas totais provenientes das células 293-T transfectadas com as construções pCDNA3-6Xmyc-hEPC2 e infectadas com os adenovírus com as construções das SMADs 2, 3, 4 e 7.....	94
Figura 21 – Análise das proteínas totais provenientes das células 293-T transfectadas com os construções pCDNA3-6Xmyc-hEPC2 e infectadas com os adenovírus com as construções das SMADs 1, 5, 6 e 8	95
Figura 22 - Super-expressão das construções pCXN2-hEPC2 e SMADs 2, 3, 4 e 7 em células 293-T	97

Figura 23 – Ensaio de imunofluorescência indireta mostrando a presença da proteína super-expressa pCXN2-hEPC2 no núcleo de células Cos-7, tratadas e não tratadas com TGF- β 1	99
Figura 24 – Ensaio de imunofluorescência indireta mostrando a presença da proteína super-expressa pCDNA3-SMAD2 no núcleo e citoplasma de células Cos-7, tratadas e não tratadas com TGF- β 1	100
Figura 25 – Ensaio de gene repórter onde a construção pCXN2-hEPC2 foi co-transfectada nas células HEPG2 com a construção CAGA-12	102
Figura 26 – “Blotting” da construção pGEX-hEPC2 utilizando o antisoro α -EPC2-II	104
Figura 27 – “Blotting” da construção pGEX-hEPC2 utilizando o antisoro α -GST	105
Figura 28 – Géis de SDS-poliacrilamida a 8% corados com Commassie Brilliant Blue, mostrando clones 4T1 pGEX-hEPC2 em diferentes condições de indução por IPTG	107
Figura 29 – Mapa do vetor pET32A e clones de pET32A-hEPC2 digeridos com <i>EcoRI</i> e <i>XhoI</i>	109
Figura 30 – Gel 8% de SDS-Poliacrilamida corado por Coomassie Brilliant Blue, mostrando a indução da expressão dos clones de pET32-hEPC2 com 1mM IPTG durante 6 horas	110

Figura 31 – Ensaio de “Western-blot” mostrando a detecção da proteína de fusão pET32-hEPC2 pelo antisoro α -EPCII	111
Figura 32 – Purificação da proteína recombinante pET32-hEPC2 sob condições desnaturantes	112

LISTA DE TABELAS

Tabela 1 – Iniciadores utilizados na amplificação do cDNA completo do gene <i>hEPC2</i>	40
Tabela 2 – Iniciadores utilizados na amplificação do cDNA completo do gene <i>hEPC1</i>	42
Tabela 3 – Limites de exon/ítron e estrutura genômica do gene <i>hEPC1</i>	45
Tabela 4 – Limites entre os exons /íntrons e estrutura genômica do gene <i>hEPC2</i>	45
Tabela 5 - Limites entre os íntrons/exons e estrutura genômica de Iso-B do gene <i>hEPC2</i>	53
Tabela 6 - Limites entre os íntrons/exons e estrutura genômica de Iso-C do gene <i>hEPC2</i>	53
Tabela 7 – Comparação entre o tamanho dos exons de <i>hEPC1</i> e <i>hEPC2</i>	64

RESUMO

A identidade na segmentação do corpo de diversos organismos, durante o desenvolvimento, é devida, em grande parte, à ação das proteínas homeóticas. Em especial, dois grupos de proteínas, Trithorax (*trxG*) e Polycomb (*PcG*) têm um papel fundamental na manutenção, respectivamente, da ativação e da repressão da transcrição gênica, associando-se à cromatina.

A importância das *PcG* nos estimulou a buscar a caracterização das proteínas humanas ortólogas ao "Enhancer of Polycomb" (*Epc*) de *Drosophila*, até então não descritas no genoma humano. Para tanto, buscamos: - obter a sequência completa e mapear o cDNA do novo gene humano homólogo ao "Enhancer of Polycomb" de *Drosophila*; - analisar sua expressão em tecidos fetais, adultos e tumorais e fazer estudos buscando sua caracterização funcional. Encontramos, mapeamos e obtivemos a sequência completa de dois genes humanos, ortólogos de *Epc1* (10p11-22) e de *Epc2* (2q21-23) de camundongo, publicando estes dados em 2001 (Camargo *et al.*, 2001).

Ambos os genes são bastante conservados entre várias espécies, sendo que o cDNA de *hEPC2* humano, por exemplo, é 94% idêntico ao *Epc2* de camundongo e possui 96% de identidade ao nível de proteína, sugerindo que a função do gene deve ter sido mantida durante a evolução. No entanto, as seqüências protéicas de *hEPC1* e *hEPC2* humanos possuem apenas 68% de identidade entre si. Portanto, é provável que após a duplicação dos parálogos, estes tenham divergido funcionalmente. A expressão de ambos os genes foi avaliada utilizando "dot-blots" contendo 76 mRNAs de amostras de tecidos fetais, adultos e tumorais, mostrando-se fraca e ubíqua. Análises *in silico* sugeriram a existência de 4 isoformas de *splicing* para *hEPC2*, as quais foram validadas por RT-PCR ou "Northern blots". Uma das isoformas (de 2.7 Kpb) se mostrou mais abundante em todas as linhagens tumorais estudadas através de análises de "Northern blot", principalmente nas linhagens de linfoma de Burkitt's Raji e na linhagem de leucemia pró-mielocítica HL-60. Esta isoforma é gerada através de um sítio alternativo de poli-adenilação, que reduz sua porção 3'UTR, retirando 4 dos 5 "elementos ricos em adenilatos e uridilatos" (AREs),

envolvidos com a degradação de mRNAs lábeis que codificam proteínas regulatórias. Estes resultados se encontram em um manuscrito recentemente submetido à publicação (anexo à tese).

Interação entre hEPC2 e SMADs e sua modulação por TGF- β

Durante a montagem da seqüência completa de *hEPC2*, verificamos que duas ESTs patenteadas mostravam alta identidade com o gene. Estas seqüências foram descritas como sendo parte de uma nova proteína de interação com as proteínas da família SMAD, envolvidas com transdução de sinais desencadeados por TGF- β . Esta citocina por sua vez, regula a proliferação, diferenciação e morte celular. Partimos para a avaliação da possível interação entre *hEPC2* e as SMADs, em colaboração com o grupo do Dr. Aristidis Moustakas, do Ludwig Institute for Cancer Research de Uppsala, Suécia.

Os resultados de co-imunoprecipitação sugeriram que as SMADs 2, 3, 4, 7 e 8 interagem com *hEPC2*, sendo que a interação entre as SMAD2, SMAD3, SMAD4 e *hEPC2* nas células tratadas com TGF- β 1, mostraram uma redução na co-imunoprecipitação. Este resultado sugere que TGF- β 1 modula negativamente a interação entre essas proteínas. Da mesma maneira, foi observada uma redução na interação de *hEPC2* com SMAD8 após o tratamento com BMP-7. Esse resultado é ainda mais destacado para as SMADs 2 e 3. Estes dados foram observados para ambas as construções de *hEPC2*, o que sugere fortemente a veracidade da interação entre estas proteínas.

A localização celular de *hEPC2*, e também sua co-localização com SMAD2 foram investigadas através de imunofluorescência indireta e confirmaram a predição do programa PSORTII, de que *hEPC2* se localiza no núcleo. No entanto, não foi possível observar a co-localização entre *hEPC2* e SMAD2.

É possível que *hEPC2* não se ligue diretamente ao DNA, necessitando se associar como parceiro de um fator de transcrição. Esta foi uma das hipóteses para a atuação de *hEPC2*, como um co-fator que se associe com

uma das SMADs e se ligue a um elemento específico de ligação a SMAD (SBE). Para investigar essa hipótese um ensaio de gene repórter foi feito utilizando uma construção de um repórter contendo 12 repetições da seqüência CAGA (seqüência específica de ligação das SMADs 2, 3 e 4) fusionado com o gene da luciferase. No entanto, este ensaio não demonstrou que a transcrição de SMAD2 é dependente de hEPC2 e o experimento deverá ser repetido.

Para confirmar a interação entre hEPC2 e as SMADs, será feito um experimento de "pull-down". Para tal o cDNA de hEPC2 foi clonado no vetor pET-32A de expressão indutível em bactérias. A proteína recombinante já foi produzida, tendo sido induzida e posteriormente purificada em condições desnaturantes.

Apesar de dezenas de genes PcG terem sido caracterizados em *Drosophila*, poucos destes genes foram estudados em mamíferos. Portanto, a descrição do gene *hEPC2* e seus transcritos alternativos, contribui para o conhecimento de PcG humanos, indicando a associação de maior expressão de uma de suas isoformas em linhagens celulares tumorais. Em relação à interação de hEPC2 com as SMADs, é interessante observar que nenhuma outra proteína foi descrita por possuir a particularidade de interagir com as SMADs de diferentes categorias. Talvez este seja um dado importante, que indique o papel singular de hEPC2 na sinalização de TGF- β 1.

ABSTRACT

The identity of body segmentation in several organisms during development is, to a large extent, due to the action of the homeotic proteins. In particular, two groups of proteins, the Trithorax (*trxG*) and Polycomb (PcG), have a major role in maintenance of respectively, transcription activation and repression, when associated to the chromatin.

The importance of PcGs has motivated us to pursue the isolation and characterization of two new human proteins that are orthologs of the "Enhancer of Polycomb" (*EPC*) of *Drosophila*. To achieve this goal we undertook the task of the cloning and mapping of complete cDNA sequence of the novel genes *hEPC1* and *hEPC2*, analyzing its expression in fetal, adult and tumoral tissues and functionally characterizing the *hEPC2* protein. In 2001, we published the mapping and cloning of the complete cDNA sequences of both genes, as being orthologs of the mouse *Epc1* (10p11-22) and *Epc2* (2q21-23), together with the strategy used to obtain the full-length cDNAs (Camargo *et al.*, 2001).

Both genes are shown to be highly conserved among several species. Thus, the human *hEPC2* cDNA is 94% identical to the mouse *Epc2* and displays 96% identity at the protein level, suggesting maintenance of its function during the evolution. However, the protein sequences of the human *hEPC1* and *hEPC2* display only 68% identity. Therefore, it is likely that they have undergone a functional divergence after their duplication. The expression of both genes was evaluated using "dot-blot" containing 76 mRNAs samples from fetal, adult and tumoral tissues and is shown to be weak and ubiquitous. "In silico" analysis suggested the existence of 4 *hEPC2* splicing isoforms that were validated by RT-PCR and/or Northern-blot. One of the isoforms (of 2.7 Kbp) is shown to be more abundant in all of the tumoral cell lines evaluated using Northern-blot analysis, mainly in the Burkitt's Raji lymphoma and in the promyelocytic leukemia HL-60. This isoform results from the use of an alternative polyadenylation site that reduces the 3'UTR, abolishing 4 of 5 "adenylates and urilates rich elements" (AREs), involved in the degradation of labile mRNAs that codify to regulatory proteins. These results have been recently submitted to publication (manuscript attached to this thesis).

Interaction between the hEPC2/SMADs and its modulation by TGF- β

During the assembly of the *hEPC2* full-length cDNA sequence, we found two patented ESTs that tagged a portion of the gene. These sequences were described as partial sequences of a “new SMAD interacting protein”, involved in signal transduction of TGF- β , a cytokine that regulates cell proliferation, differentiation and death. To evaluate this putative interaction between hEPC2 and the SMADs proteins, we begun a collaboration with the TGF- β signalling group of the Dr. Aristidis Moustakas, from the Uppsala Ludwig Institute for Cancer Research, Sweden.

The results of co-immunoprecipitation assays suggested that SMADs 2, 3, 4, 7 e 8 interact with hEPC2. Moreover, the interaction among SMAD2, SMAD3, SMAD4 and hEPC2 in cells treated with TGF- β 1 showed decreased co-immunoprecipitation. This result suggests that TGF- β 1 negatively modulates the interaction of these proteins. Likewise, we observed a reduction in hEPC2 interaction with SMAD8 upon BMP-7 treatment. This effect was even more dramatic for SMADs 2 and 3. These data were observed for both hEPC2 plasmid constructs, which strongly suggest the veracity of these proteins interaction.

The cell localization of the hEPC2 protein, as well as its co-localization with the SMAD2, were investigated through indirect immunofluorescence assay, confirming the predicted localization of hEPC2 in the cell nucleus using the PSORTII program. However, we were not able to confirm the co-localization of hEPC2 and SMAD2.

It is possible that hEPC2 does not bind directly to the DNA, requiring an association with a partner such as a transcription factor. This raises the hypothesis of hEPC2 having a role as a co-factor associated to one of the SMADs and binding to a “SMAD binding element” (SBE). To investigate this hypothesis, gene reporter assays were undertaken using a reporter construct containing 12 CAGA sequence repetitions (specific binding sequence of the SMADs 2, 3 and 4) fused to the luciferase gene. However, this assay could not demonstrate that the transcription of the SMAD is dependent on hEPC2. This experiment must be repeated.

To confirm the interaction of hEPC2 and SMADs, a pull-down assay will be performed. To this end, the coding region of hEPC2 was cloned into the pET-32A bacterial inducible expression vector. The recombinant protein was already produced, having been induced and purified under denaturing conditions.

Despite the dozens of PcG genes that were described in *Drosophila*, only a few of these genes have been characterized in mammals. Therefore, the description of the hEPC2 and its alternative transcripts is a contribution to better knowledge of the human PcGs. Regarding the hEPC2 and SMADs interaction, it's it is noteworthy that this is the first protein described to interact with SMADs of distinct categories. This may be an important indication of a unique role for hEPC2 in the TGF- β 1 signaling pathway.

PARTE I – CLONAGEM E CARACTERIZAÇÃO DOS GENES

hEPC1 E hEPC2

INTRODUÇÃO

O Projeto genoma humano do câncer como ferramenta de descoberta de novos genes humanos

Estamos observando uma transição significativa na área das pesquisas biológicas. Deixamos a era da genética clássica e clonagem posicional, para entrar na era do sequenciamento e análise de genomas completos em larga escala. Nessa nova era, o grande desafio é o aproveitamento racional e eficiente do número cada vez maior de seqüências armazenadas em bancos de dados. Em outras palavras, é preciso elucidar de forma eficiente a função desses novos genes, sua interação dentro das células, buscando compreender a organização do genoma como um todo.

O sequenciamento completo do genoma humano foi finalizado em 2001 e disponibilizado nos bancos de dados públicos (Lander *et al.*, 2001). Paralelamente, uma grande quantidade de ESTs representando genes humanos expressos em diferentes tecidos está sendo disponibilizada diariamente por diferentes grupos (Imanishi *et al.*, 2004; Ota *et al.*, 2004; Strausberg *et al.*, 2002; Strausberg *et al.*, 2001). Uma das formas mais eficientes de se explorar esses dados têm sido através do que chamamos de genômica comparativa. Em outras palavras, com ferramentas de bioinformática, busca-se aproveitar os dados gerados nos projetos genomas de organismos modelo como, por exemplo, levedura, *C. elegans* e *Drosophila*, para inferir a função biológica do novo gene. Estima-se por exemplo que um terço dos genes humanos associados a doenças tenham homólogos no genoma da levedura.

O Projeto Genoma do Câncer (HCGP), idealizado e inicialmente desenvolvido no Instituto Ludwig de Pesquisas sobre o Câncer, foi baseado na descoberta de novos genes humanos através do estudo de tumores de grande importância epidemiológica no Brasil. O HCGP foi baseado na técnica ORESTES (Open Reading Frame ESTS) (Dias Neto *et al.*, 2000), onde a informação dos genes, transcrita nas moléculas de mRNA, é acessada em sua porção codificadora central. Com mais de 1 milhão de ESTs (“expressed sequence tags”) geradas a partir de diferentes tecidos tumorais (mama, cólon, estômago, cabeça e pescoço, testículo entre outros) e tecidos normais, vários novos genes humanos foram amostrados pela primeira vez pelas ESTs deste projeto. Vários dos genes representados por estas ESTs, são candidatos à participação na tumorigênese e/ou na progressão tumoral, seja através de sua simples expressão, de alterações na sequência de DNA, ou através da alteração no nível de expressão gênica (Brentani *et al.*, 2003; Camargo *et al.*, 2001).

A exploração dos dados gerados pelo Projeto Genoma Humano do Câncer para a descoberta de novos genes humanos foi o objetivo principal que norteou esta tese de doutorado. De nosso interesse particular, foi a caracterização de novos genes envolvidos no processo de regulação epigenética, mais especificamente genes associados a regulação da expressão de genes homeóticos. Em uma busca inicial no banco de dados do projeto, selecionamos duas ESTs correspondentes a fragmentos de dois novos genes humanos, similares ao gene de *Drosophila* “Enhancer of Polycomb” que faz parte do grupo das proteínas dos Polycomb (PcG). Como estas ESTs foram gerados dentro do HCGP (Camargo *et al.*, 2001), eles também poderiam se relacionar com a tumorigênese e a progressão maligna.

Mecanismos da memória transcricional celular

A estabilidade do estado diferenciado de um organismo é a “pedra fundamental” do desenvolvimento. Alternativamente, falhas no padrão de expressão transcricional das células (que caracteriza qual o tipo celular que compõe determinado tecido) geram problemas desde a embriogênese, até o

desenvolvimento de câncer. Organismos com tecidos diferenciados precisam estabelecer padrões de expressão tecido-específicos no início do desenvolvimento e manter estes padrões por toda sua vida. Desta maneira, a viabilidade de um organismo requer que os mecanismos de manutenção da diferenciação sejam robustos.

Muitas decisões no destino das células são feitas ao nível transcricional, em resposta a estímulos durante o desenvolvimento. Alguns fatores regulatórios essenciais são expressos de maneira transitória, no entanto o padrão de expressão determinado por estes fatores necessita ser mantido. A manutenção do estado transcricional é obtida através de distintos fatores, como por exemplo, o padrão de expressão de genes homeóticos em *Drosophila melanogaster*, que necessita ser precisamente e clonalmente mantido durante o desenvolvimento, enquanto os fatores de transcrição que estabelecem esse padrão são expressos apenas durante as primeiras horas da embriogênese. A manutenção do padrão de expressão necessita de um segundo conjunto de fatores, chamados genes dos grupos Polycomb (PcG) e Trithorax (trxG) (Francis & Kingston, 2001).

A manutenção do sistema dos PcG e trxG está evolutivamente conservada desde *Drosophila* até mamíferos. Eles estão relacionados com a regulação de muitos genes-alvo além dos genes homeóticos, indicando que a manutenção epigenética de estados de ativação ou repressão deve ser um mecanismo fundamental no desenvolvimento. Os primeiros membros da família dos PcG foram identificados há 50 anos atrás, existindo pelo menos 14 membros identificados até hoje (Pirrotta, 1998). As proteínas codificadas por estes genes parecem agir em grandes complexos que reprimem a expressão gênica.

Os genes trxG foram identificados como supressores do fenótipo gerado pelos PcG (Kennison, 1995). Muitas proteínas trxG também participam em complexos regulatórios que agem ao nível da cromatina para manter a expressão gênica ativa e contrabalançar a repressão causada pelos PcG. Alguns genes possuem ambos os fenótipos dos PcG e trxG, sendo, portanto, uma grande simplificação a definição do estado determinado, onde as

proteínas trxG mantêm apenas a ativação, e as proteínas PcG apenas a repressão (Brock & van Lohuizen, 2001).

Fatores tais como, o grande grupo de proteínas PcG e trxG, o tamanho dos complexos que os contêm, e a natureza da cromatina aonde eles atuam, dificultam o entendimento do mecanismo de ação destas proteínas. No entanto, uma explosão das pesquisas em estrutura de cromatina e proteínas modificadoras de cromatina nas últimas décadas, revelou a importância destes fatores na regulação gênica. Isso também levou a um entendimento detalhado da estrutura básica da cromatina – o nucleossomo. Muitos complexos que modificam a estrutura da cromatina têm sido caracterizados, alguns dos quais estão envolvidos na manutenção dos PcG/trxG.

Infelizmente, os mecanismos de memória transcricional ainda não são claramente compreendidos. Existe informação limitada sobre a composição e as atividades desempenhadas pelos complexos protéicos. Um quadro de entendimento que vem se formando mostra que memória transcricional surge de uma combinação de diferentes mecanismos, incluindo a modificação de histonas, a alteração na posição e conformação dos nucleossomos, e a atuação específica de complexos de multi-subunidades na regulação gênica. Estes mecanismos contribuem diretamente para determinar o estado de ativação ou repressão de um gene (Francis & Kingston, 2001).

Repressão por complexos de PcG

Estudos sobre a função dos PcG se focaram em, como os complexos destas proteínas são corretamente dirigidos para genes que devem permanecer reprimidos, e como estes complexos medeiam a repressão.

O padrão de expressão de genes regulados pelas proteínas PcG, combinado com a ampla expressão das proteínas PcG, implica que outros mecanismos altamente seletivos se associem aos PcG em seu papel de mediar a repressão de transcrição. Existem dois critérios para o estabelecimento desta repressão. Primeiro, o recrutamento destas proteínas requer a presença de seqüências de DNA conhecidas como elementos responsivos a Polycombs (PRE: “Polycomb responsive elements”). A função

das seqüências PRE é provavelmente recrutar altas concentrações destas proteínas, necessárias para nuclear a estrutura repressiva da cromatina, de maneira similar a outros elementos regulatórios *cis* que recrutam seus fatores de ligação cognatos. Segundo, a repressão dos PcG parece necessitar de um gene silenciado como molde. Esta propriedade permite que o sistema dos PcG funcione como um sistema de memória que sinalize eventos regulatórios anteriores, como no caso dos genes de segmentação. Estes últimos genes são reguladores transcricionais que estabelecem padrões de expressão iniciais de genes homeóticos. Desta maneira, genes ativos não são reprimidos por PcG e permanecem ativos, enquanto que genes silenciados são estavelmente reprimidos por estas mesmas proteínas. Portanto, em transgenes contendo seqüências regulatórias PRE, a repressão parece ser estabelecida apenas nas células em que o “enhancer” não está ativo antes da ativação da função dos PcG nos embriões. Desta maneira, os padrões de expressão dos “enhancers” precoces são mantidos durante todo o desenvolvimento (Poux *et al.*, 1996; Poux *et al.*, 2001a).

O modelo mais simples que explica como as proteínas PcG reconhecem os estados transcricionalmente silenciados, diz que as características moleculares da cromatina em genes reprimidos são compatíveis com, ou estabilizam a montagem de complexos de proteínas PcG. Diferenças moleculares entre genes transcritos versus não transcritos incluem proteínas associadas, modificações de histonas, remodelagem de nucleossomos e compactação da cromatina, podendo todos estes mecanismos, afetar a ligação de proteínas nos nucleossomos. Alternativamente, o estado padrão dos PRE deve permitir a formação dos complexos de proteínas PcG, mas características de genes ativos podem prevenir sua montagem ou desestabiliza-los (Francis & Kingston, 2001) .

Os genes *Tritorax* e a ativação de transcrição

Tem sido postulado que os *trxG* contenham fatores que funcionem em diversos níveis para manter a expressão gênica. No entanto, várias proteínas *trxG* têm sido ligadas diretamente a ativação transcricional (Kennison, 1995), e

proteínas atuando em outros níveis ainda não foram identificadas. Existem três proteínas trxG nos complexos de remodelamento de cromatina: SWI/SNF Brahma (Brm), Moira (Mor) e Osa. Outra proteína trxG, GAGA, também está implicada na modulação da estrutura da cromatina. O homólogo humano de TRAP-230, Kohtalo, é um membro do complexo “mediator-like-co-activator” (Treisman, 2001), indicando uma possível ligação direta desta proteína com a maquinaria transcricional. Outras proteínas trxG residem em complexos distintos, mas suas atividades ainda não foram bem caracterizadas. No entanto, muitas destas proteínas como Kismet, Trithorax (Trx), ou discos homeóticos Ash1 e Ash2 estão associados com a cromatina (Chinwalla *et al.*, 1995; Daubresse *et al.*, 1999; Rozovskaia *et al.*, 1999). Estes dados indicam que um importante componente de ativação pelos trxG, semelhante à repressão dos PcG, ocorra ao nível da cromatina. Em princípio, este antagonismo funcional entre os dois grupos de genes pode, no entanto, envolver efeitos opostos da estrutura da cromatina.

Assim como no caso das proteínas PcG, é previsível que os complexos de trxG tenham atividades diversas que sejam sinérgicas na manutenção do estado ativado, mas a maioria destas atividades ainda não foi identificada (Francis & Kingston, 2001).

O grupo dos genes “Enhancers of trithorax and Polycomb”

Mutações em dois genes PcG, *E(z)* e *Asx* (“Additional sex combs”), exibem transformações nos membros anteriores e posteriores, de maneira consistente com um papel de ativação e repressão de *loci* homeóticos (LaJeunesse & Shearn, 1996; Milne *et al.*, 1999; Sinclair *et al.*, 1992). Um estudo recente varreu 70% de deleções no genoma de *Drosophila* que apresentaram a habilidade de tanto aumentar como suprimir fenótipos de mutações em trxG (Gildea *et al.*, 2000). Surpreendentemente, mutações em seis genes PcG aumentaram o fenótipo de mutações de trxG, sugerindo que um sub-grupo de genes PcG é necessário para a ativação, assim como para a supressão da expressão de genes homeóticos. Os autores sugerem que estes seis genes sejam renomeados como “Enhancers of trithorax and Polycomb”

(ETP), deixando na categoria de PcG apenas sete outros genes. No entanto, as evidências genéticas apresentadas acima não levam em consideração a possibilidade de que interações entre o grupo dos ETP, PcG e trxG em genes homeóticos seja indireta. Testes anteriores de interações genéticas entre as proteínas PcG (Campbell *et al.*, 1995; Cheng *et al.*, 1994) não mostrou evidência independente para a divisão funcional dos PcG e do grupo das proteínas ETP, já que mutações nos genes destes grupos aumentaram o número de alterações entre eles.

Repressão pelas proteínas PcG: da memória celular ao controle de proliferação e o câncer.

Uma ampla variedade de vias moleculares regula a resposta das células a estímulos intra e extracelulares, levando as células a se diferenciarem, proliferarem, pararem de se dividir ou entrarem em apoptose. No fim, todos estes caminhos levam a ativação ou repressão de genes específicos que irão induzir a resposta celular correta. O câncer é o resultado de um defeito nestes mecanismos moleculares, levando as células a reagirem inapropriadamente, proliferando quando e onde não deveriam. Tem se tornado claro nos últimos anos, que o “meio ambiente” da cromatina possui um papel crucial na regulação da atividade ou inatividade transcricional dos genes. Por esta razão, talvez não seja surpreendente que trabalhos sobre fatores regulatórios da cromatina relacionados ao controle de crescimento e tumorigênese venham se acumulando rapidamente. As proteínas dos grupos dos PcG e trxG possuem um papel crucial na regulação da atividade gênica ao nível da cromatina (Brock & van Lohuizen, 2001), (Ringrose & Paro, 2001), (Francis & Kingston, 2001).

À parte da heterogeneidade na composição dos complexos de proteínas PcG devido à expressão tecido-específica, pelo menos dois tipos distintos de complexos de PcG com diferentes funções/genes-alvos podem ser evidenciados tanto em *Drosophila* quanto em mamíferos (Jones *et al.*, 1998; Ng *et al.*, 2000; Sewalt *et al.*, 1998; Shao *et al.*, 1999; Tie *et al.*, 2001; van Lohuizen *et al.*, 1998). O primeiro complexo contém as proteínas EED, Enx1/EZH2 e Enx2/EZH1, que são os ortólogos de mamífero, respectivamente,

dos genes “extra sex combs” (*Esc*) e “enhancer of zeste” (*Ez*). Este complexo, chamando PcGi, parece ser necessário no início do desenvolvimento na “fase de iniciação”, quando a repressão das proteínas PcG assume o comando dos repressores atuantes iniciais, como a proteína “hunchback” em *Drosophila*. O segundo complexo também chamado complexo de manutenção (PcGm) ou PRC-1, contém muitas proteínas PcG como Bmi-1, Mel18, Mph1, M33, Ring1 e mpc2 e parece ser necessário na manutenção estável da repressão tanto em moscas, quanto em mamíferos. Baseado em descobertas recentes, um modelo foi proposto onde os repressores iniciais, como “hunchback”, recrutam o complexo PcGi para reprimir os genes-alvo, possivelmente via complexo de deacetilase dMi2 (Kehle *et al.*, 1998; van Lohuizen, 1999; Wade *et al.*, 1998; Zhang *et al.*, 1998). Subseqüentemente, o complexo PcGi, através de sua associação com as histona deacetilases, modificam a cromatina local, e via uma interação transitória recruta o complexo PcGm, resultando na manutenção de repressão (Poux *et al.*, 2001b; Tie *et al.*, 2001; van der Vlag & Otte, 1999) (Figura 2). A modificação epigenética local da cromatina pode fornecer uma explicação de como o silenciamento pelos PcG é fielmente re-estabelecido após cada replicação do DNA, onde a maioria destas proteínas se dissocia da cromatina (Buchenau *et al.*, 1998; Voncken *et al.*, 1999).

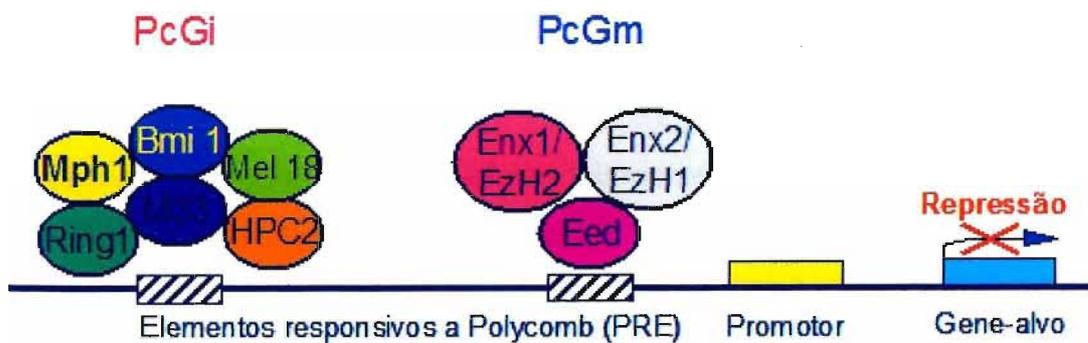


Figura 2 – Representação das proteínas componentes dos complexos PcGi e PcGm e sua atuação na repressão da transcrição gênica.

Hematopoese

A diferenciação de células hematopoéticas é regulada por fatores extracelulares como as citocinas e intracelulares como os fatores de transcrição. As decisões tomadas durante o desenvolvimento implicam se estes fatores devem ser mantidos durante as próximas gerações, enquanto que, ao mesmo tempo um grupo pequeno de células-tronco precisa ser continuamente mantido. Um conjunto de dados indica que a repressão causada pelos PcG contribui para isto e está envolvida com a regulação da hematopoese normal (Jacobs & van Lohuizen, 2002).

A expressão dos genes PcG durante a diferenciação celular da hematopoese humana parece ser finamente regulada (Lessard *et al.*, 1998), (Raaphorst *et al.*, 2000b), (Raaphorst *et al.*, 2001), (Lessard *et al.*, 1999). Enquanto EED é expresso em níveis iguais em células primitivas e maduras da medula humana, a expressão de *Bmi1* é mais alta nas células de medula mais primitivas e gradualmente declina durante a diferenciação. Em contraste, a maioria dos outros genes PcG incluindo *M33*, *mel18*, *HPH1* e *ENX1/EZH2* são pouco expressos ou não detectados nos precursores mais precoces, mas sua expressão é marcadamente aumentada com a diferenciação das células de medula (Sauvageau *et al.*, 1994).

Também durante a diferenciação das células T do timo, as proteínas BMI1 e EZH2 são diferencialmente expressas (Raaphorst *et al.*, 2001). Apesar de uma menor população de células duplamente negativas para CD4⁻ CD8⁻ na camada sub-capsular e no córtex expressarem *Bmi1*, mas não *EZH2*, a maioria expressa ambos os genes. Em contraste, a maioria dos timócitos duplamente positivos para CD4⁺ e CD8⁺ no córtex, expressam altos níveis de *EZH2*, mas nenhum ou quase nenhum *Bmi1*, indicando que a progressão dos estágios duplos negativos para os duplos positivos coincide com a perda da expressão de *Bmi1*. Nos estágios positivos únicos de CD4⁺ CD8⁺ ou CD4⁻ CD8⁻, que é a fase final da diferenciação dos timócitos, *Bmi1* não é expresso e *EZH2* é expresso em baixos níveis.

Os complexos EED/EZH2 PcGi e Bmi1/RING1 PcGm não imuno-precipitam ou imuno-localizam nos núcleos em intérfase, e além disso, eles mostram padrões de expressão mutuamente exclusivos em células humanas periféricas maduras B e T (Raaphorst *et al.*, 2000b). Células que não estão se dividindo nos folículos de manto de tonsilas humanas expressam Bmi1 e RING1 mas não EED/EZH2. No entanto, quando células B são induzidas a proliferar, a expressão de *Bmi1/RING1* é perdida e *EED/EZH2* aparece (Fukuyama *et al.*, 2000; Peytavi *et al.*, 1999; Raaphorst *et al.*, 2000b). Ao contrário, a diferenciação de centroblastos em centrócitos é acompanhada pelo re-aparecimento de *Bmi-1/RING1* e a perda de *EED/EZH2* (Raaphorst *et al.*, 2000b). Essa diminuição na regulação de *EZH2* também é observada durante a diferenciação *in vitro* de células pró-mielocíticas HL-60. Além disso, RNAs antisenses de *EZH2* suprimem a síntese de DNA em células HL-60 (Fukuyama *et al.*, 2000). Em contraste, a estimulação *in vitro* de linfoma de células do manto (MCL) com IL-10 causa um aumento na regulação de *EZH2*, e super-expressão de *EZH2* em células Ramos aumenta a sua taxa de proliferação (Visser *et al.*, 2001). Padrões de expressão similares para *Bmi-1* e *EZH2* são observados em células T periféricas, com células T que não estão se dividindo expressando *Bmi1*, mas não *EED/EZH2*, e blastos em divisão expressando *EED/EZH2* mas não *Bmi1* (Fukuyama *et al.*, 2000; Peytavi *et al.*, 1999; Raaphorst *et al.*, 2001). A ausência da expressão do complexo BMI1 PcGm em células T em divisão se correlaciona com a indução da expressão do grupo de genes *HoxB* sob a ativação da proliferação de células T (Care *et al.*, 1994). Além disso, baseado na rápida indução de antígeno-receptor em células B, os genes PcG de camundongo *Mel18*, *Bmi-1*, *M33* e *Rae28/Mph1* foram caracterizados como genes de resposta precoce imediata (Hasegawa *et al.*, 1998).

Apesar da composição molecular de complexos PcG em células hematopoéticas permanecer por ser determinada, estas análises de expressão indicam que complexos PcG de diferentes composições são formados em diferentes estágios de diferenciação hematopoética na medula óssea, timo e linfócitos de sangue periférico. Estes experimentos correlatos sugerem um papel para as proteínas PcG no desenvolvimento e proliferação de precursores

de medula humana, timócitos e células B e T. No entanto, a evidência mais convincente para esse papel essencial dos genes PcG na hematopoese, provêm do fenótipo de camundongos mutantes de genes PcG. Todos os camundongos mutantes de *Bmi1*, *Mel18*, *M33*, *Mph1* e *Eed* mostraram severas anormalidades hematopoéticas. Além de transformações no esqueleto e retardo no crescimento, mutantes de camundongo de *Bmi1*, *Mel18* e *M33* sofrem severa hipoplasia no baço e timo (Akasaka *et al.*, 1996; Akasaka *et al.*, 1997; Core *et al.*, 1997; Tokimasa *et al.*, 2001; van der Lugt *et al.*, 1994). Apesar de mutantes de células T de *Mph1* parecerem maturar quase normalmente, camundongos deficientes de *Bmi1*, *Mel18* e *M33* apresentam uma forte redução no número total de células T. As populações de linfócitos T mais maduras de duplo-positivos CD4⁺ CD8⁺ e único-positivos CD4⁺ CD8⁻ ou CD4⁻ CD8⁺ são mais severamente afetadas, causando uma relativa super-representação de populações de células T duplo-negativas CD4⁻ CD8⁻. Mutantes de camundongo de *Bmi1*, *Mel18*, *M33* e *Mph1* são também novamente defectivos em células B em desenvolvimento, com uma redução completa no número de células, que se torna mais severa com um aumento na maturação de células B (Akasaka *et al.*, 1996; Akasaka *et al.*, 1997; Core *et al.*, 1997; Tokimasa *et al.*, 2001; van der Lugt *et al.*, 1994). No entanto, nem todos os estágios de maturação de células B e T podem ser encontrados em camundongos, embora em número reduzido, sugerindo uma falha destas células em proliferar e/ou uma diminuição na taxa de sobrevivência, ao contrário de um bloco completo de diferenciação. Em concordância com isso, é observada uma resposta proliferativa defeituosa sob a ação de vários mitógenos, especialmente a citocina IL-7. A resposta mitótica e a diferenciação de precursores de células linfóides sob a estimulação de IL-7, são cruciais para a geração de um número apropriado de linfócitos maduros (Peschon *et al.*, 1994; von Freeden-Jeffry *et al.*, 1995). Apesar da sinalização à frente do receptor de IL-7 parecer intacta em mutantes de linfócitos de PcG, eles falham ao entrar em fase S sob a estimulação mitogênica com IL-7 (Akasaka *et al.*, 1997; Core *et al.*, 1997; Takihara *et al.*, 1997; van der Lugt *et al.*, 1994). Isso sugere que as proteínas PcG provavelmente agem ainda mais à frente, possivelmente ao nível dos genes-alvo *STAT*.

Interessantemente, enquanto que *Bmi1*, *Mel18*, *M33* e *Mph1*, todos componentes do complexo PcGm, são aparentemente reguladores positivos da proliferação hematopoética, *EED*, presente no complexo PcGi, tem um papel claramente oposto. Enquanto que camundongos *EED* *-/-* morrem na gastrulação, camundongos *EED* *+/-* apresentam hiper-proliferação de progenitores da medula óssea primitiva e madura, e um aumento da susceptibilidade ao desenvolvimento de tumores hematopoéticos (Faust *et al.*, 1995; Lessard *et al.*, 1999). Inter cruzamento de camundongos mutantes para *Bmi1* e *EED* revelaram uma interação genética tipo celular-específica (Lessard *et al.*, 1999). *Bmi1* parece ser dominante sob *EED* no controle da proliferação de células de medula primitivas, já que mutações em *EED* não afetaram a proliferação destas células em camundongos *Bmi1* *-/-*. No entanto, as mutações em *EED* trans-complementaram a deficiência de *Bmi1* em células de medula pré-B em camundongos *Bmi1* *+/-*. Estes dados sugerem que a proliferação de células hematopoéticas depende da contribuição relativa de complexos que não contenham *EED* e positivo em complexos contendo *Bmi1*.

Combinado com a expressão diferencial célula e tecido-específica e formação de diferentes complexos PcG, é provável que o equilíbrio entre *Bmi1* e *Mel-18* nos complexos PcG seja crítico na especificidade da atuação sob o gene-alvo e no controle de proliferação de células hematopoéticas (Jacobs & van Lohuizen, 2002).

Alvos que conectam a repressão pelos PcGs ao controle de proliferação celular

Apesar dos padrões de expressão desregulados em sub-grupos de genes *Hox* serem, em sua maioria, a causa mais provável para as transformações homeóticas vistas em diferentes PcG mutados em camundongos, para defeitos de hematopoese, a causa parece ser menos clara e uniforme (Akasaka *et al.*, 1996; Alkema *et al.*, 1995; Core *et al.*, 1997; Faust *et al.*, 1995; Kroon *et al.*, 1998; Lessard *et al.*, 1999; Sauvageau *et al.*, 1997; Schumacher *et al.*, 1996; Takihara *et al.*, 1997; Tetsu *et al.*, 1998; Thorsteinsdottir *et al.*, 1997; Tokimasa *et al.*, 2001; van der Lugt *et al.*, 1996; van der Lugt *et al.*, 1994) Os genes *Hox* parecem ser os candidatos prováveis,

devido a sua baixa-regulação sob a diferenciação de células de medula óssea primárias, e sua habilidade em induzir a hiper-proliferação destas células, firmando o estabelecimento dos genes PcG como alvo (Lessard *et al.*, 1998; Sauvageau *et al.*, 1997; Thorsteinsdottir *et al.*, 1997). No entanto, evidências diretas de que a de-repressão de genes Hox é causal das anormalidades hematopoéticas em mutantes de PcG em camundongo é ainda especulativo. Ao contrário, análise dos níveis de expressão de onze genes Hox diferentes não mostraram alterações em tecidos hematopoéticos em camundongos mutantes de *Eed*, destacando que outros genes-alvo críticos ainda não foram identificados (Lessard *et al.*, 1999).

