Lúcia Maria Armelin-Correa

Estudo da Região Promotora
do Gene do
Colágeno XVIII Humano

Instituto de Biociências

Universidade de São Paulo

São Paulo

2007

Lúcia Maria Armelin-Correa

Estudo da Região Promotora
do Gene do
Colágeno XVIII Humano

Tese apresentada ao Instituto de
Biociências da Universidade de
São Paulo, para a obtenção de
Título de Doutor em Ciências, na
Área de Genética.

Orientadora: Prof^a. Dr^a. M^a. Rita
Passos-Bueno

São Paulo

2007

Ficha Catalográfica

Armelin-Correa, Lúcia Maria
Estudo da Região Promotora do
gene do Colágeno XVIII humano
Número de páginas

Tese de Doutorado - Instituto de
Biociências da Universidade de São
Paulo. Departamento de Genética e
Biologia Evolutiva.

1. Colágeno XVIII 2. Região
Promotora 3. Fatores de Transcrição
I. Universidade de São Paulo. Instituto
de Biociências. Departamento de
Genética e Biologia Evolutiva.

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Orientadora Prof^a Dr^a M^a Rita Passos Bueno

Dedicatória

Ao meu amado pai,
meu maior exemplo
de amor a ciência.

Sonhar é fazer planos.

Viver é ter coragem de realizá-los.

Felipe Veullieme

Agradecimentos

A Deus por esta vida repleta de sonhos que aos poucos vão se tornando realidade.

A minha mãe, Lúcia, e mais uma vez ao meu pai, Egas, por serem meu porto seguro e fonte de amor incondicional. Aos meus queridos irmãos, Paula e Júnior, pelos muito encorajamentos e carinho constantes.

Ao meu marido, André, luz que entrou na minha vida e me ensinou a lutar pelos meus sonhos, estando sempre ao meu lado nos piores e melhores momentos dessa luta.

A toda a família Correa, Seu Eric, Dona Teresa, Juliana, Érica e Fábio, que me adotaram com todo amor.

A minha orientadora, a Dra Maria Rita Passos Bueno pelos ensinamentos, a confiança e a oportunidade de tornar esse sonho realidade. Por ter sempre acreditado em mim e, desde o primeiro dia, ter me acolhido, apoiado, orientado e mantido todas as portas abertas, sendo um exemplo de pesquisadora, mulher e amiga. Obrigada por TUDO!

A todos os meus professores ao longo dessa longa formação da graduação até a pós-graduação, que alimentaram e cultivaram meu amor pela Biologia.

A Cibele e Daniela amigas queridas sempre ao meu lado nesses longos anos de bancada.

A todos amigos do lab Rita, Carlos, Oscar, Flávia Errera, Guilherme, Roberto, Deinha, Fernanda, Karina, Erika K, Erika Y, Gustavo, Camila, Gerson, Kelly, Roseli, Fabiana, Nélío, Cristian, Daniel, Aline, Bia, pelo companheirismo, carinho e amizade.

A todo o pessoal do Genoma e do laboratório da Prof^a Mayana Zatz e da Prof^a Mariz. A Constância pela paciência, ajuda e carinho.

A todo o pessoal desde a época da iniciação científica por me darem a mão e por me ensinarem a andar, em especial a Toninha, Martinha, Manuela, Elisângela, Andrea Bernardino, Flávia de Paula, Agnes, Kikue, Dulci, Kelly, Ale Splendore, Eloísa, Dinamar.

Ao Dr. Chin Jia Lin e Dr^a Ângela Barbosa pelos preciosos ensinamentos, dedicação e principalmente pela amizade.

A Dr^a Mari Sogayar e Dr^a Sheila Winnischofer pelos ensinamentos e ajuda no início dos estudos funcionais.

Ao Dr. Venâncio Alves, Dr. Flair Carilho e Dr. Eymard por acreditarem no meu projeto e permitirem meu acesso a todo material e dados de pacientes.

A todos os pacientes sem os quais este trabalho não teria significado.

Ao Dr. Orlando Musso pela oportunidade de trabalhar no seu laboratório e a toda a equipe do laboratório INSERM U620 pela recepção maravilhosa e pela amizade.

A todos meus amigos que acompanharam a minha trajetória, os meus dramas, os meus fracassos e meus sucessos. Obrigada por estarem sempre ao meu lado, me apoiarem e encorajarem!

Lista de siglas e abreviaturas utilizadas:

C/EBP - CCAAT/Enhancer-Binding Protein

CHIP - Chromatin Immunoprecipitation

COL18A1 – gene da cadeia alpha 1 do colágeno XVIII humano

Col18a1- gene da cadeia alpha 1 do colágeno XVIII murino

EMSA – Eletrophoretic Mobility Shift Assay

HCC – Hepatocellular Carcinoma,

HNF3 – Hepatocyte Nuclear Factor 3

Oct-1 – Octamer binding transcription factor 1

PCR – Polymerase Chain Reaction

qPCR – quantitative Polymerase Chain Reaction

SNP – Single Nucleotide Polymorphism

SNUPE – Single Nucleotide Polymorphism Extension

Sp – Specificity proteins

YY1 - Yin and Yang 1

Sumário

Resumo	1
Abstract	2
Capítulo I. Introdução	3
Capítulo II. Casuística e Métodos	29
Capítulo III. <i>Characterization of human Collagen XVIII promoter 2: interaction of Sp1, Sp3 and YY1 with the regulatory region and a SNP that increases transcription in hepatocytes</i>	49
Capítulo IV. <i>Further characterization of COL18A1 promoter 2: Binding of Oct-1 and effect of the SNP-700T/G in the transcription levels of collagen XVIII and predisposition to HCC</i>	85
Capítulo V. <i>C/EBPα and C/EBPβ expressions correlate with that of NC11-493 in HCC and cirrhotic liver, and C/EBPβ upregulates COL18A1 transcription in vitro</i>	113
Capítulo VI. Discussão e Conclusão	143
Capítulo VII. Referências bibliográficas	148
Anexos	

O colágeno XVIII é um componente das membranas basais com diversos domínios funcionais, como a endostatina e o domínio frizzled, que têm importante papel em processos celulares como proliferação e diferenciação. *COL18A1* possui dois promotores alternativos: o promotor 1, que regula a síntese da variante NC11-303, e o promotor 2 responsável pelas variantes NC11-728 e NC11-493, expressas por hepatócitos. Existe uma variação interindividual da endostatina circulante e da expressão do colágeno XVIII no fígado. A expressão do colágeno XVIII/endostatina foi correlacionada com a progressão tanto do hepatocarcinoma (HCC), quanto da fibrose/cirrose hepática. Elucidar a regulação da expressão de *COL18A1* pode auxiliar na compreensão dessa variação interindividual e da progressão dessas doenças.

Neste trabalho demos início a caracterização do promotor 2 do *COL18A1*. Identificamos na seqüência predita como promotora cinco regiões conservadas entre humanos e camundongos. A análise *in silico* e funcional dessas regiões revelou que os fatores de transcrição, Sp1, Sp3, YY1, Oct-1, C/EBP α e C/EBP β , interagem com as mesmas. Demonstramos que C/EBP β aumenta a taxa de transcrição do promotor 2 em hepatócitos, e que existe uma correlação positiva da expressão de NC11-493 com C/EBP α e C/EBP β em tecido hepático cirrótico e tumoral. As expressões de C/EBP α em tecido hepático cirrótico e tumoral estão diretamente correlacionadas, enquanto que os níveis de NC11-493 nos tumores estão inversamente correlacionados com o tamanho dos mesmos. Mostramos a existência de diversos SNPs no promotor 2. O SNP-700T/G, funcional *in vitro*, afeta a interação de Sp3 e YY1 com essa região regulatória. A deleção da região do SNP indicou que ela possui elementos importantes para a transcrição em hepatócitos, apesar deste SNP não estar relacionado com o nível de expressão do colágeno XVIII em fígado fibrótico ou com susceptibilidade a HCC. O SNP-700T/G está em desequilíbrio de ligação com o SNPc.1135C/T, no domínio frizzled do colágeno XVIII. Não foi possível elucidar a funcionalidade do SNPs c.1135C/T *in vitro*, mas os haplótipos formados por esses dois SNPs têm diferentes frequências entre descendentes de europeus e de africanos.

Nosso trabalho traz importantes contribuições e abre novas perspectivas para a compreensão da regulação do colágeno XVIII em fígado humano, tanto em situações fisiológicas, quanto em processos fibrogênicos e tumorigênicos.

Collagen XVIII is a basal membrane component with several functional domains, such as endostatin and frizzled domains, which have important roles in cellular processes such as proliferation and differentiation. *COL18A1* has two promoter regions: promoter 1, that regulates the synthesis of NC11-303 isoform, and promoter 2, localized in intron 2, responsible for NC11-728 and NC11-493 isoforms expressed by hepatocytes. There is a large interindividual variation in circulating endostatin and in collagen XVIII liver expression. Collagen XVIII/endostatin levels were correlated with hepatocellular carcinoma (HCC) progression, as well as liver fibrosis/cirrhosis, conditions that precede HCC. Elucidating the mechanisms that regulate *COL18A1* expression in hepatocytes may help understanding its variation among individuals and liver disease stages, as well as contribute to new treatment strategies.

In this work we began to characterize *COL18A1* promoter region 2. We identified in the predicted promoter sequence five conserved regions between human and mouse. The *in silico* and functional analysis of these regions revealed that transcription factors Sp1, Sp3, YY1, Oct-1, C/EBP α and C/EBP β interact with them. We have demonstrated that C/EBP β increases promoter 2 transcription rate in hepatocytes, and that there is a positive correlation of NC11-493 expression with that of C/EBP α and C/EBP β in cirrhotic and tumor liver samples. Non-tumor and tumor C/EBP α expressions positively correlate between themselves, while NC11-493 tumor expression inversely correlates with tumor size. We also showed that there are several SNPs in *COL18A1* promoter 2 region. SNP-700T/G, functional *in vitro*, affects Sp3 and YY1 interaction with the promoter 2 region and deletion of the SNP region indicated that this sequence has important hepatocyte regulatory elements. Our results suggest that this SNP does not significantly affect *COL18A1* expression in fibrotic/cirrhotic liver and is not associated with HCC susceptibility. SNP-700T/G is in linkage disequilibrium with SNPC.1135C/T, at collagen XVIII frizzled domain. We could not elucidate SNPC.1135C/T functionality *in vitro*, but the haplotypes formed by these two SNPs have different frequencies in European and African descendants.

In conclusion, our work brings important contributions and opens new perspectives for the comprehension of collagen XVIII regulation in human liver in physiological situations, as well as in fibrotic/cirrhotic and tumorigenic process.

I.INTRODUÇÃO

I.1. COLÁGENO XVIII

I.1.1 - GENE E PROTEÍNA

Os colágenos constituem uma grande família de proteínas importantes para a estrutura da matriz extracelular. Essas proteínas têm como característica principal repetições Gly-Xaa-Yaa em sua seqüência polipeptídica que assume a conformação de uma cadeia alfa. Nessa cadeia alfa, a glicina, que possui apenas um átomo de hidrogênio em sua cadeia lateral, tem papel essencial pois, graças ao seu pequeno tamanho molecular, permite o entrelaçamento de três desses polipeptídeos e a formação de homo ou heterotrímeros. Geralmente os aminoácidos Xaa e Yaa são, respectivamente, a prolina e a hidroxiprolina que com suas estruturas em anel estabilizam a conformação da tripla hélice (Ricard-Blum & Ruggiero, 2005). Já foram descritos mais de 25 tipos de colágeno diferentes e, nos vertebrados, eles podem ser divididos em dois grandes grupos: os fibrilares, responsáveis pela extensibilidade dos tecidos através da formação de fibras ordenadas na matriz extracelular e, os não fibrilares, que mostram heterogeneidade na estrutura, localização, organização macromolecular e função (Ricard-Blum & Ruggiero, 2005).

Em 1994 o gene que codifica o colágeno XVIII humano, *COL18A1* (NCBI GeneID: 80781), foi mapeado no cromossomo 21q22.3 (Oh *et al.*, 1994a). O colágeno XVIII, assim como o colágeno XV, faz parte do subgrupo das MULTIPLEXINAS (*multiple triple-helix domains and interruptions*) dentro do grupo de colágenos não fibrilares. Esse subgrupo é caracterizado por moléculas formadas por diversos domínios colagenosos interrompidos por domínios não colagenosos, com um grande domínio globular N-terminal e outro C-terminal (Oh

et al., 1994b). Essas moléculas apresentam cadeias laterais de condroitin (colágeno XV) ou heparan sulfato (colágeno XVIII) e suas funções ainda não foram completamente elucidadas.

O gene *COL18A1* foi caracterizado perfazendo um total de 105 kb e possuindo 43 *exons*, responsáveis um peptídeo sinal, dez domínios colágeno (COL1-COL10) e onze domínios não colágeno (NC1-NC11), nomeados a partir da região C-terminal (Saarela *et al.*, 1998b). São transcritas três isoformas, que diferem quanto ao último domínio não colágeno na porção N-terminal, com 303, 493 ou 728 aminoácidos, e denominadas isoformas NC11-303, NC11-493 e NC11-728 respectivamente (Saarela *et al.*, 1998b, Elamma *et al.*, 2003). O gene possui dois promotores alternativos, separados por aproximadamente 50 kb. O promotor *upstream*, promotor 1, regula a síntese da variante NC11-303, cujo RNAm possui todos os *exons* menos o *exon* 3, enquanto que o promotor *downstream*, promotor 2, localizado no *intron* 2, é responsável pelas variantes NC11-493 e NC11-728, que não possuem apenas os *exons* 1 e 2, e diferem por um *splice* alternativo do *exon* 3. Graças a esse *splice* alternativo a isoforma NC11-728 é a única isoforma a possuir em seu domínio NC11 uma sequência de 235 aminoácidos semelhantes ao domínio frizzled dos receptores frizzled (Elamma *et al.*, 2003). Os *exons* 4-9 são responsáveis pela síntese de 301 aminoácidos do domínio NC11 comum a todas as isoformas, enquanto que os *exons* 9-43 correspondem aos demais domínios colágeno e não colágeno da proteína, também comuns a todas as isoformas (Elamma *et al.*, 2003) (Fig. 1).

Fig. 1: Diagrama da região N-terminal do gene *COL18A1*, mostrando os dois promotores distintos com seus os inícios de transcrição alternativos e o splice alternativo do *exon 3* inteiro, para síntese de NC11-303, e parcial para síntese de NC11-493 (Suzuki *et al.*, 2002).

As seqüências de aminoácidos do colágeno XVIII de humanos e camundongos apresentam uma identidade de 79% e similaridade de 95%, sendo que a estrutura dos dois genes, e suas respectivas isoformas, são muito semelhantes (Saarela *et al.*, 1998b). O gene *Col18a1*, de camundongos, também possui 2 promotores responsáveis pela transcrição de 3 isoformas: NC11-301, homóloga a isoforma humana NC11-303, NC11-517, homóloga a NC11-493 e, NC11-764, homóloga a NC11-728 (Rhen *et al.*, 1996). O promotor 1 de camundongo é responsável pela transcrição da isoforma NC11-301 e o promotor 2, upstream ao *exon 3* regula as isoformas NC11-517 e NC11-764. Assim, como nos humanos, a isoforma NC11-301 possui 42 exons, excluindo o *exon 3* e, as isoformas NC11-517 e NC11-764 possuem 41 *exons*, excluindo os *exons 1* e 2, sendo que a isoforma NC11-517 sofre um *splice* alternativo de parte do *exon 3* (Rhen *et al.*, 1996).

A maior homologia entre a proteína humana e de camundongo é observada nos últimos 184 aminoácidos, que correspondem a endostatina, um fragmento de 20-kD gerado por clivagem proteolítica da região C-terminal do colágeno XVIII (Saarela *et al.*, 1998b). Diversas proteases como MMP3, MMP7, MMP9, MMP13, MMP20, catepsina L e elastase (Wen *et al.*, 1999, Felbor *et al.*,

2000, Heljasvaara *et al.*, 2005) são responsáveis pela liberação da endostatina na matriz extracelular. A endostatina foi caracterizada como um potente inibidor de angiogênese de tumores sólidos em camundongos (O'Reilly *et al.*, 1997).

O colágeno XVIII, assim como o XV, possui na porção N-terminal um segmento homólogo a região N-terminal da trombospondina-1. Esse motivo é comum em uma série de outros colágenos (Ricard-Blum & Ruggiero, 2005). A trombospondina-1 também funciona como inibidor de angiogênese (Sund *et al.*, 2005), contudo, ainda se desconhece a função dessa seqüência nesses colágenos.

A seqüência Arg-Gly-Asp (RGD) foi identificada no domínio COL3 do colágeno XVIII, tanto de camundongo, quanto de humano (Rhen & Pihlajaniemi, 1994; Saarela *et al.*, 1998b). Essa seqüência é um sítio para ligação de integrinas e importante para adesão celular (Hersel *et al.*, 2003). O colágeno XVIII pode então estar envolvido na transdução de sinais da matriz extracelular para a célula.

A isoforma NC11-728 é a única a apresentar na porção N-terminal um segmento com 10 cisteínas, conservado entre humano, camundongo e sapo, e homólogo a porção extracelular da proteína *frizzled* de *Drosophila* (Rehn *et al.*, 1996; Elaama *et al.*, 2003). Todas as proteínas *frizzled* são caracterizadas por terem um domínio chamado de CRD, *cysteine-rich domain*, que consiste em 10 cisteínas separadas sempre com o mesmo espaçamento, sendo este domínio CRD suficiente para a ligação de proteínas Wnt (Xu & Nusse, 1998). A via de sinalização Wnt tem papel importante na regulação da proliferação, mobilidade, morfogênese e diferenciação celular (Wodarz & Nusse, 1998). Foi demonstrado que o domínio *frizzled* do colágeno XVIII é capaz de modular a sinalização de

beta-catenina em células de câncer de cólon e fígado (Quelard *et al.*, 2006), provavelmente competindo com receptores frizzled por proteínas Wnt da matriz extracelular.

Para exercerem suas funções, esses diversos domínios funcionais do colágeno XVIII necessitam sofrer proteólise e, portanto, são chamados matricriptinas (Ricard-Blum & Ruggiero, 2005). Além desses domínios, o colágeno XVIII possui também cadeias laterais de heparan-sulfato (Fig. 2), onde podem se ligar diversos fatores de crescimento com afinidade a heparina. Assim, a ação de proteases durante processos fisiológicos e patológicos gera a liberação tanto de matricriptinas, quanto de fatores de crescimento na matriz extracelular, fazendo com que o colágeno XVIII tenha um papel importante na modulação de processos como proliferação, mobilidade, morfogênese e diferenciação celular (Iozzo, 2005).

Fig. 2: Diagrama das isoformas do colágeno XVIII, mostrando seus 10 domínios colagenosos e 11 não colagenosos, assim como outros motivos funcionais e suas cadeias laterais de heparan sulfato (Criado por Oscar Suzuki).

I.1.2 – LOCALIZAÇÃO E EXPRESSÃO

Sabe-se que o colágeno XVIII é um componente comum da membrana basal dos tecidos. A isoforma NC11-303 apresenta, em adultos, forte expressão em rim, coração, pâncreas e ovário e baixa expressão em músculo esquelético, placenta, baço, próstata, testículo e intestino (Saarela *et al.*, 1998b). Em um primeiro momento foi observada, em tecidos fetais, a expressão de NC11-303 em pulmão, fígado e rim. Posteriormente, mostrou-se a expressão da isoforma curta também em retina adulta e cérebro fetal humano (Sertié *et al.*, 2000).

A isoforma NC11-493 apresenta uma expressão forte, e quase que exclusiva, em fígado fetal e coração e fígado adultos, com uma fraca expressão na próstata, testículo, ovário e intestino (Saarela *et al.*, 1998b). No fígado normal o colágeno XVIII é secretado principalmente pelos hepatócitos indo se depositar no espaço perisinusoidal, isto é, o espaço encontrado entre os hepatócitos e as células endoteliais dos capilares hepáticos, chamados de sinusóides (Fig. 3) (Saarela *et al.*, 1998a; Musso *et al.*, 1998). Acredita-se que o fígado, por apresentar uma alta expressão de colágeno XVIII e ter o papel de secretar uma série de proteínas plasmáticas, seja o responsável por fragmentos do colágeno XVIII, como a endostatina e o domínio frizzled, circulantes no sangue (Musso *et al.*, 2001a; Elamaa *et al.*, 2003).

Fig. 3: Ilustração mostrando um sinusóide hepático com trabéculas formadas pelos hepatócitos, espaço perisinusoidal com células estreladas, endotélio e lúmen do capilar hepático com células de Kupffer (Bataller e Brenner, 2005).

O RNAm da isoforma NC11-728 mostrou forte expressão no fígado, pulmão, músculo esquelético e baço fetais, assim como em pele adulta (Elaama *et al.*, 2003). Imunohistoquímica de pulmão fetal mostrou a localização de NC11-728 na membrana apical e basal do epitélio de bronquíolos fetais e, na região de ligação do músculo esquelético com o tecido conectivo em miócitos fetais (Elaama *et al.*, 2003). A isoforma NC11-728 expressa-se também no olho (Marneros & Olsen, 2005).

I.1.3 – ENVOLVIMENTO EM PATOLOGIAS

Em 1996, Sertié e col. mapearam em uma grande família brasileira consanguínea o gene que causa a síndrome de Knobloch (SK; MIM 267750) em 21q22.3. Posteriormente, foi encontrada uma mutação no sítio aceitor de *splice* do *intron* 1 do colágeno XVIII nesses pacientes (Sertié *et al.*, 2000). Essa mutação causa o *skipping* do *exon* 2 e leva a um código de parada prematuro no *exon* 4. Portanto, nessa família a mutação afeta apenas a isoforma NC11-303

dos pacientes, que devem continuar expressando normalmente NC11-493 e NC11-728. A SK tem herança autossômica recessiva e é caracterizada pela presença de alta miopia evidenciada nos primeiros anos de vida, degeneração vitreoretiniana com descolamento da retina, anomalias de mácula e encefalocele occipital. Existe uma certa variabilidade clínica, mas a perda da visão e a encefalocele occipital são as características mais marcantes do fenótipo. Alteração na pigmentação da retina e persistência de vascularização fetal nos olhos desses pacientes são características que puderam ser posteriormente reproduzidas nos camundongos *Col18a1* ^{-/-} (Fukai *et al.*, 2002), mostrando que esses animais são um bom modelo para se entender as alterações oculares em SK, assim como, para se estudar o papel do colágeno XVIII nas membranas basais de diversos tecidos. Os pacientes com mutação afetando apenas a isoforma NC11-303 apresentam níveis normais de endostatina circulante, já que a isoforma NC11-493, responsável por sua disponibilidade no sangue, não é afetada pela presença da mutação (Suzuki *et al.*, 2002). Entretanto, pacientes com mutações afetando a síntese de todas as isoformas apresentaram níveis muito baixos de endostatina circulante e perda da visão mais precoce que pacientes somente com NC11-303 afetada (Suzuki *et al.*, 2002). Esses resultados sugerem que a isoforma NC11-728, que se expressa também na retina, desempenha um papel funcional importante e, talvez, diferente do papel de NC11-303 neste tecido.

Zorick e col., 2001, encontraram uma grande variação nos níveis de endostatina circulante em indivíduos normais e um nível sérico de endostatina significativamente mais elevado em portadores de síndrome de Down que em indivíduos controles. Esses dados sugerem que essa alta taxa de endostatina

circulante pode estar relacionada com uma menor incidência de tumores sólidos nestes pacientes. Reforçando esta hipótese, foi observado que camundongos transgênicos expressando 1,6 vezes mais endostatina circulante apresentaram uma menor taxa de crescimento de tumores de melanoma que os camundongos selvagens (Sund *et al.*, 2005).

O colágeno XVIII ganhou especial atenção quando, em 1997, uma função importante da endostatina foi demonstrada por O'Reilly e col. quanto à supressão do crescimento de tumores. Em camundongos que receberam injeção subcutânea desse fragmento, foi verificada a regressão do tumor para lesões microscópicas dormentes, devido a diminuição da angiogênese e aumento da apoptose das células cancerígenas (O'Reilly *et al.*, 1997). Foi demonstrado que a endostatina age nas células endoteliais diminuindo a migração destas células, afetando de forma importante a angiogênese (Kuo *et al.*, 2001). A angiogênese é essencial para o crescimento e metástase do tumor e, ocorre através de uma relação adequada entre estimuladores e inibidores do desenvolvimento de novos capilares (Zetter, 1998). Vários estudos correlacionaram inversamente a taxa de endostatina circulante com grau de vascularização de tumores de tecido sólido (Dhar *et al.*, 2002; Teh *et al.*, 2004). Foi descrito um polimorfismo na posição 1437 da proteína do colágeno XVIII que leva a substituição, no domínio da endostatina, de um ácido aspártico, conservado entre humano e camundongo, por uma asparagina (Iughetti *et al.*, 2001). Estudos de modelagem estrutural mostraram que a troca do ácido aspártico pela asparagina não altera a estabilidade da molécula, mas altera a distribuição de cargas na sua superfície, podendo levar a uma alteração na afinidade da endostatina por outras proteínas e, assim, a uma menor capacidade de inibir a angiogênese (Iughetti *et al.*, 2001).

Apesar de Iughetti e col. terem sugerido uma maior predisposição ao câncer de próstata entre portadores do alelo polimórfico, a associação desse SNP com desenvolvimento de tumores sólidos não foi comprovada (Liu *et al.*, 2003; Nascimento *et al.*, 2004, 2006; Macpherson *et al.*, 2004, Passos-Bueno *et al.*, 2006 Anexo I). Por outro lado, existe um estudo que sugere a associação deste SNP com predisposição a tumor de mama esporádico (Lourenço *et al.*, 2006).

I.1.4 – MODELOS ANIMAIS

Recentemente foi descrita a expressão de genes homólogos ao colágeno XVIII em *Xenopus laevis* e *Caenorhabditis elegans* (Ackley *et al.*, 2001; Elamaa *et al.*, 2002). Em *X. laevis* e *C. elegans* o gene do colágeno XVIII também possui 3 isoformas e a maior homologia com o gene de camundongo, 79% e 55% respectivamente, ocorre na porção C-terminal, no domínio da endostatina (Ackley *et al.*, 2001; Elamaa *et al.*, 2002). Hanai e col., 2002, mostraram, em modelo de embriogênese de *X. laevis*, que a endostatina é capaz de inibir a duplicação de eixos induzidos por β -catenina, suprimir a transcrição de genes regulados por via de sinalização Wnt e, estimular a degradação de formas selvagens e estáveis de β -catenina por uma nova via de degradação, que não a via clássica envolvendo glicogênio sintase quinase (GSK3).

O gene homólogo ao gene do colágeno XVIII em *C. elegans* foi denominado *cle-1*, e a deleção de 231 aminoácidos da porção C-terminal, NC1, da proteína CLE-1, resultou em animais viáveis e férteis, mas com problemas de migração celular e direcionamento de axônios (Ackley *et al.*, 2001). Esses defeitos puderam ser revertidos com a expressão ectópica do domínio NC1, mas

não com expressão ectópica somente do domínio da endostatina. O domínio NC1 consiste em três elementos funcionais importantes: um domínio de associação, uma sequência sensível a proteólise e, o domínio da endostatina (Sasaki *et al.*, 1998) (Fig. 4). Na verdade a expressão ectópica somente do domínio da endostatina em vermes selvagens levou a defeitos similares aos gerados pela deleção de NC1, sugerindo que a endostatina monomérica interfere na função de trímeros de NC1 (Ackley *et al.*, 2001).

Fig. 4: Ilustração mostrando localização do domínio NC1, contendo a endostatina, na porção C-terminal da molécula do colágeno XVIII, com uma ampliação dessa porção evidenciando seus três elementos funcionais importantes: um domínio de trimerização, sítios de proteólise e, o domínio da endostatina (Criado por Oscar Suzuki).

Papéis antagônicos de NC1 e da endostatina também foram descritos para células de vertebrados em cultura, devido ao caráter de oligomerização da endostatina em NC1. Kuo e col., 2001, descrevem que trímeros ou dímeros de endostatina em NC1. Kuo e col., 2001, descrevem que trímeros ou dímeros de endostatina, gerados *in vitro*, induzem motilidade e atividade morfogênica em fibroblastos, células embrionárias renais e células de endotélio e carcinoma pulmonar, e estas alterações são antagonizadas por monômeros da molécula. A atividade do domínio NC1 está localizada, na verdade, no subdomínio da

endostatina e a oligomerização é necessária e suficiente para sua atividade motogênica. Possivelmente, há uma autoregulação negativa na atividade endógena de NC1 através da proteólise e liberação de monômeros de endostatina na matriz extracelular.

Sendo assim, o colágeno XVIII, que juntamente com o colágeno IV, é um dos componentes da membrana basal de vertebrados conservados em *C. elegans* e *Drosophila* (Hynes & Zhao, 2000), é uma molécula particularmente interessante pois o seu domínio NC1/endostatina, altamente conservado evolutivamente, está presente apenas no colágeno XV.

Apesar do grau de conservação evolutivo, e da expressão ubíqua do colágeno XVIII na membrana basal de diversos tecidos, os camundongos *knockout* para todas as isoformas desse gene, *Col18a1*^{-/-}, são saudáveis e férteis, não apresentando alterações nas membranas basais e epiteliais do coração, pulmão, fígado ou rins, talvez devido a existência de mecanismos redundantes que compensem a falta do colágeno XVIII nesses órgãos (Fukai *et al.*, 2002). Entretanto, foi encontrada uma densidade 3 vezes maior de capilares em cultura de explantes de aorta desses camundongos e, também, alterações importantes na regressão de vasos hialóides no desenvolvimento dos olhos desses animais *knockouts* (Fukai *et al.*, 2002). A adição de endostatina nessas culturas levou a uma redução significativa no número de vasos hialóides. A hipótese seria que macrófagos no vítreo do olho em desenvolvimento liberariam metaloproteinases e essas proteases clivariam o colágeno XVIII, liberando os monômeros de endostatina. Esses monômeros levariam à regressão dos vasos hialóides, e induziriam a hipóxia na região, estimulando a expressão do fator de crescimento endotelial vascular (VEGF), que por sua vez levaria a

vascularização normal da retina (Fukai *et al.*, 2002). Apesar da ação da endostatina como inibidor de angiogênese, Fukai e col., 2002, não observaram um maior crescimento de tumores em camundongos *Col18a1* ^{-/-} quando células de melanoma B16F10 ou de fibrosarcoma T241 foram injetadas nestes animais.

Posteriormente, Sund e col., 2005, mostraram que as células B16F10 produzem colágeno XVIII/endostatina endógenos, e selecionaram uma população clonal dessas células que não era capaz de produzir colágeno XVIII. Células dessa população, uma vez injetadas em camundongos *Col18a1* ^{-/-}, produziram um crescimento tumoral significativamente maior que em camundongos selvagens. Esses achados sugerem que a produção de colágeno XVIII/endostatina pelas próprias células tumorais pode compensar a falta dessas moléculas no organismo.

Esses camundongos *Col18a1* ^{-/-}, quando cruzados com camundongos de linhagem C57BL/6J, apresentaram maior incidência de hidrocefalia devido a alterações na membrana basal do epitélio do plexo coróide (Utriainen *et al.*, 2004). Além disso, camundongos com apenas uma cópia ativa do gene do colágeno XVIII, *Col18a1* ^{-/+}, que foram gerados através do cruzamento de camundongos *Col18a1* ^{-/-} com camundongos *Apoe* ^{-/-}, apresentaram maior permeabilidade vascular em vários órgãos, inclusive no fígado, mostrando que a falta desse colágeno implica em alterações em propriedades vasculares e em interações célula-matriz extracelular (Moulton *et al.*, 2004).

I.2. FIBROSE HEPÁTICA E HEPATOCARCINOMA

I.2.1 – ETIOLOGIA E PROGRESSÃO DA DOENÇA

Recentemente, foi verificado que variações na expressão de colágeno XVIII/endostatina estão associadas a quadros de fibrose/cirrose hepática e hepatocarcinoma celular (HCC; MIM 114550) (Schuppan *et al.*, 1998; Musso *et al.*, 1998; Musso *et al.*, 2001a; Musso *et al.*, 2001b, Dhar *et al.*, 2002; Tsuboi *et al.*, 2004; Hu *et al.*, 2005). A fibrose hepática resulta de um dano crônico ao fígado acompanhado de acúmulo de matriz extracelular, que é característica da maioria das doenças crônicas do fígado. Agentes como vírus da hepatite B (HBV), vírus da hepatite C (HCV), álcool (EtOH) e aflatoxina (AFB1) causam a morte dos hepatócitos e disparam um processo inflamatório no fígado, mediado por uma série de citocinas como *transforming growth factor beta 1* (TGF- β 1) e interleucina 6 (IL-6) (Bataller & Brenner, 2006). Danos crônicos ao fígado levam a ativação das células estreladas nos espaços perisinusoidais, que adquirem um fenótipo de miofibroblasto e passam a produzir intensamente colágenos, para remodelar a matriz extracelular em uma tentativa de recuperar o parênquima hepático (Fig. 5).

Fig. 5: Ilustração mostrando um sinusóide hepático durante um processo de fibrose com morte de hepatócitos, espaço perisinusoidal com infiltração de linfócitos, alterações no endotélio e células estreladas ativadas produzindo matriz extracelular (Bataller & Brenner, 2005).

Em estágios avançados da fibrose hepática o fígado contém aproximadamente 6 vezes mais matriz extracelular que o normal. Esse acúmulo de matriz extracelular é resultado de intensa síntese, e reduzida degradação da matriz, com substituição da matriz intersticial, principalmente composta por colágeno IV, por matriz fibrilar, principalmente formada por colágeno I. Essas proteínas de matriz são principalmente produzidas pelas células estreladas ativadas. Há também um intenso processo de angiogênese no local (Rosmorduc *et al.*, 1999). Com a progressão da doença, há a formação de nódulos, isto é, as fibras de colágeno depositadas começam a se encontrar e isolar grupos de hepatócitos comprometendo ainda mais a função hepática. Esses nódulos apresentam muitas vezes lesões pré-cancerosas, que acumulam uma série de alterações genéticas (Röcken & Carl-McGrath, 2001; Feitelson *et al.*, 2002) (Fig. 6). Acredita-se que um dos fatores para a progressão dessas lesões para um tumor maligno seja a ruptura no balanço angiogênico no microambiente onde encontram-se essas lesões (Folkman & Kalluri, 2004). Assim sendo, para o

desenvolvimento e progressão do tumor, é necessário acúmulo de alterações genéticas e epigenéticas em um microambiente favorável, o qual é dependente da constituição genética do indivíduo. Estima-se que cerca de 20% dos pacientes com cirrose hepática vão desenvolver hepatocarcinoma, enquanto o restante não desenvolve tumor. Ainda, a proporção de casos de HCC associados a fígado normal é muito baixa, entre 10-20% do total (Röcken & Carl-McGrath, 2001; Fattovich *et al.*, 2004). Então, a fibrose/cirrose hepática e o HCC são doenças complexas onde importantes componentes ambientais e genéticos estão envolvidos na sua manifestação clínica. Foram encontrados polimorfismos em genes como *ALDH2* e *CYP2E1*, ligados ao metabolismo do EtOH em hepatócitos, que levam a um aumento do risco de desenvolver HCC em fígado fibrótico de pacientes japoneses infectados por HCC (Kato *et al.*, 2003). Polimorfismos em diversos outros genes foram associados a diferente susceptibilidade a HCC (Taylor *et al.*, 1993; Hsieh *et al.*, 1996, Silvestri *et al.*, 2003).

O HCC é o tumor hepático mais freqüente e altamente vascularizado (Tanigawa *et al.*, 1997). É o quinto tipo de tumor em freqüência mundial, com maior incidência na África Sub-Saariana e Sudeste Asiático devido à alta incidência de infecção crônica por HBV e contaminação dos alimentos com aflatoxinas. É estimado o aparecimento de aproximadamente 500.000 novos casos de HCC por ano, com maior incidência entre homens que entre mulheres, na razão de 2:1, e comprometendo principalmente os pacientes com cirrose hepática (Röcken & Carl-McGrath, 2001; Feitelson *et al.*, 2002).

Fig. 6: Seqüência cronológica das lesões celulares que levam ao aparecimento do carcinoma hepatocelular em humanos (Adaptado de Thorgeirsson & Grisham, 2002).

Como dito anteriormente, no fígado normal o colágeno XVIII é importante componente estrutural do espaço perisinusoidal, sendo produzido principalmente pelos hepatócitos (Musso *et al.*, 1998). Um aumento da expressão deste gene está associado a quadros de fibrose e cirrose hepática, sendo que as células estreladas ativadas começam a produzir a isoforma NC11-303 e os hepatócitos dobram a síntese de RNAm de NC11-493 (Musso *et al.*, 1998; Musso *et al.*, 2001a). Com isso fica claro que as duas células usam promotores diferentes do gene do colágeno XVIII e que a regulação da isoforma NC11-493 é hepatócito específica.

I.2.2 – EXPRESSÃO DO COLÁGENO XVIII

Em 2001(b), Musso e col. mostraram que a diminuição na expressão da isoforma NC11-493 nos tumores está relacionada a hepatocarcinomas de maior tamanho, progressão e índice de recorrência. Outros estudos mostraram que a

expressão do colágeno XVIII/endostatina no parênquima hepático não tumoral era maior que nas células tumorais e, que essa expressão em tecido não tumoral estava correlacionada com a progressão do tumor (Hu *et al.*, 2005). Buscando mostrar a importância da endostatina endógena de células de HCC humano na vascularização e progressão do tumor, Tsuboi e col., 2004, injetaram anticorpos para endostatina na circulação de camundongos *nude* inoculados com linhagem de células de HCC humano, JHH-1, que produzem colágeno XVIII/endostatina. Nesses camundongos o anticorpo inibiu a ação da endostatina levando a uma maior angiogênese e crescimento da massa tumoral que em camundongos controles. Esses trabalhos sugerem que a expressão do colágeno XVIII/endostatina está correlacionada com a progressão tanto do hepatocarcinoma, quanto da fibrose/cirrose hepática, condições precursoras do HCC. Assim, compreender a regulação da expressão do colágeno XVIII em hepatócitos pode auxiliar na compreensão da progressão dessas doenças, bem como contribuir para novas estratégias de tratamento.

I.3. REGULAÇÃO DA TRANSCRIÇÃO

I.3.1. REGULAÇÃO DA TRANSCRIÇÃO

Existem vários níveis de regulação da expressão de uma proteína, (Fig. 7) sendo que a regulação da transcrição de um determinado gene seria o primeiro deles, e no caso de genes hepato-específicos parece ser o mais importante (Derman *et al.*, 1981).

Fig. 7: Representação gráfica dos níveis de expressão de uma proteína (Modificado de Alberts *et al.*, 2002).

Existem dois componentes fundamentais para a transcrição de um gene: suas seqüências de DNA regulatórias, chamadas elementos cis-regulatórios, e as proteínas regulatórias, chamadas elementos trans-regulatórios (Schrem *et al.*, 2002; Levine & Tijan, 2003).

Os elementos cis-regulatórios se dividem basicamente em três tipos: região promotora mínima, regiões ativadoras ou inibidoras da transcrição, também chamadas de *enhancers*, e insuladores (Fig. 8) (Levine & Tijan, 2003). Nesse trabalho trataremos como seqüência promotora, ou promotor de um gene, o conjunto dessas regiões regulatórias atuando na transcrição deste gene. A região promotora mínima geralmente encontra-se *upstream* e a poucas bases do início da transcrição. Ela é a seqüência de DNA onde a maquinaria transcricional, composta pela RNA polimerase II e suas proteínas acessórias, vai se ligar para iniciar a síntese do pré-RNA (Smale *et al.*, 2001; Levine & Tijan, 2003). Os primeiros promotores mínimos estudados mostraram a existência de uma seqüência TATAA padrão, também chamada TATA box, onde se liga a TATA-

binding protein (TBP), parte do complexo proteico que permite a ancoragem da RNA polimerase II no DNA (Smale *et al.*, 2001). Entretanto, posteriormente, verificou-se a existência de pelo menos duas outras diferentes seqüências que também permitem ancoragem do complexo iniciador: o *initiator element* (INR) e o *downstream promoter element* (DPE) (Levine & Tijan, 2003). Os *enhancers* de um gene podem se localizar próximos ou distantes, a mais de 100Kb, assim como, a montante, a jusante ou em *introns* deste gene, e as seqüências insuladoras protegem a região promotora de um determinado gene da ação de regiões regulatórias de outros genes (Levine & Tijan, 2003).

Nas regiões regulatórias vão se ligar uma série de proteínas, fatores de transcrição, que têm contato com seqüências de DNA específicas, aproximadamente 20 bases, chamadas de motivos transcrpcionais ou sítios de ligação. O contato do fator de transcrição com cada base não é muito forte, mas o contato desse fator com todas as 20 bases do sítio será forte e específico. Assim, quanto mais próxima a seqüência de DNA estiver da seqüência ideal do sítio de ligação específico do fator de transcrição, maior será a interação entre proteína e DNA (Alberts *et al.*, 2002). Os diferentes motivos estruturais utilizados pelas proteínas regulatórias para ler as seqüências específicas de DNA acabam levando à sua classificação em famílias de fatores de transcrição com características específicas de interação com seus sítios de ligação. Além disso, os fatores de transcrição atuam em sinergia, e a existência de vários sítios de ligação próximos uns dos outros leva ao agrupamento desses fatores em uma região regulatória, potencializando a atividade do complexo transcricional (Schrem *et al.*, 2002 ; Levine & Tijan, 2003).

