Erika Kague

Estudo da regulação transcricional do *COL18A1* e análise funcional do domínio Frizzled.

Universidade de São Paulo

Instituto de Biociências

São Paulo – 2009

Erika Kague

Estudo da regulação transcricional do *COL18A1* e análise funcional do domínio Frizzled.

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

Orientadora: Prof Dr Maria Rita Passos-Bueno

São Paulo

2009

Ficha Catalográfica

Kague, Erika
Estudo da regulação transcricional do *COL18A1* e análise funcional do domínio Frizzled.
Número de páginas 131

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

1. Colágeno XVIII 2. Sequencias conservadas não-codificates 3. Frizzled I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evelutiva.

Comissão Julgadora:

_____ _____

Prof(a). Dr(a). Prof(a). Dr(a).

_____ _____

Prof(a). Dr(a). Prof(a). Dr(a).

_____

Prof(a). Dr.(a).

Orientador(a)

Dedicatória

À minha mãe.

Amor puro e incontestável.

Epígrafe

“I'm not young enough to know everything.”

-Oscar Wilde

Agradecimentos

Durante o período de contemplação do Doutorado além de genética, ciência e pesquisa, ampliei meus conhecimentos sobre a vida. Agradeço cada momento triste e feliz, pois cada um deles, minuciosamente, me fizeram amadurecer e ver a vida de forma única. Pessoas maravilhosas me ajudaram e estiveram comigo e sou plenamente grata.

Agradeço sinceramente à Profa. Dra. Maria Rita Passos-Bueno, pelo exemplo de amor e dedicação à ciência. Pela credibilidade e oportunidades. Obrigada pela liberdade para colocar em pratica minhas idéias. Obrigada pelo apoio e compreensão em todos os momentos. Agradeço a orientação vívida e amizade.

I thank Dr. Shannon Fisher for collaboration in all my work. Thanks for introduce me to the Zebrafish community and to show me so much enthusiasm about science. Thank you for your open heart, friendship and advice. Thank you for incessant laughs. Thank you for all the support. Thank you very much for the example in science and life.

Agradeço à minha primeira orientadora, Dr Nídia Alice Pinheiro, por todo incentivo, apoio e amizade. Obrigada por me encorajar em meus primeiros passos de pesquisa científica.

Agradeço à minha mãe Rosa que tanto amo!

À Rosa mais bela
Coração palpitante, abarrotado de calor
Meiga voz, é ela
Olhos que viram e choraram dor
Mãe querida obrigada por tanto amor.

Agradeço, ao meu pai. Por me ensinar verbalmente e ativamente o significado do injusto e tortuoso. Obrigada pelo ilustre exemplo, que espero o oposto seguir. Guardarei cada uma das delicadas e singelas palavras com uma memória permanente. “Sofrimento”, vocábulo intenso..., obrigada. Obrigada por me mostrar como amar, admirar e repugnar. Me fazer ver o orgulho vivo, a ignorância insuportável, o dinheiro como a mais insignificante e miserável forma de se tratar de Valor.

Agradeço ao meu irmão travesso que amo muito. Minhas tias To e Né sempre tão presentes e atenciosas. Minha avó Lucia. Tio Massao, Mie, Irene, Peter, Akira, Leonor. Minhas primas Jéssica, Paula e Patrícia. Minha Tia Maria.

Ao meu primo e irmão, Jason, contagiando energia e alegria que tanto me ajudou e me apoiou.

Aos meus amigos, Mauricio, Erica, Roberta e Lívia que de uma forma ou outra também são minha família, e das mais engraçadas.

Aos colegas de laboratório, Lucia, Oscar, Cibele, Daniela, Roberto, Fernanda, Érika Yeh, Gerson, Gustavo, Meire, Vanessa, Rodrigo, Camila, Karina, Juliana, Guilherme, Flavia, Carlos e Kátia.

I thank my lab colleagues Seneca, Melissa, Evi, Angela, Josephin, Rob, Diana, Ellie, McCallion’s lab members Ron, David, Jimmy, Zack and Dr Andy.

I also thank my friend Evi, to be so nice and an amazing friend. My friend Russ to be such a great guide! My friend Megana and her Indian style. Thanks Paul for the help and parties.

Agradeço à Constancia, sempre ajudando em tudo. Elenice e Deyse da secretaria do departamento e Helder e Érika da secretaria da pós-graduação sempre muito dispostos em nos ajudar.

Abreviaturas utilizadas:

APC; Adenomatous poliposis coli
ATCC; American Type Culture Collection
CE; Convergência e extensão
CNSs; Seqüências não-codificantes conservadas
CRD; Domínio rico em cisteínas
DMEM; Dulbecco's Modified Eagle Medium
DNA; Acido desoxirribonucléico
dNTPs; Deoxinucleotideo trifosfato
Dsh/Dvl; Dishevelled
ENCODE; ENCyclopedia Of DNA Elements
FAP; Polipose Familiar Adenomatosa
FRZC18; Domínio Frizzled do Colágeno XVIII
FT; Fator de transcrição
G0; Geração Fundadora
G1; Primeira Geração
Gly; Glicina
GSK3; Glycogen synthase kinase 3
JNK; Kinase c-Jun N-terminal
LEF; Lymphoid enhancer binding factor
mRNA; Acido ribonucléico menssageiro
PBS; Tampão salino de fosfato
PCP; Polariadade Celular Planar
PCR; Reação em cadeia da Polimerase
PFA; Paraformaldeido
Ppt; Pipetail
RNA; Acido ribonucléico
ROCKII; Kinase II Rho-associada
sFRPs; Proteínas solúveis relacionadas ao Frizled
SK; Síndrome de Knobloch
Slb; Silberblick
SSC; Tampão de sódio citrato cloro-sódio

TCF; T-cell factor

ZFIN; The Zebrafish Model Organism Database

Sumário

Capítulo I

1. Introdução

1.1. Colágenos

Os colágenos são elementos estruturais encontrados em tecidos conectivos e intersticiais de virtualmente todos os órgãos parenquimais, representando os constituintes mais abundantes da matriz extracelular (Prockop e Kivirikko, 1995). Os membros da família dos colágenos são caracterizados por apresentarem um domínio central de colágeno (COL) formado por repetições triplas de Gly-Xaa-Yaa (as posições Xaa e Yaa são, frequentemente, ocupadas por prolina e hidroxiprolina) relacionadas com formação de estrutura de triplas hélices; no entanto os colágenos variam em tamanho, função e distribuição tecidual (Gelse, Poschl *et al.*, 2003). Domínios não-colagenosos (NC) podem ser encontrados em regiões terminais de alguns tipos de colágenos, como no colágeno XV, e ausentes em outros tipos, como no colágeno I. Cerca de 26 colágenos, geneticamente distintos, foram descritos até o momento (Myllyharju e Kivirikko, 2001). A família dos colágenos pode ser dividia em dois grupos, fibrilares e não-fibrilares. Os colágenos fibrilares (Ex: tipos I, II, III, V e XI) garantem flexibilidade e resistência aos tecidos e representam aproximadamente 90% de todos os colágenos. Os colágenos não – fibrilares (Ex: tipos IV, X, XII, XIV, XV e XVIII) apresentam grande heterogeneidade em estrutura, localização e função (Prockop e Kivirikko, 1995; Gelse, Poschl *et al.*, 2003).

1.2. Colágeno tipo VXIII

O colágeno XVIII pertence ao grupo dos colágenos não – fibrilares, formando junto com o colágeno XV o subgrupo das MULTIPLEXINAS (múltiplos domínios triplas hélices com interrupções) (Oh, Kamagata *et al.*, 1994; Oh, Warman *et al.*, 1994). O cDNA do colágeno XVIII foi clonado em 1993/94, permitindo a elucidação da cadeia α1 de sua estrutura primária (Abe, Muragaki *et al.*, 1993; Oh, Warman *et al.*, 1994; Rehn e Pihlajaniemi, 1994), esta formada por polipeptídeos com 11 domínios não - colagenosos (NC1-NC11) e 10 domínios colagenosos (COL1-COL10) (Oh, Warman *et al.*, 1994; Rehn, Hintikka *et al.*, 1994). O colágeno XVIII é uma heparan-sulfato proteoglicana e possui uma variante polipeptídica caracterizada pela presença de três domínios não-colágenos na porção N-terminal, diferentes do colágeno XV, este uma condroitil - sulfato proteoglicana (Saarela, Ylikarppa *et al.*, 1998).

O gene *COL18A1* foi mapeado no cromossomo 21q22.3, apresenta aproximadamente 105 kb e inclui 43 exons, sendo que o intron 2 corresponde a quase metade do tamanho total do gene (Oh, Warman *et al.*, 1994). Seu ortólogo em camundongos está localizado no cromossomo 10 e apresenta características similares ao humano (102 kb total e 43 exons) (Rehn, Hintikka *et al.*, 1994). Foram identificadas 3 isoformas do colágeno XVIII e a transcrição para essas variedades ocorre através de dois promotores alternativos, o promotor 1 localizado à 5´ adjacente ao exon 1 e codifica a isoforma curta do colágeno XVIII, denominada NC1-303 com 1336 aminoácidos (aa). Esta isoforma é formada pelos exons 1, 2 e 4 – 43. O promotor 2 está localizado a 5´ adjacente ao exon 3 e codifica a isoforma longa (NC1-728 com 1751 aa) e média (NC1-493 com 1516 aa), ambas formadas pelos exons 3 e

4 – 43. O exon 3 apresenta um "*splicing*" alternativo responsável pela formação de dois transcritos que variam em tamanho, os RNAs das isoformas longa e média. Visto que as isoformas curta e longa apresentam a mesma porção carboxi-terminal no exon 43, a diferença existente entre elas é o peptídeo sinal na porção amino-terminal (Elamaa, Snellman *et al.*, 2003).

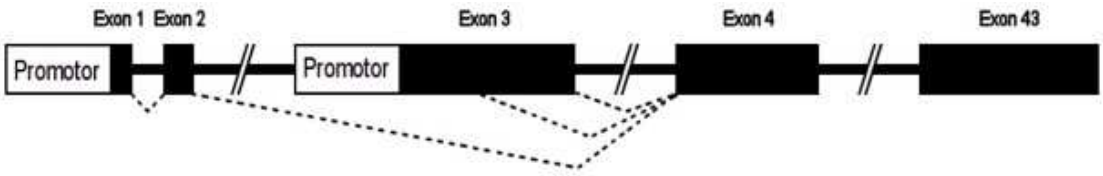


Figura 1. Esquema do gene *COL18A1*, com dois promotores, 43 exons e 3 isoformas distintas (linhas tracejadas). A isoforma longa é gerada por "splicing" no exon 3 (Figura modificada de (Suzuki, Sertie *et al.*, 2002).

Em geral, o colágeno XVIII é expresso em rim, fígado, pulmão e onipresente em membranas basais (Rehn, Hintikka *et al.*, 1994; Halfter, Dong *et al.*, 1998; Saarela, Rehn *et al.*, 1998; Saarela, Ylikarppa *et al.*, 1998). Porém, a existência de dois promotores alternativos no gene *COL18A1* possibilita a expressão diferenciada das variantes do colágeno entre os tecidos. Assim como em camundongos, a variante NC1- 303 de humanos (homóloga a variante NC1- 301 de camundongos) é encontrada predominantemente em rim e em níveis inferiores em coração, placenta e parede endotelial (Saarela, Ylikarppa *et al.*, 1998), além disso, esta isoforma também foi encontrada em retina e cérebro (Sertie, Sossi *et al.*, 2000). A variante NC1-493 é expressa principalmente em fígado e a variante NC1- 728 em pulmão, fígado e em menores concentrações em coração, cérebro e em músculo esquelético fetal (Muragaki, Timmons *et al.*, 1995; Saarela, Rehn *et al.*, 1998; Saarela, Ylikarppa *et al.*, 1998). A expressão do colágeno XVIII já foi investigada em outras espécies mais distantes evolutivamente, como galinha e zebrafish. Interessantemente, o mesmo padrão de expressão foi verificado nestas duas espécies. Em zebrafish o *col18a1* foi

detectado em placa neural anterior e células adaxiais (estágio de 5 somitos), cérebro anterior e mediano, notocorda, espinha cordal, epiderme, ductos pronéfricos, mioseptum, vesícula ótica, campo ocular e primórdios de nadadeira peitoral (estágio de 24hpf a 48hpf) (Haftek, Morvan-Dubois *et al.*, 2003). Em galinhas o *col18a1* foi detectado em retina, cápsula das lentes, epiderme, rim, pulmão, gânglio e nervos do sistema nervoso periférico, espinha cordal e vasos sanguineos (Halfter, Dong *et al.*, 1998).

A função do colágeno XVIII ainda não é bem definida em todos os tecidos em que o mesmo é expresso, no entanto, trabalhos têm demonstrado a relação deste colágeno com algumas doenças, evidenciando sua importância para o bom funcionamento de alguns tecidos e órgãos. Em 2000 Sertié e colaboradores, verificaram que uma mutação no intron 1 do gene *COL18A1* causa a Síndrome de Knobloch (SK), e neste caso há comprometimento apenas da isoforma curta. Este trabalho demonstrou pela primeira vez que o gene do *COL18A1* é o responsável pela etiologia desta síndrome e confirmou que esta doença faz parte do grupo das colagenopatias (Sertie, Quimby *et al.*, 1996; Sertie, Sossi *et al.*, 2000).

A Síndrome de Knobloch é uma desordem autossômica recessiva definida pela ocorrência de alta miopia, degeneração vitreoretiniana com descolamento da retina, anormalidades maculares e encefalocele occipital. Existe uma variabilidade clínica entre os afetados, porém, todos os pacientes apresentam anormalidades oculares severas. A encefalocele também é um dos aspectos clínicos mais observados (Czeizel, Goblyos *et al.*, 1992; Passos-Bueno, Marie *et al.*, 1994). Em 2002 foram identificadas novas mutações patogênicas relacionadas ao fenótipo da SK, sugerindo

que as outras isoformas do colágeno XVIII também poderiam desempenhar um papel crítico na organização e manutenção do olho humano (Suzuki, Sertie *et al.*, 2002).

As anormalidades oculares encontradas em pacientes com a SK sugerem um papel importante do *COL18A1* principalmente em estruturas oculares. Alem disso, defeitos em migração neuronal também foram relatados em pacientes com KS. Interessantemente, *Caenorhabditis elegans* com falta de colágeno XVIII apresentam defeitos em migração celular e direcionamento axonal (Ackley, Crew *et al.*, 2001), indicando o papel funcional do colágeno XVIII também na migração neuronal.

Camundongos nocautes para o colágeno XVIII apresentam atraso na regressão dos vasos hialóides após o nascimento e outros defeitos oculares, como atrofia das células epiteliais ciliares, descolamento do epitélio pigmentar da retina e redução da pressão intra-ocular dependente do avanço da idade (Fukai, Eklund *et al.*, 2002; Marneros e Olsen, 2003). Os camundongos nocautes (*Col18a1*$^{-/-}$) também têm ampla membrana basal dos túbulos renais, alteração na matriz mesangial dos gromérulos e elevado nível de creatinina no plasma, e ainda propuseram que o colágeno XVIII pode desempenhar um papel na suscetibilidade a danos renais associados a diabetes (Utriainen, Sormunen *et al.*, 2004). Dessa forma, o importante papel do colágeno XVIII em diferentes tecidos é incontestável, porém ainda pouco se conhece sobre sua função em cada tecido e quais os mecanismos moleculares utilizados.

1.2.1. Fragmentos clivados do Colágeno VXIII

A endostatina e o Frizzled (FRZC18) são fragmentos proteolíticos do colágeno XVIII, representando domínios de alta conservação com outras espécies. A endostatina apresenta 20- KDa e é derivada dos últimos 184 resíduos da porção C-terminal do colágeno XVIII, região com cerca de 99% de homologia entre humanos e camundongos (Saarela, Ylikarppa *et al.*, 1998). A endostatina tem capacidade de inibir a proliferação de células endoteliais, angiogênese, crescimento metastático e também inibir o crescimento em tumores primários (O'reilly, Boehm *et al.*, 1997). Pacientes com Síndrome de Down, os quais têm menor incidência de tumores sólidos, mostraram níveis plasmáticos elevados de endostatina, indicativo de ação como agente protetor para o desenvolvimento de tumores sólidos (Zorick, Mustacchi *et al.*, 2001). O polimorfismo D104N localizado no domínio da endostatina foi recentemente associado a câncer de mama esporádico (Lourenco, Cardoso-Filho *et al.*, 2006). A endostatina tem sido investigada quanto seu poder anti-angiogênico em diversos tumores e estudos terapêuticos (Wu, Ding *et al.*, 2008; Hu, Kou *et al.*, 2009; Retsky, Hrushesky *et al.*, 2009).

Uma segunda região proteolítica do colágeno XVIII de alta conservação entre humanos, camundongos e *Xenopus* é o domínio Frizzled (FRZC18), derivado da porção N-terminal da variante longa (Elamaa, Snellman *et al.*, 2003). O FRZC18 apresenta 120 aminoácidos contendo uma seqüência de 10 cisteínas invariavelmente espaçadas, “cysteine rich domain” (CRD). A região FRZC18 é homóloga a parte ligante dos receptores envolvidos na via de sinalização Wnt e proteínas solúveis relacionadas ao Frizzled (sFRPs). As sFRPs são proteínas antagonistas que ligam-se diretamente aos Wnts ou a seus receptores (Kawano e Kypta, 2003). Recentemente, o FRZC18 foi descrito como inibidor da via de sinalização Wnt/ß-catenina, com potencial de diminuir crescimento de células tumorais através de aumento de morte

celular em linhagens celulares de câncer humano com ativação de ß-catenina (Quelard, Lavergne *et al.*, 2008). E interessantemente, a endostatina quando injetada em embriões de *xenopus* leva a inibição da via de wnt/ ß-catenina através de indução de degradação citoplasmática (Hanai, Gloy *et al.*, 2002). Dessa forma, os dois fragmentos proteolíticos do colágeno XVIII demonstram capacidade de interferência na via de sinalização Wnt, mas ainda não foi verificado se modificações de concentração de ß-catenina interferiria em expressão do *COL18A1*.

1.3. Sinalização Wnt

Os Wnts compreendem uma família conservada de moléculas glicoprotéicas de sinalização envolvidas em desenvolvimento e processos biológicos, incluindo proliferação, diferenciação e polaridade celular (Logan e Nusse, 2004). A sinalização Wnt é agrupada em duas classes: canônica e não-canônica (Baksh e Tuan, 2007).

O mecanismo principal da sinalização canônica é dependente de estabilização de ß-catenina no citoplasma, ativando genes alvos dos fatores de transcrição TCF/LEF. Na ausência de sinalização Wnt, ocorre ubiquitinação de ß-catenina e sua degradação pelo proteosomo. O complexo citoplasmático de degradação que se liga a ß-catenina é consistindo em pelo menos Axina, proteína APC e GSK-3. A ativação da via canônica inicia-se com ligação de Wnt ao respectivo receptor Frizzled e a correceptores LRP5-6 ("LDL receptor related protein"), transduzindo sinal para Dishevelled (Dsh) e para Axina, levando a inibição da fosforilação de ß-catenina e inibição de sua degradação. Os níveis elevados de ß-catenina no citoplasma leva ao seu acúmulo no núcleo e formação de complexo ativador com os fatores de

transcrição TCF/LEF e consequentemente, indução transcricional de genes alvos. Mutações que promovem a ativação constitutiva da via canônica podem levar ao câncer (Logan e Nusse, 2004). O melhor exemplo conhecido é a Polipose Familiar Adenomatosa (FAP), uma doença autossômica dominante em que os pacientes apresentam pólipos no cólon e ativação exacerbada de ß-catenina (Ponti, Losi *et al.*, 2008). Mutações que levam a perda de função em Axina foram descritas em hepatocarcinomas (Taniguchi, Roberts *et al.*, 2002). Ativação da via Wnt está envolvida com numa grande variedade de tipos tumorais (Logan e Nusse, 2004; Doucas, Garcea *et al.*, 2005). Além disso, o aumento e diminuição de ß-catenina tem papel importante durante osteogêneses e adipogênese, incitando e inibindo respectivamente (Bowers e Lane, 2008; Errera, Canani *et al.*, 2008; Pederson, Ruan *et al.*, 2008).

A via não-canônica envolve outras proteínas não relacionadas com ß-catenina, e ativa vias de sinalização como a "planar cell polarity" (PCP) que guia movimentos durante a gastrulação e via de Wnt/Ca^{2+}(Topczewski, Sepich *et al.*, 2001; Katoh, 2007). Os Wnts ligam-se diretamente aos receptores Frizzled levando a alterações em localização de proteína efetoras da via, como JNK, Dvl e ROCKII ou alterações em concentrações de cálcio intracelular (Habas, Kato *et al.*, 2001; Kishida, Yamamoto *et al.*, 2004; Endo, Wolf *et al.*, 2005; Rosso, Sussman *et al.*, 2005). Durante a embriogênese a via não-canônica participa da determinação de polarização dorso-ventral dos membros. Apesar dos efeitos desta via ser menos conhecidos, estudos indicam a participação desta dia em migração neuroral. O papel do FRZC18 ainda não foi estudado na via não - canônica de Wnt, e a implicação do FRZC18 nesta via poderia ajudar a entender a função do *COL18A1* na migração neuronal, já que

pacientes com SK, com ausência da isoforma longa, apresentam defeitos em migração neuronal.

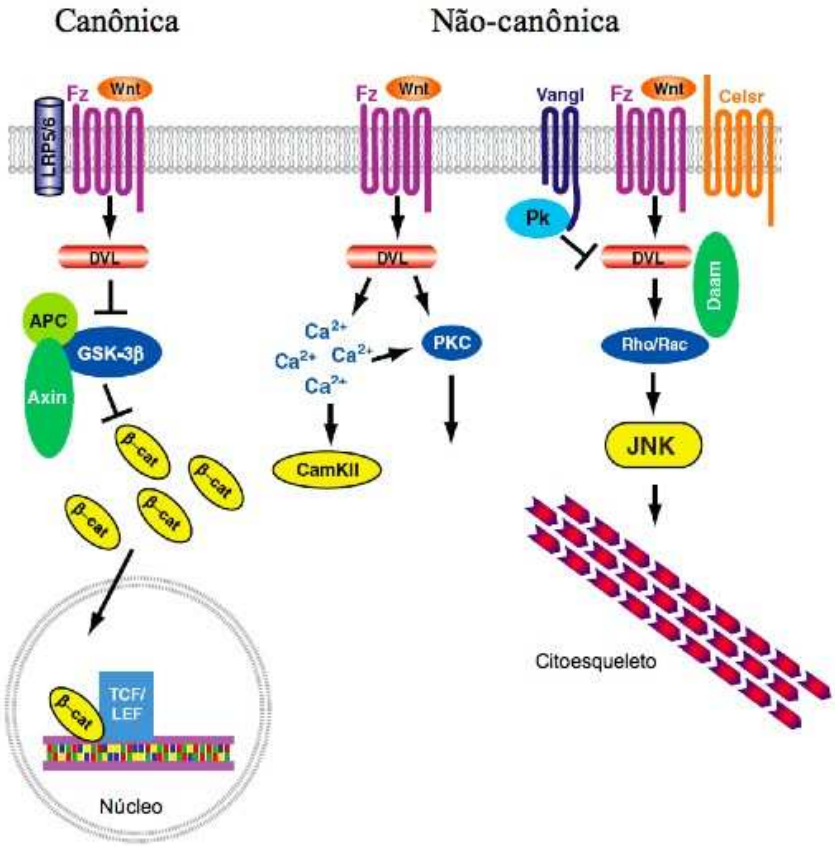


Figura 2. Via de sinalização Wnt canônica e não-canônica. (Figura modificada de (Montcouquiol, Crenshaw *et al.*, 2006).

1.4. Regulação Gênica Transcricional

Os Eucariotos empregam diversos mecanismos para controle de expressão gênica, incluindo condensação de cromatina, metilação de DNA, iniciação transcricional, "splicing" alternativo de RNA, estabilidade de mRNA, controles traducionais, diversas formas de modificações pós-traducionais, tráfego intracelular e degradação de proteína. Dentre essas categorias gerais, a regulação transcricional é o primeiro mecanismo controlador da atividade espaço-temporal dos genes e em

eucariotos a regulação é realizada através de interação de múltiplos fatores de transcrição (FT) (elementos em *trans)* com elementos *cis,* incluindo "enhancers", "silencers", "insulators" e promotores. O promotor está localizado a 5' adjacente ao sítio de início da transcrição e serve como ponto de reconhecimento por fatores de transcrição básicos de recrutamento da RNA polimerase II (Ex. TFIIA, TFIIB e TFIIF). A região controladora de genes típicos de eucariotos apresenta regiões promotoras proximais ("core promoter" - promotor basal) e distais ("enhancers" - elementos regulatórios positivos). Tipicamente, o termo promotor, se refere ao "core promoter" e as seqüências adjacentes ("enhancers proximais"). O "core promoter" é a região mínima necessária para transcrição gênica basal reconhecidas pela RNA Polimerase II, localiza-se imediatamente à 5' do sitio iniciador de transcrição e contém as seqüências canônicas suficientes pelo para reconhecimento da maquinaria transcricional e iniciação da transcrição. Os "enhancers" proximais apresentam sítios de transcrição tecido-específicos bem como os elementos distais de regulação ("enhancers").

Devido a importância de regiões promotoras e "enhancers" mutações e polimorfismos nessas regiões podem acometer a funcionalidade gênica. Mutações em sítios de ligação para fatores de transcrição nas regiões promotoras podem levar a redução transcricional de genes, como observado por Niimi e colaboradores, em estudo relacionando mutação no promotor do gene *UGRP1* com menor transcrição do mesmo e maior risco à asma (Niimi et al., 2002). Lettice e colaboradores, demonstraram que a polidactilia preaxial resulta de mutações de ponto em região reguladora "enhancer" do gene *Shh,* concluindo que mutações de ponto nestas regiões são capazes de causar anormalidades congênitas (Lettice et al., 2003). O promotor 2 do *COL18A1* foi caracterizado previamente e mostrou características de um gene

"housekeeping", com ausência de TATA box e rico em GC. Sp1, Sp3 e YY1 foram fatores de transcrição identificados para regulação deste promotor e um polimorfismo na região está relacionado com diferença de expressão em hepatócitos (Armelin-Correa, Lin *et al.*, 2005). Porém, o promotor 1 do *COL18A1* ainda não foi estudado. E nada se conhece sobre elementos regulatórios deste gene. Tais estudos são de grande importância para compreensão de variabilidade fenotípica observada em pacienntes com SK e também uma tentativa de elucidação de elementos de regulação de expressão de um gene amplamente expresso porém, com picos definidos de expressão em determinados tecidos.

Apesar de se entender a importância de seqüências regulatórias, uma porcentagem ínfima de seqüências não-codificantes funcionais foram identificadas até o momento. O projeto ENCODE ("ENCyclopedia Of DNA Elements") representa a preocupação da comunidade cientifica em melhor entender a localização de elementos regulatórios em tentativa de melhor conhecer o funcionamento da regulação de expressão gênica (The ENCODE (ENCyclopedia Of DNA Elements) Project, 2004). A identificação de elementos regulatórios não-codificantes representa um grande desafio. Uma comparação entre o genoma humano e de camundongo mostrou que 40% do genoma pode ser alinhado. E estima-se que 5% do genoma de mamíferos está sob seleção e predito para serem funcionais. Surpreendentemente, desses 5% apenas 1,5% correspondem exons e o restante é representado por seqüências não-codificantes. Ou seja, seqüências não-codificantes e conservadas correspondem em mais que o dobro de seqüências codificadoras de proteína no genoma (Pennacchio, 2003). Uma hipótese atual que seqüências de importância biológica são conservadas entre as espécies abriu uma janela de estudo e busca de regiões não - codificantes funcionais através de comparação entre seqüências de diferentes espécies,

principalmente aquelas que compartilham uma fisiologia ou biologia comum, como humanos-camundongos. Regiões onde existe grande conservação entre humanos e camundongos são interessantes para comparações com outros organismos, como galinhas e peixes como, fugu e zebrafish (Nobrega e Pennacchio, 2004). Como exemplo podemos destacar a identificação de "enhancer" específico de tecido cardíaco através de comparação entre seqüências de humano, camundongo e galinha e delimitação de uma região de alta similaridade de seqüência e após testes funcionais em camundongos a comprovação que a região regula o gene Nkx2-5 (Lien, Mcanally *et al.*, 2002). Muitos outros estudos têm sido realizados envolvendo regiões conservadas evolutivamente. Regiões pobres em genes e maiores que 500 kb chamadas de desertos gênicos podem conter elementos regulatórios com habilidade de modular a expressão gênica a distância. O gene *DACH* é separado de outros genes por desertos gênicos e para a identificação de regiões conservadas evolutivamente nesses desertos gênicos, Nóbrega e colaboradores alinharam a região com vertebrados, incluindo sapo, zebrafish e puferfish e encontraram 32 regiões conservadas. Em testes funcionais com camundongos reduziram o número para 7 regiões que dirigem a expressão de β-galactosidase em diferentes tecidos (Nobrega, Ovcharenko *et al.*, 2003). Em 2005 o grupo caracterizou "in vivo" o "enhancer" Dc2 do gene *DACH*, este "enhancer" tem identidade de 82% humano-fugu e 424 pb, sendo que apresenta um bloco de importância funcional de 144 pb (Poulin, Nobrega *et al.*, 2005). Bejerano e colaboradores alinharam seqüências de humanos, camundongos e ratos, e encontraram 481 segmentos maiores que 200 pb e com 100% de identidade entre as seqüências, estas regiões foram chamadas de regiões ultraconservadas (Bejerano, Pheasant *et al.*, 2004). Dessa forma, comparação entre seqüências de

espécies diferentes têm sido uma ferramenta importante para a identificação de regiões não-codificantes funcionais no genoma.

Para estudo funcional das regiões não-codificantes e conservadas são utilizados modelos "in vitro" e modelos "in vivo". O camundongo tem sido bastante empregado para estes estudos (Nobrega, Ovcharenko *et al.*, 2003; Poulin, Nobrega *et al.*, 2005). Galinha também é outro modelo animal que pode se utilizado (Uchikawa, Ishida *et al.*, 2003). Porém, atualmente o modelo animal que se destaca é o zebrafish (Pashos, Kague *et al.*, 2008).

1.5. Zebrafish como modelo animal

Dentre os modelos animais que se destacam atualmente para estudos funcionais, o Zebrafish (*Danio rerio*) é um dos mais utilizados. Inúmeras razões incentivam o uso deste modelo como, o seu tamanho pequeno, fácil manutenção dos aquários, rápida reprodução, transparência dos embriões e o alto número de embriões gerados por um único casal (Shin, Priest *et al.*, 2005; Fisher, Grice, Vinton, Bessling, Urasaki *et al.*, 2006). O grande número de mutantes produzidos em curto período de tempo através de mutágenos químicos e a disponibilidade do sequênciamento do genoma dessa espécie, bem como ricos bancos de dados disponíveis para consulta de seqüência e expressão gênica (Ex. ZFIN- "The Zebrafish Model Organism Database"), também fazem desde modelo um alvo ideal para estudos funcionais de domínios protéicos, genes e regiões genômicas não-codificantes, além de estudos envolvendo genética e embriologia. A disponibilidade de programas de bioinformática capazes de promover o alinhamento entre seqüências de diversas espécies e inferências de similaridade gênica nos proporcionam uma útil ferramenta de estudo de

genes humanos em outras espécies. Alinhamento entre mamíferos e espécies mais distantes evolutivamente, como peixes, podem representar potencial funcional, como já observado em diversos estudos (Shin, Priest *et al.*, 2005; Mcgaughey, Vinton *et al.*, 2008). Recentemente foi mostrado que a identificação de "enhancers" pode ser realizada utilizando o modelo de peixe sem que exista alinhamento de seqüência entre humano e teleósteos (Fisher, Grice, Vinton, Bessling e Mccallion, 2006; Fisher, Grice, Vinton, Bessling, Urasaki *et al.*, 2006), o que torna o modelo de zebrafish ainda mais interessante para identificação de elementos regulatórios funcionais.

Além de estudos funcionais envolvendo seqüências não-codificantes, o zebrafish pode ser utilizado para entendimento de vias de sinalização. A disponibilidade de mutantes para proteínas envolvidas nas vias de sinalização é uma grande vantagem deste modelo. Zebrafish mutantes *Slb/Wnt11* e *ppt/Wnt5a,* envolvidos com sinalização não-canônica, foram descritos com participação de movimentos de convergência e extensão e PCP durante a gastrulação em desenvolvimento embrionário de zebrafish (Heisenberg, Tada *et al.*, 2000; Kilian, Mansukoski *et al.*, 2003). Mutantes homozigotos para *Slb* apresentam regiões anteriores afetadas no embrião, mostrando transientemente encurtamento e alargamento do eixo corporal no final da gastrulação seguido de fusão dos olhos em estágios mais avançados de desenvolvimento. Em contraste, os mutantes homozigotos *ppt* exibem um encurtamento do eixo corporal em estágios tardios da gastrulação enquanto que a posição dos olhos não é afetada. *Slb* e *ppt* exibem funções parcialmente sobrepostas em regulação de movimentos de extensão e convergência (CE) em domínios laterais da gástrula, como em duplo nocaute *ppt*, o fenótipo homozigoto de *slb* é mais grave. Embriões de zebrafish injetados com ß-catenina ou proteínas envolvidas na via canônica levam à duplicação de eixo nos embriões,

respostas fenotípicas distintas resultantes de ativação da via canônica e não-canônica de sinalização Wnt (Kelly, Erezyilmaz *et al.*, 1995).

1.6. Objetivos

Pouco se conhece a respeito da regulação transcricional do colágeno XVIII, um constituinte universal de membranas basais e concomitantemente especificamente expresso em determinados tecidos. E com exceção do domínio da endostatina, a função que o colágeno exerce através do Frizzled continua obscura.

Dessa forma, neste trabalho procuramos identificar elementos regulatórios em *cis* do *COL18A1* utilizando-se o modelo de zebrafish; dissecar o promotor 1 do *COL18A1*; explorar se o *COL18A1* é um gene regulado pela ß-catenina e também, a investigação funcional "*in vivo"* do domínio Frizzled do colágeno XVIII na via de sinalização Wnt.

Capítulo II

2. Material e Métodos

2.1. Functionally conserved cis–regulatory elements of COL18A1 identified through zebrafish transgenesis

2.1.1. Seleção e clonagem de seqüências não-codificantes conservadas (CNSs)

Com os programas AVID (Vista browser) e PhastCons (UCSC) buscamos regiões conservadas (identidade ≥ 75%; e janela ≥ 100pb) entre humano/camundongo dispostas em até 100kb upstream do primeiro exon do *COL18A1*, regiões intrônicas e entre *COL18A1* e *SLC19A1* correspondendo a chr21:45527766-45760000, totalizando 232kb. Nomeamos as construções utilizando a abreviação CNS ("conserved non-coding sequences") seguida por + ou – (quando "downstream" ou "upstream", respectivamente, do inicio de transcrição da isoforma curta- exon 1- do *COL18A1*), e *v* quando selecionadas a partir do VISTA e *p* quando atraves do PhastCons. Todas as regiões CNSs foram amplificadas de DNA genômico Humano. Para a reação de PCR utilizamos enzima Takara LA Taq (Takara Mirus Bio). A temperatura de anelamento dos primers foi de 60˚C. Segue abaixo as construções e suas respectivas posições.

Construções	**Coordenadas (Ch21)**	**Tamanho (pb)**
CNSp - 96.9	45551381- 45552562	1181
CNSv - 97.8	45551403- 45551670	267
CNSp - 80.5	45568753- 45568986	233
CNSv - 78.7	45570340- 45570794	454
CNSp - 72.2	45575491- 45577292	1801
CNSv - 64.9	45584370- 45584612	242
CNSv - 52.9	45596190- 45596573	383
CNSp - 47.4	45601821- 45602101	280
CNSv - 40.8	45608418- 45608709	291
CNSv - 38.0	45611223- 45611488	265
CNSp - 28.7	45620217- 45620766	549
CNSv - 26.7	45622453- 45622740	287
CNSv +0.6	45650150- 45650576	427

CNSv + 3.1	45652238- 45652652	414
CNSv + 25.5	45674812- 45675078	266
CNSp + 47.8	45694764- 45697397	2633
CNSp + 51.7	45700781- 45701305	524
CNSv + 51.7	45700938- 45701276	338
CNSv + 64.8	45714075- 45714384	309
CNSp + 65.7	45714053- 45715256	1203

Tabela 1. Construções analisadas neste trabalho seguidas por suas respectivas coordenadas no cromossomo 21 e o tamanho das mesmas.

2.1.2. Injeções

As agulhas utilizadas para injeção de material genético em embriões de Zebrafish são preparadas a partir de capilares de vidro de 1,2 mm submetidos a aquecimento e extensão em Micropipette Puller Model P87 (Sutter Instrument Co.) com programa suficiente para deixar as pontas das agulhas firmes o suficientes para serem cortadas com uma lamina de aço. Apos o corte das pontas das agulhas, as mesmas são armazenadas em placas de Petri ate serem usadas.

Nas extremidades de duas agulhas é adicionado 1,5 µL da mistura de injeção (tabela 4) de uma mesma construção. Estas agulhas são suficientes para a injeção de uma construção em cerca de 200 embriões. A agulha é colocada em agulha do injetor (Pneumatic Pico Pump PV 820 injector (World Precision Instruments) e as injeções são realizadas com auxilio de lupa Stemi 2000 stereomicroscope (Carl Zeiss). Cerca de 1nL da mistura de injeção é injetada em saco vitelínico de cada embrião de estagio de 2 celulas, em região logo abaixo do blastomero, como ilustrado na figura 1 abaixo.

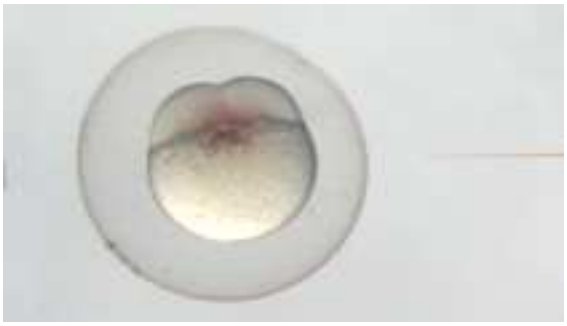


Figura 3. Injeção de material genético em embrião de Zebrafish. A injeção e realizada logo abaixo do blastomero em estagio de duas células. O vermelho fenol possibilita a visualização do material injetado (Figura cedida pela Dra Shannon Fisher).

Cada construção foi injetada em no mínimo 200 embriões. Após as injeções os embriões foram mantidos em mesma placa a 28˚C para analises posteriores.

Componentes	Quantidade (µL)
Construção (125 ng/µL)	1,0 µL
Tol2 RNA transposase (125 ng/µL)	1,4 µL
Phenol Red	0,5 µL
RNase-free Water	2,1 µL
Total	5,0 µL

Tabela 2. Mistura para injeção.

2.1.3. Análise dos embriões

Os embriões foram analisados para a expressão de EGFP de 1-5 dias após fertilização (dpf) com auxilio de estereomicroscópio Lumar V12 Stereo (Carl Zeiss). Os embriões foram tratados com tricaína, para imobilização facilitação das análises. Os embriões que apresentaram expressão de EGFP coerente com expressão endógena de *col18a1* em zebrafish foram selecionados e depois de 5 dias foram transferidos para aquários onde estão sendo mantidos para transmissão germinativa e estabelecimento de linhagem transgênica.

2.1.4. Geração de linhagem transgênica G1

Os embriões G0 foram analisados para expressão mosaica de EGFP e aqueles com expressão coerente com o *col18a1* endógeno foram selecionados e mantidos por 3

meses para estabelecimento de linhagem transgênica (G1) e identificação dos fundadores (G0). Três meses é o período mínimo necessário para maturação das células germinativas, tornando a reprodução viável. Dessa forma, o período, os peixes foram cruzados com peixes de linhagem selvagem AB, e os embriões provenientes desse cruzamento e que apresentaram expressão de EGFP foram selecionados e fotografados (G1). O esquema da metodologia encontra-se na figura 2.

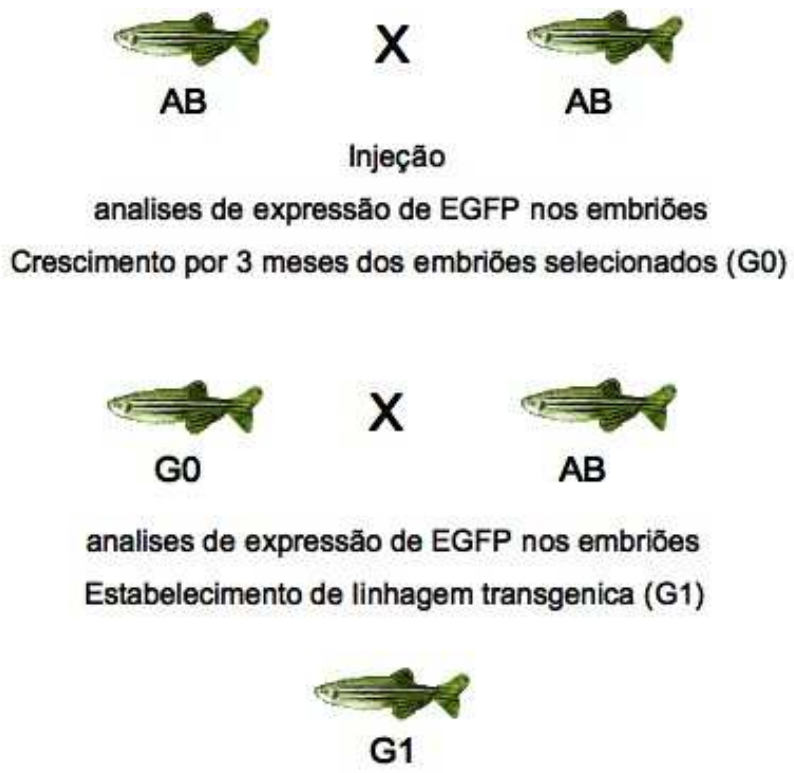


Figura 4. Geração de linhagem transgênica para expressão de EGFP para as construções de CNSs.

2.1.5. Hibridização in situ *para o* col18a1 *em Zebrafish e camundongos*

2.1.5.1. Preparação da sonda zebrafish

Amplificamos 598pb a partir do vetor cb371 proveniente do "The Zebrafish Model Organism Database" (ZFIN- http://zfin.org/cgi-bin/webdriver?MIval=aa-

ZDB_home.apg) que apresenta clonada parte da seqüência de cDNA do *col18A1* que inclui todas as isoformas. Upstream ao primer R foi incorporada seqüência de reconhecimento para RNA polimerase T7, para que pudéssemos realizar transcrição reversa desse segmento amplificado. Os primers utilizados foram: cb371F-GAATTCCTGGCACTGGATCA e cb371R-taatacgactcactatagggagTGGAAGCACATGAAGTCTGC. A temperatura de anelamento usada foi de 60˚ C. E as condições de reação foram:

Reagentes	**Quantidade (µL)**
10 X PCR Buffer	5,0
2,5 mM dNTPs	4,0
$MgSO_4$	1,0
Primer F	1,0
Primer R	1,0
Enzima Platinum Pfx (Invitrogen)	0,3
H_2O	35,7
DNA (100ng/µL)	2,0
Total	50,0

Tabela 3. Reagentes utilizados para amplificação de cDNA para preparo de sonda para *col18a1*

O produto de PCR foi cortado do gel e purificado com kit Qiagen e eluido em 20 µL. Para a transcrição reversa foram utilizados 5 µL do produto de PCR purificado e a enzima T7 RNA polimerase conforme as condições abaixo:

Reagentes	**Quantidade (µL)**
Produdo de PCR purificado	5,0
Transcription Buffer	4,0
NTP-DIG-RNA	2,0
RNase inhibitor	1,0
T7 RNA polimerase	1,0
H_2O	8,0
Total	20,0

Tabela 4. Reagentes para transcrição in vitro para preparo da sonda para col18a1.

A mistura de reagentes foi incubada a 37˚C por 2 horas. Recolhemos 1µL para analise em gel de agarose. E digerimos o restante adicionando 1 µL de DNAse por 15 minutos a 37˚ C. O RNA foi purificado com coluna SigmaSpin (Sigma) e eluído em

100 µL de H_2O. Comparamos em gel de agarose antes do tratamento com DNase e depois do tratamento para verificar o RNA e ausência de DNA.