Desregulação de PcG no câncer humano

O acúmulo de trabalhos que sugerem um papel importante das proteínas PcG no controle de proliferação celular, tem estimulado a procura por alterações na expressão de PcG em tumores humanos. Em leucemias mielóides crônicas (LMC), por exemplo, foi demonstrado que o gene *Bmi1*, parece estar super-expresso em 9 de 36 (25%) casos estudados. (Bea *et al.*, 2001). Em 4 casos, esta super-expressão foi fruto da amplificação do gene, mas para os outros 5 casos de LMCs, nenhuma alteração estrutural de *Bmi1* foi encontrada e a causa para sua expressão aumentada permanece obscura. Apesar da super-expressão de *Bmi1* ter sido exclusivamente encontrada em tumores sem alterações no *locus* de *INK41/ARF*, nenhuma diferença nos níveis de expressão de $p16^{INK4a}$ foi encontrada em tumores com e sem super-expressão de *Bmi1*, sugerindo que outro gene-alvo, além de $p16^{INK4a}$, está envolvido no desenvolvimento destes tumores. O padrão de expressão mutuamente exclusivo, característico de *EZH2* e *Bmi1*, é interrompido na doença de Hodgkin e em linfomas não-Hodgkin de células B (Raaphorst *et al.*, 2000a); (van Kemenade *et al.*, 2001). Enquanto que *Bmi1* é ausente em células B foliculares normais em divisão expressando *EZH2*, tumores de células Hodgkin Reed Sternberg e células B neoplásicas de linfomas não-Hodgkin, que originam destas células, expressam *Bmi1*. Isto sugere que estas células B falham em controlar a expressão de *Bmi1*. No caso dos linfomas não-Hodgkin,

o grau de co-expressão de *Bmi1* e *EZH2* em células ciclando foi visto por se correlacionar com o grau de malignidade.

HPH1/RAE28 estão localizados em 12p13, uma região que esta freqüentemente envolvida com anormalidades cromossômicas e apresenta perda de heterozigose (LOH) em uma variedade de doenças hematológicas humanas, incluindo leucemia linfóide aguda (LLA) (Ohta *et al.*, 2000). Interessantemente, a expressão de *HPH1* não foi detectada em 4 de 43 casos de LLA de células precursoras B, embora ambos os alelos de *HPH1* estivessem presentes (Tokimasa *et al.*, 2001).

A desregulação da expressão de PcG não se restringe a malignidades hematopoéticas. Recentemente, a super-expressão de *Bmi1* foi correlacionada com osteosarcomas de alto grau e um aumento na expressão de *Bmi1* foi observado em 58% dos tumores de pulmão de células grandes (NSCLC) (Vonlanthen *et al.*, 2001). Apesar deste estudo de 48 NSCLC, a super-expressão de *Bmi1* não pode ser associada com características do tumor e sobrevida dos pacientes, ou seja: uma significativa correlação inversa entre a expressão de *Bmi1* e *INK4a/ARF* foi observada. A super-expressão de *Bmi1* nos casos de NSCLC e em algumas das MCI super-expressando *Bmi1* não foi acompanhada pela amplificação do *locus* de *Bmi1*, indicando que outro mecanismo é responsável pela geração de sua super-expressão. Fora isso, *SCMH1*, o ortólogo humano do gene de *Drosophila* "sex comb on midleg" que compartilha domínios com o gene supressor de tumor "lethal malignant brain tumor" mapeia no *locus* 1p34, que é uma região de LOH em casos de câncer gástrico e de cólon bem diferenciados (Berger *et al.*, 1999; Bornemann *et al.*, 1996; Wismar *et al.*, 1995).

Recentemente, foi demonstrado que a proteína Polycomb 3 (HPC3) humana interage com ENL e AF9 (Garcia-Cuellar *et al.*, 2001), dois parceiros comuns de fusão da proteína *trx* humana LLM em translocações 11q23, que estão associadas com um aumento da agressividade de leucemias agudas infantis (Saha *et al.*, 1998; Tkachuk *et al.*, 1992). Esta observação sugere um papel mais direto das proteínas PcG em LLM associadas com leucemias da infância, próximo ao seu envolvimento indireto como oponentes da função dos

trxG. Ainda não está claro qual é o produto unificador que é compartilhado entre as 20 diferentes proteínas fusionadas com LLM nestas leucemias.

Utilizando análise de micro-arranjos de cDNA, (Varambally *et al.*, 2002) compararam padrões de expressão encontrados em tecidos de próstata benignos, de tumores órgão-confinados e tumores metastáticos independentes de andrógenos. Eles observaram que o mRNA correspondente ao gene "Enhancer of zeste 2" (EZH2) foi encontrado em um grupo de transcritos com expressão aumentada, que permitia distinguir tumores metastáticos de tumores localizados apenas na próstata. É interessante notar que durante a progressão maligna houve um aumento gradativo da expressão de mRNA, e do nível das proteínas em tumores benignos, órgão-confinados e tumores metastáticos, sugerindo que o aumento na expressão de *EZH2* precede o desenvolvimento dos focos metastáticos. Desta maneira, estes dados sugerem que os níveis da proteína EZH2 podem ser úteis na predição de recidivas de pacientes após a prostatectomia (Varambally *et al.*, 2002). O aumento no nível da proteína EZH2 também foi associado com maior agressividade em tumores de mama (Kleer *et al.*, 2003).

O Gene "Enhancer of Polycomb" de *Drosophila*

O gene "Enhancer of Polycomb" (*EPc*) de *Drosophila*, é um membro intrigante dos PcG; mutações têm classificado este gene como um PcG por causa de fenótipos homeóticos causados por expressão ectópica de *loci* homeóticos (Jones & Gelbart, 1990; Struhl & Akam, 1985) e porque ele aumenta o fenótipo de mutações de genes PcG conhecidos (Campbell *et al.*, 1995; Cheng *et al.*, 1994). Existem 15 PcG identificados (Pirrotta, 1997; Simon, 1995; Yamamoto *et al.*, 1997). "Supressor of zeste 2" (*Suz2*) aumenta alguns fenótipos PcG, mas ele próprio não possui um fenótipo homeótico (Adler *et al.*, 1989; Wu & Howe, 1995), talvez porque ele seja funcionalmente redundante (Brunk *et al.*, 1991). *EPc* é pouco usual entre as mutações dos PcGs, porque ele mesmo não possui um fenótipo homeótico em embriões e adultos, e apenas o gene *Abd-B* mostra modesta expressão ectópica em embriões, quando a contribuição maternal da proteína ou mRNA é removida (Soto *et al.*,

1995). No entanto, mutações em *EPC* aumentam mutações homeóticas nos genes PcG: *Pc*, *Pcl*, *ph*, *Sc*, *Scm* e *sxc*, (Campbell *et al.*, 1995; Cheng *et al.*, 1994), sugerindo que *EPC* seja importante na função dos PcG. Pode ser que assim como *Suz2*, *EPC* seja, de modo parcial, funcionalmente redundante, e, portanto, perca os efeitos homeóticos em embriões e adultos. Um outro gene, *S-adenosilmetionina sintetase*, que não foi identificado por ter fenótipo homeótico, aumenta o fenótipo de mutantes de PcG e atua como um Su(var) (Larsson *et al.*, 1996). Assim como *S-adenosilmetionina*, *EPC* pode ser necessário indiretamente para a função de PcG. Pode ser que *EPC* regule a expressão de PcG ou Su(var), ou que tenha outros efeitos indiretos. Desta maneira, a clonagem de *EPC* foi importante para o entendimento de seu papel nos PcGs e seu papel no estabelecimento de heterocromatina.

Apesar da pouca conservação de seqüência entre a maioria dos PcGs ao longo da evolução, três linhas de evidência sugerem fortemente que exista uma conservação funcional dos PcGs entre moscas e mamíferos. Primeiro, o homólogo funcional de *Pc*, *M33*, pode resgatar muitos mutantes de fenótipo de *Pc* em *Drosophila* (Muller *et al.*, 1995). Segundo, a reposição de genes PcG em camundongo leva à transformações de esqueleto axial e a expressão ectópica de genes *Hox* em embriões mutantes (Akasaka *et al.*, 1996; Core *et al.*, 1997; Takihara *et al.*, 1997; van der Lugt *et al.*, 1994). Terceiro, múltiplas cópias de *EZH2* ou *E(z)* aumenta a PEV em moscas (Laible *et al.*, 1997), sugerindo que existem mecanismos de silenciamento amplamente conservados que ligam a função dos PcG e o silenciamento da heterocromatina.

OBJETIVOS DO TRABALHO

O objetivo geral desta tese de doutorado foi explorar as potencialidades do Projeto Genoma do Câncer Humano para a descoberta de novos genes, possivelmente relacionados ao processo tumoral.

Os objetivos específicos foram:

- Obter a seqüência completa dos novos genes humanos *hEPC1* e *hEPC2*;
- Fazer o mapeamento genômico destes genes;
- Fazer a caracterização da expressão dos genes em diferentes tecidos humanos, incluindo tecidos fetal, normal e tumoral
- Caracterizar a função dos genes *hEPC1* e *hEPC2*.

MATERIAIS

REAGENTES:

Fatores de crescimento:

- TGF- β 1 (Peprotech Inc.) – estoque reconstituído em água na concentração de 50 ng/ μ l.
- BMP-7 (Creative Biomolecules) - estoque na concentração de 100ng/ μ l

Antisoros:

anti-EPCII-1 policlonal (Bethyl);

anti-EPCII-2 policlonal (Bethyl);

anti-FLAG monoclonal M2 e M5 (Sigma);

anti-myc monoclonal (Santa Cruz) e produzido pelo clone celular de hibridoma 9E10;

anti-GST policlonal (Amersham);

anti-IgG de camundongo conjugado com peroxidase (Amersham e Vector);

anti-IgG de coelho conjugado com peroxidase (Amersham e Vector);

anti-IgG de camundongo conjugado com TRITC (isotiocianato B de tetrametilrodamida) (Amersham);

anti-IgG de coelho conjugado com FITC (isotiocianato de fluoresceína) (Amersham).

SOLUÇÕES:

Todas as soluções foram feitas a partir de reagentes de grau de pureza pra análise, seguindo formulações descritas em manuais para uso em laboratório (Ausubel, 1998; Sambrook, 2001)

MÉTODOS

Clonagem do cDNA de *hEPC1* e da isoforma canônica de *hEPC2* por RT-PCR

O consenso da sequência de *hEPC1* e *hEPC2* foi obtido pelo "clustering" de ESTs humanas e de outros organismos, que mostraram similaridade com os genes correspondentes de camundongo. Iniciadores cobrindo toda a extensão da região codificante dos genes foram usados para cobrir todos os "gaps" de sequência (ver Tabelas 1 e 2). Os cDNAs completos de ambos os genes foram inicialmente amplificados gerando fragmentos com sobreposição entre si. Os moldes utilizados nas RT-PCR foram derivados de uma mistura de cDNAs de tecidos normais e tumorais de estômago (o mesmo tecido usado na geração da primeira ORESTES que mostrou similaridade com o gene *EPC*), assim como amostras de cDNAs derivadas de cérebro normal e tumoral. Duas microgramas de RNA total foram usadas em uma reação de transcrição reversa para a geração da primeira fita de cDNA, usando a enzima SuperScript II (Invitrogen) e oligo dT em um volume final de 20 μ l, utilizando o protocolo recomendado pelo fabricante. As amplificações por PCR foram feitas em reações de 25 μ l, contendo 1,5 U de Platinum Taq DNA polimerase High Fidelity (Invitrogen), 1 μ l do cDNA como molde, 5 pmol de cada iniciador, 0,25 mM de dNTP, 2 mM $MgSO_4$, 60 mM de Tris- SO_4 (pH8.9) e 18 mM $(NH_4)_2 SO_4$. O programa de ciclagem consistiu de um primeiro passo de desnaturação à 95°C por 3 minutos, seguido de 35 ciclos a 95°C por 50 segundos, anelamento a 56°C por 40 segundos para todos os iniciadores usados e extensão a 68°C por 4 minutos. Os produtos amplificados foram então reparados e clonados em "blunt" no sítio de *Sma*I de pUC18 (Amersham), ou foram diretamente sequenciados no ABI 377 (Applied Biosystems) usando a técnica de "dye-terminator". Após o sequenciamento os cromatogramas foram utilizados para a obtenção das sequências consenso completas, utilizando o pacote de programas "Phred-Phrap consed" (Ewing & Green, 1998; Ewing *et al.*, 1998; Gordon *et al.*, 1998) com "score" \geq 20.

Caracterização das variantes Iso-B e Iso-C e análise quantitativa por RT-PCR

Os iniciadores utilizados na amplificação das variantes Iso-B e Iso-C em um RT-PCR multiplex semi-quantitativo foram, respectivamente: Iso-B forward: 5' CCTGTAAGGCGGGGAGAC 3' combinado com Iso-B reverse: 5' CTGCCAGGCATTTGGGGACAC 3' e Iso-C forward: 5' GAACATCATTTACAGCGAGC 3' usado com Iso-C reverse: 5' CCGGAGACCACGTTAAAG 3'. As condições de PCR foram as mesmas usadas na amplificação de Iso-A, exceto que a temperatura de anelamento foi de 60°C. cDNAs humanos derivados de leucemia mielógena crônica, adenocarcinoma coloretal e de feto (Clontech), foram usados como molde para a reação de amplificação destas isoformas. Os produtos de PCR foram clonados em pUC18 e sequenciados. Os iniciadores desenhados para a amplificação de Iso-C também co-amplificam Iso-B com uma diferença no tamanho dos produtos. Esta reação de amplificação, usando os mesmos pares de iniciadores, permitiu a padronização de uma reação de RT-PCR semi-quantitativa. Um painel de mRNAs, derivados de diferentes tecidos foi utilizado na avaliação de Iso-B e Iso-C: glândula adrenal, medula óssea, cérebro total, cérebro fetal, fígado fetal, pulmão fetal, coração, rim, fígado, placenta, intestino delgado, medula espinhal, baço, timo, traquéia, cólon, útero e músculo esquelético (Clontech). A número de ciclos de PCR abrangendo a faixa mínima de detecção e o ponto de saturação foi de, respectivamente, 29 e 33 ciclos. Uma porção do cDNA do gene GAPDH foi usada para normalizar a quantidade de cDNA nas reações de RT-PCR e os números de ciclos foram padronizados entre 22 e 26. Os iniciadores usados na amplificação de GAPDH foram: 5' ACCACAGTCCATGCCATCAC 3' e 5' TCCACCACCCTGTTGCTGTA 3', e as condições de PCR foram as mesmas usadas na amplificação de Iso-B e Iso-C. As bandas obtidas foram analisadas por eletroforese em géis de agarose 1% corados com brometo de etídeo.

Ensaio de Northern-blot

As membranas comerciais "Human Cancer Cell Line Multiple Tissue Northern-Blot" e "Human 12-Lane Multiple Tissue Northern Blot Membranes" (Clontech) usadas na avaliação da expressão das isoformas A, B, C e D contém 2 µg de mRNA por canaleta. As hibridizações foram feitas como sugerido pelo fabricante, utilizando a solução comercial ExpressHyb (Clontech). Os resultados de todas as membranas de Northern-blot foram analisados no aparelho Storm PhosphorImager (Molecular Dynamics).

Preparo de sonda de DNA radioativa

Para marcar a sonda radioativamente, foi utilizado o kit "Ready to go" (Amersham), baseado na reação de "Random primer extension" (Feinberg & Volgestein, 1984). À mistura de reação (uma solução liofilizada contendo dATP, dGTP, dTTP, 4-8 unidades da enzima Klenow e oligonucleotídeos randômicos de 9 mers), foram adicionados 20 µl de água e os tubos foram colocados no gelo. A solução contendo o produto de PCR foi fervida por 5 minutos para desnaturação, imediatamente resfriada no gelo por pelo menos 2 minutos, e centrifugada rapidamente. Em seguida, esta solução de DNA (25 µl) foi adicionada à mistura de reação reconstituída em água (20 µl) juntamente com 5 µl (50 µCi) de [α -³²P]dCTP (3000Ci/mmol). Misturou-se gentilmente e incubou-se a reação à 37°C por 30 minutos. A sonda radioativa foi separada dos nucleotídeos livres em colunas de Sephacryl S-200 ou S-300 (Amersham) e desnaturada em banho seco a 95 °C por 5 minutos, antes de ser utilizada na hibridização. As sondas utilizadas correspondem a fragmentos do gene *hEPC2* e *hEPC1* obtidos por PCR. As variantes Iso-B e Iso-C foram avaliadas usando uma sonda similar à região compreendida entre os nucleotídeos 5 a 425 da região codificante da isoforma canônica (Iso-A). A expressão de Iso-A e D foi avaliada usando uma sonda na 3'UTR dos nucleotídeos 2,397 ao 3,352. A sonda correspondente ao controle interno 36B4 (fosfoproteína ribossomal ácida humana PO) foi preparada a partir do fragmento PstI-PstI de 0,8Kb do cDNA de

36B4, obtido por digestão do plasmídeo pGEM36B4cDNA com a respectiva enzima de restrição.

Os isótopos radioativos utilizados tanto para Northern-blot quanto para Dot-blot foi o $[\alpha^{32}]$ dCTP (3000Ci/mmol) (Amersham International plc, Buckinghamshire, England).

Análises de expressão gênica por Dot-blot

A membrana de Dot-blot comercial "Multiple tissue expression array-MTE™" (Clontech) contendo 76 mRNAs derivados de tecidos humanos e de linhagens celulares foi utilizada na análise de expressão dos genes *hEPC1* e *hEPC2*. Esta membrana possui os níveis de expressão dos mRNAs normalizados por 8 genes de expressão constitutiva ("housekeeping genes"). A expressão de *hEPC1* foi analisada utilizando uma sonda gerada por PCR dos nucleotídeos 1,991 ao 3,519. A sonda usada na análise de *hEPC2* foi a mesma usada nos Northern-blots para *IsoA/IsoD*. As condições de hibridização usadas seguiram a recomendação do fabricante.

Sequenciamento de DNA

Para tal, foi utilizado o sequenciador automático ABI Prism-377, ou ABI 3100 (Applied Biosystems. As reações de sequenciamento foram feitas utilizando o Kit "DYEnamic ET terminator cycle sequence kit" (Amersham, USA). A reação se baseia no método de Sanger modificado, ou sequenciamento pelo método de dideoxi.

Para cada molde de DNA sequenciado, foram usados 2 μ l do "sequencing reagent premix", que contém os ddNTPs marcados com substâncias fluorescentes, além de enzima adequada, dNTPs e o tampão. Para o sequenciamento, foram utilizados 100 μ M dos iniciadores utilizados na reação de amplificação. O DNA sequenciado foi derivado de produto de PCR ou de DNA plasmideal purificado em pequena escala. O volume final da reação foi de 10 μ l e as reações utilizavam o seguinte programa de ciclagem: 95°C por 20

segundos, 50°C por 15 segundos e 60°C por 1 minuto, repetindo-se este programa por 30 ciclos.

Após a reação de sequenciamento, os produtos foram precipitados para a purificação dos ddNTPs não incorporados, usando 2µl de tampão contendo acetato de sódio e EDTA e 80 µl de etanol absoluto. As amostras foram bem misturadas e incubadas no gelo em local escuro por 20 minutos. Em seguida, as amostras foram centrifugadas à 13.000 rpm por 30 minutos e o sobrenadante foi descartado. O "pellet" resultante foi lavado com etanol 75% e centrifugado por 10 minutos. O sobrenadante foi novamente descartado e o "pellet" foi seco por alguns minutos à temperatura ambiente e finalmente ressuspendido em tampão de aplicação para aplicadas ao gel ou injeção no capilar.

Análises de Bioinformática e bancos de dados

As coordenadas genômicas e o mapeamento dos genes *hEPC1* e *hEPC2* foram determinados através do Blast, utilizando o banco de dados do genoma humano (www.ncbi.nlm.nih.gov/blast) e também a ferramenta BLAT (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start&org=human>). Os alinhamentos de proteína e o filograma das proteínas ortólogas de *hEPC1* e *hEPC2* foram obtidos pelo programa Clustalw (<http://www.ebi.ac.uk/clustalw>). As análises de predição da localização celular de *hEPC2* foram feitas utilizando o program PSORT II (psort.ims.u-tokyo.ac.jp). As variantes Iso-B e Iso-C foram preditas usando análises de BlastN contra ESTs humanas, assim como o a ferramenta Aceview (www.ncbi.nlm.nih.gov/IEB/Research/Aceview). O sítio alternativo de poli-adenilação utilizado por Iso-D foi investigado através da ferramenta LTV (Ludwig Transcript Viewer, <http://cgap.nci.nih.gov/SAGE/LTviewer>) do SAGE Genie, que permite analisar a presença de "SAGE tags" no gene analisado e indicar a presença de transcritos alternativos.

Western-blot

Extração e dosagem de proteínas totais

As células foram crescidas em placas P100, 6 poços e 24 poços, nas condições desejadas para a transfecção. O meio de cultura foi removido e as culturas foram lavadas duas vezes com PBSA gelado. As placas foram colocadas sobre o gelo antes da lise. Foi acrescido 1 ml de tampão de lise às células plaqueadas em placas P100, que foram incubadas sob agitação por 20 minutos. O lisado foi então raspado das placas e colocado em tubos eppendorf previamente gelados. O lisado foi clarificado por centrifugação (15.000 rpm, 10min, 4°C) para separação da proteínas insolúveis. O lisado foi quantificado pelo método de Bradford (Bio-Rad). Pelo menos 30 µg do lisado total foi utilizado nos ensaios. Foi acrescido às amostras inibidores de proteases, como descrito abaixo. As alíquotas foram misturadas ao tampão de amostra para concentração final de 1X. As amostras foram incubadas à 95°C por 5 min e submetidas a SDS-PAGE.

Tampão de lise (solução estoque):

20mM de Tris-HCL pH 7.5

150mM NaCl

10% glicerol

1% Triton X-100

Inibidores de protease utilizados:

PMSF: estoque a 200mM, concentração final de 1mM.

Trasyolol: concentração final a 1,5%.

Ortovanadato de sódio: estoque a 100mM, concentração final de 1mM.

Leupeptina: estoque a 1m/ml, concentração final de 2µg/ml.

Pepstatina: estoque a 1mg/ml, concentração final de 2µg/ml.

Transferência úmida das proteínas do gel para a membrana de nitrocelulose

Membranas de nitrocelulose (Schleicher & Schuel BA85 ou BA83) e papéis Whatmann 3MM foram recortados com o mesmo tamanho do gel. A membrana, os papéis, o gel e as espumas do aparato de montagem de transferência foram embebidos em tampão de transferência 1X. A membrana deve ser primeiramente molhada em água estéril. A montagem foi feita colocando, na ordem: espuma embebida, 3 folhas de papel 3MM, gel, membrana, 3 folhas de papel 3MM, espuma, eliminando bolhas de ar. O cassete foi fechado e colocado na cuba com a membrana voltada para o pólo positivo. A cuba foi preenchida com tampão de transferência, que foi feita a 400 mA por 3 horas. A membrana foi corada com 0,1% Ponceau em 10% de ácido acético por 5 minutos e descorada com água.

Imunoreação

Os sítios inespecíficos da membrana foram bloqueados com 5% leite Mólico desnatado ou 5% BSA, em TBST 0,1% Tween-20 por toda a noite. A membrana foi lavada com TBST 3 vezes por 10 min. O antisoro na diluição adequada (1:200 EPC2-I (Bethyl, USA); 1:1000 anti-IgG de camundongo conjugado com peroxidase (Vector, USA); 1:25.000 anti-IgG de coelho conjugado com peroxidase (Vector, USA) foi incubado de acordo com o tempo adequado para cada antisoro (por toda a noite para EPC2-I e por volta de 45 min a 1 hora para os outros). A membrana foi lavada por 3 vezes com TBST, sob agitação. O anticorpo secundário conjugado com peroxidase foi acrescentado e incubado da mesma maneira pelo tempo adequado. Novamente a membrana foi lavada (3 vezes) com TBST por 10 min e revelada com o Kit ECL™ (Amersham) segundo instruções do fabricante. A membrana foi exposta ao filme, até a obtenção do sinal adequado.

Imunoprecipitação

Foram utilizadas cerca de 500 µg de proteína total, previamente dosadas pelo método de Bradford. À estas amostras foram acrescentados 50 µl de 50% proteína A conjugada com Sepharose (diluída em PBSA). As amostras foram colocadas em uma plataforma de rotação à 4°C, por meia hora e centrifugadas por 2.000 rpm por 5 min. Posteriormente, o sobrenadante foi transferido para um novo tubo e 10 µg do anticorpo de interesse foram adicionadas, com incubação por toda a noite sob rotação à 4°C. 30 µl de proteína G Sepharose ou proteína A Sepharose foram acrescentados às amostras, de acordo com o anticorpo utilizado. As amostras foram novamente incubadas sob rotação a 4°C por meia hora e centrifugadas à 2.000 rpm por 5 min. O sobrenadante foi descartado e o “pellet” lavado 3 vezes com tampão de lise. O “pellet” foi então ressuspensionado em 20 µl de tampão de amostra 2X e aplicado em gel SDS-PAGE, transferido para membrana de nitrocelulose e submetido ao Western-blot.

Condições de cultura e manutenção de linhagens celulares

As células foram cultivadas a 37°C em frascos e placas de plásticos descartáveis contendo DMEM suplementado com 5 ou 10% de soro fetal bovino (FBS), 1,2 g/l de bicarbonato de sódio, 25 mg/l de ampicilina e 100 mg/l de estreptomicina, em atmosfera de 5% de CO₂/95% ar para a manutenção do pH próximo ao fisiológico. As células foram sub-cultivadas sempre que atingiam 80% da densidade de saturação, utilizando para tanto, tripsina 0,1% em PBSA contendo 1mM de EDTA e tendo sido a cultura previamente lavada uma vez com solução de PBSA. Os estoques celulares foram mantidos no meio de cultivo contendo 10% DMSO (dimetilsulfóxido) a -196°C, em reservatório contendo nitrogênio líquido.

Soluções e meios de cultura para células de mamíferos:

- DMEM (“Dulbecco’s Modified Eagle Medium”), (Invitrogen, USA).
- FBS: Soro fetal bovino (Cultilab e Invitrogen)
- Tripsina: (ICN Pharmaceuticals Inc., USA; Invitrogen, UK).
- PBSA (“Phosphate Buffered Saline”): Solução salina sem cálcio ou magnésio), tamponada (pH 7,2), composta por NaCl a 140 mM; KCl a 2,7 mM, Na₂HPO₄ a 8 mM e KH₂PO₄ a 1,5 mM.

Linhagens celulares utilizadas:

293-T: células de rim embrionário humano

COS-7: células de rim de macaco

HEPG2: células de hepatoma humano

Clonagem em vetores plasmidiais

Todas as construções plasmidiais usadas nos experimentos de expressão em células de mamífero e de bactérias foram feitas utilizando o cDNA completo da isoforma canônica de *hEPC2* (Iso-A).

Para a clonagem de *hEPC2* em pCDNA3-6Xmyc foi utilizado como molde para PCR, um clone de pUC18 contendo *hEPC2*. Este clone havia sido previamente sequenciado para certificação da sequência correta do gene. Foram desenhados iniciadores contendo sítio de restrição para *EcoRI* e *XhoI*, sendo que o iniciador “sense” foi desenhado de maneira a suprimir o códon de início de tradução do cDNA, para que sua tradução incluísse o epítipo de *myc* presente no vetor. A clonagem do cDNA no vetor foi feita nos mesmos sítios de restrição para clonagem que o cDNA possuía, e de modo que a região N-terminal da proteína traduzida estivesse adjacente ao epítipo de *myc*, já que por predição, esta deve ser a região mais hidrofílica de *hEPC2*.

O cDNA de *hEPC2* foi sub-clonado no plasmídeo pGEX-4T1 também utilizando também os sítios de *EcoRI* e *XhoI*, e o inserto foi retirado de pCDNA3-6Xmyc-EPC2. Para a subclonagem de *hEPC2* no vetor de expressão pCXN2 foi feita através de digestão com *BglII* e *NotI* do vetor pCDNA3-6Xmyc-EPC2 para obtenção do inserto. O vetor pCXN2 foi digerido com a enzima

*Xho*I, tanto o inserto quanto o vetor preparado tiveram suas extremidades reparadas através do tratamento com enzima Klenow na presença de deoxinucleotídeos para, finalmente, proceder à ligação dos fragmentos com extremidades cegas. A orientação correta do inserto no sentido 5' para 3' foi confirmada por sequenciamento em alguns clones de pCXN2-hEPC2. A construção de pET-32A-hEPC2 foi feita usando o inserto de *hEPC2* retirado de pCDNA3 com as enzimas *Eco*RI e *Xho*I, e a seguir, direcionalmente clonado nos mesmos sítios do vetor pET32. Em todas as etapas de subclonagem, os fragmentos lineares de vetor plasmidial foram tratados com fosfatase alcalina de intestino de vitelo (CIAP) previamente à reação de ligação, para diminuir o número de clones não-recombinantes.

Plasmídeos utilizados:

pGEX: (Amersham) vetor de expressão bacteriano. Produz uma proteína de fusão com a glutatona S-transferase (GST). O gene da GST possui um sítio de ligação a ribossomo e um ATG, estando sob o controle do promotor PTac.

pET-32A: (Novagen) vetor de expressão indutível bacteriano. Os genes clonados nos vetores pET ficam sob o controle do forte promotor para RNA polimerase do bacteriófago T7. A expressão de T7 RNA polimerase é induzida por IPTG a partir de um operon artificial LacI-T7RNAPol integrado no genoma da cepa bactéria BL2-DE3 de *E.coli*. A série pET-32 expressa a sequência peptídica alvo fusionada com 109 aminoácidos da proteína tioredoxina. O vetor também produz um epítipo de hexahistidina que permite a detecção da proteína de fusão e sua consequente purificação em colunas de níquel imobilizado.

pCDNA3: (Invitrogen) vetor de expressão constitutiva em mamífero que possui fusionado um epítipo de uma parte da seqüência do proto-oncogene myc repetida 6 vezes.

pCXN2: vetor não comercial de expressão constitutiva em mamífero. Possui um promotor híbrido de beta-actina fusionado com o promotor precoce de citomegalovírus ("immediate early CMV promoter"), o que torna a expressão da construção bastante forte.

Transfecção celular

Por cloreto de cálcio

As células utilizadas neste procedimento se encontravam em 60% de confluência em placas P100, ou placas de 6 poços. As células eram colocadas em meio DMEM sem antibiótico. À quantidade adequada de DNA para o experimento foi misturada a 50µl de CaCl₂ 2,5 M. A esta solução foi adicionado 500 µl de tampão BBS 2X borbulhando em cada tubo com o auxílio de uma pipeta Pasteur. Os tubos foram deixados por 30 minutos à temperatura ambiente e seu conteúdo foi acrescido gentilmente às placas, que foram colocadas na estufa por 24 horas à 37°C. O meio foi removido e as células foram lavadas com PBSA e as células lisadas para a obtenção de proteínas.

Soluções utilizadas:

BBS 2X (BES-buffered saline):

50 mM BES (N,N-bis[2 – hydroxyethyl]-2-aminoethanesulfonic acid)

280mM NaCl

1,5mM Na₂HPO₄.2H₂O

pH 6,96

2,5 mM CaCl₂:

11g de CaCl₂.6 H₂O em 20 ml de H₂O destilada.

Transfecção utilizando Lipofectamina

Foi utilizado o reagente Lipofectamine 2000 (Invitrogen, USA), que possui alta eficiência de transfecção em muitos tipos celulares. Foi utilizado o protocolo sugerido pelo fabricante do produto para as transfecções transitórias.

Infecção celular utilizando adenovírus

As proteínas SMAD recombinantes utilizadas nos ensaios de co-precipitação e de imunofluorescência foram clonadas no sistema de expressão adenoviral utilizando o kit "Takara's adenovirus expression vector kit" que utiliza o método Cos-TPC (Miyake *et al.*, 1996). As construções foram feitas pelos integrantes do grupo TS (grupo de Sinalização de TGF- β), se encontrando portanto, prontas para uso. Convenientemente, adenovírus recombinantes têm sido gerados pelo uso de DNA genômico de adenovírus sem a proteína terminal (TP). O DNA genômico do adenovírus sem a TP e um plasmídeo com o gene-alvo inserido são transfectados em células por co-transfecção. No entanto, como ambas as pontas do DNA genômico do adenovírus possuem originalmente a proteína TP, e o DNA genômico sem TP possui baixa eficiência de produção de partículas virais, a geração eficiente de adenovírus recombinantes é bastante dificultada pelos métodos convencionais. O método COS-TPC emprega o uso do complexo DNA genômico com a proteína TPC. Pelo uso deste DNA-TPC e do cosmídeo dentro do qual o gene-alvo foi inserido, a eficiência de obtenção de vírus recombinantes aumenta de 10 vezes para até 100 vezes. Portanto, os adenovírus recombinantes desejados podem ser obtidos com grande eficiência. Vetores adenovirais possuem vantagens adicionais em relação a outros sistemas de expressão, como a obtenção de altos títulos virais como 10^8 - 10^9 unidades formadoras de partículas por mililitro (pfu/ml). Estes vírus podem ser concentrados para até 10^{11} pfu/ml. Por essa razão, o sistema pode transferir o gene-alvo em quase 100% de células aderentes. Além disso, uma vasta quantidade de animais podem ser infectados por adenovírus. Tanto células em proliferação como células quiescentes podem ser infectadas, além de diferentes tipos de culturas celulares, incluindo células do sistema nervoso.

As alíquotas de contendo a solução de vírus se encontravam em tubos de 2 ml congelados à -80°C , sendo que estas soluções foram alíquotadas em volume de $200\mu\text{l}$. Para sua utilização, os tubos foram descongelados à 37°C rapidamente.

As células 293-T foram escolhidas para este experimento por possuírem alta eficiência de transfecção, além disso o vetor adenoviral possui a região codificante da proteína E1A deletada do seu genoma, para que um fragmento de DNA exógeno possa ser clonado. Essa proteína é essencial para a produção e empacotamento dos vírus, e a 293T expressa constitutivamente a proteína E1A. Adicionalmente, estas células possuem o antígeno T maior sendo constitutivamente expresso, que por sua vez é responsável pela amplificação das construções de SMAD adenovirais, já que estes possuem a origem de replicação de SV40.

A infecção com as construções de SMADs e com a construção controle de LacZ foi feita 24 horas após a transfecção de *hEPC2* nas células. No dia da infecção o meio das células foi trocado de DMEM 10% FBS para DMEM 5% FBS, e seu volume foi reduzido ao máximo possível para favorecer a infecção. A diminuição do FBS é feita para evitar que fatores de crescimento presentes nesta solução alterem o resultado do experimento. Para as placas P100, o volume de meio utilizado foi de 8 ml, para P60 foi de 4 ml e para placas de 6 poços foi de 1 ml. Após o descongelamento dos estoques das soluções contendo os adenovírus, esta foi homogeneizada e o volume correspondente ao título adequado ao experimento foi gentilmente adicionado sobre o meio, de modo a ser o mais bem distribuído possível. O meio das placas foi então gentilmente homogeneizado e estas foram incubadas à 37°C por 4 horas. Após este período, o meio foi trocado, mantendo a concentração de 5%FBS, em volume de 10 ml para placas P100, 5 ml para P60 e 1,5 ml para placas de 6 poços. A incubação foi feita durante a noite e no dia seguinte, 1 hora antes da adição do fator de crescimento (TGF- β 1 ou BMP-7), o meio foi trocado para DMEM 10% FBS. Após a adição do fator de crescimento, as células foram incubadas por 2 horas na estufa, e finalmente foi feita a extração de proteínas das amostras.

Ensaio de gene repórter

Este ensaio foi feito nas células de hepatoma humano HepG2 que são utilizadas com sucesso no TSL para este experimento. Foi utilizada a construção de pCXN2-hEPC2 em 3 quantidades diferentes de DNA (0,3; 0,2 e 0,1 μ g), co-transfectada com a construção do vetor repórter pGL3-CAGA₁₂-MLP-luc contendo 12 repetições do elemento responsivo a SMAD CAGA (cedida pelo grupo sueco). As células foram plaqueadas em placas de 24 poços, e a transfecção foi feita utilizando o reagente Lipofectamine 2000 (Invitrogen), para assegurar uma alta eficiência de transfecção. Todas as condições foram testadas em triplicata e como controle foi utilizado o vetor pCXN2 vazio. No dia seguinte ao da transfecção, as células foram estimuladas com TGF- β 1 por 1 hora antes do ensaio.

O ensaio de gene repórter foi feito removendo-se cuidadosamente o meio dos poços. As culturas foram lavadas gentilmente com PBS para que permanecessem aderidas, e o máximo possível da solução foi então removida. O tampão de lise (1 ml) foi adicionado a cada poço, de maneira a cobrir todas as células. Estas foram então congeladas rapidamente em um banho de etanol contendo gelo seco, e foram descongeladas à temperatura ambiente. Este procedimento visava facilitar a lise celular. As células foram colocadas sob leve agitação para garantir sua completa cobertura pelo tampão de lise. Em seguida, as células foram então raspadas da placa e todo o líquido foi transferido para tubos de microcentrífuga colocados no gelo. Os tubos foram "vortexados" por 15 segundos, e então centrifugados a 12.000 g por 15 segundos a temperatura ambiente. O sobrenadante foi transferido para uma placa nova de 96 amostras para serem lidas no luminômetro. Foram dispensados 100 μ l do "Luciferase assay reagent" nos poços, onde já se encontravam 20 μ l do lisado celular. As amostras foram homogeneizadas e o luminômetro foi programado para aguardar 2 segundos, e então fazer a leitura da atividade da luciferase.

Imunofluorescência

As células COS-7 utilizadas nesse experimento foram plaqueadas sobre lamínulas cobertas por colágeno 0,1%, para evitar que as células despregassem durante as lavagens. O meio foi retirado e as células foram lavadas (3 vezes) com PBSA, e posteriormente fixadas com 3% paraformaldeído em PBSA por 3 minutos à temperatura ambiente. Novamente as células foram lavadas (duas vezes) em PBSA por 3 minutos, agora sob agitação. As células foram então permeabilizadas com a solução 0,5% Triton X-100 em PBSA por 10 minutos à temperatura ambiente. As células foram novamente lavadas (2 vezes) com PBSA sob agitação por 10 minutos e bloqueadas com 5% soro fetal bovino (FBS), 0.1 M de glicina em PBSA por 1,5 hora sob agitação moderada. As células foram lavadas (2 vezes) com 5% FBS por 3 minutos sob agitação. O anticorpo primário foi diluído em 5% FBS em PBSA, e colocado na ausência de agitação por 2 horas sobre as células. As culturas foram então lavadas rapidamente por 2 vezes, e mais 3 vezes por 15 minutos sob forte agitação. Novamente as células foram bloqueadas por 15 minutos em 5% FBS, sob forte agitação. O anticorpo secundário conjugado com o grupo fluorescente FITC (isotiocianato de fluoresceína) ou TRITC (isotiocianato B de tetrametilrodamida) diluído 1:1000 em PBSA 5% FBS, foram incubados por 30 minutos sem agitação. A partir desta etapa as placas são envoltas em papel alumínio para evitar a exposição à luz. As culturas foram lavadas (2 vezes) com 5% FBS por 3 minutos sob forte agitação, e colocadas novamente em PBSA 5% FBS à 4°C por toda a noite. As lamínulas foram então montadas com Fluoromount-G sobre as lâminas (Southern Biotechnology Associates) e foram observadas ao microscópio modelo Olympus Vanox-T, equipado com uma unidade de iluminação com epi-fluorescência AH2-RFL.

Expressão da construção 4T1 pGEX-hEPC2

Bactérias quimiocompetentes da linhagem DH10B foram transformadas com construção de hEPC2-4T1pGEX. Inicialmente, 3 clones transformantes foram crescidos durante toda a noite, e no dia seguinte as culturas bacterianas

foram diluídas 1:10 para crescerem até a fase exponencial. Assim que atingiram a densidade óptica igual a 0,5/0,6 foram a expressão foi induzida com 1mM de IPTG. Os clones foram então lisados em tampão Laemmli 2X e as proteínas analisadas em gel SDS-PAGE corado por Coomassie Brilliant Blue. Os resultados indicam que um dos clones parecia ter sua expressão induzida quando comparado com o vetor vazio, sugerindo que este clone estaria expressando a proteína de fusão GST-hEPC2. No entanto, várias outras tentativas falharam em evidenciar a proteína de fusão super-expressa quando comparada com a construção não induzida. Vários clones foram testados e outra linhagem de bactéria foi também testada (BL21-DE3), no entanto sem sucesso.

Diferentes temperaturas de indução de expressão (16°C e 25°C) foram testadas, assim como diferentes períodos de tempo (3 e 6 horas) ou concentração de IPTG (variando de 0,4; 0,6 e 1M), no entanto nenhuma destas variáveis foi capaz de resolver a ausência de expressão da proteína GST-hEPC2. Todos os clones foram re-sequenciados, além de os DNAs correspondentes terem sido digeridos para confirmação de que estes não haviam sofrido recombinação. Os DNAs destes clones foram levados para o TSL e novamente usados para transformação das cepas DH10B e BL21-DE3. Novamente foram testados experimentos de indução sem sucesso.

Diante destes resultados optamos por clonar o cDNA de *hEPC2* no vetor indutível de expressão em bactérias pET32A (Novagen).

