REBECA KAWAHARA

PROTEÔMICA BASEADA EM ESPECTROMETRIA DE MASSAS NA DESCOBERTA DE CANDIDATOS À BIOMARCADORES EM CÂNCER E NA IDENTIFICAÇÃO DE NOVOS ALVOS DE ADAM17

USING MASS SPECTROMETRY-BASED PROTEOMICS TO DISCOVER CANCER CANDIDATE BIOMARKERS AND TO IDENTIFY NOVEL ADAM17 SUBSTRATES

CAMPINAS
2015

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

REBECA KAWAHARA

PROTEÔMICA BASEADA EM ESPECTROMETRIA DE MASSAS NA DESCOBERTA DE CANDIDATOS À BIOMARCADORES EM CANCER E NA IDENTIFICAÇÃO DE NOVOS ALVOS DE ADAM17

USING MASS SPECTROMETRY-BASED PROTEOMICS TO DISCOVER CANCER CANDIDATE BIOMARKERS AND TO IDENTIFY NOVEL ADAM17 SUBSTRATES

Tese de doutorado apresentada ao Instituto de Biologia da Universidade Estadual de Campinas, como parte dos requisitos exigidos para obtenção do título de Doutora em Biologia Funcional e Molecular, na área de Bioquímica.

Thesis presented to the Institute of Biology .of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Functional and Molecular Biology, in Biochemistry.

Orientador: Dra. Adriana Franco Paes Leme

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELA ALUNA REBECA KAWAHARA E ORIENTADA PELA DRA. ADRIANA FRANCO PAES LEME.

Adriana Franco Paes Leme

CAMPINAS

2015

FICHA CATALOGRÁFICA

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

K179p

Kawahara, Rebeca, 1988-
Proteômica baseada em espectrometria de massas na descoberta de candidatos à biomarcadores em câncer e na identificação de novos alvos de ADAM17 / Rebeca Kawahara. – Campinas, SP : [s.n.], 2015.

Orientador: Adriana Franco Paes Leme.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Proteômica. 2. Espectrometria de massas. 3. Câncer. 4. Marcadores biológicos. 5. Fator de necrose tumoral-alfa convertase. I. Leme, Adriana Franco Paes. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Using mass spectrometry-based proteomics to discover cancer candidate biomarkers and to identify novel ADAM17 substrates
Palavras-chave em inglês:
Proteomics
Mass spectrometry
Cancer
Biological markers
Tumor necrosis factor-alpha convertase
Área de concentração: Bioquímica
Titulação: Doutora em Biologia Funcional e Molecular
Banca examinadora:
Adriana Franco Paes Leme [Orientador]
Giuseppe Palmisano
Inácio de Loiola Meirelles Junqueira de Azevedo
Vitor Marcel Faça
Márcio Chaim Bajgelman
Data de defesa: 28-04-2015
Programa de Pós-Graduação: Biologia Funcional e Molecular

Campinas, 28 de abril de 2015

BANCA EXAMINADORA

Dra. Adriana Franco Paes Leme (Orientadora) — Assinatura

Prof. Dra. Carmen Verissima Ferreira Halder — Assinatura

Prof. Dr. Giuseppe Palmisano — Assinatura

Prof. Dr. Inácio de Loiola Meirelles Junqueira de Azevedo — Assinatura

Prof. Dr. Vitor Marcel Faça — Assinatura

Dr. Celso Eduardo Benedetti — Assinatura

Prof. Dra. Fernanda Ramos Gadelha — Assinatura

Dr. Márcio Chaim Bajgelman — Assinatura

RESUMO

A espectrometria de massas aplicada à análise proteômica permite a caracterização e quantificação em larga-escala de proteínas expressas em uma determinada condição ou sistema biológico. Na presente tese, mostramos por meio de diferentes abordagens a utilidade da espectrometria de massas (1) no estudo de proteínas secretadas ou clivadas da superfície (secretoma) por células tumorigênicas como fonte de biomarcadores em câncer (2) na validação de candidatos biomarcadores em saliva humana de pacientes com câncer oral e (3) no estudo de alvos de uma metaloprotease envolvida na progressão de câncer, a ADAM17 (*A disintegrin and metalloproteinase 17*). Os resultados gerados pela espectrometria de massas juntamente com ferramentas estatísticas e de bioinformática para visualização de dados em agrupamento, *heatmaps*, rede de interação proteína-proteína, enriquecimento de vias e processos biológicos indicaram proteínas específicas direcionadas à hipótese que puderam ser validadas por métodos complementares quantitativos e funcionais. No contexto de biomarcadores em câncer, um painel de marcadores foi indicado a partir da análise do secretoma de células de carcinoma, melanoma e não cancerosas. Dentre os candidatos em carcinoma, proteínas envolvidas no sistema complemento foram validadas com expressão aumentada tanto em tecidos (proteínas C3 e CFB) quanto em saliva de pacientes com carcinoma oral de células escamosas (proteínas C3, CFB, C4B e SERPINA1). Agrina e perlecan, também observados com expressão alterada no secretoma de células de carcinoma e validados com expressão aumentada em tecidos de paciente com câncer oral, foram estudados quanto à função em processos tumorigênicos: migração, adesão, proliferação e sensibilidade à cisplatina. No estudo do degradoma da ADAM17, mimecan, perlecan e glypican-1 foram revelados como novos alvos dessa protease, e os sítios de clivagem determinados por espectrometria de massas. Além disso, funções e processos biológicos associados à clivagem de glypican-1 também foram estudados por meio da análise de seus parceiros de interação no meio extracelular seguido de ensaios funcionais. Em resumo, a espectrometria de massas esteve presente em todos os passos do método científico, desde a geração de hipóteses por meio da proteômica baseada em descoberta, até o teste e confirmação de hipóteses por meio da proteômica baseada em alvos.

ABSTRACT

Mass spectrometry applied to proteomic analysis allows large-scale characterization and quantification of proteins in specific conditions or biological systems. In this thesis, we showed different mass spectrometry-based approaches for (1) the study of cell secreted or released from the cell-surface proteins (secretome) as a source of candidate biomarkers in cancer (2) validation of candidate biomarkers in human saliva of patients with oral cancer and (3) the study of the target substrates of an important metalloprotease involved in cancer progression, the ADAM17 (A disintegrin and metalloproteinase 17). Together with statistical and bioinformatics tools, such as clustering, heatmaps, network analysis and enrichment analysis, we found specific hyphothesis-driven proteins that were further validated using complementary quantitative and functional methods. In the context of cancer biomarker, a panel of candidate biomarkers was revealed from the secretome analysis of carcinoma, melanoma and non-tumorigenic cell lines. Among the carcinoma candidates, proteins involved in complement system were validated with higher expression in tissues (C3 and CFB) and saliva (C3, CFB, C4B and SERPINA1) from patients with oral squamous cell carcinoma. The function of agrin and perlecan, which were also found with increased expression in squamous cell carcinoma tissue, were studied in tumorigenic processes: migration, adhesion, proliferation and sensibility to cisplatin. In the context of ADAM17 degradome, mimecan, perlecan and glypican-1 (GPC1) were revealed as novel substrates and their cleavage sites were determined using mass spectrometry. Besides, the function and biological processes associated with GPC1 shedding were studied by identifying the binding partners of GPC1 in the extracellular media and by performing functional assays. In summary, mass spectrometry was present in all part of the scientific method, from hyphothesis generation using discovery proteomics to hyphothesis testing using targeted proteomics.

SUMÁRIO

À Deus, pela Sua graça e misericórdia, que me deu força
e capacidade para chegar até aqui.

"Em Deus estão escondidos todos os tesouros da sabedoria e da ciência".

Colossenses 2:3

AGRADECIMENTOS

À minha orientadora, mestre e sempre "chefa", Profa. Adriana, minha profunda gratidão por me ensinar e me inspirar a arte de fazer ciência. Agradeço por cada projeto confiado, por cada minuto dedicado em corrigir, analisar, pensar, discutir, decidir e orientar; por estar sempre presente nos momentos que precisei, por me formar no conhecimento e prática em espectrometria de massas e por me mostrar que posso ir sempre além, com dedicação, planejamento e fé.

Ao meu supervisor de estágio no exterior, Dr. Michael MacCoss pela oportunidade de desenvolver parte do doutorado em um laboratório referência na área de *targeted proteomics*. Pelo grande aprendizado com os alunos e pesquisadores da *University of Washington* e pela maravilhosa experiência cultural e profissional vivenciada.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio financeiro durante os três anos de doutorado no Brasil e os três meses de estágio no exterior. À Unicamp pelo programa de pós-graduação em Biologia Funcional e Molecular. Ao Laboratório Nacional de Biociências (LNBio/CNPEM) pela infra-estrutura privilegiada concedida para que o trabalho pudesse ser desenvolvido com qualidade e excelência. À todos os professores e funcionários dessas instituições que direta ou indiretamente estiveram envolvidos com meu trabalho, muito obrigada!

À minha mãe Sra. Vera e meu pai Sr. Celso minha eterna gratidão pelo amor incondicional, pelo suporte emocional, espiritual e financeiro. Pela formação do meu caráter e educação os quais são essenciais para que eu possa alçar voos cada vez mais longes. Ao meu maninho querido Samuel agradeço por ser tão carinhoso e paciente comigo, mostrando-me que uma mente brilhante é aquela que consegue pensar diferente.

Ao amor da minha vida, Max, que me ensinou o sentido da vida que, acima do ouro e da prata, do conhecimento e da ciência e até da própria fé, é amar e ser amado.

À família Sakuma, dona Márcia e Sr. Ossamu, pelo acolhimento e cuidado todos os finais de semana, minha gratidão pelo privilégio de ter uma sogra e um sogro tão especiais!

Aos amigos do laboratório que me acompanharam e me ajudaram nessa jornada em busca das respostas às intermináveis perguntas biológicas: Sami, Dani, Anne, Fernando, Fred, Flávia Zando, Carol Carnielli, Carol Carneiro, Rute, Tati, César, Mateus, Romênia, Bianca e Flávia W. Em especial, agradeço as amigas e irmãs Sami e Dani pela paciência e compreensão aos meus

conflitos e ansiedades e por me ajudarem a crescer como pessoa, através do auto-conhecimento, do equilíbrio emocional e da aceitação.

Aos coautores e colaboradores de todos os trabalhos desenvolvidos durante o meu doutorado, em especial à Dra. Gabriela Meirelles e ao prof. Dr. Ricardo Coletta, minha gratidão por cada idéia, análise, experimento e correção.

Todos vocês foram presentes de Deus na minha vida e peça fundamental em todo o meu desenvolvimento como pessoa e cientista. Sem vocês com certeza jamais teria chegado até aqui.

MUITO OBRIGADA!!!

LISTA DE ABREVIATURAS E SIGLAS

2DE	Two-dimensional gel electrophoresis
AC	Corrente contínua
ADAM	A disintegrin and metalloproteinase
ADAMTS	A disintegrin and metalloproteinase with thrombospondin motifs
AFP	Alpha fetoprotein
CEA	Carcinoembryonic antigen
CEC	Carcinoma de células escamosas
CFB	Complement factor B
CID	Collision induced dissociation
DDA	Data-dependent acquisition
ECD	Electron capture dissociation
EGFR	Epidermal growth factor receptor
ESI	Electrospray ionization
ETD	Electron-transfer dissociation
FGF	Fibroblast growth factor
GPC1	Glypican-1
GPCRs	G protein–coupled receptors
HB-EGF	Proheparin-binding EGF-like growth factor
HCD	High energy collisional dissociation
HGF	Hepatocyte growth factor
HR/AM	High resolution/accurate mass
ICAT	Isotope-coded affinity tag
IGF	Insulin-like growth factor
iTRAQ	Isobaric tags for relative and absolute quantitation
LIT	Linear ion trap
MALDI	Matrix-assisted laser desorption/ionization
mEFs	Mouse embrionic fibroblasts
MMP	Metaloproteases de matrix
MRM	Multiple reaction monitoring
MudPIT	Multidimensional Protein Identification Technology
MβCD	Methyl-β-ciclodextrin
NSC	Nearest Shrunken Centroid
PDGF	Platelet-derived growth factor
PICS	Proteomic identification of protease cleavage sites
PMA	Phorbol 12-myristate 13-acetate
PMF	Peptide Mass Fingerprint

PRM	Parallel reaction monitoring
PSA	Prostate specific antigen
RF	Radiofrequência
RNAm	RNA mensageiro
SLRPs	Small leucine rich proteoglycans
SILAC	Stable isotope labeling by amino acids in cell culture
SIM	Single ion monitoring
Sp	Preliminary score
SRM	Monitoramento seletivo de reações
SVM-RFE	Support vector machine-recursive features elimination
TACE	Tumor necrosis factor-α-converting enzyme
TAILS	Terminal amine isotopic labeling of substrates
TGF-α	Transforming growth factor alpha
TGF-β	Transforming growth factor beta
TIMP	Tissue inhibitor of metalloproteinase
TNFα	Tumor necrosis factor alpha
TOF	Time-of-flight
VEGF	Vascular endothelial growth factor
Xcorr	Cross-correlation score
XIC	Extracted ion chromatogram

1. INTRODUÇÃO

1.1 Proteômica baseada em espectrometria de massas

O termo proteômica foi pela primeira vez definido em 1995 como conjunto de proteínas complementares ao genOMA (*PROTEin complement to a genOME*) (Wasinger, Cordwell *et al.*, 1995). O genoma humano contém aproximadamente 24.000 genes condificantes (Clamp, Fry *et al.*, 2007). No entanto, a transcrição destes para moléculas funcionais como RNA mensageiro (RNAm) e proteínas está sob um fino controle de expressão, resultando em diferentes conteúdos dessas moléculas dependendo da célula ou condição biológica (Yates, 1998).

O conjunto de transcritos (transcriptoma) de uma célula sob determinada condição pode ser estudado em larga escala utilizando técnicas como *micro-arrays* ou *RNA sequencing* (Wang, Gerstein *et al.*, 2009). Todavia, do transcrito à proteína, existem inúmeras modificações que exercem grande impacto na função final desta na célula, tais como: modificações pós-transcricionais (splicings alternativos, degradação do RNAm, etc.) e pós-traducionais de proteínas (proteólise, glicosilação, fosforilação etc). Não somente complementar ao genoma, a proteômica surgiu da necessidade de uma melhor compreensão biológica, fenotípica e funcional, proporcionada pelo conjunto de proteínas codificado pelos genes. Muito mais que identificação de proteínas, a proteômica abrange a caracterização quantitativa, funcional, estrutural e pós-traducional de proteínas bem como possibilita determinar relações entre elas (interactoma) e delas com o sistema biológico (Bensimon, Heck *et al.*, 2012).

Historicamente, a proteômica baseada em espectrometria de massas teve início com o uso da separação de proteínas por gel bidimensional (*two-dimensional gel electrophoresis, 2DE*), cujo princípio consiste em separar as proteínas de acordo com seu ponto isoelétrico pela focalização isoelétrica, na primeira dimensão e, de acordo com suas massas moleculares, na segunda dimensão, seguido ou não pela análise dos spots por espectrometria de massas para identificação das proteínas diferencialmente expressas (O'farrell, 1975). Posteriormente, a proteômica utilizando gel bidimensional foi combinada com técnicas de marcação por sondas fluorescentes, dando origem ao 2DE-DIGE (*Fluorescence Difference Gel Electrophoresis*) (Unlu, Morgan *et al.*, 1997). Nessa técnica, as amostras são marcadas separadamente com diferentes sondas fluorescentes (Cy2, Cy3, ou Cy5) e combinadas a fim de que a corrida seja realizada no mesmo gel 2DE, minimizando dessa forma a variação experimental e facilitando a comparação entre a mesma proteína proveniente de

diferentes amostras (*spot matching*) (Larbi e Jefferies, 2009). Embora a proteômica por gel bidimensional tenha trazido importantes contribuições para área, essa técnica limita-se em sensibilidade, faixa dinâmica e reprodutibilidade (Koomen, Haura *et al.*, 2008).

O rápido avanço da proteômica nos últimos anos foi possível principalmente devido ao desenvolvimento de duas tecnologias: primeiro, a emergência de novas técnicas brandas de ionização como a ionização por eletrospray (*electrospray ionization*, ESI), desenvolvida por John B. Fenn[1] e ionização a laser auxiliada por matriz (*matrix-assisted laser desorption/ionization*, MALDI), desenvolvida por Koichi Tanaka,[1] Franz Hillenkamp e Michael Karas e segundo, a miniaturização e automatização da cromatografia líquida. Juntos, essas tecnologias possibilitaram mensurar e identificar peptídeos com alta sensibilidade em amostras biológicas bastante complexas (Mallick e Kuster, 2010). Somam-se a isso o surgimento de novos espectrômetros de massas com maior sensibilidade, resolução e exatidão de massas, bem como a nova geração de algoritmos computacionais capazes de analisar dados proteômicos, permitiram que a proteômica baseada em espectrometria de massas se tornasse uma poderosa ferramenta para identificar e quantificar proteínas em larga escala.

A possibilidade de se realizar proteômica sem separação por gel de eletroforese bidimensional deu origem, em 1998, ao termo *shotgun protein analysis*[2]. Desenvolvido por Yates e colaboradores, a proteômica por *shotgun* consiste na combinação de digestão enzimática de proteínas, seguida de separação por cromatografia líquida e análise por espectrometria de massas sequencial[3] (*tandem mass spectrometry*) (Yates, 1998). Essa técnica propiciou um ganho significante em eficiência e sensibilidade na análise de misturas complexas de proteínas, pois automatizou o processo de separação de proteínas, minimizando perda de amostra, além de possibilitar utilização de cromatografia em escala nano-fluxo (Yates, 2013).

Posteriormente, técnicas de fracionamento de misturas complexas foram desenvolvidas a fim de aumentar a cobertura do proteoma. Uma das mais conhecidas, também desenvolvido pelo grupo do pesquisador Yates, é a tecnologia multidimensional de identificação de proteínas; do

[1] Ganhadores do Prêmio Nobel em Química no ano de 2002. Embora o Prêmio Nobel tenha sido dado apenas ao Koichi Tanaka para o desenvolvimento da ionização por MALDI, dois cientistas alemães Franz Hillenkamp e Michael Karas também foram essenciais para o desenvolvimento dessa técnica e merecem ser mencionados.

[2] A técnica *shotgun proteomics* se enquadra dentro da estratégia "*bottom up*", cujo método de identificação de proteínas está baseada na digestão proteolítica das amostras previamente à análise por LC-MS.

[3] Aquisição e estudo de espectros de íons produtos ou precursores originados a partir de íons selecionados em um primeiro estágio de análise de massas (Vessecchi, 2011).

inglês, *Multidimensional Protein Identification Technology* (MudPIT). Resumidamente, essa tecnologia utiliza duas dimensões de separação por cromatografia líquida: na primeira dimensão as proteínas são separadas com base na carga, em colunas de troca catiônica, e a na segunda dimensão, a separação é baseada na hidrofobicidade, em colunas de fase reversa (Washburn, 2004). Essa tecnologia trouxe uma significante melhora na faixa dinâmica e cobertura do proteoma, porém o aumento de etapas no processo, a baixa reprodutibilidade, bem como o tempo de instrumentação associado ao custo elevado para análise são ainda os principais limitantes da sua utilização. Com o desenvolvimento de técnicas de marcação de proteínas ou peptídeos (discutidas com mais detalhes na parte II da introdução), essas limitações foram minimizadas, pois proteomas complexos provenientes de amostras diferentes são analisados em uma mesma corrida cromatográfica, reduzindo a variabilidade e as etapas no processo (Walther e Mann, 2010).

Atualmente, a proteômica baseada em espectrometria de massas tem se expandido de estratégias baseada em descoberta, nas quais o objetivo é a identificação do maior número de proteínas sem o conhecimento prévio da amostra, para estratégias baseada em alvos, cujo foco é a quantificação de um pequeno conjunto de proteínas conhecidas previamente à análise (Picotti, Bodenmiller *et al.*, 2013).

Assim, a proteômica baseada em espectrometria de massas abrange uma grande diversidade de técnicas e abordagens, com diferentes objetivos e desafios técnicos e computacionais. Embora ela consista de uma poderosa tecnologia para responder perguntas biológicas, algumas aplicações requerem tecnologias sofisticadas, infraestrutura de laboratório e equipamentos adequados, além de pessoas qualificadas e treinadas. Importante salientar que o sucesso de um experimento proteômico também depende de um desenho experimental correto e de ferramentas para interpretação dos dados de forma produzir conclusões com significado biológico que direcionem a experimentos complementares apropriados (Mallick e Kuster, 2010).

A Figura 1 resume algumas aplicações em diferentes contextos e complexidades biológicas com diferentes níveis de dificuldade técnica.

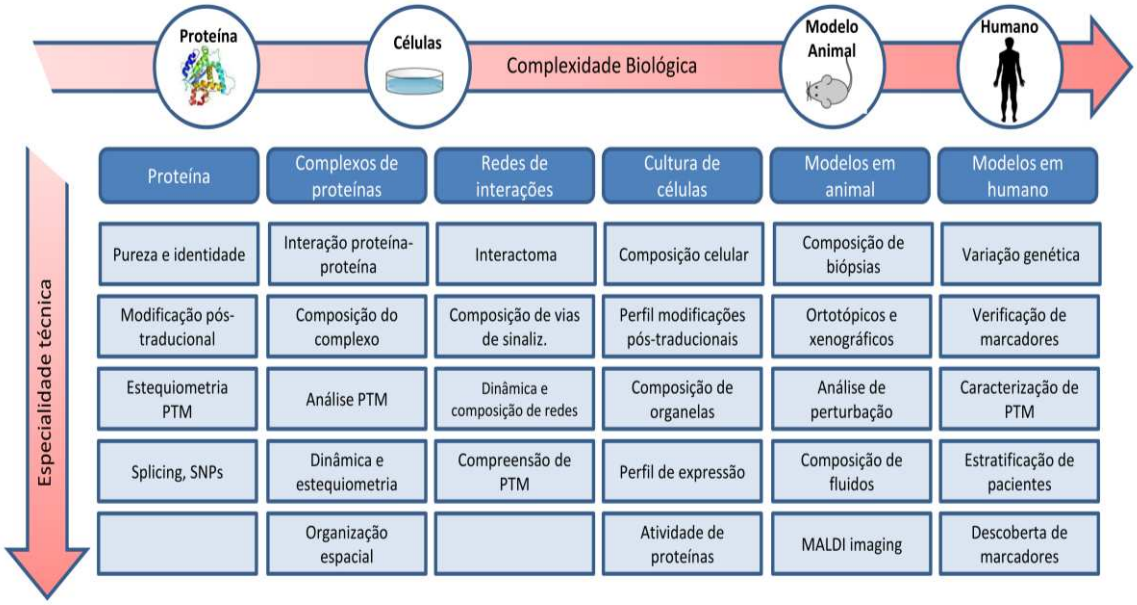

Figura 1 Aplicações da proteômica baseada em espectrometria de massas. Essa figura mostra uma comparação entre as diferentes aplicações da proteômica em termos de complexidade biológica e especialidade técnica requerida. Adaptado de Mallick e Kuster, 2010 (Mallick e Kuster, 2010).

Nesta tese, diversas dessas aplicações foram realizadas e muitas informações foram adquiridas por meio da proteômica baseada em espectrometria de massas. Da descoberta ao alvo, utilizamos a espectrometria de massas para: determinar a composição de proteínas secretadas ou clivadas no meio extracelular de cultura de células (secretoma), buscar por novos alvos de metaloprotease e determinar sítios de clivagem (degradoma), determinar parceiros de interação de proteoglicanos extracelulares (interactoma), quantificar candidatos à biomarcadores em saliva de pacientes com câncer oral, entre outros.

PARTE I: CONTEXTOS BIOLÓGICOS DESSA TESE

1.2 O câncer oral

Dados recentes do Instituto Nacional de Câncer (Instituto Nacional De Câncer, 2014) apontaram o câncer da cavidade oral como um problema de saúde pública com 11.280 casos em homens e 4.010 casos em mulheres estimados, no Brasil, no ano de 2014. Dentre as regiões brasileiras de maior incidência encontram-se o Sudeste e o Nordeste, nos quais esse tipo de neoplasia configura o quarto e nono mais frequente em homens e mulheres, respectivamente. Em termos mundiais, o câncer oral do tipo carcinoma de células escamosas (CEC) representa de 2-4% de todos os novos casos diagnosticados (Markopoulos, 2012).

Histologicamente, o CEC oral corresponde a aproximadamente 90% dos tipos de câncer da cavidade bucal, sendo as localizações de maior risco e pior prognóstico a língua e o assoalho bucal, pois são as regiões que com maior frequência desenvolvem metástase regional em linfonodos (Zini, Czerninski *et al.*, 2010).

O CEC oral é uma neoplasia derivada do epitélio escamoso estratificado da mucosa oral, podendo ou não estar associada a presença de uma lesão prévia. Uma lesão pode evoluir através de uma série de estágios histopatológicos, de uma hiperplasia benigna, passando por uma displasia epitelial e chegando a um carcinoma in situ e deste para carcinoma francamente invasivo (Shah, Begum *et al.*, 2011). Entre as lesões potencialmente malignas mais importantes para o CEC oral estão a leucoplasia e a eritroplasia (Markopoulos, 2012).

O carcinoma oral de células escamosas se origina a partir de uma combinação de fatores de riscos ambientais e genéticos. Em combinação com fatores genéticos individuais, o tabagismo, o etilismo, infecções por HPV, principalmente pelo tipo 16, e exposição à radiação UVA solar estão entre os principais fatores de risco para essa neoplasia (Feller e Lemmer, 2012).

A detecção de CEC oral é baseada em exame clínico da boca combinado com biópsia para análise histopatológica. Embora a histologia seja o padrão ouro na confirmação da presença da neoplasia, essa ferramenta ainda é limitada no acompanhamento de lesões potencialmente malignas, as quais podem ou não evoluir para uma neoplasia (Yakob, Fuentes *et al.*, 2014). Além disso, muitas lesões se encontram em regiões difíceis de visualizar em um exame clínico e em casos mais precoces passam imperceptíveis para o paciente e profissional da saúde. Por esses motivos o diagnóstico de CEC oral acontece, em muitos casos, de forma tardia, quando essa neoplasia já está

em estágios mais avançados, comprometendo o prognóstico e o tratamento. Tem-se demonstrado que a sobrevida de CEC oral diagnosticados precocemente, nos quais o estágio tumoral ainda é recente e não há evidências de metástase regional ou a distância, é de 80-90% em 5 anos, enquanto que em casos avançados a sobrevida é menor que 50% em 5 anos (Zini, Czerninski *et al.*, 2010; Feller e Lemmer, 2012).

O CEC oral possui um curso clínico muito variável devido à natureza heterogênea do tumor e às diferentes respostas individuais frente à terapia. Atualmente o sistema de estadiamento TNM (T: extensão do tumor primário, N: ausência ou presença e extensão de metástase em linfonodos regionais e M: ausência ou presença de metástase à distância) é o principal marcador de prognóstico usado para o CEC oral. Porém, muitos pacientes com o mesmo estágio da doença podem apresentar uma evolução clínica e tempo de sobrevida distintos (Costa, De Araújo Júnior *et al.*, 2005). Importante destacar que apesar dos recentes avanços em termos de prevenção e multimodalidade no tratamento do CEC oral, o prognóstico para casos mais avançados não sofreu melhoras nos últimos 20 anos (Da Silva, Ferlito *et al.*, 2011)

A presença de metástase em linfonodo é um dos fatores prognósticos mais determinantes na sobrevida do paciente. Foi evidenciado que cerca de 40% dos pacientes com CEC oral desenvolvem metástase em linfonodo. Desses, aproximadamente 25-40% sobrevivem em 5 anos. Em contraste, paciente sem metástase possuem uma sobrevida de 90% em 5 anos (Noguti, De Moura *et al.*, 2012). Outro fator importante no prognóstico é a recidiva tumoral após tratamento, a qual acomete cerca de 15-33% dos pacientes (Mucke, Wagenpfeil *et al.*, 2009). Em um estudo de 275 casos de pacientes com CEC oral (Wang, Zhang *et al.*, 2013), a sobrevida em 5 anos de pacientes com recorrência após tratamento foi de 31,8% enquanto que em pacientes não recorrentes a sobrevida nesse mesmo período foi de 79,9%.

A terapia para o CEC oral é baseada em três modalidades: cirúrgica, radioterápica e quimioterápica, sendo a escolha da estratégia terapêutica ditada pela natureza do carcinoma, incluindo sítio de origem, tamanho, características histopatológicas e envolvimento ou não de metástase (sistema TNM), e pelas condições clínicas gerais do paciente (Shah e Gil, 2009). A cirurgia é o tratamento de escolha para casos de CEC oral pequenos e localizados. Em casos avançados o tratamento cirúrgico é geralmente combinado com radioterapia e quimioterapia. Em casos de CEC oral recorrente, o tratamento de escolha consiste em quimioterapia combinada a fármacos inibidores de receptor de fator de crescimento epidermal (EGFR) (Lee, Veness *et al.*, 2005; Feller e Lemmer, 2012).

Diante do exposto, torna-se evidente a necessidade da descoberta de novos biomarcadores em CEC oral para auxiliar as atuais ferramentas de diagnóstico, sobretudo em casos precoces, bem como de prognóstico principalmente no que se refere ao curso clínico do tumor, isto é, se pode evoluir para casos agressivos (metástase) ou apresentar recidiva tumoral. Dessa forma, com o uso de biomarcadores clínicos, estratégias terapêuticas podem ser direcionadas de forma a possibilitar maior sucesso no tratamento e na melhora na sobrevida do paciente.

1.3 Saliva como fonte de biomarcadores em câncer oral

Biomarcador é definido pela FDA (*United States Food and Drug Administration*) como uma característica que é objetivamente mensurada e usada como indicador de um processos biológico normal ou patológico ou ainda em resposta a uma intervenção terapêutica. Marcadores moleculares podem ser DNA, mRNA, proteínas ou metabólitos e na prática clínica são utilizados para diagnóstico, prognóstico, monitoramento de resposta terapêutica, estadiamento de doenças, etc. Inúmeros têm sido os investimentos, considerando a ordem de importância deles, e esforços para descoberta de novos biomarcadores de interesse clínico (Biomarkers Definitions Working, 2001; Fuzery, Levin *et al.*, 2013).

Em câncer oral, a saliva tem sido bastante estudada como fonte na busca de biomarcadores, pois é um fluido fácil, simples e rápido de ser coletado, além de não invasivo. Importante ressaltar que esse fluido está em contato direto com a mucosa oral e com lesões malignas e potencialmente malignas, por isso tem se apresentado como promissora fonte de biomarcadores para diagnóstico ou prognóstico de CEC oral (Yakob, Fuentes *et al.*, 2014). Interessante notar que a saliva já foi descrita também como uma ferramenta útil para diagnóstico de câncer de mama (Bigler, Streckfus *et al.*, 2002), câncer de pulmão (Xiao, Zhang *et al.*, 2012), síndrome de Sjögren (Hu, Gao *et al.*, 2010), câncer de pâncreas (Lau, Kim *et al.*, 2013), entre outras doenças (Marti-Alamo, Mancheno-Franch *et al.*, 2012).

Recentemente, em um artigo de revisão, Yakob *et al*. (2014) (Yakob, Fuentes *et al.*, 2014) mostrou diversas ômicas aplicadas na descoberta de biomarcadores em câncer oral em saliva de pacientes. Desde os anos 90 até o presente, mais de 40 estudos foram publicados e mais de 100 diferentes constituintes foram sugeridos como potenciais biomarcadores salivares em CEC oral (revisado por Cheng, 2014) (Cheng, Rees *et al.*, 2014), dentre eles compostos inorgânicos, peptídeos, proteínas, DNA, RNAm, microRNA e metabólitos.

A saliva é um fluido secretado pelas glândulas maiores parótidas, submandibulares sublinguais e por inúmeras glândulas menores, localizadas principalmente na mucosa bucal. Esse fluido contém, além das secreções glandulares, componentes como fluído gengival, células descamativas, bactérias e produtos bacterianos, além de outros componentes variáveis. O principal componente da saliva é a água (99,5%), sendo as proteínas (0,3%) e espécies inorgânicas (0,2%) apenas uma porção bem menor no conteúdo desse fluido (Yakob, Fuentes *et al.*, 2014). Até o presente, mais de 2300 proteínas e peptídeos foram identificados na saliva humana. As mais abundantes, correspondendo a 98% do proteoma da saliva, são: α-amilase, albumina, cistatinas, histatinas, *secretory-IgA*, lactoferrina, mucinas, lisozimas, proteínas ricas em prolinas, *estaterina* e transferrina (Cheng, Rees *et al.*, 2014).

Um dos grandes desafios da análise de biomarcadores em saliva é que os candidatos biomarcadores proteicos geralmente se encontram em pequenas concentrações e requerem métodos com alta sensibilidade de detecção. Além disso, fatores como (1) a falta de padronização na coleta e no processamento e armazenamento de amostra da saliva, (2) grande variabilidade nos níveis dos potenciais biomarcadores, tanto em indivíduos sem câncer quanto com câncer e (3) presença de outras condições como inflamação, doenças e outros tipos de neoplasias, têm limitado o uso da saliva como ferramenta para análise de biomarcadores em CEC oral (Cheng, Rees *et al.*, 2014).

1.4 Secretoma: um reservatório de moléculas bioativas

O termo secretoma foi primeiramente introduzido por Tjalsma *et al*. (2000) (Tjalsma, Bolhuis *et al.*, 2000) como sendo o conjunto de proteínas liberadas pelas células, tecido ou organismo por meio de diferentes mecanismos de secreção (clássica e não clássica). Essa definição foi posteriormente estendida para ectodomínios de proteínas de membranas (ex: receptores e fatores de crescimento) liberados para o meio extracelular através de um processo denominado *shedding* (Makridakis e Vlahou, 2010). Proteínas secretadas correspondem a aproximadamente 10-15% das proteínas codificadas pelo genoma humano; dentre elas estão fatores de crescimento, citocinas, receptores de superfície celular, moléculas de adesão, proteases, proteínas de matriz extracelular e proteínas intracelulares secretadas por meio de vesículas (Karagiannis, Pavlou *et al.*, 2010).

A via clássica de secreção envolve a presença de um peptídeo sinal, o qual direciona as proteínas para o retículo endoplasmático e, em seguida, para o complexo de Golgi, de onde são transportadas por meio de vesículas até a membrana plasmática e por fim liberadas para o meio

extracelular. A via não clássica de secreção é independente do transporte pelo retículo endoplasmático e complexo de Golgi, ocorrendo principalmente através de vesículas intracelulares e exossomos (Karagiannis, Pavlou *et al.*, 2010).

A baixa concentração de proteínas secretadas (geralmente encontram-se diluídas em fluídos biológicos ou meio de cultura celular), a presença de contaminantes intracelulares liberados em processos de lise e morte celular e a presença de proteínas de alta abundância como albumina e de soro fetal bovino (presente em meio de cultura de células) são as principais dificuldades encontradas para análise de proteínas secretadas por espectrometria de massas. Apesar disso, muitas têm sido as estratégias encontradas para otimizar o preparo de amostra, seja em métodos de enriquecimento do secretoma (concentração do meio extracelular por centrifugação em membranas com filtros de massas moleculares específicas ou precipitação de proteínas), fracionamento (gel ou cromatografia líquida), remoção de proteínas muito abundantes (ex: depleção de albumina em soro) e incubação de célula em meio isento de soro fetal bovino (Makridakis e Vlahou, 2010). Concomitantemente, o desenvolvimento de espectrômetros de massas mais sensíveis, velozes, de maior resolução e exatidão de massa possibilitou o grande avanço no estudo do secretoma, sobretudo relacionado ao câncer.

O estudo do secretoma em células de câncer tem ganhado cada vez mais importância principalmente no contexto de descoberta de biomarcadores e alvos terapêuticos, já que moléculas secretadas possuem alta probabilidade de entrarem na circulação e serem detectadas em fluidos biológicos, além de serem reguladores chaves de processos tumorigênicos como invasão, metástase, proliferação e angiogênese (Xue, Lu *et al.*, 2008; Stastna e Van Eyk, 2012).

Proteínas do secretoma possuem também um papel importante no microambiente tumoral, o qual é composto não somente por células cancerosas, mas também por células do estroma como fibroblastos, células endoteliais e células do sistema imune (Pietras e Ostman, 2010). A comunicação mediada por moléculas secretadas ou clivadas da membrana, tanto pelas células do tumor quanto do estroma, modula o microambiente de forma a deixá-lo favorável ao início, crescimento e progressão do tumor. Essa sinalização depende em grande parte de (1) fatores de crescimento mitogênico, como fator de crescimento de hepatócito (HGF), fator de crescimento de transformação α (TGF-α) e fator de crescimento derivado de plaquetas (PDGF); (2) sinais inibitórios de crescimento, como o fator de crescimento de transformação β (TGF-β); e (3) fatores tróficos, como os fatores 1 e 2 de crescimento semelhantes a insulina (IGF-1 e 2), os quais favorecem a sobrevivência celular (Weinberg, 2008).

A comunicação entre as células tumorais e do estroma continuam operar mesmo depois que os tumores são formados. Para citar alguns exemplos, células epiteliais de um carcinoma frequentemente liberam PDGF, para as quais células estromais – especialmente fibroblastos, miofibroblastos e macrófagos – possuem receptores; as células estromais alternam liberando IGF-1, o qual beneficia o crescimento e sobreviência das células cancerosas ao redor (Weinberg, 2008). As proteínas secretadas pelo tumor, podem também agir de forma autócrina ou parácrina, estimulando a produção de proteases, especialmente metaloproteases, responsáveis pela degradação da matriz extracelular, migração e invasão das células tumorais à tecidos adjacentes ou corrente sanguínea (Quail e Joyce, 2013).

Assim, o crescente interesse pelo estudo do secretoma tem surgido como estratégia para elucidar os mecanismos de progressão tumoral, bem como descobrir novos biomarcadores e alvos terapêuticos (Whiteside, 2008; Hanahan e Weinberg, 2011). A proteômica baseada espectrometria de massas tem sido essencial para esse fim, seja na caracterização do secretoma de células do tumor ou do estroma bem como de fluidos biológicos. Com o uso da espectrometria de massas, inúmeros candidatos à biomarcadores com expressão alterada em diferentes tipos de câncer ou outros contextos patológicos foram levantados, os quais podem ser futuramente estudados e validados para aplicações clínicas (Stastna e Van Eyk, 2012).

1.5 Proteoglicanos e câncer

Dentro do contexto de microambiente tumoral, a matriz extracelular é responsável pela sustentação, organização, comunicação e sinalização entre células do tumor e do estroma e entre células e o meio ao redor.

A matriz extracelular é composta por três classes principais de macromoléculas: 1. proteoglicanos, 2. proteínas fibrosas (colágeno, elastina, fibronectina e laminina) e 3. glicoproteínas (geralmente com função adesiva) (Frantz, Stewart *et al.*, 2010).

Os proteoglicanos são proteínas (exceto para ácido hialurônico que não possui cadeia polipeptídica) ligadas covalentemente a cadeias de repetições de unidades dissacarídicas não ramificadas chamadas glicosaminoglicanos, os quais podem ser sulfatados (sulfato de condroitina, heparam sulfato e queratam sulfato) ou não sulfatado (ácido hialurônico). Por ser altamente carregados, proteoglicanos possuem grande afinidade por água, formando um gel que fornece suporte mecânico e resistência para os tecidos (Theocharis, Skandalis *et al.*, 2010)

Os proteoglicanos podem ser agrupados de acordo com a localização celular: no meio extracelular se encontram os *hyalectans* e *small leucine rich proteoglycans (SLRPs)*; na membrana basal, perlecan, agrina e colágeno XVIII; e na superfície celular os sindecanos e glipicanos (Edwards, 2012) (Figura 2).

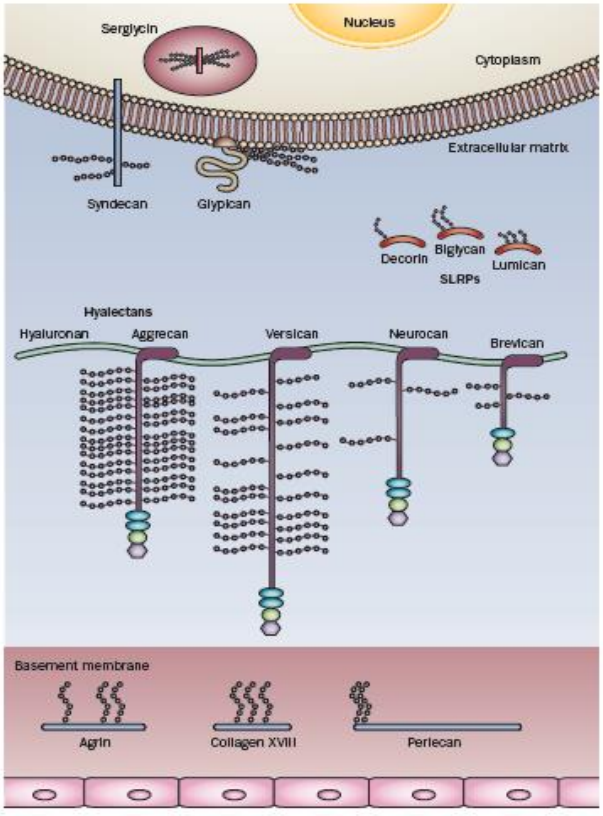

Figura 2 Classificação de proteoglicanos de acordo com a localização celular. O proteoglicano *serglycin* é encontrado em vesículas intracelular de células hematopoiéticas e endotelias. Três grupos de proteoglicanos extracelulares incluem: hyalectans (aggrecan, versican, neurocan e brevican), os quais se associam com hyaluronan na matriz extracelular; os SLRPs (decorin, biglycan e lumina) e os proteoglicanos de membrana basal (perlecan, agrina e colágeno XVIII). As duas principais famílias de proteoglicanos associados à superfície celular são os sindecanos e os glipicanos. Adaptado de Edwards, 2012.

Os proteoglicanos têm se destacado como importantes moléculas mediadoras da adesão célula-matriz extracelular, bem como reguladores da motilidade celular e da interação a fatores de crescimento como fatores de crescimento do endotélio vascular (VEGFs) e fatores de crescimento de fibroblastos (FGFs) com receptores. Há mais de 50 anos tem-se relatado o envolvimento de proteoglicanos na biologia de diversos tipos de câncer (Iozzo, Zoeller *et al.*, 2009; Iozzo e Sanderson, 2011). *Perlecan* foi encontrado com expressão aumentada em melanoma metastático

(Cohen, Murdoch *et al.*, 1994) e o silenciamento da expressão desse proteoglicano em tumor de próstata reduziu a responsividade à fator de crescimento tipo EGF ligado à heparina (HB-EGF) *in vitro* e crescimento tumoral *in vivo* (Savore, Zhang *et al.*, 2005). Sindecanos e glipicanos têm sido associados com diversos processos tumorigênicos como migração, angiogênese e proliferação (Qiao, Meyer *et al.*, 2003; Aikawa, Whipple *et al.*, 2008; Zong, Fthenou *et al.*, 2011; Aragao, Belloni *et al.*, 2012). Agrina foi encontrada com expressão aumentada em tumor de fígado (Tatrai, Dudas *et al.*, 2006; Somoracz, Tatrai *et al.*, 2010) e vias biliares (Batmunkh, Tatrai *et al.*, 2007), e sua função na progressão de câncer de fígado foi recentemente eluciada em modelos animais (Chakraborty, Lakshmanan *et al.*, 2015).

A função biológica dos proteoglicanos pode ser modulada por várias enzimas presentes no microambinete tumoral: (i) *sheddases* que clivam proteoglicanos de superfície, liberando ectodomínios funcionais (ii), proteases que clivam o core proteico dos proteoglicanos e (iii) heparanases e endosulfatases que modificam a estrutura e atividade dos heparan sulfatos e a ligação deles com fatores de crescimento (Iozzo e Sanderson, 2011).

Sindecanos por exemplo, podem ser clivados por metaloproteases de matriz (MMP7, MMP9 e MT-MMP1) ou metaloproteases de membrana como a ADAM17 (Pruessmeyer, Martin *et al.*, 2010). Uma vez solúveis no meio extracelular, fragmentos de sindecanos são capazes de exercer sinalização autócrina ou parácrina e induzir migração celular (Aragao, Belloni *et al.*, 2012). A função de perlecan também pode ser regulada através de heparanases, as quais são responsáveis por clivar cadeias de heparan sulfato em pequenos fragmentos de 5-7 kDa e dessa forma regular a biodisponibilidade e a atividade de ligantes de proteoglicanos como FGFs e VEGFs (Barash, Cohen-Kaplan *et al.*, 2010).

1.6 Metaloproteases: reguladores do microambiente tumoral

Uma das principais classes de proteínas mediadoras da modulação do microambiente durante a progressão de câncer são as metaloproteases. Elas são responsáveis não somente pela degradação da matriz extracelular - função esta primeiramente descrita a mais de 40 anos atrás - como também pela modulação e ativação, através da proteólise, de várias proteínas (citocinas, fatores de crescimento, moléculas de adesão entre outras), resultando em processos como remodelamento tecidual, inflamação, metástase, angiogênese e invasão (Gialeli, Theocharis *et al.*, 2011).

As metaloproteases pertencem à família de endopeptidases dependentes de zinco, o qual se liga a motivos conservados HEXXH no sítio ativo da enzima. Dependendo do terceiro resíduo de ligação a zinco que pode ser ácido glutâmico, histidina ou ácido aspártico, as metaloproteases são classificadas em três super-famílias, respectivamente: gluzincina, metzincina e aspzincina. Dentro da classe metzincina encontram-se as metaloproteases de matriz (MMPs) e as adamalisinas, também conhecidas como reprolisinas[4], das quais fazem parte as ADAMs (*A Disintegrin And Metalloproteinase*) e ADAMTS (*A Disintegrin And Metalloproteinase with Thrombospondin Motifs*)[5] (Wolfsberg, Straight *et al.*, 1995; Black e White, 1998; Nagase, Visse *et al.*, 2006).

Em termos estruturais, as metaloproteases são organizadas em multi-domínios. Elas compartilham semelhanças na presença de uma sequência sinal no N-terminal, um pró-domínio - contendo um grupo tiol (-SH) e um sítio de clivagem à furina - e um domínio catalítico com sítio de ligação à zinco, mas se diferem principalmente nos domínios subsequentes ao catalítico. Em geral, as MMPs possuem um domínio hemopexina e as ADAMs, um domínio do tipo disintegrina, seguido de uma região rica em cisteínas e tipo fator de crescimento epidermal (EGF), além de domínios transmembrana e citoplasmático. As ADAMTS possuem ainda um domínio trombospondina e geralmente são solúveis, pois não possuem o domínio transmembrana (Gooz, 2010; Kessenbrock, Plaks *et al.*, 2010).

As metaloproteases são sintetizadas como proteína precursora na forma de zimogênio, translocadas para o retículo endoplasmático e em seguida para o complexo de Golgi, onde sofrem um processo de maturação por meio da remoção do pró-domínio pela enzima convertase 7 (PC7) ou furina (Sternlicht e Werb, 2001). A função das MMPs e ADAMs depende principalmente do balanço entre elas e os respectivos inibidores teciduais de metaloproteases (TIMP) endógeno, os quais são capazes de regular, inibindo reversivelmente o sítio catalítico das MMPs e ADAMs.

Além da inibição pelas TIMPs, as metaloproteases podem ser reguladas em diferentes níveis: expressão gênica, compartimentalização, conversão entre forma zimogênio e ativa, modificações pós-traducionais e modulação pela interação com outras proteínas (Kessenbrock, Plaks *et al.*, 2010). Os mecanismos de regulação de metaloprotease têm sido extensivamente estudados como estratégia terapêutica para modular a atividade proteolítica e consequentemente

[4] Nomenclatura que faz referência ao fato de essas proteínas terem sido inicialmente isoladas de venenos de serpente (réptil) e do tecido reprodutivo (Wolfsberg et al., 1995)

[5] O termo ADAM remete ao nome dos dois domínios (disintegrina e metaloprotease) muito relacionados estruturalmente às metaloproteinases classe PIII de veneno de serpente, onde foi originalmente descrita (Black e White, 1998)

modular a progressão tumoral. No entanto, o uso de inibidores de metaloproteases na prática clínica não teve sucesso ao longo dos anos e isso se deve em parte devido ao papel contraditório que alguns membros das MMPs ou ADAMs exercem em favor ou contra a progressão do tumor. Exemplo disso está o estudo de Decock (2008) no qual foi mostrado um efeito anti-metastático em câncer de mama promovido pela super-expressão de MMP-8 (Decock, Hendrickx *et al.*, 2008). Em contrapartida, MMP-9 pode clivar proteínas da matriz extracelular, liberando fatores de crescimento como VEGF, TGFβ e FGF-2, os quais estimulam proliferação e migração de células endoteliais e consequentemente promover angiogênese e crescimento do tumor (Bauvois, 2012).

As ADAMs são as metaloproteases responsáveis predominantemente pela clivagem de ectodomínios de proteínas de membrana, os quais quando solúveis exercem diversas funções, sobretudo na sinalização autócrina ou parácrina mediante ligação a receptores de superfície. Segundo Reiss e Saftig (2009), 40 membros pertencentes às ADAMs foram identificados no genoma de mamíferos, sendo 21 descritos no genoma humano. Embora todas as ADAMs apresentem o domínio metaloprotease, em humanos somente 13 dos 21 genes da família codificam para proteases funcionais, sendo a ADAM10 e a ADAM17 as mais estudadas. A ADAM17, também conhecida como enzima conversora de fator de necrose tumoral α (*tumor necrosis factor-α-converting enzyme*, TACE), é a principal enzima ativadora de TNFα e conversora das pró-formas de ligantes de EGFR como HB-EGF, TGFα, anfiregulina e epiregulina (Sahin, Weskamp *et al.*, 2004). A ADAM10 por sua vez, é a principal *sheddase* de EGF e betacelulina (Duffy, Mullooly *et al.*, 2011), além de E-caderina (Maretzky, Reiss *et al.*, 2005) e mefrina-A (Herzog, Haun *et al.*, 2014).

A ativação de EGFR pela clivagem dos ligantes é um dos mecanismos mais conhecidos pelo qual a ADAM17 contribui para o desenvolvimento do tumor. Uma vez clivados, ligantes de EGFR ativam esse receptor, desencadeando uma série de vias *downstream* como p38MAPK, Akt/PKb, JNK e ERK, envolvidas em processos celulares como proliferação, migração, sobrevivência, hipertrofia celular, etc. (Ohtsu, Dempsey *et al.*, 2006) (Figura 3).

Além disso, tem-se proposto a participação da ADAM17 em um mecanismo de transativação de EGFR via receptores associados à proteína G (GPCRs) e seus agonistas. A produção de mensageiros secundários, como Ca^{+2} intracelular e proteína kinase C, regula a atividade da ADAM17 via fosforilação do domínio citoplasmático, levando à translocação da protease para superfície e consequentemente ativação e clivagem dos ligantes de EGFR (Figura 3).

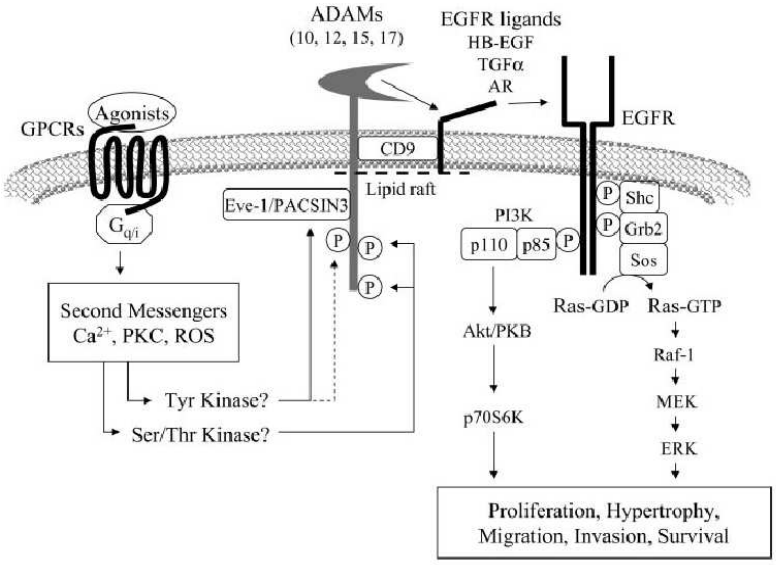

Figura 3 Mecanismo proposto de transativação de EGRF dependente de ADAMs via GPCR e as vias de sinalização em cascata. P13K, fosfatidil-inositol-3-quinase; ROS, espécies reativas de oxigênio; HB, ligante de heparina; AR, anfiregulina adaptado de Ohtsu *et al*., 2006.

A ADAM17 foi descrita super-expressa em vários tipos de câncer, incluindo cérebro, mama, gástrico, rim, fígado, pulmão, ovário, pâncreas e próstata (revisado por Murphy, 2008 (Murphy, 2008)), além de câncer de cabeça e pescoço (Stokes, Joutsa *et al.*, 2010). Em diversos modelos de célula foi mostrada que a inibição da expressão ou da atividade da ADAM17 reduz a progressão do tumor (Zheng, Jiang *et al.*, 2009; Chen, Chen *et al.*, 2013; Huang, Benaich *et al.*, 2014) e o oposto, isto é, a super-expressão promove maior crescimento tumoral, migração e angiogênese (Lin, Sun *et al.*, 2012; Zheng, Jiang *et al.*, 2012; Simabuco, Kawahara *et al.*, 2014). Além da clivagem de ligantes de EGFR, a ADAM17 também é responsável pela clivagem de moléculas de adesão como CD44, ALCAM, VCAM, L-selectina, além de receptores como Notho, IL-6R, IL-15R, TNFR e VEGFR2, liberando ectodomínios funcionais para o meio extracelular (Arribas e Esselens, 2009; Pruessmeyer e Ludwig, 2009). Por meio da clivagem desses e de outros substratos, a ADAM17 é descrita envolvida em processos inflamatórios e moduladora da adesão célula-célula e célula-matriz extracelular (Scheller, Chalaris *et al.*, 2011).

1.7 O estado da arte e os desafios na descoberta e validação de biomarcadores em câncer oral por espectrometria de massas

As recentes tecnologias capazes de identificar em larga-escala genes ou proteínas têm sido utilizadas como estratégias de primeira escolha nessa corrida em busca de novos biomarcadores, sobretudo em câncer. A proteômica corresponde a quase 10% dos artigos entre 1996 a 2009 que utilizaram dessa estratégia para busca de biomarcadores (Martins-De-Souza, 2010), sendo inúmeros os modelos de estudo desde cultura de células e modelos animais até amostras clínicas como tecido, plasma, urina e outros fluidos biológicos.

Idealmente, um biomarcador tumoral é uma molécula produzida em pequena quantidade em indivíduos saudáveis, mas em alta quantidade em pacientes com o tumor. Além disso, deve ser uma molécula facilmente detectada em fluidos biológicos de forma específica e sensível e de preferência em estágios iniciais da doença (Kulasingam e Diamandis, 2008).

Atualmente existem 12 biomarcadores tumorais usados na prática clínica e importantes exemplos são: o *carcinoembryonic antigen* (CEA) para diagnóstico e monitoramento de câncer de cólon, *prostate specific antigen* (PSA) para diagnóstico de câncer de próstata, *alpha fetoprotein* (AFP) para diagnóstico e monitoramento de câncer de fígado, CA 125 para prognóstico em câncer de ovário, CA 15-3 e CA 19-9 para monitoramento terapêutico em câncer de mama e pâncreas, respectivamente (Kulasingam e Diamandis, 2008). Interessante notar que todos esses marcadores de uso clínico foram descobertos antes do surgimento das modernas tecnologias de larga escala e mesmo diante de tantos candidatos biomarcadores apresentados nos últimos anos, nenhum foi introduzido clinicamente. Especificamente, em câncer oral ainda não existe nenhum biomarcador na prática clínica e, como mencionado anteriormente, a detecção é realizada através de exame clínico combinada com análise histopatológica de tecidos removidos por biópsia (Dionne, Warnakulasuriya *et al.*, 2014).

Em 2008, um painel de biomarcadores potenciais para detecção de CEC oral foi levantado a partir de um estudo proteômico em saliva (Hu, Arellano *et al.*, 2008). Cinco candidatos encontrados nesse trabalho (MRP14, *profiling*, CD59, catalase e M2BP foram validados por ELISA e obtiveram alta sensibilidade (90%) e especificidade (83%) para detecção de CEC. De forma semelhante, De Jong, Xie *et al*. (2010) (De Jong, Xie *et al.*, 2010) utilizaram a estratégia de análise proteômica com marcação por iTRAq (*isobaric tags for relative and absolute quantitation*) em saliva de pacientes com lesões orais pré-malignas e malignas e levantou actina e miosina como potenciais marcadores para detecção de câncer oral.

Recentemente, Hsu, Hu *et al.* (2014) (Hsu, Yu *et al.*, 2014) realizaram um estudo proteômico utilizando o secretoma de linhagem primária derivada de CEC oral e epitélio não canceroso adjacente, no qual foram levantados 19 candidatos a biomarcadores. Dentre eles, três marcadores THBS2, UFD1L e DNAJB11 foram validados em maior expressão por imunohistoquímica em tecido de paciente com CEC oral comparado com tecido adjacente não canceroso e THBS2 foi também validado em maior expressão em saliva de pacientes com CEC oral comparado com controles saudáveis.

Esses estudos demonstram a grande contribuição da proteômica baseada em espectrometria de massas na descoberta de candidatos à biomarcadores em câncer oral, entretanto nenhum candidato está sendo usado na prática clínica. Tal insucesso deve-se principalmente à falta de metodologias capazes de validar quantitativamente, em amostras clínicas, todos os candidatos (Makawita e Diamandis, 2010) e ao elevado custo associado ao longo processo no desenvolvimento, validação e aprovação de biomarcadores para uso clínico. Além disso, estratégias baseada em descoberta enfrentam desafios como complexidade e faixa dinâmica de proteínas presentes na amostra, baixa abundância da maioria dos biomarcadores mais específicos e a grande variabilidade biológica humana e da própria doença (Rifai, Gillette *et al.*, 2006). Somam-se a isso, os desafios da análise de espectrometria de massas, sobretudo em termos de ferramentas estatísticas capazes de evidenciar e priorizar proteínas diferencialmente expressas com menor índice de falsos positivos.

Atualmente, o método mais utilizado para quantificar proteínas é o ELISA (*enzyme-linked immunosorbent assay*). Porém, a disponibilidade restrita de anticorpos bem como tempo e custo elevado para análise resultam na impossibilidade avaliar a maioria dos candidatos biomarcadores levantados pelas tecnologias baseada em descoberta (Sturgeon, Hill *et al.*, 2010).