2.1.5.2. Preparação da sonda para hibridização em camundongos

Embriões tipo selvagem de camundongo foram coletados e fixados (4% paraformaldeido) 9.5 dias pós-copula (dpc). A ribo-sonda para a hibridização foi gerada através de amplificação de cDNA de camundongo utilizando-se os primers: F-AGTTCCACATCACCACAGTTCCTAT e R-TAATACGACTCACTATAGGGAGCATGTCTCACAGTAACTCTCCATCA e sintetizado com T7 RNA polimerase assim como já descrito para zebrafish. A hibridização in situ para todo o embrião foi realizada seguindo protocolos já descritos.

2.1.5.3. Hibridização

Embriões de 1 a 5dpf foram conservados em paraformaldeido (PFA). Antes de serem utilizados para a hibridização in situ.

Primeiro dia

Os embriões foram transferidos para tubos de 1,5 mL e re-hidratados por sucessivas incubações em:

75% MeOH - 25% PBS por 5 minutos

50% MeOH - 50% PBS por 5 minutos

25% MeOH – 75% PBS por 5 minutos

100% PBT (PBS/0,1% Tween 20) 4 lavagens (sob rotação) de 5 minutos cada.

Depois de re-hidratados os embriões de mais de 36h foram submetidos a digestão com Proteinase K (10 µg/mL em PBT) por 30 minutos. Em seguida foram

refixados em 4% de PFA em PBS por 20 minutos e lavados 5 vezes com 100% PBT por 5 minutos.

Enquanto as lavagens eram realizadas, foi preparado o mix de pré-hibridizaçãoo e de hibridização.

	Solução de 50% de formamida
Formamina	10 mL
20X SSC	5 mL
20% Tween 20	100 µL
1M Acido Cítrico	184 µL
Heparina, 5mg/mL	200 µL
tRNA, 50mg/mL	200 µL
Agua	4,3 mL

Tabela 5. Mix de pré- hibridização (hibridização)

Os embriões foram pré-hibridizados em 800 µL de mix de hibridização por 2-5h a 70˚C. Esta solução foi removida apos esse tempo e 200 µL de solução de hibridização contendo cerca de 200 ng de sonda de RNA foi adicionada ao tubo com os embriões a 70˚C "overnight".

Segundo Dia

Devem foram procedidas lavagens da seguinte forma:

100% HM (mix de hibridização sem heparina e sem RNAt) a 70˚C por 15 minutos

75 % HM/25% 2X SSC a 70˚C por 15 minutos

50 % HM/50% 2X SSC a 70˚C por 15 minutos

25 % HM/75% 2X SSC a 70˚C por 15 minutos

2X SSC a 70˚C por 15 minutos

0,2 2X SSC a 70˚ C por 30 minutos (duas lavagens)

75% 2X SSC/25% PBT em temperatura ambiente por 10 minutos

50% 2X SSC/50% PBT em temperatura ambiente por 10 minutos

25% 2X SSC/75% PBT em temperatura ambiente por 10 minutos

100% PBT em temperatura ambiente por 10 minutos

Em seguida o anticorpo foi bloqueado com tampão bloqueador (20% soro de cabra, 2% BMP, 150mM de N_aCl e 100mM de acido maleico pH 7,5) em temperatura ambiente por no mínimo 1 hora. Apos esse período os embriões foram incubados em solução de anticorpo "overnight'. A solução de anticorpo foi preparada a proporção 1:5000.

Terceiro Dia

Foram realizadas lavagens com :

PBT a temperatura ambiente, brevemente

PBT a temperatura ambiente , 6 vezes de 15 minutos cada (em plataforma giratória)

Tampão de coloração (100mM tris HCl pH 9,5, 50mM M_gCl_2, 100mM de N_aCl, 0,1% de Tween 20), 3 vezes de 5 minutos cada.

Os embriões foram colocados em placas de vidro com depressões , junto com solução de coloração (esta solução e sensível a luz e os embriões são mantidos no escuro nesta etapa). A solução compreende na adição de 2,25 μL de NBT e 3,5 de BCIP para cada 1 mL de tampão de coloração.

A coloração foi paralisada quando os embriões estavam devidamente corados. Para a paralisação da coloração, a solução de coloração foi removida e os embriões foram lavados em solução de paralisação (1X PBS ph5,5 e EDTA 1mM ou apenas PFA). Os embriões estão sendo mantidos em PFA a 4˚C.

2.1.6. *Analises* Post hoc

Realizamos alinhamento de todo o intervalo do *col18a1*, de humano, camundongo, galinha, stickleback e zebrafish, utilizando o MultiPipmaker e o algoritmo Shuffle-LAGAN, incorporado como parte do site do VISTA browser (identidade de 50% e janela de 100pb).

2.2. Dissection of promoter 1 of COL18A1 and feedback control of wnt signaling through regulation of COL18A1 expression

2.2.1. Amplificação e clonagem

Para a analisar a região promotora 1 do *COL18A1* foram desenhados primers com o programa Primer Express (Applied Biosystems) versão 2.0. A 5' de cada primer foi adicionado o sitio attB de recombinação compatível para o sistema de clonagem Gateway da Invitrogem. Os primers estão listados na tabela 6. Um mesmo primer R foi utilizado para a amplificação, a partir de DNA humano controle, de 10 regiões a 5' do primeiro exon do *COL18A1* (figura 3), estas regiões diminuem de tamanho em sua porção 5', simulando deleções a 5' do promotor do *COL18A1*. As condições de PCR estão listadas na tabela 2. Utilizamos 95 por 5 minutos para desnaturação inicial, 40 ciclos de 95 por 50 segundos, 60 por 40 segundos e 72 por 1:30 minutos e para extensão final 72 por 10 minutos.

Primers	Seqüência
COL18A1-attB2- R	GGGGACCACTTTGTACAAGAAAGCTGGGTGCTCTCCCCAGGACCGTCAGCG
103- F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTGCCCGGCCTCCCAGCGG
347 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCCTCGGGGCGGGACGTGT
273 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGCAGCCGCAGGTCCGCT
540 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTAGGTGCCGGCTTCCTCCTCCGTC
752 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGGCTCAGGATGCTGGCCCTGAA
919 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAGAGGGTTTCAGAGAAAGCCGTCAAGCC
1036 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTAGGATGGGTGGGGCCACGTGTG
1554 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTAGACACACAGACACACCCACTGCACTGC
1253 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTATTGGGCCTGGGAAGGCTGAGGC
1702 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTAACACCCCTGACACACACTCACACGCA

Tabela 6. Primers utilizados para amplificação de região 5' do COL18A1

Reagentes	**Quantidade (µL)**
10 X PCR Buffer	2,5
2,5 mM dNTPs	3,0
$MgSO_4$	0,5
Primer F	0,5
Primer R	0,5
Enzima Platinum Pfx (Invitrogen)	0,25
Solução Enhancer	5,0
H_2O	10,75
DNA (100ng/µL)	2,0
Total	25,0

Tabela 7. Reagentes utilizados para amplificação de região 5' do primeiro exon do COL18A1.

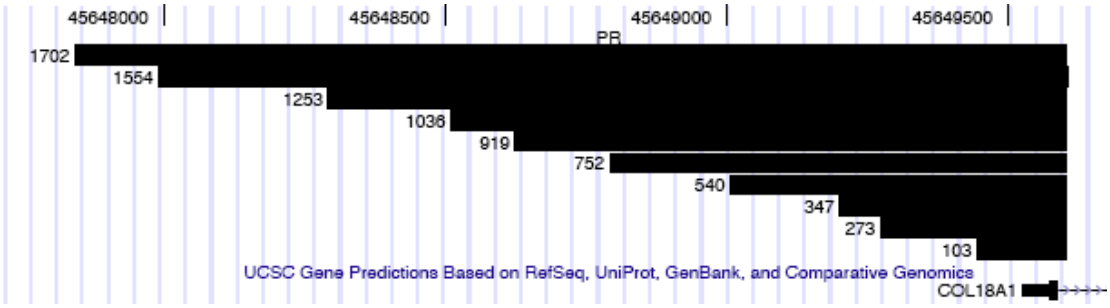


Figura 5. Estão indicados 9 fragmentos clonados upstream ao primeiro exon do COL18A1 e suas posições no cromossomo 21. O décimo fragmento não esta representado por ser o 1702 invertido.

Todas as 10 regiões amplificadas foram clonadas em vetor pGL3-Basic-gateway que apresenta gene repórter de luciferase. Sítios pra recombinação LR compatíveis para uso com sistema gateway de clonagem foram adicionados no vetor pGL3-basic.

2.2.2. Teste de linhagem celular de alta expressão de COL18A1 *e PCR em Tempo Real*

Testamos quantitativamente, através de PCR em tempo retal (Real time PCR), a expressão de *COL18A1* nas seguintes linhagens celulares: Hek293T (rim) , HELA (cérvix), SHSY (neuroblastoma), Hfob (osteoblastos) e HepG2 (hepatocarcinoma). As células com confluência de aproximadamente 95%, em garrafas de 25 cm^3, foram tripsinizadas e o RNA total dessas células foi extraído com o método de Trizol (Gibco), de acordo com o fabricante. O RNAm (1 μg) foi convertido em cDNA, utilizando-se a transcriptase reversa (SuperScript III, Invitrogen). A reação de PCR em tempo real foi realizada com o reagente SYBR Green, no 7500 Real Time PCR System e analisada com o software 7500 System SDS software v1.2 (Applied Biosystems). O RNAm alvo foi normalizado pelos controles endógenos *HPRT1* (*hypoxanthine phosphoribosyltransferase 1*) e *SDHA* (*succinate dehydrogenase complex, subunit A,*

flavoprotein). Cada amostra foi analisada em triplicata e a quantificação relativa da expressão gênica foi normalizada pela média de Hek293T, segundo o método comparativo $\Delta\Delta C_T$. A amplificação foi realizada seguindo as condições: 50°C por 2 minutos, 95°C por 10 minutos e 45 ciclos de 95°C por 15 segundos e 60° por 1 minuto. As reações foram realizas para volume total de 25 µL.

Para estudo de expressão de *COL18A1* durante a adipogenese e com níveis aumentados de ß-catenina, utilizamos no total quatro controles endógenos, alem de *SDHA* e *HPRT1,* também utilizamos *GAPDH* (Glyceraldehyde 3-phosphate dehydrogenase) e *HMBS* (Hidroxymethylbilane sinthase). Os genes alvos estudados foram: *COL18A1, Cyclin D1, cMyc,* CEBPa*, CEBPb* e *PPRA-gamma*. As analises foram realizadas com o método de ΔΔCT e normalização com valores gerados pelo Genorm.

2.2.3. Ensaio de transfecção com gene repórter de Luciferase

Os vetores foram purificados (midi-prep Qiagen) e utilizados para transfecção em duas linhagens celulares de alta expressão de *COL18A1* (Hek293T e HepG2).

As linhagens celulares HepG2 e Hek293T foram cultivadas em garrafas de 25 cm^3 com meio DMEM High Glucose, 10% de soro fetal bovino e 1% de antibiótico (ampicilina/ estreptomicina) em estufa com 5% de CO_2 como recomendado pela American Type Culture Collection (ATCC).

Quando a cultura atingiu cerca de 95% de confluência, as células foram tripsinizadas e contadas com auxílio de câmara de Neubauer. Transferimos as células para placas de 12- *Wells* 24h antes da transfecção (1×10^5 para Hek293T e 2×10^5 para HELA). Realizamos cotransfecção de 1 µg de vetor repórter para luciferase e 0,25 µg de plasmídio pRL-TK (Promega) com Lipofectamina 2000 (Invitrogen). Após 48h da

transfecção as células foram lisadas e analisadas utilizando-se o Dual-*Luciferase Repórter Assay System* (Promega) em luminômetro (LB96V Microlumat Plus-EG&Berthold). Cada transfecção foi realizada em duplicata e com três experimentos independentes. A atividade das construções foram normalizadas com o vetor repórter de luciferase isento de inserto. Para transfecções referentes à ß-catenina utilizamos Hek293T selvagem e também com expressão permanente de ß-catenina. Transfectamos cada construção em cada tipo de linhagem celular independentemente e em triplicatas. Utilizamos 500ng de pGL3-basic P1-103, P1-540, P1-1036, P2-182, P2-555 e P2-1052. A atividade de luciferase foi normalizada com ß-galactosidase (ßgal) (100ng). As construções OT-Luc (com sítios de ligação para TCF/LEF) (500ng) e ØF-Luc (com sítios mutados para TCF/LEF) (500ng) foram utilizadas para controle positivo e negativo, respectivamente.

Todos os valores de transfecção foram analisados e comparados com "One- Way ANOVA" com comparação múltipla Tukey´s utilizando o *GraphPad Prism* versão 4.00 para Windows. Utilizamos um valor mínimo significativo de $P<0,05$.

Para superexpressão de ß-catenina em Hek293T, transfectamos esta linhagem independentemente com vetor de expressão de ß-catenina e vetor de expressão de Wnt1. Transfectamos 24µg de cada vetor em garrafas de 75 cm^3. A extração de RNA foi realizada 48h depois do inicio de transfecção com o MN RNA purification kit (MN).

2.2.4. Imunoprecipitação de Cromatina (CHIP)

Realizamos CHIP com anticorpos para TCF-1 e LEF-1 (Santa Cruz), utilizando EZ ChiP ^{tm} (upstate) seguindo instruções do kit. A cromatina foi fragmentada com sonicação. Para controle positivo e negativo de experimento imunoprecipitamos a cromatina com anti-RNA Polimerase e mouse anti-IgG, respectivamente. Também

tratamos seguindo o protocolo a amostra input, que não apresenta precipitação com nenhum anticorpo. Após precipitação, realizamos PCR com primers do promotor 1 do *COL18A1* para as regiões 540 e 1036 e também para o promotor 2, região 427.

2.2.5. *Analises in silico*

Nós utilizamos o PROSCAN (http://www-bimas.cit.nih.gov/molbio/proscan/) para predição de promotor para o *COL18A1* e submetemos 2000pb referentes à seqüência a 5' adjacente ao inicio de transcrição da isoforma curta (exon 1). Cada região PE1-3 foi analisada independentemente com uma seqüência ortóloga de camundongo, com 1500 pb 5' adjacente ao início de transcrição da isoforma curta do *col18a1* com o programa MEME (http://meme.sdsc.edu/meme4_1/cgi-bin/meme.cgi), para identificação de motivos similares. Em seguida investigamos os fatores de transcrição preditos em cada motivo utilizando o TESS (http://www.cbil.upenn.edu/cgi-bin/tess). Alem disso, também procuramos fatores de transcrição conservados entre humano e camundondo com uso do CONSITE (http://asp.ii.uib.no:8090/cgi-bin/CONSITE/consite/). O Ingenuity foi utilizado para mapear possível ligação entre o *COL18A1* e a via de LEF/TCF.

2.2.6. *Construção, injeção e analises de expressão de EGFP em zebrafish*

A região PE3 foi amplificada com PCR (PE3 F-AGACACACAGACACACCCACTGCACTGC e PE3 R-GCCTCAGCCTTCCCAGGCCCAAT) com polimerase de alta fidelidade, de DNA genomico. A região amplicada foi subclonada em vetor Tol2 baseado em transposon, pGWcfosEGFP, como previamente descrito.

As injeções e analises de EGFP foram realizadas como já mencionado para analise mosaica na linhagem G0 de embriões. Este experimento foi realizado no laboratório da Dra Shannon Fisher, uPENN.

2.3. Collagen XVIII cleaved frizzled fragment involvement with non-canonical Wnt signaling

2.3.1. Construção de plasmídeos e síntese de mRNA

O domínio FRZC18 foi amplificado de DNA genômico (Roche) com os seguintes primers: F – ACCATGtgcctgcccctgccaccctcc e R- cacacagtacccatcctcctgggt. Os produtos de PCR foram clonados em vetor TOPO utilizando o kit pCR8/GW/TOPO TA Cloning kit (Invitrogen). A reação de recombinação LR (invitrogen) foi realizada para transferir o FRZc18 do vetor TOPO para o vetor destino de zebrafish do sistema gateway, pCSDest. Para o plasmídeo controle nós inserimos a seqüência de Gly-Ser-Gly (CCCAGGCCC) entre a posição 31 e 32aa da proteína FRZC18. A seqüência CCCAGGCCC foi inserida com dois passos de amplificação (a e b) utilizando-se os seguintes primers: 1aF-GGGGACAAGTTTGTACAAAAAAGCAGGCTACCATGTGCCTGCCCTGCCACCCTCC CTG e Insertion 1aR- CACCTGGGGCCTGGGCTCGCCGCTCTCGT, Insertion 1bF-ACGAGAGCGGCGGAGCCCAGGCCCCAGGTGCGGGCCGGGCAC e Insertion 1bR- GGGGACCACTTTGTACAAGAAAGCTGGGTcacacagtacccatcctcctgggt. Corte com enzima de restrição BanII de ambos os amplicons foi realizado e seguido por reação de ligação com T4 DNA ligase. O produto da ligação foi clonado em vetor pDONR 221 com enzima BP clonase (invitrogen). A reação LR foi realizada para obtermos o PCSDes-FRZC18-Mut. Os vetores pCSDes-FRZC18 e PCSDes-FRZC18-Mut foram linearizado com HpaI antes de seguirmos com a síntese de mRNA com o

kit mMESSAGE mMACHINE (Ambion). Extração com fenol:clorofórmio e precipitação com isopropanol foi utilizada para purificação de mRNA.

2.3.2. Cuidados com peixes e Injeções

Os peixes foram mantidos assim como nos trabalhos mencionados anteriormente, no laboratório da Dra. Shannon Fisher na Johns Hopkins Medical School em Baltimore/USA. Injetamos concentrações de mRNA variando de 0.01 a 0.4 ng em ≥ 100 embriões para cada concentração. Os embriões foram analisados em procura de alterações fenotípicas de 24h a 4dpf.

2.3.3. Genotipagem de Zebrafish

Cruzamos $Ppt^{+/-}$ e $slb^{+/-}$ com peixes selvagens AB. Injetamos 0.15ng de mRNA de FRZC18 nos embriões resultados do cruzamento. Os embriões que mostraram alteração fenotípica foram coletados a 3dpf e decorionados para extração de DNA. A preparação de DNA genômico foi realizada seguindo protocolos já estabelecidos (The Zebrafish Book). Os PCRs para amplificação dos sítios mutados em peixes foram realizados e seguidos de digestão com SacII e MseI para pipetail e silberblick, respectivamente. O produto de PCR que foi cortado com a enzima, representa o peixe heterozigoto para o sitio.

2.3.4. Transfecção e Western Blots

FRZC18 foi clonado no vetor Lumio Gateway (Invitrogen). Nós transfectamos células HEK 293T quando em 70% de confluência com Lipofectamina 2000 (Invitrogen). As células foram lisadas 72h após a transfecção. A fração de proteína da membrana foi isolada com o kit de preparação de proteína Mem-PER kit (Pierce), as frações citoplasmáticas e nucleares foram isoladas com NE-PER kit (Pierce). A

concentração de proteína foi medida através do método Lowry e as amostras foram normalizadas. 20 µg de proteína total foi separada por SDS-PAGE em Rgels (10% ou 4-15% Tris-HCl) e transferidas para uma membrana PVDF (Bio-Rad). A membrana foi incubada com anticorpos comerciais de coelho (anti-ß-catenin (1:100, Santa Cruz SC7199), anti-JNK/SAPK1 (1:1,000, Upstate cat. No. 06-748)), in vaca (anti-Dvl (C-19) (1:500, Santa Cruz SC7397)) e em camundongo (anti-GAPDH (1:5,000, Abcam ab 9894)). As proteínas foram detectadas com anticorpos secundários conjugados com peroxidase de (1:10,000 cada, RPN4201) e o ECL (both GE Healthcare) para detecção.

Capítulo III

3. Functionally conserved cis–regulatory elements of COL18A1 identified through zebrafish transgenesis

Erika Kague[1], Seneca Bessling[2], Josephine Lee[2], Maria Rita Passos-Bueno[1], and Shannon Fisher[3†].

1. Instituto de Biociencias/Universidade de Sao Paulo- Brazil

2. McKusick–Nathans Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD 21205

3. Department of Cell and Developmental Biology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104

†Correspondence should be addressed to S.F.

421 Curie Blvd.

BRB II/III 1213

Philadelphia, PA 19104

sfisher4@upenn.edu

Phone: 215.898.7886

FAX: 215.898.

Key words: collagen XVIII; Knobloch syndrome; zebrafish; transgenesis; cis–regulatory element; comparative genomics

Capítulo III- Functionally conserved cis–regulatory elements of *COL18A1* identified through zebrafish transgenesis

Abstract

Type XVIII Collagen is a component of basement membranes, and expressed prominently in the eye, blood vessels, liver, and the central nervous system. Homozygous mutations in *COL18A1* lead to Knobloch syndrome, characterized by ocular defects and occipital encephalocele. However, relatively little has been described on the role of type XVIII collagen in development, and nothing is known about the regulation of its tissue–specific expression pattern. We have used zebrafish transgenesis to identify and characterize cis–regulatory sequences controlling expression of the human gene. Candidate enhancers were selected from non–coding sequence associated with *COL18A1* based on sequence conservation among mammals. Although these displayed no overt conservation with orthologous zebrafish sequences, four regions nonetheless acted as tissue–specific transcriptional enhancers in the zebrafish embryo, and together recapitulated the major aspects of *col18a1* expression. Additional post–hoc computational analysis on positive enhancer sequences revealed alignments between mammalian and teleost sequences, which we hypothesize predict the corresponding zebrafish enhancers; for one of these, we demonstrate functional overlap with the orthologous human enhancer sequence. Our results provide important insight into the biological function and regulation of *COL18A1*, and point to additional sequences that may contribute to complex diseases involving *COL18A1*. More generally, we show that combining functional data with targeted analyses for phylogenetic conservation can reveal conserved cis–regulatory elements in the large number of cases where computational alignment alone falls short.

Introduction

Collagens are collectively the most abundant component of extracellular matrix(Gelse, Poschl *et al.*, 2003). Although collagens are universally present in human tissues, some types have a constrained pattern of expression. Collagen type XVIII is a nonfibrillar collagen that, together with collagen XV, comprises the subgroup of multiplexins (multiple triple helix domain with interruption)(Saarela, Ylikarppa *et al.*, 1998) (Rehn e Pihlajaniemi, 1994). *COL18A1* maps to HSA 21q22.3 and its 43 exons span 108.5 Kb(Oh, Warman *et al.*, 1994). Three isoforms are generated by the use of two promoters and alternative splicing in the third exon (Rehn, Hintikka *et al.*, 1996) (Suzuki, Sertie *et al.*, 2002) (Saarela, Ylikarppa *et al.*, 1998). The short isoform (NC11-303) is transcribed from the first promoter, upstream of exon 1, and contains exons 1, 2, and 4-43. The intermediate (NC11-493) and long (NC11-728) forms are transcribed from the second promoter, upstream of exon 3, and differ by an internal splice site within exon 3. Although collagen XVIII is a component of vessel walls and almost all basement membranes (Halfter, Dong *et al.*, 1998), the three isoforms have distinct patterns of expression. The short isoform is expressed in kidney, retina, placenta and human fetal brain (Sertie, Sossi *et al.*, 2000) (Saarela, Rehn *et al.*, 1998), whereas the intermediate and long isoforms have their highest expression levels in liver and lung, respectively(Suzuki, Sertie *et al.*, 2002) (Saarela, Rehn *et al.*, 1998). Three isoforms of collagen XVIII are also present in mice, with tissue-specific expression pattern apparently similar to human (Muragaki, Timmons *et al.*, 1995). Col18a1 expression is important for retinal pigment epithelium and hyaloid vessel maintenance, since mice lacking collagen XVIII show defects in these tissues (Fukai, Eklund *et al.*, 2002) (Hurskainen, Eklund *et al.*, 2005). However, its expression has not been examined in whole mouse during embryogenesis. In zebrafish embryos from

segmentation to hatching, *col18a1* is expressed in pronephric ducts, otic vesicle, forebrain, tectum, epidermis, spinal cord, pectoral fin bud, myoseptum and notochord (Haftek, Morvan-Dubois *et al.*, 2003). Expression of *col18a1* after hatching, or at any stage in liver, gut endoderm, or blood vessels has not been described.

Collagen XVIII has multiple recognizable domains and is likely to play a role in diverse biological processes. Endostatin, a 20 kDa C–terminal proteolytic product of collagen XVIII, acts as an inhibitor of angiogenesis and tumor growth (O'reilly, Boehm *et al.*, 1997) (Marneros e Olsen, 2001) (Marneros e Olsen, 2005) (Abdollahi, Hahnfeldt *et al.*, 2004). Collagen XVIII also has two N-terminal domains: trombospondin-1 like, of unknown function; and frizzled, a cleaved fragment related to the frizzled family of cell surface Wnt receptors (Elamaa, Snellman *et al.*, 2003) (Quelard, Lavergne *et al.*, 2008). The zebrafish *diwanka* gene, required for motor axon guidance, encodes a lysyl glycosyltransferase whose main target relevant to axon guidance is thought to be myotomal collagen XVIII (Schneider e Granato, 2006). The *C. elegans* homologue of *col18a1* is also required for axon guidance (Ackley, Crew *et al.*, 2001), suggesting an ancient conserved role for collagen XVIII in neural development.

Further insights into *COL18A1* function have arisen from the finding that null mutations in *COL18A1* cause Knobloch syndrome (KS), an autosomal recessive disorder characterized by high myopia, vitreoretinal degeneration with retinal detachment, macular abnormalities, and occipital encephalocele, which is the pathognomonic feature of the syndrome (Sertie, Sossi *et al.*, 2000) (Suzuki, Sertie *et al.*, 2002). Inter– and intra–familial clinical variability is present, but all patients have ocular abnormalities that usually lead to bilateral blindness. Some patients also display generalized hyperextensibility of the joints, unilateral duplicated renal collecting

system, epilepsy, and evidence of altered neuronal migration (Suzuki, Sertie *et al.*, 2002) (Passos-Bueno, Suzuki *et al.*, 2006). Such variability may reflect the complexity and pleitropy of *COL18A1* function. Alterations in collagen XVIII levels have been associated with complex disorders such as hepatocarcinoma(Musso, Theret *et al.*, 2001) (Musso, Rehn *et al.*, 2001) (Hu, Huang *et al.*, 2005) (Armelin-Correa, Lin *et al.*, 2005), obesity (Errera, Canani *et al.*, 2008) among others (Lourenco, Cardoso-Filho *et al.*, 2006) (Iizasa, Chang *et al.*, 2004), and therefore, the understanding of its regulatory control might contribute to new strategies for treatment of these disorders. To date, only the function of the proximal second promoter of *COL18A1* has been examined, and more distant cis–regulatory sequences have not been described (Armelin-Correa, Lin *et al.*, 2005).

Ascribing function to non–coding sequence is a primary challenge arising from the sequencing of the human genome (The ENCODE (ENCyclopedia Of DNA Elements) Project, 2004). Functional non-coding sequences can activate transcription and participate in control of spatiotemporal gene expression and also play a role in diseases(Grice, Rochelle *et al.*, 2005; Antonellis, Huynh *et al.*, 2008). The availability of additional genome sequences has created the opportunity for cross species alignment and the determination, with the help of computational tools, of conserved sequences between genomes (Ahituv, Rubin *et al.*, 2004; Nobrega e Pennacchio, 2004; Poulin, Nobrega *et al.*, 2005; Woolfe, Goodson *et al.*, 2005). The hypothesis that functional elements are under negative selection suggests that highly conserved sequences will likely be functional, although in reality the correlation between degree of conservation and function for cis–regulatory elements is unclear. We have previously shown that moderate conservation, confined to mammalian lineages, is often predictive of function conserved to teleosts, making the zebrafish a feasible

model to test function of many mammalian regulatory sequences (Fisher, Grice, Vinton, Bessling e Mccallion, 2006; Pashos, Kague *et al.*, 2008) (Antonellis, Huynh *et al.*, 2008).

We have examined conserved non–coding sequences (CNSs) in a 232 kb region encompassing *COL18A1* to identify functional cis–regulatory elements. Using a transposon-based transgenic assay in zebrafish, we identified five enhancer elements that control transcription consistent with endogenous *col18a1*, in tissues. Therefore the enhancers drive expression spatiotemporal specific expression in tissues including retina, kidney, blood vessels, gut, cartilage and liver.

Methods

Identification of conserved non-coding sequences

Conserved non-coding sequences were selected *in silico* from intronic and flanking regions of *COL18A1* (100kb upstream from translation start site and between *COL18A1* and *SLC19A1*), a 232 kb genomic sequence corresponding to chr21:45527766-45760000. First we used VISTA Browser (parameters ≥75%, ≥100bp) aligning human and mouse *Col18a1* orthologues. And second for a multi-species alignment, we used PhastCons, from UCSC Browser. (Need to say which species were used for PhastCons). Conserved sequences were named CNS*, where * denotes distance (kb) and relative position (+ or -, 5' or 3', respectively from translation start site of the short isoform); the genome coordinates of sequences selected for analysis are given in Table 1. All CNSs were amplified by PCR, with high fidelity DNA polymerase, from human genomic DNA and subcloned into a Tol2 transposon based vector, pGWcfosEGFP, as previously described (Fisher, Grice, Vinton, Bessling, Urasaki *et al.*, 2006).

Post-hoc computational analysis was performed with MultiPipmaker (Elnitski, Riemer *et al.*, 2005), using default parameters, as outlined in the Results section. We also used the glocal alignment program Shuffle–LAGAN, as implemented through the VISTA web server; for alignments other than human to mouse, the parameters were relaxed to 50% identity over 100 bases.

In situ Hybridization

Zebrafish embryos were collected from wild–type matings at stages from prim-6, or 24 hours post fertilization (hpf) to 5 days post fertilization (dpf) and fixed in 4% paraformaldehyde. Wild-type mouse embryos at 9.5 days post copulation (dpc) were collected and fixed in 4% paraformaldehyde. A partial cDNA clone of zebrafish *col18a1* (plasmid Cb371; GenBank accession # CB923547) was used to amplify a riboprobe template with primers: F- GAATTCCTGGCACTGGATCA and R-TAATACGACTCACTATAGGGAGTGGAAGCACATGAAGTCTGC. The riboprobe template for mouse in situ was generated by PCR amplification from mouse cDNA with primers: F- AGTTCCACATCACCACAGTTCCTAT and R-TAATACGACTCACTATAGGGAGCATGTCTCACAGTAACTCTCCATCA Both digoxygenin labeled probes were synthesized with T7 RNA polymerase Whole-mount situ hybrydizations were performed according to standard protocols (Thisse e Thisse, 2008).

Injections and EGFP expression analyses

Fish were cared for following standard protocols (Kimmel, Ballard *et al.*, 1995) (Westerfield, 2000). Each purified plasmid construct was injected into ≥ 200 zebrafish AB embryos at the 2–cell stage, as previously described (Fisher, Grice, Vinton, Bessling, Urasaki *et al.*, 2006). G0 embryos were screened for mosaic EGFP expression from 24 hpf to 5 dpf using a Zeiss V12 Stereomicroscope, and imaged with

AxioVision 4.5 software. Embryos were raised to screen for germline transmission of the transgenes. Mosaic expression patterns were confirmed by examination of G1 embryos; at least 4 independent lines were examined for each construct.

Results

Col18a1 has a complex pattern of expression in zebrafish embryos

To obtain a detailed picture of *col18a1* expression during zebrafish embryogenesis, we performed whole embryo *in situ* hybridization at stages from 24hpf to 5dpf, with a probe designed to detect all isoforms. In agreement with previous descriptions (Haftek, Morvan-Dubois *et al.*, 2003), we observed *col18a1* expression at 24 hpf in epidermis, pronephric ducts, spinal cord, somite boundaries (myoseptum), otic vesicle, forebrain, and eye primordium (Figure 1A and 1B). At 2 dpf *col18a1* showed a similar pattern, with additional expression seen in anterior notochord, branchial arches, and pectoral fin bud (Figure 1C). In the eyes, *col18a1* expression was primarily confined to retinal pigment epithelium. By 3 dpf expression was also seen in gut, cartilage and blood vessels (Figure 1D, 1E and 1F), and by 4-5 dpf in the liver.

Col18a1 expression in whole mouse embryos

To determine if *Col18a1* expression in mouse is similar to zebrafish, we carried out *in situ* hybridizations on 9.5 day post coitus (dpc) whole mouse embryos. *Col18a1* was expressed in epidermis, otic vesicle, retinal primordium, cranial blood vessels, aorta, forebrain, brain ventricles, and pharyngeal arches (Figure 2). This overall pattern is similar to zebrafish, with expression in retina, otic vesicle and blood vessels observed in both organisms.

Selection of conserved non-coding sequences (CNS) for functional analysis

To select conserved non-coding sequences for analysis, we examined a total of 232 kb, corresponding to intronic sequences of *COL18A1* and 100 kb upstream of the first exon (Figure 3). We identified 12 CNSs based on a pair-wise alignment between human and mouse loci (Vista browser, ≥ 75% identity and ≥ 100 bp) named CNSv, and 8 CNS based on a multiple alignment with orthologues in mammals (PhastCons), named CNSp (Fig. 3; Table 1). Twelve CNSs are localized upstream of *COL18A1* translation start site and 8 CNSs are located in intronic regions. Three VISTA CNSs partially overlapped three CNSs seleted with PhastCons (CNSv -97.8/CNSp -96.9; CNSv +51.7/CNSp +51.7 and CNSv +64.8/CNSp +65.7), as illustrated in Figure 4. Using either algorithm, the non–coding sequences associated with human *COL18A1* failed to show detectable conservation with orthologous regions in teleosts.

CNSs regulate reporter gene expression consistent with zebrafish *col18a1*

To test the regulatory function of the identified CNSs, each sequence was cloned into a vector containing the heterologous minimal *cFos* promoter and *egfp* coding sequence. Each construct was injected in ≥200 zebrafish embryos at the 2–cell stage, which were screened for mosaic EGFP expression from 24 h to 5 dpf. Of the twenty constructs analyzed, five regulated transcription consistent with endogenous *col18a1* expression in zebrafish. Three of the CNSs are localized upstream of the translation start site (CNSv -97.8; CNSp-96.9; CNSv-78.7) and two in intronic regions of *COL18A1* (CNSp +47.8; CNSv+64.8). Fish carrying these constructs were established as transgenic lines and studied in detail in non–mosaic embryos. Each CNS regulated a distinct pattern of tissue–specific expression, as detailed below.

CNSv -97.8/ CNSp -96.9 (Retina and Pronephric Ducts)

CNSv -97.8 and CNSp -96.9 were identified using the Vista and PhastCons algorithms, respectively, and overlapped substantially. Transgenes containing the two

elements led to indistinguishable patterns of *egfp* expression. At 24 hpf, strong expression was seen in retina, forebrain–midbrain junction, otic vesicle, pharyngeal arches and pronephric ducts (Figure 4A-D). Expression in pronephric ducts, retina and forebrain–midbrain junction was transient, and no longer detectable by 2 dpf (Figure 4E and 4F). After 2 dpf, EGFP was expressed in cartilage of the craniofacial skeleton, pectoral fins, and otic capsule.

CNSv -78.7 (Endoderm derivatives)

CNSv -78.7 drove expression in the entire brain and in the pharyngeal arches at 24 hpf (Figure 5A), but at later stages expression in the brain decreased (Figure 5B). Cartilage expression was observed clearly from 2 dpf to 5 dpf (Figure 5B, 5D-F). In addition, expression was observed in gut endoderm at 5 dpf (Figure 5C).

CNSp +47.8 (Blood Vessels and Retinal Pigment Epithelium)

Blood vessels in adult tissues are a prominent site of *COL18A1* expression. A single element, CNSp +47.8, led to expression in blood vessels in the embryo, beginning at 2 dpf in the eyes and pharyngeal arches (Figure 6A). At 3dpf EGFP expression was observed widely in blood vessels, including the dorsal aorta, dorsal ciliary vein, intersegmental vessels, and craniofacial vessels (Figure 6B-E). In addition, CNSp +47.8 drove EGFP expression in cartilage, retinal pigment epithelium, and floor plate (Figure 6C).

CNSv +64.8 (Liver, Spinal Cord and Myoseptum)

At 24 hpf, CNSv +64.8 transgenic embryos showed EGFP expression in the tectum, otic vesicle, pectoral fin bud and pharyngeal arches (Figure 7A). The early expression was transient, and at 2 dpf expression was seen only in cartilage (Figure 7B). At 3 dpf, expression was seen also in spinal cord and myoseptum (Figure 7D and 7F), and by 5 dpf expression was observed in liver (Figure 7E).

Functional differences in overlapping elements

There were three pairs of CNSs identified independently by VISTA and PhastCons, which partially overlapped (Figure 8). As shown above, CNSv -97.8 and CNSp -96.9 (267bp and 1181bp, respectively) regulated indistinguishable patterns of expression (Figure 8A). A second pair of elements, CNSv +51.7 and CNSp +51.7 (338bp and 524bp, respectively) regulated no detectable expression (Figure 8B). And surprisingly a third pair, CNSv +64.8 and CNSp +65.7 (309bp and 1203bp, respectively), showed functional divergence (Figure 8C). Only the smaller construct resulted in reporter gene expression, suggesting that the additional sequence present in CNSp +65.7 contains a negative regulatory element.

***Post hoc* analysis reveals human–zebrafish sequence similarities**

The initial analysis to identify conserved non–coding regions, performed with VISTA and PhastCons, failed to detect any conservation extending from human to teleosts. We subsequently performed alignment of the entire *Col18a1* interval, from human, mouse, chicken, stickleback, and zebrafish, using MultiPipmaker(Schwartz, Elnitski *et al.*, 2003; Elnitski, Riemer *et al.*, 2005). Pipmaker is designed to detect alignments with high sensitivity, but does not calculate significance or likelihood of conservation. Therefore, it is more likely to return false positive alignment results, but in the setting where functional sequences have already been identified, we hypothesized that it might be better able to detect orthologous zebrafish enhancer elements. Indeed, regions of sequence similarity were detected either within or very near several of the tested regions, including sequences adjacent to two of the functional enhancers, CNS -96.9 and +47.8 (data not shown). However, many additional alignments were detected of repeat sequences, likely to be spurious.

We also performed alignments from the above species with the Shuffle-LAGAN algorithm, incorporated as part of the VISTA browser site. Shuffle-LAGAN is among the class of "glocal" alignment programs, integrating features of both global and local applications, and is particularly designed to detect sequence similarities despite small–scale rearrangements such as inversions, translocations, duplications, and deletions. We performed an alignment of intergenic sequences from human, mouse, chicken, stickleback, and zebrafish, centered around exons 2 and 3 and including the positive enhancers CNS +47.8 and +64.8; for alignments more distant than human to mouse, the standard parameters were relaxed to 50% identity over 100 bp. We detected alignments of non–coding sequences extending to teleosts, with stickleback for CNS +64.8 and with zebrafish for CNS +47.8 (Fig. 9A). Alignments of upstream regions failed to detect similarity beyond ~40 kb upstream of the human gene, despite the presence of two positive enhancer sequences farther upstream. However, there is a break in conserved synteny between mammals and teleosts upstream of *COL18A1* (the adjacent annotated genes are *POFUT2* and *cdtsp1* respectively), and associated rearrangements may obscure alignments.

Identification of an orthologous zebrafish enhancer

A portion of the alignment between zebrafish and CNS +47.8 was repetitive sequence, with many close matches in the both the human and zebrafish genomes (Fig. 9A). We amplified the non–repetitive portion of the aligned sequence and evaluated its regulatory function in mosaic transgenic zebrafish. It was a strong positive transcriptional enhancer, with expression at 24 hpf in epidermis and skeletal muscle, frequent sites of spurious expression in mosaic analysis. Interestingly, the human CNS +47.8 similarly displayed a high level of background expression in mosaic G0 embryos (data not shown), although the expression in G1 established lines was tissue–specific

(Fig. 7). We also observed clear patterns of tissue–specific EGFP expression for zebrafish CNS +88. At 24 hpf, 51% (97/190) of injected embryos had heart endothelial cells strongly expressing EGFP (Fig. 9B). In addition, we observed clones of EGFP–expressing floorplate (Fig 9C), cartilage (Fig. 9D, E) and blood vessel endothelial cells (Fig. 9F-I), demonstrating substantial functional overlap with the human CNS+47.8 enhancer.

Discussion

Our analysis of zebrafish *col18a1* demonstrated a complex and dynamic pattern of expression during embryogenesis. In agreement with previous descriptions(Haftek, Morvan-Dubois *et al.*, 2003), we observed expression in pronephric ducts, otic vesicle, epidermis, spinal cord, notochord, forebrain, and the tectum in the first three days. We also observed transient expression in the retina at 24 hpf, although eye expression at later stages was limited to the retinal pigment epithelium. Importantly, at later stages we describe expression in cartilage, blood vessels, gut, and liver, not previously reported in zebrafish but consistent with expression in other species (Halfter, Dong *et al.*, 1998; Fukai, Eklund *et al.*, 2002; Marneros e Olsen, 2003). Although *Col18a1* expression was known for some specific tissues (Muragaki, Timmons *et al.*, 1995; Fukai, Eklund *et al.*, 2002), its overall pattern of expression in mouse embryogenesis had not been described. We show that the expression at 9.5 dpc is largely similar to zebrafish at comparable stages. The substantial conservation of expression pattern across vertebrates argues that the regulatory mechanisms governing *Col18a1* expression are likely to be conserved as well.

We employed transgenesis in the zebrafish to functionally evaluate potential enhancer elements regulating *COL18A1* transcription. Out of twenty sequences tested,

we identified five non–coding elements associated with the human gene, from four independent regions, that regulate transcription consistent with endogenous zebrafish *col18a1*. Together, the expression patterns controlled by these enhancer regions constitute the major components of the expression pattern in the embryo, including expression in retina, retinal pigment epithelium, pronephric ducts, liver, forebrain, gut, blood vessels and cartilage. Thus our screening strategy has proven an efficient approach to enumerate the regulatory regions associated with a complex expression pattern, over a large physical segment of the genome. Importantly, only one of the functional enhancers was identified by both VISTA and PhastCons. Additionally, the two algorithms identified an overlapping pair of sequences, only the shorter one of which was a functional enhancer, which suggests the presence of a negative regulatory element in the longer sequence. These discrepancies highlight the importance of employing multiple approaches in comparative sequence analysis.

We observed substantial overlap in function of the *COL18A1* enhancer elements; most notably, all directed expression to cartilage, although at varying levels and with different timing. The general principle of multiple enhancer elements with overlapping function has been observed previously for other genes with complex expression patterns. In the case of collagen XVIII, the biological significance of expression in cartilage is unclear, since Knobloch Syndrome patients do not have any apparent cartilage phenotype.

The three splicing forms of *COL18A1* show differential expression in adult human tissues. In particular, the short form, transcribed from the upstream promoter, is the predominant form in retina and kidney, while the intermediate and long forms, transcribed from the downstream promoter, predominate in liver. Interestingly, our data reveals transcriptional regulation that mirrors the tissue specific expression

pattern, with CNS -97.8/-96.9 regulating expression in retina and pronephric ducts, CNS +47.8 in retinal pigment epithelium, and CNS +64.8 in liver. This suggests that specific enhancer elements interact differentially with the two *COL18A1* promoters.

We have previously demonstrated the utility of relying on moderate levels of sequence conservation, i.e. among mammals, to identify putative enhancer elements for *in vivo* testing (Fisher, Grice, Vinton, Bessling e Mccallion, 2006; Fisher, Grice, Vinton, Bessling, Urasaki *et al.*, 2006). In this study, although the algorithms used did not detect non–coding conservation at the *COL18A1* locus from human to teleosts, the human sequences functioned appropriately in zebrafish transgenics. We could not unambiguously identify conserved transcription factor binding sites by sequence inspection, likely due to known variations on consensus sequences as well as incomplete databases for binding sites. However, given the conserved function of the human sequences in zebrafish, the most parsimonious explanation is that the same or related transcription factors bind the orthologous elements in the two species. Additional *post hoc* analysis did reveal detectable sequence similarities in several regions between human and zebrafish. However, these computational strategies would be difficult to apply on a genome–wide scale, or even at a single locus, with no prior knowledge of functional enhancers.