Fig. 8: Representação gráfica dos elementos básicos que atuam na regulação da transcrição de um gene.

As proteínas regulatórias se dividem basicamente em dois grupos: as proteínas que se ligam diretamente a sequência de DNA do promotor e as que não se ligam diretamente (Fig. 8). Essas últimas proteínas regulatórias não reconhecem diretamente a sequência de DNA, mas vão se ligar a proteínas ligadas ao DNA, estabilizando o complexo repressor ou ativador da transcrição, sendo chamadas de co-repressores ou co-ativadores conforme a situação (Schrem *et al.*, 2002). Já as proteínas que se ligam diretamente à sequência de DNA vão atuar ativando ou reprimindo a transcrição basicamente de duas formas: facilitando ou dificultando a montagem da maquinaria transcricional da RNA polimerase II, ou alterando a estrutura da cromatina da região promotora, o que vai permitir maior ou menor acessibilidade dos fatores de transcrição e da maquinaria transcricional à sequência de DNA em questão (Schrem *et al.*, 2002).

Existe uma série de fatores de transcrição que são chamados de ubíquos como *Specificity protein 1* (Sp1), *Yin and Yang 1* (YY1) e *Octamer factor 1* (Oct1), estão presentes na maioria das células, participando da transcrição da maioria

dos genes. Mas existem fatores de transcrição que se encontram em uma grande quantidade em determinados tipos celulares e/ou determinados contextos fisiopatológico. No fígado há uma série de fatores de transcrição que estão com quantidade enriquecida, como as famílias de *hepatocyte nuclear proteins* (HNFs) e de *CAAT enhancer binding proteins* (C/EBP) (Krivan & Wasserman, 2001, Schrem *et al.*, 2002). Essas famílias já foram relacionadas com a transcrição de uma série de genes hepato-específicos, como a albumina e, portanto, são fortes candidatas a atuarem na regulação de genes preferencialmente expressos por hepatócitos.

I.3.2 – POLIMORFISMOS DE BASE ÚNICA (SNPs) EM REGIÕES REGULADORAS

Como vimos anteriormente, a sequência de bases do DNA que forma o sítio de ligação de um determinado fator de transcrição vai determinar sua maior ou menor afinidade por esse sítio e, alterações nessas seqüências podem alterar esta afinidade e/ou criar afinidade por outros fatores de transcrição. Então, polimorfismos de base única em regiões cis-regulatórias podem ter um impacto considerável em um fenótipo, a partir do momento que podem alterar a afinidade de fatores de transcrição por estas regiões, alterando o nível de transcrição e, podendo levar a uma mudança na disponibilidade de uma determinada proteína. Acredita-se que esses SNPs tenham um papel importante na evolução das espécies por levarem a uma modulação leve de determinados traços fenotípicos e por representarem muitas vezes mutações co-dominantes (Wray, 2007). Estudos mostram que há uma importante contribuição de polimorfismos genéticos nas diferenças interindividuais de expressão gênica (Yan *et al.*, 2002, Cheung *et al.*, 2003; Morley *et al.*, 2004) e que existe uma grande quantidade de

promotores humanos com SNPs funcionais, isto é, SNPs capazes de alterar a atividade desses promotores (Hoogendoorn *et al.*, 2003, Masotti *et al.*, 2005 Anexo II). Estudos relacionados com polimorfismos em regiões promotoras verificaram que SNPs no promotor de determinados genes estão associados a diferentes riscos a determinadas doenças, como retinopatia diabética (Errera *et al.*, 2007) e infecção crônica por HBV (Hohler *et al.*, 1998), por exemplo.

As variações interindividuais da endostatina circulante e do colágeno XVIII no fígado podem estar relacionadas a SNPs no promotor 2 deste gene, e a identificação e análise funcional desses SNPs são importantes para a melhor compreensão da susceptibilidade a doenças complexas, como a fibrose/cirrose e o HCC.

I.3.3 – REGIÕES NÃO CODIFICANTES E CONSERVADAS EVOLUTIVAMENTE

O seqüenciamento do genoma humano tornou possível o estudo de qualquer região de DNA, mas um dos grande desafios atuais é a identificação de seqüências regulatórias. Acredita-se que regiões não codificantes funcionais do genoma acumulem mutações mais lentamente que as não funcionais por estarem sofrendo pressão seletiva. Assim, desenvolveu-se a técnica de *phylogenetic footprinting* que consiste no processo de definir espécies adequadamente distantes filogeneticamente, alinhar seqüências de DNA a montante de genes ortólogos e identificar segmentos conservados que possivelmente vão funcionar como *enhancers* de algum gene (Pennacchio & Rubin, 2001; Wasserman & Sandelin, 2004).

Além de conservadas, essas seqüências de DNA não codificadoras devem possuir também agrupamentos de prováveis sítios de ligação de fatores de transcrição, formando os chamados módulos cis-regulatórios, aumentando, a probabilidade de serem funcionais. Krivan e Wasserman, 2001, utilizaram a bioinformática na identificação de elementos regulatórios de genes hepato-específicos, mostrando que para diminuir o número de falsos positivos na busca de regiões regulatórias deve-se levar em consideração a propriedade dos fatores de transcrição de interagirem de forma cooperativa e a taxa de conservação filogenética da região reguladora em questão. Atualmente, estão disponíveis gratuitamente na internet softwares de apoio à busca de elementos transcrpcionais tais como o *Transcription Element Search System TESS* (Schug & Overton, 1997) e outros com funcionalidades adicionais tais como o *rVista* (Loots *et al.*, 2002), que busca regiões conservadas, e o *Cluster Buster* (Frith *et al.*, 2003), que busca agrupamentos de elementos transcrpcionais.

I.3. 4 – PROMOTORES DO COLÁGENO XVIII

O fato do gene do colágeno XVIII possuir dois promotores alternativos, separados por aproximadamente 50kb, repete-se em outros genes da família dos colágenos como em $\alpha 2(I)$, $\alpha 1(III)$, $\alpha 6(IV)$, $\alpha 2(VI)$ e $\alpha 1(IX)$ (Rehn *et al.*, 1996). Com isso, diferentes isoformas podem ser expressas em tecidos específicos usando diferentes controles de transcrição.

Em camundongos, as regiões *upstream* aos *exons* 1 e 3 do *Col18a1* foram analisadas por ferramentas de predição de elementos transcrpcionais. A região que corresponderia ao promotor 1 não tem seqüências TATAA e CCAAT, mas contém três GC *boxes* consecutivos, que são típicos de genes

housekeeping. Todos os sítios de início de transcrição da isoforma NC11-301 estão *downstream* a esses GC boxes (Rehn *et al.*, 1996). Isso estaria de acordo com a baixa expressão constitutiva da variante NC11-301 do colágeno XVIII em vários tecidos. O promotor 2 de *Col18a1*, apesar de não possuir TATAbox, possui dois sítios CCAAT próximos ao início da transcrição e diversos sítios prováveis para Sp1 e HNF3 ao longo de 3000pb a montante do exon3 (Liétard *et al.*, 2000). Foram realizados ensaios de luciferase e *mobility shift assays* que mostraram que tanto Sp1, quanto HNF3, exercem papel importante na ativação do promotor murino em hepatócitos. Foi também sugerido que a combinação de um sítio de ligação de HNF3 com um sítio NF1 próximo exerceria uma regulação inibitória e tecido específica de *Col18a1*. Esses dados estão de acordo com a idéia que a arquitetura de regiões promotoras de genes específicos do fígado se caracteriza pela existência de elementos transcricionais comuns e hepato-específicos, espaçados por curtas seqüências de nucleotídeos, permitindo que a transcrição seja controlada através da sinergia de ambos transativadores, comuns e específicos (Krivan & Wasserman, 2001). Foi observado que extrato de proteínas nucleares de tumores de HCC, em relação ao fígado normal, têm maior afinidade por essa seqüência HNF3-NF1 de *Col18a1*^{-/-}, sugerindo que nesses tumores existe alteração na expressão de fatores de transcrição (Liétard *et al.*, 2000).

Os promotores do gene do colágeno XVIII humano ainda não foram caracterizados. Atualmente não estão descritos, em humanos, SNPs nos 3,2kb anteriores ao exon 1, que corresponderia a região do promotor 1, mas já existem SNPs descritos na região de 3,2kb upstream do exon 3, que corresponderia ao promotor 2.

I.4. OBJETIVOS

Assim, diante da grande variação individual na expressão de endostatina, da alta disponibilidade do colágeno XVIII no fígado, sua importância na angiogênese e das alterações de expressão que sofre na cirrose/fibrose hepática e no HCC, levantamos a hipótese que alterações polimórficas na região promotora deste gene em humanos poderiam estar associadas a diferentes níveis de expressão das isoformas NC11-493 e NC11-728 no parênquima hepático, conferindo aos indivíduos, conforme seu genótipo, diferente susceptibilidade a fibrose/cirrose hepática e/ou hepatocarcinoma.

Esta tese de doutorado teve então como principais objetivos:

- a) caracterizar a região promotora 2 do gene do colágeno XVIII humano, elucidando suas regiões regulatórias e os fatores de transcrição que a ela se ligam;
- b) caracterizar os polimorfismos na região promotora 2 do colágeno XVIII na população brasileira e verificar sua funcionalidade *in vitro*;
- c) verificar se estes polimorfismos são responsáveis por parte da variabilidade da expressão de NC11-493 e NC11-728 no fígado humano e se correspondem a alelos de risco ao HCC.

II. CASUÍSTICA E MÉTODOS

II.1 – PACIENTES E CONTROLES

O presente projeto foi realizado em colaboração do Prof. Dr. Flair José Carilho, do Dr. José Eymard Moraes de Medeiros, ambos do Departamento de Gastroenterologia do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (FMUSP), e do Prof. Dr. Venâncio Alves do Departamento de Patologia do Hospital das Clínicas FMUSP. Os pacientes selecionados nesse serviço apresentavam hepatocarcinoma celular.

Foram estudadas amostras de DNA de 90 pacientes (74 descendentes de europeus e 16 descendentes de africanos). O grupo controle foi constituído por 258 indivíduos (121 descendentes de europeus e 137 descendentes de africanos) sem histórico de HCC, recrutados no Hemocentro do Hospital das Clínicas da FMUSP. Os indivíduos foram classificados usando critério que inclui: cor de pele, tipo de cabelo, formato do nariz e da boca (Krieger *et al.*, 1965). Para os indivíduos controles foi explicado o objetivo da pesquisa e requisitada a assinatura de termo de consentimento, conforme resolução 196 do Conselho Nacional de Ética Brasileira. Esse material encontra-se estocado no laboratório da Profa. Dra. Ma Rita Passos-Bueno.

Durante dez meses parte do projeto foi realizado na unidade INSERM U620, em Rennes na França, com a colaboração do Dr. Orlando Musso e o financiamento da organização francesa EGIDE.

II.2 – CONSIDERAÇÕES ÉTICAS

O presente projeto passou pelo Comitê de Ética do Hospital das Clínicas FMUSP.

II.3- MÉTODOS

II.3.1- EXTRAÇÃO DE DNA

A extração do DNA genômico, a partir de linfócitos, foi realizada com amostras de 5-10 ml de sangue total coletado em tubos contendo 0,4 ml de EDTA 5 % com anticoagulante. A extração foi realizada de acordo com a metodologia descrita por Miller e col., 1988, que obedece as seguintes etapas:

1. Transferir cada amostra de sangue para um tubo polipropileno (Falcon) e completá-lo com 50 ml de solução de lise de membrana celular (NH_4Cl 1550 mM; KHCO_3 100 mM; EDTA 10 mM pH 7,4), sob leve agitação.
2. Manter o tubo no gelo durante 30 minutos.
3. Centrifugar durante 15 minutos a 1800 rpm.
4. Desprezar o sobrenadante e lavar o precipitado com 10 ml da solução de lise.
5. Centrifugar novamente por 5 minutos a 2.000 rpm.
6. Desprezar o sobrenadante e ressuspender o precipitado em 3 ml de solução de lise dos núcleos (Tris-HCl 100 mM pH 8,0; NaCl 4M; EDTA 20 mM pH 8,2).
7. Acrescentar 50 μl de proteinase K na concentração de 10 mg/ ml e 300 μl de SDS 10%. Agitar levemente.
8. Incubar os tubos a 37°C por um período de 3-24 horas.
9. Após a incubação, adicionar 1 ml de NaCl 6M no tubo e misturar vigorosamente.
10. Centrifugar por 20 minutos a 2.500 rpm.
11. Transferir o sobrenadante para outro tubo Falcon limpo e centrifugar por mais 15 minutos a 2500 rpm.

12. Transferir o sobrenadante para um tubo de vidro. Precipitar o DNA adicionando duas vezes o volume de etanol absoluto. Inverter o tubo algumas vezes.

13. Recolher o DNA com auxílio de um bastão de vidro e lavá-lo com etanol 70%.

14. Depositar o DNA em tubo *Eppendorf* e dissolvê-lo em 200-800 μ l de TE⁻⁴ (Tris-HCl 10 mM; EDTA 0,1 mM; pH final de 7,5).

15. Incubar a 65°C durante 30 minutos e em seguida armazenar a 4°C.

A quantificação do DNA foi feita com auxílio de um espectrofotômetro (Ultrospec 3000PRO - Amersham Biosciences) utilizando comprimento de onda igual a 260 nm. Para leitura, foram utilizados 2 μ l de DNA diluídos em 498 μ l H₂O MilliQ. Normalmente o rendimento da técnica acima descrita é de aproximadamente de 300-400 μ g de DNA por 10 ml de sangue.

II.3.2 – PCR

O método de PCR (*Polymerase Chain Reaction*) é uma técnica que permite a amplificação *in vitro* de seqüências específicas de DNA, usando a enzima DNA polimerase, os quatro nucleotídeos e um par de *primers* em fitas de DNA complementares (Saiki *et al.*, 1985). Os *primers* são usados para direcionar a síntese de DNA em ciclos repetidos. Em cada ciclo, ambas as fitas servem como moldes para a formação de fitas novas, obtendo-se assim um aumento exponencial do produto. Desta forma, tem sido permitido o estudo de seqüências de interesse pelos métodos convencionais de análise de DNA.

As condições de PCR de todos os DNA amplificados neste trabalho seguem o seguinte protocolo:

1. 50-100 ng de DNA total

2. 10 mM Tris-HCl, 50 mM KCl, 1,5 mM MgCl₂, pH9,0 à temperatura ambiente (10X PCR Buffer - Amersham Biosciences)
3. 200 µM de cada um dos 4 dNTPs (ATP, CTP, GTP e TTP)
4. 0,50 µM de cada *primer*
5. 0,5 unidades da enzima Taq DNA polimerase (Amersham Biosciences)
6. Completar com H₂O para um volume final de 20 µl

Nas seguintes condições de tempo:

94°C	-	4 minutos	
94°C	-	40 segundos	} 35 ciclos
Annealing-	-	40 segundos	
72°C	-	40 segundos	
72°C	-	6 minutos	

O sucesso da amplificação foi verificado por meio de eletroforese em gel de agarose 1,5% a 150V por uma hora. Aproximadamente 5 µl de cada amostra de produto de PCR a ser analisada foi misturado a 1 µl de tampão de corrida (sucrose 40%; bromofenol *bleu* 0,25%) e aplicados no gel juntamente com um produto de DNA de tamanho conhecido. Após a corrida, o gel foi corado com brometo de etídeo (0,5 µg/ml) e visualizado em um transiluminador de luz ultravioleta, a fim de confirmarmos a presença de uma única banda do tamanho esperado.

Os termocicladores utilizados foram da *Perkin Elmer* e PTC-150 *MJ Research*.

Padronizamos as condições da PCR para triar SNPs ao longo de 1487pb a montante do exon 3, correspondendo ao promotor 2 do colágeno XVIII. As

condições ótimas de *annealing* e os *primers* utilizados para triagem de polimorfismos no promotor 2 do colágeno XVIII estão descritos nas Tabela 1.

Tabela 1. *Primers* desenhados para amplificação de 600pb 5' ao exon 3 do gene COL18A1.

Nome	Ann.	Tam.	FOWARD 5'-3'	REVERSE 5'-3'
P2COL01	64 °C	254 pb	TAA ATC CTT GGG TGC TGG AAC C	AAG AGC AGC AGC AGG ATG TGG
P2COL02	64 °C	245 pb	TGC CTC AGT GCT TCT GGG ATG	GCA GAA GGA AAG AAG GCA CCA G
P2COL03	64 °C	258 pb	AGC AGC TCT GAG AGG GTC TGA G	GGC TGC TGA GTC TCT GGA TTG
P2COL04	60 °C	227 pb	CTA ACC GGA TCC TTT TGT GG	GCC CAA GTC GTT AAA CAT CC
P2COL05	60 °C	236 pb	AGG GCA TAA GGC ACA CTC AG	CTC AGC CCT TCC TGA CCT TC
P2COL06	64 °C	248 pb	GAT GCA CAT GTG TGA ACG TG	GGA TGT GAG TGG CAT GTG AG
P2COL07	62 °C	291 pb	ACC GAT GGC TAC AGC ATG TC	AGC CAG GCA GTT TAC ACA GG
P2COL08	62 °C	242 pb	CTG GGC ATC TCT GAG ACG AG	AAA CAG CCC CAC AAC CTG TC
P2COL09	60 °C	178 pb	CCA CAA CAG AGT TCC CTT CC	AAT CCC CAC ACT GCT CAA AC

Para P2COL02, P2COL03 e P2COL06 o PCR é do tipo “*touchdown*” e consiste das seguintes etapas:

95°C	-	2 minutos	
94°C	-	20 segundos	} 12 ciclos
70°C	(-0,5°C a cada ciclo)-	30 seg.	
72°C	-	40 segundos	
94°C	-	20 segundos	} 23 ciclos
64°C	-	30 segundos	
72°C	-	40 segundos	
72°C	-	10 minutos	

II.3.3 - TRIAGEM DE SNPs

A triagem de polimorfismos nos indivíduos controles foi iniciada com a técnica de polimorfismo conformacional de fita simples, ou SSCP (*Single Strand Conformation Polymorphism*) (Orita *et al.*, 1989), e posteriormente começamos a realizar somente a técnica de cromatografia líquida denaturante de alta performance, ou dHPLC (*denaturing High Performance Liquid Chromatography*) (Xiao & Oefner, 2001). As amostras que apresentaram um padrão diferente do normal foram posteriormente seqüenciadas.

II.3.3.1 – MÉTODO DE SSCP

O método de SSCP (*Single Strand Conformation Polymorphism*) baseia-se na mobilidade eletroforética diferencial de um fragmento de DNA em fita simples em géis não desnaturantes, em função não só do seu tamanho, mas também da sua seqüência de nucleotídeos. Além disso, a mobilidade eletroforética é sensível à temperatura. As duas moléculas de fitas simples de DNA de cada produto de PCR desnaturado assumem uma conformação tridimensional a qual é dependente da seqüência primária. Assim, a troca de apenas um par de base, uma inserção ou uma deleção podem ser detectadas pela presença no gel de bandas que migram diferentemente entre as amostras (Orita *et al.*, 1989).

Os produtos amplificados foram submetidos a uma análise por SSCP em gel de acrilamida não desnaturante (MDETM). As amostras de PCR foram misturadas ao corante de formamida, ou SSCP-loading buffer (980 mL/L formamida, 10 mmol/L EDTA, 0,25 g/L xileno cianol FF, e 0,25 g/L bromofenol blue) na proporção 1:1, desnaturados à 95°C por 5 minutos, mantidas no gelo e

aplicadas no gel de poliacrilamida. Estes géis foram submetidos à eletroforese à 8W por 12-24 horas, dependendo do tamanho do fragmento de PCR, à temperatura ambiente, e corados por meio da coloração com nitrato de prata (Bassan *et al.*, 1991) modificada, como a seguir:

1. Fixar o DNA no gel em soluções de etanol 10% por 10 minutos e ácido nítrico 1% por 3 minutos,
2. Lavar o gel em água destilada,
3. Corar em nitrato de prata (AgNO_3 0,2%; formaldeído 0,055%) por 20 minutos,
4. Lavar o gel em água destilada,
5. Revelar em solução 1 (Na_2CO_3 2%; formaldeído 0,55%), seguida pela solução 2 (Na_2CO_3 2%; formaldeído 0,0275%),
6. Bloquear a reação de coloração em solução de ácido acético 10%,
7. Lavar o gel em água destilada.

Após 24 horas secando à temperatura ambiente, o gel era fotografado em um filme apropriado (*Typon-Promega*) com auxílio de um megacoscópio e arquivado. Quando observada alteração no padrão de migração das bandas o produto de PCR era sequenciado automaticamente como descrito no item II.3.3.3.

II.3.3.2 – MÉTODO DE dHPLC

Esse método é normalmente referido como desnaturação em cromatografia líquida de alta performance, ou dHPLC (Oefner & Underhill, 1995). Ele permite uma detecção automatizada de alterações no DNA.

Esta metodologia, que baseia-se na detecção de mutações por meio de análise de heteroduplex de cromatografia modulado por temperatura (Kuklin *et al.*, 1997), é comercializada pela *Transgenomic Inc.*

A triagem de mutações por meio da técnica de dHPLC baseia-se na retenção diferencial de moléculas de heteroduplex de DNA em relação à moléculas de homoduplex em um cartucho de retenção de ácido nucléico sob condições de desnaturação parcial do DNA. Os fragmentos de PCR de heteroduplex contêm uma quantidade maior de regiões não pareadas em comparação aos fragmentos homoduplex, sendo assim, são retidos por menos tempo do que seus análogos homoduplex no cartucho da fase estacionária (Xiao & Oefner, 2001).

Esta metodologia automatizada é composta por um cartucho de copolímeros não-porosos de estireno-divinilbenzeno (*DNASep^R*, *Transgenomic*) e por soluções de triethylammonio acetato (TEAA) e de acetonitrila. Inicialmente, o DNA é carregado juntamente com uma solução constante de TEAA (100 mM) até o cartucho (*DNASep^R*, *Transgenomic*) de retenção do DNA, fase estacionária. Esta solução (TEAA) é responsável pela ligação do DNA de dupla fita (DNAds) ao cartucho. Durante a corrida, quantidades gradativamente crescentes de acetonitrila são adicionada à fase móvel. Esta substância é o solvente orgânico responsável pela liberação do complexo DNAds-TEAA do cartucho. Os fragmentos de DNA liberados são detectadas por um feixe de luz ultra-violeta. Estes sinais são convertidos em valores digitais, apresentados como cromatogramas, nos quais os picos corresponderiam aos fragmentos de DNA (protocolo da *Transgenomic Inc.*).

Esta é uma metodologia automatizada, sendo necessária uma manutenção regular do aparelho, *Wave^R Nucleic Acid Fragment Analysis System, Transgenomic Inc.* Além disto, este aparelho possui um programa, *Wavemaker*, que calcula e sugere temperaturas parciais ideais para a detecção de mutações ao longo da seqüência de DNA estudada, assim como, as porcentagens gradativas de solvente orgânico necessárias para a liberação de cada seqüência de DNA (protocolo da *Transgenomic Inc.*).

Durante a padronização de cada produto de PCR, é necessário, fazer um teste de qualidade das condições de PCR. Nesta fase, o produto de PCR é aplicado sob condições não desnaturantes no dHPLC. O produto ideal é aquele no qual se verifica apenas um pico no cromatograma (indicando que existem apenas fragmentos de DNA de mesmo tamanho).

Em seguida, as condições teóricas calculadas pelo programa *Wavemaker* devem ser validadas com testes práticos, de preferência usando-se amostras com mutações já conhecidas. Nesta fase, após a amplificação do DNA, uma reação adicional para geração de heteroduplex é necessária (95°C a 4 minutos, seguidos por redução lenta e gradativa da temperatura de 95°C à 25°C, por aproximadamente 45 minutos).

A relação de temperaturas no dHPLC para análise de polimorfismos no promotor 2 do colágeno XVIII é descrita na Tabela 2.

Tabela 2. Temperaturas usadas para análise dos *amplicons* por *dHPLC*.

Nome	Temperaturas
P2COL01	62 e 63 °C
P2COL02	62 e 63 °C
P2COL03	61 e 63 °C
P2COL04	62 e 64 °C
P2COL05	63 e 64 °C
P2COL07	63 e 64 °C
P2COL08	63 e 64 °C
P2COL09	61 e 62 °C

II.3.3.3 – SEQÜENCIAMENTO AUTOMÁTICO

Para o seqüenciamento, o produto de PCR foi quantificado por meio de eletroforese em gel de agarose 1,5% e TBE 1X utilizando como padrão, uma amostra de concentração conhecida (*Low DNA MassTM Ladder* da Invitrogen, ou um outro produto de PCR conhecido). Aproximadamente 40 ng do produto da PCR eram purificados por meio de adição de 10 unidades da enzima Exonuclease I (*USB*) e de 5 unidades da enzima *Shrimp Alkaline Phosphatase* (*Amersham Biosciences*), seguida de incubação a 37°C por 30 minutos e inativação da enzima a 80°C por 15 minutos.

O produto purificado foi aliquoteado em 2 novos tubos de PCR, aos quais foram posteriormente adicionados 4 µl de *Premix* (*Amersham Biosciences*). Em um dos tubos, adicionavam-se 1 µl (5 µM) do *primer Forward* e no outro 1 µl (5 µM) do *primer Reverse*, previamente usados na respectiva reação de PCR. O volume final de 10 µl era completado com H₂O. A reação de seqüenciamento foi realizada nas seguintes condições: 24 ciclos de desnaturação à 95°C por 10 segundos, *annealing* à 50°C por 10 segundos e extensão à 60°C por 1 minutos.

O produto da reação de seqüência foi então precipitado de acordo com as seguintes etapas:

1. Acrescentar 1 µl de acetato e 27,5 µl de etanol absoluto;
2. Centrifugar a 12.000 rpm por 15 minutos;
3. Retirar o sobrenadante, acrescentar 150 µl de etanol 70%;
4. Centrifugar a 12 rpm por 10 minutos;
5. Retirar o sobrenadante;
6. Deixar no gelo os tubos com a tampa aberta por 2 minutos para que todo resíduo de etanol evapore;
8. Acrescentar 10 µl de *Loading Solution* (Amersham);
9. Passar os tubos por 10 segundos no vórtex.

Inicialmente, as amostras de PCR foram seqüenciados no equipamento *ABI Prism™ Sequencer 377* (versão 3.0 *Applied Biosystems*) usando o kit *BigDye Terminator Cycle Sequencing* (PE Biosystems, Foster City, CA) para caracterização da natureza da alteração. Posteriormente usamos o equipamento *MegaBACE™ 1000 DNA Analysis Systems* (Amersham Biosciences).

II.3.3.4 – MÉTODO DE SNUPE

Este método denominado SNUpe (*Single Nucleotide Primer Extension-Amersham Biosciences*) é empregado para análise de SNPs (*Single Nucleotide Polymorphisms*). Ele consiste da utilização de um único *primer* com cerca de 20 bases, sendo que a última base deste antecede ao nucleotídeo que pode estar mutado. O *premix*, com o qual esta reação é feita, contém além de tampões e termosequenases, também dideoxynucleotídeos marcados com diferentes fluorocromos.

O método se baseia no princípio de que dideoxynucleotídeos quando incorporados, cessam a síntese da molécula, o que ocorre logo após a

incorporação de apenas uma base na posição 3' do *primer*, que corresponde a posição da base polimórfica. O dideoxinucleotídeo adicionado, por estar marcados cada base com diferente fluorocromo, emite uma fluorescência quando excitado por um feixe de laser e terá uma cor de acordo com a base adicionada. Por exemplo, se for uma citosina emitirá um pico de cor azul, e se for adenina, um pico de cor verde. Este método foi empregado utilizando-se o seqüenciador automático (MegaBACE™ 1000 DNA Analysis Systems) e seu *software* correspondente (Amersham Biosciences). As condições para esta reação são as seguintes: 5 ng do produto amplificado, 4 µL de premix, 2 pmoles de *primer* sintetizado para a reação de SNuPE e água milli-Q, em um volume final de 10 µL.

II.3. 4 – ANÁLISE *IN SILICO*

II.3. 4.1 – PROGRAMAS DE PREDIÇÃO DE FATORES DE TRANSCRIÇÃO

Os programas para predição de elementos regulatórios em seqüências de nucleotídeos se baseiam em matrizes de sítios consenso de ligação de fatores de transcrição. O programa analisa uma determinada seqüência, verificando o grau de homologia com essas matrizes, e apresenta um *score* de similaridade. Essas análises orientam os estudos funcionais de regiões regulatórias. Os programas mais utilizados estão listados abaixo e são de uso livre.

- GENOMATIX: Genomatix Software GmbH (<http://www.genomatix.de/>).
- TESS (www.cbil.upenn.edu/cgi-bin/tess/)
- Tfsearch (www.cbrc.jp/research/db/TFSEARCH.html)

II.3. 4.2 – GENÔMICA COMPARATIVA

Utilizamos o programas VISTA (<http://genome.lbl.gov/vista/index.shtml>) disponível gratuitamente na internet para a identificação de regiões conservadas. Além de alinhar seqüências genômicas de diversas espécies, este programa prediz sítios conservados de ligação fatores de transcrição.

II.3. 5 – ESTUDO FUNCIONAL DA REGIÃO PROMOTORA DE COL18A1

Iniciamos o estudo funcional da região promotora utilizando ensaios de gene repórter firefly luciferase (*luc+*). O vetor utilizado foi o pGL3-Basic Vector (Promega) que não possui promotor eucariótico e possibilita a clonagem de seqüências regulatórias a montante do gene repórter *luc+*. O gene *luc+* expressa a luciferase quando a seqüência regulatória em teste é ativada nas células transfectadas com esse plasmídeo.

II.3. 5.1 – CLONAGEM

Construímos diversos vetores com os primeiros 1487pb do promotor 2 do colágeno XVIII clonados no vetor pGL3-Basic, descritos na Tabela 3.

Tabela 3. Construções utilizadas no estudo do promotor 2 de *COL18A1*.

Construções	Descrição
pGL3-1487inv	Região de 1487pb a montante do exon 3, correspondendo ao promotor 2, clonada no sentido 3'-5'
pGL3-1487	Região de 1487pb a montante do exon 3, clonada no sentido 5'-3'
pGL3-1487/700G	Região de 1487pb a montante do exon 3, clonada no sentido 5'-3', com uma guanina na posição correspondendo a base -700 do promotor.
pGL3-1310	Região de 1310pb a montante do exon 3, clonada no sentido 5'-3'
pGL3-1052	Região de 1052pb a montante do exon 3, clonada no sentido 5'-3'
pGL3-555	Região de 555pb a montante do exon 3, clonada no sentido 5'-3'
pGL3-458	Região de 458pb a montante do exon 3, clonada no sentido 5'-3'
pGL3-186	Região de 186pb a montante do exon 3, clonada no sentido 5'-3'
pGL3-186mut	Região de 186pb a montante do exon 3, clonada no sentido 5'-3', com sítio de ligação para C/EBP β mutado descrito no Capítulo V.
pGL3-D700	Região de 1487pb a montante do exon 3, clonada no sentido 5'-3', com deleção de 41pb correspondendo a sonda usada em EMSA para estudo de fatores de transcrição do SNP-700 descrita no Capítulo III.

Os insertos dos vetores foram obtidos amplificando-se a região em questão com *primers* com cauda *polylinker* para as enzimas de restrição Bgl II e Hind III, que permitem a inserção correta do inserto no vetor pGL3-Basic (Cap III Material e Métodos). A quantidade de 1ug produto de PCR purificado e 1ug do vetor pGL3-Basic foram digeridos com as enzimas de restrição Bgl II e Hind III (37° C por 15h).

O vetor pGL3-Basic foi desfosforilado após a digestão e, juntamente com os insertos digeridos, foi aplicado em gel de agarose 0,8% para extração e purificação das bandas por coluna GFX. Esse inserto e vetor purificados foram

quantificados por meio de eletroforese em gel de agarose 0,8% e TBE 1X utilizando como padrão, uma amostra de concentração conhecida (*Low DNA MassTM Ladder* da Invitrogen, ou um outro produto de PCR conhecido)

Para a reação de ligação foram colocados em tubo *Eppendorf* o volume de inserto e de vetor purificados correspondendo a proporção de 300ng:100ng respectivamente. Esse volume foi então reduzido em centrífuga *Speed Vac* (*Eppendorff[®]*) a 15uL finais, onde acrescentou-se 1uL da enzima T4 ligase (*Gibco*), 2uL de tampão ligase com ATP (10x) e 2uL de Água Milli-Q. A reação ocorreu em banho seco a 15°C por 15h aproximadamente. Após esse período, a T4 ligase foi inativada no banho seco a 65°C por 20 min.

Dois microlitros da reação de ligação foram eletroporados em bactérias eletrocompetentes (*Top10*) e estas foram plaqueadas em meio seletivo com ampicilina. Após 12h, as colônias que cresceram foram submetidas a PCR com *primers* (Promega) flanqueando o sítio de clonagem do vetor pGL3-Basic. O produto de PCR das colônias foi aplicado em gel de agarose 0,8% para seleção das que incorporaram o inserto com tamanho adequado. Aquelas que incorporaram o inserto adequado tiveram o produto de PCR purificado por Exonuclease I e Shrimp Alkaline Phosphatase (37°C por 1h, 80°C por 15 min) e foram seqüenciadas com os *primers* do vetor e com *primers* usados para a triagem de polimorfismos na região promotora descritos no item II.3.2 – PCR.

II.3. 5.2 – ENSAIO DE LUCIFERASE

Para os primeiros ensaios de luciferase, 1µg da construção contendo o gene repórter da luciferase e 0,5µg de pRL-TK (*Promega*) foram transfectados em duplicatas, em placas de 12 wells, com *Lipofectamine* 2000 (Invitrogen), de

acordo com o protocolo do fabricante. O vetor pRL-TK possui o gene da *Renilla* luciferase, que é controlado pelo promotor constitutivo da *timidina kinase* do vírus *herpes simplex*. Esse vetor é usado como um controle interno para normalizar a eficiência de transfecção. As células são lisadas 48h após a transfecção, e a expressão do gene repórter é medida utilizando-se o *Dual-Luciferase Reporter Assay System* (Promega) num luminômetro de placa (LB96V *Microlumat Plus* - EG&Berthold). Os diferentes níveis de luminescência indicam a atividade da região regulatória clonada. Os controles do ensaio são: o controle negativo, o grupo de células transfectadas com o vetor pGL3-Basic “vazio”, e o controle positivo outro grupo transfectado com o vetor pGL3-SV40, que possui o promotor do *SimianVirus* com alta atividade na maioria de linhagens celulares de mamíferos.

Os valores foram expressos como atividade relativa da região candidata. Essa atividade relativa é definida como: valor da *firefly luciferase* obtido com a construção repórter normalizado pelo valor da *renilla luciferase* obtido vetor pRL-TK, esse valor é posteriormente dividido pela atividade normalizada de um plasmídeo referência do ensaio, geralmente pGL3-Basic “vazio”, considerado com atividade igual a 1. Os dados de luciferase foram gerados pela média dos resultados de ao menos 3 experimentos independentes. A significância estatística (*P*-value) foi calculada usando One-Way ANOVA com posterior Teste Múltiplo de Comparação de Tukey com a ajuda do software Graph Prism version 4.0 (Graph Pad Software, San Diego California USA, www.graphpad.com). O teste de Mann Whitney foi utilizado para comparar as atividades dos vetores pGL3-1487 e pGL3-1487/700G. O nível de significância adotado foi de 0.05.

II.3.5.3 – CULTURA DE CÉLULAS

As linhagens celulares utilizadas são linhagens imortalizadas vendidas pela American Type Cell Culture (ATCC) como as linhagens HepG2 (hepatoma humano), Hep3B (hepatoma humano), HCT116 (câncer de cólon) e HEK293T (epitélio renal humano) ou cedidas pela unidade INSERM U620 como Huh7 (hepatoma humano) e HepaRG (hepatoma humano). Essas linhagens foram cultivadas conforme descrito nos capítulos III, IV e V.

II.3.6 – ESTUDOS PROTÉICOS

II.3.6.1 – ELETROPHORETIC MOBILITY SHIFT ASSAY (EMSA)

Preparamos extrato nuclear de HepG2 de acordo com o protocolo de Read e col., 1998, e a concentração de proteínas foi determinada usando o reagente Bradford (Bio-Rad). Diversas sondas correspondendo a seqüências com prováveis sítios de ligação para fatores de transcrição nos 1487pb do promotor 2 de *COL18A1*, foram preparadas, anelando cada uma com sua respectiva sonda complementar. Essas sondas estão listadas no Capítulo III. Antes do anelamento, as sondas senso foram marcadas com [$\gamma^{32}\text{P}$]ATP e T4 polynucleotide kinase (New England), seguido de purificação usando colunas Sephadex G50 (Amersham Pharmacia Biotech). As reações de ligação de proteínas foram realizadas em 20 ou 25 μL de tampão (25mM Tris-HCl, pH 7,5, 50mM KCl, 1mM dithiothreitol, 5% glicerol), com 300-500 cpm de sonda marcada, 1 μg de poly (dl:dC) e 5 μg de extrato nuclear de HepG2. As misturas foram incubadas por 30 min à temperatura ambiente na presença ou na ausência de competidores específicos ou não específicos não-marcados. Para os ensaios de *supershift*, a reação de ligação foi pré-incubada com os

respectivos anticorpos Sp1, Sp3, YY1, PEA3, Ets1 e Oct1 (Santa Cruz Biotechnology, Inc). Os complexos proteína-DNA foram analisados em eletroforese vertical (gel de acrilamida 6% não-desnaturante) com o tampão de corrida 0,5X Tris-borate-EDTA.

II.3.6.2 – WESTERN BLOT

A linhagem celular Hep3B foi transfectada com doses crescentes do vetor de expressão MSV-C/EBP β e o vetor pcDNA3 sempre completando um total de 50ug de plasmídeo transfectado. As células foram lavadas com PBS 1X, raspadas da placa com um tampão de lise (b-glycerophosphate 60mM, MOPS 25mM pH 7.2, EGTA 15mM, MgCl₂ 15mM, NaF1mM, DTT 2mM, ortovanato de sódio 1mM, PMSF 1mM and inibidor de protease 1%) e sonicadas. A concentração de proteínas foi determinada pelo método de Lowry (Lowry *et al.*, 1951). Para a detecção de C/EBP β foram separados 30ug de proteína em eletroforese de gel 7,5% de bis-acrilamida e transferidos para membrana de nitrocelulose. Após a transferência a membrana foi bloqueada em temperatura ambiente com 3% de albumina de soro bovino em PBST (1X tampão de solução fosfato salina e 0.01% Tween20).

A membrana foi então incubada por 14 horas em câmara fria com anti-C/EBP β (1/500) (Santa Cruz, Santa Cruz, California) em 3% de albumina de soro bovino em PBST. Depois foi realizada incubação por 1 hora com anticorpo secundário conjugado com peroxidase *horseradish*. As bandas foram então reveladas utilizando SuperSignal West Dura Extended Duration Substrate (Pierce, Rockford, IL).

II.3.6.3 – COIMUNOPRECIPITAÇÃO DE CROMATINA (CHIP)

A coimunoprecipitação de cromatina é uma técnica para se detectar *in vivo* a ligação específica de proteínas, como fatores de transcrição, ao DNA (Kuo & Allis, 1999). Células HepaRG confluentes, cultivadas por 10 dias, foram tratadas como descrito anteriormente (Gripon et al., 2002). As células foram fixadas com formaldeído 1% e a co-imunoprecipitação da cromatina foi realizada utilizando o kit CHIP-IT™ Chromatin Immunoprecipitation Kits & Shering Kits (Active Motif, Rixensart, Belgium) segundo orientações do fabricante. Os anticorpos C/EBP α e C/EBP β sc-150 (Santa Cruz) foram utilizados para coimunoprecipitar o DNA que posteriormente foi utilizado para reação de PCR. Os primers usados para amplificação do DNA foram -345F: CCCAATCCAGAGACTCAGCAG e -142R: CACTCTTGCAGAAGGAAAGAAGG e; -163F: CTGGTGCCTTCTTTCCTTCTG e +43R: AGAGCAGCAGCAGGATGTG.

Como controles negativos foram utilizadas reações de PCR tendo como *templates* o resultado da coimunoprecipitação com anticorpo α SMA (Santa Cruz) ou apenas água Milliq. Como controle positivo foi utilizada a cromatina total antes da coimunoprecipitação.

II.3.7 – PCR QUANTITATIVO (qPCR)

O RNA total foi isolado de amostras de tecido de fígado fibrótico/cirrótico e tumoral de pacientes com HCC utilizando o kit SV Total RNA Isolation System (Promega, Madison, WI) seguindo as orientações do fabricante e armazenadas a -80°C. 1 μ g de RNA foi usado para reação de transcrição reversa em cDNA utilizando o kit High-Capacity cDNA Archive (Applied Biosystems, Foster City,

CA). Os produtos de RT foram diluídos em água para 200ul e, 15ng de cDNA foram utilizadas para a reação de PCR quantitativo. Os primers foram desenhados com a ajuda do programa Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) e as seqüências selecionadas estão descritas no Capítulo IV e V. Os pares de primers foram testados por curva de dissociação para garantir a especificidade em cada reação, isto é ausência de dímeros de primers e outros produtos não específicos. O qPCR foi realizado utilizando SYBR Green I (Applied Biosystems), em um sistema de detecção de seqüências ABI Prism 7300 (Applied Biosystems). As curvas de amplificação foram lidas com a ajuda do software ABI Prism 7300 SDS (Applied Biosystems) utilizando o método comparativo de gatilho de ciclos (*comparative cycle threshold method*). A normalização dos *threshold cycle* (Ct) dos genes de interesse foi realizada com Ct do 18S da amostra, e expressa como: $\Delta Ct = Ct \text{ gene de interesse} - Ct \text{ 18S}$. A expressão relativa do gene dos pacientes foi expressa como $\Delta\Delta Ct = \Delta Ct \text{ gene de interesse} - \Delta Ct \text{ gene da amostra calibradora}$, e então calculado o $2^{-\Delta\Delta Ct}$, segundo a metodologia publicada (Livak and Schmittgen, 2001). Cada reação foi realizada em duplicata e a variabilidade entre os ΔCt utilizados sempre inferior a 10%. A amostra calibradora utilizada foi o produto de RT PCR do RNAm de linhagem celular HepaRG, gentilmente cedida pelo Dr. C. Guguen-Guillouzo. As células foram cultivadas como descrito anteriormente (Gripon et al., 2002) e após 2 semanas, quando atingiram a confluência, o RNAm foi extraído como descrito anteriormente.