O monitoramento seletivo de reações (SRM) e o monitoramento de reações paralelas (PRM)[6] em espectrometria de massas tem surgido como uma promissora estratégia para superar esses desafios, pois é capaz de quantificar com alta reprodutibilidade, sensibilidade, acurácia e velocidade várias proteínas ao mesmo tempo (Schiess, Wollscheid *et al.*, 2009; Picotti e Aebersold, 2012). No entanto, para o desenvolvimento de um ensaio por SRM ou PRM confiável e bem sucedido são necessários alguns requisitos como: instrumento adequado, otimização do método (*dwell time*, *cycle time*, energia de colisão, tempo de gradiente na cromatografia líquida), determinação de peptídeos proteotípicos e análise de dados (Stergachis, Maclean *et al.*, 2011;

[6] As técnicas de SRM e PRM serão detalhadas na segunda parte dessa revisão bibliográfica

Bereman, Maclean *et al.*, 2012). Tais fatores ainda limitam a aplicação dessa técnica de maneira ampla, embora diversos esforços estão sendo realizados para desenvolver e padronizar métodos adequados bem como ferramentas estatísticas e computacionais para análise desses tipos de dados gerados (Keshishian, Addona *et al.*, 2009; Stergachis, Maclean *et al.*, 2011; Bereman, Maclean *et al.*, 2012; Chang, Picotti *et al.*, 2012; Carr, Abbatiello *et al.*, 2014; Welinder, Jonsson *et al.*, 2014).

1.8 O estado da arte e os desafios no estudo do degradoma de metaloproteases

Proteases são responsáveis por hidrolizar irreversivelmente ligações peptídicas de proteínas, resultando em uma modificação pós-traducional importante que regula diversos processos fisiológicos e patológicos na célula. A maioria das informações disponíveis sobre enzimas derivam de técnicas *in vitro*, as quais são capazes determinar a estrutura molecular, cinética enzimática ou especificidade do sítio ativo (Doucet e Overall, 2008). Todavia, essas técnicas são limitadas em revelar funções de proteases dentro de um contexto biológico, uma vez que não levam em consideração a complexa rede de regulação e interação nas quais estão envolvidas. Essa rede é composta não somente de protease, como também inibidores, ativadores, moduladores, substratos e produtos de clivagem, além outros interactores que podem regular sua função (Doucet, Butler *et al.*, 2008).

Com o desenvolvimento da proteômica baseada em espectrometria de massas, tornou-se possível o estudo global de proteases em um contexto mais amplo, considerando a protease como parte de um sistema. Chamou-se de degradômica essa nova área dentro da proteômica que visa decifrar a função de proteases em um sistema biológico pela determinação do repertório de substratos, sítios de clivagens, interactores e reguladores de atividade (Doucet, Butler *et al.*, 2008).

Nos últimos cinco anos, muitas estratégias que utilizam espectrometria de massas foram desenvolvidas para o estudo do degradoma de proteases (Schilling e Overall, 2007; Impens, Colaert *et al.*, 2010). A maioria desses métodos utiliza o princípio de marcação no novo N-terminal (N-terminômica) ou C-terminal (C-terminômica) gerado pela clivagem proteolítica e determinação por espectrometria de massas da sequência do peptídeo contendo essa marcação. De acordo com a nomenclatura proposta por Berger e Schechter (Berger e Schechter, 1970), os resíduos C-terminais gerados após a clivagem são numerados como Px', iniciando com P1' no resíduo onde ocorreu a clivagem e os resíduos N-terminais são numerados como Px, sendo o resíduo onde ocorreu a clivagem numerado como P1 (Exemplo: N-terminal–P3–P2–P1 | P1'–P2'–P3' – C-terminal). Essas

técnicas permitem não somente determinar os substratos de uma determinada protease como também os sítios de clivagem, os quais são importantes para elucidar a especificidade enzima-substrato. (Van Den Berg e Tholey, 2012).

O grupo do pesquisador Dr. Christopher M. Overall é um dos mais reconhecidos na área de estudo do degradoma de metaloproteases, publicando importantes trabalhos de desenvolvimento de metodologias para determinar em larga escala alvos de MMPs e sítios de clivagens. A proteômica quantitativa[7] utilizando marcação por *isotope-coded affinity tag* (ICAT) revelou novos substratos para MT1-MMP em células de carcinoma de mama (MDA-MB-231) (Tam, Morrison *et al.*, 2004). Posteriormente, a marcação por iTRAQ em células de fibroblastos selvagem e *knockout* para $MMP2^{-/-}$ foi realizada e novos substratos também identificados para essa protease (Dean e Overall, 2007). Em 2010, Overall e colaboradores desenvolveram uma metodologia para enriquecimento do N-terminoma gerado pela clivagem por MMPs (Kleifeld, Doucet *et al.*, 2010), chamada de marcação isotópica do amino terminal de substratos (*terminal amine isotopic labeling of substrates*, TAILS). Essa técnica consiste em comparar dois proteomas na presença e ausência de uma determinada protease pela marcação do novo N-terminal gerado pela clivagem, seguido de enriquecimento dos peptídeos marcados previamente à análise por espectrometria de massas. Resumidamente, cultiva-se uma determinada linhagem celular na presença ou ausência de uma protease. Após evento de clivagem, ambos proteomas são desnaturados, reduzidos e submetidos a marcação do N-terminal por reagentes (iTRAQ ou dimetilação) com isótopos leves (controle) ou pesados (protease). Esses proteomas são combinados e digeridos por tripsina e os peptídeos que não contêm o N-terminal marcado são removidos da mistura por meio de um polímero que reage com N-terminais livres. Os peptídeos marcados no N-terminal, agora enriquecidos, são analisados por espectrometria de massas e comparados na presença e ausência da protease. Com a utilização dessa metodologia, Prudova *et al*., (Prudova, Auf Dem Keller *et al.*, 2010) foram capazes de determinar a especificidade no sítio de clivagem de MMP2 e MMP9 e Starr *et al*. (Starr, Bellac *et al.*, 2012) identificaram novos alvos para MT6-MMP utilizando secretoma de fibroblastos após incubação com MT6-MMP recombinante.

A busca por novos alvos da ADAM17 também tem sido de grande interesse, uma vez que, como mostrado anteriormente, essa protease está envolvida em diversos processos tumorigênicos e inflamatórios (Arribas e Esselens, 2009). Tucher *et al*. (2014) determinaram *in vitro* os sítios de clivagens para ADAM17 e ADAM10 utilizando a metodologia PICS (*Proteomic*

[7] Mais detalhes dessas técnicas de marcação em proteômica quantitativa se encontram na segunda parte dessa revisão bibliográfica

Identification of Protease Cleavage Sites) também desenvolvida por Overall (Schilling e Overall, 2008). Essa técnica consiste em obter uma biblioteca de peptídeos derivada de um proteoma digerido com tripsina ou Lys-C ou glu-C, bloquear todas as aminas por metilação e, em seguida, incubar com a protease de interesse. A clivagem do peptídeo pela protease gera um novo N-terminal, que é posteriormente marcado com biotina e enriquecido por afinidade com estreptoavidina. Esses peptídeos marcados com biotina são analisados por espectrometria de massas e avaliados quanto ao sítio de clivagem. Utilizando uma biblioteca de peptídeos provenientes do proteoma de levedura, Tucher e colaboradores mostraram que a ADAM17 possui preferência para resíduos hidrofóbicos no sítio de clivagem do substrato, sendo valina e alanina os mais frequentes em P1' e P1, respectivamente (Tucher, Linke *et al.*, 2014).

Embora muitos avanços tenham sido alcançados no estudo de degradoma de proteases utilizando espectrometria de massas, sobretudo na identificação de novos alvos e sítios de clivagens em larga escala, o grande desafio ainda é a elucidação do papel funcional da clivagem desses substratos em diferentes contextos celulares.

PARTE II: FUNDAMENTOS DE ESPECTROMETRIA DE MASSAS PARA ANÁLISE DE PROTEÍNAS

1.9 Espectrometria de Massas para análise de proteínas: instrumentação

Fundamentalmente, a espectrometria de massas mede a relação massa-carga (*m/z*) de íons em fase gasosa (Han, Aslanian *et al.*, 2008). Por definição, um espectrômetro de massas consiste em uma fonte de íons, um analisador de massas, um detector e um sistema de aquisição de dados. O estudo de analitos na forma de íons em fase gasosa possui duas características fundamentais: 1. o movimento desses íons podem ser precisamente controlados por um campo magnético e/ou elétrico (proporcional a *m/z*) e 2. Íons em fase gasosa são transmitidos com alta eficiência para sistemas detectores e multiplicadores de elétrons, possibilitando uma análise com alta sensibilidade (Kinter, 2000).

A ionização por *eletrospray* (ESI) e a ionização a laser auxiliada por matriz (MALDI) são as duas técnicas mais utilizadas para volatilizar e ionizar proteínas ou peptídeos na análise por espectrometria de massas (Gstaiger e Aebersold, 2009). Em ESI, os analitos são dissolvidos em uma solução ácida – sendo por isso facilmente acoplado a cromatografia líquida - que ao passar por um fino capilar sob um alto potencial elétrico, é disperso em pequenas gotículas carregadas, as quais evaporam e transferem o analito em sua forma ionizada para o espectrômetro de massas (Aebersold e Mann, 2003). Uma característica importante da ionização por *eletrospray* é que os peptídeos gerados são na maioria das vezes multiprotonados, resultado da associação de prótons com a porção amino terminal ou resíduos básicos dos peptídeos como lisina, histidina e arginina (Kinter, 2000). Na ionização por MALDI, as proteínas ou peptídeos são co-cristalizados com uma matriz, normalmente um ácido orgânico, e em seguida submetidos à incidência de um laser que promove a sublimação e transferência da matriz, e consequentemente da espécie não volátil, para a fase gasosa (Figeys, 2005).

Uma vez ionizados e transferidos para dentro do espectrômetro de massas, os íons são separados de acordo com suas razões massa/carga (*m/z*) no analisador de massas (Glish e Vachet, 2003). Nessa etapa, parâmetros importantes como sensibilidade, resolução e exatidão de massa determinam a qualidade do dado (espectro) gerado, bem como caracterizam diferentes tipos de analisadores de massas (Domon e Aebersold, 2006).

Entre esses parâmetros, a resolução em um espectro de massas para um pico correspondente a íons de carga unitária e de massa m, pode ser expressa como m/Δm, onde Δm é definida como a largura total do pico na metade de sua altura máxima (*full width at half maximum, FWHM*). Na prática, quanto maior a resolução, maior a separação entre picos de *m/z* próximas e maior a exatidão na determinação da *m/z*. A exatidão de massa, por sua vez, pode ser expressa como diferença (erro) entre a massa exata (teórica) e a massa medida (observada) em relação à massa teórica, expressa em ppm (partes por milhão) (Kinter, 2000).

Existem quatro tipos básicos de analisadores de massas atualmente utilizados em pesquisa proteômica: 1. Quadrupolo; 2. Armadilha de íons (*ion trap*); 3. Tempo de vôo (*Time-of-flight*, TOF) e 4. Orbitrap.

O analisador do tipo quadrupolo consiste em quatro barras paralelas, arranjadas em dois pares opostos. Um dos pares apresenta um potencial elétrico aplicado de + (U+ Vcos(wt)) e o outro par apresenta um potencial elétrico aplicado de - (U+ Vcos(wt)), onde U é uma voltagem de corrente contínua (AC) e o termo Vcos(wt) é uma voltagem de corrente alternada (ou de radiofrequência, RF). A aplicação de uma voltagem de corrente continua (AC) e alternada (RF) gera um campo eletrostático oscilante, afetando a trajetória centralizada dos íons. O quadrupolo age como um filtro de massas; em se fazendo variar essas voltagens, somente íons com uma determinada *m/z* irão atravessar o centro do quadrupolo, enquanto que os outros íons serão desviados da trajetória central. As vantagens desse tipo de analisador de massas são: alta velocidade de varredura e fácil acoplamento para análise em tandem. No entanto, a característica de filtrar íons faz do quadrupolo um analisador de massas discriminatório, resultando em baixa sensibilidade, quando utilizado no modo varredura, pelo uso ineficiente dos íons (Kinter, 2000). Além disso, a resolução proporcionada pelo quadrupolo é baixa, também chamada de resolução unitária (Glish e Vachet, 2003), pois é capaz de resolver íons com diferença de 1 *m/z*.

De forma similar ao quadrupolo, o analisador de massas do tipo armadilha de íons 3D (ion trap 3D) consiste em dois eletrodos hiperbólicos terminais (“tampa”) e um eletrodo em forma de anel, conectados a uma voltagem RF e AC. Os íons são aprisionados no eletrodo anel, sob determinada voltagem RF e oscilam em trajetórias estáveis e concêntricas. Em seguida, um potencial RF é aplicado, provocando uma desestabilização da trajetória (proporcional à *m/z*) e expulsão do íon para fora do *trap*, onde é detectado. A forma compacta do *ion trap* e o aprisionamento de íons (evitando a sua perda, isto é, análise não discriminatória) são fatores que contribuem para o aumento da sensibilidade desse tipo de analisador quando comparado com o quadrupolo. Porém, ainda assim a resolução é baixa, além de não possibilitar análise de valores de

m/z inferiores a 1/3 do precursor (*1/3 rule*), na condição de fragmentação do peptídeo (*low-mass cut-off*). Além disso, o ion trap 3D possui limitado número de íons que podem ser aprisionados sem que haja interações entre eles (repulsão eletrostática), resultando em desvio de órbita dentro do trap ("*mass shift*") (Dass, 2007). O ion trap 3D foi modificado para um arranjo quadrupolar - denominado ion trap linear (*linear ion trap*, LIT) -, no qual o íon é confinado radialmente por um campo de radiofrequência (RF) bidimensional (2DE) e axialmente pela interrupção dos potenciais aplicados na extremidade dos eletrodos. Essa configuração permitiu aumentar a eficiência de injeção e a capacidade de armazenamento do íon, reduzindo o *chemical shift* e o *low-mass cut-off*. A utilização do LIT não se limitou apenas ao armazenamento de íons, mas tem sido extensivamente aplicado em instrumentos híbridos, por ser capaz de isolar íons e realizar experimentos em tandem (Domon e Aebersold, 2006).

O analisador de massa do tipo tempo de vôo (*time-of-flight*) baseia-se no princípio de que as velocidades de dois íons, criados no mesmo instante, com a mesma energia cinética, variarão conforme a *m/z* dos íons – os mais leves viajarão e atingirão o detector mais rápido que íons com *m/z* mais altas. Esse tipo de analisador opera de modo pulsado, de forma que os íons criados são acelerados e em seguida atravessam uma região (sem campo elétrico) para medida do tempo de vôo. A ionização do tipo MALDI foi a que melhor se ajustou com esse tipo de analisador, uma vez que, ao contrário da ESI, é produzida de forma pulsada. Outra vantagem é que o analisador TOF permite detecção de íons com grande faixa de *m/z* (de *m/z* 50 a 5000), além de possibilitar varreduras rápidas e não discriminatórias. A principal desvantagem é a baixa resolução devida principalmente à dispersão de energia cinética dos íons na fonte, verificada nos analisadores TOF mais antigos. Essa limitação foi contornada com o uso de espelhos eletrostáticos ou aceleração ortogonal, podendo chegar a resoluções mais altas (Aebersold e Mann, 2003).

O analisador do tipo orbitrap possui alta performance analítica em termos de resolução, sensibilidade, exatidão de massas, faixa dinâmica, tamanho e custo reduzidos (Hu, Noll *et al.*, 2005). Embora o orbitrap seja uma versão modificada armadilha de Kingdon[8], descoberta em 1923, seu uso como analisador de massas somente foi desenvolvido em 1999 pelo Dr. Alexander A. Makarov (Makarov, 1999). A estrutura de um orbitrap consiste em um eletrodo externo no formato de barril e um eletrodo central coaxial fusiforme. Ambos os eletrodos são conectados de forma independente com fontes elétricas, as quais produzem um campo com potencial de distribuição do

[8] Um dispositivo de captura de íons que consiste de um eletrodo externo do tipo barril e um eletrodo interno coaxial fusiforme, os quais formam um campo eletrostático com potencial de distribuição quadro-logarítmica.

tipo quadro-logarítmica[9]. Uma vez injetados de forma deslocada ao equador (z=0), os íons presos em órbita começam a oscilar no sentido axial e rotacional em um movimento do tipo harmônico ao longo do eletrodo axial. A frequência de oscilação do íon no sentido axial é independente da energia e posição do íon, porém inversamente proporcional à raiz quadrada da *m/z*. O movimento dos íons entre os eletrodos induzem uma corrente elétrica que pode ser medida e posteriormente transformada em espectro de massas através da equação de transformada de Fourier (Hu, Noll *et al.*, 2005).

Por fim, na composição do espectrômetro de massas, temos o detector, cuja função é a de detectar, amplificar o sinal da corrente de íons que vem do analisador e transferir o sinal para o sistema de processamento de dados. Os dois tipos principais de detectores são os multiplicadores de elétrons e os fotomultiplicadores. No detector do tipo multiplicador, várias placas coletoras de íons em série são utilizadas e mantidas em potencial crescente. Quando o íon choca à superfície do multiplicador, dois elétrons são ejetados e em seguida atraídos pela segunda placa gerando, cada um, novos dois elétrons. Esse processo continua até chegar à extremidade do multiplicador de elétrons, resultando em uma corrente elétrica que é analisada e registrada pelo sistema de dados. O ganho típico da amplificação da corrente gerada pela cascata de elétrons é da ordem de 1.000.000 de vezes. A placa de conversão fotomultiplicadora é similar a uma multiplicadora de elétrons, na qual inicialmente os íons atingem uma placa, resultando na emissão de elétrons. Todavia, os elétrons gerados atingem uma tela de fósforo, resultando na liberação fótons. Estes, por sua vez, são detectados por uma fotomultiplicadora, que opera em cascada como uma multiplicadora de elétrons (Dass, 2007).

A Figura 4 resume, de forma ilustrativa, os componentes de um espectrômetro de massas. Embora existam outros tipos de fontes ionizadores, analisadores de massas e detectores, os que foram aqui descritos são os mais utilizados para aplicação na análise proteômica.

[9] Equação que descreve o movimento dos íons no eixo z (axial), r (radial) e φ (rotacional) sobre a presença de um campo elétrico.

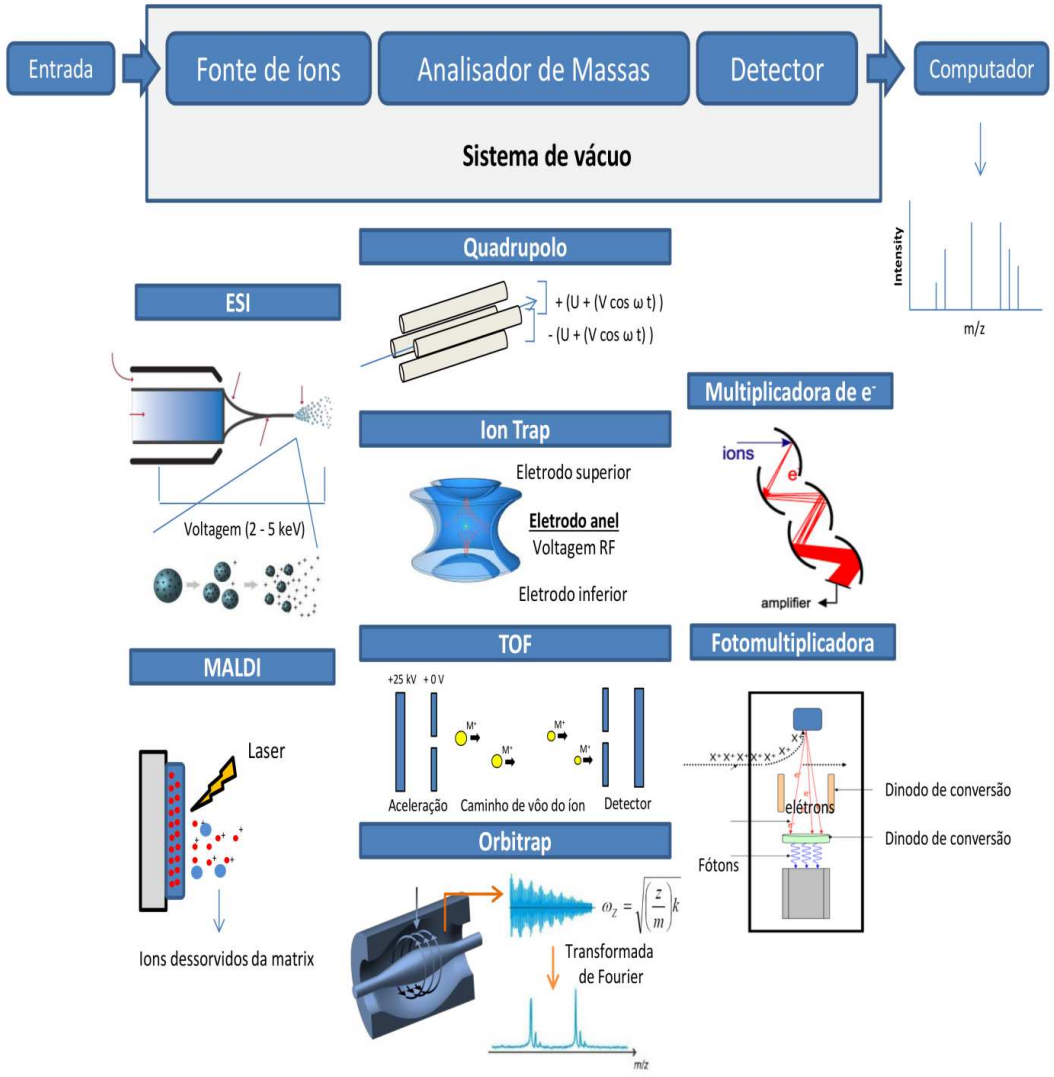

Figura 4 Componentes de um espectrômetro de massas. Diferentes tipos de fonte de ionização, analisadores de massas e detectores mais utilizados para análise proteômica são mostrados. Fonte: O AUTOR.

1.10 Fragmentação de peptídeos

Inicialmente a identificação de peptídeos era realizada unicamente pela informação relativa à *m/z* dos peptídeos oriundos da digestão enzimática (*Peptide Mass Fingerprint* – PMF), no qual, nesse caso, apenas um estágio de análise de massas era utilizado. Com o surgimento da espectrometria sequencial (tandem), a identificação do peptídeo passou a ser realizada utilizando tanto a informação do primeiro analisador de massas (MS), o qual isola uma determinada *m/z*

(chamado de precursor), como a informação dos íons fragmentos, também chamados de íons produtos (MS/MS), obtidos no segundo analisador de massas após evento de fragmentação (Yates, 1998).

A fragmentação envolve um aumento da energia interna dos íons levando-os à dissociação. A forma mais comumente utilizada para fragmentar peptídeos é a dissociação induzida por colisão (*collision induced dissociation* – CID). Nesse processo, os peptídeos ionizados e isolados no primeiro analisador de massas são acelerados (com uma energia cinética de 10 eV a 50 eV) para múltiplas colisões com um gás inerte (hélio, argônio ou nitrogênio). A energia cinética dessas colisões é convertida em energia vibracional e liberada na forma de reações de fragmentação direcionadas pelos sítios protonados (Kinter, 2000; Glish e Vachet, 2003).

A teoria do próton móvel é descrita como a energia interna adquirida que induz a transferência intramolecular dos prótons dentro de cada peptídeo, culminando na desestabilização das ligações do esqueleto polipeptídico e formação de padrões de MS/MS com intensidades heterogêneas. A mobilidade do próton depende da carga do peptídeo, do N-terminal e de resíduos de aminoácidos básicos na sequência (lisina, arginina e histidina), os quais produzem diferentes estruturas protonadas e, portanto diferentes sítios direcionados à fragmentação (Kinter, 2000). A reação de fragmentação se inicia através da formação de um intermediário cíclico no sítio protonado, seguido de reações que resultam na formação de dois íons-fragmentos, que são classificados como íons que retêm a carga residual (próton) no lado N-terminal (gerando fragmentos -a, -b e -c, dependendo da ligação que é fragmentada); íons que retém a carga residual (próton) na região C-terminal (gerando os fragmentos -x, -y -z, dependendo da ligação que é fragmentada), segundo a nomenclatura proposta por Roepstorff–Fohlmann–Biemann (Roepstorff e Fohlman, 1984) (Figura 5A). Em peptídeos gerados por digestão tríptica, a fragmentação por CID favorece principalmente a formação de íons da série b e y. Embora essa técnica seja universalmente utilizada para sequenciar peptídeos, ela gera informações limitadas para peptídeos grandes (>15 aa) e proteínas intactas, além de não preservar modificações pós-traducionais.

Outra técnica de fragmentação, introduzida por McLafferty em 1998, é a dissociação por captura de elétrons (*electron capture dissociation*, ECD) (Zubarev, 1998). Nessa técnica, moléculas com múltiplas protonações interagem com elétrons de baixa energia, resultando na quebra da ligação peptídica N-Cα e formação de fragmentos principalmente do tipo c e z. A fragmentação por ECD preserva modificações pós-traducionais além de permitir um espectro de MS/MS mais informativo e homogêneo, com maior cobertura de sequência, sendo, portanto mais

utilizado em análises de proteínas intactas sem uso de digestão enzimática[10]. De forma similar ao ECD, a dissociação por transferência de elétrons (*electron-transfer dissociation*, ETD), desenvolvida em 2004 por Syka, Coon *et al*., utiliza ânions com afinidade eletrônica relativamente baixa. Esses elétrons são transferidos para cátions multiprotonados de peptídeos, resultando na fragmentação do esqueleto peptídico e formação de íons da série c e z (Syka, Coon *et al.*, 2004; Han, Aslanian *et al.*, 2008).

Ainda mais recente, a fragmentação por dissociação induzida por alta energia de colisão (h*igh energy collisional dissociation*, HCD) foi pela primeira vez implementada em um instrumento comercial da empresa Thermo (LTQ Orbitrap XL) em 2007 (Olsen, Macek *et al.*, 2007). Essa tecnologia é bastante similar a CID, porém utiliza uma maior energia de colisão, com menor tempo de ativação (~0.1 ms para HCD, comparado com ~10 a 30 ms para CID). A fragmentação por HCD também gera íons predominantemente do tipo b e y, porém pode gerar também íons da série a e outros íons menores. Além disso, a fragmentação por HCD utiliza uma câmara do tipo octapolo (e não do tipo *ion trap*) e a detecção dos fragmentos é realizada em alta resolução no orbitrap, possibilitando um espectro mais informativo principalmente na região de baixa *m/z* (exemplo: íons a2/b2, y1/y2, íons imônios e íons repórteres de marcação isóbara, por exemplo, por iTRAq e TMT) (Jedrychowski, Huttlin *et al.*, 2011; Shao, Zhang *et al.*, 2014).

1.11 Experimentos em espectrometria sequencial

Uma vez entendido os principais processos de fragmentação, podemos discutir os diferentes tipos de experimentos que configuram a espectrometria de massas sequencial. Quando a análise envolve dois analisadores separados fisicamente, isto é, íons de determinada *m/z* são selecionados em uma seção do equipamento, dissociados em uma região intermediária (célula de colisão) e os íons produtos gerados são transmitidos para outro analisador, a espectrometria sequencial é chamada de espectrometria de massas sequencial no espaço (*tandem mass spectrometry in space*). No entanto, utilizando analisador do tipo *ion trap*, esses eventos de isolamento do precursor, fragmentação e detecção dos íons produtos podem ser realizados no mesmo espaço físico, mas em intervalos de tempo sequenciais. Nesse caso, conceitua-se como

[10] Chamamos a abordagem proteômica de *top-down* quando proteínas são analizadas sem digestão enzimática.

espectrometria de massas sequencial no tempo (*tandem mass spectrometry in time*), sendo possível realizar múltiplos estágios de MS (MS^n) com um único analisador (Kinter, 2000).

Dependendo da informação que se deseja obter a respeito do analito, diferentes experimentos podem ser realizados em espectrometria de massas sequencial: 1. Varredura de íons produtos; 2. Varredura de íons precursores; 3. Varredura de perda neutra e 4. Monitoramento seletivo de reações (Figura 5 B-E) (Domon e Aebersold, 2006).

A primeira e mais comum dentre os tipos de experimento de espectrometria de massa sequencial, a varredura de íons produtos (p*roduct ion scanning*), tem como objetivo a determinação de espectros de fragmentos para identificação da sequência de aminoácidos do peptídeo precursor. Nesse experimento, o primeiro analisador é configurado para selecionar um íon precursor (MS1). Este, por sua vez, é selecionado para evento de fragmentação em uma câmara de colisão e os íons produtos resultantes analisados em um segundo analisador de massas (MS2) (Figura 5B) (Domon e Aebersold, 2006; Nogueira e Domont, 2014).

Na varredura de íons precursores (*precursor ion scanning*), fixa-se um determinado fragmento e faz-se uma varredura de íons precursores que foram capazes de gerar esse fragmento. Esse método é geralmente utilizado para determinar um conjunto de peptídeos de uma amostra que contém determinado grupo funcional (ex: glicosilação) (Figura 5C). Já na varredura por perda neutra (*neutral loss scanning*), tanto o precursor quanto o fragmento são escaneados, de modo que somente aqueles que produzem uma determinada perda neutra após evento de fragmentação serão detectados (ex: fosforilação) (Figura 5D). Por fim, podemos configurar os analisadores de massas para selecionar apenas determinado precursor e fragmento específicos, correspondentes a uma proteína ou analito de interesse. Esse tipo de experimento chama-se monitoramento seletivo de reações (*selected reaction monitoring*), e é uma importante estratégia utilizada para quantificação de peptídeos (Figura 5E) (Domon e Aebersold, 2006; Nogueira e Domont, 2014).

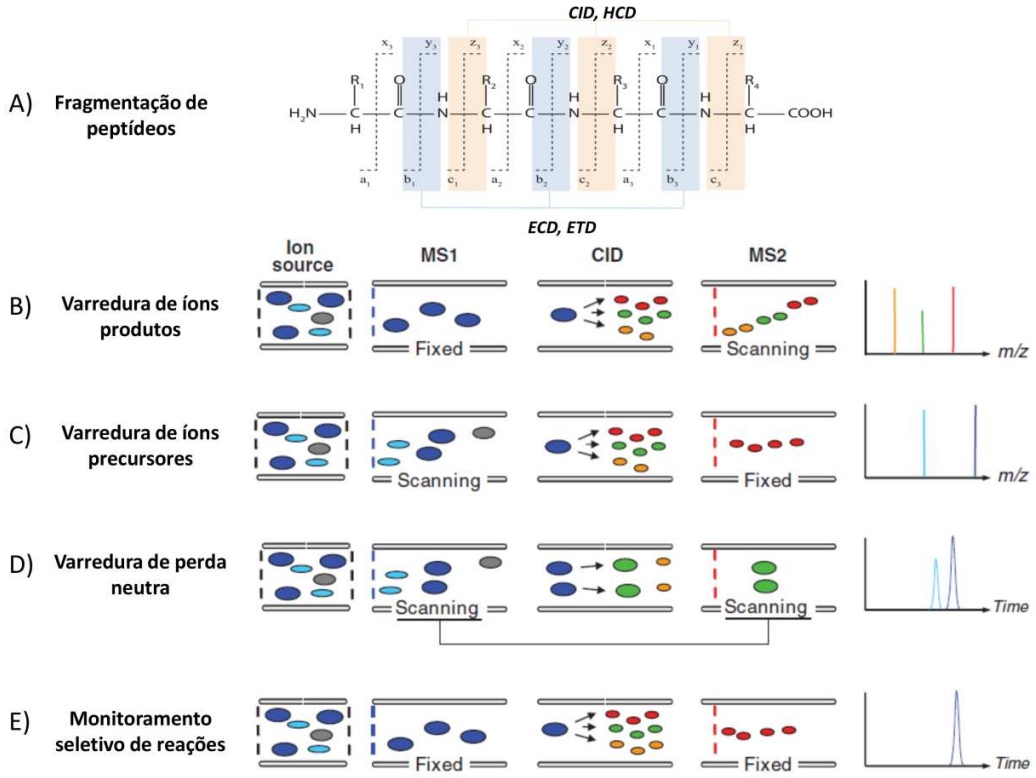

Figura 5 Representação esquemática da fragmentação de peptídeos e dos experimentos realizados em espectrometria de massas sequencial. A) Série de íons gerados pela fragmentação, segundo lado N-terminal ou C-terminal que o íon retém a carga residual (próton). Em rosa estão destacados os principais íons formados pela fragmentação do tipo CID e HCD. Em azul os principais íons formados por ECD e ETD. B) A varredura de íons produtos é o tipo de experimento mais comum em proteômica. O primeiro analisador é fixado para selecionar um precursor por vez (MS1). O íon selecionado é fragmentado e os íons fragmentos resultantes são detectados após passar pelo segundo analisador de massas (MS2). C) Na varredura de íons precursores, o segundo analisador de massa é fixado em uma determinada *m/z* de forma a transmitir somente o sinal do precursor (MS1) que gera o fragmento dessa *m/z* especificada (MS2). D) Na varredura de perda neutra, tanto o precursor quanto o fragmento são escaneados, de modo que somente aqueles que produzem uma determinada perda neutra após evento de fragmentação serão detectados. E) No monitoramento seletivo de reações, o espectrômetro é configurado para selecionar um precursor e um fragmento específico referente a um peptídeo previamente determinado. O par precursor e fragmento do peptídeo é chamado de transição. Adaptado de Domon e Aebersold, 2006.

1.12 Análise de dados proteômicos

Uma vez os espectros de MS e MS/MS gerados, softwares de busca podem ser utilizados para identificar as proteínas em bancos de dados. A etapa inicial do processamento de dados de MS é a detecção do pico, a qual envolve a diferenciação entre sinal e ruído e extração

precisa das massas de cada pico no espectro, utilizando para isso características como intensidade, padrão isotópico e tempo de retenção de precursores e fragmentos. Em seguida, essa lista de massas é submetida a algoritmos de busca, cujo objetivo final é a identificação da sequência de peptídeo que melhor correlaciona ao espectro observado (Sadygov, Cociorva *et al.*, 2004).

Os programas mais empregados para identificação de proteínas em bancos de dados a partir de dados de MS são o Sequest, o Mascot e o Andromeda.

O Sequest é um programa que utiliza um modelo descritivo para determinar matematicamente a sobreposição entre o espectro teórico e experimental. Dois scores são gerados para assegurar a qualidade da sobreposição, sendo o primeiro chamado de *preliminary score* (Sp) e o segundo *cross-correlation score* (Xcorr). O Sp score leva em consideração a soma das intensidades dos íons fragmentos que tiveram um *match* com sequências preditas do banco de dados bem como a largura do peptídeo e a continuidade da série de íons gerados pela fragmentação. Já o Xcorr se baseia em um espectro teórico gerado pelos íons fragmentos da série b e y de cada sequência. O espectro teórico e o observado são correlacionados a fim de obter similaridades entre os espectros. Em seguida, o maior Xcorr obtido para a sequência de aminoácidos é comparado com o Xcorr da sequência subsequente que obteve *match* com o espectro observado. Essa diferença, chamada de ΔCn, é um parâmetro muito importante para determinar se o *match* é único, assegurando dessa forma a confiabilidade da identificação. Na literatura, diferentes critérios são usados para classificar uma identificação como confiável ou não. De forma geral, estes valores são: Xcorr > 3,75 para peptídeos com carga +3; Xcorr > 2,2 para peptídeos com carga +2 e Xcorr > 1,9 para peptídeos com carga +1. Em todos os casos descritos, ΔCn>0,10 é exigido para que a determinação seja considerada suficientemente confiável (Eng, Mccormack *et al.*, 1994; Sadygov, Cociorva *et al.*, 2004).

O programa Mascot e MaxQuant pertencem a um grupo de métodos que usam modelos probabilísticos para identificação de proteínas. No Mascot, os valores de *m/z* dos fragmentos preditos são comparados com os fragmentos experimentais sendo que, neste caso, a comparação se inicia com base nos íons b e y mais intensos. A probabilidade de o valor de *m/z* de um fragmento teoricamente obtido coincidir, de maneira randômica, com o valor de *m/z* de um fragmento obtido experimentalmente é calculada, levando em consideração o tamanho do banco de dados, e expressa como um score = $- \log_{10}(P)$. Assim, quanto maior for o score obtido, menor é a probabilidade de que este resultado seja um resultado ao acaso (Perkins, Pappin *et al.*, 1999).

Utilizando o mesmo princípio probabilístico do Mascot, o programa MaxQuant usa um algoritmo, chamado Andromeda, para identificação de peptídeos em banco de dados de sequência.

A grande vantagem desse programa é que ele é um software aberto (www.maxquant.org) e disponível para ser utilizado localmente em qualquer computador. Além disso, a ferramenta Andromeda possibilita lidar com dados provenientes de espectros de alta exatidão de massa e resolução e possui um segundo algoritmo de busca para identificação de peptídeos co-fragmentados, resultando em um aumento de proteínas identificadas em misturas complexas (Cox, Neuhauser *et al.*, 2011).

1.13 Fundamentos da proteômica quantitativa

Inicialmente, a proteômica baseada em espectrometria de massas era essencialmente qualitativa. Um típico experimento proteômico resultava em uma lista de proteínas identificadas, sem nenhuma informação a respeito da abundância, distribuição ou estequiometria. No entanto, o grande desafio da atual proteômica é torná-la uma disciplina quantitativa, ou seja, capaz de revelar alterações em abundância de proteínas em diferentes condições ou ao longo do tempo.

Existem duas estratégias para comparar abundância de proteínas entre amostras, utilizando dados de espectrometria de massas: 1. Contagem espectral (*spectral counting*) e 2. Intensidade do peptídeo. À essas abordagens de comparação quantitativa de proteínas sem o uso de marcação isotópica, denomina-se *label-free* (Neilson, Ali *et al.*, 2011). A primeira se baseia no princípio de que quanto mais abundante a proteína está na amostra, maior será número de vezes que determinado(s) peptídeo(s) dessa proteína será selecionado e fragmentado, portanto maior será o número de espectros desse peptídeo (Mann e Kelleher, 2008). Embora essa abordagem seja largamente utilizada, a acurácia na quantificação de proteínas pouco abundantes, ou seja, com baixas contagens de espectros, é baixa. Se uma proteína é identificada em diferentes amostras com apenas uma contagem de espectro, por exemplo, ela será considerada igualmente abundante entre as amostras, porém as intensidades desse peptídeo podem ser diferentes (Schulze e Usadel, 2010). Além disso, peptídeos gerados pela clivagem proteolítica possuem diferentes propriedades físico-químicas (tamanho, carga, hidrofobicidade) e capacidade de ionização, as quais influenciam a detectabilidade e intensidade de sinal na análise por espectrometria de massas (Ong e Mann, 2005). Portanto, a comparação entre diferentes peptídeos de uma mesma proteína limita a acurácia na quantificação.

Alguns programas de análise de dados proteômicos reportam como dado quantitativo a intensidade do peptídeo, a qual é obtida pela área sobre a curva de um cromatograma de íons

extraídos (*extracted ion chromatogram*, XIC)[11]. Essa abordagem possui maior acurácia na quantificação em relação à contagem espectral quando se compara a intensidade de um mesmo peptídeo entre as amostras. Porém, a presença de interferentes na amostra (detergente ou proteínas muito abundantes na matriz), a não reprodutibilidade da corrida cromatográfica e o método estocástico de análise dependente dos dados acarretam erros e variabilidade no processo, limitando a quantificação (Ong e Mann, 2005).

O desenvolvimento de técnicas de marcação isotópica de proteínas ou peptídeos trouxe grandes avanços para a proteômica quantitativa. Peptídeos provenientes de diferentes amostras são marcados com diferentes isótopos e submetidos a análise por LC-MS na mesma corrida cromatográfica. A marcação isotópica produz uma a diferença de massa (idealmente de pelo menos 4 Da para quantificação baseada em precursor) entre o peptídeo marcado e não marcado, os quais são posteriormente relacionados à amostra de origem (Ong, Foster *et al.*, 2003).

A marcação isotópica pode ser realizada em diferentes estágios do preparo da amostra e com diferentes reagentes: (i) externamente (com o uso de peptídeos sintéticos que são adicionados na amostra); (ii) quimicamente e (iii) metabolicamente (Figura 6).

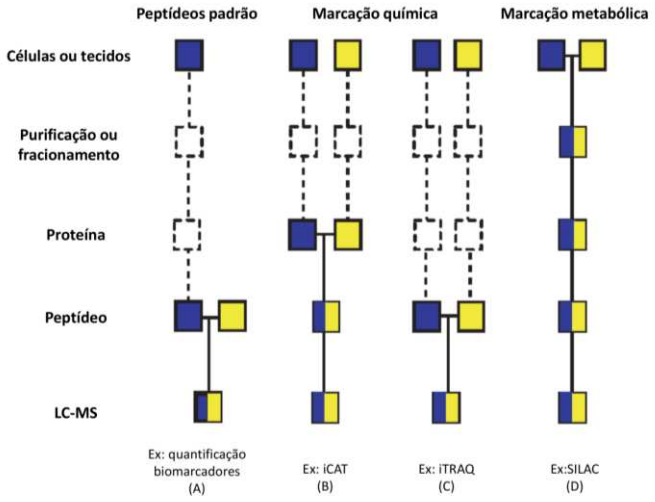

Figura 6 Representação esquemática dos estágios de incorporação da marcação isotópica. A figura mostra as etapas de um experimento em proteômica quantitativa, começando pela obtenção da amostra proveniente, por exemplo, de células ou tecidos, passando pela purificação, digestão de proteínas e análise por LC-MS/MS. As cores azul e amarela das caixas representam as condições e estágios onde as proteínas são marcadas com diferentes isótopos e comparadas pela espectrometria de massas. A linha horizontal refere-se ao estágio onde as amostras são combinadas. Abaixo de cada diagrama encontram-se exemplos de aplicações que utilizam cada uma das estratégias. Adaptado de Ong e Mann, 2005.

[11] O XIC é definido como intensidade de sinal ao longo do tempo de eluição de um peptídeo

A maneira mais simples, em termos de processamento, de introduzir peptídeos marcados isotopicamente é a síntese química desses peptídeos e a adição de quantidades conhecidas na amostra na forma de um padrão interno (Figura 6A). Essa abordagem tem sido bastante aplicada para quantificação "absoluta" de proteínas e geralmente requer análise por monitoramento seletivo de reações (SRM). Apesar de consistir na forma mais reprodutível e sensível de quantificar proteínas, a síntese desses padrões precisa ser baseada em peptídeos pré-conhecidos ou pré-selecionados (Ong e Mann, 2005; Rifai, Gillette *et al.*, 2006; Bantscheff, Schirle *et al.*, 2007).

Outra estratégia de marcação é a modificação química através da reação entre peptídeos e reagentes marcados com diferentes isótopos. Exemplos desse tipo de marcação mais utilizados são o ICAT (*isotope-coded affinity tag*) e o iTRAQ (*isobaric tag for relative and absolute quantitation*). O ICAT consiste em um reagente que apresenta um grupo reativo direcionado a resíduos de cisteína, oito átomos de hidrogênio (leve) ou deutério (pesado) e um resíduo de biotina. As proteínas são desnaturadas, reduzidas e marcadas com o reagente ICAT (^{1}H ou ^{2}H) (Gygi, Rist *et al.*, 1999; Steen e Mann, 2004). Os proteomas são combinados, digeridos e os peptídeos marcados enriquecidos através da captura por afinidade biotina-avidina. Após análise por espectrometria de massas, os peptídeos marcados e não marcados são diferenciados em nível de MS (precursor), através da diferença de massa produzida pela marcação das cisteínas com diferentes isótopos do ICAT (Figura 6B).

O reagente iTRAQ possui um grupo reativo a aminas primárias dos peptídeos, um grupo de "equilíbrio" (equilibra a massa do íon repórter) e um grupo repórter (produz fragmentos com massas 113,114, 115, 116, 117, 118, 119 e 121 Da). Cada amostra, após a digestão, é marcada com um reagente contendo um íon repórter específico e, posteriormente, combinadas para serem analisadas em uma mesma corrida cromatográfica. A quantificação é realizada comparando-se as intensidades dos diferentes íons repórteres no espectro de fragmentação (Ross, Huang *et al.*, 2004; Yan e Chen, 2005) (Figura 6C).

Por fim, na marcação metabólica, a incorporação é realizada na etapa mais inicial do processo, isto é, no proteoma de células metabolicamente ativas, sendo, portanto, teoricamente, a estratégia de maior acurácia e reprodutibilidade. Desenvolvida pelo grupo do Dr. Mathias Mann, a análise da razão de isótopos estáveis em aminoácidos em cultura de células, do inglês "*stable isotope labeling by amino acids in cell culture, SILAC*" consiste no cultivo de populações de células em meios de cultura diferenciais: um contendo aminoácidos no seu isótopo de maior abundância (Ex: Arg-$^{12}C_6$ ou Lys-$^{12}C_6$ *light medium*) e outro contendo aminoácidos marcados estavelmente com

um isotópo mais pesado (ex: Arg-$^{13}C_6$ ou Lys-$^{13}C_6$, *heavy medium*) (Ong e Mann, 2006). Essas células são cultivadas até completa incorporação do aminoácido pesado, geralmente observada após cinco repiques. Em seguida, as proteínas são extraídas, quantificadas e quantidades iguais de cada condição (leve e pesada) são combinadas e submetidas ao mesmo processo de fracionamento, digestão e análise por espectrometria de massas. A diferença de *m/z* correspondente ao número de isótopos pesados incorporados no peptídeo em relação ao mesmo peptídeo não marcado permitirá a comparação quantitativa em intensidade da proteína proveniente de cada condição. Uma vez que os proteomas de determinadas condições são combinados no início do processo, os erros de quantificação introduzidos durante o preparo de amostra, fracionamento e análise por espectrometria de massas irão afetar ambos proteomas igualmente. Outra vantagem é que combinando dois proteomas, o tempo de preparo de amostra e análise por MS é reduzido consideravelmente. As desvantagens são que essa metodologia é aplicada principalmente para cultura de células imortalizadas, e o resultado em termos de proteínas identificadas e quantificadas é geralmente menor que proporcionadas por técnicas *label-free*. Essa última desvantagem ocorre principalmente devido ao aumento da complexidade do espectro de MS e à quantificação ser realizada baseada no par *heavy-light* do peptídeo, isto é se em uma das condições o peptídeo não é identificado, não haverá quantificação (Figura 6D) (Bantscheff, Schirle *et al.*, 2007).

1.14 Proteômica baseada em descoberta e proteômica baseada em alvos

A proteômica baseada em espectrometria de massas evoluiu de tal modo que atualmente é possível a identificação de mais de 4000 proteínas em uma única corrida, sem fracionamento, em um gradiente de 4 horas (Nagaraj, Kulak *et al.*, 2012). Da mesma forma, a proteômica quantitativa possibilitou o estudo da alteração de inúmeros sistemas biológicos, tanto em termos de expressão, quanto de modificações pós-traducionais e interações com outras proteínas.

A proteômica por *shotgun* descrita anteriormente é aplicada sobretudo em estudos nos quais buscamos conhecer as proteínas que estão com a expressão alterada sob determinado estímulo ou condição e, dessa forma, levantar hipóteses quanto aos mecanismos moleculares por trás dos processos biológicos ou respostas fenotípicas das células. Tendo em vista que não sabemos a priori as proteínas que iremos identificar, tem-se denominado essa estratégia de proteômica baseada em descoberta (*discovery-based proteomics*). Nesta abordagem, o foco está na identificação e

quantificação relativa de um maior número possível de proteínas ou modificações pós-traducionais em um sistema biológico (Marx, 2013).

Em um experimento de proteômica baseada em descoberta, o espectrômetro é operado principalmente em um modo de aquisição dependente dos dados (*data-dependent acquisition*, DDA), no qual se utiliza a informação de um espectro de varredura (*survey scan* ou *full MS scan*) para selecionar os íons para fragmentação em um segundo experimento de espectrometria de massas sequencial. Dentre os tipos de aquisição por DDA, o mais utilizado é o método de aquisição em ciclos continuados de *full MS scans* seguidos de isolamento dos íons mais abundantes para subsequentes fragmentações (MS/MS scans), gerando o que chamamos de "seleção aleatória ou estocástica dos íons". Após a obtenção das *m/z* dos fragmentos e da informação da massa do precursor, a proteína pode ser identificada, utilizando uma busca contra um banco de dados (Domon, 2010). Embora essa estratégia seja poderosa para identificar com acurácia milhares de proteínas em uma amostra complexa, o modo de aquisição dependente dos dados é tendencioso para detecção dos picos mais abundantes, isto é, peptídeo de baixa abundância podem não ser selecionado para fragmentação. Além disso, por trabalhar com um elevado tempo de ciclo, a estratégia por DDA limita-se na quantidade de pontos por pico obtida, comprometendo a reprodutibilidade quantitativa quando a área do XIC é extraída.Vale lembrar que a seleção aleatória dos ions em DDA também compromete a reprodutibilidade na identificação dos peptídeos, sobretudo quando a cromatografia anterior à entrada no espectrômetro não é consistente (Picotti, Bodenmiller *et al.*, 2013).

Nesse contexto, tem emergido outra abordagem, chamada de proteômica baseada em alvos (*targeted proteomics*), cujo objetivo é a identificação e quantificação de um conjunto ou painel de proteínas direcionadas à responder determinada hipótese biológica (Gillette e Carr, 2013; Picotti, Bodenmiller *et al.*, 2013). O monitoramento seletivo de reações é a estratégia de escolha para esse tipo de objetivo, pois opera em um modo independente dos dados e oferece alta reprodutibilidade, seletividade, sensibilidade e acurácia quantitativa. A maioria dos experimentos de SRM é realizada em espectrômetros do tipo triplo quadrupolo – essencialmente constituído de dois analisadores de massas quadrupolo, separados por uma câmara de colisão, os quais atuam como um filtro de massas capazes de estabilizar somente íons de *m/z* específicos. O íon precursor e fragmento são escolhidos a partir da informação da *m/z* do peptídeo, em geral único para a proteína de interesse, com seu respectivo fragmento (transição) e o instrumento é programado para selecionar, fragmentar e quantificar somente os peptídeos de interesse (Bereman, Maclean *et al.*, 2012; Picotti e Aebersold, 2012). Vale lembrar que a partir do monitoramento de várias transições (dando origem

ao termo *multiple reaction monitoring*, MRM), diferentes proteínas podem ser quantificadas ao mesmo tempo – capacidade de multiplexação (*multiplexing capacity*), possibilitando assim uma metodologia para quantificação de proteínas mais rápida, barata, com maior acurácia e eficiência quando comparada aos tradicionais métodos baseado em anticorpo (ELISA, Western blot, etc.) (Lange, Picotti *et al.*, 2008). *Targeted proteomics* permite também uma maior cobertura do proteoma, pois é capaz de detectar proteínas com até 100 cópias por células, resultado de uma maior sensibilidade proporcionada pelo monitoramento somente de transições específicas da proteína de interesse (Picotti, Bodenmiller *et al.*, 2009).

A introdução de equipamentos de alta resolução e exatidão de massas (*high resolution/accurate mass*, HR/AM) acoplados com analisadores de rápida velocidade de varredura (como *ion trap* ou quadrupolo) permitiu o surgimento de uma alternativa ao SRM para realização de proteômica baseada em alvos. No chamado monitoramento de reações paralelas (*parallel reaction monitoring*, PRM), precursores de *m/z* específicos referentes aos peptídeos de interesse previamente estabelecidos são isolados no primeiro analisador e após fragmentação todos os íons fragmentos são detectados em um espectro de MS/MS, ao invés de um único fragmento por transição como realizado no SRM convencional. A aquisição do espectro do precursor (MS) é realizada em alta resolução e exatidão de massa, resultando em análises mais seletivas pela capacidade de separação de íons com *m/z* muito próximas bem como maior confiabilidade na identificação do peptídeo de interesse (Gallien, Duriez *et al.*, 2013).

Uma das aplicações mais encontradas em *targeted proteomics* está a quantificação de candidatos biomarcadores para validação e verificação destes em uso clínico (Whiteaker, Lin *et al.*, 2011). Utilizando peptídeos marcados isotopicamente é possível quantificar proteínas com alta acurácia e reprodutibilidade intra e inter-laboratórios. Além disso, tem-se destacado o uso de *targeted proteomics* na quantificação de proteínas específicas de vias metabólicas, eventos de fosforilação e estudo de vias de sinalização, proteínas presentes em complexos (interactoma) entre outros (Picotti e Aebersold, 2012). Em resumo, a proteômica baseada em alvos tem emergido como essencial no desenvolvimento do método científico, uma vez que pode ser usada para testar hipóteses biológicas e trazer respostas com bases em medidas reprodutíveis e quantitativas de proteínas. É importante enfatizar que as estratégias baseadas em descobertas são de extrema importância para gerar hipóteses, porém é fundamental que essas sejam testadas e comprovadas por métodos quantitativos (Marx, 2013).

2. ORGANIZAÇÃO DA TESE

Esta tese abrange os resultados de cinco sub-projetos realizados durante o doutorado e mostrados na forma de manuscritos publicados ou submetidos. De forma geral, o objetivo dessa tese é mostrar aplicações da espectrometria de massas como ferramenta na descoberta e validação de candidatos à biomarcadores em câncer (capítulos 1, 2 e 3) e no estudo do degradoma da ADAM17 (capítulos 4 e 5). Um resumo do objetivo e conteúdo de cada capítulo é apresentado a seguir.

CAPITULO 1

O objetivo principal do trabalho intitulado *Integrative Analysis to Select Cancer Candidate Biomarkers to Targeted Validation* foi o desenvolvimento de um *pipeline* para seleção de um painel de candidatos à biomarcadores em câncer a partir de dados de proteômica baseada em descoberta. Nesse trabalho, métodos estatísticos de seleção de atributos (Beta-binomial, *Nearest Shrunken Centroid*, NSC e *Support Vector Machine-Recursive Features Elimination*, SVM-RFE) foram propostos e testados em um conjunto de dados provenientes do secretoma de duas linhagens de células não cancerosas (HaCaT, queratinócitos humanos e HEK293, células de rim embrionário) e quatro linhagens de câncer, divididas em duas classses: carcinoma (SCC-9, carcinoma oral de células escamosas proveniente de língua e A431 carcinoma de células escamosas proveniente de pele) e melanoma (A2058, melanoma derivado de sítio metastático e SK-MEL-28, melanoma derivado da pele). Utilizando o *pipeline*, um painel de 137 e 271 candidatos biomarcadores foram selecionados para as classe de carcinoma e melanoma, respectivamente. A expressão de alguns deles foram validados por western blot e imunohistoquímica e duas proteínas do sistema complemento, C3 e CFB, foram avaliadas por *targeted proteomics* em saliva de pacientes normais e com câncer oral.

CAPITULO 2

A partir dos resultados obtidos e apresentados no capítulo 1, selecionamos 14 proteínas candidatas a biomarcadores em carcinoma para serem validadas em 30 amostras de saliva pacientes, sendo 10 deles controles e 20 de pacientes com câncer oral utilizando a técnica monitoramento seletivo de reações (SRM) em espectrômetro de massa do tipo triplo-quadrupolo. Esse trabalho foi realizado na *University of Washington*, em colaboração com o pesquisador Dr. Michael MacCoss,

durante três meses de Bolsa Estágio de Pesquisa no Exterior (BEPE) com o apoio financeiro da FAPESP. O trabalho intitulado "*A targeted proteomic strategy for the verification of oral cancer candidate biomarkers in human saliva*" apresenta um protocolo detalhado de análise de proteínas por SRM, desde a seleção de peptídeos proteotípicos até a análise estatística de dados de SRM. Este estudo apresentou um painel de proteínas com maior expressão em saliva de pacientes com CEC oral, sendo cinco proteínas CFB, C3, C4B, SERPINA1 e LRG1 com expressão também associada a maior risco em câncer oral.

CAPITULO 3

Neste capítulo buscou-se avaliar a função de dois proteoglicanos de matriz extracelular, agrina e perlecan, em câncer oral, observados previamente com expressão alterada no secretoma de células da classe carcinoma no estudo apresentado no capítulo 1. Agrina e perlecan foram estudados quanto à expressão em tecidos de pacientes normais e com CEC e quanto à função em processos tumorigênicos. Por meio do silenciamento gênico por siRNA dessas duas proteínas, eventos de adesão, migração, proliferação e sensibilidade à cisplatina foram avaliados em três linhagens celulares: SCC-9 (originadas de carcinoma de células escamosas proveniente de língua), SCC-9 LN1 (originada de células SCC-9 metastáticas isoladas de linfonodos) e A431 (originadas de carcinoma epidermóide). Esse trabalho foi publicado na revista Plos One (v.9, n.12 doi: 10.1371/journal.pone.0115004) com o título "*Agrin and perlecan mediate tumorigenic processes in oral squamous cell carcinoma*".

CAPÍTULO 4

O objetivo do trabalho "*Deciphering the Role of the ADAM17-Dependent Secretome in Cell Signaling*" publicado na revista *Journal of Proteome Research* (v.13, n.4, doi: 10.1021/pr401224u), foi a caracterização funcional e proteica do secretoma de células provenientes de camundongos (mEFs, "*Mouse embrionic fibroblasts*") selvagem (*Wild-type*) e *knockout* para ADAM17. Inicialmente, ensaios funcionais mostraram que o secretoma na condição selvagem, ou seja, na presença de ADAM17 induziu proliferação e migração de células. Para avaliar os substratos que poderiam estar sendo modulados pela ADAM17, foi realizada a caracterização do secretoma na presença e ausência da ADAM17, utilizando duas estratégias de proteômica quantitativa *label-free* e SILAC. Os resultados obtidos revelaram importantes vias de sinalização

associadas ao secretoma enriquecido da ADAM17 bem como novos alvos de matriz extracelular que essa enzima é capaz de clivar diretamente e/ou regular a expressão.

CAPÍTULO 5

O trabalho "*Mass spectrometry-based proteomics revealed Glypican-1 as a novel ADAM17 substrate"* apresentado no capítulo 5 teve como objetivo a descoberta de novos alvos da ADAM17 em câncer oral. Para isso duas estratégias de proteômica baseada em espectrometria de massas foram realizadas: 1. Biotinilação de proteínas de superfície de células SCC-9 seguido de incubação com ADAM17 recombinante produzida em células de inseto e 2. Análise do secretoma de células SCC-9 controle e estavelmente silenciadas (shRNA) para ADAM17. A partir dessas abordagens, glypican-1 (GPC1) foi revelado como um novo alvo da ADAM17 e diversos ensaios foram realizados para validar que a clivagem é dependente de metaloprotease e mediada pela ADAM17. O estudo da função do GPC1 solúvel foi realizado pela determinação de parceiros de interação e o fenótipo em eventos de adesão, migração e proliferação avaliados na super-expressão de GPC1 (inteiro e fragmento originado do sítio de clivagem pela ADAM17) e no tratamento com ativadores e inibidores de metaloproteases.

CAPÍTULO 1

Integrative Analysis to Select Cancer Candidate Biomarkers to Targeted Validation

Rebeca Kawahara#, Gabriela V. Meirelles#, Henry Heberle, Romênia R. Domingues, Daniela C. Granato, Sami Yokoo, Rafael R. Canevarolo, Flavia V. Winck, Ana Carolina P. Ribeiro, Thaís Bianca Brandão, Paulo R. Filgueiras, Karen S. P. Cruz, José Alexandre Barbuto, Ronei J. Poppi, Rosane Minghim, Guilherme P. Telles, Felipe Paiva Fonseca, Jay W. Fox, Alan R. Santos-Silva, Ricardo D. Coletta, Nicholas E. Sherman, Adriana F. Paes Leme

# these authors contributed equally to this work.

ABSTRACT

Targeted proteomics has flourished as the method of choice for prospecting for and validating potential candidate biomarkers in many diseases. However, challenges still remain due to the lack of standardized routines that can prioritize a limited number of proteins to be further validated in human samples. To help researchers identify candidate biomarkers that best characterize their samples under study, a well-designed integrative analysis pipeline, comprising MS-based discovery, feature selection methods, clustering techniques, bioinformatic analyses and targeted approaches was performed using discovery-based proteomic data from the secretomes of three classes of human cell lines (carcinoma, melanoma and non-cancerous). Three feature selection algorithms, namely, Beta-binomial, Nearest Shrunken Centroids (NSC), and Support Vector Machine-Recursive Features Elimination (SVM-RFE), indicated a panel of 137 candidate biomarkers for carcinoma and 271 for melanoma, which were differentially abundant between the tumor classes. We further tested the strength of the pipeline in selecting candidate biomarkers by immunoblotting, human tissue microarrays, label-free targeted MS and functional experiments. In conclusion, the proposed integrative analysis was able to pre-qualify and prioritize candidate biomarkers from discovery-based proteomics to targeted MS.