Following the *post hoc* analysis, we have functionally tested a zebrafish sequence and demonstrated enhancer activity. We find substantial functional equivalence with the orthologous human enhancer, including expression in blood vessels, a biologically important site of *col18a1* expression. . More generally, we show that combining functional data with targeted analyses for phylogenetic conservation can reveal conserved cis–regulatory elements in the large number of cases where computational alignment alone falls short.

There are settings in which variations in expression levels or patterns of *COL18A1* may be clinically important, such as regulating neovasculogenesis through the endostatin domain, and predisposition for obesity in diabetes type 2 and for hepatocarcinoma (Armelin-Correa, Lin *et al.*, 2005; Errera, Canani *et al.*, 2008). Therefore, identification of the regulatory regions will provide insight into normal and pathogenic regulation of *COL18A1* expression.

Acknowledgments

We would like to thank Andy McCallion and Ron Miller for mouse embryos; Lucia Armelin, Oscar Takeo Suzuki and Cibelle Masotti for initial lab help and scientific discussions; Melissa Biddelecomb for her kindness and help with fish care; and Constancia G. Urbani for secretarial help. This study was supported by Fapesp (Cepid).

References

Abdollahi, A., P. Hahnfeldt*, et al.* Endostatin's antiangiogenic signaling network. Mol Cell, v.13, n.5, Mar 12, p.649-63. 2004.

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.

Ahituv, N., E. M. Rubin*, et al.* Exploiting human--fish genome comparisons for deciphering gene regulation. Hum Mol Genet, v.13 Spec No 2, Oct 1, p.R261-6. 2004.

Antonellis, A., J. L. Huynh*, et al.* Identification of neural crest and glial enhancers at the mouse Sox10 locus through transgenesis in zebrafish. PLoS Genet, v.4, n.9, Sep, p.e1000174. 2008.

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.

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.

Elnitski, L., C. Riemer*, et al.* MultiPipMaker: comparative alignment server for multiple DNA sequences. Curr Protoc Bioinformatics, v.Chapter 10, Apr, p.Unit10 4. 2005.

Errera, F. I., L. H. Canani*, et al.* COL18A1 is highly expressed during human adipocyte differentiation and the SNP c.1136C > T in its "frizzled" motif is associated with obesity in diabetes type 2 patients. An Acad Bras Cienc, v.80, n.1, Mar, p.167-77. 2008.

Fisher, S., E. A. Grice*, et al.* Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science, v.312, n.5771, Apr 14, p.276-9. 2006.

_____. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat Protoc, v.1, n.3, p.1297-305. 2006.

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.

Gelse, K., E. Poschl*, et al.* Collagens--structure, function, and biosynthesis. Adv Drug Deliv Rev, v.55, n.12, Nov 28, p.1531-46. 2003.

Grice, E. A., E. S. Rochelle*, et al.* Evaluation of the RET regulatory landscape reveals the biological relevance of a HSCR-implicated enhancer. Hum Mol Genet, v.14, n.24, Dec 15, p.3837-45. 2005.

Haftek, Z., G. Morvan-Dubois*, et al.* Sequence and embryonic expression of collagen XVIII NC1 domain (endostatin) in the zebrafish. Gene Expr Patterns, v.3, n.3, Jun, p.351-4. 2003.

Halfter, W., S. Dong*, et al.* Collagen XVIII is a basement membrane heparan sulfate proteoglycan. J Biol Chem, v.273, n.39, Sep 25, p.25404-12. 1998.

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.

Hurskainen, M., L. Eklund*, et al.* Abnormal maturation of the retinal vasculature in type XVIII collagen/endostatin deficient mice and changes in retinal glial cells due to lack of collagen types XV and XVIII. FASEB J, v.19, n.11, Sep, p.1564-6. 2005.

Iizasa, T., H. Chang*, et al.* Overexpression of collagen XVIII is associated with poor outcome and elevated levels of circulating serum endostatin in non-small cell lung cancer. Clin Cancer Res, v.10, n.16, Aug 15, p.5361-6. 2004.

Kimmel, C. B., W. W. Ballard*, et al.* Stages of embryonic development of the zebrafish. Dev Dyn, v.203, n.3, Jul, p.253-310. 1995.

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.

Marneros, A. G. e B. R. Olsen. The role of collagen-derived proteolytic fragments in angiogenesis. Matrix Biol, v.20, n.5-6, Sep, p.337-45. 2001.

_____. Age-dependent iris abnormalities in collagen XVIII/endostatin deficient mice with similarities to human pigment dispersion syndrome. Invest Ophthalmol Vis Sci, v.44, n.6, Jun, p.2367-72. 2003.

_____. Physiological role of collagen XVIII and endostatin. FASEB J, v.19, n.7, May, p.716-28. 2005.

Muragaki, Y., S. Timmons*, et al.* Mouse Col18a1 is expressed in a tissue-specific manner as three alternative variants and is localized in basement membrane zones. Proc Natl Acad Sci U S A, v.92, n.19, Sep 12, p.8763-7. 1995.

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. 2001.

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. 2001.

Nobrega, M. A. e L. A. Pennacchio. Comparative genomic analysis as a tool for biological discovery. J Physiol, v.554, n.Pt 1, Jan 1, p.31-9. 2004.

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.

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. 1994.

Pashos, E. E., E. Kague*, et al.* Evaluation of cis-regulatory function in zebrafish. Brief Funct Genomic Proteomic, Sep 27. 2008.

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.

Poulin, F., M. A. Nobrega*, et al.* In vivo characterization of a vertebrate ultraconserved enhancer. Genomics, v.85, n.6, Jun, p.774-81. 2005.

Quelard, D., E. Lavergne*, et al.* A cryptic frizzled module in cell surface collagen 18 inhibits Wnt/beta-catenin signaling. PLoS ONE, v.3, n.4, p.e1878. 2008.

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.

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. 1998.

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. 1998.

Schneider, V. A. e M. Granato. The myotomal diwanka (lh3) glycosyltransferase and type XVIII collagen are critical for motor growth cone migration. Neuron, v.50, n.5, Jun 1, p.683-95. 2006.

Schwartz, S., L. Elnitski*, et al.* MultiPipMaker and supporting tools: Alignments and analysis of multiple genomic DNA sequences. Nucleic Acids Res, v.31, n.13, Jul 1, p.3518-24. 2003.

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.

The ENCODE (ENCyclopedia Of DNA Elements) Project. Science, v.306, n.5696, Oct 22, p.636-40. 2004.

Thisse, C. e B. Thisse. High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat Protoc, v.3, n.1, p.59-69. 2008.

Westerfield, M. The Zebrafish Book. A guide for the laboratory use of zebrafish (Danio rerio). Eugene: University of Oregon Press. 2000

Woolfe, A., M. Goodson*, et al.* Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol, v.3, n.1, Jan, p.e7. 2005.

Tables and Figures Legends

Table I. CNSs coordinates in chromosome 21and their respective size.

Figure 1. Zebrafish *col18a1* expression detected by whole embryo *in situ* hybridization at 24 hpf (A, B); 2 dpf (C); and 3 dpf (D-F). Dynamic expression was seen in neural structures including forebrain and retina (A), spinal cord (B) and tectum (C). Expression was also seen in pronephric duct and pronephros (B, F); myospetum (B); epidermis (C); otic vesicle (C); pharyngeal arches (C); craniofacial cartilages and blood vessels (D); retinal pigment epithelium (E); pectoral fins (F); and gut endothelium (F). Abbreviations used in all figures: BV (blood vessels), Ca (cartilage), FB (forebrain), Gu (gut endothelium), My (myoseptum), OV (otic vesicle), PA (pharyngeal arches), PD (Pronephric Duct), PF (pectoral fin), Pr (pronephros), Re (retina), RPE (retinal pigment epithelium), SC (spinal cord), Te (tectum).

Figure 2. Conserved pattern of *Col18a1* expression in mouse embryogenesis. *In* situ hybridization for *Col18a1* was performed in mouse embryos at 11.5 dpc. Expression was observed in forebrain, pharyngeal arches, and spinal cord (A), and in retina, blood vessels, otic vesicle, and aorta (B, C).

Figure 3. Interval containing *COL18A1* represented in the UCSC Genome Browser. CNSv (red bars) and CNSp (blue bars) sequences are shown in custom tracks. Conservation peaks from Vista Browser are illustrated at the top, and PhastCons Conserved elements shown at the bottom.

Figure 4. Conserved noncoding sequences regulate EGFP expression in zebrafish embryos consistent with endogenous *col18a1*. Embryos transgenic for either CNSv -97.8 or CNSp -96.9 had EGFP expression at 24 hpf in forebrain, retina, tectum, otic vesicle, and pronephric duct (A-D). At 3 dpf, EGFP was expressed in cartilage of the head, pectoral fins, and otic vesicle (E, F).

Figure 5. CNSv -78.7 regulated expression at 24hpf embryos in the brain and pharyngeal arches (A). Expression was seen in craniofacial, otic capsule, and pectoral fin cartilages (B, D-F) at 3dpf (B, D), 4dpf (E) and 5dpf (F). GFP expression was also seen in gut endothelium at 5 dpf (C).

Figure 6. CNSp +47.8 regulated GFP expression in blood vessels at 24hpf and also clearly in BV of pharyngeal arches at 3dpf (B). RPE expression (C) and Ao expression is brightly observed at 3dpf (D and E).

Figure 7. CNSv +64.8 regulated GFP expression at 24hpf in tectum, otic vesicle, pharyngeal arches (A), and spinal cord (B), and at 2 dpf in myoseptum (C). Expression was also seen in craniofacial and and pectoral fin cartilage at 2 (D) and 3 (E) dpf, and in liver at 5 dpf (F).

Figure 8. Sequence length discrepancy between CNSv and CNSp sequences can show difference in driving EGFP expression. In A, both constructs drove the same EGFP expression, CNSv -97.8 and CNSp -96.9. However in B, only the shorter construct functioned CNSv +64.8.

Figure 9.

Table I.

Construcs	**Coordinates (Ch 21)**	**Size (bp)**
CNSp - 96.9	45551381- 45552562	1181
CNSv - 97.8	45551403- 45551670	267
CNSp - 80.5	45568753- 45568986	233
CNSv - 78.7	45570340- 45570794	454
CNSp - 72.2	45575491- 45577292	1801
CNSv - 64.9	45584370- 45584612	242
CNSv - 52.9	45596190- 45596573	383
CNSp - 47.4	45601821- 45602101	280
CNSv - 40.8	45608418- 45608709	291
CNSv - 38.0	45611223- 45611488	265
CNSp - 28.7	45620217- 45620766	549
CNSv - 26.7	45622453- 45622740	287
CNSv +0.6	45650150- 45650576	427
CNSv + 3.1	45652238- 45652652	414
CNSv + 25.5	45674812- 45675078	266
CNSp + 47.8	45694764- 45697397	2633
CNSp + 51.7	45700781- 45701305	524
CNSv + 51.7	45700938- 45701276	338
CNSv + 64.8	45714075- 45714384	309
CNSp + 65.7	45714053- 45715256	1203

Figure 1.

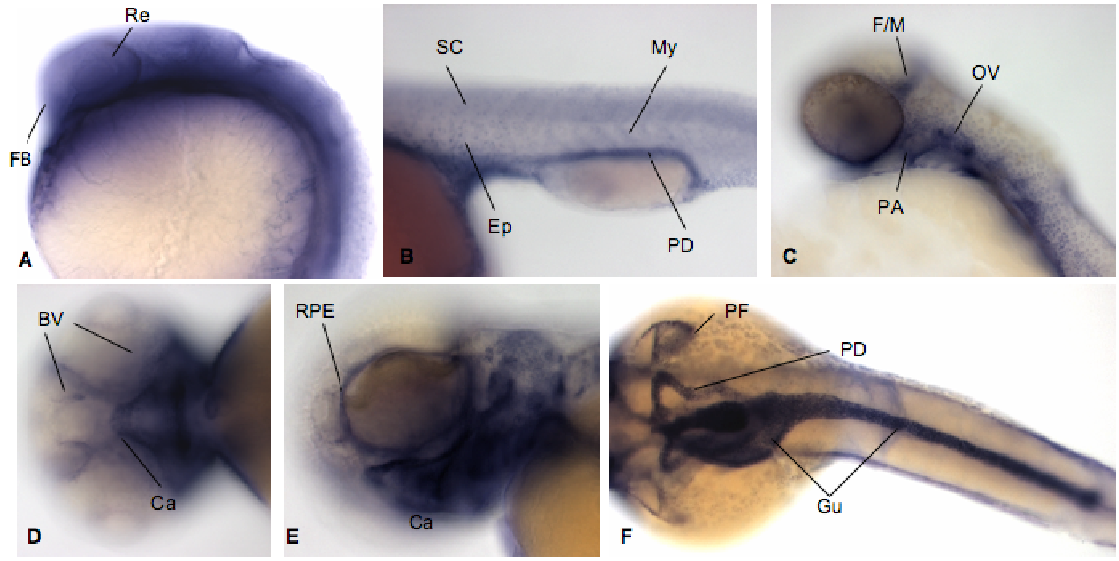


Figure 2.

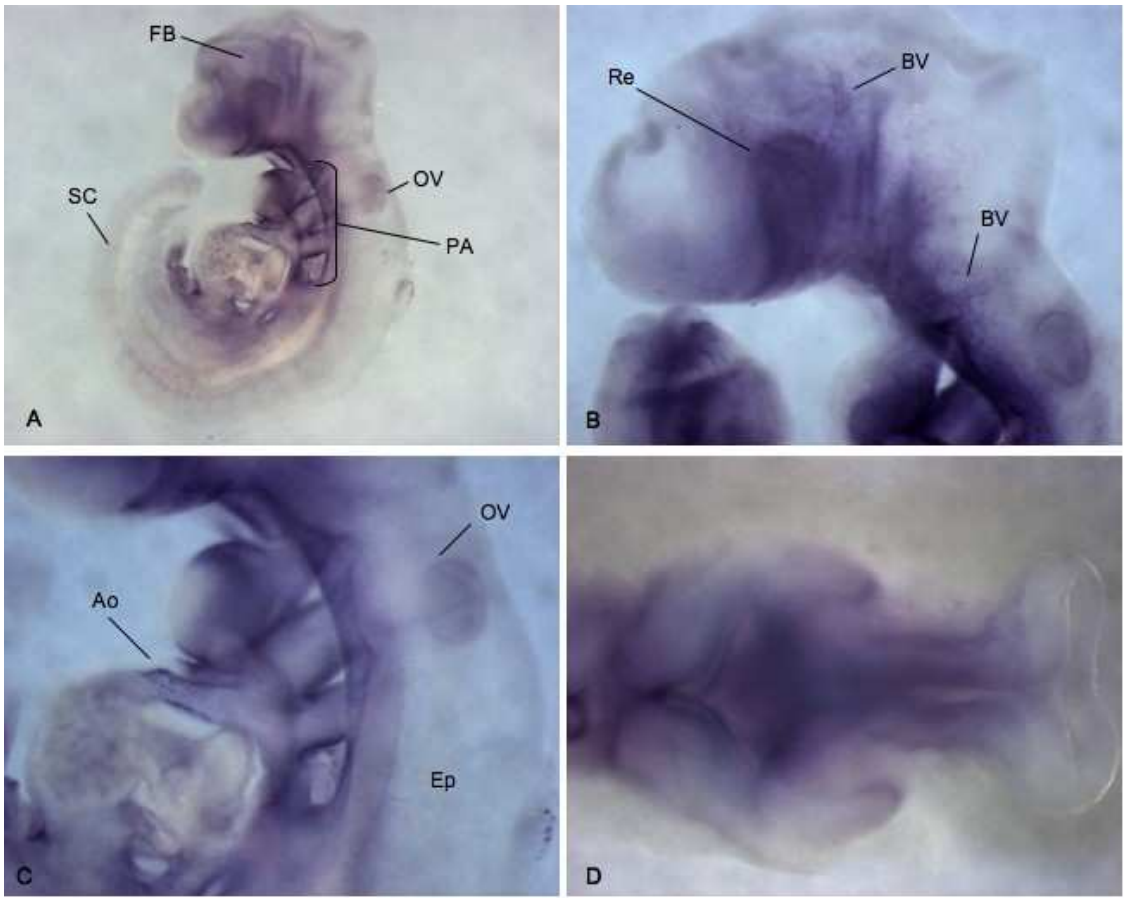


Figure 3

.

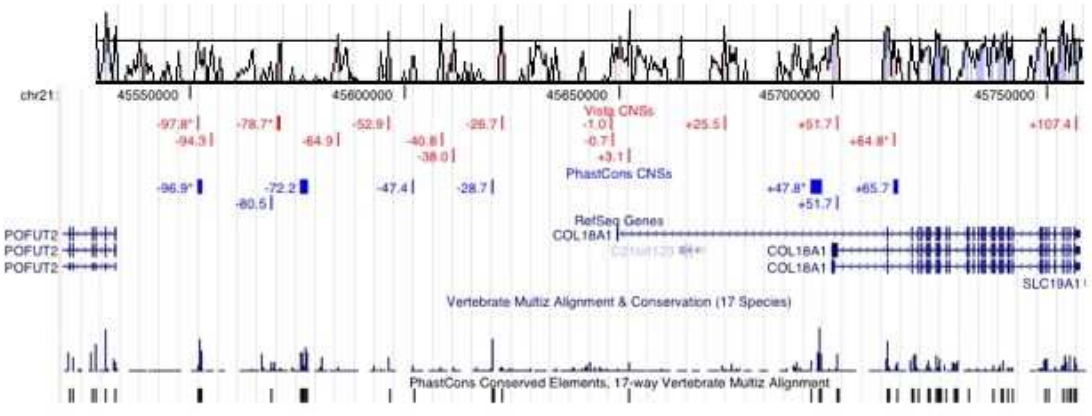


Figure 4.

Figure 5.

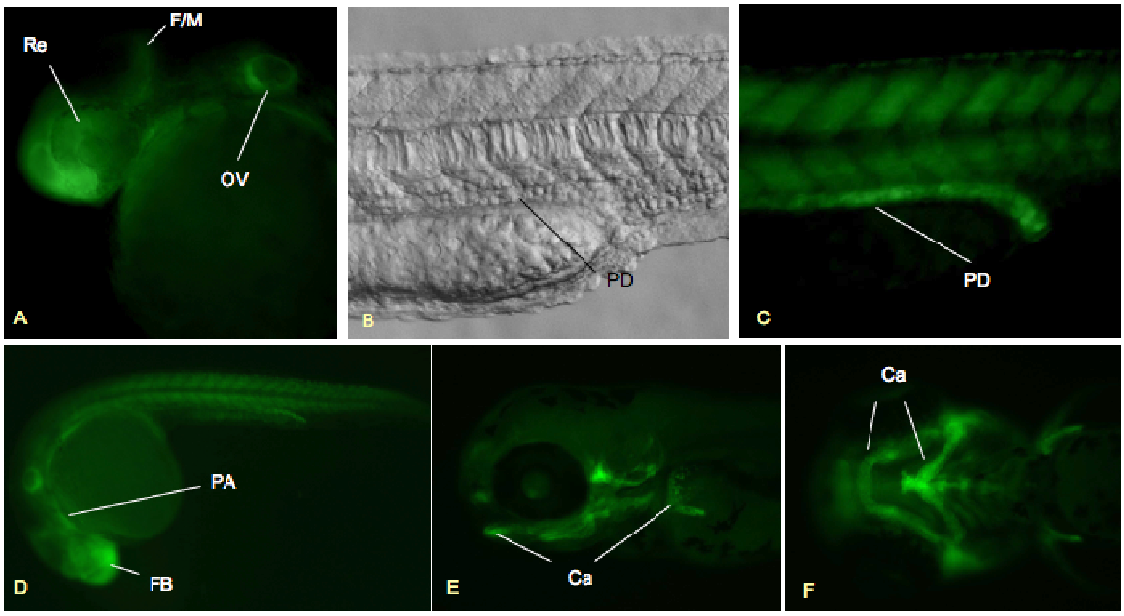


Figure 6.

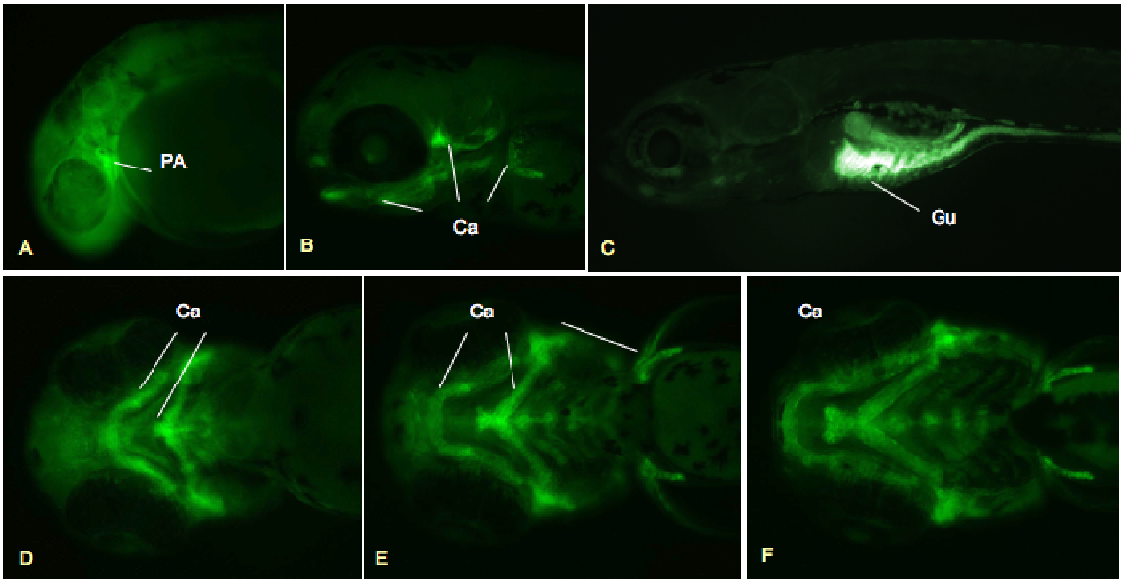


Figure 7.

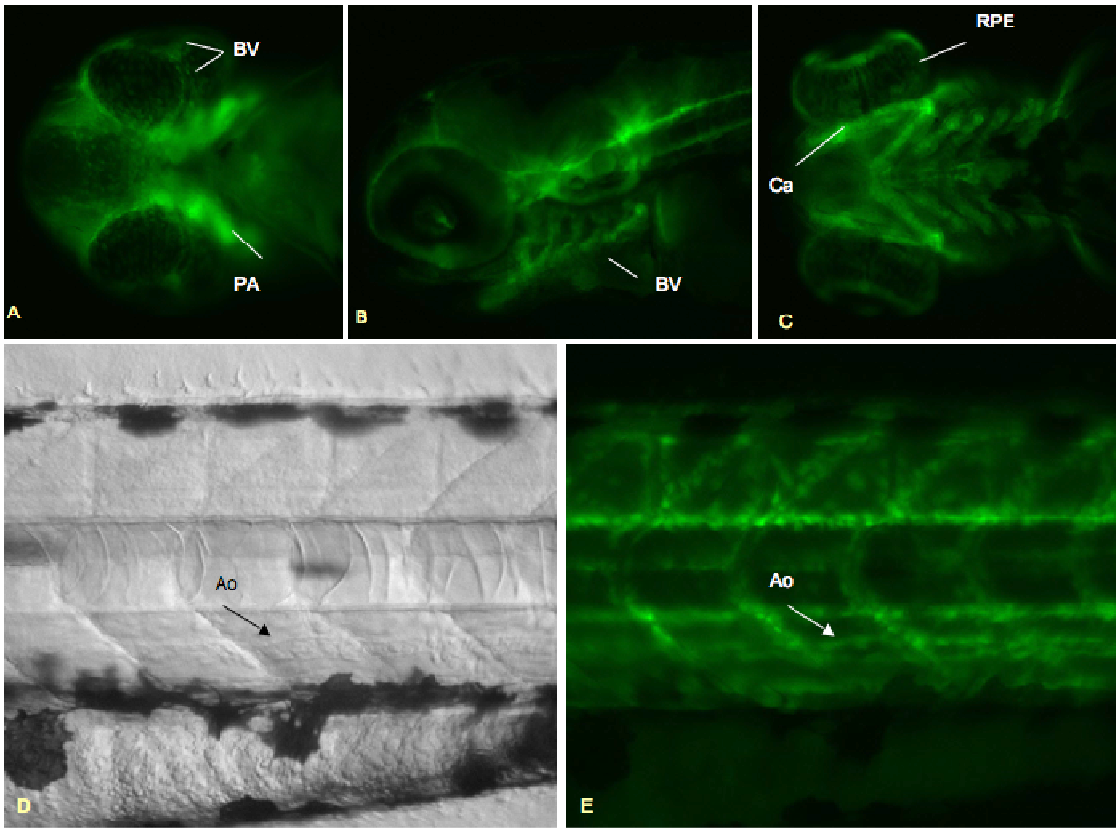


Figure 8.

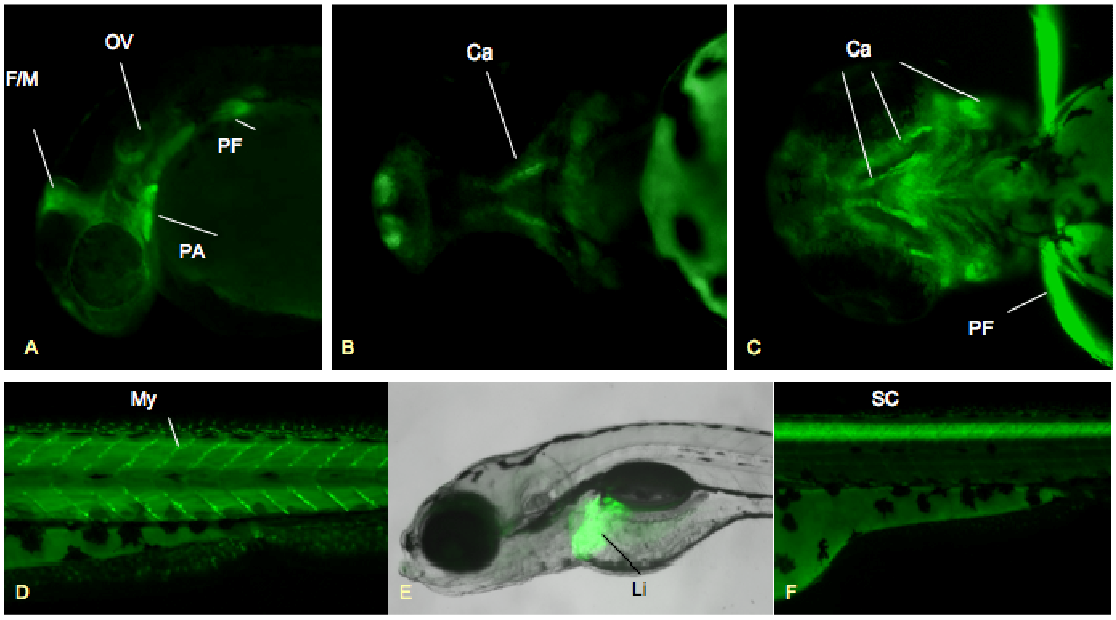
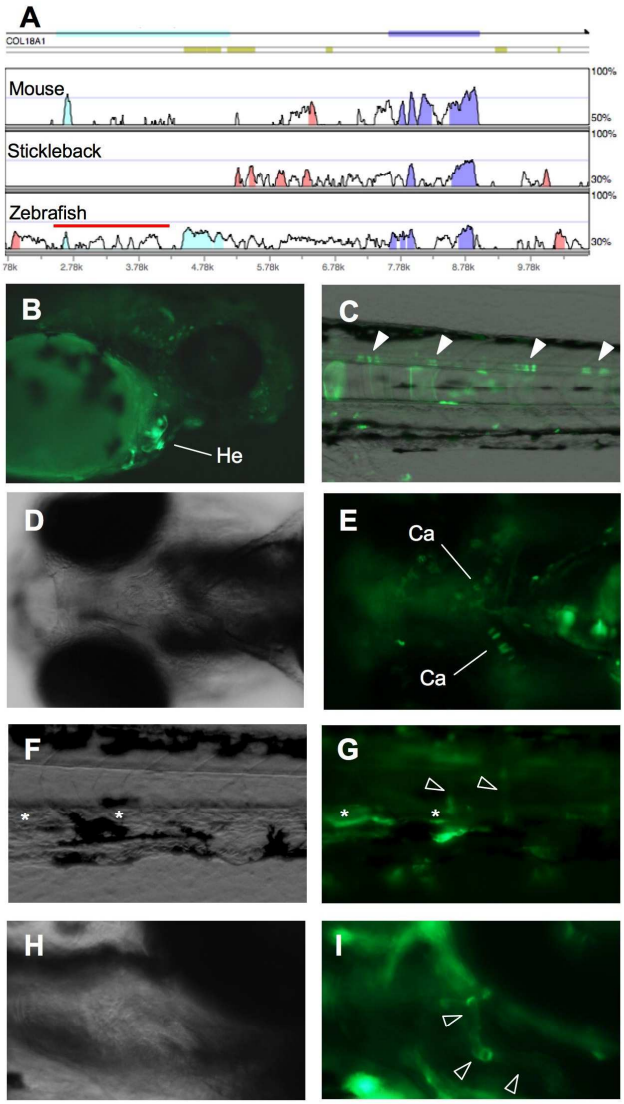


Figure 9.


A
COL18A1
Mouse
Stickleback
Zebrafish
100%
50%
100%
30%
100%
30%
78k
2.78k
3.78k
4.78k
5.78k
6.78k
7.78k
8.78k
9.78k
B
He
C
D
E
Ca
Ca
F
G
H
I

Capítulo IV

4. Dissection of *COL18A1* promoter 1 and control of wnt signaling through regulation of *COL18A1* expression

Erika Kague[1]; Shannon Fisher[2]; Angela Maria Costa[3]; Andrea Sertie[1], Gui Hu[2] and Maria Rita Passos-Bueno[1†].

1- Centro de Estudos do Genoma Humano, Instituto de Biociências, Universidade de São Paulo, Brazil

2- Department of Cell and Developmental Biology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104.

3- Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal.

†- Correspondence should be addressed to M.R.P.B. Current address: Rua do Matão 277 sala 200, Depto. Genética e Biologia Evolutiva, Instituto de Biociências, Cidade Universitaria, São Paulo, Brazil, CEP: 05508-090; Phone (55) 11 30919910; Fax (55) 11 30917419; passos@ib.usp.br.

Capítulo IV- Dissection of *COL18A1* promoter 1 and control of wnt signaling through regulation of *COL18A1* expression

Abstract

Collagen XVIII is a heparan sulfate proteoglycan and a component of basement membranes. *COL18A1* presents 43 exons and transcribes three different isoforms with the use of two promoters. Each isoform has its own pattern of expression and their regulation is not yet well understood. Characterization of promoter 2 that transcribes intermediate and long isoforms has been previously characterized in human and mouse; however none is know about promoter 1. Null mutations in *COL18A1* leads to Knobloch Syndrome, an autosomal recessive disorder characterized for myopia, retina detachment and occipital encephalocele. Collagen XVIII generates two cleaved fragments, endostatin and FRZC18 derived from its C and N-terminal, respectively. Endostatin function is mostly associated as an angiogenic inhibitor, however endostatin and FRZC18 have been both investigated to be involved as antagonists of Wnt signaling as well. The effects of ß-catenin in regulation of *COL18A1* is not know. Here, we dissect *COL18A1* promoter 1 and identified 3 promoter enhancer elements. We attested ß-catenin and TCF/LEF transcription factors as regulators of *COL18A1* expression in an inverse manner. Our results may tray to explain the opposite expression profile observed in processes such adipogenesis, where decrease of Wnt signaling is accompanied with increase of *COL18A1* expression.

Keywords: *COL18A1,* promoter, ß-catenin, TCF/LEF, adipogenesis.

Capítulo IV- Dissection of *COL18A1* promoter 1 and control of wnt signaling through regulation of *COL18A1* expression

Introduction

Collagens constitute a large group of extracellular matrix proteins that preserve triple helical sequence of repeated Gly-X-Y motifs in their primary structure (Prockop e Kivirikko, 1995). Types XVIII and XV belong to the multiplexin subgroup of non–fibrillar collagens, however type XV is a chondroitin sulfate proteoglycan and type XVIII a heparan sulfate proteoglycan (Oh, Kamagata *et al.*, 1994; Oh, Warman *et al.*, 1994; Saarela, Ylikarppa *et al.*, 1998). *COL18A1* was mapped to ch21q22.3 and encompasses 108 Kb that accommodates 43 exons, but the distance between the second and the third exon comprises almost half of the entire gene length (Saarela, Ylikarppa *et al.*, 1998). Three different isoforms are transcribed from *COL18A1* by the use of two promoters and an alternative splicing in its third exon. The three encoded collagen XVIII proteins differ only by their signal peptides and by part of the N-terminal region of NC11 domain. Short isoform (NC11-303) is transcribed from the first promoter localized upstream of exon 1, and it is produced by exons 1, 2, 4-43. Intermediate (NC11-493) and long (NC11-728) forms are transcribed from the second promoter, upstream of exon 3, and they differ by an internal splice site within exon 3 (Elamaa, Peterson *et al.*, 2002). *COL18A1* is expressed in basement membranes of a variety of tissues, but the isoforms have expression tissue-specificity, that is assured by the two promoters and different enhancer elements. The short isoform is highly expressed in kidney, retina, placenta and human fetal brain, whereas the intermediate and long isoforms have their highest expression levels in liver and lung, respectively (Rehn, Hintikka *et al.*, 1994; Halfter, Dong *et al.*, 1998; Saarela, Rehn *et al.*, 1998; Saarela, Ylikarppa *et al.*, 1998). Human

and mouse promoter 2 of collagen XVIII gene had been previously studied (Kvist, Latvanlehto *et al.*, 1999; Armelin-Correa, Lin *et al.*, 2005), nevertheless promoter 1 dissection had not yet been performed.

Collagen XVIII has been investigated in a variety of species, and its expression showed up early during embryogenesis in a complex pattern (Halfter, Dong *et al.*, 1998; Saarela, Ylikarppa *et al.*, 1998; Haftek, Morvan-Dubois *et al.*, 2003; Utriainen, Sormunen *et al.*, 2004). Homozygous null mutations in *COL18A1* leads to Knobloch syndrome, characterized by high myopia, retinal detachment, macular degeneration and occipital encefalocele (Czeizel, Goblyos *et al.*, 1992; Passos-Bueno, Marie *et al.*, 1994; Suzuki, Sertie *et al.*, 2002). Some patients also present other relevant phenotypes such as neuronal abnormal migration and recurrent epilepsy and kidney malformation (Passos-Bueno, Suzuki *et al.*, 2006). *COL18A1* importance in embryo development is not yet completely understood. *COL18A1* have most recently been associated with some complex common disorders, as hepatocarcinoma and obesity(Musso, Rehn *et al.*, 2001; Musso, Theret *et al.*, 2001; Armelin-Correa, Lin *et al.*, 2005; Errera, Canani *et al.*, 2008; Quelard, Lavergne *et al.*, 2008). These imply that this collagen XVIII might have an important functional effect not only during embryonic development but also in adulthood.

Two major conserved domains had been demonstrated to be cleaved products from collagen XVIII. Endostatin, a 40 KDa proteolytic from C-terminal domain, has its function implied in angiogenesis inhibition and it has been largely related with many pathogeneses (O'reilly, Boehm *et al.*, 1997). It was described in xenopus embryos the potential of endostatin to inhibit ß-catenin stimulating its degradation of both wild type and stabilized forms (Hanai, Gloy *et al.*, 2002). The Frizzeld domain (FRZC18) is a N-terminal fragment derived from only long isoform of collagen XVIII

and corresponds to a domain of 10 invariable spaced cysteines (cysteine rich domain-CRD) (Elamaa, Snellman *et al.*, 2003). The FRZC18 was showed to inhibit ß-catenin (Quelard, Lavergne *et al.*, 2008) and we recently described its role in Wnt non-canonical signaling as well. Therefore, both cleaved fragments have been somehow related to antagonize the Wnt signaling pathway with wnt signaling.

Wnt family encompasses glycoproteins conserved signaling molecules involved in a surfeit of developmental and cell biological processes, including cell proliferation, differentiation and polarity (Logan e Nusse, 2004). In Wnt/ ß-catenin signaling Wnts interact with Frizzled receptors (Fz) and with coreceptors such as low-density lipoprotein receptor-related protein-5 and-6 (LRP5 and 6) and promotes ß-catenin stabilization in the cytoplasm, with subsequent activation of TCF/LEF leading to expression of target genes (Logan e Nusse, 2004). In the absence of Wnt signaling, a cytoplasmic degradation complex, consisting of at least axin, adenomatous polyposis coli (APC) protein, glycogen syntase kinase 3 (GSK-3) leads to ß-catenin ubiquitination and its degradation by the proteosome. Components of Wnt signaling have been related with inhibition of adipogenesis, however not much is understood in this process. C-myc, c-jun, Cyclin D, VEGF, FGF4, FGF18, E-cadherin, TCF, LEF, Dfz3, Frizzled 7, Dickkopf, Axin2 (He, Sparks *et al.*, 1998; Mann, Gelos *et al.*, 1999; Zhang, Gaspard *et al.*, 2001; Quelard, Lavergne *et al.*, 2008) are examples of target genes. Wnt signaling can promote expression of several Wnt pathway components, indicating that feedback control is a key feature of Wnt signaling regulation.

Thus, we attempted to dissect the promoter 1 of *COL18A1* and analyze if wnt/ß-catenin pathway could regulate *COL18A1* expression since it has endostatin and FRZC18 fragments as its antagonists. *COL18A1* promoter 1 showed characteristics reminiscent of housekeeping and developmental genes and similarities

with *COL18A1* promoter 2. Both promoters of *COL18A1* are regulated by ß-catenin and TCF/LEF and its expression is decreased with ß-catenin overexpression. *COL18A1* was also showed to have an increase of its expression at 1 day of adipogenic differenciation. It highlights a Wnt signaling as a pathway that *COL18A1* could function and play a role during embryogenesis and in adipogenic differentiation process.

Methods

Deletion constructs for promoter analysis

We used a high fidelity DNA polymerase enzyme (Pfx, invitrogen) to amplify nine distinct regions upstream of the translation start site of exon 1 of *COL18A1*. All the fragments had a single reverse primer, as illustrated in table 1, and they were cloned using the Gateway system (invitrogen) and a pGL3-basic modified with attb sites as the destination vector (pGL3-basic-GW). The longer fragment was cloned also in inverted position.

Cell Culture and transfections

Hek293T, HepG2, Hela, Hfob and SHSY cell lines were cultured as recommended by American Type Culture Collection (ATCC).

For promoter study we selected two cell lines (Hek293T and HepG2) that showed highest *COL18A1* expression among five lines studied (Hek293T, HepG2, Hela, Hfob and SHSY). Expression analyses were carried out with Real-Time PCR. All transfections were carried out using Lipofectamine 2000 (Invitrogen) as recommended. For promoter 1 deletion analyzes we used Hek293T and HepG2 cell lines, what were transferred to 12-wells plates (1×10^5 cells/well) 24h before transfection. We cotransfected 1 µg of pGL3-Basic-GW + promoter fragment and 0.25 µg of pRL-TK (promega). After 48h of transfection we performed cell lyses followed by analysis with Dual-Luciferase Reporter Assay System (Promega) in Luminometer (LB96V Microlumat Plus- EG&Berthold). Each transfection was performed in duplicate and three independent experiments. Luciferase activity was normalized with the empty vector. Transfections of pGL3-basic P1-103, P1-540, P1-

1036, P2-182, P2-555 and P2-1052 (500ng each) in Hek293T wild cells and in Hek293T that constitutively expressed ß-catenin were proceeded independently and in triplicates. Luciferase activity was normalized with ß-galactosidase (ßgal) (100ng). We used OT-Luc (carrying TCF/LEF binding sites) (500ng) and ØF-Luc (carrying TCF/LEF mutated binding sites) (500ng) as positive and negative controls, respectively. The data was analyzed with "One-Way ANOVA" with multiple Tukey's comparisons in GraphPad Prism (4.00). We used $P<0.05$. Student's T-test was used for statistical analyses.

Distinguished plates of Hek293T cell lines were transfected with 24 µg of vectors expressing *WNT1a* and *ß-catenin*. RNA extraction was carried out 48h post-transfection with MN RNA purification kit (MN).

Real-Time PCR

Total RNA was extracted from cells lines cultures using Trizol (Gibco) or RNa extraction kit (MN), as standard protocols. 1µg of RNA was used to synthesis of cDNA with SuperScript III (Invitrogen). Real Time PCR reaction was carried out with SYBR Green, and placed in 7500 Real Time PCR System. The data was analyzed with 7500 System SDS software v1.2 (Applied Biosystems). The PCR conditions were carried out as followed : 50°C for 2 minutes, 95°C for 10 minutes and 45 cicles of 95°C for 15 seconds and 60° for 1 minute. We used a total volume of 25 µL.

HPRT1 (hypoxanthine phosphoribosyltransferase 1), *SDHA* (succinate dehydrogenase complex, subunit A, flavoprotein), *GAPDH* (Glyceraldehyde 3-phosphate dehydrogenase) and *HMBS* (Hidroxymethylbilane sinthase) were used as

endogenous controls. Target genes were *COL18A1* and *cMyc*. For adipogenesis study we also included *CEBPa, CEBPb* and *PPRAgamma*. Primers are listed in table I. Analyses were performed with ΔΔCT method and normalization based on Genorm (Vandesompele et al., 2002) normalization calculated values in each experiment.

Chromatin Imunoprecipitation (CHIP)

We performed CHIP using EZ ChiP ^{tm} (upstate) and TCF-1 and LEF-1 antibodies (Santa Cruz) as recommended. Chromatin fragmentation was standardizing with sonication. As positive and negative control we used imunoprecipitation with anti-RNA Polimerase and mouse anti-IgG, respectively. The input sample was prepared as suggested. PCRs with primers for both promoters of *COL18A1* were used for final detection of DNA fragment precipitation. To promoter 1 and 2 amplification we used regions P1-540 and P1-752 and P2-427 (Armelin-Correa, Lin *et al.*, 2005), respectively.

In silico analyzes

We used PROSCAN (http://www-bimas.cit.nih.gov/molbio/proscan/) to predict the *COL18A1* promoter. PE1-3 isolated sequences and 1500 pb upstream of transcription start site of mouse *col18a1* short isoform were submitted as query in MEME (http://meme.sdsc.edu/meme4_1/cgi-bin/meme.cgi) for identification of motifs. Prediction of TF binding sites was determined with TESS (http://www.cbil.upenn.edu/cgi-bin/tess). To identify conserved TF between two orthologous sequences we used CONSITE (http://asp.ii.uib.no:8090/cgi-bin/CONSITE/consite/). Ingenuity was used to map possible link between *COL18A1* and TCF/LEF transcription factors.

Construct, injections and EGFP expression analyses

PE3 region was amplified by PCR (PE3 F-AGACACACAGACACACCCACTGCACTGC and PE3 R-GCCTCAGCCTTCCCAGGCCCAAT) with high fidelity DNA polymerase, from human genomic DNA and subcloned into a Tol2 transposon based vector, pGWcfosEGFP, as previously described (Fisher, Grice, Vinton, Bessling, Urasaki *et al.*, 2006).

Fish were cared for following standard protocols (Kimmel, Ballard *et al.*, 1995) (Westerfield, 2000). Purified plasmid construct was injected into ≥ 200 zebrafish AB embryos at the 2–cell stage, as previously described (Fisher, Grice, Vinton, Bessling, Urasaki *et al.*, 2006). G0 embryos were screened for mosaic EGFP expression from 24 hpf to 5 dpf (days post fertilization) using a Zeiss V12 Stereomicroscope, and imaged with AxioVision 4.5 software.

In vitro adipogenic differentiation

Adipose tissue-derived mesenchymal stem cells were obtained from lipoaspirates of abdominal subcutaneous fat. The cells were isolated and cultured as described previously (Zuk et al., 2002; Costa et al., 2007; Bueno, 2007). Their differentiation into adipocytes was induced by treatment of semi-confluent cells for 14 days with DMEM-HG supplemented with 10% FBS (Hyclone), 1μM dexamethasone, 100 M indomethacin, 500 μM 3-isobutyl-1-methylxanthine (IBMX), and 10 μg/ml insulin as standard protocol (ZUK et al., 2002). To assess neutral lipid content at day 14 after exposure to inducers of differentiation, mature adipocytes were washed with PBS, fixed with 4% paraformaldehyde and stained with Oil Red O. A

successful differentiation process was considered when it was observed that the majority of cells accumulated lipid vesicles in the cytoplasm. The use of human lipoaspirates to establish stem cell populations was accordingly to the guidelines of our local Ethical Committee. RNA extraction was conducted with trizol (Gibco) reagent at time 0 (no adipogenic induction medium), 12 hours, 1, 3, 7 and 14 days of the differentiation process.