Expressão e purificação da proteína de fusão His-hEPC2

Dez colônias transformadas com pET-32A-hEPC2, cuja sequência do gene e das pontas do vetor foram confirmadas por sequenciamento, foram transformadas na cepa de *Escherichia coli* BL21(DE3). Esta cepa é ideal para o uso com sistemas de expressão baseados no promotor T7 de bacteriófagos, como o pET. A BL21(DE3) carrega o lisógeno DE3 lambda. DE3 expressa a T7 RNA polimerase a partir do promotor de *lacUV5*, que permite a indução das proteínas recombinantes por IPTG. Amostras de 10 ml de cultura de cada um dos clones foram crescidas a 37°C durante toda à noite em meio LB contendo

ampicilina (100µg/ml). No dia seguinte, estas culturas foram diluídas 1:10 em 5ml de LB contendo ampicilina e foram novamente crescidas até a $OD_{600}=0,6$. IPTG, na concentração final de 1mM, foi adicionado às culturas para a indução da expressão da proteína His-hEPC2. Um pequeno volume destas culturas (500 µl) foi retirado antes da indução, para ser utilizado como controle na análise de indução de expressão. Culturas de BL21(DE3) transformadas com o vetor pET32A vazio também foram crescidas para serem utilizadas como controle, e parte da cultura também foi induzida. Outro controle utilizado foi a cultura da cepa BL21(DE2) não-transformada. O tempo de indução inicial foi de 6 horas. Após a indução, as culturas foram centrifugadas para que as células fossem concentradas. O sobrenadante foi descartado, e as células foram ressuscitadas em 1,25 ml (1/4 do volume original) de tampão Tris-HCl (150mM); NaCl (50mM), pH 7,4. Amostras (20 µl) das suspensões foram então lisadas em tampão Laemli, desnaturados a 95°C por 5 minutos e os tubos foram imediatamente colocados no gelo por 2 minutos. Cerca de 10 µl das amostras foram aplicados e analisados em géis de SDS-poliacrilamida à 8%. Após a corrida, os géis foram corados com Coomassie Brilliant Blue, para a verificação da expressão da construção pET32A-hEPC2. Após a constatação positiva da expressão dos clones de pET32A-hEPC2, o próximo passo foi fazer a purificação da proteína.

Inicialmente, a purificação da proteína pET32-hEPC2 foi feita em condições nativas em colunas de cromatografia de afinidade de NiSo₄, já que a proteína de fusão possui um epítipo de 6 histidinas que possuem afinidade por cátions divalentes (ex: Ni²⁺). Para a obtenção do máximo possível da proteína purificada, o volume inicial de cultura induzida foi de 20 ml. A partir deste momento todos os procedimentos do protocolo são feitos à 4°C, para evitar a degradação das proteínas. A cultura foi sedimentada e o "pellet" de bactérias foi ressuscitado em tampão Tris-HCl (150mM) pH 7,4, NaCl (50mM) em um quarto do volume original. A solução foi então sonicada em 10 pulsos de 10 segundos, com intervalo de 20 segundos usando 65% da amplitude do sonicador. A sonicação tem como objetivo solubilizar eventuais corpos de inclusão de proteína. A amostra foi então centrifugada a 12.000 rpm por 40

minutos. Como a amostra foi sonicada, o esperado é que a proteína entre em solução e fique no sobrenadante. Este sobrenadante é então filtrado para retirada de partículas insolúveis. A coluna de afinidade e todas as soluções utilizadas neste protocolo foram preparadas de acordo com as instruções do manual do sistema pET. Foi utilizada uma coluna de vidro de 5 ml da (Amersham) usando 3 ml da resina que é preparada para ser impregnada com o NiSo₄. Após a passagem da solução do sobrenadante da cultura de bactérias na coluna, todos os eluatos foram coletados para avaliação em géis de SDS-policrilamida. Os resultados mostraram que a proteína estava se ligando ao NiSo₄, mas estava sendo eluída nas lavagens da coluna, previamente a eluição com o agente competidor imidazol em concentração mais elevada (1M). A concentração final de imidazol no tampão de lavagem é de 60mM, mas mesmo assim essa concentração foi suficiente para desligar a proteína do NiSo₄. Geralmente, a maioria das proteínas é eluída entre 100-400mM de imidazol, sugerindo uma fraca ligação da proteína pET32-hEPC2 na coluna. A partir deste dado, algumas variáveis foram testadas no intuito de melhorar a ligação da proteína na coluna, como o aumento da concentração do NiSo₄ de 50 mM para 300 mM, aumentar o volume da coluna de 3 ml para 5 ml, aumentar o volume da cultura induzida para 50 ml e aumentar o volume de todos os tampões (ligação, lavagem e de NiSo₄) para 10 vezes o volume da coluna. Além disso, foi acoplada à coluna uma bomba peristáltica programada para controlar o fluxo de passagem das soluções para 0,5 ml por minuto.

Como nenhuma destas variáveis melhorou a ligação da proteína na resina com NiSo₄, optamos por fazer a purificação da proteína em condições desnaturantes. Para fazer este experimento todos os tampões são acrescidos do agente desnaturante uréia a 6M (concentração final). Após a eluição da proteína, a amostra foi dialisada para que esta pudesse re-adquirir a sua forma nativa ("re-folding"). A diálise foi feita lentamente para se evitar mudanças bruscas na concentração do eluato que pudessem afetar a conformação proteica. Após a diálise, a amostra foi concentrada usando uma coluna Centricon pl-20, com especificações para filtrar proteínas menores de 30 KDa. A proteína pET32-hEPC2 se mostrou presente no eluato, co-purificada com algumas proteínas menores contaminantes de menor intensidade.

Tampão de ligação 8X:

40 mM imidazol

4M NaCl

160 mM Tris-Hcl, pH 7,9

Tampão de lavagem 8X:

480 mM imidazol

4M NaCl; 160 mM Tris-Hcl, pH 7,9

Tampão de eluição 4X:

4M imidazol

2M NaCl

80 mM Tris-HCl, pH 7,9

Tampão de carregamento 8X:

400mM NiSO₄

RESULTADOS

Estratégia para a obtenção da sequência completa dos cDNAs de *hEPC1* e *hEPC2*.

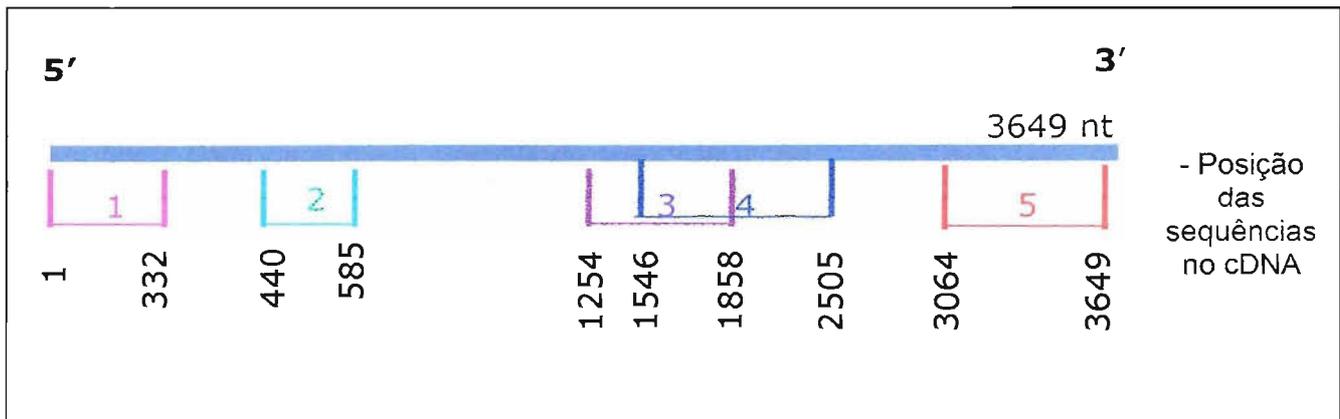
No início das análises deste trabalho, uma busca no banco de dados do Projeto Genoma Humano do Câncer (HCGP, financiado pela FAPESP e pelo Instituto Ludwig de Pesquisa sobre o Câncer) permitiu a identificação de duas seqüências ORESTES similares à sequência do gene *EPc* de *Drosophila*: BE149158 e AW809208. Tais ESTs humanas foram derivadas, respectivamente, de tumores de cabeça e pescoço e de estômago. Análises do tipo BLAST feitas contra o banco de dados “High-Throughput Genome Sequences” – HTGS, do Genbank, mostraram que as ESTs apresentavam similaridade com PACs mapeados em diferentes cromossomos. Desta maneira, verificou-se que BE149158 representa um fragmento de um gene mapeado no cromossomo 10 e AW809208 representa um gene mapeado no cromossomo 2. Os autores, que caracterizaram o gene *EPc* de *Drosophila* (Stankunas *et al.*, 1998), descrevem também nesse artigo, a existência de dois ortólogos de *EPc* em camundongo e sugerem a presença de dois ortólogos humanos. No entanto, seus dados se basearam apenas em um mapeamento de *hEPC1* e *hEPC2* através de FISH (“Fluorescence in situ hybridization”). Porém, esses dados não foram confirmados por nossas análises, já que os autores afirmam que *hEPC2* se localiza no locus 22q13.3, enquanto que na realidade o gene está em 2q23.1 (dado obtido através de Blast contra o genoma humano).

De posse destas informações, partimos para a obtenção das seqüências completas destes genes. Buscamos montar contigs virtuais, utilizando seqüências genômicas e de ESTs humanas (presentes no Genbank, e no banco de dados do HCGP), sendo que estas foram ordenadas em relação aos genes ortólogos de camundongo. Usando a sequência da EST AW809208 contra estes bancos de dados, buscou-se seqüências que apresentassem similaridade com esta EST. Inicialmente, foi formado um “contig” a 3’ do gene *Epc2* de camundongo, com as seqüências de cDNA FLJ11148 fis, clone

PLACE1006704 (ID:AK002010) e o clone de cDNA tj82b10.x1 Soares_NSF_F8_9W_OT_PA_P_S1 (ID: AI458426). No entanto, na porção mais a 5' do transcrito, a EST AW809208 não apresentou sobreposição com nenhuma outra sequência. Na porção 5' do gene putativo um "contig" foi montado com a sequência 13 da Patente WO9855512 (ID: AX002206) e o clone de cDNA wg68a04.x1 Soares_NSF_F8_9W_OT_PA_P_S1 (ID: AI761763.1) (Figura 3). Para fechar os "gaps" da sequência e confirmar a sequência prevista "in silico", oligonucleotídeos foram desenhados flanqueando toda a extensão do gene putativo (Tabela 1). Para confirmação da sequência completa da região codificante do gene, oligonucleotídeos foram desenhados baseando-se no alinhamento das sequências acima citadas com o PAC RP11-548K3. Todos amplificam regiões de aproximadamente 400 a 500 pares de bases, tornando possível o sequenciamento direto (usando iniciadores "sense" e "antisense") de toda a extensão dos produtos amplificados. Todos os fragmentos gerados por RT-PCR possuem sobreposição de sequência, permitindo a montagem correta da sequência do gene.

Tabela 1 – Iniciadores utilizados na amplificação do cDNA completo do gene *hEPC2*

Iniciadores	Posição dos iniciadores no cDNA	Sequência dos iniciadores	Tamanho dos produtos de PCR (nt)
EPC2.1F EPC2.1R	-64 a -44 nt 262 a 182 nt	5'ggacatgcctgatctcaacg 3' 5'ttgcaaaggcttaatttcca 3'	346
EPC2.2F EPC2.2R	151 a 170 nt 490 a 510 nt	5'aaacagccaaaacagttcat 3' 5'tcttctccgaaaggcaacat 3'	360
EPC2.3F EPC2.3R	356 a 376 nt 748 a 767 nt	5'gctaaacgaagatgattacc 3' 5'gctggagtggcatataactc 3'	411
EPC2.4F EPC2.4R	656 a 676 nt 1030 a 1049 nt	5'cttaaccttagaagttgtgg 3' 5'ggatcattttcttcttccgg 3'	393
EPC2.5F EPC2.5R	911 a 929 nt 1352 a 1372 nt	5'catgtcagcaaccactcc 3' 5'cattggtccgagagaaattag 3'	503
EPC2.6F EPC2.6R	1265 a 1284 nt 1688 a 1707 nt	5'cacagaacatgaccagtcc 3' 5'ctggtgagaggaatgctgag 3'	442
EPC2.7F EPC2.7R	1569 a 1588 nt 2018 a 2037 nt	5'cagcaagaatggcatacag 3' 5'gttcatgggaagtgtctgag 3'	469
EPC2.8F EPC2.8R	1886 a 1905 nt 2389 a 2406 nt	5'gaacctctaaaacattatac 3' 5'gtccatttaggtccattg 3'	521
EPC2.9F EPC2.9R	2247 a 2267 nt 2663 a 2683 nt	5'cttaaatggaatagcagagac 3' 5'ggatttatttttacaagatac 3'	437
EPC2.10F EPC2.10R	2568 a 2586 nt 2883 a 2900 nt	5'ggattttaaaaaataattc 3' 5'cgatccatccctttcccc 3'	333
EPC2.11F EPC2.11R	2766 a 2785 nt 3184 a 3201 nt	5'cacttggttcattttgtttc 3' 5'ctgggatggctcgatatac 3'	429
EPC2.12F EPC2.12R	3019 a 3037 nt 3436 a 3454 nt	5'gcactctacgacatttaaac 3' 5'gcagttttcacttatcaag 3'	391



- 1- (AI761763.1) cDNA wg68a04.x1 Soares_NSF_F8_9W_OT_PA_P_S1
- 2- (AW809208) MR4-ST0118-261099-012-g02
- 3- (AX002206) sequência 13 da Patente WO9855512
- 4- (AK002010) cDNA FLJ11148 fis, clone PLACE1006704
- 5- (AI458426) cDNA tj82b10.x1 Soares_NSF_F8_9W_OT_PA_P_S1

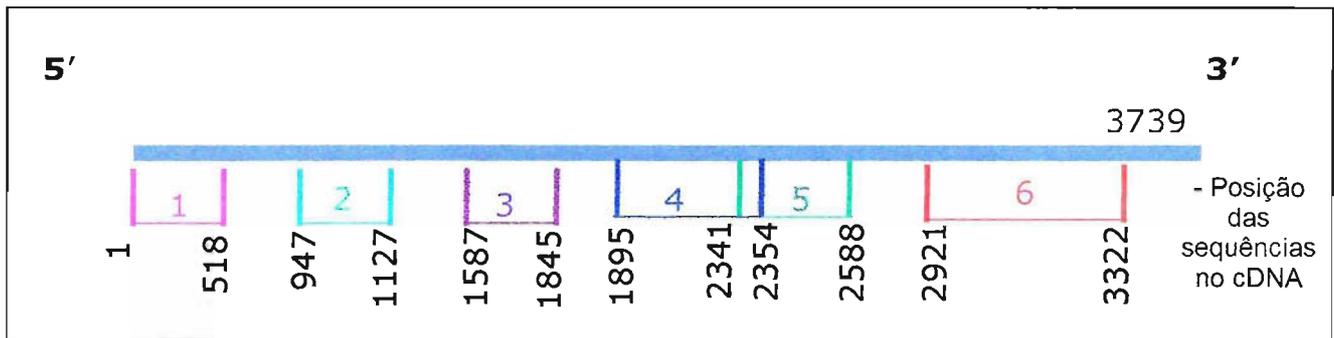
Figura 3 - Representação das sequências nucleotídicas utilizadas na montagem do transcrito completo de *hEPC2*

A mesma abordagem foi utilizada para a obtenção da sequência codificante do gene *hEPC1* a partir da EST BE149158. A sequência desta EST foi usada para montar um “cluster” de sequências à 3’ do gene putativo, junto com sequências que possuíam pequenos “gaps” entre si: MR3-HT0487-290100-102-a07 (ID: BQ349493); yv38f05.s1 Soares fetal liver spleen 1NFLS (ID:N54966.1); ne28b08.s1 NCI_CGAP_Co3 (ID: AA494420) e oc38e11.s1 NCI_CGAP_GCB1 (ID:AA808029). Estes “gaps” foram resolvidos através de alinhamentos com a sequência do PAC RP11-166N17, que posteriormente foi confirmada por RT-PCR seguida de sequenciamento. A EST AI733045.1 - oj36d02.x5 NCI_CGAP_Lu5 apresentava similaridade na porção 5’ deste gene. Utilizando a sequência de *EPC1* de camundongo, buscamos a região do PAC RP11-166N17 dos nucleotídeos 109561-109741 para desenhar oligonucleotídeos que permitissem a união do “cluster” a 3’ com a EST a 5’

(Figura 4). Da mesma maneira que em *hEPC2*, desenhamos oligonucleotídeos cobrindo toda a região codificante do gene *hEPC1* (incluindo as extremidades 3' e 5' não traduzidas) (Tabela 2). Os fragmentos gerados por estes oligonucleotídeos também produziram produtos de PCR com porções de sequência comuns entre si, permitindo sua ordenação.

Tabela 2 – Iniciadores utilizados na amplificação do cDNA completo do gene *hEPC1*

Iniciadores	Posição dos iniciadores no cDNA	Sequência dos iniciadores	Tamanho dos produtos de PCR (nt)
EPC1.1F EPC1.3R	5 a 145 nt 1011 a 1024 nt	5' aacatcatcttcagcggg 3' 5' cagcggcagacgatg 3'	897
EPC1.2F EPC1.3R	318 a 337 nt 1011 a 1024 nt	5' tgatttggattctgaagatg 3' 5' cagcggcagacgatg 3'	707
EPC1.4F EPC1.6R	979 a 998 nt 1673 a 1690 nt	5' cggaaatatgaaaagaagcc 3' 5' ttactttgcgtagaggtag 3'	713
EPC1.5F EPC1.8R	1625 a 1643 nt 2204 a 2221 nt	5' ggagtataaaccgaacagg 3' 5' cctaggtatatgtcgtgc 3'	617
EPC1.7F EPC1.10R	1991 a 2009 nt 3042 a 3062 nt	5' gcttacacctcagtagtac 3' 5' aagatccccatttaatagtagtac 3'	1092
EPC1.9F EPC1.11R	2910 a 2921 nt 3481 a 3499 nt	5' ggtataacataacttggttcc 3' 5' cgatatttgtggttctggc 3'	590



- 1- (AI733045.1) oj36d02.x5 NCI_CGAP_Lu5
- 2- PAC RP11-166N173- HT0252-271099-017-g02
- 3- (BQ349493) MR3-HT0487-290100-102-a07
- 4- (N54966.1) yv38f05.s1 Soares fetal liver spleen 1NFLS
- 5- (AA494420) ne28b08.s1 NCI_CGAP_Co3
- 6- (AA808029) oc38e11.s1 NCI_CGAP_GCB1

Figura 4 - Representação das sequências nucleotídicas utilizadas na montagem do transcrito completo de *hEPC1*.

A montagem de ambos os genes foi feita utilizando o pacote de programas Phred-Phrap/Consed (Ewing & Green, 1998; Ewing *et al.*, 1998; Gordon *et al.*, 1998), onde o Phred utiliza os cromatogramas gerados pelo sequenciador automático de DNA (ABI/377 – Applied Biosystems), fazendo o “base calling” e atribuindo valores para cada uma das bases, que são proporcionais à qualidade obtida no sequenciamento. O programa Phrap utiliza os dados gerados pelo Phred para a montagem dos contigs de sequência de acordo com a similaridade e a qualidade dos dados de diferentes leituras. Já o Consed consiste em uma interface gráfica que permite a visualização pelo usuário, dos dados gerados.

Estrutura genômica de *hEPC1* e *hEPC2*

Para a delimitação das regiões correspondentes aos introns e exons de ambos os genes, as sequências foram inicialmente comparadas aos fragmentos genômicos contidos no banco de dados HTGS do Genbank. Posteriormente, com a finalização do sequenciamento do genoma humano, as sequências foram novamente re-analisadas através do BLAST contra o genoma humano e também utilizando a ferramenta de análise BLAT. A sequência codificante de *hEPC1* apresenta 3.739 nucleotídeos (nt) distribuídos em 14 exons (espalhados em 139 Kb do *locus* 10p11.22), com uma janela aberta de leitura de 2.292 nt, que codificam para 763 aminoácidos (Tabela 3). *hEPC2* apresenta 3.649 nt com uma janela aberta de leitura de 2.421 nt que codificam para 807 aminoácidos. O gene também possui 14 exons, espalhados em 207 Kb do *locus* 2q23.1. (Tabela 4). As sequências dos sítios aceptores, assim como dos sítios doadores de “splicing” do tipo respectivamente AG e GT, foram identificadas em todos os exons de ambos os genes. Como descrito para a maioria dos genes de mamíferos, em *hEPC1* e *hEPC2*, o último exon (aquele que contém o exon mais a 3´) é o maior. As sequências obtidas para *hEPC1* e *hEPC2* foram depositadas nos bancos de dados públicos do NCBI recebendo os respectivos códigos de acesso: **AF286905** e **AF286904**.

Tabela 3 – Limites de exon e íntron e estrutura genômica do gene *hEPC1*

Exon	Tamanho do exon (nt)	Tamanho do íntron (nt)	Posição no cDNA	Posição no genoma humano	Sítio acceptor	Sítio doador
1	126		1-126	32671550-32671425		AATGgtaagtttgtaa
2	160	72641	127-286	32598865-32598706	TTTtaggaacatcatct	CAGCgtaagtaatgat
3	145	12034	287-432	32586671-32586526	GTcagcttttagtttg	GCAGgtacaacttcat
4	208	397	433-639	32586128-32585922	TTTtagccoagtcagtot	AAAAGtaagtgaaat
5	148	343	640-788	32585578-32585430	TGCagaatcgcaaaaa	AGAGgtaagtgtatatt
6	159	1173	789-948	32584256-32584097	TTTtaggtataaatttgg	TAAggtgattaatgtt
7	176	3938	949-1125	32580208-32580032	TGCagcaagataaagc	CCAGgtacaaaatacaa
8	89	81	1126-1215	32579949-32579860	TAAaggttttgtctgg	TGCTgtaagtttccaga
9	148	83	1216-1364	32579776-32579628	TTAagcctcacttaga	GAAGgtaaggtatttc
10	352	1643	1365-1717	32577984-32577632	AAtaggtcttactgg	TTTGgtatgttgaactg
11	118	11416	1718-1836	32566215-32566097	TTCagcatttacagcc	ACAGgtgagggatttg
12	141	995	1837-1978	32565101-32564960	TTTtaggttttgtttc	TCAggtaaagaataaag
13	363	108	1979-2342	32564851-32564488	TCTaggagtatacaag	CAAGgtgagtagttat
14	1379	2418	2343-3722	32562069-32560685	AACagggaaaaatcatg	

Tabela 4 – Limites entre os exons /íntrons e estrutura genômica do gene *hEPC2*

Exon	Tamanho do exon (nt)	Tamanho do íntron (nt)	Posição no cDNA	Posição no genoma humano	Sítio acceptor	Sítio doador
1	172 nt		1-172	149613337-49613508		ATCGgtagggactoga
2	160 nt	45042	173-332	149658552-149658711	TCCaggaacatcattt	CAGCgtaagtttgtta
3	145 nt	53249	333-478	149711960-149712105	TCTagcttttaatacta	TCAGgtactgtaccat
4	208 nt	10158	479-685	149722265-149722471	TTTtagcttgtaaacact	AAAGgtaatgtgcaaaa
5	150 nt	7647	686-834	149730120-149730268	AATagaatcgtaagaa	AAAGgtaacattgctc
6	135 nt	737	835-967	149731008-149731140	TCAagataaccatttgg	TAAGgtgaatattgtc
7	195 nt	2153	968-1159	149733293-149733484	AACagcaccctcatca	ACAGgtgcttgtttta
8	91 nt	4001	1160-1249	149737489-149737578	TCTaggtattgtoccc	TGCTgtaagaactttt
9	146 nt	1552	1250-1395	149739130-149739275	TGCagcctcgtttgga	GAAGgtgaagtatttg
10	345 nt	104	1396-1739	149739382-149739725	TTCagggtcataatgg	TCAGgtgaagctgttc
11	140 nt	10254	1740-1876	149749982-149750118	ACCagtaacaggggggt	ACAGgtaaactgctct
12	162 nt	1814	1877-2036	149751935-149752094	TATagggctcaagcac	TCAGgtacagctgggg
13	338 nt	906	2037-2370	149753006-149753339	TTCaggaacctctaaa	CAAGgtgtgtgtgtgt
14	1264 nt	1302	2371-3633	149754643-149755905	CATagagagaaccacg	

As proteínas codificadas por *hEPC1* (ID: AF286905) e *hEPC2* (ID: AF286904) são bastante conservadas durante a evolução, desde *Arabidopsis thaliana*, passando por levedura, *C. elegans*, *Drosophila*, camundongo e homem (Figura 5). Esta conservação na sequência protéica, que é pouco usual para um membro da família dos genes Polycomb (PcG), se mostra maior na região dos domínios descritos como EPCa, EPCb e EPCc. Na região de 54 aminoácidos de EPCa de *Drosophila*, 85% são idênticos e 93% são similares aos homólogos de mamíferos.

```

hEPC2 -----
mEPC1-II SSSSVDSVPRENHESEKPALNNIADNTVAMEVTMSKLSFRARALDAAKPLPIYRGKMDPD 840
hEPC1 -----
mEPC1-I -----
fEPC -----MSKLSFRARALDASKPLPVFRCELDLPD 27
dEPC -----MSKLSFRARHLDPKQMPIYLAELDPD 27
cEPC -----MATTSKAFRARALDSNRSMTVYWGHELDPD 29
yEPC SGRSTRRSGRSAHDDGLDSFSKGDGSGAGASAGSSNSRFRHRKISVKQHLKIYLPNDLKH 76

hEPC2 -----MEKEESEHHLQRAISAQVVFREKKES---MVI PVPEAESNVN 40
mEPC1-II LNDCVSINRAVPQMPTGMEKEESEHHLQRAISAQVVFREKKES---MVI PVPEAESNVN 897
hEPC1 -----MEHHLQRAISAQVYGEKRDN---MVI PVPEAESNIA 34
mEPC1-I -----MEHHLQRAISAQVYGEKRDN---MVI PVPEAESNIA 34
fEPC LHEYASINRAVPQMPTGMEKEESEHHLQRAISAHQVYGEKREN---MVI PVPEAESNIP 84
dEPC LPEYSAINRAVPQMPSGMEKEESEHHLQRAICTGLIIPTP-----EVLQTDQP 76
cEPC LSECSVGNRAVTPMPSGMEKEEQEKHLQEAIAAQASTSGIQLNHVIPTPKVDRVEDQR 89
yEPC LDKDELQQREVVETETGVEKNEEKVHLHRILQMGSGHTKHKDY-----IPTPDASMTWN 131
• ***. :
•

hEPC2 YYNRLYKGEFKQPKQFIHIQPFN-LDNEQPDYDMDSEDETLNRLNRK---MEIKPLQFE 96
mEPC1-II YYNRLYKGEFKQPKQFIHIQPFN-LDNEQPDYDMDSEDETLNRLNRK---MEIKPLQFE 953
hEPC1 YYESIYPGEFKMPKQLIHIQPFPS-LDAEQPDYDLSEDEVFNKLNK---MDICPLQFE 90
mEPC1-I YYESIYPGEFRMPKQLIHIQPFPS-LDAEQPDYDLSEDEVFNKLNK---MDICPLQFE 90
fEPC YYDSLYPGEFKMPKQLIHIQPFPS-LDTEQPDYDLSDDDMFVNKLNK---MELSOLQFE 140
dEPC FYDAYYPDYKMPRQMIHQPLG-LDTEVPDYDMSADEDWLSQQ-QR---LELTELKFE 131
cEPC YHSTYHNKMKHRSKYIKVHAWQALERDEPEYDYDTEDEAWLSDH---THIDPRVLE 143
yEPC EYDKFYTGSGFQETTSYIKFSATV-EDCCGTNYNMDERDETFLNEQVNGSSDILTEDEFE 190
: . : . * : . : : : : : : : : : : : : : : : *

hEPC2 IMIDRLEKASSNQLVTL---QEAKLLNE--DDYLIKAVYDYWVRKRKRCRGPS----- 145
mEPC1-II IMIDRLEKASSNQLVTL---QEAKLLNE--DDYLIKAVYDYWVRKRKRCRGPS----- 1002
hEPC1 EMIDRLEKSGSQPVGL---QEAKLLKE--DDELIREVYEWIKRKRKRCRGPS----- 139
mEPC1-I EMIDRLEKSGSQPVSL---QEAKLLKE--DDELIREVYEWIKRKRKRCRGSS----- 139
fEPC EMIDRLEKSGSQLVSL---PEAKLLKE--DDELIKVFDYWSRKRKNSKANS----- 189
dEPC QMMDRLEKSSGQTVVTL---NEAKSLNQ--DDETSISVYDYWLNKRLKMQHP----- 179
cEPC KIFPTVESHSSSETQIAS---EDSVINLHKSLSDSIVYEIYEWLKRKTSAAATSGCVGVG 200
yEPC ILCSSFHAIHERQPFSLMDPESILSFEELKPTLIKSDMADFNLRLNQLNHEINS---HKT 247
: . . * : : : : : : : : : : : : : : : :

hEPC2 -LIPQIKQEKR-DGSTNNDPYVAFRRRTEKMQTRKRNKRNDEASYEKMLKLRREFSRAITI 203
mEPC1-II -LIPQIKQEKR-DGSTNNDPYVAFRRRTEKMQTRKRNKRNDEASYEKMLKLRREFSRAITI 1060
hEPC1 -LIPSVKQEKR-DGSSTNDPYVAFRRRTEKMQTRKRNKRNDEASYEKMLKLRDLRAVTI 197
mEPC1-I -LIPLVKQEKR-DGSSTNDPYVAFRRRTEKMQTRKRNKRNDEASYEKMLKLRDLRAVTI 197
fEPC -LIPTVKQEKR-DGSSTNDPYVAFRRRTEKMQTRKRNKRNDEASYEKMLKLRDLRAVTI 247
dEPC -LILTVKTESR-PGASSNPPYLAFFRRRTEKMQTRKRNKRNDEASYEKMLKLRDLQRAITI 237
cEPC GLIPRVRTCEKRDGQGVINPYVAFRRRAEKMQTRKRNKRNDEDSYEKILKLVHDMASKAQL 260
yEPC HFITQFDPVSMNTRPLIQLIEKFGSKIYDYWRERKIEVNGYEIFPQLKFERPGEKEEID 307
:* . : : * : . . : : : . ** : : . :

hEPC2 LEMIKRREKTKRELLHLTLEVVEKRYHLGDYGG--EILNEVKISR-SEKELYATPAT-LH 259
mEPC1-II LEMIKRREKTKRELLHLTLEVVEKRYHLGDYGG--EILNEVKVNR-SEKELYASPAT-LH 1116
hEPC1 LEMIKRREKTKRELLHLTLEIMEKRYNLGDYNG--EIMSEVMAQRQPMKPTYAIPIT 255
mEPC1-I LEMIKRREKTKRELLHLTLEIMEKRYNLGDYSG--EIMSEVMAQRQPMKPTYAIPIT 255
fEPC LEMVRRREKTKRELLHLTLEVFEKRNFMDFGG--EVMAEVLAEQALVRP-QIIPLVPLT 304
dEPC LEMVRRREKTKRDHLKMTVNI FEKRVEMRDFNG--AVYSELNSQYKNTRPAYNPLYTNQY 295
cEPC FDMTARREKQKLALIDMSEILAKRMEMSDFGGSPSSFNITEKIRAAATLEVVKPPLAE 320
yEPC PYVCFRRREVRHRKTRRIDILNSQ-----RLRALHQELKNAKDLALLVAKREN 356
: ** . : : . . : : : : : : : : : : :

hEPC2 D-----IKQYDFHSSDEDEFPPQVLSPVSEPEEENDPDGPCAFFFFRAGCYYAPRLDQAN- 373
mEPC1-II D-----IKQYDFQSSDEDEFPPQVLSPVSEAEENDPDGSCAFFRRRAGCYYAPRLDQANN 1231
hEPC1 D-----LNQYDFPSSDEEPLSQVLSGSSAEEDNDPDGPFYAFRRKAGCYYAPHLDQTN 372
mEPC1-I D-----LNQYDFPSSDEEPLSQVLSGSSAEEDNDPDGPFYAFRRKAGCYYAPHLDQTN 371
fEPC D-----LNQYDFPSSDEEPLSQVLSGSSAEEDNDPDGAYAFRRKSGCYYAPRQDCVGG 414
dEPC QSASPAANDSIADSEEDYFGAGAQNGLGSESEETPFYAFRRRSSCVYLPTR-QRDGR 474
cEPC R-----FAFKRRRGCVYRAALT-----VYNVPTAPATVPPVQTQAAV 421
yEPC E-----AELRKAELKRAAAAAAAKAKNNKRNQLLEDKSSRLTKQQQQQLLQQQQQQQN 462

```

```

hEPC2      HSCENSEL-----ADLDKLRYPHCLTTLTVPR---RCIGFARRRIGRGGRVIMDRISTE 424
mEPC1-II   HMCENSEL-----ADLDKLRYPHCLTTLTVPR---RCIGFVRRRIGRGGRVIMDRISTE 1282
hEPC1      WPWTSPKD-----GGLGDVRYRYCLTTLTVPQ---RCIGFARRRVGRGGRVLLDRAHSD 423
mEPC1-I    WPWTSPKD-----GGLGDVRYRYCLTTLTVPQ---RCLGFARRRVGRGGRVLLDRAHSD 422
fEPC      WPWCGPGE-----DGLAEARFRYSLTTLTVPR---RCLGMARRRVGRGGRVLLDRAHTD 465
dEPC      YPWDSADEDMAPSGACSDAKYRYTLTSLNYPRP---RCIGFARRRLGRGGRILLDRATTN 531
cEPC      ASSSSSKS-----TDMVPSNMKFFETFVRDSQDSVSRSLGFVRRRMGRGGRVVFDMPRN 476
yEPC      ALKTENGK-----QLANASSSSTSQPITSHVYVKLPSSKIPDIVLEDVDALLNSKEKNA 516

                                     :       .       :       :       . : : . :
hEPC2      HDPVLKQIDPEMLNS----- 439
mEPC1-II   HDPVLKQIDPEMLNG----- 1297
hEPC1      YDSVFHHLDLEMLS----- 437
mEPC1-I    YDSMFHHLDLMLNS----- 436
fEPC      YDNI FHGLDPEVLHPTH----- 483
dEPC      FDDFWSQLDYTVMESVTVNKCRDKIKQRDPDQLCKPSSPPTVVDLVPAGKPVGEGHPVPE 591
cEPC      RD----- 478
yEPC      RKFVQEKME----- 525

.
hEPC2      -----FSSSSQTIDFSSNFSRTNASSKHCENRL-----SL 469
mEPC1-II   -----FSSSSQTIDFSSNFSRTNASSKPCENRL-----SL 1327
hEPC1      -----SPQSPVNOFANTSETNTSDKSFSK-----DL 464
mEPC1-I    -----SPQSPVNOFANTSEPNTSDRSSSK-----DL 463
fEPC      -----PSPRRLTPPPPTTSRSPATDKFASTSETNTSDRISSSFDPSSLSSPSTDL 533
dEPC      LIKEEPSDTEKPVAATESGRESKERELEEKFENDVDDEFVASDDENVSRGLGIYSSAVTL 651
cEPC      -----DNDERTSTDPWAEYCVADSSRTFRARNS-----SLGT 510
yEPC      -----KRKIEDADVFFNLTDPPFNPVFDMSLPKN-----FSTS 558

                                     .       :       .
hEPC2      SEILSNIRSCRLLQCFQPR----- 487
mEPC1-II   SEILSNIRSCRLLQCFQPR----- 1345
hEPC1      SQILVNIKSCRWRHFRPR----- 482
mEPC1-I    SQILVDIKSCRWRHFRPR----- 481
fEPC      SQILLSTKACRWRHFRPR----- 551
dEPC      PQSCADLRQKRRLRKKQLRELNAAKRLKRSSEAVATAEDVQLEDEDTKPDLRPLPRA 711
cEPC      EEETDDLSPKSLYPARNS----- 528
yEPC      NVPFASIASSKFQIDRSFY----- 578

                                     . :       :
hEPC2      -----LLNLQDSSEECTS----- 501
mEPC1-II   -----LLNVQDISEECTS----- 1359
hEPC1      -----TPSLHDSINDELSC----- 496
mEPC1-I    -----TPSLPDSISGELSS----- 495
fEPC      -----TLPLHELDNAHPIF----- 565
dEPC      YLQRLALILGQKLVKVEEDNTEVSENVAPELLEMNDPEQLQPPRANHQQHRSNHNNNNNN 771
cEPC      -----RFAFNDETEREW----- 542
yEPC      -----SHLPEYLKGISDDIR----- 593

```

Figura 5 – Alinhamento mostrando os domínios de *EPC* em diferentes espécies. Em vermelho, o domínio EPcA; em azul o domínio EPcB e em verde o domínio EPcC.

Legenda: *hEPC2*: *EPC2* humano; *mEPC2*: *EPC2* de camundongo; *mEPC1-I*: Isoforma-I de *EPC1* de camundongo; *mEPC1-II*: Isoforma-II de *EPC1* de camundongo; *fEPC*: *EPC* de *Fugu ruprides*; *dEPC*: *EPC* de *Drosophila melanogaster*; *cEPC*: *EPC* de *Caenorhabditis elegans*; *yEPC*: *EPC* de levedura.

Análise dos domínios protéicos de hEPC2

Análises de domínios de proteína foram feitas na sequência de aminoácidos predita de hEPC2, usando as ferramentas do banco de dados de domínios conservados (Conserved Domain Database – CDD: www.ncbi.nlm.nih.gov/Structure/CDD) do NCBI e o PSORT II (www.psort.nibb.ac.jp). Os resultados do CDD indicam a presença de um domínio de “PHD zinc finger” nos aminoácidos 51 a 104, podendo indicar que hEPC2 pode agir como um fator de transcrição (Figura 6). Este dado foi reforçado pelos resultados das análises do PSORT II, que sugeriram a presença de 7 possíveis sinais de localização nuclear: 180-RKRK-183; 198-KRDGSTNNDPYVAFRRR- 214; 224-RKNDEASYEKMLKLRRE-240; 238-RKNDEASYEKMLKLRRE-254; 338-RKNDEASYEKMLKLRRE-344, 339-RKNDEASYEKMLKLRRE-342 and 397-RKNDEASYEKMLKLRRE-403. Estes dados sugerem que a localização sub-celular mais provável da proteína hEPC2 seja o núcleo (91%).



Figura 6 – Domínios preditos pelo CDD para a seqüência de aminoácidos de hEPC2. Em vermelho o domínio que caracteriza os genes “Enhancer of Polycomb”; em azul o domínio EpcC; em cinza o domínio de “PHD zinc finger protein”.

Análise de similaridade da proteína hEPC2 entre diferentes espécies

Aparentemente, sequências similares ao gene *hEPC2* se encontram no genoma de vários eucariotos. A identidade entre as seqüências dos ortólogos ao nível de aminoácidos varia, respectivamente, de 24% e 26% em levedura e *C. elegans*, até 37% em *Drosophila*, 47% em *Fugu rubripes* e 94% em camundongo (Ids: AK31025 e AK 047947). As regiões mais conservadas de *hEPC2*, assim como de *hEPC1* (Shimono *et al.*, 2000) são os domínios descritos em *Drosophila* como EPcA, EPcB e EPcC, principalmente o domínio EPcA (Stankunas *et al.*, 1998). As sequências dos parálogos *hEPC1* e *hEPC2* apresentam uma identidade de 53% ao nível de proteína. No entanto, a similaridade entre as seqüências de nucleotídeos destes genes só é significativa nos domínios EPcA (74% de identidade) e EPcB (80% de identidade), sugerindo uma divergência antiga entre os dois genes. Uma comparação entre os EPcs humano e de camundongo, mostra que a conservação é mais alta entre os ortólogos do que entre os parálogos, reforçando que o evento de duplicação gênica pode ter precedido a divergência funcional entre *hEPC1* e *hEPC2* (Figura 7).

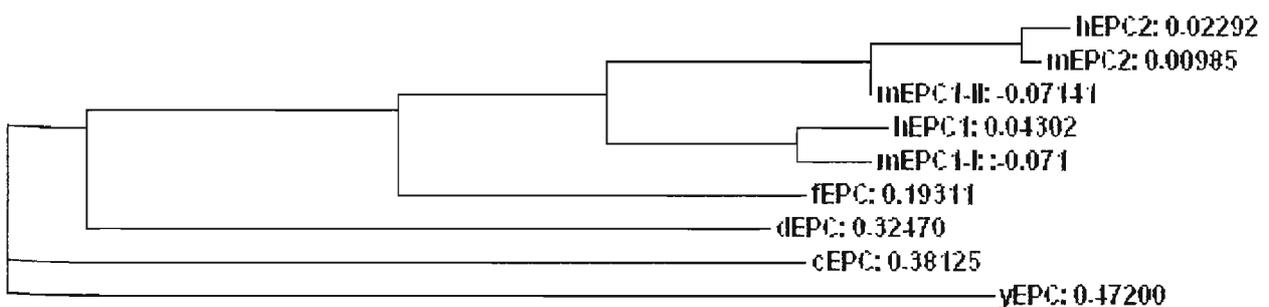
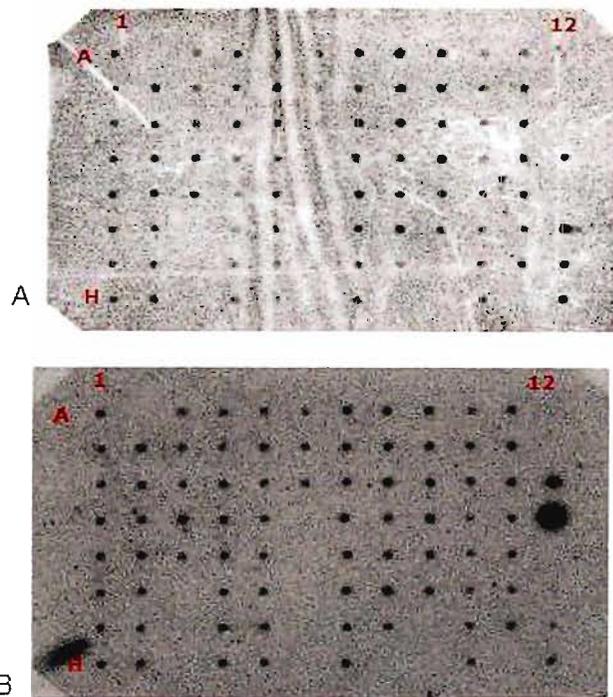


Figura 7 – Filograma mostrando a similaridade entre a seqüência de aminoácidos dos homólogos de *EPc* em diferentes espécies.