Characterization of human Collagen XVIII promoter 2: interaction of Sp1, Sp3 and YY1 with the regulatory region and a SNP that increases transcription in hepatocytes.

Lucia M. Armelin-Correa¹, Chin J. Lin^{2,3}, Angela Barbosa³, Kelly Bagatini¹, Sheila M. B. Winnischofer⁴, Mari C. Sogayar⁴, Maria Rita Passos-Bueno¹

¹Department of Genetics and Evolution Biology , Bioscience Institute, University of São Paulo, São Paulo, Brazil;

²Department of Pathology, Hospital das Clínicas da FMUSP, University of São Paulo, São Paulo, Brazil;

²Endocrinology Division, Hospital das Clínicas da FMUSP, University of São Paulo, São Paulo, Brazil;

⁴Department of Biochemistry, Chemistry Institute, University of São Paulo, São Paulo, Brazil

Corresponding author:

Dr. Maria Rita Passos-Bueno

Centro de Estudos do Genoma Humano – Depto. de Biologia

Instituto de Biociências - Universidade de São Paulo

Rua do Matão, 277 sala 200

São Paulo, SP, Brazil CEP: 05508-900

Telephone: +55-11-3091-9910 FAX: +55-11-3091-7419

E-mail address: passos@ib.usp.br

Publicado na revista Matrix Biology, 2005, vol. 24, pag. 500-509.

Abbreviations: SNP – Single Nucleotide Polymorphism, Sp – Specificity proteins, YY1 - Yin and Yang 1, C/EBP - CCAAT/Enhancer-Binding Protein, HNF3 – Hepatocyte Nuclear Factor 3, TF – Transcription factor, PCR – Polymerase Chain Reaction, EMSA – Eletrophoretic Mobility Shift Assay

Keywords

Collagen XVIII; transcriptional regulation; functional SNP; Sp1; Sp3; YY1

Resumo

Diferentes níveis de expressão do colágeno XVIII foram associados a diversos processos patológicos como câncer, fibrose hepática, retinopatia diabética e doença de Alzheimer. Compreender a regulação da transcrição do colágeno XVIII pode elucidar vias relacionadas com a progressão dessas doenças. O promotor 2 do gene *COL18A1* é pouco estudado, e é responsável pela transcrição desse gene em diversos tecidos adultos como fígado, olhos e cérebro. Este estudo teve como objetivo a caracterização dos elementos cis-regulatórios que interagem com o promotor 2 de *COL18A1* e, a identificação de SNPs nessa região em diferentes grupos étnicos. Nossos resultados mostraram que existem cinco regiões conservadas (I a V) entre o promotor 2 de humanos e camundongos e, que o promotor mínimo de *COL18A1* está localizado entre os nucleotídeos -186 e -21. Sp1 e Sp3 se ligam a região conservada I e V, enquanto que Sp3 e YY1 interagem com a região II. Nós verificamos que o SNP-700T/G faz parte de dois haplótipos comuns, com frequência diferentes entre descendentes de europeus e africanos. O alelo G aumenta a atividade do promotor e a ligação de um fator de transcrição ainda desconhecido. O SNP-700T/G afeta a interação de Sp3 e YY1 com o promotor 2 de *COL18A1* e, o nucleotídeo -700 é parte do sítio de ligação de um fator de transcrição ainda desconhecido, envolvido na expressão deste gene em hepatócitos. Além disso nós confirmamos o envolvimento de Sp1 na regulação deste gene.

Abstract

Different levels of Collagen XVIII expression have been associated with several pathological processes such as cancer, liver fibrosis, diabetic retinopathy and Alzheimer's disease. Understanding the transcriptional regulation of Collagen XVIII might elucidate some pathways related to the progression of these diseases. The promoter 2 of COL18A1 gene is poorly understood and is responsible for the transcription of this gene in several adult tissues such as liver, eyes and brain. This study focused upon characterization of cis-regulatory elements interacting with human COL18A1 promoter 2 and identification of SNPs in this region in different ethnic groups. Our results show that there are five conserved regions (I to V) between human and mouse promoter 2 and that the human COL18A1 core promoter is located between nucleotides -186 and -21. Sp1 and Sp3 bind to conserved regions I and V, while Sp3 and YY1 interact with region II. We have verified that the SNP at position -700 (T>G) is embedded in two common haplotypes, which have different frequencies between European and African descendents. The allele -700G increases transcription and binding for a still unknown transcription factor. SNP -700 affects Sp3 and YY1 interaction with this region, even though it is not part of these transcription factors predicted binding sites. Therefore, our results show for the first time that Sp3 and YY1 interact with human COL18A1 promoter 2, and that nucleotide -700 is part of a binding motif for a still unknown TF that is involved in the expression of this gene in hepatocytes. In addition, we also confirm the involvement of Sp1 in the regulation of this gene.

1. Introduction

Type XVIII collagen is a member of the MULTIPLEXIN collagen family characterized by polypeptide chains with multiple triple-helical domains separated and flanked by non-triple-helical regions (Oh et al., 1994; Rehn and Pihlajaniemi, 1994). This collagen, a homotrimer molecule made of three alpha1 chains, is a heparan sulfate proteoglycan constituent of epithelial and endothelial basement membranes of a wide variety of tissues (Saarela et al., 1998a). Human type XVIII collagen, encoded by the COL18A1 gene, has three isoforms, which differ in their N-terminal non-collagenous domain (NC11), namely: NC11-303, NC11-493 and NC11-728. The COL18A1 gene spans 105Kbp and contains 43 exons and 2 putative promoter regions, namely: promoter 1 located 5' from exon 1, and promoter 2 within intron 2 (Fig. 1) (Saarela et al., 1998b; Elamaa et al., 2003). The NC11-303 isoform is associated with promoter 1 and the other two isoforms have their transcription regulated by the promoter 2. Variant NC11-728 is expressed in various human fetal tissues (Elamaa et al., 2003) and adult brain and retina (Suzuki et al., 2002), while NC11-493 is highly expressed in liver (Saarela et al., 1998b). Liétard et cols (2000) have shown that the mouse Col18a1 promoter 2 is up regulated by Sp1 and HNF3 and, that it has a silencer-like element that may be regulated by HNF3 and NF1/CTF interactions with this regulatory region, but, so far, data on human COL18A1 regulatory regions are limited to the fact that there are two distinct promoters (Saarela et al., 1998b; Elamaa et al., 2003).

Although COL18A1 function is not completely understood, it is well known that its deficiency causes several ocular problems and neuronal cell migration disorders in patients with Knobloch syndrome (Sertié et al., 2000; Suzuki et al., 2002; Kliemann et al., 2003). Type XVIII collagen expression varies in many

pathological conditions such as cancer (Musso et al., 2001), liver fibrosis (Musso et al., 1998), Alzheimer's disease (van Horssen et al., 2002) and diabetic retinopathy (Noma et al., 2002). Also, the 184-aminoacid proteolytic carboxyterminus fragment of Collagen XVIII, called endostatin, has been characterized as a potent endogenous inhibitor of angiogenesis (O'Reilly et al., 1997). Interestingly, there is a wide variation in serum levels of endostatin, mainly originated from NC11-493 proteolysis, in a normal human population (Zorick et al., 2001). The characterization of the regulatory mechanisms underlying the COL18A1 gene expression will help elucidate the variability of Collagen XVIII and endostatin levels both in physical and pathological conditions. Therefore, in the present study we have focused upon characterization of cis-regulatory elements interacting with human COL18A1 promoter 2 and identification of functional SNPs in this region in different ethnic groups.

2. Results

2.1. Functional analysis of the COL18A1 promoter 2

Alignment of approximately 1.5Kb of the human and 3.7Kb of the mouse Collagen XVIII promoter 2 showed five regions with more than 60% identity over a minimum of 40bp (Fig. 1).

To elucidate the role of these conserved regions in the transcription of COL18A1 in hepatocytes, six promoter deletion constructs were generated (Fig. 2). The constructs pGL3-1310, representing deletion of the first half of region V, and pGL3-1052, representing deletion of the second half of region V and region IV, showed no significant ($p=0,793$ and $0,966$ respectively) differences in transcriptional activity when compared with the largest construct, pGL-1487. The pGL3-555 construct caused a significant ($p<0.001$) increase in transcriptional activity, suggesting the presence of cis-repressing elements between nucleotides -1052 and -555 . Removal of the sequence spanning -555 to -458 , which includes conserved region III, decreased the promoter activity ($p=0.005$). Deletion of nucleotides from -458 to -186 produced an increase in promoter activity, but this increase was not statistically significant ($p=0,332$). These results suggest that COL18A1 has a core promoter located between nucleotides -186 and -21 , an enhancer element between nucleotides -555 and -458 and a repressor element between nucleotides -1052 and -555 . The -1487 bp fragment, cloned in the reverse orientation, did not present transcriptional activity (data not shown).

2.2. Nucleoprotein complexes formed by the promoter 2

Since our *in silico* analysis has identified potential binding sites for Sp1/Sp3, HNF3, C/EBP and YY1 in the COL18A1 promoter region (Fig. 1),

EMSAs were performed using probes representing clusters in regions I, II, III and IV (Fig. 1).

Three specific complexes (A, B and C) were formed with probe –94/–46 (Fig. 3A, lane 6). Anti-Sp1 antibody was able to abrogate complex B (Fig. 3A, lane 7) whereas an antibody directed against Sp3 abrogated complexes A and C (Fig. 3A, lane 8).

The –181/-131 probe generated two specific complexes (A and B) (Fig. 3B, lane 5). Antibody directed against YY1 abrogated complex B (Fig. 3B, lane 7) and antibody against Sp3 abrogated complex A (data not shown). No super-shift or complex abrogation was observed when an antibody against Sp1 was used (Fig 3B, lanes 6).

EMSA analysis with probe –1300/–1268, which corresponds to part of conserved region V with binding sites for Sp1/Sp3 (Fig.1), yielded at least two specific complexes (A and B) (Fig. 3C, lane 6). Even though anti-Sp1 antibody did not abrogate any complex, it generated a super-shift (Fig. 3C, lane 7), suggesting that Sp1 interacts with this region. Antibody directed against Sp3 abrogated complex B (Fig. 3C, lane 8).

EMSA analysis with probes –517/-470, -1382/-1346 and –1469/-1430 did not show any super-shift or complex abrogation when antibodies against HNF3 and C/EBP were used (data not shown).

Therefore, Sp1 and Sp3 bind to the COL18A1 promoter 2 from nucleotide –94 to –46, and from –1300 to –1268, while YY1 and Sp3 bind to the region between nucleotides –181 and –131.

2.3. Polymorphisms in the human COL18A1 promoter 2

We have tested the –700 T/G SNP (GenBank accession No. rs755548) in 258 individuals (121 European and 137 African descendents) and observed a significant difference ($p=0.0066$) in genotypic frequencies between the two ethnic groups (Table 1).

Screening of COL18A1 promoter 2 in DNA samples of 48 European and 46 African descendents revealed also 3 other new SNPs, -251 C/T, -1428 C/T and -1528 G/A. All SNPs are in Hardy-Weinberg equilibrium. Seven putative haplotypes were identified in the two ethnic groups (Table 2). Among the European descendents, the most frequent haplotype is Hap1 (54%), with Hap5 and Hap6 being exclusive to this group. African descendents present Hap2 as the most frequent haplotype (55%) and Hap7 is found only in this group. Hap1 and Hap 2, sequences differ only at the –700 T/G SNP.

2.4. Functional analysis of the –700T>G SNP

We selected the -700 SNP to perform functional studies due to its high and different frequency among the two ethnic groups. The pGL3-1487 and pGL3-1487/-700G constructs were used to this end. The pGL3–1487/–700G construct showed an increase of 39% in transcriptional activity when compared to pGL3–1487 ($p<0.01$, Fig. 4).

For EMSA analysis probes –700T and –700G were selected. Both probes produced at least four specific complexes (A, B, C and D) (Fig. 5A, lane 4 and Fig. 5B, lane 4), but the –700G probe formed an evident extra specific complex X (Fig. 5B, lane 4) while a very weak band in the corresponding position was observed with the –700T probe (Fig. 5C, lane 4). Sp3 completely abrogated complex B for

both probes (Fig. 5A and 5B, lanes 6), while YY1 caused abrogation of complex C for both probes (Fig. 5C, lanes 3 and 6).

When the –700T labeled probe was incubated with 20-fold molar excess of the –700G unlabeled probe, complexes B and D were abolished (Fig. 5D, lane 6), while complexes A and C persisted even with 50-fold molar excess of the –700G unlabeled probe (Fig. 5D, lane 7). In contrast, when the –700G labeled probe was co-incubated with 20-fold molar excess of the –700T unlabeled probe, only complex D was abolished (Fig. 5D, lane 13), while the other complexes persisted. Complexes C and D were abolished with 50-fold molar excess of the –700T probe, while complex A, X and B persisted (Fig. 5D, lane 14).

These findings suggest that complex X is specific for the –700G probe. Nuclear proteins interacting with complexes A and D seem to have equal affinity for both probes, while those forming complex B have greater affinity for the –700G probe and those responsible for complex C display greater affinity for the –700T probe. Finally, complex B contains Sp3 and complex C contains YY1.

3. Discussion

The alignment of the human COL18A1 and the mouse Col18a1 promoter 2 sequences revealed five conserved regions, which might be functional if we take into account that conservation across different species of non-coding regions implies in functional constraints (Loots et al., 2000; Woolfe et al., 2004; Nobrega et al., 2004).

Human type XVIII collagen core promoter 2 contains two conserved regions, namely: region II and I. Sp1 and Sp3 bind to the conserved region I (Fig. 1), which contains three Sp1/Sp3 consensus binding site (Fig 1), including a highly conserved CCCAC sequence representing the inverse of the GT1 motif used to clone and identify the Sp3 TF (Hagen et al., 1992). Sp1 and Sp3 are members of the Sp family of TFs, which contain three zinc-fingers that bind GC-boxes and promote transcription of several genes (Suske et al., 1999). Sp1 has been described to strongly activate the mouse Col18a1 transcription in HepG2 cells (Liétard et al., 2000) and to participate in collagen type IV gene expression (Fischer et al., 1993), while Sp3 competes for the same binding site as Sp1 (Hagen et al., 1992), and can either repress or activate a given promoter by recruiting different isoforms (Kennett et al., 1997). Sp3 and YY1 bind to region II (Fig. 1), which includes two conserved binding sites for Sp3, and also a CCTTCTT sequence identical to the consensus YY1 repressor sequence (Shrivastava and Calame, 1994). YY1 is a zinc finger protein that belongs to the GLI-Kruppel gene family (Shi et al., 1991) and participate in the transcription of alpha 1 and alpha 2 collagen type I gene (Miao et al., 1999; Riquet et al., 2001) and genes with TATA-less promoters (Karantzoulis-Fegaras et al., 1999; Johansson et al., 1998). Indeed, a TATA-box was not found in the COL18A1 core promoter region, in

agreement with the mouse TATA-less promoter (Liétard et al., 2000). It is important to notice that the alignment between human and mouse promoter sequence evidence that the two species transcription initiation site and core promoter do not correspond (Fig.1).

Our transfection experiments suggest that COL18A1 has a positive regulatory element between nucleotides –555 and -458, possibly in conserved region III, but EMSA experiments did not confirm the predicted binding sites for C/EBP. Other cis-regulatory elements may be responsible for the increase in transcription promoted by this region.

Sp1 and Sp3 interact with region V between nucleotides –1300 and –1268 (Fig. 1), however, this region together with region IV have no significant impact in the transcriptional regulation of COL18A1 in hepatocytes. Our results also indicate that the sequence comprised between nucleotides –1052 and –555 possibly contains silencer elements for the human liver isoform of Collagen XVIII. The corresponding mouse region, even though displaying low identity with the human sequence, also contains negative regulatory elements (Liétard et al., 2000). The finding that Sp3 binds to nucleotides between –720 and –680 is in accordance with the binding site predicted *in silico* for this TF (Fig. 1). Even though YY1 also binds to this region, our *in silico* analysis did not show any YY1 consensus sequence. We suggest that the GCATTCC sequence, nucleotides from -697 to –690, could be an YY1 inhibitory binding site considering the low stringency level of this TF repressor consensus sequence (Shrivastava and Calame, 1994). Therefore, it will be important to define the roles of Sp3 and YY1 in Collagen XVIII transcriptional regulation as these two TFs can function either as positive or as

negative regulatory factors (Thomas and Seto, 1999; Weill et al., 2003; Li et al., 2004).

No predicted binding sites for HNF3 and C/EBP in probes -1469/-1430 and -1382/-1346, could be confirmed by our EMSA experiments. C/EBP and HNF3 have been described as important TFs for liver-specific gene transcription (Krivan and Wasserman, 2001). HNF3 has also been shown to drive mouse Col18a1 expression in liver cells (Liétard et al., 2000; Fig.1). Maybe there are other binding sites for these TFs in human promoter 2 sequence.

According to our findings, the 1487 bp of COL18A1 promoter 2 presents a SNP frequency of 1/396bp, a frequency intermediary to the SNP frequency that was previously described for intergenic and conserved non-coding region (Banerjee et al., 2002). We have shown in this report that the SNP at position –700 (T>G) is functional. This SNP is embedded in haplotypes Hap1 and Hap2, which differ by only one base, namely, a guanine (G) at position –700 instead of a thymine (T), and showed different frequencies among ethnic groups. These haplotype differences reinforce the notion that individuals with African ancestry can be distinguished from others based on anatomical classification (Parra et al., 2004). One could speculate that haplotype Hap2, which is most frequent in African descendents, is ancestral to Hap1, and that the G to T change at position –700 has led to loss of a positive transcriptional element. In fact, the G allele significantly increased both promoter transcriptional activity and affinity for a particular nuclear protein, resulting in the formation of a new protein-DNA complex. Our EMSA results suggest that Sp3 and YY1 affinity for this region is affected by the SNP -700, even though this SNP is not part of these TFs predicted

binding sites. *In silico* analysis revealed that SNP-700 may create or disrupt diverse TFs binding sites.

In conclusion, our studies suggest that Sp1, Sp3 and YY1 regulate transcription of COL18A1 gene in human liver. Also a polymorphism at position – 700 of promoter 2, whose frequency varies according to the ethnic group, can increase transcription rate and affinity for a particular TF, whose binding site may overlap with those of Sp3 and YY1. Additional studies will be required to better characterize the regulatory factors involved in transcriptional regulation of Collagen XVIII liver isoform, as well as to identify which are the binding factors affected by the presence of the -700G allele.

4. Experimental Procedures

4.1. *In silico* analysis of the human Collagen XVIII promoter 2

We used the Lagan Toolkit (http://lagan.stanford.edu/lagan_web/lagan.shtml) in order to align the candidate promoter 2 of human COL18A1 gene, extending from nucleotides -1487 to +43 (bases from 13126 to 14654, GenBank accession No. BX322561), with murine Col18a1 promoter 2 (bases from 1 to 3751, GenBank accession No. U34607). RVista (<http://genome.lbl.gov/vista/rvista/submit.shtml>) and TESS (<http://www.cbil.upenn.edu/tess>) were used to search for binding sites for TFs Sp1/Sp3, HNF3, C/EBP and YY1. Transcription initiation has been previously described for the human liver isoform at position -21 (GenBank accession No. NM030582) and for the mouse liver isoform at position -311 (Liétard et al. 2000, GenBank accession No. U34607). Nucleotide positions are numbered relative to the first methionine, where the A of the ATG codon is +1 and the base immediately preceding -1 (Den Dunnen and Antonarakis, 2001).

4.2. Plasmid constructs

A PCR product, extending from -1487 to +43 of the COL18A1 promoter 2, was amplified from human DNA, cloned into the pGL3-Basic vector (Promega), and designated as pGL3-1487. Deletional constructs using endonucleases or PCR were generated using pGL3-1487 as a template, namely: pGL3-1310, pGL3-1052, pGL3-555, pGL3-458, pGL3-186 and a pGL3-1487 reverse. pGL3-1487/-700G was generated by PCR amplification and cloning of a DNA sample containing a G nucleotide at position -700. Primers used are listed in Table 3. All plasmid constructs, except for pGL3-1487 reverse, had a variable 5'-end but a common 3'-

end. The integrity of all constructs was assessed by DNA sequence analysis, using an automated Megabace 1000 (GE Healthcare Bio-Sciences).

4.3. Cell culture and transient transfection assays

HepG2 cells were obtained from the American Type Culture Collection (ATCC) and were maintained as described elsewhere (Miao et al., 1999). Twenty-four hours before transfection, cells were plated onto 12-well plates (1×10^5 /well). To detect the promoter activity of the 5'-flanking region of the COL18A1 gene, 1µg of the luciferase reporter plasmid and 0.5µg of the pRL-TK plasmid (Promega) were co-transfected, in duplicate wells, with *Lipofectamine* 2000 (Invitrogen). The pRL-TK plasmid (Promega) was used to normalize the transfection efficiency. Cells were harvested 48h after transfection, and expression of the reporter genes was analyzed using the Dual-Luciferase Reporter Assay System (Promega) in a microplate luminometer (LB96V *Microlumat Plus* - EG&Berthold). The transcriptional activity of the COL18A1 constructs was normalized relative to the activity of the pGL3-1487 construct and expressed as fold change relative to luciferase activity. The luciferase data were generated by pooling the results of at least 3 experiments. One-way ANOVA with Tukey's Multiple Comparison post test was performed to compare the 6 deletion constructs using GraphPad Prism version 4.00 for Windows (GraphPad Software, San Diego California USA, www.graphpad.com). The Mann Whitney test was used to compare activities of pGL3-1487 and pGL3-1487/700G. The minimal significance level adopted was 0.05.

4.4. Eletrophoretic mobility shift assay (EMSA)

Nuclear extracts from HepG2 were prepared as described elsewhere (Read, 1998). Protein concentration was determined by the Bradford method (Bio-Rad) using bovine serum albumin as a standard. Double-stranded DNA probes were prepared by annealing the desired sense and antisense oligonucleotides (Table 4). Sense oligonucleotides were 5' end-labeled with [$\gamma^{32}\text{P}$] ATP and T4 polynucleotide kinase (New England), followed by purification using Sephadex G50 spin columns (GE Healthcare Bio-Sciences). A non-specific competitor (NS Comp) was used to test the specificity of the complexes (Table 4). Protein binding reactions were carried out in 15 μL of buffer (20mM Hepes pH7.9, 80mM KCl, 5mM Mg_2Cl , 2% (v/v) Ficoll, 5% (v/v) Glycerol, 0.1mM EDTA, 1mM Dithiothreitol) with $0.5 - 1 \times 10^6$ cpm of labeled probe, 1 μg of poly dl-dC, 50 μg of Bovine Serum Albumin and 5 μg of HepG2 nuclear protein. The mixtures were incubated for 30 min at room temperature in the presence or in the absence of non-labeled competitor oligonucleotides. For super-shift assays, the binding reaction mixture was pre-incubated with the following Santa Cruz Biotechnology antibodies: Sp1(sc-420), Sp3 (sc-644), HNF3beta (sc-6554), C/EBPbeta (sc-746) and YY1 (sc-1703). The protein-DNA complexes were resolved by electrophoresis (6% non-denaturing polyacrylamide gel) in running buffer 0.5X TBE (44.58mM Tris Base, 44.48mM Boric Acid, 1.25 mM $\text{Na}_2\text{EDTA} \cdot 2\text{H}_2\text{O}$), visualized using the STORM 860 phosphorimager scanner (GE Healthcare Bio-Sciences) and analysed using the Image Quant[®] 5.2 software.

4.5. Subjects

A sample of individuals from the Brazilian population was obtained among healthy blood donors at Hospital das Clínicas São Paulo, Brazil. These individuals consist

of 121 European (mean age 55 ± 10 SD) and 137 African descendents (mean age 45 ± 15 SD), categorized using standard criteria, which include: color, type of hair, nose and mouth formats (Krieger et al., 1965). A standard protocol was used for DNA extraction from blood samples (Miller et al., 1988). The local Ethics Committee approved the study protocol. All samples were obtained upon signed informed consent.

4.6. Genotyping

Analysis of 1583bp 5' of COL18A1 exon 3 was performed employing Denaturing High Performance Liquid Chromatography (Transgenomic) in accordance to manufacturer protocols. Primers used are described on Table 5. Samples showing abnormal peaks were sequenced using an automated Megabace 1000 (GE Healthcare Bio-Sciences). Once a SNP was characterized, screening was carried out using the Single Nucleotide Primer Extension (SNUPE™ GE Healthcare Bio-Sciences) protocol. Gene and genotype frequencies in affected and normal individuals were compared by contingency table analysis using χ^2 statistics and Fisher's exact test. The minimal significance level adopted was 0.05.

4.7. Haplotypes

Using the Arlequin Ver 2.000 software (<http://lgb.unige.ch/arlequin/>) possible population haplotypes were built for the two ethnic groups (47 European and 43 African descendents) and their frequencies were estimated. Linkage disequilibrium and Hardy-Weinberg equilibrium tests were also performed.

Acknowledgments

The authors thank Dr. B. B. Mendonça for laboratory facilities and Dr. Novack and Dr. Bachega for antibodies. The authors are indebted to C. Masotti and O. T. Suzuki for valuable suggestions. FAPESP, CNPq, FINEP and PRP-USP supported this work.

References

- Banerjee, P., Bahlo, M., Schwartz, J.R., Loots, G.G., Houston, K.A., Dubchak, I., Speed, T.P., Rubin EM. 2002. SNPs in putative regulatory regions identified by human mouse comparative sequencing and transcription factor binding site data. *Mamm. Genome* 13, 554-557.
- Den Dunnen, J.T., Antonarakis, S.E., 2001. Nomenclature for the description of human sequence variations. *Hum. Genet.* 109, 121-124.
- Elamaa, H., Snellman, A., Rehn, M., Autio-Harmainen, H., Pihlajaniemi, T., 2003. Characterization of human type XVIII collagen gene and proteolytic processing and tissue location of the variant containing a frizzled motif. *Matrix Biol.* 22, 427-442.
- Fischer, G., Schmidt, C., Opitz, J., Cully, Z., Kuhn, K., Poschl, E., 1993. Identification of a novel sequence element in the common promoter region of human collagen type IV genes, involved in the regulation of divergent transcription. *Biochem. J.* 292, 687-695.
- Hagen, G., Muller, S., Beato, M., Suske, G., 1992. Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. *Nucleic. Acids Res.* 20,5519-5525.

- Johansson, E., Hjortsberg, K., Thelander, L., 1998. Two YY-1-binding proximal elements regulate the promoter strength of the TATA-less mouse ribonucleotide reductase R1 gene. *J. Biol. Chem.* 273, 29816-29821.
- Karantzoulis-Fegaras, F., Antoniou, H., Lai, S.L., Kulkarni, G., D'Abreo, C., Wong, G.K., Miller, T.L., Chan, Y., Atkins, J., Wang, Y., Marsden, P.A., 1999. Characterization of the human endothelial nitric-oxide synthase promoter. *J. Biol. Chem.* 274, 3076-3093.
- Kennett, S.B., Udvardia, A.J., Horowitz, J.M., 1997. Sp3 encodes multiple proteins that differ in their capacity to stimulate or repress transcription. *Nucleic Acids Res.* 25, 3110-3117.
- Kliemann, S.E., Waetge, R.T., Suzuki, O.T., Passos-Bueno, M.R., Rosenberg, S., 2003. Evidence of neuronal migration disorders in Knobloch syndrome: clinical and molecular analysis of two novel families. *Am. J. Med. Genet. A.* 119, 15-19.
- Krieger, H., Morton, N.E., Mi, M.P., Azevedo, E., Freire-Maia, A., Yasuda, N., 1965. Racial admixture in north-eastern Brazil. *Ann. Hum. Genet.* 29, 113-125.
- Krivan, W., Wasserman, W.W., 2001. A predictive model for regulatory sequences directing liver-specific transcription. *Genome Res.* 11, 1559-1566.
- Li, L., He, S., Sun, J.M., Davie, J.R., 2004. Gene regulation by Sp1 and Sp3. *Biochem. Cell. Biol.* 82, 460-471.
- Liétard, J., Théret, N., Rehn, M., Musso, O., Dargère, D., Pihlajaniemi, T., Clément, B., 2000. The promoter region of the long variant of Collagen XVIII, the precursor of endostatin, contains liver-specific regulatory elements. *Hepatology* 32, 1377-1385.

- Loots, G.G., Locksley, R.M., Blankespoor, C.M., Wang, Z.E., Miller, W., Rubin, E.M., Frazer, K.A., 2000. Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* 288, 136-140.
- Miao, K., Potter, J.J., Anania, F.A., Rennie-Tankersley, L., Mezey, E., 1999. Identification of two repressor elements in the mouse alpha 2(I) collagen promoter. *Arch. Biochem. Biophys.* 361, 7-16.
- Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16, 1215.
- Musso, O., Rehn, M., Theret, N., Turlin, B., Bioulac-Sage, P., Lotrian, D., Campion, J.P., Pihlajaniemi, T., Clement, B., 2001. Tumor progression is associated with a significant decrease in the expression of the endostatin precursor Collagen XVIII in human hepatocellular carcinomas. *Cancer Res.* 61, 45-49.
- Musso, O., Rehn, M., Saarela, J., Theret, N., Lietard, J., Hintikka, Lotrian, D., Campion, J.P., Pihlajaniemi, T., Clement, B., 1998. Collagen XVIII is localized in sinusoids and basement membrane zones and expressed by hepatocytes and activated stellate cells in fibrotic human liver. *Hepatology* 28, 98-107.
- Nobrega, M.A., Pennacchio, L.A., 2004. Comparative genomic analysis as a tool for biological discovery. *J. Physiol.* 554,31-39.
- Noma, H., Funatsu, H., Yamashita, H., Kitano, S., Mishima, H.K., Hori, S., 2002. Regulation of angiogenesis in diabetic retinopathy: possible balance between vascular endothelial growth factor and endostatin. *Arch. Ophthalmol.* 120, 1075-1080.
- Oh, S.P., Kamagata, Y., Muragaki, Y., Timmons, S., Ooshima, A., Olsen, B.R., 1994. Isolation and sequencing of cDNAs for proteins with multiple domains of

- Gly-Xaa-Yaa repeats identify a distinct family of collagenous proteins. *Proc. Natl. Acad. Sci. U. S. A.* 91, 4229-4233.
- O'Reilly, M.S., Boehm, T., Shing, Y., Fukai, N., Vasios, G., Lane, W.S., Flynn, E., Birkhead, J.R., Olsen, B.R., Folkman, J., 1997. Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. *Cell* 88, 277-285.
- Parra, E.J., Kittles, R.A., Shriver, M.D., 2004. Implications of correlations between skin color and genetic ancestry for biomedical research. *Nat. Genet.* 36, S54-60.
- Read, M., 1998. Electrophoretic mobility shift assay (EMSA) In: Dockerty K, ed. *Essential Techniques*. New York: Wiley & Sons pp. 7-11.
- Rehn, M., Pihlajaniemi, T., 1994. Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen. *Proc. Natl. Acad. Sci.* 91, 4234-4238.
- Riquet, F.B., Tan, L., Choy, B.K., Osaki, M., Karsenty, G., Osborne, T.F., Auron, P.E., Goldring, M.B., 2001. YY1 is a positive regulator of transcription of the Col1a1 gene. *J. Biol. Chem.* 276, 38665-38672.
- Saarela, J., Rehn, M., Oikarinen, A., Autio-Harmainen, H., Pihlajaniemi, T., 1998a. The short and long forms of type XVIII collagen show clear tissue specificities in their expression and location in basement membrane zones in humans. *Am. J. Pathol.* 153, 611-626.
- Saarela, J., Ylikärppä, R., Rehn, M., Purmonen, S., Pihlajaniemi, T., 1998b. Complete primary structure of two variant forms of human type XVIII collagen and tissue-specific differences in expression of the corresponding transcripts. *Matrix Biol.* 16, 319-328.
- Sertié, A.L., Sossi, V., Camargo, A.A., Zatz, M., Brahe, C., Passos-Bueno, M.R., 2000. Collagen XVIII, containing an endogenous inhibitor of angiogenesis and

- tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome) *Hum. Mol. Genet.* 9, 2051-2058.
- Shi, Y., Seto, E., Chang, L.S., Shenk, T., 1991. Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. *Cell.* 67, 377-388.
- Shrivastava, A., Calame, K., 1994. An analysis of genes regulated by the multi-functional transcriptional regulator Yin Yang-1. *Nucleic Acids Res.* 22, 5151-5155.
- Suske, G., 1999. The Sp-family of transcription factors. *Gene.* 238, 291-300.
- Suzuki, O.T., Sertie, A.L., Der Kaloustian, V.M., Kok, F., Carpenter, M., Murray, J., Czeizel, A.E., Kliemann, S.E., Rosemberg, S., Monteiro, M., Olsen, B.R., Passos-Bueno, M.R., 2002. Molecular analysis of collagen XVIII reveals novel mutations, presence of a third isoform, and possible genetic heterogeneity in Knobloch syndrome. *Am. J. Hum. Genet.* 71, 1320-1329.
- Thomas, M.J., Seto, E., 1999. Unlocking the mechanisms of TF YY1: are chromatin modifying enzymes the key? *Gene* 236, 197-208.
- van Horssen, J., Wilhelmus, M.M., Heljasvaara, R., Pihlajaniemi, T., Wesseling, P., de Waal, R.M., Verbeek, M.M., 2002. Collagen XVIII: a novel heparan sulfate proteoglycan associated with vascular amyloid depositions and senile plaques in Alzheimer's disease brains. *Brain Pathol.* 12, 456-62.
- Weill, L., Shestakova, E., Bonnefoy, E., 2003. Transcription factor YY1 binds to the murine beta interferon promoter and regulates its transcriptional capacity with a dual activator/repressor role. *J. Virol.* 77, 2903-2914.
- Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W.,

Edwards, Y.J., Cooke, J.E., Elgar, G., 2004. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 3:e7

Zorick, T.S., Mustacchi, Z., Bando, S.Y., Zatz, M., Moreira-Filho, C.A., Olsen, B., Passos-Bueno, M.R., 2001. High serum endostatin levels in Down syndrome: implications for improved treatment and prevention of solid tumours. Eur. J. Hum. Genet. 9, 811-814.

		Probe -1469/-1430		Region V
human: -1487	AAATCACAGCAGCAGCAGT	TTTCAAGCAAGAGACTTCTCTGGGCATCTCTGAGAC	-1428	
>>>>			<<<<	
mouse: -3634	AAATCACATTAGGA----	AATCTTCAAGCACACTG--TTCCTGGGCATTCTCAAGC	-3583	
		Probe -1382/-1346		
human: -1427	GAGTTTGAAGCAGTGTGGGGATTCTGTTTGGGAGCACGGCTCAGT	GACATGCAGATGTTT	-1368	
>>>>			<<<<	
mouse: -3582	AGACGCAATCATTGTGGGGACTAACTTGGAGATGTAGAGCTCACAGGCA--	CAGATGTGG	-3525	
		Probe -1300/-1268*		
human: -1367	GCTTGTGGAATGCCTGCGCCTG	CCCTCGAGAAGGAGGTGGCACTCAGCTGCTTGAACAA	-1308	
>>>>			<<<<	
mouse: -3524	GCCTATGG-----	CTGCCTCACTCTTAGAGAGGGACTTGGTATCCAGCTGTTTGCAGCTG	-3470	
		Probe -1300/-1268*		
human: -1307	GGCTGGTCCCAAGCAGCCAGCCCTGCCCTGTGGACAGCAATCGA	CCGATGGCTACAGC	-1248	
>>>>			<<<<	
mouse: -3469	GACTCTGAGCCAAACGGCACAGCCCTGCCCTTTAGACAGGCATTGATTGATGTGATGT-		-3411	
		Region IV		
human: -1247	ATGTCCTGACCTGGGACAGCTGGGGACAGGT-TGTGGGGCTGTTTCGTACAGCTGCTG		-1189	
>>>>			<<<<	
mouse: -3410	-----	GGGACAGTTGGGGACATGCCTTTGTGGCTGTTTCATCACA----	-3368	
		Region IV		
human: -1188	TCACATTAACCTGACCCAGCCCTGGGGAGGGTGGCTGCAACCCAGCCGCAC---	AGCC	-1132	
>>>>			<<<<	
mouse: -3367	AAGCATTTGCC--	ACTGCAGCCTTGGGGAGGGTGGCAGTAACCTGTCCACACCACAGCT	-3310	
		Region IV		
human: -1131	TGGAGGAGGCGAGGTGACGAGGGACTCCAGCTGGGGAGCTGGCGTGACAGCACTGAGG		-1072	
>>>>			<<<<	
mouse: -3309	TGGAGAAAGTGGAGCCAGGAGGGACAGCTGGCTGAGGCTGTAGCCTGGTACATGCTACAT		-3250	
		Region IV		
human: -732	TTGTGGACGGGGAGACTGA-----		-714	
>>>>			<<<<	
mouse: -1462	CCGTGGATGTTGACTTTGATTAGATACAGAACGTCTCTGAAATTCGAAACCCAGCTACA		-1403	
		Probe -700*		
human: -713	-----	GGCGGGAGGAGTATG	-697	
>>>>			<<<<	
mouse: -1402	TATCCAGAAGTCTTGATGTTGGAGGGGCTGGGAAGGCATGTGGTCAGGTGGCCAGGCAA		-1343	
		Region IV		
human: -696	-----	CATTCTGCTCCCTGGTCCCGGGCTCCGCTCTCTGCTGGGTCTCCT	-654	
>>>>			<<<<	
mouse: -1342	GAGGCAGGAGGAGACTCCCTTCCCTTCCCCCCCCCCCCCCCCCGTGGGAGTTCAT		-1283	
		Region III		
		Probe -517/-470		
human: -517	-----	CCATTGCAACAGTGGAAATGGCACTGCAGCAAGC	-484	
>>>>			<<<<	
mouse: -502	GGACACCCAGCCTTCTCTAAATCAACAACCTGTAAACACAAGAATTGGGACTGCAGCAAGC		-443	
		Region III		
human: -483	TGAGCATTTTTTCGGTTCCAAGTTTCCCAACAACAATGGCTCCAGGCCAGGCGCCTCTG		-424	
>>>>			<<<<	
mouse: -442	TGAGCGTTCTTCAAACCTG-GGTTTCATCATAGAATGGTGGCCAGCAGCAGG-ACTTCCG		-385	
		Region III		
human: -423	GGCGTTTCAGGTGAGGGGGAGCAGGGGCTGGGGCCGCTGCCTCAGTCTTCTGGGAT		-364	
>>>>			<<<<	
mouse: -384	GGTGTTC-----	TCCATGGAGGCCCTAGGATCTGGGGCGCGGTTTCCATGTCCCCGAG--	-331	
		Region III		
human: -363	GGGTCTCTGGCTTGGAGCCCAATCCA-----	GAGACTCAGCAGCCCTGC	-319	
>>>>			<<<<	
mouse: -330	--GACCTTGGCCAGGCCCCACGCCACACCCACTCTCAGCAGAGACTCAGCAGCTCTA-		-274	
		Region II		
human: -318	CCCGGAGAGCCTAGCGGAGCAGTTGGGAACCTGTGGACACTGCACAGCAGCCCT-C		-260	
>>>>			<<<<	
mouse: -273	-----	AGTTACACAAGCTGGAGTTCGGCTGGGTGTCTGTTACCCCATCTTTC	-227	
		Region II		
human: -259	TGTCCTCAGGAGCCTGGGGCTGCTGCAAGGCCGTGGGAAGAAGTTATAAATCCTTGGG		-200	
>>>>			<<<<	
mouse: -226	TGCTCTGGCTGCTCTGGAGTGCAGCCAGG-----	AAGGAGTTAAA-----	-181	
		Region II		
human: -199	TGCTGGAACCCCGGGTGCCTGGG-CACCTTGATTCCAAGGCTGGTGCCTTCTTCTCTCT		-141	
>>>>			<<<<	
mouse: -180	ATCTGAAGGTACCGTTTGGTGGATCACCTTGATTCTCAGGCTGGG--	TTCATTCCCTTCT	-123	
		Region I		
human: -140	GCAAGAGTGGGGGTGGAAGACAGCCGGCCAGAGGCTGGGAAGCCTGCACAGGGAGGGC		-81	
>>>>			<<<<	
mouse: -122	GCCAGAGGGAAGCGGTGGAG-----	GGAAGGAGACTGGGAACC-----	-77	
		Region I		
human: -80	AGGAGGGCTGAGCCTGGCTAGGCCACCTCCAGGCACAGAGGCCCTCCGCCGCCGCA		-21	
>>>>			<<<<	
mouse: -76	GGGAGGGTGGGGACCTGGCAGGCCACCTCCAGGACCACAGGAGCC--	GGCCCAGCGCA	-17	
		Region I		
human: -20	GCTCCAGCCGACTGCCCGATGGCTCCCTACCCCTGTGGTGCCACATCTGTGCTGCTGC		+40	
>>>>			<<<<	
mouse: -18	GAGGC--	TCTCACTGCCCTGATGGCTCCGACCCAGCAGACGCTCTGCTGCTGCTGC	+40	

Figure 1. Major features and *in silico* analysis of bases –1487 to +40 of the human COL18A1 promoter. Alignment between human and mouse Collagen XVIII promoter 2 showing in boxes conserved regions with more than 60% identity over a minimum 40bp in windows of 40bp. EMSA probes used are in gray boxes and Sp1/Sp3, HNF3, C/EBP and YY1 predicted binding sites are underlined. The mouse HNF3 and NF1/CTF binding sites are also underlined. Probes with super-shift or complex abrogation have an asterisk. Transcription initiation is indicated with an arrow for human (-21) and mouse (-311) sequences, the –700 SNP are indicated in a box. Nucleotide positions are numbered relative to the first methionine, where the A of the ATG codon is +1 and the base immediately preceding –1.