Results

Analysis of *COL18A1* promoter 1

To study the activity of the *COL18A1* promoter 1 we built ten 5'-deletion-promoter-luciferase reporter gene constructs, wherein the putative region was fused upstream of luciferase reporter gene (pGL3-basic-GW). Length of the fragments varied from 1700 to 103pb (Figure 1A). The total region length was selected with the caveat of including the PROSCAN predicted promoter localized at 600pb 5' from the translation start site.

We performed transient transfections with each of 10 constructs in Hek293T and in HepG2. In terms of luciferase activity the minimal fragment analyzed capable to lead to expression is situated between -103 and -273pb (core). The expression level was increased significantly (over fourfold) with the constructs containing -540 and-752bp, -919 and -1036bp in both cell lines, corresponding respectively to PE1 and PE2 (Figure 1A, B and C). However construct containing -1235 and -1554bp (named PE3) showed fourfold luciferase activity increase in HepG2 cell but not in Hek293 (Figure 1B and C). The results suggest a minimal core promoter region extending from 103 to 273 and 3 proximal enhancer elements (PE1-3). Decrease of luciferase activity implies presence of silencer-like elements in such regions. Instead of TATA

and CAAT boxes, the promoter 1 of *COL18A1* has a high G + C content, moreover it is situated in a CpG island, with 23.1% of CpG and 66% of G or C. The CpG island extends for 1703pb including the promoter, exon 1 and 2 and part of intron 2 of *COL18A1* (UCSC browser).

In silico analyzes provide insights for TF binding sites in *COL18A1* promoter 1

Trying to delimitate important sequences or transcription factors that could play a role in regulation of identified promoter 1 regions (core, PE1-3), we used MEME to find constrained motifs sequences present in both promoter 1 regions (PE1-3) and the mouse correspondent orthologous sequence of 1500pb upstream of *col18a1* exon1. We identified two, two and three constrained motifs in PE1, PE2 and PE3, respectively. In only PE3 the three constrained motifs identified had the same sequential order in human and mouse sequences. Prediction of TF binding sites with TESS program is showed in Table IIA. PE3 showed TCF/LEF binding sites, however it was not predicted in any of the motifs sequences of PE1 and 2. We also compared PEs and core promoter 1 of *COL18A1* with the same mouse orthologous sequence using CONSITE. It identifies conserved TFs between two sequences. The transcription factors shared between human and mouse in each the regions are shown in Table IIB. TCF/LEF is also predicted just in PE3 region. And fourth, we aligned the core promoter 1 with core promoter 2 using CONSITE and they showed to share MZF, Pax5 (Bsap), Sp1 and c-ETS TFs. Analysis of 1700pb sequences relative to promoter 1 study, TCF/LEF TFs are predicted to bind .

PE3 showed enhancer activity *in vivo*

We functionally tested PE3 region for enhancer activity using zebrafish embryos and analyzing them for mosaic EGFP expression. PE3 leaded to mosaic

EGFP expression in notochord and forebrain at 2dpf embryos (Figure 2), showing a positive activity of this region and its potential as *COL18A1* proximal enhancer element. Liver expression was not detected.

Evaluation of TCF1/LEF1 binding to *COL18A1* promoter

We showed interaction between TCF/ LEF and the *COL18A1* promoters performing Chromatin Imunoprecipitation (CHIP) (Figure 3A and B). PCRs products were detected for both promoters when precipitated with TCF-1 and LEF-1 antibodies. Despite the core promoter of promoter 1 does not have any predicted TCF/ LEF binding sites, PE3 and *COL18A1* promoter 2 showed predicted TCF/LEF binding sites. We used Ingenuity program to look for a network between *COL18A1* and TCF/LEF, and it showed a network through *CTNNB1* (catenin beta-1), as illustrated in figure 4 *COL18A1* corroborating to *COL18A1* regulation by ß-catenin signaling. However the effect of this interaction is unknown.

Decreased expression of *COL18A1* in cells with active ß-catenin signaling

In attempt to understand the effect of canonical wnt signaling activation on the regulation of *COL18A1* expression, as increase of ß-catenin could lead to upregulation or dowregulatio of *COL18A1*, we first transfected Hek293T cell line that overexpressed ß-catenin constitutively and wild Hek293T cell line with constructs of both *COL18A1* promoters (promoter 1: P1-103, P1-540 and P1-1036; promoter 2: P2-182, P2-555 and P2-1052). A statistic decrease of promoter activity was observed in cells overexpressing ß-catenin transfected with constructs P1-1036 and P2-182 (Figure 5). Both have predicted TCF/LEF binding sites. However, *COL18A1* expression regulation cannot be generalized with this experiment and the possibilities

of distal enhancers regulate activity of proximal regulatory regions of *COL18A1* and play a majority role in its regulation cannot be rejected. Therefore, for a global view of the effect of ß-catenin in *COL18A1* expression regulation we transfected Hek293T with vectors expressing ß-catenin and with WNT1 and performed Real Time-PCR of *COL18A1* and *cMyc*. Cells transfected with ß-catenin and with WNT showed decrease of *COL18A1* expression and the expected increase of *cMyc* (Figure 6).

Increase of *COL181* expression during adipogenesis

One of the biological processes where a decrease of ß-catenina and an increase of *COL18A1* expression happen concomitantly is during adipogenesis.

To evaluate *COL18A1* expression variation during adipogenesis, we submitted to adipogenic differentiation two different lines of human stem cells derived from adipose tissue, and measured expression of *COL18A1* and CEBPa*, CEBPb* and *PPRA-gamma* (genes involved in adipogenesis) at 0, 12 h, 1 day, 3 days, 7 days and 14 days. CEBPb showed high expression levels after 12h of differentiation induction and the same remain until 3 days, at 7 days we observed a little decrease in its expression. However CEBPa*,* increases its expression after 7 days of differentiation. *PPAR-gamma* showed a progressive increase in expression levels. *COL18A1* had a high increase of expression at 1 day of adipogenic differentiation and, interestingly at 14 days its expression decreased to lower values in comparison of that anticipating differentiation induction (Figure 7).

Discussion

Here we describe sequences responsible for *COL18A1* promoter 1 activity. A sufficient sequence capable to lead to promoter activity is localized between -103 and

-273pb (core promoter). An outline of the promoter 1 is represented by peaks and valleys of luciferase activity corresponding to positive (core promoter and enhancers) and negative regulatory regions (insulators), respectively. The *COL18A1* promoter 1 does not present a TATA and CAAT box, but it has a high G + C, besides it is located in a CpG island, showing characteristics of housekeeping gene.

The hypothesis of neutral selection in preserving functional sequence around the genome has been widely applicable. Therefore, in attempt to discuss about important TF that regulates *COL18A1* promoter 1 sequence we aligned each identified region with orthologous counterparts and looked for conserved predicted TF binding sites. TF such as AP-2, NF-1, Sp1, TCF1, LEF1 and PPAR-gamma are predicted to bind in promoter putative region. AP-2 is expressed in neural crest during mouse embryogenesis. NF-1 is known as a liver high expressed TF. PPAR-gamma is one of known TF related with adipogenesis differentiation, a process where an increase of *COL18A1* is observed (Errera, Canani *et al.*, 2008). Both the core promoter and proximal enhancers of promoter 1 have Sp1 conserved predicted binding sites. Sp1 is involved in gene expression in the early development, and it is found in housekeeping and widely expressed genes in eukaryotes. Considering expression of the short collagen XVIII form and *COL18A1* importance in the eyes, liver, brain and adipogenesis process, AP-2, NF-2 and PPAR-gamma are potential TFs to specifically regulate *COL18A1* promoter. PE3 showed higher activity in HepG2 cell line indicating cell-type specificity and NF1 is interestingly predicted to this region. PE3 showed *in vivo* enhancer activity to regulate *COL18A1* as a proximal enhancer element.

We had previously described and characterized the promoter 2 of the *COL18A1,* which showed a core promoter localized in the position -20 to -182 and

controlled by Sp1/Sp3 and YY1. Here, we aligned both core promoters and speculated conserved TF. Both promoters shared Sp1, MZF and Pax5. The similarity between both promoters could help to understand basal transcription the by different isoforms.

TCF/LEF predicted sites are found in both *COL18A1* promoter sequences but not in the core promoter of short isoform. Chromatin imunoprecipitation with TCF1 and LEF1 confirmed their binding to the promoter 1 and 2 of *COL18A1*. ß-catenin can lead to activation and also inhibition of target genes transcription through TCF/ LEF triggering different processes during embryogenesis; therefore it might participate in regulation of *COL18A1* expression. The activity of both *COL18A1* promoters when ß-catenin was overexpressed in Hek293T cell lines showed a downregulation in promoter activity. In a global view of expression, *COL18A1* also showed a decrease of expression when with ß-catenin. This data indicates regulation of *COL18A1* expression by the Wnt-canonical signaling.

In consequence of transcription activation of *COL18A1* isoforms there are production of endostatin and Frizzled, both functioning as antagonists of wnt signaling. Wnt signaling feedback control though *COL18A1* is possible. In this scenario, activation of wnt signaling and consequently activation of TCF/LEF transcription factors incite decrease in endostatin and Frizzled levels, triggering to increase of Wnt signaling activity. The feedback proposed here might play an important role during central processes. In embryogenesis, Wnt signaling (as one of the major active pathways) and *COL18A1* could act in synchrony. During carcinogenesis process, endostatin levels differences, mainly its downregulation, might lead to cancer endorsement by decrease of two processes: angiogenesis and degradation of *ß*-catenin in wnt signaling. Moreover, a biological scenario where it is

observed an opposite balance between wnt and *COL18A1* occurs during adipogenesis process. In this case lower levels of ß-catenin leads to a high expression of *COL18A1* contributing to the decrease of ß-catenin through endostatin and FRZC18, mainly at day 1 during *in vitro* adipogenenic differentiation.

In conclusion, our work elucidated characteristics of *COL18A1* promoter 1 and its downregulation by binding with TCF/LEF highlighting a new process of feedback regulation of wnt signaling pathway with involvement of *COL18A1*. It can help to understand biological processes such as adipogenic differentiation and cancer.

Acknowledgments

We would like to thank Fapesp/CEPID and CNPq for financial support.

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.

Czeizel, A. E., P. Goblyos*, et al.* The second report of Knobloch syndrome. Am J Med Genet, v.42, n.6, Apr 1, p.777-9. 1992.

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.* COL18A1 is highly expressed during human adipocyte differentiation and the SNP c.1136C > T in its "frizzled" motif is associated with obesity in diabetes type 2 patients. An Acad Bras Cienc, v.80, n.1, Mar, p.167-77. 2008.

Fisher, S., E. A. Grice*, et al.* Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat Protoc, v.1, n.3, p.1297-305. 2006.

Haftek, Z., G. Morvan-Dubois*, et al.* Sequence and embryonic expression of collagen XVIII NC1 domain (endostatin) in the zebrafish. Gene Expr Patterns, v.3, n.3, Jun, p.351-4. 2003.

Halfter, W., S. Dong*, et al.* Collagen XVIII is a basement membrane heparan sulfate proteoglycan. J Biol Chem, v.273, n.39, Sep 25, p.25404-12. 1998.

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.

He, T. C., A. B. Sparks*, et al.* Identification of c-MYC as a target of the APC pathway. Science, v.281, n.5382, Sep 4, p.1509-12. 1998.

Kimmel, C. B., W. W. Ballard*, et al.* Stages of embryonic development of the zebrafish. Dev Dyn, v.203, n.3, Jul, p.253-310. 1995.

Kvist, A. P., A. Latvanlehto*, et al.* Complete exon-intron organization and chromosomal location of the gene for mouse type XIII collagen (col13a1) and comparison with its human homologue. Matrix Biol, v.18, n.3, Jun, p.261-74. 1999.

Logan, C. Y. e R. Nusse. The Wnt signaling pathway in development and disease. Annu Rev Cell Dev Biol, v.20, p.781-810. 2004.

Mann, B., M. Gelos*, et al.* Target genes of beta-catenin-T cell-factor/lymphoid-enhancer-factor signaling in human colorectal carcinomas. Proc Natl Acad Sci U S A, v.96, n.4, Feb 16, p.1603-8. 1999.

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. 2001.

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. 2001.

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.

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.

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. 1994.

Passos-Bueno, M. R., S. K. Marie*, et al.* Knobloch syndrome in a large Brazilian consanguineous family: confirmation of autosomal recessive inheritance. Am J Med Genet, v.52, n.2, Aug 15, p.170-3. 1994.

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.

Prockop, D. J. e K. I. Kivirikko. Collagens: molecular biology, diseases, and potentials for therapy. Annu Rev Biochem, v.64, p.403-34. 1995.

Quelard, D., E. Lavergne*, et al.* A cryptic frizzled module in cell surface collagen 18 inhibits Wnt/beta-catenin signaling. PLoS ONE, v.3, n.4, p.e1878. 2008.

Rehn, M., E. Hintikka*, et al.* Primary structure of the alpha 1 chain of mouse type XVIII collagen, partial structure of the corresponding gene, and comparison of the alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen chain. J Biol Chem, v.269, n.19, May 13, p.13929-35. 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. 1998.

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. 1998.

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.

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.

Westerfield, M. The Zebrafish Book. A guide for the laboratory use of zebrafish (Danio rerio). Eugene: University of Oregon Press. 2000

Zhang, X., J. P. Gaspard*, et al.* Regulation of vascular endothelial growth factor by the Wnt and K-ras pathways in colonic neoplasia. Cancer Res, v.61, n.16, Aug 15, p.6050-4. 2001.

Tables and Figures Legends

Table I. List of primers used for amplification of promoter 1 regions and used in Real-Time PCR.

Table II. Predicted Transcription Factors (TF) shared between human and mouse sequence in 1500pb upstream of transcription start site of *COL18A1* exon 1 and between both core promoters of *COL18A1* (Table A). In table B, list of predicted TF for each motif of promoter regions.

Figure 1. Promoter 1 study using deletions constructs and transfections. All the constructs are indicated in A, each of them cloned in a pGL3-Basic plasmid, +1 indicates transcription start site. B and C show luciferase activity measured after transfections in Heck293T and in HepG2, respectively.

Figure 2. Mosaic EGFP expression in G0 zebrafish embryos driven by injection of PE3 region. Expression in forebrain (asterisk) and notochord (triangles) were perfect observed at 2dpf embryos. It triggers to functional activity of PE3 in vivo.

Figure 3. Promoter 1 and 2 sequences precipitated with TCF1 and LEF1 antibodies in CHIP experiments. In A, ØHaeIII marker (lane 1), positive control (lane 2), negative control (lane 3) and input (lane 4). In B, precipitation with TEF1 (lanes 1 and 2) and TCF1 (lanes 3 and 4); and amplification of P1-540 (lanes 1 and 3) and P1 (lanes 2 and 4). Amplification of P2 is not shown.

Figure 4. Network shown between TCF/LEF and *COL18A1* generated with Ingenuity program. This network turns possible through CTNNB1 (ß-catenin). It reinforce TCF/LEF as possible regulators of *COL18A1* expression.

Figure 5. Promoter 1 and 2 regions showed activity decrease when in constantly expression of ß-catenin cell lines. P1-540. P1-1036 and P2-186 showed decrease in promoter activity with $p<0,05$ with ß-catenin. However, P2-1052 had an increase with $p<0,05$ in the same cells.

Figure 6. Real-Time PCR showed a decrease in *COL18A1* expression (black bar) when Hek293T were transfected with ß-catenin or Wnt1 mammalian expression vectors. *cMyc* were used as a positive control, and it expression increased only with Wnt1 transfection.

Figure 7. Real-Time PCR were performed during adipogenesis differentiation and expression of genes involved in the process, such as CEBPa (A) and b (B) and PPRA-gamma (C) were measured as controls, and *COL18A1* expression was also analyzed (D). *COL18A1* showed a relative increase of expression during adipogenesis process, mainly at 1 day after start of induction. This gene might have important skills during early steps of adipogenic differentiation.

Table I.

Primers *COL18A1* promoter 1	**Sequence 5'-3'**
COL18A1-attB2- R	GGGGACCACTTTGTACAAGAAAGCTGGGTGCTCTCCCCAGGACCGTCAGCG
103 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTGCCCGGCCTCCCAGCGG
347 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCCTCGGGGCGGGACGTGT
273 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGCAGCCGCAGGTCCGCT
540 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTAGGTGCCGGCTTCCTCCTCCGTC
752 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGGCTCAGGATGCTGGCCCTGAA
919 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAGAGGGTTTCAGAGAAAGCCGTCAAGCC
1036 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTAGGATGGGTGGGGCCACGTGTG
1554 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTAGACACACAGACACACCCACTGCACTGC

1253 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTATTGGGCCTGGGAAGGCTGAGGC
1702 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTAACACCCCTGACACACACTCACACGCA
Primers Real Time-PCR	**Sequence 5'-3'**
HPRT1 F	TGACACTGGCAAAACAATGC
HPRT1 R	GGTCCTTTTCACCCAGCAAGC
HMBS F	GGCAATGCGGCTGCAA
HMBS R	GGTACCCACGCGAATCAC
SDHA F	TGGGAACAAGAGGGCATCTG
SDHA R	CCACCACTGCATCAAATTCA
GAPDH F	TGCACCACCAACTGCTTAGC
GAPDH R	GGCATGGACTGTGGTCATGA
COL18 F	CAGCAATGTGTTTGCTGAGTC
COL18 R	TCAAACGGAAACTGCCCT
cMYC F	GCGACTCTGAGGAGGAACAAG
cMYC R	CCAGACTCTGACCTTTTGCCA
C/EBPa F	AAGAACAGCAACGAGTACCGG
C/EBPa R	CATTGTCACTGGTCAGCTCCA
C/EBPß F	CGCCGCCTGCCTTTAAATCC
C/EBPß R	AGCCAAGCAGTCCGCCTCG
PPRAg F	GGGGGTGATGTGTTTGAACT
PPRAg R	GCGATCTTGACAGGAAAGACA

Table II.

A

Region	TF shared with mouse (CONSITE)	TF shared with core promoter 2 (CONSITE)
Core Promoter 1	Sp1, Pax5(Bsap), AP2	MZF, Pax5, Sp1, c-ETS
PE 1	c-ETS,RUSH1-alfa,RoR-alfa, Sp1, AP2, GATA1, Thing-E47, deltaEF1, FREAC-3, MZF, Pax2, Gklf	
PE 2	Spz1,MZF, AP2	
PE 3	Sox5, 9 and 17, SRY, HFH, RUSH1-alfa, NHF, Pbx, EN-1, GATA, Nkx,TCF/LEF, PPAR-gamma, Pax2, Sp1, c-ETS, FREAC, Yin-Yang, MEF2, Gklf, Gfi	

B

Region	Motives found also in mouse (MEME)	TFs predicted (TESS)
PE1	motive 2	
	motive 3	NF-E
PE2	motive 1	ADR1, NF-1
	motive 2	RFX2, NF-X3, E12, myogenin
PE3	motive 1	MEF-2, YY1,FEF, POU1f1a, TBP, SEF4,YY1,AP-3, NF-1, ETF, DBF4, TFIIA, TFIIB, TFIID, TMF, UBP-1, POU2F1
	motive 2	NF-1, ABF-2, LBP-1, TFIID, RFX2, NF-X3
	motive 3	TCF/LEF, Dof2, NMB1, TFIID, PBF

Capítulo IV- Dissection of *COL18A1* promoter 1 and control of wnt signaling through regulation of *COL18A1* expression

Figure 1.

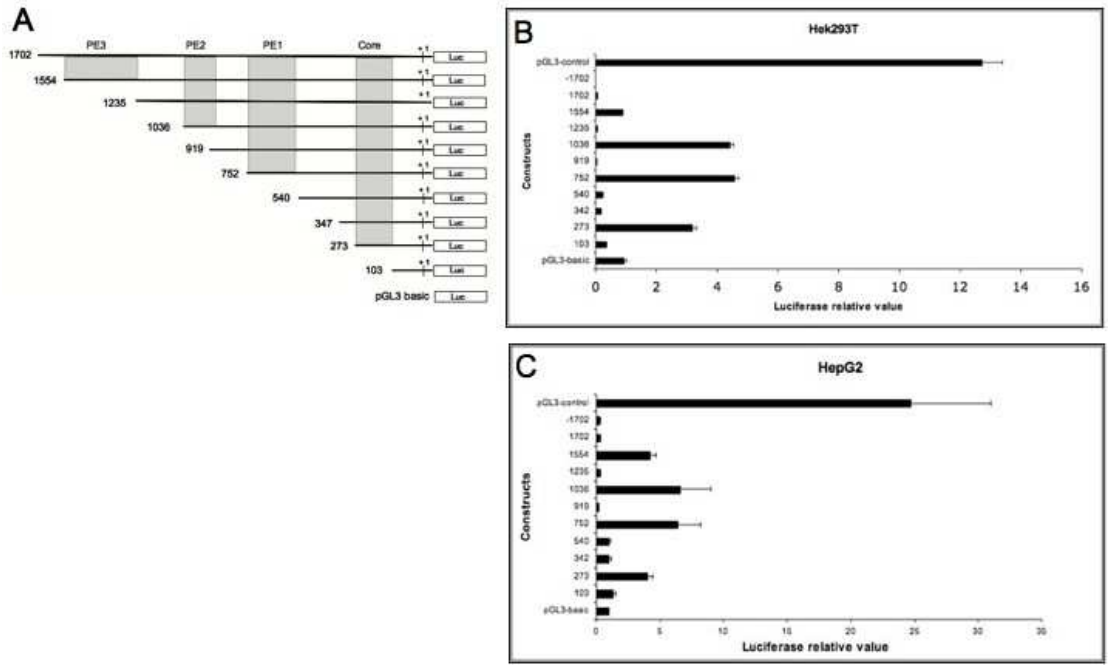


Figure 2.

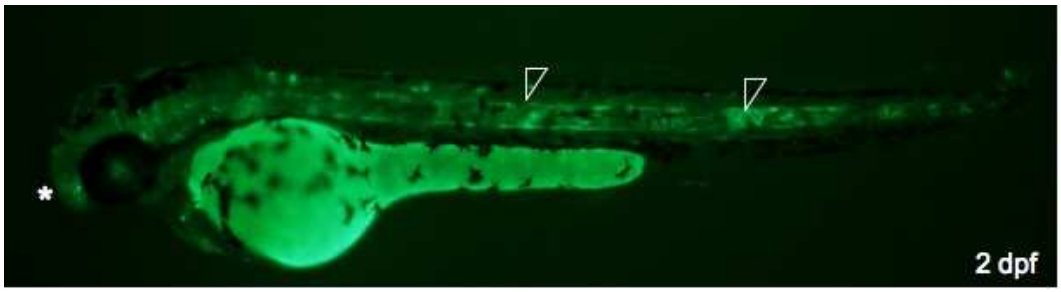


Figure 3.

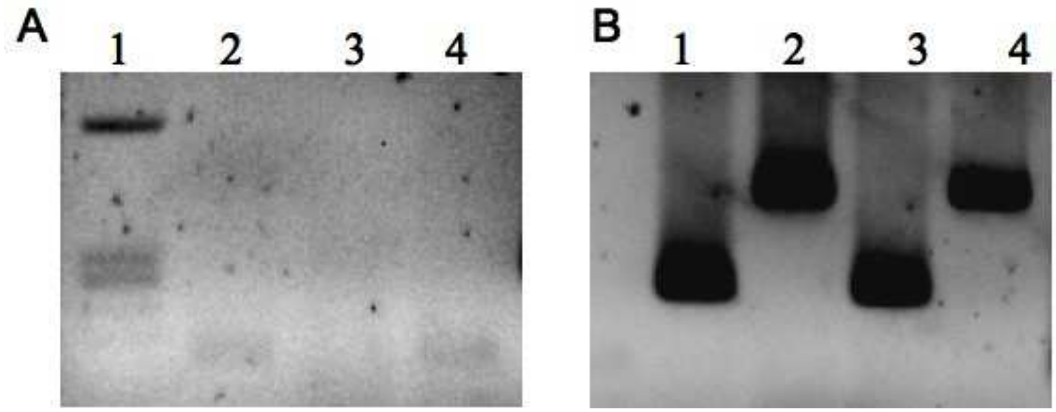


Figure 4.

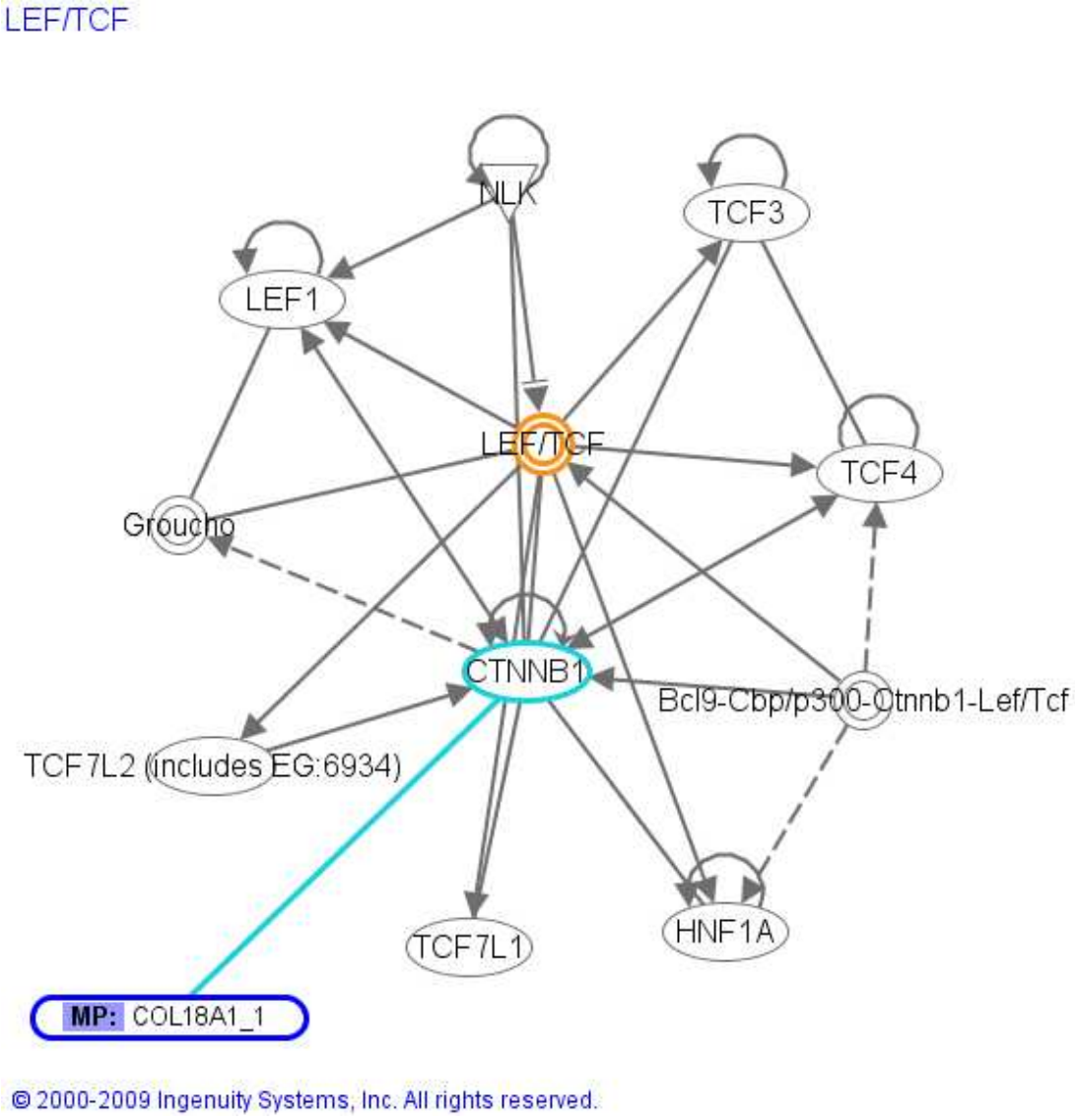


Figure 5.

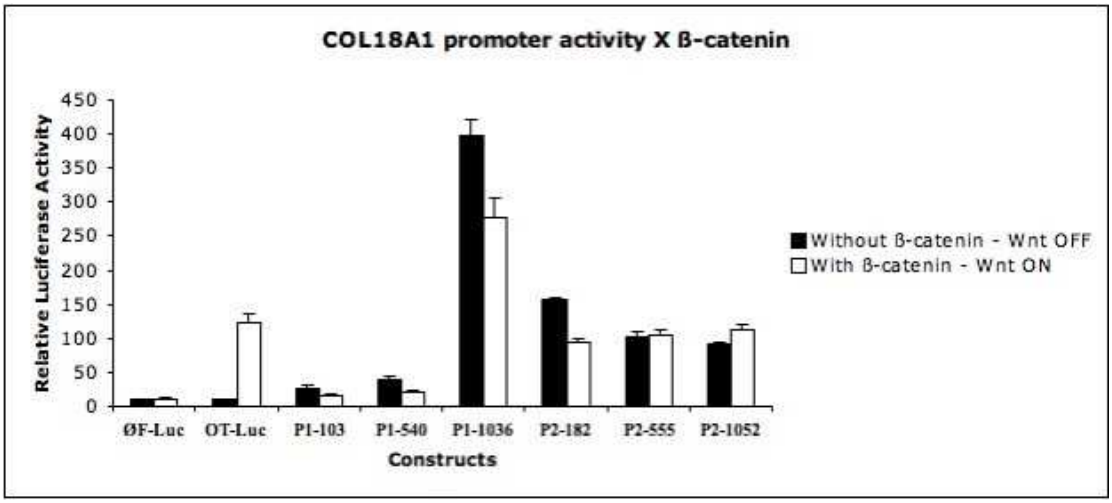


Figure 6.

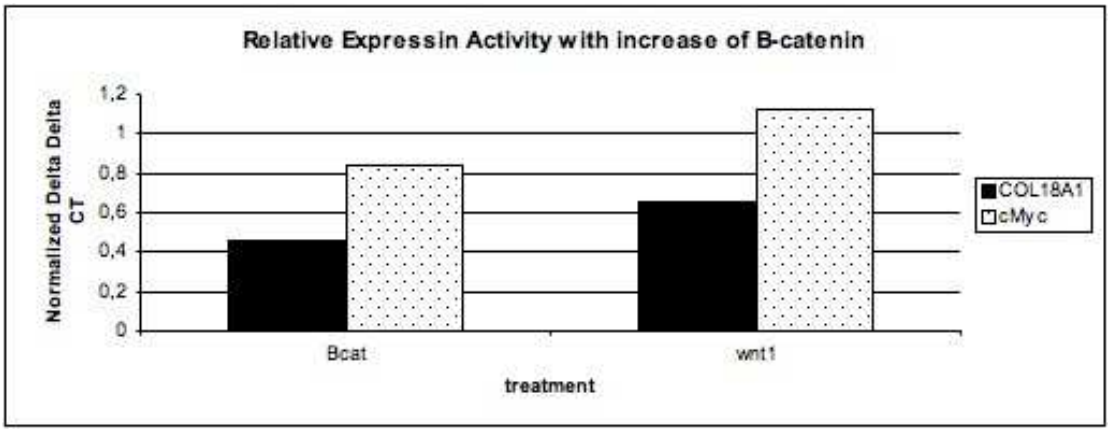
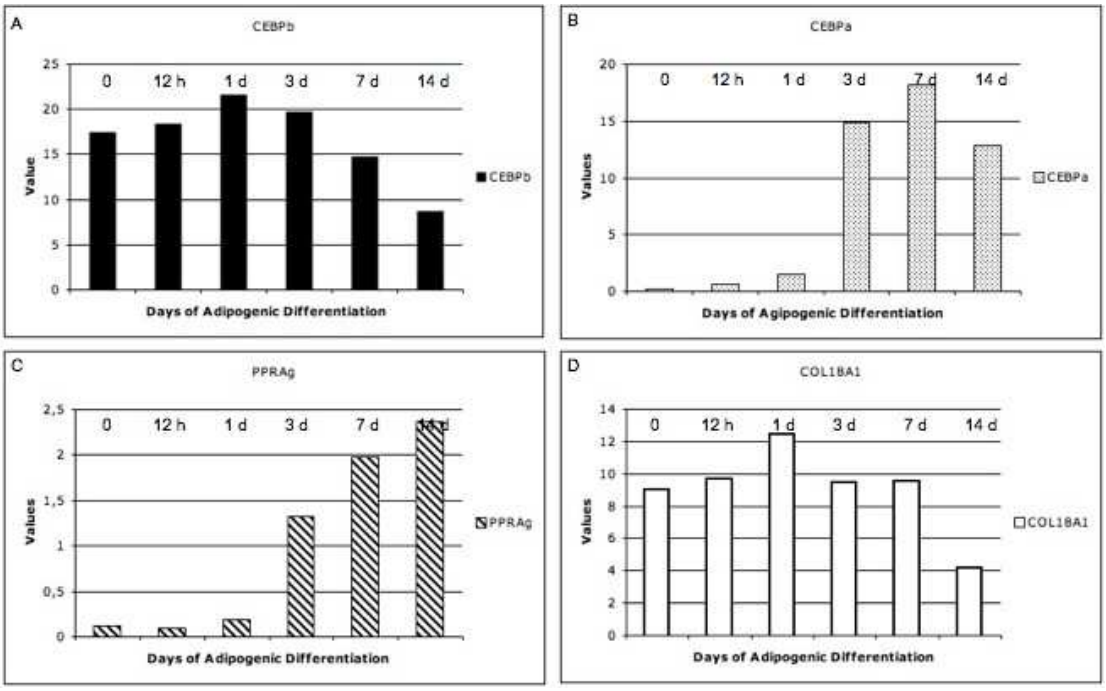


Figure 7.


A
CEBPb
0 12 h 1 d 3 d 7 d 14 d
Value
CEBPb
Days of Adipogenic Differentiation
B
CEBPa
0 12 h 1 d 3 d 7 d 14 d
Values
CEBPa
Days of Agipogenic Differentiation
C
PPRAg
0 12 h 1 d 3 d 7 d 14 d
Values
PPRAg
Days of Adipogenic Differentiation
D
COL18A1
0 12 h 1 d 3 d 7 d 14 d
Values
COL18A1
Days of Adipogenic Differentiation

Capítulo V

5. Collagen XVIII cleaved frizzled fragment inhibits non-canonical Wnt signaling

Erika Kague[1]; Shannon Fisher[2]; Lucia Maria Armelin-Correa[1] and Maria Rita Passos-Bueno[1†].

4- Centro de Estudos do Genoma Humano, Instituto de Biociencias, Universidade de São Paulo, Brazil

5- McKusick–Nathans Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD 21205; current address: Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, PA 19104

† - Corresponding author. Current address: Rua do Matão 277 sala 200, Depto. Genética e Biologia Evolutiva, Instituto de Biociências, Cidade Universitaria, São Paulo, Brazil, CEP: 05508-090; Phone (55) 11 30919910; Fax (55) 11 30917419; passos@ib.usp.br.

Capítulo V- Collagen XVIII cleaved frizzled fragment inhibits non-canonical Wnt signaling

Abstract

Collagen type XVIII is a heparan sulfate proteoglycan component of basement membranes. Homozygous null mutations in *COL18A1* leads to high myopia, retinal detachment and occipital encephalocele (Knobloch Syndrome). Some patients also have neuronal migration abnormalities. The longest of the three transcribed isoforms of *COL18A1* can generate a small proteolytic fragment that contains a motif (FRZC18) with sequence similarity to secreted Frizzled related proteins. *In vitro* data suggests that FRZC18 can antagonize canonical Wnt signaling, but its *in vivo* function is not completely understood. We studied the role of FRZC18 in Wnt signalin, using the zebrafish embryo as an *in vivo* model. We show that overexpression of FRZC18 induces shortened tail and fusion of the eyes, phenotypes similar to those observed in mutants of the non–canonical Wnts, *slb/wnt11* and *ppt/wnt5a*. FRZC18 also leads to broadened notochord and somites, indicative of disturbance in gastrulation movements. The effects of FRZC18 overexpression were enhanced in *slb* and *ppt* heterozygous mutants. Conversely, a mutation in FRZC18 predicted to interfere with its binding to Wnts also reduced its ability to mimic the *slb* and *ppt* mutant phenotypes. Our data supports a role of FRZC18 *in vivo* as an antagonist of non-canonical Wnt signaling, acting through *Wnt11* and *Wnt5a.*

Keywords: Collagen XVIII, FRZC18, Wnt signaling, *slb*, *ppt*, Knobloch Syndrome

Capítulo V- Collagen XVIII cleaved frizzled fragment inhibits non-canonical Wnt signaling

Introduction

Collagen type XVIII is a heparan sulfate proteoglycan expressed in basement membranes in a wide variety of tissues (Halfter, Dong *et al.*, 1998; Saarela, Rehn *et al.*, 1998; Saarela, Ylikarppa *et al.*, 1998). Collagens XVIII and XV make up the multiplexin subgroup of non–fibrillar collagens (Oh, Warman *et al.*, 1994; Rehn e Pihlajaniemi, 1994; Saarela, Ylikarppa *et al.*, 1998). *COL18A1* maps to ch21q22.3, comprises 43 exons and transcribes 3 different isoforms by the use of two promoters and alternative splicing in the third exon. The three encoded collagen XVIII proteins differ only by their signal peptides and by part of the N-terminal NC11 domain. The short isoform (NC11-303) is transcribed from the first promoter, localized upstream of exon 1, and is formed by exons 1, 2, 4-43. The intermediate (NC11-493) and long (NC11-728) forms are transcribed from the second promoter, upstream of exon 3, and they differ by an internal splice site within exon 3 (Elamaa, Peterson *et al.*, 2002). Homozygous null mutations in *COL18A1* lead to Knobloch Syndrome, a rare autosomal recessive disorder characterized by high myopia, vitreoretinal degeneration, macular alteration and occipital encephalocele (Sertie, Sossi *et al.*, 2000; Suzuki, Sertie *et al.*, 2002). Besides these general clinical features, patients lacking only the long and intermediate isoforms also have abnormal neuronal cell migration and recurrent epilepsy, processes not yet completely understood (Suzuki, Sertie *et al.*, 2002; Passos-Bueno, Suzuki *et al.*, 2006).

At least two cleaved fragments are derived from collagen XVIII, endostatin and the frizzled domain. Endostatin is a 20kD C-terminal cleaved product that can inhibit angiogenesis and endothelial cell migration and proliferation (O'reilly, Boehm

et al., 1997). Endostatin is common to all three variants of collagen XVIII, but the frizzled domain (FRZC18) is derived only from the N-terminal region of the long isoform. The 35kD FRZC18 fragment is characterized by a conserved region of 120 amino acids containing 10 invariantly spaced cysteines, the cysteine rich domain (CRD) (Elamaa, Snellman *et al.*, 2003; Quelard, Lavergne *et al.*, 2008). This domain shows homology with soluble frizzled related proteins (sFRPs), which are antagonists that directly bind to Wnts (Kawano e Kypta, 2003).

Wnts encompass a family of glycoproteins, conserved signaling molecules involved in a surfeit of developmental and cell biological processes, including cell proliferation, differentiation and polarity (Heisenberg, Tada *et al.*, 2000; Logan e Nusse, 2004). Canonical Wnt signaling is dependent on ß-catenin protein stabilization in the cytoplasm, with subsequent activation of transcription of TCF/LEF target genes (Logan e Nusse, 2004). Non-canonical Wnt signaling is independent of ß-catenin, activating signaling pathways such as the planar-cell-polarity (PCP) pathway that guides cell movements during gastrulation, and the Wnt/Ca^{2+} pathway (Veeman, Axelrod *et al.*, 2003); recently, non-canonical Wnt signaling has also been related to neuronal cell migration and polarity (Carreira-Barbosa, Concha *et al.*, 2003; Lange, Mix *et al.*, 2006). Wnts bind directly to Frizzled transmembrane receptors; canonical Wnts also require a coreceptor, a transmembrane molecule of the LRP (LDL receptor related protein) class, LRP5 or 6 (Holmen, Salic *et al.*, 2002; Mi e Johnson, 2005). Each Frizzled CRD has specific affinities for particular Wnts, and it is believed that signal transduction depends on the strength of this interaction. The function of FRZC18 is not clear; recently Quérlard et al., have suggested through *in vitro* experiments that it can inhibit canonical Wnt signaling in liver tumor cells, possibly through interaction with Wnt3a (Quelard, Lavergne *et al.*, 2008), but they did not

assay for possible effects on non–canonical Wnt signaling or tested this hypothesis in *in vivo* experiments

The examination of Wnt signaling in the zebrafish has been greatly enhanced by the isolation of mutations affecting both the canonical (Kelly, Erezyilmaz *et al.*, 1995; Kelly, Chin *et al.*, 2000) and non–canonical pathways (Heisenberg, Tada *et al.*, 2000; Lele, Bakkers *et al.*, 2001; Marlow, Topczewski *et al.*, 2002; Kilian, Mansukoski *et al.*, 2003; Ulrich, Krieg *et al.*, 2005). Mutants in *slb/wnt11* and *ppt/wnt5a* , both non-canonical Wnts, have defects in convergent extension movements and the PCP pathway (Heisenberg, Tada *et al.*, 2000; Kilian, Mansukoski *et al.*, 2003). *Slb* homozygous mutants are predominantly affected in anterior regions of embryos, showing a transiently shortened and broadened body axis at the end of gastrulation followed by fusion of the eyes at later developmental stages (Heisenberg, Tada *et al.*, 2000). In contrast, *ppt* homozygous mutant embryos exhibit a shortened body axis from late gastrulation stages onwards while the position of the eyes is only mildly affected (Kilian, Mansukoski *et al.*, 2003). *Slb* and *ppt* have partially overlapping function in regulating convergent extension movements in lateral domains of gastrula, and in the absence of *ppt*, the *slb* homozygous phenotype is strongly enhanced (Heisenberg e Tada, 2002).

We have overexpressed FRZC18 in zebrafish embryos to test its participation in Wnt signaling pathways *in vivo*. Excess FRZC18 induced shortened tail and fusion of the eyes, similar to *slb* and *ppt* mutants, and these phenotypes were enhanced in the setting of reduced *slb* and *ppt*. We propose that FRZC18 plays a predominant role as an antagonist of non–canonical Wnt signaling in the embryo, through interactions with Wnt5a and Wnt11.

Results

Overexpression of FRZC18 mRNA in zebrafish embryos

To investigate the role of FRZC18 in Wnt signaling, we injected varying amounts of mRNA encoding FRZC18 into zebrafish embryos at the 1–cell stage. Injections of more than 400 pg led to embryo death. Injection of 300 pg resulted in 66% affected embryos (n = 500), exhibiting either a shortened tail or fusion of the eyes observed after 24 hpf (Figure 1). There was considerable variability in phenotype (Figure 2). Although the majority of affected embryos had both eye and tail defects, a small percentage presented only eyes or only tail defects (5.9 % and 3.18%, respectively). The malformation of the anterior axis also showed variability, including embryos with reduced interocular distance, complete fusion of the eyes, and also embryos with no eyes (Figure 3).

To confirm a requirement for Wnt binding in generation of the phenotypes, we introduced a mutation in the FRZC18, based on previously work defining contact sites for binding between Wnts and frizzled CRD domains (Dann, Hsieh *et al.*, 2001). Based on spacing of the conserved cysteines, we introduced a mutation predicted to interfere with Wnt binding. Injections with the FRZC18-mut resulted in only 18% of embryos affected (n=470) in contrast to 66% affected when injected with the FRZC18 (Figure 4).

FRZC18 and its relation with wnt11 and wnt5

The phenotypes caused by overexpression of FRZC18 were reminiscent of those seen in mutants for the non–canonical Wnt genes *slb/wnt11* (Heisenberg, Tada *et al.*, 2000; Lele, Bakkers *et al.*, 2001) and *ppt/wnt5a* (Lele, Bakkers *et al.*, 2001;

Kilian, Mansukoski *et al.*, 2003). To test for direct interaction, we injected FRZC18 RNA into embryos heterozygous for mutations in *slb* or *ppt*, in an amount that causes only occasional phenotypes in wild–type embryos. Embryos displaying a phenotype were genotyped for the presence of the mutated alleles. In the setting of reduced *slb* dosage, 63% (32/52) of embryos with fused eyes were heterozygous for *slb*. Similarly, in the injected *ppt* clutch, 68% (67/99) of embryos with a shortened tail were heterozygous for *ppt*. Heterozygotes for both mutations normally appear wild–type, strongly suggesting that FRZC18 is directly impinging on non–canonical Wnt signaling in the zebrafish embryo.