Legenda: *hEPC2*: *EPC2* humano; *mEPC2*: *EPC2* de camundongo; *mEPC1-I*: Isoforma-I de *EPC1* de camundongo; *mEPC1-II*: Isoforma-II de *EPC1* de camundongo; *fEPC*: *EPc* de *Fugu rubripes*; *dEPC*: *EPc* de *Drosophila melanogaster*; *cEPC*: *EPc* de *Caenorhabditis elegans*; *yEPC*: *EPc* de levedura.

Análise da expressão de *hEPC1* e *hEPC2*

Para a determinação do padrão de expressão dos transcritos canônicos de *hEPC1* e *hEPC2*, foi utilizada uma membrana comercial de dot-blot (Multiple Tissue Expression Array, Clontech) contendo 76 amostras de mRNAs de diferentes tecidos, incluindo tecidos fetais, normais e tumorais. As sondas utilizadas nestes experimentos foram desenhadas na região 3'UTR dos genes, de modo a não possuírem similaridade significativa entre si. A expressão de ambos os genes *hEPC1* e *hEPC2* se mostrou fraca e ubíqua em todas as amostras avaliadas (Figura 8). Os controles de DNA humano e de DNA de *E.coli* que se mostraram positivos nos dois experimentos, devem ser provavelmente devido a alta similaridade entre estes DNAs genômicos e as sondas de cDNA usadas nos experimentos.



Cérebro total		Substância negra	Coração	Esôfago	Cólon transverso	Rim	Pulmão	Fígado	Leucemia HL-60	Cérebro fetal	RNA total de levedura
Córtex cerebral	Cerebelo direito	Núcleo acúmbens	Aorta	Estômago	Cólon descendente	Músculo esquelético	Placenta	Pâncreas	Hela S3	Coração fetal	tRNA de levedura
Lobo frontal	Corpo caloso	Tálamo	Átrio esquerdo	Duodeno	Reto	Baço	Bexiga	Glândula adrenal	Leucemia K-562	Rim fetal	RNA de E.coli
Lobo parietal	Amígdala	Glândula pituitária	Átrio direito	Jejuno		Timo	Útero	Tireóide	Leucemia MOLT-4	Fígado fetal	DNA de E.coli
Lobo occipital	Núcleo caudato	Medula espinhal	Ventrículo esquerdo	Íleo		Linfócitos de sangue periférico	Próstata	Glândula salivar	Linfoma Burkitt Raji	Baço fetal	Poli r(a)
Lobo temporal	Hipocampo		Ventrículo direito	Íleoceco		Linfonodo	Testículo	Glândula mamária	Linfoma Burkitt Daudi	Timo fetal	DNA Cot-1 humano
p.g.* do córtex cerebral	Medula oblongata		Septo interventricular	Apêndice		Medula óssea	Ovário		Adenocarcinoma colorretal SW480	Pulmão fetal	DNA humano 100 ng
ponte	Putâmen		Ápice do coração	Cólon ascendente		Traquéia			Carcinoma de pulmão A549		DNA humano 500 ng

Figura 8 – Análise de expressão de *hEPC1* (A) e *hEPC2* (B) em 76 amostras de mRNA de tecidos fetais, adultos e de linhagens celulares tumorais (Multiple tissue cDNA [MTC] panel, Clontech).

Identificação de variantes de “splicing” de *hEPC2*

O artigo que descreve o gene *EPC* em *Drosophila* sugere a presença de isoformas menores dos ortólogos de camundongo (*Epc1* e *Epc2*). Um transcrito menor de *hEPC1* humano também foi sugerido por (Shimono *et al.*, 2000). No entanto, nem sua seqüência, nem sua função foram determinadas. Baseado nessas indicações, foram utilizadas ferramentas de Bioinformática para a investigação da possível existência de transcritos alternativos de *hEPC2*. O uso do Acembly, uma ferramenta que prediz a existência de variantes de “splicing” disponibilizada na página da web do NCBI, sugeriu a existência de duas outras isoformas de *hEPC2*. Um destes transcritos, chamado Isoforma B (Iso-B), foi predito como tendo 1,686 nt distribuídos em três exons. Os dois primeiros exons possuem as mesmas coordenadas genômicas da isoforma canônica de *hEPC2* (chamada Iso-A), e o último exon de 1.302 nt mapeia em uma região genômica distinta (Tabela 5). A longa região 3' não traduzida (UTR) é descrita pelo programa Acembly por ser maior que 95% das 3'UTRs encontradas nos transcritos humanos (Lander *et al.*, 2001). A outra isoforma, identificada como Isoforma C (Iso-C), foi predita por ter 1,435 nt distribuídos em dois exons. O primeiro exon de Iso-C corresponde ao segundo exon de ambas Iso-A e Iso-B, e o último exon possui um “gap”, com uma perda de 79 nucleotídeos em relação a Iso-B (Tabela 6). A região N-terminal da seqüência predita de aminoácidos de Iso-C está incompleta, já que a janela de leitura da região 5'UTR permanece aberta.

Tabela 5 - Limites entre os íntrons/exons e estrutura genômica de Iso-B do gene *hEPC2*

Exon	Tamanho do exon	Tamanho do intron	Posição no cDNA	Posição no genoma humano	Sítio acceptor	Sítio doador
1	172 nt		1-172	149613337-149613508		ATCGgtagggactcga
2	160 nt	45042	173-332	149658552-149658711	TCCaggaacatcattt	CAGCgtaagtttgta
3	1353 nt	1169	333-1686	149659880-149661233	TGCagtgaaattgcat	

Tabela 6 - Limites entre os íntrons/exons e estrutura genômica de Iso-C do gene *hEPC2*

Exon	Tamanho do exon	Tamanho do intron	Posição no cDNA	Posição no genoma humano	Sítio acceptor	Sítio doador
1	160 nt		1-160	149658552-149658711		CAGCgtaagtttgta
2	1265 nt	1247	161-1435	149659959-149661233	GCCaggcatttgggga	

Para a confirmação da existência das isoformas B e C e avaliação de sua expressão, foi utilizada uma membrana de Northern-blot contendo 12 amostras de mRNAs derivados de tecidos normais (Clontech, USA). Uma sonda contendo 500 nt de seqüência comum entre as isoformas foi utilizada. Usando esta abordagem duas bandas, sendo uma de 3.6 Kb (Iso-A) e outra de aproximadamente 2.8 Kb foram detectadas. No entanto, bandas correspondentes a Iso-B e Iso-C não foram detectadas, mesmo após uma longa exposição de 72 horas (Figura 9).

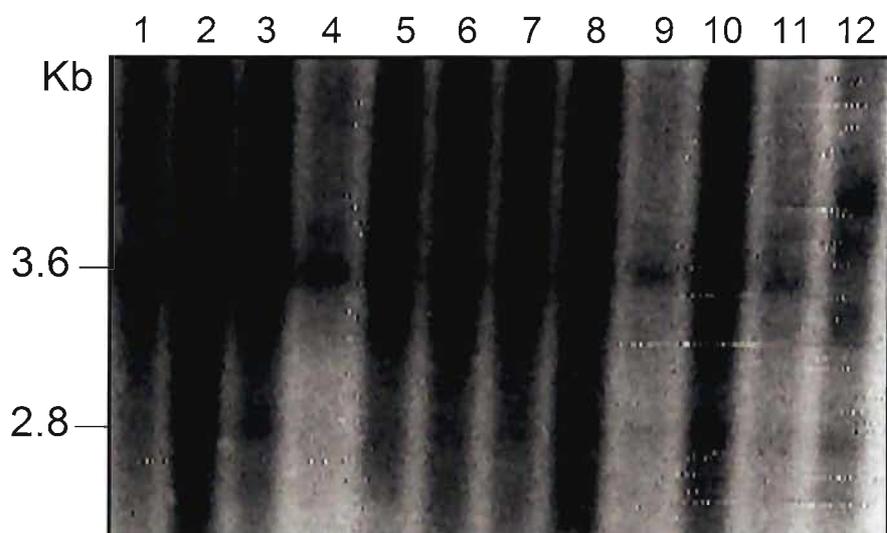


Figura 9 - Northern-blot de tecidos normais (Multiple tissue Northern-blot, Clontech) ensaiados para a análise de Iso-B e Iso-C. A banda predominante (3.6 Kb) representa a isoforma canônica de *hEPC2*. A membrana foi superexposta por 72 horas. A ordem das amostras de mRNA na membrana é: 1- cérebro; 2-coração; 3- músculo esquelético; 4-cólon; 5-timo; 6-baço; 7-rim; 8-fígado; 9-placenta; 10-intestino delgado; 11-pulmão; 12- leucócitos de sangue periférico.

Para testar a veracidade das predições de Iso-B e Iso-C, um experimento de RT-PCR foi planejado, e as isoformas foram avaliadas em um painel de cDNAs não tumorais. Entre as Isoformas B e C existe uma diferença de tamanho de 79 nt entre o primeiro e o segundo exon de Iso-C, permitindo uma análise semi-quantitativa por RT-PCR. Para a normalização das amostras de mRNAs/cDNAs, foi feita a amplificação de uma porção do cDNA do GAPDH. Diferentes números de ciclos de PCR foram usados na amplificação das isoformas para garantir que as reações ocorreriam na fase exponencial, para desta maneira, permitir a estimativa dos níveis de expressão relativa das distintas isoformas entre os diferentes tecidos (Figura 10). As predições do Aceview foram confirmadas com sucesso e duas bandas foram observadas na maioria dos tecidos analisados, As regiões diferentes entre os cDNAs das isoformas foram confirmadas por sequenciamento.

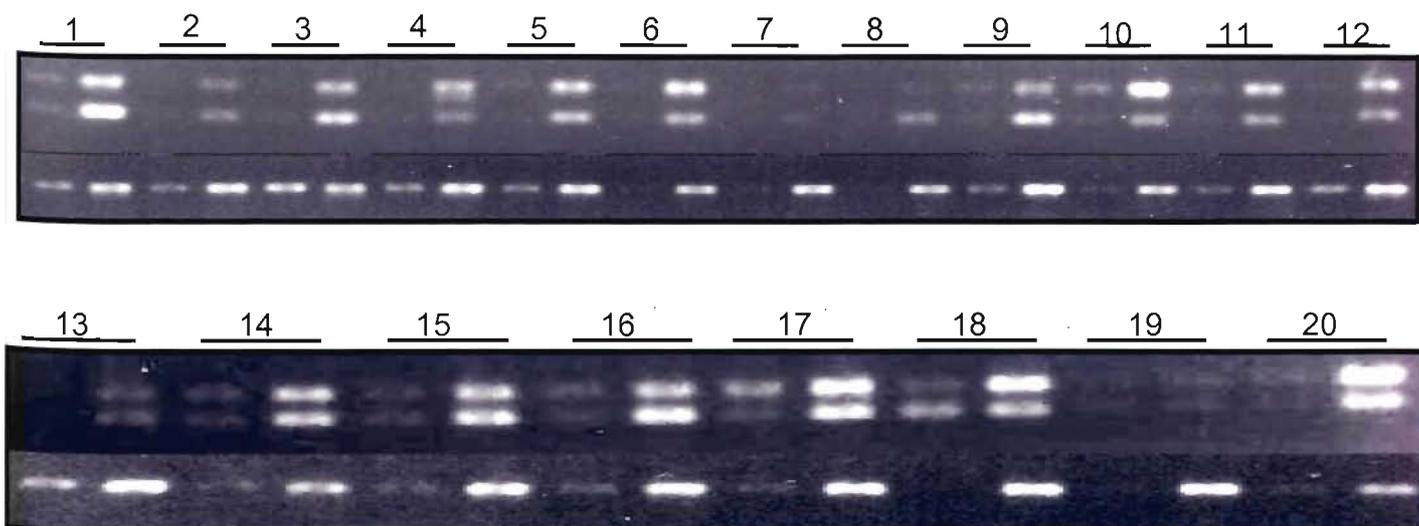


Figura 10 - RT-PCR semi-quantitativo de Iso-B e Iso-C. O duplete de bandas mostra Iso-B (banda mais alta) e Iso-C (banda mais baixa) co-amplificados com, respectivamente, 29 e 33 ciclos de amplificação. A banda mais baixa da figura corresponde à amplificação de GAPDH com 22 e 26 ciclos de PCR.

Legenda: 1: Cérebro total; 2: Coração; 3: Músculo esquelético; 4: Cólon; 5: Timo; 6: Baço; 7: Rim; 8: Fígado; 9: Intestino delgado; 10: Placenta; 11: Pulmão; 12: Medula óssea; 13: Cérebro fetal; 14: Fígado fetal; 15: Medula espinhal; 16: Testículo; 17: Útero; 18: Traquéia; 19: Glândula salivar; 20: Glândula adrenal.

As RT-PCRs mostraram a presença das bandas correspondentes à amplificação de Iso-B e Iso-C em quase todos os tecidos testados, exceto nas amostras de rim e glândula salivar, onde ambas as isoformas estão ausentes.

Análise de Bioinformática e identificação da isoforma D de *hEPC2*

Considerando a detecção de uma banda mais evidente no Northern-blot utilizado na investigação de Iso-B e Iso-C, foram feitas análises *in silico* para avaliação de um transcrito putativo adicional de *hEPC2*. Algumas análises de Bioinformática foram feitas no intuito de investigar a existência e a estrutura deste transcrito, como análise dos íntrons de *hEPC2* contra o banco de ESTs humanas, na tentativa de encontrar ESTs que sugerissem a presença de diferentes transcritos. Essa abordagem e outras como o uso da ferramenta Acembly não sugeriram a presença deste transcrito menor de *hEPC2*. No entanto, análises utilizando a ferramenta BLAT revelaram a existência de um mRNA (AK001433) e 5 ESTs 3' (BM6677033; AU150585, CA338207, BF679822; AI052446) com similaridade com *hEPC2*. Quando comparadas com a seqüência canônica de *hEPC2*, estas seqüências mostraram uma 3'UTR mais curta, compatível como o tamanho da banda observada no Northern-blot de análise de Iso-B e Iso-C. Desta maneira, estas seqüências sugerem a existência de um transcrito alternativo de *hEPC2* que possivelmente utiliza um sítio alternativo de poli-adenilação. Para análise do uso do sítio alternativo de poli-adenilação, foi utilizada a ferramenta de análise SAGE Genie (<http://cgap.nci.nih.gov/SAGE>). SAGE é um método desenvolvido para a análise quantitativa da expressão gênica, através do sequenciamento de fragmentos curtos (usualmente de 10 nt) de um grande número de moléculas de mRNAs representando seqüências concatenadas. O programa SAGE-Genie usa um novo método analítico de alinhamentos confiáveis de seqüências SAGE (SAGE tags) com genes conhecidos. Esta análise permite a visualização da expressão gênica em tecidos ou bibliotecas individuais. Usando a opção chamada "Ludwig Transcript Viewer", é possível observar sítios alternativos de poli-adenilação, além de prováveis sítios internos de ligação de oligo-dT (que levariam a uma falsa representação de transcritos mais curtos, que poderiam

ser confundidos com o uso de sítios alternativos de poli-adenilação). Através do uso desta ferramenta para a análise de Iso-A, observamos a existência de 4 SAGE-tags virtuais, sendo que dois deles se mostraram reais, tendo sido muito freqüentemente encontrados em diversas bibliotecas SAGE. Estes dois SAGE “tags” gerados através de restrição com a enzima *Hinf*III, foram identificadas na 3’UTR do gene *hEPC2*. O “tag” correspondente a isoforma mais longa, Iso-A, é CAAATTTAAT localizado na posição 3,059 do transcrito, com uma freqüência de 12 “tags” por 200,000 e utiliza o sítio de poli-adenilação AATAAA, encontrado 542 nt após do SAGE “tag”. O segundo “tag”, TTTACATTTT, está localizado na posição 2,804 de Iso-A e possui 16 “tags” por 200,000. As seqüências de cDNA correspondentes a isoforma mais curta aparentemente utilizam um sinal de poli-adenilação com a mesma seqüência do anterior, e este se localiza a 12 nucleotídeos do “tag” e 15 nucleotídeos à frente se encontra o final do transcrito. A mesma membrana de Northern-blot usada nas análises de expressão de Iso-B e Iso-C, além de outra membrana contendo mRNA de 8 linhagens tumorais foram usadas para confirmar a existência deste transcrito menor de *hEPC2*. A sonda usada nestes experimentos é feita por uma seqüência exclusiva de Iso-A e Iso-D, correspondendo a uma porção da 3’UTR, comum a ambos os transcritos.

As análises de Northern-blot revelaram a presença de um transcrito de 3.6 Kb e novamente uma banda de aproximadamente 2.8 Kb, compatível com o tamanho de Iso-A e com o transcrito menor de *hEPC2*, possivelmente derivado do uso de um sítio alternativo de poli-adenilação (Figura 11). Esta isoforma, denominada isoforma D (Iso-D) é expressa em todos os tecidos testados, incluindo tecidos normais e linhagens celulares tumorais. Em tecidos normais, Iso-A é mais expressa na amostra de mRNA de coração e músculo esquelético quando comparada com Iso-D, mas em rim, Iso-D é mais abundante do que Iso-A. No entanto, em amostras de cérebro, cólon, timo, baço, fígado, placenta, pulmão e leucócitos de sangue periférico, ambas as isoformas foram fracamente expressas e são muito similares em abundância. Curiosamente, uma alta expressão de Iso-D foi observada em todas as linhagens tumorais avaliadas, incluindo HeLa S3, leucemia mielóide crônica K-562, leucemia linfoblástica MOLT-4, adenocarcinoma coloretal SW480,

carcinoma de pulmão A549 e melanoma G-361. No entanto, em linfoma de Burkkit e leucemia promielocítica HL-60, a expressão de Iso-D foi marcadamente mais alta.

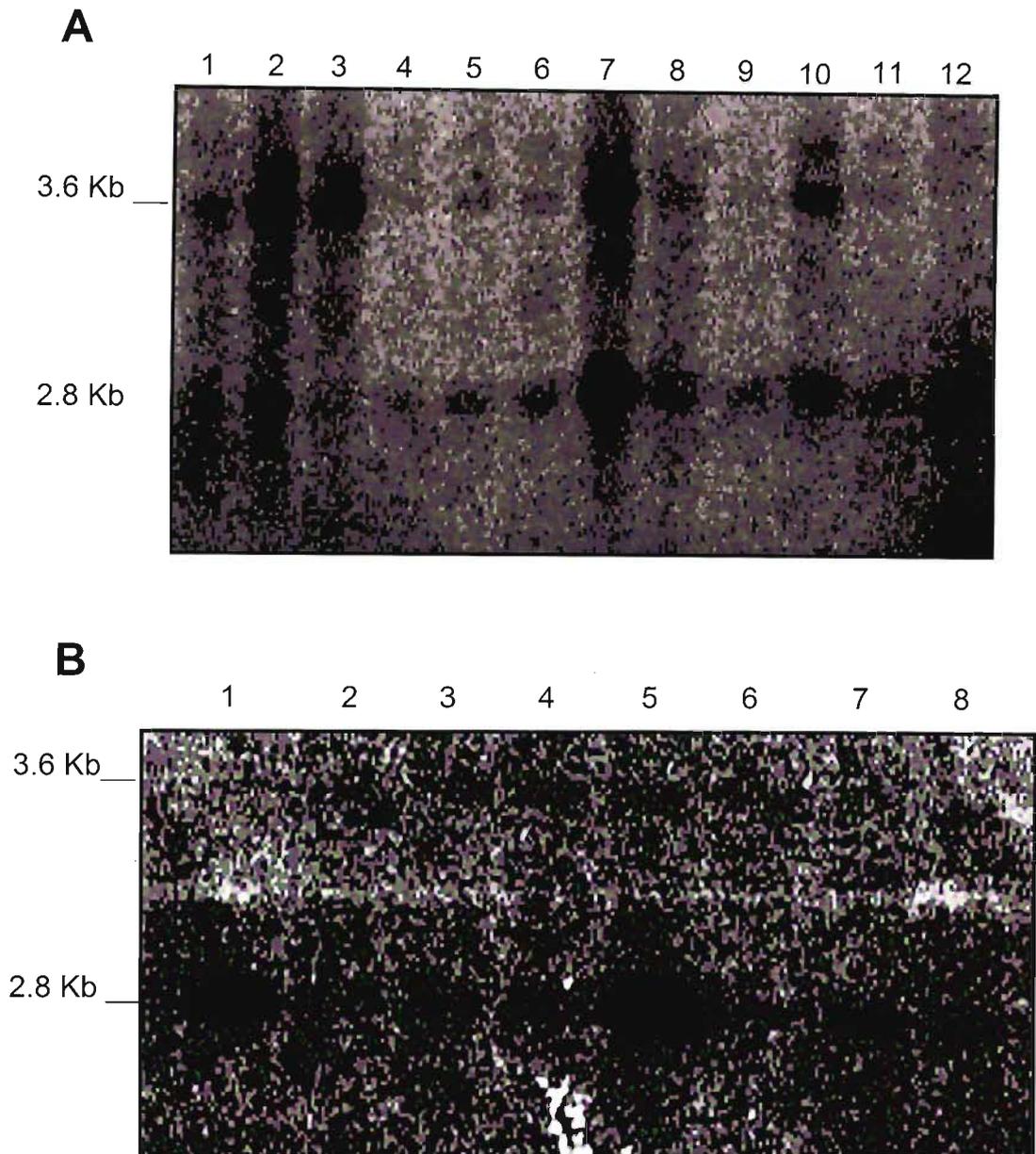
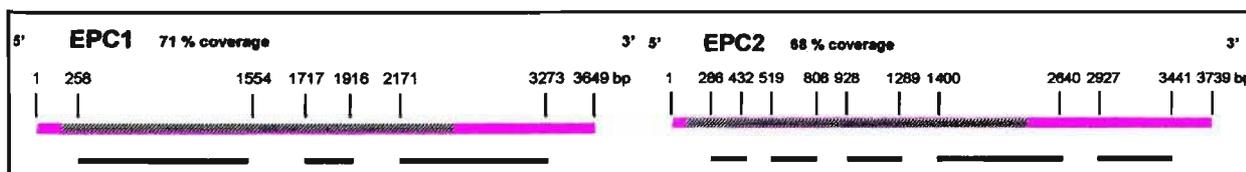


Figura 11 - Northern-blots de tecidos normais (A) e linhagens celulares tumorais (B) sondadas para *hEPC2*. Ambas as figuras mostram a presença de Iso-A e Iso-D, com respectivamente 3.6 Kb and 2.8 Kb. Em **A** os tecidos são: 1-cérebro; 2-coração; 3-músculo esquelético; 4-cólon; 5-timo; 6- baço; 7-rim; 8-fígado; 9-placenta; 10-intestino delgado; 11-pulmão; 12-leucócitos de sangue periférico. Em **B** as linhagens celulares tumorais humanas são: 1-leucemia promielocítica HL-60; 2-He-la S3; 3-leucemia mielóide crônica K-562; 4-leucemia linfoblástica MOLT-4; 5- linfoma de Burkitt's Raji; 6- adenocarcinoma coloretal

DISCUSSÃO

hEPC1 e *hEPC2* representam dois dos primeiros genes humanos completos seqüenciados no Brasil. A seqüência completa destes transcritos foi obtida a partir de dados fragmentados, derivados de ORESTES do Projeto Genoma do Câncer Humano (HCGP). Pode-se observar que as ORESTES representantes de fragmentos destes genes, geradas durante o HCGP, oferecem uma boa cobertura do transcrito, especialmente de sua região codificadora (Figura 12). A distribuição espacial destas ESTs nestes transcritos foi utilizada para demonstrar que a coleção de seqüências ORESTES, aliadas às demais ESTs disponíveis poderia ser de grande valor para determinar a seqüência primária de novos transcritos humanos, com especial ênfase em transcritos raros, como é o caso destes genes. Esta abordagem foi publicada em nosso primeiro artigo, onde pela primeira vez demonstramos a seqüência completa de *hEPC1* e *hEPC2* (Camargo *et al.*, 2001). O artigo completo se encontra como o **Anexo 1** desta tese.

Figura 12 - Representação da cobertura de contigs de ORESTES em relação aos cDNAs dos genes *hEPC1* e *hEPC2* (retirada de Camargo *et al.*, 2001).



Genes ortólogos de *hEPC1* e *hEPC2* são encontrados em muitos organismos, sugerindo uma importante função biológica que presumivelmente foi mantida ao longo da evolução. Apenas o gene *E(z)* da família dos PcG havia sido descrito anteriormente por possuir homólogo em levedura. É sugerido na literatura que talvez os PcG conservados em levedura e nematódeos tenham co-evoluído de uma função primitiva de repressão transcricional para a função de regulação da cromatina, a qual se tornou mais elaborada em eucariotos superiores e adaptada a regular diferentes alvos em diferentes organismos (Stankunas *et al.*, 1998).

A descrição de *EPc* de *Drosophila* mostra que o gene é ubiquamente expresso e se liga a cromossomos politênicos em aproximadamente 100 sítios, apenas um terço dos sítios de ligação dos PcG. Interessantemente, *EPc* não é detectado no cromocentro da heterocromatina, apoiando o modelo em que *EPc* possui um papel funcional (e não estrutural) na heterocromatina, e a conclusão de que existe uma menor sobreposição de mecanismos de formação de heterocromatina e repressão mediada pelos PcG do que previamente suposto.

Assim como todas as proteínas PcG até agora caracterizadas, exceto *E(z)* (Carrington & Jones, 1996), a ligação de *EPc* é discreta, não existindo ligação inespecífica ("background"), sugerindo que *EPc* funciona regulando *loci* específicos. No entanto, mesmo que os sítios de ligação de *EPc* nos cromossomos mostrem sobreposição limitada com proteínas PcG individuais, pelo menos metade dos sítios de ligação de *EPc* são compartilhados com sítios caracterizados de ligação de PcG. Estes dados sugerem que *EPc* é realmente um membro da família dos PcG (Stankunas *et al.*, 1998).

Já que *EPc* pode se ligar aos *loci* alvo na ausência de proteínas PcG conhecidas, *EPc* pode interagir com um diferente sub-grupo de proteínas de cromatina em relação às proteínas PcG previamente caracterizadas. Apenas sete sítios de ligação politênicos são comuns entre todas as proteínas PcG, enquanto que o número de ligação, compartilhados e únicos, de proteínas PcG identificados até então totaliza 240, dando forte embasamento para a heterogeneidade dos complexos de PcG existente. Parece provável que *EPc* funcione como um membro repressivo de complexos PcG em *loci* que apresentam sítios de ligação que se sobrepõem com aqueles de outras proteínas PcG, sendo provável que *EPc* tenha uma função repressiva similar em *loci*-alvo não ligados a proteínas PcG conhecidas, mas isso ainda precisa ser melhor demonstrado. Não se pode assumir que os *loci* de ligação de proteínas PcG em cromossomos politênicos sejam de repressão funcional, porque existem evidências de proteínas PcG que se ligam mesmo quando os genes são transcritos (Strutt & Paro, 1997)) e proteínas PcG que são necessárias para a ativação de transcrição (Fauvarque *et al.*, 1995).

Experimentos mostrados por (Stankunas *et al.*, 1998), não sugerem as possibilidades de que *EPc* afete diretamente a compartimentalização nuclear

da cromatina em compartimentos ativos e inativos (Schlossherr *et al.*, 1994; Strouboulis & Wolffe, 1996), ou que EPc seja necessário para o estabelecimento de uma arquitetura necessária à existência de repressão. No entanto, a observação de que EPc é uma proteína de cromatina de distribuição limitada, torna provável que ela regule genes, podendo portanto, ter um efeito indireto na arquitetura nuclear. A ligação discreta de EPc nos cromossomos politênicos torna improvável que esta proteína tenha um papel geral na estrutura cromossômica ou da cromatina, que afete indiretamente a formação da heterocromatina e repressão pelas proteínas PcG, como a S-adenosilmetionina. Têm sido sugerido que EPc tem um efeito indireto na formação da heterocromatina, e portanto no efeito posicional variegação. Provavelmente por regular genes necessários para a formação da heterocromatina.

Interessantemente, o gene ortólogo de levedura de *EPc*, nomeado *EPL1* ("Enhancer of Polycomb like protein") foi identificado como um componente do complexo de histona acetiltransferase NuA4 (Galarneau *et al.*, 2000). O complexo humano TIP60, homólogo de NuA4, inclui uma das proteínas homólogas de EPc (o código de acesso desta proteína não é descrito no trabalho). Este homólogo humano de *EPc* também se encontra presente em p400, um complexo multiprotéico com forte homologia com o complexo SWI2/SNF2 (Fuchs *et al.*, 2001). Já que se suspeita que EPc promove o silenciamento do estado transcricional da cromatina e formação de heterocromatina, é possível que a presença desta proteína humana homóloga a EPc no complexo p400 possa ter um efeito inibitório. Já o complexo NuA4 está relacionado com ativação transcricional *in vitro* e o gene *EPL1* é descrito como essencial para a viabilidade celular, o que reforça o papel desta proteína como um regulador transcricional epigenético e sugere que a função da proteína pode ter sido mantida nos homólogos eucarióticos (Galarneau *et al.*, 2000). De fato, a função dos PcG parece ser bem conservada durante a evolução. Quando o gene PcG Ying Yang 1 (YY1), que codifica para a proteína de vertebrados homóloga à proteína de *Drosophila* "pheihomeotic" (PHO), foi inserida em *Drosophila*, sendo que esta proteína foi capaz de compensar o fenótipo de PHO, reprimindo a transcrição de uma maneira consistente com a

função de PcG in vivo (Atchison *et al.*, 2003). Portanto, apesar da presença de apenas dois domínios conservados entre as proteínas parálogas de *Drosophila* e a humana, sua função parece ser essencialmente a mesma. Outro exemplo é do gene *M33*, o homólogo de camundongo do gene de *Drosophila* Polycomb (*Pc*), que foi capaz de resgatar o fenótipo de mutantes de *Pc* em moscas *Drosophila* (Muller *et al.*, 1995).

Os genes *hEPC1* e *hEPC2* são bastante conservados durante a evolução, desde levedura, passando por *Arabidopsis thaliana*, *C. elegans*, *Drosophila*, camundongo e homem. Esta conservação na seqüência protéica é pouco habitual para um membro da família dos genes Polycomb (PcG), e se mostra maior na região dos domínios descritos como EPcA, EPcB e EPcC. Na região de 54 aminoácidos de EPcA de *Drosophila*, 85% são idênticos e 93% são similares aos homólogos de mamíferos.

No entanto, uma comparação entre as seqüências de nucleotídeos de *hEPC1* e *hEPC2* mostra que estes genes não são tão similares entre si como seus ortólogos, o que pode indicar que estas proteínas sofreram uma divergência funcional. O alinhamento da seqüência da proteína EPc de *Drosophila* com as seqüências preditas de aminoácidos de *hEPC1* e *hEPC2* mostra que *hEPC1* é um pouco mais similar a EPc (com 39% de identidade e 57% de positivos) que *hEPC2* (17% de identidade e 53% de positivos). Este dado provavelmente sugere que o papel ancestral do gene *EPc* pode ser mais conservado em *hEPC1* do que em *hEPC2*. Uma evidência da divergência funcional entre *hEPC1* e *hEPC2* é a ausência de um domínio “zinc finger” em *hEPC1*.

Apesar da possível divergência de função entre *hEPC1* e *hEPC2*, é provável que algumas propriedades tenham sido mantidas entre estas proteínas, especialmente em relação à estrutura dos domínios EPcA, e EPcC (Stankunas *et al.*, 1998). Os exons 2, 3 e 4 de *hEPC1* e *hEPC2* têm exatamente o mesmo tamanho e estão contidos no domínio EPcA, sugerindo a conservação funcional desta região em comum entre as proteínas (Tabela 7). Interessantemente, uma proteína chamada BR140 foi recuperada quando o domínio EPcA foi usado para uma busca por proteínas similares. Esta proteína possui uma significativa conservação quando comparada com uma pequena

porção da seqüência de aminoácidos de EPcA, sugerindo que este domínio seja modular e que suas diferentes porções podem ter funções distintas (Stankunas *et al.*, 1998). A predição de um domínio de "PHD zinc finger" dentro do domínio EPcA de EPC2 é um achado que também pode corroborar a hipótese de organização modular de EPcA.

Tabela 7 – Comparação entre o tamanho dos exons de *hEPC1* e *hEPC2*

Exon	Tamanho dos exons de <i>hEPC1</i> (nt)	Tamanho dos exons de <i>hEPC2</i> (nt)
1	126	172
2	160	160
3	145	145
4	208	208
5	148	150
6	159	135
7	176	195
8	89	91
9	148	146
10	352	345
11	118	140
12	141	162
13	363	338
14	1379	1264

Comparação da expressão de *hEPC1* e *hEPC2* em diferentes tecidos humanos

A expressão de *hEPC1* e *hEPC2* foi investigada em diferentes tecidos através do uso de uma membrana comercial de "dot-blot" (Multiple Tissue Expression Array - Clontech), contendo 76 amostras de mRNAs derivadas de tecidos e linhagens celulares fetais, normais e tumorais. Curiosamente, a expressão de ambos os genes se mostrou fraca e ubíqua em todos os tecidos avaliados, não revelando portanto, diferenças significativas entre o padrão de expressão dos genes parálogos.

Uma consequência das varreduras de vários ensaios de duplo híbrido com proteínas PcG nos anos recentes, mostra que as proteínas PcG parecem existir aos pares em vertebrados. (Satijn & Otte, 1999a). O gene de *Drosophila* Ph possui os homólogos humanos *HPH1* e *HPH2* (Gunster *et al.*, 1997). Estes homólogos humanos possuem seus ortólogos em camundongo, respectivamente, *Rae-28/Mph1* e *Mph2* (Alkema *et al.*, 1997a; Hemenway *et al.*, 1998; Nomura *et al.*, 1994). Este exemplo pode ser estendido para todas as proteínas PcG de vertebrados conhecida até então. *HPC2* é o um homólogo humano de Pc (Satijn *et al.*, 1997), que é mais relacionado como ortólogo murino *MPc2* (Alkema *et al.*, 1997b) e o ortólogo de *Xenopus* de Pc *XPc* (Reijnen *et al.*, 1995), que o ortólogo murino de Pc, *M33* (Pearce *et al.*, 1992). De maneira contrária, *M33* é mais similar ao homólogo humano de Pc, *CBX1/HPC1* (Gecz *et al.*, 1995; Satijn *et al.*, 1997). *Bmi1* possui um similar próximo, *mel-18* (Tagawa *et al.*, 1990), que apresenta identidade maior que 95%. Assim como *Bmi1*, foi demonstrado que *mel-18* interage com *Rae-28/Mph1* (Alkema *et al.*, 1997a). *EZH2* é altamente relacionado com seu parálogo *EZH1* (Laible *et al.*, 1997). Finalmente, os dois homólogos de *RING* foram identificados. *RING1* humano e *dinG* são homólogos, respectivamente, à *Ring1A* e *Ring1B* de camundongo (Freemont *et al.*, 1991; Lovering *et al.*, 1993; Schoorlemmer *et al.*, 1997). O que pode significar esta aparente redundância?

Análises de expressão dos pares de proteínas PcG acima mencionados mostraram que estas proteínas são diferencialmente distribuídas em tecidos e linhagens celulares (Buchenau *et al.*, 1998; Gunster *et al.*, 1997; Lessard *et al.*, 1998). Por exemplo, o gene *HPH2* é expresso em virtualmente todos os tecidos testados em um nível uniforme, enquanto que *HPH1* tem uma distribuição tecidual bem restrita, limitada a testículo, ovário e timo (Gunster *et al.*, 1997). Também em linhagens celulares, *HPH2* possui uma expressão uniforme em todas as linhagens, enquanto que *HPH1* é expresso em altos níveis em apenas algumas das linhagens em que *HPH2* é expresso (Gunster *et al.*, 1997). Observações similares foram feitas para *HPC1* e *HPC2* (Satijn *et al.*, 1997). Estes dados sugerem que, dependendo do tipo celular ou tecido, os complexos PcG podem conter tanto ambos, como apenas uma das proteínas do par. Por exemplo, em testículo, ambos *HPH1* e *HPH2* são expressos em altos níveis,

portanto é provável que o complexo PcG contenha ambas as proteínas. De maneira oposta, em baço, apenas *HPH2* é expresso, e conseqüentemente, o complexo PcG provavelmente possui como componente apenas *HPH2* (Gunster *et al.*, 1997).

Esta especulação pode ser entendida de outra maneira, ou seja: os pares de proteínas são também capazes de interagir com diferentes proteínas através de regiões não conservadas. Dois exemplos serão discutidos a seguir.

Os pares de homólogos de *RING1* em vertebrados, *RING1* e *dinG* (ambos humanos), *Ring1A* e *Ring1B* (ambos de camundongo), mostram um alto nível de similaridade de seqüência. Eles possuem um motivo de "RING finger" em comum e são ambos capazes de interagir com os homólogos de Pc e Bmi-1. No entanto, análises de interação proteína-proteína feitas com a proteína de camundongo Ring1B mostrou que esta proteína é capaz de interagir com o homólogo de camundongo de Ph, MPh2 (Hemenway *et al.*, 1998). Em contraste, as proteínas RING1/Ring1A não foram capazes de interagir com os homólogos de Ph, HPH1 e HPH2 (Satijn & Otte, 1999b). Este domínio de interação entre MPh2 e Ring1B está localizado nas regiões central e C-terminal da proteína Ring1b (Hemenway *et al.*, 1998). De maneira importante, esta região não é conservada entre Ring1A e RingB, o que explica porque Ring1A não interage com HPH1 e HPH2. Este achado implica que quando Ring1A e Ring1B são expressos em diferentes tecidos, eles são capazes de recrutar diferentes proteínas PcG, como Mph2/HPH2 por Ring1B neste caso.

Outro exemplo provém de uma varredura utilizando ensaios de duplo-híbrido, que foi feita com os homólogos de *Pc* em vertebrados, *Xpc* e *HPC2*. Estes ortólogos interagem com a região de ligação C-terminal da proteína (CtBP) (Sewalt *et al.*, 1999). CtBP é uma proteína com uma função desconhecida que foi primeiramente identificada como uma proteína que se liga a região C-terminal da proteína de adenovírus E1A (Schaeper *et al.*, 1995). O ortólogo de *Drosophila* de CtBP (dCtBP) interage com diversas proteínas de segmentação, que atuam como repressores, através de um motivo que também está envolvido com interações entre CtBP e E1A (Schaeper *et al.*, 1995). CtBP se liga a HPC2 e XPc através da mesma região conservada de

motivos de aminoácidos, PI/LDL. Este motivo está presente apenas em Xpc, MPc2 e HPC2, os ortólogos relacionados a Pc que representam uma das classes dos homólogos de Pc. Os motivos de ligação de CtBP estão ausentes no homólogo de camundongo Pc. M33 e HPC1, que representam a outra classe de homólogos de Pc. Conseqüentemente, CtBP não interage com M33 ou HPC1 (Sewalt *et al.*, 1999). Já que *HPC1* e *HPC2* possuem padrões de expressão altamente exclusivos, isto implica que apenas nos tipos celulares onde *HPC2* é expresso, a proteína CtBP pode ser recrutada para o complexo PcG.

O significado de CtBP interagir com a classe específica dos homólogos *Pc* pode envolver uma função-alvo dos complexos específicos de PcG. Foi achado que a proteína dCtBP interage com os repressores e as proteínas de ligação ao DNA: Knirps, Snail e Hairy (Nibu *et al.*, 1998; Poortinga *et al.*, 1998). É concebível que as proteínas CtBP alvejem os complexos PcG em *loci* particulares na cromatina, que contêm sítios de ligação para os repressores como os homólogos de Knirps e Hairy. Isto pode, no entanto, envolver apenas um complexo que contenha HPC2 e não HPC1, já que CtBP interaje com HPC2, e não com HPC1 (Satijn & Otte, 1999a).

Desta forma, em vertebrados, existem várias possibilidades de criar complexos PcG com diferentes composições. Primeiramente, as proteínas PcG existem em pares altamente relacionados que possuem diferentes perfis de expressão. Isto permite o estabelecimento de complexos de PcG com diferentes composições, dependendo do tipo celular. Em segundo lugar, estes pares diferem em regiões não conservadas, permitindo o recrutamento de diferentes proteínas PcG e proteínas repressoras. Ring1B recruta Mph2, Ring1A não. HPC2 recruta CtBP, HPC1 não. Este dado apresenta uma camada a mais de complexidade (Satijn & Otte, 1999a). Em *Drosophila*, diferentes proteínas, incluindo Pc, Ph e Psc se ligam em padrões de sobreposição em cromossomos politênicos (Rastelli *et al.*, 1993; Zink & Paro, 1989). Apesar do diferente padrão de ligação de proteínas PcG em cromossomos politênicos sugerirem a existência de complexos PcG com diferentes composições, proteínas PcG de *Drosophila* parecem não existir aos pares. Como mostrado anteriormente, um alto nível de complexidade e

flexibilidade existe nos sistemas de vertebrados. Quais são as implicações funcionais da existência de complexos PcG com diferentes composições? Uma possibilidade é que isto determine a especificidade de ligação aos diferentes genes-alvo (Satijn & Otte, 1999a).