INTRODUCTION

Discovery-based proteomics has been known as the most powerful tool for globally profiling proteomes and has been employed to mine biomarkers and therapeutic targets in many clinical conditions [1-4]. However, the contribution of novel molecules in clinical practice has been disappointing, and several reasons for failure have arisen in the long processes of biomarker and therapeutic target validation [5-7].

Recently, targeted proteomics has succeeded as the method of choice to overcome the drawbacks in validating and verifying potential biomarkers and therapeutic targets [6,8-10]. Nevertheless, discovery-based proteomics can provide a large contribution in generating hypothesis-driven targets based on shotgun proteomics data [2,11-14]. In addition to the bottleneck of discovery strategies such as the technical limitations of peptide quantification, undersampling, stochastic sampling process, and dynamic range [5,7], there is a limited ability to use unbiased and robust methods to treat large-scale data as a whole when aiming to determine novel candidate biomarkers and therapeutic targets.

Ideally, for candidate biomarker outcomes in proteomics, the list of thousands of proteins identified by the discovery methods must be reduced into a smaller subset of features that will provide the maximal discriminating power between the conditions of optimal sensitivity and specificity. Many methods have already been proposed to compare the protein abundance in label-free shotgun proteomics with the aim of finding evidence for candidate biomarkers in proteomics datasets. Most of these methods are based on p-values that were derived from *t*-test [15,16], ANOVA [17], Fisher's exact test [18,19], etc. However, although these methods point to differences in protein abundance individually across conditions, they are limited in analyzing sets of data that contain multiple classes as well as providing an optimal feature set that capture the maximal variance in the data. In this work, we aimed to retrieve ranked lists of candidate biomarkers, which are considered here to be proteins that change in abundance on average between the different biological sample classes. A combination of three different methods was tested: a univariate method, Beta-binomial, a semi-multivariate method, Nearest Shrunken Centroids (NSC), and a multivariate method, Support Vector Machine-Recursive Features Elimination (SVM-RFE).

The mentioned methods were selected based on the following main reasons: (1) Beta-binomial is a univariate statistical method that was described by Pham et al. [20] to test the significance of differential protein abundances that were expressed in spectral counts in mass spectrometry-based proteomics. Moreover, experimental results from the same work showed that

the Beta-binomial test performs favorably in comparison with other methods (e.g., Fisher's exact test, G-test, *t*-test and local-pooled-error technique) on several datasets in terms of both the true detection rate and the false positive rate and can also be applied in experiments with one or more replicates and in multiple condition comparisons; (2) NSC has already been shown to have the best performance compared to different univariate and multivariate methods in the previous work by Christin et al. [21]; (3) SVM-RFE is based on a machine-learning technique that has a completely different approach compared to NSC and was chosen as a complementary method to test both the results and the performances. NSC and SVM-RFE were combined to a double cross-validation step to define a final optimal set of discriminating proteins for distinguishing the three secretome classes with strictly low errors. Therefore, all of the three methods have already been separately tested and benchmarked for proteomics datasets, but they have not been used together in the same pipeline in which both the initial and final datasets were compared by different clustering techniques (heat map/hierarchical clustering and neighbor joining clustering) and silhouette coefficients. Furthermore, the final ranked lists of proteins were compared in a Venn diagram to be finally evaluated/validated by targeted proteomics in our proposed discovery-to-targeted pipeline.

In summary, the pipeline described in this work was tested on well-controlled data obtained from the secretomes of human melanoma (A2058 and SK-MEL-28), skin- and tongue-derived carcinoma (A431 and SCC-9, respectively) and non-cancerous (HaCaT and HEK293) cell lines. The MS-based discovery step was based on a routine shotgun analysis, which was followed by data analysis using the three mentioned approaches (Beta-binomial, NSC and SVM-RFE). These feature selection methods indicated that there was a panel of 137 proteins for carcinoma and 271 proteins for melanoma that were differentially abundant in these cell types. These selected proteins were then investigated by bioinformatics analyses, such as protein-protein interaction networks construction, enrichment analysis and literature curation. A protein network anticipated a potentially important role for the set of candidate biomarkers in the carcinoma, which was especially related to the complement and coagulation cascades, whereas in melanoma, the pathways associated with the cell cycle, cell adhesion and ubiquitin-mediated proteolysis were highlighted as being among the most altered in this pathologic condition. We further tested the strength of the pipeline in selecting candidate biomarkers by immunoblotting, human tissue microarrays, label-free targeted MS and functional experiments. It is noteworthy that the proteins Complement Factor B (CFB) and Complement C3 (C3) were found in significantly increased levels in oral squamous cell carcinoma (OSCC), compared to the adjacent normal tissue, and in human saliva from oral squamous cell carcinoma (OSCC) patients, using the pseudoSRM approach. Moreover, CFB knockdown

decreased both the migration in the skin-derived epidermoid carcinoma (A431) cell line and chemotaxis in human macrophages. Furthermore, the pipeline was also applied to a published proteomics dataset of prostate cancer [22], and the results were compared with the approaches that were previously used.

In conclusion, we suggest that our proposed integrative analysis based on a discovery-to-targeted pipeline is especially valuable to better characterize candidate biomarkers for targeted MS verification.

RESULTS

A novel experimental pipeline has been proposed in this study to provide the bridge between discovery MS and targeted MS. This pipeline comprises four steps: MS-based discovery, feature selection analyses, bioinformatic tools to boost the extraction of biological information and targeted validation (Fig. 1). As a proof of concept, melanoma (A2058 and SK-MEL-28), skin and tongue-derived carcinoma (A431 and SCC-9, respectively) and non-cancerous cell lines (HaCaT and HEK293) had the protein content of their secretome collected, concentrated, trypsin digested and analyzed by LC-MS/MS. State-of-the-art univariate and multivariate methods were then employed to identify the most differentially abundant proteins among the three classes. A bioinformatics platform compiled these data into integrative networks that revealed cancer-specific biological information. These networks were able to characterize both carcinoma and melanoma cell archetypes and to point out pathways that could be potentially altered in each condition. Protein expression by tissue array in carcinoma and melanoma patients' samples and by saliva samples, as well as gene silencing and functional experiments in cell lines provided validation for the proposed pipeline. Along with these findings, we also described the results that were obtained when the same pipeline was applied to an external dataset, which was a published study on prostate cancer [22]; these findings reinforced the effectiveness of our approach.

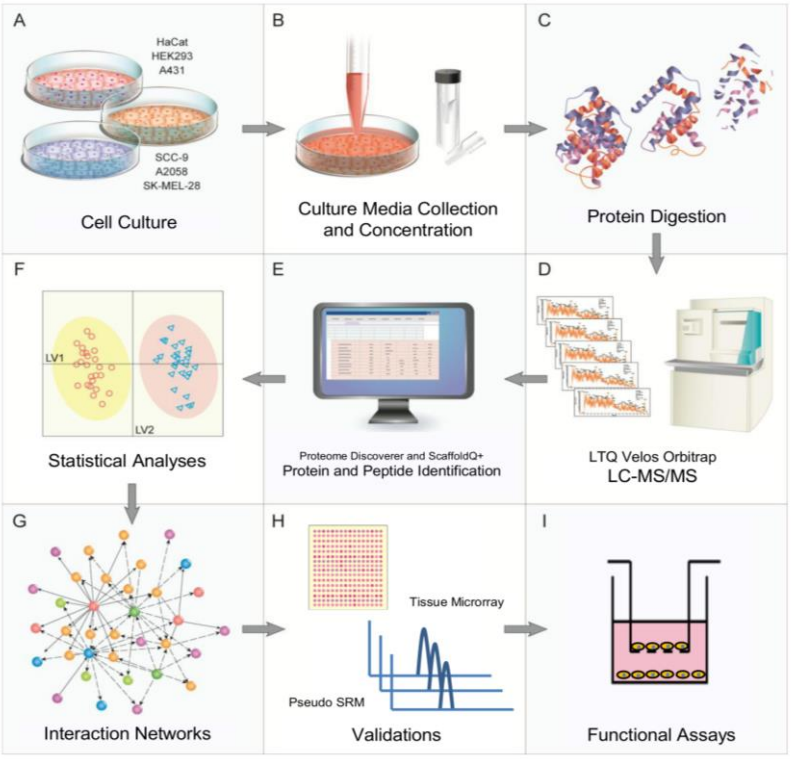

Figure 1: Experimental workflow and overview of the proteomics and bioinformatics analyses, validations and functional assays.

Data Analyses

Label-free Quantitation in Data Dependent Analysis

A list of 2,574 proteins with less than 1% FDR was generated by the Scaffold Q+ software from three biological replicates (Supplementary Table S1 and S2). From this total, 877 proteins presented spectral counts ≤ 2 and were discarded from the following steps, leaving 1,697 remaining proteins for the subsequent analyses. The number of proteins that were identified in each experiment is shown in Table 1, whereas the number of proteins that were exclusive or shared by the cell lines is available in Supplementary Fig S1.

Table 1: Number of proteins identified per experiment in each cell line.

Cell Line	Number of Proteins Identified by Mass Spectrometry		
	Exp.1	Exp.2	Exp.3
HaCaT	2015	2197	2201
HEK293	1690	1904	1861
A431	1879	1781	1884
SCC-9	2036	2213	2180
A2058	1950	1926	1974
SK-MEL-28	1660	1554	1770
Total		2574	
Total Spectra		151,221	

Clustering and feature selection analyses of proteomics data

An unsupervised hierarchical clustering performed with the 1,697 proteins mentioned above segregated the samples into two main classes, one that was composed exclusively by melanoma cell lines and the other that was composed by carcinoma and non-cancerous cells (Fig 2A). Interestingly, the basal cluster segregated the cells according to their tissue of origin: from the epithelium-derived cell lines (SCC-9, A431 and HaCaT), from the skin-derived melanoma cells (SK-MEL-28 and A2058) and from the human kidney non-cancerous cells (HEK293), although a perfect group segregation for either non-cancerous or cancer cell lines was not observed. The result of this exploratory, unsupervised analysis indicated that melanoma's secretome is radically different from that produced by carcinoma and non-cancerous cells. That finding is probably due to the considerable similarities that are found between carcinoma and non-cancerous cell secretome, despite their obvious differences.

Aiming to evoke the most prominent dissimilarities among the groups, univariate, semi-multivariate and multivariate analyses were conducted, including the Beta-binomial, NSC and SVM-RFE methods, respectively. The models retrieved 601, 130 and 13 proteins, respectively, that were differentially abundant among the three secretome classes. These proteins were further associated with each class after a decision boundary step (Supplementary Table S3).

Both the SVM-RFE and NSC methods had their performance assessed in terms of double cross-validation errors, accuracy, sensitivity and specificity. These models presented 5.5% and 0% errors in double cross-validation, and 94.4% and 100% accuracy, respectively. Regarding the sensitivity and specificity, SVM-RFE showed 83.3% sensitivity for carcinoma and 100% for the other classes and 91.7% specificity for non-cancerous and 100% for the other classes, whereas NSC exhibited 100% sensitivity and specificity for all of the three classes (Supplementary Table S4).

The candidate biomarkers retrieved from the feature selection analyses were also used to perform the hierarchical clustering and heat maps again using the MetaboAnalyst platform (Fig. 2B). By this later analysis considering only the selected features, the same-class cell lines were clustered together, which confirms the set of retrieved candidate biomarkers as good discriminating proteins for distinguishing the three secretome classes (Fig. 2B). From this set, the Beta-binomial, NSC and SVM-RFE models retrieved 135, 32 and 4 characteristic proteins for carcinoma and 269, 78 and 6 proteins for melanoma, respectively (Supplementary Table S5).

Furthermore, the final ranked lists of 601, 130 and 13 candidate biomarkers for all three classes, which resulted from the Beta-binomial, NSC and SVM-RFE models, respectively, were compared by a Venn diagram and by the Jaccard similarity coefficient. This comparison showed that the SVM-RFE optimal feature subset is almost completely shared by the NSC and Beta-binomial models (12 out of 13 proteins) and that the NSC optimal feature subset is almost completely shared by the Beta-binomial model (128 out of 130 proteins) (Fig 2C). Moreover, the NSC and Beta-binomial output rankings were similar starting from the top-ranked proteins, and all three models appeared to have a similar and stable behavior from the 10^{th} to the 130^{th} protein in their output ranked lists (Fig 2D). Notably, the SVM-RFE model was able to discriminate the three classes based on the smallest set of only 13 proteins (gene names: C3, CLU, MEGF10, MMP8, BANF1, VIM, APEX1, CA2, TACSTD2, KRT8, TNC, C1R and IGFBP7), of which only BANF1 was not retrieved by the other two methods. In contrast, as expected for a univariate method, the Beta-binomial model yielded the largest set of differentially abundant proteins, covering all of the proteins that were retrieved by the two multivariate methods (except for two proteins from NSC). Notably, using only the 12 candidate biomarkers retrieved by the three feature selection methods, a perfect segregation among the carcinoma, melanoma and non-cancerous classes was also observed (Fig 2E). The complete ranked protein lists that resulted from the three methods are available in Supplementary Table S3; the plots showing the spectral count distribution in the melanoma, carcinoma and non-cancerous cells for the 130 candidate biomarkers retrieved from NSC can be found in Supplementary Fig. S2.

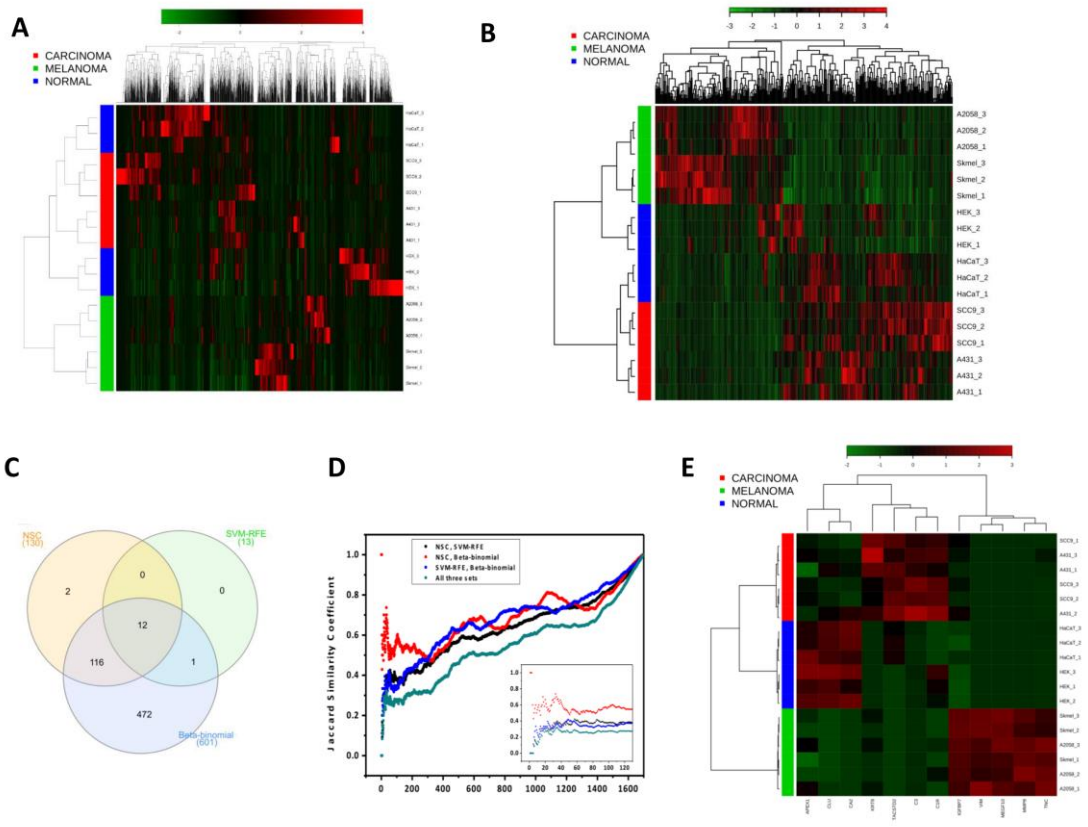

Figure 2: Comparison of the three feature selection methods (Beta-binomial, SVM-RFE and NSC) used to identify differentially abundant proteins among carcinoma, melanoma and non-cancerous cells. (A) Clustering of the whole secretome dataset before applying feature selection methods. From the 2,574 proteins identified and quantified by spectral counts, 1,697 (65.9%) compose the heat map. The 877 remaining proteins exhibited ≤ 2 spectral counts and were excluded from the analysis. (B) Clustering after applying feature selection methods. 603 significant differentially abundant proteins among melanoma, carcinoma and non-cancerous classes selected by Beta-binomial, NSC and SVM-RFE analyses compose the heat map. (C) Venn diagram showing the intersections among the optimal feature subsets (N) retrieved by the three methods. (D) Jaccard similarity coefficient vs. the optimal feature subset (N) retrieved by each method. (E) Clustering of the 12 significant differentially abundant proteins among melanoma, carcinoma and non-cancerous classes identified in the intersection of Beta-binomial, NSC and SVM-RFE analyses. The secretome dataset is composed by non-cancerous cells (HaCaT and HEK293), carcinoma (A-431 and SCC-9) and melanoma (A2038 and SK-MEL-28) cell lines.

In addition to the hierarchical clustering and heat map analysis, similarity trees were constructed from a Euclidean distance matrix of the 18 samples considered for the feature selection analyses. Figure 3 shows that the Neighbor Joining (NJ) trees were capable of showing the most similar elements of the set, which were present in the same or in nearby branches. In this work, a

reasonable separation of the three classes was found when the whole dataset was considered in the NJ tree construction (Fig. 3A) (silhouette coefficient, SC > 0.2). However, as was also shown by the previous unsupervised hierarchical clustering and heat map analysis for the whole dataset, the melanoma samples were the only ones that clustered together in the same or nearby branches connected to the same node, separated from the carcinoma and non-cancerous samples, which were distributed in different branches and did not show a perfect segregation in their respective classes. On the other hand, as expected, there was an improvement in the NJ clustering and silhouette coefficients that were calculated after feature selection, considering only the candidate biomarkers that were retrieved from each model (Fig. 3B-D). For instance, if a labeled dataset has a silhouette coefficient that is closer to 1 (ranging from -1 to 1), then the classes are almost homogeneous and different from each other and classifiers will probably perform well in constructing a good model with a low double cross-validation error. Consequently, this finding also means that sets of good discriminating features (proteins) among the classes could be retrieved by feature selection analysis.

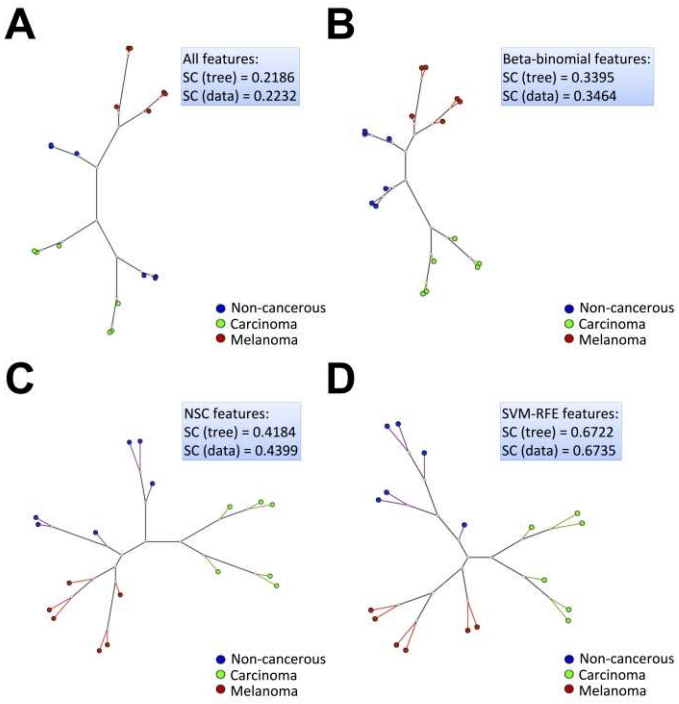

Figure 3: Neighbor joining (NJ) clustering calculated from a Euclidean distance matrix of the secretome dataset samples, considering (A) all features (1,697 proteins), (B) Beta-binomial (601 proteins), (C) NSC (130 proteins) and (D) SVM-RFE (13 proteins) features. SC (tree) stands for silhouette coefficient calculated from the NJ tree and SC (data) stands for silhouette coefficient calculated directly from the original data of each analysis.

Besides the feature selection methods described above, the univariate ANOVA test was also performed in our data to compare our results to a classical statistical method. In total, ANOVA retrieved 875 differentially abundant proteins ($p<0.05$, Supplementary Table S3). The ANOVA result corroborates the results obtained by the three methods proposed in our pipeline, which can be observed by the intersections in a second Venn diagram built for the four output lists of candidate biomarkers (Supplementary Fig. S3). However, it brought over 384 exclusive proteins from a total of 987 proteins selected by the four methods (~40%), which is a large percentage for ANOVA to be considered in the pipeline as a method that could contribute with an optimal set of features for selecting candidates. Moreover, when we compare the rank index of candidate biomarkers retrieved by the three feature selection methods proposed in the pipeline to the rank index given by ANOVA, we observed that the reduced list of candidates selected by the three approaches were not the top candidates chosen by ANOVA (Supplementary Table S3).

The same feature selection analyses were also performed for a published proteomics dataset on prostate cancer [22] to validate our approach. The output final ranked lists of candidate biomarkers that resulted from each method (Supplementary Tables S6-S10) were analyzed by a Venn diagram, which showed that five candidates that were validated/verified by different approaches in the original work by Kim et al. [22] were also identified in the intersections of the Venn diagram (Supplementary Fig. S4 and Supplementary Table S9), which reinforces the effectiveness of the proposed discovery-to-targeted pipeline.

Bioinformatics analyses

To evaluate the protein interaction profile within each tumor class, protein-protein interaction networks were constructed using the IIS software for the candidate biomarkers obtained by the feature selection methods and estimated to be associated to carcinoma or melanoma classes (Supplementary Table S5). The networks represent a "snapshot" of the secretome of both classes, which illustrate the proteins that most probably play a role in the secretome regulation of each tumor type (Fig. 4, Supplementary Tables S11 and S12). Our network analysis showed direct connections between the identified candidate biomarkers and the proteins that are involved in enriched KEGG pathways ($p \leq 0.05$), highlighting the most important pathways that are likely to be activated/inhibited in each disease. Accordingly, the networks suggested a potentially important role for carcinoma biomarkers in focal adhesion, regulation of actin cytoskeleton, ECM-receptor interaction, glutathione metabolism, glycolysis/gluconeogenesis and, especially, complement and coagulation cascades, which were not enriched among the melanoma biomarkers (Fig. 4A). In

contrast, focal adhesion, cell cycle, regulation of actin cytoskeleton, ECM-receptor interaction, cell adhesion molecules, glycolysis/gluconeogenesis and ubiquitin-mediated proteolysis were identified as being significantly enriched ($p \leq 0.05$) pathways in the melanoma secretome; these pathways presented at least one candidate biomarker that participates in each of them. Focal adhesion and ECM-receptor interaction, especially, appear to have a relevant role in melanoma due to the outstanding, differential abundance of the proteins that belong to these two pathways (Fig. 4B). All of the significant enriched pathways that were extracted from both carcinoma and melanoma networks are listed in Supplementary Tables S11 and S12, respectively.

Furthermore, to verify whether the candidate biomarkers for carcinoma and melanoma had been previously described as related to cancer or to some biomarker application, an Ingenuity (IPA) biomarker filter module analysis and a search in the Human Protein Atlas Database were performed.

The IPA biomarker analysis retrieved 45 (32%) proteins of the candidate biomarkers identified in our study for carcinoma and 76 (28%) for melanoma, which have been previously described as being strongly associated with cancer and/or involved in biomarker applications. Likewise, the Human Protein Atlas Database retrieved 32 (23%) and 60 (22%) proteins of the carcinoma and melanoma candidates, respectively, which were found to be previously associated with cancer (Supplementary Table S5).

These analyses were also performed for the set of candidate biomarkers that were retrieved by the feature selection methods applied to the prostate cancer proteomics dataset published by Kim et al. [22]. Interestingly, from the 47 proteins that were identified in the intersection of the three methods (Supplementary Fig. S4 and Supplementary Table S9), IPA determined that 13 (28%) proteins had been associated with some biomarker application, whereas 17 (36%) proteins had already been described as candidate cancer biomarkers according to the Human Protein Atlas Database (Supplementary Table S10).

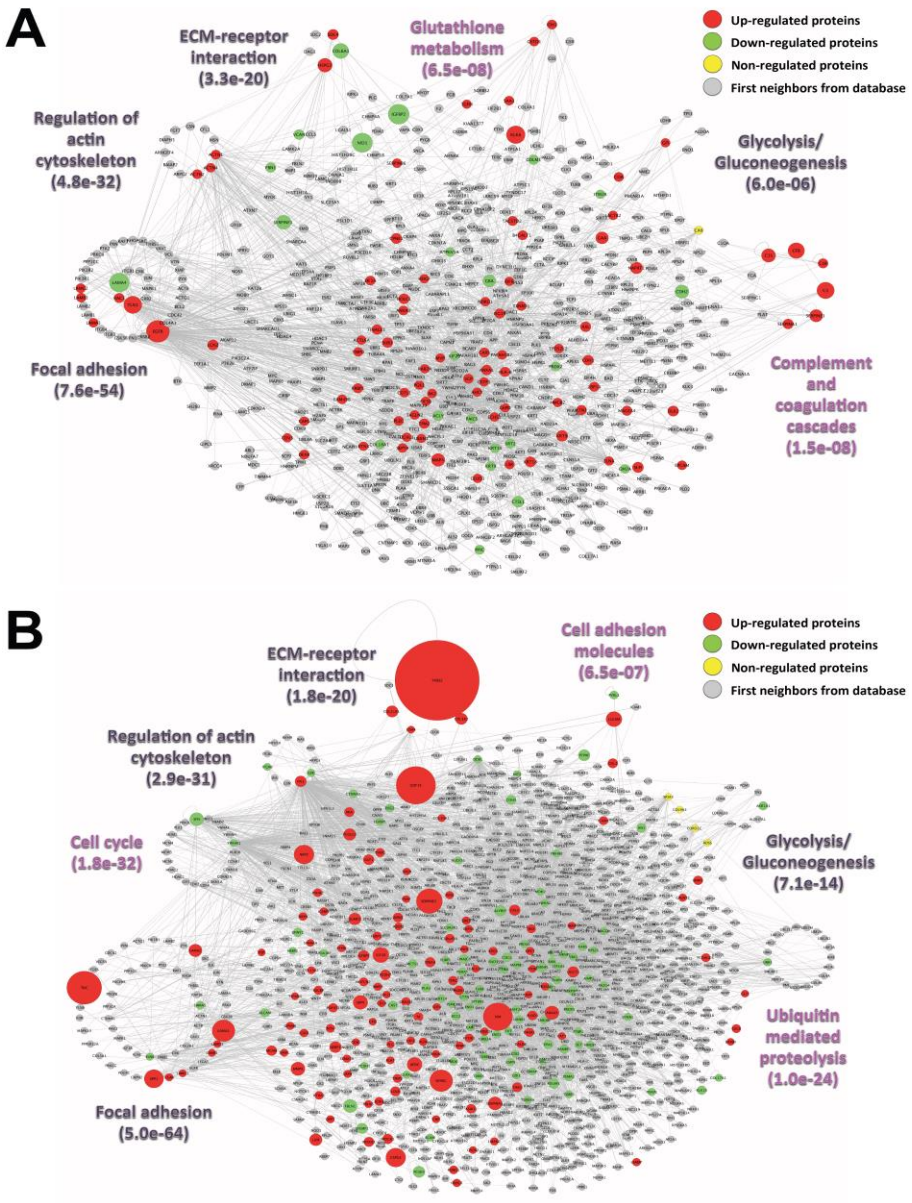

Figure 4: Interaction networks of the identified (A) carcinoma and (B) melanoma candidate biomarkers by Beta-binomial, NSC and SVM-RFE analyses. The selected most relevant enriched KEGG pathways ($p \leq 0.05$) among the up-regulated (red), down-regulated (green), non-regulated (yellow) and background intermediary proteins (grey) from the IIS database are depicted by clustering with a circular layout proteins involved in each respective pathway. Clusters were assigned only to pathways containing more than three proteins with at least one protein from the proteome dataset (disease pathways or pathways specific for defined cell types were not considered); proteins belonging to more than one pathway were assigned to the pathway clusters with the best enrichment p-values; some proteins were also assigned to different pathway clusters based on complementary data from the Uniprot database. In magenta, pathway clusters exclusive of each network; in black, pathway clusters in common. The node sizes of up, down and non-regulated proteins are proportional to their fold change ($-1.3 \geq$ fold change ≥ 1.3, compared to the non-cancerous class). The protein-protein networks were built using the IIS software and visualized using Cytoscape.

Validation of the expression of candidate markers for melanoma and carcinoma

Based on the available commercial antibodies, six up-regulated proteins retrieved by Beta-binomial, NSC and/or SVM-RFE models were chosen to be validated by immunoblotting. The overexpression of Fibronectin (FN1), Tenascin-C (TNC) and Growth/differentiation factor 15 (GDF15) in melanoma cell lines and of Complement factor B (CFB), Talin-1 (TLN1) and Epidermal growth factor receptor (EGFR) in carcinoma cell lines was confirmed in the conditioned media of the six cell lines (Supplementary Fig. S5).

To further investigate whether those candidate markers were clinically associated with tumors, we used tissue microarrays with human melanoma samples to examine TNC and GDF15 expression. Both TNC and GDF15 were found in the cytoplasm of the nevoid cells, with significantly higher expression levels in the tumor cells compared with normal cells (Fig. 5-A and 5-B). Interestingly, the expression of GDF15 was significantly higher in metastatic than in primary melanomas (one-way ANOVA, $p < 0.0001$).

The expression of CFB was limited to the cytoplasm of the basal and suprabasal layers of the normal oral tissue, whereas broad positivity was found in the tumor cells (Fig. 5-C). Considering the intensity levels, the expression of CFB was significantly higher in tumors compared with normal mucosa (Mann Whitney U test, $p = 0.0057$, Fig. 5-C). Similarly, C3 was found in the cytoplasm of the epithelial cells, but the intensity was significantly higher in tumor cells compared to normal keratinocytes (Mann Whitney U test, $p = 0.016$, Fig. 5D). Immunoreactivity for C3 was also observed in inflammatory and endothelial cells.

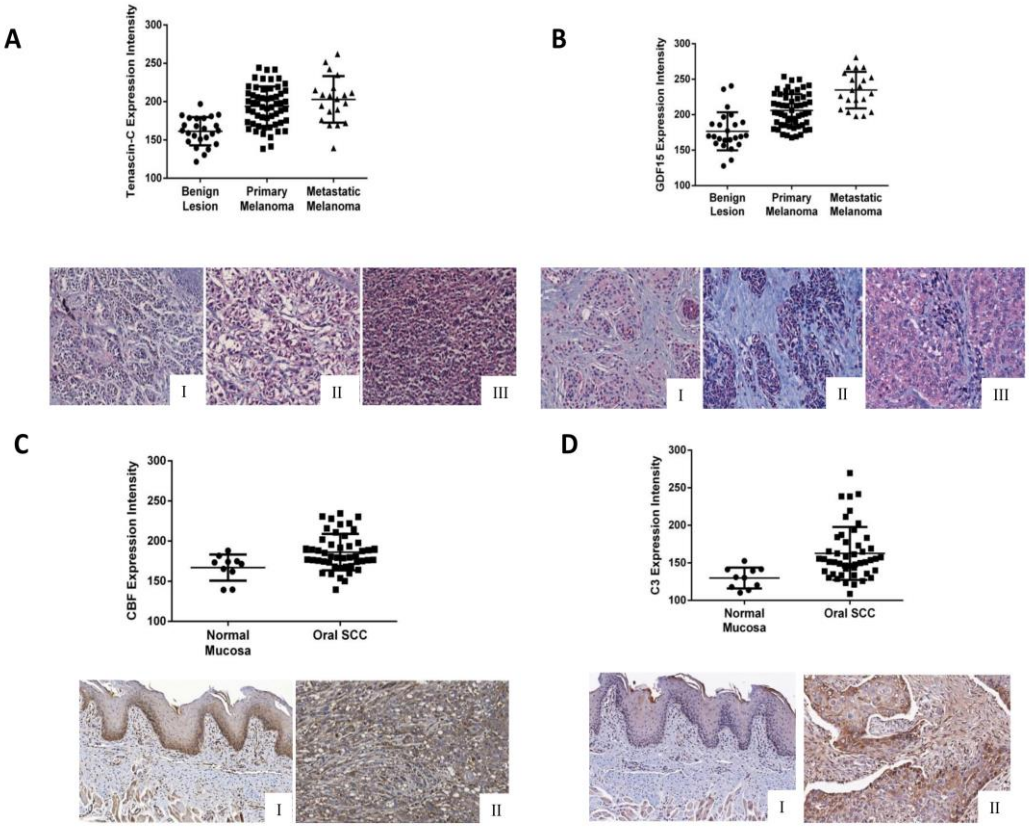

Figure 5: Validation of the higher expression of (A) tenascin-C and (B) GDF15 (I-Benign lesion; II-Primary Melanoma; III-Metastatic Melanoma) on melanoma cancer tissue microarrays and (C) CFB and (D) C3 (I- Normal Mucosa; II- Oral SCC) on carcinoma cancer tissue microarrays. Tenascin-C showed statistically significant expression among the categories benign lesion, primary melanoma and metastatic melanoma, but not between primary melanoma and metastatic melanoma (One-way ANOVA, benign lesion vs. primary melanoma, $p < 0.0001$; benign lesion vs. metastatic melanoma, $p < 0.0009$; primary melanoma vs. metastatic melanoma, $p = 0.1748$). GDF15 showed statistically significant expression among the categories benign lesion, primary melanoma and metastatic melanoma (One-way ANOVA, benign lesion vs. primary melanoma, $p < 0.0001$; benign lesion vs. metastatic melanoma, $p < 0.0001$; primary melanoma vs. metastatic melanoma, $p < 0.0001$). CFB and C3 showed higher expression in OSCC compared with normal mucosa (Mann Whitney U, $p = 0.009$ and $p = 0.0005$, respectively).

Label-free targeted MS

To further test the strength of the pipeline in selecting candidate biomarkers that were retrieved by all of the methods, we prioritized two candidates from the carcinoma secretome to have their abundance assessed in the saliva of Oral Squamous Cell Carcinoma (OSCC) patients, as a first step toward biomarker evaluation in clinical samples. We believe that saliva is a promising biofluid for investigation due to the ease of its collection and its direct contact with oral cancer lesions. The

samples were collected from OSCC patients, who were divided into two groups: patients who had undergone surgical resection (named as "no lesion", n=7) and those who had active oral malignant lesion (named as "lesion", n=10) at the time of the saliva collection (Supplementary Table S13). Saliva samples from healthy individuals were also used as a control (n=7).

We validated both C3 and CFB, and C3 was selected for being top ranked in the three feature selection analysis results (SVM-RFE rank index=1; NSC rank index=13; Beta-binomial rank index=15) as well as for being assigned to the complement and coagulation cascades pathway, an enriched (p-value = 1.52e-08) carcinoma-exclusive KEGG pathway that is based on complementary data from the Uniprot database. Regarding CFB, it was simultaneously retrieved by NSC (rank index=33) and Beta-binomial (rank index=31), and very importantly, it takes part in the same pathway as C3. In addition, both CFB and C3 have not been previously reported to be related to cancer biomarkers, according to the IPA biomarker analysis and The Human Protein Atlas.

We selected two peptides for each protein based on the following criteria: uniqueness, high relative abundance, MS/MS spectral quality, experimental observation of proteomic data repositories (PeptideAtlas) and DDA analysis performed using LTQ Orbitrap Velos. The targeted proteomics were performed using selected ion monitoring (SIM) of each targeted peptide in high mass resolution for quantitation, followed by scheduled MS/MS for confirming targeted peptide sequences (Supplementary Table S14). The peak area of each targeted peptide was extracted using the Xcalibur software (Supplementary Table S15) and normalized to the angiotensin internal standard (Supplementary Table S16), spiked in all of the samples to a final concentration of 5 fmol/µl to correct run-to-run variations.

The averages of the normalized intensities of each peptide in each sample were visualized in a scatter plot graph, and ANOVA followed by Tukey's test was performed to evaluate the statistical significance among the conditions (Fig. 6). It was observed that the saliva from the OSCC patients with lesions had a significantly higher normalized intensity of the precursor area of both CFB and C3 compared to healthy subjects with respect to all of the peptides evaluated (Fig. 6A-D). Additionally, the C3 peptide, IPIEDGSGEVVLSR, and the CFB peptide, YGLVTYATYPK, both showed a significant difference between the patients without a lesion and with a lesion (Fig. 6B and 6C). When the sum of the three transitions of each peptide (normalized by the sum of the three transitions of the angiotensin internal peptide) was considered, similar results were found (Supplementary Fig. S6). The extracted ion current peak area from MS1 and the three MS/MS transitions as well as the CV% of each replicate are shown in Supplementary Tables S15 and S16.

The performance of the method was evaluated using angiotensin spiked in the HEK cell lysate digest (500 ng) in five different concentrations, for which each sample was run in triplicate. Good linearity (R=0.998) (Supplementary Fig. S7) and CV<15% were observed at three concentration points (Supplementary Table S17).

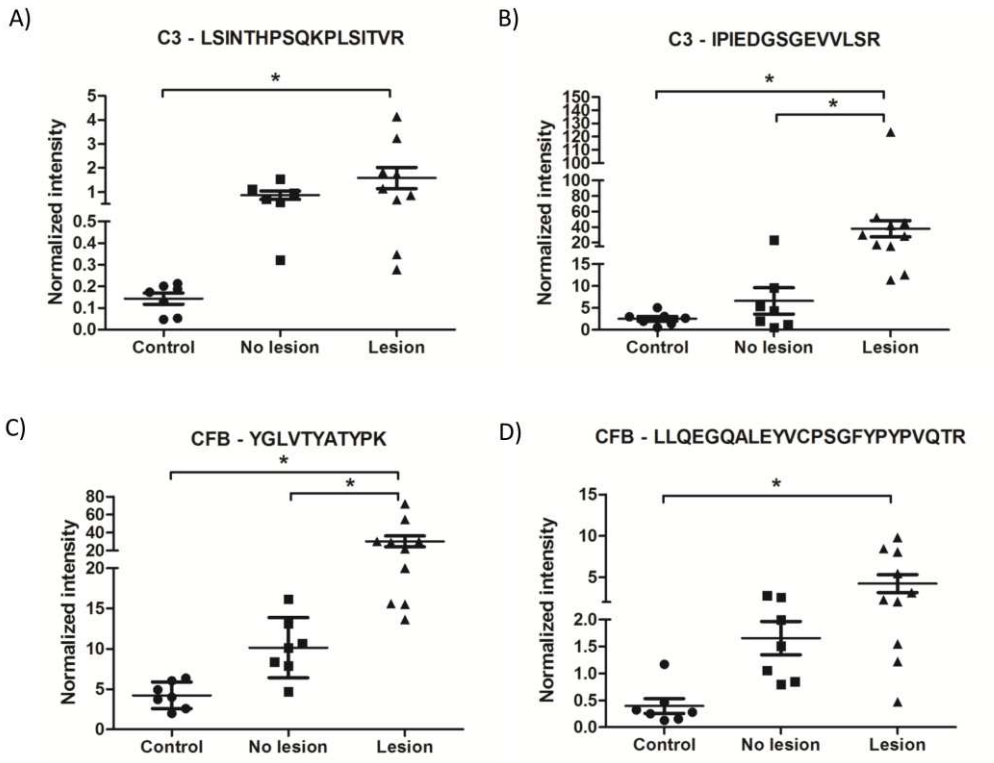

Figure 6: CFB and C3 peptides showed higher normalized intensities in OSCC saliva samples than in healthy saliva samples. PseudoSRM analytical approach for peptides of C3 (precursor m/z 631.05, +3; 735.89, +2) and CFB (precursor m/z 638.33, +2; 939.13, +3) normalized with 5 fmol/µl of angiotensin (m/z 432.89, +3) as an internal reference peptide. These data represent two technical replicates of saliva samples from healthy patients (n=7), saliva samples from patients who undergone surgical resection of OSCC (named no lesion, n=7) and saliva samples from patients with active OSCC lesion without any treatment (named lesion, n=10) (ANOVA followed by Tukey's test). The normalization to the internal reference peptide was performed for each run.

CFB knockdown decreased the migration of A431 cells and impaired the chemoattraction of human macrophages

The final approach that was used to explore the strength of the pipeline was to perform functional assays, which was chosen because of the implication that CFB could have in biological processes that are related to cancer.

It is well known that complement proteins are considered to be powerful proinflammatory molecules in the body [23], and recently, C3 was evidenced as a key player in the production and activation of ovarian cancer growth and progression [24]; however, there is still no evidence associated with CFB in oral tumorigenic processes. Therefore, we performed the knockdown of CFB in the A431 cell line using siRNA, and we first evaluated the effect of this protein in cell migration. As observed in Fig. 7A, CFB knockdown decreased the migration of A431 cells compared with mock and control siRNAs (one-way ANOVA followed by Tukey's test, n=2, $p<0.001$).

Furthermore, CFB is a protein that is secreted by macrophages, fibroblasts, endothelial cells and tumor cells [23]. Therefore, we evaluated the paracrine effect that CFB depletion in tumor cells could exert on macrophage chemotaxis. To accomplish this goal, macrophages were placed in the upper chamber of a transwell plate, whereas A431 cells that were treated either with mock, control siRNA or siRNA against CFB were laid in the lower chamber of the same plate. Macrophage migration through the transwell was significantly reduced in CFB knockdown A431 cells, which suggested that the presence of this protein in the conditioned medium had the ability to modulate macrophage taxis (Fig. 7B, one-way ANOVA followed by Tukey's test, n=2, $p<0.001$). Cell knockdown for CFB was confirmed by qRT-PCR (Fig. 7C). Together, these experiments showed that CFB protein plays a role in tumorigenic processes such as macrophage chemotaxis and cell migration.

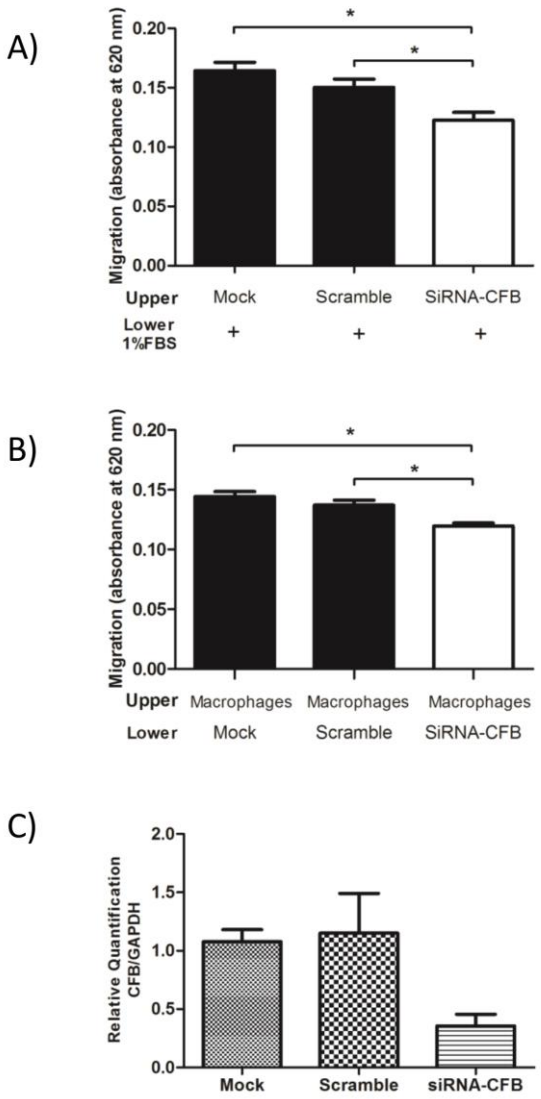

Figure 7: CFB knockdown decreased the migration of skin-derived epidermoid carcinoma (A431) cells and reduced the chemotaxis of human macrophages. (A) A431/untreated (mock), A431/control (scrambled) and A431/siRNA CFB cells were seeded in serum-free media in the upper chamber of a 96-well transwell plates. RPMI media, which was supplemented with 1% FBS, was added in the lower chamber (n=2, triplicate, one-way ANOVA followed by Tukey's test, * $p < 0.05$). (B) Chemotaxis of human macrophages was reduced when were seeded in the upper chamber, and A431 cells treated with mock, control siRNA and siRNA against CFB were added in the lower chamber of the transwell (n=2, triplicate, a one-way ANOVA followed by Tukey's test, *$p < 0.05$). C) Real-time quantitative PCR confirms the inhibition of the expression of CFB after transient transfections with siRNA in A431 cells. The data were normalized with the (glyceraldehyde-3-phosphate dehydrogenase gene was used as internal reference). Each bar represents mean ± SD of three independent experiments.

DISCUSSION

This study introduced an integrative analysis based on a pipeline that combines MS-based discovery followed by feature selection methods, clustering, Venn diagram, network analyses, and targeted approaches to generate reliable hypothesis-driven targets based on shotgun proteomics, to provide a bridge between discovery MS and targeted MS.

Well-controlled proteomic data from the secretomes of three classes of human cell lines were analyzed with respect to the protein content of their secretomes using discovery-based proteomics. To retrieve ranked lists of candidate biomarkers, a combination of a univariate method (Beta-binomial), a semi-multivariate method (NSC) and a multivariate method (SVM-RFE) was tested. The great advantage of the feature selection methods used in this work is that NSC and SVM-RFE models summarize thousands of features into a few key components that capture the maximal variance in the data. Together with the Beta-binomial model, which was used to test the significance of differential protein abundances expressed in spectral counts, the three ranked lists of candidate biomarkers were retrieved and compared using the Jaccard similarity coefficient and a Venn diagram, to be further evaluated by bioinformatic analyses (interaction networks, pathway enrichment and biomarker investigation) and targeted proteomics. Moreover, both the initial and final datasets were compared by different clustering techniques (heat map/hierarchical clustering and neighbor joining clustering) and silhouette coefficients, which showed an improvement in both the clustering and silhouettes after feature selection and served as a proof-of-concept that the set of retrieved candidates was constituted by good discriminating proteins for distinguishing the three secretome classes.

Our approach proved to be of great value in tracking potentially promising candidate biomarkers from proteomics data, since many of these proteins have already been demonstrated to be associated with cancer. For example, both the IPA and Human Protein Atlas Database analyses retrieved, respectively, 32% and 23% of the carcinoma candidates and 28% and 22% of the melanoma candidates that were previously found to be associated with cancer (Supplementary Table S5).

To further explore the biological role of these findings, we integrated the proteomics data into networks that highlighted the direct connections between the selected candidates and their possible roles in each disease (Fig. 4). Despite the highly complex and dynamic nature of network biology [25], our interaction networks enabled us to easily sum up all of the proteomics data and decipher the main cellular contexts of the candidate biomarkers in carcinomas and melanomas. The

proteins were clustered in highly enriched pathways and were visualized by their relative abundances through node colors and sizes; most of the candidates were found to be related to cell-cell communication and interactions. Specifically, this analysis retrieved exclusive pathways for the carcinoma candidates, such as complement and coagulation cascades (Fig. 4A), and for the melanoma candidates, such as cellular functions associated with the cell cycle, cell adhesion and ubiquitin-mediated proteolysis (Fig. 4B).

In the final steps of the proposed pipeline, we tested the promising proteins CFB and C3 as candidate carcinoma biomarkers, which in addition to being associated with the enriched complement and coagulation cascade pathway, were validated using immunoblotting, tissue microarrays and retrieved in the intersections of the feature selection methods.

In the first approach, we have indeed found a higher expression of CFB and C3 proteins using a label-free pseudoSRM analysis of human saliva from Oral Squamous Cell Carcinoma (OSCC) patients in comparison with healthy individuals (Fig. 6). Because saliva is simple to collect and process, it may lead to a useful clinical tool for the noninvasive prognosis of oral cancer in the future [13,26]. It is important to highlight that oral cancer, primarily OSCC, is the sixth most common cancer and is an important public health concern worldwide [27], with low 5-year survival rate due to the compounding factors of late detection and lack of truly effective therapies [28,29].

Although complement components are primarily synthesized locally by many cell types, including macrophages, fibroblasts and endothelial cells [30], some neoplastic cells have also been shown to synthesize and secrete components of the C system [31-33]; however, the role of the complement system in tumor cells remains controversial. Recently, an autocrine effect of complement proteins has been shown; specifically, C3 and C5 are secreted by ovarian cancer cells on tumor growth [24]. It is also well known that the complement system contributes to inflammation, mainly through C3a and C5a, which are the most powerful proinflammatory anaphylatoxins in the body [23,34] and to immunosuppression through components such as C3, C4 and C5a [35]. Interestingly, the adopting characteristics that involve the inflammatory state and the ability to avoid the immune system have been emerging as hallmarks in cancer [36].

Because no evidence was shown regarding the function of CFB in cancer cells, in the second approach, we explored the functional role of CFB in tumorigenic processes, such as cell migration and chemotaxis. The CFB knockdown in the skin-derived epidermoid carcinoma (A431) cells decreased the ability of the cells to migrate and the chemotaxis of human macrophages (Fig.

7A/B), which suggests that, in addition to a higher expression in OSCC tissues and saliva, CFB might mediate these events in carcinomas.

Furthermore, we applied our pipeline for a published label-free proteomic dataset [22], which previously reported the identification of 133 significantly differentially expressed proteins in extracapsular and organ-confined prostate cancer direct-EPS fluids using a hierarchical Bayesian statistical algorithm known as QSpec. Among these proteins, five proteins were validated/verified using different methods (ELISA, Western blot and SRM-MS). Using the feature selection methods proposed in our pipeline, the same five proteins validated by Kim et al. [22] were also found in the intersections of our Venn diagram analysis (SFN, MME, TGM4, TIMP1 and PARK7, Table E9), reinforcing the effectiveness of our approach.

In conclusion, the proposed integrative analysis based on a discovery-to-targeted pipeline was able to pre-qualify potential candidates from discovery-based proteomics to targeted MS and can contribute to the next phases of biomarker development in translational initiatives to drive either patient stratification, decision making or intervention.

MATERIALS AND METHODS

Cell culture

SCC-9 cells (squamous cell carcinoma, a tumor cell line originated from a human tongue squamous cell carcinoma) were obtained from the American Type Culture Collection (ATCC, Manassas, VA) and cultured in DMEM/Ham's F12 medium (Cultilab), supplemented with 10% fetal bovine serum (FBS), antibiotics and 0.4 μg/ml hydrocortisone. Human keratinocyte HaCaT (immortalized, but not transformed, epithelial cell line), Human embryonic kidney HEK293 and human melanoma A2058 cell lines (isolated from a metastatic site in a skin-derived lymph node) were maintained in DMEM containing 10% FBS and antibiotics. Human melanoma SK-MEL-28 cells (malignant skin-derived melanoma cell line) and human epidermoid carcinoma A431 (skin-derived epidermoid carcinoma cell line) were grown in Roswell Park Memorial Institute (RPMI)−1640 medium supplemented with 10% FBS and antibiotics. All cells were maintained at 37°C in a 5% CO_2 atmosphere.

Sample preparation for MS

Label-free Discovery Proteomics: Cells at 80% confluence (two 15-cm dishes per condition per experiment) were gently washed three times in phosphate buffered saline (PBS) and incubated in a serum-free medium (20 ml per dish) for 24 h at 37 °C. After collection of the conditioned media EDTA and PMSF (Phenylmethylsulfonyl fluoride) were added at a final concentration of 1 mM. Cell debris and intact cells were eliminated by centrifugation at 4,000 rpm (Eppendorf Centrifuge 5810R) for 5 min at 4 °C and the conditioned media were subsequently concentrated using a 3000-Dalton centrifugal filter (Millipore, Billerica, MA) at 4,000 x *g* at 4°C. Protein concentrations were determined using a Bradford assay (Bio-Rad, Hercules, CA, USA). Proteins (80 μg) were treated with a final concentration of 1.6 M urea, following reduction (5 mM dithiothreitol, 25 min at 56°C), alkylation (14 mM iodoacetamide, 30 min at room temperature in the dark) and digestion with trypsin (1:50, w/w). The reaction was stopped with 1% TFA and desalted with Sep-pack cartridges (Waters). The samples were dried in a vacuum concentrator, reconstituted in 0.1% formic acid and analyzed by LC-MS/MS. Three independent experiments were performed for each cell line.

Label-free Targeted Proteomics: The saliva was collected from healthy individuals (n=7), patients who underwent surgical resection (named as no lesion, n=7) and patients with active oral malignant lesion (named as lesion, n=10). Individuals were asked to first rinse their mouth with

5 ml of drinking water and to harvest the saliva into a glass receptacle. Saliva was then aliquot in 2 ml tubes and immediately frozen at -80 °C. All patients and volunteers enrolled signed a formulary stating their awareness and consent for the study, approved by the Research Ethics Committee of Faculdade de Odontologia de Piracicaba, Universidade Estadual de Campinas, UNICAMP, Piracicaba, Brazil.

Proteins were extracted by homogenizing the 100 µl of whole saliva with 100 µl of a solution containing 100 mM Tris-HCl, pH 7.5, 8 M urea, 2 M thiourea containing Protease Inhibitor Cocktail cOmplete Mini Tablets (Roche, Auckland New Zealand), 5 mM EDTA, 1 mM PMSF and 1 mM DTT. Samples were sonicated for 10 min and centrifuged at 10,000 x *g* for 5 min. Protein concentrations were determined using a Bradford assay (Bio-Rad, Hercules, CA, USA). Five fmol/µl of angiotensin synthetic peptide (precursor m/z 432.8998, +3, DRVYIHPFHL, Sigma-Aldrich) were added to each peptide mixture (600 ng of total protein) as an internal reference peptide.

Mass spectrometric analysis

Label-free Discovery Proteomics: An aliquot containing 2.2 µg of peptides was analyzed on an ETD-enabled LTQ Orbitrap Velos Mass Spectrometer (Thermo Fisher Scientific) connected to a nanoflow liquid chromatography column (LC-MS/MS) by an EASY-nLC System (Proxeon Biosystem) through a Proxeon nanoelectrospray ion source. Peptides were separated by a 2-90% acetonitrile gradient in 0.1% formic acid using a pre-column EASY-Column (2 cm x id 100 µm, 5 µm particle size), and an analytical column PicoFrit Column (20 cm x ID75 µm, 5 µm particle size, New Objective), at a flow of 300 nl/min over 212 min. The nanoelectrospray voltage was set to 1.7 kV, and the source temperature was 275°C. All instrument methods for the LTQ Orbitrap Velos were set up in the data-dependent analysis (DDA) mode. The full scan MS spectra (m/z 300-2000) were acquired in the Orbitrap analyzer after accumulation to a target value of $1e^6$. The resolution in the Orbitrap was set to *r*= 60,000. The 20 most intense peptide ions with charge states ≥ 2 were sequentially isolated to a target value of 5,000 and fragmented in the linear ion trap by low-energy CID (normalized collision energy of 35%). The signal threshold for triggering an MS/MS event was set to 500 counts. Dynamic exclusion was enabled with an exclusion size list of 500, an exclusion duration of 60 s, and a repeat count of 1. An activation q= 0.25 and an activation time of 10 ms were used.

Label-free Targeted Proteomics: An aliquot containing 600 ng of digested proteins from saliva containing 5 fmol/µl of spiked synthetic angiotensin peptide was analyzed on an LTQ

Orbitrap Velos mass spectrometer as previously described. Briefly, peptides were separated by a 2-90% acetonitrile gradient in 0.1% formic acid using an analytical column PicoFrit Column (20 cm x ID75 μm, 5 μm particle size, New Objective), at a flow of 300 nl/min over 80 min. The resolution in the Orbitrap was set to *r*= 60,000. The AGC target was $1.00e^5$ for SIM scans in the Orbitrap mass analyzer and $1.00e^4$ for MS/MS scans in the ion trap mass analyzer. The wide SIM windows were defined between 15 amu over the mass range of the selected C3 and CFB peptides. Targeted MS/MS was performed in the linear ion trap using global scheduled inclusion lists. The data were analyzed using the Xcalibur software (Thermo Fisher Scientific) to determine the extracted ion current peak area for MS1 and three transitions for each targeted peptide. Two technical replicates from each sample were performed. Each peptide was normalized by dividing the individual peptide (precursor area or the sum of three ion transitions/peptide) by the individual angiotensin reference peptides (precursor area or the sum of three ion transitions/peptide).

Data analysis

Label-free quantitation in data dependent analysis: Peak lists (msf) were generated from the raw data files using the Proteome Discoverer software version 1.3 (Thermo Fisher Scientific) with the Sequest search engine and searched against the Human International Protein Database (IPI) v. 3.86 (91,522 sequences; 36,630,302 residues), with the following parameters: carbamidomethylation as the fixed modification, oxidation of methionine as the variable modification, one trypsin missed cleavage and a tolerance of 10 ppm for precursor and 1 Da for fragment ions. All datasets were processed using the workflow feature in the Proteome Discoverer software, and the resulting search data were further analyzed in the software ScaffoldQ+v.3.3.1. The scoring parameters (Xcorr and Peptide Probability) in the ScaffoldQ+ software were set to obtain a false discovery rate (FDR) of less than 1%, using the number of total spectra output from the ScaffoldQ+ software. A normalization criterion, the quantitative value, was applied to the spectral counts [37,38].

All mass spectrometric raw and msf files associated with this study are available for download via FTP from the PeptideAtlas data repository by accessing the following link: http://www.peptideatlas.org/PASS/ PASS00388.

Feature selection analyses of proteomics data

Heat map and hierarchical clustering analyses

Files containing the identified proteins and their spectral counts were used for the clustering and heat maps generation, as well as to perform the feature selection analyses. Heat maps and hierarchical clustering were constructed in the web-based chemometrics platform MetaboAnalyst 2.0 using the Pearson distance measure. For this specific analysis, protein spectral counts were previously z-score transformed.

Neighbor joining trees

In order to evaluate how similar the three classes were when considering their spectral counts distribution within the samples, the secretome dataset was analyzed using the neighbor joining (NJ) clustering method [39]. The phenetic trees were constructed from an Euclidean distance matrix using the VisPipeline software (http://vicg.icmc.usp.br/infovis2/Tools), developed at Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, USP, São Carlos, Brazil. The silhouette coefficients [40] were also calculated for the 18 secretome dataset samples (both the raw data and their NJ clustering) using the VisPipeline software. The closest to 1 the silhouette coefficient (ranging from -1 to 1), the more efficient is data clusterization. Silhouette coefficients were also calculated after feature selection, in order to check for the coefficients improvement.