FRZC18 interfering in gastrulation movements

In an attempt to verify if FRZC18 has any effect in the gastrulation process, we performed in situ hybridization for *myoD* in AB embryos at 10-14 somite stage submitted to injection of FRZC18 or FRZC18-mut.

As observed in figure 5 (A-E), embryos injected with FRZC18 mRNA showed defects in somite formation and they presented a broadened notochord. In some embryos this defect showed to be very severe and the somite did not develop (Figure 5F). The control construct, FRZC18-mut, did not show any defects and it appeared similar like to non-injected embryos.

FRZC18 overexpression does not interfere in ß-Catenin.

Trying to understand which downstream proteins are related with overexpression of FRZC18, we transfected Hek293T cells with FRZC18 (Lumio-FRZC18) and we investigated proteins involved with canonical Wnt signaling (ß-catenin (ß-cat)) and non- canonical Wnt patway (disheveled (dvl) and Jun NH2-

terminal Kinase (JNK)). Except for dvl that was tested only on the membrane all the others were evaluated in all cell compartments. We did not observe difference in quantity or distribution of ß-cat, dvl and JNK proteins in cells transfected with FRZC18 as those transfected with empty vector only (figures 6).

Discussion

We have used the zebrafish embryo to evaluate the *in vivo* function of the frizzled domain of collagen XVIII and its role in regulating Wnt signaling. Overexpression of FRZC18 during early embryogenesis mimics the phenotypes of mutants in *slb* and *ppt*, the zebrafish orthologues of *Wnt11* and *Wnt5a* respectively, and the primary non–canonical Wnts signals in the early embryo. During gastrulation, embryos with excess FRZC18 displayed a shortened anterior–posterior axis, broad notochord, and widened somites, hallmarks of a defect in convergent extension. Later, they showed at a high frequency defects in tail extension, as seen in *ppt* mutants, and a narrowing of interocular distance, characteristic of *slb* mutants. There was no apparent difference in the category of defects induced by varying levels of injected FRZC18 RNA, suggesting that FRZC18 is equally capable of affecting signaling from both Wnts.

Two lines of evidence support a specific interaction between FRZC18 and the non–canonical Wnts. Structure–function correlations on another frizzled domain identified cysteine residues required for interaction with its cognate Wnt (Dann, Hsieh *et al.*, 2001). Based on these results, we introduced a mutation into FRZC18 predicted to interfere with Wnt binding; the mutant FRZC18 was much less effective at inducing both the early and late phenotypes associated with reduction in non–canonical Wnt signaling. We also examined the effect of FRZC18 overexpression in

the setting of reduced Wnt signaling. Injection of RNA amounts that produce little phenotype in wild–type embryos generated defects at high frequency in both *slb* and *ppt* heterozygotes.

Expression of *col18a1* in the early embryo overlaps with key areas of Wnt signaling, including the anterior neural plate during somitogenesis (Haftek, Morvan-Dubois *et al.*, 2003), consistent with FRZC18 playing an *in vivo* role in regulating Wnt signaling. In addition to regulating convergence and extension movements at gastrulation, non–canonical Wnt signaling plays a key role in neuronal migration and neural plate closure (Carreira-Barbosa, Concha *et al.*, 2003; Lange, Mix *et al.*, 2006). Interestingly, Knobloch Syndrome patients lacking all isoforms of collagen XVIII display defects in neuronal migration, while those with mutations affecting only the short isoform do not (Suzuki, Sertie *et al.*, 2002; Passos-Bueno, Suzuki *et al.*, 2006). We speculate that the phenotypic difference may be due in part to the former group of patients lacking FRZC18, with a subsequent upregulation of non–canonical Wnt signaling during neural patterning.

In the early zebrafish embryo, canonical Wnt signaling is required for dorsal axis induction and posterior domain specification, and the phenotypes associated with a reduction in signaling have been well characterized (Funayama, Fagotto *et al.*, 1995; Kelly, Erezyilmaz *et al.*, 1995). At the levels of RNA introduced, we saw no evidence of an effect on canonical Wnt signaling in zebrafish and also no difference in ß-catenin concentrations were observed in our cell culture. It has been previously reported that overexpression of FRZC18 in cell culture reduced canonical Wnt signaling (Quelard, Lavergne *et al.*, 2008). However, that report did not examine any downstream effectors of the non–canonical pathway. Several explanations may reconcile the apparent conflict. First, it is possible that FRZC18 is capable of

interacting with canonical Wnts, although at a lower affinity. There is also evidence that some Wnt proteins, including Wnt3 and Wnt5a, are capable of activating both arms of the pathway, depending on the repertoire of available receptors and other modulating proteins (Habas, Kato *et al.*, 2001; Westfall, Brimeyer *et al.*, 2003; Kishida, Yamamoto *et al.*, 2004; Endo, Wolf *et al.*, 2005). We have not observed differences of Dvl and JNK (members of the non-canincal signaling) in cell compartments in comparison with cells overexpressing of FRZC18; it indicates that other non-canonical proteins might be downstream effectors of FRZC18 action in the non-canonical Wnt signaling. Further examination of the role of FRZC18 in the zebrafish embryo, where both canonical and non–canonical arms of the Wnt signaling pathway can be readily assayed, should help to resolve the issue.

In summary, we have focused on the investigation of the functionality of FRZC18 through overexpression in zebrafish embryos. We conclude that FRZC18 is regulating non-canonical Wnt signaling through interactions with Wnt11 and Wnt5. FRZC18 also disrupted convergent extension movements during gastrulation, similar to mutants for members of the non-canonical Wnt pathway. Analysis of downstream proteins did not show an effect on the canonical Wnt signaling pathway. FRZC18 participation in non-canonical Wnt signaling might help further studies to understand abnormal neuronal cell migration in some Knobloch patients and in cell migration processes during embryogenesis.

Materials and Methods

Plasmid constructs and mRNA synthesis

Capítulo V- Collagen XVIII cleaved frizzled fragment inhibits non-canonical Wnt signaling

The collagen XVIII frizzled domain (FRZC18) sequence was amplified from human genomic DNA (Roche) with the following primers: F – accatgtgcctgcccctgccaccctcc and R- cacacagtacccatcctcctgggt. PCR product was cloned into the pCR8/GW/TOPO vector (Invitrogen). LR recombination (Invitrogen) was performed to transfer the FRZC18 to the Gateway destination pCSDest, to generate pCS-FRZC18. To generate the negative control plasmid, we inserted Gly-Ser-Gly (CCCAGGCCC) between codons 31 and 32 in the FRZC18 sequence as follows. Two amplifications were performed, using the FRZC18 topo clone as a template and the following primers: 1aF-ggggacaagtttgtacaaaaaagcaggctaccatgtgcctgccctgccaccctccctg and 1aR-cacctggggcctgggctcgccgctctcgt, 1bF-acgagagcggcggagcccaggccccaggtgcgggccgggcac and 1bR- ggggaccactttgtacaagaaagctgggtcacacagtacccatcctcctgggt. Both amplicons were digested with BanII and ligated together. The ligation product was cloned into pDONR 221 with BP clonase (Invitrogen), and an LR reaction performed to obtain pCS-FRZC18-mut. Both plasmids were linearized with HpaI and in vitro synthesis of capped mRNA performed with the mMESSAGE mMACHINE kit (Ambion).

Fish care and Injections

Wild–type AB fish and carriers of the *pipetail* (*ppt*) and *silberblick* (*slb*) alleles were maintained in accordance with standard conditions (Westerfield, 2000) and the regulations of the Johns Hopkins IACUC. For injections, mRNAs were diluted in sterile water containing 0.1% phenol red, and injected in a volume of 1-5nL into embryos at the 1-2 cell stage. Injected and non–injected control embryos were analyzed for phenotype from 24h to 4dpf.

In situ Hybridization

Zebrafish embryos were collected at 10 to 14 somite stage (14-16hpf) and fixed in 4% paraformaldehyde. Digoxin labeled antisense *myoD* riboprobe was synthesized using T7 RNA polymerase, and whole-mount situ hybrydizations performed according to standard protocols.

Zebrafish genotyping

Fish heterozygous for *pipetail* (*ppt*) or *silberblick* (*slb*) mutations were mated with AB fish. Embryos were injected with 0.15ng of FRZC18 mRNA. Embryos showing phenotype alteration were collected at 3 dpf and genomic DNA prepared using standard protocols (Westerfield, 2000). After PCR amplification, DNA was digested with diagnostic restriction enzymes.

Transfection and Protein Blots

FRZC18 was cloned into a mammalian Lumio Gateway vector (Invitrogen). We transfected HEK 293T cells at 70% confluence with Lipofectamine 2000 (Invitrogen). Cells were harvested 72h after transfection. The membrane protein fraction was isolated with the Mem-PER protein preparation kit (Pierce), and the cytoplasmic and nuclear fractions were isolated with the NE-PER kit (Pierce). Protein content was measured by the Lowry method and samples were normalized. 20 μg of total protein was separated by SDS-PAGE on precast Rgels (10% or 4-15% Tris-HCl) and transferred to a PVDF membrane (Bio-Rad). The membrane was probed with commercial antibodies raised in rabbit (anti-ß-catenin (1:100, Santa Cruz SC7199) and anti-JNK/SAPK1 (1:1,000, Upstate cat. No. 06-748)), in goat (anti-Dvl (C-19) (1:500, Santa Cruz SC7397)) and in mouse (anti-GAPDH (1:5,000, Abcam ab 9894)).

Proteins were detected with secondary antibodies conjugated to horseradish peroxidase (1:10,000 each, RPN4201) and the ECL detection kit (both GE Healthcare).

Acknowledgments

We would like to thank Dr Jantji M Gerdes for helping with western blots and discussion of ideas. We thank Seneca Bessling and Melissa Biddelecomb for her kindness and help with fish care; Constancia G. Urbani for secretarial help. This study was financially supported by Fapesp/CEPID and CNPq

References

Carreira-Barbosa, F., Concha, M.L., Takeuchi, M., Ueno, N., Wilson, S.W. and Tada, M. (2003) Prickle 1 regulates cell movements during gastrulation and neuronal migration in zebrafish. Development 130, 4037-46.

Dann, C.E., Hsieh, J.C., Rattner, A., Sharma, D., Nathans, J. and Leahy, D.J. (2001) Insights into Wnt binding and signalling from the structures of two Frizzled cysteine-rich domains. Nature 412, 86-90.

Elamaa, H., Peterson, J., Pihlajaniemi, T. and Destree, O. (2002) Cloning of three variants of type XVIII collagen and their expression patterns during Xenopus laevis development. Mech Dev 114, 109-13.

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, 427-42.

Endo, Y., Wolf, V., Muraiso, K., Kamijo, K., Soon, L., Uren, A., Barshishat-Kupper, M. and Rubin, J.S. (2005) Wnt-3a-dependent cell motility involves RhoA activation and is specifically regulated by dishevelled-2. J Biol Chem 280, 777-86.

Funayama, N., Fagotto, F., McCrea, P. and Gumbiner, B.M. (1995) Embryonic axis induction by the armadillo repeat domain of beta-catenin: evidence for intracellular signaling. J Cell Biol 128, 959-68.

Habas, R., Kato, Y. and He, X. (2001) Wnt/Frizzled activation of Rho regulates vertebrate gastrulation and requires a novel Formin homology protein Daam1. Cell 107, 843-54.

Haftek, Z., Morvan-Dubois, G., Thisse, B., Thisse, C., Garrone, R. and Le Guellec, D. (2003) Sequence and embryonic expression of collagen XVIII NC1 domain (endostatin) in the zebrafish. Gene Expr Patterns 3, 351-4.

Halfter, W., Dong, S., Schurer, B. and Cole, G.J. (1998) Collagen XVIII is a basement membrane heparan sulfate proteoglycan. J Biol Chem 273, 25404-12.

Heisenberg, C.P. and Tada, M. (2002) Zebrafish gastrulation movements: bridging cell and developmental biology. Semin Cell Dev Biol 13, 471-9.

Heisenberg, C.P., Tada, M., Rauch, G.J., Saude, L., Concha, M.L., Geisler, R., Stemple, D.L., Smith, J.C. and Wilson, S.W. (2000) Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. Nature 405, 76-81.

Holmen, S.L., Salic, A., Zylstra, C.R., Kirschner, M.W. and Williams, B.O. (2002) A novel set of Wnt-Frizzled fusion proteins identifies receptor components that activate beta -catenin-dependent signaling. J Biol Chem 277, 34727-35.

Kawano, Y. and Kypta, R. (2003) Secreted antagonists of the Wnt signalling pathway. J Cell Sci 116, 2627-34.

Kelly, C., Chin, A.J., Leatherman, J.L., Kozlowski, D.J. and Weinberg, E.S. (2000) Maternally controlled (beta)-catenin-mediated signaling is required for organizer formation in the zebrafish. Development 127, 3899-911.

Kelly, G.M., Erezyilmaz, D.F. and Moon, R.T. (1995) Induction of a secondary embryonic axis in zebrafish occurs following the overexpression of beta-catenin. Mech Dev 53, 261-73.

Kilian, B., Mansukoski, H., Barbosa, F.C., Ulrich, F., Tada, M. and Heisenberg, C.P. (2003) The role of Ppt/Wnt5 in regulating cell shape and movement during zebrafish gastrulation. Mech Dev 120, 467-76.

Kishida, S., Yamamoto, H. and Kikuchi, A. (2004) Wnt-3a and Dvl induce neurite retraction by activating Rho-associated kinase. Mol Cell Biol 24, 4487-501.

Lange, C., Mix, E., Rateitschak, K. and Rolfs, A. (2006) Wnt signal pathways and neural stem cell differentiation. Neurodegener Dis 3, 76-86.

Lele, Z., Bakkers, J. and Hammerschmidt, M. (2001) Morpholino phenocopies of the swirl, snailhouse, somitabun, minifin, silberblick, and pipetail mutations. Genesis 30, 190-4.

Logan, C.Y. and Nusse, R. (2004) The Wnt signaling pathway in development and disease. Annu Rev Cell Dev Biol 20, 781-810.

Marlow, F., Topczewski, J., Sepich, D. and Solnica-Krezel, L. (2002) Zebrafish Rho kinase 2 acts downstream of Wnt11 to mediate cell polarity and effective convergence and extension movements. Curr Biol 12, 876-84.

Mi, K. and Johnson, G.V. (2005) Role of the intracellular domains of LRP5 and LRP6 in activating the Wnt canonical pathway. J Cell Biochem 95, 328-38.

O'Reilly, M.S., Boehm, T., Shing, Y., Fukai, N., Vasios, G., Lane, W.S., Flynn, E., Birkhead, J.R., Olsen, B.R. and Folkman, J. (1997) Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. Cell 88, 277-85.

Oh, S.P., Warman, M.L., Seldin, M.F., Cheng, S.D., Knoll, J.H., Timmons, S. and Olsen, B.R. (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, 494-9.

Passos-Bueno, M.R., Suzuki, O.T., Armelin-Correa, L.M., Sertie, A.L., Errera, F.I., Bagatini, K., Kok, F. and Leite, K.R. (2006) Mutations in collagen 18A1 and their relevance to the human phenotype. An Acad Bras Cienc 78, 123-31.

Quelard, D., Lavergne, E., Hendaoui, I., Elamaa, H., Tiirola, U., Heljasvaara, R., Pihlajaniemi, T., Clement, B. and Musso, O. (2008) A cryptic frizzled module in cell surface collagen 18 inhibits Wnt/beta-catenin signaling. PLoS ONE 3, e1878.

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 U S A 91, 4234-8.

Saarela, J., Rehn, M., Oikarinen, A., Autio-Harmainen, H. and 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, 611-26.

Saarela, J., Ylikarppa, R., Rehn, M., Purmonen, S. and Pihlajaniemi, T. (1998b) 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, 319-28.

Sertie, A.L., Sossi, V., Camargo, A.A., Zatz, M., Brahe, C. and 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, 2051-8.

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. and 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, 1320-9.

Ulrich, F., Krieg, M., Schotz, E.M., Link, V., Castanon, I., Schnabel, V., Taubenberger, A., Mueller, D., Puech, P.H. and Heisenberg, C.P. (2005) Wnt11 functions in gastrulation by controlling cell cohesion through Rab5c and E-cadherin. Dev Cell 9, 555-64.

Veeman, M.T., Axelrod, J.D. and Moon, R.T. (2003) A second canon. Functions and mechanisms of beta-catenin-independent Wnt signaling. Dev Cell 5, 367-77.

Westerfield, M. (2000) The Zebrafish Book. A guide for the laboratory use of zebrafish (Danio rerio). University of Oregon Press, Eugene.

Westfall, T.A., Brimeyer, R., Twedt, J., Gladon, J., Olberding, A., Furutani-Seiki, M. and Slusarski, D.C. (2003) Wnt-5/pipetail functions in vertebrate axis formation as a negative regulator of Wnt/beta-catenin activity. J Cell Biol 162, 889-98.

Figure Legends

Figure 1. Overexpression of FRZC18 leads to shortened tail and fusion of the eyes. A- No injected wild type fish (Wt). B- Embryo injected with 0.2 ng of FRZC18 showed posterior body axis defect and abnormal notochord (arrow). C and D- 0.3ng of FRZC18. Embryos had fusion of the eyes and shortened tail, similar with silberblick and pipetail mutants.

Figure 2. Injection of FRZC18 leading to different classes of affected embryos. The arrow indicates increase in severity of affected embryos and the different classes of phenotypes observed. Normal embryos to eyeless and shortened tail embryos were found as result of the FRZC18 mRNA injection.

Figure 3. Embryos injected with FRZC18 show variable intereyes distance. A- No injected wild-type (Wt). B-D- FRZC18 injected embryos. B- Some embryos showed proximity of the eyes, without complete fusion. C- Others had complete eyes fusion. D - Embryo with only one eye.

Figure 4. Higher percentage of embryos is affected when injected with FRZC18. The percentage of embryos and the construct injected are represented in y and x-axis, respectively. Affected embryos are shown in black bars and normal embryos in dashed bars. A higher percentage of affected embryos is observed in FRZC18 injected group, when compared with no-injected and FRZC18-mut groups.

Figure 5. In situ hybridization for MyoD of zebrafish embryos at 10-12 somites stage. A- No-injected embryo. B- Embryos injected with 0.3ng of FRZC18-mut do not show notochord or somites defects. C-F- Embryos injected with 0.3ng of FRZC18 showed different severity defects. In C,somites are not completely formed. In D, notochord is broadened and somites have incompletely formation. In E, the notochord is very broadened and somites have defects in formation. F. The more severe defect in the embryo, with just a sign of notochord and somites, even though notochord seems to be broadened.

Figure 6. Western blots to understand proteins affected by FRZC18 antagonism in wnt signaling. No differences in ß-catenin or JNK were observed in the compartments analyzed between FRZC18 and EV (Empty Vector) transfections.

Capítulo V- Collagen XVIII cleaved frizzled fragment inhibits non-canonical Wnt signaling

Figure 1.

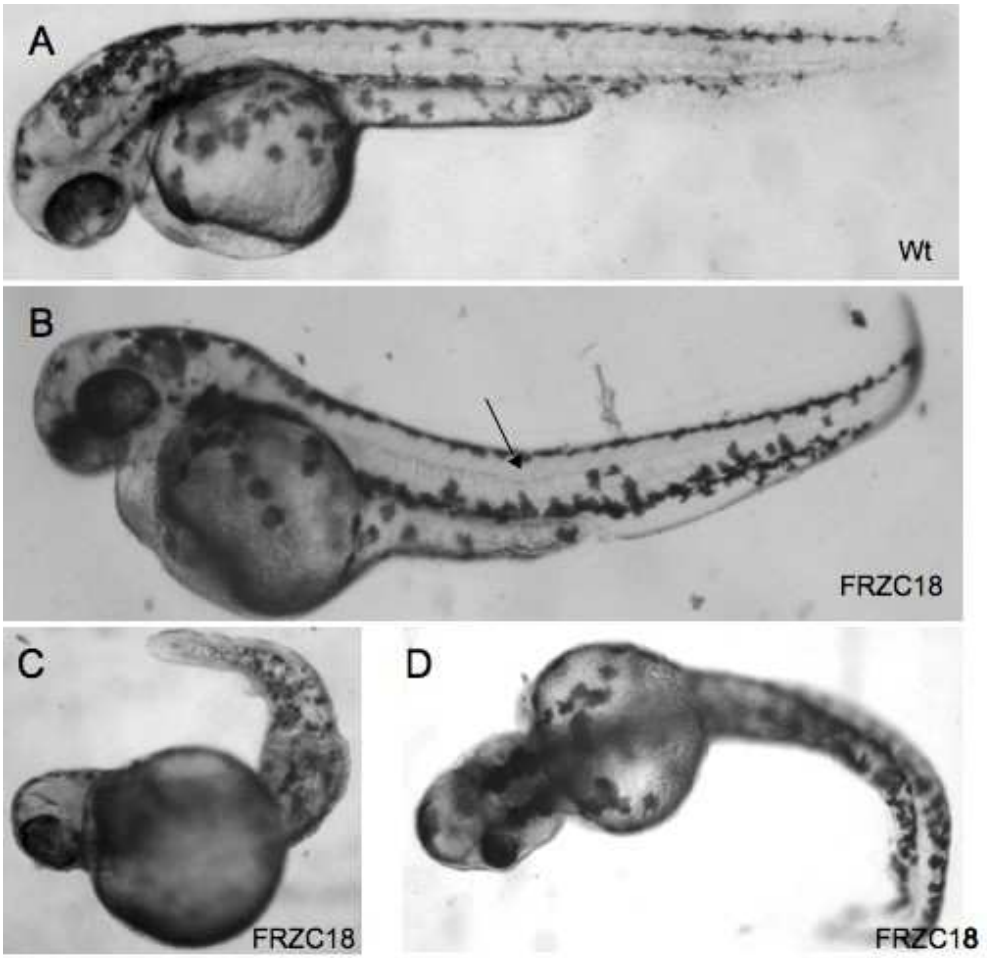


Figure 2.

Figure 3.

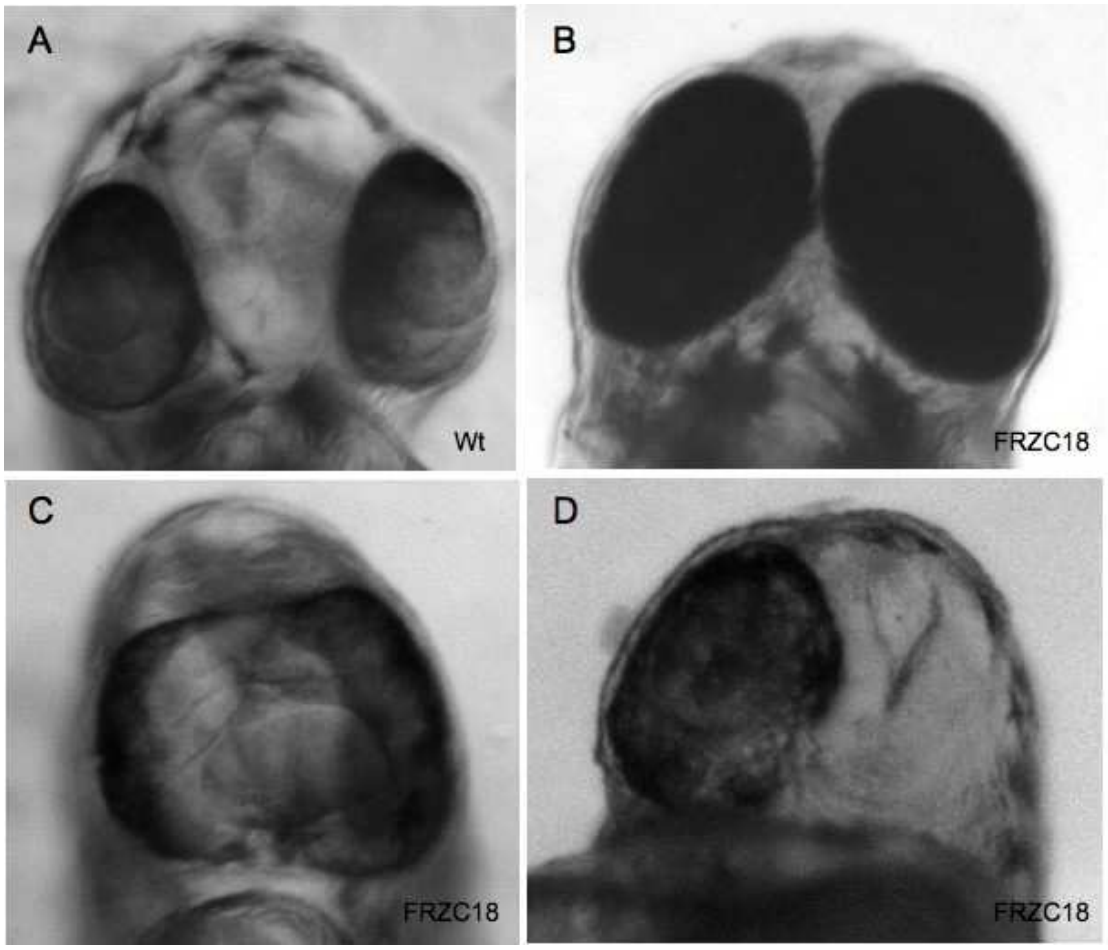
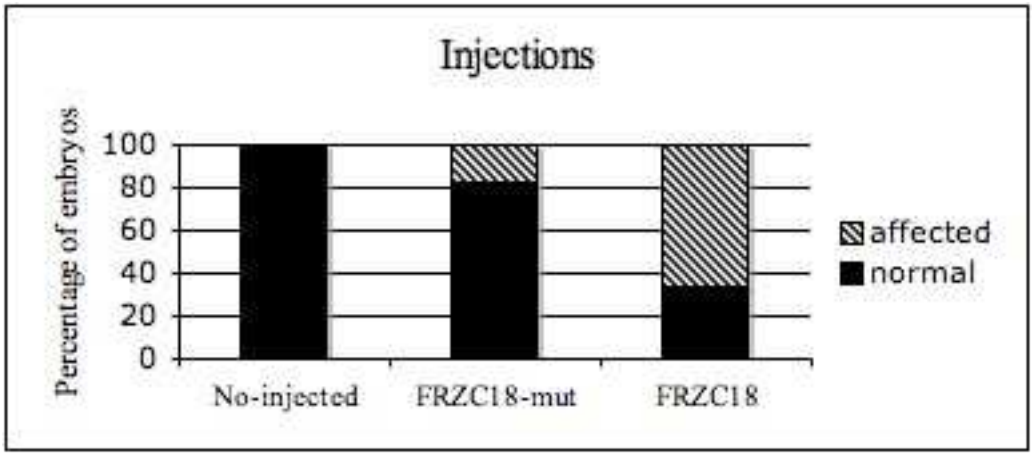


Figure 4.



Capítulo V- Collagen XVIII cleaved frizzled fragment inhibits non-canonical Wnt signaling

Figure 5.

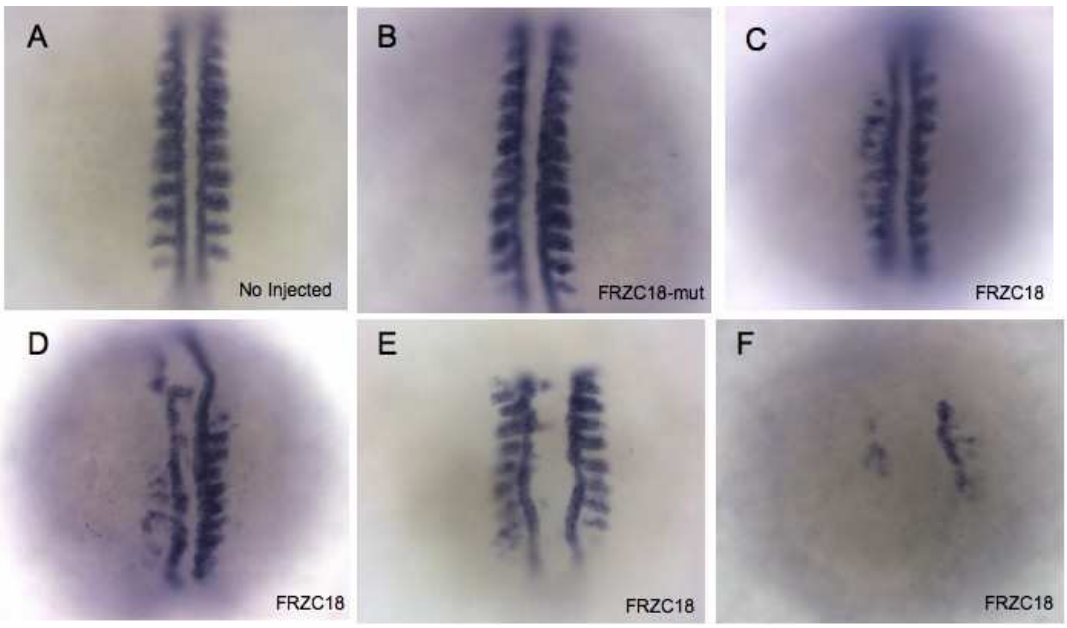
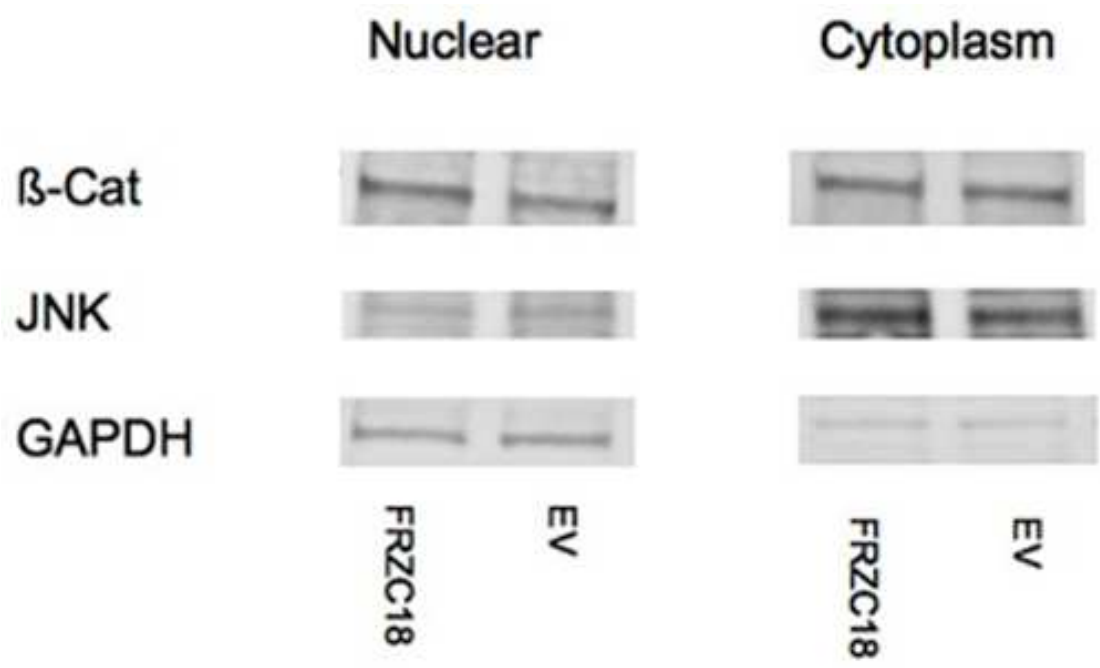


Figure 6.



Capítulo VI

6. Discussão

O colágeno XVIII é um importante componente da maioria das membranas basais do corpo e apresenta um complexo modelo de expressão, com participação em diferentes tecidos e órgãos durante o desenvolvimento embrionário e na fase adulta (Halfter, Dong *et al.*, 1998; Elamaa, Peterson *et al.*, 2002; Kawashima, Watanabe *et al.*, 2003). O *COL18A1* apresenta dois promotores que dirigem a transcrição de três isoformas distintas com picos de expressões em diferentes tecidos. Mutações que levam à perda de função do gene *COL18A1* causam a Síndrome de Knobloch, cuja as características fenotipicas marcantes envolvem problemas oculares, como alta miopia e descolamento de retina e também a presença de encefalocele occipital (Sertie, Sossi *et al.*, 2000; Suzuki, Sertie *et al.*, 2002). A endostatina como produto proteolítico do colágeno XVIII é um potente inibidor de angiogenese e tem sido investigada para utilização em terapias gênicas (O'reilly, Boehm *et al.*, 1997; Hong, Lee *et al.*, 2004). O colágeno XVIII também foi recentemente vinculado ao processo de adipogenese com aumento de expressão do mesmo durante a diferenciação adipogênica e predisposição em patogêneses como diabetes tipo 2 e hepatocarcinoma (Musso, Rehn *et al.*, 2001; Errera, Canani *et al.*, 2008). A participação deste colágeno como um antagonista na via de Wnt através de fragmentos proteolíticos, endostatina e Frizzled, é uma questão recente e molecularmente interessante (Hanai, Gloy *et al.*, 2002; Quelard, Lavergne *et al.*, 2008).

Visto o importante papel do *COL18A1* e inúmeras questões pertinentes a sua complexa função e distribuição tecidual, uma melhor elucidação da regulação deste gene e sua participação funcional na via de Wnt são extremamente relevantes. Desta forma, neste trabalho procuramos entender a participação *in vivo* do FRZC18 e a regulação transcricional do *COL18A1,* estudando não apenas as regiões basais de regulação transcricional, mas investigando também seqüências reguladoras de caráter genômico, os enhancers distais.

O promotor 1 *COL18A1,* até então ainda não caracterizado, apresenta uma região mínima capaz de promover a transcrição basal de luciferase entre -103 e -275pb do sitio de inicio da transcrição. Esta região, ou promotor basal, se assemelha muito com a região correspondente do promotor 2 do *COL18A1* (Armelin-Correa, Lin et al., 2005). Ambas seqüências são ricas em sítios de ligação para o fator de

transcrição Sp1 e são ausentes de TATA ou CAAT boxes, característica de genes "housekeeping" (CAP IV). A semelhança entre os promotores basais poderia ajudar a entender a sobreposição de expressão basal referente a cada isoforma. Além do promotor basal, identificamos três seqüências que aumentam a atividade do promotor, PE1-3. A região PE3 mostrou-se funcional em testes *in vivo*, confirmando a região como um enhancer proximal. Apesar de PE3 induzir relativamente uma maior atividade em células HepG2 e indicando regulação tecido-específico, expressão em fígado não foi observada *in vivo*. Análises de conservação entre a seqüência promotora ortóloga de camundongos seguidas de predição para fatores de transcrição, delimitamos alguns fatores de transcrição que poderiam desempenhar papel regulador nos PEs, entre eles TCF/LEF, PPRA-gamma, AP-2 e NF3 que estão envolvidos em via de Wnt, adipogenese, formação de crista neural e alta expressão em fígado, respectivamente. Mostramos que os fatores de transcrição TCF/LEF podem se ligar a região promotora do *COL18A1,* e que aumento de ß-catenina leva a redução de atividade dos promotores do *COL18A1* e redução de sua expressão. Dessa forma, essa relação inversa entre colágeno XVIII e ß-catenina nos ajuda a melhor entender e especular um pouco sobre a regulação molecular em processos como a adipogenese e também o câncer. Durante a adipogenese é observado uma redução na expressão de Wnt1 (Ross, Hemati *et al.*, 2000), um ativador de ß-catenina. E concomitantemente, observa-se um aumento de expressão de *COL18A1* (CAP IV). Este fato corrobora a regulação negativa do *COL18A1* pela ß-catenina. Alem disso, diferenças em níveis de expressão de *COL18A1* está diretamente ligada a níveis de endostatina e FRZC18, ambos antatonistas de ß-catenina. Em cânceres onde a via de Wnt está comprometida com aumento exacerbado de ß-catenina, poderíamos imaginar uma redução de *COL18A1* e consequentemente, redução de endostatina e aumento de crescimento tumoral. Dessa forma, o estabelecimento de ligação entre o colágeno XVIII como um gene alvo regulado pela via de Wnt ajuda a entender processos biológicos importantes e complexos em que relativamente pouco se conhece molecularmente.

O modelo complexo de expressão do *COL18A1* estende-se desde nemátodos, até humanos. O colágeno XVIII desempenha papel em migração neuronal em nemátodos e em migração de axônios motores em zebrafish também foi descrito (Ackley, Crew *et al.*, 2001; Schneider e Granato, 2006). Sua expressão em galinhas muito se assemelha à encontrada em zebrafish e em camundongos (Halfter, Dong *et al.*, 1998; Haftek, Morvan-Dubois *et al.*, 2003). Nós descrevemos expressão de

COL18A1 em tecidos embrionários de zebrafish que ainda não haviam sido ralatados, como fígado, cartilagem e vasos sanguíneos (CAP III). Em embriões de camundongos o *COL18A1* apresenta o mesmo padrão de expressão observado em galinhas e em zebrafish, o que indica o complexo modelo de expressão do *COL18A1* e conservação de regulação da expressão desse gene ao longo da evolução. Nós identificamos quatro distintos elementos regulatórios em cis do *COL18A1* (CAP III) que somando-se dirigem expressão em tecidos de zebrafish representando quase completamente a expressão endógeno deste gene, incluindo ducto pronefrico, fígado, retina, vasos sanguíneos, cartilagem, intestino, notocorda, cérebros anterior, mediano e posterior, arcos faringeais e vesícula ótica. A expressão dirigida por cada elemento sugere regulação isoforma-específica, já que CNSv -97.8/ CNSp -96.9 leva expressão em ducto pronefrico enquanto que CNSv +64.8 em fígado, onde as isoformas curtas e media/longa são altamente expressas, respectivamente. CNSp +47.8 dirige expressão em vasos sangüíneos e pode ser uma seqüência alvo para estudos relacionados com aumento de expressão de endostatina. Os resultados também indicam que os enhancers responsáveis por regulação tecido-específico estão dispostos distintamente do promotor do gene.

As construções testadas funcionalmente como enhancers foram selecionadas de acordo com conservação entre mamíferos e não apresentaram conservação com espécies distantes evolutivamente, como fugo, stickeback e zebrafish. Porém, interessantemente, as seqüências funcionaram em zebrafish. Submetemos as regiões já identificadas funcionalmente em analises *post hoc* utilizando uma ferramenta com maior sensibilidade porém, incapaz de calcular significância ou a conservação da vizinhança e que geralmente fornece maior número de falso-positivos. Identificamos conservação com zebrafish nas seqüências das regiões CNS -96.9 e CNSp +47.8. Testamos funcionalmente a seqüência de zebrafish ortóloga a CNSp +47.8, e obtivemos expressão mosaica em vasos sanguíneos, resultados que se sobrepõe à respectiva construção humana. Este resultado indica que as regiões funcionais sem conservação detectada com primeira analise, ainda apresentam seqüências detectadas por fatores de transcrição comuns e conservados entre zebrafish e mamíferos e podem funcionar em zebrafish, complementando também trabalho posterior realizado por Fisher e colaboradores (Fisher, Grice, Vinton, Bessling e Mccallion, 2006; Fisher, Grice, Vinton, Bessling, Urasaki *et al.*, 2006).

Alem de analizarmos a regulação transcricional do *COL18A1* identificando enhancers e caracterizando o promotor 1 do gene, também testamos *in vivo* o papel do domínio FRZC18. Através de superexpressão deste domínio em embriões de zebrafish, relacionamos o mesmo com a via não –canônica de Wnt. A via canônica, dependende de ß-catenina, é a mais conhecida e estudada e pouco se entende sobre a via não-canônica, porem alterações nas divergentes vias são claramente diferenciadas fenotipicamente em embriões de zebrafish. Defeitos na via canônica de Wnt em embriões de zebrafish, promovem duplicação aterior de eixo (Kelly, Erezyilmaz *et al.*, 1995). Enquanto que alteração na via não-canônica relaciona-se com encurtamento de eixo posterior e redução de distancia interocular nos embriões (Heisenberg, Tada *et al.*, 2000; Kilian, Mansukoski *et al.*, 2003; Ulrich, Concha *et al.*, 2003). Nossos resultados indicam participação de FRZC18 na via não-canônica de Wnt (CAP V), sendo que os embriões apresentaram características fenotipicas semalhantes a mutantes wnt11 e wnt5, slb e ppt, respectivamente. Isto implica participação em movimento de gastrulação durante a embriogenese. Observamos defeitos marcantes de gastrulação dos embriões, quando injetados com FRC18, incluindo enlargamento de notocorda e somitos. Interessantemente, não observamos alteração na via canônica de Wnt que envolve ß-catenina. Recentemente, o mesmo domínio foi relacionado funcionalmente como antagonista de ß-catenina em linhagens celulares de hepatocarcinoma (Quelard, Lavergne *et al.*, 2008). Dessa forma, alem da via de ß-catenina o colágeno XVIII também participa da via não- canônica, onde proteínas como ROCKII e JNK estão vinculadas e a reorganização do citoesqueleto é o resultado efetivo da via. Como participante de matriz extracelular em membranas basais o colágeno XVIII estruturalmente poderia levar a alterações no citoesqueleto através de interação com Wnts. A via não-canônica também está envolvida em migração neuronal. Foram relatados pacientes com síndrome de Knobloch com falta apenas das isoformas media e longa que apresentam defeitos em migração neuronal e casos de epilepsia (Passos-Bueno, Suzuki *et al.*, 2006). O envolvimento do *COL18A1* na via não-canônica poderia representar um mecanismo integrado na migração neuronal.

Procuramos esclarecer alguns dos elementos de regulação do *COL18A1* e entender funcionalmente o papel do FRZC18 na via de sinalização Wnt, porem muitas questões ainda permanecem obscuras. Nao ainda mecanismos de interação com

moléculas Wnt ou como o FRZC18 funciona como antagonista da via de sinalização Wnt ainda perma

Capítulo VII

7. Referências Bibliográficas

Abdollahi, A., P. Hahnfeldt, *et al.* Endostatin's antiangiogenic signaling network. Mol Cell, v.13, n.5, Mar 12, p.649-63. 2004.

Abe, N., Y. Muragaki, *et al.* Identification of a novel collagen chain represented by extensive interruptions in the triple-helical region. Biochem Biophys Res Commun, v.196, n.2, Oct 29, p.576-82. 1993.

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.

Ahituv, N., E. M. Rubin, *et al.* Exploiting human--fish genome comparisons for deciphering gene regulation. Hum Mol Genet, v.13 Spec No 2, Oct 1, p.R261-6. 2004.

Antonellis, A., J. L. Huynh, *et al.* Identification of neural crest and glial enhancers at the mouse Sox10 locus through transgenesis in zebrafish. PLoS Genet, v.4, n.9, Sep, p.e1000174. 2008.

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.

Baksh, D. e R. S. Tuan. Canonical and non-canonical Wnts differentially affect the development potential of primary isolate of human bone marrow mesenchymal stem cells. J Cell Physiol, v.212, n.3, Sep, p.817-26. 2007.

Bejerano, G., M. Pheasant, *et al.* Ultraconserved elements in the human genome. Science, v.304, n.5675, May 28, p.1321-5. 2004.

Bowers, R. R. e M. D. Lane. Wnt signaling and adipocyte lineage commitment. Cell Cycle, v.7, n.9, May 1, p.1191-6. 2008.

Carreira-Barbosa, F., M. L. Concha, *et al.* Prickle 1 regulates cell movements during gastrulation and neuronal migration in zebrafish. Development, v.130, n.17, Sep, p.4037-46. 2003.

Czeizel, A. E., P. Goblyos, *et al.* The second report of Knobloch syndrome. Am J Med Genet, v.42, n.6, Apr 1, p.777-9. 1992.

Dann, C. E., J. C. Hsieh, *et al.* Insights into Wnt binding and signalling from the structures of two Frizzled cysteine-rich domains. Nature, v.412, n.6842, Jul 5, p.86-90. 2001.

Doucas, H., G. Garcea, *et al.* Changes in the Wnt signalling pathway in gastrointestinal cancers and their prognostic significance. Eur J Cancer, v.41, n.3, Feb, p.365-79. 2005.