Mais uma vez, diferentes composições de complexos PcG direcionados para diferentes genes-alvo são eventos-regulatórios chave, que provavelmente, resultam em diferentes alterações fenotípicas pleiotrópicas. Outra classe de potenciais genes-alvo de PcG de vertebrados é formado por oncogenes e genes supressores de tumor. É considerável que a interferência dos níveis de expressão de quase todas as proteínas PcG, resulte de alguma maneira em transformação celular (Satijn & Otte, 1999a). O gene *Bmi-1* foi isolado como um oncogene que coopera com o proto-oncogene *c-myc* (Haupt *et al.*, 1991; van Lohuizen *et al.*, 1991a). Camundongos transgênicos super-expressando *Bmi-1* desenvolvem linfomas (Alkema *et al.*, 1995) e *Bmi-1* é capaz de transformar fibroblastos de roedores (Cohen *et al.*, 1996; Jacobs *et al.*, 1999). Em contraste, seu parceiro relacionado, *mel-18*, apresenta efeitos opostos. Apenas quando a expressão de *mel-18* é interrompida pela super-expressão de RNA antisense, um fenótipo tumorigênico aparece nos fibroblastos. Isso pode sugerir que *mel-18* seja um gene supressor de tumor. Quando *mel-18* é super-expresso em camundongos transgênicos, a diminuição na regulação do proto-oncogene *c-myc* é observada (Kanno *et al.*, 1995); (Tetsu *et al.*, 1998).

Da mesma maneira, efeitos opostos têm sido observados por interferência com os níveis de expressão das proteínas humanas HPC2 e RING1. A super-expressão de *HPC2* resulta na repressão do oncogene *c-myc* e a super-expressão do dominante negativo de *HPC2* resulta no aumento da expressão de *c-myc* (Satijn *et al.*, 1997; Satijn & Otte, 1999b). A super-expressão do dominante negativo de *HPC2* resulta em transformação celular. Por outro lado, super-expressão de *RING1* resulta em transformação celular e indução de tumores em camundongos "nude". Isto é acompanhado por um aumento dos níveis de expressão dos oncogenes *c-jun* e *c-fos*, mas não por *c-myc* (Satijn & Otte, 1999b). Estes resultados podem sugerir que *HPC2* atue como um gene supressor de tumor e *RING1* como um oncogene. No entanto, ainda não está claro como estas observações podem ser explicadas ao nível

molecular. Mesmo se por acaso *c-myc* for um gene-alvo dos PcG, isto ainda não é reconhecido. Os efeitos aparentemente opostos das proteínas PcG na transformação celular são prováveis por emergir da ação de diferentes composições de complexos de proteínas PcG em genes regulatórios importantes, que são diferentes genes-alvo de diferentes complexos de PcG (Satijn & Otte, 1999a).

Como estas observações podem ser transpostas para *hEPC1* e *hEPC2* ainda não se sabe, no entanto, é interessante notar que a análise de expressão de ambos os genes nos tecidos avaliados mostrou resultados muito similares. Este dado não se mostra em concordância com os dados observados entre os pares *Ring1B/Ring1A* e *HPC2/HPC1*. No entanto, um dado que pode ser relevante nessa especulação, e que diferencia *hEPC1* e *hEPC2* dos outros de pares de genes citados, é o fato de *hEPC1* e *hEPC2* serem integrantes do grupo dos ETP. Talvez, a característica destas proteínas atuem ao nível de ativação e repressão de transcrição, explique o porque destes genes serem expressos na maioria dos tecidos humanos avaliados.

O ortólogo humano de *hEPC1*

Durante o trabalho da obtenção da sequência completa do gene *hEPC1* (Shimono *et al.*, 2000) publicaram um artigo resultante de uma busca por proteínas que interagissem com a proteína alvo de seu estudo: “RET finger protein (RFP)”. Através da metodologia de duplo híbrido, o ortólogo humano de *EPC1* foi descrito por interagir com RFP. As proteínas RFP pertencem à família dos “large B-box RING finger protein”, cujos membros contêm um motivo tripartido que consiste em um “Ring finger”, um domínio “coiled-coil” e 3 hélices definidas, e além disso, uma região carboxi-terminal conhecida como domínio RFP (Berg, 1990; Cao *et al.*, 1997; Freemont *et al.*, 1991; Reddy *et al.*, 1992; Takahashi *et al.*, 1988). Diversas proteínas da família “RING finger protein” já foram descritas, entre elas PML, BMI1/Mel-18, RING1 e KAP-1 (Borden, 1998; Borden, 2000; Saurin *et al.*, 1996). “Ring finger proteins” parecem atuar na formação e arquitetura de grandes complexos de proteínas, que por sua vez, contribuem em diversos processos celulares como oncogênese, apoptose,

desenvolvimento e ubiquitinação (Borden, 1998; Borden, 2000; Saurin *et al.*, 1996). Por exemplo, PML, BMI1/Mel-18, RING1 e RFP estão envolvidas em oncogênese pela formação de proteínas de fusão (Borrow *et al.*, 1990; de The *et al.*, 1991; Hasegawa *et al.*, 1996; Kakizuka *et al.*, 1991; Takahashi *et al.*, 1988) ou pela desregulação nos níveis de expressão de certos oncogenes (Kanno *et al.*, 1995); (Satijn & Otte, 1999b; Tetsu *et al.*, 1998).

Neste trabalho, (Shimono *et al.*, 2000), descrevem que RFP interage com *hEPC1*, sendo que esta interação é mediada pelo domínio "coiled-coil" de RFP além do domínio EPcA e a região carboxi-terminal de *hEPC1*. Como a localização das proteínas associadas está no núcleo, é sugerido que elas estejam envolvidas no silenciamento gênico. *EPc* é classificado como um membro da família dos Polycomb, o que implica em sua função como repressor transcricional. No entanto, este estudo revelou que a atividade repressiva de *hEPC1* é relativamente fraca e que, surpreendentemente, o domínio EPcA é um forte ativador transcricional. Além de *EPc*, o gene "Enhancer of Zeste (*EZH*)" da família dos PcG também possui uma função dupla de manutenção da atividade transcricional por proteínas da família dos genes Trithorax (*TrxG*) (LaJeunesse & Shearn, 1996). No entanto, ele se associa com *Esc* de *Drosophila* (cujo homólogo humano é *EED*), e funciona como um repressor de transcrição. Estes dados sugerem que ambos, *EPc* e *EZH*, tenham funções diferentes das outras proteínas da família dos PcG. Os autores chamam atenção para o fato dos homólogos de camundongo de *hEPC1* e *hEPC2* possuírem transcritos mais longos e menores (curiosamente, Stankunas, *et al* 1998 só sugerem a presença de um transcrito menor em *Epc1*), e que os transcritos menores podem agir como ativadores de transcrição como os *TrxG*, enquanto que os transcritos maiores funcionariam como repressores transcricionais da família dos PcG.

Desta maneira, diante da descrição do gene *hEPC1* e de seu estudo por um outro grupo, optamos por focar nossos estudos no gene *hEPC2*, cuja descrição e estudo ainda são inéditos na literatura.

***hEPC1* e *hEPC2* como membros do grupo de “Enhancer of Polycomb and Trithorax Proteins” (ETP)**

hEPC1 e *hEPC2* são membros do grupo de proteínas ETP, que são descritas por possuírem dupla atividade transcricional, de repressão e ativação (Brock & van Lohuizen, 2001). Realmente, *hEPC1* foi descrito por ser uma proteína que interage com a proteína RFP, atuando em ambas ativação e repressão de transcrição (Shimono *et al.*, 2000). Portanto, é plausível extrapolar que os domínios EPcA (ativação de transcrição) e EPcC (repressão de transcrição) de *EPC2* possam ter funções similares aos domínios correspondentes de *EPC1*.

Em mamíferos, a maioria dos genes PcG está presente em pares altamente relacionados (Schumacher & Magnuson, 1997). Análises recentes feitas em camundongos deficientes dos pares de genes *Bmi-1* e *Mel18*, revelaram que estes genes atuam sinergisticamente e de maneira dose-dependente para a própria repressão individual de genes-alvo durante o desenvolvimento. Diferentes genes-alvo necessitam de níveis diferentes de proteínas PcG. Interessantemente, a ausência de ambos *Bmi-1* e *Mel18* leva a uma redução significativa no nível de alguns genes homeóticos, enfatizando a necessidade deste grupo de genes ETPs de mamífero para a ativação gênica (Brock & van Lohuizen, 2001). Consistente com a idéia de que genes PcG de mamíferos podem ativar alvos, *EED* (PcG) e *Bmi-1* (ambos do grupo ETP) possuem efeitos opostos em precursores hematopoéticos (Lessard *et al.*, 1999).

Um número de modelos pode ser imaginado para explicar o papel do grupo das proteínas ETP na ativação e na repressão. As proteínas ETP podem ser sub-unidades comuns de diferentes complexos necessários tanto na ativação quanto na repressão de *loci*-alvo. Existe uma boa evidência de proteínas ETP sendo membros de complexos de repressão. Recentemente, um complexo protéico chamado PRC1 que contém as proteínas Pc, Ph, Psc e Scm foi purificado de embriões de *Drosophila* (Shao *et al.*, 1999). Co-imunoprecipitação, co-fracionamento e evidência de interação proteína-proteína em mamíferos, sustentam a idéia de que existe um complexo de

proteínas contendo BMI-1, HPH, HPC2 e RING (Satijn & Otte, 1999a), (),(). Em outros organismos, estes complexos contêm proteínas dos PcG e ETP. Da mesma maneira, outro complexo PcG caracterizado contém as proteínas PcG Esc e a proteína ETP E(z) (Ng et al., 2000), ou respectivamente, seus homólogos em mamífero, EED e EZH/Enx1 (), (Sewalt et al., 1998). Sugestivo de funções separadas, este último complexo é mais antigo porque seus homólogos existem em plantas e *Caenorhabditis elegans*, que parecem não possuir outros homólogos de PcG.

Até hoje, nem as proteínas do grupo dos PcG nem dos ETP foram identificadas como membros dos complexos trxG, mas pelo menos uma proteína trxG pode ser necessária na manutenção da repressão. "Trithorax-like" (Trl) codifica para o fator GAGA, um ativador transcricional geral necessário para a expressão de diversos genes em *Drosophila*, incluindo os genes homeóticos. O fator GAGA tem um sinergismo com fatores de remodelamento de nucleossomo contendo ISWI no promotor (Tsukiyama et al., 1995), no entanto, mutações em GAGA causa a desrepressão do PRE Fab-7 (Hagstrom et al., 1997), tendo sido sugerido que Trl poderia ser re-classificado como um ETP (Gildea et al., 2000). Além disso, o fator GAGA se associa com PREs *in vivo* (Strutt et al., 1997). Uma possibilidade tentadora é a de que GAGA seja necessário para o remodelamento da cromatina, que permite o acesso subsequente dos PREs às proteínas PcG e trxG (Strutt et al., 1997). Alguns dos fatores GAGA co-imunoprecipitam com Pc e Psc, sugerindo que este é um constituinte dos complexos de PcG, e não simplesmente necessário para o remodelamento da cromatina, de modo a permitir o acesso dos complexos PcG (Horard et al., 2000). O nível e a atividade das proteínas ETP pode ser altamente regulado. Neste caso, interações dos ETPs com os PcG e os trxG, podem trazer os mecanismos de discriminação necessários para ligar o elemento responsivo de um *locus*-alvo de um estado neutro para ativado ou reprimido. Se tais complexos ETP se formarem apenas transitoriamente, ou agirem dentro de uma pequena janela do desenvolvimento, eles podem ter escapado de serem detectados previamente (Brock & van Lohuizen, 2001).

As evidências sugerem que as proteínas PcG e ETPs podem agir de maneira gene-específica e tipo celular específico. Os dados genéticos apontam

claramente para a possibilidade de que existem múltiplos complexos PcG que, por analogia com os diferentes complexos de remodelamento-trxG, podem contribuir apenas com propriedades enzimáticas únicas e funcionais. A caracterização de complexos contendo proteínas do grupo PcG e/ou ETP será um importante passo na tentativa do entendimento dos papéis destas proteínas na manutenção da ativação e repressão (Brock & van Lohuizen, 2001).

A existência de quatro isoformas transcricionais de *hEPC2* pode estar implicada no seu papel de regulação de transcrição tanto ao nível gene-específico, quanto no celular-específico. É importante notar que Iso-B e Iso-C apenas retêm o domínio EPcA, o bloco que é similar ao domínio de ativação transcricional de *hEPC1*. Outro dado que é interessante de se notar, é que o terceiro exon de Iso-B codifica para 16 aminoácidos que não apresentam similaridade com as outras isoformas, nem com outras proteínas humanas conhecidas. No entanto, tanto Iso-A (transcrito canônico) quanto Iso-D apresentam os domínios EPcA e EpcC (Figura 13), podendo apresentar também, atividades de repressão.

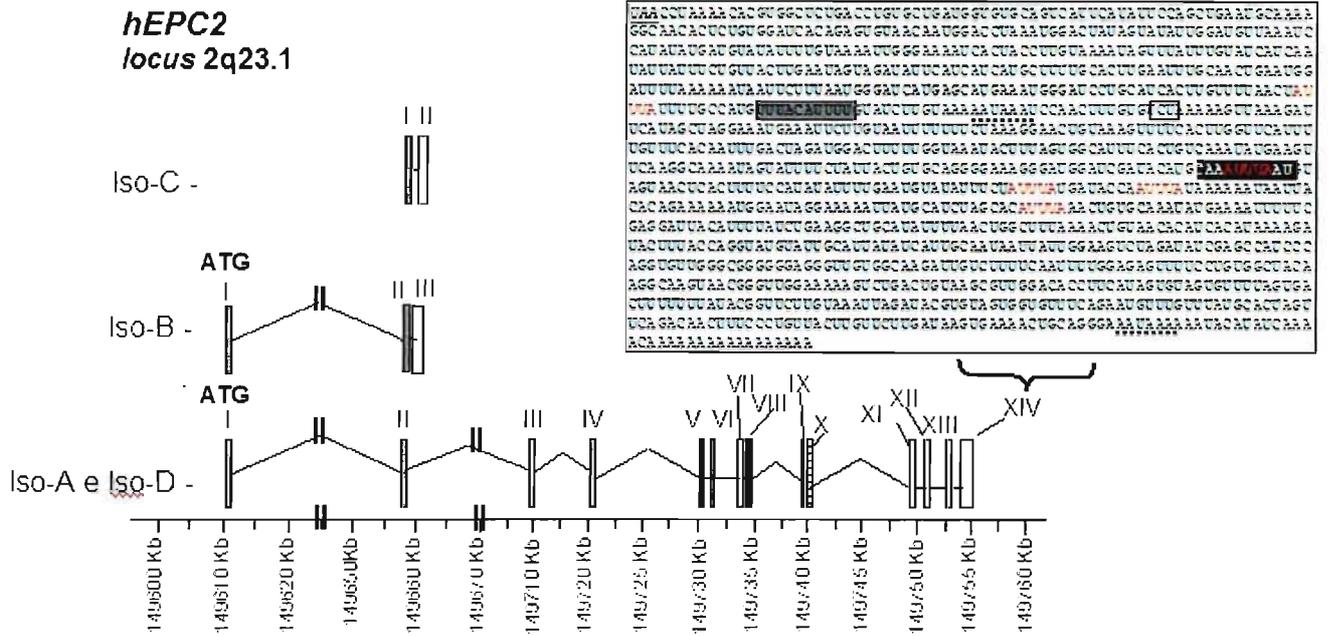


Figura 13 – Estrutura genômica das isoformas A, B C and D do gene humano *hEPC2*. As caixas numeradas representam os exons. As linhas que conectam as caixas representam os íntrons. ATG indica o exon que contem o códon de iniciação de tradução. As caixas em cinza representam o domínio EPcA. As caixas em preto representam o domínio EPcB e as caixas hachuradas representam o domínio EPcC. Uma ampliação no exon XIV mostra a 3'UTR de Iso-A e Iso-D. A sequência em itálico e sublinhada representa o códon de parada. A caixa representa o fim do transcrito Iso-D. As sequências em vermelho representam possíveis AREs de Iso-D e Iso-A. As linhas pontilhadas representam os sítios de poli-adenilação. A caixa em cinza representa o “Sage tag” de Iso-D e a caixa em preto o “Sage tag” de Iso-A.

O equilíbrio das isoformas de *hEPC2* pode ser importante durante o desenvolvimento, em diferentes tecidos, durante o controle de ciclo celular, ou mesmo em condições patológicas envolvendo crescimento celular e diferenciação. Uma evidência para esta hipótese é o achado da regulação diferencial de Iso-B e Iso-C no diferentes tecidos avaliados (Figura 8). Apesar de não haver uma grande variação na abundância de Iso-B e Iso-C, diferentes razões na expressão relativa destes transcritos foram freqüentemente encontradas, para diferentes tecidos, assim com em amostras fetais e adultas do mesmo tecido. Por exemplo, no cérebro total, ambas as isoformas são mais abundantes no tecido adulto do que no tecido fetal, especialmente Iso-C. Em fígado, o oposto foi observado, sendo que a expressão de ambas as variantes é mais alta no feto quando comparada com o tecido adulto. No entanto, o que é conclusivo em relação à expressão de Iso-B e Iso-C em todos os tecidos avaliados, é o baixo nível de expressão destas variantes. De maneira oposta, pode-se observar a maior expressão de Iso-D em todas as linhagens celulares tumorais testadas.

O uso de um sítio alternativo de poli-adenilação em Iso-D

Iso-D possui a mesma região codificante de Iso-A, no entanto, possui uma redução em sua região 3' não traduzida (3'UTR) devido ao uso de um sítio mais interno, alternativo, de poli adenilação (Figura 10). Como seqüências presentes na região 3' UTR podem modular a estabilidade do transcrito, este pode ser um mecanismo para controlar a estabilidade e portanto, a abundância deste transcrito de *hEPC2*.

A região 3'UTR está envolvida na regulação da clivagem de transcritos, na poli-adenilação, e no controle da poliadenilação alternativa, a exportação nuclear, o nível de tradução dos transcritos, e em muitos casos determina a sua estabilidade. Em relação à desregulação desta região, eventos como a degradação dos transcritos podem ter como causa a diminuição da cauda poli-A ou a de-adenilação independente da clivagem endonucleolítica (Pesole *et al.*, 2000). As seqüências regulatórias conhecidas como elementos ricos em

adenilatos e uridilatos (ARE) têm sido encontrados em vários mRNAs lábeis que codificam proteínas regulatórias, como proto-oncogenes, fatores de crescimento e seus receptores, mediadores inflamatórios e citocinas (Grzybowska *et al.*, 2001). Os elementos ricos em AU variam em tamanho de 50 a 150 nt e, usualmente, contêm uma ou algumas cópias dos pentâmeros AUUUA ou dos nonâmeros UUAUUUA. Geralmente, os AREs são divididos em não-AUUUA (ex: *c-jun*) e contendo-AUUUA. Este último grupo foi dividido em duas classes: - classe I, ex: *c-fos*; - classe II, fator estimulante de colônias granulo-monocíticas (GM-CSF). Estas classes se diferenciam em seu mecanismo regulatório e nos padrões de de-adenilação (Grzybowska *et al.*, 2001). A função primária dos AREs é determinar a degradação seletiva dos mRNAs, no entanto, este processo de degradação mediada pelos AREs é auto-regulada: sob condições de stress, por estimulação celular ou durante a transformação oncogênica mediada pela estabilização de mRNAs contendo ARE (Conne *et al.*, 2000; Pesole *et al.*, 2000).

Perturbações na 3'UTR, como mutações em um alelo, podem resultar em um efeito dominante negativo relacionado a várias patologias. Algumas doenças humanas são causadas por mutações nestas sequências ou pela ligação de proteínas regulatórias nesta região (Conne *et al.*, 2000). A ocorrência de AREs defectivos causa a estabilização de mRNAs normalmente lábeis, resultando em uma maior regulação da sua expressão e em alterações metabólicas que levam a doenças como linfoma de células do manto, artrite crônica e desordens inflamatórias. O aumento na estabilidade dos transcritos pode ser causado também, pela elevação dos níveis de ligação de proteínas regulatórias (como no caso da proteína ELAV-like em neuroblastoma). Outro mecanismo, a supressão da tradução de p53 associado com a presença de sequência Alu-E na região 3'UTR, têm sido indicada (Grzybowska *et al.*, 2001). A desregulação do proto-oncogene *CCDN1* (ciclina D1) está relacionada com o desenvolvimento de vários tumores, incluindo sua super-expressão na vasta maioria das MCL e translocações associadas a leucemias do tipo (11q13). Rearranjos na 3'UTR do gene *CCDN1* levam à transcrição de um mRNA menor que 1,5 Kb, observado em tumores primários e em linhagens celulares. Este transcrito apresenta a mesma região codificante do transcrito de 4,5 Kb que

está presente em mRNA de tecidos normais. No entanto, o transcrito de 1,5 Kb usa um sítio alternativo de poli adenilação, ou possui parte de sua 3'UTR deletada. Estas duas alterações resultam em um mRNA mais estável, que é também mais expresso (Rimokh *et al.*, 1994).

A análise da 3'UTR de Iso-A mostrou a presença de cinco motivos ARE do tipo (AUUUA). No entanto, quatro destes motivos são perdidos quando ocorre o uso do sítio alternativo de poli-adenilação utilizado na Iso-D. A maior abundância desta isoforma em tecidos tumorais pode significar que a proteína correspondente seja produzida em maior quantidade, fruto de um mRNA mais estável, sugerindo desta maneira, que esta isoforma, assim como *CCDN1*, pode ser mais abundante nas linhagens celulares tumorais devido a maior estabilidade do mRNA desta variante. Adicionalmente, o ritmo de transcrição e degradação de Iso-A nas linhagens celulares tumorais pode ser diferencialmente regulado, já que sua expressão é mais baixa quando comparada com os tecidos normais avaliados. O gene *c-fos* possui três motivos AUUUA em sua 3'UTR, sendo que a deleção de cada um deles aumenta gradualmente a estabilização do transcrito, levando a transformação de proto-oncogene para um oncogene (Schiavi *et al.*, 1992).

Curiosamente, foi observada uma expressão mais alta de Iso-D nas linhagens de leucemia promielocítica HL-60 e em linfoma de Burkitt Raji (Figura 9). Alguns PcG foram descritos por estarem associados com a linfomagênese (Lessard & Sauvageau, 2003), como o gene *Bmi-1* ("B lymphoma Mo-MLV insertion region"), que se encontra amplificado em linfomas de células do manto (Bea *et al.*, 2001), ou o gene "Enhancer of Zeste" (*EZH2*), que é super-expresso em linfomas (Visser *et al.*, 2001). Ambos, mRNA e proteínas, do gene *EZH2* também foram observados com expressão aumentada em pacientes com câncer de próstata metastático, quando comparados com pacientes com tumores órgão-confinados ou benignos (Varambally *et al.*, 2002). O aumento no nível da proteína *EZH2* também foi associado com maior agressividade em câncer de mama (Kleer *et al.*, 2003).

A diferenciação de precursores de linfócitos na medula óssea é associada com profundas alterações na transcrição gênica dos PcG (Lessard *et al.*, 1998). As células precursoras transcrevem preferencialmente o gene PcG *BMI-1*, e a

transcrição de outros genes PcG ocorre em estágios tardios de diferenciação de células B, enquanto que a transcrição de *Bmi-1* gradualmente desaparece (Lessard *et al.*, 1998). A evidência mais convincente do papel dos genes PcG na linfopoiese vem no entanto, de estudos com camundongos mutantes. A super-expressão de BMI-1 em camundongos transgênicos resultou em um aumento na linfo-proliferação e por fim, no desenvolvimento de linfomas de células B (Jacobs *et al.*, 1999) (van Lohuizen *et al.*, 1991b). Em contraste, camundongos com uma deleção dos genes PcG *Bmi-1* e *Mel-18* desenvolvem hipoplasia enquanto que a perda do gene PcG *EED* resultou em um aumento da linfo-proliferação (van der Lugt *et al.*, 1994) (Akasaka *et al.*, 1996) (Lessard *et al.*, 1999). Estas observações sugerem que os dois complexos de PcG têm papéis opostos na regulação da hematopoiese. Recentemente foi demonstrado que a expressão dos dois complexos PcG, refletidos pela detecção de *Bmi-1/RING1* e *EZH2/EED* são mutuamente exclusivas em diferentes estágios de diferenciação das células B nos centros de germinação (Lessard *et al.*, 1998; Lessard *et al.*, 1999; van Lohuizen, 1999). Estes resultados mostraram que a expressão dos genes PcG é estritamente regulada durante o desenvolvimento folicular das células B e sugerem um papel das proteínas PcG nas reações celulares nos centros germinativos.

A expressão das proteínas PcG *Bmi-1* e *EZH2* em células T periféricas maduras também é mutuamente exclusiva e associada com o *status* proliferativo destas células. Células T periféricas maduras que se encontram quiescentes expressam *Bmi-1*, enquanto que blastos em divisão expressam *EZH2*. A expressão destas proteínas PcG em estágios específicos das células T sugere que os genes PcG contribuem para a diferenciação destas células. Provavelmente, estas proteínas refletem a estabilização da expressão de um tipo celular específico e a irreversibilidade na escolha da linhagem celular (Raaphorst *et al.*, 2001)

Os padrões de expressão de *Bmi-1* e *EZH2* foram analisados também em uma variedade de linfomas de células B não-Hodgkin (B-NHLs) incluindo linfomas linfocítios, linfomas foliculares, linfomas de células B grandes, linfomas celulares ocultos e linfoma de Burkitt. Em contraste ao padrão mutuamente exclusivo de *Bmi-1* e *EZH2* em folículos reativos, as células neoplásicas dos B-

NHLs de graus intermediários e avançados mostraram forte co-expressão de *Bmi-1* e *EZH2*. Este padrão foi coerente com a expressão de *Mib-1/K1-67*, um marcador de proliferação celular. Células neoplásicas de B-NHLs de baixo grau de malignidade tiveram a expressão de *Bmi-1* baixo/*EZH2* positivo (centroblastos neoplásicos) ou *Bmi-1* baixo/*EZH2* negativo (centrócitos positivos). Estas observações mostram que B-NHLs de baixo grau, intermediário e alto grau possuem associação com um aumento na co-expressão de *Bmi-1* e *EZH2*, cuja expressão normal é mutuamente exclusiva. Esse padrão de expressão é provavelmente causado por uma falha da baixa expressão de *Bmi-1* em células neoplásicas em divisão, porque a expressão de *Bmi-1* é ausente na divisão normal de células B. Estas observações estão em concordância com os achados dos estudos de *Bmi-1* em camundongos transgênicos. A extensa co-expressão de *Bmi-1* e *EZH2*, correlacionada com o grau clínico e a presença da expressão de *Mib-1/Ki-67*, sugerem que a expressão irregular destes genes é um evento precoce na formação de B-NHL. Esses dados demonstram o papel da expressão anormal de genes PcG na gênese de linfomas humanos (Folkert J et al, 2001).

(Sellers & Loda, 2002), sugerem a intrigante possibilidade de que a amplificação ou super-expressão de PcG, como *Bmi-1* e *EZH2*, possam ser participantes na transformação celular como repressores transcricionais. Esta repressão pode atingir muitos genes que conferem vantagem no crescimento celular e podem culminar com o surgimento de tumores e sua progressão. Portanto, é possível extrapolar que a super-expressão de Iso-D possa levar a uma diminuição da expressão de outros genes, incluindo da variante Iso-A. No entanto, é importante enfatizar que Iso-D está presente também nos tecidos normais analisados, sugerindo a necessidade de outros fatores para a super-expressão desta isoforma nas linhagens tumorais. Como observado para outras proteínas PcG, as diferentes variantes de EPC2 podem compor diferentes complexos em paralelo nas células.

A descrição das distintas isoformas de *hEPC2* e a associação de um destes transcritos com o câncer é um dado único dentro da literatura dos PcG. Desvendar o complexo de proteínas que se associam com *hEPC2* será um importante passo no entendimento de sua função, assim como do papel

particular de cada uma das diferentes isoformas. O segundo bloco desta tese investiga este aspecto, buscando decifrar os possíveis parceiros moleculares de hEPC2.

PARTE II: INTERAÇÃO ENTRE hEPC2 E AS SMADS E SUA MODULAÇÃO POR TGF- β

INTRODUÇÃO

As proteínas SMAD como sinalizadoras da via de TGF- β

Nas análises iniciais da sequência nucleotídica de *hEPC2* utilizando a ferramenta BlastN contra o banco Nr, duas sequências parciais descritas na patente WO 9855512 (sequências 11 e 13), mostraram identidade com duas regiões distintas do gene *hEPC2*. Utilizando o banco de dados Derwent, disponível através do "site" da FAPESP, foi possível acessar os dados de descrição desta patente. Os fragmentos patenteados foram obtidos através de ensaios de duplo híbrido e são descritos como polipeptídeos que interagem com proteínas da família SMAD, conhecidos como reguladores transcricionais envolvidos no desenvolvimento, pois atuam na transdução de sinal de TGF- β (fator de crescimento tumoral) (Moustakas *et al.*, 2001).

O programa genético de uma célula em um organismo multicelular deve ser estritamente controlado em função do organismo como um todo. Nos últimos 20 anos, a família dos polipeptídeos secretórios "Transforming Growth Factor- β " (TGF- β), emergiu como uma fonte sinalizadora exercendo esse tipo de controle. Esta família inclui vários tipos de TGF- β , as "bone morphogenetic proteins" (BMPs), os "Nodals", as Activinas, o hormônio anti-Mülleriano, e muitos outros fatores estruturalmente relacionados em vertebrados, insetos e nematódeos (Massague & Wotton, 2000).

Produzidos em diversos tipos celulares, estes fatores regulam a migração celular, adesão, multiplicação, diferenciação e morte durante toda a vida do organismo, além de desempenhar um importante papel no desenvolvimento e em numerosas doenças, tais como o câncer. Muitas destas respostas resultam de alterações na expressão de genes-chave.

Os membros da família dos TGF- β possuem diferentes efeitos dependendo do tipo ou estado da célula. Por exemplo, na mesma ferida cicatrizante, TGF- β pode estimular ou inibir a proliferação celular dependendo se o alvo é um fibroblasto ou um queratinócito (Ashcroft *et al.*, 1999). Em células epiteliais mamárias, TGF- β causa parada de crescimento ou comportamento metastático de acordo com o nível da atividade de Ras oncogênico presente nas células (Oft *et al.*, 1996). BMP4 humano e seu ortólogo, DPP, sinaliza para a dorsalização em moscas (Padgett *et al.*, 1993), e formação dos ossos em vertebrados. Os membros da família dos TGF- β são hormônios multifuncionais, e a natureza de seus efeitos depende do que se chama "contexto celular".

É plausível que a transdução de sinal de TGF- β tenha que ser numerosa e complexa para justificar sua diversidade de respostas. Ao contrário, o desvendamento de um sistema simples que faz a mediação de diversas respostas de TGF- β foi elucidado. O sistema envolve uma família de receptores de membrana da família das proteínas quinases e uma família de receptores de substrato, as proteínas SMAD, que vão ao núcleo, onde agem como fatores de transcrição. O ligante TGF- β se liga e permite a montagem de um complexo receptor que ativa as SMADs, e as SMADs montam complexos de multi sub-unidades que regulam a transcrição (Heldin *et al.*, 1997; Massague, 1998) (Figura 14).

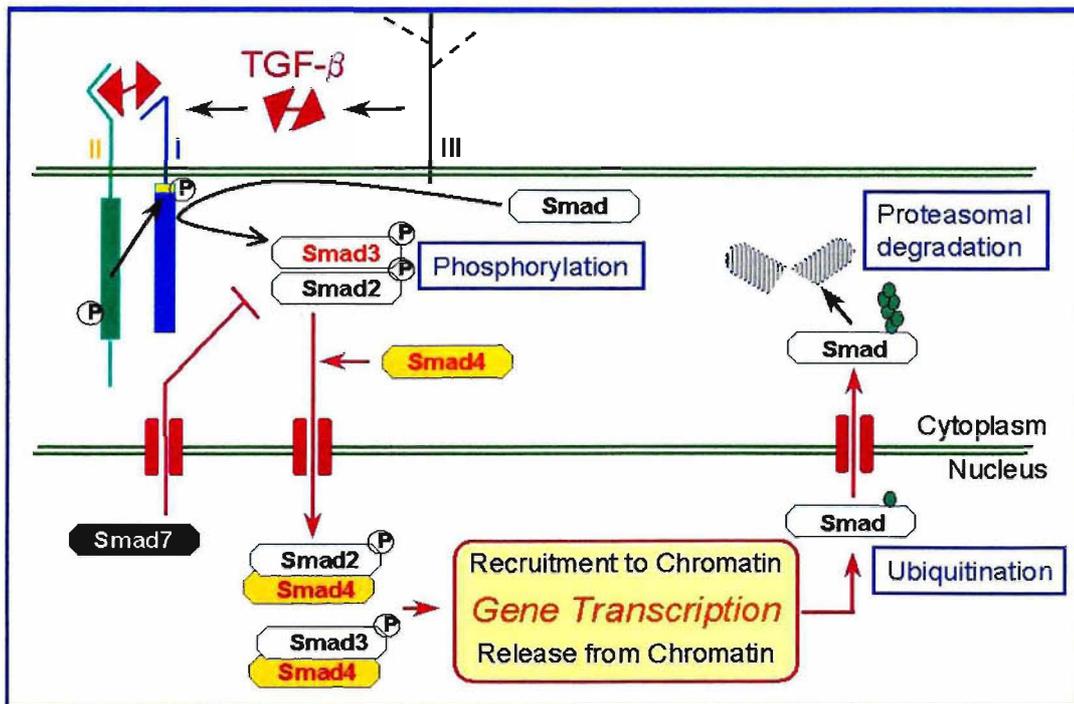


Figura 14 – Transdução do sinal de TGF-β pelas proteínas SMADs (retirado do site do grupo “TGF-β signaling” (www.licr.uu.se/groups/), do Ludwig Institute for Cancer Research, Uppsala/Suécia).

Portanto, dois passos são suficientes para carrear o estímulo hormonal aos genes. Como um sistema tão simples pode mediar uma variedade de repostas celulares tão específicas? Um complexo SMAD se une no núcleo a uma série de proteínas-parceiras, que são específicas de um tipo celular específico, em cada conjunto de condições particulares. Estes parceiros determinam a sequência de DNA que o complexo SMAD vai se ligar, os co-ativadores ou co-repressores transcricionais que irão recrutar os outros fatores de transcrição com os quais irão cooperar, e ainda quanto tempo esta ação vai durar (Moustakas *et al.*, 2001). A mistura de parceiros das SMADs e os reguladores presentes em uma determinada célula no momento da estimulação por TGF-β irão decidir o resultado da resposta, e definir em termos moleculares, o contexto celular. A identificação destes parceiros é portanto crítica para o entendimento da atuação de TGF-β (Massague & Wotton, 2000).

RESULTADOS

Interação entre hEPC2 e as SMADs

Como os genes da família dos Polycomb (PcG) são importantes repressores da expressão gênica e desempenham um papel importante na manutenção da diferenciação celular (Francis & Kingston, 2001), é possível que hEPC2 possa agir na via de TGF- β como um co-repressor ou co-ativador, associado as SMADs. Portanto, nossa proposta de estudo foi examinar a possível interação entre hEPC2 (um dos ortólogos humanos de EPC) e as SMADs, já que essa investigação poderia ter impacto no entendimento da regulação gênica que define a diferenciação, e sua conseguinte subversão nas células cancerosas.

Para investigar a possível interação de hEPC2 com as proteínas SMAD, foi montada uma colaboração com o grupo do Dr. Aristidis Moustakas, chefe do grupo de sinalização de TGF- β (TSL) do Ludwig Institute for Cancer Research, Uppsala, Suécia, e o grupo da Dra. Mari Cleide Sogayar, do Laboratório de Biologia Celular e Molecular do Instituto de Química da USP (LBCM). A proposta de estudo era confirmar inicialmente a interação hEPC2 e SMADs, através de ensaios de co-imunoprecipitação e "pull-down". Posteriormente, seria definido o papel transcricional de hEPC2 na via de sinalização de TGF- β .

co-Imunoprecipitação (IP) de hEPC2

Inicialmente, a região codificante de Iso-A foi clonada no vetor pCDNA3 para expressão em mamífero (originais da Invitrogen, USA), que possui fusionado um epítipo de uma parte da seqüência do proto-oncogene *myc* repetida 6 vezes (vetor modificado, gerado pelo grupo sueco) (Figura 15A e 15B). O cDNA de *hEPC2* foi clonado logo após o epítipo no sentido 5'→3', pois a proteína traduzida na porção amino-terminal é hidrofílica, sendo portanto, possivelmente mais exposta. Para a clonagem, foram desenhados oligonucleotídeos para a deleção do códon de iniciação, o qual foi substituído pela seqüência de fusão do vetor.

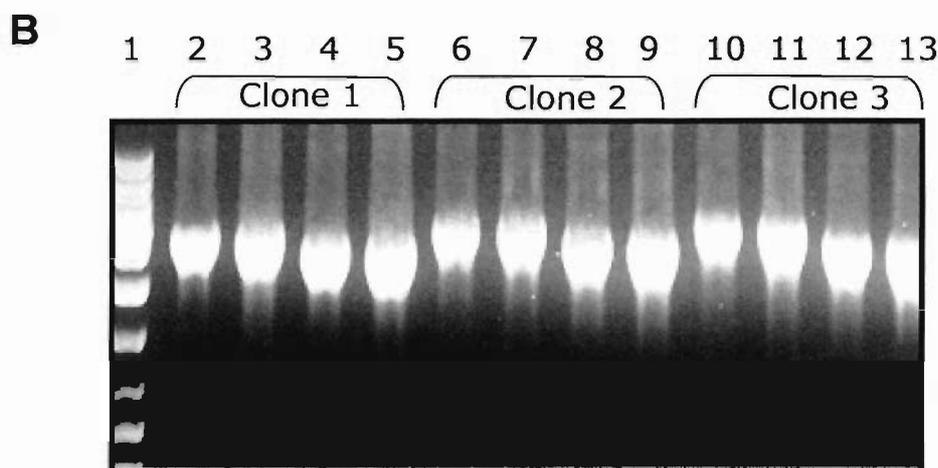
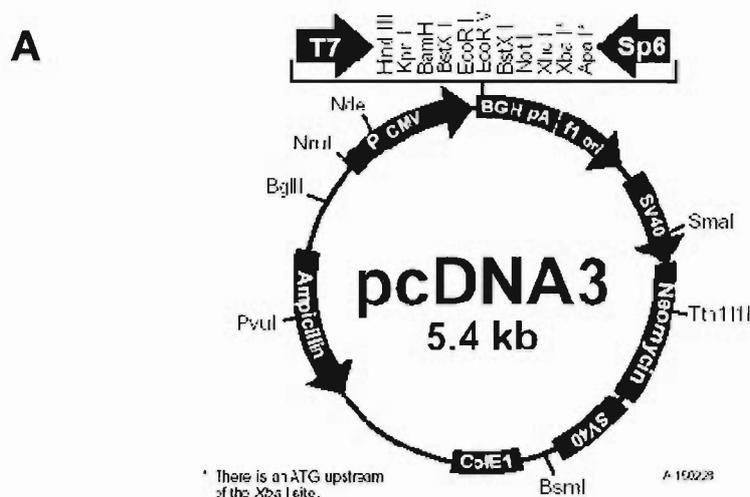


Figura 15 – A: Mapa do vetor de expressão constitutiva pCDNA3 (Invitrogen), que possui um epítipo de 6X myc. B: Amplificação do inserto de *hEPC2* clonado em pCDNA3-6Xmyc, com diferentes combinações de oligonucleotídeos. 1 – Peso molecular 1 Kb Gene Ruler Mix (Fermentas, UK); 2, 6 e 10– Amplificação de pCDNA3-6Xmyc-*hEPC2* com os oligonucleotídeos T7 + *hEPC2R**; 3, 7 e 11– Amplificação de pCDNA3-6Xmyc-*hEPC2* com os oligonucleotídeos *hEPC2F**+SP6; 4, 8 e 12– Amplificação de pCDNA3-6Xmyc-*hEPC2* com os oligonucleotídeos T7+ SP6; 5, 9 e 13– Amplificação de pCDNA3-6Xmyc-*hEPC2* com os oligonucleotídeos *hEPC2F*+*hEPC2R*.
* - *hEPC2F* e *hEPC2R*, oligonucleotídeos específicos para *hEPC2*, respectivamente “sense” e “antisense”

O gene também foi clonado no vetor para expressão indutível em bactérias pGEX-4T1 (Figura 16A e 16B), que ao ser expresso produz uma proteína de fusão glutationa S-transferase (GST)-hEPC2. Da mesma maneira, a porção amino-terminal da proteína será expressa fusionada a GST. Após a clonagem da isoforma canônica de *hEPC2* nos vetores acima citados, foi feito sequenciamento das construções para assegurar que a seqüência se encontrasse sem erros.