Figure 2. Functional analysis of human COL18A1 promoter 2. Graphic representation showing activity profiles of the reporter luciferase 5'-3' deletion constructs in HepG2 cells, with a schematic representation of conserved regions deleted in each construct. Data from these transient transfections are expressed as fold luciferase activity relative to pGL3-1487 and represent mean ± S.E.. Asterisks corresponds to p<0.001.

A

B

Cold -181/-131 probe	-	-	+	+	-	-	-
Non-specific competitor	-	-	-	-	+	-	-
Anti Sp1	-	-	-	-	-	+	-
Anti YY1	-	-	-	-	-	-	+
	1	2	20x	100x	100x	6	7

C

Figure 3. Electrophoretic Mobility Shift Experiments with double-stranded ^{32}P end-labeled probes. (A) Labeled probe -94/-46 was incubated in the absence (lane 1) or in the presence of HepG2 cells nuclear extract (lane 2) under the following conditions: 20-, 100- and 200-fold molar excess of unlabeled -94/-46 probe (lanes 3 to 5), 100-fold molar excess of non-specific unlabeled probe (lane 6) or antibodies against Sp1, Sp3 and YY1 (lanes 7 to 9). (B) Labeled

probe –181/-131 was incubated in the absence (lane 1) or in the presence of HepG2 cells nuclear extract (lane 2) under the following conditions: 20-, 100-fold molar excess of unlabeled –181/-131 probe (lanes 3 and 4), 100-fold molar excess of non-specific unlabeled probe (lane 5) or antibodies against Sp1 and YY1 (lanes 6 and 7). (C) Labeled probe –1300/-1268 was incubated in the absence (lane 1) or in the presence of HepG2 cells nuclear extract (lane 2) under the following conditions: 20-, 100- and 200-fold molar excess of unlabeled –1300/-1268 probe (lanes 3 to 5), 100-fold molar excess of non-specific unlabeled probe (lane 6) or antibodies against Sp1, Sp3 and YY1 (lanes 7 to 9).

Figure 4. SNP-700 functional analysis. Activity profiles of *COL18A1* promoter reporter luciferase construct pGL3-1487, with base T at position -700, and pGL3-1487/700G, with base G at position –700. Data from these HepG2 transient transfections are expressed as fold luciferase activity relative to pGL3-1487 and represent mean ± S.E. Asterisks corresponds to $p < 0.01$.

A

B

Figure 5. Electrophoretic Mobility Shift Experiments with labeled probes –700T and –700G. A) Double-stranded, ^{32}P end-labeled –700T probe incubated in absence (lane 1) or in the presence of HepG2 cells nuclear extract (lane 2) under the following conditions: a 100-fold molar excess of unlabeled –700T probe (lane 3), 100-fold molar excess of non-specific unlabeled probe (lane 4) or antibodies against Sp1 and Sp3 (lanes 5 and 6). B) Double-stranded, ^{32}P

end-labeled –700G probe incubated in absence (lane 1) or in the presence of HepG2 cells nuclear extract (lane 2) under the following conditions: a 100-fold molar excess of unlabeled –700G probe (lane 3), 100-fold molar excess of non-specific unlabeled probe (lane 4) or antibodies against Sp1 and Sp3 (lanes 5 and 6). C) Double-stranded, ³²P end-labeled –700T and –700G probes incubated in absence (lanes 1 and 4) or in the presence of HepG2 cells nuclear extract (lane 2 and 5) and anti-YY1 antibody (lanes 3 and 6) D) Double-stranded, ³²P end-labeled –700T and –700G probes incubated in absence (lanes 1 and 8) or in the presence of HepG2 cells nuclear extract (lane 2 and 9), –700T probe was incubated in the presence of a 1-, 5-, 10-, 20- and 50-fold molar excess of unlabeled –700G probe (lanes 3 to 7) and –700G probe was incubated in the presence a 1-, 5-, 10-, 20- and 50-fold molar excess of unlabeled –700T probe (lanes 10 to 14).

Table 1. SNP–700 genotype distribution between European and African descendents

Genotypes	European descendents	African descendents	Total
GG	18 (15%)	41 (30%)	59 (23%)
TG	66 (55%)	70 (51%)	136 (53%)
TT	37 (31%)	26 (19%)	63 (24%)
Total	121 (100%)	137 (100%)	258 (100%)

Table 2. Putative haplotypes in *COL18A1* promoter 2

Haplotype	-251	-700	-1428	-1546	% European descendents	% African descendents
Hap1	C	T	C	G	0.53790	0.39535
Hap2	C	G	C	G	0.32979	0.54651
Hap3	T	T	C	G	0.06849	0.02326
Hap4	C	G	C	A	0.03657	0.01163
Hap5	C	G	T	G	0.02128	0.00000
Hap6	T	G	C	A	0.00598	0.00000
Hap7	C	T	T	G	0.00000	0.02326

Table 3. Primer pairs used to construct vectors

Region	Primer Set
-1487 to +43	5'- GCA GAT CTA AAT CAC AGC AGC AGC - 3' 5'- TAA AGC TTA GAG CAG CAG CAG GAT G - 3'
-1310 to +43	5'- ATA GAT CTC AAG GCT GGG TC - 3' 5'- TAA AGC TTA GAG CAG CAG CAG GAT G - 3'
-1052 to +43	5'- GCA GAT CTT GAA ACC TGG GTG TGT - 3' 5'- TAA AGC TTA GAG CAG CAG CAG GAT G - 3'
- 458 to +43	5'- GCA GAT CTC CCA CAA CAC AAT GGC - 3' 5'- TAA AGC TTA GAG CAG CAG CAG GAT G - 3'

*Bold sequence refers to restriction endonucleases sites

Table 4. Oligonucleotides used in EMSA analyses

Probe	Sequence (5'-3'; sense strand)
-94/-46	5'- GCA CAG GGG AGG GCA GGA GGG CGT GAG CCT GGC TAG CCC CAC CTC CAG G - 3'
-181/-131	5'- TGG GCA CCT TGA TTC CAA GGC TGG TGC CTT CTT TCC TTC TGC AAG AGT GG - 3'
-517/-470	5'- CCA TTG CAA CAG TGG AAT TGG CAC TGC AGC AAG CTG AGC ATT TTT TCG - 3'
-1300/-1268	5'- TCC CAA GCA GCC CAG CCC TGC CCT GTG GAC AGC - 3'
-1382/-1346	5'- GAC ATG CAG ATG TTT GCT TGT GGA ATG CCT GCG CCT G - 3'
-1469/-1430	5'- TCA GTT TTC AAG CAA AGA GAC TTC TCT GGG CAT CTC TGA G - 3'
-700T	5'- AGA CTG AGG CCG GGA AGG AGT ATG CAT TCC TGC CCT GGT CC - 3'
-700G	5'- AGA CTG AGG CCG GGA AGG AGG ATG CAT TCC TGC CCT GGT CC - 3'
NS Comp	5'- CAA GGC AAG AGA TAA CAC AAA GTC AAG GTG AAG ATC A - 3'

Table 5. Primer pairs used to screen the COL18A1 promoter 2

Region	Primer Set
-210 to +43	5'- TAA ATC CTT GGG TGC TGG AAC C - 3' 5'- AG AGC AGC AGC AGG ATG TGG - 3'
-383 to -139	5'- TGC CTC AGT GCT TCT GGG ATG - 3' 5'- GCA GAA GGA AAG AAG GCA CCA G - 3'
-580 to -323	5'- AGC AGC TCT GAG AGG GTC TGA G - 3' 5'- GGC TGC TGA GTC TCT GGA TTG - 3'
-746 to -500	5'- CTA ACC GGA TCC TTT TGT GG - 3' 5'- GCC CAA GTC GTT AAA CAT CC - 3'
-865 to -629	5'- AGG GCA TAA GGC ACA CTC AG - 3' 5'- CTC AGC CCT TCC TGA CCT TC - 3'
-1027 to -780	5'- GAT GCA CAT GTG TGA ACG TG - 3' 5'- GGA TGT GAG TGG CAT GTG AG - 3'
-1262 to -972	5'- ACC GAT GGC TAC AGC ATG TC - 3' 5'- AGC CAG GCA GTT TAC ACA GG - 3'
-1444 to -1203	5'- CTG GGC ATC TCT GAG ACG AG - 3' 5'- AAA CAG CCC CAC AAC CTG TC - 3'
-1583 to -1406	5'- CCA CAA CAG AGT TCC CTT CC - 3' 5'- AAT CCC CAC ACT GCT CAA AC - 3'

Further characterization of COL18A1 promoter 2: Binding of Oct-1 and effect of the SNP-700T/G in the transcription levels of collagen XVIII and predisposition to HCC

Lúcia Maria Armelin-Correa ¹, Flávia Imbrosi Errera¹, Venâncio Alves², José Eymard de Medeiros³, Flair José Carilho³, Venâncio Alves⁴, Delphine Quelard⁵, Orlando Musso⁵, Maria Rita Passos-Bueno¹

¹Human Genome Center, Department of Genetics and Evolution Biology , Bioscience Institute, University of São Paulo, São Paulo, Brazil;

²Department of Pathology, Hospital das Clínicas da FMUSP, University of São Paulo, São Paulo, Brazil;

³Gastroenterology Division, Hospital das Clínicas da FMUSP, University of São Paulo, São Paulo, Brazil;

⁴Pathology Division, Hospital das Clínicas da FMUSP, University of São Paulo, São Paulo, Brazil;

⁵INSERM U620, Rennes I University, Rennes, France

Corresponding author:

Dr. Maria Rita Passos-Bueno

Centro de Estudos do Genoma Humano – Depto. de Biologia

Instituto de Biociências - Universidade de São Paulo

Rua do Matão, 277 sala 200

São Paulo, SP, Brazil CEP: 05508-900

Telephone: +55-11-3091-9910 FAX: +55-11-3091-7419

E-mail address: passos@ib.usp.br

Abbreviations: HCC – Hepatocellular Carcinoma, SNP – Single Nucleotide Polymorphism, Oct-1 – Octamer binding transcription factor 1, qPCR – quantitative Polymerase Chain Reaction, EMSA – Eletrophoretic Mobility Shift Assay, SNUPE – Single Nucleotide Polymorphism Extension

Keywords

collagen XVIII; transcriptional regulation; functional SNP; Oct-1

Resumo

O colágeno XVIII é um componente da membrana basal de vários órgãos tais como o fígado, os olhos e o cérebro. Alterações na expressão do colágeno XVIII foram descritas em várias doenças hepáticas como a fibrose/cirrose e o HCC, mas sua regulação da transcrição é ainda pouco conhecida. Nós caracterizamos previamente a ligação de Sp1/Sp3 e YY1 com o promotor 2 de *COL18A1* e identificamos um SNP funcional na posição -700T/G, que está inserido em dois haplótipos, com diferentes frequências entre descendentes de europeus e africanos. Neste estudo nos prosseguimos com a caracterização do promotor 2 de *COL18A1* descrevendo a ligação do fator de transcrição Oct-1 a região do SNP (bases de -680 a -720) e a deleção desta seqüência leva a uma atividade do promotor semelhante a do alelo G em HepG2. Mostramos, também, que o SNP-700 está em desequilíbrio de ligação com um SNP não sinônimo do domínio frizzled do colágeno XVIII, SNPc.1135C/T. Nós não encontramos correlação do SNP-700T/G com susceptibilidade a HCC ou expressão do colágeno XVIII em fígado fibrótico/cirrótico.

Abstract

Collagen XVIII is a basement membrane component of several tissues such as liver, eyes and brain. Changes in collagen XVIII expression have been described in several hepatic diseases such as fibrosis/cirrhosis and HCC, but its transcriptional regulation is still poorly understood. We have previously characterized that Sp1/Sp3 and YY1 interact with *COL18A1* promoter 2 and identified a functional SNP at position -700T/G, which is embedded in two haplotypes, with different frequencies between European and African descendants. In this study we further characterized *COL18A1* promoter 2 by describing that the transcription factor Oct-1 binds the SNP-700 sequence (bases from -680 to -720) and deletion of this sequence leads to a comparable promoter transcription activity as the G allele in HepG2. We also show that SNP-700 is in linkage disequilibrium with a non-synonymous SNP in collagen XVIII frizzled domain, SNPC.1135C/T. We did not find any correlation of SNP-700T/G with HCC susceptibility or, with collagen XVIII expression in fibrotic/cirrhotic liver tissue.

1. Introduction

Collagen XVIII, a component of the endothelial and epithelial basement membrane of several tissues, including liver (Saarela et al., 1998a, Musso et al., 2001a), is a heparan sulfate proteoglycan and member of the Multiplexin (multiple triple-helix domains with interruptions) family of collagens. The proteolytic cleavage of its C-terminal region liberates a 184-aminoacid fragment, called endostatin, a potent inhibitor of endothelial cell migration and proliferation, and, therefore, angiogenesis (Oh et al., 1994; Rehn and Pihlajaniemi, 1994; O'Reilly et al., 1997). Collagen XVIII is transcribed by the *COL18A1* gene, located on chromosome 21q22.3, which has 43 exons spanning along 105Kb (Saarela et al., 1998a and 1998b). In the liver sinusoids this collagen is expressed as two isoforms: NC11-493 and NC11-728, which are transcribed by promoter 2, localized upstream exon 3. The pre-mRNA suffers an alternative splice of part of exon 3 giving rise to the NC11-493 isoform. The non spliced mRNA is responsible for the NC11-728 isoform which exclusively contains 235 aminoacids of the Frizzled like domain (Elamaa et al., 2003). Quelard and cols, 2006, have shown that collagen XVIII frizzled domain was able to inhibit Wnt- β -catenin signalling pathway *in vitro*. The Wnt- β -catenin signalling pathway is involved in the progression of several types of cancer, including the hepatocellular carcinoma HCC (de La Coste et al., 1998).

Collagen XVIII and endostatin expression levels were also associated with HCC progression (Dhar et al., 2002; Musso et al, 2001a; Musso et al, 2001b). The genetic model currently accepted for HCC is of a complex disease with several environmental and genetic factors involved (Feo et al., 2006, Kato et al., 2005). Many genes would participate in the development of the disease

and variations at each of these genes would have a small contribution to its susceptibility or progression. Single nucleotide polymorphisms (SNPs) are an important source of genetic variation, leading to subtle phenotypic changes which are important for species evolution (Wray, 2007) and also for the development of complex diseases. It would thus be important to evaluate if SNPs in *COL18A1* represent at-risk alleles for HCC. Our group has described two SNPs, SNP-700T/G (Armelin-Correa et al., 2005) and SNPc.1135C/T (Errera et al., submitted, Anexo III), which are apart by less than 2000 bp in human collagen type XVIII core promoter and coding region, respectively. The G allele of SNP -700T/G confers increased transcription rate *in vitro* for *COL18A1* promoter 2 in HepG2 cell line (Armelin-Correa et al., 2005), while SNPc.1135C/T involves the substitution of a conserved threonine, at codon 379, in the collagen XVIII frizzled domain, to a methionine (Errera et al., submitted, Anexo III).

Therefore the objectives of this study were: to further analyse the role of SNP-700 in collagen XVIII expression *in vitro* and *in vivo*, to verify the functionality of SNPc.1135 in colon cancer cell line, to verify if these two loci are in linkage disequilibrium and to analyse their association with HCC.

2. Materials and Methods

2.1. Patients/control samples and DNA extraction

Blood samples of 90 Brazilian patients with HCC and 77 controls were obtained at Hospital das Clínicas São Paulo, Brazil. The samples of HCC patients consist of 74 European descendants (16 females, mean age 60 ± 16 SD; 58 males, mean age 57 ± 12 SD) and of 16 African descendants (4 females, mean age 56 ± 10 SD; 12 males, mean age 57 ± 6 SD). The samples of controls consist of 40 European descendants (23 females, mean age 68 ± 9 SD; 17 males, mean age 66 ± 7 SD) and of 37 African descendants (14 females, mean age 63 ± 13 SD; 23 males, mean age 50 ± 14 SD). The ethnic groups were defined on the basis of self-classification and subjective classification (skin color, nose and lip shapes, hair texture and information about family ancestry). A standard protocol was used for DNA extraction from blood samples (Miller et al., 1988). The local Ethics Committee approved the study protocol. All samples were obtained upon signed informed consent.

Non-tumor liver tissues of 35 French patients, with European ancestry, with hepatocellular carcinoma were obtained after informed consent, in compliance with French laws and the national Ethics Committee: 19 of these patients underwent hepatectomy and 16 underwent liver resection. Their mean age (32 men and 3 women) was 60.2 ± 9 years. The samples were flash frozen in nitrogen-cooled isopentane at the earliest possible time after resection, and stored at -80°C as previously described (Musso et al., 1998). Total DNA was isolated from distant non-tumor liver tissue from the 35 French patients using DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA) following manufacturer protocol and stored at -20°C .

2.2. Genotyping

SNPs c.1135C/T (Thre379Met) and -700T/G were analyzed by SNUPE (*Single Nucleotide Primer extension*) (SNUPE™ GE Healthcare BioSciences) method, for Brazilian samples. The corresponding DNA regions of *COL18A1* gene were amplified using PCR buffer, 0.2mM dNTPs, 0.6μM primer and 0,2U Taq polimerase, in 35 cycles with 100 ng of DNA, according to the following steps: 5 minutes denaturation at 95°C, 10 minutes extension at 72°C, and using the following primers and respective annealing temperature: c.1135C/T exon 3F: 5'-CCCAACCACCTCCACCACGAGA-3' and exon 3R: 5'-CTGGATGAACCGCAGCCGATGT-3' (SNPc.1135 - 66°C); -865F 5'-AGGGCATAAGGCACACTCAG-3' and -629R 5'-CTCAGCCCTTCCTGACCTTC-3'; (SNP-700 - 60°C). PCR products were submitted to SNUPE reaction and then analyzed in automatic sequencer (MegaBACE™ 1000 DNA Analysis Systems), using the correspondent *software* and following the User's guide instructions (GE Health Science) and 2 pmols of either oligonucleotide c.1135C>T: 5'-CGAGGAAGGGGTGGCAGTGC-3', or -700: 5'-GGACCAGGGCAGGAATGCAT-3'.

The genotyping of SNP-700T/G, for French patients, was performed using Applied Biosystems designed probe for SNP rs755548, 5'-CGGGGAGACTGAGGCCGGGAAGGAG[G/T]ATGCATTCCTGCCCTGGTCCC GGGC-3'. Reaction was performed using 0.050μg de DNA, 7.5 μl of Taq Man Universal Mix (Applied Biosystems), 0.75 μl 20X Assay Mix (Applied Biosystems) and q.s.q. MilliQ water to 15.0 μl of reaction on an ABI Prism 7300

sequence detection system (Applied Biosystems). Allelic discrimination was performed with the ABI Prism 7300 SDS software (Applied Biosystems).

Allelic and genotype frequencies in affected and normal individuals were compared by contingency table analysis using χ^2 statistics and Fisher's exact test. The minimal significance level adopted was 0.05.

2.3. Quantitative RT-PCR

Total RNA was isolated from the 35 French samples distant non-tumor liver tissue using SV Total RNA Isolation System (Promega, Madison, WI) following manufacturer protocol and stored at -80°C. 1ug of RNAs were reversed-transcribed into cDNA using a High-Capacity cDNA Archive kit (Applied Biosystems, Foster City, CA). RT products were diluted in water to 200 ul and 15 ng of cDNA were used for quantitative PCR (qPCR) reaction. Primers were designed with Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and selected sequences were as follows: NC11-493 forward 5'-TCCAGCTGTGGAATGACACT-3', NC11-493 reverse 5'-AGTCTCCAGGACCAGGGTTT-3', NC11-728 forward 5'-GCTTCTCTCTCCTCCTTGCTG-3', NC11-728 reverse 5'-CAACAAGCCCGTAGGAACAT-3', 18S forward 5'-GTAACCCGTTGAACCCCAT-3', 18S reverse 5'-CCATCCAATCGGTAGTAGCG.

Primer pairs were tested by dissociation curve to warrant specificity in each reaction tube (absence of primer dimers and other non-specific products). qPCR was performed using intercalation of SYBR Green I (Applied Biosystems), on an ABI Prism 7300 sequence detection system (Applied

Biosystems). The amplification curves were read with the ABI Prism 7300 SDS software (Applied Biosystems) using comparative cycle threshold method. Normalisation of the threshold cycle (Ct) of the genes of interest was done against the relative quantity of 18S and expressed as $\Delta Ct = Ct \text{ gene of interest} - Ct \text{ 18S}$. Relative gene expression among patients as $\Delta\Delta Ct = \Delta Ct \text{ gene of interest} - \Delta Ct \text{ calibrator gene}$ were then calculated as $2^{-\Delta\Delta Ct}$, following published methodology (Livak and Schmittgen, 2001). Each reaction was performed in duplicate and variability in ΔCt was inferior to 10%. RT product from the cell line HepaRG was used as calibrator. HepaRG cell line was kindly provided by Dr. C. Guguen-Guillouzo and cultured as previously described (Gripon et al., 2002). Cells were plated and 2 weeks after, when reaching confluence, the RNA was extracted as previously described.

2.4 Haplotypes

Using the SNPalyseTMDynacon Ver 6.000 software possible population haplotypes were built for the two ethnic groups and their frequencies were estimated. Linkage disequilibrium and Hardy-Weinberg equilibrium tests were also performed. Level of significance adopted was 0.05.

2.5. Plasmids

Plasmids pGL3-1487T, pGL3-1487G, pGL3-186 were previously constructed by us (Armelin-Correa et al., 2005). pGL3-Basic (Promega, Madison, WI) was used as empty Firefly luciferase vector. Construct pGL3-700Del was generated by site direct mutagenesis. Deletion of 41bp (Genebank accession number BX322561, deleted bases from 13893 to 13933),

corresponding to the EMSA probes -700T (Armelin-Correa et al., 2005), was achieved using PCR primer pairs amplifying pGL3-1487T construction from bases from -1487 to -720 and bases from -680 to +43 in two independent PCR reactions. Amplified DNA from each PCR reaction (reaction 1: from -1487 to -720, reaction 2: from -680 to +43) was purified and another PCR reaction was performed using as template products from reaction 1 and reaction 2. The amplified product, without bases from -720 to -680, was purified and ligated in a pGL3-Basic plasmid (Promega). Plasmid DNAs were purified on Qiagen columns (Qiagen, Valencia, CA) and quantified by optical density₂₆₀. Construct was named pGL3-700Del and was confirmed by restriction digestion with HindIII and BglII (Promega) and DNA sequencing.

The expression vector pSecTag-FZ was kindly provided by Dr.T. Pihlajaniemi (University of Oulu, Finland) and construct pSecTag-FZ379 was generated by us by site direct mutagenesis. The TCF-luciferase reporter vector Super-8-TOP (S8+), a TCF-responsive reporter containing 8 copies of the wild type TCF-binding site, was kindly provided by Pr.R. Moon (University of Washington School of Medicine, Seattle, WA). The SFRP5 expression vector was kindly provided by Dr. S. Baylin (John Hopkins Medical Institutions, Baltimore, Maryland) (Suzuki et al., 2004). The plasmids pSecTag (Invitrogen) and pcDNA3.1/HisA (His) (Invitrogen) were used as empty expression vector.

2.6. Cell Culture and Transfection

HepG2, human hepatoma cell line, and HCT116, human colon cancer cell line, both from American Type Cell Culture ATCC, were maintained in DMEM medium from Invitrogen supplemented with 10% FCS, 50 units/ml

penicillin, 50 ug/ml streptomycin. HepG2 transfection was done as previously described (Armelin-Correa et al., 2005).

Hep3B, human hepatoma cell line, from American Type Cell Culture ATCC, and Huh7, human hepatoma cell line, provided by INSERM U620, were maintained in Williams' E medium from Invitrogen supplemented with 10% FCS, 50 units/ml penicillin, 50 ug/ml streptomycin, 1mg/ml albumin, 1 ug/ml insulin, 2mM glutamine and 5×10^{-4} hydrocortisone hemisuccinate. Twenty-four hours before transfection, Hep3B and Huh7 cell lines were plated onto 24-well plates in density of 0.3×10^5 cells/well. To detect the promoter activity of the 5'-flanking region of the *COL18A1* gene, 0.025 µg of the luciferase reporter plasmid was transfected, in triplicate wells, with Lipofectamine and Plus Reagent (Invitrogen). 0.002 µg of the Renilla luciferase pRL-Null plasmid (Promega) was used to normalize the transfection efficiency. Cells were harvested 48h after transfection, and expression of reporter genes was analysed using the Dual-Luciferase Reporter Assay System (Promega). The transcriptional activity of the *COL18A1* constructs was normalized relative to the activity of the pGL3-Basic plasmid and expressed as fold change relative to luciferase activity. The luciferase data was generated by pooling the results of at least 2 experiments in triplicates.

HCT116 cell line was plated onto 24-well plates in density of 0.5×10^5 cells/well twenty-four hours before transfection. To detect the S8+ promoter activity 15 ng of the luciferase reporter plasmid and 250ng of expression plasmids pSecTag-FRZ, pSecTag-FRZ379, His-Sfrp5 as well as the empty expression vectors pSecTag and His, were co-transfected, in triplicate wells, with Lipofectamine and Plus Reagent (Invitrogen). Transfection was normalized

by $\mu\text{g}/\mu\text{l}$ of protein by the Lowry method (Lowry et al., 1951). Cells were harvested 48h after transfection, and expression of reporter genes was analysed using the Dual-Luciferase Reporter Assay System (Promega). The transcriptional activity was normalized relative to the activity of the either empty vector pSecTag or His, according to the cloning vector used for each protein expression, and expressed as fold change relative to luciferase activity. The luciferase data was generated by pooling the results of at least 2 experiments in triplicates. Luciferase activity was analysed performing Unpaired t test using Graph Pad (<http://www.graphpad.com/quickcalcs/>) and the minimal significance level adopted was 0.05.

2.7. Electrophoretic mobility shift assay (EMSA)

Nuclear extracts from HepG2 were prepared as described elsewhere (Read, 1998). Protein concentration was determined by the Bradford method (Bio-Rad) using bovine serum albumin as a standard. Double-stranded DNA probes were prepared by annealing the desired sense and antisense oligonucleotides -700T and -700G previously described (Armelin-Correa et al., 2005). Sense oligonucleotides were 5'-end labeled with [$\gamma^{32}\text{P}$] ATP and T4 polynucleotide kinase (New England), followed by purification using Sephadex G50 spin columns (GE Healthcare BioSciences). A non-specific competitor (NS Comp) (Armelin-Correa et al., 2005) was used to test the specificity of the complexes. Protein binding reactions and electrophoresis were carried out as previously described (Armelin-Correa et al., 2005). For super-shift assays, the following Santa Cruz Biotechnology antibodies were used: Ets-1 (c-20)X, PEA3 (h-120)X, Oct-1 (c-21)X.

3. Results

3.1. Functional analysis of the of SNP-700T/G and OCT1 binding site

We have previously described a polymorphism in human collagen XVIII promoter 2 region, SNP-700 T/G, that is functional in HepG2 cell line (Armelin-Correa et al, 2005). To confirm the difference in transcription rate of the two alleles, we transfected the promoter 2 constructs of *COL18A1* containing the different genotypes, pGL3-1487T and pGL3-1487G, in Hep3B and Huh7 cell lines (Fig 1A, B). Transfection of Hep3B with the G allele construct showed a significant ($p<0.001$) higher luciferase activity than with the T allele, while in Huh7 the G allele had a significant ($p<0.001$) lower luciferase activity than the T allele.

To further explore the functionality of SNP-700T/G, we have deleted a sequence corresponding to 41bp, including the SNP-700, from construct 1487T. This new construction was called Del700 and it showed a luciferase activity significantly ($p<0.001$) higher than construction 1487T in HepG2 transfection (Fig.2).

We then searched for putative transcription factors binding sites in this 41bp deleted sequence with either allele T or G at position -700. *In silico* analysis indicated that the T allele would preferentially bind Oct-1, while the G allele would preferentially bind PEA3 and Ets-1. We performed EMSA analysis with probes -700T and -700G. Both probes produced at least four specific complexes (A, B, C and D) (Fig. 3 A, B) and probe -700G the complex X previously described (Armelin-Correa et al., 2005). Oct-1 produced an extra complex just over complex D in both probes (Fig. 3 A, B) while the other two antibodies, Ets1 and Pea3, had no effects in any of the complexes. This result

suggests that Oct-1 binds the SNP-700 sequence and that this binding is not abolished or altered by allele G.

3.2. Functional analysis of the of SNPc.1135C/T

To perform a functional study of SNPc.1135C/T we transfected the HCT116 cell line, which is originated from a human colon carcinoma and has a constitutively activated Wnt- β -catenin signalling pathway, with a S8+ vector and an expression vector containing either the SNPc.1135 C allele, named vector FRZ, or the SNPc.1135 T allele, named vector FRZ379. An expression vector containing the Sfrp5 sequence was used as a positive control of Wnt- β -catenin signalling pathway inhibition. The FRZ construction inhibited the Wnt- β -catenin signalling pathway as previously described (Quelard et al., 2006) and we found no significant difference in the activity of the FRZ379 construction (Fig. 4). This result suggests that the SNPc.1135C/T has no evident effect in collagen XVIII frizzled domain capacity to inhibit the Wnt- β -catenin signalling pathway in HCT116 cell line.

3.3. Linkage disequilibrium between SNP-700T/G and SNPc.1135C/T

We have genotyped 40 European descendants and 37 African descendants for SNP-700T/G and SNPc.1135C/T. Four putative haplotypes were identified in the two ethnic groups (Table 1). Hap1 is the most common haplotype among European descendants and Hap2 is the most common among African descendants. Hap4 could only be found among African descendants. The two SNPs are in linkage disequilibrium in both populations (European descendants $D'=1$; African descendants $D'=0.93$).

3.4. Frequency and expression of SNP-700T/G among HCC patients

As SNP-700T/G and SNPe.1135C/T showed evident linkage disequilibrium and only SNP-700T/G has been proved until now to be functional, we have genotyped 109 HCC patients with European ancestry (74 from Brazil and 35 from France) and 16 patients with African ancestry for SNP-700T/G. We found that patients' genotypic and allelic frequencies are very similar to that of controls (Table 2), therefore SNP-700T/G seems not to be associated with predisposition to HCC.

We next evaluated if the SNP-700 T/G was associated with *in vivo* transcriptional levels of *COL18A1*. We have performed qPCR to analyse the expression of NC11-493 and NC11-728 in non-tumor liver samples of 35 of these HCC patients with European ancestry. GG samples showed higher mean NC11-728 levels than TT samples, but not with statistical significance (Table 3).

4. Discussion

Collagen XVIII is a heparan sulfate proteoglycan that generates at least two matricryptins, namely endostatin and frizzled fragment, which influence angiogenesis and cell proliferation respectively (O'Reilly et al., 1997; Elamaa et al., 2003; Quelard et al., 2006). SNPs that interfere in the function or quantity of these two proteins might have important functional physiological consequences for individuals depending on the environmental insults that they are exposed to. We have thus analysed the functionality of two collagen XVIII SNPs one in promoter 2, intron 2, SNP-700T/G, and another in the frizzled domain, exon 3, SNPC.1135C/T.

Functional studies of SNP-700T/G have previously shown that allele G, more frequent among African descendants, presents an increased transcription rate in HepG2 cell line when compared to allele T (Armelin-Correa et al., 2005). In this present report, we further confirm that the SNP-700T/G contributes to *in vitro* variability of expression levels of the collagen XVIII isoforms. Interestingly, the effect of this nucleotide substitution is not always associated with a positive regulation in the transcription (Fig. 1 A, B), as this effect was not observed in Huh7 cells. Curiously, Huh7 cell line, which also derived from hepatoma, has higher expression of collagen XVIII than HepG2 and Hep3B (Musso O, personal communication). These data suggests that these different hepatoma cell lines have different expression of transcription factors affecting *COL18A1* transcriptional regulation.

We also observed that deletion of the SNP-700 segment led to an increased transcription rate in HepG2 cell line. This result thus confirms that this region contains an important regulatory element, and also suggests that the T

allele creates a binding site for a transcription factor N, with negative effects on collagen XVIII promoter 2 in HepG2 cell line (Fig. 2). Since the luciferase activity from construction Del700 was similar to that previously described for construction pGL3-1487G (Armelin-Correa et al., 2005), we hypothesize that this binding site for N is abolished or attenuated by the SNP-700 G allele, thus conferring this allele an increased transcription rate in HepG2 cell line. Maybe Huh7 cell line does not express this transcription factor N, but another transcription factor with positive effect in the promoter activity and stronger affinity for SNP-700T allele than for G allele. The construction Del700 should be also transfected in Huh7 and Hep3B cell lines to further explore these hypothesis.

Our results also show that Oct-1 binds the SNP-700 sequence regardless of which allele is at position -700, probably at the motif GIATGCAT at collagen XVIII promoter 2 positions -702 to -694. Oct-1 is a ubiquitous transcription factor member of the homeodomain group of transcription factors. Oct-1 has been previously described to cooperate with HNF-1 to activate the transcription of the Hepatitis B virus preS1 promoter in hepatoma cell line (Zhou and Yen, 1991).

To perform SNPC.1135C/T functional studies we transfected collagen XVIII Frizzled domain expression vectors in a colon cancer cell line with constitutively activated Wnt- β -catenin signalling pathway. Our results suggest that the SNPC.1135C/T has no evident effect in collagen XVIII frizzled domain capacity to inhibit the Wnt- β -catenin signaling pathway in a colon cancer cell line (Fig. 4). Even though, we found a negative result with this cell line we believe that it would be important to test the functional effect of this SNP in

other cell differentiation models that are highly dependent on Wnt- β -catenin signaling pathway and with high collagen XVIII expression.

We analysed the haplotype frequencies and linkage disequilibrium between these two SNPs, SNP-700T/G and SNPc.1135C/T, in samples of European and African descendants. The haplotypes formed by these SNPs have different frequencies between the two ethnic groups and one haplotype was exclusively found among African descendants (Table 1). Although rare haplotypes, as the haplotype Hap4 here reported, may inflate the D' value, the SNP -700T/G and SNPc.1135C/T are so close to each other, less than 2000bp, that the size of the analyzed samples are adequate to evaluate the linkage disequilibrium between these loci. The D' values found in both populations for these SNPs are at the highest and lowest limit (European descendants $D'=1$; African descendants $D'=-0.93$) strongly suggesting that the SNP-700T/G and SNPc.1135C/T are in linkage disequilibrium in these populations and that their most frequent haplotypes are different.

The genotypic and allelic frequencies of SNP-700T>G were very similar between European descendants from Brazil or France (Brazilians: GG= 18%, CT=51%, TT=31%; Frenchs: GG=17%, CT=40%, TT=43%, $p=0.09$), and we also did not observe significant difference between controls and HCC. Therefore, our data suggest that SNP-700T/G is not associated with HCC clinical manifestation (Table 2). The lack of correlation between the SNP -700 alleles and *in vivo* collagen XVIII transcription levels in hepatocytes is in agreement with this negative association studies finding, and suggests that this SNP may have an *in vivo* low effect in collagen XVIII expression.

Therefore this study reinforces the functionality of SNP-700T/G in collagen XVIII transcriptional regulation and demonstrates that Oct-1 binds this SNP region. Also we show that SNP-700 T allele is in linkage disequilibrium with SNPC.1135 C allele and this haplotype is more frequent among European descendants, while the haplotype SNP-700 G allele with SNPC.1135 T allele is more frequent among African descendants. Further studies should elucidate these haplotypes functions in collagen XVIII physiological and pathological roles.

Acknowledgments

The authors are indebted to Dr Elise Lavergne, C. Masotti and O. T. Suzuki for valuable suggestions. FAPESP, CNPq, FINEP and PRP-USP supported this work.

5. References

- Armelin-Correa, L. M., C. J. Lin, *et al.* Characterization of human collagen XVIII promoter 2: interaction of Sp1, Sp3 and YY1 with the regulatory region and a SNP that increases transcription in hepatocytes. Matrix Biol, v.24, n.8, Dec, p.550-9. 2005.
- De La Coste, A., B. Romagnolo, *et al.* Somatic mutations of the b-catenin gene are frequent in mouse and human hepatocellular carcinomas. Proc. Natl. Acad. Sci., v.95, July, p.8847-51, 1998.
- Dhar, D. K., T. Ono, *et al.* Serum endostatin predicts tumor vascularity in hepatocellular carcinoma. Cancer, v.95, n.10, Nov 15, p.2188-95. 2002.
- Elamaa, H., A. Snellman, *et al.* Characterization of the human type XVIII collagen gene and proteolytic processing and tissue location of the variant containing a frizzled motif. Matrix Biol, v.22, n.5, Sep, p.427-42. 2003.
- Feo, F., M.R. De Miglio, *et al.* Hepatocellular carcinoma as a complex polygenic disease. Interpretative analysis of recent developments on genetic predisposition. Biochim. et Biophys. Acta, n.1765, p.126-47, 2006.

Gripon, P., S. Rumin, *et al.* Infection of a hepatoma cell line by hepatitis B virus. Proc Natl Acad Sci U S A, v.99, n.24, p.15655-60. 2002.

Kato, S., T. Tajiri, *et al.* Genetic polymorphisms of aldehyde dehydrogenase 2, cytochrome p450 2E1 for liver cancer risk in HCV antibody-positive japanese patients and the variations of CYP2E1 mRNA expression levels in the liver due to its polymorphism. Scand J Gastroenterol, v.38, n.8, Aug, p.886-93. 2003.

Livak, K. J. e T. D. Schmittgen. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods, v.25, n.4, Dec, p.402-8. 2001.

Lowry, O. H., N. J. Rosebrough, *et al.* Protein measurement with the Folin phenol reagent. J Biol Chem, v.193, n.1, Nov, p.265-75. 1951.

Miller, S. A., D. D. Dykes, *et al.* A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res, v.16, n.3, Feb 11, p.1215. 1988.

Musso, O., M. Rehn, *et al.* Collagen XVIII is localized in sinusoids and basement membrane zones and expressed by hepatocytes and activated stellate cells in fibrotic human liver. Hepatology, v.28, n.1, Jul, p.98-107. 1998.

Musso, O., N. Theret, *et al.* Tumor hepatocytes and basement membrane-Producing cells specifically express two different forms of the endostatin precursor, collagen XVIII, in human liver cancers. Hepatology, v.33, n.4, Apr, p.868-76. 2001a.

Musso, O., M. Rehn, *et al.* Tumor progression is associated with a significant decrease in the expression of the endostatin precursor collagen XVIII in human hepatocellular carcinomas. Cancer Res, v.61, n.1, Jan 1, p.45-9. 2001b.

Oh, S. P., Y. Kamagata, *et al.* Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of collagenous proteins. Proc Natl Acad Sci U S A, v.91, n.10, May 10, p.4229-33. 1994.

O'reilly, M. S., T. Boehm, *et al.* Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. Cell, v.88, n.2, Jan 24, p.277-85. 1997.

Quelard, D., E. Laverne, *et al.* The frizzled module of COL18A1 induces tumor cell death through inhibition of the Wnt-B-catenin signaling pathway. Abstracts for American Society of Liver Disease. Hepatology Oct, p. 588A. 2006.

Read, M. Electrophoretic mobility shift assay (EMSA) In: Dockerty K, ed. Essential Techniques. New York, Wiley & Sons p. 7-11, 1998.

Rehn, M. e T. Pihlajaniemi. Alpha 1(XVIII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen. Proc Natl Acad Sci U S A, v.91, n.10, May 10, p.4234-8. 1994.

Saarela, J., M. Rehn, *et al.* The short and long forms of type XVIII collagen show clear tissue specificities in their expression and location in basement membrane zones in humans. Am J Pathol, v.153, n.2, Aug, p.611-26. 1998a.

Saarela, J., R. Ylikarppa, *et al.* Complete primary structure of two variant forms of human type XVIII collagen and tissue-specific differences in the expression of the corresponding transcripts. Matrix Biol, v.16, n.6, Jan, p.319-28. 1998b.

Suzuki, H., D. N. Watkins, *et al.* Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. Nat Genet. 36: 417-22 p. 2004.

Wray, G. A. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet, v.8, n.3, Mar, p.206-16. 2007.

Zhou, D. X. e T. S. Yen. The ubiquitous transcription factor Oct-1 and the liver-specific factor HNF-1 are both required to activate transcription of a hepatitis B virus promoter. Mol Cell Biol, v.11, n.3, Mar, p.1353-9. 1991.

Fig. 1. Functional analysis of SNP-700 in human *COL18A1* 1487pb promoter 2 region. Graphic representation showing fold of luciferase activity of reporter constructs pGL3-1487T and pGL3-1487G. Hep3B (**A**) and Huh7 (**B**) cells were transfected with 0.025 μ g of pGL3-1487T or pGL3-1487G constructs and 0.002 μ g of pRL-Null vector. Firefly luciferase activity was normalized to Renilla luciferase activity of each sample. Data from 2 transient transfections in triplicates are expressed as fold luciferase activity relative to 0.025 μ g of pGL3-Basic, and represent mean $\pm$ SD.