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.

Elnitski, L., C. Riemer*, et al.* MultiPipMaker: comparative alignment server for multiple DNA sequences. Curr Protoc Bioinformatics, v.Chapter 10, Apr, p.Unit10 4. 2005.

Endo, Y., V. Wolf*, et al.* Wnt-3a-dependent cell motility involves RhoA activation and is specifically regulated by dishevelled-2. J Biol Chem, v.280, n.1, Jan 7, p.777-86. 2005.

Errera, F. I., L. H. Canani*, et al.* COL18A1 is highly expressed during human adipocyte differentiation and the SNP c.1136C > T in its "frizzled" motif is associated with obesity in diabetes type 2 patients. An Acad Bras Cienc, v.80, n.1, Mar, p.167-77. 2008.

Fisher, S., E. A. Grice*, et al.* Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science, v.312, n.5771, Apr 14, p.276-9. 2006.

_____. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat Protoc, v.1, n.3, p.1297-305. 2006.

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.

Funayama, N., F. Fagotto*, et al.* Embryonic axis induction by the armadillo repeat domain of beta-catenin: evidence for intracellular signaling. J Cell Biol, v.128, n.5, Mar, p.959-68. 1995.

Gelse, K., E. Poschl*, et al.* Collagens--structure, function, and biosynthesis. Adv Drug Deliv Rev, v.55, n.12, Nov 28, p.1531-46. 2003.

Grice, E. A., E. S. Rochelle*, et al.* Evaluation of the RET regulatory landscape reveals the biological relevance of a HSCR-implicated enhancer. Hum Mol Genet, v.14, n.24, Dec 15, p.3837-45. 2005.

Habas, R., Y. Kato*, et al.* Wnt/Frizzled activation of Rho regulates vertebrate gastrulation and requires a novel Formin homology protein Daam1. Cell, v.107, n.7, Dec 28, p.843-54. 2001.

Haftek, Z., G. Morvan-Dubois*, et al.* Sequence and embryonic expression of collagen XVIII NC1 domain (endostatin) in the zebrafish. Gene Expr Patterns, v.3, n.3, Jun, p.351-4. 2003.

Halfter, W., S. Dong, *et al.* Collagen XVIII is a basement membrane heparan sulfate proteoglycan. J Biol Chem, v.273, n.39, Sep 25, p.25404-12. 1998.

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.

He, T. C., A. B. Sparks, *et al.* Identification of c-MYC as a target of the APC pathway. Science, v.281, n.5382, Sep 4, p.1509-12. 1998.

Heisenberg, C. P. e M. Tada. Zebrafish gastrulation movements: bridging cell and developmental biology. Semin Cell Dev Biol, v.13, n.6, Dec, p.471-9. 2002.

Heisenberg, C. P., M. Tada, *et al.* Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. Nature, v.405, n.6782, May 4, p.76-81. 2000.

Holmen, S. L., A. Salic, *et al.* A novel set of Wnt-Frizzled fusion proteins identifies receptor components that activate beta -catenin-dependent signaling. J Biol Chem, v.277, n.38, Sep 20, p.34727-35. 2002.

Hong, S. Y., M. H. Lee, *et al.* Adeno-associated virus mediated endostatin gene therapy in combination with topoisomerase inhibitor effectively controls liver tumor in mouse model. World J Gastroenterol, v.10, n.8, Apr 15, p.1191-7. 2004.

Hu, B., L. Kou, *et al.* Bifidobacterium longum as a delivery system of TRAIL and endostatin cooperates with chemotherapeutic drugs to inhibit hypoxic tumor growth. Cancer Gene Ther, Feb 20. 2009.

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.

Hurskainen, M., L. Eklund, *et al.* Abnormal maturation of the retinal vasculature in type XVIII collagen/endostatin deficient mice and changes in retinal glial cells due to lack of collagen types XV and XVIII. FASEB J, v.19, n.11, Sep, p.1564-6. 2005.

Iizasa, T., H. Chang, *et al.* Overexpression of collagen XVIII is associated with poor outcome and elevated levels of circulating serum endostatin in non-small cell lung cancer. Clin Cancer Res, v.10, n.16, Aug 15, p.5361-6. 2004.

Katoh, M. Comparative integromics on non-canonical WNT or planar cell polarity signaling molecules: transcriptional mechanism of PTK7 in colorectal cancer and that of SEMA6A in undifferentiated ES cells. Int J Mol Med, v.20, n.3, Sep, p.405-9. 2007.

Kawano, Y. e R. Kypta. Secreted antagonists of the Wnt signalling pathway. J Cell Sci, v.116, n.Pt 13, Jul 1, p.2627-34. 2003.

Kawashima, H., N. Watanabe*, et al.* Collagen XVIII, a basement membrane heparan sulfate proteoglycan, interacts with L-selectin and monocyte chemoattractant protein-1. J Biol Chem, v.278, n.15, Apr 11, p.13069-76. 2003.

Kelly, C., A. J. Chin*, et al.* Maternally controlled (beta)-catenin-mediated signaling is required for organizer formation in the zebrafish. Development, v.127, n.18, Sep, p.3899-911. 2000.

Kelly, G. M., D. F. Erezyilmaz*, et al.* Induction of a secondary embryonic axis in zebrafish occurs following the overexpression of beta-catenin. Mech Dev, v.53, n.2, Oct, p.261-73. 1995.

Kilian, B., H. Mansukoski*, et al.* The role of Ppt/Wnt5 in regulating cell shape and movement during zebrafish gastrulation. Mech Dev, v.120, n.4, Apr, p.467-76. 2003.

Kimmel, C. B., W. W. Ballard*, et al.* Stages of embryonic development of the zebrafish. Dev Dyn, v.203, n.3, Jul, p.253-310. 1995.

Kishida, S., H. Yamamoto*, et al.* Wnt-3a and Dvl induce neurite retraction by activating Rho-associated kinase. Mol Cell Biol, v.24, n.10, May, p.4487-501. 2004.

Kvist, A. P., A. Latvanlehto*, et al.* Complete exon-intron organization and chromosomal location of the gene for mouse type XIII collagen (col13a1) and comparison with its human homologue. Matrix Biol, v.18, n.3, Jun, p.261-74. 1999.

Lange, C., E. Mix*, et al.* Wnt signal pathways and neural stem cell differentiation. Neurodegener Dis, v.3, n.1-2, p.76-86. 2006.

Lele, Z., J. Bakkers*, et al.* Morpholino phenocopies of the swirl, snailhouse, somitabun, minifin, silberblick, and pipetail mutations. Genesis, v.30, n.3, Jul, p.190-4. 2001.

Lien, C. L., J. Mcanally*, et al.* Cardiac-specific activity of an Nkx2-5 enhancer requires an evolutionarily conserved Smad binding site. Dev Biol, v.244, n.2, Apr 15, p.257-66. 2002.

Logan, C. Y. e R. Nusse. The Wnt signaling pathway in development and disease. Annu Rev Cell Dev Biol, v.20, p.781-810. 2004.

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.

Mann, B., M. Gelos*, et al.* Target genes of beta-catenin-T cell-factor/lymphoid-enhancer-factor signaling in human colorectal carcinomas. Proc Natl Acad Sci U S A, v.96, n.4, Feb 16, p.1603-8. 1999.

Marlow, F., J. Topczewski*, et al.* Zebrafish Rho kinase 2 acts downstream of Wnt11 to mediate cell polarity and effective convergence and extension movements. Curr Biol, v.12, n.11, Jun 4, p.876-84. 2002.

Marneros, A. G. e B. R. Olsen. The role of collagen-derived proteolytic fragments in angiogenesis. Matrix Biol, v.20, n.5-6, Sep, p.337-45. 2001.

_____. Age-dependent iris abnormalities in collagen XVIII/endostatin deficient mice with similarities to human pigment dispersion syndrome. Invest Ophthalmol Vis Sci, v.44, n.6, Jun, p.2367-72. 2003.

_____. Physiological role of collagen XVIII and endostatin. FASEB J, v.19, n.7, May, p.716-28. 2005.

Mcgaughey, D. M., R. M. Vinton, *et al.* Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at phox2b. Genome Res, v.18, n.2, Feb, p.252-60. 2008.

Mi, K. e G. V. Johnson. Role of the intracellular domains of LRP5 and LRP6 in activating the Wnt canonical pathway. J Cell Biochem, v.95, n.2, May 15, p.328-38. 2005.

Montcouquiol, M., E. B. Crenshaw, 3rd, *et al.* Noncanonical Wnt signaling and neural polarity. Annu Rev Neurosci, v.29, p.363-86. 2006.

Muragaki, Y., S. Timmons, *et al.* Mouse Col18a1 is expressed in a tissue-specific manner as three alternative variants and is localized in basement membrane zones. Proc Natl Acad Sci U S A, v.92, n.19, Sep 12, p.8763-7. 1995.

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. 2001.

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. 2001.

Myllyharju, J. e K. I. Kivirikko. Collagens and collagen-related diseases. Ann Med, v.33, n.1, Feb, p.7-21. 2001.

Nobrega, M. A., I. Ovcharenko, *et al.* Scanning human gene deserts for long-range enhancers. Science, v.302, n.5644, Oct 17, p.413. 2003.

Nobrega, M. A. e L. A. Pennacchio. Comparative genomic analysis as a tool for biological discovery. J Physiol, v.554, n.Pt 1, Jan 1, p.31-9. 2004.

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.

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.

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. 1994.

Pashos, E. E., E. Kague*, et al.* Evaluation of cis-regulatory function in zebrafish. Brief Funct Genomic Proteomic, Sep 27. 2008.

Passos-Bueno, M. R., S. K. Marie*, et al.* Knobloch syndrome in a large Brazilian consanguineous family: confirmation of autosomal recessive inheritance. Am J Med Genet, v.52, n.2, Aug 15, p.170-3. 1994.

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.

Pederson, L., M. Ruan*, et al.* Regulation of bone formation by osteoclasts involves Wnt/BMP signaling and the chemokine sphingosine-1-phosphate. Proc Natl Acad Sci U S A, v.105, n.52, Dec 30, p.20764-9. 2008.

Pennacchio, L. A. Insights from human/mouse genome comparisons. Mamm Genome, v.14, n.7, Jul, p.429-36. 2003.

Ponti, G., L. Losi*, et al.* Wnt pathway, angiogenetic and hormonal markers in sporadic and familial adenomatous polyposis-associated juvenile nasopharyngeal angiofibromas (JNA). Appl Immunohistochem Mol Morphol, v.16, n.2, Mar, p.173-8. 2008.

Poulin, F., M. A. Nobrega*, et al.* In vivo characterization of a vertebrate ultraconserved enhancer. Genomics, v.85, n.6, Jun, p.774-81. 2005.

Prockop, D. J. e K. I. Kivirikko. Collagens: molecular biology, diseases, and potentials for therapy. Annu Rev Biochem, v.64, p.403-34. 1995.

Quelard, D., E. Lavergne*, et al.* A cryptic frizzled module in cell surface collagen 18 inhibits Wnt/beta-catenin signaling. PLoS ONE, v.3, n.4, p.e1878. 2008.

Rehn, M., E. Hintikka*, et al.* Primary structure of the alpha 1 chain of mouse type XVIII collagen, partial structure of the corresponding gene, and comparison of the alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen chain. J Biol Chem, v.269, n.19, May 13, p.13929-35. 1994.

_____. 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.

Retsky, M. W., W. J. Hrushesky, *et al.* Hypothesis: primary antiangiogenic method proposed to treat early stage breast cancer. BMC Cancer, v.9, p.7. 2009.

Ross, S. E., N. Hemati, *et al.* Inhibition of adipogenesis by Wnt signaling. Science, v.289, n.5481, Aug 11, p.950-3. 2000.

Rosso, S. B., D. Sussman, *et al.* Wnt signaling through Dishevelled, Rac and JNK regulates dendritic development. Nat Neurosci, v.8, n.1, Jan, p.34-42. 2005.

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. 1998.

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. 1998.

Schneider, V. A. e M. Granato. The myotomal diwanka (lh3) glycosyltransferase and type XVIII collagen are critical for motor growth cone migration. Neuron, v.50, n.5, Jun 1, p.683-95. 2006.

Schwartz, S., L. Elnitski, *et al.* MultiPipMaker and supporting tools: Alignments and analysis of multiple genomic DNA sequences. Nucleic Acids Res, v.31, n.13, Jul 1, p.3518-24. 2003.

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.

Shin, J. T., J. R. Priest, *et al.* Human-zebrafish non-coding conserved elements act in vivo to regulate transcription. Nucleic Acids Res, v.33, n.17, p.5437-45. 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.

Taniguchi, K., L. R. Roberts, *et al.* Mutational spectrum of beta-catenin, AXIN1, and AXIN2 in hepatocellular carcinomas and hepatoblastomas. Oncogene, v.21, n.31, Jul 18, p.4863-71. 2002.

The ENCODE (ENCyclopedia Of DNA Elements) Project. Science, v.306, n.5696, Oct 22, p.636-40. 2004.

Thisse, C. e B. Thisse. High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat Protoc, v.3, n.1, p.59-69. 2008.

Topczewski, J., D. S. Sepich, *et al.* The zebrafish glypican knypek controls cell polarity during gastrulation movements of convergent extension. Dev Cell, v.1, n.2, Aug, p.251-64. 2001.

Uchikawa, M., Y. Ishida, *et al.* Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. Dev Cell, v.4, n.4, Apr, p.509-19. 2003.

Ulrich, F., M. L. Concha, *et al.* Slb/Wnt11 controls hypoblast cell migration and morphogenesis at the onset of zebrafish gastrulation. Development, v.130, n.22, Nov, p.5375-84. 2003.

Ulrich, F., M. Krieg, *et al.* Wnt11 functions in gastrulation by controlling cell cohesion through Rab5c and E-cadherin. Dev Cell, v.9, n.4, Oct, p.555-64. 2005.

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.

Veeman, M. T., J. D. Axelrod, *et al.* A second canon. Functions and mechanisms of beta-catenin-independent Wnt signaling. Dev Cell, v.5, n.3, Sep, p.367-77. 2003.

Westerfield, M. The Zebrafish Book. A guide for the laboratory use of zebrafish (Danio rerio). Eugene: University of Oregon Press. 2000

Westfall, T. A., R. Brimeyer, *et al.* Wnt-5/pipetail functions in vertebrate axis formation as a negative regulator of Wnt/beta-catenin activity. J Cell Biol, v.162, n.5, Sep 1, p.889-98. 2003.

Woolfe, A., M. Goodson, *et al.* Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol, v.3, n.1, Jan, p.e7. 2005.

Wu, J., D. Ding, *et al.* Sustained delivery of endostatin improves the efficacy of therapy in Lewis lung cancer model. J Control Release, Nov 19. 2008.

Zhang, X., J. P. Gaspard, *et al.* Regulation of vascular endothelial growth factor by the Wnt and K-ras pathways in colonic neoplasia. Cancer Res, v.61, n.16, Aug 15, p.6050-4. 2001.

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.

Capítulo V- Collagen XVIII cleaved frizzled fragment inhibits non-canonical Wnt signaling

Anexos

Anexo 1

Molecular Vision 2009; 15:801-809 <http://www.molvis.org/molvis/v15/a82>
Received 7 November 2008 | Accepted 15 April 2009 | Published 23 April 2009

Novel pathogenic mutations and skin biopsy analysis in Knobloch syndrome

Oscar Suzuki,[1] Erika Kague,[1] Kelly Bagatini,[1] Hongmin Tu,[2] Ritva Heljasvaara,[2] Lorenza Carvalhaes,[3] Elisandra Gava,[3] Gisele de Oliveira,[3] Paulo Godoi,[4,5] Glaucius Oliva,[4] Gregory Kitten,[3] Taina Pihlajaniemi,[2] Maria-Rita Passos-Bueno[1]

[1]Centro de Estudos do Genoma Humano, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil; [2]Collagen Research Unit, Biocenter and Department of Medical Biochemistry and Molecular Biology, University of Oulu, Oulu, Finland; [3]Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; [4]Departamento de Física e Informática, Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, Brazil; [5]Departamento de Físico-Química, Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, Brazil

Purpose: To facilitate future diagnosis of Knobloch syndrome (KS) and better understand its etiology, we sought to identify not yet described *COL18A1* mutations in KS patients. In addition, we tested whether mutations in this gene lead to absence of the *COL18A1* gene product and attempted to better characterize the functional effect of a previously reported missense mutation.

Methods: Direct sequencing of *COL18A1* exons was performed in KS patients from four unrelated pedigrees. We used immunofluorescent histochemistry in skin biopsies to evaluate the presence of type XVIII collagen in four KS patients carrying two already described mutations: c.3277C>T, a nonsense mutation, and c.3601G>A, a missense mutation. Furthermore, we determined the binding properties of the mutated endostatin domain p.A1381T (c.3601G>A) to extracellular matrix proteins using ELISA and surface plasmon resonance assays.

Results: We identified four novel mutations in *COL18A1*, including a large deletion involving exon 41. Skin biopsies from KS patients revealed lack of type XVIII collagen in epithelial basement membranes and blood vessels. We also found a reduced affinity of p.A1381T endostatin to some extracellular matrix components.

Conclusions: *COL18A1* mutations involved in Knobloch syndrome have a distribution bias toward the coding exons of the C-terminal end. Large deletions must also be considered when point mutations are not identified in patients with characteristic KS phenotype. We report, for the first time, lack of type XVIII collagen in KS patients by immunofluorescent histochemistry in skin biopsy samples. As a final point, we suggest the employment of this technique as a preliminary and complementary test for diagnosis of KS in cases when mutation screening either does not detect mutations or reveals mutations of uncertain effect, such as the p.A1381T change.

Knobloch syndrome (KS; OMIM 267750) is an autosomal recessive disorder characterized by high myopia, macular abnormalities, vitreoretinal degeneration, retinal detachment, and occipital encephalocele [1-4]. The spectrum of clinical variability is not completely known due to the small number of cases reported in the literature [5]. Most KS cases are caused by null mutations in the *COL18A1* gene (chr21q22.3), which comprises 43 exons and transcribes three different isoforms by the use of two promoters and alternative splicing in the third exon [4,6]. The three encoded collagen XVIII proteins differ only by their signal peptides and by part of the N-terminal region of the NC11 domain. The short isoform (NC11–303) is transcribed from the first promoter, upstream of exon 1, and is encoded by exons 1, 2, 4–43. The intermediate (NC11–493) and long (NC11–728) forms are transcribed from the second promoter, upstream of exon 3, and they differ by use of an internal splice site within exon 3. Collagen XVIII is a component of basement membranes [7]; however, the abundance of its isoforms varies: the short isoform is predominant in most tissues, including heart, kidney, retina, and fetal brain, while the intermediate and long isoforms are highly expressed in the liver [3,4,6]. The C-terminal domain of type XVIII collagen can be cleaved to form endostatin, which functions as a potent angiogenesis inhibitor that influences endothelial cell proliferation, migration, apoptosis and tubulogenesis [8-11]. Endostatin binds to several extracellular matrix (ECM) components, including laminin-1, fibulin-1, fibulin-2, nidogen-2, perlecan, heparan sulfate, and fibronectin [12-14].

To date, 12 rare mutations have been described in KS patients with the typical clinical features of the disease [3-5, 15-17]. All mutations, with the exception of a missense

Correspondence to: Dr. Marita-Rita Passos-Bueno, Centro de Estudos do Genoma Humano, Instituto de Biociências, Universidade de São Paulo, Departamento de Genética e Biologia Evolutiva, Rua do Matão, 277, Sala 200, São Paulo, SP 05508-090, Brazil; Phone: 55-11-3091-9910; FAX: 55-11-3091-7419;
email: passos@ib.usp.br

change (p.A1381T; numbered according to GenBank cDNA AF018081.1) of still unclear functional effect on the protein [18,19], are predicted to create premature stop codons. These mutations possibly lead to a lack of the protein, even though the complete lack of collagen XVIII has still not been demonstrated. *COL18A1* mutations are distributed along the gene in regions common to all known isoforms, except for c. 12–2A>T, which only affects the short isoform. The description of pathogenic mutations in a larger number of KS patients and further functional analysis of the mutation p.A1381T may better characterize the spectrum of mutations along the *COL18A1* gene.

Screening the entire gene for mutations is still time consuming. We have previously shown that endostatin plasma measurements do not allow a precise diagnosis, as patients with null mutations had positive levels of this fragment [4], and no other methods to confirm KS diagnosis were tested. New strategies that might allow the screening of a larger number of patients, including those without the full phenotype of the syndrome, could help delineate the clinical variability.

In the present work we describe four new families with KS patients and the respective responsible mutations in *COL18A1*. Consistent with previous results, we find heterogeneity for mutation sites leading to this syndrome. We also assessed the immunohistochemical expression of type XVIII collagen in skin biopsies of KS patients carrying nonsense mutations to test the potential of this method as a screening tool for a larger number of patients. Finally, we evaluated in situ and in vitro effects of endostatin p.A1381T by testing skin biopsies for the presence of type XVIII collagen in KS patients carrying this mutation and evaluating the binding properties of the mutated endostatin to other ECM components.

METHODS

COL18A1 mutations analysis: Five KS patients were referred to the Centro de Estudos do Genoma Humano at the University of São Paulo. Two were siblings and the other three were not related. They presented with typical KS phenotypic characteristics: high myopia detected early in childhood (usually before 1 year of age) and occipital encephalocele, as described elsewhere [2,20]. No other alterations were observed.

We screened for mutations in the *COL18A1* gene by direct sequencing of exons and flanking intronic regions, as previously described [4]. All mutations were named according to the nomenclature suggested by den Dunnen and Antonarakis [21]. The numbering followed the short isoform (GenBank AF018082.1), except where indicated.

Skin biopsies: After informed consent was obtained, 5 mm skin punch biopsies were taken from the forearms of four other KS patients, who were from two unrelated families. Clinical and molecular analysis of these patients has been described in detail by Suzuki et al. and Kliemann et al. [4,18]. Control skin samples were obtained from two patients that were undergoing surgery and had no suspicion of KS. Informed consent was also obtained from these patients. This project was approved by the Ethical Committee of the Institute of Biosciences, University of São Paulo (São Paulo, Brazil).

Tissue samples were fixed according to a procedure described elsewhere [22]. Briefly, samples were immediately fixed and cryosubstituted in a −70 °C solution of 80% methanol and 20% dimethyl sulfoxide for 5–7 days, transferred to −20 °C for 1–2 days, and then brought to room temperature. Samples were rinsed 3X in absolute ethanol and embedded in Paraplast Plus following standard protocols. In addition to maintaining morphological structure, this method also appears to preserve the antigenicity to a higher degree than aldehyde fixatives.

Immunofluorescent histochemistry: Xylene was used to dewax 7 µm-thick sections. These were then rehydrated through a graded series of ethanol into phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4). Next, 1 mg/ml testicular hyaluronidase (Sigma, St. Louis, MO) in 50 mM acetate buffer, pH 5, was added at room temperature for 30 min. This was followed by several washes in PBS. Blocking was achieved using 2% BSA (BSA) in PBS at room temperature for 1 h followed by an overnight incubation at 4 °C with primary antibodies diluted in PBS containing 0.1% BSA and 0.01% Tween-20. A rabbit polyclonal antibody against type XVIII collagen (QH48.18, “anti-all”) which recognizes all three isoforms of type XVIII collagen [23] was used at a dilution of 1:300. Rabbit anti-human type IV collagen polyclonal antibody (Rockland, Gilbertsville, PA), an antibody against type IV collagen, was diluted 1:400 before it was used as a basement membrane marker. After several rinses in PBS, sections were incubated for 1 h at room temperature with 1:300 Cy3-conjugated goat anti-rabbit IgG secondary antibody (Jackson Laboratory, Bar Harbor, ME). After several rinses in PBS, sections were mounted in a mixture of 10% 1.0 M Tris-HCl, pH 9.0 and 90% glycerol and analyzed using a laser scanning confocal microscope (Zeiss 510META; Carl Zeiss AG, Oberkochen, Germany).

The distribution patterns and levels of expression of types IV and XVIII collagens were analyzed using pseudocolor images made with the thermo lookup table in the Confocal Assistant 4.02 freeware program. This tool utilizes gray-scale pixel values to generate pseudocolor images that allow the discrimination of small differences in the level of immunofluorescence.

Molecular modeling: Human endostatin structure (1BNL) was obtained from the Protein Data Bank. the p.A1381T change was created and analyzed using GRASP [24] and Insight-II/Discover software (Molecular Simulations Inc., San Diego, CA).

Recombinant endostatin production: The coding region of the endostatin domain was cloned into the pET-15b vector (Novagen, Madison, WI). p.A1381T endostatin was produced by PCR-based site directed mutagenesis.

Endostatin production and purification were based on a procedure published elsewhere [25]. Briefly, recombinant human endostatin expression in *E. coli* was induced using 1 mM IPTG (isopropyl-beta-D-thiogalactopyranoside) for 3 h. Cells were isolated by centrifugation and lysed by freeze thaw cycle in GUMCAC-0-buffer, which contained 6 M guanidine-HCl, 0.5 M NaCl, and 20 mM Tris-HCl, pH 7.9. This was followed by sonication in the presence of 15 mM β-mercaptoethanol. Lysate was loaded onto a ProBond™ column (Invitrogen, Carlsbad, CA) that had been previously equilibrated with URMCAC-0 buffer, which contained 8 M urea, 0.5 NaCl, and 20 mM Tris-HCl, pH 7.9. Endostatin was eluted with a gradient concentration of imidazole (0 to 500 mM) and dialyzed against 4 M urea, 0.1 M NaCl, 1 mM reduced glutathione, 0.1 mM oxidized glutathione, 20 mM Tris-HCl pH 7.9 (16 h at 4 °C). This was followed by a dialysis against 1 M urea, 0.1 M NaCl, 0.1 mM reduced glutathione, 0.01 mM oxidized glutathione, 20 mM Tris-HCl pH 7.9 (6 h at 4 °C), and a final dialysis step against PBS pH 6.9 (16 h at 4 °C). Human endostatin was then loaded onto a HiTrap-SP column (Amersham Biosciences, Uppsala, Sweden) and eluted using a NaCl gradient (0 to 1.5 M) in PBS pH 6.9. Buffer was exchanged by dialysis against 0.1 M NaCl and 20 mM Tris-HCl (pH 7.4) for 16 h at 4 °C.

Binding assays: Surface plasmon resonance (SPR) assays were performed with a BIAcore® 3000 instrument (Biacore AB, Uppsala, Sweden). Extracellular matrix proteins, the laminin-1-nidogen-1 complex, perlecan, and fibulin were immobilized on CM5 sensor chips (Biacore AB) as described previously [26]. Mouse laminin-1-nidogen-1 complex and mouse perlecan were prepared from the mouse Engelbreth-Holm-Swarm tumor [27]. The recombinant microfibrillar component fibulin-1 was prepared following the procedure published by Sasaki et al. [28]. Binding assays were performed in triplicates in 0.02 M Tris-HCl, 0.11 M NaCl containing 0.05% P-20 surfactant (Biacore AB) at a flow rate of 20 µl/min. The association phase was monitored for 3 min, and the dissociation curve was recorded for 10 min. The bulk effects were subtracted using the reference control surfaces. The chips were regenerated by treatment with 1 M NaCl for 30 s. For the calculation of kinetics constants, the sensorgrams at concentrations of 0, 10, 50, 200, and 400 nM were fitted globally to the 1:1 Langmuir model with BIAevaluation software version 3.1.

ELISA assays were based on a procedure used by Rehn et al. [25]. Laminin, type IV collagen, nidogen-1, fibulin-1, and 10 µg/ml heparin-BSA were coated onto the surface of microtiter wells at 4 °C overnight. All the other steps were performed at room temperature. The wells were blocked with 5% nonfat milk in 0.05 M Tris-HCl, pH 7.4, 0.11 M NaCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$ (TBS-Ca/Mg) for 1 h and then washed with 0.05% Tween-20 in TBS-Ca/Mg, and incubated for 3 h with endostatin as a soluble ligand diluted in 5% nonfat milk TBS-Ca/Mg. After thoroughly washing, the samples were incubated for 1 h with the antibody HES.6 against the human endostatin domain [29] diluted in TBS-Ca/Mg-5% milk, followed by washing with 0.05% Tween-20 in TBS-Ca/Mg, and incubation with the secondary antibody conjugated with horseradish peroxidase. Endostatin bound to the immobilized proteins was detected by adding 5-aminosalicylic acid (Sigma) in the presence of 0.01% H_2O_2. Detection was performed at 490 nm.

RESULTS

Identification of new mutations in COL18A1 leading to KS: Mutation screening of the *COL18A1* gene was performed in five KS patients, who were from four unrelated families (KS14–17). All the patients presented with congenital high myopia and occipital encephalocele, typical KS phenotypic characteristics. Three of the families are Brazilian (KS14, KS16, and KS17) and one is North American (KS15).

We identified four mutations not previously described: one insertion, two out-of-frame deletions in exons coding the C-terminal region of the protein, and a splice site mutation. All of these possibly lead to premature stop codons (Table 1). The mutation that occurs at the acceptor splice site of intron 7, c.929–2A>G, involves a highly conserved base and is predicted to disrupt splicing according to computational analysis (NetGene2 version 2.4 [30]; NNSPLICE version 0.9 [31]). We have detected a mutation in only one of the alleles in families KS14 and KS15, and these patients may be compound heterozygotes. Even though the KS16 patient was initially identified as a homozygous carrier of the mutation c.3514_3515delCT, we found only her mother harbors this mutation. Paternity was confirmed with the analysis of five microsatellite markers (D22S944, S1623, S1638, S1648, S1709). This patient also appeared homozygous for another mutation in this exon (c.3570G>A), which is present in the mother in heterozygosity but it is not present in the father, who carries only the G allele of this SNP. Therefore, we concluded that the father carries a deletion of at least exon 41 and the patient is actually hemizygous for the c.3514_3515delCT mutation and is a compound heterozygote for mutations in the *COL18A1* gene.

p.A1381T mutation leads to a weaker binding affinity to some of the ECM proteins: To better understand if the rare p.A1381T endostatin mutation has the potential to disrupt its function, we performed ELISA and SPR assays to evaluate the fragment's ability to bind to other ECM proteins. Our molecular modeling analysis of the p.A1381T endostatin suggested that it was possible to accommodate the threonine side chain in this position (Figure 1). ELISA assays showed

Table 1. Mutations identified in KS patients.

Patient	Mutation	Region	Affected Isoforms	Description of mutation
KS1	c.12–2A>T (homozygous)	Intron 1	Short isoform	[3]
KS3	c.2969–2978delCAGGGCCCCC (maternal)	Exon 36	All isoforms	[4]
	c.3514–3515delCT (paternal)	Exon 41		
KS4	c.1238–1239insA (maternal)	Exon 10	All isoforms	[4]
	c.3514–3513delCT (paternal)	Exon 41		
KS5	c.3514–3515delCT (maternal)	Exon 41	All isoforms	[4]
	c.2105delT (paternal)	Exon 23		
KS8	c.12–2A>T (homozygous)	Intron 1	Short isoform	[4]
KS9	c.3277C>T (homozygous)	Exon 40	All isoforms	[4]
KS10	c.2416C>T (homozygous)	Exon 18	All isoforms	[5], [17]
KS11	c.3769G>A (p.D1437N *) (maternal)	Exon 42	All isoforms	[15]
	c.2823_2824insC (paternal)	Exon 35		
KS12	c.3601G>A (p.A1381T *) (homozygous)	Exon 41	All isoforms	[18]
KS13	c.3544+3A>C (homozygous)	Intron 36	All isoforms	[16]
KS14	c.2673_2674insC (heterozygous)	Exon 33	All isoforms	Novel mutation
KS15	c.2824_2831delGGCCCCCC (heterozygous)	Exon 35	All isoforms	Novel mutation
KS16	c.3514–3515delCT (maternal)	Exon 41	All isoforms	Novel mutation
	exon 41 deleted (paternal)			Novel mutation
KS17	c.2673_2674insC (maternal)	Exon 33	All isoforms	Novel mutation
	c.929–2A>G (paternal)	Intron 7		Novel mutation

List of published COL18A1 mutations identified in Knobloch syndrome patients and the exon numbers of the mutation sites. Asterisk represents position based on the NC11–493 isoform (AF018081.1).

the binding properties of the p.A1381T endostatin to be similar to the wild type endostatin, without any significant differences between them. SPR experiments showed that the p.A1381T mutation leads to a somewhat weaker binding affinity to some of the ECM proteins that were tested (Figure 2; Table 2). The differences between wild type and mutant endostatin binding to laminin-1-nidogen-1 complex and to fibulin-1 were small but significant (p=0.0017 and p=0.0011, respectively). Perlecan binding was weak for both p.A1381T and wild type endostatins and the difference was not significant.

Skin biopsies from KS patients do not show COL18A1 expression: Null mutations that affect all type XVIII collagen isoforms must lead to complete lack of the protein. To confirm the absence of type XVIII collagen and to evaluate the possibility of distinguishing KS patients from normal individuals by the analysis of type XVIII collagen in skin biopsies, we performed fluorescent immunohistochemical staining for type XVIII collagen in skin biopsies from two KS individuals carrying the previously described nonsense mutation c.3277C>T. We observed a complete lack of type XVIII collagen expression in the tissue sections, while control samples displayed high levels of staining in the epithelial and endothelial basement membranes (Figure 3). The lack of expression of type XVIII collagen was especially evident in pseudocolor, thermo images in which even minor amounts of immunofluorescently labeled material can be visualized. Although another basement membrane protein, type IV collagen, showed a similar distribution pattern in both control and KS groups, its immunostaining intensity in KS samples was noticeably lower (Figure 3).

To understand the cause of KS in individuals carrying the change p.A1381T we also performed immunohistochemistry from skin biopsies of two patients. Surprisingly, they also showed lack of type XVIII collagen staining (Figure 3).

DISCUSSION

KS is thought to be a rare condition: only 42 patients from 15 unrelated families have been described [2,4,5,16-18,20, 32-36]. Among these families, only 12 different *COL18A1* mutations have been identified [3-5,15-17]. Although some clustering has been observed in exon 41, there is heterogeneity of mutation site distribution. Here we describe four unrelated families harboring five different mutations (Table 1); all of the patients are compound heterozygotes. KS16 represents an unusual case, where the chromosome inherited from the father has an undetermined deletion in *COL18A1* that includes exon 41. This mechanism may elucidate cases already reported where *COL18A1* mutations were not detected [4]. Sequencing the 43 exons of this gene is laborious for small study centers to confirm the KS diagnosis, so the identification of biases in mutation distribution could help in the molecular diagnosis. Mutations identified so far indicate that exons 30 through 42 of *COL18A1* are more frequently mutated, as 5 (83.3%) out of 6 mutations here identified were located in these exons, as well as 9 (64.3%) out of 14 previously identified mutations. Interestingly, 5 (26.3%) out of 19 mutations previously identified are located in exon 41, pointing to this exon as the most frequently affected. For this reason, we suggest that

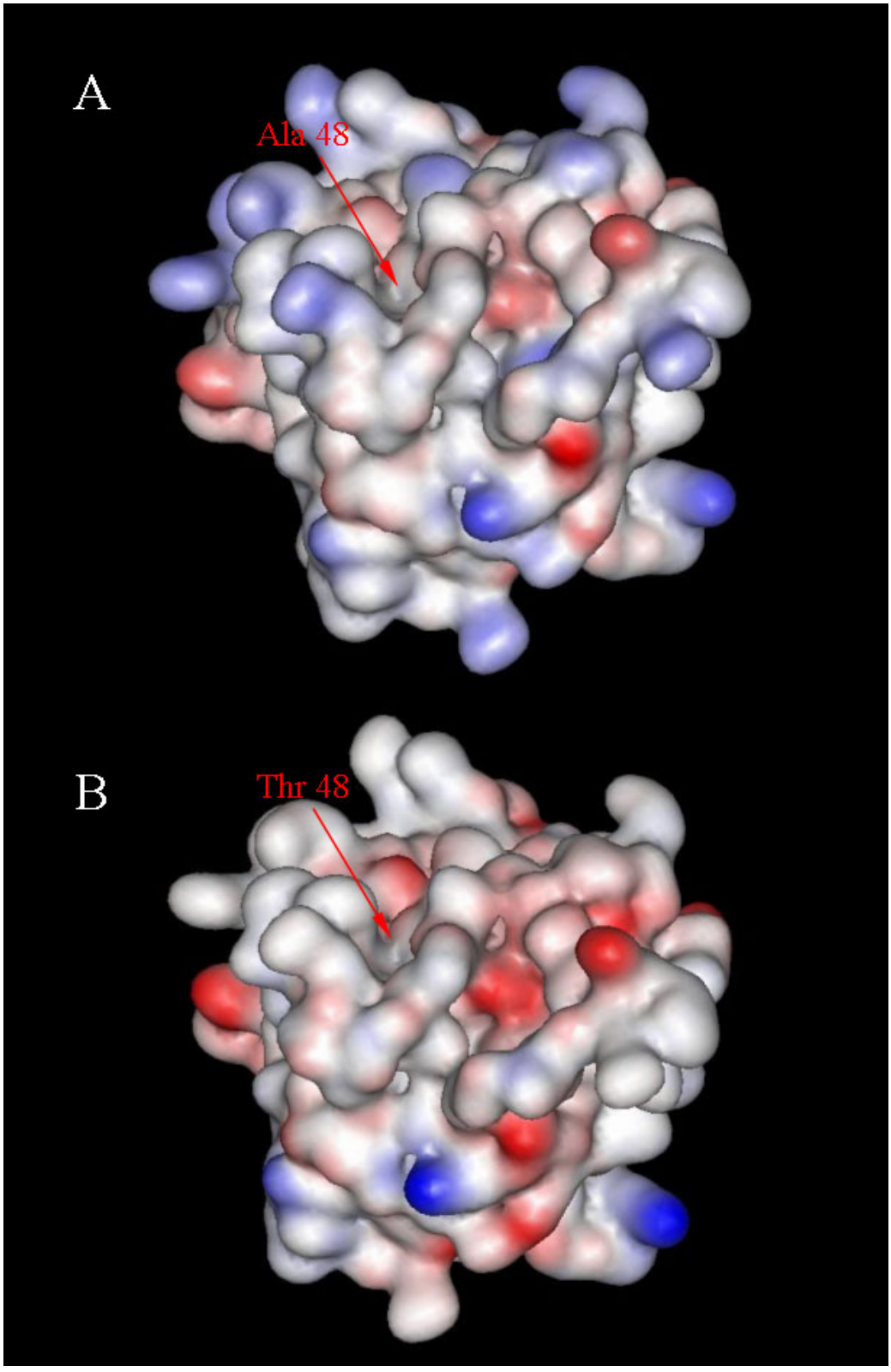


Figure 1. Molecular modeling of wild type and p.A1381T endostatins. Analysis of the electrostatic surface of wild type (**A**) and p.A1381T (**B**) endostatins. Areas shown in red represent a negative potential while the blue areas are positive. The position 48 of the endostatin domain (Protein Data Bank entry 1BNL) corresponds to the position 1381 of the NC11–493 isoform (GenBank AF018081.1).

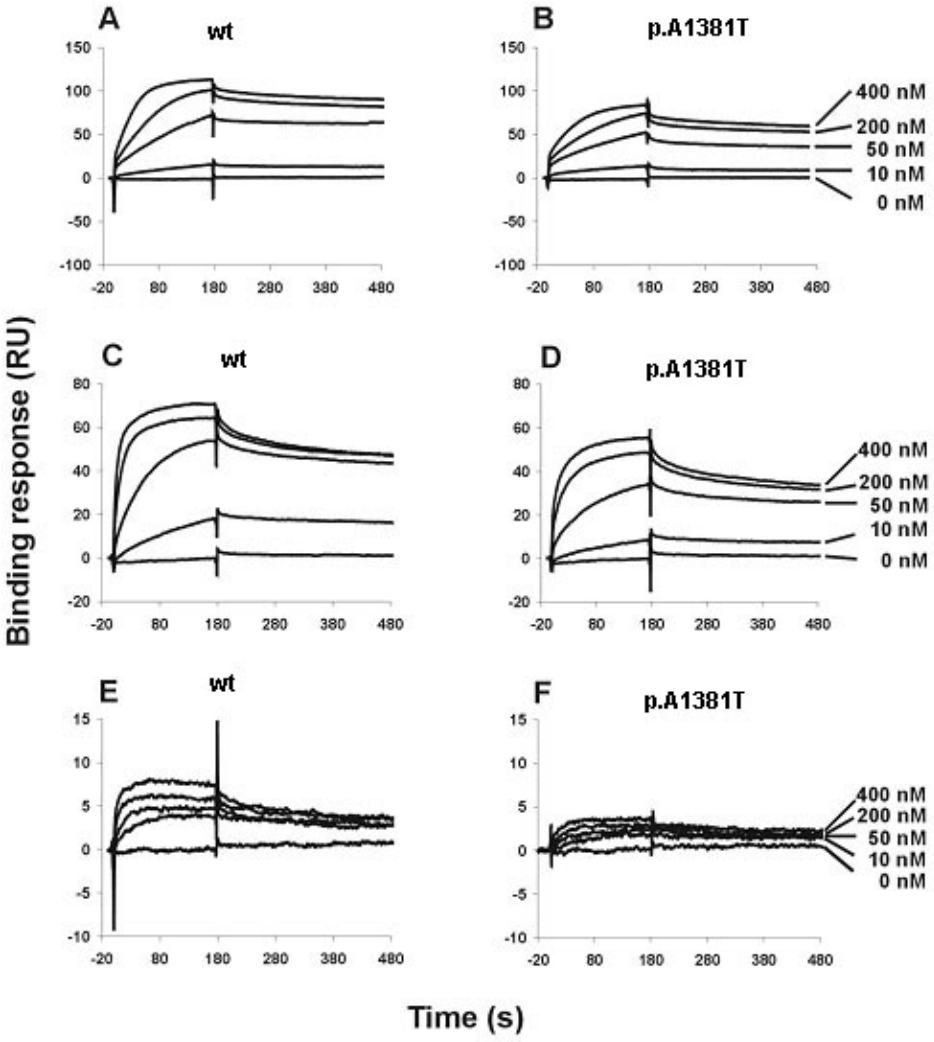


Figure 2. SPR sensorgrams of endostatin interactions with ECM components. Endostatin variants (wild type, mutant p.A1381T) were diluted to a concentration series in 0.02 M Tris-HCl, 0.11 M NaCl containing 0.05% P-20, and then injected into the sensor chips immobilized with laminin-1-nidogen-1 complex (**A**=wt, **B**=p.A1381T), fibulin (**C**=wt, **D**=p.A1381T), and perlecan (**E**=wt, **F**=p.A1381T) at 25 °C with a flow rate of 20 µl/min. Sensorgrams show binding of various concentrations of endostatin to the coated sensor surfaces. The association curves were monitored for 3 min, and the dissociation phases were recorded for 10 min but presented for 5 min. All the kinetics studies were performed three times independently at concentrations of 0–400 nM, and the data were analyzed with BIAevaluation software version 3.1 using the 1:1 Langmuir binding model.

Table 2. Binding affinities (K_D) of wild type and mutant endostatins to ECM components

Sample	Laminin-1-nidogen-1 complex	Fibulin	Perlecan
Wild type	3.48±0.24 nM	3.10±0.39 nM	5.53±1.53 nM
p.A1381T	6.17±0.27 nM	6.95±0.25 nM	5.19±0.92 nM

The kinetics constants were obtained by fitting the sensorgrams to the 1:1 Langmuir model.

mutation screening in KS patients should start with these exons. It is possible that KS is less rare, and the spectrum of clinical variability is wider than predicted.