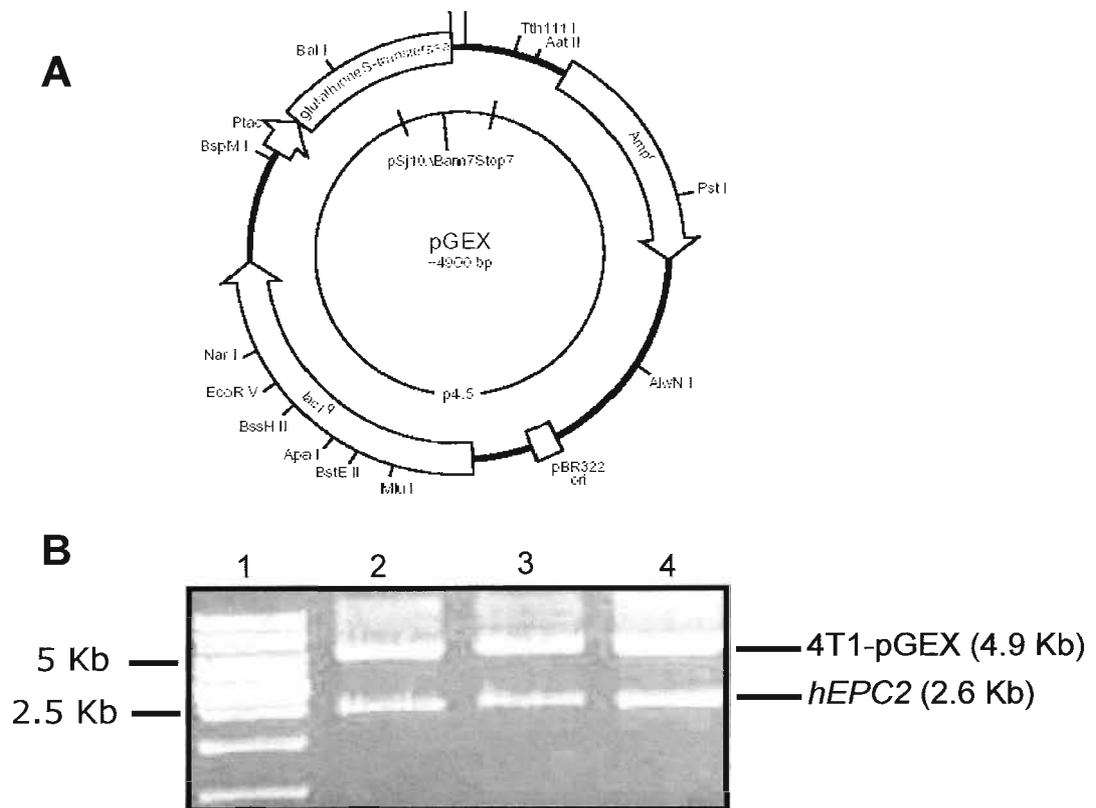


Figura 16 – **A**: Mapa do vetor de expressão indutível em bactéria pGEX (Amersham). **B**: 1: Padrão de peso molecular 1Kb Gene Ruler Mix (Fermentas); 2 a 4: Clones 1, 2 e 3 de 4T1-pGEX-hEPC2 digeridos com *EcoRI* e *XhoI*.

Para a investigação da interação entre hEPC2 e as SMADs, uma das metodologias utilizadas foi a co-imunoprecipitação. O vetor pCDNA3 possui um epítipo contendo seis repetições do proto-oncogene myc, gerando uma proteína de fusão que pode ser detectada tanto por um anticorpo anti-myc, quanto no caso por um anticorpo específico para hEPC2.

Além desta construção, a porção codificante de *hEPC2* também foi clonada no vetor de expressão em mamífero não comercial pCXN2 (Niwa *et al.*, 1991) (Figura 17 A e B).

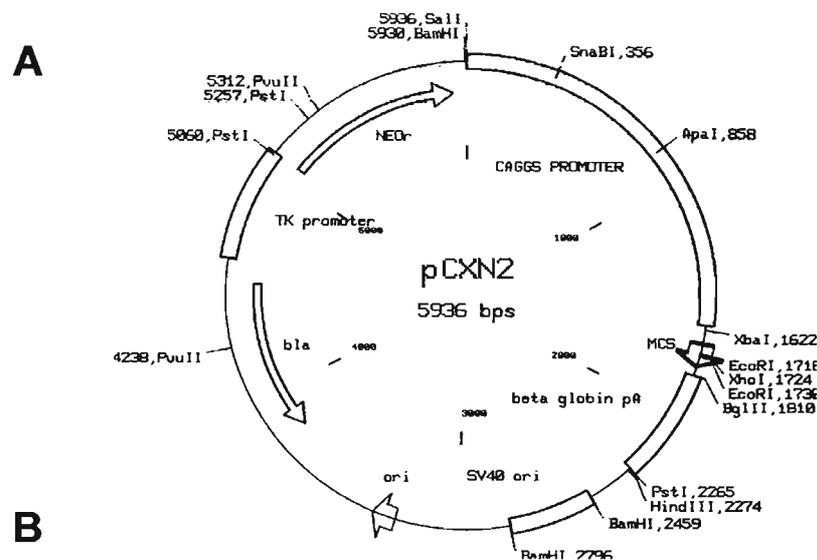


Figura 17 – A: Mapa do vetor pCXN2. **B:** 1: Padrão de peso molecular 1 Kb Gene Ruler (Fermentas); 2 a 6: clones de pCXN2-hEPC2 digeridos com *BglII* e *XbaI*; 7: pCXN2 vazio digerido com *SallI*.

Dois peptídeos distintos e parciais de hEPC2 foram gerados e utilizados para imunizar coelhos e gerar anti-soros policlonais α -hEPC2 (Bethyl, Montgomery) (EPC2I: NH₂- RTNASSKHCENRLSL -COOH e EPC2II: NH₂-SAASVALLNTSKNGI -COOH), já que estes reagentes seriam de extrema importância nos estudos de interação entre hEPC2 e as proteínas SMAD.

Super-expressão de pCXN2-hEPC2 e 6X myc pCDNA3-hEPC2

Já que o bom funcionamento de todos estes reagentes seria de extrema importância nos experimentos a serem executados no TSL (Laboratório de Sinalização de TGF- β), o primeiro ensaio realizado foi a super-expressão transitória das construções pCXN2-hEPC2 e 6X myc pCDNA3-hEPC2 nas células de rim de macaco Cos-7 e nas células de rim embrionário humano 293-T. Estas células foram escolhidas por exibirem uma alta eficiência de transfecção utilizando o protocolo de Ca₃(PO₄)₂ (Sambrook, 2001). A transfecção foi feita usando 0,5, 2,5 e 5 μ g de DNA, visando testar a expressão das construções de uma maneira dose-dependente. Os lisados protéicos foram então imunoprecipitados utilizando o anti-soro EPC2-II, que também foi usado no Western-blotting (Figura 18).

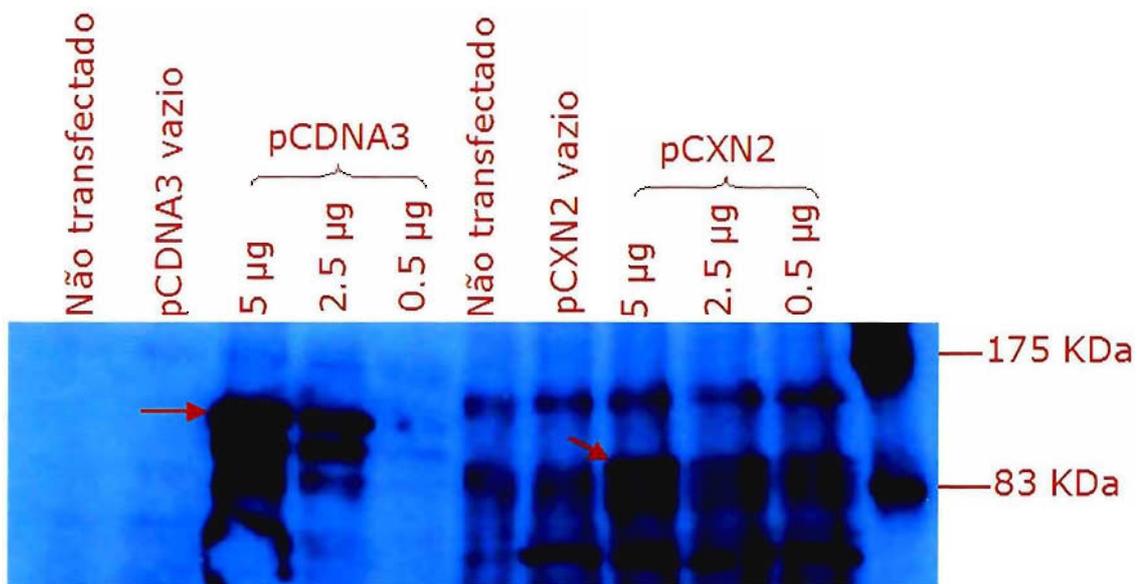


Figura 18 - Super-expressão das construções pCDNA3-6Xmyc-hEPC2 e pCXN2-hEPC2, co-transfectados transitoriamente em células 293-T. As amostras foram imunoprecipitadas e reveladas com o antisoro EPC2-II

A abordagem de fazer o ensaio de IP ao invés da análise direta por Western-blot foi devida a uma maior sensibilidade que seria obtida na detecção de hEPC2. A expressão de ambas as construções foi detectada com sucesso, denotando deste modo, que os dois plasmídeos recombinantes poderiam ser usados nos experimentos subsequentes, assim como o antisoro α -EPC2-II. As mesmas amostras utilizadas no IP foram avaliadas por Western-blot usando os antisoros α -EPC2-I e α -EPC2-II. No entanto, apenas o antisoro α -EPC2-II foi eficiente na detecção de hEPC2, especialmente com a construção 6X myc pCDNA3-hEPC2.

Investigação da interação de hEPC2 e SMADs por IP

Foi feito um ensaio de IP para testar a interação de hEPC2 com as SMADS, utilizando inicialmente a construção 6Xmyc pCDNA3-EPC2 que se mostrou mais fortemente expressa no experimento piloto. Foi investigado também, se os fatores de crescimento TGF- β 1 e BMP-7, ambos pertencentes a

super-família de TGF- β , modulariam a interação putativa entre as proteínas hEPC2 e SMADs. Como construções das SMADs 1, 2, 3, 4, 5, 6, 7 e 8 já se encontravam disponíveis em vetores adenovirais, estas foram utilizados nos experimentos.

As células 293-T foram escolhidas para este experimento por possuírem alta eficiência de transfecção. Adicionalmente, estas células possuem o antígeno T maior de SV40 sendo constitutivamente expresso, que por sua vez é responsável pela amplificação das construções de SMAD adenovirais.

As células foram transfectadas com 6Xmyc pCDNA3-hEPC2 em um dia, e no dia seguinte foram infectadas com as construções adenovirais de SMADs. No terceiro dia, as células infectadas com as SMADs 2, 3, 4 e 7 foram tratadas com TGF- β 1 por duas horas antes das células serem lisadas. Da mesma maneira as condições infectadas com as SMADs 1, 5, 6 e 8 foram tratadas com BMP-7. Controles tratados somente com o veículo carreador das citocinas foram feitos em duplicatas. As imunoprecipitações foram feitas com antisoros anti-flag, que é o epítipo presente nas construções das SMADs e os Western-blots foram analisados com o antisoro α -EPC2-II.

Os resultados das co-imunoprecipitações sugerem que as SMADs 2, 3, 4, 7 e 8 podem interagir com hEPC2. Quando SMAD2, SMAD3, SMAD4 e hEPC2 foram co-expressas e as células tratadas com TGF- β 1, uma redução na co-imunoprecipitação foi observada. Esse resultado sugere que TGF- β 1 modula negativamente a interação entre essas proteínas. Da mesma maneira foi observada uma redução na interação de hEPC2 com SMAD8 após o tratamento com BMP-7 (Figuras 19 e 20). No entanto, quando os lisados protéicos destes mesmos experimentos foram analisados por Western-blot utilizando antisoros α -EPC2-II e α -myc, observa-se uma maior expressão de hEPC2 nas condições onde hEPC2 é super-expresso e na ausência de tratamento com o fator de crescimento TGF- β 1 (Figura 21 A e B). Isso não é observado para SMAD8 (Figura 22 A).

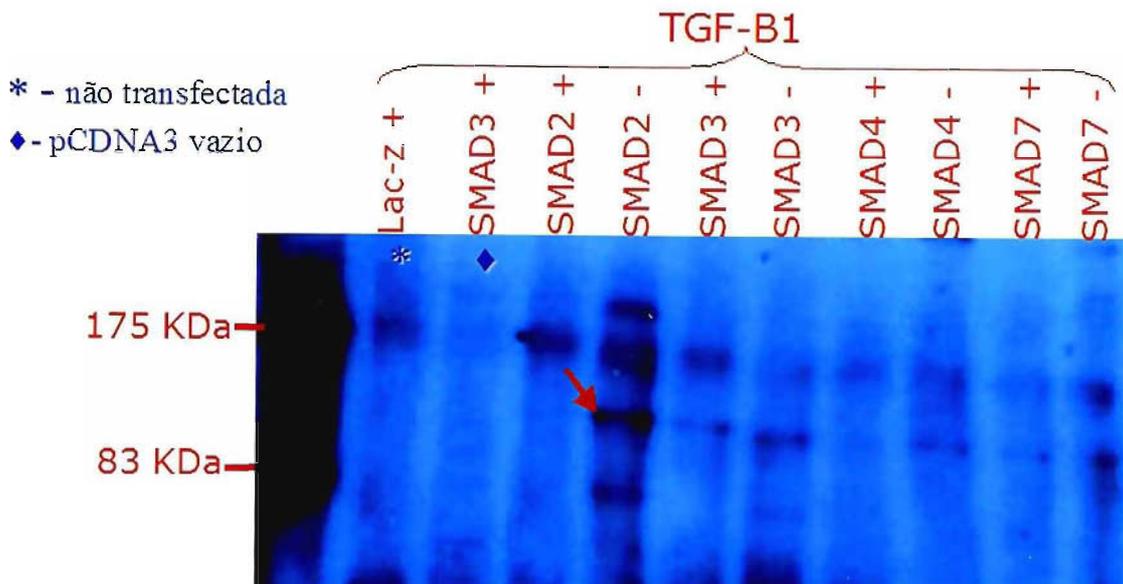


Figura 19 - Super-expressão das construções pCDNA3-6Xmyc-hEPC2 e SMADs 2, 3, 4 e 7 em células 293-T. As células foram transfectadas com a construção pCDNA3-6Xmyc-hEPC2 e infectadas 24 horas depois com os adenovírus contendo as construções das SMADs descritas acima das canaletas. As células foram tratadas com TGF- β 1 por 2 horas após 24 horas da infecção (+), ou apenas com o veículo carreador de TGF- β 1 (-). As amostras foram imunoprecipitadas com o antisoro monoclonal α -flag (M2, Sigma) (epítipo das construções de SMAD) e o "blotting" foi feito com antisoro α -EPC2II. Na primeira canaleta o controle foi feito com lisado proteico não transfectado, mas infectado com a construção contendo o gene Lac-Z e tratado com TGF- β 1. Na segunda canaleta o controle foi feito com células transfectadas com o vetor pCDNA3 vazio, e infectado com a construção de SMAD3, tratadas com TGF- β 1.

* - não transfectado

◆ - pCDNA3 vazio

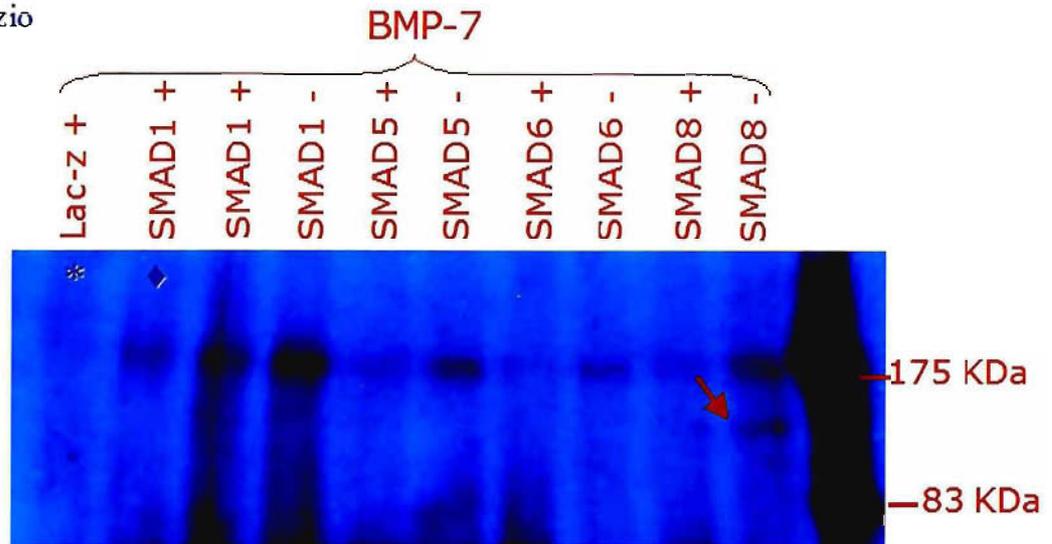


Figura 20 - Super-expressão das construções pCDNA3-6Xmyc-hEPC2 e SMADs 1, 5, 6 e 8 em células 293-T. As células foram transfectadas com a construção pCDNA3-6Xmyc-hEPC2 e infectadas 24 horas depois com os adenovírus contendo as construções das SMADs descritas acima das canaletas. As células foram tratadas com BMP-7 por 2 horas após 24 horas da infecção (+), ou apenas com o veículo carreador de BMP-7 (-). As amostras foram imunoprecipitadas com o antisoro monoclonal α -flag (M2, Sigma) (epítipo das construções de SMAD) e o "blotting" foi feito com antisoro α -EPC2II. Na primeira canaleta o controle foi feito com lisado protéico não transfectado, mas infectado com a construção contendo o gene Lac-Z e tratado com BMP-7. Na segunda canaleta o controle foi feito com células transfectadas com o vetor pCDNA3 vazio, e infectado com a construção de SMAD1, tratadas com BMP-7.

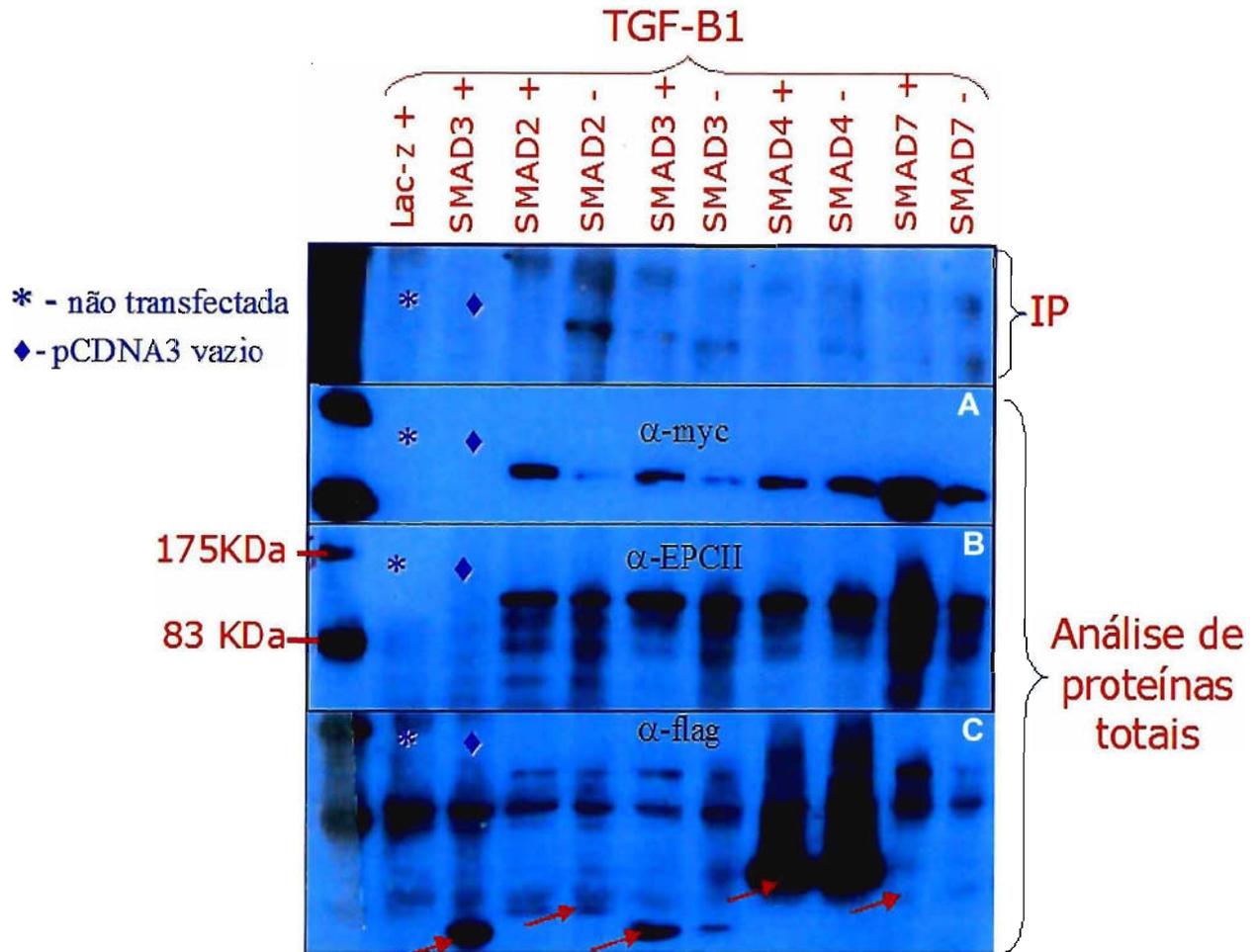


Figura 21 – Análise das proteínas totais provenientes das células 293-T transfectadas com as construções pCDNA3-6Xmyc-hEPC2 e infectadas com os adenovírus com as construções das SMADs 2, 3, 4 e 7. Abaixo da figura da co-imunoprecipitação está a análise das proteínas totais: A: com antisoro α -myc (epítipo de pCDNA3-6X myc-hEPC2); B: com antisoro α -EPC2-II; C: com antisoro α -FLAG (epítipo das SMADs).

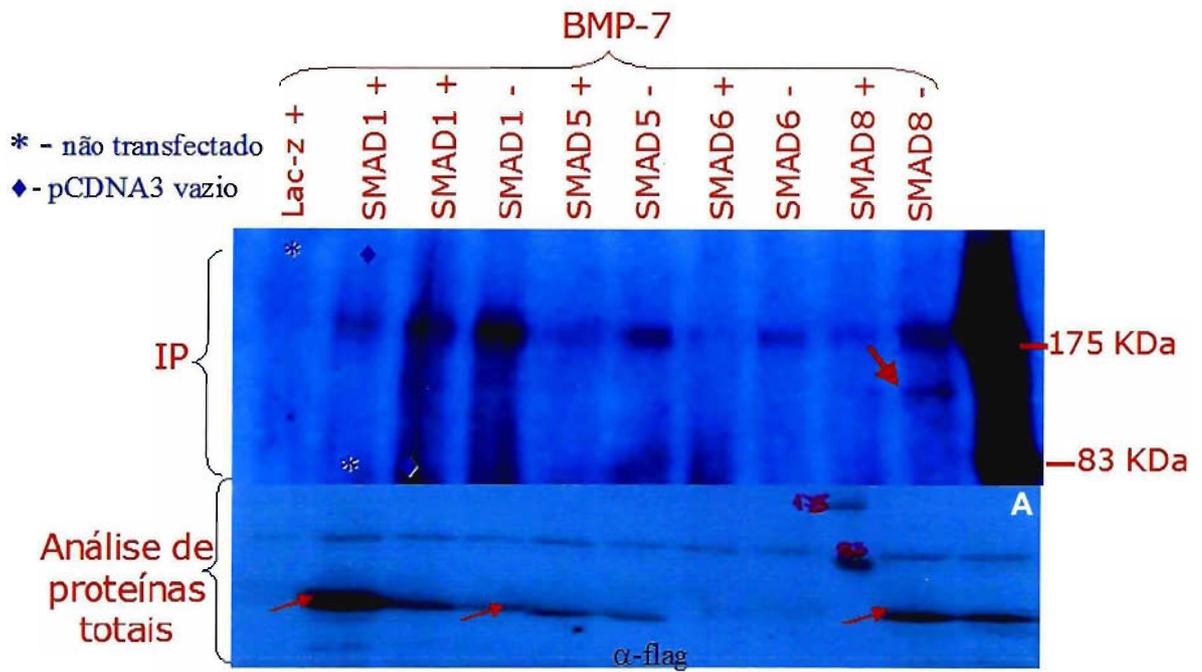


Figura 22 – Análise das proteínas totais provenientes das células 293-T transfectadas com as construções pCDNA3-6Xmyc-hEPC2 e infectadas com os adenovírus com as construções das SMADs 1, 5, 6 e 8. Abaixo da figura da co-imunoprecipitação está a análise das proteínas totais: A: com antisoro α -FLAG (epítipo das SMADs).

Portanto, apesar de ser marcante o aumento da expressão de hEPC2 nas condições tratadas com TGF- β 1, a interação entre as SMADs e hEPC2 é exacerbada nestas mesmas amostras. Esse resultado é ainda mais destacado para as SMAD2 e 3. No entanto, ao analisar os lisados protéicos com o antisoro anti-flag, foi possível observar que a expressão das construções adenovirais de SMAD não se mostraram normalizadas (Figura 21 C). SMAD2 se mostrou fracamente expressa, assim como SMAD3 e SMAD 7. No entanto a expressão de SMAD4 se mostrou fortíssima.

Para a confirmação dos resultados de interação de hEPC2 com algumas das SMADs, o mesmo experimento de IP foi feito utilizando-se a construção pCXN2-hEPC2. Como os resultados relativos as SMADs da via de TGF- β 1 se mostraram mais interessantes, somente estas condições foram repetidas. O experimento foi feito da mesma maneira do anterior e novamente foi possível observar claramente a interação entre as SMADs 2, 3, 4 e 7, sendo que as condições onde as SMADs 2 e 3 foram tratadas com TGF- β 1, houve uma nítida redução da interação entre as proteínas. No entanto, a interação entre SMAD4 e hEPC2 não se mostrou modulável pela citocina como observado no experimento anterior. Desta vez, a expressão das SMADs foi normalizada, descartando a hipótese de artefato na modulação de TGF- β 1 na interação das proteínas (Figura 23).

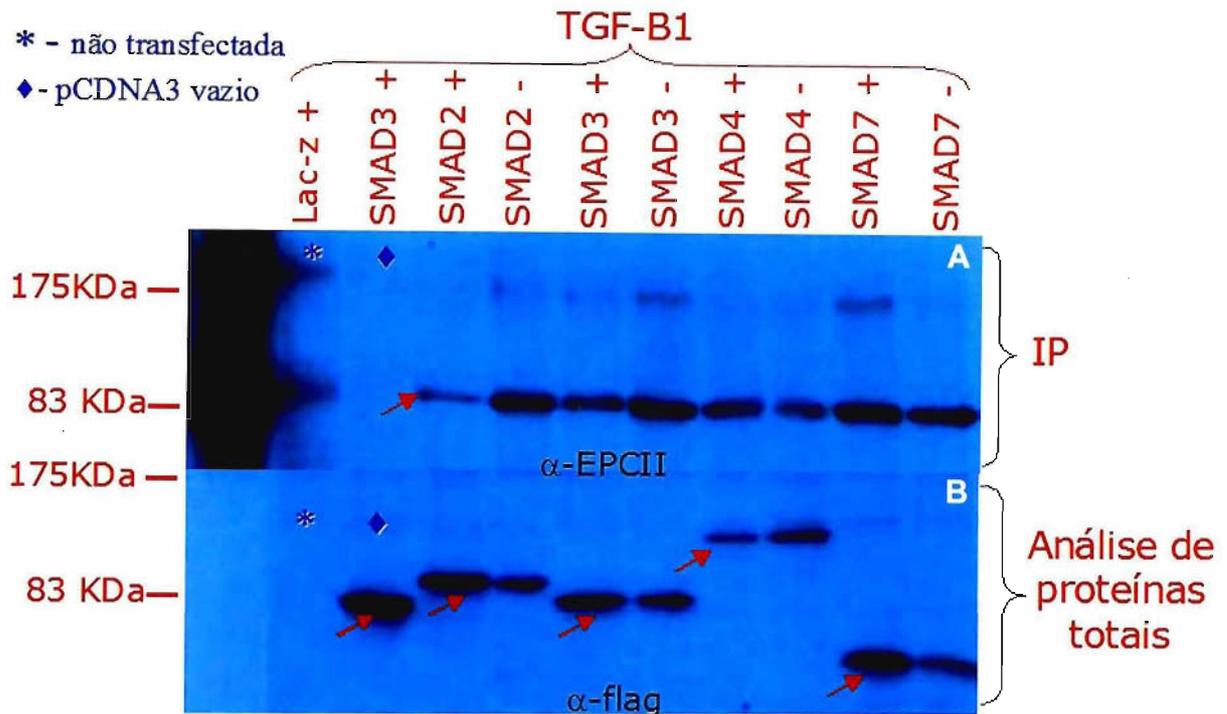


Figura 23 - Super-expressão das construções pCXN2-hEPC2 e SMADs 2, 3, 4 e 7 em células 293-T. **A:** As células foram transfectadas com a construção pCXN2-hEPC2 e infectadas 24 horas depois com os adenovírus contendo as construções das SMADs descritas acima das canaletas. As células foram tratadas com TGF- β 1 por 2 horas após 24 horas da infecção (+), ou apenas com o veículo carreador de TGF- β 1 (-). As amostras foram imunoprecipitadas com o antisoro monoclonal α -flag (M2, Sigma) (epítipo das construções de SMAD) e o "blotting" foi feito com antisoro α -EPC2II. Na primeira canaleta o controle foi feito com lisado protéico não transfectado, mas infectado com a construção contendo o gene Lac-Z e tratado com TGF- β 1. Na segunda canaleta o controle foi feito com células transfectadas com o vetor pCDNA3 vazio, e infectado com a construção de SMAD3, tratadas com TGF- β 1. **B:** Análise das proteínas totais correspondentes ao experimento de co-imunoprecipitação com o antisoro α -FLAG.

Investigação da localização celular de hEPC2

Para confirmar a provável localização nuclear de hEPC2, e também para observar se haveria alteração na localização destas proteínas quando as células fossem tratadas com TGF- β 1, foi realizado um ensaio de imunofluorescência indireta (IF). Inicialmente, este experimento foi feito usando as células 293-T, com as quais os experimentos de IPs haviam sido feitos. No entanto, a aderência destas células ao substrato não é muito forte, e durante as lavagens do experimento de IF, a maioria das células se desprezaram das lâminas. Portanto, foram utilizadas células Cos-7, já que estas células possuem alta eficiência de transfecção e aderem bem às lâminas de vidro, sendo portanto adequadas para este procedimento. As células plaqueadas em lâminas colocadas em placas de 6 poços foram co-transfectadas com as construções de pCXN2-hEPC2 e pCDNA3-SMAD2 (disponível no TSL). Como verificamos que diversas SMADs interagem com hEPC2, foi feita a opção de investigar apenas uma delas neste experimento. A escolha foi pela SMAD2, já que nos IPs, utilizando ambas as construções de hEPC2, foi observado um efeito mais forte de modulação da interação de hEPC2 e SMAD2 por TGF- β 1. As amostras foram transfectadas no primeiro dia do experimento, e no dia seguinte foram tratadas com TGF- β 1 por duas horas antes da IF. Duplicatas foram hibridizadas com antisoros primários α -EPC2, α -flag (tag de SMAD2), e combinando ambos os anticorpos. Foram feitos também, controles com o vetor pCDNA3 vazio incubado com α -EPC2 e testada a detecção da proteína hEPC2 endógena. Anticorpos secundários anti-camundongo conjugado com TRITC e anti-coelho conjugado com FITC foram incubados, respectivamente, com as condições α -flag e α -EPC2. Os resultados confirmaram a localização de hEPC2 no núcleo das células, mas só foi possível detectar a proteína nas condições de super-expressão. Não houve diferença na localização da proteína hEPC2 do citoplasma para o núcleo quando moduladas por TGF- β 1 (Figura 24). Da mesma maneira, não foi possível observar a translocação de SMAD2 do citoplasma para o núcleo quando as células foram tratadas com TGF- β 1 (Figura 25).

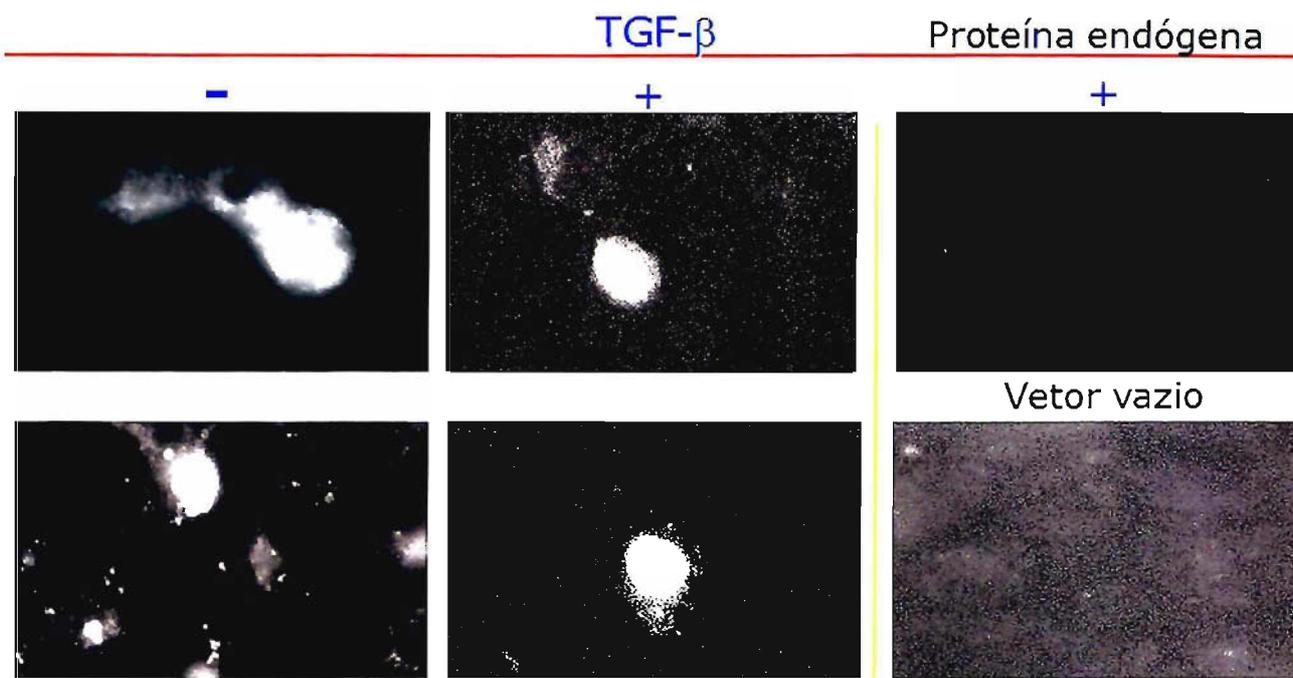


Figura 24 – Ensaio de imunofluorescência indireta mostrando a presença da proteína super-expressa pCXN2-hEPC2 no núcleo de células Cos-7, tratadas (+) e não tratadas (-) com TGF- β 1. A proteína hEPC2 endógena não foi detectada. Como controle da especificidade de detecção da proteína hEPC2, foi feita transfecção do vetor pCXN2 vazio. Visualizado no microscópio Olympus Vanox-T.

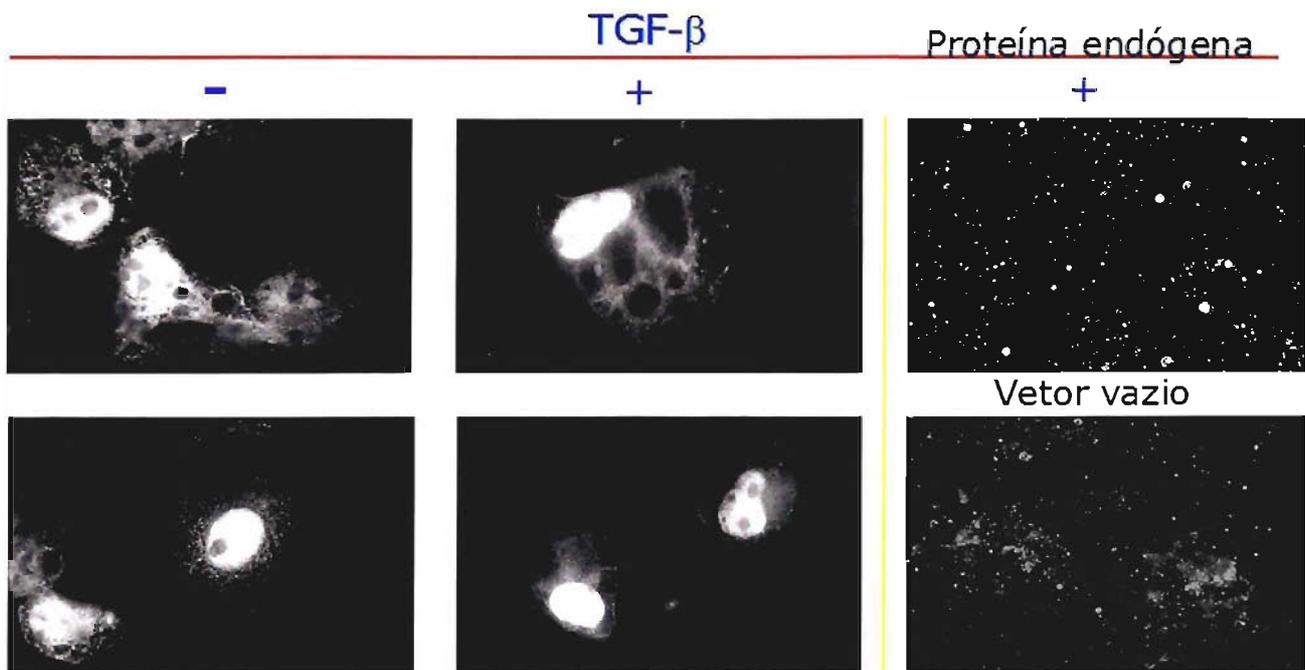


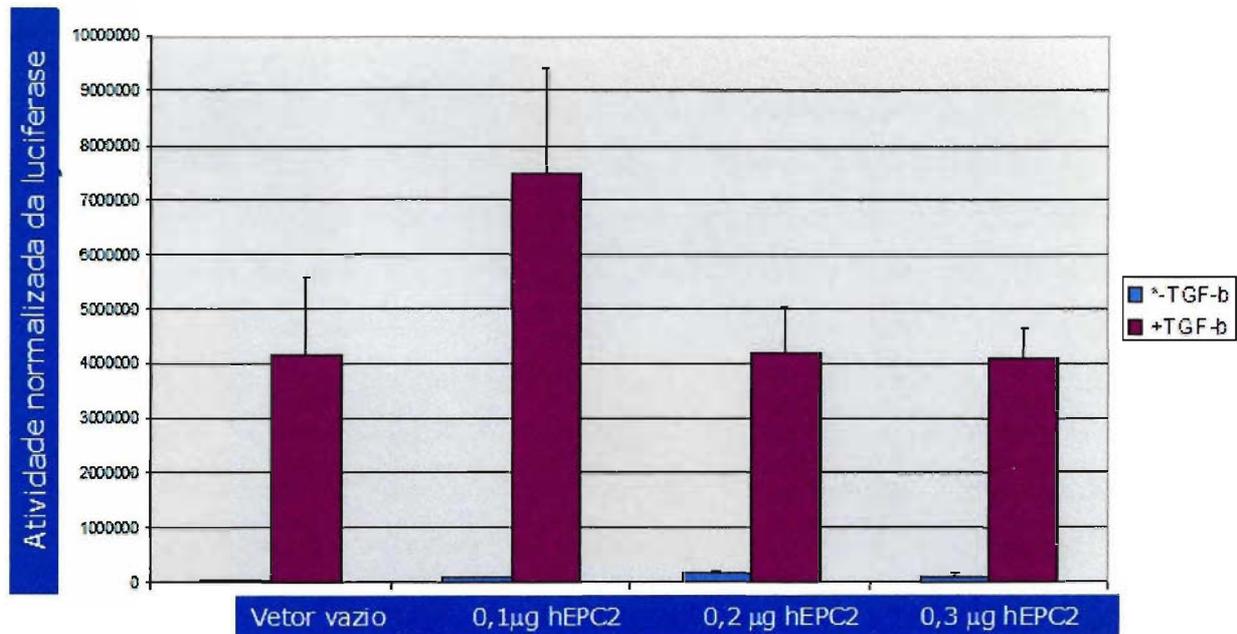
Figura 25 – Ensaio de imunofluorescência indireta mostrando a presença da proteína super-expressa pCDNA3-SMAD2 no núcleo e citoplasma de células Cos-7, tratadas (+) e não tratadas (-) com TGF- β 1. A proteína SMAD2 endógena não foi detectada. Como controle da especificidade de detecção da proteína SMAD2, foi feita a transfecção do vetor pCDNA3 vazio. Visualizado no microscópio Olympus Vanox-T.