A.

Fig.2. Functional analysis of SNP-700 in human *COL18A1* promoter 2. Graphic representation is showing the activity profiles of construct Del700, carrying a deletion of 40bp including the SNP-700, and construct 1487T. Luciferase activity was normalized by Renilla Luciferase for each sample. Data from these two HepG2 transient transfections in triplicates are expressed as fold luciferase activity relative to pGL3-1487T and represent mean $\pm$ SD.

A.

Fig.3. EMSA assay with labeled probes -700T and -700G. **A.** Double-stranded, ^{32}P end-labeled -700T probe incubated in absence (line 1) or presence of HepG2 cells nuclear extract (line 2) under the following conditions: a 100 molar excess of unlabeled NS Comp probe (lane 3), antibody against PEA3 (lane 4), Oct1 (lane 5) and Ets1 (lane 6). Asterisk shows supershift of complex D at the presence of antibody against Oct-1. **B.** Double-stranded, ^{32}P end-labeled -700G probe incubated in absence (line 1) or presence of HepG2 cells nuclear extract (line 2) under the following conditions: a 100 molar excess of unlabeled NS Comp probe (lane 3), antibody against PEA3 (lane 4), Oct1 (lane 5) and Ets1 (lane 6). Asterisk shows supershift of complex D at the presence of antibody against Oct-1.

Fig.4. Functional analysis of SNPc.1135 in human *COL18A1*. Graphic representation is showing the activity profiles of reporter S8+ vector in HCT116 cell line transiently transfected with pSecTag-FRZ, pSecTag-FRZ379, His-Sfrp5, Empty Vector His and Empty Vector pSec. Luciferase activity was normalized to ug/ul of protein for each sample. Data from 2 transfections in triplicates are expressed as fold luciferase activity relative to the respective empty vector and represent mean $\pm$ SD.

Table 1

Putative haplotypes in *COL18A1* promoter 2 and exon 3, for 40 European descendants and 37 African descendants.

Haplotypes	SNP-700	SNPc.1135	% European descendants	% African descendants
Hap 1	T	C	0.57500	0.40415
Hap 2	G	T	0.36250	0.48523
Hap 3	G	C	0.06250	0.09585
Hap 4	T	T	0.00000	0.01477

Table 2

SNP-700 T/G genotype distribution among Brazilian HCC patients and controls

Genotypes	Controls European ancestry	Patients European ancestry	Controls African ancestry	Patients African ancestry
GG	4 (10 %)	19 (17 %)	13 (35 %)	6 (37 %)
TG	26 (65 %)	52 (48 %)	17 (46 %)	7 (44 %)
TT	10 (25 %)	38 (35 %)	7 (19 %)	3 (19 %)
Total	40 (100 %)	109 (100 %)	37 (100%)	16 (100 %)

Table 3

SNP-700 genotype and NC11-493 and NC11-728 expression values among 35 patients

Genotypes	Patients European ancestry	Genotypic Frequency	C11-493	C11-728
GG	6	17%	3.42 ± 1.62	5.82 ± 2.49
TG	14	40%	3.30 ± 1.37	4.53 ± 2.41
TT	15	43%	3.39 ± 1.47	4.06 ± 2.48
Total	35	100%		

C11-493 and C11-728 expression values are represented by the mean and standard deviation in each genotype.

C/EBP α and C/EBP β expressions correlate with that of NC11-493 in HCC and cirrhotic liver, and C/EBP β upregulates *COL18A1* transcription *in vitro*.

Lucia Armelin-Correa², Maria Rita Passos-Bueno², Delphine Quelard¹, Bruno Turlin³, Bruno Clément¹, Orlando Musso¹.

¹INSERM U620, Rennes I University, Rennes, France

²Human Genome Center, Department of Genetics and Evolution Biology , Bioscience Institute, University of São Paulo, São Paulo, Brazil;

³Anatomic Pathology Laboratory, Pontchaillou Hospital, Rennes.

Corresponding author:

Dr. Maria Rita Passos-Bueno

Centro de Estudos do Genoma Humano – Depto. de Biologia

Instituto de Biociências - Universidade de São Paulo

Rua do Matão, 277 sala 200

São Paulo, SP, Brazil CEP: 05508-900

Telephone: +55-11-3091-9910 FAX: +55-11-3091-7419

E-mail address: passos@ib.usp.br

Abbreviations: HCC – Hepatocellular Carcinoma, C/EBP – CAAT enhancer binding protein, qPCR – quantitative Polymerase Chain Reaction, CHIP - Chromatin Immunoprecipitation

Keywords

collagen XVIII; C/EBPs; transcriptional regulation; liver cirrhosis; liver fibrosis; hepatocellular carcinoma

Resumo

C/EBP α e C/EBP β são fatores de transcrição enriquecidos no fígado e envolvidos com o ciclo celular e a tumorigênese dos hepatócitos. Nós descrevemos previamente prováveis sítios de ligação de *CCAAT/enhancer-binding proteins* (C/EBPs) no promotor 2 de *COL18A1* (Armelin-Correa et al., 2005). Nós mostramos nesse trabalho que a expressão de C/EBP α e C/EBP β está correlacionada com a de NC11-493 em fígado fibrótico/cirrótico e em HCC, e ambos fatores de transcrição se ligam ao promotor 2 de *COL18A1*. Nossos resultados sugerem que a expressão de NC11-493 em fígado fibrótico/cirrótico não prediz sua expressão no tumor e confirma a correlação inversa entre a expressão de NC11-493 e tamanho do tumor. Ensaios de luciferase mostraram que o sítio C/EBP β do promotor 2 de *COL18A1* responde a doses crescentes do vetor de expressão de C/EBP β . Uma vez este sítio mutado a atividade transcricional deste promotor mínimo e sua transativação por C/EBP β ficam seriamente prejudicadas em linhagens de hepatócitos. Estes resultados mostram o papel importante dos C/EBPs na transcrição de *COL18A1* e confirmam a correlação entre a expressão desse gene e a progressão do HCC.

Abstract

C/EBP α and C/EBP β are liver enriched transcription factors involved in hepatocytes cell cycle and tumorigenesis. We have previously described putative binding sites for CCAAT/enhancer-binding proteins (C/EBPs) in *COL18A1* promoter 2 region (Armelin-Correa et al., 2005). We show in this report that C/EBP α and C/EBP β expression correlate with that of NC11-493 in fibrotic/cirrhotic and HCC liver samples, and that both transcription factors bind the *COL18A1* promoter 2 region. Our results also suggest that NC11-493 expression in fibrotic/cirrhotic liver is not predictive of its expression in HCC and confirms the previously described inverse correlation between tumor size and NC11-493 expression in hepatocellular carcinoma. Luciferase assays showed that the C/EBP β binding site in *COL18A1* core promoter 2 region responds to increasing doses of C/EBP β expression vector and, once this binding site is mutated, the core promoter transcription capacity and its transactivation by C/EBP β are significantly impaired in hepatocytes cell lines. These results thus point out the important role of C/EBPs in *COL18A1* transcription and confirm the correlation between this gene expression and HCC progression.

1. Introduction

Collagen XVIII is an extracellular matrix protein that, together with type XV collagen, forms the subgroup of multiplexins (multiple triple-helix domains with interruptions) within the collagen family (Oh et al., 1994; Rehn and Pihlajaniemi, 1994). Human collagen XVIII gene (*COL18A1*), located on 21q22.3, has 43 exons spanning along 105Kbp and has two promoter regions responsible for the transcription of 3 isoforms, which differ only by their N-terminal non-collagenous domain. Isoform NC11-303 is associated with promoter 1, localized upstream to exon 1, and contains all 43 exons but exon 3. Isoforms NC11-493 and NC11-728 are associated with promoter 2, localized upstream exon 3, and only exons 1 and 2 are not included in their transcripts. An alternative splice of the mRNA sequence, responsible for 235 aminoacids of the Frizzled like domain, is the only difference between the NC11-493 and the NC11-728 isoforms. All three isoforms may suffer proteolytic cleavage of their C-terminal region liberating a 184-aminoacid fragment, called endostatin, a potent inhibitor of angiogenesis, endothelial cell migration and proliferation (O'Reilly et al., 1997). The three isoforms differ in tissue expression. NC11-303 isoform is present in the endothelial and epithelial basement membrane of several tissues such as kidney and muscle (Saarela et al., 1998a). NC11-728 expression was described in fetal liver, lung, skeletal muscle, spleen, and adult skin (Elamaa et al, 2003). NC11-493 isoform has a striking high expression in fetal and adult liver (Saarela et al., 1998a and 1998b). Null mutations in *COL18A1*, leading to deficiency of NC11-303 or of all isoforms, cause severe ocular problems and occipital encephalocele, which are the main clinical features of Knobloch Syndrome (KS), a rare autosomal recessive condition. However, it is very interesting that KS patients do not have any liver or kidney

major malformations or functional deficiencies despite absence of this collagen and endostatin (Sertié et al, 2000; Suzuki et al., 2002). It is possible that the phenotypic effect of the lack of the collagen XVIII in some organs, such as liver and kidney, depends on the presence of adverse injuries, such as infection and cancer development.

Hepatocytes are the major source of collagen XVIII in normal liver, forming a continuous layer along the perisinusoidal spaces. These cells, under fibrogenic stimulus, increase *COL18A1* transcription levels (Musso et al., 1998; Jia et al., 2001). Liver fibrosis, incited by chronic liver injuries caused by ethanol abuse, HCV/HBV infection, cholestasis or metabolic syndromes, may progress to cirrhosis, and most of the hepatocellular carcinoma (HCC) cases occur in association with liver fibrosis/cirrhosis (Di Bisceglie et al, 1998). Musso and cols, 2001(b) have described that HCC patients with higher *COL18A1* expression had better prognosis. The environmental or genetic factors involved in the regulation of the *COL18A1* expression levels under these circumstances are still unknown.

We have previously described putative binding sites for CCAAT/enhancer-binding proteins (C/EBPs) in *COL18A1* promoter 2 region (Armelin-Correa et al., 2005). C/EBPs (C/EBP α , C/EBP β , C/EBP δ , C/EBP γ , C/EBP ϵ , and C/EBP ζ) are members of the basic region leucine zipper subfamily of transcription factors and must form homo- or heterodimers to stabilize the protein-DNA complex (Ramji & Foka, 2002). These liver enriched transcription factors are involved in hepatocytes cell cycle and tumorigenesis (Schrem et al., 2002). Studies have suggested the involvement of C/EBP α in liver tumorigenesis (Xu et al, 2001; Tan et al, 2005). Tomizawa and cols have shown that C/EBP α and C/EBP β were downregulated in HCC (Tomizawa et al, 2003).

Therefore, our goals were to investigate if there is a correlation between *COL18A1* expression level and C/EBP α and C/EBP β in human fibrotic/cirrhotic liver and HCC samples and, if so, to determine if C/EBP α and C/EBP β are involved in NC11-493 and NC11-728 transcriptional regulation.

2. Materials and Methods

2.1. Patients and tissue samples

Non-tumor and tumor liver tissues of patients with hepatocellular carcinoma were obtained after informed consent, in compliance with French laws and the national Ethics Committee. We obtained 75 human hepatic tissues from 55 patients (49 males and 6 females, mean age 60 ± 9 years SD). The liver fibrogenic activity could be identified for 49 patients, according to standard criteria (Anthony et al., 1978): 2 had severe fibrotic activity, 9 moderate, 16 mild and 22 patients had no fibrotic activity. The 2 patients with severe fibrotic activity and the 9 patients with moderate activity were grouped and classified as intense fibrotic activity. The METAVIR score classified the 75 samples: 44 cirrhosis, 9 fibroses, 2 histologically normal and 20 matched hepatocellular carcinoma samples (Bedossa et al, 1996). Samples are listed in Anexo III. The samples were flash frozen in nitrogen-cooled isopentane at the earliest possible time after resection, and stored at -80°C as previously described (Musso et al, 1998).

2.2. Quantitative RT-PCR

Total RNA was isolated from distant non-tumor and tumor liver tissue using SV Total RNA Isolation System (Promega, Madison, WI) following manufacturer protocol and stored at -80°C. 1ug of RNAs were reversed-transcribed into cDNA using a High-Capacity cDNA Archive kit (Applied Biosystems, Foster City, CA). RT products were diluted in water to 200 ul and 15 ng of cDNA were used for quantitative PCR (qPCR) reaction. Primers were designed with Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and selected sequences are in Table 1.

Primer pairs were tested by dissociation curve to warrant specificity in each reaction tube (absence of primer dimers and other non-specific products). qPCR was performed using intercalation of SYBR Green I (Applied Biosystems), on an ABI Prism 7300 sequence detection system (Applied Biosystems). The amplification curves were read with the ABI Prism 7300 SDS software (Applied Biosystems) using comparative cycle threshold method. Normalisation of the threshold cycle (Ct) of the genes of interest was done against the relative quantity of 18S and expressed as $\Delta Ct = Ct \text{ gene of interest} - Ct \text{ 18S}$. Relative gene expression among patients as $\Delta\Delta Ct = \Delta Ct \text{ gene of interest} - \Delta Ct \text{ calibrator gene}$ were then calculated as $2^{-\Delta\Delta Ct}$, following published methodology (Livak and Schmittgen, 2001). Each reaction was performed in duplicate and variability in ΔCt was inferior to 10%. RT product from the cell line HepaRG was used as calibrator. HepaRG cell line was kindly provided by Dr. C. Guguen-Guillouzo and cultured as previously described (Gripon et al., 2002). Cells were plated and 2 weeks after, when reaching confluence, the RNA was extracted as previously described.

2.3. Plasmids

Plasmids pGL3-186wt was previously constructed by us (Armelin-Correa et al., 2005). The pGL3-Basic (Promega) plasmid was used as empty Firefly luciferase vector. Construct pGL3-186mut was generated by site direct mutagenesis of pGL3-186. Mutation of the C/EBP β binding site was performed by means of two independent PCR reactions with primers 186F (GCTAGCCCGGGTGCTGGGCAC) and 186R (TAAAGCTTAGAGCAGCAGCAGGATGTGGC), flanking the -186 and +43 fragment of pGL3-186wt and, primers 186mutF (AGGCTCCGGAGCCTGCACAGGGGA) and 186mutR (TCCGGAGCCTCTGGCCGGCTGTCTT), which overlap the C/EBP β binding site and contain base substitutions changing the target sequence into a AccIII restriction site. Amplified DNA from PCR reactions (reaction 1: primers 186F and 186mutR, reaction 2: primers 186R and 186mutF) were purified and another PCR reaction was performed using as template products from reaction 1 and reaction 2. The amplified DNA, with the C/EBP β mutated site was digested with XmaI and HindIII (Promega), purified and ligated in a pGL3-Basic plasmid (Promega). Plasmid DNAs were purified on NucleoBond Midiprep kit columns (Macherey-Nagel, Easton, PA) and quantified by optical density. Construct was named pGL3-186mut and was confirmed by restriction digestion with AccIII (Promega) and DNA sequencing. The expression vector MSV- C/EBP β was kindly provided by Dr.S.L. McKnight (University of Texas Southwestern Medical Center at Dallas, USA) and pcDNA3 (Invitrogen) was used as empty expression vector.

2.4. Cell Culture and Transfection

Hep3B and Huh7 cell lines were maintained in Williams' E medium from Invitrogen supplemented with 10% FCS, 50 units/ml penicilin, 50 ug/ml streptomycin, 1mg/ml albumine, 1 ug/ml insulin, 2mM glutamine and 5×10^{-4} hydrocortisone hemisuccinate. Hek293T cell line was maintained in Dulbecco's Modified Media (D-MEM) low glicose from Invitrogen supplemented with 10% FBS, 50 units/ml penicilin, 50 ug/ml streptomycin.

Twenty-four hours before transfection Hep3B and Huh7 cell lines were plated onto 24-well plates in density of 0.3×10^5 cells/well, and Hek293T in density of 0.75×10^5 cells/well. To detect the promoter activity of the 5'-flanking region of the COL18A1 gene, 0.025 μ g of the luciferase reporter plasmid and varying amounts of MSV-C/EBP β plasmids (0-0.050 μ g), as well as the required pcDNA3 vector dose to complete always 0.050 μ g of total plasmid transfected, as indicated in the figures, were co-transfected, in triplicate wells, with Lipofectamine and Plus Reagent (Invitrogen). 0.002 μ g of the Renilla luciferase pRL-Null plasmid (Promega) was used to normalize the transfection efficiency when cells were not co-transfected with MSV- C/EBP β plasmids. When cells were co-transfected with MSV-C/EBP β plasmids transfection was normalized by ug/ul of protein by the Lowry method (Lowry et al., 1951). Cells were harvested 48h after transfection, and expression of reporter genes was analysed using the Dual-Luciferase Reporter Assay System (Promega). The transcriptional activity of the COL18A1 constructs was normalized relative to the activity of the pGL3-Basic plasmid and expressed as fold change relative to luciferase activity. The luciferase data was generated by pooling the results of at least 2 experiments in triplicates.

2.5. Western Blot

Hep3B cells were transfected with increasing doses of the MSV-C/EBP β expression vector (0 μ g, 0.010 μ g, 0.025 μ g and 0.050 μ g) and the required pcDNA3 vector dose to complete always 0.050 μ g of total plasmid transfected. Cells were washed by phosphate-buffer saline. Cells were harvested with a lyses buffer (b-glycerophosphate 60mM, MOPS 25mM pH 7.2, EGTA 15mM, MgCl₂ 15mM, NaF1mM, DTT 2mM, Sodium orthovanate 1mM, PMSF 1mM and protease inhibitor 1%) and soniquated. The protein concentration was determined by Lowry method (Lowry et al., 1951). For C/EBP β protein detection 30 μ g of protein was separated by 7.5% bis-acrylamide gel electrophoresis and transferred to a nitrocelulose membrane. After transfer, the membrane was blocked at room temperature for 1h30min with 3% bovine serum albumin in PBST (1x phosphate-buffer saline and 0.01% Tween 20). The membrane was then incubated with anti-C/EBP β sc-150 (1/500) (Santa Cruz, Santa Cruz, California) in 3% bovine serum albumin in PBST overnight at cold room. After incubation with horseradish peroxidase-conjugated secondary antibody at room temperature for 1 hour, the bands were developed using SuperSignal West Dura Extended Duration Substrate (Pierce, Rockford, IL).

2.6. *In silico* analysis of the human collagen XVIII promoter

We used Transcription Element Search System TESS (<http://www.cbil.upenn.edu/cgi-bin/tess/tess>) in order to search for binding sites for C/EBPs in pGL3-186. Nucleotide positions are numbered relative to the first methionine, where the A of the ATG codon is +1 and the base immediately preceding -1 (Den Dunnen & Antonarakis, 2001).

2.7. Statistical Analysis

Statistical Analysis of qPCR data were performed using the software Analyse-it® (<http://www.analyse-it.com>). Luciferase activity was analysed using Graph Pad (<http://www.graphpad.com/quickcalcs/>) and the minimal significance level adopted was 0.05.

2.8. Chromatin Immunoprecipitation (CHIP)

Confluent HepaRG cells were treated as previously described (Gripon et al., 2002). CHIP was performed with the CHIP-IT™ Chromatin Immunoprecipitation Kits & Shering Kits (Active Motif, Rixensart, Belgium) following manufacturer protocol. The anti Small Muscle Actin- α and anti- C/EBP α and anti- C/EBP β sc-150 antibody (Santa Cruz) were used to coimmunoprecipitate the DNA subjected to PCR. Primers used to amplify the DNA were -345F: CCCAATCCAGAGACTCAGCAG and -142R: CACTCTTGCAGAAGGAAAGAAGG and; -163F: CTGGTGCCTTCTTTCCTTCTG and +43R: AGAGCAGCAGCAGGATGTG.

3. Results

3.1. C/EBP α and C/EBP β expression positively correlate with that of NC11-493 in fibrotic human liver.

Expression of NC11-493, C/EBP α and C/EBP β were analysed in 55 non-tumor samples, while that of NC11-728 in 35 of these samples. The expression of these genes had a non-gaussian distribution (Shapiro-Wilk test $p < 0.05$) and NC11-728, C/EBP α and C/EBP β expression also showed positive skews (Skewness test

$p < 0.01$). Samples presented a wide interindividual variation of C/EBP α expression, while the NC11-493, NC11-728 and C/EBP β expression were more homogeneous (Fig 1A). We classified the samples in quartiles according to their NC11-493 expression. Thirteen samples were above the 75th percentile, 28 samples were within the 25th and 75th boundaries, and 14 were below the 25th percentile (range, 1.28-9.69; median, 4.06; 25-75th percentiles, 2.82-5.59 respectively). The three groups were designated high, moderate, and low non-tumor NC11-493 expression groups, and we observed a significant difference among their average NC11-493 expression (Kruskal Wallis test $p < 0.0001$) (Fig. 1B). When the expression of NC11-728, C/EBP α and C/EBP β were compared among the three groups the relative levels of these mRNAs were statistically different among groups (Kruskal Wallis test $p < 0.01$) (Fig. 1B). There is a positive correlation of NC11-493 expression with that of NC11-728, C/EBP α and C/EBP β and, of C/EBP α and C/EBP β (Table 2). No correlation was found between NC11-728 and C/EBP α , or between NC11-728 and C/EBP β expression. Samples with mild or intense fibrotic activity presented a significant (Mann-Whitney $p < 0.05$) higher C/EBP α expression than samples with no fibrotic activity (Fig 1C), but no difference was found for NC11-493, NC11-728 and C/EBP β and fibrosis activity. Besides, there was no difference in non-tumor NC11-493, NC11-728, C/EBP α and C/EBP β expression for different METAVIR classifications ($p > 0.05$).

3.2. C/EBP α and C/EBP β expression positively correlate with that of NC11-493 in HCC.

We further analysed the correlation among NC11-493, C/EBP α and C/EBP β in the 20 matched tumor samples. In these samples NC11-493 and

C/EBP α had a symmetric normal distribution, and C/EBP β a non gaussian distribution with a positive skew (Shapiro-Wilk test $p < 0.0001$, Skewness test $p < 0.0001$). Tumor samples presented a wide interindividual variation of C/EBP α expression, while the NC11-493 and C/EBP β expression were more homogeneous (Fig 2A, B, C). NC11-493 expression in tumor samples also showed positive correlation with that of C/EBP α and C/EBP β (Table 3) and negative correlation with tumor size (Fig 2D).

We did not find a significant difference in the expression of these transcripts in tumor and non-tumor tissue ($p > 0.05$, Fig 2A, B, C), although C/EBP α expression in these samples showed a positive correlation between HCC and fibrotic/cirrhotic samples ($R = 0.47$, $p = 0.038$, Table 3). No difference in tumor NC11-493, C/EBP α and C/EBP β expression was found among different METAVIR or fibrosis activity groups.

3.3. C/EBP β positively regulates collagen XVIII promoter 2 in hepatocytes *in vitro*.

Chromatin immunoprecipitation assays showed that C/EBP α and C/EBP β binds the DNA sequence of promoter 2 of *COL18A* between bases -163 and +43 and that C/EBP β also binds bases -345 and -142 (Fig. 3 A, B).

In order to elucidate the role of C/EBP β in collagen XVIII transcription we transiently transfected Hep3B and Huh7 cell lines with the pGL3-186 construct, containing the core promoter 2 of *COL18A1*, and increasing doses of a C/EBP β expressing vector. We confirmed the activity of mouse C/EBP β expression vector in Hep3B cell line by Western Blot (Fig 4A). Increasing doses of C/EBP β stimulated *COL18A1* core promoter transcription in hepatome cell lines and also in Hek293T (Fig 4 B, C, D).

To further explore the role of C/EBP β in the core promoter of *COL18A1*, we have mutated its putative binding site, comprising bases from -99 to -105 in human *COL18A1* promoter 2, by site-directed mutagenesis (Fig 5A). The mutant construction showed a significant decrease ($p<0.0001$) in the promoter activity when compared to wild type pGL3-186 construct, and the transcription activation by C/EBP β expression vector was significantly ($p<0.005$) reduced in Hep3B and Huh7 cell lines (Fig 5 B,C).

4. Discussion

Alterations in the expression of collagen XVIII and in its proteolytic product, endostatin, have been associated with liver fibrosis and HCC (Musso et al., 1998; Musso et al., 2001b; Jia et al., 2001; Dhar et al., 2002). Understanding the factors that modulate collagen XVIII levels in hepatocytes will help to elucidate pathways involved in these diseases.

We show in this report that NC11-493 expression correlates with NC11-728, CEBP α and C/EBP β both in fibrotic/cirrhotic and HCC liver samples (Table 2, 3) and, therefore, samples with high NC11-493 have also higher NC11-728, CEBP α and C/EBP β expression levels, while those with low NC11-493 expression levels have respectively lower transcription profile of the other three transcripts (Fig 1B). Even though there was no significant difference in NC11-493 expression between tumor and non tumor samples (Friedman test, $p=0.18$), as suggested by Hu and cols, 2005, HCC samples did show a slight lower NC11-493 expression than fibrotic/cirrhotic samples (Fig 2A). Our results thus suggest that NC11-493 expression in fibrotic/cirrhotic liver is not predictive of its expression in HCC and

confirm the inverse correlation of NC11-493 expression and tumor size previously described (Musso et al., 2001b, Fig 2D) .

The correlation between NC11-493 and NC11-728 was expected, as the two collagen XVIII isoforms originate from the same pre mRNA. On the other hand, it was unexpected that only NC11-493 correlates with C/EBP α and C/EBP β expression levels. The liver has the highest described expression of the NC11-493 isoform in the body (Saarela et al, 1998b), while the NC11-728 expression in this and other tissues is very low (Suzuki et al., 2002; Elamaa et al., 2003; Fig 1 A). So, our data further support that the splice machinery responsible for the NC11-493 mRNA is strongly active in hepatocytes. As processing of pre-mRNA occurs co-transcriptionally (Kornblihtt et al., 2004), it is possible that the NC11-728 expression levels are no longer representative of the collagen XVIII promoter 2 activity thus, we were not able to correlate the relative NC11-728 expression level with that of C/EBPs. Another hypothesis is based on the cross-talk of transcription factors during transcription and splice (Kornblihtt et al., 2004). C/EBPs, when bound to collagen XVIII promoter 2, would either have a dual function, participating in the transcription and the splice machinery of NC11-493 mRNA, or would help to recruit splice proteins that favour the processing of NC11-493 mRNA.

CEBP α , which presented the widest range of expression levels, was the only of the 4 transcripts analysed that was correlated with fibrotic activity (Fig. 1C); in addition, a positive correlation of the CEBP α levels between fibrotic/cirrhotic tissue and HCC was observed (Table 3). We did not find a correlation between tumor size and C/EBP α or C/EBP β expression as previously suggested by Tomizawa and cols, 2003. In opposition to Xu and cols, 2001, 60% of samples showed higher C/EBP α expression in tumor than non-tumor samples (Fig 2B),

while 65% of showed lower C/EBP β expression in tumor than non-tumor samples (Fig 2C). C/EBP α and C/EBP β are key transcription factors for hepatocytes (Schrem et al., 2002), therefore they may be important factors regulating collagen XVIII transcription in fibrosis or HCC. These preliminary data suggest that it will be important to further explore the role of this TF in the regulation of *COL18A1* gene.

CHIP assays further confirmed that C/EBP α and C/EBP β bind *COL18A1* promoter 2 (Fig 3A,B). We show that the C/EBP β binding site in *COL18A1* core promoter 2 region responds to increasing doses of C/EBP β expression (Fig. 4 B, C, D), and that once this binding site is mutated, the core promoter transcription capacity and its transactivation by C/EBP β are significantly impaired in hepatocyte cell lines (Fig 5 B, C). The promoter transactivation by C/EBP β is not completely abolished since this transcription factor has low binding site specificity and, still has some sequence affinity even though its core binding sequence has been mutated (Krivan & Wasserman, 2001). Therefore, our data confirms the C/EBP β conserved binding site in the human *COL18A1* promoter 2, bases from -99 to -105, that has a 100% identity with a motif in the mouse *Col18a1* promoter 2, bases from 87 to 93 (Fig 5A). Since different C/EBPs family members interact with common recognition sequence, these results strongly suggest that C/EBP α and C/EBP β are important for *COL18A1* transcription in hepatocytes and are responsible for its tissue specific expression. These C/EBP binding sites neighbour Sp1, Sp3 and YY1 sites, characterizing a cluster of transcription factors in *COL18A1* promoter 2 (Armelin-Correa et al., 2005). Further experiments will elucidate how the interactions among these factors regulate collagen XVIII in hepatocytes.

We here describe for the first time the correlation of NC11-493 with C/EBP α and C/EBP β expression in fibrotic/cirrhotic and HCC liver samples, and the role of C/EBP β in *COL18A1* transcriptional regulation. Therefore, the decrease in C/EBP β expression in HCC tumor may explain part of the decrease of collagen XVIII expression in HCC dedifferentiated tumors (Musso et al., 2001b). Additional studies will have to be developed to elucidate the role of the C/EBPs isoforms and their interactions with Sp1/Sp3 and YY1 in collagen XVIII transcription in hepatocytes.

Acknowledgments

The authors are indebted to Dr Elise Lavergne, C. Masotti and O. T. Suzuki for valuable suggestions. FAPESP, CNPq, FINEP and PRP-USP supported this work.

5. References

Armelin-Correa, L. M., C. J. Lin, *et al.* Characterization of human collagen XVIII promoter 2: interaction of Sp1, Sp3 and YY1 with the regulatory region and a SNP that increases transcription in hepatocytes. Matrix Biol, v.24, n.8, Dec, p.550-9. 2005.

Anthony P.P, K.G. Ishak, *et al.* The morphology of cirrhosis. Recommendations on definition, nomenclature and classification by a working group sponsored by the World Health Organization. J. Clin. Pathol., v.31, p.395-414. 1978.

Bedossa P. and T. Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. Hepatology, v. 24, n.2, Aug, p.289-93. 1996.

Den Dunnen, J.T., S.E. Antonarakis. Nomenclature for the description of human sequence variations. Hum. Genet. v.109, p.121-24. 2001.

Dhar, D. K., T. Ono, *et al.* Serum endostatin predicts tumor vascularity in hepatocellular carcinoma. Cancer, v.95, n.10, Nov 15, p.2188-95. 2002.

Di Bisceglie, A.M., R.L. Carithers, *et al.* Hepatocellular carcinoma – Meeting Report. Hepatology, v.28, n.4, p.1161-5. 1998.

Elamaa, H., A. Snellman, *et al.* Characterization of the human type XVIII collagen gene and proteolytic processing and tissue location of the variant containing a frizzled motif. Matrix Biol, v.22, n.5, Sep, p.427-42. 2003.

Gripon, P., S. Rumin, *et al.* Infection of a hepatoma cell line by hepatitis B virus. Proc Natl Acad Sci U S A, v.99, n.24, p.15655-60. 2002.

Hu, T. H., C. C. Huang, *et al.* Increased endostatin/collagen XVIII expression correlates with elevated VEGF level and poor prognosis in hepatocellular carcinoma. Mod Pathol, v.18, n.5, May, p.663-72. 2005.

Jia, J., M. Bauer, *et al.* Modulation of collagen XVIII/endostatin expression in lobular and biliary rat liver fibrogenesis. J. of Hepatology, v.35, p.386-91.2001.

Kato, S., T. Tajiri, *et al.* Genetic polymorphisms of aldehyde dehydrogenase 2, cytochrome p450 2E1 for liver cancer risk in HCV antibody-positive japanese patients and the variations of CYP2E1 mRNA expression levels in the liver due to its polymorphism. Scand J Gastroenterol, v.38, n.8, Aug, p.886-93. 2003.

Kornblihtt, A.R., M. de la Mata *et al.* Multiple links between transcription and splicing. RNA, v.10, n.10, p.1489-98. 2004.

Krivan, W. e W. W. Wasserman. A predictive model for regulatory sequences directing liver-specific transcription. Genome Res, v.11, n.9, Sep, p.1559-66. 2001.

Livak, K. J. e T. D. Schmittgen. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods, v.25, n.4, Dec, p.402-8. 2001.

Lowry, O. H., N. J. Rosebrough, *et al.* Protein measurement with the Folin phenol reagent. J Biol Chem, v.193, n.1, Nov, p.265-75. 1951.

Miller, S. A., D. D. Dykes, *et al.* A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res, v.16, n.3, Feb 11, p.1215. 1988.

Musso, O., M. Rehn, *et al.* Collagen XVIII is localized in sinusoids and basement membrane zones and expressed by hepatocytes and activated stellate cells in fibrotic human liver. Hepatology, v.28, n.1, Jul, p.98-107. 1998.

Musso, O., N. Theret, *et al.* Tumor hepatocytes and basement membrane-Producing cells specifically express two different forms of the endostatin precursor, collagen XVIII, in human liver cancers. Hepatology, v.33, n.4, Apr, p.868-76. 2001a.

Musso, O., M. Rehn, *et al.* Tumor progression is associated with a significant decrease in the expression of the endostatin precursor collagen XVIII in human hepatocellular carcinomas. Cancer Res, v.61, n.1, Jan 1, p.45-9. 2001b.

Oh, S. P., Y. Kamagata, *et al.* Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of collagenous proteins. Proc Natl Acad Sci U S A, v.91, n.10, May 10, p.4229-33. 1994.

O'reilly, M. S., T. Boehm, *et al.* Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. Cell, v.88, n.2, Jan 24, p.277-85. 1997.

Quelard, D., E. Lavergne, *et al.* The frizzled module of COL18A1 induces tumor cell death through inhibition of the Wnt-B-catenin signaling pathway. Abstracts for American Society of Liver Disease. Hepatology Oct, p. 588A. 2006.

Ramji, D. P. e P. Foka. CCAAT/enhancer-binding proteins: structure, function and regulation. Biochem J, v.365, n.Pt 3, Aug 1, p.561-75. 2002.

Read, M. Electrophoretic mobility shift assay (EMSA) In: Dockerty K, ed. Essential Techniques. New York, Wiley & Sons p. 7-11, 1998.

Rehn, M. e T. Pihlajaniemi. Alpha 1(XVIII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen. Proc Natl Acad Sci U S A, v.91, n.10, May 10, p.4234-8. 1994.

Saarela, J., M. Rehn, *et al.* The short and long forms of type XVIII collagen show clear tissue specificities in their expression and location in basement membrane zones in humans. Am J Pathol, v.153, n.2, Aug, p.611-26. 1998a.

Saarela, J., R. Ylikarppa, *et al.* Complete primary structure of two variant forms of human type XVIII collagen and tissue-specific differences in the expression of the corresponding transcripts. Matrix Biol, v.16, n.6, Jan, p.319-28. 1998b.

Schrem, H., J. Klempnauer, *et al.* Liver-enriched transcription factors in liver function and development. Part I and Part II. Pharmacol Rev, v.54, n.1, Mar, p.129-330. 2002.

Sertie, A. L., V. Sossi, *et al.* Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome). Hum Mol Genet, v.9, n.13, Aug 12, p.2051-8. 2000.

Suzuki, O. T., A. L. Sertie, *et al.* Molecular analysis of collagen XVIII reveals novel mutations, presence of a third isoform, and possible genetic heterogeneity in Knobloch syndrome. Am J Hum Genet, v.71, n.6, Dec, p.1320-9. 2002.

Tan, E.O., S.C. Hooi, *et al.* CCAAT/Enhancer binding protein alpha knock-in mice exhibit early liver glycogen storage and reduced susceptibility to hepatocellular carcinoma. Cancer Res. v.65, n.22, Nov, p.10330-7. 2005.

Tomizawa, M., K. Watanabe, *et al.* Down-regulated expression of the CCAAT/enhancer binding protein alpha and beta genes in human hepatocellular carcinoma: a possible prognostic marker. Anticancer Res, v.23, n.1A, Jan-Feb, p.351-4. 2003.

Xu, L., L. Hui, *et al.* Expression profiling suggested a regulatory role of liver-enriched transcription factors in human hepatocellular carcinoma. Cancer Res, v.61, n.7, Apr 1, p.3176-81. 2001.

A

	Median	IQR	95% CI of Median	
NC11-493 - 493NT	4,07	2,76	2,94	4,90
CEBPa - CEBPaNT	3,95	8,15	2,83	8,15
CEBPb - CEBPbNT	0,83	0,79	0,70	1,05
NC11-728 - 728NT	3,80	2,40	3,25	4,59

B

C

Fig. 1. The relative mRNA levels of NC11-493, NC11-728, C/EBP α and C/EBP β in human non tumor hepatic tissues. The levels of mRNAs were measured in duplicates by QPCR and expressed as relative changes in gene expression between fibrotic tissue and HepaRG cell line. **A.** Box and Whisker plots representation of the relative mRNA levels of NC11-493, C/EBP α and C/EBP β in 55 non-tumor samples and NC11-728 in 35 non-tumor samples. The central line in the box represents the median, and the upper and lower lines of the box represent the 75th and 25th quartiles, respectively. The lines outside the box represent the range of measurements. The sign (+) represents extreme measurements, while the sign (o) represents atypical measurements. Even though the relative expression of NC11-728 is close to that of NC11-493, this is due to the normalization process through HepaRG gene expression. The actual NC11-728 expression (Ct mean $\pm$ SD 26.32 $\pm$ 0.78) in all samples was approximately 32 times lower than that of NC11-493 (Ct mean $\pm$ SD 21.34 $\pm$ 0.76). **B.** Bar graphs representing the mean $\pm$ SD of mRNA levels of non-tumoral NC11-493, NC11-728, C/EBP α and C/EBP β for low, moderate and high non-tumoral NC11-493 expression groups. 13 samples were above the 75th percentile, 28 samples were within the 25th and 75th percentiles boundaries, and 14 below the 25th percentile. The NC11-728 expression was measure for 14 samples of the low, 17 of the moderate and of 4 high expression group. **C.** Bar graphs representing the mean $\pm$ SD of mRNA levels of non-tumoral C/EBP α for patients with severe, mild, moderate and absent fibrose activity. The asterisk represents (***) p<0.0001 and (**) p<0.01 (*) p<0.05.

A

NC11-493	Median	IQR	95% CI of Median	
493NT	5,59	1,58	4,90	6,06
493T	4,50	3,61	2,87	6,36

B

CEBPa	Median	IQR	95% CI of Median	
CEBPaNT	11,18	5,18	8,86	13,45
CEBPaT	12,77	10,15	9,04	18,57

C

CEBPb	Median	IQR	95% CI of Median	
CEBPbNT	1,28	0,61	0,99	1,57
CEBPbT	1,10	0,68	0,82	1,26

D

Spearman rank correlation rs statistic	-0,48
2-tailed p	0,0331

Fig. 2. The relative mRNA levels of NC11-493 **(A)**, C/EBP α **(B)** and C/EBP β **(C)** in 20 human non-tumor (NT) and matched tumor (T) hepatic tissues. The levels of mRNAs were measured in duplicates by QPCR and expressed as relative changes in gene expression between tissue and HepaRG cell line. Box and Whisker plots representation of the relative mRNA levels of NC11-493, C/EBP α and C/EBP β . The central line in the box represents the median, and the upper and lower lines of the box represent the 75th and 25th quartiles, respectively. The lines outside the box represent the range of measurements. The sign (+) represents extreme measurements, while the sign (o) represents atypical measurements **D.** Spearman Rank correlation between NC11-493 expression in tumor samples and their size.

Fig. 3. Binding of C/EBP α and C/EBP β to the -163/+43 bp **(A)** and the -345/-142 bp **(B)** regions of *COL18A1* core promoter 2 was assessed by CHIP analysis of confluent HepRG cell line. Lanes correspond to size standard (lane Kb), input DNA (lane 1), mock antibody, anti-SMA alpha (lane 2), antibody anti-C/EBP α (lane 3), antibody anti-C/EBP β (lane 4) and water (lane 5).

A.

B.

C.

D.

Fig.4. Functional analysis of CEBP β binding site in human *COL18A1* core promoter 2. **A.** Western blot showing control protein HSC70 and human and mouse C/EBP β protein detection in Hep3B cells transfected with increasing doses of the MSV- C/EBP β expression vector. **B, C, D.** Graphic representation showing the effects of C/EBP β overexpression in activity profiles of reporter luciferase constructs in Hep3B (**B**), Huh7 (**C**) and Hek293T (**D**) cells respectively. Luciferase activity was normalized to μ g/protein for each sample. Data from 2 transient transfections in triplicates are expressed as fold luciferase activity relative to pGL3-Basic cotransfected with empty expression vector, and represent mean $\pm$ SD

Fig.5. Functional analysis of site directed mutagenesis of the putative CEBP β binding site in human *COL18A1* core promoter 2. **A** Alignment between human and mouse Collagen XVIII promoter 2 showing in a box the conserved putative CEBP β binding site predicted in silico, and arrows indicated the bases mutated (GGG $\rightarrow$ CCG). **B, C.** Graphic representation showing the effects of 0.025 μ g MSV-CEBP β overexpression in activity profiles of 0.025 μ g pGL3-186 wild type (pGL3-186 wt) or pGL3-186 mutant (mut) constructs in Hep3B (**B**) and Huh7 (**C**) cells lines respectively. Luciferase activity was normalized to μ g/protein for each sample. Data from 2 transient transfections in triplicates are expressed as fold luciferase activity relative to 0.025 μ g pGL3-Basic cotransfected with 0.025 μ g empty expression vector, and represent as mean $\pm$ SD.

Table 1

Primers used used for quantitative PCR (qPCR) reaction.