Identification of candidate biomarkers

The univariate Beta-binomial model was used to test the significance of protein differential abundances expressed in spectral counts in our label-free mass spectrometry-based proteomics dataset. The Beta-binomial model was constructed using a software package implemented in R according to Pham et al. [20].

In addition, protein spectral counts were submitted to other two different approaches: the semi-multivariate Nearest Shrunken Centroids (NSC) and the multivariate Support Vector Machine-Recursive Features Elimination (SVM-RFE). The NSC and the SVM-RFE models were also performed using software packages implemented in R according to Tibshirani et al. [41] and Guyon et al. [42], respectively. For both methods, a double cross-validation procedure was applied to define the optimal feature (protein) subsets (N) from the ranked proteins lists (independently ranked by each method). In the case of the SVM-RFE model, the optimal feature subset was the smallest set that provided the minimum mean classification error, whereas for the NSC model it was the subset that minimized the classification error and maximized the sum of true class probabilities [21].

The double cross-validation procedure was developed and implemented in R based on the work of Christin et al. [21]. Both feature selection methods had their performance assessed in terms of accuracy, sensitivity and specificity, using the caret package implemented in R [43]. The output final ranked lists of candidate biomarkers that resulted from each model (N defined by $p < 0.05$, in the case of the Beta-binomial model, or by double cross-validation, in the case of NSC and SVM-RFE models) were also compared with each other using the Jaccard similarity coefficient and a Venn diagram and considered for further analyses.

To compare our results to a classical statistical method, besides the methods described above, the univariate ANOVA test was also performed in our data using the ScaffoldQ+ software, with N defined by $p < 0.05$.

The same feature selection analyses using the three methods were performed for a published proteomics dataset of prostate cancer [22] in order to validate our proposed pipeline. The output final ranked lists of candidate biomarkers that resulted from each model were also compared by a Venn diagram.

Double cross-validation

The double cross-validation (DCV) is a type of statistical validation stricter than the cross-validation (CV), as in DCV a CV is performed within another CV. A CV error is an inappropriate estimate of the prediction error of the model, since this error is not based on an independent test set, as all data – both test and training samples – are used at once. Therefore, in order to avoid overly optimistic performance estimates, a "nested" CV scheme was performed in the DCV to estimate the prediction error, in which the parameter optimization is executed in an internal loop (inner loop) and the prediction error is estimated in an external loop (outer loop) on a completely independent set of samples [21,44]

NSC (Nearest Shrunken Centroids)

The NSC double cross-validation was developed as follows:

Each inner loop (from a total of five) which is associated to an outer loop (from a total of six) calculates one model and one ranking using 90 threshold values. Many variables are discarded based on the threshold values and on the data set. A test data set associated to each inner loop is used to test each model and threshold combination. For each model the threshold that gives the minimum error with respect to the test data set and maximum probability score is recorded.

At the end of each outer loop, the maximum threshold among the five inner loops thresholds, that gives the minimum error and maximum probability overall is selected (resulting in the minimum N).

The thresholds for each outer loop are then applied to the training data set for model calculation and are challenged with the respective test data sets. The outer loop threshold is then selected by the minimum error, followed by the maximum probability and by the maximum threshold value.

The double cross-validation error is the average of the errors of the inner thresholds applied to their respective outer loops.

The NSC double cross-validation was repeated 100 times, and the final double cross-validation error was calculated as the closest value to the average of all 100 errors. N was then selected from this final double cross-validation error, followed by the maximum probability and by the maximum threshold value associated with the final error.

SVM-RFE (Support Vector Machine-Recursive Features Elimination)

In the case of SVM-RFE, the majority of calculated Ns in the inner loops were extremely small, ranging between one and four. Thus, when applying these Ns in the outer loops, models with only a few variables (approximately four) were constructed, and when using the test data set of the outer loops, errors ended up being extremely large (around 50%). Therefore, a different approach for the SVM-RFE double cross-validation was developed as follows:

Instead of using the minimum N from the inner loops of each outer loop, the mean error from the five inner loops **is** calculated for each value of N. Then, a minimum N that also minimizes e(N) is selected, where e(N) is the mean error of the inner loops when the first N proteins are considered.

The process is repeated for the six outer loops, again selecting the minimum N minimizing e(N). In general, the value of N presenting minimal error is selected over 30 training subsets.

Because the value of N is calculated by the average of the five inner loops, the ranking is also calculated based on such loops. For this calculation, the vectors W[i] of weights of each variable for each inner loop are multiplied. Thus, the ranking within an outer loop is the product vector W' = W[1]W[2]W[3]W[4]W[5].

This weight vector W' will define the ranking for each outer loop. This rank, with N selected from the corresponding inner loop, is used to construct the SVM-RFE model for a loop and to calculate its error.

The double cross-validation error is the average of the errors of the six outer loops.

The SVM-RFE double cross-validation was repeated 100 times, and the final double cross-validation error was calculated as the closest value to the average of all 100 errors. N was then selected as the value associated with this final double cross-validation error.

Estimation of protein classes

After the feature selection analyses, aiming to characterize the retrieved proteins as candidate biomarkers of each cell type, the secretome class in which each protein abundance changed the most among all classes was estimated. For that, a routine was developed and implemented in R, as further described. Proteins were associated to one of the three classes (melanoma, carcinoma or non-cancerous) using decision boundaries based on protein average spectral count values found in each class. After calculating the average values (hence, one value for each class, e.g., $x \leq y \leq z$), two boundaries were established as the mathematical intermediate point between the average values (B1=(x+y)/2 and B2=(y+z)/2). The average values were then associated to their closest boundary and classes were assigned index 1 if closest to the boundary with the lowest value, or 2 in the opposite case. In the example above, x would always be associated to B1 and z to B2; y attribution, however, would depend on how far x and z were from y. If y was associated to B1, both x and y classes were labeled with index 1; z class, otherwise, would score 2 and would be the class in which the corresponding protein abundance changed the most compared to the other two classes. If y was associated to B2, y and z would be labeled with index 2, whereas x would score 1 and would be the class in which the corresponding protein abundance is the most altered (Supplementary Table S3, "Class boundary index" column). These indexes were used to classify proteins in the dataset, considering that proteins are candidate biomarkers of the secretome classes where the average abundance values are the most altered compared to the other classes.

Bioinformatics analyses

To explore the biological significance of the variables that greatly contributed to the characterization of each tumor class, protein-protein interaction networks were constructed using the Integrated Interactome System (IIS) software [45], developed at Laboratório Nacional de Biociências, CNPEM, Campinas, Brazil, for the candidate biomarkers identified by either the Beta-

binomial, NSC and SVM-RFE models, and further estimated to be associated to the carcinoma or melanoma classes (Table E4). Enrichment analyses were performed in the networks using the IIS software for the curated pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [46]. Significantly enriched KEGG pathways ($p \leq 0.05$) for proteins of carcinoma and melanoma secretomes were assigned as clusters in the networks and different colors and sizes were attributed to proteins proportionally to their fold change compared to the non-cancerous secretome class ($-1.3 \geq FC \geq 1.3$). Zero values were replaced by one in order to calculate the fold change. The resultant networks were visualized using the Cytoscape 2.8.2 software [47].

To evaluate whether the candidate biomarkers for melanoma and carcinoma were previously described related to cancer or to some biomarker application, a biomarker analysis was performed using the Ingenuity Systems Pathway software (IPA; Ingenuity Systems, Redwood City, CA). The Ingenuity biomarker filter module analysis was performed based on the following criteria: biofluids – "all", disease – "cancer", species – "human", and biomarker application – "all". Moreover, the Human Protein Atlas [48] was used to determine whether the retrieved candidates were previously indicated as cancer biomarkers.

Immunoblotting

Proteins (5 µg) in the conditioned media from HaCaT, SCC-9, A431, A2058, SK-MEL-28 and HEK293 cell lines were separated under disulfide reducing conditions using SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto nitrocellulose membranes. The membranes were blocked in 5% dry milk in Tris-Tween buffered saline (TTBS). Membranes were then incubated overnight at 4°C with the following antibodies: anti-fibronectin (1:1000, Abcam), anti-tenascin-C (1:1000, Abcam), anti-GDF15 (1:1000, Abcam), anti-talin-1 (1:1000, Abcam), anti-EGFR (1:5000, Santa Cruz) and anti-CFB (1:1000, Abcam). Membranes were washed, incubated in horseradish peroxidase conjugated secondary antibodies and developed using enhanced chemiluminescence detection according to the manufacturer's instructions (Amersham Biosciences).

Tissue array immunohistochemistry and statistical analysis

High density tissue microarrays were obtained from Biomax (OR601a and ME1004a). The presence of Complement Factor B (CFB) and Complement Component 3 (C3) was analyzed in 10 cancer-adjacent normal tissues and in 47 primary oral squamous cell carcinomas by immunohistochemistry using the streptavidin-biotin peroxidase complex (Dako). For tenascin-C

and GDF15, 20 benign nevoid lesions (intradermal and compound nevus), 50 primary melanomas and 20 metastatic melanomas were subjected to immunohistochemical analysis with phosphatase alkaline/permanent red-based method (Dako). Protein quantification was assessed with the aid of Aperio Scanscope CS® Slide Scanner and the Pixel Count V9 algorithm software (Aperio Technologies, Vista, CA; USA). By using specific input parameters, the percentage of cytoplasm positivity was calculated and classified as weak, moderate and strong, according to its staining intensity. Each category received an intensity score, 1 to weak, 2 to moderate and 3 to strong staining. The final score of each tissue sample was calculated as the sum of the percentage of each category multiplied by their respective intensity scores as the following formula: Score = (%weak x 1) + (%moderate x 2) + (%strong x 3). Mann Whitney U and one-way ANOVA (p-value < 0.05) were used to compare protein quantities between groups.

Small interfering RNA transfection

For silencing CFB gene, $3x10^5$ skin-derived epidermoid carcinoma (A431) cells were seeded in a six-well culture plate and transfected with 50 nM small interfering RNA (siRNA) duplex (sc-44510, Santa Cruz) and Lipofectamine 2000 according to the manufacturer's instructions (Invitrogen). Random stealth siRNA duplexes coding for nonfunctional RNAs served as control (sc-37007, Santa Cruz). After 72 h of incubation at 37 °C and 5% CO_2 atmosphere, transfection success was evaluated by real-time quantitative PCR and the cells have proceeded immediately for cell migration assay as described below.

Real-time Quantitative PCR

Skin-derived epidermoid carcinoma (A431) cells had their total RNA extracted by TRIzol reagent (Invitrogen Corporation), and 2 µg of total RNA were used for retro-transcription with a First-Strand cDNA Synthesis Kit (GE Healthcare). Real-time quantitative PCR for CFB was performed using a SYBR Green PCR Master Mix (Applied Biosystems), and dissociation curves were generated to confirm the specificity of the products. The threshold cycle (CT) values of the targeted gene were normalized relative to the glyceraldehyde-3-phosphate dehydrogenase gene expression, and relative expression ratios were calculated using the $2^{-\Delta\Delta}$ Ct method. Three independent experiments were performed in triplicates. The following PCR primers were used: CFB forward 5'-TCTCAG TCATTCGCCCTTCA-3' and reverse 5'-CCTACGCTGACCTTGAT-3'; GAPDH forward 5'-GAAGGTGAAGGTCGGAGTCAAC-3' and reverse 5'-CAGAGTTAAAA GCAGCCCCTGGT-3'.

***In vitro* differentiation of macrophages derived from monocytes**

Peripheral blood mononuclear cells (PBMCs) were collected from healthy volunteers through apheresis, performed in a Trima Accel System (Cobe BCT, Denver, CO, USA), at the Hospital Alemão Oswaldo Cruz, São Paulo, Brazil, after informed consent of donors. This procedure was approved by the Research Ethics Committee of the same institution. For the enrichment of mononuclear cells, the product of apheresis was submitted to a separation with Ficoll-Paque (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) for 30 min at 900 x *g* at 18°C. Mononuclear cells were collected and centrifuged at 600 x *g* for 10 min at 18 °C with RPMI-1640 medium (Gibco, Grand Island, NY, EUA) and washed twice with RPMI-1640 medium at 300 x *g* and 200 x *g*, respectively. The mononuclear cells were then resuspended in R10 supplemented with 1% of Antibiotic-Antimycotic (Gibco, Grand Island, NY, EUA), seeded in a 24-well plate ($3x10^6$ cells/ml suspension) and incubated at 37 °C and 5% CO_2 atmosphere for 2 h. After this period, non-adherent cells were removed and cell medium was supplemented with 50 ng/ml GM-CSF (PeproTech, Rocky Hill, NJ, USA). These cells were maintained in culture for seven days for differentiation of monocytes into macrophages.

Transwell migration assay

Untreated (mock), control siRNA-transfected (scramble) and CFB siRNA-transfected skin-derived epidermoid carcinoma (A431) cells ($3x10^5$ cells) allowed to migrate for 16 h toward the lower chamber containing RPMI medium supplemented with 1% FBS. Two independent experiments were performed in triplicate.

For the co-culture assay, macrophage cells ($7.5x10^4$ cells) were added in the upper chamber, and either mock, scrambled or CFB siRNA-transfected A431 cells ($7.5x10^4$ cells) were added into the transwell plate lower chamber in 150 µl of serum-free RPMI-1640. At the end of the assay, the remaining cells at the top chamber were removed using a cotton swab, whereas the cells at the bottom of the insert filter were fixed with 10% formaldehyde for 10 min, washed with PBS and stained with 1% toluidine blue solution in 1% borax for 5 min. The dye was eluted in 1% SDS and absorbance was measured at 620 nm. Two independent experiments were performed in triplicate.

REFERENCES

1 Kulasingam, V. & Diamandis, E. P. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature clinical practice. Oncology* **5**, 588-599, doi:10.1038/ncponc1187 (2008).

2 Wu, C. C. *et al.* Candidate serological biomarkers for cancer identified from the secretomes of 23 cancer cell lines and the human protein atlas. *Molecular & cellular proteomics : MCP* **9**, 1100-1117, doi:10.1074/mcp.M900398-MCP200 (2010).

3 Chen, R., Pan, S., Brentnall, T. A. & Aebersold, R. Proteomic profiling of pancreatic cancer for biomarker discovery. *Molecular & cellular proteomics : MCP* **4**, 523-533, doi:10.1074/mcp.R500004-MCP200 (2005).

4 Shimwell, N. J. *et al.* Combined proteome and transcriptome analyses for the discovery of urinary biomarkers for urothelial carcinoma. *British journal of cancer* **108**, 1854-1861, doi:10.1038/bjc.2013.157 (2013).

5 Rifai, N., Gillette, M. A. & Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature biotechnology* **24**, 971-983, doi:10.1038/nbt1235 (2006).

6 Whiteaker, J. R. *et al.* A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nature biotechnology* **29**, 625-634, doi:10.1038/nbt.1900 (2011).

7 Makawita, S. & Diamandis, E. P. The bottleneck in the cancer biomarker pipeline and protein quantification through mass spectrometry-based approaches: current strategies for candidate verification. *Clinical chemistry* **56**, 212-222, doi:10.1373/clinchem.2009.127019 (2010).

8 Picotti, P. *et al.* High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nature methods* **7**, 43-46, doi:10.1038/nmeth.1408 (2010).

9 Picotti, P., Bodenmiller, B. & Aebersold, R. Proteomics meets the scientific method. *Nature methods* **10**, 24-27, doi:10.1038/nmeth.2291 (2013).

10 Gillette, M. A. & Carr, S. A. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nature methods* **10**, 28-34, doi:10.1038/nmeth.2309 (2013).

11 Chang, K. P. *et al.* Identification of PRDX4 and P4HA2 as metastasis-associated proteins in oral cavity squamous cell carcinoma by comparative tissue proteomics of microdissected specimens using iTRAQ technology. *Journal of proteome research* **10**, 4935-4947, doi:10.1021/pr200311p (2011).

12 de Jong, E. P. *et al.* Quantitative proteomics reveals myosin and actin as promising saliva biomarkers for distinguishing pre-malignant and malignant oral lesions. *PloS one* **5**, e11148, doi:10.1371/journal.pone.0011148 (2010).

13 Hu, S. *et al.* Salivary proteomics for oral cancer biomarker discovery. *Clinical cancer research : an official journal of the American Association for Cancer Research* **14**, 6246-6252, doi:10.1158/1078-0432.CCR-07-5037 (2008).

14 Sepiashvili, L. *et al.* Potentially novel candidate biomarkers for head and neck squamous cell carcinoma identified using an integrated cell line-based discovery strategy. *Molecular & cellular proteomics : MCP* **11**, 1404-1415, doi:10.1074/mcp.M112.020933 (2012).

15 van der Post, S. & Hansson, G. C. Membrane Protein Profiling of Human Colon Reveals Distinct Regional Differences. *Molecular & cellular proteomics : MCP*, doi:10.1074/mcp.M114.040204 (2014).

16 Simabuco, F. M. *et al.* ADAM17 mediates OSCC development in an orthotopic murine model. *Molecular cancer* **13**, 24, doi:10.1186/1476-4598-13-24 (2014).

17 Liu, N. Q. *et al.* Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. *Journal of mammary gland biology and neoplasia* **17**, 155-164, doi:10.1007/s10911-012-9252-6 (2012).

18 Granato, D. C. *et al.* Integrated proteomics identified up-regulated focal adhesion-mediated proteins in human squamous cell carcinoma in an orthotopic murine model. *PloS one* **9**, e98208, doi:10.1371/journal.pone.0098208 (2014).

19 Kulasingam, V. & Diamandis, E. P. Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets. *Molecular & cellular proteomics : MCP* **6**, 1997-2011, doi:10.1074/mcp.M600465-MCP200 (2007).

20 Pham, T. V., Piersma, S. R., Warmoes, M. & Jimenez, C. R. On the beta-binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. *Bioinformatics* **26**, 363-369, doi:10.1093/bioinformatics/btp677 (2010).

21 Christin, C. *et al.* A critical assessment of feature selection methods for biomarker discovery in clinical proteomics. *Molecular & cellular proteomics : MCP* **12**, 263-276, doi:10.1074/mcp.M112.022566 (2013).

22 Kim, Y. *et al.* Identification of differentially expressed proteins in direct expressed prostatic secretions of men with organ-confined versus extracapsular prostate cancer. *Molecular & cellular proteomics : MCP* **11**, 1870-1884, doi:10.1074/mcp.M112.017889 (2012).

23 Rutkowski, M. J., Sughrue, M. E., Kane, A. J., Mills, S. A. & Parsa, A. T. Cancer and the complement cascade. *Molecular cancer research : MCR* **8**, 1453-1465, doi:10.1158/1541-7786.MCR-10-0225 (2010).

24 Cho, M. S. *et al.* Autocrine effects of tumor-derived complement. *Cell reports* **6**, 1085-1095, doi:10.1016/j.celrep.2014.02.014 (2014).

25 Bensimon, A., Heck, A. J. & Aebersold, R. Mass spectrometry-based proteomics and network biology. *Annual review of biochemistry* **81**, 379-405, doi:10.1146/annurev-biochem-072909-100424 (2012).

26 Bonne, N. J. & Wong, D. T. Salivary biomarker development using genomic, proteomic and metabolomic approaches. *Genome medicine* **4**, 82, doi:10.1186/gm383 (2012).

27 Leemans, C. R., Braakhuis, B. J. & Brakenhoff, R. H. The molecular biology of head and neck cancer. *Nature reviews. Cancer* **11**, 9-22, doi:10.1038/nrc2982 (2011).

28 Argiris, A., Karamouzis, M. V., Raben, D. & Ferris, R. L. Head and neck cancer. *Lancet* **371**, 1695-1709, doi:10.1016/S0140-6736(08)60728-X (2008).

29 da Silva, S. D. *et al.* Advances and applications of oral cancer basic research. *Oral oncology* **47**, 783-791, doi:10.1016/j.oraloncology.2011.07.004 (2011).

30 Macor, P. & Tedesco, F. Complement as effector system in cancer immunotherapy. *Immunology letters* **111**, 6-13, doi:10.1016/j.imlet.2007.04.014 (2007).

31 Bjorge, L. *et al.* Ascitic complement system in ovarian cancer. *British journal of cancer* **92**, 895-905, doi:10.1038/sj.bjc.6602334 (2005).

32 Kim, D. Y., Martin, C. B., Lee, S. N. & Martin, B. K. Expression of complement protein C5a in a murine mammary cancer model: tumor regression by interference with the cell cycle. *Cancer immunology, immunotherapy : CII* **54**, 1026-1037, doi:10.1007/s00262-005-0672-7 (2005).

33 Gollapalli, K. *et al.* Investigation of serum proteome alterations in human glioblastoma multiforme. *Proteomics* **12**, 2378-2390, doi:10.1002/pmic.201200002 (2012).

34 Rutkowski, M. J. *et al.* The complement cascade as a mediator of tissue growth and regeneration. *Inflammation research : official journal of the European Histamine Research Society ... [et al.]* **59**, 897-905, doi:10.1007/s00011-010-0220-6 (2010).

35 Markiewski, M. M. *et al.* Modulation of the antitumor immune response by complement. *Nature immunology* **9**, 1225-1235, doi:10.1038/ni.1655 (2008).

36 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

37 Aragao, A. Z. *et al.* Novel processed form of syndecan-1 shed from SCC-9 cells plays a role in cell migration. *PloS one* **7**, e43521, doi:10.1371/journal.pone.0043521 (2012).

38 Didangelos, A. *et al.* Extracellular matrix composition and remodeling in human abdominal aortic aneurysms: a proteomics approach. *Molecular & cellular proteomics : MCP* **10**, M111 008128, doi:10.1074/mcp.M111.008128 (2011).

39 Paiva, J. G., Florian-Cruz, L., Pedrini, H., Telles, G. P. & Minghim, R. Improved similarity trees and their application to visual data classification. *IEEE transactions on visualization and computer graphics* **17**, 2459-2468, doi:10.1109/TVCG.2011.212 (2011).

40 Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**, 53-65 (1987).

41 Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6567-6572, doi:10.1073/pnas.082099299 (2002).

42 Guyon, I. W., J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Machine learning* **46**, 389-422, doi:10.1023/A:1012487302797 (2002).

43 Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **26**, 1-26 (2008).

44 Smit, S., Hoefsloot, H. C. & Smilde, A. K. Statistical data processing in clinical proteomics. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **866**, 77-88, doi:10.1016/j.jchromb.2007.10.042 (2008).

45 Carazzolle, M. F. *et al.* IIS--Integrated Interactome System: a web-based platform for the annotation, analysis and visualization of protein-metabolite-gene-drug interactions by integrating a variety of data sources and tools. *PloS one* **9**, e100385, doi:10.1371/journal.pone.0100385 (2014).

46 Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic acids research* **36**, D480-484, doi:10.1093/nar/gkm882 (2008).

47 Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431-432, doi:10.1093/bioinformatics/btq675 (2011).

48 Ponten, F., Schwenk, J. M., Asplund, A. & Edqvist, P. H. The Human Protein Atlas as a proteomic resource for biomarker discovery. *Journal of internal medicine* **270**, 428-446, doi:10.1111/j.1365-2796.2011.02427.x (2011).

ACKNOWLEDGEMENTS

This work was supported by FAPESP Grants: 2009/54067-3, 2010/19278-0, 2011/22421-2 and 2009/53839-2 and CNPq Grants: 470567/2009-0, 470549/2011-4, 301702/2011-0 and 470268/2013-1.

CAPÍTULO 2

A targeted proteomic strategy for the verification of oral cancer candidate biomarkers in human saliva

Rebeca Kawahara, James Bollinger, César Rivera, Ana Carolina P. Ribeiro, Thaís Bianca Brandão, Adriana F. Paes Leme, Michael MacCoss

ABSTRACT

Head and neck cancers, including oral squamous cell carcinoma (OSCC), are the sixth most common malignancy in the world and are characterized by poor prognosis and a low survival rate. Saliva is oral fluid with intimate contact with OSCC. Besides non-invasive, simple, and rapid to collect, saliva is a potential source of biomarkers. In this study, we build an SRM assay that targets fourteen OSCC candidate biomarker proteins, which were evaluated in a set of clinically-derived saliva samples. Using Skyline software package, we demonstrated a statistically significant higher abundance of the C1R, LCN2, SLPI, FAM49B, TAGLN2, CFB, C3, C4B, LGR1, SERPINA1 candidate biomarkers in the saliva of OSCC patients. Furthermore, our study also demonstrated that CFB, C3, C4B, SERPINA1 and LRG1 are associated with the risk of developing OSCC. Overall, this study successfully used targeted proteomics to measure in saliva a panel of biomarker candidates for OSCC.

Keywords: Selected reaction monitoring, saliva, oral cancer, skyline

INTRODUCTION

Head and neck cancers are the sixth most common malignant tumors worldwide [1]. Oral cancer is the most frequent subtype among these with oral squamous cell carcinoma (OSCC) annually affecting over 300,000 people worldwide [2]. Despite advancements in oral cancer prevention and multimodality treatments, the 5-year survival rate of OSCC patients is less than 50% and the prognosis of advanced cases has not changed considerably over the past 20 years [3,4]. Currently, oral cancer diagnosis depends on a thorough oral examination and the most important prognostic factor for OSCC patients is still the clinical stage of disease (TNM stage) at the time of intervention [2,5].

Scientists are continually searching for novel OSCC biomarker panels. Saliva is a great choice matrix for this task as it is an easy-to-obtain body fluid that can be collected noninvasively [6]. Previous studies have identified a large number of putative biomarker candidates in human saliva [7-11]. Unfortunately, none of these candidates have passed through the development and validation stage to gain utility in clinical practice [12]. The bottleneck in this pipeline is attributed, in part to a lack of robust methods to validate a majority of the candidates generated by discovery-phase studies [13,14]. ELISA is currently the most commonly used analytical platform for the validation of biomarker candidates derived from bodily fluids. However, antibody availability restricts analysis to a small subset of potential candidates and the cost and low throughput of these methods ultimately renders a large majority of candidates untestable [15].

Selected reaction monitoring (SRM) tandem mass spectrometry has emerged as a technique capable of precisely and quantitatively measuring peptides in complex biological samples [16]. Due to its high specificity, sensitivity, precision, and multiplexing capabilities, SRM provides an attractive alternative for the validation of candidate biomarkers [17]. However, there are many challenges still associated with the development of a quantitative SRM assay. Among these is the challenge of identifying the optimal set of proteotypic peptides for each protein target. Proteotypic is defined here as a peptide that is unique to a particular gene product, devoid of single nucleotide polymorphisms, readily detectable by the mass spectrometers, and has salient structural features with characteristic MS/MS fragmentation patterns. Furthermore, there are computational challenges related to assay development, analytical validation, data management, peak integration, data quality assessment and biostatistical analysis. Collectively, these challenges have hindered widespread use of the SRM methodology.

The current work presents a successful pipeline for the measurement of a panel of candidate biomarkers for oral cancer in human saliva via an SRM assay. Fourteen candidate biomarkers derived from a previous discovery-based proteomics study (Kawahara et al, 2014, under review) were measured in a set of 30 independent clinically-derived human saliva samples. For the selection of proteotypic peptides, we used an empirical approach that is based on the use of analytical standards (*in vitro*-synthetized fusion proteins, recombinant proteins or native proteins isolated from plasma). For each protein standard, we considered every doubly-charged monoisotopic tryptic peptide from 7 to 23 amino acids in length. For each peptide, we considered every singly-charged monoisotopic y-ion from y_n to y_{n-3} [18]. Skyline, a vendor-neutral tool [19], was used extensively for MRM assay development, method export, peak detection and peak integration as well as all subsequent statistical analysis. Overall, this study used targeted proteomics to study a panel of candidate biomarkers in human saliva collected from healthy patients and patients with OSCC.

EXPERIMENTAL SECTION

Proteins and peptides

Human complement C4 (Cat# P151-0) and C3c (Cat# pro-555) derived from human plasma were purchased from BBI Solutions and Prospec, respectively. Recombinant SAA1 (Cat# cyt-657) and SERPINA1 (Cat# pro-529) were purchased from Prospec and recombinant CFB (Cat# C329) was purchased from Novo Protein.

Heavy [$^{13}C_6{}^{15}N_2$]-lysine–labeled LLLEYLEEK and IEAIPQIDK peptides from *Schistosoma japonicum (SJ)* GST (Thermo Scientific) and VSEADSSNADWVTK and YGLVTYATYPK peptides from CFB (SynPeptide) were obtained to use as internal standards. Similarly, unlabeled LLLEYLEEK and IEAIPQIDK peptides were obtained to use as calibration standards. A mixture containing 15 heavy isotope-labeled peptides was used as retention time standards (Pierce® Retention Time Calibration Mixture, Thermo), quality control and normalization.

Recombinant proteins production using *in vitro* expression system

The recombinant proteins, native proteins isolated from human plasma and *in vitro*-synthesized proteins were used to empirically derive optimal proteotypic peptides and fragment ions as previously described [18] with some modifications.

For the nine *in-vitro* synthetized proteins (C1R, TINAG, SLPI, SERPINE1, LRG1, LCN2, TAGLN2, ANXA1 and FAM49B), clones containing an in-frame fused C-terminal SJ GST tag were obtained from the Arizona State University Biodesign Institute plasmid repository. Each bacterial plasmid was grown overnight in 5 ml of Luria broth supplemented with 100 μgml−1 ampicillin (LB-amp). Plasmid DNA was extracted using the manufacture mini-prep protocol with the exception of an additional wash with PE buffer (Qiagen). All plasmid stocks were Sanger sequenced (Genewiz) using an M13 priming site upstream of the T7 promoter to confirm the identity of the insert and to check for contamination of the plasmid stocks. For *in vitro* protein synthesis, 1 μg of mini-prepped plasmid DNA was used directly in the 1-Step Human Coupled IVT Kit (Thermo Scientific) according to manufacturer's protocols with a few minor modifications. Briefly, a mix containing HeLa lysate, Accessory Proteins, and a proprietary Reaction Mix was added to 1 μg of plasmid DNA in a 25-μl transcription reaction supplemented with 0.3 μl RNase inhibitors (Thermo Scientific) and the reaction was incubated at 30 °C for 3 h/200 rpm.

Recombinant Protein Purification and digestion

To enrich GST-fusion protein, we used 200 µl of glutathione sepharose 4B beads (GE), washed 3 times with 1 ml 1× Dulbecco's phosphate-buffered saline (DPBS; Gibco) and resuspended in 1.25 ml of 1× DPBS. A 125-µl aliquot of the washed bead slurry was added to each 1.5 ml-tubes containing the HeLa lysate reaction such that each tube received 20 µl of packed beads. Completed translation reactions were added to the beads and the bead-protein mixture was rocked end-over-end for 16 h at 4 °C.

The bead-protein mixture was washed twice with 150 µl of wash buffer (1× DPBS supplemented with 863 mM NaCl) and twice with 150 µl of 50 mM ammonium bicarbonate (pH 7.8) each. After the last wash, the beads were resuspended in 50 µl Elution Buffer (0.05% PPS silent surfactant (Protein Discovery), 5 fmol/µl heavy isotope–labeled SJ GST peptides LLLEYLEEK and IEAIPQIDK (Thermo Scientific), and 50 mM ammonium bicarbonate (pH 7.8)). Ten microliters of each enriched protein sample was added to 4 µl 4× LDS buffer (Invitrogen) and saved for immunoblotting.

Bead bound protein samples (25 µl) were diluted 1:1 in 0.05% PPS silent surfactant/50 mM ammonium bicarbonate (pH 7.8), boiled for 5 min at 95 °C, reduced with 5 mM dithiothreitol for 30 min at 60 °C, and alkylated with 15 mM iodoacetamide for 30 min at 25 °C in the dark. Alkylation reactions were quenched by bringing the DTT concentration up to 15 mM and proteins were subsequently digested with 1 µg trypsin (Promega) at 37 °C for 2 h while shaking at 1200 rpm. Digestion progress was terminated via the addition of formic acid to 1.6% and samples were incubated for 1 h at room temperature to facilitate hydrolysis of the PPS silent surfactant.

Immunoblotting of *in-vitro* synthetized proteins

Undigested protein extracts from each of the *in vitro* translation reactions were boiled in 1× LDS buffer (Invitrogen), resolved on a 4–12% Bis-Tris denaturing/reducing SDS-PAGE (Invitrogen) and transferred onto a nitrocellulose membrane (Bio-Rad) for immunoblotting. Membranes were then blocked with 5% non-fat dry milk (Safeway) in PBS-tween buffer and probed with a primary antibody targeting SJ GST (GE 27-4577-01) as an epitope. Incubations with the primary antibody were done at 4 °C overnight using a 1:1000 dilution. Secondary incubations were performed in 5% non-fat dry milk in PBS-tween using 1:10000 diluted peroxidase-conjugated rabbit anti-goat IgG (H+L) (Pierce). Membranes were visualized using an ECL plus western blotting kit (Amersham) and detected with radiographic film (Thermo).

SRM assay development and refinement

All samples were analyzed with a TSQ-Vantage triple-quadrupole instrument (Thermo Scientific) using either a nanoLC separation system (Eksigent) or a nanoACQUITY UPLC (Waters). A 3-µl aliquot of each sample was separated on a self-packed 15-cm x 75 µm inner diameter column (Polymicro Technologies) with ReproSil-Pur C18-AQ (3 µm, particle size and 120 Å pore size). Peptides were separated using a 7-min gradient from 5% acetonitrile in 0.1% formic acid to 40% acetonitrile in 0.1% formic acid at a flow of 750 nl/min. The gradient was followed by 2 min linear gradient from 40% acetonitrile in 0.1% formic acid to 68% acetonitrile in 0.1 % formic acid also at 750 nl/min and washed for 3 min at 95% acetonitrile in 0.1% formic acid and finally re-equilibrated to initial conditions at 750 nl/min in 5% acetonitrile in 0.1% formic acid for 6 min. Ions were isolated in both Q1 and Q3 using 0.7 FWHM resolution. Peptide fragmentation was performed using Argon at 1.5 mTorr in Q2 using calculated peptide specific collision energies as described previously [20]. Data were acquired using a dwell time of 10 ms.

For each analytical standard, we initially monitored all monoisotopic, +2 charge states for every fully tryptic peptide ranging from 7 to 23 amino acids in length. In addition, we monitored the heavy and light forms of the SJ GST peptides LLLEYLEEK and IEAIPQIDK. For each precursor, the y_3 to y_{n-1} fragment ions were considered in their monoisotopic, +1 charge state. All cysteines were monitored as carbamidomethyl cysteines.

After mass spectrometry analysis, the resulted raw files were imported into Skyline (http://skyline.maccosslab.org) [19]. Chromatographic data from each peptide were manually analyzed to determine the quality of the peptide signal and peak shape and the four best transitions for each peptide were selected based on the intensity and fragment m/z referentially higher than precursor m/z.

The samples were incubated in the 4 ^{o}C auto-sampler for 48 h to evaluate the peptide stability. The selected transitions refined in the first step were then submitted to another mass spectrometry run and the data were similarly analyzed in Skyline. The peptides that demonstrated more than 50% decrease in the peak area in integrated signal intensity after 48 h were deemed insufficiently unstable and excluded from further consideration.

Creating a chromatogram Library using PANORAMA

All the results from the method development using the recombinant, native or *in vitro*-synthetized proteins were uploaded to PANORAMA daily (https://daily.panoramaweb.org) and re-imported back into Skyline as a chromatogram library. Chromatogram libraries provide a way to

store targeted assays that have been curated in Skyline and reuse them in the future for measuring proteins and peptides in other samples. Chromatogram libraries capture physiochemical and experimental properties such as the relative ion abundances, fragmentation patterns, measured chromatograms, and iRT retention time information, and optimal instrument parameters such as collision energy.

Clinical Samples

The saliva was collected from healthy individuals (n=8), OSCC patients who underwent surgical resection (named as no lesion, n=9), and OSCC patients with active oral malignant lesions (named as lesion, n=13). All enrolled patients and volunteers signed a formulary stating their awareness and consent for the study as it was approved by the ethics review board of Instituto do Câncer do Estado de São Paulo, Octavio Frias de Oliveira, ICESP, São Paulo, SP, Brazil and Plataforma Brasil. Individuals were initially asked to rinse their mouth with 5 ml of drinking water. Saliva was then harvested into a clean glass receptacle, aliquoted in 2 ml eppendorf tubes and immediately stored at -80 °C. Clinic pathological data, such as sex, anatomical site of the primary tumor, TNM classification of malignant tumors (TNM), WHO differentiation degree, and treatment, were collected from patient's charts and listed in Table 2.

Human saliva sample preparation and digestion

Protein extraction was performed by homogenizing the 100 µl of whole saliva with 100 µl of a solution containing 100 mM Tris-HCl, pH 7.5, 8 M urea, 2 M thiourea containing Protease Inhibitor Cocktail Complete Mini Tablets (Roche, Auckland New Zealand), 5 mM EDTA, 1 mM PMSF and 1 mM DTT. Samples were sonicated for 10 min and then centrifuged at 10,000 x *g* for 5 min. Protein concentration was determined for each sample with a Bradford assay (Bio-Rad, Hercules, CA, USA) and a 10 µg aliquot of each was submitted to reduction (5 mM dithiothreitol, 25 min at 56°C), followed by alkylation (14 mM iodoacetamide, 30 min at room temperature in the dark) and digestion with trypsin (1:50, w/w). The reaction was stopped with 1% TFA and desalted with stage-tips C18 resin. The samples were dried in a vacuum concentrator and reconstituted in 0.1% formic acid.

Method refinement in the human saliva matrix

The saliva samples were ressuspended to a final concentration of 0.33 µg/µl and three different samples from OSCC patients were pooled. The refined transitions for the 14 proteins

(C1R, TINAG, SLPI, SERPINE1, LRG1, LCN2, TAGLN2, ANXA1, FAM49B, CFB, C4B, C3, SERPINA1 and SAA1) were exported from Skyline as transition list and imported into Xcalibur as multiple methods. Multiple runs were performed as described before with the pool of human saliva, using a gradient of 30 min from 5% to 40% acetonitrile in 0.1% formic acid at a flow of 250 nl/min. The gradient was followed by steeper 5-min linear gradient from 40% to 60% acetonitrile in 0.1% formic acid also at 250 nl/min. The column was then washed for 5 min at 95% acetonitrile in 0.1% formic acid and finally re-equilibrated to initial conditions at 500 nl/min. Peptide fragmentation was performed at 1.5 mTorr in Q2. Data were acquired using a scan width of 0.002 m/z and a dwell time of 10 ms.

Standard curve of synthetic peptides

A dilution curve of the unlabeled SJ GST peptides, LLLEYLEEK and IEAIPQIDK, was spiked with 5 fmol/µl of each heavy isotope–labeled LLLEYLEEK and IEAIPQIDK peptides. The calibration points used were 0, 0.1, 0.25, 5, 12.5 and 40 fmol/µl each of the light LLLEYLEEK and IEAIPQIDK peptides. Each calibration point was diluted into 25 fmol/µl Bovine Quality Control (QC) standard mix (Michrome) in 2% acetonitrile and 0.1% formic acid in water. Peptide standards were measured in a single replicate and a linear regression of the data points was used to calibrate the SJ GST peptide light-to-heavy ratio of all other samples.

Similarly, an eight-point calibration curve was prepared using the aforementioned CFB recombinant protein digest in a theoretical concentration range of 0, 0.1, 1, 6.25, 12.5, 25, 50 and 100 fmol/µl with heavy isotope-labeled YGLVTYATYPK (4 fmol/µl) and VSEADSSNADWVTK (18 fmol/µl) peptides from CFB, heavy isotope-labeled LLLEYLEEK and IEAIPQIDK (5 fmol/µl each) peptides from SJ GST as well as the Thermo Peptide Retention Time standards (6.7 fmol/µl). The curve was measured in triplicate and the results were analyzed using Skyline External Tool Quasar Quantitative Statistics to determine limits of detection and quantification, calculate mean and coefficient of variation for all transitions of each CFB heavy peptides.

SRM in the human saliva

An aliquot containing 1.2 µg of digested proteins from saliva spiked with 18 fmol/µl of VSEADSSNADWVT[$^{13}C_6$]K from CFB, 4 fmol/µl of YGLVTYATYP[$^{13}C_6$]K, **5 fmol/µl** of SJ GST heavy IEAIPQIDK peptide and 6.7 fmol/µl of Pearce Retention Time Calibration Mixture. Samples were analyzed on a TSQ-Vantage triple-quadrupole instrument (Thermo Scientific) as previously described. Briefly, the peptides were separated on a 20-cm column in a gradient of 30

min from 5 to 40% acetonitrile in 0.1% formic acid at a flow of 250 nl/min. The gradient was followed by steeper 5-min linear gradient from 40% to 60 % acetonitrile in 0.1 % formic acid also at 250 nl/min. The column was then washed for 5 min at 95 % acetonitrile in 0.1 % formic acid and finally re-equilibrated to initial conditions at 500 nl/min. Two technical replicates from each sample were performed. Peptide fragmentation was performed at 1.5 mTorr in Q2. Data was acquired using a scan width of 0.002 m/z and a dwell time of 10 ms.

SRM data analysis

Data sets were imported into Skyline, and peaks were automatically integrated and manually inspected.

The confidence of the peak identity was assessed by 1) calculating a dot-product correlation between the peak intensities of the transitions for the peptide of interest and the library spectrum for the same peptide and 2) calculating the retention time prediction obtained from the regression of RT by iRT values of five standards peptides (SSAAPPPPPR, IGDYAGIK, SAAGAFGPELSR, SFANQPLEVVYSK, LTILEELR) out of the 15 Pearce Retention Time Calibration Mixture peptides spiked and monitored in all samples. The following equation was used to predict RTs for the targeted peptides: $RT = 0.09 \times iRT + 18.51$ ($R= 0.9998$).

Statistical analysis was performed using MSstats [21-23], an external tool implemented directly in Skyline [24]. Samples were annotated into three respective conditions (Control, No lesion and Lesion) and individual biological replicates. The number of samples of each condition was: 8 for control, 9 for no lesion and 13 for lesion. Each samples was analyzed in two technical replicates. For more details see Supplemental Material and Methods.

RESULTS AND DISCUSSION

In this work we demonstrated the use of human saliva collected from healthy patients or patients with oral squamous cell carcinoma to measure candidate biomarkers using selected reaction monitoring (SRM) assay. The candidate biomarkers were selected from a previous discovery-based proteomic study that used support vector machine-recursive features elimination (SVM-RFE), nearest shrunken centroids (NSC) and beta-binomial model tested in a well-controlled data obtained from the secretome of three classes of carcinoma (including OSCC originated from a human tongue) melanoma and non-cancerous cell lines. Among 2,574 proteins identified in the secretome, the statistical models indicated a panel of 137 candidate biomarkers for carcinoma and 271 for melanoma, which were differentially abundant between the tumor classes. Among the candidate biomarkers for carcinoma, fourteen proteins were selected based on the availability of protein analytical standards for method development as well as cellular localization in the extracellular milieu (Kawahara et al., 2015 under review) and used for verification in human saliva using targeted proteomics.

Selection of proteotypic peptides

Our first task was to select an optimal set of proteotypic peptides of each protein to develop an SRM assay for human saliva. We utilized a previously described empirical approach for this task that is based on the use of analytical standards [18]. In the current report, we utilized *in-vitro* synthetized proteins, commercial recombinant proteins and commercial native proteins that were isolated from plasma.

Figure 1 illustrates our workflow for SRM method development. Using a fast gradient (7 minutes) at a relative high flow rate for nano-flow UPLC (750 nl/min), we were able to expedite method development during the arduous step of monitoring every transition for every tryptic peptide. In this step, 2,908 transitions, corresponding to 391 peptides were monitored using multiples methods, with 50 maximum transitions per sample injection, resulting in approximately 58 runs.

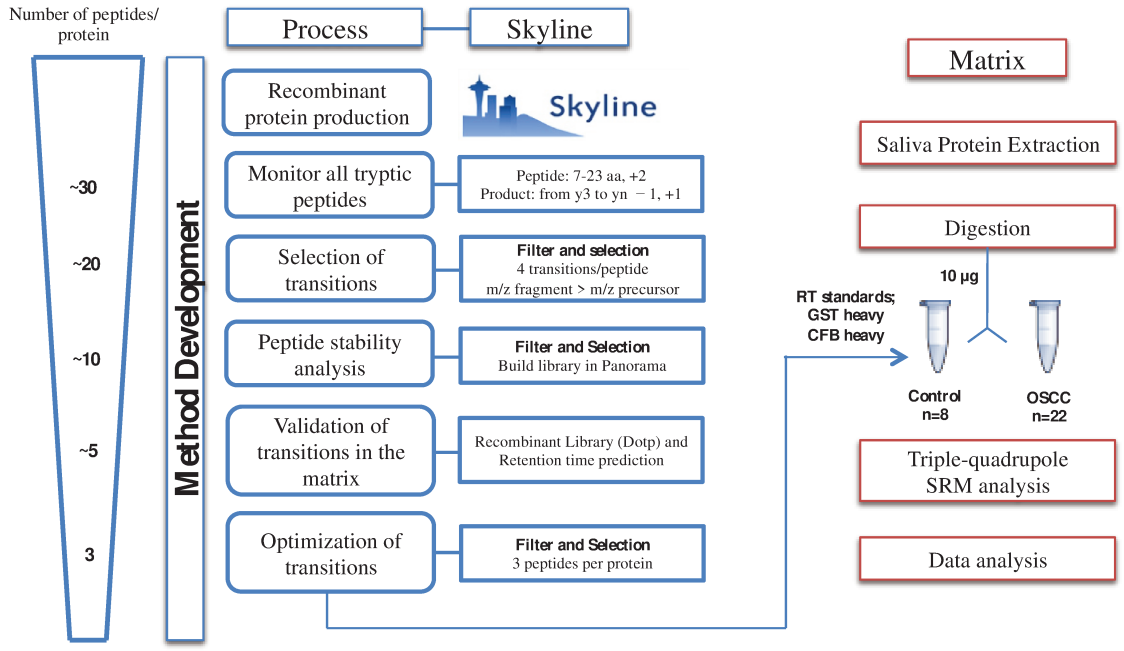

Figure 1 Experimental workflow of SRM assay development using empirical measurement for selection of proteotypic peptides.

For quantification of the *in-vitro* synthetized proteins, we used the ratio of the SJ GST peptides from the C-terminal fusion protein to the spiked-in heavy isotope-labeled peptides. Observed ratios were fit into a calibration curve of the light and heavy isotope–labeled peptides (Supplemental Figure 1A). Each of the tested proteins produced at least 500 fmoles per 25 µl of the in vitro reaction (Supplemental Figure 1A and B).

For each protein, we selected peptides and their respective fragment ions using Skyline. Peptides with non-existent or ambiguous chromatograms were excluded. Three to four transitions were chosen from the remaining peptides based on SRM intensity and fragment m/z higher than precursor m/z, preferentially. Stability of the peptides was evaluated after 48 h in the auto-sampler and the peptides that showed 50% decrease in the peak area were excluded. The final transitions obtaining after monitoring the standard proteins in these two steps were 650 that correspond to 169 peptides.

Obtained SRM results were exported to Panorama for the purpose of building a chromatogram library. This library was used as a reference to assign confidence in the identification of targeted peptides in the human saliva matrix. Skyline utilizes said library to calculate the dot-product (dotp) correlation for the ratio of the observed SRM peak intensities of a peptide in a specific biological matrix versus the library spectrum.

The filtered set of peptides and transitions were then monitored in a pool of digested saliva matrix. Assay refinement at this stage was based on those peptides and transitions that demonstrated the optimal signal-to-noise ratios. Ultimately, we selected three peptides per protein which showed good signal and good dotp (>0.8).

Analytical assay performance using CFB heavy peptides and global standards

The refinement process for optimal SRM peptides also guided our selection of peptides to be synthetized in a heavy isotope labeled analog. Figure 2A shows the peak area comparison of all tryptic peptides from 7 to 23 amino acids in length measured using the CFB recombinant protein. The first, fourth and fifth peptides, indicated by the colored arrows, were the peptides observed in highest intensity within the saliva matrix. Heavy isotope analogs for the peptides VSEADSSNADWVTCK and YGLVTYATYPCK (red arrows) were purchased from a commercial source. The relative signal intensities of the fragment ions generated from endogenous peptides in saliva were compared to those from the heavy labeled peptides in Figure 2B and C.

Calibration curves were constructed by dilution of digested recombinant CFB in known concentration range from 0.1 to 100 fmol/µl into constant amount of CFB heavy peptides VSEADSSNADWVTCK and YGLVTYATYPCK were added at a final concentration of 18 and 4 fmol/µl, respectively, together with 6.7 fmol/µl of Pearce Retention Time Calibration Mixture and 5 fmol/µl of SJ GST heavy peptides.

The limit of detection (LOD) and the limit of quantification (LOQ) for both CFB peptides normalized by the respective labeled peptide were calculated using Quasar Quantitative Statistics [24], an external tool implementable in Skyline. The complete data used for calculating the LOD, LOQ and CVs are presented in Supplemental Tables 1 and 2. The calculated LOD and LOQ for VSEADSSNADWVTK peptide were 0.90 fmol/µl and 2.7 fmol/µl, respectively. The calculated LOD and LOQ for YGLVTYATYPK were 1.8 fmol/µl and 5.5 fmol/µl, respectively.

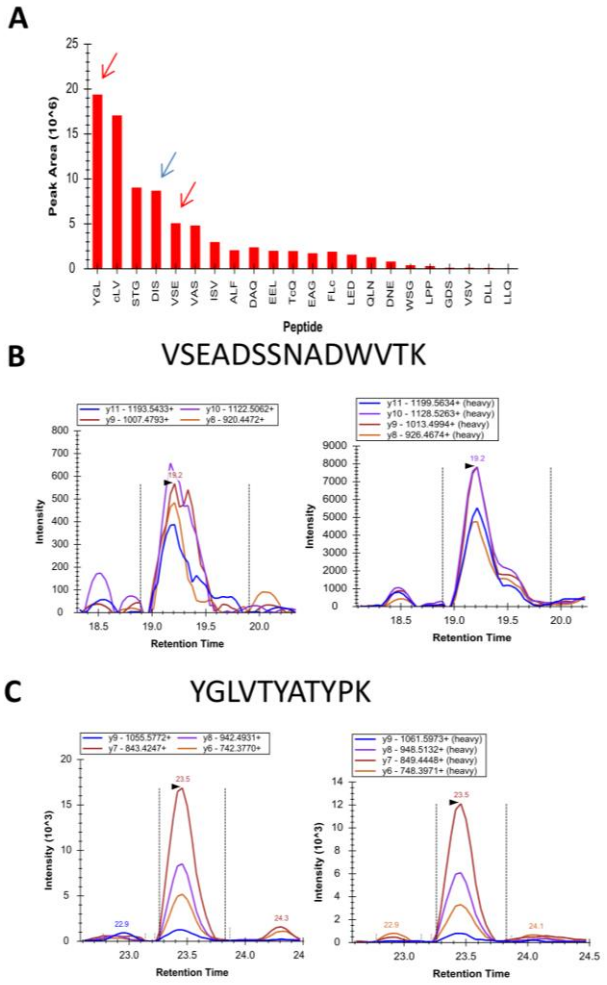

Figure 2 Analytical assay performance using CFB heavy peptides and global standards. A) Recombinant CFB tryptic peptide intensities from 7 to 23 amino acids in length. The arrows show the proteotypic peptides measured in the saliva. The red arrows show peptides synthetized in the labeled form. B) Resulting SRM mass spectra (left, endogenous specific transitions for the different peptides corresponding to the protein used for quantification and right, corresponding heavy peptides) of VSEADSSNADWVTK and C) YGLVTYATYPK.

Using the same calibration curve of recombinant CFB, we compared the performance of our SRM method when the peptide VSEADSSNADWVTK and YGLVTYATYPK were normalized by the respective heavy peptides or by the global standards. Linear regression for the two peptide sequences was higher than 0.99 (R^2 values) within the investigated concentration range when normalizing by either the heavy peptides or by the global standards (Supplemental Figure 2).

The use of internal reference peptides from a targeted protein of interest has been previously described [25]. Our study demonstrates that external reference peptides can be used as an alternative method to internal reference peptides for the normalization of signal intensities. Besides,

assay performance was compared with stable isotope labeling strategies, considering linearity and reproducibility. It is important to highlight that the use of retention time standards, as external reference peptides, is easy to implement, cost effective, and retains the potential to normalize for ion suppression across multiple different points in the chromatogram.

SRM analysis of OSCC candidate biomarkers in human saliva

For the optimized SRM assay, three peptides were monitored per protein with three to four transitions monitored per peptide. In total, 14 proteins, 48 peptides, 188 transitions (cycle time approximately 2s in a non-scheduled method) were considered and monitored in each 65-min run. Among the 48 peptides, six peptides were used as global standards, including five peptides from the Retention Time Calibration Mixture (SSAAPPPPPR, IGDYAGIK, SAAGAFGPELSR, SFANQPLEVVYSK, LTILEELR) as well as the SJ GST heavy isotope-labeled peptide IEAIPQIDK. The global standards were chosen to represent different points in the chromatography run. They were used to assess the quality of data over multiple replicates and sample injections. We used several features in Skyline to view the SRM data and evaluate the reproducibility and quality of data based on the retention time and peak areas of the global standards and labeled peptides. Using these peptides as global standards, we observed CVs ranging from 7.5-25% (Table 1, Supplemental Table SM4).

Table 1 Peptides used for normalization and the respective CV calculated for all runs.

Protein Name	Peptide	CV%
CFB (heavy)	YGLVTYATYPK	15.5%
CFB (heavy)	VSEADSSNADWVTK	13.8%
GST	IEAIPQIDK	19.50%
Pierce RT	SSAAPPPPPR	7.50%
Pierce RT	IGDYAGIK	13.40%
Pierce RT	SAAGAFGPELSR	21.00%
Pierce RT	SFANQPLEVVYSK	19.80%
Pierce RT	LTILEELR	25.00%

All targeted peptides were manually analyzed to ensure correct peak picking. For that, we used the dot-product (dotp) correlation as well as the retention time prediction based on a linear regression of the retention times from the five peptides from the Retention Time Calibration Mixture.

We also used the advanced refinement tool in Skyline to automatically excluded peptides with dotp scores lower than 0.8. With the exception of SAA1, SLPI and TINAG, at least

two peptides with dotp scores > 0.8 were detected for each protein that was within the retention time prediction window (2 min) in all samples and in both technical replicates. For SAA1 and TINAG no peptides were detected above the minimum dotp threshold of 0.8 and were thus excluded from statistical analysis. For the SLP1 protein, only the peptide CLDPVDTPNPTR remained after the advanced refinement. A representative result showing one of the three peptides selected for each protein and the relative contribution of each fragment ion both in the library as well in the saliva matrix is displayed in Figure 3. The final complete list of targeted proteins, peptides and transitions used in this study is shown in Supplemental Table 3. Furthermore, a detailed report showing the dotp and retention time results for all peptides is available in Supplemental Table 4.

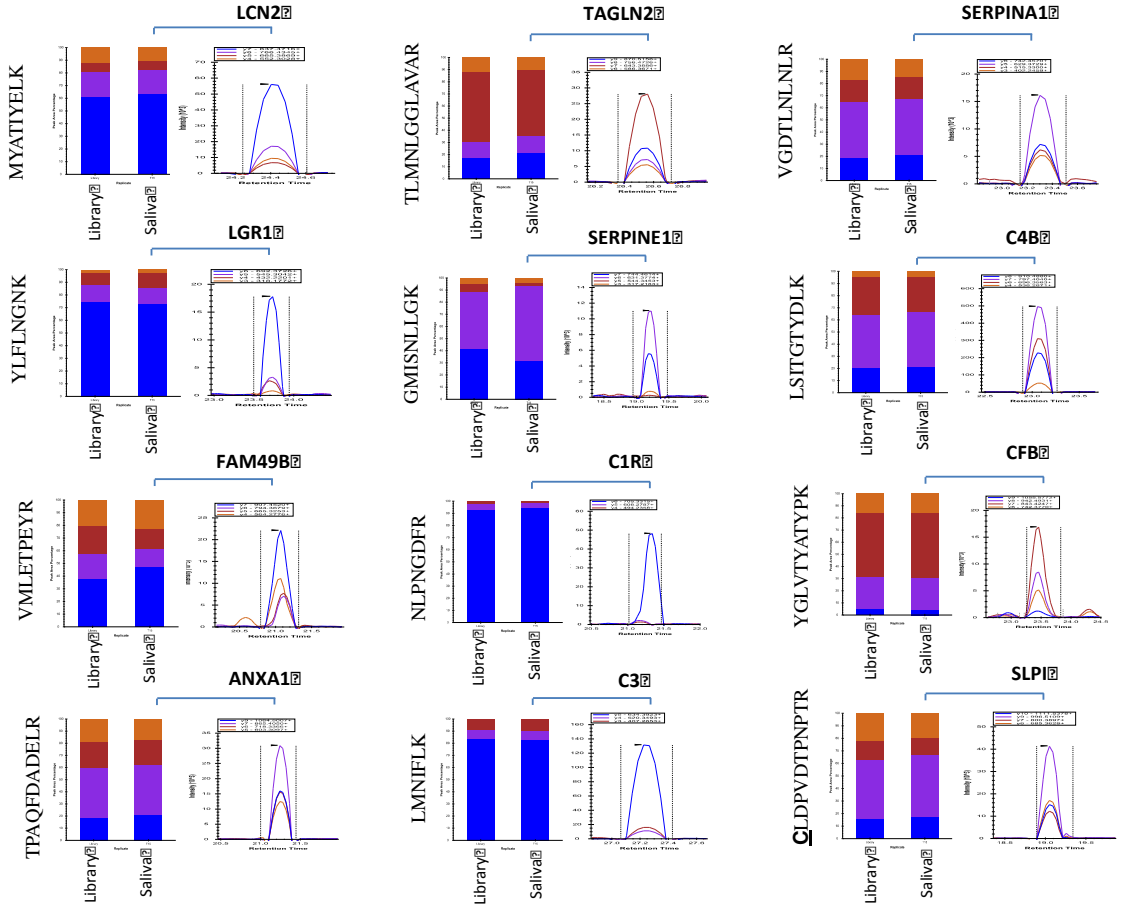

Figure 3 Proteotypic peptides identified using *in vitro*–synthesized protein monitored in human saliva. The relative contribution of each fragment ion to each peptide peak is displayed as different colors. The library contains the result from the recombinant, *in vitro*-synthetized or isolated from plasma proteins. The graph represents endogenous specific transitions for the different peptides measured in the human saliva.

Statistical analysis using MSstats

MSstats is tool for quantitative proteomics that is based on a family of linear mixed-effects model. It represents all the quantitative measurements for targeted protein analysis to facilitate detection of systematic changes in protein abundance between conditions with high sensitivity and specificity all at a controlled false discovery rate [23].

After peak identification and quantitation in Skyline, the results were exported as a .csv file containing the following variables: ProteinName, PeptideSequence, PrecursorCharge, FragmentIon, ProductCharge, IsotopeLabelType, Condition, BioReplicate, Run, and Intensity.

Using the annotation feature present in Skyline, the experiment was designed as a "GroupComparison" experiment. We initially configured the conditions present in the study: Control (n=8, healthy patients), No lesion (n=9, patients who undergone surgical resection) and Lesion (n=13, patients who had active oral malignant lesion at the time of saliva collection). A unique identifier for each biological replicate in the experiment was annotated as "Bioreplicate." The two technical replicates of each biological replicate were assigned with the same unique identifier. The intensity value was considered the transition area without any transformation. The complete MSstats input used in this analysis at a level of both protein and peptide and the results are shown in Supplemental Tables 5 and 6, respectively.