Investigação se hEPC2 afeta a transcrição dependente de SMAD

É possível que hEPC2 não se ligue diretamente ao DNA, necessitando se associar como parceiro de um fator de transcrição. Esta foi uma das hipóteses para a atuação de hEPC2, como um co-fator que se associasse com uma das SMADs e aumentasse a transcrição ao nível do elemento responsivo específico de ligação a SMAD (SBE), quando estas proteínas se encontrassem complexadas. Para investigar esta hipótese foi feito um ensaio de gene repórter, utilizando uma construção de um repórter contendo 12 repetições da sequência CAGA (sequência específica de ligação das SMADs 2, 3 e 4) fusionado com o gene da luciferase.

As triplicatas das células transfectadas com 0,1 µg de pCXN2-hEPC2 e tratadas com TGF-β1, mostraram uma ativação do promotor CAGA-12 maior do que o controle do vetor vazio tratado com TGF-β1. No entanto, esse resultado não foi reprodutível nas condições transfectadas com 0,2 e 0,3 µg de DNA, cujo valor de ativação foi igual ao do controle (Figura 26).

Efeito da expressão de hEPC2 no promotor CAGA-12



* O experimento foi feito em triplicata em células HEPG2

Figura 26 – Ensaio de gene repórter onde a construção pCXN2-hEPC2 foi co-transfectada nas células HEPG2 com a construção CAGA-12. As amostras foram transfectadas com quantidades crescentes de pCXN2-hEPC2 (0,1; 0,2 e 0,3 µg de DNA) e foram tratadas ou não com TGF-β1.

Super expressão das construções pGEX- hEPC2

A porção codificante de *hEPC2* foi clonada no vetor de expressão indutível em bactérias pGEX-4T1, que produz a proteína-alvo fusionada com GST (glutathione S-transferase). O constructo foi feito para ser usado em ensaio de "pull-down", também visando testar a interação *in vitro* de hEPC2/SMADs. A técnica consiste em utilizar a proteína de interesse como "isca" para identificar a proteína putativa associada ("presa"). A isca, previamente produzida e purificada em um sistema de expressão (neste caso bacteriano) é imobilizada, servindo como um suporte de afinidade secundário que permite a interação com a "presa". As proteínas "presa" podem ser obtidas de várias fontes incluindo proteínas purificadas, lisados celulares, ou proteínas produzidas por transcrição/tradução *in vitro*.

Um dos clones testados de pGEX-hEPC2 mostrou-se induzido por IPTG. No entanto, várias outras tentativas falharam em mostrar a proteína de fusão super-expressa quando comparada com a construção não induzida. Vários clones foram testados e outra linhagem de bactéria foi testada (BL21-DE3), no entanto sem sucesso. Um experimento de Western-blot usando antisoro α -EPC2 mostrou uma banda de aproximadamente 84 KDa presente no vetor vazio, assim como em 3 clones pGEX-hEPC2 nas condições induzida e não-induzida (Figura 27). Provavelmente essa banda é inespecífica pois o tamanho é menor do que o esperado (que é 117 KDa) e ela também está presente nos controles. No entanto, esta mesma banda se mostrou mais forte nas condições induzidas, sugerindo uma possível modulação por IPTG. Ao fazer "blotting", agora com antisoro α -GST, uma banda de aproximadamente 70KDa apareceu apenas nas amostras não-induzidas. Os controles de BL21-DE3 não-transformada e transformada com o vetor vazio nas condições induzidas e não induzidas, ficaram limpos (Figura 28). No entanto, é inexplicável o porque da banda se mostrar presente apenas nos clones não-induzidos e com um tamanho menor do que o esperado. Provavelmente, se trata de mais uma banda inespecífica.

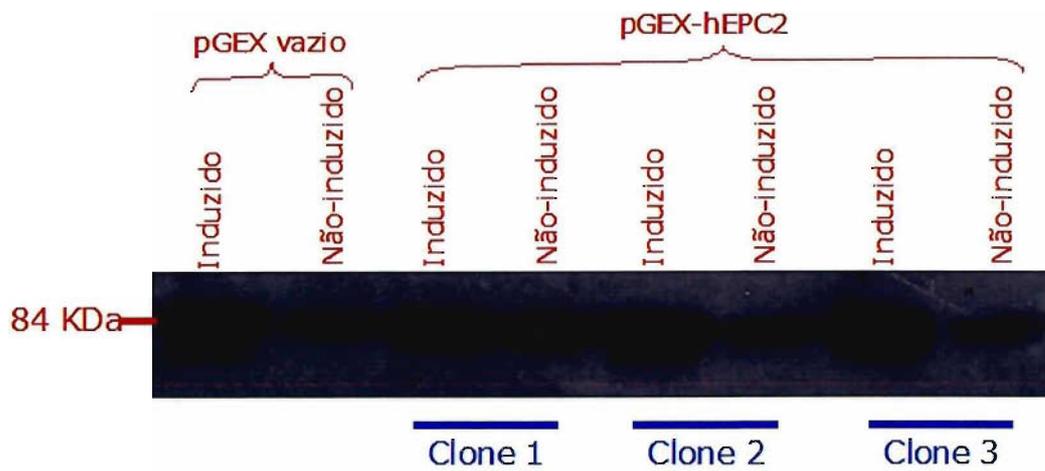


Figura 27 – “Blotting” da construção pGEX-hEPC2 utilizando o antígeno α -EPC2-II. As culturas de bactéria BL21-DE3 transformadas com o vetor pGEX vazio e três clones de 4T1 pGEX-hEPC2, foram induzidas e não-induzidas por IPTG a 1 mM.

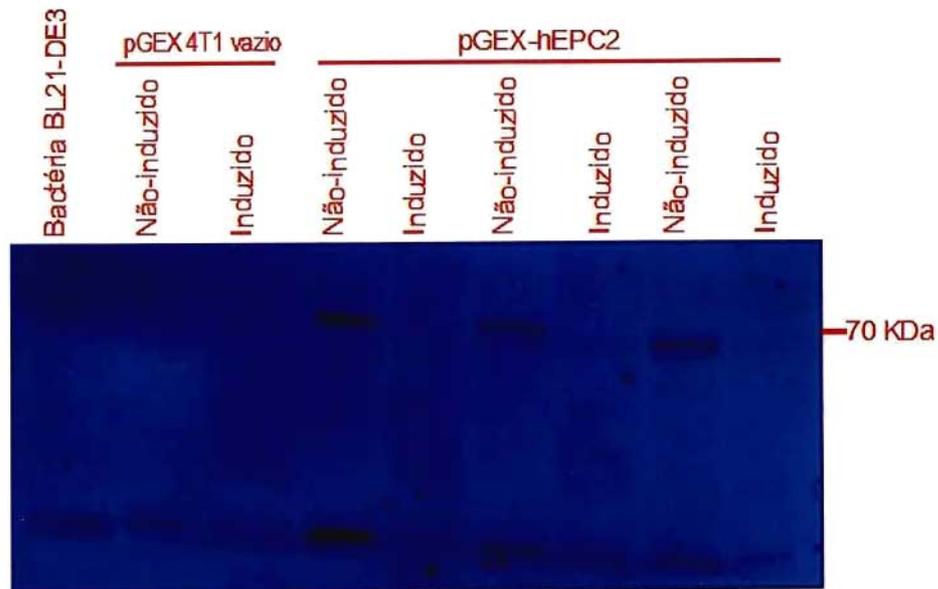


Figura 28 – “Blotting” da construção pGEX-hEPC2 utilizando o antisoro α -GST (Amersham). As culturas de bactéria BL21-DE3 não transformadas, transformada com o vetor pGEX vazio e três clones de 4T1 pGEX-hEPC2, foram induzidas ou não, por IPTG a 1 mM.

No intuito de solucionar o problema, foram testadas induções dos clones com diferentes concentrações de IPTG (0,4; 0,6 e 1M) e por diferentes períodos de tempo (3 e 6 horas). Como a proteína poderia estar formando corpos de inclusão, e por isso não estar sendo detectada nos géis, as culturas foram induzidas, e posteriormente o “pellet” foi sonicado. No entanto, mais uma vez, a proteína não se mostrou presente (Figura 29). Todos os clones foram re-sequenciados, além de os DNAs correspondentes terem sido digeridos para confirmação de que estes não haviam sofrido recombinação. Os clones que estavam sendo crescidos à temperatura de 37°C em todos os experimentos anteriores, foram também crescidos à 16 °C e à 25 °C por 6 horas. Essa abordagem visava evitar a formação de corpos de inclusão, mas novamente a proteína não se mostrou expressa.

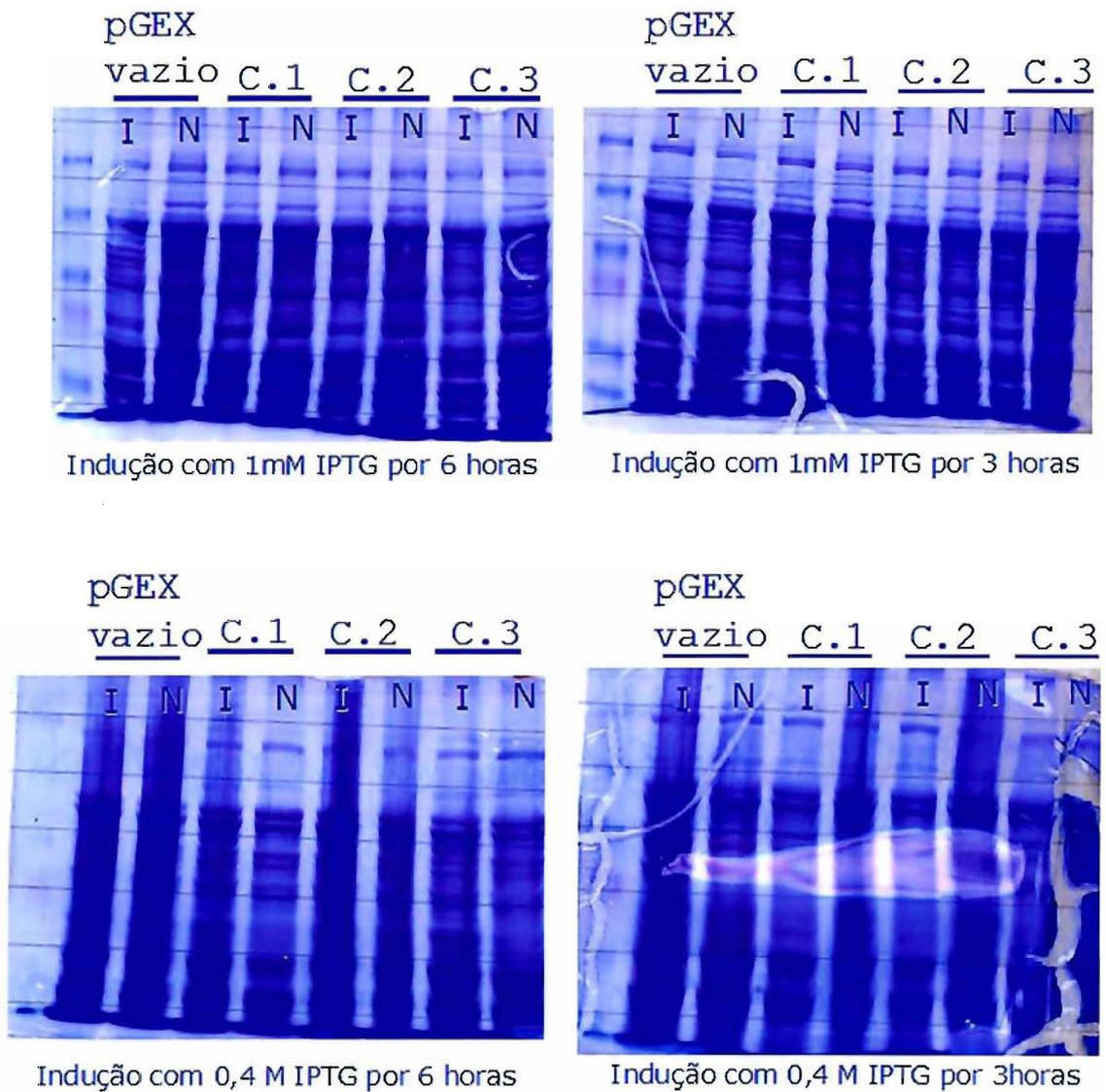


Figura 29 – Géis de SDS-poliacrilamida a 8% corados com Comassie Brilliant Blue, mostrando os clones 1, 2 e 3 de 4T1 pGEX-hEPC2 em diferentes condições de indução por IPTG. A concentração do IPTG foi testada neste experimento (1mM e 0,4 mM), assim como o tempo de indução de expressão (3 e 6 horas).

Os DNAs destes clones foram levados para o TSL e re-transformados nas cepas bacterianas DH10B e BL21-DE3. Novamente, foram testados experimentos de indução sem sucesso. Existem outros casos documentados em grupos de discussão da área que tiveram insucesso na expressão de proteínas de fusão nos vetores pGEX.

Diante destes resultados optamos por clonar o cDNA de *hEPC2* no vetor indutível de expressão em bactérias pET32A (Novagen) (Figura 30).

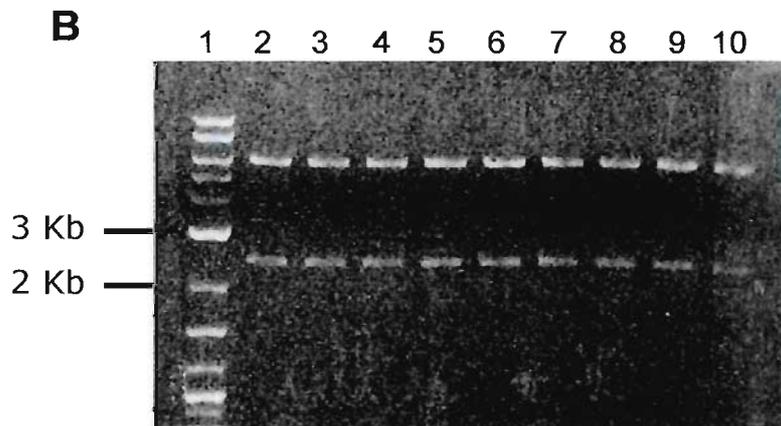
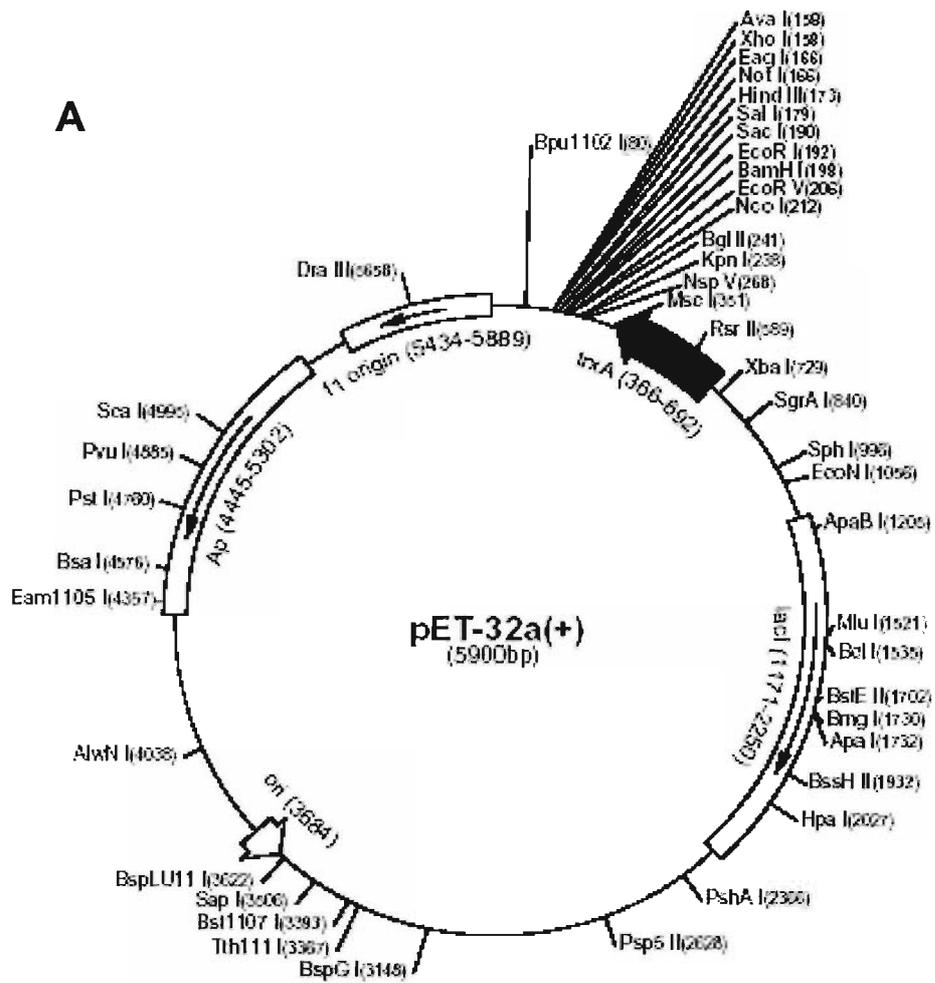


Figura 30 – A: Mapa do vetor pET32A (Novagen). B: 1: Padrão de peso molecular; 2 a 10: Clones de pET32A-hEPC2 digeridos com *EcoRI* e *XhoI*, a banda mais baixa (aproximadamente 2,7 Kb) corresponde a hEPC2.

Super expressão e purificação da proteína recombinante pETA- hEPC2

O cDNA de *hEPC2* foi então clonado no pET32A ainda no TSL, sequenciado no LBCM e transformado em BL32-DE3. Nove clones de pET32-hEPC2 em *E.coli* BI21-DE3 foram isolados, crescidos em cultura líquida até a densidade ótica de 0,6 e então induzidos com 1mM IPTG por 6 horas. As culturas foram separadas em "pellet" e sobrenadante e estas amostras foram solubilizadas e desnaturadas em tampão Laemli e visualizadas em géis desnaturantes de SDS-poliacrilamida corados com Coomassie Brilliant Blue. Neste primeiro experimento, ainda não foi possível observar a indução da proteína de fusão pET32-hEPC2, provavelmente porque muita proteína foi aplicada ao gel, especialmente nas amostras de "pellet". No experimento seguinte, a indução foi feita da mesma maneira, no entanto com as amostras de "pellet" mais diluídas. Desta maneira, foi possível visualizar a indução da expressão de pET32-hEPC2 em todos os 8 clones testados, sendo que os controles do vetor vazio e do lisado da bactéria BI21-DE3 não apresentaram a banda observada nas amostras dos clones induzidos por IPTG (Figura 31). A banda presente nestas amostras possui um tamanho compatível com o peso molecular esperado de 117 KDa.

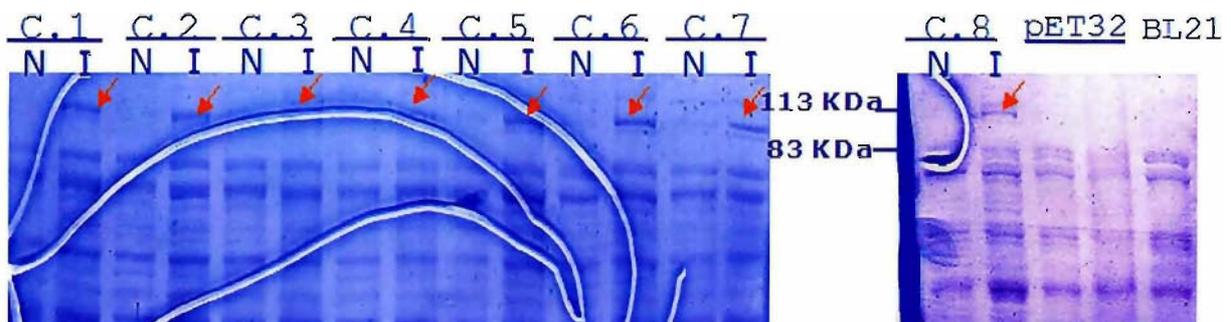


Figura 31 – Gel 8% de SDS-Poliacrilamida corado por Coomassie Brilliant Blue, mostrando a indução da expressão dos clones de pET32-hEPC2 com 1mM IPTG durante 6 horas. A proteína recombinante se encontra indicada pelas setas.

Legenda: C – clone; N – não induzido; I – induzido.

Para confirmar que a banda presente nas amostras induzidas dos clones realmente se referia a expressão da construção pET32-hEPC2, foi feito um ensaio de Western-blot usando o antisoro α -EPCII. Este experimento permitiu confirmar que a banda observada nos géis corados com Comassie Blue Brilliant se referia à proteína EPC2 (Figura 32).

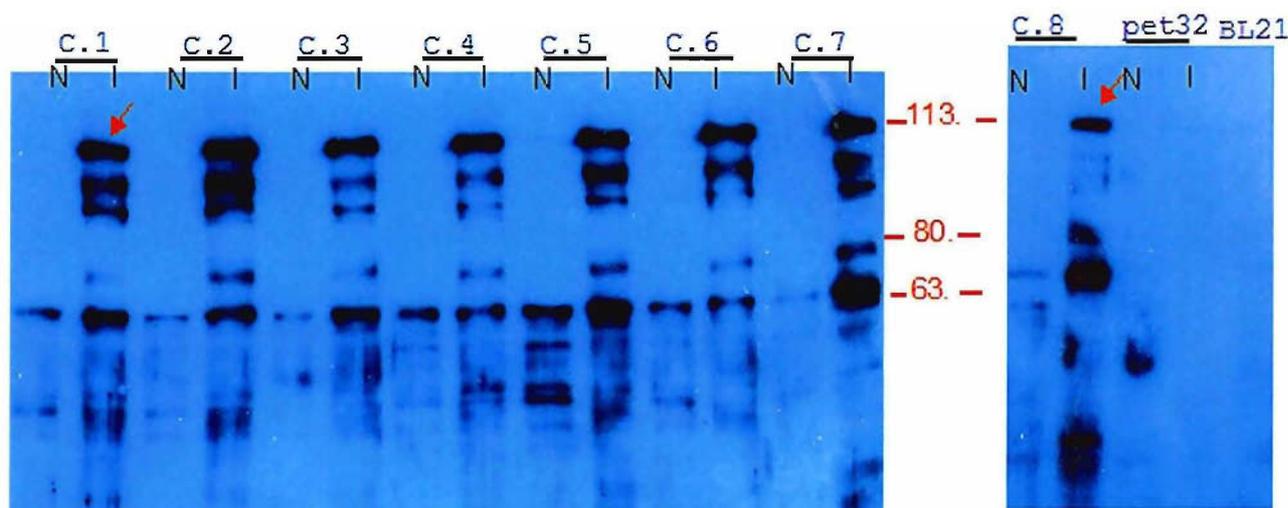


Figura 32 – Ensaio de “Western-blot” mostrando a detecção da proteína de fusão pET32-hEPC2 (indicada pela seta) pelo antisoro α -EPCII.

Legenda: C – clone; N – não induzido; I – induzido.

O próximo passo foi partir para a purificação da proteína pET32-hEPC2 em condições nativas em colunas de cromatografia de afinidade de NiSO₄. Durante o processo de purificação da proteína, foi possível observar que a proteína estava se ligando fracamente à resina impregnada com NiSO₄, já que durante as lavagens a proteína já estava sendo eluída da coluna. Algumas variáveis foram testadas no intuito de melhorar a ligação da proteína na coluna, no entanto nenhuma delas funcionou. Portanto, a próxima opção foi tentar a purificação da proteína em condições desnaturantes, já que a proteína recombinante não depende da estrutura terciária para sua interação com o NiSO₄. Além disso, através desta medida o epítipo de histidina poderia se tornar mais exposto permitindo uma ligação mais eficiente.

A purificação da proteína pET32-hEPC2 em condições desnaturantes funcionou e esta se mostrou presente no eluato, co-purificada com algumas proteínas menores contaminantes de menor intensidade (Figura 33). É possível que estas proteínas menores sejam produto de degradação da própria pET32-hEPC2, no entanto a intensidade destas contaminantes não deve interferir no ensaio de "pull-down".

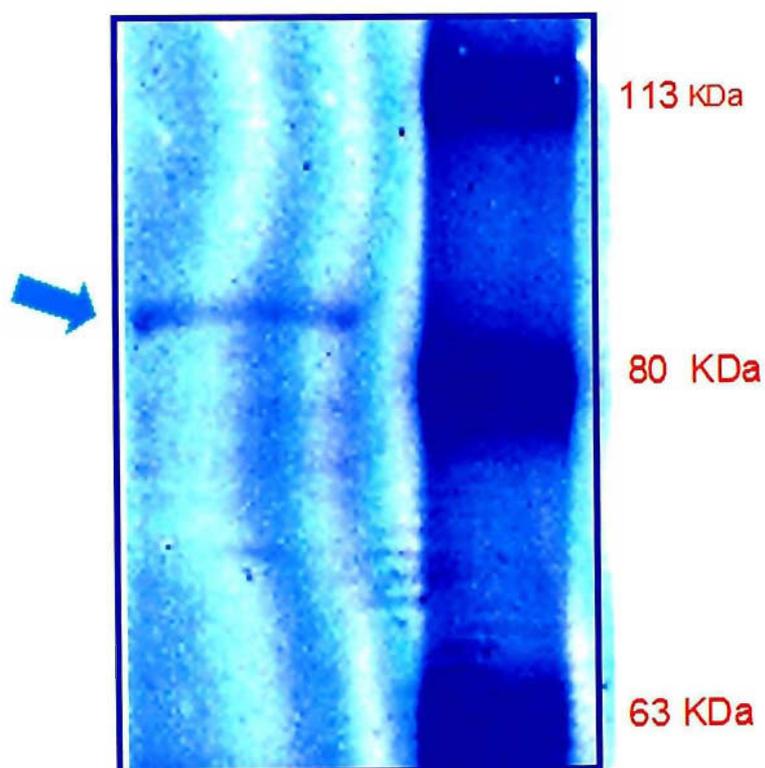


Figura 33 – Purificação da proteína recombinante pET32-hEPC2 (indicada pela seta), sob condições desnaturantes.

DISCUSSÃO

Interação de hEPC2 com as proteínas SMADs e modulação do complexo por TGF- β

O gene *EPC* de *Drosophila* é conhecido por ser um supressor de variegação por efeito de posicionamento, um mecanismo que reprime a expressão gênica através da extensão da eucromatina dentro de regiões genômicas de heterocromatina (Stankunas *et al.*, 1998). Portanto, *EPC* pode regular a compactação da cromatina, levando assim, à modulação da expressão gênica. Como os genes da família dos Polycomb (PcG) são importantes repressores da expressão gênica e desempenham um papel importante na manutenção da diferenciação celular (Francis & Kingston, 2001), é possível que hEPC2 possa agir na via de TGF- β como um co-repressor associado as SMADs. No entanto, *hEPC1* e *hEPC2* são também classificados dentro do grupo dos ETP, o que acrescenta a possibilidade de hEPC2 poder também agir como um ativador transcricional (Brock & van Lohuizen, 2001). Ao contrário de hEPC1, hEPC2 possui um domínio de "PHD zinc finger", o que reforça que a proteína se ligue diretamente ao DNA, agindo como um fator de transcrição e podendo, desta maneira, atuar como um parceiro das SMADs. Outro dado que pode indicar a interação das SMADs com proteínas PcG é a recente publicação do grupo do Dr. Moustakas, mostrando que a proteína da família dos PcG YY1 interage com as SMADs 1, 2, 3 e 4 e inibe a atividade transcricional das SMADs de uma maneira gene-específica (Kurisaki *et al.*, 2003).

As SMADs são classificadas em 3 categorias: - As SMADs receptoras (R-SMADs): 1, 2, 3, 5 e 8; - A SMAD mediadora comum (Co-SMAD): 4; e as SMADs inibitórias (I-SMADs): 6 e 7. Ambas as SMADs receptoras e SMAD4 são capazes de ativar a transcrição, e esta função reside primariamente dentro do domínio MH2. Foi mostrado que a ativação transcricional pelas R-SMADs pode ocorrer, pelo menos em parte, por sua habilidade em recrutar os co-ativadores gerais p300 e CBP. p300 e CBP são proteínas grandes com

diferentes regiões para a interação com diferentes fatores de transcrição. p300 e CBP possuem atividade de histona acetil-transferase (HAT), sugerindo que seu recrutamento pelo complexo SMAD possa aumentar a transcrição de genes-alvo pela alteração da estrutura do nucleossomo e desta forma, pelo remodelamento da cromatina (Moustakas *et al.*, 2001). A interação é diretamente mediada pelo domínio MH2 das R-SMADs (Shi, 2001). Como no caso da interação entre p300 e SMAD4, que foi mapeada na porção N-terminal do domínio MH2 (de Caestecker *et al.*, 2000).

Têm sido proposto que as proteínas SMAD ativam a transcrição através do alívio da ação de repressores transcricionais. Membros da família Hox de proteínas que possuem o homeodomínio, reprimem a transcrição quando ligadas aos seus sítios de DNA cognatos. Evidências recentes sugerem que um complexo de SMAD ativado por BMP é capaz de aliviar a repressão da osteopontina pela interação direta com Hoxc-8 (Shi *et al.*, 1999). Interessantemente, esta de-repressão parece ocorrer pela proteína de ligação SMAD1 e pelo desligamento de Hoxc-8 do DNA. Outro repressor, SIP1 ("SMAD interacting protein 1"), foi identificado por sua habilidade de interagir com o domínio MH2 de SMAD1 (Verschueren *et al.*, 1999). SIP1 se liga ao DNA através de dois conjuntos separados de "zinc fingers", possuindo um homeodomínio e uma região de interação com SMAD (Remacle *et al.*, 1999). A ligação de SMAD pode bloquear a habilidade de SIP1 de reprimir a transcrição ou remover SIP1 do DNA, aliviando assim a repressão. A proteína "zinc-finger" repressora transcricional Evi1, que também existe como uma versão truncada gerada pelo rearranjo cromossomal de MDS1/Evi1 em leucemia, foi descrita por antagonizar a sinalização de SMAD pela ligação a SMAD3 e interferir com sua ligação no DNA (Kurokawa *et al.*, 1998). A proteína Brinker de *Drosophila* é provavelmente um repressor transcricional, e sua ação é prevenida pela sinalização de SMAD (Campbell & Tomlinson, 1999; Jazwinska *et al.*, 1999); (Minami *et al.*, 1999). Ainda não é claro se isto é via interações diretas com as SMADs, ou através da sinalização destas proteínas.

Um complexo de SMAD ligado ao DNA tem a opção de recrutar não só co-ativadores, mas também co-repressores. A proteína contendo homeodomínio TIGF reprime a transcrição dependente de SMAD em parte pelo

recrutamento de histonas de-acetilases (HDACs) (Wotton *et al.*, 1999a). TGIF interage com SMAD2/3 de uma maneira indutível por TGF- β , resultando no recrutamento de TGIF por elementos responsivos a SMAD no DNA. O recrutamento de TGIF e HDACs associadas resulta na repressão da transcrição mediada por SMADs. Isso pode afetar genes que são normalmente ativados pela sinalização de TGF- β . Para tais genes, o nível de expressão de TGIF parece ajustar o máximo de resposta transcricional para TGF- β , em parte competindo com p300 pela ligação com o complexo SMAD. O homeodomínio de TGIF por si só se liga ao elemento responsivo RXR (Bertolino *et al.*, 1995), podendo reprimir a transcrição deste elemento independentemente das SMADs (Wotton *et al.*, 1999b). Não está claro se a função de ligação ao DNA de TGIF, está envolvida no contexto do complexo transcricional de SMAD.

Outros co-repressores da via de TGF- β são as proteínas c-Ski e SnoN. Ski foi originalmente descoberta como um produto do oncogene (*v-Ski*), que causa transformação em fibroblastos de embriões de frango e hipotrofia muscular em camundongo. Seu parceiro celular c-Ski e a proteína relacionada Sno-N, foram descobertos mais tarde por serem co-repressores que recrutam HDAC via a proteína adaptadora N-CoR (Luo *et al.*, 1999). C-Ski e Sno-N foram identificados por interagirem com as SMADs 3 e 4, tendo sido descrito que ambos agem como co-repressores das SMADs 2/3 (Akiyoshi *et al.*, 1999; Luo *et al.*, 1999; Sun *et al.*, 1999a). Em contraste com a interação das SMADs com TGIF, que é induzido pela estimulação de TGF- β , a interação das SMADs com Ski e SnoN é observada sob condições basais e desaparece durante as primeiras horas de estimulação com TGF- β (Stroschein *et al.*, 1999); (Sun *et al.*, 1999b). Este efeito parece ser tipo celular-dependente, podendo ser mediado pela indução de TGF- β e sua degradação mediada por proteassoma.

A regulação da transcrição por TGF- β depende portanto, da habilidade das SMADs em recrutar proteínas com diferentes atividades remodeladoras de cromatina. TGIF pode atuar como um regulador negativo de complexos SMADs ativados por agonistas, enquanto que Ski e SnoN podem servir para proteger contra transativação independente de agonistas pelo nível basal das SMADs. Uma onda de expressão de Sno-N algumas horas após a adição de TGF- β

pode contribuir para a finalização da resposta de TGF- β (Stroschein *et al.*, 1999).

Apesar da lista de fatores de transcrição que interagem com as SMADs ser grande, é interessante observar que até agora nenhuma outra proteína foi descrita por possuir a particularidade de interagir com SMADs de diferentes categorias (Moustakas *et al.*, 2001). No entanto hEPC2 parece interagir com as R-SMADs (2, 3 e 8), Co-SMAD (4) e I-SMAD (7). Talvez, este seja mais um dado que indica o papel singular de hEPC2 na sinalização de TGF- β 1. Outro dado intrigante observado na interação entre hEPC2 e as SMADs, é a modulação desta interação por TGF- β , especialmente entre hEPC2 e as SMADs 2 e 3. Na presença da citocina, a interação entre hEPC2 e as SMADs 2, 3, 4, 7 e 8 se mostra mais fraca, quando foi utilizado a construção pCDNA3-6Xmyc-hEPC2. Este mesmo resultado é observado utilizando a construção pCXN2-hEPC2 para a interação entre hEPC2 e as SMADs 2 e 3. Ainda não se sabe qual o significado desta modulação negativa da interação entre hEPC2 e as SMADs, no entanto, será importante observar se esta modulação da interação também é observada em outros tipos celulares. Isso porque é bastante provável que, sendo a regulação de transcrição por TGF- β um processo dinâmico, estas interações dependam do contexto celular. Outro ponto que deverá ser investigado é o mapeamento dos domínios de interação entre hEPC2 e as SMADs, uma vez que estes resultados possam ajudar a desvendar qual é a atuação destas proteínas. A investigação da interação das outras isoformas de hEPC2, como Iso-B e Iso-C, que possuem apenas o domínio EPcA (que está relacionado com ativação de transcrição em hEPC1), também será um estudo interessante. Será que as proteínas codificadas por estas isoformas também interagem com as SMADs?

Transcrição dependente de SMADs

Pela associação com parceiros de ligação ao DNA, formando complexos de composições específicas, as SMADs podem adquirir alta afinidade e interações seletivas com o DNA. Os domínios de ligação ao DNA das SMADs e

de seus parceiros no mesmo complexo, serão capazes de agir sinergisticamente se suas sequências cognatas estiverem presentes na distância correta de cada um em um promotor alvo (Massague & Wotton, 2000).

A fusão do domínio MH2 das SMADs a um domínio de ligação ao DNA heterólogo, como por exemplo Gal4p, demonstrou que, uma vez recrutado ao DNA, o complexo de SMAD é capaz de ativar a transcrição (Liu *et al.*, 1996). Ambos R-SMADs e Co-SMADS podem se ligar ao DNA via o domínio MH1 (Kim *et al.*, 1997; Shi *et al.*, 1998). A ligação ótima é obtida através da sequência CAGAC (Shi *et al.*, 1998; Zawel *et al.*, 1998). Alguns "SMAD binding elements (SBE)" estão frequentemente presentes nas regiões responsivas de genes-alvo de TGF- β , Activina ou BMP. A estrutura tri-dimensional do domínio MH1 de SMAD3 ligado ao seu DNA cognato, mostra que o domínio monômero MH1 se liga precisamente à metade desta sequência (Shi *et al.*, 1998). De modo a sugerir que, em vários casos, é necessária a associação de outra proteína com as SMADs, para que o complexo seja capaz de ativar ou reprimir especificamente a expressão do gene-alvo.

Calcula-se que a sequência CAGAC, que é o elemento responsivo a SMAD, esteja presente no genoma, na média, uma vez a cada 1024 bp, ou por volta de uma vez na região regulatória de qualquer gene de tamanho médio. Se a ligação ao SBE for suficiente para a ativação transcricional dependente de SMAD, uma proteína SMAD poderá levar a ativação não-seletiva de um número massivo de genes. No entanto, este não é o caso (Massague & Wotton, 2000). A ativação de genes-alvo somente via ligação em um SBE não é possível por duas razões. Primeiramente, a afinidade de um domínio MH1 de uma SMAD com um SBE é da ordem de $10^{-7}M$ (Shi *et al.*, 1998), o que é muito fraco para uma ligação efetiva *in vivo* sem o envolvimento de contatos adicionais com o DNA. Sabe-se que são necessários muitos SBEs, formados por concatêmeros artificiais, para conseguir a ativação de uma SMAD em um gene repórter (Zawel *et al.*, 1998). Em segundo lugar, a ligação de SMAD em um SBE perde a seletividade, já que SMADs 1, 3 e 4 tem uma afinidade similar ao SBE. Isto não é surpreendente porque a sequência " β -hairpin" é idêntica em todos os R-SMADs e altamente conservada em SMAD4. Portanto, contatos

adicionais parecem ser necessários para a especificidade, e alta afinidade de ligação do complexo SMAD a um gene-alvo. Uma região regulatória de jun-B contendo múltiplas cópias de SBE foi mostrada, quando testados como dímeros, por se ligar a SMADs e permitir a ativação de um gene repórter quando co-transfectado com SMADs. No entanto, a habilidade de TGF- β em ativar o promotor natural de junB através deste elemento permanece por ser confirmado, assim como o envolvimento de outras proteínas como co-fatores de SMADs endógenos nesta resposta ainda não foi descoberto (Massague & Wotton, 2000).

A interação das SMADs com fatores de ligação ao DNA, resultando em uma forte ligação do complexo ao DNA, foi primeiramente descrita pela proteína de *Xenopus* FAST-1 (Chen *et al.*, 1996). FAST-1 é conhecido por mediar a ativação de um painel de genes homeobox no desenvolvimento do mesoderma (Watanabe & Whitman, 1999). Na transdução de sinal de BMP, a proteína OAZ foi identificada como um co-fator de ligação ao DNA, que se associa com um complexo SMAD1-SMAD4 ativado e permite o reconhecimento e ativação do gene homeobox Xent-2 (Hata *et al.*, 2000).

Pouco se conhece a respeito do mecanismo de atuação dos complexos multiméricos das proteínas PcG na manutenção da repressão da expressão dos genes-alvo. No entanto, sabe-se que em *Drosophila* estas proteínas se ligam a elementos regulatórios específicos chamados Elementos Responsivos a Polycomb (PRE). Cerca de 100 genes, incluindo os homeóticos, são potencialmente regulados pelos PcG e Trx, mas apenas 10 PRE foram caracterizados até agora (Ringrose *et al.*, 2003).

Já que para as SMADs ativarem a expressão gênica específica em determinado contexto celular elas geralmente se associam a outras proteínas, nossa hipótese de estudo foi investigar se a associação de hEPC2 com as SMADs, afetaria sua transcrição de uma maneira dose-dependente. Para iniciar estes experimentos optamos por testar a interação entre hEPC2 e SMAD2, pois dentre as SMADs esta foi a que mostrou os efeitos mais dramáticos de interação com hEPC2, e de modulação por TGF- β .

O experimento não se mostrou conclusivo, já que as condições utilizando diferentes concentrações de DNA da construção pCXN2-hEPC2, não

tiveram o mesmo resultado de ativação do promotor. Apenas a condição onde as células foram transfectadas com 0,1 µg de pcXN2-hEPC2 (testadas em triplicata) e tratadas com TGF-β mostrou uma ativação da construção do promotor repórter. Este experimento será re-feito, além disso é possível que sua repetição em outra linhagem celular, ou utilizando outra construção de promotor, resulte em resultados mais conclusivos. Outro ponto que também deve ser testado é a associação de hEPC2 com as outras SMADs.

Localização nuclear de hEPC2

As proteínas da família dos PcG geralmente se localizam no núcleo das células, mais especificamente no compartimento da pericromatina situado na superfície de domínios condensados de cromatina (Cmarko *et al.*, 2003). Foi mostrado anteriormente, que estes compartimentos são sítios nucleares onde a maioria dos pré-mRNAs são sintetizados. Este dado indica que genes ativos e silenciados estão proximamente associados no espaço, sugerindo que o silenciamento causado pelos PcG é um evento local, ao invés de afetar grandes domínios de cromatina.

O ensaio de imunofluorescência indireta permitiu confirmar a localização de hEPC2 no núcleo das células, como foi predito pelos sinais de localização nuclear obtidos pela análise da seqüência da proteína hEPC2 pelo programa PSORTII (ver parte I, Análise dos domínios protéicos de hEPC2).