Primer	Sequence 5'-3'
NC11-493 forward	TCCAGCTGTGGAATGACACT
NC11-493 reverse	AGTCTCCAGGACCAGGGTTT
NC11-728 forward	GCTTCTCTCTCCTCCTTGCTG
NC11-728 reverse	GAGAGTCCTTGGCTGTCTGG
CEBP α forward	GAGGAGGGGAGAATTCTTGG
CEBP α reverse	CATTTCCAAGGCACAAGGT
CEBP β forward	CGTGTGTACACGGGACTGAC
CEBP β reverse	CAACAAGCCCGTAGGAACAT
18S forward	GTAACCCGTTGAACCCCAT
18S reverse	CCATCCAATCGGTAGTAGCG

Table 2

Spearman rank order correlations between collagen XVIII isoforms, NC11-493 and NC11-728, and C/EBP α and C/EBP β in non-tumor human hepatic tissues.

Non Tumor Samples	R	p<
NC11-493/NC11-728	0.66	0.0001
NC11-493/ C/EBP α	0.59	0.0001
NC11-493/ C/EBP β	0.58	0.0001
C/EBP α /C/EBP β	0.32	0.05

Spearman correlation values are expressed as correlation coefficients (R) and p values were calculated for each test.

Table 3

Spearman rank order correlations between collagen XVIII isoforms, NC11-493 and C/EBP α and C/EBP β in tumor human hepatic tissues.

Tumor Samples	R	p<
NC11-493/ C/EBP α	0.70	0.01
NC11-493/ C/EBP β	0.52	0.05
C/EBP α /C/EBP β	0.63	0.01
C/EBP α Tumor/ C/EBP α Non-Tumor	0.47	0.05

Spearman correlation values are expressed as correlation coefficients (R) and p values were calculated for each test.

VI. DISCUSSÃO E CONCLUSÃO

O colágeno XVIII é um importante gene de desenvolvimento que faz parte da membrana basal de diversos tecidos (Saarela *et al.*, 1998a e 1998b, Passos-Bueno *et al.*, 2006 Anexo I). Este gene atua em diversos processos celulares, como migração e proliferação celular, e fisiológicos, como angiogênese (O'Reilly *et al.*, 1997; Quelard *et al.*, 2006). Ele se expressa como 3 isoformas, reguladas por 2 promotores (Saarela *et al.*, 1998a; Elamaa *et al.*, 2003). Compreender a regulação da transcrição desse colágeno é um passo fundamental para elucidar os mecanismos envolvidos em sua expressão, tanto em situações fisiológicas, quanto patológicas. Nesse trabalho esclarecemos alguns componentes envolvidos na regulação da transcrição de *COL18A1* em hepatócitos humanos.

Iniciamos o estudo descrevendo a existência de 5 regiões conservadas entre humanos e camundongos no promotor 2 de *COL18A1* (Cap. III; Fig. 1). Essas regiões conservadas, possivelmente, sofrem ou sofreram pressão seletiva e, portanto, contém prováveis sítios de ligação para fatores de transcrição que atuam na regulação da expressão de *COL18A1*.

Fig. 1. Representação gráfica de 1487pb do promotor 2 de *COL18A1*, indicando exon 3 do gene contendo SNPC.1135C/T, principais posições referentes a construções com vetor repórter utilizadas, flecha indicando início da transcrição a 21pb, asterisco indicando posição do SNP-700T/G, regiões conservadas entre humano e camundongo (RC I –V), sítios de ligação e seus respectivos fatores de transcrição e chaves delimitando as principais regiões regulatórias identificadas em hepatócitos.

Analizando essas regiões potencialmente funcionais descrevemos uma série de interações entre os fatores de transcrição Sp1, Sp3, YY1 (Cap. III), Oct-1 (Cap. IV), C/EBP α e C/EBP β (Cap. V) com o promotor 2 do colágeno XVIII humano (Fig 1). Nas regiões conservadas I e II, que correspondem ao promotor mínimo do colágeno XVIII, encontramos interação de Sp1, Sp3, YY1, C/EBP α e C/EBP β constituindo um agrupamento de fatores de transcrição que atuariam cooperativamente. Demonstramos que C/EBP β é um ativador da transcrição de *COL18A1* em hepatócitos (Cap. V). Já foi descrita uma sinergia entre C/EBP β e Sp1 na regulação do gene *Cyp2d5* em ratos, onde a ativação da transcrição por C/EBP β é completamente dependente de Sp1, através de interação proteína-proteína (Lee *et al.*, 1994, Lee *et al.*, 1997). YY1 também interage com C/EBP β na regulação da transcrição do papiloma vírus humano (Bauknecht *et al.*, 1996). Será importante explorar então o papel da interação e sinergia entre C/EBP α , C/EBP β , Oct-1, Sp1, Sp3 e YY1 na regulação da transcrição do colágeno XVIII tanto no fígado, quanto em outros tecidos.

O fato de fatores de transcrição como Sp1, Sp3, YY1 serem ubíquos não explicaria a expressão hepato-específica da isoforma NC11-493, mas por outro lado, explicaria a expressão da isoforma NC11-728, também regulada pelo promotor 2, em diversos outros órgãos além do fígado, como no olho e na pele de adultos e no músculo esquelético, pulmão e baço fetais humanos (Elaama *et al.*, 2003). Assim, Sp1, Sp3, YY1 poderiam estar participando de uma atividade transcricional basal do promotor 2 do colágeno XVIII em vários órgãos e, no fígado, atuariam em sinergia com fatores enriquecidos em hepatócitos, como C/EBP α , C/EBP β , HNF3 e HNF1, estes sim responsáveis por uma alta taxa de transcrição desse promotor nessas células. Foi demonstrado que HNF3 e HNF1

atuam na regulação do colágeno XVIII no promotor de camundongo em hepatócitos (Liétard *et al.*, 2000) e, apesar de não termos encontrado sítios de ligação de HNF3 na região do promotora estudada, estes podem estar localizados em seqüências mais distantes.

Além de demonstrarmos que existem pelo menos dois sítios de ligação para C/EBPs no promotor 2 do colágeno XVIII, encontramos correlação positiva da expressão da isoforma NC11-493 com a expressão de C/EBP α e C/EBP β em tecido hepático cirrótico e tumoral (Cap. V). Essa correlação é bastante interessante já que ambos os fatores de transcrição estão relacionados com vários processos fisiopatológicos hepáticos, como por exemplo diferenciação, proliferação celular, processos inflamatórios e tumorigênese (Ramji & Foka, 2002). Alterações na expressão de C/EBP α e C/EBP β podem contribuir com a diminuição da expressão de *COL18A1* durante a progressão de HCC (Musso *et al.*, 2001b), já que essa progressão leva a dediferenciação das células tumorais e perda da expressão de uma série de fatores de transcrição hepato-específicos, entre eles C/EBP α e C/EBP β (Xu *et al.*, 2001; Tomizawa *et al.*, 2003). Esses dois fatores são codificados por genes sem *íntron*, que para exercerem suas funções devem formar homo ou heterodímeros. Pelo menos dois sítios de início de tradução no RNAm de C/EBP α e C/EBP β são responsáveis por suas diferentes isoformas. As isoformas de C/EBP α são p42 (polipeptídeo com 42KDa), que possui 3 sítios de ativação da transcrição, e p30 (polipeptídeo com 30KDa), com apenas 1 sítio de ativação da transcrição e, portanto, menor potencial ativador. As isoformas de C/EBP β são LAP (*liver activating protein*) isoforma ativadora da transcrição e LIP (*liver inhibitory protein*) isoforma inibidora da transcrição (Schrem *et al.*, 2004). Vários estudos sugerem que a razão entre as diferentes

isoformas de C/EBP α e C/EBP β , assim como seus padrões de dimerização, têm papel importante na expressão de genes em hepatócitos (Buck *et al.*, 1994; Luedde *et al.*, 2004). Estudos são necessários para elucidar como as razões entre p42/p30 e LAP/LIP afetam a expressão de *COL18A1* em hepatócitos em processos fisiológicos, na fibrose/cirrose hepática e no surgimento e progressão do HCC. Para se determinar a importância de C/EBP α e C/EBP β na transcrição do colágeno XVIII seria interessante verificar-se a expressão de *Col18a1* em camundongos knockout para *C/ebp α* e *C/ebp β* (Lee *et al.*, 1997b; Tanaka *et al.*, 1997).

Também mostramos nesse trabalho a existência de diversos SNPs no promotor 2 do colágeno XVIII, sendo o SNP-700T/G funcional *in vitro* (Cap. III e IV). Entretanto, apesar de termos identificado, utilizando linhagem celular HepG2, 3 fatores de transcrição, Oct-1, Sp3 e YY1, que se ligam aos 41pb vizinhos a esse polimorfismo, não conseguimos encontrar um fator de transcrição que se ligue exclusivamente a um dos alelos. No Capítulo III sugerimos que YY1 tem maior afinidade pelo alelo T, enquanto Sp3 tem maior afinidade pelo alelo G. YY1 e Sp3 já foram descritos tanto como inibidores como ativadores da transcrição (Kennett *et al.*, 1997; Thomas & Seto, 1999). Além disso, mostramos que a deleção desses 41pb leva a uma atividade do promotor semelhante a atividade do promotor contendo o alelo G, em linhagem celular HepG2 (Capítulo IV). Estes dados mostram que este segmento contém uma região reguladora importante e, sugerem que o alelo T facilita a ligação de um complexo inibidor da transcrição do promotor em HepG2, enquanto o alelo G dificultaria a formação desse complexo. Assim, essa diferente afinidade de YY1 e Sp3 pelos diferentes alelos e suas interações com Oct-1 poderiam explicar parte da alteração da atividade

transcricional do SNP-700T/G em HepG2. Além disso, a linhagem celular Huh7 apresentou a atividade oposta dos alelos do SNP-700T/G em relação as linhagens HepG2 e Hep3B. A análise de fatores de transcrição diferencialmente expressos por essas linhagens pode ajudar a elucidar melhor a regulação desse SNP em hepatócitos.

No Capítulo IV descrevemos o desequilíbrio de ligação entre o SNP-700T/G e o SNPc.1135C/T, no domínio frizzled do colágeno XVIII. Não foi possível elucidar a funcionalidade do SNPs c.1135C/T *in vitro*, mas os haplótipos formados por esses SNPs tem diferentes frequências entre descendentes de europeus e de africanos. Apesar de não termos encontrado associação do SNP-700T/G com susceptibilidade a HCC ou com níveis de expressão de NC11-493 em fígado fibrótico, a análise da correlação dos diferentes haplótipos com a expressão do colágeno XVIII em fígado normal, fibrótico/cirrótico e tumoral, além da associação com número, progressão, grau de vascularização e malignização de nódulos hepáticos pode ajudar a esclarecer se o colágeno XVIII atua no processo de tumorigênese hepática. A indução de inflamação e hepatocarcinoma em camundongo *Col18a1*^{-/-} (Fukai *et al.*, 2002) pode auxiliar na compreensão do papel deste gene em patologias hepáticas.

Nosso trabalho traz importantes contribuições e abre novas perspectivas para a compreensão da regulação do colágeno XVIII no fígado, tanto em situações fisiológicas, quanto em processos fibrogênicos e tumorigênicos.

VII. REFERÊNCIAS BIBLIOGRÁFICAS

Ackley, B. D., J. R. Crew, *et al.* The NC1/endostatin domain of *Caenorhabditis elegans* type XVIII collagen affects cell migration and axon guidance. J Cell Biol, v.152, n.6, Mar 19, p.1219-32. 2001.

Alberts B, *et al.* Control of Gene Expression. In: Alberts B, *et al.* Molecular Biology of the Cell. New York, Garland Science p. 375-466. 2002

Bassam, B.S., G. Caetano-Anolles, *et al.* Fast sensitive silver staining of DNA in polyacrilamide gels. Anal. Biochem. v.196, p.80-3. 1991.

Battaller, R. e D. A. Brenner. Liver fibrosis. J Clin Invest, v.115, n.2, Feb, p.209-18. 2005.

Bauknecht, T., R. H. See, *et al.* A novel C/EBP beta-YY1 complex controls the cell-type-specific activity of the human papillomavirus type 18 upstream regulatory region. J Virol, v.70, n.11, Nov, p.7695-705. 1996.

Buck, M., H. Turler, *et al.* LAP (NF-IL-6), a tissue-specific transcriptional activator, is an inhibitor of hepatoma cell proliferation. Embo J, v.13, n.4, Feb 15, p.851-60. 1994.

Cheung, V. G., L. K. Conlin, *et al.* Natural variation in human gene expression assessed in lymphoblastoid cells. Nat Genet, v.33, n.3, Mar, p.422-5. 2003.

Derman, E., K. Krauter, *et al.* Transcriptional control in the production of liver-specific mRNAs. Cell, v.23, n.3, Mar, p.731-9. 1981.

Dhar, D. K., T. Ono, *et al.* Serum endostatin predicts tumor vascularity in hepatocellular carcinoma. Cancer, v.95, n.10, Nov 15, p.2188-95. 2002.

Elamaa, H., J. Peterson, *et al.* Cloning of three variants of type XVIII collagen and their expression patterns during *Xenopus laevis* development. Mech Dev, v.114, n.1-2, Jun, p.109-13. 2002.

Elamaa, H., A. Snellman, *et al.* Characterization of the human type XVIII collagen gene and proteolytic processing and tissue location of the variant containing a frizzled motif. Matrix Biol, v.22, n.5, Sep, p.427-42. 2003.

Errera, F. I., L. H. Canani, *et al.* Functional vascular endothelial growth factor - 634G>C SNP is associated with proliferative diabetic retinopathy: a case-control study in a Brazilian population of European ancestry. Diabetes Care, v.30, n.2, Feb, p.275-9. 2007.

- Fattovich, G., T. Stroffolini, *et al.* Hepatocellular carcinoma in cirrhosis: incidence and risk factors. Gastroenterology, v.127, n.5 Suppl 1, Nov, p.S35-50. 2004.
- Feitelson, M. A., B. Sun, *et al.* Genetic mechanisms of hepatocarcinogenesis. Oncogene, v.21, n.16, Apr 11, p.2593-604. 2002.
- Felbor, U., L. Dreier, *et al.* Secreted cathepsin L generates endostatin from collagen XVIII. Embo J, v.19, n.6, Mar 15, p.1187-94. 2000.
- Folkman, J. e R. Kalluri. Cancer without disease. Nature, v.427, n.6977, Feb 26, p.787. 2004.
- Frith, M. C., M. C. Li, *et al.* Cluster-Buster: Finding dense clusters of motifs in DNA sequences. Nucleic Acids Res, v.31, n.13, Jul 1, p.3666-8. 2003.
- Fukai, N., L. Eklund, *et al.* Lack of collagen XVIII/endostatin results in eye abnormalities. Embo J, v.21, n.7, Apr 2, p.1535-44. 2002.
- Gripon, P., S. Rumin, *et al.* Infection of a human hepatoma cell line by hepatitis B virus. Proc Natl Acad Sci U S A, v.99, n.24, Nov 26, p.15655-60. 2002.
- Grompe, M. The rapid detection of unknown mutations in nucleic acids. Nat Genet, v.5, n.2, Oct, p.111-7. 1993.
- Hanai, J., J. Gloy, *et al.* Endostatin is a potential inhibitor of Wnt signaling. J Cell Biol, v.158, n.3, Aug 5, p.529-39. 2002.
- Heljasvaara, R., P. Nyberg, *et al.* Generation of biologically active endostatin fragments from human collagen XVIII by distinct matrix metalloproteases. Exp Cell Res, v.307, n.2, Jul 15, p.292-304. 2005.
- Hersel, U., C. Dahmen, *et al.* RGD modified polymers: biomaterials for stimulated cell adhesion and beyond. Biomaterials, v.24, n.24, Nov, p.4385-415. 2003.
- Hohler, T., A. Kruger, *et al.* A tumor necrosis factor-alpha (TNF-alpha) promoter polymorphism is associated with chronic hepatitis B infection. Clin Exp Immunol, v.111, n.3, Mar, p.579-82. 1998.
- Hoogendoorn, B., S. L. Coleman, *et al.* Functional analysis of human promoter polymorphisms. Hum Mol Genet, v.12, n.18, Sep 15, p.2249-54. 2003.
- Hsieh, L. L., R. C. Huang, *et al.* L-myc, GST M1 genetic polymorphism and hepatocellular carcinoma risk among chronic hepatitis B carriers. Cancer Lett, v.103, n.2, Jun 5, p.171-6. 1996.

Hu, T. H., C. C. Huang, *et al.* Increased endostatin/collagen XVIII expression correlates with elevated VEGF level and poor prognosis in hepatocellular carcinoma. Mod Pathol, v.18, n.5, May, p.663-72. 2005.

Hynes, R. O. e Q. Zhao. The evolution of cell adhesion. J Cell Biol, v.150, n.2, Jul 24, p.F89-96. 2000.

Iozzo, R. V. Basement membrane proteoglycans: from cellar to ceiling. Nat Rev Mol Cell Biol, v.6, n.8, Aug, p.646-56. 2005.

Iughetti, P., O. Suzuki, *et al.* A polymorphism in endostatin, an angiogenesis inhibitor, predisposes for the development of prostatic adenocarcinoma. Cancer Res, v.61, n.20, Oct 15, p.7375-8. 2001.

Kato, S., T. Tajiri, *et al.* Genetic polymorphisms of aldehyde dehydrogenase 2, cytochrome p450 2E1 for liver cancer risk in HCV antibody-positive Japanese patients and the variations of CYP2E1 mRNA expression levels in the liver due to its polymorphism. Scand J Gastroenterol, v.38, n.8, Aug, p.886-93. 2003.

Kennett, S. B., A. J. Udvardi, *et al.* Sp3 encodes multiple proteins that differ in their capacity to stimulate or repress transcription. Nucleic Acids Res, v.25, n.15, Aug 1, p.3110-7. 1997.

Krieger, H., N. E. Morton, *et al.* Racial admixture in north-eastern Brazil. Ann Hum Genet, v.29, n.2, Nov, p.113-25. 1965.

Krivan, W. e W. W. Wasserman. A predictive model for regulatory sequences directing liver-specific transcription. Genome Res, v.11, n.9, Sep, p.1559-66. 2001.

Kuklin, A., K. Munson, *et al.* Detection of single-nucleotide polymorphisms with the WAVE DNA fragment analysis system. Genet Test, v.1, n.3, p.201-6. 1997.

Kuo, C. J., K. R. Lamontagne, Jr., *et al.* Oligomerization-dependent regulation of motility and morphogenesis by the collagen XVIII NC1/endostatin domain. J Cell Biol, v.152, n.6, Mar 19, p.1233-46. 2001.

Kuo, M. e C. D. Allis. In vivo cross-linking and immunoprecipitation for studying dynamic protein:DNA associations in a chromatin environment. Methods, v.19, p.425-33. 1999.

Lee, Y. H., S. C. Williams, *et al.* The ability of C/EBP beta but not C/EBP alpha to synergize with an Sp1 protein is specified by the leucine zipper and activation domain. Mol Cell Biol, v.17, n.4, Apr, p.2038-47. 1997.

Lee, Y. H., B. Sauer, *et al.* Disruption of the c/ebp alpha gene in adult mouse liver. Mol Cell Biol, v.17, n.10, Oct, p.6014-22. 1997.

Lee, Y. H., M. Yano, *et al.* A novel cis-acting element controlling the rat CYP2D5 gene and requiring cooperativity between C/EBP beta and an Sp1 factor. Mol Cell Biol, v.14, n.2, Feb, p.1383-94. 1994.

Levine, M. e R. Tjian. Transcription regulation and animal diversity. Nature, v.424, n.6945, Jul 10, p.147-51. 2003.

Lietard, J., N. Theret, *et al.* The promoter of the long variant of collagen XVIII, the precursor of endostatin, contains liver-specific regulatory elements. Hepatology, v.32, n.6, Dec, p.1377-85. 2000.

Liu, T. C., C. T. Peng, *et al.* Association of endostatin D104N with leukemia. Kaohsiung J Med Sci, v.19, n.1, Jan, p.1-5. 2003.

Livak, K. J. e T. D. Schmittgen. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods, v.25, n.4, Dec, p.402-8. 2001.

Loots, G. G., I. Ovcharenko, *et al.* rVista for comparative sequence-based discovery of functional transcription factor binding sites. Genome Res, v.12, n.5, May, p.832-9. 2002.

Lourenco, G. J., C. Cardoso-Filho, *et al.* A high risk of occurrence of sporadic breast cancer in individuals with the 104NN polymorphism of the COL18A1 gene. Breast Cancer Res Treat, v.100, n.3, Dec, p.335-8. 2006.

Lowry, O. H., N. J. Rosebrough, *et al.* Protein measurement with the Folin phenol reagent. J Biol Chem, v.193, n.1, Nov, p.265-75. 1951.

Luedde, T., M. Duderstadt, *et al.* C/EBP beta isoforms LIP and LAP modulate progression of the cell cycle in the regenerating mouse liver. Hepatology, v.40, n.2, Aug, p.356-65. 2004.

Macpherson, G. R., A. S. Singh, *et al.* Genotyping and functional analysis of the D104N variant of human endostatin. Cancer Biol Ther, v.3, n.12, Dec, p.1298-303. 2004.

Marneros, A. G. e B. R. Olsen. Physiological role of collagen XVIII and endostatin. Faseb J, v.19, n.7, May, p.716-28. 2005.

Masotti, C., L.M. Armelin-Correa *et al.* A functional SNP in the promoter region of TCOF1 is associated with reduced gene expression and YY1 DNA-protein interaction. Gene, v.359, Oct 10, p.44-52. 2005.

Miller, S. A., D. D. Dykes, *et al.* A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res, v.16, n.3, Feb 11, p.1215. 1988.

Morley, M., C. M. Molony, *et al.* Genetic analysis of genome-wide variation in human gene expression. Nature, v.430, n.7001, Aug 12, p.743-7. 2004.

Moulton, K. S., B. R. Olsen, *et al.* Loss of collagen XVIII enhances neovascularization and vascular permeability in atherosclerosis. Circulation, v.110, n.10, Sep 7, p.1330-6. 2004.

Musso, O., M. Rehn, *et al.* Collagen XVIII is localized in sinusoids and basement membrane zones and expressed by hepatocytes and activated stellate cells in fibrotic human liver. Hepatology, v.28, n.1, Jul, p.98-107. 1998.

Musso, O., N. Theret, *et al.* Tumor hepatocytes and basement membrane-Producing cells specifically express two different forms of the endostatin precursor, collagen XVIII, in human liver cancers. Hepatology, v.33, n.4, Apr, p.868-76. 2001a.

Musso, O., M. Rehn, *et al.* Tumor progression is associated with a significant decrease in the expression of the endostatin precursor collagen XVIII in human hepatocellular carcinomas. Cancer Res, v.61, n.1, Jan 1, p.45-9. 2001b.

Nascimento, H., G. J. Lourenco, *et al.* D104N polymorphism in endostatin, an angiogenesis inhibitor, in acute and chronic myeloid leukaemia. Leuk Res, Oct 24. 2006.

Nascimento, H., C. S. Rodrigues Coy, *et al.* A polymorphism in the angiogenesis inhibitor, endostatin, in sporadic colorectal adenocarcinoma. Int J Colorectal Dis, v.19, n.5, Sep, p.499-501. 2004.

Ofner, P.J. e P.A. Underhill. Comparative DNA sequencing by denaturing high-performance liquid chromatography (dHPLC). Am. J. Hum. Genet. v.57, p.A266. 1995.

Oh, S. P., M. L. Warman, *et al.* Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localization of the alpha 1(XVIII) collagen gene to mouse chromosome 10 and human chromosome 21. Genomics, v.19, n.3, Feb, p.494-9. 1994a.

Oh, S. P., Y. Kamagata, *et al.* Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of

collagenous proteins. Proc Natl Acad Sci U S A, v.91, n.10, May 10, p.4229-33. 1994b.

O'reilly, M. S., T. Boehm, *et al.* Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. Cell, v.88, n.2, Jan 24, p.277-85. 1997.

Orita, M., H. Iwahana, *et al.* Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci U S A, v.86, n.8, Apr, p.2766-70. 1989.

Passos-Bueno, M. R., O. T. Suzuki, *et al.* Mutations in collagen 18A1 and their relevance to the human phenotype. An Acad Bras Cienc, v.78, n.1, Mar, p.123-31. 2006.

Pennacchio, L. A. e E. M. Rubin. Genomic strategies to identify mammalian regulatory sequences. Nat Rev Genet, v.2, n.2, Feb, p.100-9. 2001.

Quelard, D., E. Lavergne, *et al.* The frizzled module of COL18A1 induces tumor cell death through inhibition of the Wnt-B-catenin signaling pathway. Abstracts for American Society of Liver Disease. Hepatology Oct, p. 588A. 2006.

Ramji, D. P. e P. Foka. CCAAT/enhancer-binding proteins: structure, function and regulation. Biochem J, v.365, n.Pt 3, Aug 1, p.561-75. 2002.

Read, M. Electrophoretic mobility shift assay (EMSA) In: Dockerty K, ed. Essential Techniques. New York, Wiley & Sons p. 7-11, 1998.

Rehn, M., E. Hintikka, *et al.* Characterization of the mouse gene for the alpha 1 chain of type XVIII collagen (Col18a1) reveals that the three variant N-terminal polypeptide forms are transcribed from two widely separated promoters. Genomics, v.32, n.3, Mar 15, p.436-46. 1996.

Rehn, M. e T. Pihlajaniemi. Alpha 1(XVIII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen. Proc Natl Acad Sci U S A, v.91, n.10, May 10, p.4234-8. 1994.

Ricard-Blum, S. e C. Faye. [Collagens associated with basement membranes and their matricryptins]. J Soc Biol, v.199, n.4, p.321-8. 2005.

Rocken, C. e S. Carl-Mcgrath. Pathology and pathogenesis of hepatocellular carcinoma. Dig Dis, v.19, n.4, p.269-78. 2001.

Rosmorduc, O., D. Wendum, *et al.* Hepatocellular hypoxia-induced vascular endothelial growth factor expression and angiogenesis in experimental biliary cirrhosis. Am J Pathol, v.155, n.4, Oct, p.1065-73. 1999.

Saarela, J., M. Rehn, *et al.* The short and long forms of type XVIII collagen show clear tissue specificities in their expression and location in basement membrane zones in humans. Am J Pathol, v.153, n.2, Aug, p.611-26. 1998a.

Saarela, J., R. Ylikarppa, *et al.* Complete primary structure of two variant forms of human type XVIII collagen and tissue-specific differences in the expression of the corresponding transcripts. Matrix Biol, v.16, n.6, Jan, p.319-28. 1998b.

Saiki, R. K., S. Scharf, *et al.* Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science, v.230, n.4732, Dec 20, p.1350-4. 1985.

Sasaki, T., N. Fukai, *et al.* Structure, function and tissue forms of the C-terminal globular domain of collagen XVIII containing the angiogenesis inhibitor endostatin. Embo J, v.17, n.15, Aug 3, p.4249-56. 1998.

Schrem, H., J. Klempnauer, *et al.* Liver-enriched transcription factors in liver function and development. Part I and Part II. Pharmacol Rev, v.54, n.1, Mar, p.129-330. 2002.

Schug, J. e G. C. Overton. Modeling transcription factor binding sites with Gibbs Sampling and Minimum Description Length encoding. Proc Int Conf Intell Syst Mol Biol, v.5, p.268-71. 1997.

Schuppan, D., T. Cramer, *et al.* Hepatocytes as a source of collagen type XVIII endostatin. Lancet, v.352, n.9131, Sep 12, p.879-80. 1998.

Sertie, A. L., M. Quimby, *et al.* A gene which causes severe ocular alterations and occipital encephalocele (Knobloch syndrome) is mapped to 21q22.3. Hum Mol Genet, v.5, n.6, Jun, p.843-7. 1996.

Sertie, A. L., V. Sossi, *et al.* Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome). Hum Mol Genet, v.9, n.13, Aug 12, p.2051-8. 2000.

Silvestri, L., L. Sonzogni, *et al.* CYP enzyme polymorphisms and susceptibility to HCV-related chronic liver disease and liver cancer. Int J Cancer, v.104, n.3, Apr 10, p.310-7. 2003.

Smale, S. T. Core promoters: active contributors to combinatorial gene regulation. Genes Dev, v.15, n.19, Oct 1, p.2503-8. 2001.

Sund, M., Y. Hamano, *et al.* Function of endogenous inhibitors of angiogenesis as endothelium-specific tumor suppressors. Proc Natl Acad Sci U S A, v.102, n.8, Feb 22, p.2934-9. 2005.

Suzuki, O. T., A. L. Sertie, *et al.* Molecular analysis of collagen XVIII reveals novel mutations, presence of a third isoform, and possible genetic heterogeneity in Knobloch syndrome. Am J Hum Genet, v.71, n.6, Dec, p.1320-9. 2002.

Tanaka, T., N. Yoshida, *et al.* Defective adipocyte differentiation in mice lacking the C/EBPbeta and/or C/EBPdelta gene. Embo J, v.16, n.24, Dec 15, p.7432-43. 1997.

Tanigawa, N., C. Lu, *et al.* Quantitation of sinusoid-like vessels in hepatocellular carcinoma: its clinical and prognostic significance. Hepatology, v.26, n.5, Nov, p.1216-23. 1997.

Taylor, J. A., D. A. Bell, *et al.* L-myc proto-oncogene alleles and susceptibility to hepatocellular carcinoma. Int J Cancer, v.54, n.6, Jul 30, p.927-30. 1993.

Teh, S. H., A. D. Hill, *et al.* Raised plasma endostatin levels correlate inversely with breast cancer angiogenesis. J Surg Res, v.116, n.1, Jan, p.165-71. 2004.

Thomas, M. J. e E. Seto. Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? Gene, v.236, n.2, Aug 20, p.197-208. 1999.

Thorgeirsson, S. S. e J. W. Grisham. Molecular pathogenesis of human hepatocellular carcinoma. Nat Genet, v.31, n.4, Aug, p.339-46. 2002.

Tomizawa, M., K. Watanabe, *et al.* Down-regulated expression of the CCAAT/enhancer binding protein alpha and beta genes in human hepatocellular carcinoma: a possible prognostic marker. Anticancer Res, v.23, n.1A, Jan-Feb, p.351-4. 2003.

Tsuboi, S., K. Nouse, *et al.* Anti-endostatin monoclonal antibody enhances growth of human hepatocellular carcinoma cells by inhibiting activity of endostatin secreted by the transplanted cells in nude mice. Int J Oncol, v.25, n.5, Nov, p.1267-71. 2004.

Utriainen, A., R. Sormunen, *et al.* Structurally altered basement membranes and hydrocephalus in a type XVIII collagen deficient mouse line. Hum Mol Genet, v.13, n.18, Sep 15, p.2089-99. 2004.

Wasserman, W. W. e A. Sandelin. Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet, v.5, n.4, Apr, p.276-87. 2004.

Wen, W., M. A. Moses, *et al.* The generation of endostatin is mediated by elastase. Cancer Res, v.59, n.24, Dec 15, p.6052-6. 1999.

Wodarz, A. e R. Nusse. Mechanisms of Wnt signaling in development. Annu Rev Cell Dev Biol, v.14, p.59-88. 1998.

Wray, G. A. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet, v.8, n.3, Mar, p.206-16. 2007.

Xiao, W. e P. J. Oefner. Denaturing high-performance liquid chromatography: A review. Hum Mutat, v.17, n.6, Jun, p.439-74. 2001.

Xu, L., L. Hui, *et al.* Expression profiling suggested a regulatory role of liver-enriched transcription factors in human hepatocellular carcinoma. Cancer Res, v.61, n.7, Apr 1, p.3176-81. 2001.

Xu, Y. K. e R. Nusse. The Frizzled CRD domain is conserved in diverse proteins including several receptor tyrosine kinases. Curr Biol, v.8, n.12, Jun 4, p.R405-6. 1998.

Yan, H., W. Yuan, *et al.* Allelic variation in human gene expression. Science, v.297, n.5584, Aug 16, p.1143. 2002.

Zetter, B. R. Angiogenesis and tumor metastasis. Annu Rev Med, v.49, p.407-24. 1998.

Zorick, T. S., Z. Mustacchi, *et al.* High serum endostatin levels in Down syndrome: implications for improved treatment and prevention of solid tumours. Eur J Hum Genet, v.9, n.11, Nov, p.811-4. 2001.

ANEXO I

Mutations in Collagen 18A1 (*COL18A1*) and their relevance to the human phenotype

MARIA RITA PASSOS-BUENO¹, OSCAR T. SUZUKI¹, LUCIA M. ARMELIN-CORREA¹,
ANDRÉA L. SERTIÉ¹, FLAVIA IMBROISI V. ERRERA¹, KELLY BAGATINI¹,
FERNANDO KOK^{1,2} and KATIA R.M. LEITE³

¹Centro de Estudos do Genoma Humano, Departamento de Genética e Biologia Evolutiva,
Instituto de Biociências, USP, Rua do Matão, 277, 05508-900 São Paulo, SP, Brasil

²Serviço de Neurologia Infantil, Faculdade de Medicina, USP,
Av. Dr. Arnaldo, 455, 01246-903 São Paulo, SP, Brasil

³Laboratório de Patologia Cirúrgica e Molecular, Hospital Sírio Libanês,
Rua Dona Adma Jafet, 91, 01308-050 São Paulo, SP, Brasil

*Manuscript received on May 12, 2005; accepted for publication on May 16, 2005;
presented by LUCIA MENDONÇA PREVIATO*

ABSTRACT

Collagen XVIII, a proteoglycan, is a component of basement membranes (BMs). There are three distinct isoforms that differ only by their N-terminal, but with a specific pattern of tissue and developmental expression. Cleavage of its C-terminal produces endostatin, an inhibitor of angiogenesis. In its N-terminal, there is a frizzled motif which seems to be involved in Wnt signaling. Mutations in this gene cause Knobloch syndrome (KS), an autosomal recessive disorder characterized by vitreoretinal and macular degeneration and occipital encephalocele. This review discuss the effect of both rare and polymorphic alleles in the human phenotype, showing that deficiency of one of the collagen XVIII isoforms is sufficient to cause KS and that null alleles causing deficiency of all collagen XVIII isoforms are associated with a more severe ocular defect. This review besides illustrating the functional importance of collagen XVIII in eye development and its structure maintenance throughout life, it also shows its role in other tissues and organs, such as nervous system and kidney.

Key words: *COL18A1*, collagen XVIII, Knobloch syndrome, eye development, neuronal cell migration, craniosynostosis, polymorphisms, D1437N.

INTRODUCTION

Type XVIII collagen is a non-fibrillar proteoglycan collagen which, together with collagen XV, form the subgroup of multiplexins (multiple triple-helix domains with interruptions) within the collagen family (Rehn and Pihlajaniemi 1994, Oh et al. 1994, Saarela et al. 1998). This protein forms homotrimers

and contains an N-terminal non-collagenous domain (NC-11), 10 collagenous repeats alternated with 9 non-collagenous repeats and a C-terminal non-collagenous region, NC-1 (Saarela et al. 1998). The N-terminal region presents thrombospondin-1-like and frizzled motifs. This last motif is characterized by a conserved region of 120 amino acids containing a sequence of 10 invariantly spaced cysteines, the cysteine-rich domain or CRD (Xu and Nusse 1998). The frizzled domain binds Wnt signaling

Correspondence to: Maria Rita Passos-Bueno
E-mail: passos@ib.usp.br; passosbuenom@terra.com.br

molecules, which are involved in cell proliferation, polarity and differentiation in various developmental processes (Wodarz and Nusse 1998). However, the function of the frizzled motif in collagen XVIII molecule is still not known (Lin et al. 2001, Elamaa et al. 2003). The C-terminal region NC1, which contains a trimerization domain and the last 184 amino acid residues, can release endostatin (20 kDa) and endostatin-like fragments by its cleavage with elastase and/or cathepsin L (Wen et al. 1999, Felbor et al. 2000, John et al. 2005). These fragments can be detected in blood plasma and tissue extracts (Sasaki et al. 1998), during physiological or pathological processing of collagen XVIII (Zorick et al. 2001, Strik et al. 2001, Dhar et al. 2002). Endostatin, first identified as a potent inhibitor of endothelial cell proliferation and migration, with the ability to reduce tumor growth, seems to play several other functions, including apoptosis induction (Dhanabal et al. 1999, Dixelius et al. 2000, Schmidt et al. 2004). Endostatin/collagen XVIII also acts in neuronal and renal epithelial cell migration (Ackley et al. 2001, Karihaloo et al. 2001, Lin et al. 2001) and interacts with at least two cell surface receptors (glypican and integrins), but its downstream targets are still unknown (Rehn et al. 2001, Karumanchi et al. 2001, Wickstrom et al. 2002, Abdollahi et al. 2004, Wickstrom et al. 2004). Besides the C-terminal proteolytic fragments, there is also a soluble type XVIII collagen fragment derived from the NC11 domain, which contains the frizzled motif, with still unknown function (Elamaa et al. 2003). Most of the research on collagen XVIII was originally focused on endostatin because of its great potential to inhibit tumor angiogenesis. But it is also known that proteoglycans in the extracellular space play a major role in chemical signaling between cells as they can provide a reservoir for various secreted signaling molecules such as members of wingless/wnt, TGF- β and FGF families and cell adhesion molecules-like (Lin et al. 2001, Aricescu et al. 2002). Therefore, collagen XVIII in the extracellular matrix may be important for defining plas-

ticity of epithelial and endothelial cells in different organs and may be involved in determining cell polarity, influencing cell metabolism and differentiation (Lin et al. 2001).

Collagen XVIII is encoded by the *COL18A1* gene, mapped to 21q22.3. This gene has 43 exons, which are transcribed in 3 distinct isoforms by use of two promoters and of an alternative splicing of the third exon. The three proteins differ only by their signal peptides and by part of the N-terminal region of the NC11 domain, and contain 303 (NC11-303), 493 (NC11-493) and 728 (NC11-728) amino acid residues. NC11-303 mRNA is transcribed from the promoter located upstream of exon 1, and it does not contain exon 3. NC11-493 and NC11-728 mRNAs are both transcribed by the promoter located in intron 2 and do not include exons 1 and 2. Isoform NC11-728 and NC11-493 are encoded by the entire exon 3 and exons 4 through 43, but NC11-493 variant uses an internal splice donor site within exon 3, thus lacking 235 amino acid residues (Elamaa et al. 2003). The NC11-493 variant mRNA is mainly expressed in fetal and adult liver, whereas the NC11-303 variant was found at the highest levels in the fetal kidney (Saarela et al. 1998). NC11-728, is the only isoform containing the frizzled domain, it is also expressed in several tissues, with the highest levels in liver and lung (Elamaa et al. 2003). NC11-303 and NC11-728 are also expressed in human fetal brain and retina (Sertié et al. 2000, Suzuki et al. 2002). It seems that the expression pattern of the collagen XVIII isoforms change during development (Lin et al. 2001), suggesting that these molecules play important roles during development.

In 2000, we demonstrated that a mutation in the *COL18A1* gene causes an autosomal recessive disorder, Knobloch syndrome, showing for the first time the functional importance of the collagen XVIII molecule in humans (Sertié et al. 2000). We will focus this review on the distribution of pathogenic and polymorphic mutations along the *COL18A1* gene, their corresponding phenotypes and

the possible functional roles of collagen XVIII based on the human and mice knockout phenotype.

PATHOGENIC MUTATIONS AND KNOBLOCH SYNDROME (KS)

PHENOTYPE

Knobloch syndrome (KS; MIM 267750) is an autosomal recessive condition defined by the occurrence of high myopia, usually evidenced up to the first year of life, vitreoretinal degeneration with retinal detachment, macular abnormalities and occipital encephalocele. Clinical variability is present but, to date, all patients have severe ocular abnormalities that usually lead to bilateral blindness. The occipital encephalocele is also a major clinical feature that has been described in 23 of 25 reported cases; its size varies among affected patients and sometimes it is noticed only through CT scan (reviewed in Suzuki et al. 2002). Altered retinal pigmentation is a prominent finding and more recently it was described a case of persistent fetal vasculature in the eye of a patient with KS (Seaver et al. 1993, Passos-Bueno et al. 1994, Kliemann et al. 2003, Duh et al. 2004).

Heterotopic nodules in the lateral ventricles and neuronal migration abnormalities in three KS patients have recently been noticed (Kliemann et al. 2003, Fernando Kok, personal communication). In two of these patients epilepsy was also present, suggesting that these clinical complication might be more common in KS patients than in the general population. In one patient, sagittal craniosynostosis was also present; confirmation of this association in additional cases will expand the spectrum of clinical variability in KS.

Other minor clinical abnormalities, such as lens subluxation, cataracts, flat nasal bridge, mid-face hypoplasia, bilateral epicanthic folds, hypoplasia of the right lung with anomalous pulmonary return, cardiac dextroversion, generalized hyperextensibility of the joints, unusual palmar crease and unilateral duplicated renal collecting system, have been observed in a few cases (reviewed in Suzuki

et al. 2002). The spectrum of clinical variability of KS is still not defined since most of each of the above mentioned clinical alterations were reported only once in single families and before the identification of the gene.

IDENTIFICATION OF MUTATIONS IN THE *COL18A1* AS CAUSATIVE OF KS

Based on the study of a single large inbred Brazilian family with 12 individuals affected with KS, we assigned the disease gene to 21q22.3 and demonstrated that a homozygous mutation at the AG consensus acceptor splice site of *COL18A1* intron 1 (IVS1-2A>T) causes Knobloch syndrome (Sertié et al. 2000). This mutation predicts the skipping of exon 2 and creation of a stop codon in exon 4 and, therefore, the truncation of only the $\alpha 1$ (XVIII) collagen NC11-303 (Sertié et al. 2000). Patients with this mutation present endostatin plasma levels within the normal range (Suzuki et al. 2002). This is expected as the isoform NC1-493, which produces most of the circulating endostatin, should not be altered by the presence of the IVS1-2A>T mutation (Sertié et al. 2000, Suzuki et al. 2002).