Before importing the file into R studio, rows containing missing values were excluded and all peptides were labeled as light. We performed statistical analysis in two steps. Initially, we summarized all features into protein-level conclusions. Secondly, we considered each feature separately at the peptide level. For protein-level modeling and analysis, the protein id was assigned in the ProteinName column and was the same for all peptides derived from the same protein. For peptide-level comparisons, we alternatively used a unique peptide id in the Protein Name column.

Data were imported into R studio and pre-processed using Log2 followed by normalization to the global standard peptides. This normalization first equalizes endogenous intensities of global standard proteins across runs and applies the same between-run shifts to the remaining endogenous proteins in the experiment [23].

All data were then visualized for exploratory data analysis using dataProcessplots. It was observed from the QC plot that the median intensities of all reference transitions from the global standards across all proteins were equal between runs (Figure 4A and Supplemental Figure 3).

Using the profile plot, we observed the profile of each peptide from each protein across all runs. Profile plot helps to identify potential sources of variation (both variation of interest and

nuisance variation) for each protein. Most of the proteins studied demonstrated a very similar profile plots for all peptides considered in the analysis (Figure 4B and Supplemental Figure 4).

Condition plot was used to visualize potential systematic differences in protein intensities between conditions (Figure 4C and Supplemental Figure 5). Clear differences among the conditions are observed for the C3, C4B, CFB, FAM49B, LCN2, LRG1, SERPINA1 and TAGLN2 proteins, although a more refined model-based estimation was further conducted.

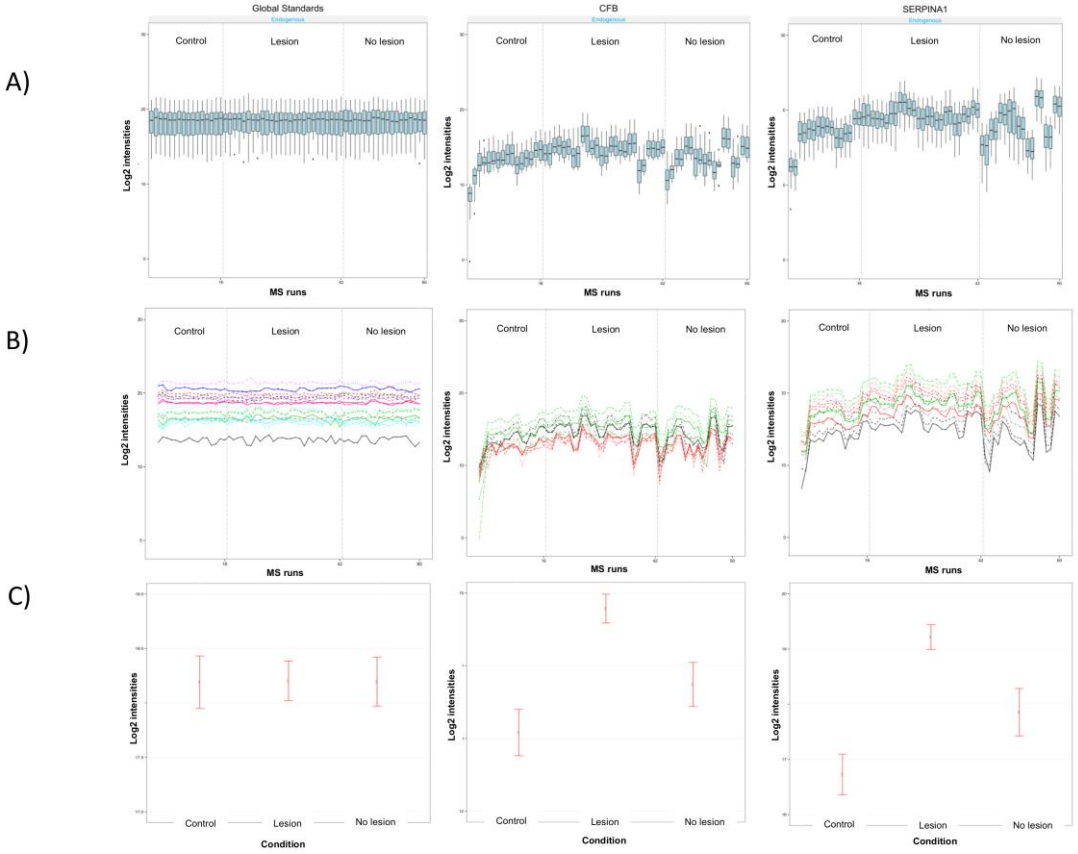

Figure 4 Visualization for explanatory data analysis using MSstats. A) Examples of quality control (QC) plots for Global Standards, SERPINA1 and CFB proteins. X-axis: run. Y-axis: log-intensities of transitions. B) Examples of profile plots for Global Standards, SERPINA1 and CFB proteins. Line colors indicate peptides and line types indicate transitions. C) Examples of condition plots for Global Standards, SERPINA1 and CFB proteins. X-axis: condition. Y-axis: log ratio of endogenous over reference intensities of each transition in a run. Dots indicate the mean of log ratio for each condition. Error bars have confidence intervals with 0.95 significant level for each condition. The plots visualize the differences between conditions, which are of the main biological interest.

The statistical model was used to evaluate each protein for evidence of differential abundance between the conditions. This model takes into account the experimental design and considers the available sources of variation. The output for a differential abundance test is a table

with the following columns: Protein, Label (of the comparison), log2 fold change (log2FC), standard error of the log2 fold change (SE), Student's *t*-test (Tvalue), degree of freedom of the Student's *t*-test (DF), raw p-values (p-value), and p-values adjusted for comparisons across multiple proteins using the approach by Benjamini and Hochberg (adjusted p-value). The cutoff of the adjusted p-value corresponds to the cutoff of the False Discovery Rate [26].

We performed three comparisons of our three conditions simultaneously: control versus OSCC lesion, control versus OSCC no lesion, and OSCC no lesion versus OSCC lesion. The complete MSstats statistical results of each comparison at the level of protein and peptide are shown in Supplemental Tables 7 and 8, respectively.

Volcano plot at a protein and peptide level

Volcano plots visualize the outcome of one comparison between conditions for all the proteins and combine the information on statistical and practical significance.

At a protein-level, the comparison between control vs OSCC lesion showed statistical difference with a FCcutoff>1.5 for all proteins except for ANXA1 and SERPINE1 (Figure 5A). Alternatively, at the peptide level, only one of the peptides from ANXA1 and SERPINE1 failed to demonstrate a statistical difference. One peptide from C1R showed no statistical difference at the peptide level as well (Figure 5B).

In the comparison of control versus OSCC no lesion, ANXA1, C3, LCN2 and SERPINE1 did not demonstrate a statistical difference at a protein level. At a peptide level, the result was similar for the peptides from C3 (2 peptides), SERPINE1 (2 peptides), LNC2 (2 peptides) and ANXA1 (2 peptides). However, other peptides from TAGLN2 (2), C1R, SERPINA1 (1) CFB (2) also did not demonstrate a difference. One peptide from ANXA1 and one peptide from LCN2 were down-regulated and fourteen additional peptides were up-regulated: LRG1 (3), CFB (1), FAM49B (2), TAGLN2 (1), SLPI (1), C4B (3), C3 (1), C1R (1), SERPINA1 (2) (Supplemental Figure 6).

In the OSCC no lesion versus OSCC lesion conditions, we observed that four proteins did not have significant changes in abundance: SLPI, SERPINE1, C1R and FAM49B. The remaining eight proteins were up-regulated in the OSCC lesion condition. Similar results were observed at a peptide level. Peptides from FAM49B (2), SERPINE1 (2), C1R (1) and SLP (1) and TAGLN2 (2) did not demonstrate a difference at the peptide level (Supplemental Figure 7).

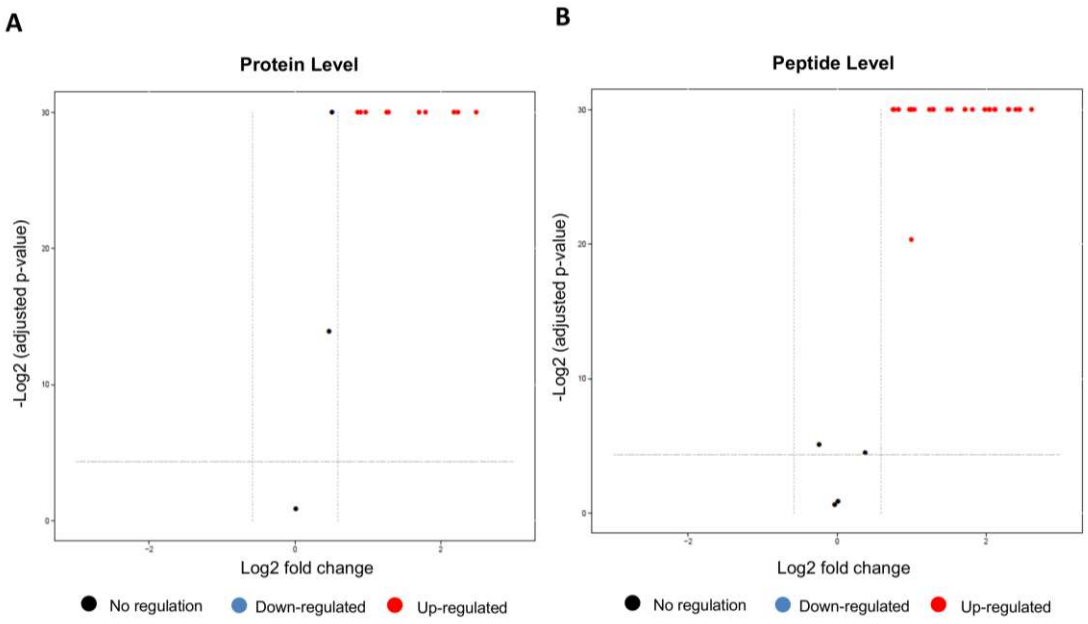

Figure 5 Volcano plot of the comparison Control vs Lesion condition. X-axis: practical significance, model-based estimate of log-fold change. Y-axis: statistical significance, FDR-adjusted p-values on the negative log2 scale. The dashed line represents the FDR cutoff (default sig=0.05) and fold-change FCcutoff=1.5. A) Results at a protein level. Up regulated in OSCC (red): C1R, LCN2, SLPI, FAM49B, TAGLN2, CFB, C3, C4B, LRG1 and SERPINA1. No regulation (black):ANXA1, SERPINE1 and Global Standards. B) Results at a peptide level. Up-regulated in OSCC (red): ANXA1 (peptides TPA, NAL); CFB (peptides YGL, VSE, DIS); C3 (peptides LMN, SSL, TFI); C4B (peptides TTN, VGD, GSF); C1R (peptide DYF); LCN2 (peptides MYA, TFV, SYP); LRG1 (peptides DLL, YLF, VAA); FAM49B (peptides MSL, VML); SERPINA1 (peptides LSI, SVL, AVL); SERPINE1 (peptide FII), SLPI (peptide CLD); TAGLN2 (peptides IQA, TLM, NVI). No regulation (black): ANXA1 (peptide ALY); SERPINE1 (peptide GMI) and C1R (peptide NLP).

Comparison between relative quantification of CFB using labeled peptide or global standards

The ratio of the YGLVTYATYPK and VSEADSSNADWVTK peptides to their heavy isotope labeled analogs was compared to the ratio of these same peptides to the set of global standards. Integrated signal intensities were exported using Skyline (Supplemental table 4 and 9) and compared by utilizing statistical significance among the three conditions (control, OSCC no lesion, and OSCC lesion) with a Student's *t*-test.

Both normalization approaches showed statistical significance between control and lesion condition for the two CFB peptides (Student's *t*-test, $p<0.05$, Supplemental Figure 8). Both normalization approaches resulted in no statistically significant differences between the control versus OSCC no lesion and OSCC no lesion versus OSCC lesion comparison.

Although SRM with stable isotope-labeled standards is considered to be the "gold standard" MS-based quantitation method [27,28], our results demonstrate that comparative analytical assay performance can be obtained by utilizing external reference standards.

However, it is important to remember that there are different "fit-for-purpose" approaches for targeted proteomics [29]. For comparative analysis, as in the case of this study, using semi-quantitative measurements of proteins, protein forms, or peptides in biological systems (Tier 3), label-free measurement can be used with additional chromatographic and mass spectrometric information to establish confidence in peptide assignment. If the purpose is clinical bioanalysis/ diagnostic, laboratory test (Tier 1) or research use assays for quantifying proteins, peptides, and posttranslational modifications (Tier 2), best practice dictates the use of stable isotope-labeled internal standards for each targeted analyte.

Risk analysis of OSCC, using the targeted peptides measured by SRM

To explore the relevance of the protein expression in clinical application, we used the peptide levels for each protein evaluated in this study to perform risk association to OSCC analysis. The normalized ratios for targeted peptides to global standards from 31 peptides, considering only those filtered by dotp>0.8, in 30 samples are shown in the Supplemental Table 4.

The clinicophatological variables revealed that patients were mainly diagnosed at advanced clinical stage (most have lymph nodes involved) with a predominantly moderate/poor WHO differentiation degree and active lesions at the time of saliva collection (Table 2).

Table 2 Clinical and pathological characteristics of OSCC patients (n=22).

Features	OSCC n (%)
Sex	
Female	3 (13.6)
Male	19 (86.4)
TNM[1]	
Primary tumor (T)	
1	2 (15.4)
2	4 (30.8)
3	2 (15.4)
4	5 (38.5)
Regional lymph nodes (N)	
0	3 (23.1)
1	4 (30.8)
2	6 (46.2)
3	0
Metastasis (M)	
M0	2 (15.4)
Mx	11 (84.6)
WHO differentiation degree[2]	
Well	5 (29.4)
Moderate	7 (41.2)
Poor	5 (29.4)
Active lesion[1]	
No	9 (40.9)
Yes	13 (59.1)
Survival state	
Alive	14 (63.6)
Deceased	8 (36.4)

[1]n=22, [2]n=17

Using univariate factor binary logistic regression, we observed an association between the CFB_1 (YGLVTYATYPK), C3_3 (TFISPIK), C4B_1 (TTNIQGINLLFSSR), C4B_3 (GSFEFPVGDAVSK), LGR1_1 (DLLLPQPDLR), LRG1_2 (YLFLNGNK) and SERPINA1_3 (LSITGTYDLK) peptides with OSCC risk. For all these peptides, the odds ratio for OSCC was 12.50 (95% of confidence interval (CI) and p=0.03) (Table 3, Supplemental Material and Methods).

Interestingly, four out of five of these proteins (C3, CFB, C4B and SERPINA1) are involved in the complement and coagulation cascade, which was previously demonstrated to be enriched by network analysis using the carcinoma candidate biomarkers (KEGG pathway, $p < 0.05$) (Kawahara et al., 2014).

The complement system has recently received increased attention in biomedical research. In light of evidence that complement proteins can facilitate cellular proliferation and contribute to the inflammatory state of the cancer [30,31], there is increasing speculation that there is a relationship between complement proteins and cancer. In previous work, we demonstrated the role

of CFB in tumor cell migration and macrophage chemotaxis (Kawahara et al., 2014). Recently, Cho et al. 2014 [32] also described a role for the complement system in enhancing cancer growth and C3 was implicated as a key player for the activation of ovarian cancer growth and progression. Furthermore, Kim et al. (2014) suggested C3 as a plasma biomarker for the measurement disease progression in neuroblastoma patients [33] and Lee et al. (2014) showed CFB as biomarker candidate for pancreatic ductal adenocarcinoma [34]. Thus, a panel of complement proteins has emerged as potential candidates associated with cancer development. Our study suggests that these candidates are present in high abundance in saliva of OSCC patients and that they were also associated with OSCC risk. Collectively, our results open the potential of novel avenues for OSCC follow-up studies.

Table 2 Risk analysis of OSCC, using the peptides intensities normalized by global standards measured by SRM.

Peptides[a]	Peptide Sequence	Odds ratio[b]	IC 95%	p-value
ANXA1_1	TPAQFDADELR	0.231	0.38-1.413	0.113
ANXA1_2	NALLSLAK	0.231	0.38-1.413	0.113
ANXA1_3	ALYEAGER	1	0.198-5.045	1
CFB_1	YGLVTYATYPK	12.250	1.268-118.361	0.030*
CFB_2	VSEADSSNADWVTK	4.333	0.708-26.531	0.113
CFB_3	DISEVVTPR	4.333	0.708-26.531	0.113
C1R_1	DYFIATCK	4.333	0.708-26.531	0.113
C1R_2	NLPNGDFR	0.231	0.038-1.413	0.113
C3_1	LMNIFLK	4.333	0.708-26.531	0.113
C3_2	SSLSVPYVIVPLK	4.333	0.708-26.531	0.113
C3_3	TFISPIK	12.250	1.268-118.361	0.030*
C4B_1	TTNIQGINLLFSSR	12.250	1.268-118.361	0.030*
C4B_2	VGDTLNLNLR	1.846E9	0.0-iv	0.998
C4B_3	GSFEFPVGDAVSK	12.250	1.268-118.361	0.030*
LCN2_1	MYATIYELK	2	0.381-10.511	0.413
LCN2_2	TFVPGCQPGEFTLGNIK	1	0.198-5.045	1
LCN2_3	SYPGLTSYLVR	1	0.198-5.045	1
LGR1_1	DLLLPQPDLR	12.250	1.268-118.361	0.030 *
LGR1_2	YLFLNGNK	12.250	1.268-118.361	0.030*
LGR1_3	VAAGAFQGLR	1.846E9	0.0-iv	0.998
FAM49B_1	MSLFYAEATPMLK	2	0.381-10.511	0.413
FAM49B_2	VMLETPEYR	1	0.198-5.045	1
SERPINA1_1	LSITGTYDLK	1.612E9	0.0-iv	0.998
SERPINA1_2	SVLGQLGITK	8.4	0.879-80.265	0.065
SERPINA1_3	AVLTIDEK	12.250	1.268-118.361	0.030 *
SERPINE1_1	FIINDWVK	0.5	0.095-2.628	0.413
SERPINE1_2	GMISNLLGK	0.867	0.164-4.579	0.866
SLPI	CLDPVDTPNPTR	0.952	0.179-5.081	0.954
TAGLN2_1	IQASTMAFK	1.0	0.198-5.045	1.000
TAGLN2_2	TLMNLGGLAVAR	1.0	0.198-5.045	1.000

TAGLN2_3	NVIGLQMGTNR	1.0	0.198-5.045	1.000

(a) Categorized as 'low' and 'high'. (b) Odds ratio for Low/High. (iv) Indeterminate value. * $p<0.05$

CONCLUSION

In summary, this study presented a targeted proteomic approach for the measurement of OSCC candidate biomarkers in human saliva. We demonstrated a statistically significant higher abundance of ten candidate biomarkers in saliva of OSCC patients and among them, CFB, C3, C4B, SERPINA1 and LRG1 are associated with the risk of developing OSCC.

REFERENCES

(1) Siegel, R.; Naishadham, D.; Jemal, A.: Cancer statistics, 2013. *CA: a cancer journal for clinicians* **2013**, *63*, 11-30.

(2) Argiris, A.; Karamouzis, M. V.; Raben, D.; Ferris, R. L.: Head and neck cancer. *Lancet* **2008**, *371*, 1695-709.

(3) Mydlarz, W. K.; Hennessey, P. T.; Califano, J. A.: Advances and Perspectives in the Molecular Diagnosis of Head and Neck Cancer. *Expert opinion on medical diagnostics* **2010**, *4*, 53-65.

(4) da Silva, S. D.; Ferlito, A.; Takes, R. P.; Brakenhoff, R. H.; Valentin, M. D.; Woolgar, J. A.; Bradford, C. R.; Rodrigo, J. P.; Rinaldo, A.; Hier, M. P.; Kowalski, L. P.: Advances and applications of oral cancer basic research. *Oral oncology* **2011**, *47*, 783-91.

(5) Hall, S. F.; Groome, P. A.; Rothwell, D.; Dixon, P. F.: Using TNM staging to predict survival in patients with squamous cell carcinoma of head & neck. *Head & neck* **1999**, *21*, 30-8.

(6) Bonne, N. J.; Wong, D. T.: Salivary biomarker development using genomic, proteomic and metabolomic approaches. *Genome medicine* **2012**, *4*, 82.

(7) Hu, S.; Arellano, M.; Boontheung, P.; Wang, J.; Zhou, H.; Jiang, J.; Elashoff, D.; Wei, R.; Loo, J. A.; Wong, D. T.: Salivary proteomics for oral cancer biomarker discovery. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2008**, *14*, 6246-52.

(8) de Jong, E. P.; Xie, H.; Onsongo, G.; Stone, M. D.; Chen, X. B.; Kooren, J. A.; Refsland, E. W.; Griffin, R. J.; Ondrey, F. G.; Wu, B.; Le, C. T.; Rhodus, N. L.; Carlis, J. V.; Griffin, T. J.: Quantitative proteomics reveals myosin and actin as promising saliva biomarkers for distinguishing pre-malignant and malignant oral lesions. *PloS one* **2010**, *5*, e11148.

(9) Chang, K. P.; Yu, J. S.; Chien, K. Y.; Lee, C. W.; Liang, Y.; Liao, C. T.; Yen, T. C.; Lee, L. Y.; Huang, L. L.; Liu, S. C.; Chang, Y. S.; Chi, L. M.: Identification of PRDX4 and P4HA2 as metastasis-associated proteins in oral cavity squamous cell carcinoma by comparative tissue proteomics of microdissected specimens using iTRAQ technology. *Journal of proteome research* **2011**, *10*, 4935-47.

(10) Sepiashvili, L.; Hui, A.; Ignatchenko, V.; Shi, W.; Su, S.; Xu, W.; Huang, S. H.; O'Sullivan, B.; Waldron, J.; Irish, J. C.; Perez-Ordonez, B.; Liu, F. F.; Kislinger, T.: Potentially novel candidate biomarkers for head and neck squamous cell carcinoma identified using an integrated cell line-based discovery strategy. *Molecular & cellular proteomics : MCP* **2012**, *11*, 1404-15.

(11) Hsu, C. W.; Yu, J. S.; Peng, P. H.; Liu, S. C.; Chang, Y. S.; Chang, K. P.; Wu, C. C.: Secretome Profiling of Primary Cells Reveals That THBS2 Is a Salivary Biomarker of Oral Cavitfy Squamous Cell Carcinoma. *Journal of proteome research* **2014**.

(12) Kulasingam, V.; Diamandis, E. P.: Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature clinical practice. Oncology* **2008**, *5*, 588-99.

(13) Makawita, S.; Diamandis, E. P.: The bottleneck in the cancer biomarker pipeline and protein quantification through mass spectrometry-based approaches: current strategies for candidate verification. *Clinical chemistry* **2010**, *56*, 212-22.

(14) Rifai, N.; Gillette, M. A.; Carr, S. A.: Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature biotechnology* **2006**, *24*, 971-83.

(15) Sturgeon, C.; Hill, R.; Hortin, G. L.; Thompson, D.: Taking a new biomarker into routine use--a perspective from the routine clinical biochemistry laboratory. *Proteomics. Clinical applications* **2010**, *4*, 892-903.

(16) Picotti, P.; Rinner, O.; Stallmach, R.; Dautel, F.; Farrah, T.; Domon, B.; Wenschuh, H.; Aebersold, R.: High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nature methods* **2010**, *7*, 43-6.

(17) Whiteaker, J. R.; Lin, C.; Kennedy, J.; Hou, L.; Trute, M.; Sokal, I.; Yan, P.; Schoenherr, R. M.; Zhao, L.; Voytovich, U. J.; Kelly-Spratt, K. S.; Krasnoselsky, A.; Gafken, P. R.; Hogan, J. M.; Jones, L. A.; Wang, P.; Amon, L.; Chodosh, L. A.; Nelson, P. S.; McIntosh, M. W.; Kemp, C. J.; Paulovich, A. G.: A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nature biotechnology* **2011**, *29*, 625-34.

(18) Stergachis, A. B.; MacLean, B.; Lee, K.; Stamatoyannopoulos, J. A.; MacCoss, M. J.: Rapid empirical discovery of optimal peptides for targeted proteomics. *Nature methods* **2011**, *8*, 1041-3.

(19) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J.: Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966-8.

(20) Maclean, B.; Tomazela, D. M.; Abbatiello, S. E.; Zhang, S.; Whiteaker, J. R.; Paulovich, A. G.; Carr, S. A.; Maccoss, M. J.: Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Analytical chemistry* **2010**, *82*, 10116-24.

(21) Surinova, S.; Huttenhain, R.; Chang, C. Y.; Espona, L.; Vitek, O.; Aebersold, R.: Automated selected reaction monitoring data analysis workflow for large-scale targeted proteomic studies. *Nature protocols* **2013**, *8*, 1602-19.

(22) Chang, C. Y.; Picotti, P.; Huttenhain, R.; Heinzelmann-Schwarz, V.; Jovanovic, M.; Aebersold, R.; Vitek, O.: Protein significance analysis in selected reaction monitoring (SRM) measurements. *Molecular & cellular proteomics : MCP* **2012**, *11*, M111 014662.

(23) Choi, M.; Chang, C. Y.; Clough, T.; Broudy, D.; Killeen, T.; Maclean, B.; Vitek, O.: MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **2014**.

(24) Broudy, D.; Killeen, T.; Choi, M.; Shulman, N.; Mani, D. R.; Abbatiello, S. E.; Mani, D.; Ahmad, R.; Sahu, A. K.; Schilling, B.; Tamura, K.; Boss, Y.; Sharma, V.; Gibson, B. W.; Carr, S. A.; Vitek, O.; MacCoss, M. J.; MacLean, B.: A framework for installable external tools in Skyline. *Bioinformatics* **2014**, *30*, 2521-3.

(25) Sherrod, S. D.; Myers, M. V.; Li, M.; Myers, J. S.; Carpenter, K. L.; Maclean, B.; Maccoss, M. J.; Liebler, D. C.; Ham, A. J.: Label-free quantitation of protein modifications by pseudo selected reaction monitoring with internal reference peptides. *Journal of proteome research* **2012**, *11*, 3467-79.

(26) Benjamini, Y. H., Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* **1995**, *57*, 289-300.

(27) Lange, V.; Picotti, P.; Domon, B.; Aebersold, R.: Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular systems biology* **2008**, *4*, 222.

(28) Keshishian, H.; Addona, T.; Burgess, M.; Kuhn, E.; Carr, S. A.: Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Molecular & cellular proteomics : MCP* **2007**, *6*, 2212-29.

(29) Carr, S. A.; Abbatiello, S. E.; Ackermann, B. L.; Borchers, C.; Domon, B.; Deutsch, E. W.; Grant, R. P.; Hoofnagle, A. N.; Huttenhain, R.; Koomen, J. M.; Liebler, D. C.; Liu,

T.; MacLean, B.; Mani, D. R.; Mansfield, E.; Neubert, H.; Paulovich, A. G.; Reiter, L.; Vitek, O.; Aebersold, R.; Anderson, L.; Bethem, R.; Blonder, J.; Boja, E.; Botelho, J.; Boyne, M.; Bradshaw, R. A.; Burlingame, A. L.; Chan, D.; Keshishian, H.; Kuhn, E.; Kinsinger, C.; Lee, J. S.; Lee, S. W.; Moritz, R.; Oses-Prieto, J.; Rifai, N.; Ritchie, J.; Rodriguez, H.; Srinivas, P. R.; Townsend, R. R.; Van Eyk, J.; Whiteley, G.; Wiita, A.; Weintraub, S.: Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Molecular & cellular proteomics : MCP* **2014**, *13*, 907-17.

(30) Rutkowski, M. J.; Sughrue, M. E.; Kane, A. J.; Mills, S. A.; Parsa, A. T.: Cancer and the complement cascade. *Molecular cancer research : MCR* **2010**, *8*, 1453-65.

(31) Rao, S. K.; Pavicevic, Z.; Du, Z.; Kim, J. G.; Fan, M.; Jiao, Y.; Rosebush, M.; Samant, S.; Gu, W.; Pfeffer, L. M.; Nosrat, C. A.: Pro-inflammatory genes as biomarkers and therapeutic targets in oral squamous cell carcinoma. *The Journal of biological chemistry* **2010**, *285*, 32512-21.

(32) Cho, M. S.; Vasquez, H. G.; Rupaimoole, R.; Pradeep, S.; Wu, S.; Zand, B.; Han, H. D.; Rodriguez-Aguayo, C.; Bottsford-Miller, J.; Huang, J.; Miyake, T.; Choi, H. J.; Dalton, H. J.; Ivan, C.; Baggerly, K.; Lopez-Berestein, G.; Sood, A. K.; Afshar-Kharghan, V.: Autocrine effects of tumor-derived complement. *Cell reports* **2014**, *6*, 1085-95.

(33) Kim, P. Y.; Tan, O.; Diakiw, S. M.; Carter, D.; Sekerye, E. O.; Wasinger, V. C.; Liu, T.; Kavallaris, M.; Norris, M. D.; Haber, M.; Chesler, L.; Dolnikov, A.; Trahair, T. N.; Cheung, N. K.; Marshall, G. M.; Cheung, B. B.: Identification of plasma complement C3 as a potential biomarker for neuroblastoma using a quantitative proteomic approach. *Journal of proteomics* **2014**, *96*, 1-12.

(34) Lee, M. J.; Na, K.; Jeong, S. K.; Lim, J. S.; Kim, S. A.; Lee, M. J.; Song, S. Y.; Kim, H.; Hancock, W. S.; Paik, Y. K.: Identification of human complement factor B as a novel biomarker candidate for pancreatic ductal adenocarcinoma. *Journal of proteome research* **2014**, *13*, 4878-88.

CAPÍTULO 3

Agrin and perlecan mediate tumorigenic processes in oral squamous cell carcinoma

Rebeca Kawahara, Daniela C. Granato, Carolina M. Carnielli, Nilva Cervigne, Carine Ervolino de Oliveria, César Rivera Martinez, Sami Yokoo, Felipe Paiva Fonseca, Alan R. Santos-Silva, Edgard Graner, Ricardo D. Coletta and Adriana F. Paes Leme

PLOS ONE, v.9, n.12, 2014
(doi: 10.1371/journal.pone.0115004)

ABSTRACT

Oral squamous cell carcinoma is the most common type of cancer in the oral cavity, representing more than 90% of all oral cancers. The characterization of altered molecules in oral cancer is essential to understand molecular mechanisms underlying tumor progression as well as to contribute to cancer biomarker and therapeutic target discovery. Proteoglycans are key molecular effectors of cell surface and pericellular microenvironments, performing multiple functions in cancer. Two of the major basement membrane proteoglycans, agrin and perlecan, were investigated in this study regarding their role in oral cancer. Using real time quantitative PCR (qRT-PCR), we showed that agrin and perlecan are highly expressed in oral squamous cell carcinoma. Interestingly, cell lines originated from distinct sites showed different expression of agrin and perlecan. Enzymatically targeting chondroitin sulfate modification by chondroitinase, oral squamous carcinoma cell line had a reduced ability to adhere to extracellular matrix proteins and increased sensibility to cisplatin. Additionally, knockdown of agrin and perlecan promoted a decrease on cell migration and adhesion, and on resistance of cells to cisplatin. Our study showed, for the first time, a negative regulation on oral cancer-associated events by either targeting chondroitin sulfate content or agrin and perlecan levels.

INTRODUCTION

Head and neck cancers are the sixth most common malignancy in the world, accounting for more than 500,000 new cases every year [1]. Oral squamous cell carcinoma (OSCC) is the most prevalent cancer occurring in this area [2]. Despite advancements in prevention and multimodality treatments, oral cancer is still characterized by poor prognosis and a low survival rate [3-5].

Long-standing as well as recent data implicate tumor extracellular matrix (ECM) as a significant contributor to tumor progression [6,7]. However, the entire process orquestrated by interactions between cancer cells and ECM remains poorly understood. One of the major constituents of the ECM, the proteoglycans (PGs), is markedly altered during malignant transformation and tumor progression. Their role is associated with a number of tumorigenic processes, including control of cell growth and survival, induction of apoptosis, adhesion, and invasion [8-10]. Among the main heparan sulfate PGs (HSPG), identified in basement membrane, are agrin and perlecan, which not only were reported to be overexpressed in some cancers, such as prostate cancer [11], hepatocellular carcinoma [12] and breast cancer [13], but also had their function associated with tumorigenic events [10,14,15]. Though, no evidence was reported regarding their role in oral cancer.

Perlecan is a large proteoglycan (400-500 kDa) harboring five distinct structural domains, to which long chains of heparan sulfate and/or chondroitin sulfate are attached [16]. This molecule is present in all vascularized tissues with a distribution that is primarily confined to basement membranes [17,18]. Also, other studies have also identified perlecan in the stromal spaces of various pathophysiological conditions [19-21].

Agrin shares a rather intriguing multimodular organization with perlecan, but more complexity to agrin can be added due to at least four sites of alternative splicing [22]. The amino acid sequence of agrin encodes a protein with a molecular size of 220 kDa, but the observed molecular weight is around 400 kDa due to the long heparan sulfate (HS) and chondroitin sulfate (CS) glycosaminoglycans (GAGs) attached to the core protein [23]. Although originally discovered in the neuromuscular junctions, agrin has been observed in numerous other tissues, and it is described as highly expressed in hepatocellular carcinomas [12,24,25] and cholangiocellular carcinomas [12,24]. Nevertheless, little is known about its role at locations other than the neuromuscular junctions, and even less information is known about its role in tumor tissues.

In the present study, we focused on understanding the role of the proteoglycans agrin and perlecan in oral cancer. First, we sought to validated the overexpression of agrin and perlecan in oral cancer tissues compared to normal tissues and in cell lines with different site of origin: oral squamous carcinoma originated from human tongue (SCC-9), oral squamous carcinoma SCC-9 isolated from lymph nodes (SCC-9 LN-1) and a skin-derived squamous carcinoma (A431). Next, we showed that oral squamous carcinoma cell line had a reduced ability to adhere to extracellular matrix proteins and increased sensibility to cisplatin when treated with chondroitinase. By specific target agrin and perlecan protein levels with siRNA, we showed that OSCC cells have decreased cell adhesion and migration and increased sensibility to cisplatin treatment. Overall, our findings opened new avenues to better understand the role of agrin and perlecan, as well as their involvement in carcinogenesis, which may offer a novel approach to cancer therapy by targeting the tumor microenvironment.

MATERIALS AND METHODS

Cell culture

SCC-9 cells (a tumor cell line originated from a human tongue squamous cell carcinoma) were obtained from the American Type Culture Collection (ATCC, Manassas, VA) and cultured in DMEM/Ham's F12 medium (Cultilab), supplemented with 10% fetal bovine serum (FBS), antibiotics and 0.4 µg/ml hydrocortisone.

Metastatic oral squamous cell carcinoma (SCC-9) cells were isolated from lymph nodes (LN-1) originating the cell line SCC-9 LN-1 [26] and cultured as recommended for SCC-9. Human epidermoid carcinoma, A431, were grown in Roswell Park Memorial Institute (RPMI) −1640 medium supplemented with 10% FBS and antibiotics. All cell lines were maintained at 37°C in a 5% CO_2 atmosphere.

Analysis of mRNA Expression levels

Fresh-frozen OSCC samples (n=16) and normal oral mucosa (n=16) were used for perlecan and agrin mRNA levels quantification, using qRT-PCR. All patients and volunteers enrolled signed a formulary stating their awareness and consent for the study, approved by the Research Ethics Committee of Faculdade de Odontologia de Piracicaba, Universidade Estadual de Campinas, UNICAMP, Piracicaba, Brazil. Clinical pathological data, such as sex, age, anatomical site of the primary tumor, clinical stage and histopathological grade were collected from patient's charts and showed in Table S1. Brief, total mRNA was isolated from fresh-frozen tissue samples using mirVana™ miRNA Isolation Kit (Ambion), according to the manufacturer's protocol. First-strand cDNAs were synthesized from 2µg of DNase-treated total RNA using the SuperScript® II Reverse Transcription Kit (Invitrogen, Carlsbad, CA, USA). Briefly, diluted cDNA product (1:3) was used to perform a quantitative reverse-transcription polymerase chain reaction (qRT-PCR) using SYBR Green PCR Master Mix (Applied Bios stems) in a 7900 Real-Time PCR System (Applied Biosystems, Foster, CA, USA). The following primers were used to determine mRNA expression levels: Perlecan forward 5'- AAT GCGCTGGACACATTCG-3' and reverse 5'-ATTCACCAGGGCTCGGAAATA-3'; Agrin forward 5'TTGTCGAGTA CCTCAACGCT-3' and reverse 5'-CAGGCTCAGTTCAAAGTCGT-3'. PPIA (cyclophilin A) was used as the reference gene, forward 5' GCTTTGGGTCCAGGAATGG3' and reverse 5' GTTGTCCACAGTCAGCAAT GGT3'. The PCR cycles were 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. Each reaction was performed in triplicate and analyzed individually. The

results were calculated using $2^{-\Delta\Delta Ct}$ relative quantification method; relative quantification was normalized to the pooled normal oral tissues, used as reference control.

For analysis of mRNA expression levels in A431, SCC-9 LN-1 and SCC-9 cell lines, total RNA was obtained using the TRIzol reagent (Invitrogen Corporation) and 2 µg of total RNA were used for retro-transcription using a First-Strand cDNA Synthesis Kit (GE Healthcare). Real-time quantitative PCR for agrin and perlecan was performed using the SYBRH Green PCR Master Mix (Applied Biosystems), and the dissociation curves were performed to confirm the specificity of the products. The threshold cycle (CT) values of the targeted gene were normalized relative to glyceraldehyde-3-phosphate dehydrogenase gene, and relative expression ratios were calculated using the 2-ΔΔ Ct method. Two independent experiments were performed with triplicates.

Immunohistochemistry

High density tissue microarrays were obtained from Biomax (OR601a). The presence of agrin was analyzed in 10 cancer-adjacent normal tissues and in 47 primary oral squamous cell carcinomas by immunohistochemistry using the streptavidin-biotin peroxidase complex (Dako). Protein quantification was assessed with the aid of Aperio Scanscope CS Slide Scanner and the Pixel Count V9 algorithm software (Aperio Technologies, Vista, CA; USA). By using specific input parameters, the percentage of cytoplasm positivity was calculated and classified as weak, moderate and strong, according to its staining intensity. Each category received an intensity score, 1 to weak, 2 to moderate and 3 to strong staining. The final score of each tissue sample was calculated as the sum of the percentage of each category multiplied by their respective intensity scores as the following formula: Score = (%weak x 1) + (%moderate x 2) + (%strong x 3). Clinical pathological data, such as sex, age, anatomical site of the primary tumor, clinical stage and histopathological grade were collected from patient's charts and showed in Table S2.

Chondroitinase treatment

Chondroitinase ABC from Proteus vulgaris (Sigma) was reconstituted in a 0.01% bovine serum albumin (BSA) aqueous solution to a final concentration of 6 U/ml. The treatment was performed by diluting the stock chondroitinase in serum free media to a final concentration of 0.1 U/ml for 4h at 37^{o}C. After treatment, SCC-9 LN-1 cell lines were submitted to adhesion and migration assays.

In vitro cell migration and adhesion assays

SCC-9, SCC-9 LN-1 or A431 cells, transfected with control siRNA (scramble, sc-44510, Santa Cruz) or specific siRNA against agrin (sc-29652, Santa Cruz) or perlecan (sc-44010, Santa Cruz), were submitted either to cell migration or adhesion assays. Briefly, $3x10^5$ cells were plated in 6 well-plate and, after 24 h, the oligoribonucleotides were transfected with lipofectamine 2000, according to the manufacturer (Invitrogen).

For the in vitro cell migration assay, cells were harvested after 72 h, plated in the upper chambers of the transwell (HTS Transwell 96-Well Plate, Corning), and allowed to migrate towards the lower chamber containing RPMI medium supplemented with 1% FBS for 16 h. At the end of the assay, the non-migrated cells at the top chamber were removed using a cotton swab, and the cells at the bottom of the insert filter were fixed with 10% formaldehyde for 10 min, followed by PBS washing and then staining with 1% toluidine blue solution in 1% borax, for 5 min. The dye was eluted using 1% SDS, and the absorbance of stained cells was measured at 620 nm. Three independent experiments were performed in triplicates.

SCC-9, SCC-9 LN-1 or A431 siRNAs-transfected cells were also submitted to cell adhesion assay, as described by Aragão et al [27]. After 72 h, cells were seeded in a 6-well plate ($2x10^5$) and incubated for 24h. Then, were washed twice, incubated in serum-free media for 4h, and seeded in a Matrigel (2 µg per well; BD Biosciences) coated 96-well plate, previously three times PBS-washed and blocked with 3% BSA during 2 h. The adhesion was evaluated during 1 h in serum-free media supplemented with 3% BSA. The wells were washed 3 times, and cells fixed with 10% formaldehyde. Cells were then stained with toluidine blue solution, and absorbance measured at 620 nm, as described above. Three independent experiments were performed with three replicates.

In vitro viability assay

SCC-9, SCC-9 LN-1 or A431 siRNA-transfected cells (scramble, Agrin or Perlecan) were seeded onto 96-well plates and incubated at $37^oC/5\%CO_2$ for two days. MTT (12 mM tetrazolium 3-(4,5- dimethylthiazol-2)-2,5-diphenyltetrazolium bromide) was added, and cells were kept at 37 °C for 4 h, in the dark. The media was removed, and 100 µl of 1 N HCl/isopropanol (1:25) was added into each well, followed by gentle agitation at room temperature for 15 min. Finally, the absorbance was measured at 595 nm. Three independent experiments were performed in triplicates.

Drug Sensitivity Assay (MTT)

Cisplatin [cis-diammineplatinum(II) dichloride] (Sigma-Aldrich) was dissolved in 0.150 M NaCl. Aliquots were stored at -20 °C for up to a maximum of three months, and thawed immediately before use.

Cells ($1x10^4$) were seeded in 96-well plates and allowed to adhere overnight at 37 °C. Briefly, following treatments of cells with cisplatin (0; 1; 5; 7.5; 10; 25; 50; 100 µM) for 48 h, MTT reagent [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] was added to each well, and incubated for 4 h at 37°C, in the dark. The media were removed, 100 µl of 1 N HCl/isopropanol (1:25) was added in each well, and incubated for 15 min at room temperature under gentle agitation. Finally, absorbance was measured at 595 nm. Three independent experiments were performed in triplicates.

Statistical Analysis

Statistical analysis of gene expression validation by qRT-PCR and immunohistochemistry was performed using the non-parametric Kruskal-Wallis followed by Dunn's test and Mann-Whitney U test, respectively. For the functional assays adhesion, migration and proliferation, Student's *t*-test or ANOVA followed by Tukey's test was used; p-values<0.05 were set as statistically significant. For drug sensitivity assay, a nonlinear regression curve fit (one phase exponential decay) was used to analyze cisplatin dose response experiments and determine the IC50. All the statistical analysis for mRNA expression analysis and functional assays were performed in GraphPad Prism v6.01.

RESULTS

Agrin and Perlecan are up-regulated in OSCC samples

We started this study by evaluating the mRNA levels of agrin and perlecan, using qRT-PCR in an independent cohort of OSCC samples and normal oral mucosa. Higher expression levels of agrin and perlecan were observed in OSCC samples compared to controls (Figure 1A, n=16, Kruskal-Wallis followed by Dunn's test, $p<0.05$, Table S1).

We also performed immunohistochemistry analysis in commercial tissue microarrays containing 10 cancer-adjacent normal tissues and 47 primary OSCCs. Normal epithelial cells showed weak reactivity for the antibody against agrin, but a broad positivity was found in the

neoplastic cells. Furthermore, there was a significantly higher expression of agrin in tumor samples, compared to controls (Figure S1, Mann-Whitney U test, $p<0.0005$, Table S2).

The mRNA levels of agrin and perlecan were evaluated in cell lines with different sites of origin: oral squamous carcinoma cell line (SCC-9) isolated from tongue, oral squamous carcinoma cell line (SCC-9) isolated from lymph nodes (SCC-9 LN-1) and a skin-derived squamous carcinoma (A431). Whereas the mRNA levels of agrin were higher in SCC-9, the mRNA levels of perlecan were higher in metastatic SCC-9 LN-1 cell lines. A431 showed an intermediary expression for both agrin and perlecan (Figure 1 B/C).

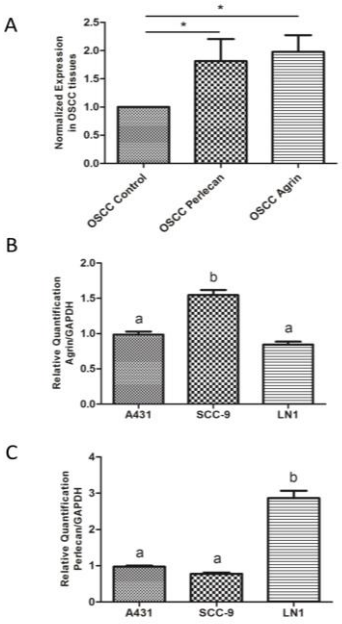

Figure 1 mRNA expression levels of agrin and perlecan. (A) Validation of higher expression of agrin and perlecan by qRT-PCR in OSCC tumor tissues**.** Agrin and Perlecan showed higher mRNA expression levels in human OSCC tumor tissues compared to control tissues by qRT-PCR (Kruskal-Wallis followed by Dunn's test, n = 16, * $p<0.05$). (B) Agrin showed higher mRNA expression levels in SCC-9 compared to A431 and SCC-9 LN-1 cell lines. (C) Perlecan showed higher mRNA expression levels in SCC-9 LN-1 compared with A431 and SCC-9 cell lines. The data were normalized with glyceraldehyde-3-phosphate dehydrogenase gene, used as internal reference). Each bar represents means ± SD of at least two independent experiments in triplicates (one-way ANOVA followed by Tukey's test. Different letters indicate statistically difference at $p<0.05$).

Targeting chondroitin sulfate modification reduced the ability of SCC-9 LN-1 cells to adhere to extracellular matrix and decreased the cell resistance to cisplatin

To further understand the general role of chondroitin sulfate post-translational modification in tumorigenic processes such as cell adhesion and migration, SCC-9 LN-1 cells were

treated with 0.1 U/ml of chondroitinase for 4 h in serum free media, and evaluated for the ability of these cells to adhere to ECM proteins (Matrigel) and migrate. We showed that SCC-9 LN-1 cells treated with chondroitinase had lower ability to adhere to ECM proteins compared to the control (Figure 2A, n=3, Student's *t*-test, $p<0.05$), whereas no significant effect was observed in the migration of SCC-9 LN-1 cells treated with chondroitinase (Figure 2B, n=3, Student's *t*-test, $p>0.05$).

Considering previous studies demonstrated the function of heparan sulfate modifications in drug uptake [28], we also evaluated the effect of chondroitinase treatment in the cisplatin cell resistance calculated by the IC50. For that, viability of the SCC-9, SCC-9 LN-1 and A431 cells were determined in the absence (vehicle) or presence of 1, 5, 7.5, 10, 25, 50 or 100 µM of cisplatin. The same curve was constructed with the addition of 0.1 U/ml of chondroitinase. The IC50 of SCC-9 LN-1, SCC-9 and A431 was 36.31 µM, 21.5 µM and 8.24 µM, respectively. When chondroitinase was added, SCC-9 LN-1, SCC-9 and A431 cell lines exhibited a drop in the IC50 (25.62 µM for SCC-9 LN-1, 16.88 µM for SCC-9 and 4.52 µM for A431), decreasing resistance to cisplatin in about 1.5, 1.3 and 1.8 fold for SCC-9 LN-1, SCC-9 and A431 cells, respectively (Figure 2C).

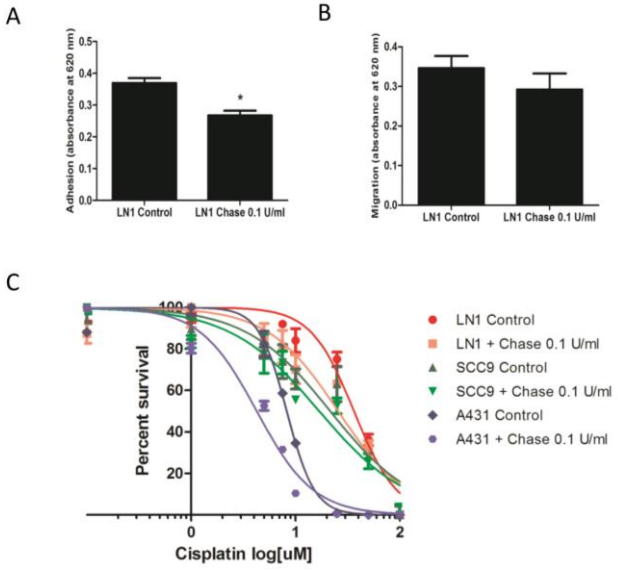

Figure 2 Treatment with chondroitinase ABC decreased SCC-9 LN-1 cell adhesion to extracellular matrix, but not SCC-9 LN-1 cell migration and increased sensibility of SCC-9 LN-1, SCC-9 and A431 cells to cisplatin. (A) SCC-9 LN-1 had a lower ability to adhere to extracellular matrix proteins (Matrigel) after treatment with 0.1 U/ml of chondroitinase for 4 h/37 °C in serum free media (n=3, triplicate, Student's *t*-test, * indicates $p<0.05$). (B) SCC-9 LN-1 was treated with 0.1 U/ml for 4h/37 °C in serum free media were

seeded in the upper chamber of 96-well transwell plates (n=3, triplicate). RPMI media, which was supplemented with 1% FBS, was added in the lower chamber. (C) SCC-9 LN-1**,** SCC-9 and A431 cells treated with increasing concentrations of cisplatin (0 – 100 µM) for 48 h in the presence of 0.1 U/ml of chondroitinase showed increased sensibility to cisplatin, calculated by a non-linear regression of a dose-response curves (log[µM cisplatin] vs normalized response). Data are expressed as means ± SD from one independent experiment.

Expression-associated phenotype of agrin and perlecan in adhesion, migration and proliferation event

In order to investigate the role of agrin and perlecan in oral carcinogenesis, we first evaluated the ability of the cells to adhere to extracellular matrix proteins. First, SCC-9, SCC-9 LN-1 and A431 cells lines were scramble-siRNA, agrin-siRNA or perlecan-siRNA transfected, and knockdown of agrin and perlecan were confirmed by qRT-PCR (Figure S2). We observed that targeted agrin knockdown decreased the adhesion of SCC-9 and SCC-9 LN-1 cells to the Matrigel (Figure 3 A and B, n=3, Student's *t*-test, $p<0.05$), but no significant effect was observed for agrin knockdown in A431 cells. The knockdown of perlecan decreased the adhesion of SCC-9 LN-1 and A431 cells (Figure 3 C and D, n=3, Student's *t*-test, $p<0.05$), but no significant effect was observed for perlecan knockdown in SCC-9 cells.

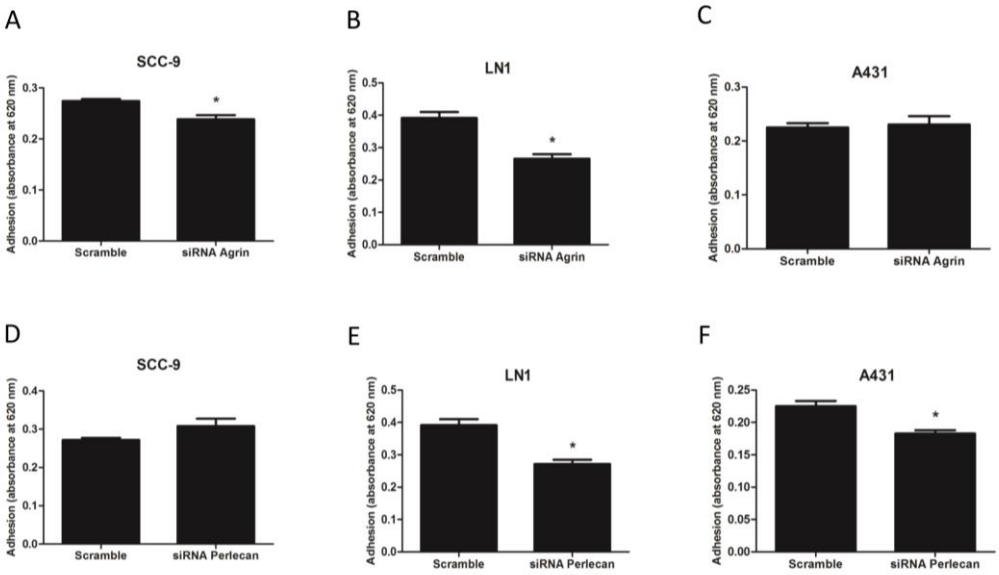

Figure 3 The role of agrin and perlecan in cell adhesion. Knockdown of agrin decreased adhesion to Matrigel in SCC-9 (n=2, A) and SCC-9 LN-1 (n=3, B), while no difference was observed in A431 (n=3, C). When perlecan was silenced by siRNA no difference was observed in SCC-9 adhesion to Matrigel (n=2, D) but a significant reduction was observed in SCC-9 LN-1 (n=3, E) and A431 (n=3, F) adhesion to Matrigel (Student's *t*-test, * indicates $p<0.05$).

The role of agrin and perlecan in cell migration was determined using transwell chamber. First, SCC-9, SCC-9 LN-1 and A431 cells treated with control siRNA (scramble) and siRNA against agrin or perlecan were plated in the upper chambers, and allowed to migrate towards the lower chamber containing medium supplemented with 1% FBS. For agrin silenced cells, migration was significantly diminished in SCC-9, SCC-9 LN-1 and A431 cells compared to scrambled control (Figure 4 A-C, n=3, Student's *t*-test, $p<0.001$). For perlecan silenced cells, migration was significantly diminished in SCC-9 LN-1 and A431 cells compared to scrambled control.

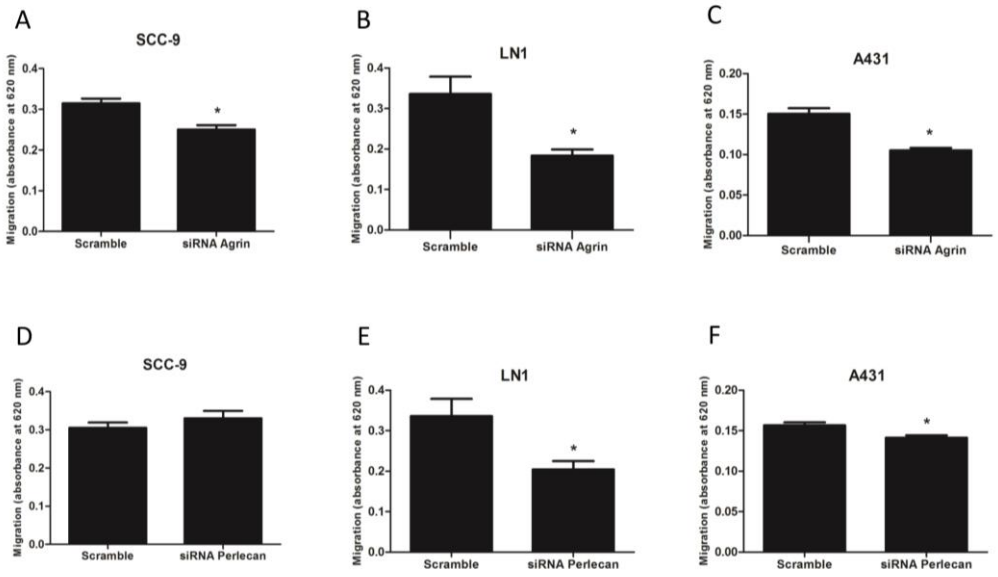

Figure 4 The role of agrin and perlecan in cell migration. Knockdown of agrin decreased migration of SCC-9 (n=2, A), SCC-9 LN-1 (n=3, B) and A431 (n=3, C) cell lines. When perlecan was silenced by siRNA no difference was observed in SCC-9 migration (n=2, D), but a significant reduction was observed in SCC-9 LN-1 (n=3, E) and A431 (n=3, F) migration (Student's *t*-test, * indicates $p<0.05$).

Cell viability was tested using MTT assay in the presence of 10% FBS, and we verified that SCC-9 and SCC-9 LN-1-agrin knockdown had a significant reduction in cell viability (Figure 5A, n=3, One-way ANOVA, followed by Tukey's test, $p<0.001$). In the A431 knockdown cells no significant difference was observed for agrin- or perlecan-siRNA, compared to the scrambled-siRNA (Figure 5B).

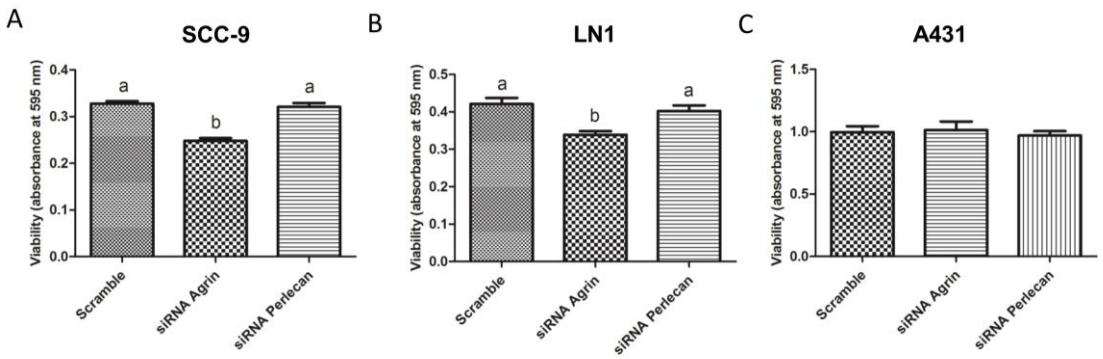

Figure 5 The role of agrin and perlecan in cell viability. The viability of SCC-9 (n=2, A) and SCC-9 LN-1 (n=3, B) was significantly reduced after siRNA-knockdown of agrin, but no difference in viability was observed in perlecan knockdown. The viability of A431 was not altered neither by agrin knockdown nor by perlecan knockdown (n=3, C) (One-way ANOVA followed by Tukey's test, different letters indicate statistically difference at $p<0.05$).

The role of agrin and perlecan in cisplatin cell resistance

We sought to determine the cisplatin cell resistance in SCC-9, SCC-9 LN-1 and A431 when agrin or perlecan was silenced in these cell lines. It was observed that the siRNA-knockdown of agrin promoted a reduction in the cisplatin cell resistance in all cell lines used in this study: SCC-9 (2.4 fold), SCC-9 LN-1 (3.8 fold) and A431 (1.7 fold) (Figure S3). On the other hand, when perlecan was silenced, no significant difference was observed in cisplatin cell resistance for SCC-9, SCC-9 LN-1 and A431 cell lines compared with scrambled (Figure S3).

At a concentration of 10 μM of cisplatin, SCC-9, SCC-9 LN-1 and A431 cell lines showed a significant reduction in cell viability when agrin expression was silenced (Figure 6 A and B, n=3, ANOVA followed by Tukey's test, $p<0.05$). A significant reduction in cell viability was also observed just in A431 perlecan-knockdown cells treated with 10 μM of cisplatin (Figure 6C, n=3, ANOVA followed by Tukey's test, $p<0.05$).