To find an additional way to detect type XVIII collagen disruption in KS patients and gain a deeper understanding of the KS etiology, we performed protein measurements in situ and characterization of a mutation with an unknown effect in vitro. We evaluated for the first time the immunohistochemical presence of type XVIII collagen in KS patients. We have analyzed the distribution pattern of type XVIII collagen in skin biopsies of four KS patients: two siblings with the known pathogenic nonsense mutation c. 3277C>T and two other siblings with the missense endostatin mutation c.3601G>A (p.A1381T). As expected for carriers of null mutations, we were unable to detect type XVIII collagen in any of the tested KS patients, including those with typical KS clinical features that harbor the mutation p.A1381T, of uncertain effect, confirming that the primary molecular defect is in collagen XVIII. Interestingly, the KS samples also show a lower expression level of type IV collagen when compared with the control samples. Since endostatin/collagen XVIII binds to several ECM components [12-14], it is possible that

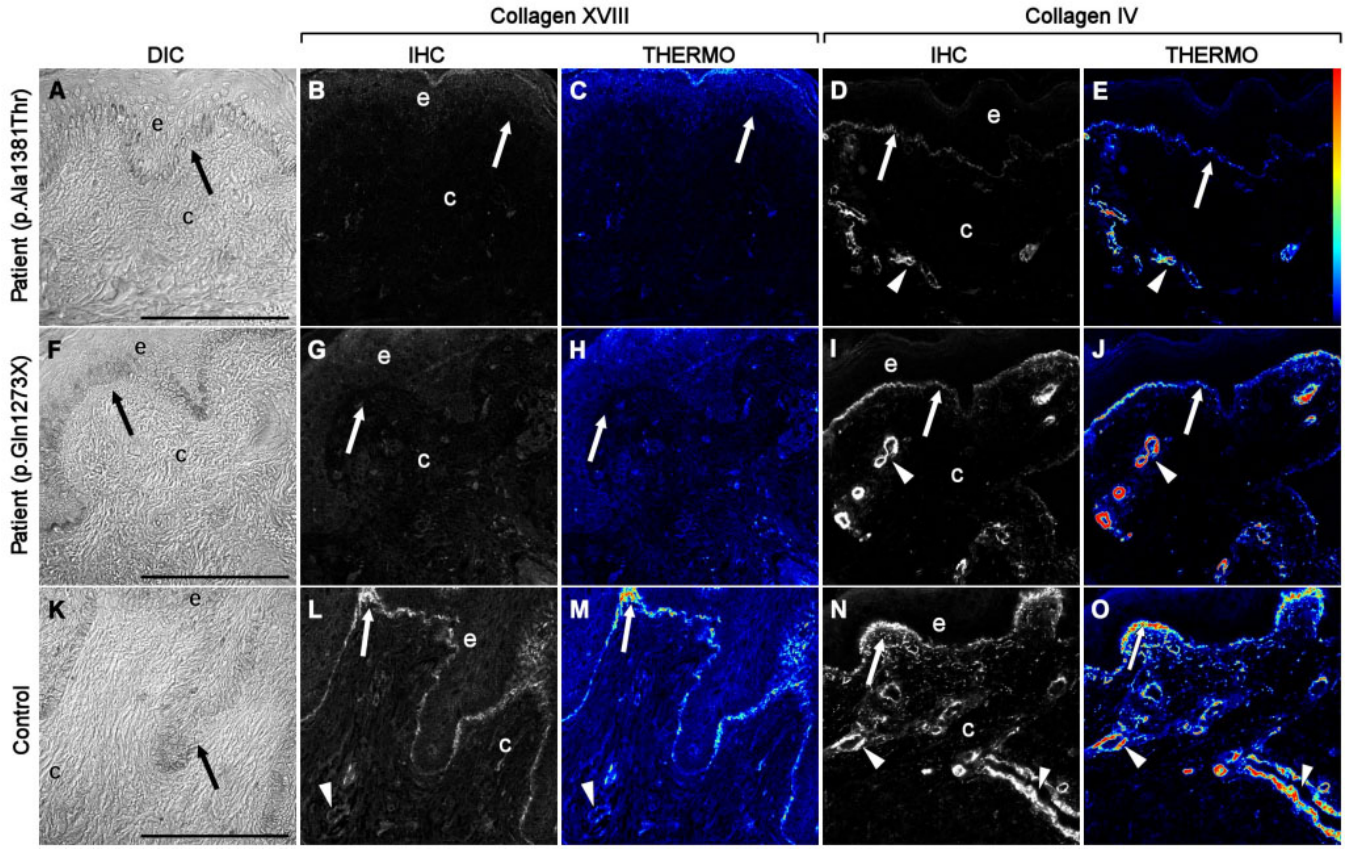


Figure 3. Immunofluorescent localization of type XVIII collagen in KS and control skin samples. Differential interference contrast (DIC) images (**A, F, K**) show that the epithelial and connective tissues obtained from skin biopsies were intact and well preserved after being prepared using cryofixation and cryosubstitution techniques. Immunolocalization analyses of KS patients carrying the p.A1381T change (**B, C**) and the nonsense mutation c.3277C>T (NC11–493 isoform position p.G1273X, GenBank AF018081.1; **G, H**) were negative for expression of collagen XVIII as compared to controls (**L, M**). Pseudocolor, thermo images showed that collagen XVIII expression was present in control samples (**M**) but undetectable in KS samples (**C, H**). Although the distribution pattern of type IV collagen was similar in both control (**N, O**), and KS (**D, E, I, J**) groups, the immunostaining intensity was noticeably lower in KS samples. The thermo color bar to the right of panel **E** indicates immunofluorescence staining intensity: red, higher levels; blue/black, lower levels. Arrows point to basement membrane, and arrowheads mark blood vessels. Scale bars equal 100 µm.

a lack of this protein may lead to changes in the overall organization and stability of the ECM, especially in the basement membrane where it is normally expressed.

We also tested the p.A1381T endostatin for binding affinity to a selection of ECM proteins. Even though the molecular modeling of the protein did not predict a significant modification, p.A1381T has a weaker affinity to some of the tested ECM components when compared with wild-type endostatin. Indeed, Stahl et al. (2005) [19] used a recombinant human p.A1381T endostatin in a mammalian expression system and suggested that this change leads to a reduced folding efficiency. They based this suggestion on the residue position and the observed 40% reduction in p.A1381T endostatin production. Considering that KS is probably caused by a complete disruption of at least the isoform NC11–303, the current data does not allow us to conclude that the p.A1381T residue change is sufficient to lead to the KS phenotype or is pathogenic, but this hypothesis seems unlikely. Although in silico analysis indicates that the c. 3601G>A change does not create new splice sites (data not shown), we cannot exclude the possibility that it affects splicing in vivo and leads to creation of a premature stop codon, since no type XVIII collagen was detected in the skin biopsies of these patients. It is also possible that the KS in this family is caused by a yet to be detected mutation or larger deletion in *COL18A1*.

We have previously shown that endostatin serum levels measured by an ELISA assay were lower in KS patients as compared with controls [4]. Although we were able to discriminate patients and controls through this method, we raised the hypothesis that the antibody used in the assay is cross-reactive. This is evidenced by the positive endostatin levels in KS patients that are carriers of null mutations, which should lead to a complete or nearly complete lack of type XVIII collagen. The immunohistochemistry results confirm our prediction that null mutations in the *COL18A1* gene lead

to a severe depletion of the protein. Therefore, immunohistochemical assessment of type XVIII collagen for diagnosing KS may be more precise than endostatin measurements through ELISA assays. The mutation screening of *COL18A1* gene is still expensive and laborious, and the effect of some detected mutations may be uncertain, as is the case for the endostatin p.A1381T change. Even when a possible pathogenic mutation is found, the immunohistochemical analysis of skin biopsies can be an important tool to confirm the diagnosis. Finally, we show for the first time that it is possible to diagnose KS through the direct detection of the protein. In turn, this could allow the screening of a larger number of patients with a possible diagnosis of KS and could help in understanding the phenotypic spectrum of this syndrome.

ACKNOWLEDGMENTS

We thank Päivi Pirilä for her assistance on the SPR assays, Dr. Luis Eduardo Soares Netto and his students for the help with endostatin expression, Mrs. Juliana F. Lucatelli for help in sequencing analysis, Daniela Hipólito Carvalho for help with the preparation of samples for immunohistochemistry, Bjorn Olsen for providing the *COL18A1*cDNA plasmid, and Dr. Shannon Fisher for proofreading the manuscript. This work was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo - CEPID, and Conselho Nacional de Desenvolvimento Científico e Tecnológico.

REFERENCES

1. Cohen MM, Lemire RJ. Syndromes with cephaloceles. Teratology 1982; 25:161-72. [PMID: 7101196]
2. Passos-Bueno MR, Marie SK, Monteiro M, Neustein I, Whittle MR, Vainzof M, Zatz M. Knobloch syndrome in a large Brazilian consanguineous family: confirmation of autosomal recessive inheritance. Am J Med Genet 1994; 52:170-3. [PMID: 7802003]
3. Sertié AL, Sossi V, Camargo AA, Zatz M, Brahe C, Passos-Bueno MR. 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 2000; 9:2051-8. [PMID: 10942434]
4. Suzuki OT, Sertié AL, Der Kaloustian VM, Kok F, Carpenter M, Murray J, Czeizel AE, Kliemann SE, Rosemberg 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 2002; 71:1320-9. [PMID: 12415512]
5. Passos-Bueno MR, Suzuki OT, Armelin-Correa LM, Sertié AL, Errera FIV, Bagatini K, Kok F, Leite KRM. Mutations in collagen 18A1 and their relevance to the human phenotype. An Acad Bras Cienc 2006; 78:123-31. [PMID: 16532212]
6. Saarela J, Ylikärppä 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 1998; 16:319-28. [PMID: 9503365]
7. Halfter W, Dong S, Schurer B, Cole GJ. Collagen XVIII is a basement membrane heparan sulfate proteoglycan. J Biol Chem 1998; 273:25404-12. [PMID: 9738008]
8. Dhanabal M, Ramchandran R, Waterman MJ, Lu H, Knebelmann B, Segal M, Sukhatme VP. Endostatin induces endothelial cell apoptosis. J Biol Chem 1999; 274:11721-6. [PMID: 10206987]
9. Ergün S, Kilic N, Wurmbach JH, Ebrahimnejad A, Fernando M, Sevinc S, Kilic E, Chalajour F, Fiedler W, Lauke H, Lamszus K, Hammerer P, Weil J, Herbst H, Folkman J. Endostatin inhibits angiogenesis by stabilization of newly formed endothelial tubes. Angiogenesis 2001; 4:193-206. [PMID: 11911017]
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 1997; 88:277-85. [PMID: 9008168]
11. Yamaguchi N, Anand-Apte B, Lee M, Sasaki T, Fukai N, Shapiro R, Que I, Lowik C, Timpl R, Olsen BR. Endostatin inhibits VEGF-induced endothelial cell migration and tumor growth independently of zinc binding. EMBO J 1999; 18:4414-23. [PMID: 10449407]
12. Hohenester E, Sasaki T, Olsen BR, Timpl R. Crystal structure of the angiogenesis inhibitor endostatin at 1.5 A resolution. EMBO J 1998; 17:1656-64. [PMID: 9501087]
13. Sasaki T, Fukai N, Mann K, Göhring W, Olsen BR, Timpl R. Structure, function and tissue forms of the C-terminal globular domain of collagen XVIII containing the angiogenesis inhibitor endostatin. EMBO J 1998; 17:4249-56. [PMID: 9687493]
14. Sasaki T, Larsson H, Tisi D, Claesson-Welsh L, Hohenester E, Timpl R. Endostatins derived from collagens XV and XVIII differ in structural and binding properties, tissue distribution and anti-angiogenic activity. J Mol Biol 2000; 301:1179-90. [PMID: 10966814]
15. Menzel O, Bekkeheien RCJ, 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 2004; 23:77-84. [PMID: 14695535]
16. Keren B, Suzuki OT, Gérard-Blanluet M, Brémond-Gignac D, Elmaleh M, Titomanlio L, Delezoide A, Passos-Bueno MR, Verloes A. CNS malformations in Knobloch syndrome with splice mutation in *COL18A1* gene. Am J Med Genet A 2007; 143A:1514-8. [PMID: 17546652]
17. Williams TA, Kirkby GR, Williams D, Ainsworth JR. A phenotypic variant of Knobloch syndrome. Ophthalmic Genet 2008; 29:85-6. [PMID: 18484314]
18. Kliemann SE, Waetge RTL, Suzuki OT, Passos-Bueno MR, Rosemberg S. Evidence of neuronal migration disorders in Knobloch syndrome: clinical and molecular analysis of two novel families. Am J Med Genet A 2003; 119A:15-9. [PMID: 12707952]
19. Stahl S, Gaetzner S, Mueller TD, Felbor U. Endostatin phenylalanines 31 and 34 define a receptor binding site. Genes Cells 2005; 10:929-39. [PMID: 16115201]
20. Czeizel AE, Göblyös P, Kustos G, Mester E, Paraicz E. The second report of Knobloch syndrome. Am J Med Genet 1992; 42:777-9. [PMID: 1554013]

21. den Dunnen JT, Antonarakis SE. Nomenclature for the description of human sequence variations. Hum Genet 2001; 109:121-4. [PMID: 11479744]
22. Carvalhaes LS, Gervásio OL, Guatimosim C, Heljasvaara R, Sormunen R, Pihlajaniemi T, Kitten GT. Collagen XVIII/endostatin is associated with the epithelial-mesenchymal transformation in the atrioventricular valves during cardiac development. Dev Dyn 2006; 235:132-42. [PMID: 16170784]
23. 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 1998; 153:611-26. [PMID: 9708820]
24. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991; 11:281-96. [PMID: 1758883]
25. Rehn M, Veikkola T, Kukk-Valdre E, Nakamura H, Ilmonen M, Lombardo C, Pihlajaniemi T, Alitalo K, Vuori K. Interaction of endostatin with integrins implicated in angiogenesis. Proc Natl Acad Sci USA 2001; 98:1024-9. [PMID: 11158588]
26. Tu H, Sasaki T, Snellman A, Göhring W, Pirilä P, Timpl R, Pihlajaniemi T. The type XIII collagen ectodomain is a 150-nm rod and capable of binding to fibronectin, nidogen-2, perlecan, and heparin. J Biol Chem 2002; 277:23092-9. [PMID: 11956183]
27. Timpl R, Paulsson M, Dziadek M, Fujiwara S. Basement membranes. Methods Enzymol 1987; 145:363-91. [PMID: 3110560]
28. Sasaki T, Kostka G, Göhring W, Wiedemann H, Mann K, Chu ML, Timpl R. Structural characterization of two variants of fibulin-1 that differ in nidogen affinity. J Mol Biol 1995; 245:241-50. [PMID: 7844816]
29. Heljasvaara R, Nyberg P, Luostarinen J, Parikka M, Heikkilä P, Rehn M, Sorsa T, Salo T, Pihlajaniemi T. Generation of biologically active endostatin fragments from human collagen XVIII by distinct matrix metalloproteases. Exp Cell Res 2005; 307:292-304. [PMID: 15950618]
30. Brunak S, Engelbrecht J, Knudsen S. Prediction of human mRNA donor and acceptor sites from the DNA sequence. J Mol Biol 1991; 220:49-65. [PMID: 2067018]
31. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. J Comput Biol 1997; 4:311-23. [PMID: 9278062]
32. Cohen MM, Knobloch WH, Gorlin RJ. A dominantly inherited syndrome of hyaloideoretinal degeneration, cleft palate and maxillary hypoplasia (Cervenka syndrome). Birth Defects Orig Artic Ser 1971; 7:83-6. [PMID: 5173248]
33. Seaver LH, Joffe L, Spark RP, Smith BL, Hoyme HE. Congenital scalp defects and vitreoretinal degeneration: redefining the Knobloch syndrome. Am J Med Genet 1993; 46:203-8. [PMID: 8484411]
34. Wilson C, Aftimos S, Pereira A, McKay R. Report of two sibs with Knobloch syndrome (encephalocoele and viteroretinal degeneration) and other anomalies. Am J Med Genet 1998; 78:286-90. [PMID: 9677068]
35. Sniderman LC, Koenekoop RK, O'Gorman AM, Usher RH, Sufrategui MR, Moroz B, Watters GV, Der Kaloustian VM. Knobloch syndrome involving midline scalp defect of the frontal region. Am J Med Genet 2000; 90:146-9. [PMID: 10607954]
36. Duh EJ, Yao Y, Dagli M, Goldberg MF. Persistence of fetal vasculature in a patient with Knobloch syndrome: potential role for endostatin in fetal vascular remodeling of the eye. Ophthalmology 2004; 111:1885-8. [PMID: 15465551]

The print version of this article was created on 16 April 2009. This reflects all typographical corrections and errata to the article through that date. Details of any changes may be found in the online version of the article.

Anexo 2

Briefings in Functional Genomics and Proteomics Advance Access published September 27, 2008
 doi:10.1093/bfgp/eln045

Evaluation of *cis*-regulatory function in zebrafish

Evanthia E. Pashos, Erika Kague and Shannon Fisher

Abstract

As increasing numbers of vertebrate genomes are sequenced, comparative genomics offers tremendous promise to unveil mechanisms of transcriptional gene regulation on a large scale. However, the challenge of analysing immense amounts of sequence data and relating primary sequence to function is daunting. Several teleost species occupy crucial niches in the world of comparative genomics, as experimental model organisms of wide utility and living roadmaps of molecular evolution. Extant species have evolved after a teleost-specific genome duplication, and offer the opportunity to examine the evolution of thousands of duplicate gene pairs. Transgenesis in zebrafish is being increasingly employed to functionally examine non-coding sequences, from fish and mammals. Here, we discuss current approaches to the study of gene regulation in teleosts, and the promise of future research.

Keywords: *zebrafish; genomics; transgenesis; cis-regulatory elements*

INTRODUCTION

The completion of the human genome sequence, followed successively by multiple vertebrate genomes, has ushered in the field of comparative genomics. It is estimated that 5% of the human genome is under negative selection, accumulating mutations at less than the neutral rate, while only 1.5% can be attributed to gene-coding sequences [1]. We presume that selective pressure acts to conserve functional sequences; while we can reliably recognize most coding exons from primary sequence, there are few good rules to predict the function of conserved non-coding sequences.

One important function of non-coding sequence elements is to regulate transcription of nearby protein-coding genes, although it is unclear what fraction of conserved non-coding sequences are *cis*-regulatory elements (CREs). Efficient methods to assay the function of potential CREs are important, among other reasons, to guide future improvements to the computational algorithms designed to detect them. As part of the ENCODE project, cell culture-based assays are being employed on a large scale to assay for transcriptional regulatory activity in 1% of the human genome [2, 3]. However, these cannot capture the complex tissue and stage-specific gene regulation required for normal development. While transgenic assays in mouse are highly informative, it is difficult to scale up such experiments to produce the volume of information needed to test and improve predictive algorithms.

The increasing use of non-mammalian vertebrate model organisms, and development of tools to manipulate them genetically, offers an alternative. Among these other model systems, zebrafish possess features that make it ideal for the study of gene regulation in development. They have proven easy to maintain and breed in the laboratory, allowing large-scale genetic screens as well as convenient access to embryos. The embryos develop rapidly and externally, and are transparent during early

Corresponding author. Shannon Fisher, BRB II/III 1213, 421 Curie Blvd., Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, PA 19104, USA. Tel:(215)-898-7886; Fax: (215)-898-9871; E-mail: sfisher4@upenn.edu

Evanthia E. Pashos is a PhD candidate in the Human Genetics Graduate Program at The Johns Hopkins University of School of Medicine, and is studying the role of transcriptional regulation in differentiation of pancreatic cell types.

Erika Kague is a PhD candidate in the Departamento de Genética e Biologia Evolutiva, Instituto de Biociências at the Universidade de São Paulo, and is studying regulatory elements associated with human *COL18A1* and their contribution to human disease.

Shannon Fisher is a member of the Department of Cell and Developmental Biology at the University of Pennsylvania School of Medicine. She has been developing approaches in zebrafish to study the function of regulatory sequences, to understand their role in normal development and their contribution to human disease.

development, making them ideal vehicles for the use of fluorescent proteins as reporters of gene transcription. The zebrafish genome sequence is available as high-quality draft, and the most recent assembly is incorporated into the Ensembl and UCSC genome browsers, facilitating the application of powerful alignment algorithms. Finally, recent improvements in transgenesis efficiency in zebrafish have made feasible the type of large-scale screens of putative CREs that are beyond the reach of all but the largest mouse transgenesis efforts. In the current review, we discuss the methods available to analyse CREs in zebrafish, particularly on a large scale; the use of multi-species conservation as a guide to selecting putative CREs for analysis; the challenges of applying computational algorithms to the zebrafish or other teleost genomes; and promising future directions in the use of zebrafish to study regulatory sequences.

DEVELOPMENTAL GENE EXPRESSION DATA

A crucial prerequisite for the study of gene regulation in any embryo is a detailed knowledge of normal developmental gene expression. Among teleost model systems, zebrafish has the most extensive accumulation of publications, and the published expression data is being systematically curated on ZFIN (The Zebrafish Model Organism Database) [4]. In addition, several efforts to screen cDNA clones by whole-mount *in situ* hybridization [5–11] have led to direct submission data for almost 10 000 clones, resulting in an extensive online database of gene expression patterns. Similar efforts have been undertaken for medaka [12, 13], but these have not yet reached the scale of the ZFIN database. However, they do provide a valuable source of comparative data, illustrating aspects of gene expression likely to be conserved among teleosts.

MODES OF TRANSGENESIS

Large-scale efforts to characterize CREs, needed to validate computational prediction algorithms, require methods more efficient than mouse transgenesis. Two basic approaches have been employed in zebrafish to examine large numbers of CREs, based either on DNA injections or on transposon vectors; each approach has specific advantages. Initial reports of transgenesis in teleosts, through injection of plasmid DNA directly into the embryo, were rather inefficient [14–18]. Despite technical limitations, important early studies of gene regulatory elements were carried out in both zebrafish and medaka [19, 20]. Generally these have relied on piecing together expression patterns from large numbers of highly mosaic injected embryos; unthinkable in mouse, these experiments are tedious but feasible in zebrafish because of the ease of injecting large numbers of embryos. Remarkably, a putative enhancer can be co-injected with a separate DNA fragment containing a minimal promoter and reporter gene, and accurately report tissue-specific transcription [21]. Recent studies illustrate the utility of co-injection assays to examine large collections of potential CREs [22, 23]. One significant advantage of this approach is that the sequences for testing need not be cloned into vectors for analysis, greatly reducing the initial effort. However, the injected embryos are highly mosaic, with only a few cells per embryo typically expressing the reporter gene, which can complicate the analysis of regulatory elements that drive expression in small domains. An improvement in both the number of fluorescent cells per embryo and percentage of fluorescent embryos in transient assays has been reported with the use of the Gal4VP16/UAS system [24]. This observation is likely due to amplification of the fluorescent signal rather than an enhancement of transgene integration. Nevertheless, the use of this expression system might prove useful for regulatory elements that have a modest effect on expression levels.

Recent alternative approaches to transgenesis in fish have resulted in increases, sometimes dramatic, in the rates of germline transmission [25–30]. In the case of vectors based on the Tol2 transposon, the improvement in germline transmission is correlated with an increase in the prevalence of somatic expression in mosaic injected embryos compared to plasmid injections [31–33]. This makes it possible in most cases to discern expression patterns from individual or small numbers of injected embryos.

REPORTERS OF TRANSCRIPTIONAL ACTIVITY

Bacterial *lacZ* is a convenient and robust reporter gene used in many organisms, including in many of the earliest reported examples of zebrafish transgenesis. The activity of its product enzyme β-galactosidase is usually assayed in fixed tissues,

but can also be used in conjunction with fluorescent substrates to report on gene expression in live embryos [20, 34]. Its use for generating stable transgenic lines in zebrafish has been largely superseded by the advent of green fluorescent protein (GFP) and other fluorophores and the demonstration of their successful use in live zebrafish [35–40]. The use of fluorescent proteins in conjunction with the detailed imaging possible on live zebrafish embryos has facilitated sophisticated examination of early developmental events. The array of available fluorescent proteins, now available, offer many potentially useful features: faster activation and brighter fluorescence compared to native GFP [41], a wide range of colors [42], activity in monomer form [43] and the property of activating fluorescence [44] or shifting excitation and emission wavelengths in response to light of a specific wavelength [45, 46]. For the purpose of monitoring transcriptional activity, one particularly useful variant is destabilized GFP [47], with a half life considerably shorter than that of native GFP or EGFP.

For analysis of CREs, the primary benefit of using fluorophores as reporter genes is the ease of the initial survey of transcriptional activity. Any time course of expression in mice requires generation and sacrifice of multiple litters, while similar studies can be carried out on a single clutch of live fish embryos. This type of serial analysis is more practical in established transgenic lines or injected embryos with a low degree of mosaicism, such as generated by Tol2 vectors. For the highly mosaic embryos generated by plasmid co-injections, it may be advantageous to use *lacZ* as the reporter gene, fix and stain embryos at selected time points, and have a large collection of fixed specimens for detailed analysis. However, this type of study, as well, may be performed with a fluorescent protein as a marker [23].

CONSERVATION AS A GUIDE TO REGULATORY FUNCTION

Prior to the availability of genome sequence data, the most logical (and often the only feasible) strategy to identify CREs for vertebrate genes was to test sequences immediately upstream of the transcription start site. The availability of multiple vertebrate genome sequences has simplified the cloning and functional analysis of elements far removed from gene-coding sequences. More importantly, the rise of the comparative genomics field has led to genome-wide identification of sequences conserved across evolution. Now it is both feasible and tempting to examine sequences quite distant from gene-coding regions for the cis-regulatory function.

The recent increase in completed vertebrate genome sequences has led to the development of increasingly sophisticated tools for the alignment of two or more sequences and evaluation of the significance of conserved sequence stretches [48–52]. The widely used mVISTA algorithm constructs pairwise alignments with a common reference sequence and slides a window of fixed width along the aligned sequences, seeking sequence stretches that satisfy certain criteria of length and identity (commonly $\geq$70% identity over $\geq$100 bp). The key feature of this approach, the use of a fixed sliding window, risks overlooking conserved sequences when the window is too large and an inability to discriminate between conserved and non-conserved sequences when the window is too small.

In contrast, strategies based on hidden Markov models do not require sliding windows of a fixed size; they utilize continuous Markov models of substitution and allow nearly all parameters to be estimated from the data by maximum likelihood, resulting in the efficient prediction of conserved sequences. An example of this approach, the PhastCons algorithm [50], analyses multi-species alignments generated using the MULTIZ threaded blockset alignment strategy [53] rather than iterative pairwise alignments to a common reference. It implements a phylogenetic hidden Markov model [54, 55] that examines both how a site changes within an alignment and how this process changes from one site to another. PhastCons is implemented as a track on the UCSC Genome Browser [56] and is easily applicable to genome sequences displayed on the browser. The results of PhastCons analysis are presented as LOD scores, calculating the likelihood that a sequence is evolutionarily conserved. This provides an objective criterion for prioritizing the analysis of large numbers of potential enhancer sequences.

Despite the ready availability of comparative genomic predictions, searches for regulatory elements frequently rely on the analysis of extended promoter fragments, followed often by *post hoc* analysis to identify conserved motifs or regions harboring transcription factor binding sites. More rarely, comparative genomic information has been used prospectively to choose sequences for analysis

in mice [57, 58], zebrafish [24, 31, 32, 59] or other organisms [60, 61]. However, many questions remain about the utility of such information, most importantly what is the 'ideal' evolutionary distance across which to compare sequences, and what degree of conservation has the best predictive value.

EXTREME CONSERVATION ONLY PREDICTS A FRACTION OF REGULATORY FUNCTION

Genome-wide sequence comparisons among vertebrates have led to an appreciation of the large amount of conserved non-coding sequence, including 'ultraconserved' sequences, (identical over at least 200 bp among human, mouse and rat [62]) and other highly conserved sequences (>70% identical between human and fugu over ≥100 bp [23]). There is evidence that these highly conserved elements are over-represented in sequences associated with developmentally important genes, and that a high proportion function as CREs in zebrafish [23]. Nonetheless, they comprise a tiny fraction of predicted CREs in the human genome, and there are numerous examples of important genes in development whose expression patterns are well conserved, but are not associated with highly conserved non-coding sequences. Importantly, it is unclear what the predictive value is of ultraconservation; in a recent analysis of ultraconserved elements in transgenic mice, ultraconservation provided no enhancement in the detection of CREs compared to a lower degree of conservation [63].

CONSERVED FUNCTION WITHOUT SEQUENCE CONSERVATION

Some of the earliest examples of zebrafish transgenesis involved the use of regulatory sequences from mammalian genes [20, 64]. These early studies suggested conserved function of CREs across a wide evolutionary range. However, they were performed in the absence of complete sequence data or comparative genomics analysis. We can now appreciate that the non-coding sequence used for the *HOXA2* study [20] is in fact well conserved among vertebrates; the 3.5 kb upstream sequence includes a smaller stretch displaying detectable conservation among all sequenced vertebrates (Figure 1A). In contrast, the 1 kb segment upstream of *GAP43*, while showing detectable conservation among mammals, does not show significant alignment extending to teleosts (Figure 1B).

Despite early success, additional examples of the use of mammalian regulatory elements in zebrafish transgenics were largely absent from the literature until recently. Mammalian sequences highly conserved with orthologous teleost sequences could reasonably be predicted to act as CREs in zebrafish transgenesis assays. More surprisingly, though, human non-coding sequences, conserved among mammals but not found to have homologues in zebrafish, can act as functional CREs in transgenic zebrafish [32, 33]. Such observations might be explained by the inability of current algorithms to detect constrained sequences across great evolutionary distances, since at the level of primary sequence enhancer function is presumed to reside in individual short transcription factor binding sites. Other possible explanations include prevalence of diffuse enhancer architecture that imposes only loose restrictions on the underlying sequence, or by equivalently functioning enhancers produced by convergent evolution.

An important issue little addressed in the literature to date is to what extent conservation, even across more moderate evolutionary distances, is a predictor of regulatory function. To rigorously address this question, large genomic segments must be tested in an unbiased manner for function as CREs. A recent study has done this for the interval surrounding the zebrafish *phox2b* gene, and suggests that a substantial fraction of CREs is missed by relying on standard criteria for conservation [31]. An important caveat to extrapolation of these results is the evolutionary distance between zebrafish and other sequenced teleosts; among the other teleost genomes that have been or are being sequenced (fugu, tetraodon, medaka, stickelback), all are considerably farther in evolutionary distance from zebrafish than mouse is from human [65, 66].

LOCAL REARRANGEMENTS CONFOUND GLOBAL ALIGNMENT ALGORITHMS

Algorithms for detecting sequence conservation, especially when applied to entire genome alignments, can be confounded by local genomic rearrangements. Such rearrangements can alter the topological relationship of a CRE to its cognate

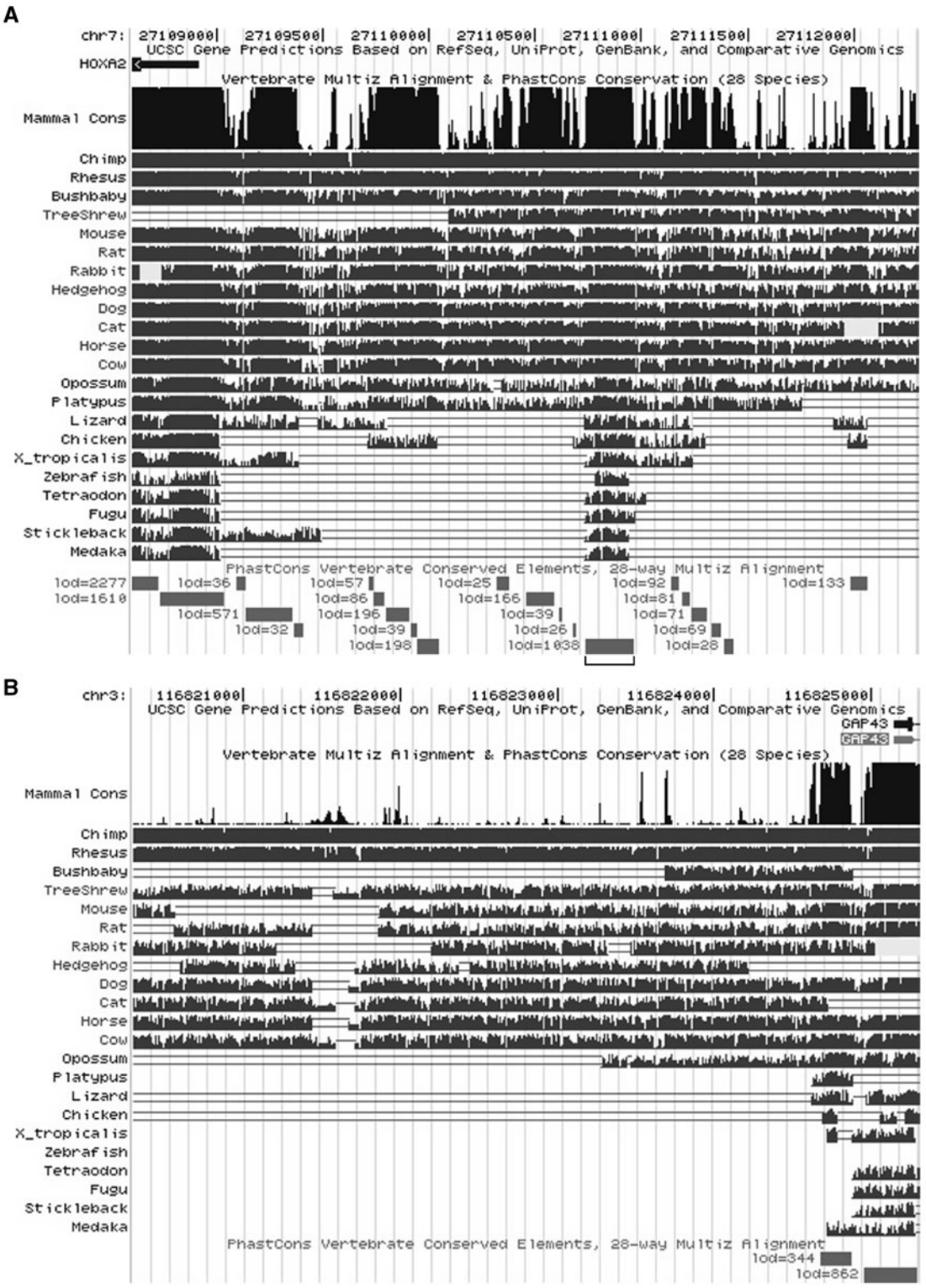


Figure 1: Retrospective analysis of conservation for mammalian sequences used in zebrafish transgenesis. (**A**) The upstream genomic region of the *HOXA2* gene, corresponding to that used previously in transgenic zebrafish [20], displayed in the UCSC Genome Browser [64]. Below are displayed the tracks showing the MULTIZ alignment of 28 vertebrate species, and the likelihood of conservation calculated based on that alignment using the PhastCons algorithm. Note the sequence showing detectable alignment with all vertebrates, including teleosts, indicated by the bracket below the PhastCons track. (**B**) A similar depiction of the region upstream of *GAP43*, also previously used in transgenic zebrafish [60]. In contrast with the *HOXA2* region, there are no sequences alignable between mammals and teleosts.

gene, while preserving its regulatory function. One recent report sought to determine the frequency with which such rearrangements, or CRE shuffling, could be observed in evolution, and the degree to which they might lead to the failure to identify functional CREs through sequence conservation [22]. The authors identified conserved non-coding sequences by comparison among mammalian genomes. For a subset of these, they used an alignment algorithm designed to account for rearrangements [67] to find conserved elements associated with the orthologous genes in three teleost genomes. Most significantly, there was evidence of shuffling for nearly three-fourths of analysed elements. Furthermore, a high proportion of tested elements demonstrated regulatory activity in transgenic zebrafish. Taken together, these results suggest that current genome comparisons between mammals and teleosts often fail to identify conserved non-coding sequence elements, many of which are functional CREs.

Although shuffling of CREs may be more prevasive than commonly appreciated, there is also evidence that some CREs are constrained in their position relative to their associated gene. A recent examination of highly conserved (>95%) elements in the human genome found statistically significant conservation of both position and distance relative to the cognate gene, at least across mammalian genomes [68]. These data raise the interesting possibility that relative position and distance to the regulated gene are constrained for a subset of regulatory elements, perhaps represented by the most highly conserved sequences.

DUPLICATE GENE EVOLUTION IN TELEOSTS

The genomes of extant teleost species serve as a valuable resource to examine evolution of duplicate genes, because of a whole-genome duplication event that took place shortly before the radiation of teleost species. Current estimates place the duplication event approximately 300–350 million years ago [69, 70]. In the intervening time, the majority of duplicate genes have been lost, with as many as 30% of mammalian genes represented by duplicate orthologues in the zebrafish genome [71, 72]. The widely postulated model for preservation of duplicate genes in the teleost genome is based on the hypothesis of subfunctionalization, proposed by Force *et al*. [73]. According to the model, complementary degenerative mutations within independent regulatory elements of the duplicate genes would lead to partitioning of the gene expression pattern of the ancestral allele and subsequent selective pressure to preserve both duplicates. Subfunctionalization could occur by two routes; in qualitative subfunctionalization, a CRE for one duplicate gene would acquire a complete loss-of-function mutation, and the second locus would subsequently acquire a mutation in a different CRE. In contrast, quantitative subfunctionalization would result from fixation of mutations reducing expression of both duplicates.

In a recent study providing evidence for the subfunctionalization hypothesis, Woolfe *et al*. [74] aligned genomic sequence around seven pairs of fugu co-orthologues with corresponding mammalian orthologous regions to look for differential presence of conserved elements between co-orthologues. The authors concluded that all seven pairs of fugu co-orthologues were associated with conserved elements that were partitioned differentially to the co-orthologues, consistent with the model of qualitative subfunctionalization. An important correlate to this computational evidence will be the functional testing of differentially conserved non-coding elements, to confirm that they represent CREs responsible for the differential expression patterns of the co-orthologues.

It is worth noting that although the teleost-specific whole-genome duplication has created an ideal paradigm in which to study duplicate gene evolution, it also complicates genomic sequence comparisons with teleosts. The duplication event closely paralleled the teleost radiation and was followed by independent evolution of genome structures in separate teleost lineages, leading to difficulty finding non-coding sequences conserved among teleosts at duplicate loci. In addition, the whole-genome alignment strategies used to compare mammalian and teleost genomes do not make allowance for the existence of two valid alignments in the teleost sequence corresponding to a single mammalian interval; thus in many cases the second alignment will be missed.

FUTURE APPROACHES TO CRE IDENTIFICATION

One promising approach to the prediction of CREs is based on the hypothesis that enhancers will contain

clusters of transcription factor binding sites, which may be recognized by spatial characteristics and evolutionary conservation. Further, it is hypothesized that genes having a similar expression pattern will be regulated by a common set of transcription factors, reflected by shared composition of motifs within their CREs. One recent study [59] identified five enhancers associated with four zebrafish genes expressed in the midline. A comparison of these enhancers revealed a pair of shared short sequence motifs required for their normal function in regulating midline expression. The requirement for the shared motifs was confirmed by *in vivo* analysis of mutated endogenous enhancers, and by construction of a functional synthetic enhancer containing both motifs.

Del Bene *et al*. [75] initially took a broader approach, starting with a genome-wide prediction of binding sites for the transcription factor Ath5, localized within 5 kb upstream of all annotated genes. The initial prediction was refined by inclusion of conservation data, among mammals and between mammals and teleosts. The final list was validated by examination of endogenous expression of selected predicted target genes and functional assays of sequences containing predicted Ath5 binding sites. The study was confined to sequences 5 kb upstream of annotated genes, but the basic principle of combining transcription factor binding site predictions, phylogenetic conservation data and functional validation could be applied to the examination of more distant regulatory elements.

CONCLUSIONS

As recent technical improvements have made zebrafish transgenesis more efficient, it has become feasible to analyse regulatory elements on a larger scale than is practical for most laboratories working with transgenic mice. There are several circumstances in which it will be highly advantageous to perform these analyses on mammalian sequences. One example is to perform a preliminary survey of a large number of potential enhancers, with only the most promising of these to be moved into transgenic mice. Another is to expand our database of information, in general, about *cis*-regulatory sequences, to better allow us to find and predict enhancers from primary sequence alone. Finally, in the case of duplicate teleost genes, there are significant barriers to identify conserved sequences initially. It may be more straightforward to identify functional enhancers from the human genes first, and then turn to the teleost sequences to identify orthologous elements.

The studies summarized here have focused on one general approach to the understanding of regulatory sequences, the introduction of short isolated sequences as a part of randomly integrated transgenes. As useful as the information thus gained can be, it misses aspects of enhancer function that depend on the context of the entire regulatory landscape. Also, studies on isolated enhancers can indicate their sufficiency of regulatory function, but not their necessity in the context of other enhancers regulating the same gene. One approach to address these questions would be to combine transgenesis with large DNA vectors, such as BAC clones with recombineering, to mutate or delete specific short sequences. Although feasible in zebrafish, BAC transgenesis is technically much more challenging and seldom employed. An alternative in mouse is to mutate an enhancer *in situ* through homologous recombination. Although the technology to perform targeted mutagenesis in zebrafish has not been available previously, recent successes with zinc finger nuclease-mediated mutagenesis [76, 77] suggest that such experiments may be feasible in zebrafish in the near future.

Key Points

- The zebrafish is an important emerging model system to functionally evaluate CREs. Most significantly, efficient methods of transgenesis in zebrafish allow rapid evaluation of large numbers of potential CREs.
- Evolutionary conservation can serve as an important guide to identify potential CREs for *in vivo* testing. However, there are significant limitations to the utility of sequence conservation identification with current common approaches.
- Multi-faceted approaches using a combination of sequence conservation, phylogenetic information, transcription factor binding site predictions and *in vivo* validation will be most successful in the identification of CREs.

References

1. Waterston RH, Lindblad-Toh K, Birney E, *et al*. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002;**420**:520–62.
2. Birney E, Stamatoyannopoulos JA, Dutta A, *et al*. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007;**447**:799–816.
3. Zhang ZD, Paccanaro A, Fu Y, *et al*. Statistical analysis of the genomic distribution and correlation of regulatory

elements in the ENCODE regions. *Genome Res* 2007;**17**: 787–97.