No entanto, não foi possível observar a co-localização das proteínas hEPC2 e SMAD2, mesmo quando estas foram tratadas com TGF-β1. Porém, mesmo SMAD2, que é sabidamente translocada do citoplasma para o núcleo quando ativada por TGF-β1, não sofreu tal efeito neste experimento. Talvez a justificativa para este resultado seja um excesso da proteína SMAD2 super-expressa, que não foi passível de ser modulada por TGF-β1. Esta também pode ser a explicação da não observação da co-localização de hEPC2 e SMAD2. Este experimento deverá ser repetido com uma variação na quantidade das construções de hEPC2 e de SMAD2 que serão super-expressas, para a obtenção de resultados mais conclusivos.

CONCLUSÕES

- A sequência codificante de *hEPC1* apresenta 3.739 nt distribuídos em 14 exons, espalhados em 139 Kb do *locus* 10p11.22, com uma janela aberta de leitura de 2.292 nt, que codificam para 763 aminoácidos.
- *hEPC2* apresenta 3.649 nt com uma janela aberta de leitura de 2.421 nt que codificam para 807 aminoácidos. O gene também possui 14 exons, espalhados em 207 Kb do *locus* 2q23.1.
- *hEPC2* é uma proteína de localização nuclear de acordo com o programa PSORT II, que prediz 7 sinais de localização nuclear. Este dado foi confirmado pelo ensaio de imunofluorescência indireta.
- Os resultados das análises de *hEPC2* pelo programa “Conserved Domain Database” indicam a presença de um domínio de “PHD zinc finger” nos aminoácidos 51 a 104, que pode indicar que *hEPC2* possa agir como um fator de transcrição.
- Sequências similares a *hEPC1* e *hEPC2* são encontradas nos genomas de vários eucariotos, sugerindo uma função importante destes genes que foi mantida durante a evolução.
- As sequências dos parálogos *hEPC1* e *hEPC2* apresentam uma identidade de 53% ao nível de proteína. No entanto, a similaridade à nível de nucleotídeos destes genes só é significativa nos domínios EPcA (74% de identidade) e EPcB (80% de identidade), sugerindo uma divergência antiga entre os dois genes.

- Uma comparação entre os *EPCs* humano e de camundongo, mostrou que a conservação é mais alta entre os ortólogos do que entre os parálogos, reforçando que o evento de duplicação gênica pode ter precedido divergência funcional entre *hEPC1* e *hEPC2*.
- Os resultados de análise da expressão gênica de *hEPC1* e *hEPC2* mostrou que ambos os genes são ubíqua e fracamente expressos na maioria dos tecidos analisados, dentre estes tecidos fetais, adultos e tumorais.
- *hEPC2* possui, além do transcrito canônico (Iso-A), 3 transcritos alternativos. Iso-B e Iso-C foram descobertos através de análises através do programa Acembly. Iso-B possui 1,686 nt distribuídos em três exons e Iso-C tem 1,435 nt distribuídos em dois exons.
- A existência destas isoformas foi confirmada através de RT-PCR semi-quantitativa em um painel de cDNAs, onde Iso-B e Iso-C foram co-amplificadas.
- Iso-B e Iso-C foram amplificadas em quase todas as amostras testadas, exceto em amostras de rim e glândula salivar. As ampliações mostraram que ambas as isoformas possuem uma oscilação de expressão entre os tecidos testados.
- Cinco seqüências, dentre elas 1 mRNA e 4 ESTs, quando comparadas com a seqüência canônica de *hEPC2* mostraram uma 3'UTR mais curta, sugerindo a existência de uma nova isoforma de *hEPC2*. O tamanho destas seqüências é compatível como o tamanho da banda observada no Northern-blot de análise de Iso-B e Iso-C.

- Através de análises utilizando a ferramenta SAGE Genie, foi possível identificar duas SAGE “tags” nestas sequências, uma delas correspondentes a Iso-A, na posição 3,059 do transcrito, com uma frequência de 12 “tags” por 200,000. A outra SAGE “tag” no nucleotídeo 2,804 de Iso-A , com 16 “tags” por 200,000.
- As análises de Northern-blot revelaram a presença de um transcrito de 3.6 Kb e uma banda de aproximadamente 2.8 Kb, compatível com o tamanho de Iso-A e com o transcrito menor de *hEPC2*, chamado Iso-D.
- Iso-D se mostrou expresso em todos os tecidos normais e tumorais avaliados por Northern-blot. No entanto, Iso-D se mostrou mais abundante nas linhagens celulares tumorais, principalmente em linfoma de Burkkit e leucemia promielocítica HL-60.
- Análises de Blast contra o banco nr (não redundante) mostraram duas ESTs similares à *hEPC2*, que foram descritas por serem seqüências parciais de peptídeos que interagem com as proteínas SMAD. Estas proteínas são transdutoras de sinal de TGF- β .
- Ensaio de co-imunoprecipitação entre *hEPC2* e as SMADs 1, 2, 3, 4, 5, 6, 7 e 8 confirmaram a interação entre *hEPC2* e as SMADs 2, 3, 4, 7 e 8.
- As células co-expressando SMAD2, SMAD3, SMAD4 e *hEPC2*, quando tratadas com TGF- β 1, mostraram uma redução na co-imunoprecipitação. Sugerindo que TGF- β 1 modula negativamente a interação entre estas proteínas. Este dado foi observado para ambas as construções para expressão constitutiva de *hEPC2* em mamíferos.

- Ensaio de imunofluorescência indireta mostraram que a proteína hEPC2 se localiza no núcleo de células Cos-7.
- Não foi possível observar a co-localização de hEPC2 e SMAD2 neste mesmo ensaio.
- Ensaio de gene repórter não demonstrou que a transcrição de SMAD2 é afetada por hEPC2 de uma maneira dependente.
- Futuros ensaios de “pull down” serão feitos visando a confirmação da interação de hEPC2 e as SMADs, e mapeamento dos domínios desta interação. Para tal, a construção hEPC2 foi clonado no vetor de expressão indutível em bactérias pGEX32A.
- A proteína recombinante gerada por esta construção foi expressa e purificada em condições desnaturantes.

BIBLIOGRAFIA

- Aalfs J. D., and Kingston R. E. (2000). What does 'chromatin remodeling' mean? *Trends Biochem Sci* **25**: 548-55.
- Adler P. N., Charlton J., and Brunk B. (1989). Genetic interactions of the suppressor 2 of zeste region genes. *Dev Genet* **10**: 249-60.
- Akasaka T., Kanno M., Balling R., Mieza M. A., Taniguchi M., and Koseki H. (1996). A role for mel-18, a Polycomb group-related vertebrate gene, during theanterior-posterior specification of the axial skeleton. *Development* **122**: 1513-22.
- Akasaka T., Tsuji K., Kawahira H., Kanno M., Harigaya K., Hu L., Ebihara Y., Nakahata T., Tetsu O., Taniguchi M., and Koseki H. (1997). The role of mel-18, a mammalian Polycomb group gene, during IL-7-dependent proliferation of lymphocyte precursors. *Immunity* **7**: 135-46.
- Akiyoshi S., Inoue H., Hanai J., Kusanagi K., Nemoto N., Miyazono K., and Kawabata M. (1999). c-Ski acts as a transcriptional co-repressor in transforming growth factor-beta signaling through interaction with smads. *J Biol Chem* **274**: 35269-77.
- Alberts B., and Sternglanz R. (1990). Gene expression. Chromatin contract to silence. *Nature* **344**: 193-4.
- Alkema M. J., van der Lugt N. M., Bobeldijk R. C., Berns A., and van Lohuizen M. (1995). Transformation of axial skeleton due to overexpression of bmi-1 in transgenic mice. *Nature* **374**: 724-7.
- Ausubel F. M. B., R. ; Kingston, R. E. ; Moore, D. D.; Seidman, J. G.; Smith, J. A.; Struhl, K. (1998). "Current Protocols in Molecular Biology," John Wiley & Sons.
- Barges S., Mihaly J., Galloni M., Hagstrom K., Muller M., Shanower G., Schedl P., Gyurkovics H., and Karch F. (2000). The Fab-8 boundary defines the distal limit of the bithorax complex iab-7 domain and insulates iab-7 from initiation elements and a PRE in the adjacent iab-8 domain. *Development* **127**: 779-90.

- Bea S., Tort F., Pinyol M., Puig X., Hernandez L., Hernandez S., Fernandez P. L., van Lohuizen M., Colomer D., and Campo E. (2001). BMI-1 gene amplification and overexpression in hematological malignancies occur mainly in mantle cell lymphomas. *Cancer Res* **61**: 2409-12.
- Berger J., Kurahashi H., Takihara Y., Shimada K., Brock H. W., and Randazzo F. (1999). The human homolog of Sex comb on midleg (SCMH1) maps to chromosome 1p34. *Gene* **237**: 185-91.
- Bertolino E., Reimund B., Wildt-Perinic D., and Clerc R. G. (1995). A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *J Biol Chem* **270**: 31178-88.
- Boivin A., and Dura J. M. (1998). In vivo chromatin accessibility correlates with gene silencing in *Drosophila*. *Genetics* **150**: 1539-49.
- Bornemann D., Miller E., and Simon J. (1996). The *Drosophila* Polycomb group gene Sex comb on midleg (*Scm*) encodes a zinc finger protein with similarity to polyhomeotic protein. *Development* **122**: 1621-30.
- Brentani H., Caballero O. L., Camargo A. A., da Silva A. M., da Silva W. A., Jr., Dias Neto E., et al. (2003). The generation and utilization of a cancer-oriented representation of the human transcriptome by using expressed sequence tags. *Proc Natl Acad Sci U S A* **100**: 13418-23.
- Brock H. W., and van Lohuizen M. (2001). The Polycomb group--no longer an exclusive club? *Curr Opin Genet Dev* **11**: 175-81.
- Brunk B. P., Martin E. C., and Adler P. N. (1991). *Drosophila* genes Posterior Sex Combs and Suppressor two of zeste encode proteins with homology to the murine *bmi-1* oncogene. *Nature* **353**: 351-3.
- Buchenau P., Hodgson J., Strutt H., and Arndt-Jovin D. J. (1998). The distribution of polycomb-group proteins during cell division and development in *Drosophila* embryos: impact on models for silencing. *J Cell Biol* **141**: 469-81.
- Bulger M., and Groudine M. (1999). Looping versus linking: toward a model for long-distance gene activation. *Genes Dev* **13**: 2465-77.
- Bunker C. A., and Kingston R. E. (1994). Transcriptional repression by *Drosophila* and mammalian Polycomb group proteins in transfected mammalian cells. *Mol Cell Biol* **14**: 1721-32.

- Busturia A., Wightman C. D., and Sakonju S. (1997). A silencer is required for maintenance of transcriptional repression throughout *Drosophila* development. *Development* **124**: 4343-50.
- Camargo A. A., Samaia H. P., Dias-Neto E., Simao D. F., Migotto I. A., Briones M. R., Costa F. F., Nagai M. A., Verjovski-Almeida S., et al (2001). The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. *Proc Natl Acad Sci U S A* **98**: 12103-8.
- Campbell R. B., Sinclair D. A., Couling M., and Brock H. W. (1995). Genetic interactions and dosage effects of Polycomb group genes of *Drosophila*. *Mol Gen Genet* **246**: 291-300.
- Care A., Testa U., Bassani A., Tritarelli E., Montesoro E., Samoggia P., Cianetti L., and Peschle C. (1994). Coordinate expression and proliferative role of HOXB genes in activated adult T lymphocytes. *Mol Cell Biol* **14**: 4872-7.
- Chen X., Rubock M. J., and Whitman M. (1996). A transcriptional partner for MAD proteins in TGF-beta signalling. *Nature* **383**: 691-6.
- Cheng N. N., Sinclair D. A., Campbell R. B., and Brock H. W. (1994). Interactions of polyhomeotic with Polycomb group genes of *Drosophila melanogaster*. *Genetics* **138**: 1151-62.
- Chinwalla V., Jane E. P., and Harte P. J. (1995). The *Drosophila* trithorax protein binds to specific chromosomal sites and is co-localized with Polycomb at many sites. *Embo J* **14**: 2056-65.
- Cmarko D., Verschure P. J., Otte A. P., van Driel R., and Fakan S. (2003). Polycomb group gene silencing proteins are concentrated in the perichromatin compartment of the mammalian nucleus. *J Cell Sci* **116**: 335-43.
- Core N., Bel S., Gaunt S. J., Aurrand-Lions M., Pearce J., Fisher A., and Djabali M. (1997). Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. *Development* **124**: 721-9.
- Daubresse G., Deuring R., Moore L., Papoulas O., Zakrajsek I., Waldrip W. R., Scott M. P., Kennison J. A., and Tamkun J. W. (1999). The *Drosophila* kismet gene is related to chromatin-remodeling factors and is required for both segmentation and segment identity. *Development* **126**: 1175-87.

- de Caestecker M. P., Yahata T., Wang D., Parks W. T., Huang S., Hill C. S., Shioda T., Roberts A. B., and Lechleider R. J. (2000). The Smad4 activation domain (SAD) is a proline-rich, p300-dependent transcriptional activation domain. *J Biol Chem* **275**: 2115-22.
- Dias Neto E., Correa R. G., Verjovski-Almeida S., Briones M. R., Nagai M. A., da Silva W., Jr., Zago M. A., Bordin S., Costa F. F., Goldman G. H., Carvalho A. F., Matsukuma A., Baia G. S., Simpson D. H., Brunstein A., de Oliveira P. S., Bucher P., Jongeneel C. V., O'Hare M. J., Soares F., Brentani R. R., Reis L. F., de Souza S. J., and Simpson A. J. (2000). Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc Natl Acad Sci U S A* **97**: 3491-6.
- Dietzel S., Niemann H., Bruckner B., Maurange C., and Paro R. (1999). The nuclear distribution of Polycomb during *Drosophila melanogaster* development shown with a GFP fusion protein. *Chromosoma* **108**: 83-94.
- Ewing B., and Green P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**: 186-94.
- Ewing B., Hillier L., Wendl M. C., and Green P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**: 175-85.
- Faust C., Schumacher A., Holdener B., and Magnuson T. (1995). The eed mutation disrupts anterior mesoderm production in mice. *Development* **121**: 273-85.
- Fauvarque M. O., Zuber V., and Dura J. M. (1995). Regulation of polyhomeotic transcription may involve local changes in chromatin activity in *Drosophila*. *Mech Dev* **52**: 343-55.
- Francastel C., Schubeler D., Martin D. I., and Groudine M. (2000). Nuclear compartmentalization and gene activity. *Nat Rev Mol Cell Biol* **1**: 137-43.
- Francis N. J., and Kingston R. E. (2001). Mechanisms of transcriptional memory. *Nat Rev Mol Cell Biol* **2**: 409-21.
- Fukuyama T., Otsuka T., Shigematsu H., Uchida N., Arima F., Ohno Y., Iwasaki H., Fukuda T., and Niho Y. (2000). Proliferative involvement of ENX-1, a putative human polycomb group gene, in haematopoietic cells. *Br J Haematol* **108**: 842-7.

- Garcia-Cuellar M. P., Zilles O., Schreiner S. A., Birke M., Winkler T. H., and Slany R. K. (2001). The ENL moiety of the childhood leukemia-associated MLL-ENL oncoprotein recruits human Polycomb 3. *Oncogene* **20**: 411-9.
- Gasser S. M. (2001). Positions of potential: nuclear organization and gene expression. *Cell* **104**: 639-42.
- Gaunt S. J., and Singh P. B. (1990). Homeogene expression patterns and chromosomal imprinting. *Trends Genet* **6**: 208-12.
- Gerasimova T. I., Byrd K., and Corces V. G. (2000). A chromatin insulator determines the nuclear localization of DNA. *Mol Cell* **6**: 1025-35.
- Gildea J. J., Lopez R., and Shearn A. (2000). A screen for new trithorax group genes identified little imaginal discs, the *Drosophila melanogaster* homologue of human retinoblastoma binding protein 2. *Genetics* **156**: 645-63.
- Gordon D., Abajian C., and Green P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res* **8**: 195-202.
- Grunstein M. (1998). Yeast heterochromatin: regulation of its assembly and inheritance by histones. *Cell* **93**: 325-8.
- Hata A., Seoane J., Lagna G., Montalvo E., Hemmati-Brivanlou A., and Massague J. (2000). OAZ uses distinct DNA- and protein-binding zinc fingers in separate BMP-Smad and Olf signaling pathways. *Cell* **100**: 229-40.
- Hasegawa M., Tetsu O., Kanno R., Inoue H., Ishihara H., Kamiyasu M., Taniguchi M., and Kanno M. (1998). Mammalian Polycomb group genes are categorized as a new type of early response gene induced by B-cell receptor cross-linking. *Mol Immunol* **35**: 559-63.
- Heldin C. H., Miyazono K., and ten Dijke P. (1997). TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature* **390**: 465-71.
- Imanishi T., Itoh T., Suzuki Y., O'Donovan C., Fukuchi S., Koyanagi K. O., Barrero R. A., Tamura T., et al. (2004). Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biol* **2**: E162.

- Jacobs J. J., and van Lohuizen M. (2002). Polycomb repression: from cellular memory to cellular proliferation and cancer. *Biochim Biophys Acta* **1602**: 151-61.
- Jazwinska A., Kirov N., Wieschaus E., Roth S., and Rushlow C. (1999). The Drosophila gene brinker reveals a novel mechanism of Dpp target gene regulation. *Cell* **96**: 563-73.
- Jones C. A., Ng J., Peterson A. J., Morgan K., Simon J., and Jones R. S. (1998). The Drosophila esc and E(z) proteins are direct partners in polycomb group-mediated repression. *Mol Cell Biol* **18**: 2825-34.
- Jones R. S., and Gelbart W. M. (1990). Genetic analysis of the enhancer of zeste locus and its role in gene regulation in Drosophila melanogaster. *Genetics* **126**: 185-99.
- Kehle J., Beuchle D., Treuheit S., Christen B., Kennison J. A., Bienz M., and Muller J. (1998). dMi-2, a hunchback-interacting protein that functions in polycomb repression. *Science* **282**: 1897-900.
- Kennison J. A. (1995). The Polycomb and trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. *Annu Rev Genet* **29**: 289-303.
- Kleer C. G., Cao Q., Varambally S., Shen R., Ota I., Tomlins S. A., Ghosh D., Sewalt R. G., Otte A. P., Hayes D. F., Sabel M. S., Livant D., Weiss S. J., Rubin M. A., and Chinnaiyan A. M. (2003). EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* **100**: 11606-11.
- Kim J., Johnson K., Chen H. J., Carroll S., and Laughon A. (1997). Drosophila Mad binds to DNA and directly mediates activation of vestigial by Decapentaplegic. *Nature* **388**: 304-8.
- Kroon E., Kros J., Thorsteinsdottir U., Baban S., Buchberg A. M., and Sauvageau G. (1998). Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *Embo J* **17**: 3714-25.
- Kyba M., and Brock H. W. (1998). The Drosophila polycomb group protein Psc contacts ph and Pc through specific conserved domains. *Mol Cell Biol* **18**: 2712-20.

- Kurisaki K., Kurisaki, A. Valcourt, U., Terentiev A.A., Pardali K., Ten Dijke P., Heldin C.H., Ericsson J. and Moustakas A (2003). Nuclear factor YY1 inhibits transforming growth factor beta- and bone morphogenetic protein-induced cell differentiation. *Mol Cell Biol* **23**: 4494-510.
- Kurokawa M., Mitani K., Irie K., Matsuyama T., Takahashi T., Chiba S., Yazaki Y., Matsumoto K., and Hirai H. (1998). The oncoprotein Evi-1 represses TGF-beta signalling by inhibiting Smad3. *Nature* **394**: 92-6.
- Laible G., Wolf A., Dorn R., Reuter G., Nislow C., Lebersorger A., Popkin D., Pillus L., and Jenuwein T. (1997). Mammalian homologues of the Polycomb-group gene Enhancer of zeste mediate gene silencing in *Drosophila* heterochromatin and at *S. cerevisiae* telomeres. *Embo J* **16**: 3219-32.
- LaJeunesse D., and Shearn A. (1996). E(z): a polycomb group gene or a trithorax group gene? *Development* **122**: 2189-97.
- Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J., Devon K., Dewar K., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.
- Larsson J., Zhang J., and Rasmuson-Lestander A. (1996). Mutations in the *Drosophila melanogaster* gene encoding S-adenosylmethionine synthetase [corrected] suppress position-effect variegation. *Genetics* **143**: 887-96.
- Lessard J., Baban S., and Sauvageau G. (1998). Stage-specific expression of polycomb group genes in human bone marrow cells. *Blood* **91**: 1216-24.
- Lessard J., Schumacher A., Thorsteinsdottir U., van Lohuizen M., Magnuson T., and Sauvageau G. (1999). Functional antagonism of the Polycomb-Group genes *eed* and *Bmi1* in hemopoietic cell proliferation. *Genes Dev* **13**: 2691-703.
- Liu F., Hata A., Baker J. C., Doody J., Carcamo J., Harland R. M., and Massague J. (1996). A human Mad protein acting as a BMP-regulated transcriptional activator. *Nature* **381**: 620-3.
- Locke J., Kotarski M. A., and Tartof K. D. (1988). Dosage-dependent modifiers of position effect variegation in *Drosophila* and a mass action model that explains their effect. *Genetics* **120**: 181-98.

- Luo K., Stroschein S. L., Wang W., Chen D., Martens E., Zhou S., and Zhou Q. (1999). The Ski oncoprotein interacts with the Smad proteins to repress TGFbeta signaling. *Genes Dev* **13**: 2196-206.
- Massague J. (1998). TGF-beta signal transduction. *Annu Rev Biochem* **67**: 753-91.
- Massague J., and Wotton D. (2000). Transcriptional control by the TGF-beta/Smad signaling system. *Embo J* **19**: 1745-54.
- McCall K., and Bender W. (1996). Probes of chromatin accessibility in the Drosophila bithorax complex respond differently to Polycomb-mediated repression. *Embo J* **15**: 569-80.
- Milne T. A., Sinclair D. A., and Brock H. W. (1999). The Additional sex combs gene of Drosophila is required for activation and repression of homeotic loci, and interacts specifically with Polycomb and super sex combs. *Mol Gen Genet* **261**: 753-61.
- Minami M., Kinoshita N., Kamoshida Y., Tanimoto H., and Tabata T. (1999). brinker is a target of Dpp in Drosophila that negatively regulates Dpp-dependent genes. *Nature* **398**: 242-6.
- Moustakas A., Souchelnytskyi S., and Heldin C. H. (2001). Smad regulation in TGF-beta signal transduction. *J Cell Sci* **114**: 4359-69.
- Muller J. (1995). Transcriptional silencing by the Polycomb protein in Drosophila embryos. *Embo J* **14**: 1209-20.
- Muller J., Gaunt S., and Lawrence P. A. (1995). Function of the Polycomb protein is conserved in mice and flies. *Development* **121**: 2847-52.
- Nakashima K., Yanagisawa M., Arakawa H., Kimura N., Hisatsune T., Kawabata M., Miyazono K., and Taga T. (1999). Synergistic signaling in fetal brain by STAT3-Smad1 complex bridged by p300. *Science* **284**: 479-82.
- Ng J., Hart C. M., Morgan K., and Simon J. A. (2000). A Drosophila ESC-E(Z) protein complex is distinct from other polycomb group complexes and contains covalently modified ESC. *Mol Cell Biol* **20**: 3069-78.
- Oft M., Peli J., Rudaz C., Schwarz H., Beug H., and Reichmann E. (1996). TGF-beta1 and Ha-Ras collaborate in modulating the phenotypic plasticity and invasiveness of epithelial tumor cells. *Genes Dev* **10**: 2462-77.

- Ohta H., Tokimasa S., Zou Z., Funaki S., Kurahashi H., Takahashi Y., Kimura M., Matsuoka R., Horie M., Hara J., Shimada K., and Takihara Y. (2000). Structure and chromosomal localization of the RAE28/HPH1 gene, a human homologue of the polyhomeotic gene. *DNA Seq* **11**: 61-73.
- Orlando V., Jane E. P., Chinwalla V., Harte P. J., and Paro R. (1998). Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *Embo J* **17**: 5141-50.
- Orlando V., and Paro R. (1993). Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin. *Cell* **75**: 1187-98.
- Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., et al. (2004). Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet* **36**: 40-5.
- Padgett R. W., Wozney J. M., and Gelbart W. M. (1993). Human BMP sequences can confer normal dorsal-ventral patterning in the *Drosophila* embryo. *Proc Natl Acad Sci U S A* **90**: 2905-9.
- Paro R. (1990). Imprinting a determined state into the chromatin of *Drosophila*. *Trends Genet* **6**: 416-21.
- Paro R., and Hogness D. S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proc Natl Acad Sci U S A* **88**: 263-7.
- Peschon J. J., Morrissey P. J., Grabstein K. H., Ramsdell F. J., Maraskovsky E., Gliniak B. C., Park L. S., Ziegler S. F., Williams D. E., Ware C. B., and et al. (1994). Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice. *J Exp Med* **180**: 1955-60.
- Peytavi R., Hong S. S., Gay B., d'Angeac A. D., Selig L., Benichou S., Benarous R., and Boulanger P. (1999). HEED, the product of the human homolog of the murine eed gene, binds to the matrix protein of HIV-1. *J Biol Chem* **274**: 1635-45.
- Pikaart M. J., Recillas-Targa F., and Felsenfeld G. (1998). Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev* **12**: 2852-62.

- Pirrotta V. (1997). PcG complexes and chromatin silencing. *Curr Opin Genet Dev* **7**: 249-58.
- Pirrotta V. (1998). Polycomb the genome: PcG, trxG, and chromatin silencing. *Cell* **93**: 333-6.
- Pirrotta V., and Rastelli L. (1994). White gene expression, repressive chromatin domains and homeotic gene regulation in *Drosophila*. *Bioessays* **16**: 549-56.
- Poux S., Kostic C., and Pirrotta V. (1996). Hunchback-independent silencing of late Ubx enhancers by a Polycomb Group Response Element. *Embo J* **15**: 4713-22.
- Poux S., McCabe D., and Pirrotta V. (2001a). Recruitment of components of Polycomb Group chromatin complexes in *Drosophila*. *Development* **128**: 75-85.
- Poux S., Melfi R., and Pirrotta V. (2001b). Establishment of Polycomb silencing requires a transient interaction between PC and ESC. *Genes Dev* **15**: 2509-14.
- Raaphorst F. M., Otte A. P., van Kemenade F. J., Blokzijl T., Fieret E., Hamer K. M., Satijn D. P., and Meijer C. J. (2001). Distinct BMI-1 and EZH2 expression patterns in thymocytes and mature T cells suggest a role for Polycomb genes in human T cell differentiation. *J Immunol* **166**: 5925-34.
- Raaphorst F. M., van Kemenade F. J., Blokzijl T., Fieret E., Hamer K. M., Satijn D. P., Otte A. P., and Meijer C. J. (2000a). Coexpression of BMI-1 and EZH2 polycomb group genes in Reed-Sternberg cells of Hodgkin's disease. *Am J Pathol* **157**: 709-15.
- Raaphorst F. M., van Kemenade F. J., Fieret E., Hamer K. M., Satijn D. P., Otte A. P., and Meijer C. J. (2000b). Cutting edge: polycomb gene expression patterns reflect distinct B cell differentiation stages in human germinal centers. *J Immunol* **164**: 1-4.
- Remacle J. E., Kraft H., Lerchner W., Wuytens G., Collart C., Verschueren K., Smith J. C., and Huylebroeck D. (1999). New mode of DNA binding of multi-zinc finger transcription factors: deltaEF1 family members bind with two hands to two target sites. *Embo J* **18**: 5073-84.

- Reuter G., and Spierer P. (1992). Position effect variegation and chromatin proteins. *Bioessays* **14**: 605-12.
- Ringrose L., and Paro R. (2001). Remembering silence. *Bioessays* **23**: 566-70.
- Rozovskaia T., Tillib S., Smith S., Sedkov Y., Rozenblatt-Rosen O., Petruk S., Yano T., Nakamura T., Ben-Simchon L., Gildea J., Croce C. M., Shearn A., Canaani E., and Mazo A. (1999). Trithorax and ASH1 interact directly and associate with the trithorax group-responsive bxd region of the Ultrabithorax promoter. *Mol Cell Biol* **19**: 6441-7.
- Saha V., Young B. D., and Freemont P. S. (1998). Translocations, fusion genes, and acute leukemia. *J Cell Biochem Suppl* **30-31**: 264-76.
- Sambrook J. R., David W (2001). "Molecular cloning: a laboratory manual," Cold Spring Harbor Laboratory Press.
- Sass G. L., and Henikoff S. (1998). Comparative analysis of position-effect variegation mutations in *Drosophila melanogaster* delineates the targets of modifiers. *Genetics* **148**: 733-41.
- Satijn D. P., and Otte A. P. (1999). Polycomb group protein complexes: do different complexes regulate distinct target genes? *Biochim Biophys Acta* **1447**: 1-16.
- Saurin A. J., Shiels C., Williamson J., Satijn D. P., Otte A. P., Sheer D., and Freemont P. S. (1998). The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. *J Cell Biol* **142**: 887-98.
- Sauvageau G., Lansdorp P. M., Eaves C. J., Hogge D. E., Dragowska W. H., Reid D. S., Largman C., Lawrence H. J., and Humphries R. K. (1994). Differential expression of homeobox genes in functionally distinct CD34+ subpopulations of human bone marrow cells. *Proc Natl Acad Sci U S A* **91**: 12223-7.
- Sauvageau G., Thorsteinsdottir U., Hough M. R., Hugo P., Lawrence H. J., Largman C., and Humphries R. K. (1997). Overexpression of HOXB3 in hematopoietic cells causes defective lymphoid development and progressive myeloproliferation. *Immunity* **6**: 13-22.

- Shi X., Yang X., Chen D., Chang Z., and Cao X. (1999). Smad1 interacts with homeobox DNA-binding proteins in bone morphogenetic protein signaling. *J Biol Chem* **274**: 13711-7.
- Shi Y. (2001). Structural insights on Smad function in TGFbeta signaling. *Bioessays* **23**: 223-32.
- Shi Y., Wang Y. F., Jayaraman L., Yang H., Massague J., and Pavletich N. P. (1998). Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. *Cell* **94**: 585-94.
- Sewalt R. G., van der Vlag J., Gunster M. J., Hamer K. M., den Blaauwen J. L., Satijn D. P., Hendrix T., van Driel R., and Otte A. P. (1998). Characterization of interactions between the mammalian polycomb-group proteins Enx1/EZH2 and EED suggests the existence of different mammalian polycomb-group protein complexes. *Mol Cell Biol* **18**: 3586-95.
- Schlossherr J., Eggert H., Paro R., Cremer S., and Jack R. S. (1994). Gene inactivation in *Drosophila* mediated by the Polycomb gene product or by position-effect variegation does not involve major changes in the accessibility of the chromatin fibre. *Mol Gen Genet* **243**: 453-62.
- Schumacher A., Faust C., and Magnuson T. (1996). Positional cloning of a global regulator of anterior-posterior patterning in mice. *Nature* **384**: 648.
- Shao Z., Raible F., Mollaaghababa R., Guyon J. R., Wu C. T., Bender W., and Kingston R. E. (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**: 37-46.
- Shimono Y., Murakami H., Hasegawa Y., and Takahashi M. (2000). RET finger protein is a transcriptional repressor and interacts with enhancer of polycomb that has dual transcriptional functions. *J Biol Chem* **275**: 39411-9.
- Simon J. (1995). Locking in stable states of gene expression: transcriptional control during *Drosophila* development. *Curr Opin Cell Biol* **7**: 376-85.
- Sinclair D. A., Campbell R. B., Nicholls F., Slade E., and Brock H. W. (1992). Genetic analysis of the additional sex combs locus of *Drosophila melanogaster*. *Genetics* **130**: 817-25.

- Sinclair D. A., Clegg N. J., Antonchuk J., Milne T. A., Stankunas K., Ruse C., Grigliatti T. A., Kassis J. A., and Brock H. W. (1998a). Enhancer of Polycomb is a suppressor of position-effect variegation in *Drosophila melanogaster*. *Genetics* **148**: 211-20.
- Sinclair D. A., Milne T. A., Hodgson J. W., Shellard J., Salinas C. A., Kyba M., Randazzo F., and Brock H. W. (1998b). The Additional sex combs gene of *Drosophila* encodes a chromatin protein that binds to shared and unique Polycomb group sites on polytene chromosomes. *Development* **125**: 1207-16.
- Soto M. C., Chou T. B., and Bender W. (1995). Comparison of germline mosaics of genes in the Polycomb group of *Drosophila melanogaster*. *Genetics* **140**: 231-43.
- Stankunas K., Berger J., Ruse C., Sinclair D. A., Randazzo F., and Brock H. W. (1998). The enhancer of polycomb gene of *Drosophila* encodes a chromatin protein conserved in yeast and mammals. *Development* **125**: 4055-66.
- Strausberg R. L., Feingold E. A., Grouse L. H., Derge J. G., Klausner R. D., Collins F. S., Wagner L., et al. (2002). Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci U S A* **99**: 16899-903.
- Strausberg R. L., Greenhut S. F., Grouse L. H., Schaefer C. F., and Buetow K. H. (2001). In silico analysis of cancer through the Cancer Genome Anatomy Project. *Trends Cell Biol* **11**: S66-71.
- Stroschein S. L., Wang W., and Luo K. (1999). Cooperative binding of Smad proteins to two adjacent DNA elements in the plasminogen activator inhibitor-1 promoter mediates transforming growth factor beta-induced smad-dependent transcriptional activation. *J Biol Chem* **274**: 9431-41.
- Struhl G., and Akam M. (1985). Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of *Drosophila*. *Embo J* **4**: 3259-64.
- Strutt H., Cavalli G., and Paro R. (1997). Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *Embo J* **16**: 3621-32.

- Strutt H., and Paro R. (1997). The polycomb group protein complex of *Drosophila melanogaster* has different compositions at different target genes. *Mol Cell Biol* **17**: 6773-83.
- Sun Y., Liu X., Eaton E. N., Lane W. S., Lodish H. F., and Weinberg R. A. (1999a). Interaction of the Ski oncoprotein with Smad3 regulates TGF-beta signaling. *Mol Cell* **4**: 499-509.
- Sun Y., Liu X., Ng-Eaton E., Lodish H. F., and Weinberg R. A. (1999b). SnoN and Ski protooncoproteins are rapidly degraded in response to transforming growth factor beta signaling. *Proc Natl Acad Sci U S A* **96**: 12442-7.
- Takahara Y., Tomotsune D., Shirai M., Katoh-Fukui Y., Nishii K., Motaleb M. A., Nomura M., Tsuchiya R., Fujita Y., Shibata Y., Higashinakagawa T., and Shimada K. (1997). Targeted disruption of the mouse homologue of the *Drosophila* polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects. *Development* **124**: 3673-82.
- Tetsu O., Ishihara H., Kanno R., Kamiyasu M., Inoue H., Tokuhisa T., Taniguchi M., and Kanno M. (1998). mel-18 negatively regulates cell cycle progression upon B cell antigen receptor stimulation through a cascade leading to c-myc/cdc25. *Immunity* **9**: 439-48.
- Thorsteinsdottir U., Sauvageau G., Hough M. R., Dragowska W., Lansdorp P. M., Lawrence H. J., Largman C., and Humphries R. K. (1997). Overexpression of HOXA10 in murine hematopoietic cells perturbs both myeloid and lymphoid differentiation and leads to acute myeloid leukemia. *Mol Cell Biol* **17**: 495-505.
- Tie F., Furuyama T., Prasad-Sinha J., Jane E., and Harte P. J. (2001). The *Drosophila* Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development* **128**: 275-86.
- Tkachuk D. C., Kohler S., and Cleary M. L. (1992). Involvement of a homolog of *Drosophila* trithorax by 11q23 chromosomal translocations in acute leukemias. *Cell* **71**: 691-700.
- Tokimasa S., Ohta H., Sawada A., Matsuda Y., Kim J. Y., Nishiguchi S., Hara J., and Takihara Y. (2001). Lack of the Polycomb-group gene rae28

- causes maturation arrest at the early B-cell developmental stage. *Exp Hematol* **29**: 93-103.
- Treisman J. (2001). Drosophila homologues of the transcriptional coactivation complex subunits TRAP240 and TRAP230 are required for identical processes in eye-antennal disc development. *Development* **128**: 603-15.
- Tschiersch B., Hofmann A., Krauss V., Dorn R., Korge G., and Reuter G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *Embo J* **13**: 3822-31.
- Tumbar T., Sudlow G., and Belmont A. S. (1999). Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *J Cell Biol* **145**: 1341-54.
- van der Lugt N. M., Alkema M., Berns A., and Deschamps J. (1996). The Polycomb-group homolog Bmi-1 is a regulator of murine Hox gene expression. *Mech Dev* **58**: 153-64.
- van der Lugt N. M., Domen J., Linders K., van Roon M., Robanus-Maandag E., te Riele H., van der Valk M., Deschamps J., Sofroniew M., van Lohuizen M., and et al. (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes Dev* **8**: 757-69.
- van der Vlag J., and Otte A. P. (1999). Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nat Genet* **23**: 474-8.
- van Kemenade F. J., Raaphorst F. M., Blokzijl T., Fieret E., Hamer K. M., Satijn D. P., Otte A. P., and Meijer C. J. (2001). Coexpression of BMI-1 and EZH2 polycomb-group proteins is associated with cycling cells and degree of malignancy in B-cell non-Hodgkin lymphoma. *Blood* **97**: 3896-901.
- van Lohuizen M. (1999). The trithorax-group and polycomb-group chromatin modifiers: implications for disease. *Curr Opin Genet Dev* **9**: 355-61.
- van Lohuizen M., Tijms M., Voncken J. W., Schumacher A., Magnuson T., and Wientjens E. (1998). Interaction of mouse polycomb-group (Pc-G)

proteins Enx1 and Enx2 with Eed: indication for separate Pc-G complexes. *Mol Cell Biol* **18**: 3572-9.

Varambally S., Dhanasekaran S. M., Zhou M., Barrette T. R., Kumar-Sinha C., Sanda M. G., Ghosh D., Pienta K. J., Sewalt R. G., Otte A. P., Rubin M. A., and Chinnaiyan A. M. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**: 624-9.

Verschueren K., Remacle J. E., Collart C., Kraft H., Baker B. S., Tylzanowski P., Nelles L., Wuytens G., Su M. T., Bodmer R., Smith J. C., and Huylebroeck D. (1999). SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. *J Biol Chem* **274**: 20489-98.

Visser H. P., Gunster M. J., Kluin-Nelemans H. C., Manders E. M., Raaphorst F. M., Meijer C. J., Willemze R., and Otte A. P. (2001). The Polycomb group protein EZH2 is upregulated in proliferating, cultured human mantle cell lymphoma. *Br J Haematol* **112**: 950-8.

von Freeden-Jeffry U., Vieira P., Lucian L. A., McNeil T., Burdach S. E., and Murray R. (1995). Lymphopenia in interleukin (IL)-7 gene-deleted mice identifies IL-7 as a nonredundant cytokine. *J Exp Med* **181**: 1519-26.

Voncken J. W., Schweizer D., Aagaard L., Sattler L., Jantsch M. F., and van Lohuizen M. (1999). Chromatin-association of the Polycomb group protein BMI1 is cell cycle-regulated and correlates with its phosphorylation status. *J Cell Sci* **112 (Pt 24)**: 4627-39.

Vonlanthen S., Heighway J., Altermatt H. J., Gugger M., Kappeler A., Borner M. M., van Lohuizen M., and Betticher D. C. (2001). The bmi-1 oncoprotein is differentially expressed in non-small cell lung cancer and correlates with INK4A-ARF locus expression. *Br J Cancer* **84**: 1372-6.

Wade P. A., Jones P. L., Vermaak D., and Wolffe A. P. (1998). A multiple subunit Mi-2 histone deacetylase from *Xenopus laevis* cofractionates with an associated Snf2 superfamily ATPase. *Curr Biol* **8**: 843-6.

Watanabe M., and Whitman M. (1999). FAST-1 is a key maternal effector of mesoderm inducers in the early *Xenopus* embryo. *Development* **126**: 5621-34.

- Wismar J., Loffler T., Habtemichael N., Vef O., Geissen M., Zirwes R., Altmeyer W., Sass H., and Gateff E. (1995). The *Drosophila melanogaster* tumor suppressor gene lethal(3)malignant brain tumor encodes a proline-rich protein with a novel zinc finger. *Mech Dev* **53**: 141-54.
- Wotton D., Lo R. S., Lee S., and Massague J. (1999a). A Smad transcriptional corepressor. *Cell* **97**: 29-39.
- Wotton D., Lo R. S., Swaby L. A., and Massague J. (1999b). Multiple modes of repression by the Smad transcriptional corepressor TGIF. *J Biol Chem* **274**: 37105-10.
- Wu C. T., and Howe M. (1995). A genetic analysis of the Suppressor 2 of zeste complex of *Drosophila melanogaster*. *Genetics* **140**: 139-81.
- Yamamoto Y., Girard F., Bello B., Affolter M., and Gehring W. J. (1997). The cramped gene of *Drosophila* is a member of the Polycomb-group, and interacts with mus209, the gene encoding Proliferating Cell Nuclear Antigen. *Development* **124**: 3385-94.
- Zawel L., Dai J. L., Buckhaults P., Zhou S., Kinzler K. W., Vogelstein B., and Kern S. E. (1998). Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol Cell* **1**: 611-7.
- Zhang Y., LeRoy G., Seelig H. P., Lane W. S., and Reinberg D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95**: 279-89.
- Zhou J., Ashe H., Burks C., and Levine M. (1999). Characterization of the transvection mediating region of the abdominal-B locus in *Drosophila*. *Development* **126**: 3057-65.