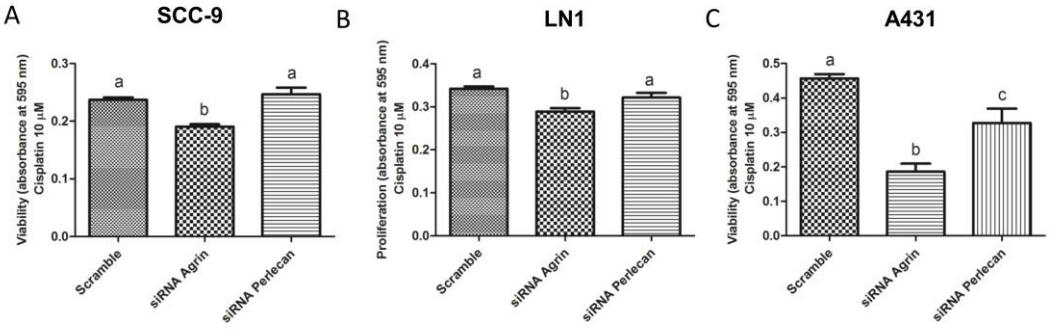

Figure 6 Viability of SCC-9 LN-1 and A431 cell lines in 10 µM of cisplatin treatment. At 10 µM of cisplatin, the viability of SCC-9 (n=3, A), SCC-9 LN-1 (n=3, B) and A431 (n=3C) was significantly reduced when agrin was silenced. Only in A431 cell line the same result of decreasing in viability was observed for perlecan siRNA-knockdown (C) (One-way ANOVA followed by Tukey's test, different letters indicate statistically difference at $p<0.05$).

DISCUSSION

Proteoglycans (PGs), essential macromolecules of the tumor microenvironment, have their expression altered during malignant transformation and tumor progression [9]. Agrin and perlecan are two of the major HSPG identified in the basement membrane, and their functional roles in modulation of cancer growth have been reviewed elsewhere [22]. In this study, we showed for the first time that agrin and perlecan are highly expressed in OSCCs, and the function of these proteins in oral cancer associated processes was investigated. Not only the expression of agrin and perlecan was shown to be higher in OSCC tissues compared to control tissues, but also their expression might be associated with different sites of origin, where higher expression of agrin was found in cell line originated from primary site (SCC-9), whereas higher expression of perlecan was found in cell line originated from metastatic site (SCC-9 LN-1). The spread of cancer cells from a primary tumor to form metastases at distant sites is a complex process that remains poorly defined [29], but it has been reported to involve detachment of cells from the tumor tissue, regulation of cell motility and invasion, proliferation and evasion through the lymphatic system or blood vessels [30]. Efforts have been made to elucidate tumor-related proteins that could influence the appearance of metastases in oral squamous cell carcinoma [31-33], which occur in about 40% of patients with oral cancer [34]. Therefore, the study of altered molecules in cell lines originated from distinct sites is essential to understand the molecular basis of this process.

GAGs polysaccharide chains are the main contributors to the proteoglycan functional properties and essential part of the matured proteoglycan molecules [35]. Besides, it was reported that glycans play a crucial role at various pathophysiological steps of cancer progression [36], especially by acting as co-receptors to stabilize growth-factor receptor signaling complexes and enhancing integrin-mediated cell adhesion, motility and intracellular signaling. The disruption of GAGs modification by heparanase was shown to facilitate tumor cell invasion [37] angiogenesis and metastasis [38]. In order to further understand the role of chondroitin sulfate modification in oral cancer, SCC-9 LN-1 cell lines were treated with chondroitinase and tested in adhesion and migration processes. Interestingly, when SCC-9 LN-1 cells were treated with chondroitinase, they had reduced ability to adhere to extracellular matrix proteins. It is important to mention that this event may be a result of the disruption of many cell surface chondroitin sulfate proteoglycans, such as syndecans, chondroitin sulfate proteoglycan 4, betaglycan, neuropilin-1, receptor protein tyrosine phosphatase, integrin and VEGFR-2, which were also previously demonstrated to be overexpressed in cancer [39].

It is well established that based upon their direct involvement in cell–cell and cell–ECM interactions, PGs have been strongly implicated in the regulation of cell movement [40]. However, how PGs actually affect this process is only partially understood and in some instances, controversial. In this study, we have demonstrated that agrin and perlecan play a role in the oral cancer cell movement by silencing agrin and perlecan, which promoted a strongly reduced in the ability of SCC-9 and SCC-9 LN-1 cell line to migrate and to adhere to matrigel.

Perlecan has been associated with the induction of cellular proliferation, differentiation and angiogenesis by interacting with a number of growth factors including FGFs 1, 2, 7, 9, and 18; hepatocyte growth factor, platelet derived growth factors-AA and -BB, and VEGF [14,41]. Perlecan also exhibits adhesive [42] or anti-adhesive [43] properties presumably by differentially affecting surface receptors such as α2β1 integrin [44,45]. Agrin is also able to interact with integrins [46], however there is still very limited acknowledgement on how agrin signals through the integrin receptors and how these interactions influence cell behavior.

Therapeutic PGs- and GAG-targeting modifications have been considered as anti-invasion and tumor-specific drug delivery potential approaches [47]. It was reported that both heparan sulfate and chondroitin sulfate modifications are able to interact with cisplatin and mediate entry-pathway [28,48]. Platinum-based chemotherapy has been used for treating a wide variety of solid tumors, including lung, head and neck, ovarian, cervical, and testicular cancers for over three decades [49]. However, the emergence of drug resistance may limit the effectiveness of platinating agents in solid tumors [50-52], including OSCCs [53]. Recently, novel extracellular matrix cisplatin-resistant biomarkers from epithelial ovarian carcinoma (EOC) were identified using secretome analysis from EOC cell lines [54]. In our study, we showed new evidences of increasing cisplatin-sensibility by disrupting chondroitin sulfate modification, or reduction of agrin and perlecan proteins levels, suggesting it may have a potential target for therapeutic intervention.

In summary, we have identified a relevant role of agrin and perlecan in oral cancer cell adhesion, migration and cisplatin cell resistance, opening new perspectives for further investigations and targeting innovative and/or complementary therapeutic strategies.

ACKNOWLEDGMENTS

This work was supported by FAPESP Grants 2009/54067-3, 2010/19278-0, 2011/22421-2, and CNPq Grants 470549/2011-4, 301702/2011-0 and 470268/2013-1. We are thankful to Rafael Canevarolo from Laboratório Nacional de Biociências for suggesting some experiments.

REFERENCES

1. Argiris A, Karamouzis MV, Raben D, Ferris RL (2008) Head and neck cancer. Lancet 371: 1695-1709.
2. Belcher R, Hayes K, Fedewa S, Chen AY (2014) Current treatment of head and neck squamous cell cancer. J Surg Oncol.
3. Mydlarz WK, Hennessey PT, Califano JA (2010) Advances and Perspectives in the Molecular Diagnosis of Head and Neck Cancer. Expert Opin Med Diagn 4: 53-65.
4. da Silva SD, Ferlito A, Takes RP, Brakenhoff RH, Valentin MD, et al. (2011) Advances and applications of oral cancer basic research. Oral Oncol 47: 783-791.
5. Pereira MC, Oliveira DT, Landman G, Kowalski LP (2007) Histologic subtypes of oral squamous cell carcinoma: prognostic relevance. J Can Dent Assoc 73: 339-344.
6. van Kempen LC, Ruiter DJ, van Muijen GN, Coussens LM (2003) The tumor microenvironment: a critical determinant of neoplastic evolution. Eur J Cell Biol 82: 539-548.
7. Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, et al. (2012) The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Mol Cell Proteomics 11: M111 014647.
8. Barash U, Cohen-Kaplan V, Dowek I, Sanderson RD, Ilan N, et al. (2010) Proteoglycans in health and disease: new concepts for heparanase function in tumor progression and metastasis. FEBS J 277: 3890-3903.
9. Theocharis AD, Skandalis SS, Tzanakakis GN, Karamanos NK (2010) Proteoglycans in health and disease: novel roles for proteoglycans in malignancy and their pharmacological targeting. FEBS J 277: 3904-3923.
10. Iozzo RV, Sanderson RD (2011) Proteoglycans in cancer biology, tumour microenvironment and angiogenesis. J Cell Mol Med 15: 1013-1031.
11. Edwards IJ (2012) Proteoglycans in prostate cancer. Nat Rev Urol 9: 196-206.
12. Tatrai P, Dudas J, Batmunkh E, Mathe M, Zalatnai A, et al. (2006) Agrin, a novel basement membrane component in human and rat liver, accumulates in cirrhosis and hepatocellular carcinoma. Lab Invest 86: 1149-1160.
13. Mundhenke C, Meyer K, Drew S, Friedl A (2002) Heparan sulfate proteoglycans as regulators of fibroblast growth factor-2 receptor binding in breast carcinomas. Am J Pathol 160: 185-194.
14. Sharma B, Handler M, Eichstetter I, Whitelock JM, Nugent MA, et al. (1998) Antisense targeting of perlecan blocks tumor growth and angiogenesis in vivo. J Clin Invest 102: 1599-1608.
15. Savore C, Zhang C, Muir C, Liu R, Wyrwa J, et al. (2005) Perlecan knockdown in metastatic prostate cancer cells reduces heparin-binding growth factor responses in vitro and tumor growth in vivo. Clin Exp Metastasis 22: 377-390.
16. Iozzo RV, Cohen IR, Grassel S, Murdoch AD (1994) The biology of perlecan: the multifaceted heparan sulphate proteoglycan of basement membranes and pericellular matrices. Biochem J 302 (Pt 3): 625-639.
17. Noonan DM, Fulle A, Valente P, Cai S, Horigan E, et al. (1991) The complete sequence of perlecan, a basement membrane heparan sulfate proteoglycan, reveals extensive similarity with laminin A chain, low density lipoprotein-receptor, and the neural cell adhesion molecule. J Biol Chem 266: 22939-22947.
18. Murdoch AD, Liu B, Schwarting R, Tuan RS, Iozzo RV (1994) Widespread expression of perlecan proteoglycan in basement membranes and extracellular matrices of human tissues

as detected by a novel monoclonal antibody against domain III and by in situ hybridization. J Histochem Cytochem 42: 239-249.
19. Mishra M, Chandavarkar V, Naik VV, Kale AD (2013) An immunohistochemical study of basement membrane heparan sulfate proteoglycan (perlecan) in oral epithelial dysplasia and squamous cell carcinoma. J Oral Maxillofac Pathol 17: 31-35.
20. Ida-Yonemochi H, Ikarashi T, Nagata M, Hoshina H, Takagi R, et al. (2002) The basement membrane-type heparan sulfate proteoglycan (perlecan) in ameloblastomas: its intercellular localization in stellate reticulum-like foci and biosynthesis by tumor cells in culture. Virchows Arch 441: 165-173.
21. Ikarashi T, Ida-Yonemochi H, Ohshiro K, Cheng J, Saku T (2004) Intraepithelial expression of perlecan, a basement membrane-type heparan sulfate proteoglycan reflects dysplastic changes of the oral mucosal epithelium. J Oral Pathol Med 33: 87-95.
22. Iozzo RV, Zoeller JJ, Nystrom A (2009) Basement membrane proteoglycans: modulators Par Excellence of cancer growth and angiogenesis. Mol Cells 27: 503-513.
23. Winzen U, Cole GJ, Halfter W (2003) Agrin is a chimeric proteoglycan with the attachment sites for heparan sulfate/chondroitin sulfate located in two multiple serine-glycine clusters. J Biol Chem 278: 30106-30114.
24. Somoracz A, Tatrai P, Horvath G, Kiss A, Kupcsulik P, et al. (2010) Agrin immunohistochemistry facilitates the determination of primary versus metastatic origin of liver carcinomas. Hum Pathol 41: 1310-1319.
25. Batmunkh E, Tatrai P, Szabo E, Lodi C, Holczbauer A, et al. (2007) Comparison of the expression of agrin, a basement membrane heparan sulfate proteoglycan, in cholangiocarcinoma and hepatocellular carcinoma. Hum Pathol 38: 1508-1515.
26. Agostini M, Almeida LY, Bastos DC, Ortega RM, Moreira FS, et al. (2014) The Fatty Acid synthase inhibitor orlistat reduces the growth and metastasis of orthotopic tongue oral squamous cell carcinomas. Mol Cancer Ther 13: 585-595.
27. Aragao AZ, Belloni M, Simabuco FM, Zanetti MR, Yokoo S, et al. (2012) Novel processed form of syndecan-1 shed from SCC-9 cells plays a role in cell migration. PLoS One 7: e43521.
28. Silva H, Frezard F, Peterson EJ, Kabolizadeh P, Ryan JJ, et al. (2012) Heparan sulfate proteoglycan-mediated entry pathway for charged tri-platinum compounds: differential cellular accumulation mechanisms for platinum. Mol Pharm 9: 1795-1802.
29. Terp MG, Lund RR, Jensen ON, Leth-Larsen R, Ditzel HJ (2012) Identification of markers associated with highly aggressive metastatic phenotypes using quantitative comparative proteomics. Cancer Genomics Proteomics 9: 265-273.
30. Noguti J, De Moura CF, De Jesus GP, Da Silva VH, Hossaka TA, et al. (2012) Metastasis from oral cancer: an overview. Cancer Genomics Proteomics 9: 329-335.
31. O'Donnell RK, Kupferman M, Wei SJ, Singhal S, Weber R, et al. (2005) Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. Oncogene 24: 1244-1251.
32. Chang KP, Yu JS, Chien KY, Lee CW, Liang Y, et al. (2011) Identification of PRDX4 and P4HA2 as metastasis-associated proteins in oral cavity squamous cell carcinoma by comparative tissue proteomics of microdissected specimens using iTRAQ technology. J Proteome Res 10: 4935-4947.
33. Chai YD, Zhang L, Yang Y, Su T, Charugundla P, et al. (2014) Discovery of potential serum protein biomarkers for lymph-node metastasis in oral cancer. Head Neck.
34. Feller L LJ (2012) Oral Squamous Cell Carcinoma: Epidemiology, Clinical Presentation and Treatment. Journal of Cancer Therapy 3: 263-268.

35. Afratis N, Gialeli C, Nikitovic D, Tsegenidis T, Karousou E, et al. (2012) Glycosaminoglycans: key players in cancer cell biology and treatment. FEBS J 279: 1177-1197.
36. Fuster MM, Esko JD (2005) The sweet and sour of cancer: glycans as novel therapeutic targets. Nat Rev Cancer 5: 526-542.
37. Vlodavsky I, Friedmann Y (2001) Molecular properties and involvement of heparanase in cancer metastasis and angiogenesis. J Clin Invest 108: 341-347.
38. Goldshmidt O, Zcharia E, Abramovitch R, Metzger S, Aingorn H, et al. (2002) Cell surface expression and secretion of heparanase markedly promote tumor angiogenesis and metastasis. Proc Natl Acad Sci U S A 99: 10031-10036.
39. Pantazaka E, Papadimitriou E (2014) Chondroitin sulfate-cell membrane effectors as regulators of growth factor-mediated vascular and cancer cell migration. Biochim Biophys Acta 1840: 2643-2650.
40. Cattaruzza S, Perris R (2005) Proteoglycan control of cell movement during wound healing and cancer spreading. Matrix Biol 24: 400-417.
41. Zhou Z, Wang J, Cao R, Morita H, Soininen R, et al. (2004) Impaired angiogenesis, delayed wound healing and retarded tumor growth in perlecan heparan sulfate-deficient mice. Cancer Res 64: 4699-4702.
42. Whitelock JM, Graham LD, Melrose J, Murdoch AD, Iozzo RV, et al. (1999) Human perlecan immunopurified from different endothelial cell sources has different adhesive properties for vascular cells. Matrix Biol 18: 163-178.
43. Klein G, Conzelmann S, Beck S, Timpl R, Muller CA (1995) Perlecan in human bone marrow: a growth-factor-presenting, but anti-adhesive, extracellular matrix component for hematopoietic cells. Matrix Biol 14: 457-465.
44. Bix G, Iozzo RV (2005) Matrix revolutions: "tails" of basement-membrane components with angiostatic functions. Trends Cell Biol 15: 52-60.
45. Hayashi K, Madri JA, Yurchenco PD (1992) Endothelial cells interact with the core protein of basement membrane perlecan through beta 1 and beta 3 integrins: an adhesion modulated by glycosaminoglycan. J Cell Biol 119: 945-959.
46. Martin PT, Sanes JR (1997) Integrins mediate adhesion to agrin and modulate agrin signaling. Development 124: 3909-3917.
47. Yip GW, Smollich M, Gotte M (2006) Therapeutic value of glycosaminoglycans in cancer. Mol Cancer Ther 5: 2139-2148.
48. Zhang JS, Anraku M, Kadowaki D, Imai T, Suenaga A, et al. (2011) Spectroscopic studies of interactions of chondroitin sulfates with cisplatin. Carbohydr Res 346: 631-637.
49. Rabik CA, Dolan ME (2007) Molecular mechanisms of resistance and toxicity associated with platinating agents. Cancer Treat Rev 33: 9-23.
50. Galluzzi L, Senovilla L, Vitale I, Michels J, Martins I, et al. (2012) Molecular mechanisms of cisplatin resistance. Oncogene 31: 1869-1883.
51. Dasari S, Bernard Tchounwou P (2014) Cisplatin in Cancer therapy: Molecular mechanisms of action. Eur J Pharmacol.
52. Barr MP, Gray SG, Hoffmann AC, Hilger RA, Thomale J, et al. (2013) Generation and characterisation of cisplatin-resistant non-small cell lung cancer cell lines displaying a stem-like signature. PLoS One 8: e54193.
53. Eberle KE, Sansing HA, Szaniszlo P, Resto VA, Berrier AL (2011) Carcinoma matrix controls resistance to cisplatin through talin regulation of NF-kB. PLoS One 6: e21496.
54. Teng PN, Wang G, Hood BL, Conrads KA, Hamilton CA, et al. (2014) Identification of candidate circulating cisplatin-resistant biomarkers from epithelial ovarian carcinoma cell secretomes. Br J Cancer 110: 123-132.

CAPÍTULO 4

Deciphering the role of the ADAM17-enriched secretome in cell signaling

Rebeca Kawahara; Renato Niyama Lima; Romênia R. Domingues; Bianca Alves Pauletti; Michelle Assis; Ana Carolina Migliorini Figueira; Adriana Franco Paes Leme

JOURNAL OF PROTEOME RESEARCH, v. 13, n.4, p.2080-93, 2014

(doi: 10.1021/pr401224u)

ABSTRACT

ADAM17 has been initially identified as the main sheddase responsible for releasing the soluble form of a variety of cell-surface proteins, including growth factors, cytokines, cell adhesion molecules and receptors, most of which are associated with the pathological processes, including cancer and inflammation. However, the function and composition of ADAM17-dependent secretome on a proteome-wide scale is poorly understood. In this study, we observed that the ADAM17-dependent secretome plays an important role in promoting cell proliferation and migration. To further demonstrate the repertoire of proteins involved in this cross-talk, we employed mass spectrometry-based proteomics using non-metabolic and metabolic labeling approaches to explore the secretome composition of wild-type and ADAM17-/- knockout mouse embryonic fibroblasts (mEFs) cells. Bioinformatic analyses indicated the differential regulation of 277 soluble proteins in ADAM17-dependent secretome as well as novel direct ADAM17 cleavage substrates, such as mimecan and perlecan. Furthermore, we found that the ADAM17-dependent secretome promoted an opposite regulation of ERK and FAK pathways as well as PPARγ downstream activation. These findings demonstrated fine-tuning of cell signaling rendered by the soluble molecules mediated by ADAM17.

Keywords: ADAM17, secretome, label-free, SILAC, signaling, proliferation, migration

INTRODUCTION

Proteolysis of membrane and extracellular matrix proteins contribute to the bioavailability of soluble factors in the extracellular media and metalloproteases are particularly important in this process [1,2]. However, the physiological and pathological substrates of the contributing proteases have remained incompletely elucidated. Thus, many efforts have been put forth to identify the substrate repertoire and cleavage sites that indirectly or directly characterize the function of metalloproteases [3-9].

ADAM17 (known as tumor necrosis factor-α converting enzyme, TACE) is a zinc metalloproteinase that mediates shedding of multiple cell surface proteins. It has been proposed that TACE is critical for the shedding of a large number of cell surface proteins, including several growth factors of the epidermal growth factor (EGF) family (transforming growth factor-α (TGF-α), heparin binding EGF-like growth factor (HBEGF), amphiregulin), in addition to transforming growth factor-R, TNF receptors I and II, adhesion molecules (VCAM, ALCAM, NCAM, CD44), among others [10,11]. Many of these ectodomains have critical roles in carcinogenesis, inflammation, and other pathologies [11-13]. Through the release of soluble growth factors, a variety of signaling pathways has been reported to be modulated by ADAM-17 such as EGFR [14-17], ERK/MAPK[18], IL-6R [19,20] and TGFβ [21] pathways.

Using ADAM17 silencing, some studies have demonstrated the role of this enzyme in promoting cell proliferation and migration in cancer cell lines [15,22-24]. Nevertheless, the broad ADAM17 effect on the set of molecules present in the extracellular media on a proteome-wide scale as well as their role in cell signaling and biological events is poorly understood.

The secretome of a cell consists of soluble, secreted proteins and the membrane ectodomains that are proteolytically released by sheddases [25]. The secretome is considered critical mediators of the cell communication and organization, which not only directly affect the target cell phenotype but can also modify communication between cells and the extracellular environment [26,27].

In this study, considering the observation that the ADAM17-dependent secretome plays an important role in melanoma migration and mouse embryonic fibroblast (mEF) cells proliferation, we used in-depth secretome analysis to gain insights into the soluble proteins involved in these events. To achieve this, we used label-free and SILAC quantitative proteomic approaches on wild-type and ADAM17-/- knockout mEF secretomes. Using this strategy, we identified multiple changes in the extracellular proteome that were induced by ADAM17. Downstream effects

inferred by the secretome network analysis revealed important opposite regulation of ERK and FAK pathways as well as PPARγ transcription factor activation. Finally, the ADAM17-dependent secretome analysis enabled the identification of mimecan and perlecan as novel direct ADAM17 substrates.

EXPERIMENTAL SECTION

Cell lines and SILAC cell culture

SV40-transformed mEF (mouse embryonic fibroblast) cells were cultured in DMEM supplemented with 10% FBS and antibiotics at 37°C in 5% CO_2 atmosphere [28]. Wild-type mEF cells (WT) and ADAM17$^{-/-}$ mEF cells (KO) were kindly provided by Dr. Carl Blobel (Weill Medical College of Cornell University). Human Epidermoid Carcinoma A-431 (an epidermoid carcinoma cell line originated from skin) and B16-F10 melanoma murine cell lines were grown in Roswell Park Memorial Institute (RPMI) −1640 medium supplemented with 10% FBS and antibiotics at 37°C in a 5% CO_2 air atmosphere.

For SILAC labeling, wild-type cells were grown in Dulbecco's modified Eagle's medium containing 0.398 mM of Heavy L-Arg-$^{13}C_6{}^{15}N_4$ (Cambridge Isotope Laboratories) and 0.798 mM of L-Lys -$^{13}C_6{}^{15}N_2$, (Cambridge Isotope Laboratories) and ADAM17-/- mEFs were cultured in DMEM unlabeled with Arg and Lys natural amino acids (Light) supplemented with 10% dialyzed fetal bovine serum (FBS; Invitrogen), for at least five doubling times and tested for full incorporation.

Proliferation assay

WT and KO mEFs cells were seeded onto 96-well plates and incubated for two days. MTT (12 mM tetrazolium 3-(4,5-dimethylthiazol-2)-2,5-diphenyltetrazolium bromide) was added, and the cells were incubated for 4 h at 37°C, in the dark. The media were removed, and 100 µl of HCl 1 N and isopropanol (1:25) was added into each well and incubated for 15 min at room temperature under gentle agitation. Finally, the absorbance was measured at 595 nm. Three independent experiments were performed in triplicate.

Transwell migration assay

WT and KO mEFs cells were plated in the upper chambers of 8 mm pore transwells (HTS Transwell 96-Well Plate, Corning) after a starvation period of 16 h. The cells were allowed to

migrate towards the lower chamber containing different extracellular media conditions. At the end of the assay, the cells at the top chamber were removed using a cotton swab, and the cells at the bottom of the insert filter were fixed with 10% formaldehyde for 10 min, washed with PBS and stained with 1% toluidine blue solution in 1% borax for 5 min. The dye was eluted using 1% SDS, and the absorbance was measured at 620 nm. Three independent experiments were performed in triplicate.

Sample preparation for MS

Cells were grown at 80% confluence, gently washed three times in phosphate-buffered saline (PBS) and incubated for 24 h in serum-free medium. The media were collected, and a final concentration of 1 mM EDTA and 1 mM PMSF were added to the media. Cell debris was eliminated by centrifugation at 4,000 x *g* for 5 min at 4°C and subsequently concentrated using a 3000-Dalton centrifugation filter (Millipore, Billerica, MA) at 4,000 x *g* at 4°C. Protein concentrations were determined using a Bradford assay (Bio-Rad, Hercules, CA, USA). Proteins (80 µg) were treated with a final concentration of 1.6 M urea, following reduction (5 mM dithiotreitol, 25 min at 56°C), alkylation (14 mM iodoacetamide, 30 min at room temperature in the dark) and digestion with trypsin (1:50, w/w). The reaction was stopped with 0.4% TFA and desalted using Sep-pack cartridges (Waters). The samples were then dried in a vacuum concentrator and reconstituted in 0.1% formic acid. Three biological replicates were performed using the label-free method, and one biological replicate with three technical runs were performed using the SILAC method.

Mass spectrometric analysis

Nanoflow nLC-MS/MS: An aliquot containing 2.2 µg of proteins was analyzed on an ETD-enabled LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) connected to a nanoflow liquid chromatography (LC-MS/MS) instrument by an EASY-nLC system (Proxeon Biosystem) with a Proxeon nanoelectrospray ion source. Peptides were separated with a 2-90% acetonitrile gradient in 0.1% formic acid using an analytical column PicoFrit Column (20 cm x ID75 µm, 5 µm particle size, New Objective) at a flow of 300 nl/min over 194 min. The nanoelectrospray voltage was set to 2.2 kV and the source temperature was 275°C. All of the instrument methods were set up in the data-dependent acquisition mode. The full scan MS spectra (m/z 300-1600) were acquired in the Orbitrap analyzer after accumulation to a target value of $1x10^6$. The resolution in the Orbitrap was set to r= 60,000 and the 20 most intense peptide ions with

charge states ≥ 2 were sequentially isolated to a target value of 5,000 and fragmented in the linear ion trap using low-energy CID (normalized collision energy of 35%). The signal threshold for triggering an MS/MS event was set to 500 counts. Dynamic exclusion was enabled with an exclusion size list of 500, exclusion duration of 60 s, and a repeat count of 1. An activation q= 0.25 and activation time of 10 ms were used.

Protein identification and quantitative analysis

The raw files were processed using the MaxQuant version 1.2.7.4 [29] and the MS/MS spectra were searched using the Andromeda search engine [30] against the Uniprot Mouse Protein Database (release July 11th, 2012; 69,711 entries). The initial maximal allowed mass tolerance was set to 20 ppm for precursor and then it was set to 6 ppm in the main search and to 0.5 Da for fragment ions. Enzyme specificity was set to trypsin with a maximum of two missed cleavages. Carbamidomethylation of cysteine (57.021464 Da) was set as a fixed modification, and oxidation of methionine (15.994915 Da) and protein N-terminal acetylation (42.010565 Da) were selected as variable modifications. The minimum peptide length was set to 6 amino acids. For SILAC data analyses, the same parameters were used including heavy label Arg10 and Lys8. For protein quantification, a minimum of two ratio counts was set and the 'requantify' and 'match between runs' functions were enabled. The false discovery rates (FDRs) of peptide and protein were both set to 0.01.

Bioinformatic analyses were performed using Perseus v.1.2.7.4 [29], which is available in the MaxQuant environment. First, reverse and contaminant entries were excluded from further analysis. Label-free quantification was performed using the normalized spectral protein intensity (LFQ intensity). Data obtained from three independent experiments from each condition were annotated into the WT and KO groups. For analysis of differentially expressed proteins, the data were converted into log2 and Student's *t*-test analysis was applied on the WT and KO groups. For the SILAC approach, the protein ratio intensity between Heavy and Light were used to compare differential protein expression in the secretomes from WT and KO cells. The protein ratios were calculated from the median of all normalized peptide ratios using only unique peptides or peptides assigned to the protein group with the highest number of peptides (razor peptides).

All mass spectrometric raw files and search parameter settings associated with this study are available for downloading via FTP from the PeptideAtlas data repository by accessing the following link: http://www.peptideatlas.org/PASS/ PASS00318.

Bioinformatic analysis

Differentially expressed proteins from label-free experiments were uploaded onto the Ingenuity Pathways (IPA; Ingenuity Systems, Redwood City, CA) Knowledge Base as a tab-delimited text file of gene names. Biological networks were generated using their Knowledge Base for interactions between mapped Focus Genes (user's list) and all other gene objects stored in the knowledge base. A detailed description of IPA can be found on the Ingenuity Systems website (www.ingenuity.com).

Transcription factors likely to regulate gene expression were identified using the Expression2kinases (X2K) software [31] Chip-X database.

Immunoblotting

Protein samples were subjected to 10% SDS-PAGE under reduction conditions and transferred to nitrocellulose membranes. The membranes were blocked in 5% dry milk in Tris-Tween-buffered saline (TTBS). The membranes were then incubated overnight at 4°C with the following antibodies: anti-basement heparin sulfate proteoglycan 2 (1:1000, Abcam), anti-mimecan (1:1000, RD system), anti-Erk (1:1,000; Santa Cruz Biotech), anti-phospho-Erk (1:1,000; Santa Cruz Biotech), anti-FAK (1:1,000; Santa Cruz Biotech), and anti-phospho-FAK (1:1,000; Santa Cruz Biotech). The membranes were washed and incubated in horseradish peroxidase-conjugated secondary antibodies and developed using enhanced chemiluminescence detection according to the manufacturer's instructions (Amersham Biosciences). The band densitometry was analyzed using ImageJ software (http://rsb.info.nih.gov/ij/).

Generation of the stably ADAM17 knockdown A-431 cell line

The lentiviral particle production and transduction of A-431 cells for ADAM17 knockdown were achieved with the Mission® shRNA Vector pLKO.1-puro System using shRNA plasmid pLKO.1-Neo-CMV-tGFP (Sigma-Aldrich, Munich, Germany). The experimental procedures were performed according to the manufacturers' instructions. After transduction, G418 antibiotic was added to the cultures at a final concentration of 0.8 mg/ml and incubated for approximately one week until the complete death of untransfected cells.

Real-time Quantitative PCR

Total RNA was obtained using the TRIzol reagent (Invitrogen Corporation) and 2 µg of total RNA were used for retro-transcription using a First-Strand cDNA Synthesis Kit (GE

Healthcare). Real-time quantitative PCR for mimecan and perlecan was performed using the SYBRH Green PCR Master Mix (Applied Biosystems), and the dissociation curves were performed to confirm the specificity of the products. The threshold cycle (CT) values of the targeted gene were normalized relative to glyceraldehydes-3-phospate dehydrogenase gene, and relative expression ratios were calculated using the 2-ΔΔ Ct method. Three independent experiments were performed with triplicates. The following PCR primers were used: Mimecan forward 5' - GGACCATAACGACCTGGAATC-3' and reverse 5'-CTCCCGAATGTAACGAGTGTC-3'; Perlecan forward 5'- AAT GCGCTGGACACATTCG-3' and reverse 5'- ATTCACCAGGGCTCGGAAATA-3'; GAPDH forward 5'-GAAGGTGAAGGTCGGAGTCAAC-3' and reverse 5'- CAGAGTTAAAA GCAGCCCCTGGT-3'.

Slot blot

A-431 scramble and knockdown for ADAM17 cells were washed three times in phosphate-buffered saline (PBS) and incubated for 24 h in the serum-free medium. The media were collected, and a final concentration of 1 mM PMSF and 1 mM EDTA was added to the media. Briefly, cell debris were eliminated by centrifugation at 4,000×*g* for 5 min at 4°C and subsequently concentrated using a 3,000 Da centrifugal filter (Millipore, Billerica, MA) at 4,000×*g* at 4°C. Conditioned media (20 µg) from A-431 scramble and knockdown for ADAM17 were applied to the nitrocellulose membrane under vacuum in an immunodot apparatus (Slot blot, Amersham Biosciences, Freiburg, Germany). The membranes were blocked for 1 h with 5% dry milk in TTBS. Human basement heparan sulfate proteoglycan were evaluated by incubating the membranes overnight at 4°C with anti-basement heparan sulfate proteoglycan (1:1000, Abcam). The loading control was evaluated using Ponceau S staining.

PPARγ transcription assay

Transactivation assay was performed as described [32] under some modifications. WT and KO mEF cells were seeded onto 24-well plates at a density of 1×10^5 cells/well and grown in DMEM (supplemented with 10% FBS and antibiotics) under 5% CO_2 at 37°C. Cell transfections used the following plasmids were CMV-PPAR, full-length PPAR cloned upstream of the CMV promoter; DR1-LUC; PPAR response element (DR-1) followed by *Firefly* luciferase gene reporter; pRL, transfection control of *Renilla* luciferase cloned upstream of the CMV promoter; and pBlue, the empty plasmid pBluescript as a control.

After 24 h, 400 ng of each plasmid (full-length PPAR or empty plasmid pBlue, DR1-LUC, pRL) were mixed with Lipofectamine 2000 (Invitrogen), and then transfected, following the manufacturer's protocol. The plasmid pRL was added into all of the conditions as a control of transfection efficiency. Four hours later, Rosiglitazone was added at 10 µM/well. On the next day, the cells were lysed using the Kit Dual-Luciferase Assay System (Promega) and the luciferase activities were measured using the GLOMAX multi-detection system (Promega). Firefly Luciferase activities were normalized using Renilla Luciferase activities. Each condition was performed in triplicate. The results were reported as the fold induction of normalized luciferase activity ± standard error of the mean.

Zymography

The conditioned media subject to MS analysis were also evaluated on its gelatinolytic activity. Protein samples (10 µg) from WT and KO mEF cells were subjected to 1-D electrophoresis on 12% SDS-polyacrylamide gels containing 1 mg/ml gelatin under nonreducing conditions, and the gelatinolytic activity was determined as previously described [33]. Gels were stained with Coomassie blue and destained. Gelatin digestion was identified as clear bands against a blue background. The band densitometry was analyzed using ImageJ software (http://rsb.info.nih.gov/ij/)

Analysis of potential substrate cleavage by ADAM17

For cleavage site analysis, recombinant mimecan (residue from Ala20 to Phe298, R&D system) and perlecan (residue from Arg3684 to Ser4391, which correspond the C-terminal of the heparan sulfate proteoglycan perlecan named endorepellin, R&D system) were incubated with or without recombinant 0.01 µg ADAM17 (residue Arg215-Asn671, R&D Systems) in PBS at a final volume of 20 µl for 16 h at 37°C. Neo-amino termini of perlecan were labeled with 20 mM of EZ-Link Sulfo-NHS-SS-biotin (Pierce Chemical, Rockford, IL) for 2 h and neo-amino termini of mimecan were labeled with 30 mM formaldehyde and 15 mM cyanoborohydride sodium for 24 h [34]. Both strategies were stopped by the addition of 50 mM Tris-HCl, pH 7.5 and subjected to 10% SDS-PAGE under reducing conditions. The gel was silver stained, and the cleavage product bands were removed and subjected to an in-gel trypsin digestion protocol [35]. The resulting peptides were subsequently analyzed using LC-MS/MS as previously described, using a gradient over 45 min.

Peak lists (msf) were generated from the raw data files using Proteome Discoverer version 1.3 (Thermo Fisher Scientific) with the Sequest search engine and searched against Uniprot mouse mimecan sequence (Q62000). Searches were performed using the following parameters:

semi-C-terminal trypsin cleavage specificity with up to two missed cleavages, cysteine carbamidomethylation and dimethylation of lysine ε-amines (+28.03130 Da) as fixed modifications, and N-terminal dimethylation (+28.03130 Da) and methionine oxidation as variable modifications. For perlecan analysis, the search was performed against the Uniprot human perlecan sequence (P98160). Trypsin was defined to cleave semi-C-terminal, cysteine carbamidomethylation and biotinylation of lysine ε-amines (+88.00 Da) as fixed modifications, and N-terminal biotinylation (+88.000 Da) and methionine oxidation as variable modifications. The output of the search was filtered using Xcorr cutoffs (+1>1.8, +2>2.00, +3>2.50 and +4>3.50).

Statistical Analysis

For the functional experiments, ANOVA followed by Bonferroni multiple test and Student's *t*-test were used with a significance level established at 0.05 (GraphPad Prism version 5 for Windows).

RESULTS

Functional characterization of wild-type and ADAM17 -/- mEF secretome in migration and proliferation events

The effect of ADAM17-dependent secretome composition on the modulation of migration and proliferation events was evaluated using functional assays with concentrated conditioned media obtained from WT and KO mEF cells. Firstly, we demonstrated that ADAM17 knockout mEF cells have significantly diminished migration and proliferation capabilities using 1% FBS stimulus (Fig 1A and 1B, n=3, triplicate, ANOVA followed by Bonferroni multiple test, $p<0.05$). Secondly, we exchanged the conditioned media to confirm whether the differential secretome would be able to modulate these events. We showed that the ADAM17-dependent secretome had a higher capability to stimulate migration in the B16-F10 melanoma murine cell line compared to the secretome originating from KO mEF cells (Fig 1C, n=2, triplicate, ANOVA followed by Bonferroni multiple test, $p<0.05$). We also demonstrated that the WT secretome enhanced proliferation in KO knockout cells, and the KO secretome inhibited wild-type cells proliferation (Fig 1D, n=1, sextuplicate, ANOVA followed by the Bonferroni multiple test, $p<0.05$).

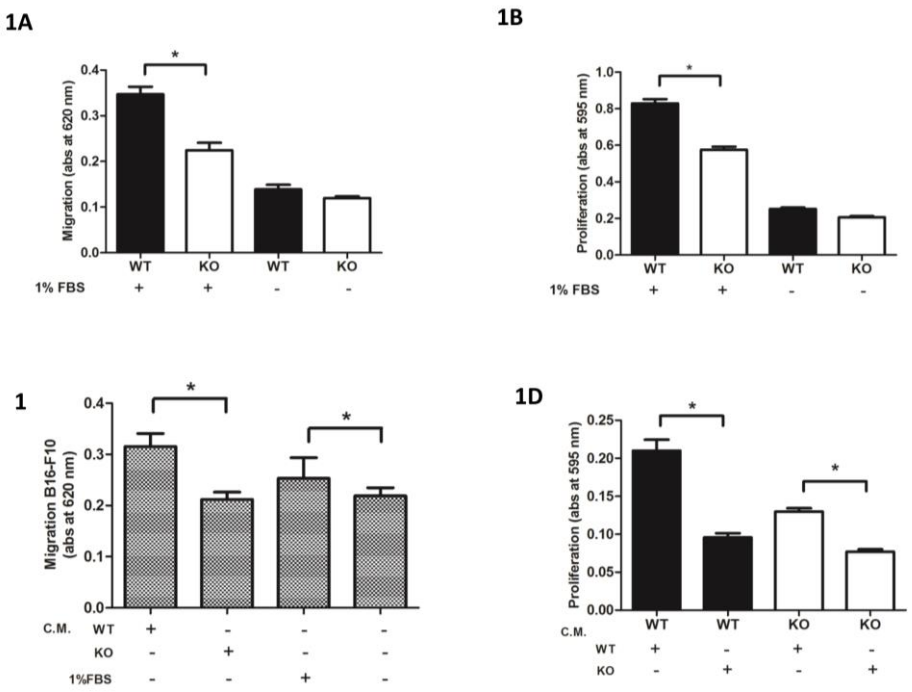

Figure 1 ADAM17-dependent secretome plays a role in cell migration and proliferation. (A) The migration of wild-type and ADAM17−/− mEF cells was evaluated using a transwell assay. Cells were seeded in serum-free media in the upper chamber of 96-well transwell plates, and media with or without 1% FBS were added in the lower chamber (n = 3, ANOVA followed by Bonferroni multiple test, *p < 0.05). (B) Wild-type and ADAM17−/− mEF cells were seeded in 96-well plates. After 48 h, cell viability was measured by MTT assay (n = 3, ANOVA followed by Bonferroni multiple test, *p < 0.05). (C) B16-F10 cells were seeded in serum-free media in the upper chamber of 96-well transwell plates. The 10 µg of concentrated conditioned media (CM) from wild-type or ADAM17−/− mEF cells was added in serum-free media in the lower chamber (n = 2, Student's t test, *p < 0.05). (D) Wild-type and ADAM17−/− mEF cells were seeded in 96-well plates with 5 µg of concentrated conditioned media (CM) from wildtype.
or ADAM17−/− mEF cells. After 48 h, cell viability was measured using a MTT assay (n = 1, ANOVA followed by Bonferroni multiple test,*p < 0.05).

Proteomic analysis revealed differential ADAM17-dependent secretome composition

To identify the proteins that were able to affect migration and proliferation, we used mass spectrometry-based proteomic approaches. Secretome analysis was performed according to the experimental design shown in fig 2A. We identified 1,781 proteins in label-free strategy and 751 proteins in SILAC strategy with a FDR less than 1% (Supplemental Table SM1 to SM5). Correlation analyses between all of the individual replicates, both in the label-free and SILAC methods, resulted in R-values of at least 0.93, indicating high reproducibility among the samples (Supplemental Figure SM1 and SM2).

Label-free quantitative proteomics were selected to explore the in-depth proteomic differential composition between the WT and KO secretomes. For that, filter valid values were applied for proteins with at least three valid values in total row, resulting in 1,029 proteins. Among them, 926 proteins were in common between the two conditions and 54 and 49 were exclusive in WT and KO, respectively (Fig 2B). We used the normalized spectral protein intensity (LFQ intensity) provided by the MaxQuant algorithm, and then converted the values into log2 values to perform Student's *t*-test analysis. Overall, 277 proteins showed statistically significant expression (Student's *t*-test, $p<0.05$, Supplemental Table SM2). Among these proteins, 179 proteins were down-regulated, and 98 were up-regulated in the ADAM17-dependent secretome from wild-type cells. Hierarchical clustering of significantly altered proteins was performed using the Z-score calculation on the log2 intensity values and it was represented as a heat map (Fig 2C).

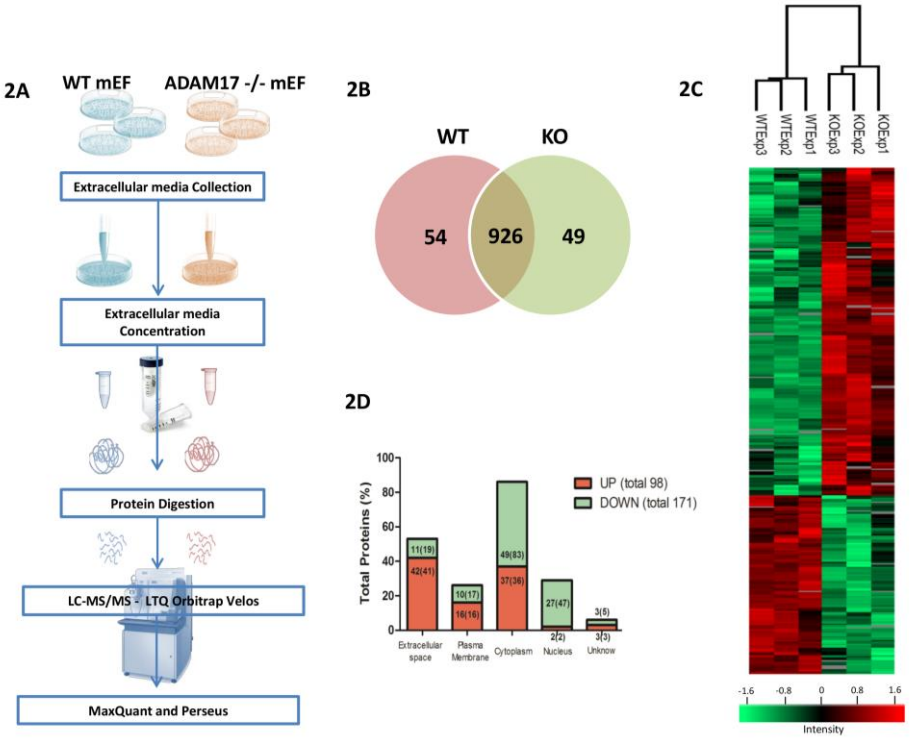

Figure 2 Bioinformatic analyses of differential expressed proteins in WT and KO mEF secretomes given by label-free quantification. A) Experimental workflow for secretome-based mass spectrometry analysis. B) Venn diagram showing the number of proteins identified by MS exclusive or in common between Wild-type and ADAM17 -/- mEF secretomes. C) Clustering of significantly up- and down-regulated proteins (Student's *t*-test, $p<0.05$) in Wild-type compared with ADAM17 -/- mEF secretome is shown as a Heat map, applying the Euclidian distance method and average linkage. D) The subcellular location of proteins was obtained based on the annotation available in the Ingenuity knowledge base. The corresponding percentage of proteins annotated to a subcellular localization category and the corresponding number of proteins (between brackets) is shown for up-regulated proteins (red) and down-regulated proteins (green) in WT secretome.

As a complementary analysis, in our SILAC experiments, KO cells were labeled with light DMEM medium (Arg0; Lys0), whereas WT cells were labeled with heavy medium (Arg10; Lys8). We applied the SILAC-based approach as a confirmatory strategy of the identification and quantification of differentially expressed proteins of WT and KO secretomes. From 751 proteins identified by this strategy, 498 showed H/L ratios in at least one of the replicates. Since only technical replicates were analyzed, for SILAC experiments we considered 1.5-fold change to perform data analysis [35-37]. Among the identified proteins, 344 proteins were up-regulated and 25 proteins were down-regulated. It is important to point out that the differences in the number of proteins identified using SILAC and label-free approaches can be associated with increased spectral complexity in SILAC samples, which leads to fewer identifications once fractionation of samples was not performed [29,36].

Location, biological function and interaction network of differentially expressed proteins and validation

Differentially expressed proteins obtained from the label-free analysis were analyzed in terms of their subcellular localization using IPA (Fig 2D). Up-regulated proteins in the wild-type secretome showed an enrichment of proteins from the extracellular space and plasma membrane, which represented respectively 42% (41 proteins) and 16% (16 proteins) of 98 up-regulated proteins in the dataset compared to 11% (19) and 10% (17) of 171 down-regulated proteins (Fig 2D), indicating an effect of the presence of ADAM17 on the secretome.

Because the ADAM17-dependent secretome modulates cell proliferation and migration, we aimed to identify the proteins that contribute to these events. For that, differentially expressed proteins identified in WT and KO secretomes were analyzed in Ingenuity software regarding migration and proliferation biological function related-proteins. Table 1 shows 51 and 31 up-regulated proteins in WT secretome significantly enriched in cell proliferation (p= 2.92E-10) and cell movement (p= 4.51E-07) biological function and 70 down-regulated proteins related with cell death (p=2.22E-9). The complete results given by biological function IPA analysis are available in supplemental Table SM6 and SM7.

TABLE 1 Biological functions of up and down-regulated proteins in WT mEF secretome given by IPA.

Category	Function Annotation	p-Value	Molecules	# Molecules
Up-regulated Proteins in WT mEF secretome				
Cellular Growth and Proliferation	proliferation of cells	2.92E-10	AKR1B10,AXL,B4GALT1,C3,CCT5,CCT7,CDH2,Clu,COL1A1,COL4A1,COL4A2,COL6A1,COL6A2,COL6A3,CREG1,CSF1,CST3,CTSD,EIF4A1,EPHB4,FBLN5,FBN1,FSCN1,GAS1,GPC1,HSPG2,LOX,LRP1,LTBP1,LTBP3,MFGE8,NAP1L1,NCAM1,NRP1,PIK3IP1,POSTN,PSAP,PSMD2,PTPRK,PXDN,SERPINB9,SERPINH1,SLC3A2,SPARC,STMN1,TCN2,TGFB1,THBS1,TNC,VCAM1,VCAN	51
Cellular Movement	cell movement	4.51E-07	AXL,B4GALT1,BGN,C3,CDH2,CHST1,COL1A1,COL4A1,COL4A2,CSF1,FBLN5,FBN1,FSCN1,GPC1,LOX,LRP1,MATN2,NCAM1,NRP1,PIK3IP1,POSTN,PTPRK,SDC4,SLC3A2,SPARC,STMN1,TGFB1,THBS1,TNC,VCAM1,VCAN	31
Down-regulated Proteins in WT mEF secretome				
Cell Death and Survival	cell death	2.22E-09	AIMP1,APRT,ARHGDIA,CAPRIN1,CCL13,CDC37,CFH,Cxcl12,DCN,DPYSL3,EFEMP1,EIF3H,ENO1,EZR,FDPS,FIS1,FKBP4,FUBP1,GSN,HDGF,HINT1,HLA-B,HMGB1,HMGB2,HMOX1,HNRNPC,HSP90AA1,HSPA1A/HSPA1B,HSPA8,IGFBP6,LDLR,LGALS3,LGALS3BP,LMNA,LMNB1,MAN2A1,MAP4,MAPK1,MGP,MSN,NCL,NOLC1,NPM1,PAFAH1B2,PARK7,PCNA,PDCD6IP,PLEC,PNP,PPIA,PPP2R4,PRDX5,PRDX6,PROS1,RBBP4,RDX,RHOA,RTN4,SOD1,SPP1,TAGLN2,TPM3,UBA1,UBE2I,UBQLN1,VCL,VIM,YBX1,YWHAG,ZYX	70

Biological network analysis was also performed using the up and down-regulated proteins present in WT and KO secretome data. The top two most significant pathways generated in each condition were selected and merged to obtain global views (Supplemental Figure SM3). It was observed that the ERK1/2 pathway is predicted to be activated by up-regulated proteins present in the WT secretome, while down-regulated proteins lead to the inhibition of the focal adhesion pathway. To provide proof-of-concept for these data, we validated the activation of ERK1/2 and inhibition of the FAK pathway by analyzing their phosphorylation using intracellular extracts obtained from WT and KO mEF cells. We confirmed an increase in ERK1/2 phosphorylation and decrease in FAK phosphorylation in WT compared to KO cells (Fig 3A, 3B and 3C, n=3, Student's *t*-test, $p<0.05$).

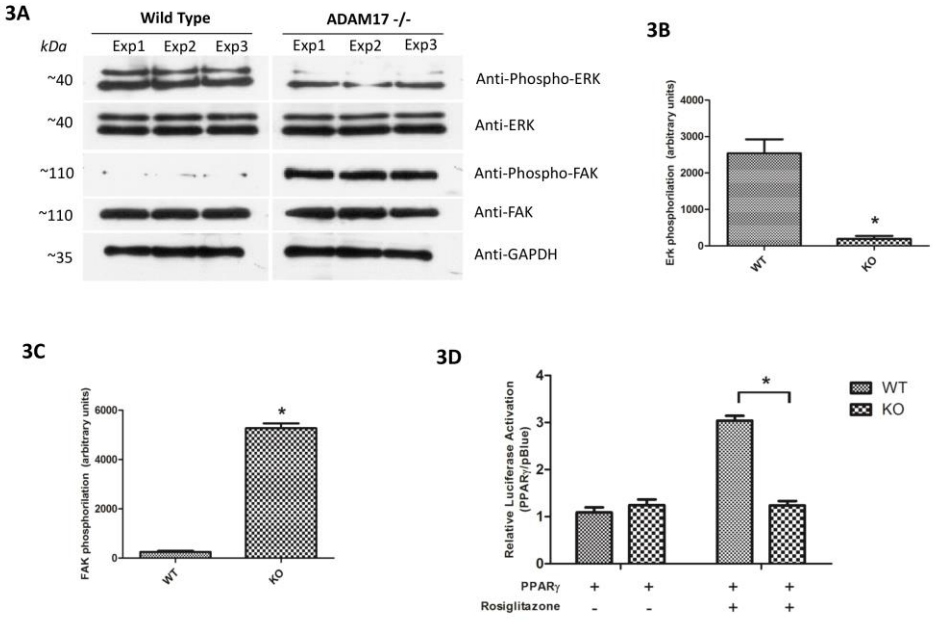

Figure 3. Wild-type cells have increased in ERK phosphorylation and decreased in FAK phosphorylation. (A) The activation of the ERK pathway and the inhibition of the FAK pathway in wild-type compared with ADAM17−/− mEF cells were validated by immunoblotting using anti-phospho-ERK, anti-phospho-FAK, anti-ERK, anti-FAK, and anti-GAPDH antibodies in intracellular extracts. (B) Densitometry analysis of ERK phosphorylation levels was calculated by band intensity using ImageJ software (n = 3, Student's t test, *p < 0.05). (C) Densitometry analysis of FAK phosphorylation levels was calculated by band intensity using ImageJ software (n = 3, Student's t test *p < 0.05). (D) PPARγ activation is increased in WT mEF cells. PPARγ transcriptional activity was determined using luciferase promoter activity assay. WT and KO cells were transfected with PPARγ (or pBlue as control), DR1 firefly luciferase, and renilla luciferase. After 4 h of transfection, the cells were treated with 10 μM rosiglitazone for 18 h or vehicle DMSO. PPARγ activation was evaluated by fold-increase of normalized
luciferase activity of WT or KO mEF cells transfected with PPARγ/ control pBlue treated or not with rosiglitazone (n = 3, two-way ANOVA followed by Bonferroni multiple test, *p < 0.001).

The regulatory mechanisms associated with the transcription factor binding motifs of the altered protein expression in the secretome were evaluated using the X2K software. The most significantly enriched regulation transcription factor found in the ADAM17-dependent secretome data set was PPARγ (p=1.41E-08), which consists of 33 proteins (Table 2, Supplemental Table SM8). We also evaluated each protein regarding its cellular localization using IPA and we observed that most of the proteins with PPARγ binding motifs localize in extracellular space (54%) (Table 2). This result prompted to the evaluation of whether PPARγ was more or less activated in the presence of ADAM17. We observed that once stimulated with the PPAR agonist Rosiglitazone, only WT

cells were able to produce a significantly stimulatory response (Fig 3D, n=3, two-way ANOVA followed by Bonferroni multiple test, $p<0.001$).

TABLE 2 Top five enriched transcription factor (TF) for up-regulated proteins in WT mEF secretome using X2X software. The cellular localization of each protein is shown in brackets. PM: plasma membrane; ES: extracellular space; C: cytoplasm; N: nucleus; U: unknown.

TF	Target/ Input	P-value	Genes
PPARG-20887899	33	1.41E-08	AXL (PM); CD109 (PM); CDH2 (PM); GAS1 (PM); NRP1 (PM); PTPRK (PM); SLC3A2 (PM); COL12A1 (ES); COL4A2 (ES); COL5A1 (ES); COL5A2 (ES); COL6A1 (ES); COL6A2 (ES); CST3 (ES); FSTL1 (ES); HSPG2 (ES); LTBP3 (ES); Masp1 (ES); MATN2 (ES); PSAP (ES); PXDN (ES); SERPINH1 (ES); SPARC (ES); TCN2 (ES); THBS1 (ES); ACO2 (C); B4GALT1 (C); CCT5 (C); DDAH1 (C); FKBP10 (C); FSCN1 (C); GANAB (C); HADH (C)
SUZ12-20075857	29	1.02E-04	CD109 (PM); CDH2 (PM); GAS1 (PM); GPC1 (PM); NCAM1 (PM); NRP1 (PM); COL12A1 (ES); COL4A1 (ES); COL4A2 (ES); COL5A1 (ES); CSF1 (ES); FBLN5 (ES); FBN1 (ES); FSTL1 (ES); LOX (ES); LTBP1 (ES); LTBP3 (ES); OLFML2B (ES); PXDN (ES); SERPINH1 (ES); TGFB1 (ES); THBS1 (ES); VCAN (ES); CHST1 (C); CPE (C); DDAH1 (C); FKBP10 (C); NAGLU (C); PREP (C)
E2F1-18555785	28	1.23E-04	WDR61 (U); EPHB4 (PM); NUDT2 (PM); PTPRK (PM); SDC4 (PM); SLC3A2 (PM); CREG1 (N); CST3 (ES); FSTL1 (ES); MATN2 (ES); ACO2 (C); CCT5 (C); Clu (C); CTSD (C); DDAH1 (C); EIF4A1 (C); EXT1 (C); EXT2 (C); FSCN1 (C); FUCA1 (C); GALC (C); GANAB (C); GNS (C); HADH (C); HSP90B1 (C); MDH2 (C); STMN1 (C); VCP (C)
CREM-20920259	24	0.1231575	PIK3IP1 (U); WDR61 (U); CDH2 (PM); FAT1 (PM); NUDT2 (PM); SLC3A2 (PM); NAP1L1 (N); ACO2 (C); B4GALT1 (C); CCT5 (C); CCT7 (C); DNAJB11 (C); EIF4A1 (C); EXT1 (C); FSCN1 (C); GALC (C); GANAB (C); GM2A (C); GNS (C); HADH (C); MDH2 (C); STMN1 (C); UXS1 (C); VCP (C)
SOX2-18692474	23	3.81E-04	CD109 (PM); CDSN (PM); EPHB4 (PM); GAS1 (PM); NCAM1 (PM); SDC4 (PM); SLC3A2 (PM); COL4A1 (ES); COL4A2 (ES); CST3 (ES); FSTL1 (ES); MATN2 (ES); OLFML2B (ES); SERPINH1 (ES); B4GALT1 (C); CPE (C); EIF4A1 (C); EXT1 (C); FSCN1 (C); HADH (C); PREP (C); UXS1 (C); VCP (C)

ADAM17-dependent secretome exhibits increased gelatinolytic activity

Considering the identification of a large number of degraded extracellular matrix proteins, such as collagens and glycoproteins (Table 3), we evaluated the gelatinolytic activity in the ADAM17-dependent secretome. By subjecting the secretome from WT and KO mEF cells to nonreducing SDS-PAGE separation in a gel containing gelatin, we observed higher gelatinolytic

activity in ~100 kDa (visualized as a colorless band after Coomassie blue staining of the gels) in the secretome originating from WT cells (Fig 4A and 4B, n=3, Student's *t*-test, p<0.05).

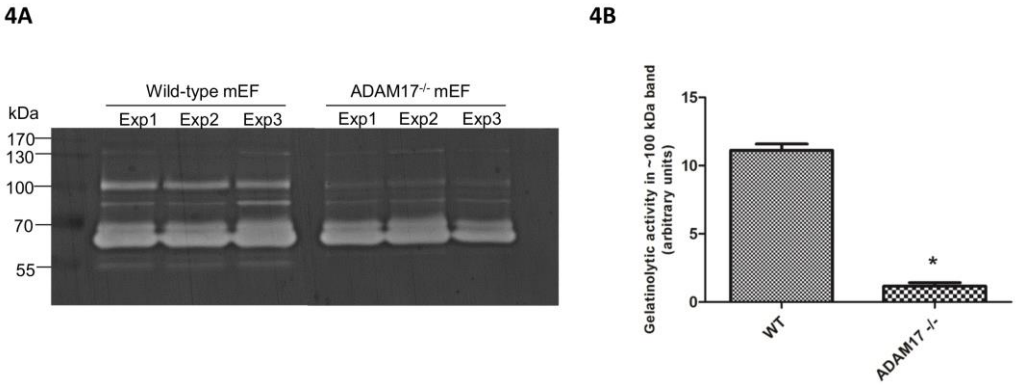

Figure 4 ADAM17-dependent secretome has increased gelatinolytic activity. A) Gelatinolytic activity is increased in the secretome originated from WT cells. B) Gelatinolytic activity levels of ~100 kDa band in figure A were calculated by band intensity using ImageJ software (n=3, Student's *t*-test, *p<0.05). Numbers on the left indicate molecular mass marker mobility.