Pathogenic mutations in the *COL18A1* gene were found in 6 out of 9 other unrelated families with KS (3 Brazilian, 3 north-American and 3 European), thus confirming that mutations in this gene causes KS. A total of 7 different pathogenic changes (Table I) were identified: three patients were compound heterozygotes and two homozygotes (Suzuki et al. 2002, unpublished data). Endostatin plasma levels, measured in 2 patients with null mutations, showed levels in the lower range, suggesting that these alterations lead to absence of the protein (Suzuki et al. 2002).

We did not identify mutations in three unrelated probands with the major typical clinical signs of the syndrome but, endostatin plasma levels, measured in two of them, showed values in the lower range (Suzuki et al. 2002, Kliemann et al. 2003). These data, therefore, suggest that the KS phenotype can be caused by an unusual mutational mechanism in *COL18A1* or to non-allelic genetic

TABLE I
Pathogenic mutations found in KS patients.
Nucleotide numbers are based on the reference sequence AF018081.

Family	Pathogenic mutations detected
KS1	c.12-2A>T* (homozygous)
KS3	c.2969-2978delCAGGGCCCCC (maternal); c.3514-3515delCT (paternal)
KS4	c.1238-1239insA (maternal); c.3514-3515delCT (paternal)
KS5	c.3514-3515delCT (maternal); c.2105delC (paternal)
KS8	IVS1-2A>T (homozygous)
KS9	c.3277C>T (homozygous)
KS10	c.2416C>T (homozygous)

*mutation position based on the mRNA sequence AF018082.

heterogeneity, as also suggested by others (Menzel et al. 2004).

We also did not find pathogenic mutations in KS patient who has high myopia associated with several other clinical features that were not typical of the syndrome: a scalp defect in the frontal region, developmental delay, telecanthus, hypertelorism, and high arched palate (Sniderman et al. 2000, Suzuki et al. 2002). This patient presented endostatin normal plasma levels. This finding suggests that this patient's phenotype is probably caused by mutations at another locus, and that mutations in *COL18A1* cause very specific ocular and neuronal cell migration abnormalities. Therefore, the minimum clinical features that define the diagnose of KS are high myopia and occipital encephalocele.

GENOTYPE-PHENOTYPE CORRELATION IN KS

All pathogenic mutations in the *COL18A1* gene to date probably lead to mRNA decay or to a truncated protein thus the disease might be caused by deficiency of a functional protein. So, these mutations represent human knockout models for the *COL18A1* gene: homozygotes for the IVS1-2T>A represent a knockout of only the short isoform, while the other mutations represent knockout of all collagen XVIII isoforms and endostatin.

Patients with deficiency of only the short isoform present the main clinical features of the syndrome, suggesting that the lack of this isoform is enough to cause the phenotype. It also seems that lack of only the short isoform is associated with a less severe ocular prognosis. These observations imply that collagen XVIII isoforms may play important functional roles in the human eye. The ocular alterations observed in KS patients are complex and, for obvious reasons, it is difficult to obtain a detailed histology and electron microscopic characterization of a KS patient's eye. Thus, the histopathological and the molecular changes resulting from null *COL18A1* alleles in humans remain unknown and the study of mice models will be of great importance to elucidate them. Indeed, there has been a growing number of data suggesting that the Knockout mice of *col18a1(col18a1-/-)* constitute a promising model to study KS. These mice are viable and fertile, do not present occipital encephalocele but show abnormal visual function due to several ocular alterations which worsen with increasing age (Fukai et al. 2002, Marneros and Olsen 2003, Marneros et al. 2004). Developmental defects in hyaloid vessel regression affecting postnatal levels of VEGF expression in the neural retina were observed in these mice (Fukai et al. 2002). Interestingly, Duh et al. (2004) described persistence of

fetal vasculature in a patient with clinical features of Knobloch syndrome associated with no detectable plasma levels of endostatin. The iris basement membrane of the *col18a1*^{-/-} mice was also shown to be altered (Marneros and Olsen 2003, Marneros et al. 2004), further confirming that iris modifications in some KS patients are due to the deficiency of collagen XVIII (Passos-Bueno et al. 1994, Duh et al. 2004). Another important finding in the mice *col18a1*^{-/-} is the age-dependent accumulation of electron-dense deposits between the retinal pigment epithelium (RPE) and Bruch's membrane. These deposits contain excess basement membrane material and they were associated with an abnormal vitamin A metabolism in the retinal pigment epithelium with reduced rhodopsin levels (Marneros et al. 2004). These findings further confirm that endostatin/*col18a1* plays a significant role in blood vessels remodeling and morphogenesis in the eye.

Variability of the size of the occipital alteration is commonly observed both in intra and inter-familial cases, but it does not have an obvious correlation with the site of the mutation in *COL18A1* gene.

Pathogenic mutations leading to deficiency of all collagen XVIII isoforms were found in two out of 3 KS patients who, besides minimum KS clinical features, also had abnormal neuronal cell migration and recurrent epilepsy. None of the 12 patients, all in middle age, with lack of only the short form of collagen XVIII has presented epilepsy; thus, we could speculate that the median and/or long collagen XVIII isoforms, in opposition to the short form, play an important role in neuronal cell migration. Alterations in neuronal cell migration have also been observed in *col18a1* knockout animal models (Ackley et al. 2001). The current data, therefore, strongly support an important functional role of collagen XVIII in neuronal development. It is possible that the occipital encephalocele actually represents abnormal cell migration instead of a defect in bone or suture development at the occipital region.

Col18a1^{-/-} mice, depending on the genetic background, develop a significant enlargement of the skull associated with severe hydrocephalus. This clinical alteration presents incomplete penetrance and wide expressivity, with some mice presenting just enlargement of the ventricles visible only through MRI (Utriainen et al. 2004). This phenotype was not noticed in KS patients, maybe due to the small number of cases studied to date.

One of the patients (homozygous for the mutation c.2416 C > T; numbered according to the mRNA sequence AF018081, deposited in GenBank) has a duplicated kidney (data not reported). Lin et al. (2001) have shown that collagen XVIII expression is involved in kidney and lung morphogenesis, through still unknown mechanisms. Although no major kidney defect was observed among the mutant mice, they showed broadened kidney tubular basal membrane, altered mesangial matrix of the glomerulus and elevated serum creatinine levels (Utriainen et al. 2004). These findings further support the importance of collagen XVIII in kidney development and function. Therefore, KS patients should be carefully examined for functional and morphological kidney defects.

None of the patients with mutations in the *COL18A1* gene have any major liver defect, which is surprising because one of the isoforms (NC11-493) is highly expressed in this organ.

POLYMORPHIC CHANGES

We identified twenty polymorphic variants in the *COL18A1* gene (Iughetti et al. 2001, NCBI dbSNP database); of these, only two represent missense changes (I841V and D1437N). D1437N (previously referred as D104N) change is located in the COOH-terminal globular domain NC1 of collagen XVIII, the encoding region of endostatin. The aspartic acid at position 1437 (or amino acid residue 104 of endostatin) is conserved in human and mouse endostatin/collagen XVIII, as well as, in the endostatin-like molecule produced from the NC1 domain of collagen XV (Iughetti et al. 2001). The

structural modeling analysis of the human mutated 1437N endostatin suggests that this amino acid residue is located at the surface of the molecule and does not change endostatin stability but can modify the charged distribution surrounding the mutated aminoacid. (Iughetti et al. 2001).

It has been postulated that endostatin/collagen XVIII play important roles in tumor growth due to its anti-angiogenic activity. Indeed, we have observed that Down syndrome (DS) patients, who have a decreased incidence of solid tumors, showed higher serum levels of endostatin (Zorick et al. 2001). Still others have shown that patients with a better prognosis of hepatocellular carcinoma presented higher collagen XVIII expression in tumor cells (Musso et al. 2001) and lower production of the molecule by the surrounding stroma (Hu et al. 2004). Therefore, endostatin and collagen XVIII are good candidate proteins to be studied in the susceptibility of tumor growth. In this regard, we have hypothesized that the SNP D1437N (or D104N/endostatin) could represent an at-risk susceptibility allele for progression or aggressiveness of solid tumors. In a case-control study including 181 prostate cancer cases and 198 non-cancer individuals, we observed that the heterozygous N104 individuals have a 2.5 times increased chance of developing prostate cancer as compared with homozygous D104 subjects (Iughetti et al. 2001). However, this association was not confirmed by us in another sample (data not published) as well as by others including different solid tumors (Ortega et al. 2003, Liu et al. 2003, Nascimento et al. 2004, Macpherson et al. 2004). It is possible that the variation of the genotypic frequencies observed in our original report just represent racial stratification. Further studies will be necessary to understand the role of collagen XVIII/endostatin and predisposition to solid tumors.

CONCLUSIONS

Collagen XVIII, encoded by the *COL18A1* gene at 21q22.3, is a very interesting molecule, which produces at least two proteolytic fragments: one de-

rived from its C-terminal, endostatin and the other from its N-terminal, containing the frizzled motif. There are at least 3 distinct isoforms, which only differ at their N-terminal region; each of them present a specific tissue and development pattern of expression and possibly perform distinct functions.

Null alleles in the collagen XVIII lead to Knobloch Syndrome (KS), a rare autosomal recessive disorder mainly characterized by high myopia, vitreoretinal degeneration, macular alteration and occipital encephalocele. Iris and retinal pigment epithelium alteration seem also to be important features of the syndrome. The complete spectrum of the disease is still not established, but high myopia up to the first year of life and occipital encephalocele are the two minimal clinical criteria to classify a patient as KS. Collagen XVIII is a critical component of the basement membrane of the iris, vitreo and retina and its presence is fundamental for normal eye development during embryogenesis; this protein also seems to play important functional roles in neuronal cell migration and as a component of basement membrane of kidney. Polymorphisms in collagen XVIII are common, but their relationship to susceptibility of solid tumors are still controversial.

ACKNOWLEDGMENTS

The authors would like to thank the patients and their family for your enthusiasm and collaboration in the research. Authors are supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Centro de Pesquisa, Inovação e Difusão (CEPID), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

RESUMO

Colágeno XVIII, uma proteoglicana, é um componente das membranas basais (MBs). Existem três isoformas distintas que diferem apenas na região N-terminal, mas que apresentam um padrão específico de expressão nos diferentes tecidos e durante o desenvolvimento. A clivagem da região C-terminal produz endostatina, um inibidor de angiogênese. Na sua região N-terminal, há um motivo

“frizzled”, o qual parece estar envolvido com a sinalização de Wnt. Mutações no gene *COL18A1* causam a síndrome de Knobloch (SK), uma condição de herança autossômica recessiva caracterizada por degeneração vítreo – retiniana, degeneração de mácula e encefalocele occipital. Esta revisão discute o efeito tanto de alelos raros como polimórficos no fenótipo, mostrando que deficiência de uma das isoformas de colágeno XVIII é suficiente para causar SK e que alelos nulos causando deficiência de todas as isoformas de colágeno XVIII estão associadas a alterações oculares mais graves. Esta revisão, além de ilustrar a importância funcional do colágeno XVIII no desenvolvimento do olho e na manutenção de sua estrutura, também mostra que esta proteína tem um papel funcional importante em outros tecidos e órgão, como no sistema nervoso central e rim.

Palavras-chave: COL18A1, colágeno XVIII, síndrome de Knobloch, desenvolvimento ocular, migração de célula neuronal, polimorfismos, D1437N.

REFERENCES

- ABDOLLAHI A, HAHNFELDT P, MAERCKER C, GRONE HJ, DEBUS J, ANSORGE W, FOLKMAN J, HLATKY L AND HUBER PE. 2004. Endostatin's antiangiogenic signalling network. *Mol Cell* 13: 649–663.
- ACKLEY BD, CREW JR, ELAMAA H, PIHLAJANIEMI T, KUO CJ AND KRAMER JM. 2001. The NC1/endostatin domain of *Caenorhabditis elegans* type XVIII collagen affects cell migration and axon guidance. *J Cell Biol* 152: 1219–1232.
- ARICESCU AR, MCKINNELL IW, HALFTER W AND STOKER AW. 2002. Heparan sulfate proteoglycans are ligands for receptor protein tyrosine phosphatase sigma. *Mol Cell Biol* 22: 1881–1892.
- DHANABAL M, RAMCHANDRAN R, WATERMAN MJ, LU H, KNEBELMANN B, SEGAL M AND SUKHATME VP. 1999. Endostatin induces endothelial cell apoptosis. *J Biol Chem* 274: 11721–11726.
- DHAR DK, ONO T, YAMANOI A, SODA Y, YAMAGUCHI E, RAHMAN MA, KOHNO H AND NAGASUE N. 2002. Serum endostatin predicts tumor vascularity in hepatocellular carcinoma. *Cancer* 95: 2188–2195.
- DIXELIUS J, LARSSON H, SASAKI T, HOLMQVIST K, LU L, ENGSTROM A, TIMPL R, WELSH M AND CLAESSON-WELSH L. 2000. Endostatin-induced tyrosine kinase signalling through the Shb adaptor protein regulates endothelial cell apoptosis. *Blood* 95: 3403–3411.
- DUH EJ, YAO YG, DAGLI M AND GOLDBERG MF. 2004. Persistence of fetal vasculature in a patient with Knobloch syndrome: potential role for endostatin in fetal vascular remodeling of the eye. *Ophthalmology* 111: 1885–1888.
- ELAMAA H, SNELLMAN A, REHN M, AUTIO-HARMAINEN H AND PIHLAJANIEMI T. 2003. Characterization of the human type XVIII collagen gene and proteolytic processing and tissue location of the variant containing a frizzled motif. *Matrix Biol* 22: 427–442.
- FELBOR U, DREIER L, BRYANT RA, PLOEGH HL, OLSEN BR AND MOTHES W. 2000. Secreted cathepsin L generates endostatin from collagen XVIII. *EMBO J* 19: 1187–1194.
- FUKAI N ET AL. 2002. Lack of collagen XVIII/endostatin results in eye abnormalities. *EMBO J* 21: 1535–1544.
- HU TH, HUANG CC, WU CL, LIN PR, LIU SY, LIN JW, CHUANG JH AND TAI MH. 2004. Increased endostatin/collagen XVIII expression correlates with elevated VEGF level and poor prognosis in hepatocellular carcinoma. *Mod Pathol* [Epub ahead of print].
- IUGHETTI P ET AL. 2001. A polymorphism in endostatin, an angiogenesis inhibitor, predisposes for the development of prostatic adenocarcinoma. *Cancer Res* 61: 7375–7378.
- JOHN H, RADTKE K, STANDKER L AND FORSSMANN WG. 2005. Identification and characterization of novel endogenous proteolytic forms of the human angiogenesis inhibitors restin and endostatin. *Biochim Biophys Acta* 1747: 161–170.
- KARIHALOO A ET AL. 2001. Endostatin regulates branching morphogenesis of renal epithelial cells and ureteric bud. *Proc Natl Acad Sci USA* 98: 12509–12514.
- KARUMANCHI AS ET AL. 2001. Cell surface glypicans are low-affinity endostatin receptors. *Mol Cell* 7: 811–822.

- KLIEMANN SE, WAETGE RT, SUZUKI OT, PASSOS-BUENO MR AND ROSEMBERG S. 2003. Evidence of neuronal migration disorders in Knobloch syndrome: clinical and molecular analysis of two novel families. *Am J Med Genet A* 119: 15–19.
- KUO CJ ET AL. 2001. Oligomerization-dependent regulation of motility and morphogenesis by the collagen XVIII NC1/endostatin domain. *J Cell Biol* 152: 1233–1246.
- LIN Y, ZHANG S, REHN M, ITARANTA P, TUUKKANEN J, HELJASVAARA R, PELTOKETO H, PIHLAJANIEMI T AND VAINIO S. 2001. Induced repatterning of type XVIII collagen expression in ureter bud from kidney to lung type: association with sonic hedgehog and ectopic surfactant protein C. *Development* 128: 1573–1585.
- LIU TC, PENG CT, LIN SF, CHANG CS, CHEN TP AND CHANG JG. 2003. Association of endostatin D104N with leukemia. *Kaohsiung J Med Sci* 19: 1–5.
- MACPHERSON GR, SINGH AS, BENNETT CL, VENZON DJ, LIEWEHR DJ, FRANKS ME, DAHUT WL, KANTOFF PW, PRICE DK AND FIGG WD. 2004. Genotyping and Functional Analysis of the D104N Variant of Human Endostatin. *Cancer Biol Ther* 3 [Epub ahead of print].
- MARNEROS AG AND OLSEN BR. 2003. Age-dependent iris abnormalities in collagen XVIII/endostatin deficient mice with similarities to human pigment dispersion syndrome. *Invest Ophthalmol Vis Sci* 44: 2367–2372.
- MARNEROS AG ET AL. 2004. Collagen XVIII/endostatin is essential for vision and retinal pigment epithelial function. *EMBO J* 23: 89–99.
- MENZEL O ET AL. 2004. Knobloch syndrome: novel mutations in COL18A1, evidence for genetic heterogeneity, and a functionally impaired polymorphism in endostatin. *Hum Mutat* 23: 77–84.
- MUSSO O, REHN M, THERET N, TURLIN B, BIOULAC-SAGE P, LOTRIAN D, CAMPION JP, PIHLAJANIEMI T AND CLEMENT B. 2001. Tumor progression is associated with a significant decrease in the expression of the endostatin precursor collagen XVIII in human hepatocellular carcinomas. *Cancer Res* 61: 45–49.
- NASCIMENTO H, RODRIGUES COY CS, NAVARRO GOES JR, FERREIRA COSTA F AND PASSOS LIMA CS. 2004. A polymorphism in the angiogenesis inhibitor, endostatin, in sporadic colorectal adenocarcinoma. *Int J Colorectal Dis* 19: 499–501.
- OH SP, WARMAN ML, SELDIN MF, CHENG SD, KNOLL JH, TIMMONS S AND OLSEN BR. 1994. Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localization of the alpha 1(XVIII) collagen gene to mouse chromosome 10 and human chromosome 21. *Genomics* 19: 494–499.
- ORTEGA MM, NASCIMENTO H, COSTA FF, TEORI MT AND LIMA CS. 2003. A polymorphism in the angiogenesis inhibitor, endostatin, in multiple myeloma. *Leuk Res* 27: 93–94.
- PASSOS-BUENO MR, MARIE SK, MONTEIRO M, NEUSTEIN I, WHITTLE MR, VAINZOF M AND ZATZ M. 1994. Knobloch syndrome in a large Brazilian consanguineous family: confirmation of autosomal recessive inheritance. *Am J Med Genet* 52: 170–173.
- REHN M AND PIHLAJANIEMI T. 1994. Alpha 1(XVIII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen. *Proc Natl Acad Sci USA* 91: 4234–4238.
- REHN M, VEIKKOLA T, KUKK-VALDRE E, NAKAMURA H, ILMONEN M, LOMBARDO C, PIHLAJANIEMI T, ALITALO K AND VUORI K. 2001. Interaction of endostatin with integrins implicated in angiogenesis. *Proc Natl Acad Sci USA* 98: 1024–1029.
- SAARELA J, YLIKARPPA R, REHN M, PURMONEN S AND PIHLAJANIEMI T. 1998. Complete primary structure of two variant forms of human type XVIII collagen and tissue-specific differences in the expression of the corresponding transcripts. *Matrix Biol* 16: 319–328.
- SASAKI T, FUKAI N, MANN K, GOHRING W, OLSEN BR AND TIMPL R. 1998. Structure, function and tissue forms of the C-terminal globular domain of collagen XVIII containing the angiogenesis inhibitor endostatin. *EMBO J* 17: 4249–4256.
- SCHMIDT A, WENZEL D, FERRING I, KAZEMI S, SASAKI T, HESCHELER J, TIMPL R, ADDICKS K, FLEISCHMANN BK AND BLOCH W. 2004. Influence of endostatin on embryonic vasculo- and angiogenesis. *Dev Dyn* 230: 468–480.

- SEAVER LH, JOFFE L, SPARK RP, SMITH BL AND HOYME HE. 1993. Congenital scalp defects and vitreoretinal degeneration: redefining the Knobloch syndrome. *Am J Med Genet* 46: 203–208.
- SERTIÉ AL, SOSSI V, CAMARGO AA, ZATZ M, BRAHE C AND PASSOS-BUENO MR. 2000. Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome). *Hum Mol Genet* 9: 2051–2058.
- SNIDERMAN LC, KOENEKOOP RK, O'GORMAN AM, USHER RH, SUFRATEGUI MR, MOROZ B, WATERS GV AND DER KALOUSTIAN VM. 2000. Knobloch syndrome involving midline scalp defect of the frontal region. *Am J Med Genet* 90: 146–149.
- STRIK HM, SCHLUESENER HJ, SEID K, MEYERMANN R AND DEININGER MH. 2001. Localization of endostatin in rat and human gliomas. *Cancer* 91: 1013–1019.
- SUZUKI OT ET AL. 2002. Molecular analysis of collagen XVIII reveals novel mutations, presence of a third isoform, and possible genetic heterogeneity in Knobloch syndrome. *Am J Hum Genet* 71: 1320–1329.
- UTRIAINEN A, SORMUNEN R, KETTUNEN M, CARVALHAES LS, SAJANTI E, EKLUND L, KAUPPINEN R, KITTEN GT AND PIHLAJANIEMI T. 2004. Structurally altered basement membranes and hydrocephalus in a type XVIII collagen deficient mouse line. *Hum Mol Genet* 13: 2089–2099.
- WEN W, MOSES MA, WIEDERSCHAIN D, ARBISER JL AND FOLKMAN J. 1999. The generation of endostatin is mediated by elastase. *Cancer Res* 59: 6052–6056.
- WICKSTROM SA, ALITALO K AND KESKI-OJA J. 2002. Endostatin associates with integrin $\alpha 5\beta 1$ and caveolin-1, and activates Src via a tyrosyl phosphatase-dependent pathway in human endothelial cells. *Cancer Res* 62: 5580–5589.
- WICKSTROM SA, ALITALO K AND KESKI-OJA J. 2004. An endostatin-derived peptide interacts with integrins and regulates actin cytoskeleton and migration of endothelial cells. *J Biol Chem* 279: 20178–20185.
- WODARZ A AND NUSSE R. 1998. Mechanisms of Wnt signalling in development. *Annu Rev Cell Dev Biol* 14: 59–88.
- XU YK AND NUSSE R. 1998. The Frizzled CRD domain is conserved in diverse proteins including several receptor tyrosine kinases. *Curr Biol* 8: R405–406.
- ZORICK TS, MUSTACCHI Z, BANDO SY, ZATZ M, MOREIRA-FILHO CA, OLSEN B AND PASSOS-BUENO MR. 2001. High serum endostatin levels in Down syndrome: implications for improved treatment and prevention of solid tumours. *Eur J Hum Genet* 9: 811–814.

ANEXO II

A functional SNP in the promoter region of *TCOF1* is associated with reduced gene expression and YY1 DNA–protein interaction

Cibele Masotti ^a, Lucia M. Armelin-Correa ^a, Alessandra Splendore ^b, Chin J. Lin ^{c,d},
Angela Barbosa ^{c,e}, Mari C. Sogayar ^f, Maria Rita Passos-Bueno ^{a,*}

^aInstituto de Biociências, Universidade de São Paulo, Departamento de Biologia, Rua do Matão, 277/Sala 200, 05508-900 São Paulo/SP, Brazil

^bInstituto Nacional do Câncer, Divisão de Genética, Rio de Janeiro, Brazil

^cLaboratório de Hormônios e Genética Molecular LIM/42, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Departamento de Clínica Médica, São Paulo, Brazil

^dHospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Departamento de Patologia, São Paulo, Brazil

^eCentro de Biotecnologia, Instituto Butantan, São Paulo, Brazil

^fInstituto de Química, Universidade de São Paulo, Departamento de Bioquímica, São Paulo, Brazil

Received 20 April 2005; received in revised form 10 June 2005; accepted 22 June 2005

Available online 18 August 2005

Received by A.J. van Wijnen

Abstract

Treacher Collins syndrome (TCS) is an autosomal dominant craniofacial malformation caused by null mutations in the *TCOF1* gene. High inter and intra familial clinical variability, ranging from mild malar hypoplasia to perinatal death due to airway collapse is observed, but, to date, no genotype–phenotype correlation has been reported. Considering haploinsufficiency as the molecular mechanism underlying the disease, we have hypothesized that mutations in the promoter region of the gene, which has never been previously characterized, in *trans* with a pathogenic mutation, could modulate the phenotype. Therefore, the aims of the present study were to determine the *TCOF1* gene's core promoter and to identify mutations in this region that could contribute to the phenotypic variation observed in this syndrome. We have delimited the minimal promoter to a region of less than 150 bp, with 63% of identity among 5 different species. We screened 1.2 kbp of the *TCOF1* 5' flanking sequence in the DNA obtained from 21 patients and 51 controls and identified four new single nucleotide polymorphisms (SNPs), one of which (–346C>T), was proved to be functional, as it decreased the promoter activity by 38%. Electrophoretic mobility shift assay (EMSA) analysis demonstrated that the –346T allele impairs DNA-binding to the YY1 transcription factor. This promoter variant represents a candidate allele to explain the clinical variability in patients bearing TCS.

© 2005 Elsevier B.V. All rights reserved.

Keywords: *TCOF1* promoter; Clinical variability in Treacher Collins syndrome; Functional SNP; Modifier allele

1. Introduction

Treacher Collins syndrome (TCS; OMIM 154500) is a craniofacial disorder that affects the development of structures derived from the embryonic first and second branchial arches (Poswillo, 1975). TCS is inherited in an autosomal dominant pattern with 60% of the cases resulting from de novo mutations (Splendore et al., 2003). TCS has variable expressivity and high penetrance, since only one unequivocal case of non-penetrance has been reported to date (Dixon et al., 1994). The syndrome

Abbreviations: *TCOF1*, Treacher Collins–Franceschetti syndrome gene; YY1, Ying Yang 1 transcription factor; EMSA, Electrophoretic Mobility Shift Assay; SNP, Single Nucleotide Polymorphisms; SSCP, Single Strand Conformation Polymorphism; dHPLC, Denaturing High-Performance Liquid Chromatography; (prime), denotes a truncated gene at the indicated side; UBF, Upstream Binding Factor; bp, basepairs; kb, 1000 basepairs; cpm, counts per minute.

* Corresponding author. Tel.: +55 11 3091 7740; fax: +55 11 3091 7419.

E-mail address: passos@ib.usp.br (M.R. Passos-Bueno).

has an incidence of 1:50,000 live births and the common clinical features include down slanting palpebral fissures with lower eyelid coloboma, hypoplasia of the mandible and the zygomatic complex, malformed ears and conductive hearing loss due to atresia of the external ear canal (Gorlin et al., 2001).

The gene associated with the syndrome, *TCOF1*, contains 28 exons and several alternative splices (The Treacher Collins Collaborative Group, 1996; Edwards et al., 1997; So et al., 2004). The *TCOF1* product, a nucleolar protein named treacle, plays a role in ribosome biogenesis and interacts with hNop56p in pre-ribosomal ribonucleoprotein complexes and also with the RNA polymerase I transcription factor UBF (Hayano et al., 2003; Valdez et al., 2004). More than 118 different pathogenic mutations were reported in the coding region of *TCOF1*, most of which are family-specific deletions or duplications causing a premature stop codon (Splendore et al., 2005). Absence of the truncated protein in fibroblasts from TCS patients suggests that null mutations in *TCOF1* probably lead to nonsense mediated mRNA decay (Isaac et al., 2000). Therefore, haploinsufficiency has been proposed as the molecular mechanism underlying the disorder. Identification of families with no pathogenic mutation in the coding region of *TCOF1* has suggested the possibility of genetic heterogeneity or the existence of different mechanisms leading to the syndrome (Isaac et al., 2000; Splendore et al., 2000, 2002, 2005).

TCS has a high intra and interfamilial phenotypic variation, ranging from perinatal death, due to compromised airways, to individuals that cannot be diagnosed on clinical grounds alone. Several attempts to evaluate the clinical variability in TCS have demonstrated no genotype–phenotype correlation in the syndrome (Edwards et al., 1997; Gladwin et al., 1996; Dixon, 1996). It has also been shown that clinical variability does not depend on the type or location of the mutation, sex or on whether the case is a sporadic or a familial one (Splendore et al., 2000; Teber et al., 2004). A recent study has demonstrated that the genetic background has a major effect on the phenotype of *Tcof1*^{+/-} mice, suggesting that variations in other genes in the same or related pathways affect the function of the protein in different individuals (Dixon and Dixon, 2004).

We hypothesized that clinical variability can be determined by the degree of *TCOF1* expression regulated by the wild-type allele. Several possible mechanisms could account for variations in treacle's expression, including mutations in the promoter region of the *TCOF1*. In this regard, we have hypothesized that mutations in *cis*-acting elements of the wild-type allele could modulate the phenotype in TCS patients. In the present work, we aimed at delimitating the *TCOF1* gene's minimal promoter region, identifying functional SNPs in 1.2 kbp of the 5' upstream region and evaluate any genotype–phenotype correlation in TCS patients.

2. Materials and methods

2.1. Patients and controls

Twenty-one Brazilian unrelated TCS patients with previously identified pathogenic mutation (Splendore et al., 2000, 2002, 2003, 2005) were screened for sequence variation in the promoter region of *TCOF1*. For each identified variation, at least 51 controls were screened. For the –346C>T SNP, 13 additional patients and 162 controls were included in the screening, since this SNP initially exhibited a higher frequency (>20%) in patients as compared to controls. We also screened three other patients (2 familial cases and 1 sporadic) for whom no pathogenic mutation in the coding region of the gene was detected through sequencing or Southern blot analysis (Splendore et al., 2005).

DNA was extracted according to standard techniques (Miller et al., 1988; Richards et al., 1993). The study protocols were previously approved by the ethical committee of our institution and informed consent was obtained from both patients and control subjects or from their legal tutors.

2.2. Mutation screening of the *TCOF1* promoter region

The 5' flanking sequence from the human *TCOF1* gene was obtained from the Homo sapiens chromosome 5 genomic contig (GenBank accession no. NT_029289) after a BLASTn (<http://www.ncbi.nlm.nih.gov/blast/>) search, using the exon 1 sequence (NM_000356). Four overlapping segments (P1, P2, P3 and P4), corresponding to 1241 bp of the upstream sequence relative to the first methionine (NT_029289), were amplified with primers available upon request.

PCR products were analyzed through SSCP and dHPLC (Transgenomic). Samples showing mobility shifts or altered peaks were sequenced in both directions in an automated sequencer (MegaBACE 1000, GE Healthcare Bio-Sciences).

To verify whether the –1025G>C and –948G>A SNPs were in *cis*, PCR products were cloned with the TOPO TA™ Cloning Kit (Invitrogen Corporation). Once we observed that they were in linkage disequilibrium, we have screened only the –1025G>C transition in the control population, by PCR, followed by digestion with the *HincII* restriction enzyme (New England Biolabs). The –346C>T SNP was screened through Single Nucleotide Primer Extension (SNUPE™ Amersham Pharmacia Biotech) with the 5'-agatgagtaaaacgcagacc-3' primer.

All the promoter mutations detected in this study were named relative to the start codon of the *TCOF1* gene (<http://www.genomic.unimelb.edu.au/mdi/mutnomen>; Den Dunnen and Antonarakis, 2000). The +1 position is located 76 nucleotides downstream of the 5' end of *TCOF1* gene (position 109000247 in the reference sequence NT_029289,

TCOF1 from 10900171 to 10942791). Hardy–Weinberg equilibrium in patients and controls was verified by χ^2 test at a 0.05 significance level.

2.3. Plasmid constructs

The luciferase reporter gene assays were performed using plasmid constructs bearing the *TCOF1* gene 5' flanking region: pGL3-1200wt (from –1241 to +56), pGL3-600wt (–571 to +56), pGL3-400wt (–369 to +56), pGL3-300wt (–258 to +56) and pGL3-200wt (–151 to +56) (Fig. 1). The 1200 bp (–1241 to +56 bp) and 600 bp (–571 to +56 bp) genomic DNA fragments were amplified from a DNA sample of a control subject with oligonucleotides containing 5' terminal restriction sites to the enzymes *Bgl*II or *Hind*III. They were inserted into the pGL3-Basic vector (Promega), in either sense (pGL3-1200wt) or anti-sense orientations (pGL3-1200wtAS). The pGL3-1200wt plasmid was used to generate four 5' deleted constructs either by digestion with *Sma*I (pGL3-300wt) or by PCR amplification using primers with a 5' terminal restriction site for the *Hind*III enzyme (pGL3-600wt, pGL3-400wt and pGL3-200wt; primer sequences available upon request). The –346C>T variant was also chosen for functional studies (pGL3-600pol). The pGL3-600pol vector's fragment was generated by PCR amplification of the same pGL3-600wt region, but using a patient's DNA sample containing the T variant in position –346. All constructs were verified by direct sequencing.

2.4. Cell culture and transient transfection assays

Human immortalized hepatocytes (HepG2—HB8065—obtained from American Type Culture Collection) were grown in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum and 1% ampicillin/streptomycin, maintained in a 5% CO₂ incubator at 37 °C. Twenty-four hours before transfection, cells were plated onto 12-well plates (1 × 10⁵/well) without antibiotics.

Samples (1 µg) of each luciferase reporter plasmid were transfected, in duplicate wells, with *Lipofectamine* 2000 (Invitrogen), according to the manufacturer's instructions. The pRL-TK plasmid (0.5 µg/well; Promega) was used to normalize the transfection efficiency. Cells were harvested 48 h after transfection, and the reporter gene expression was analyzed using the Dual-Luciferase Reporter Assay System (Promega) in a microplate luminometer (LB96V *Microlumet Plus*—EG&Berthold). The *TCOF1* gene promoter-fragments' activities were normalized relative to the activity of the pGL3-1200wt vector and expressed as relative luciferase unit. Luciferase data were generated by pooling the results of at least three experiments performed in duplicate. Statistical significance (*p*-value) was determined using Kruskal–Wallis Test, followed by Dunn's adjustment for comparing all vectors and Friedman Test to compare the pGL3-600wt and 600 pol vectors. The significance level adopted was 0.05.

2.5. In silico analysis

We used the prediction web softwares TESS MASTER ("TESS: Transcription Element Search Software on the WWW", <http://www.cbil.upenn.edu/tess>) and TFSEARCH ("TFSEARCH: Searching Transcription Factor Binding Sites", <http://www.rwcp.or.jp/papia/>) to search for promoter elements. Alignment of the *TCOF1* promoter sequences was performed using the UCSC (University of California-Santa Cruz Genome Bioinformatics, <http://genome.cse.ucsc.edu/>) Genome Browser with the following assemblies: human May 2004, chimpanzee Nov 2003, mouse May 2004, rat Jun 2003 and dog Jul 2004.

2.6. EMSA

Nuclear extracts from HepG2 cells were prepared as described (Read, 1998) and the protein concentration was determined using the Bradford's method. Two 51 bp double-

Fig. 1. Schematic representation of the different *TCOF1* promoter fragments inserted in the pGL3-Basic vectors. The –600 bp pol fragment differs from its wild-type version (–600 bp wt) only by the –346C>T nucleotide change. The anti-sense construct is a negative control (pGL3-1200wtAS).

phoscreen and the autoradiography captured using a STORM 860 scanner (GE Healthcare Bio-Sciences) and analyzed with the Image Quant[®] (5.2) software.

3. Results

3.1. Identification of SNPs in the *TCOF1* gene promoter region

Screening of 1.2 kbp of the *TCOF1* promoter sequence in 21 TCS patients and 51 controls revealed the presence of 5 SNPs. Two of these SNPs were identified only in the control group (–106A>C and –332T>C), with a frequency lower than 2%. The three others (–948G>A, –1025G>C and –346C>T), with a frequency higher than 10% and in Hardy–Weinberg equilibrium ($p>0.05$; Table 1), were identified in both patients and controls and with no statistical difference between the groups ($p>0.05$; Table 1). The –106A>C SNP was the only one previously described (dbSNP/GenBank accession number rs3806951). All individuals who displayed the –948G>A were also heterozygous for the –1025G>C SNP, suggesting that they are in linkage disequilibrium (LD). Indeed, we observed that both variations were in the same chromosome. On the other hand, the –346C>T variant was not found in LD with the –948G>A and –1025G>C SNPs.

The three patients with no detectable pathogenic mutation in the coding region of *TCOF1* did revealed no alterations in the promoter region by direct sequencing.

3.2. Species-conserved SNPs in the proximal *TCOF1* promoter

Computational analysis and visual inspection of the 1.2 kbp *TCOF1*'s promoter region revealed the absence of the TATAA-box element, and no initiator or downstream promoter element consensus sequence. In silico analysis also indicated more than 15 consensus sites for the ubiquitous Sp1 transcription factor and a high G/C content (75%) in the segment from position –200 bp to the first ATG. We also verified that the region from –152 to –1 bp shares 63% identity among the 5 species studied (Fig. 2). The –332T>C change is conserved only between human and chimpanzee, while the other four SNPs are located in conserved nucleotides in all species analyzed (human, mouse, rat, dog and chimpanzee; Fig. 2 and data not shown).

3.3. Functional analysis

All the vectors tested, except for the negative control pGL3-1200AS, showed a significantly higher ($p<0.05$) activity than the promoter-less pGL3-Basic vector ($p<0.05$;

Fig. 3. Functional activity of *TCOF1* promoter-reporter constructs. Plasmids containing serial deletions of the *TCOF1* promoter (A) and the wild-type and the polymorphic (–346C>T variant) 600 bp constructs (B) were transiently transfected into HepG2 cells. Data are means of at least two-independent transfections, each performed in triplicate, normalized to the pGL3-1200wt construct's activity (relative means $\pm$ SD).

Fig. 4. The $-346C>T$ SNP and nuclear protein affinity to the *TCOF1* promoter. A: Gel shift evidencing the DNA–protein specific complexes 1 and 2 for wild-type (WT) and polymorphic (POL) probes: lanes 1 and 10, free labeled probes; 2 and 11, labeled probes and HepG2 nuclear extracts (NE); 3, 4 and 5, WT labeled probe, NE and specific competitor (WT non-labeled probe) 10 $\times$, 30 $\times$ and 100 $\times$, respectively; 6, NE and 100 $\times$ non-specific competitor; 7 and 8, WT labeled probe, NE and antibodies to Sp1 and Sp3, respectively; 9 and 12, WT and POL labeled probes, NE and antibody to YY1 (abrogation of Complex 2). B: Competition gel shift (WT $\times$ POL). Lanes 1 and 8, free labeled probes; 2 and 9, labeled probes and NE; 3 to 6 and 10 to 13, indicated specific competitor in increasing concentrations; 7 and 14, labeled probes, NE and 50 $\times$ non-specific competitor. Complex 2 generated by wild-type probe persisted even in the presence of 30 $\times$ molar excess of non-labeled polymorphic competitor. By comparison, as few as 3 $\times$ molar excess of wild-type competitor was able to inhibit the polymorphic probe-generated complex 2; C: quantification of complexes 1 and 2 (YY1) in each concentration of specific competitor. Probes average intensity was calculated based in three competition gel shifts and considered as the sum above background.

Fig. 3A). The shortest construct, pGL3-200wt, displayed an activity 150% higher than that of the pGL3-1200wt construct. Since no statistical difference was observed between the transcriptional activities of pGL3-200wt and 300wt vectors, the 200 bp fragment presumably contains the minimal promoter. The pGL3-200wt's insert contains 207 bp, of which 76 bp are not transcribed, 75 bp belong to the 5' untranslated region and 56 bp are part of exon 1 (Fig. 1).

Comparing all constructs with the pGL3-1200wt vector, pGL3-600wt is not significantly different, but pGL3-400wt, pGL3-300wt and pGL3-200wt present higher activity ($p < 0,05$; Fig. 3A). These data suggest that the region between –1200 and –400 might comprise a repressor element. Comparison of the polymorphic 600 bp construct (pGL3-600pol) with its wild-type counterpart evidenced that the –346C>T variant is able to significantly decrease the reporter gene transcription by 38% ($p < 0,05$; Fig. 3B).

3.4. Impact of –346C>T SNP in DNA–protein interaction

EMSA evidenced the formation of two specific DNA–protein complexes (1 and 2) for both wild-type and polymorphic probes (Fig. 4A). A third complex was shown to be non-specific, since co-incubation with a non-specific probe in a molar excess of 100× abolished its formation (Fig. 4A, lane 6). Addition of antibody against YY1 led to abrogation of only complex 2 in both the wild-type and the polymorphic probes (Fig. 4A, lanes 9 and 12), suggesting that YY1 participates in complex 2 formation. Western blot analysis showed that HepG2 cells express the YY1 protein (data not shown). Sp1 and Sp3 do not seem to be involved in the formation of complexes 1 and 2 (Fig. 4A, lanes 7 and 8).

In the EMSA experiments, the complexes formed with the polymorphic probe were consistently less intense than the wild-type ones, suggesting that the –346C>T SNP affects nuclear protein binding. Through competition assays

between the two probes (Fig. 4B) we observed that complexes 1 and 2 display less intense bands in the presence of both wild-type and polymorphic competitors (4B, lanes 3–6 and 10–13). In comparison with complex 1, we also observed that complex 2 requires higher concentrations of competitors to diminish its band intensity (Fig. 4C). These data indicate that the complexes contain sequence-specific DNA-binding proteins and that the –346C>T SNP resides in a functionally important nucleotide, perhaps more relevant for complex 2.