4. ZFIN: The Zebrafish Model Organism Database. http://zfin.org/cgi-bin/webdriver?MIval=aa-ZDB_home.apg (9 August 2008, date last accessed).
5. Expression of the Zebrafish Genome during Embryogenesis (NIH R01 RR15402). http://zfin.org/cgi-bin/webdriver?MIval=aa-pubview2.apg&OID=ZDB-PUB-010810-1 (9 August 2008, date last accessed).
6. Thisse B, Thisse C. Fast Release Clones: a High Throughput Expression Analysis, 2004. http://zfin.org/cgi-bin/webdriver?MIval=aa-pubview2.apg&OID=ZDB-PUB-040907-1 (9 August 2008, date last accessed).
7. Thisse B, Thisse C. High Throughput Expression Analysis of Zf-Models Consortium Clones, 2005. http://zfin.org/cgi-bin/webdriver?MIval=aa-pubview2.apg&OID=ZDB-PUB-051025-1 (9 August 2008, date last accessed).
8. Thisse B, Thisse C. Expression from: Unexpected Novel Relational Links Uncovered by Extensive Developmental Profiling of Nuclear Receptor Expression, 2008. http://zfin.org/cgi-bin/webdriver?MIval=aa-pubview2.apg&OID=ZDB-PUB-080220-1 (9 August 2008, date last accessed).
9. Thisse B, Wright GJ, Thisse C. Embryonic and Larval Expression Patterns from a Large Scale Screening for Novel Low Affinity Extracellular Protein Interactions. 2008. http://zfin.org/cgi-bin/webdriver?MIval=aa-pubview2.apg&OID=ZDB-PUB-080227-22 (9 August 2008, date last accessed).
10. Submission and Curation of Gene Expression Data. http://zfin.org/cgi-bin/webdriver?MIval=aa-pubview2.apg&OID=ZDB-PUB-031103-24 (9 August 2008, date last accessed).
11. Kudoh T, Tsang M, Hukriede NA, *et al*. A gene expression screen in zebrafish embryogenesis. 2001.
12. Medaka EST. Database. http://medaka.lab.nig.ac.jp/est_index.html (9 August 2008, date last accessed).
13. Medaka Expression Pattern Database. http://ani.embl.de:8080/mepd/ (9 August 2008, date last accessed).
14. Houdebine LM, Chourrout D. Transgenesis in fish. *Experientia* 1991;**47**:891–7.
15. Stuart GW, McMurray JV, Westerfield M. Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos. *Development* 1988;**103**:403–12.
16. Stuart GW, Vielkind JR, McMurray JV, *et al*. Stable lines of transgenic zebrafish exhibit reproducible patterns of transgene expression. *Development* 1990;**109**:577–84.
17. Culp P, Nusslein-Volhard C, Hopkins N. High-frequency germ-line transmission of plasmid DNA sequences injected into fertilized zebrafish eggs. *Proc Natl Acad Sci USA* 1991;**88**: 7953–7.
18. Ozato K, Wakamatsu Y, Inoue K. Medaka as a model of transgenic fish. *Mol Mar Biol Biotechnol* 1992;**1**: 346–54.
19. Winkler C, Hong Y, Wittbrodt J, *et al*. Analysis of heterologous and homologous promoters and enhancers in vitro and in vivo by gene transfer into Japanese medaka (Oryzias latipes) Xiphophorus. *Mol Mar Biol Biotechnol* 1992; **1**:326–37.
20. Westerfield M, Wegner J, Jegalian BG, *et al*. Specific activation of mammalian Hox promoters in mosaic transgenic zebrafish. *Genes Dev* 1992;**6**:591–8.
21. Muller F, Chang B, Albert S, *et al*. Intronic enhancers control expression of zebrafish sonic hedgehog in floor plate and notochord. *Development* 1999;**126**:2103–16.
22. Sanges R, Kalmar E, Claudiani P, *et al*. Shuffling of cis-regulatory elements is a pervasive feature of the vertebrate lineage. *Genome Biol* 2006;**7**:R56.
23. Woolfe A, Goodson M, Goode DK, *et al*. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* 2005;**3**:e7.
24. Hadrys T, Punnamoottil B, Pieper M, *et al*. Conserved co-regulation and promoter sharing of hoxb3a and hoxb4a in zebrafish. *Dev Biol* 2006;**297**:26–43.
25. Thermes V, Grabher C, Ristoratore F, *et al*. I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mech Dev* 2002;**118**:91–8.
26. Davidson AE, Balciunas D, Mohn D, *et al*. Efficient gene delivery and gene expression in zebrafish using the sleeping beauty transposon. *Dev Biol* 2003;**263**:191–202.
27. Kawakami K, Shima A, Kawakami N. Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc Natl Acad Sci USA* 2000; **97**:11403–8.
28. Hori H, Suzuki M, Inagaki H, *et al*. An active Ac-like transposable element in teleost fish. *J Mar Biotechnol* 1998;**6**: 206–7.
29. Kawakami K, Koga A, Hori H, *et al*. Excision of the tol2 transposable element of the medaka fish, Oryzias latipes, in zebrafish, Danio rerio. *Gene* 1998;**225**:17–22.
30. Koga A, Suzuki M, Inagaki H, *et al*. Transposable element in fish. *Nature* 1996;**383**:30.
31. McGaughey DM, Vinton RM, Huynh J, *et al*. Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at phox2b. *Genome Res* 2008;**18**: 252–60.
32. Fisher S, Grice EA, Vinton RM, *et al*. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science* 2006;**312**:276–9.
33. Fisher S, Grice EA, Vinton RM, *et al*. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat Protoc* 2006;**1**:1297–305.
34. Lin S, Yang S, Hopkins N. lacZ expression in germline transgenic zebrafish can be detected in living embryos. *Dev Biol* 1994;**161**:77–83.
35. Higashijima S, Okamoto H, Ueno N, *et al*. High-frequency generation of transgenic zebrafish which reliably express GFP in whole muscles or the whole body by using promoters of zebrafish origin. *Dev Biol* 1997;**192**:289–99.
36. Moss JB, Price AL, Raz E, *et al*. Green fluorescent protein marks skeletal muscle in murine cell lines and zebrafish. *Gene* 1996;**173**:89–98.
37. Peters KG, Rao PS, Bell BS, *et al*. Green fluorescent fusion proteins: powerful tools for monitoring protein expression in live zebrafish embryos. *Dev Biol* 1995;**171**:252–7.
38. Finley KR, Davidson AE, Ekker SC. Three-color imaging using fluorescent proteins in living zebrafish embryos. *Biotechniques* 2001;**31**:66–70, 72.
39. Amsterdam A, Lin S, Hopkins N. The Aequorea victoria green fluorescent protein can be used as a reporter in live zebrafish embryos. *Dev Biol* 1995;**171**:123–9.

40. Amsterdam A, Lin S, Moss LG, *et al*. Requirements for green fluorescent protein detection in transgenic zebrafish embryos. *Gene* 1996;**173**:99–103.
41. Zhang G, Gurtu V, Kain SR. An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. *Biochem Biophys Res Commun* 1996;**227**:707–711.
42. Shaner NC, Campbell RE, Steinbach PA, *et al*. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* 2004;**22**:1567–72.
43. Campbell RE, Tour O, Palmer AE, *et al*. A monomeric red fluorescent protein. *Proc Natl Acad Sci USA* 2002;**99**: 7877–82.
44. Patterson GH, Lippincott-Schwartz J. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science* 2002;**297**:1873–7.
45. Gurskaya NG, Verkhusha VV, Shcheglov AS, *et al*. Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nat Biotechnol* 2006;**24**:461–5.
46. Ando R, Hama H, Yamamoto-Hino M, *et al*. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc Natl Acad Sci USA* 2002;**99**:12651–6.
47. Li X, Zhao X, Fang Y, *et al*. Generation of destabilized green fluorescent protein as a transcription reporter. *J Biol Chem* 1998;**273**:34970–5.
48. Cooper GM, Sidow A. Genomic regulatory regions: insights from comparative sequence analysis. *Curr Opin Genet Dev* 2003;**13**:604–10.
49. Margulies EH, Blanchette M, Haussler D, *et al*. Identification and characterization of multi-species conserved sequences. *Genome Res* 2003;**13**:2507–18.
50. Siepel A, Bejerano G, Pedersen JS, *et al*. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;**15**:1034–50.
51. Brudno M, Poliakov A, Salamov A, *et al*. Automated whole-genome multiple alignment of rat, mouse, and human. *Genome Res* 2004;**14**:685–92.
52. Schwartz S, Kent WJ, Smit A, *et al*. Human-mouse alignments with BLASTZ. *Genome Res* 2003;**13**:103–7.
53. Blanchette M, Kent WJ, Riemer C, *et al*. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* 2004;**14**:708–15.
54. Siepel A, Haussler D. Combining phylogenetic and hidden Markov models in biosequence analysis. *J Comput Biol* 2004; **11**:413–28.
55. Siepel A, Haussler D. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol Biol Evol* 2004;**21**:468–88.
56. UCSC Genome Browser. http://genome.ucsc.edu/cgi-bin/hgGateway (9 August 2008, date last accessed).
57. Pennacchio LA, Ahituv N, Moses AM, *et al*. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 2006;**444**:499–502.
58. Poulin F, Nobrega MA, Plajzer-Frick I, *et al*. In vivo characterization of a vertebrate ultraconserved enhancer. *Genomics* 2005;**85**:774–81.
59. Rastegar S, Hess I, Dickmeis T, *et al*. The words of the regulatory code are arranged in a variable manner in highly conserved enhancers. *Dev Biol* 2008;**318**:366–77.
60. Allende ML, Manzanares M, Tena JJ, *et al*. Cracking the genome's second code: enhancer detection by combined phylogenetic footprinting and transgenic fish and frog embryos. *Methods* 2006;**39**:212–9.
61. de la Calle-Mustienes E, Feijoo CG, Manzanares M, *et al*. A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. *Genome Res* 2005;**15**:1061–72.
62. Bejerano G, Pheasant M, Makunin I, *et al*. Ultraconserved elements in the human genome. *Science* 2004;**304**:1321–5.
63. Visel A, Prabhakar S, Akiyama JA, *et al*. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet* 2008;**40**:158–60.
64. Reinhard E, Nedivi E, Wegner J, *et al*. Neural selective activation and temporal regulation of a mammalian GAP-43 promoter in zebrafish. *Development* 1994;**120**:1767–75.
65. Volff JN. Genome evolution and biodiversity in teleost fish. *Heredity* 2005;**94**:280–94.
66. Steinke D, Salzburger W, Meyer A. Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. *J Mol Evol* 2006;**62**: 772–84.
67. Brudno M, Chapman M, Gottgens B, *et al*. Fast and sensitive multiple alignment of large genomic sequences. *BMC Bioinformatics* 2003;**4**:66.
68. Sun H, Skogerbo G, Chen R. Conserved distances between vertebrate highly conserved elements. *Hum Mol Genet* 2006; **15**:2911–22.
69. Postlethwait JH, Yan YL, Gates MA, *et al*. Vertebrate genome evolution and the zebrafish gene map. *Nat Genet* 1998;**18**:345–9.
70. Taylor JS, Van de Peer Y, Braasch I, *et al*. Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc Lond B Biol Sci* 2001;**356**:1661–79.
71. Postlethwait JH, Woods IG, Ngo-Hazelett P, *et al*. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* 2000;**10**:1890–902.
72. Woods IG, Wilson C, Friedlander B, *et al*. The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res* 2005;**15**:1307–14.
73. Force A, Lynch M, Pickett FB, *et al*. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 1999;**151**:1531–45.
74. Woolfe A, Elgar G. Comparative genomics using Fugu reveals insights into regulatory subfunctionalization. *Genome Biol* 2007;**8**:R53.
75. Del Bene F, Ettwiller L, Skowronska-Krawczyk D, *et al*. In vivo validation of a computationally predicted conserved Ath5 target gene set. *PLoS Genet* 2007;**3**:1661–71.
76. Doyon Y, McCammon JM, Miller JC, *et al*. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol* 2008;**26**:702–8.
77. Meng X, Noyes MB, Zhu LJ, *et al*. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* 2008;**26**:695–701.

Anexo 3

Anais da Academia Brasileira de Ciências (2008) 80(1): 167-177
(Annals of the Brazilian Academy of Sciences)
ISSN 0001-3765
www.scielo.br/aabc

COL18A1 is highly expressed during human adipocyte differentiation and the SNP c.1136C>T in its "frizzled" motif is associated with obesity in diabetes type 2 patients

FLAVIA I.V. ERRERA[1,2], LUÍS H. CANANI[3], ERIKA YEH[1], ÉRIKA KAGUE[1], LUCIA M. ARMELIN-CORRÊA[1], OSCAR T. SUZUKI[1], BALDUÍNO TSCHIEDEL[5], MARIA ELIZABETH R. SILVA[4], ANDRÉA L. SERTIÉ[1] and MARIA RITA PASSOS-BUENO[1]

[1]Centro de Estudos do Genoma Humano, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, Rua do Matão 277, 05509-900 São Paulo, SP, Brasil
[2]Faculdade de Medicina, Escola Superior de Ciências da Santa Casa de Vitória (EMESCAM), Av. Nossa Senhora da Penha, 2190, 29045-102 Vitória, ES, Brasil
[3]Departamento de Endocrinologia, Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul Rua Ramiro Barcellos, 2350, 90035-903 Porto Alegre, RS, Brasil
[4] Laboratório de Investigação Médica, LIM-18, Hospital das Clínicas da Universidade de São Paulo Faculdade de Medicina, Av. Dr. Arnaldo, 455, 01246-903 São Paulo, SP, Brasil
[5]Sessão de Endocrinologia, Hospital Conceição, Av. Francisco Trein, 596, 90013-901 Porto Alegre, RS, Brasil

Manuscript received on October 3, 2007; accepted for publication on January 21, 2008;
presented by ROGÉRIO MENEGHINI

ABSTRACT

Collagen XVIII can generate two fragments, NC11-728 containing a frizzled motif which possibly acts in Wnt signaling and Endostatin, which is cleaved from the NC1 and is a potent inhibitor of angiogenesis. Collagen XVIII and Wnt signaling have recently been associated with adipogenic differentiation and obesity in some animal models, but not in humans. In the present report, we have shown that *COL18A1* expression increases during human adipogenic differentiation. We also tested if polymorphisms in the Frizzled (c.1136C>T; Thr379Met) and Endostatin (c.4349G>A; Asp1437Asn) regions contribute towards susceptibility to obesity in patients with type 2 diabetes (113 obese, BMI $\geq$30; 232 non-obese, BMI $<$ 30) of European ancestry. No evidence of association was observed between the allele c.4349G>A and obesity, but we observed a significantly higher frequency of homozygotes c.1136TT in obese (19.5%) than in non-obese individuals (10.9%) [P = 0.02; OR = 2.0 (95%CI: 1.07-3.73)], suggesting that the allele c.1136T is associated to obesity in a recessive model. This genotype, after controlling for cholesterol, LDL cholesterol, and triglycerides, was independently associated with obesity (P = 0.048), and increases the chance of obesity in 2.8 times. Therefore, our data suggest the involvement of collagen XVIII in human adipogenesis and susceptibility to obesity.

Key words: *COL18A1*, obesity, adipogenesis, endostatin, frizzled.

INTRODUCTION

COL18A1, located at chromosome 21q22.3, was first shown to have an important role in humans when we demonstrated that homozygous null mutations in this gene cause Knobloch syndrome (KS) (Sertié et al. 2000). KS is an autosomal recessive condition characterized by vitreo and retinal degeneration, macular abnormalities and occipital encephalocele (KS; OMIM 267750). The importance of collagen XVIII in retinal angiogenesis was further confirmed through the observation of ocular delayed hyaloid vessel regression in knockout col18a-/- mice (Fukai et al. 2002).

Collagen XVIII is an important proteoglycan of

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

most basement membranes of the body which belongs to a distinct class of nonfibrillar collagens, called multiplexins. This protein forms homotrimers and contains a 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 (Oh et al. 1994, Saarela et al. 1998). *COL18A1* can be transcribed as 3 distinct isoforms by use of two promoters and an alternative splicing of the third exon. The three encoded collagen XVIII proteins differ only by their signal peptides and by part of the N-terminal region of the NC11 domain, which contain respectively 303 (NC11-303), 493 (NC11-493) and 728 (NC11-728) amino acid residues (Saarela et al. 1998, Elamaa et al. 2003). Collagen XVIII presents high expression levels in liver, kidney, retina and adipose tissue and the predominant isoform varies according to the tissue type (Saarela et al. 1998, Suzuki et al. 2002, Elamaa et al. 2003, NCBI UniGene Hs.517356).

At least two functional fragments can be released from collagen XVIII: Endostatin, a 20 kDa fragment cleaved from the NC1 region- common to the three collagen XVIII isoforms (ES-C18) (O'Reilly et al. 1997, John et al. 2005) and the N-terminal Frizzled cysteine-rich domain cleaved from the NC11-728 variant (Frizzled-C18) (Elamaa et al. 2003). ES-C18 is a potent inhibitor of endothelial cell proliferation and migration with the ability to reduce tumor growth (O'Reilly et al. 1997, Sasaki et al. 1998). Collagen XVIII and ES-C18 play several other functions, including apoptosis induction (Dhanabal et al. 1999, Schmidt et al. 2004) and maintenance of vascular permeability (Moulton et al. 2004). The Frizzled-C18 motif shows homology to secreted frizzled related proteins (sFRP). These proteins can block Wnt signaling either by interacting with Wnts to prevent them from binding to frizzled receptors or forming nonfunctional complexes with them (Wodarz and Nusse 1998). The function of the soluble Frizzled-C18 motif is still unknown, but it is possibly involved in Wnt signaling as sFRPs does (Elaama et al. 2003, Quelard et al. 2006).

Although the molecular mechanisms regulated by collagen XVIII and its NC11 and NC1 functional fragments are still being investigated, it is unquestionable the functional role of collagen XVIII in angiogenesis. Therefore, variability in quantity or quality of collagen XVIII and its related fragments could be associated with the predisposition of human disorders dependent on angiogenesis (Passos-Bueno et al. 2006). Substantial evidence has established that growth of solid tumors is angiogenesis dependent whereas angiogenesis inhibition prevents it and can lead to regression of the lesions (Udagawa et al. 2002, Folkman 2002). Although we and others were not able to confirm the association between the SNP Asp 1437 Asn (D104N) at the endostatin region and prostate cancer (Iughetti et al. 2001, Macpherson et al. 2004, Passos-Bueno et al. 2006), several reports have shown that lower ES-C18 levels have been correlated with a worse tumor prognosis (Musso et al. 2001) and that the SNP D104N might be associated with breast cancer (Lourenço et al. 2006, Balasubramanian et al. 2007). Therefore, analysis of *COL18A1* gene SNPs in diseases in which angiogenesis plays a role can provide further insights in the functional importance of this molecule in humans.

Adipose tissue is very unique as it can grow and regress throughout adulthood. Its growth, which can arise through increases in cell size (due to accumulation of triglycerides) and/or cell number (due to proliferation of preadipocytes and their subsequent differentiation into mature adipocytes), is largely angiogenesis dependent (Rupnick et al. 2002, Cao 2007, Nishimura et al. 2007). Collagen XVIII, which is abundant in adipose tissue (NCBI UniGene Hs.517356), is involved in angiogenesis control and it has been shown that its transcription levels increase during bovine adipocyte differentiation (Inoue-Murayama et al. 2000). However, the function of collagen XVIII in the molecular control of adipogenesis and obesity is very poorly characterized and it has not yet been investigated in humans. Therefore, in this study we have first evaluated if *COL18A1* mRNA levels vary during *in vitro* adipogenesis of human adipose tissue-derived stem cells. In addition, we also aimed to verify if collagen XVIII would have any effect on the predisposition of obesity, as this is known to be a major risk factor for several diseases such as type 2 *diabetes mellitus* (DM2) and cardiovascular disorders. To achieve this goal, we tested if SNP c.1136C>T (Thr379Met), which is being here described for the first time, and SNP c.4349G>A (Asp1437Asn) are associated

with obesity in a cohort of Brazilian patients with DM2. These SNPs are missense mutations in the coding region of the human *COL181A* and are located respectively in the NC11-728 (Frizzled-C18) and Endostatin (ES-C18) domains of the protein.

MATERIALS AND METHODS

In vitro Adipogenic Differentiation

Adipose tissue-derived mesenchymal stem cells were obtained from lipoaspirates of abdominal subcutaneous fat. The cells were isolated and cultured as described previously (Zuk et al. 2002, Costa et al. 2008, Bueno 2007). Their differentiation into adipocytes was induced by treatment of semi-confluent cells for 14 days with DMEM-HG supplemented with 10% FBS (Hyclone), 1μM dexamethasone, 100μM indomethacin, 500μM 3-isobutyl-1-methylxanthine (IBMX), and 10μg/ml insulin as standard protocol (Zuk et al. 2002). To assess neutral lipid content at day 14 after exposure to inducers of differentiation, mature adipocytes were washed with PBS, fixed with 4% paraformaldehyde and stained with Oil Red O. A successful differentiation process was considered when it was observed that the majority of cells accumulated lipid vesicles in the cytoplasm. The use of human lipoaspirates to establish stem cell populations was accordingly to the guidelines of our local Ethical Committee.

Quantitative Real-time PCR (qRT-PCR)

RNA extraction was conducted with trizol (Gibco) reagent at days 0 (no adipogenic induction medium), 3, 7 and 14 of the differentiation process. 5μg of total RNA was treated with 5 units of RQ1 RNase- Free DNAse (Promega). cDNAs were synthesized from 1μg of total RNA using SuperScript II (Invitrogen). Quantitative real time PCR was performed with SYBR Green PCR master mix in an ABI Prism 7100 system (Applied Biosystems) and approximately 50 ng of cDNA was used to final reaction volume of 25μL. The primers used in this study are listed in Table I; the *COL18A1* primers amplify all isoforms. The PCR conditions were: 94°C for 15 seconds, 60°C for 30 seconds, and 72°C for 30 seconds for 45 cycles. The four samples acquired from adipogenic differentiation (time points 0, 3, 7 and 14 days) were submitted to qRT-PCR and the threshold cycle (ct) for each individual PCR product was calculated by the instrument software. ΔCt was calculated with the difference between sample Cts of each time point of differentiation and reference sample, which was the time 0 of differentiation. And $\Delta\Delta$Cts was calculated based on the efficiency of each primer. GeNorm (Vandesompele et al. 2002) was used to obtain the normalization factor for each sample based on the 2 best endogenous controls that the program selects. Among the 4 internal control genes analyzed (*GAPDH*, *SDHA*, *HMBS* and *HPRT1*) GeNorm program selected *GAPDH* and *SDHA* to generate the normalization factor used to normalize the samples $\Delta\Delta$Cts.

DM2 Patients

We have conducted a case-control study with a total of 345 unrelated patients with DM2. One hundred thirty five patients were ascertained at the Endocrinology Department, Hospital das Clínicas – USP, 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 (Canani et al. 2005, Errera et al. 2007). Briefly, patients with DM2 were identified from a multicentric study that started recruiting patients in southern Brazil in 2002. That project originally 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 (mainly from Portugal, Spain, Italy and Germany). Patients of African ancestry or those who defined themselves as having mixed or other ancestry were not included. The ethical committees of the participant Institutions approved this study and blood samples were drawn only after signature of the informed consent. Patients underwent a standardized evaluation consisting of a questionnaire, physical examination and laboratory tests. Diagnosis of DM2 was based on the guidelines of the Expert Committee report of the American Diabetes Association (American Diabetes Associa-

TABLE I
Sequence of the primers used in the Quantitative Real-Time PCR experiment.

Gene	Sense Primer ($5' \rightarrow 3'$)	Antisense Primer ($5' \rightarrow 3'$)
SDHA	TGGGAACAAGAGGGCATCTG	CCACCACTGCATCAAATTCATG
HPRT1	TGACACTGGCAAAACAATGCA	GGTCCTTTTCACCAGCAAGCT
HMBS	GGCAATGCGGCTGCAA	GGGTACCCACGCGAATCAC
GAPDH	CCATCTTCCAGGAGCGAGAT	TTCTCCATGGTGGTGAAGAC
LPL	CTGGCATTGCAGGAAGTCTG	GCATCATCAGGAGAAAGACGA
COL18A1	AAGGACGAGCTGCTGTTTCC	TTGCCGTCAAAGGAGAAGATG

tion, 2005). 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/height^2 (kg/m^2). Obesity was classified according Word Health Organization (WHO) criteria (World Health Organization 2000). Patients with BMI $\geq$ 30 kg/m^2 were considered cases and patients with BMI $<$ 30 kg/m^2 as controls.

Peripheral Blood Collection, DNA Extraction and Biochemical Dosages

Genomic DNA was extracted from peripheral blood leucocytes with salting-out procedure as previously described (Miller et al. 1988). In diabetic patients, glucose was determined by a glucose oxidase method; creatinine by the Jaffé reaction; A1c test by an ion-exchange HPLC procedure (Merck-Hitachi L-9100 Glycated hemoglobin analyzer; reference range: 2.7-4.3%; Merck, Darmastadt, 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.1136C>T (Thr379Met) and c.4349G>A (Asp1437Asn) 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, 0.2 mM dNTPs, 0.6μM of each primer; 0.2 U Taq polymerase, in 35 cycles with 100 ng of DNA, according to the following steps: 5 minutes denaturation at 95°C, 10 minutes extention at 72°C, and using the following primers and respective annealing temperature: exon $3F$: cccaaccacctccaccacgaga and exon $3R$: ctggatgaaccgcagccgatgt (66°C); $42F$: acaaacacccacacccatc and $42R$: gaaactgcagataggaggccc (59°C). PCR products were submitted to SNuPE reaction and then analyzed in automatic sequencer (MegaBACE^{TM} 1000 DNA Analysis Systems), using the correspondent *software* and following the User's guide instructions (General Electrics Health Science) and 2 pmols of oligonucleotide [c.1136C>T: CGAGGAAGGGGTGGCAGTGC, and c.4349G>A: GGGCACGCATCTTCTCCTTT (5'-3')].

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 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. Sample size power was calculated with PS Power and Sample Size Calculations (Version 2.1.30) (Dupont and Plummer 1998).

RESULTS

COL18A1 Gene Expression During Human Adipogenesis

On day 14 after inducing adipocyte differentiation of adipose tissue-derived human mesenchymal stem cells (hMSCs), we observed that the majority of cells accumulated lipid droplets in the cytoplasm (Fig. 1). In addition, *LPL*, an adipogenic marker, was quantified as a control of the differentiation experiment and it showed extreme high expression levels after the induction of the adipogenesis process (Fig. 2). Therefore, we considered that the stem cell population successfully differentiated into adipocytes.

COL18A1 expression was analyzed at four time points (0, 3, 7 and 14 days) during the differentiation induction and we observed relatively higher expression of this gene at day 3 of the differentiation processes, when it increased 2,5 fold in comparison with undifferentiated cells. At days 7 and 14 the expression was 1.4 and 0.85 fold-higher than that of untreated stem cells (time 0), respectively (Fig. 2).

Association Study Between *COL18A1* and Obesity

The main clinical features of the patients are depicted in Table II. Among the 345 patients analyzed, 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. The genotype distribution for SNPs c.1136C>T and c.4349G>A were in HWE (Table III; $P > 0.05$). There was no difference in the genotype or allelic frequencies of the SNP c.4349G>A between obese and non-obese patients ($P = 0.11$). However, a borderline association between SNP c.1136C>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) for obesity. This effect was more pronounced if we exclude individuals with overweight, that is those with BMI between 28 and 30 kg/m^2 [TT genotype frequency in controls (BMI < 27 kg/m^2, 12/140) was 8.5% ($\chi^2 = 6.38$ and $P = 0.01$)]. In a logistical model, after controlling for cholesterol, triglycerides, LDL cholesterol levels, the TT genotype still held associated with the presence of obesity ($P = 0.048$; OR 2,8 CI 95% 1.01-7.74).

DISCUSSION

In the present work, we have shown for the first time that human collagen XVIII mRNA levels increase during differentiation of human mesenchymal stem cells (hMSCs) into adipocytes. Its transcription profile observed during the differentiation process suggests that *COL18A1* expression increases at an earlier stage of hMSCs adipogenesis, which is comparable to other well known genes associated with adipocyte differentiation, such as C/EBPβ and C/EBPδ (Rosen 2005). These results are in accordance with those observed in the differentiation of bovine preadipocytes obtained from intramuscular fat (Inoue-Murayama et al. 2000). Our results thus indicate that collagen XVIII may play an important role in human adipogenesis and the higher increase of *COL18A1* expression in the early phase of adipogenic differentiation point to the importance of this gene in the adipogenic commitment of the cells. It was observed that collagen XVIII expression levels did not vary during adipogenesis of the mouse preadipocyte cell line 3T3-L1 (Inoue-Murayama et al. 2000). It is therefore possible that the effect of collagen XVIII in adipocyte differentiation depends on the origin of the preadipocyte or it is species related. It is worth to note that we have used primer pairs that recognize the three collagen XVIII isoforms to quantify its expression in human adipocity differentiation, but it would be important to verify if a specific or all collagen XVIII isoforms are involved in this process.

We have also evaluated whether there was an association between *COL18A1* gene polymorphisms c.1136C>T and c.4349G>A, located respectively in the NC11- 728/Frizzled-C18 and NC-1/ES-C18 domains of collagen XVIII, 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 (Passos-Bueno et al. 2006) and we did not observe any evidence of association

TABLE II
Clinical variables in DM2 patients according to presence of obesity.

	Controls (n= 232)	Cases (n= 113)	*P*
Gender Female: Male	122:110	77:36	
DM duration (years)	12.78 ± 7.7451	12.81 ± 7.17	0.97
Age of DM Diagnosis (years)	48.81 ± 10.33	44.64 ± 9.62	0.07
Current age (years)	60.80 ± 9.60	60.02 ± 8.34	0.47
BMI (kg/m2)	25.83 ± 2.53	34.31 ± 4.64	—
Waist circumference (cm)	92.88 ± 9.29	105.36 ± 12.40	0.00
Hip circumference (cm)	98.48 ± 7.54	114.48 ± 15.33	0.00
Glucose (mg/dl)	172.47 ± 68.43	175.34 ± 66.73	0.78
LDL- cholesterol (mg/dl)	130.27 ± 46.19	147.70 ± 48.05	0.01
HDL-cholesterol (mg/dl)	45.82 ± 12.70	44.06 ± 12.00	0.24
Cholesterol (mg/dl)	202.97 ± 47.14	223.81 ± 53.41	0.02
Creatinine (mg/dl)	1.25	1.29	0.80
Tryglycerides	174.08 (70-238)	216.98 (89-260.8)	0.02
A1c test (%)	7.13 ± 2.36	7.49 ± 2.53	0.31
Systolic Arterial Pressure (mmHg)	142.89 ± 24.02	144.54 ± 20.85	0.62
Diastolic Arterial Pressure (mmHg)	85.35 ± 13.66	87.11 ± 11.93	0.33

TABLE III
Genotypic and allelic distribution for SNP c.1136C>T and SNP c.4349G>A according to body mass index.

	Body Mass Index	
	> 30 kg/m^2 (n = 113)	< 30Kg/m^2 (n = 232)
SNP c.1136C>T		
Genotypes		
CC - n (%)	46 (40.7)	111 (47.8)
CT - n (%)	45 (39.8)	96 (41.4)
TT - n (%)	22 (19.5)	25 (10.8)
Alleles		
C	60.6%	68.5%
T	39.4%	31.5%
SNP c.4349G>A		
Genotypes		
GG - n (%)	102 (90.3)	198 (85.4%)
GA - n (%)	11 (9.7)	33 (14.2%)
AA - n (%)	0 (0)	1 (0.4%)
Alleles		
G	95.1%	92.5%
A	4.9%	7.5%

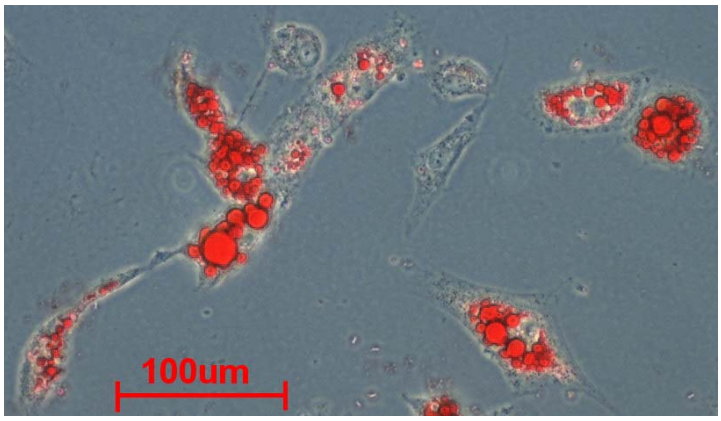


Fig. 1 – Oil red staining of the cells after 14 days of the beginning of adipogenic differentiation showing accumulated lipid vesicles in the cytoplasm.

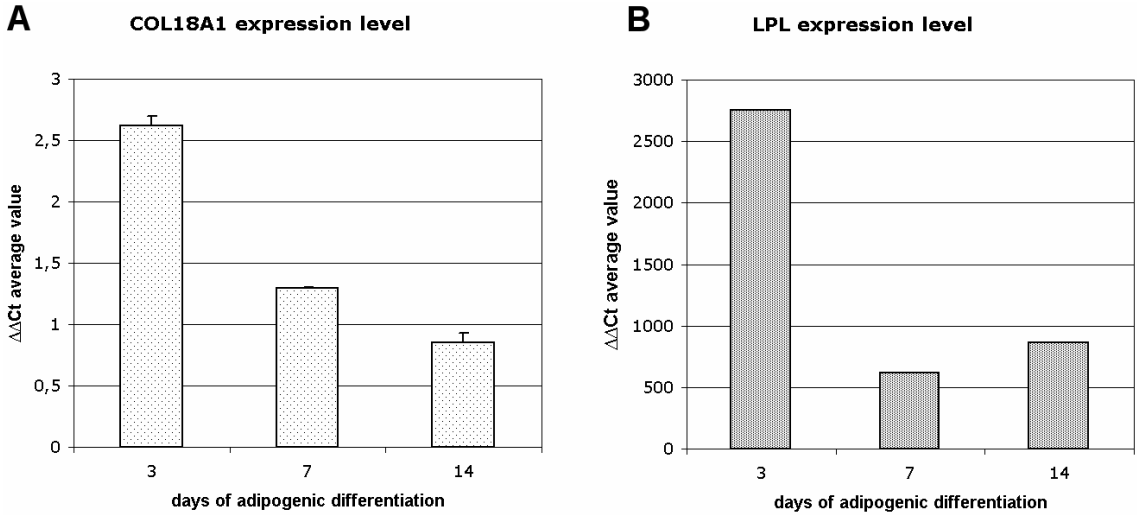


Fig. 2 – A) Quantitative real time PCR showing expression of *COL18A1* during human adipogenic differentiation *in vitro*. B) *LPL* was quantified as a control of the differentiation experiment and, as expected, it showed high expression levels after the beginning of the adipogenesis process.

between the SNP c.4349G>A and obesity. However, we observed consistent evidence of association between SNP c.1136C>T and obesity.

Our results suggest that the allele c.1136T when in homozygosis increases about two-fold the chance of a DM2 patient to develop obesity. When overweight DM2 patients were excluded from the control group, we found a more pronounced difference in c.1136T frequency between cases and controls, further supporting the association between this SNP and obesity in DM2 patients. The importance of TT genotype for obesity was still confirmed by logistic regression analysis with the following 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 CT heterozygotes or CC homozygotes. Hence, our data suggest that TT genotype acts as an independent factor for predisposition to obesity in DM2 patients. The observed frequency of TT genotype in our obese cohort (19.5%) was very close to the 22% estimate to achieve an acceptable power level with the sample size available, thus, suggesting that this association might not represent a false positive result. However, although both cases and controls are of EA

and ascertained in a common cohort of patients, we still cannot exclude the possibility that these results are due to population stratification.

It is well established that proteins with frizzled motif act in Wnt signaling through different pathways in various processes, including adipogenesis, obesity, insulin secretion and DM2 (Ross et al. 2000, Bennett et al. 2002, Kennel and MacDougald 2005, Guo et al. 2006). It has been suggested that the N-terminal portion of collagen XVIII containing the Frizzled-C18 motif could be involved in Wnt signaling (Elaama et al. 2003, Vainio et al. 2003, Marneros and Olsen 2005) and, more recently, Quelard et al. (2006) showed that the Frizzled-C18 module acts as an extracellular inhibitor of Wnt-B-catenin signaling by quenching Wnts in the hepatocarcinoma tumor. The SNP c.1136C>T/Thr379Met is being described in the present work for the first time. In a comparative genomic analysis we verified that threonine at position 379 is conserved among mammals (human, rat, mouse, cow, cat) and is substituted for another polar aminoacid with an uncharged side chain (serine) in other vertebrates (chicken, xenopus, zebrafish, medaka, fugu, stickleback), suggesting that this residue might be important for normal protein function. The frizzled motif of the NC11-728 has a cysteine-rich domain, that contains 10 cysteine-conserved sites, and threonine 379 is located very close to the third cysteine residue. Therefore, it is possible that the presence of a methionine instead of a threonine can partially impair the function of this domain in preadipocyte proliferation or differentiation through a direct or indirect effect of the Wnt pathway.

Recently, we verified that allele c.1136T is in linkage disequilibrium with the G allele of the polymorphism –700G>T, which confers higher transcription levels to this collagen in hepatocytes (Armelin-Correa et al. 2005). Therefore, 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.

It has not been reported obesity in patients with null mutations in the collagen XVIII or in the knockout col18-/- mice (Suzuki et al. 2002, Fukai et al. 2002). However, it is possible that the obesity phenotype is induced only in the presence of high food intake or depending on the genetic background of the individual. 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 (Zorick et al. 2001).

COL18A1 and its receptor glypican has been identified as promising candidate genes to be involved with body weight and adiposity in DM2 mice (Brown et al. 2005, Gesta et al. 2006), thus supporting our finding that collagen XVIII is important in adipogenesis and in obesity of DM2 patients.

In conclusion, the present study shows for the first time that collagen XVIII mRNA levels increase during human adipogenesis and that the SNP c.1136C>T in the frizzled motif of the *COL18A1* doubles the chance of a patient with type 2 diabetes to develop obesity. Considering the functional role of collagen XVIII in angiogenesis and that it may be involved in Wnt signaling, we could speculate that this collagen might act in the regulation of adipose tissue mass through different mechanisms: by controlling angiogenesis in fat tissue and by stimulating the differentiation of mesenchymal stem cell into mature adipocytes. Therefore, although preliminary, our present data may be a starting point towards understanding this novel role of collagen XVIII in humans.

ACKNOWLEDGMENTS

We are grateful to Dr. Carlos Maranduba for helping in patients interviews and molecular analysis, to Dr. Jorge Gross, Dr. Israel Roisenberg, Dr. Daniel Gianella and to Dr. Hugo Lisbôa and his team for kindly providing part of the sample of diabetic patients analyzed in this study, and to Natassia Vieira for helping with the human adipose-derived mesenchymal stem cells characterization, isolation, and differentiation. This research has been 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 Desenvolvimento Científico e Tecnológico (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 Nível Superior (Fundação CAPES).

RESUMO

Colágeno XVIII pode gerar dois fragmentos, um correspondendo à região NC11-728 contendo o motivo "frizzled", o qual possivelmente atua na sinalização Wnt, e outro correspondendo a Endostatina, que é clivada a partir da região NC1 e é uma potente inibidora de angiogênese. Colágeno XVIII e a via de sinalização Wnt foram recentemente associados à diferenciação adipogênica e obesidade em alguns modelos animais, porém ainda não em humanos. No presente trabalho, mostramos que os níveis de expressão gênica do *COL18A1* aumentam durante o processo de diferenciação adipogênica em humanos. Também testamos se polimorfismos localizados no motivo "Frizzled" (c.1136C>T; Thr379Met) e na região da Endostatina (c.4349G>A; Asp1437Asn) contribuem na predisposição a obesidade em pacientes com diabetes tipo 2. (113 obesos, BMI $\geq$ 30; 232 não-obesos, BMI $<$ 30) de ancestralidade Européia. Nenhuma evidência de associação entre o alelo c.4349G>A e obesidade foi observada, contudo, observamos uma freqüência significativamente maior de homozigotos c.1136TT em obesos (19.5%) do que em não-obesos (10.9%) [P = 0.02; OR = 2.0 (95%CI: 1.07-3.73)], sugerindo que o alelo c.1136T está associado com obesidade conforme um modelo recessivo. Este genótipo manteve-se associado à obesidade (P = 0.048) mesmo após o controle das variáveis colesterol, LDL e triglicérides, e confere um risco 2.8 vezes maior de obesidade. Portanto, nossos dados sugerem o envolvimento do colágeno XVIII em adipogênese humana e predisposição a obesidade.

Palavras-chave: *COL18A1*, obesidade, adipogênese, endostatina, "frizzled".

REFERENCES

AMERICAN DIABETES ASSOCIATION. 2005. Diagnosis and classification of *diabetes mellitus*. Diabetes Care [WINDOWS-1252?] 28: S37–S42.

ARMELIN-CORREA LM ET AL. 2005. 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.

BALASUBRAMANIAN SP, CROSS SS, GLOBE J, COX A, BROWN NJ AND REED MW. 2007. Endostatin gene variation and protein levels in breast cancer susceptibility and severity. BMC Cancer 7: 107–114.

BENNETT CN ET AL. 2002. Regulation of Wnt signaling during adipogenesis. J Biol Chem 277: 30998–31004.

BROWN AC, OLVER WI, DONNELLY CJ, MAY ME, NAGGERT JK, SHAFFER DJ AND ROOPENIAN DC. 2005. Searching QTL by gene expression: analysis of diabesity. BMC Genet 6: 12–20.

BUENO DF. 2007. Uso de células tronco adultas para estudo da etiopatogenia das fissuras lábio palatinas e bioengenharia de tecidos. Tese de doutorado, Instituto de Biociências, Universidade de São Paulo, SP, Brasil.

CAO Y. 2007. Angiogenesis modulates adipogenesis and obesity. J Clin Invest 117: 2362–2368.

CANANI LH ET AL. 2005. The fatty acid-binding protein-2 A54T polymorphism is associated with renal disease in patient with type 2 diabetes. Diabetes 54: 3326–3330.

COSTA AM, BUENO DF, MARTINS MT, KERKIS I, KERKIS A, FANGANIELLO RD, CERRUTI H, ALONSO N AND PASSOS-BUENO MR. 2008. Reconstruction of Large Cranial Defects in Non Immunosuppressed Experimental Design With Human Dental Pulp Stem Cells. J Craniofac Surg 19: 204–210.

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.

DUPONT WD AND PLUMMER JR WD. 1998. Power and sample size calculations for studies involving linear regression. Control Clin Trials 19: 589–601.

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.

ERRERA FI ET AL. 2007. 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 30: 275–279.

FOLKMAN J. 2002. Role of angiogenesis in tumor growth and metastasis. Semin Oncol 29 (6 Suppl 16): 15–18.

FUKAI N ET AL. 2002. Lack of collagen XVIII/endostatin results in eye abnormalities. EMBO J 21: 1535–1544.

GESTA S ET AL. 2006. Evidence for a role of developmental genes in the origin of obesity and body fat distribution. Proc Natl Acad Sci USA 103: 6676–6681.

GUO YF ET AL. 2006. 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.

INOUE-MURAYAMA M, SUGIMOTO Y, NIIMI Y AND ASO H. 2000. Type XVIII collagen is newly transcribed during bovine adipogenesis. Differentiation 65: 281–285.

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.

KENNELL JA AND MACDOUGALD AO. 2005. Wnt signaling inhibits adipogenesis through beta-catenin-dependent and -independent mechanisms. J Biol Chem 280: 24004–24010.

LOURENÇO GJ ET AL. 2006. High risk of occurrence of sporadic breast cancer in individuals with the 104NN polymorphism of the *COL18A1* gene. Breast Cancer Res Treat 100: 335–338.

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: 1298–1303.

MARNEROS AG AND OLSEN BR. 2005. Physiological role of collagen XVIII and endostatin. Faseb J 19(7): 716–728.

MILLER SA, DYKES DD AND POLESKY HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16: 1215.

MOULTON KS, OLSEN BR, SONN S, FUKAI N, ZURAKOWSKI D AND ZENG X. 2004. Loss of collagen XVIII enhances neovascularization and vascular permeability in atherosclerosis. Circulation 110: 1330–1336.

MUSSO O ET AL. 2001. Tumor hepatocytes and basement membrane-producing cells specifically express two different forms of the endostatin precursor, collagen XVIII, in human liver cancers. Hepatology 33: 868–876.

NISHIMURA S ET AL. 2007. Adipogenesis in Obesity Requires Close Interplay Between Differentiating Adipocytes, Stromal Cells, and Blood Vessels. Diabetes 56: 1517–1526.

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.

O'REILLY MS ET AL. 1997. Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. Cell 88: 277–285.

PASSOS-BUENO MR ET AL. 2006. Mutations in Collagen 18A1 (*COL18A1*) and their relevance to the human phenotype. Anais Acad Bras Cienc 78: 123–132.

QUELARD D ET AL. 2006. The frizzled module of col18a1 induces tumor cell death through inhibition of the wnt-bcatening signaling pathway. Hepatology, AASLD Abstract, 588A p.

ROSEN ED. 2005. The transcriptional basis of adipocyte development. Prostaglandins Leukot Essent Fatty Acids 73: 31–34.

ROSS SE ET AL. 2000. Inhibition of adipogenesis by Wnt signaling. Science 289: 950–953.

RUPNICK MA ET AL. 2002. Adipose tissue mass can be regulated through the vasculature. Proc Natl Acad Sci USA 99: 10730–10735.

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.

SERTIÉ AL, SOSSI V, CAMARGO AAA, 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.

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.

UDAGAWA T, FERNANDEZ A, ACHILLES E, FOLKMAN J AND AMATO RJ. 2002. Persistance of microscopic human cancers in mice: alterations in the angiogenic balance accompanies loss of tumor dormancy. Faseb J 16: 1361–1370.

VAINIO S, LIN Y AND PIHLAJANIEMI T. 2003. Induced Repatterning of Type XVIII Collagen Associates with Ectopic Sonic Hedgehog and Lung Surfactant C Gene Expression and Changes in Epithelial Epigenesis in the Ureteric Bud. J Am Soc Nephrol 14 (Suppl 1): S3–8.

VANDESOMPELE J, DE PRETER K, PATTYN F, POPPE B, VAN ROY N, DE PAEPE A AND SPELEMAN F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3(7): RESEARCH0034.

WHO. 2000. Obesity: preventing and managing the global epidemic. In: REPORT ON A WHO CONSULTATION. Geneva, World Health Organization, Technical Report Series 894.

WODARZ A AND NUSSE R. 1998. Mechanisms of Wnt signaling in development. Annu Rev Cell Dev Biol 14: 59–88.

ZORICK TS ET AL. 2001. High serum endostatin levels in Down syndrome: implications for improved treatment and prevention of solid tumours. Eur J Hum Genet 9: 811–814.

ZUK PA ET AL. 2002. Human adipose tissue is a source of multipotent stem cells. Mol Biol Cell 13: 4279–4795.