Potential novel proteins directly or indirectly modulated by ADAM17

Proteins that were up-regulated or exclusive in the WT secretome were identified using label-free and SILAC strategies (Table 3). We observed many up-regulated proteins in the WT secretome that have not been previously reported as ADAM17 substrates, particularly extracellular matrix proteins and membrane proteins, although it is possible that the shedding of these substrates may have been promoted by other proteases activated in the presence of ADAM17. We considered as potential direct or indirect ADAM17 substrates when the proteins are statistically significant (Student's *t*-test, p<0.05, Label-free method) and has intensity ratio >1.5 (SILAC method); exclusive proteins in WT in label-free method were also included. For this analysis, we excluded the cytoplasmic and nucleus proteins.

We selected two proteins, mimecan and perlecan, where suitable antibodies and recombinant proteins were available and which showed higher protein abundance in WT mEF cells compared to KO mEF cells using LFQ intensity quantification and SILAC labeling (Table 3). First, the higher protein abundance of mimecan in the WT secretome compared to the KO secretome was confirmed using immunoblotting (Fig 5A). Interestingly, a double band was observed, suggesting that it was a proteolytic cleavage. Since human antibody perlecan is specie specific and do not react with mouse secretome, we validated the protein abundance of perlecan using a human A-431 cell

line that was silenced by lentiviral ADAM17 shRNA (Fig 5B). In fact, we could observe that the secretome from A-431 knockdown for ADAM17 had lower perlecan protein abundance compared with mock and scramble cell lines.

To investigate the effect of ADAM17 on mimecan and perlecan protein abundance, the mRNA levels of mimecan and perlecan were evaluated using RT-qPCR. Quantitative PCR analysis showed a significant decrease in mimecan mRNA levels in the ADAM17 knockout cell line (Fig 5C, n=3, triplicate, Student's *t*-test, $p<0.05$), while no difference was observed in the mRNA levels of perlecan in A-431 mock, scramble and shRNA-ADAM17 cell lines (Fig 5D, n=3, triplicate, ANOVA, $p>0.05$).

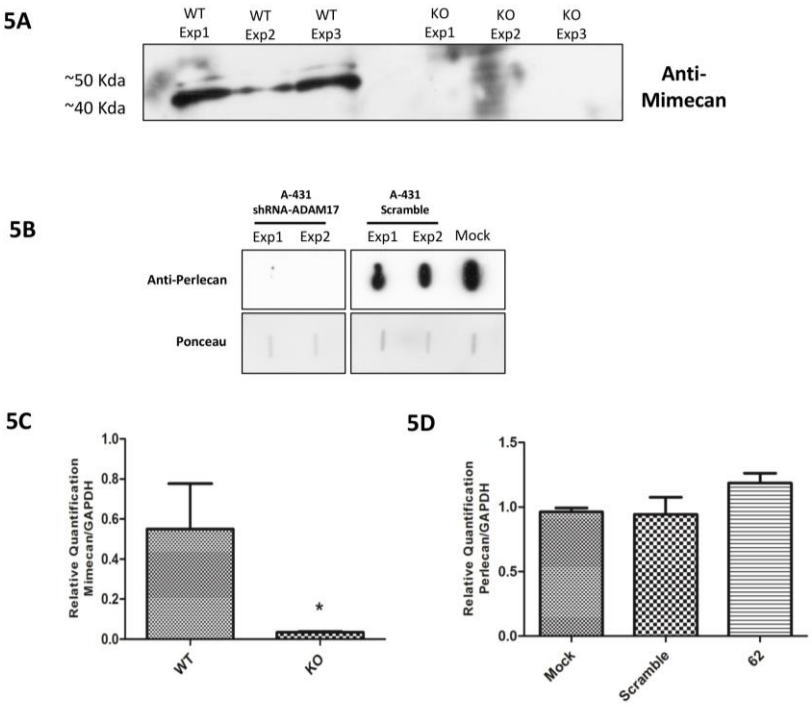

Figure 5 Novel ADAM17 substrates revealed by quantitative proteomics. (A) Mimecan expression in Wild-type and ADAM17 -/- mEF secretome by immunoblotting (IB). (B) Analysis of perlecan expression in ADAM17 knockdown (shRNA), scramble and mock A-431 secretome by slot blotting. Ponceau S staining was used as loading control. Relative mRNA expression levels of mimecan (C) and perlecan (D) were measured by the real-time quantitative PCR. The data were normalized with glyceraldehyde-3-phosphate dehydrogenase gene. Student's *t*-test indicated differences in the mimecan expression level between Wild-type and ADAM17 -/- mEF cell lines. No statistical difference was observed in perlecan mRNA levels among A-431 Mock, scramble and shRNA-ADAM17 cell lines. Columns represent mean ± SD (n = 3) and * indicates $p<0.05$.

Table 3 Known and potential proteins modulated by ADAM17 identified by label-free and SILAC quantitative proteomics in conditioned media from WT and KO mEF cells.

Uniprot ID	Protein Name	Gene name	Label-Free*			SILAC#			Refe or Location
			Exp 1	Exp 2	Exp 3	Exp 1	Exp 2	Exp3	
Known ADAM17 direct substrates identified in this study									
O08712;Q3UK97	Tumor necrosis factor receptor superfamily member 11B	Tnfrsf11b	Only WT	Only WT	Only WT				[72]
P14719;P14719-2;Q5D095;G3UYU3	Interleukin-1 receptor-like 1	Il1rl1	Only WT	Only WT	Only WT				[73]
P13595;F7C5V8;E9QB01;P13595-2;F6RLF6;E9Q589;P13595-3;F7C5W6;P13595-4;F7CRA9;F8VQH8	Neural cell adhesion molecule 1	Ncam1	3.5	5.8	3.8				[74]
O35988;Q3U5S6	Syndecan-4	Sdc4	2.9	2.5	6.6	NaN	NaN	2.2	[49]
P35822;B2RRF0;Q8C5G0	Receptor-type tyrosine-protein phosphatase kappa	Ptprk	2.1	2.4	3.6	3.7	4.5	NaN	[75]
P07141;P07141-2;D3Z090;F6RNW8;D3YTW1	Macrophage colony-stimulating factor 1	Csf1	2.2	2.4	3.2	3.3	3.1	3.0	[76]
Q00993;Q6PE80;F6YPR4	Tyrosine-protein kinase receptor UFO	Axl	1.3	1.4	2.9				[77]
Q8C8K1;P54761;E9PWK7	Ephrin type-B receptor 4	Ephb4	1.5	1.6	1.8				[78]
P29533;Q3UPN1;P29533-2;Q544V4	Vascular cell adhesion protein 1	Vcam1	2.0	1.8	1.4	NaN	NaN	1.4	[79]
A2APM2;P15379;E9QKL7;P15379-7;E9QKL8;P15379-8;A2APM1;P15379-4;E9QKL9;E9QKM0;P15379-9;P15379-10;E9QKM1;P15379-13;P15379-	CD44 antigen	Cd44				2.3	2.6	2.3	[80]

12;Q80X37;E9QKM8;P15379-5;P15379-11;A2APM6;A2APM4;P15379-6;A2APM3;P15379-3;A2APM5;Q3U8S1;P15379-2									
Potential proteins modulated by ADAM17 identified in this study									
P28653;Q3TNY9	Biglycan	Bgn	5.1	4.8	6.4	6.1	6.3	6.3	Extracellular Space
Q8CG16;F6QBW6;Q8CFG9;F8WHL5	Complement C1r-A subcomponent	C1ra;C1rb	1.7	2.3	1.8	2.3	3.1	2.4	Extracellular Space
E9Q6C2;Q8CG14;D3YUK7;E9Q493	Complement C1s-A subcomponent	C1s;C1sa	6.1	6.6	6.5	5.3	3.9	4.3	Extracellular Space
P15116;D3YYT0	Cadherin-2	Cdh2	1.9	5.1	4.1	NaN	3.6	3.2	Plasma Membrane
Q7TPC1	Corneodesmosin	Cdsn	15.6	Only WT	21.1	20.3	14.9	40.1	Plasma Membrane
Q60847;Q60847-5;Q608472;E9PX70;Q60847-3;Q60847-4	Collagen alpha-1(XII) chain	Col12a1	Only WT	5.2	18.8	6.3	5.7	4.7	Extracellular Space
P11087;F8WGB7;P11087-2	Collagen alpha-1(I) chain	Col1a1	27.9	50.3	51.0	12.3	9.6	10.2	Extracellular Space
P02463	Collagen alpha-1(IV) chain;Arresten	Col4a1	2.4	5.2	7.0	1.4	1.6	1.9	Extracellular Space
B2RQQ8;P08122	Collagen alpha-2(IV) chain;Canstatin	Col4a2	3.2	4.0	3.8	1.4	1.6	1.5	Extracellular Space
B1AWB9;O88207	Collagen alpha-1(V) chain	Col5a1	4.4	12.3	13.2	4.0	2.7	3.3	Extracellular Space
Q3U962	Collagen alpha-2(V) chain	Col5a2	2.4	4.6	5.4	3.8	4.4	4.8	Extracellular Space
Q04857	Collagen alpha-1(VI) chain	Col6a1	4.1	4.0	3.3	4.6	4.5	4.5	Extracellular Space
Q02788;D3Z7D5	Collagen alpha-2(VI) chain	Col6a2	11.1	8.1	11.9	3.5	3.7	3.9	Extracellular Space
E9Q8G0;D3YWD1	Collagen alpha-3(VI) chain	Col6a3	Only	Only	Only	6.1	3.0	8.6	Extracellular Space

			WT	WT	WT				
E9PZD8;G3X8Q5; Q61147;G3X9T8	Ceruloplasmin	Cp	6.8	8.5	13.0	4.3	4.1	3.9	Extracellular Space
A2APX2;P21460; A2APX3	Cystatin-C	Cst3	2.0	2.6	3.6	2.2	2.3	1.9	Extracellular Space
P50608;Q543D2	Fibromodulin	Fmod	Only WT	Only WT	Only WT	6.1	8.9	8.8	Extracellular Space
Q9WVH9	Fibulin-5	Fbln5	5.9	5.3	7.4	15.1	NaN	11.4	Extracellular Space
A2AQ53;Q61554	Fibrillin-1	Fbn1	7.9	12.7	6.0	4.6	5.2	6.0	Extracellular Space
Q62356	Follistatin-related protein 1	Fstl1	2.2	3.8	3.9	3.9	4.6	4.5	Extracellular Space
Q01721	Growth arrest-specific protein 1	Gas1	1.9	1.9	3.6	5.6	4.7	NaN	Plasma Membrane
E9PZ16;Q3UHH3	Basement membrane-specific heparan sulfate proteoglycan core protein	Hspg2	2.8	7.2	5.4	1.8	1.7	1.7	Extracellular Space
Q6GU68	Immunoglobulin superfamily containing leucine-rich repeat protein	Islr	1.5	3.4	2.7	NaN	2.2	2.6	Extracellular Space
Q6PE62;P47878	Insulin-like growth factor-binding protein 3	Igfbp3	Only WT	Only WT	Only WT	NaN	1.975	NaN	Extracellular Space
Q6PE55	Platelet-derived growth factor receptor-like protein	Pdgfrl	Only WT	Only WT	Only WT	NaN	NaN	4.7297	Plasma Membrane
Q91ZX7	Prolow-density lipoprotein receptor-related protein 1	Lrp1	1.7	2.6	2.4	1.8	1.7	NaN	Plasma Membrane
Q8CG19;Q8CG19-2; Q8CG19-3;B1B1E2	Latent-transforming growth factor beta-binding protein 1	Ltbp1	2.2	4.0	2.0	2.9	3.4	2.9	Extracellular Space
F8VQ06;E9QJZ6; Q61810;Q61810-2	Latent-transforming growth factor beta-binding protein 3	Ltbp3	4.1	5.1	2.5	2.9	NaN	2.0	Extracellular Space
O08746;O08746-2	Matrilin-2	Matn2	2.1	1.9	1.3	NaN	2.9	3.1	Extracellular Space
P21956;P21956-2	Lactadherin	Mfge8	7.5	7.9	10.3	6.3	5.1	3.8	Extracellular Space
Q543C5;Q62000	Mimecan	Ogn	23.0	35.7	58.7	17.0	12.7	19.9	Extracellular

									Space
Q3V1G4	Olfactomedin-like protein 2B	Olfml2b	6.0	7.7	5.4	3.1	3.6	3.3	Extracellular Space
O35103	Osteomodulin	Omd	57.6	20.9	Only WT	4.9	NaN	NaN	Extracellular Space
Q62009;Q62009-2; Q62009-3;Q62009-4; Q62009-5	Periostin	Postn	102.4	259.6	71.7	7.6	7.0	10.7	Extracellular Space
Q8BFQ1;E9PZ00; Q3UFE8;Q61207	Sulfated glycoprotein 1	Psap	2.6	2.8	3.8	4.4	4.9	5.1	Extracellular Space
P35822;B2RRF0	Receptor-type tyrosine-protein phosphatase kappa	Ptprk	2.1	2.4	3.6	3.7	4.5	NaN	Plasma Membrane
Q3UQ28	Peroxidasin homolog	Pxdn	3.6	4.2	2.8	2.9	3.1	3.3	Extracellular Space
P19324	Serpin H1	Serpinh1	5.2	4.2	2.2	7.0	7.5	7.0	Extracellular Space
P07214;Q5NCU5; Q5NCU4	SPARC	Sparc	2.3	4.4	3.7	2.9	2.7	2.8	Extracellular Space
Q80YX1;E9QKW5; Q80YX1-3;E9QM59; Q80YX1-4;E9QM60; Q80YX1-5	Tenascin	Tnc	6.7	4.7	3.1	4.3	4.5	5.4	Extracellular Space
E9PYH0;E9QQ15; Q62059;E9QMK2; Q62059-3	Versican core protein	Vcan	56.6	32.4	16.3	NaN	6.2	7.5	Extracellular Space

NaN= not a number.
* = LFQ intensity ratio of proteins identified in the conditioned medium from Wild-type/ADAM17-/- mEFs cells.
# = H/L intensity ratio of proteins identified in the conditioned medium from Wild-type/ADAM17-/- mEFs cells.

Although the ADAM17-dependent secretome has an effect on mimecan mRNA levels, we considered the potential effect on cleavage due to the higher intensity ratio observed in MS analysis as well as the double band observed in immunoblotting. ADAM17 cleavage products were visualized in silver-stained SDS-PAGE, after incubating mimecan and the C-terminal of the heparan sulfate proteoglycan perlecan recombinant proteins, named endorepellin, with ADAM17 recombinant protease (Fig 6A). N-terminal labeling approach was used to determine the cleavage site of these proteins, which were generated by ADAM17. Perlecan and mimecan recombinant proteins were cleaved into a product that lacked the first 124 and 70 residues, respectively (Supplemental Table SM9 and SM10). Perlecan showed residues of glycine and leucine in P1'and P2' and mimecan showed residues of leucine and glutamine in P1' and P2'. Manual verification of the N-terminal labeled peptide of mimecan and perlecan spectra are shown in fig 6B and 6C.

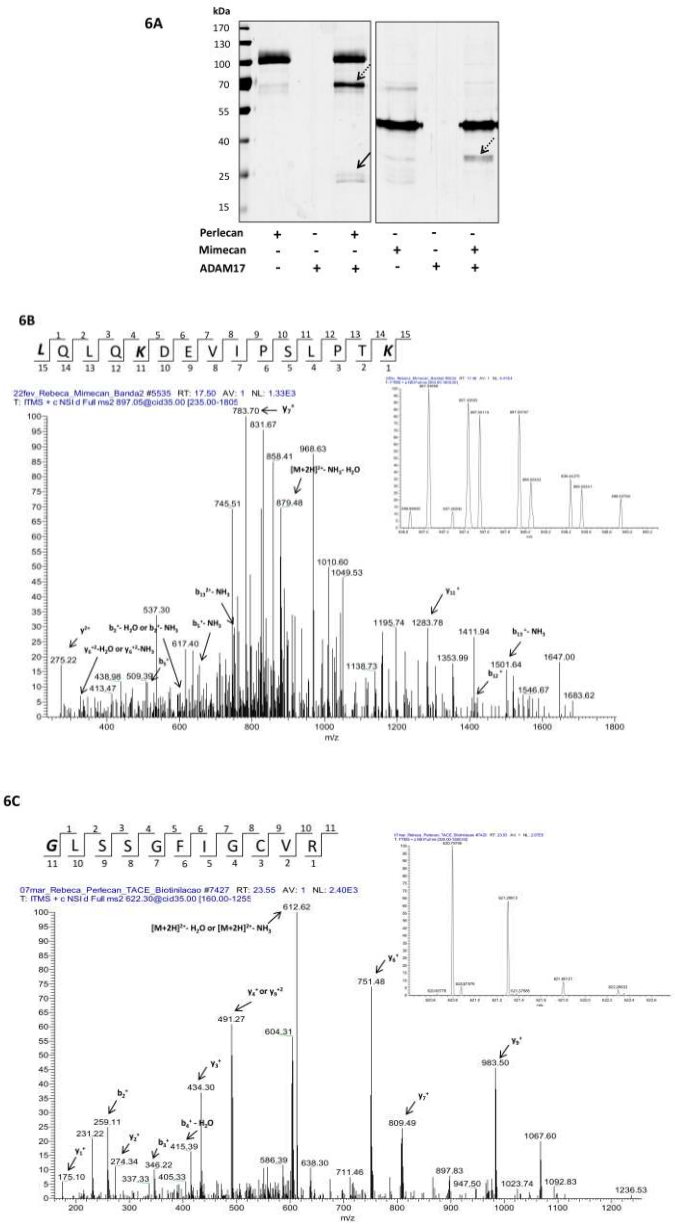

Figure 6 Proteolytic activity of ADAM17 upon mimecan and perlecan and analysis of their cleavage sites. (A) Mimecan (residue Ala20-Phe298) and perlecan (residue Arg3684Ser4391 corresponding the C-terminal end of the heparan sulfate proteoglycan Perlecan, named endorepellin) were incubated at a 1:100 (w/w) enzyme-to-substrate ratio with or without ADAM17 for 16 h at 37°C. Arrows indicate product of hydrolysis of mimecan and perlecan and dashed arrows indicate the product of hydrolysis where the neo-amino termini were identified by MS. (B) Manual verification of N-terminal labeled peptide of mimecan MS/MS spectrum for b and y ions, m/z 897.0499, +2, xcorr: 2.09 (corresponding to residue from 71 to 85, Uniprot Mouse Q62000). (C) Manual verification of N-terminal labeled peptide of perlecan MS/MS spectrum for b and y series, m/z 620.79749, +2, xcorr 3.12 (corresponding to residue from 3810 to 3821, Uniprot Human P98160).

DISCUSSION

Metalloproteases are important in numerous aspects of biology, ranging from cell proliferation, differentiation and remodeling of the extracellular matrix to vascularization and cell migration. Proteases belonging to the 'A Disintegrin And Metalloproteinase' (ADAM) family are membrane-anchored proteases that are able to cleave the extracellular domains of membrane-bound proteins in a process known as 'ectodomain shedding'. Through the shedding of numerous growth factors, cytokines, receptors, and cell adhesion molecules, ADAM17 has been reported to be an indispensable regulator of several physiological cellular events, but it is also associated with cancer development and other diseases [10,38,39].

The elucidation of protease substrate degradomes is essential for understanding the function of proteolytic pathways in the protease network and how proteases regulate cell function [40,41]. In addition, a new concept has emerged of the regulation of proteolysis on substrate-dependent levels [42]. Thus, understanding how protease activity affects the global state, in a system biological context, is essential to reveal the biological function of proteases in homeostasis and disease conditions.

In this report, we demonstrated that the secretome composition of wild-type (WT) and ADAM17 knockout (KO) mEF cell lines differentially modulated critical cellular processes. The secretome is a crucial component for the intercellular communication and the cross-talk between tumor and stroma has a key influence on tumor progression [43-45], especially by disrupting the homeostasis of the extracellular matrix via either proteolysis or an alteration of its architecture due to changes in its molecular composition and/or stoichiometry [46,47]

Our first evidence of the distinct protein secretome composition of WT and KO cells was the promoting effect of ADAM17 on migration in melanoma cell lines (Fig 1C). We next demonstrated that the ADAM17-dependent secretome enhanced proliferation when the KO and WT media were exchanged (Fig 1D). Moreover, the lack of ADAM17 resulted in a secretome composition that reduced cell proliferation in WT cells. Previously studies have shown that overexpression of ADAM17 in MCF-7 breast cancer cells increased in vitro invasion and proliferation, whereas down-regulation of ADAM17 expression in MDA-MB-435 cells decreased invasion and proliferation [48]. Consistent with this finding, it was also demonstrated that ADAM17 promoted the MDA-MB-231 malignant phenotype via increased proliferation, invasion and angiogenesis [49] and knockdown of ADAM17 expression using siRNA decreased endothelial cell proliferation and invasion in vitro [22]. More recently, Das et al, 2012 [23] demonstrated that when

ADAM17 was silenced in MC38CEA cells, in vivo tumor growth and in vitro cell motility were significantly diminished. Although the general mechanism of this effect is known to associate to ADAM17 substrates present in the extracellular environment, the panel of specific proteins that dictate these events has not yet been reported.

Thus, to identify the players that may be involved in the intercellular cross-talk, we employed mass spectrometry-based proteomics using label-free and SILAC quantitative methods. Firstly, known ADAM17 substrates, whose cleavage had previously associated with tumorigenic process were found, such as: syndecan-4 [50], TNFR [51] and CD44 [52]. Secondly, using network analysis, the panel of up-regulated proteins identified in the ADAM17-dependent secretome was analyzed regarding biological process and pathways involved with proliferation and migration. The network analysis indicated that ERK1/2 was as the major signaling hub (Fig 3A) and its predicted activation was also validated in cells (Fig 3 A/B). It is well known that the ERK pathway is one of the most important signaling pathways for cell proliferation and migration, and it has been shown that its activation is ADAM17-mediated EGFR-transactivation[53-55].

In addition, down-regulated proteins in WT cells predicted inhibition of the FAK pathway, which was an interesting effect associated with an ADAM17-dependent secretome (Fig 3A/C). Interestingly, Zheng et al, 2009 [56] showed that inhibition of FAK and dephosphorylation of its Y397 residue promoted cell migration, invasion, and metastasis, and it was mediated by the Ras-induced Fgd1-Cdc42-PAK1-MEK-ERK signaling cascade. This effect can be also correlated with the higher proliferation and migration observed in mEF cells and melanoma cells (Fig 1C/D).

In this study, we verified a large number of degraded extracellular matrix proteins, such as collagens, proteoglycans and glycoproteins (Table 3), which might be associated with higher gelatinolytic activity in the ADAM17-dependent secretome (Fig 4A/B). Some studies have demonstrated that ADAM17 is related with positive regulation of MMP-2 [57] and MMP-9 [58]. Moreover, Xiao et al, 2012 [59] showed that ADAM17 triggered prostate cancer cell invasion by up-regulating the expression of MMP-2 and MMP-9 via activation of the EGFR-MEK-ERK signal pathway.

Considering that the reduced and increased secretion or cleavage of proteins might be a secondary effect of the absence of ADAM17, we investigated whether differentially expressed protein sets can be controlled using a common transcription factor. Many protein sets, which represent proteins controlled by a common transcription factors (TFs) were found, potentially suggesting a role of ADAM17 in downstream effects. PPARγ was found to regulate 33 up-regulated proteins present in the wild-type mEF secretome (Table 2). For the first time, we also showed that

PPARγ was more active in wild-type cells and most likely regulated these up-regulated proteins (Fig 3D). Interestingly, PPARγ plays an important role in carcinogenesis and is positively regulated by ERK and MAP kinase [60]. It was shown that PPARγ induces epithelial to mesenchymal transformation via Rho GTPase-dependent activation of ERK1/2 [61] as well as invasion in esophageal cancer [62].

Several known and potentially novel ADAM17 substrates were highly expressed in the wild-type secretome (Table 3). We selected two candidate substrates, mimecan and perlecan, to further evaluate the direct or indirect ADAM17 cleavage. Mimecan, a member of the small leucine-rich proteoglycan gene family, is located in the extracellular matrix and is important not only for the regulation of the structure of the matrix but also for the regulation of proinflammatory signaling, cell cycle and growth factor activities [63-65]. Perlecan is a heparan sulfate proteoglycan that is present in all basement membranes and other tissues, such as cartilage, plays important roles in development and tissue functions and is associated with various diseases. Perlecan binds to basement membrane components, such as laminins and collagen IV and growth factors, such as FGFs and VEGFs, which are critical for growth factor signaling and angiogenesis [66-69]. Although the cleavage of mimecan by MMP12 [70] and perlecan by stromelysin, collagenase, heparitinase I and plasmin [71] were reported, we showed that both mimecan and perlecan were proteolytically processed by ADAM17 (Fig 6A). The conserved P1' and P2' cleavage site of leucine was observed for the mimecan and perlecan products (Fig 6A/B/C), respectively, supporting our findings in cleavage site specificity as previously reported for ADAM17 and other metalloproteinases [34,72]. However, future studies are required to determine the biological role of mimecan and perlecan products. Moreover, although both proteins showed product cleavage when incubated with recombinant ADAM17, mRNA expression analysis for mimecan indicated that in addition to cleavage, its expression profile was also affected by the absence of ADAM17 (Fig 5C).

Overall, the ADAM17-dependent secretome demonstrated an extensive role in cell proliferation and migration phenotypes, which might be associated with fine-tuned signaling pathways dictated by the direct or indirect effect of ADAM17 expression in a complex network.

ABBREVIATIONS

mEF, Mouse embryonic fibroblasts; TACE, tumor necrosis factor-α converting enzyme; ADAM17, Disintegrin and metalloproteinase domain-containing protein 17; ALCAM, Activated leukocyte cell adhesion molecule; ANOVA, analysis of variance; CD44, Extracellular matrix receptor III; CV, conditioned media; DR1, Direct repeat-1; EGF, epidermal growth factor; ERK, extracellular signal-regulated protein kinase; FAK, focal adhesion kinase; FBS, fetal bovine serum; FDR, false discovery rate; FGF, fibroblast growth factor; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; H/L, heavy/light; HBEGF, heparin binding EGF-like growth factor; IB, immunoblotting; IPA, Ingenuity Pathways; KO, ADAM17 knockout; LFQ, label-free quantification; MAP, molecule activity predictor; MAPK mitogen-activated protein kinase; MTT, tetrazolium 3-(4,5-dimethylthiazol-2)-2,5-diphenyltetrazolium bromide; NCAM, Neural cell adhesion molecule; PBS, phosphate-buffered saline; PPAR, Peroxisome proliferator-activated receptor; RL, renilla luciferase; RT-qPCR, real time quantitative PCR; shRNA, small hairpin RNA; SILAC, Stable isotope labeling by amino acids in cell culture; SV40, Simian vacuolating virus 40; TGF-α, transforming growth factor-α; TNF, tumor necrosis factor; TTBS, Tris-Tween-buffered saline; VCAM, Vascular cell adhesion protein; VEGF, Vascular endothelial growth factor; WT, Wild-type; X2K, Expression2kinases.

ACKNOWLEDGEMENTS

This work was supported by FAPESP Grants: 2009/54067-3, 2010/19278-0, 2011/22421-2 and 2009/53839-2 and CNPq Grants: 470567/2009-0, 470549/2011-4 and 301702/2011-0.

We are thankful to Dr. Carl Blobel from Weill Medical College of Cornell University for donating the WT mEF cells and ADAM17$^{-/-}$ mEF cells; Dr. Jay W. Fox from University of Virginia for suggesting some experiments; Dr. Kleber Gomes Franchini for anti-phospho-ERK, anti-ERK, anti-phospho-FAK and anti-FAK antibody donation.

NOTES

The authors declare no competing financial interest.

REFERENCE

(1) Chang, C.; Werb, Z.: The many faces of metalloproteases: cell growth, invasion, angiogenesis and metastasis. *Trends Cell Biol* 2001, *11*, S37-43.

(2) Seals, D. F.; Courtneidge, S. A.: The ADAMs family of metalloproteases: multidomain proteins with multiple functions. *Genes Dev* 2003, *17*, 7-30.

(3) Dean, R. A.; Overall, C. M.: Proteomics discovery of metalloproteinase substrates in the cellular context by iTRAQ labeling reveals a diverse MMP-2 substrate degradome. *Molecular & cellular proteomics : MCP* 2007, *6*, 611-23.

(4) Bech-Serra, J. J.; Santiago-Josefat, B.; Esselens, C.; Saftig, P.; Baselga, J.; Arribas, J.; Canals, F.: Proteomic identification of desmoglein 2 and activated leukocyte cell adhesion molecule as substrates of ADAM17 and ADAM10 by difference gel electrophoresis. *Molecular and cellular biology* 2006, *26*, 5086-95.

(5) Starr, A. E.; Bellac, C. L.; Dufour, A.; Goebeler, V.; Overall, C. M.: Biochemical characterization and N-terminomics analysis of leukolysin, the membrane-type 6 matrix metalloprotease (MMP25): chemokine and vimentin cleavages enhance cell migration and macrophage phagocytic activities. *The Journal of biological chemistry* 2012, *287*, 13382-95.

(6) Kuhn, P. H.; Koroniak, K.; Hogl, S.; Colombo, A.; Zeitschel, U.; Willem, M.; Volbracht, C.; Schepers, U.; Imhof, A.; Hoffmeister, A.; Haass, C.; Rossner, S.; Brase, S.; Lichtenthaler, S. F.: Secretome protein enrichment identifies physiological BACE1 protease substrates in neurons. *The EMBO journal* 2012, *31*, 3157-68.

(7) Prudova, A.; auf dem Keller, U.; Butler, G. S.; Overall, C. M.: Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. *Molecular & cellular proteomics : MCP* 2010, *9*, 894-911.

(8) auf dem Keller, U.; Prudova, A.; Eckhard, U.; Fingleton, B.; Overall, C. M.: Systems-level analysis of proteolytic events in increased vascular permeability and complement activation in skin inflammation. *Science signaling* 2013, *6*, rs2.

(9) Jefferson, T.; Auf dem Keller, U.; Bellac, C.; Metz, V. V.; Broder, C.; Hedrich, J.; Ohler, A.; Maier, W.; Magdolen, V.; Sterchi, E.; Bond, J. S.; Jayakumar, A.; Traupe, H.; Chalaris, A.; Rose-John, S.; Pietrzik, C. U.; Postina, R.; Overall, C. M.; Becker-Pauly, C.: The substrate degradome of meprin metalloproteases reveals an unexpected proteolytic link between meprin beta and ADAM10. *Cellular and molecular life sciences : CMLS* 2013, *70*, 309-33.

(10) Arribas, J.; Esselens, C.: ADAM17 as a therapeutic target in multiple diseases. *Curr Pharm Des* 2009, *15*, 2319-35.

(11) Scheller, J.; Chalaris, A.; Garbers, C.; Rose-John, S.: ADAM17: a molecular switch to control inflammation and tissue regeneration. *Trends Immunol* 2011, *32*, 380-7.

(12) Pruessmeyer, J.; Ludwig, A.: The good, the bad and the ugly substrates for ADAM10 and ADAM17 in brain pathology, inflammation and cancer. *Semin Cell Dev Biol* 2009, *20*, 164-74.

(13) van der Vorst, E. P.; Keijbeck, A. A.; de Winther, M. P.; Donners, M. M.: A disintegrin and metalloproteases: molecular scissors in angiogenesis, inflammation and atherosclerosis. *Atherosclerosis* 2012, *224*, 302-8.

(14) Blobel, C. P.: ADAMs: key components in EGFR signalling and development. *Nature reviews. Molecular cell biology* 2005, *6*, 32-43.

(15) Franovic, A.; Robert, I.; Smith, K.; Kurban, G.; Pause, A.; Gunaratnam, L.; Lee, S.: Multiple acquired renal carcinoma tumor capabilities abolished upon silencing of ADAM17. *Cancer research* 2006, *66*, 8083-90.

(16) Hall, K. C.; Hill, D.; Otero, M.; Plumb, D. A.; Froemel, D.; Dragomir, C. L.; Maretzky, T.; Boskey, A.; Crawford, H. C.; Selleri, L.; Goldring, M. B.; Blobel, C. P.: ADAM17 controls endochondral ossification by regulating terminal differentiation of chondrocytes. *Molecular and cellular biology* 2013, *33*, 3077-90.

(17) Breshears, L. M.; Schlievert, P. M.; Peterson, M. L.: A disintegrin and metalloproteinase 17 (ADAM17) and epidermal growth factor receptor (EGFR) signaling drive the epithelial response to

Staphylococcus aureus toxic shock syndrome toxin-1 (TSST-1). *The Journal of biological chemistry* 2012, *287*, 32578-87.

(18) Fan, H.; Derynck, R.: Ectodomain shedding of TGF-alpha and other transmembrane proteins is induced by receptor tyrosine kinase activation and MAP kinase signaling cascades. *The EMBO journal* 1999, *18*, 6962-72.

(19) Garbers, C.; Janner, N.; Chalaris, A.; Moss, M. L.; Floss, D. M.; Meyer, D.; Koch-Nolte, F.; Rose-John, S.; Scheller, J.: Species specificity of ADAM10 and ADAM17 proteins in interleukin-6 (IL-6) trans-signaling and novel role of ADAM10 in inducible IL-6 receptor shedding. *The Journal of biological chemistry* 2011, *286*, 14804-11.

(20) Althoff, K.; Reddy, P.; Voltz, N.; Rose-John, S.; Mullberg, J.: Shedding of interleukin-6 receptor and tumor necrosis factor alpha. Contribution of the stalk sequence to the cleavage pattern of transmembrane proteins. *European journal of biochemistry / FEBS* 2000, *267*, 2624-31.

(21) Malapeira, J.; Esselens, C.; Bech-Serra, J. J.; Canals, F.; Arribas, J.: ADAM17 (TACE) regulates TGFbeta signaling through the cleavage of vasorin. *Oncogene* 2011, *30*, 1912-22.

(22) Gooz, P.; Gooz, M.; Baldys, A.; Hoffman, S.: ADAM-17 regulates endothelial cell morphology, proliferation, and in vitro angiogenesis. *Biochem Biophys Res Commun* 2009, *380*, 33-8.

(23) Das, S.; Czarnek, M.; Bzowska, M.; Mezyk-Kopec, R.; Stalinska, K.; Wyroba, B.; Sroka, J.; Jucha, J.; Deneka, D.; Stoklosa, P.; Ogonek, J.; Swartz, M. A.; Madeja, Z.; Bereta, J.: ADAM17 silencing in mouse colon carcinoma cells: the effect on tumoricidal cytokines and angiogenesis. *PloS one* 2012, *7*, e50791.

(24) Saftig, P.; Reiss, K.: The "A Disintegrin And Metalloproteases" ADAM10 and ADAM17: novel drug targets with therapeutic potential? *Eur J Cell Biol* 2011, *90*, 527-35.

(25) Makridakis, M.; Vlahou, A.: Secretome proteomics for discovery of cancer biomarkers. *Journal of proteomics* 2010, *73*, 2291-305.

(26) Jacobs, J. M.; Waters, K. M.; Kathmann, L. E.; Camp, D. G., 2nd; Wiley, H. S.; Smith, R. D.; Thrall, B. D.: The mammary epithelial cell secretome and its regulation by signal transduction pathways. *Journal of proteome research* 2008, *7*, 558-69.

(27) Karagiannis, G. S.; Pavlou, M. P.; Diamandis, E. P.: Cancer secretomics reveal pathophysiological pathways in cancer molecular oncology. *Mol Oncol* 2010, *4*, 496-510.

(28) Le Gall, S. M.; Maretzky, T.; Issuree, P. D.; Niu, X. D.; Reiss, K.; Saftig, P.; Khokha, R.; Lundell, D.; Blobel, C. P.: ADAM17 is regulated by a rapid and reversible mechanism that controls access to its catalytic site. *J Cell Sci* 2010, *123*, 3913-22.

(29) Cox, J.; Mann, M.: MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 2008, *26*, 1367-72.

(30) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M.: Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* 2011, *10*, 1794-805.

(31) Chen, E. Y.; Xu, H.; Gordonov, S.; Lim, M. P.; Perkins, M. H.; Ma'ayan, A.: Expression2Kinases: mRNA profiling linked to multiple upstream regulatory layers. *Bioinformatics* 2012, *28*, 105-11.

(32) Batista, F. A.; Trivella, D. B.; Bernardes, A.; Gratieri, J.; Oliveira, P. S.; Figueira, A. C.; Webb, P.; Polikarpov, I.: Structural insights into human peroxisome proliferator activated receptor delta (PPAR-delta) selective ligand binding. *PloS one* 2012, *7*, e33643.

(33) Paes Leme, A. F.; Kitano, E. S.; Furtado, M. F.; Valente, R. H.; Camargo, A. C.; Ho, P. L.; Fox, J. W.; Serrano, S. M.: Analysis of the subproteomes of proteinases and heparin-binding toxins of eight Bothrops venoms. *Proteomics* 2009, *9*, 733-45.

(34) Paes Leme, A. F.; Escalante, T.; Pereira, J. G.; Oliveira, A. K.; Sanchez, E. F.; Gutierrez, J. M.; Serrano, S. M.; Fox, J. W.: High resolution analysis of snake venom metalloproteinase (SVMP) peptide bond cleavage specificity using proteome based peptide libraries and mass spectrometry. *Journal of proteomics* 2011, *74*, 401-10.

(35) Paes Leme, A. F.; Sherman, N. E.; Smalley, D. M.; Sizukusa, L. O.; Oliveira, A. K.; Menezes, M. C.; Fox, J. W.; Serrano, S. M.: Hemorrhagic activity of HF3, a snake venom metalloproteinase: insights from the proteomic analysis of mouse skin and blood plasma. *Journal of proteome research* 2012, *11*, 279-91.

(36) Colzani, M.; Schutz, F.; Potts, A.; Waridel, P.; Quadroni, M.: Relative protein quantification by isobaric SILAC with immonium ion splitting (ISIS). *Molecular & cellular proteomics : MCP* 2008, *7*, 927-37.

(37) Zhang, G.; Neubert, T. A.: Comparison of three quantitative phosphoproteomic strategies to study receptor tyrosine kinase signaling. *Journal of proteome research* 2011, *10*, 5454-62.

(38) Duffy, M. J.; Mullooly, M.; O'Donovan, N.; Sukor, S.; Crown, J.; Pierce, A.; McGowan, P. M.: The ADAMs family of proteases: new biomarkers and therapeutic targets for cancer? *Clin Proteomics* 2011, *8*, 9.

(39) Mochizuki, S.; Okada, Y.: ADAMs in cancer cell proliferation and progression. *Cancer science* 2007, *98*, 621-8.

(40) Lopez-Otin, C.; Overall, C. M.: Protease degradomics: a new challenge for proteomics. *Nature reviews. Molecular cell biology* 2002, *3*, 509-19.

(41) Schilling, O.; Overall, C. M.: Proteomic discovery of protease substrates. *Curr Opin Chem Biol* 2007, *11*, 36-45.

(42) Dang, M.; Armbruster, N.; Miller, M. A.; Cermeno, E.; Hartmann, M.; Bell, G. W.; Root, D. E.; Lauffenburger, D. A.; Lodish, H. F.; Herrlich, A.: Regulated ADAM17-dependent EGF family ligand release by substrate-selecting signaling pathways. *Proceedings of the National Academy of Sciences of the United States of America* 2013, *110*, 9776-81.

(43) Whiteside, T. L.: The tumor microenvironment and its role in promoting tumor growth. *Oncogene* 2008, *27*, 5904-12.

(44) De Boeck, A.; Hendrix, A.; Maynard, D.; Van Bockstal, M.; Daniels, A.; Pauwels, P.; Gespach, C.; Bracke, M.; De Wever, O.: Differential secretome analysis of cancer-associated fibroblasts and bone marrow-derived precursors to identify microenvironmental regulators of colon cancer progression. *Proteomics* 2013, *13*, 379-88.

(45) Chen, S. T.; Pan, T. L.; Juan, H. F.; Chen, T. Y.; Lin, Y. S.; Huang, C. M.: Breast tumor microenvironment: proteomics highlights the treatments targeting secretome. *Journal of proteome research* 2008, *7*, 1379-87.

(46) Kessenbrock, K.; Plaks, V.; Werb, Z.: Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* 2010, *141*, 52-67.

(47) Gallagher, P. G.; Bao, Y.; Prorock, A.; Zigrino, P.; Nischt, R.; Politi, V.; Mauch, C.; Dragulev, B.; Fox, J. W.: Gene expression profiling reveals cross-talk between melanoma and fibroblasts: implications for host-tumor interactions in metastasis. *Cancer research* 2005, *65*, 4134-46.

(48) McGowan, P. M.; Ryan, B. M.; Hill, A. D.; McDermott, E.; O'Higgins, N.; Duffy, M. J.: ADAM-17 expression in breast cancer correlates with variables of tumor progression. *Clin Cancer Res* 2007, *13*, 2335-43.

(49) Zheng, X.; Jiang, F.; Katakowski, M.; Zhang, Z. G.; Lu, Q. E.; Chopp, M.: ADAM17 promotes breast cancer cell malignant phenotype through EGFR-PI3K-AKT activation. *Cancer Biol Ther* 2009, *8*, 1045-54.

(50) Pruessmeyer, J.; Martin, C.; Hess, F. M.; Schwarz, N.; Schmidt, S.; Kogel, T.; Hoettecke, N.; Schmidt, B.; Sechi, A.; Uhlig, S.; Ludwig, A.: A disintegrin and metalloproteinase 17 (ADAM17) mediates inflammation-induced shedding of syndecan-1 and -4 by lung epithelial cells. *The Journal of biological chemistry* 2010, *285*, 555-64.

(51) Rowlands, D. J.; Islam, M. N.; Das, S. R.; Huertas, A.; Quadri, S. K.; Horiuchi, K.; Inamdar, N.; Emin, M. T.; Lindert, J.; Ten, V. S.; Bhattacharya, S.; Bhattacharya, J.: Activation of TNFR1 ectodomain shedding by mitochondrial Ca2+ determines the severity of inflammation in mouse lung microvessels. *J Clin Invest* 2011, *121*, 1986-99.

(52) Stamenkovic, I.; Yu, Q.: Shedding light on proteolytic cleavage of CD44: the responsible sheddase and functional significance of shedding. *J Invest Dermatol* 2009, *129*, 1321-4.

(53) Maretzky, T.; Evers, A.; Zhou, W.; Swendeman, S. L.; Wong, P. M.; Rafii, S.; Reiss, K.; Blobel, C. P.: Migration of growth factor-stimulated epithelial and endothelial cells depends on EGFR transactivation by ADAM17. *Nature communications* 2011, *2*, 229.

(54) Xu, P.; Derynck, R.: Direct activation of TACE-mediated ectodomain shedding by p38 MAP kinase regulates EGF receptor-dependent cell proliferation. *Molecular cell* 2010, *37*, 551-66.

(55) Chen, X.; Chen, L.; Chen, J.; Hu, W.; Gao, H.; Xie, B.; Wang, X.; Yin, Z.; Li, S.; Wang, X.: ADAM17 promotes U87 glioblastoma stem cell migration and invasion. *Brain research* 2013.

(56) Zheng, Y.; Xia, Y.; Hawke, D.; Halle, M.; Tremblay, M. L.; Gao, X.; Zhou, X. Z.; Aldape, K.; Cobb, M. H.; Xie, K.; He, J.; Lu, Z.: FAK phosphorylation by ERK primes ras-induced tyrosine dephosphorylation of FAK mediated by PIN1 and PTP-PEST. *Molecular cell* 2009, *35*, 11-25.

(57) Wang, X.; Oka, T.; Chow, F. L.; Cooper, S. B.; Odenbach, J.; Lopaschuk, G. D.; Kassiri, Z.; Fernandez-Patron, C.: Tumor necrosis factor-alpha-converting enzyme is a key regulator of agonist-induced cardiac hypertrophy and fibrosis. *Hypertension* 2009, *54*, 575-82.

(58) Li, Y. Q.; Yan, J. P.; Xu, W. L.; Wang, H.; Xia, Y. J.; Wang, H. J.; Zhu, Y. Y.; Huang, X. J.: ADAM17 mediates MMP9 expression in lung epithelial cells. *PloS one* 2013, *8*, e51701.

(59) Xiao, L. J.; Lin, P.; Lin, F.; Liu, X.; Qin, W.; Zou, H. F.; Guo, L.; Liu, W.; Wang, S. J.; Yu, X. G.: ADAM17 targets MMP-2 and MMP-9 via EGFR-MEK-ERK pathway activation to promote prostate cancer cell invasion. *International journal of oncology* 2012, *40*, 1714-24.

(60) Chana, R. S.; Lewington, A. J.; Brunskill, N. J.: Differential effects of peroxisome proliferator activated receptor-gamma (PPAR gamma) ligands in proximal tubular cells: thiazolidinediones are partial PPAR gamma agonists. *Kidney international* 2004, *65*, 2081-90.

(61) Chen, L.; Necela, B. M.; Su, W.; Yanagisawa, M.; Anastasiadis, P. Z.; Fields, A. P.; Thompson, E. A.: Peroxisome proliferator-activated receptor gamma promotes epithelial to mesenchymal transformation by Rho GTPase-dependent activation of ERK1/2. *The Journal of biological chemistry* 2006, *281*, 24575-87.

(62) Takahashi, H.; Fujita, K.; Fujisawa, T.; Yonemitsu, K.; Tomimoto, A.; Ikeda, I.; Yoneda, M.; Masuda, T.; Schaefer, K.; Saubermann, L. J.; Shimamura, T.; Saitoh, S.; Tachibana, M.; Wada, K.; Nakagama, H.; Nakajima, A.: Inhibition of peroxisome proliferator-activated receptor gamma activity in esophageal carcinoma cells results in a drastic decrease of invasive properties. *Cancer science* 2006, *97*, 854-60.

(63) Iozzo, R. V.: The family of the small leucine-rich proteoglycans: key regulators of matrix assembly and cellular growth. *Crit Rev Biochem Mol Biol* 1997, *32*, 141-74.

(64) Gu, X. S.; Lei, J. P.; Shi, J. B.; Lian, W. L.; Yang, X.; Zheng, X.; Qin, Y. W.: Mimecan is involved in aortic hypertrophy induced by sinoaortic denervation in rats. *Mol Cell Biochem* 2011, *352*, 309-16.

(65) Kampmann, A.; Fernandez, B.; Deindl, E.; Kubin, T.; Pipp, F.; Eitenmuller, I.; Hoefer, I. E.; Schaper, W.; Zimmermann, R.: The proteoglycan osteoglycin/mimecan is correlated with arteriogenesis. *Mol Cell Biochem* 2009, *322*, 15-23.

(66) Zoeller, J. J.; Whitelock, J. M.; Iozzo, R. V.: Perlecan regulates developmental angiogenesis by modulating the VEGF-VEGFR2 axis. *Matrix Biol* 2009, *28*, 284-91.

(67) Patel, V. N.; Knox, S. M.; Likar, K. M.; Lathrop, C. A.; Hossain, R.; Eftekhari, S.; Whitelock, J. M.; Elkin, M.; Vlodavsky, I.; Hoffman, M. P.: Heparanase cleavage of perlecan heparan sulfate modulates FGF10 activity during ex vivo submandibular gland branching morphogenesis. *Development* 2007, *134*, 4177-86.

(68) Ishijima, M.; Suzuki, N.; Hozumi, K.; Matsunobu, T.; Kosaki, K.; Kaneko, H.; Hassell, J. R.; Arikawa-Hirasawa, E.; Yamada, Y.: Perlecan modulates VEGF signaling and is essential for vascularization in endochondral bone formation. *Matrix Biol* 2012, *31*, 234-45.

(69) Hohenester, E.; Yurchenco, P. D.: Laminins in basement membrane assembly. *Cell Adh Migr* 2013, *7*, 56-63.

(70) Barascuk, N.; Vassiliadis, E.; Zheng, Q.; Wang, Y.; Wang, W.; Larsen, L.; Rasmussen, L. M.; Karsdal, M. A.: Levels of Circulating MMCN-151, a Degradation Product of Mimecan, Reflect Pathological Extracellular Matrix Remodeling in Apolipoprotein E Knockout Mice. *Biomarker insights* 2011, *6*, 97-106.

(71) Whitelock, J. M.; Murdoch, A. D.; Iozzo, R. V.; Underwood, P. A.: The degradation of human endothelial cell-derived perlecan and release of bound basic fibroblast growth factor by stromelysin, collagenase, plasmin, and heparanases. *The Journal of biological chemistry* 1996, *271*, 10079-86.

(72) Caescu, C. I.; Jeschke, G. R.; Turk, B. E.: Active-site determinants of substrate recognition by the metalloproteinases TACE and ADAM10. *Biochem J* 2009, *424*, 79-88.

(73) Peschon, J. J.; Slack, J. L.; Reddy, P.; Stocking, K. L.; Sunnarborg, S. W.; Lee, D. C.; Russell, W. E.; Castner, B. J.; Johnson, R. S.; Fitzner, J. N.; Boyce, R. W.; Nelson, N.; Kozlosky, C. J.; Wolfson, M. F.; Rauch, C. T.; Cerretti, D. P.; Paxton, R. J.; March, C. J.; Black, R. A.: An essential role for ectodomain shedding in mammalian development. *Science* 1998, *282*, 1281-4.

(74) Reddy, P.; Slack, J. L.; Davis, R.; Cerretti, D. P.; Kozlosky, C. J.; Blanton, R. A.; Shows, D.; Peschon, J. J.; Black, R. A.: Functional analysis of the domain structure of tumor necrosis factor-alpha converting enzyme. *The Journal of biological chemistry* 2000, *275*, 14608-14.

(75) Kalus, I.; Bormann, U.; Mzoughi, M.; Schachner, M.; Kleene, R.: Proteolytic cleavage of the neural cell adhesion molecule by ADAM17/TACE is involved in neurite outgrowth. *J Neurochem* 2006, *98*, 78-88.

(76) Ruhe, J. E.; Streit, S.; Hart, S.; Ullrich, A.: EGFR signaling leads to downregulation of PTP-LAR via TACE-mediated proteolytic processing. *Cell Signal* 2006, *18*, 1515-27.

(77) Rovida, E.; Paccagnini, A.; Del Rosso, M.; Peschon, J.; Dello Sbarba, P.: TNF-alpha-converting enzyme cleaves the macrophage colony-stimulating factor receptor in macrophages undergoing activation. *J Immunol* 2001, *166*, 1583-9.

(78) Guo, L.; Eisenman, J. R.; Mahimkar, R. M.; Peschon, J. J.; Paxton, R. J.; Black, R. A.; Johnson, R. S.: A proteomic approach for the identification of cell-surface proteins shed by metalloproteases. *Molecular & cellular proteomics : MCP* 2002, *1*, 30-6.

(79) Weskamp, G.; Mendelson, K.; Swendeman, S.; Le Gall, S.; Ma, Y.; Lyman, S.; Hinoki, A.; Eguchi, S.; Guaiquil, V.; Horiuchi, K.; Blobel, C. P.: Pathological neovascularization is reduced by inactivation of ADAM17 in endothelial cells but not in pericytes. *Circ Res* 2010, *106*, 932-40.

(80) Garton, K. J.; Gough, P. J.; Philalay, J.; Wille, P. T.; Blobel, C. P.; Whitehead, R. H.; Dempsey, P. J.; Raines, E. W.: Stimulated shedding of vascular cell adhesion molecule 1 (VCAM-1) is mediated by tumor necrosis factor-alpha-converting enzyme (ADAM 17). *The Journal of biological chemistry* 2003, *278*, 37459-64.

(81) Stoeck, A.; Keller, S.; Riedle, S.; Sanderson, M. P.; Runz, S.; Le Naour, F.; Gutwein, P.; Ludwig, A.; Rubinstein, E.; Altevogt, P.: A role for exosomes in the constitutive and stimulus-induced ectodomain cleavage of L1 and CD44. *Biochem J* 2006, *393*, 609-18.

CAPÍTULO 5

Mass spectrometry-based proteomics revealed Glypican-1 as a novel ADAM17 substrate

Rebeca Kawahara; Daniela Campos Granato; Sami yokoo; Daniel Maragno Trindade; and Adriana Franco Paes Leme

CAPSULE

Background: ADAM17 is a major sheddase involved in the regulation of many biological processes in cancer.

Results: Glypican-1 is a direct substrate of ADAM17 and its function can be modulated by cleavage.

Conclusion: The function of glypican-1 is modulated by ADAM17.

Significance: The function of ADAM17 can be also studied by its substrates.

ABSTRACT

ADAM17 (a disintegrin and metalloproteinase 17) is a plasma membrane metalloprotease involved in proteolytic release of the extracellular domain of many cell surface molecules, a process known as ectodomain shedding. Through this process, ADAM17 is implicated in several aspects of tumor growth and metastasis in a broad range of tumors, including head and neck squamous cell carcinomas (HNSCC). However, inhibition of metalloproteases as a therapeutic approach has failed, since there is a limited knowledge of individual protease degradome as well as function of cleaved substrates in a cellular context. In this study, glypican-1 (GPC1) was revealed as a new substrate for ADAM17 and its shedding was confirmed to be metalloprotease-dependent, induced by a pleiotropic agent (PMA) and physiologic ligand (EGF), and inhibited by marimastat. Considering GPC1 localized preferentially in cholesterol-rich lipid rafts in the membrane, we demonstrated that the depletion of cholesterol modulated the localization of GPC1 in the cell surface, enhancing PMA-induced GPC1 shedding. Besides, we showed that GPC1 is a direct target of ADAM17 which was validated using ADAM17 knockdown and mass spectrometry-based cleavage site approach. The soluble GPC1 partners pointed to biological functions and pathways related mainly to cellular movement, adhesion and proliferation, events that were also modulated by GPC1 cleavage. Altogether, we showed that the function of GPC1 may be modulated by proteolysis signaling.

INTRODUCTION

Proteolytic cleavage of plasma membrane proteins resulting in ectodomain release is a process known as 'shedding'. This process represents a common molecular mechanism for regulating the biological activity of a range of cell-surface proteins (1). Members of ADAM family have emerged as major ectodomain shedding proteinases responsible for the production of soluble, functional protein ectodomains that can initiate or inhibit autocrine and paracrine signaling (2). ADAM17, also known as tumor necrosis factor-α (TNF) converting enzyme (TACE) (3), is a primary sheddase of several biologically important membrane-anchored proteins, including TNF, epithelial cell growth factors, cell adhesion molecules, and ligands of EGFR (e.g. soluble transforming growth factor-α (TGF-α); amphiregulin, and heparin-binding epidermal growth factor (HB-EGF)) (4,5). The shedding of EGFR-ligands, growth factors and cytokines by ADAM17 has been associated with development of cancer (6-8) and other diseases such as inflammation, septic shock and rheumatoid arthritis (9-11).

Among cancer types, ADAM17 is involved in the development of squamous cell carcinoma of the head and neck (HNSCC) which is the sixth most common cancer worldwide (12). Besides ADAM17 being overexpressed in HNSCC (13,14), its activity has been associated with advanced clinical stage of HNSCC as well as tumors recurrence (15). Recently, our group (16) showed that ADAM17 is implicated in oral cancer development, such as increasing in tumor size and proliferation in an orthotopic murine tumor model and Huang, et al. (17) demonstrated the importance of ADAM17 in HNSCC cell proliferation and migration. The major mechanism by which ADAM17 contributes to tumor progression is related to the activation of growth factor receptors of EGFR, which is widely studied oncogene in head and neck tumors and a therapeutic target in oral squamous cell carcinoma (OSCC) (18). Although ADAM17 is essential in the activation of EGFR signaling, therapeutic strategies using ADAM17 inhibitors have failed clinically (19). Therefore, there is still a need for deeper investigation regarding the role of metalloproteases in cancer, especially in the context of repertoire substrates and the function of their cleavage.

To explore the repertoire of ADAM17 substrates in HNSCC we performed two mass spectrometry-based proteomics strategies: first, we used cell-surface biotinylation of HNSCC cell line, the squamous cell carcinoma cell line derived from tongue (SCC-9), followed by incubation with recombinant ADAM17. Second, we performed label-free analysis of the secretome of ADAM17-scrambled and knockdown SCC-9 cell lines. Both approaches showed glypican-1 (GPC1) as a novel substrate candidate for ADAM17. We also demonstrated that the cleavage is

induced by phorbol 12-myristate 13-acetate (PMA) and epidermal growth factor (EGF) while it is inhibited by broad-spectrum inhibitor of metalloproteases, marimastat, in a cell-based assay. Moreover, we showed that depletion of cholesterol increased PMA-induced GPC1 shedding, possibly by modulating GPC1 localization in the cell surface. The involvement of ADAM17 in GPC1 shedding was also validated in the secretome of SCC-9 and A431 cell lines by western blot and by N-termini labeling followed by MS.

To evaluate the functional role of soluble GPC1, its interaction partners were identified using co-immunoprecipitation (IP) followed by MS. Using bioinformatics analyses we showed an involvement of GPC1 complex in overrepresented processes such as proliferation, adhesion, and cell movement, which were indeed validated to be modulated by GPC1 up-regulation or treatment with PMA and marimastat. Further, cleaved GPC1 negatively regulated cell migration. Taken together, this study underscored GPC1 as a novel substrate of ADAM17 and revealed that the function of GPC1 may be modulated by proteolysis signaling.

EXPERIMENTAL PROCEDURES

Mammalian human cell lines

The human OSCC (oral squamous cell carcinoma) cell line, SCC-9, was obtained from American Type Culture Collection (ATCC), and cultured as recommended. SCC-9 cells are originated from human squamous carcinoma of the tongue. Human Epidermoid Carcinoma A431 (epidermoid carcinoma cell line originated from skin) cell line was grown in Roswell Park Memorial Institute (RPMI) −1640 medium supplemented with 10% FBS and antibiotics at 37°C in a 5% CO2 air atmosphere.

Generation of stable ADAM17 knockdown in SCC-9 and A431 cell line

The lentiviral particle production and transduction of SCC-9 and A431 cells for ADAM17 knockdown were achieved with the Mission shRNA Vector pLKO.1-puro System using shRNA plasmid pLKO.1-Neo-CMV-tGFP (Sigma-Aldrich, Munich, Germany). The experimental procedures were performed according to the manufacturers' instructions. After transduction, G418 antibiotic was added to the cultures at a final concentration of 0.8 mg/ml and incubated for approximately 1 week until the complete death of untransduced cells.

Recombinant ADAM17 baculovirus production

The extracellular domain sequence of ADAM17 (Met1 to Asn671, Uniprot access P78536), including signal peptide, pro-domain, metalloproteases domain, disintegrin domain, cysteine-rich region and EGF like domain, was amplified from HEK293 cDNA library using ADAM17 forward primer 5'- AAGCTAGCATGAGGCAGTCTCTCCTATTCCTG-3' and ADAM17 reverse primer 5'- TTGAATTCGTTGTCTGCTAAAAACTTTCCAAAAAG-3' and subcloned into vector pGEM-T, resulting in plasmid pGEM-T-ADAM17. This insert was subcloned into pcDNA 4/TO/myc-His-A by using BamHI and EcoRI sites. After that, pcDNA 4/TO/myc-His-ADAM17 was digested with BamHI and PmeI, generating ADAM17 sequence in fusion with His tag and subcloned into pFastBAC Dual+EGFP (pFBDg), which had the EGFP cDNA cloned under p10 promoter, as previously described (20). This last construct was used to generate recombinant baculovirus using the Bac-to-bac system (Invitrogen) according to the manufacturers' instructions.