3.5. –346C>T segregation in TCS Brazilian patients

We were able to verify that the pathogenic mutation was in *cis* with the –346C>T SNP in one familial case (family II; Fig. 5) and in *trans* in two familial cases (families I and III, Fig. 5). We observed that in these three genealogies that the affected heterozygous mother always had a milder phenotype than her offsprings, who presented abnormal external ears and hearing impairment, besides hypoplasia of the mandible and the zygomatic complex. The –346T>C SNP was also present in six other cases with a *de novo* pathogenic mutation, for which we were unable to determine its origin (Splendore et al., 2003); consequently, we were not able to establish the phase of the SNP in relation of the pathogenic mutation. It is of note that all of these sporadic cases also present a severe phenotype.

4. Discussion

This report represents the first analysis of the promoter region of the *TCOF1* gene, with identification of a novel functional SNP in the regulatory region. We have delimited the *TCOF1* minimal promoter to a region of 151 bp, which contains more than 15 potential Sp1 binding sites and lacks the consensus TATAA-box element.

Fig. 5. TCS familial pedigrees and the –346C>T SNP segregation. The genotype for the SNP is represented by CC and TT for homozygous and CT for heterozygous individuals. Familial cases are informative to determine whether the pathogenic mutation is in *cis* or in *trans* with the (T) allele. The patients who had inherited the pathogenic mutation in the coding region of *TCOF1* and the T variant from the affected parent probably have both alterations in *cis* (family II). In families I and III, the (T) allele seems to be in *trans* with the pathogenic mutation.

The absence of a pathogenic mutation in 1200 bp of the 5' upstream sequence in three patients without mutations in the coding region of *TCOF1* further supports genetic heterogeneity of TCS. However, we cannot rule out that the cause of the disease in these patients is due to mutations in distant regulatory regions, as has been shown for preaxial polydactyly and for lactase persistence (Lettice et al., 2003; Olds and Sibley, 2003).

We have detected three SNPs in the promoter region of TCS and we have also demonstrated that the $-346C>T$ variant is functional. The $-346T$ allele decreased by 38% the reporter gene transcription, which might have important physiological effects (Yan et al., 2002). This SNP is interfering with YY1 and other transcription factors' anchorage and stability. Although located 11 bp away from the YY1 consensus sequence, it is possible that the decrease in *TCOF1* promoter activity caused by the $-346T$ allele is caused by a YY1's reduced DNA-binding affinity to this region. YY1 is a transcription factor with four zinc-finger motifs and functions as a repressor, activator or initiator of transcription in promoters lacking the consensus TATAA-box element (OMIM *600013; Smale, 1997). Numerous in vitro and in vivo studies have suggested that YY1 controls expression of developmentally regulated genes, including those involved in early mouse craniofacial and limb morphogenesis (Donohoe et al., 1999; Tan et al., 2002). These data together with our in vitro results suggest that YY1 may play a major role in *TCOF1* regulation and, in this case, the SNP $-346C>T$ might have important implications in the transcription levels of this gene.

It is noteworthy that the individuals with the functional SNP in *trans* with the pathogenic mutation have a mild phenotype, although the sample is too small to yield any genotype–phenotype correlation. We must consider that the phenotype variation might reflect the underlying complexity of haplotypic structure involved in clinical variability. In addition, we cannot exclude that gene expression can be modulated by stochastic or environmental effects. Retinoic acid, a well-known teratogen, had been demonstrated to generate a TCS-like phenotype in mice as well as other craniofacial anomalies (Poswillo, 1975; Sulik et al., 1987). It is also important to observe that all of our probands present a typical and severe form of TCS, possibly due to ascertainment bias. Therefore, considering the rarity of the disease, the relatively low frequency of the $-346C>T$ SNP and the ascertainment bias, other approaches will be necessary to assess the effect of this polymorphism in the TCS phenotype.

Variability of the transcriptional level through genetic and environmental factors of the wild allele in carriers of autosomal dominant mutations, particularly those leading to haploinsufficiency of the protein product, can be a general mechanism to explain phenotypic variation in Mendelian disorders. Identification of a functional SNP in *cis*-regulatory elements and delimitation of the minimal *TCOF1* promoter constitute the first step in the elucidation

of the complex mechanisms involved in the modulation of the TCS phenotype.

Acknowledgments

We would like to thank Dr. Mendonça for the use of her laboratory facilities, Dr. Novak and Dr. Bachega for Sp3 and Sp1 antibodies. We would also like to thank Oscar T. Suzuki and Sheila M. B. Winnischofer for helpful discussions, Constância Gotto for secretarial assistance. This work was supported by grants from Fundação de Amparo à Pesquisa de São Paulo (FAPESP) and Conselho Nacional de Pesquisa (CNPq).

References

- Den Dunnen, J.T., Antonarakis, S.E., 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Human Mutat.* 15 (1), 7–12.
- Dixon, M.J., 1996. Treacher Collins syndrome. *Hum. Mol. Genet.* 5, 1391–1396.
- Dixon, J., Dixon, M.J., 2004. Genetic background has a major effect on the penetrance and severity of craniofacial defects in mice heterozygous for the gene encoding the nucleolar protein treacle. *Dev. Dyn.* 229, 907–914.
- Dixon, M.J., Marres, H.A.M., Edwards, S.J., Dixon, J., Cremers, C.W.R.J., 1994. Treacher Collins syndrome: correlation between clinical and genetic linkage studies. *Clin. dysmorph.* 3, 96–103.
- Donohoe, M.E., Zhang, X., McGinnis, L., Biggers, J., Li, E., Shi, Y., 1999. Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. *Mol. Cell. Biol.* 19, 7237–7244.
- Edwards, S.J., Gladwin, A.J., Dixon, M.J., 1997. The mutational spectrum in Treacher Collins syndrome reveals a predominance of mutations that create a premature-termination codon. *Am. J. Hum. Genet.* 60, 515–524.
- Gladwin, A.J., et al., 1996. Treacher Collins syndrome may result from insertions, deletions or splicing mutations, which introduce a termination codon into the gene. *Hum. Mol. Genet.* 5, 1533–1538.
- Gorlin, R.J., Cohen Jr., M.M., Hennekam, R.C.M., 2001. *Syndromes of the Head and Neck*, 4th ed. Oxford University Press, Oxford, pp. 799–802.
- Hayano, T., Yanagida, M., Yamauchi, Y., Shinkawa, T., Isobe, T., Takahashi, N., 2003. Proteomic analysis of human Nop56p-associated pre-ribosomal ribonucleoprotein complexes. Possible link between Nop56p and the nucleolar protein treacle responsible for Treacher Collins syndrome. *J. Biol. Chem.* 278 (36), 34309–34319.
- Isaac, C., et al., 2000. Characterization of the nucleolar gene product, treacle, in Treacher Collins syndrome. *Mol. Biol. Cell.* 11, 188–192.
- Lettice, L.A., et al., 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* 12 (14), 1725–1735.
- Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16, 1215.
- Olds, L.C., Sibley, E., 2003. Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a *cis* regulatory element. *Hum. Mol. Genet.* 12 (18), 2333–2340.
- Poswillo, D., 1975. The pathogenesis of Treacher Collins syndrome (mandibulofacial dysostosis). *Br. J. Oral Surg.* 13, 1–26.
- Read, M., 1998. Electrophoretic mobility shift assay (EMSA). In: Docherty, K. (Ed.), *Essential Techniques*. Wiley & Sons, New York, US.

- Richards, B., et al., 1993. Multiplex PCR amplification from the *CFTR* gene using DNA prepared from buccal brushes/swabs. *Hum. Mol. Genet.* 2, 159–163.
- Smale, S.T., 1997. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim. Biophys. Acta* 1351 (1–2), 73–88.
- So, R.B., Gonzales, B., Henning, D., Dixon, J., Dixon, M.J., Valdez, B.C., 2004. Another face of the Treacher Collins syndrome (*TCOF1*) gene: identification of additional exons. *Gene* 328, 49–57.
- Splendore, A., et al., 2000. High mutation detection rate in *TCOF1* among Treacher Collins syndrome patients reveals clustering of mutations in 16 novel pathogenic changes. *Human Mutat.* 16, 315–322.
- Splendore, A., Jabs, E.W., Passos-Bueno, M.R., 2002. Screening of *TCOF1* in patients from different populations: confirmation of mutational hot spots and identification of a novel missense mutation that suggests an important functional domain in the protein treacle. *J. Med. Genet.* 39, 493–495.
- Splendore, A., Jabs, E.W., Félix, T.M., Passos-Bueno, M.R., 2003. Parental origin of mutations in sporadic cases of Treacher Collins syndrome. *Eur. J. Hum. Genet.* 11 (9), 718–722.
- Splendore, A., Fanganiello, R.D., Masotti, C., Morganti, L.S., Passos-Bueno, M.R., 2005. *TCOF1* mutation database: novel mutation in the alternatively spliced exon 6A and update in mutation nomenclature. *Hum. Mutat.* 25 (5), 429–434.
- Sulik, K.K., Johnston, M.C., Smiley, S.J., Speight, H.S., Jarvis, B.E., 1987. Mandibulofacial dysostosis (Treacher Collins syndrome): a new proposal for its pathogenesis. *A.m. J. Med. Genet.* 27 (2), 359–372.
- Tan, D.P., et al., 2002. YY1 activates *Msx2* gene independent of bone morphogenetic protein signaling. *Nucleic Acids Res.* 30 (5), 1213–1223.
- Teber, O.A., et al., 2004. Genotyping in 46 patients with tentative diagnosis of Treacher Collins syndrome revealed unexpected phenotypic variation. *Eur. J. Hum. Genet.* 12 (11), 879–890.
- The Treacher Collins Collaborative Group, 1996. Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome. *Nat. Gen.* 12, 130–136.
- Valdez, B.C., Henning, D., So, R.B., Dixon, J., Dixon, M.J., 2004. The Treacher Collins syndrome (*TCOF1*) gene is involved in ribosomal DNA gene transcription by interacting with upstream binding factor. *Proc. Natl. Acad. Sci.* 101 (29), 10709–10714.
- Yan, H., Weishi, Y., Velculescu, V.E., Vogelstein, B., Kinzler, K.W., 2002. Allelic variation in human gene expression. *Science* 297 (16), 1143.

ANEXO III

Evidences that collagen XVIII is a new obesity gene based on a case-control study of *COL18A1* SNPs in type 2 diabetes patients

Short title: COL18A1 SNPs in type 2 diabetes patients

Flavia Imbroisi V. Errera, PhD^{1,2}, Luís Henrique Canani, MD, PhD³, Maria Elizabeth Rossi Silva, MD, PhD⁴, Erika Yeh, BSc¹, Oscar T. Suzuki, PhD¹, Lucia Maria Armelin-Correa, BSc¹, Balduino Tschiedel, MD, PhD⁵, Israel Roisenberg, MD, PhD⁶, Jorge Gross, MD, PhD^{3,7}, Maria Rita Passos-Bueno, PhD¹

¹ Centro de Estudos do Genoma Humano, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências da Universidade de São Paulo, São Paulo

² Departamento de Morfologia, Escola Superior de Ciências da Santa Casa de Misericórdia, Vitória, Espírito Santo

³ Divisão de Endocrinologia, Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil

⁴ Laboratório de Investigação Médica LIM-18, Hospital das Clínicas, Faculdade de Medicina da Universidade de São Paulo, São Paulo

⁵ Departamento de Oftalmologia, Hospital das Clínicas, Faculdade de Medicina da Universidade de São Paulo, São Paulo

⁶ Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil

⁷ Divisão de Endocrinologia, Hospital Conceição, Rio Grande do Sul, Porto Alegre, Brasil

Correspondence to Maria Rita Passos-Bueno, Rua do Matão, 277, Departamento de Biologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brasil, CEP: 05508-900, Tel: 55-11-3091-9910, Fax: 55-11-3091-7419, passos@ib.usp.br

Submitted to Diabetes Care 23 April 2007

Abstract

Objective: Collagen XVIII is a basement membrane (BM) proteoglycan with a “frizzled” (FZ-C18) N-terminal motif, apparently involved with Wnt signaling, and endostatin (ES-C18), a C-terminal proteolytic fragment, which is a potent inhibitor of angiogenesis. Most recently, *COL18A1* and Wnt signaling have been associated to adipocyte differentiation and obesity. Therefore, we have undertaken this work to verify if polymorphisms (SNPs) in the Frizzled (c.1135C>T; Thr379Met) and Endostatin (c.4349G>A; Asp1437Asn) regions contribute to susceptibility to obesity in patients with type 2 diabetes.

Methods: We have conducted a case-control study with a total of 345 unrelated Brazilian patients with type 2 diabetes. The SNPs in *COL18A1* were analyzed, by SNUPE assay (*Single Nucleotide Primer extension*), in 113 obese (BMI $\geq$ 30, cases) and 232 non-obese patients (BMI <30, controls) of European Ancestry (EA).

Results: We observed a significantly higher frequency of c.1135TT homozygotes in obese individuals (22/113; 19.5%) than in non-obese individuals (25/232; 10.9%) [$P=0.02$; OR=2,0 (95%CI: 1,07-3,73)], suggesting that the c.1135T allele is associated to obesity in a recessive model. Still, this genotype was independently associated with obesity ($P=0.048$), and increases the chance of obesity by 2.8-fold, after controlling for other variables such as cholesterol, LDL cholesterol, and triglycerides. No evidence of association was observed between the SNP c.4349G>A and obesity.

Conclusion: These data suggest an association of collagen XVIII with obesity in type 2 diabetes patients, pointing out *COL18A1* gene as a promising new obesity gene.

Introduction

Obesity is a worldwide epidemic health problem that confers risks for several diseases, including type diabetes 2. In this context, individuals with central obesity (fat accumulated in visceral depths) are at a higher risk of developing complications than those with peripheral obesity (fat distributed subcutaneously) [1-3]. Obesity is a multifactorial condition where both genetic and environmental factors play an important role in the manifestation of the phenotype. There is currently an intense effort in searching for obesity susceptibility genes and, interestingly, there are a growing number of evidences showing the involvement of developmental genes in obesity and fat distribution [4].

COL18A1, an important developmental gene located at chromosome 21q22.3, can be transcribed as 3 distinct isoforms by the use of two promoters and of an alternative splicing of the third exon [5,6,7]. The three encoded collagen XVIII proteins differ only by their signal peptides and by part of the N-terminal region of the NC11 domain, that contain respectively 303 (NC11-303), 493 (NC11-493) and 728 (NC11-728) amino acid residues [8]. Collagen XVIII is an important proteoglycan of most basement membranes of the body, with higher expression levels in liver, kidney, retina and adipose tissue; the predominant isoform also varies according to the tissue type [6-9; NCBI UniGene Rs.517356]

Endostatin (ES-C18), a 20 kDa proteolytic fragment released from the C-terminal region (NC1) of collagen XVIII, has been the focus of intense cancer research in the past 10 years as it is a potent inhibitor of endothelial cell proliferation and migration with the ability to reduce tumor growth [10]. Collagen

XVIII and ES-C18 play several other functions, including apoptosis induction [11,12] and maintenance of vascular permeability [13]. The downstream molecular pathways involved in these processes are still unknown, but an involvement of ES-C18 with the coreceptor glypican and with Wnt signaling has been proposed [14,15]. Besides ES C18, there is also a soluble type XVIII collagen fragment derived from the NC11 domain, which contains a frizzled (FZ) motif with still unknown function [8]. This motif is characterized by the presence of 10 invariantly spaced cysteines and, therefore, called the cysteine-rich domain (CRD). Secreted Wnt molecules bind the CRD portion of transmembrane FZ receptors and trigger intracellular signalling pathways acting in a series of cellular process, such as proliferation and differentiation [16]. The interaction of Wnt proteins with FZ surface receptors is mediated by the coreceptors *LRP5* and *LRP6* [17]. Several secreted proteins containing FZ domains (sFZRP-secreted *Frizzled related protein*) can, on the other hand, antagonize Wnt signaling, and it has been suggested that FZ C18 might be one of these proteins [8]. Therefore, it is possible that collagen XVIII FZ domain (FZ-C18), as a proteolytic fragment, binds Wnt proteins in the extracellular matrix and competes with FZ receptors, modulating a series of cellular signalling pathways.

Wnt signaling has been shown to be involved in adipogenesis [18,19] and in glucose-induced insulin secretion [20-21] and, therefore, genes related to the WNT signaling pathway might confer susceptibility to type 2 diabetes and obesity. Indeed, it has been reported that genetic alterations in *WNT5B* are associated with type 2 diabetes, while in *WNT10B* and *LPR5* have been associated with obesity [22-24]. However, to our knowledge, no association

studies between genes containing FZ motif and obesity or diabetes have yet been reported.

In view of this information and considering the expression pattern of collagen XVIII and its role in cell proliferation and in angiogenesis, we considered *COL18A1* as a very attractive susceptibility candidate gene to obesity. Therefore, we tested if the SNPs c.1135C>T (Thr379Met; FZ motif of *COL18A1* – FZ-C.18) and c.4349G>A (Asp1437Asn; endostatin region- ES-C.18) were associated to obesity in a cohort of type 2 diabetes patients.

Material and methods

Patients

We have conducted a case-control study with a total of 345 unrelated patients with type 2 diabetes. One hundred thirty five patients were ascertained at the Endocrine Service, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, while the remaining 210 patients are enrolled in a large collaborative study of Diabetes at the University of Rio Grande do Sul, and detailed clinical description can be found elsewhere [25]. Briefly, patients with type 2 diabetes were identified from a multicentric study that started recruiting patients in southern Brazil in 2002. That project aimed to study risk factors for chronic complications of diabetes. It included four centers located at general hospitals in the state of Rio Grande do Sul, namely Grupo Hospitalar Nossa Senhora da Conceição, Hospital São Vicente de Paula, Hospital Universitário de Rio Grande, and Hospital de Clínicas de Porto Alegre. In this study, the ethnical groups were defined based on self-classification and subjective classification (color skin, nose and lip shapes, hair texture and information

about family ancestry). All patients are of European ancestry (EA, mainly from Portugal, Spain, Italy and Germany). Patients of African ancestry (AA) or those who defined themselves as having mixed or other ancestry were not included. The ethical committees of the participating Institutions approved this study and data and blood collection was drawn only after signature of the informed consent.

Patients underwent a standardized evaluation consisting of a questionnaire, physical examination and laboratory tests. Diagnosis of type 2 diabetes was based on the guidelines of the Expert Committee report of the American Diabetes Association [26]. Hypertension was defined as blood pressure $\geq 140/90$ mmHg, or use of antihypertensive medication. Weight without shoes and in light outdoor clothes and height were measured, and body mass index (BMI) was calculated as weight/height² (kg/m²). Obesity was classified according World Health Organization (WHO) criteria [27]. Patients with BMI ≥ 30 kg/m² were considered cases and patients with BMI < 30 kg/m² as controls.

Peripheral blood collection, DNA extraction and biochemical dosages

Genomic DNA was extracted from peripheral blood leucocytes by salting-out procedure as previously described [28]. In diabetic patients, glucose was determined by a glucose oxidase method; creatinine by the Jaffé reaction; A_{1c} test by an ion-exchange HPLC procedure (Merck-Hitachi L-9100 Glycated hemoglobin analyzer; reference range: 2.7-4.3 %; Merck, Darmstadt, Germany), and triglyceride and cholesterol levels by enzymatic methods. Albuminuria was measured by immunoturbidimetry (SERA-PAK® immuno Microalbuminuria; Bayer, Tarrytown, NY, mean intra- and interassay CV: 4.5%

and 7.6%, respectively). Total cholesterol, HDL cholesterol and triglycerides were measured by standard enzymatic methods.

Analysis of SNPs in the *COL18A1* gene coding regions of *Frizzled* and *Endostatin*

SNPs c.1135C>T (Thre379Met/frizzled domain) and c.4349 G>A (Asp1437Asn/ endostatin region) were analyzed by SNuPE (*Single Nucleotide Primer extension*-Amersham Pharmacia) method. Exons 3 (coding region of frizzled domain) and 42 (coding region of endostatin domain) of *COL18A1* gene were amplified using PCR buffer, 0.2mM dNTPs, 0.6µM primer and 0,2U Taq polimerase, in 35 cycles with 100 ng o DNA, according to the following steps: 5 minutes denaturation at 95°C, 10 minutes extention at 72°C, and using the following primers (5'-3') and respective annealing temperature: exon 3F: CCCAACCACCTCCACCACGAGA and exon 3R: CTGGATGAACCGCAGCCGATGT. (66°C); 42F: ACAAACACCCACACCCATC and 42R: GAAACTGCAGATAGGAGGCC (59°C). PCR products were submitted to SNuPE reaction and then analyzed in automatic sequencer (MegaBACE™ 1000 DNA Analysis Systems), using the correspondent *software* and following the User's guide instructions (General Eletrics Health Science) and 2 pmols of oligonucleotide [c.1135C>T: CGAGGAAGGGGTGGCAGTGC, and c.4349G>A: GGGCACGCATCTTCTCCTTT].

Statistical Analysis

Statistical analysis was performed using the SPSS package (SPSS for Windows, version 10.0). Continuous variables were expressed by mean and

standard deviation and the groups were compared by unpaired Student's t-test. Allele frequencies were determined by gene counting, and departures from the Hardy–Weinberg equilibrium (HWE) were verified using the χ^2 test. Genotype and allele frequency were compared among different groups by χ^2 Test or Fisher's exact test. A *P* value <0.05 was considered statistically significant. Logistic regression analysis was performed to assess the independent role of the *COL18A1* genotypes in obesity. Data is shown as mean $\pm$ SD.

Results

Characterization of the patient sample

The main clinical features of the patients are depicted in Table 1. In the total sample, the mean duration of type 2 diabetes was 13.78 ± 7.78 years. Classification for obesity was obtained for 345 patients: 113 were classified as obese (BMI: 34.31 ± 4.64 ; cases) and 232 as non-obese (BMI: 25.83 ± 2.53 ; controls). Cases also had higher waist and hip circumferences as compared to controls (*P* <0.05). They also presented higher triglycerides, total and LDL cholesterol (*P* <0.05). There were no differences for other clinical or laboratorial characteristics.

Association Study between *COL18A1* and Obesity

The genotype distribution for SNPs c.1135C>T and c. 4349G>A were in HWE in the type 2 diabetes patient samples (Table 2; *P* >0.05). There was no difference in the genotype or allelic frequencies of the SNP c.4349 G>A among obese and non-obese patients (*P* = 0.11). However, a borderline association between SNP c.1135C>T genotypes and obesity was found (*P*=0.07). Assuming a recessive model (TT vs. CT+CC), the genotype TT was significantly

more frequent in obese diabetic patients (22/113, 19.5%) than in non-obese diabetic patients (25/232, 10.5%; $\chi^2=4.73$ and $P = 0.02$). In this model, patients with TT genotype had an odds ratio (OR) of 2.57 (95% CI: 1.21-5.47) of presenting obesity. This effect was more pronounced if overweight type 2 diabetes patients (BMI between 28 and 30 kg/m²) are excluded from the non-obese control group [TT genotype frequency in non-obese, BMI <27 kg/m², 12/140 or 8.5% ; $\chi^2=6.38$ and $P = 0.01$].

In a logistical model, after controlling for cholesterol, triglycerides, HDL cholesterol levels, the TT genotype still held association with the presence of obesity ($P = 0.048$; OR 2,8 CI 95% 1.01- 7.74).

Discussion

In the present work, we have evaluated whether there was an association between *COL18A1* gene polymorphisms c.1135C>T and c.4349G>A, located in two important collagen XVIII functional domains, and obesity in type 2 diabetes patients. The allele frequency and genotype distribution for the SNP c.4349G>A in our sample is similar to those previous reported in the literature [30]. We did not observe any evidence of association of this SNP and obesity. The functional effect of this SNP, which leads to the substitution of a conserved aspartic acid at residue 1437 to asparagine in the ES domain, remains inconclusive [29-32].

The SNP c.1135C>T (Thr379Met) is described in the present work for the first time. Codon 379 is located within the *Frizzled* (FZ) motif, present only in the collagen XVIII NC11-728 isoform. We performed a comparative genomic analysis including 6 species (human, chimpanzee, dog, mouse, rat chicken, fugu and zebrafish) and we verified that the amino acid threonine at position

379 is highly conserved in vertebrates from *Fugu* to man, suggesting that this residue might be important for normal protein function.

We did not find evidence of association between any of these SNPs and type 2 diabetes, so the presence of diabetes is not causing any distortion in their allelic distribution (data not shown; available on request). In contrast, our results suggest that the allele c.1135T (Thr379Met) when in homozygosis increases about two-fold the chance that a type 2 diabetes patient develops obesity. When overweight type 2 diabetes patients were excluded from the control group, we found a more pronounced difference in c.1135T frequency between cases and controls, further supporting the association between this SNP and obesity in type 2 diabetes patients. The importance of TT genotype for obesity was still confirmed by logistic regression analysis with variables total cholesterol levels, LDL and triglycerides. If these variables are taken into account, the chance that a TT homozygote develops obesity is increased by 2,8 times when compared to TC heterozygotes or CC homozygotes. Hence, our data suggest that TT genotype acts as an independent factor for predisposition to obesity in type 2 diabetes patients. The expression of COL18A1 significantly increases during adipocyte differentiation [33] and, two recent reports have indicated, directly or indirectly, collagen XVIII as diabetesity-relevant candidate gene. Brown et al [34] have identified COL18A1 out of the 71 most promising diabetesity genes selected through the use of the data mining tool ExQuest and 29 QTL loci related to body weight, adiposity or type 2 diabetes in mouse. Gesta et al [4], verified that glypican, one of the receptors of ES C18 [15], was among the 198 genes differentially expressed in mice and human intra

abdominal and subcutaneous adipose tissue. These findings thus reinforce the importance of collagen XVIII in the predisposition to obesity.

Little is known about the function of FZ in collagen XVIII, but it is well established that proteins with FZ domain act in Wnt signaling through different pathways in various processes, including adipogenesis, obesity, insulin secretion and type 2 diabetes [19,20,22,24,36] The substitution of a threonine to a methionine in position 379 may lead to a functional alteration in the protein. The methionine, for example, can enhance ligand-binding affinity or decrease its specificity; in either way this allele would alter the function of this collagen XVIII fragment (FZ-C.18) in the Wnt signaling pathway activation. As the number of pre-adipocytes, as well as, their differentiation into adipocytes are important factors for obesity we cannot yet infer which is the effect of SNP c.1135C<T in collagen XVIII function in these processes. We cannot exclude the possibility that the mutated allele is in linkage disequilibrium with another one, which would be actually the responsible for the phenotype. Recently, we verified that allele c.1135T is in linkage disequilibrium with the G allele of the polymorphism -700G>A, which confers higher transcription levels to this collagen [37].

Obesity has not been reported in patients with null mutations in the collagen XVIII [29], or in the knockout col18^{-/-} mice [38]. However, it is possible that the phenotype is induced only by high food intake, or depends on the individual genetic background. In contrast, obesity is very common in Down syndrome patients, who have three functional copies of COL18A1 gene, as the consequence of trissomy of chromosome 21 [39].

Type 2 diabetes is a very complex metabolic disorder and it is very difficult to disentangle the primary and secondary causes of the disease in most of the patients. Therefore, considering that collagen XVIII is present in most of the basement membrane of all tissues, and that deficiency of this protein is a risk factor to atherosclerosis [13], it is also possible that this SNP is more related to altered lipid or vascular metabolism than to adipogenesis itself.

In conclusion, the present study suggests a role of collagen XVIII SNPc.1135C<T in obesity in type 2 diabetes patients. We cannot extrapolate our results to any type of obesity since all the patients included in the present study have type 2 diabetes, and visceral obesity is a risk factor for this disease. It will be certainly important to test this SNP in obese individuals, and type 2 diabetes patients, of different populations.

Acknowledgments

We are grateful to Dr. Carlos Maranduba, to Dr. Daniel Gianella and to Dr. Hugo Lisbôa and his team for kindly providing part of the sample of diabetic patients. This research was supported by Fundação de Auxílio à Pesquisa do Estado de São Paulo/Centro de Excelência em Pesquisa, Inovação e Difusão (FAPESP/CEPID), Conselho Nacional de Pesquisa (CNPq) and Hospital de Clínicas de Porto Alegre. LHC is the recipient of a postdoctoral (ProDoc) grant from Fundação de Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (Fundação CAPES).

References

1. Lean ME: Pathophysiology of obesity. *Proc Nutr Soc* 59: 331–336, 2000
2. Mauriege P, Despres JP, Moorjani S, Prud'Homme D, Lamarche B, Bouchard C, Nadeau A, Tremblay A, Lupien PJ: Abdominal and femoral adipose tissue lipolysis and cardiovascular disease risk factors in men. *Eur J Clin Invest* 23: 729–740, 1993
3. Kissebah AH, Krakower GR: Regional adiposity and morbidity. *Physiol Rev* 74, 761–811, 1994
4. Gesta S, Bluher M, Yamamoto Y, Norris AW, Berndt J, Kralisch S, Boucher J, Lewis C, Kahn CR: Evidence for a role of developmental genes in the origin of obesity and body fat distribution. *Proc Natl Acad Sci USA* 103:6676-6681, 2006
5. Oh SP, Warman ML, Seldin MF, Cheng SD, Knoll JH, Timmons S, Olsen BR: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localization of the alpha-1 (XVIII) collagen gene to mouse chromosome 10 and human chromosome 21. *Genomics* 19: 494-499, 1994
6. Saarela J, Rehn M, Oikarinen A, Autio-Harmainen H, Pihlajaniemi T: The short and long forms of type XVIII collagen show clear tissue specificities in their expression and location in basement membrane zones in humans. *Am J Pathol* 153: 611– 626, 1998a.
7. Saarela J, Ylikarppa R, Rehn M, Purmonen S, Pihlajaniemi T: Complete primary structure of two variant forms of human type XVIII collagen and tissue-specific differences in the expression of the corresponding transcripts. *Matrix Biol* 16: 319-328, 1998
8. Elamaa H, Snellman A, Rehn M, Autio-Harmainen H, Pihlajaniemi T: Characterization of the human type XVIII collagen gene and proteolytic processing and tissue location of the variant containing a frizzled motif. *Matrix Biol* 22: 427-442, 2003
9. Suzuki OT, Sertie AL, Der Kaloustian VM, Kok F, Carpenter M, Murray J, Czeizel AE, Kliemann SE, Rosenberg S, Monteiro M, Olsen BR, Passos-Bueno MR: Molecular analysis of collagen XVIII reveals novel mutations, presence of a third isoform, and possible genetic heterogeneity in Knobloch syndrome. *Am J Hum Genet* 71:1320-1329, 2002
10. O'Reilly MS, Boehm T, Shing Y, Fukai N, Vasios G, Lane WS, Flynn E, Birkhead JR, Olsen BR, Folkman J: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. *Cell* 88: 277-285, 1997

11. Dhanabal M, Ramchandran R, Waterman MJ, Lu H, Knebelmann B, Segal M, Sukhatme VP: Endostatin induces endothelial cell apoptosis. *J Biol Chem* 274: 11721-11726, 1999
12. Schmidt A, Wenzel D, Thorey I, Werner S, Fleischmann BK, Bloch W: Endostatin down-regulates soluble guanylate cyclase (sGC) in endothelial cells in vivo: influence of endostatin on vascular endothelial growth factor (VEGF) signaling. *Endothelium* 12:251-257, 2005.
13. Moulton KS, Olsen BR, Sonn S, Fukai N, Zurakowski D, Zeng X: Loss of collagen XVIII enhances neovascularization and vascular permeability in atherosclerosis. *Circulation* 110:1330-1336, 2004
14. Hanai J, Gloy J, Karumanchi SA, Kale S, Tang J, Hu G, Chan B, Ramchandran R, Jha V, Sukhatme VP, Sokol S: Endostatin is a potential inhibitor of Wnt signaling. *J Cell Biol* 158:529-539, 2002
15. Karumanchi SA, Jha V, Ramchandran R, Karihaloo A, Tsiokas L, Chan B, Dhanabal M, Hanai JI, Venkataraman G, Shriver Z, Keiser N, Kalluri R, Zeng H, Mukhopadhyay D, Chen RL, Lander AD, Hagihara K, Yamaguchi Y, Sasisekharan R, Cantley L, Sukhatme VP: Cell surface glypicans are low-affinity endostatin receptors. *Mol Cell* 7:811-822, 2001
16. Wodarz A, Nusse R: Mechanisms of Wnt signaling in development. *Annu Rev Cell Dev Biol* 14:59-88, 1998
17. Tamai K, Semenov M, Kato Y, Spokony R, Liu C, Katsuyama Y, Hess F, Saint-Jeannet JP, He X: LDL-receptor-related proteins in Wnt signal transduction. *Nature* 407: 530-535, 2000
18. Ross SE, Hemati N, Longo KA, Bennett CN, Lucas PC, Erickson RL, MacDougald OA: Inhibition of adipogenesis by Wnt signaling. *Science* 289: 950-953, 2000
19. Bennett CN, Ross SE, Longo KA, Bajnok L, Hemati N, Johnson KW, Harrison SD, MacDougald OA: Regulation of Wnt signaling during adipogenesis. *J Biol Chem* 277:30998-31004, 2002
20. Fujino T, Asaba H, Kang MJ, Ikeda Y, Sone H, Takada S, Kim DH, Ioka RX, Ono M, Tomoyori H, Okubo M, Murase T, Kamataki A, Yamamoto J, Magoori K, Takahashi S, Miyamoto Y, Oishi H, Nose M, Okazaki M, Usui S, Imaizumi K, Yanagisawa M, Sakai J, Yamamoto TT: Low-density lipoprotein receptor-related protein 5 (LRP5) is essential for normal cholesterol metabolism and glucose-induced insulin secretion. *Proc Natl Acad Sci USA* 100:229-234, 2003

21. Kennell JA, MacDougald AO: Wnt signaling inhibits adipogenesis through beta-catenin-dependent and -independent mechanisms. *J Biol Chem* 280:24004-24010, 2005
22. Kanazawa A, Tsukada S, Sekine A, Tsunoda T, Takahashi A, Kashiwagi A, Tanaka Y, Babazono T, Matsuda M, Kaku K, Iwamoto Y, Kawamori R, Kikkawa R, Nakamura Y, Maeda S: Association of the gene encoding wingless-type mammary tumor virus integration-site family member 5B (WNT5B) with type 2 diabetes. *Am J Hum Genet* 75:832-843, 2004
23. Christodoulides C, Scarda A, Granzotto M, Milan G, Dalla Nora E, Keogh J, De Pergola G, Stirling H, Pannacciulli N, Sethi JK, Federspil G, Vidal-Puig A, Farooqi IS, O'Rahilly S, Vettor R: WNT10B mutations in human obesity. *Diabetologia* 49:678-684, 2006.
24. Guo YF, Xiong DH, Shen H, Zhao LJ, Xiao P, Guo Y, Wang W, Yang TL, Recker RR, Deng HW: Polymorphisms of the low-density lipoprotein receptor-related protein 5 (LRP5) gene are associated with obesity phenotypes in a large family-based association study. *J Med Genet* 43:798-803, 2006
25. Canani LH, Capp C, Ng DP, Choo SG, Maia AL, Nabinger GB, Santos K, Crispim D, Roisemberg I, Krolewski AS, Gross JL: The fatty acid-binding protein-2 A54T polymorphism is associated with renal disease in patient with type 2 diabetes. *Diabetes* 54:3326-30, 2005
26. American Diabetes Association. Diagnosis and classification of Diabetes Mellitus. *Diabetes Care* 28: S37-S42, 2005
27. *Obesity: preventing and managing the global epidemic. Report on a WHO consultation.* Geneva, World Health Organization, 2000 (WHO Technical Report Series, No. 894).
28. Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* 16:1215, 1988
29. Passos-Bueno MR, Suzuki OT, Armelin-Correa LM, Sertie AL, Errera FI, Bagatini K, Kok F, Leite KR: Mutations in Collagen 18A1 (COL18A1) and their relevance to the human phenotype. *Anais Acad Bras Ciências* 78:123-132, 2006
30. Iughetti P, Suzuki O, Godoi PH, Alves VA, Sertie AL, Zorick T, Soares F, Camargo A, Moreira ES, di Loreto C, Moreira-Filho CA, Simpson A, Oliva G, Passos-Bueno MR: A polymorphism in endostatin, an angiogenesis inhibitor, predisposes for the development of prostatic adenocarcinoma. *Cancer Res* 61: 7375-7378, 2001

31. Menzel O, Bekkeheien RC, Reymond A, Fukai N, Boye E, Kosztolanyi G, Aftimos S, Deutsch S, Scott HS, Olsen BR, Antonarakis SE, Guipponi M: Knobloch syndrome: novel mutations in COL18A1, evidence for genetic heterogeneity, and a functionally impaired polymorphism in endostatin. *Hum Mutat* 23:77-84, 2004
32. Macpherson GR, Singh AS, Bennett CL, Venzon DJ, Liewehr DJ, Franks ME, Dahut WL, Kantoff PW, Price DK, Figg WD: Genotyping and Functional Analysis of the D104N Variant of Human Endostatin. *Cancer Biol Ther* 3:1298-1303, 2004
33. Inoue-Murayama, M, Sugimoto Y, Niimi Y, Aso H: Type XVIII collagen is newly transcribed during bovine adipogenesis. *Differentiation* 65:281-285, 2000
34. Brown AC Olver WI, Donnelly CJ, May ME, Naggert JK, Shaffer DJ, Roopenian DC: Searching QTL by gene expression: analysis of diabetes. *BMC Genet* 6:12, 2005
35. Hu E, Zhu Y, Fredrickson T, Barnes M, Kelsell D, Beeley L, Brooks D: Tissue restricted expression of two human Frzbs in preadipocytes and pancreas. *Biochem Biophys Res Commun* 247:287-293, 1998
36. Kumar S, Leontovich A, Coenen MJ, Bahn RS: Gene expression profiling of orbital adipose tissue from patients with Graves' ophthalmopathy: a potential role for secreted frizzled-related protein-1 in orbital adipogenesis. *J Clin Endocrinol Metab* 90:4730-4735, 2005
37. Armelin-Correa LM, Lin CJ, Barbosa A, Bagatini K, Winnischofer SM, Sogayar MC, Passos-Bueno MR: Characterization of human Collagen XVIII promoter 2: interaction of Sp1, Sp3 and YY1 with the regulatory region and a SNP that increases transcription in hepatocytes. *Matrix Biol* 24:550-559, 2005
38. Fukai N, Eklund L, Marneros AG, Oh SP, Keene DR, Tamarkin L, Niemela M, Ilves M, Li E, Pihlajaniemi T, Olsen BR: Lack of collagen XVIII/endostatin results in eye abnormalities. *EMBO J* 21:1535-1544, 2002
39. Zorick TS, Mustacchi Z, Bando SY, Zatz M, Moreira-Filho CA, Olsen B, Passos-Bueno MR: High serum endostatin levels in Down syndrome: implications for improved treatment and prevention of solid tumours. *Eur J Hum Genet* 9: 811-814, 2001

Table 1. Clinic variables in type 2 diabetes patients according to presence of obesity (mean $\pm$ SD).

	Controls (n= 232)	Cases (n=113)	P
Gender Female: Male	122:110	77:36	
DM duration (years)	12.78 $\pm$ 7.7451	12.81 $\pm$ 7.17	0.97
Age of DM Diagnosis (years)	48.81 $\pm$ 10.33	44.64 $\pm$ 9.62	0.07
Current age (years)	60.80 $\pm$ 9.60	60.02 $\pm$ 8.34	0.47
BMI (kg/m²)	25.83 $\pm$ 2.53	34.31 $\pm$ 4.64	0.00
Waist circumference (cm)	92.88 $\pm$ 9.29	105.36 $\pm$ 12.40	0.00
Hip circumference (cm)	98.48 $\pm$ 7.54	114.48 $\pm$ 15.33	0.00
Glucose (mg/dl)	172.47 $\pm$ 68.43	175.34 $\pm$ 66.73	0.78
LDL- cholesterol (mg/dl)	130.27 $\pm$ 46.19	147.70 $\pm$ 48.05	0.01
HDL-cholesterol (mg/dl)	45.82 $\pm$ 12.70	44.06 $\pm$ 12.00	0.24
Cholesterol (mg/dl)	202.97 $\pm$ 47.14	223.81 $\pm$ 53.41	0.02
Creatinine (mg/dl)	0.92 (0.2-6.7) *	0.95 (0.4-7.0) *	0.74
Tryglycerides	174.08 (70-238) *	216.98 (89-260.8) *	0.03
A1c test (%)	7.13 $\pm$ 2.36	7.49 $\pm$ 2.53	0.31
Systolic Arterial Pressure (mmHg)	142.89 $\pm$ 24.02	144.54 $\pm$ 20.85	0.62
Diastolic Arterial Pressure (mmHg)	85.35 $\pm$ 13.66	87.11 $\pm$ 11.93	0.33

* Values do not have a normal distribution and are represented by median and lowest and highest limits

Table 2. Genotypic and allelic distribution for SNP c.1135C>T SNP and c.4349G>A according to body mass index.

Body Mass Index		
	>30 kg/m ²	<30 kg/m ²
	(n= 113)	(n= 232)
<u>SNP c.1135C>T</u>		
Genotypes		
CC	46 (40.7%)	111 (47.8%)
CT	45 (39.8%)	96 (41.4%)
TT	22 (19.5%)	25 (10.8%)
Alleles		
C	60.6%	68.5%
T	39.4%	31.5%
<u>SNP c.4349G>A</u>		
Genotypes		
GG	102 (90.3%)	198 (85.4%)
GA	11 (9.7%)	33 (14.2%)
AA	0 (0%)	1 (0.4%)
Alleles		
G	95.1%	92.5%
A	4.9%	7.5%