Expression and purification of ADAM17-671

Sf9 cells (Invitrogen), used exclusively for the production of baculovirus, were cultivated in SF900II SFM (Gibco) supplemented with 10% FBS and 1× PenStrep (Gibco) at 27˚C. High Five™ (Invitrogen) cells were adapted to grow in suspension culture in Express Five™ serum free media (Gibco) supplemented with 20 mM L-Glutamine (Gibco) and 1× PenStrep (Gibco) and maintained in incubator (ThermoForma) at 28°C, 200 rpm. For recombinant ADAM17-671 production, High Five cells were scaled up from the stock culture to a cell density of 1×10^6 in a 2L Erlenmeyer flasks containing 500 mL of Express Five media and incubated at 26°C at 140 rpm. 24 h post-inoculation, cells were infected with the recombinant baculo virus produced in Sf-9 cells, at a multiplicity of infection (m.o.i.) of 1 plaque-forming unit (pfu) per cell. High Five culture media were harvested after 72 hours post-infection by centrifugation at 1500 × rpm for five minutes and cell-free supernatant containing extracellular ADAM17-671 was used for purification. The purification was performed using Ni-sepharose High Performance (GE Healthcare), equilibrated with 20 mM Tris-HCl pH 7.5; 250 mM NaCl and 20 mM imidazole (affinity buffer) and incubated for 2 h with the extracellular media. After that, the resin was washed with affinity buffer and eluted with affinity buffer containing increasing concentrations of imidazole 100 mM, 200 mM, 500 mM, 1 M. Recombinant ADAM17 was visualized using Coomassie blue SDS-PAGE and immunoblotting (anti-His).

ADAM17 activity assays using fluorogenic peptides and cell-based assay

The enzymatic activity of recombinant ADAM17 was measured at 37°C by monitoring the increased fluorescence intensity upon degradation of the fluorogenic peptides, derived from the pro tumor necrosis factor-α (TNF-α), Mca-Pro-Leu-Ala-Gln-Ala-Val-Dpa-Arg-Ser-Ser-Ser-Arg-NH_2 (RD system), at 320 nm and 405 nm. The standard assay mixture contained 25 mM TrisHCl, 2.5 µM $ZnCl_2$, 0.005% Brij-35 (w/v), pH 9.0.

The activity of recombinant ADAM17 was also tested in HEK293 cells stably transfected with HB-EGF-AP. Briefly, these cells were seeded in 24-well plates (Corning) and incubated for 24 h, followed by starvation of 4 h. Crescent volumes (5, 10 and 20 µl) of the solution containing the purified recombinant ADAM17 were added in 500 µl of DMEM phenol-free medium for 1 h/37^{o}C. Cell-conditioned medium was harvested and mixed 1:1 with a solution of the alkaline phosphatase substrate (0.5 M Tris-HCl, pH 9.5, containing 5 mM p-nitrophenyl phosphate disodium, 1 mM diethanolamine, 50 µM $MgCl_2$, 150 mM NaCl, 5 mM EDTA) in 96-well plates. The reactions were incubated at room temperature overnight in the dark, followed by quantification of alkaline phosphatase activity by measuring the absorbance at 405 nm.

Cell surface biotinylation and incubation with recombinant ADAM17-671

SCC-9 cells were rinsed three times with phosphate-buffered saline (PBS) and incubated with 0.5 mg/ml EZ-Link Sulfo-NHS-SS-biotin (Thermo Fisher Scientific, Bonn, Germany) for 30 min at 4^{o}C on a rocking platform. After washing three times with PBS, the cells were incubated with 20 mM glycine in PBS for 10 min to quench the residual biotin and treated with or without recombinant ADAM17-671 in DMEM phenol/serum-free media for 16 h at 37 ^{o}C. After that, the extracellular media were collected, centrifuged for 5 min at 500 g, and incubated with streptavidin-agarose beads for 2 h at 4 ^{o}C while rotating. Upon reduction with 50 mM DTT, the biotinylated proteins were released from the beads and the eluted material was diluted to a final concentration of 10 mM DTT, followed by alkylation (14 mM iodoacetamide, 30 min at room temperature in the dark) and digestion with trypsin (1:50, w/w) (Promega). The reaction was stopped with 0.4% TFA and desalted using Sep-Pak cartridges (Waters). The samples were dried in a vacuum concentrator and reconstituted in 20 µl of 0.1% formic acid. One experiment was performed for each condition.

Analysis of biotinylated proteins by LC-MS/MS and data analysis using MaxQuant and Perseus

An aliquot of 4.5 µl of the resulting peptide mixture was analyzed on an ETD-enabled LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) connected to a nanoflow liquid chromatography (LC-MS/MS) instrument by an EASY-nLC system (Proxeon Biosystem) with a Proxeon nanoelectrospray ion source. Peptides were separated with a 2-90% acetonitrile gradient in 0.1% formic acid using an analytical column PicoFrit Column (20 cm x ID75 µm, 5 µm particle size, New Objective) at a flow of 300 nl/min over 45 min. The nanoelectrospray voltage was set to 2.2 kV and the source temperature was 275°C. All of the instrument methods were set up in the data-dependent acquisition mode. The full scan MS spectra (m/z 300-1600) were acquired in the Orbitrap analyzer after accumulation to a target value of $1x10^6$. The resolution in the Orbitrap was set to r= 60,000 and the 20 most intense peptide ions with charge states ≥ 2 were sequentially isolated to a target value of 5,000 and fragmented in the linear ion trap using low-energy CID (normalized collision energy of 35%). The signal threshold for triggering an MS/MS event was set to 1000 counts. Dynamic exclusion was enabled with an exclusion size list of 500, exclusion duration of 60 s, and a repeat count of 1. An activation q= 0.25 and activation time of 10 ms were used.

The raw files were processed using the MaxQuant version 1.2.7.429 (21) and the MS/MS spectra were searched using the Andromeda search engine (22) against the Uniprot Human Protein Database (release July 11, 2012; 69,711 entries). The initial maximal allowed mass tolerance was set to 20 ppm for precursor and then set to 6 ppm in the main search and to 0.5 Da for fragment ions. Enzyme specificity was set to trypsin with a maximum of two missed cleavages.

Carbamidomethylation of cysteine (57.02 Da) was set as a fixed modification, and oxidation of methionine (15.99 Da), thioacylation of amino termini (88.03 Da) and protein N-terminal acetylation (42.01 Da) were selected as variable modifications. Bioinformatics analysis was performed using the software Perseus v.1.2.7.4 (21) available in the MaxQuant environment and reverse and contaminant entries were excluded from further analysis. Peptide intensity values were considered to relatively compare the abundance of proteins present in the condition incubated with ADAM17 and control.

Secretome of ADAM17-scrambled and knockdown SCC-9 cells

SCC-9 scramble cells and knockdown for ADAM17 were grown at 80% confluence, gently washed three times in PBS and incubated for 24 h in serum-free medium. The media were

collected, and a final concentration of 1 mM EDTA and 1 mM PMSF were added to the media. Cell debris was eliminated by centrifugation at 4,000 x *rpm* for 5 min at 4°C and subsequently concentrated using a 3000-Dalton centrifugation filter (Millipore, Billerica, MA) at 4,000 x *rpm* at 4°C. Protein concentrations were determined using a Bradford assay (Bio-Rad, Hercules, CA, USA). Two independent experiments were performed.

Mass spectrometric analysis of SCC-9 scramble and ADAM17-knockdown and data analysis using Proteome Discoverer and ScaffoldQ+

Proteins (10 µg) were treated with a final concentration of 1.6 M urea, following reduction (5 mM dithiothreitol, 25 min at 56°C), alkylation (14 mM iodoacetamide, 30 min at room temperature in the dark) and digestion with trypsin (1:50, w/w). The reaction was stopped with 0.4% TFA and desalted using C18 Stage tips. The samples were then dried in a vacuum concentrator and reconstituted in 0.1% formic acid. The MS analysis was performed as described before, in a 2–90% acetonitrile gradient over 194 min.

Peak lists (msf) were generated from the raw data files using the Proteome Discoverer software version 1.3 (Thermo Fisher Scientific) with the Sequest search engine and searched against the Uniprot database, with the following parameters: carbamidomethylation as the fixed modification, oxidation of methionine as the variable modification, one trypsin missed cleavage and a tolerance of 10 ppm for precursor and 1 Da for fragment ions. All datasets were processed using the workflow feature in the Proteome Discoverer software, and the resulting search data were further analyzed in the software ScaffoldQ+v.3.3.1. The scoring parameters (Xcorr and Peptide Probability) in the ScaffoldQ+ software were set to obtain a false discovery rate (FDR) of less than 1%. Using the number of total spectra output from the ScaffoldQ+ software, we identified the differentially expressed proteins using spectral counting. A normalization criterion, the quantitative value, was applied to the spectral counts. Student's *t*-test was applied to evaluate differential expressed proteins presented in the secretomes at $p<0.05$.

Immunoblotting

Protein samples were subjected to 10% SDS-PAGE under reduction conditions and transferred to nitrocellulose membranes. The membranes were blocked in 5% dry milk in Tris-Tween-buffered saline (TTBS). The membranes were then incubated overnight at 4°C with anti-glypican-1 (1:1,000; Santa Cruz Biotech), anti-GAPDH (1:1,3000, Bethyl) and anti-ADAM17 (1:1,000, Millipore). The membranes were washed and incubated in horseradish peroxidase-

conjugated secondary antibodies and developed using enhanced chemiluminescence detection according to the manufacturer's instructions (Amersham Biosciences).

Analysis of mRNA Expression levels

Total RNA was obtained using the TRIzol reagent (Invitrogen Corporation) and 2 µg of total RNA were used for retro-transcription using a First-Strand cDNA Synthesis Kit (GE Healthcare). Real-time quantitative PCR (RT_qPCR) was performed using the SYBRH Green PCR Master Mix (Applied Biosystems), and the dissociation curves were performed to confirm the specificity of the products. The threshold cycle (CT) values of the target gene were normalized relative to glyceraldehyde-3-phospate dehydrogenase gene, and relative expression ratios were calculated using the 2-ΔΔ Ct method. Two independent experiments were performed with triplicates. The following PCR primers were used: Glypican-1 forward 5'-GCCGAAATGTGCTCAAGGGC-3' and reverse 5'-ACGCTGCCGATGACACTCTCC-3'; GAPDH forward 5′-GAAGGTGAAGGTCGGAGTCAAC-3′ and reverse 5′-CAGAGTTAAAAGCAGCCCCTGGT-3′; ADAM17 forward 5'-GGACCCCTTCCCAAATAGCA-3' and reverse 5'-ATGGTCCGTGAGATCCTCAAA-3'; ADAM10 forward 5'-CTCCAAGTAGTAATCCAAAGTT-3' and reverse 5'-CTGACGCTGGGGTTGCTG-3'.

Generation of GPC1constructions

Human GPC1 cDNA cloned into the pcDNA3 (kindly provided by Dr. Andreas Frield) was used as template for GPC1 sequence amplification with the following primers: GPC1 Forward 5'-CCGGAATTCCGGGACCCGGCCAGCAAGAGCCGG-3' and GPC1 reverse 5'-CCGCTCGAGCGGTTACCGCCACCGGGGCCTGGC-3'. This sequence was first cloned into pGEM-T and then subcloned into pcDNA3.1+/CMV-AP-Flag, which was constructed from a plasmid encoding HB-EGF-AP kindly provided by Dr. Michael R. Freeman (Department of Surgery, Harvard Medical School, Boston) (23). GPC1 Δ389-558 (fragment), corresponding to the Met1 to Asp388, was constructed using the following primers containing BamHI and EcoRI: forward 5'- CCGGGATCCCGGATGGAGCTCCGGGCCCGAGGCTGGTG-3' and reverse 5' CCGGAATTCCGGTTACTTATCGTCGTCATCCTTGTAGTCGTCGCGGAGCTGGGC-3'.

Stable transfections

A431 cells were stably selected for expression of the construct encoding GPC1 fused to alkaline phosphatase and flag tag, termed "Flag-AP-GPC1. Briefly, cells were transfected with pcDNA 3.1+/Flag-AP-GPC1 using Lipofectamine PLUS (Invitrogen) following manufacturer's instructions. After transfection, G418 (Geneticin) antibiotic was added to cultures at a concentration of 0.6 mg/ml and incubated for about 2 weeks, until complete death of untransfected cells.

Transient transfections

A431 cells were transiently transfected with empty vector pcDNA 3.1+/Flag or pcDNA 3.1+/ GPC1 Δ389-558 (fragment) using the Lipofectamine and Plus reagent (Invitrogen), following manufacturer's instruc instructions.

Shedding assays

A431 cells stably transfected with Flag-AP-GPC1 were seeded into 24-well plates (Corning) at $1x10^5$ cells/well and incubated for 24 h. At the following day, cells were starved for 4 h, treated or not with PMA (100 ng/ml), EGF (100 ng/ml) or marimastat (2μM) or PMA (100 ng/ml) and marimastat (2μM) at the same time for 1h in a phenol and serum free medium. For the cholesterol depletion assay, A431 cells stably transfected with Flag-AP-GPC1 were pre-treated with methyl-β-cyclodextrin (5 mM) in a serum free media, followed by activation with PMA (100 ng/ml) for 1 h. The cleavage of Flag-AP-GPC1 was measured after overnight incubation as described above using 100 μl of conditioned media collected from each well and 100 μl of AP buffer.

Analysis of glypican-1 cleavage site by LC-MS/MS

For cleavage site analysis, recombinant glypican-1 purified from A431 cell line stably transfected with Flag-AP-GPC1 was incubated with or without 0.01 μg of recombinant ADAM17 (residue Arg215-Asn671, R&D Systems) in PBS at a final volume of 20 μl for 16 h at 37^oC. Neo-amino termini of glypican-1 were labeled with 20 mM of EZ-Link Sulfo-NHS-SS-biotin (Pierce Chemical, Rockford, IL) for 2 h. The reaction was stopped by the addition of 50 mM Tris-HCl, pH 7.5 and subjected to in-solution trypsin digestion protocol as described before. The resulting peptides were subsequently analyzed using LC-MS/MS as previously described, using a 2-90% acetonitrile gradient over 43 min.

Peak lists (msf) were generated from the raw data files using Proteome Discoverer version 1.3 (Thermo Fisher Scientific) with the Sequest search engine and searched against Uniprot

human glypican-1 sequence (P35052). Searches were performed using the following parameters: semi-C-terminal trypsin cleavage specificity with up to two missed cleavages, cysteine carbamidomethylation and thioacylation of amino termini (88.03 Da) as fixed modifications, and N-terminal thioacylation of amino termini (88.03 Da) and methionine oxidation as variable modifications.

GPC1 binding partners identified by LC-MS/MS

To identify soluble GPC1 binding partners, A431 cells stably transfected with Flag-AP-GPC1 or A431 mock were seeded in 150-mm dishes. After reaching 90% confluence, cells were gently washed three times in PBS and incubated for 24 h in serum-free medium. The media were collected, and a final concentration of 1 mM EDTA and 1 mM PMSF were added to the media. The media were incubated at 4 °C overnight under gentle agitation with 150 µl of anti-FLAG M2 affinity gel (Sigma), previously equilibrated with TBS. After that, the beads were washed three times with TBS, and the immunocomplexes were eluted with FLAG peptide (Sigma) at a final concentration of 150 ng/µl during 2 h at 4 °C, under agitation. The immunocomplexes were reduced (5 mM dithiothreitol, 25 min at 56 °C), alkylated (14 mM iodoacetamide, 30 min at room temperature in the dark), digested with trypsin (Promega) and analyzed by LC-MS/MS as described before, using a 2-90% acetonitrile gradient over 43 min. Two independent experiments were performed.

The raw files were processed using the MaxQuant version 1.2.7.4 as described before with carbamidomethylation of cysteine set as a fixed modification, and oxidation of methionine and protein N-terminal acetylation chosen as variable modifications. Bioinformatics analysis was performed using the software Perseus v.1.2.7.4 available in the MaxQuant environment and reverse and contaminant entries were excluded from further analysis. Protein intensity was considered as expression values and only proteins exclusively identified in both Flag-AP-GPC1 independent experiments compared with the mock control were considered for further analysis.

Interaction networks of GPC1 partners

The list of proteins exclusively identified in the two independent experiments in the IP with Flag-AP-GPC1 was submitted to interaction networks analysis using Ingenuity Pathways (IPA; Ingenuity Systems, Redwood City, CA). For that, we considered only extracellular and plasma membrane proteins assigned according to Ingenuity database. Biological networks were generated using their Knowledge Base for interactions between mapped Focus Genes (user's list)

and all other gene objects stored in the knowledge base. Molecular Activity Predictor (MAP) tool was performed to predict the upstream and/or downstream effects of activation or inhibition of molecules in a network. Functional analysis of the networks was performed to identify the biological functions that were most significant to the genes in the network. The significance of functional enrichment was computed by a Fisher's exact test at $p < 0.05$.

Database for Annotation, Visualization and Integrated Discovery (DAVID bioinformatics, david.abcc.ncifcrf.go) (24) resources was used to find significantly enriched biological processes, using Gene Ontology (GO) classification and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Fisher's exact test was used to check for significant over-representation ($p<0.05$) of GO terms and KEGG pathways in the protein list against the Homo sapiens genome. Biological process and KEGG pathways were considered significant if the P-value was ≤ 0.05. Furthermore, Benjamin multiple testing was performed to globally correct the p-value controlling family-wide false discovery rate (p-value ≤ 0.1).

Functional Assays

Cell adhesion assay

A431 control cells and stably transfected with Flag-AP-GPC1 or transient transfected with empty vector and GPC1 Δ389-558 were submitted to adhesion assay as described by Aragão et al. (25). Briefly, $3x10^5$ cells were plated in 6-well plate and incubated overnight at 37^{o}C, while a 96-well plate was coated with Matrigel™ (2 µg per well; BD Biosciences) at 4 ^{o}C. The following day, cells were starved for 4 h, treated or not with PMA (100 ng/ml) or marimastat (2 µM) or PMA (100 ng/ml) and marimastat 2µM at the same time for 1 h in serum-free medium. Serum-free media supplemented with the GPC1 and partners eluted from the immunoprecipitation of Flag-AP-GPC1 and of control mock A431 cells (10% v/v) were also used as treatment for 1h. Cells were washed with PBS, trypsinised and seeded at a density of $1x10^4$ cells/well in the coated 96-well plate, previously washed three times with PBS and blocked with 3% BSA (bovine serum albumin) during 2 h. The adhesion was evaluated during 1 h in serum-free media supplemented with 3% BSA. The wells were washed 3 times and cells were fixed with 10% formaldehyde. Cells were stained with 1% toluidine blue containing 1% borax for 5 min. The dye was eluted using 100 µl 1% SDS and the absorbance was measured at 620 nm. Three independent experiments were performed with triplicate.

Transwell migration assay

A431 control cells and stably transfected with Flag-AP-GPC1 or transient transfected with empty vector and GPC1 Δ389-558 were treated or not with PMA (100 ng/ml) or marimastat (2 µM) or PMA (100 ng/ml) and marimastat 2µM at the same time in a phenol and serum-free medium for 1 h and plated in the upper chambers of 8 mm pore transwells (HTS Transwell 96-Well Plate, Corning). The cells were allowed to migrate towards the lower chamber containing 1% FBS or serum-free media supplemented with the GPC1 and partners eluted from the immunoprecipitation of Flag-AP-GPC1 and of control mock A431 cells (10% v/v). At the end of the assay, the cells at the top chamber were removed using a cotton swab, and the cells at the bottom of the insert filter were fixed with 10% formaldehyde for 10 min, washed with PBS and stained with 1% toluidine blue solution in 1% borax for 5 min. The dye was eluted using 1% SDS, and the absorbance was measured at 620 nm. Three independent experiments were performed in triplicate.

Bromodeoxyuridine-labeling (BrdU) index

A431 control cells and stably transfected with Flag-AP-GPC1 or transient transfected with empty vector and GPC1 Δ389-558 were plated in 96-well plates at a density of 10^4 cells per well in medium containing 10% of FBS. After 48 h, the cells were washed with PBS and cultured in serum-free medium for 8 h. Following serum starvation, cells were incubated for 16h in serum-free medium containing the following treatments: PMA (100 ng/ml) or marimastat (2 µM) or PMA (100 ng/ml) and marimastat 2µM at the same time or serum-free media supplemented with the GPC1 and partners eluted from the immunoprecipitation of Flag-AP-GPC1 and of control mock A431 cells (10% v/v). After that, proliferation rates were determined by measuring BrdU incorporation into DNA, (Cell Proliferation ELISA BrdU Colorimetric, Roche Applied Science, Germany). Briefly, BrdU antigen was added to the cultures in 1:10 dilution and kept for 2 h at 37°C in 5% CO_2. After incubation, the medium was removed and manufacturer's protocol was followed. Absorbance was measured at 450 nm with correction at 690 nm. Two experiments were performed with triplicate.

RESULTS

Expression and purification of recombinant ADAM17-671

The extracellular domain of ADAM17 (Met1 to Asn671), corresponding to the sequence of prodomain, metalloprotease, disintegrin-like, cysteine-rich and EGF-like domain, in fusion with a C-terminal 6-His tag was cloned and expressed in baculovirus expression system. After affinity purification using Ni sepharose resin, the recombinant ADAM17-671 was visualized by Coomassie blue stained SDS-PAGE and immunoblotted against histidine tag (Figure 1 A and B). Although the predicted molecular mass of the mature recombinant ADAM17-671 form, which lacks the transmembrane and cytoplasmic domains, is 52 kDa, the observed band under reduction condition, is ~70 kDa due to glycosylation modifications, as previously reported (26). The observed 100 kDa band corresponds to the form of ADAM17-671 with its prodomain.

We also evaluated the activity of the recombinant ADAM17-671 using AP report assay or quenched fluorescent substrate. Increased activity, corresponding to the shedding of the HB-EGF-AP, was observed with increasing the concentrations of recombinant ADAM17 added to the media of HEK293 cells, stably transfected with the chimeric protein (Figure 1C). In addition, increased fluorescence activity was observed by increasing the concentration of recombinant ADAM17 (Figure 1D). Both mature and pro-mature forms of ADAM17-671 were confirmed by mass spectrometry (not shown).

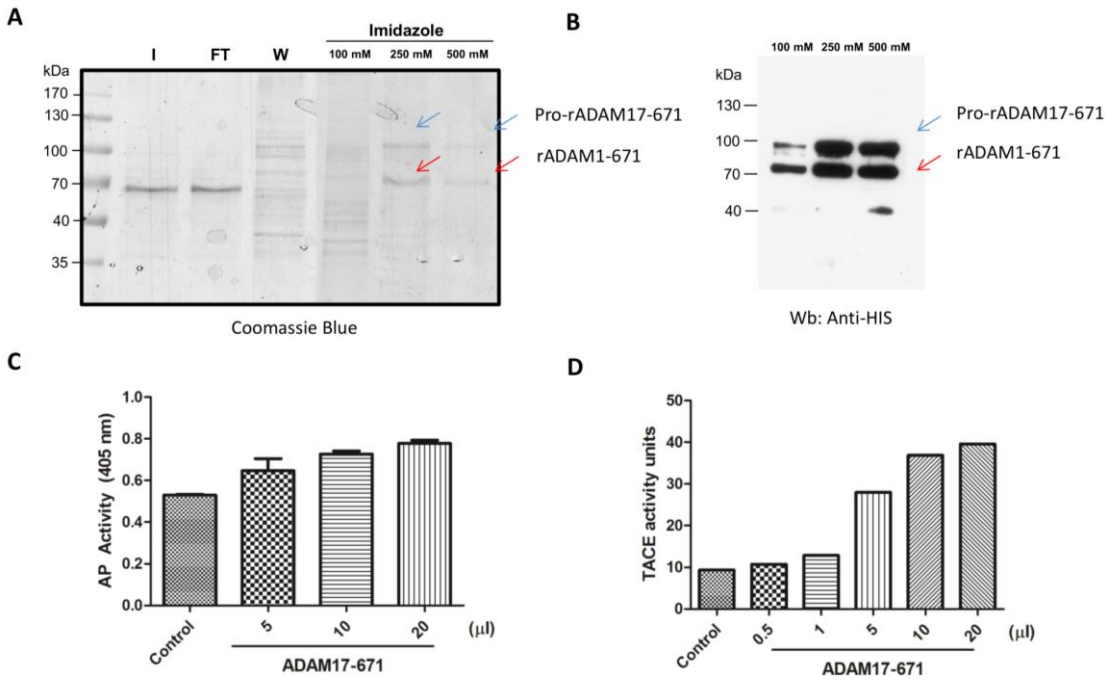

Figure 1 Expression and purification of recombinant ADAM17 extracellular domain. A) The extracellular domain of ADAM17, corresponding to Met1 to Asn 671 (rADAM17), was expressed using baculovirus system. After purification using anti-His resin, the proteins were visualized by Coomassie blue stained gels. I: input; FT: flow through; W: wash; elution (100 mM, 250 mM and 500 mM imidazole) fractions. Blue arrow corresponds to the ADAM17 with prodomain and red arrow to mature recombinant ADAM17. B) Immunoblotting anti-His of elution fractions (100 mM, 250 mM and 500 mM imidazole). C) ADAM17 activity assay using HEK293 stably transfected with HB-EGF-AP. Crescent volume of recombinant ADAM17expressed and purified from insect cells was added in the conditioned media of HEK293 stably transfected with HB-EGF-AP. Supernatant was collected and evaluated by AP assay. D) ADAM17 activity was also measured by cleavage of a fluorogenic substrate Mca-P-L-A-Q-A-V-Dpa-R-S-S-S-R-NH2 using as crescent volume of recombinant ADAM17.

Identification of ADAM17 substrates using mass spectrometry-based proteomic approaches

We performed two mass spectrometry-based strategies in order to find new ADAM17 substrates in SCC-9 cells (Figure 2). First, by using cell surface-labeled with NHS-SS-Sulfo-Biotin incubated with extracellular domain of ADAM17 recombinant protein (Figure 2A), we identified 187 proteins with a FDR less than 1% (Supplemental Table SM1). Among them, 46 proteins were found exclusively in the extracellular media of SCC-9 incubated with ADAM17 recombinant protein and 31 proteins showed intensity ratio > 1.5 between the condition incubated with ADAM17 and control (Supplemental Table SM2).

The second strategy consisted in analyzing the conditioned media from SCC-9 cells scramble and knockdown for ADAM17 (Figure 2B). A list of 1262 proteins with less than 1% FDR was generated by the Scaffold Q+ software from two biological replicates of each condition (Supplemental Table SM3). Overall, 84 proteins showed differential levels (Student's *t*-test, $p < 0.05$, Supplemental Table SM4) between the two conditions. Among them, 31 proteins decreased in abundance, and 53 were increased abundance in the SCC-9 scrambled cells.

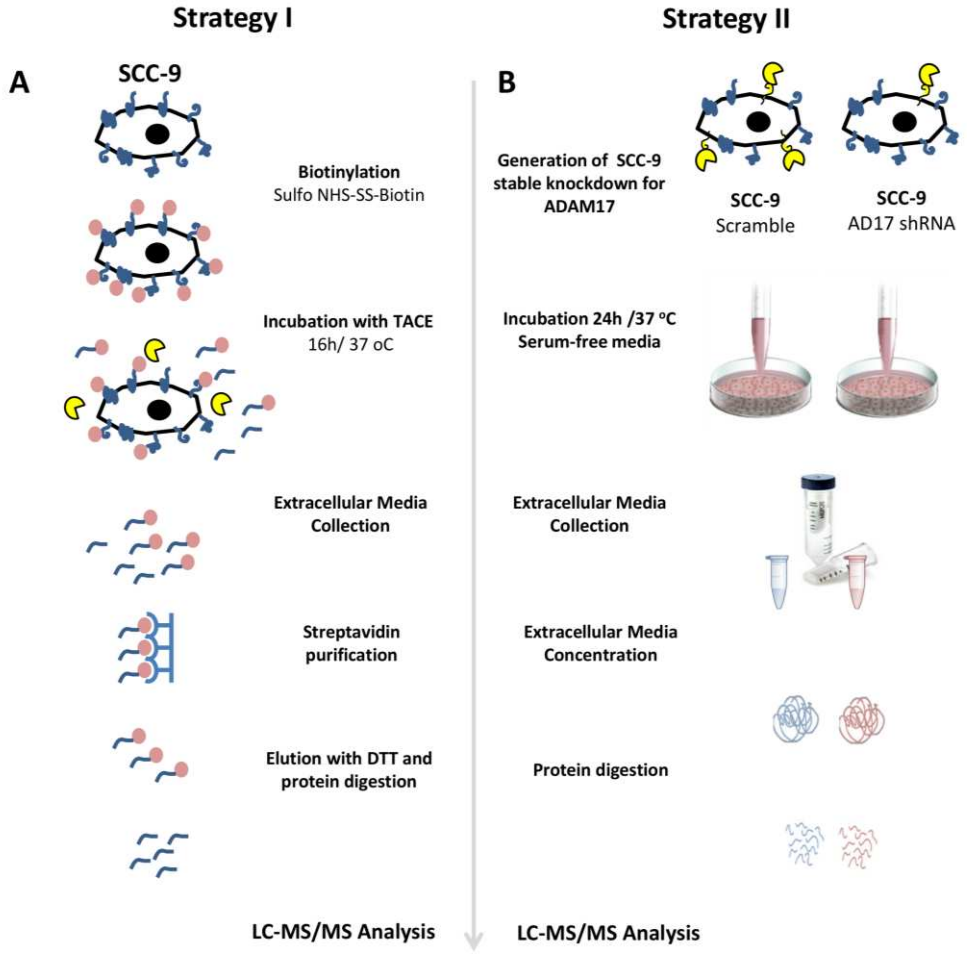

Figure 2 Strategies used for ADAM17 substrate discovery. A) Biotinylation strategy: cells were grown until confluence and incubated with 0.5 mg/ml EZ-Link Sulfo-NHS-SS-biotin. After quenching with glycine, the cells were incubated with recombinant ADAM17 for 16h/37^{o}C. The extracellular media were collected and the biotinylated proteins were purified on streptavidin-beads. Upon reduction with DTT, the biotinylated proteins were released from the beads. The eluted material was submitted to trypsin digestion protocol and analyzed by LC-MS/MS. B) Secretome analysis of scramble and knockdown of ADAM17 using SCC-9 cells. Cells were cultured for 24h in serum-free media and the extracellular media were collected, concentrated, digested with trypsin and analyzed by LC-MS/MS.

The differential abundant proteins showed in both strategies were analyzed regarding extracellular and plasma membrane localization using signal P (classical secretion pathway) and TMHMM (transmembrane helices) predictors. In the first strategy, from 77 proteins, which were considered with differential abundance in the presence of ADAM17, 25 (29%) showed at least one transmembrane helices prediction and 40 (52%) proteins showed signal P prediction (Supplemental Table SM2). In the second strategy, from the 84 proteins with differential abundance (Student's *t*-test, $p < 0.05$) between scrambled and knockdown for ADAM17 SCC-9 cells, 11 (13%) proteins showed at least one transmembrane helices prediction and 23 (27%) showed signal P prediction (Supplemental Table S4). For novel candidate substrates modulated by ADAM17 we considered only proteins with extracellular and plasma membrane localization, which were displayed in Table 1. Many known ADAM17 substrates were found exclusively or up-regulated in the extracellular media of SCC-9 cells incubated with recombinant ADAM17, such as: L1CAM (27), ALCAM (28), CD44 (29), SDC4 (30) and AXL (31) (Table1)

Cadherin-13, glypican-1, dystroglycan and urokinase-type plasminogen activator were found in both strategies as novel substrate candidates for ADAM17, since they were observed only when incubated with recombinant ADAM17 and with increased abundance in the medium of scrambled SCC9 compared to shADAM17-SCC9 (Table 1). In addition, considering glypican-1 was also found in increased abundance in the medium of wild-type compared to knockout ADAM17 mEFs cells (32), we selected this protein for validation as direct target and to evaluate the functional role in a biological context.

Table 1 ADAM17 degradome in SCC-9 cells identified using two mass spectrometry-based strategies

Protein names	Gene names	Uniprot Accession	**Strategy I**	**Strategy II**
			ADAM17/Control[#]	Scrambled/shAD17*
Known ADAM17 substrates				
Neural cell adhesion molecule L1	L1CAM (27)	P32004	Only ADAM17	
CD166 antigen	ALCAM (28)	Q13740	8.04	
CD44 antigen	CD44 (29)	P16070	6.84	
Syndecan-4;Syndecan	SDC4 (30)	P31431	2.23	
Tyrosine-protein kinase receptor UFO	AXL (31)	P30530	Only ADAM17	
ADAM17 modulated substrate predicted to have transmembrane domain by TMHMM				
Cadherin-13	CDH13	P55290	Only ADAM17	Only Scrambled
Dystroglycan;Alpha-dystroglycan;Beta-dystroglycan	DAG1	Q14118	Only ADAM17	1.6
Urokinase-type plasminogen activator	PLAU	P00749	Only ADAM17	Only Scrambled
Neutral alpha-glucosidase AB	GANAB	Q14697	Only ADAM17	
NKG2D ligand 2	ULBP2	Q9BZM5	Only ADAM17	
Integrin beta-1	ITGB1	P05556	Only ADAM17	
ICOS ligand	ICOSLG	O75144	Only ADAM17	
Midkine	MDK	P21741	Only ADAM17	
Disintegrin-like and metalloproteinase domain-containing protein 17**	ADAM17	P78536	Only ADAM17	
Cadherin-1;E-Cad/CTF1;E-Cad/CTF2;E-Cad/CTF3	CDH1	P12830	Only ADAM17	
Ephrin-B2	EFNB2	P52799	Only ADAM17	
4F2 cell-surface antigen heavy chain	SLC3A2	P08195	30.2	
Prostaglandin F2 receptor negative regulator	PTGFRN	Q9P2B2	12.4	
Receptor-type tyrosine-protein phosphatase F	PTPRF	P10586	5.2	
Cadherin-3	CDH3	P22223	2.5	
Thrombomodulin	THBD	P07204	2.0	

Large neutral amino acids transporter small subunit 1	SLC7A5	Q01650	1.8	
5-nucleotidase	NT5E	P21589	Only ADAM17	
Basigin	BSG	P35613	3.0	
CD9 antigen	CD9	P21926	Only ADAM17	
Gap junction alpha-1 protein;Gap junction protein	GJA1	P17302	Only ADAM17	
Neutral amino acid transporter B(0)	SLC1A5	Q15758	Only ADAM17	
Tyrosine-protein phosphatase non-receptor type 1 OS=Homo sapiens GN=PTPN1 PE=1 SV=1	PTN1	P18031		Only Scrambled
Lipolysis-stimulated lipoprotein receptor OS=Homo sapiens GN=LSR PE=1 SV=4	LSR	Q86X29		10.9
Isoform 2 of Syntenin-1 OS=Homo sapiens GN=SDCBP	SDCB1	O00560-2		5.9
Isoform 2 of Tissue factor OS=Homo sapiens GN=F3	TF	P13726-2		Only Scrambled
Isoform Short of Beta-1,4-galactosyltransferase 1 OS=Homo sapiens GN=B4GALT1	B4GT1	P15291-2		1.56
Isoform 2 of Integrin alpha-3 OS=Homo sapiens GN=ITGA3	ITA3	P26006-1		Only Scrambled
ADAM17 modulated substrate predicted to be extracellular by Signal P				
Glypican-1;Secreted glypican-1	GPC1	P35052	Only ADAM17	1.7
Reticulocalbin-1	RCN1	Q15293	Only ADAM17	
Protein disulfide-isomerase A3;Thioredoxin	PDIA3	P30101	Only ADAM17	
Prostate stem cell antigen	PSCA	O43653	Only ADAM17	
Insulin-like growth factor-binding protein 7	IGFBP7	Q16270	Only ADAM17	
Integrin beta-4	ITGB4	P16144	Only ADAM17	
Endoplasmin	HSP90B1	P14625	Only ADAM17	
Laminin subunit gamma-2	LAMC2	Q13753	66.9	
78 kDa glucose-regulated protein	HSPA5	P11021	15.5	
CD109 antigen	CD109	Q6YHK3	7.7	
Plasminogen activator inhibitor 1	SERPINE1	P05121	4.2	
Beta-2-microglobulin;Beta-2-microglobulin form pI 5.3	B2M	P61769	3.8	
Glucosidase 2 subunit beta	PRKCSH	P14314	3.7	
Fibroblast growth factor-binding protein 1	FGFBP1	Q14512	1.6	

Serum amyloid A protein	SAA1	P0DJI8	1.5	
26S proteasome non-ATPase regulatory subunit 7 OS=Homo sapiens GN=PSMD7 PE=1 SV=2	PSD7	P51665		Only Scramble
Alpha-1-antitrypsin OS=Homo sapiens GN=SERPINA1 PE=1 SV=3	A1AT	P01009		Only Scramble
Isoform 2 of EGF-like repeat and discoidin I-like domain-containing protein 3 OS=Homo sapiens GN=EDIL3	EDIL3	O43854-2		Only Scramble
Complement C1r subcomponent OS=Homo sapiens GN=C1R PE=1 SV=2	C1R	P00736		2.5

* p value <0.05 (Student's *t*-test test applied to normalized spectral counts). Ratio between protein spectral counts from scramble and knockdown for ADAM17 SCC-9 cells.
** Recombinant ADAM17
[#] Ratio between protein intensity of SCC-9 incubated with recombinant ADAM17 and SCC-9 control.
Transmembrane helices and secretion protein prediction according to TMHMM and SignalP server, respectively.

PMA or EGF-induced shedding of GPC1 is metalloprotease-dependent and its shedding is enhanced by cholesterol depletion

To validate whether GPC1 shedding is metalloprotease-dependent, the chimera Flag-AP-GPC1 was stably transfected in A431 cells and the overexpression was confirmed by RT-qPCR (Figure 3A). The release of GPC1 was significantly induced after PMA (100 ng/ml) or EGF (100 ng/ml) treatments (Figure 3B and 3C, n=3, ANOVA followed by Tukey test, $p<0.05$). On the other hand, as expected, the treatment with the metalloprotease inhibitor marimastat reduced both PMA and EGF-stimulated as well as constitutive shedding of GPC1, suggesting that a metalloprotease is involved in this process (Figure 3B and 3C).

Since previous studies reported that glypican is localized preferentially in cholesterol-rich lipid rafts (33,34), we hypothesized whether depletion of cholesterol modulates the PMA-induced shedding. We observed that when A431 cells were pre-treated with 5 mM of methyl-beta-cyclodextrin (MβCD) for 30 min, not only the constitutive GPC1 shedding was increased but also the PMA-induced shedding. It was observed that without MβCD pre-treatment, the fold increase in PMA-induced GPC1 shedding was 1.3, whereas with cholesterol depletion the fold increase was 2.0 (Figure 3D). Moreover, we showed that the disruption of cholesterol by MβCD did not change the total amount of GPC1, but exposed it on cell surface (Figure 3E) as it was detected after cell-surface biotinylation.

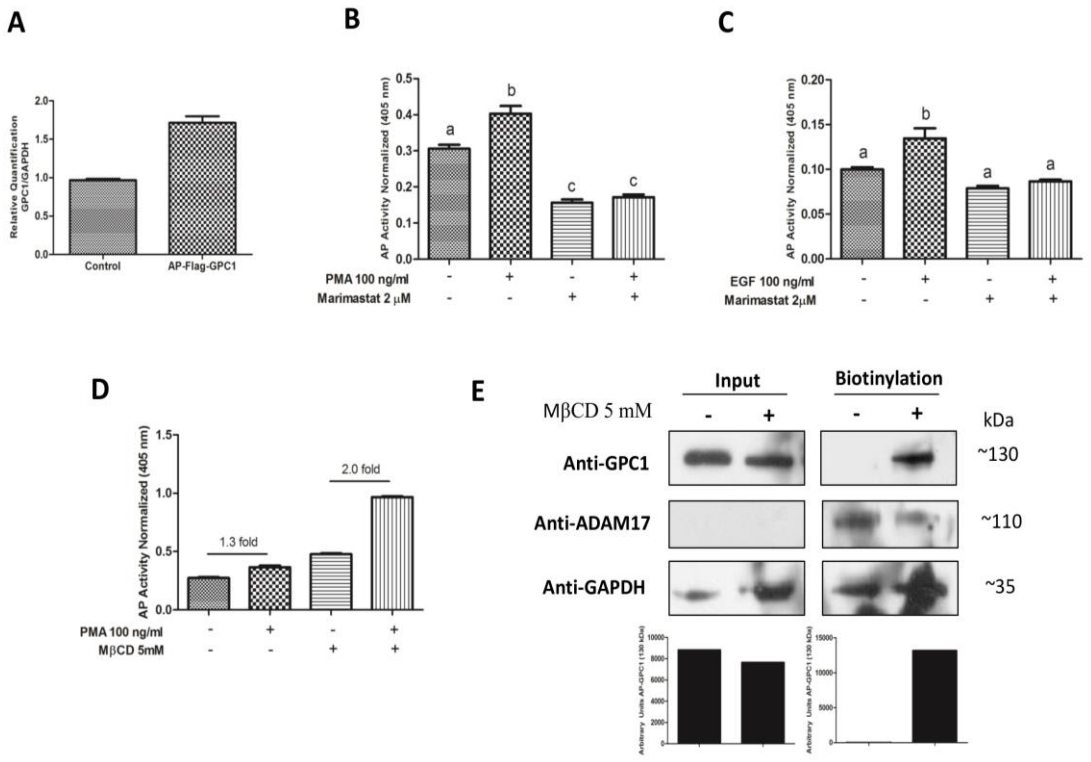

Figure 3 GPC1 shedding assays. A) Overexpression of Flag-AP-GPC1 in A431 cells was confirmed by RT-qPCR. B) A431 cells stably expressing Flag-AP-GPC1 were cultured in 24-well plate and after 24 h, the cells were submitted to a starvation period of 4 h followed by A) PMA (100 ng/ml) or Marimastat (2μM) treatment for 1 h or C) EGF (100 ng/ml) or Marimastat (2μM) treatments for 1 h. Supernatant was collected and evaluated by AP assay. Shedding activity is presented as a ratio of the AP activity in the supernatant and the cell lysate (n=3, one-way ANOVA followed by Tukey test. Different letters indicate statistically difference at $p<0.05$). D) A431 cells stably expressing Flag-AP-GPC1 were treated with 5 mM of methyl-beta-cyclodextrin for 30 min, followed by treatment with PMA (100 ng/ml) for 1 h. E) A431 Flag-AP-GPC1 cells were treated with 5 mM of methyl-beta-cyclodextrin for 30 min, followed by labeling with biotin reagent. Cell surface Flag-AP-GPC1 was detected by immunoblotting of cell surface biotinylated proteins only after treatment with MβCD 5 mM. Flag-AP-GPC1 was detected in the input (total extract) in both conditions and ADAM17 was detected only after cell surface enrichment. Anti-GAPDH antibody was used as loading control. Densitometry analysis of the 130 kDa band, corresponding to the recombinant Flag-AP-GPC1 was calculated using ImageJ software.

ADAM17 is responsible for GPC1 shedding

Knockdown of ADAM17 expression by shRNA in SCC-9 and A431 was confirmed by immunoblotting of membrane-enriched extracts (Figure 4A) and RTqPCR (Figure 4B). To test whether ADAM17 is required for direct GPC1 shedding, we compared endogenous levels of GPC1 present in the media of ADAM17 scrambled or knockdown SCC-9 and A431 cells. As shown in figure 4C, bands at ~55 kDa, ~50 kDa and ~40 kDa were observed in increased abundance in SCC-9 and A431 scrambled media compared to SCC-9 and A431 shRNA-AD17 media, suggesting a

proteolytic processing of GPC1. Immunoblotting analysis of intracellular extracts from scrambled and shRNA-AD17 of SCC-9 and A431 cells showed similar abundance of GPC1 between the two conditions and a band at ~70 kDa corresponding to the expected full-length molecular weight of GPC1 (Figure 4D).

To investigate the effect of the modulation of ADAM17 on glypican-1 mRNA expression, RT-qPCR was performed. No significant difference was observed in GPC1 mRNA expression between scrambled and knockdown SCC-9 and A431 cells for ADAM17 (Figure 4E). This result indicated that the difference in GPC1 protein abundance observed in the secretomes was not caused by the difference in expression but likely due to a post-transcriptional event.

Moreover, it was observed that although mRNA and protein levels of ADAM17 were significantly reduced after transduction with shRNA-AD17 lentivirus in SCC-9 and A431 cell lines, no difference was observed in mRNA levels of ADAM10 (Figure 4F), suggesting that the difference in GPC1 protein abundance is mostly triggered by ADAM17.

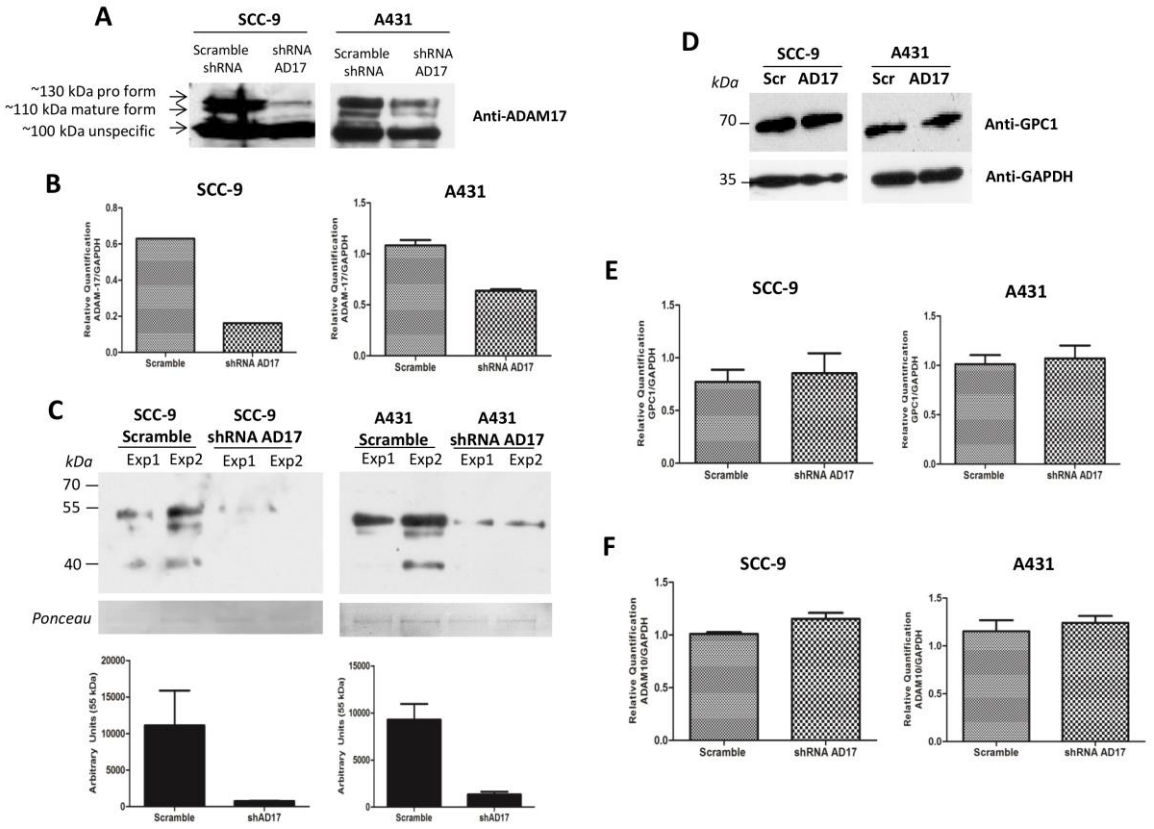

Figure 4 The cleavage of Glypican-1 is mediated by ADAM17. Knockdown of ADAM17 expression by shRNA in SCC-9 andA431 cell line was confirmed by immunoblotting of membrane-enriched extracts (A) and by quantitative RT-PCR (B). C) The extracellular concentrated media (5µg) from SCC-9 or A431 stably transfected with a scrambled small hairpin RNA or shRNA targeting ADAM17 were analyzed by immunoblotting with an antibody against endogenous Glypican-1, which recognizes epitope sequence Asp287 to Arg360. Ponceau S staining was used as loading control. Densitometry analysis of GPC1 at 55 kDa levels was calculated by band intensity using ImageJ software. D) Cell lysate extracts were analyzed by immunoblotting with an antibody against endogenous Glypican-1. Anti-GAPDH antibody was used as loading control. Relative mRNA expression levels of GPC1 (E) and ADAM10 (F) were measured by the RT-qPCR in scramble or ADAM17-shRNA SCC-9 or A431 cells. The data were normalized with glyceraldehyde-3-phosphate dehydrogenase gene.

Biotin N-terminal labeling approach was used to in vitro determine the GPC1 cleavage site by ADAM17. Using Flag-AP-GPC1 stably transfected A431cells, full-length GPC1 was immunoprecipitated from intracellular extracts and incubated with recombinant ADAM17. Neo N-terminal was labeled with biotin and after trypsin digestion the peptides were sequenced using mass spectrometry analysis. One peptide from GPC1, with residues of valine and glutamine in P1' and P2', was found only in the condition incubated with recombinant ADAM17 (Figure 5A and Supplemental Table SM5). The glypican-1 fragment resulting from cleavage has a theoretical

molecular weight of 40 kDa (residue Asp24 to Asp388), which was observed in the media of scrambled SCC-9 and A431 (Figure 4C). Besides, the sequence coverage obtained from the five peptides identified in the medium of SCC-9 incubated with recombinant ADAM17 (strategy I at the discovery phase, not shown), is consistent with the GPC1 cleavage site generated by ADAM17. Manual verification of the N-terminal labeled peptide of glypican-1 spectra is shown in figure 5B.

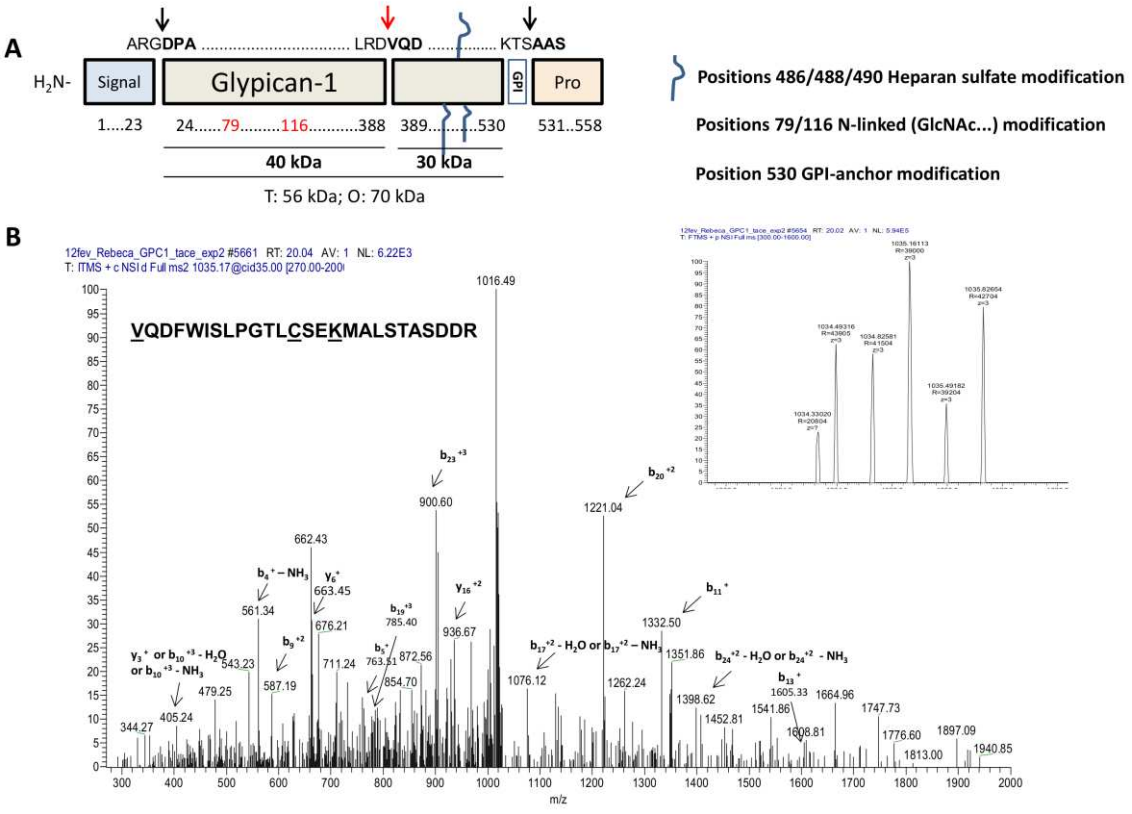

Figure 5 Glypican-1 cleavage site analysis. A) Squematic representation of glypican-1 structure. Black arrows indicate signal peptide and prodomain cleavage sites, according to Uniprot (P35052). Red arrow indicates the cleavage site generated by ADAM17. The positions of post-translational modifications are showed in left. T:theoretical; O:observed B) N-termini peptide of GPC1 resulted from ADAM17 cleavage. Manual verification of MS/MS spectrum for b and y ion series, m/z 1035.16113, +3, xcorr: 1.42 (corresponding to residue 388−414, Uniprot Human P35052). Underlined residues in the sequence are modified: V1-Biotin (88.03 Da), C13-Carbamidomethyl (57.02146 Da), K16-Biotin (88.03 Da).

Extracellular GPC1 binding partners revealed cell adhesion, migration and proliferation-associated functions

After validating the proteolytic processing of GPC1 by ADAM17, we sought to determine the biological role of soluble extracellular GPC1. For that, we performed

immunoprecipitation of soluble recombinant Flag-AP-GPC1, stably transfected in A431, followed by elution with FLAG peptide, trypsin digested and LC-MS/MS analysis. Two experiments were performed with a negative control (Supplemental Table SM6). Filtering criteria were applied, in which only proteins exclusively identified in Flag-AP-GPC1 samples in the two independent experiments were considered. A total of 125 proteins were identified exclusively in GPC1 complexes (Supplemental Table SM7). By using IPA, subcellular localization was assigned to each gene name in the list, and only extracellular space (11 genes) and plasma membrane (21 genes) proteins were considered for network analysis (Supplemental Table SM7). It was observed that the extracellular GPC1 interactome led to predict downstream effects in activation of ERK, PKC(s), NFκB and focal adhesion kinase (Figure 6). Moreover, the genes in the network functional analysis pointed to cellular movement as the top enriched biological function (p=4.24E-17-7.81E-03), encompassing 22 genes of the interactome input list (Supplemental Table SM8).

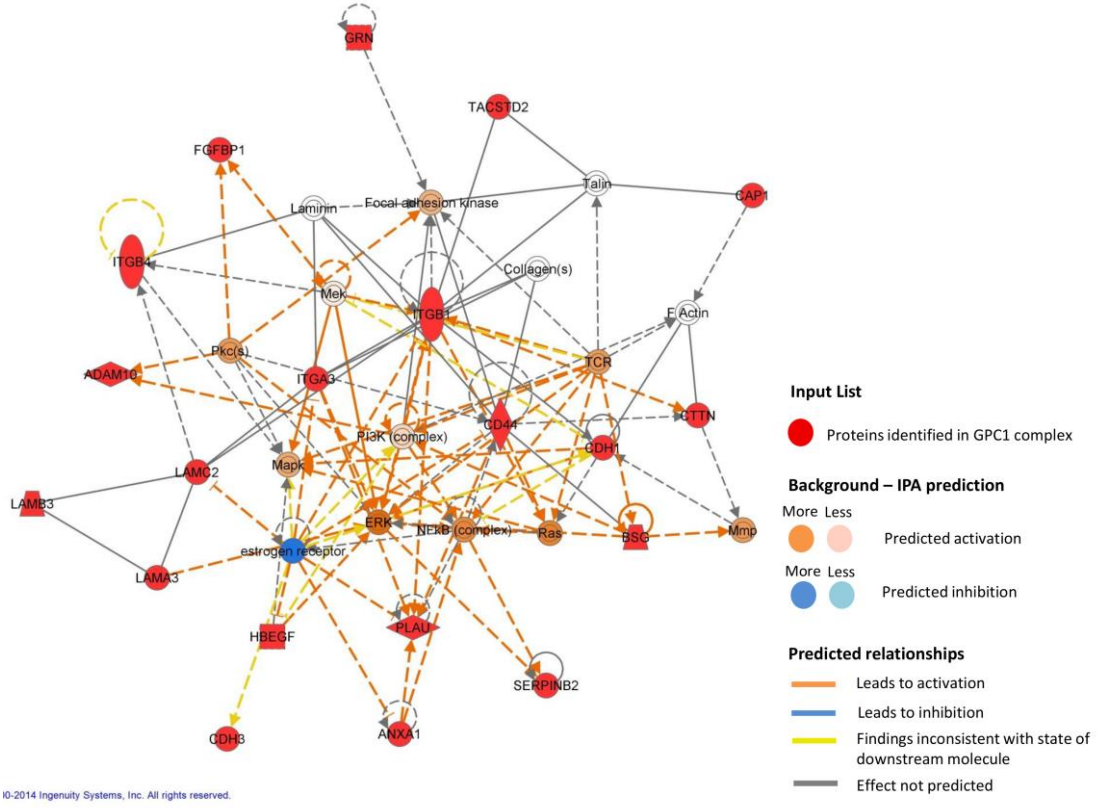

Figure 6 Network analysis of the GPC1 interactome. The list of 33 exclusive proteins predicted to be localized in plasma membrane and extracellular milieu identified in GPC1 complex was uploaded into the IPA and used to generate biological networks. Analysis of predicted pathway activation or inhibition was performed using molecule activity predictor (MAP) tool. The nodes in red are proteins identified by mass spectrometry as part of GPC1 complex. Protein abbreviations are listed in table SM7.

Besides, enrichment analysis revealed ECM-receptor interaction (p=1.12E-08) and focal adhesion (p= 6.24E-05) as the most enriched KEGG pathway present in the GPC1 interactome. Gene ontology based on biological processes also showed a significant enrichment in cell adhesion (p= 4.88E-07), biological adhesion (p= 4.94E-07), wound healing (p=3.34E-05), cell surface receptor linked signal transduction (p=1.01E-04) and regulation of cell proliferation (p=1.04E-04), among others (Supplemental Table SM9)

As a proof-of-concept, the ability of cells to adhere to extracellular matrix proteins, to migrate and to proliferate was tested in control and overexpressing GPC1 A431 cells. We observed that A431 overexpressing GPC1 had a reduced ability to adhere to Matrigel, but increased migration and proliferation (Figure 7A-C). However, when cells overexpressing GPC1 were treated with PMA (100 ng/ml), which increases the soluble GPC1, the cell adhesion was increased, but it does not affect the migration and proliferation. As expected, marimastat (2 μM) has an opposed effect inducing migration and proliferation (Figure 7 D-F), but not changing cell adhesion. Moreover, we performed the same functional assays using A431 cells treated or not with GPC1 complexes. Although no significant difference was observed in adhesion and proliferation assays, when A431 cells were treated with GPC1 complexes we observed a negative effect in cell migration, indicating GPC1 has an exclusive role when it is soluble (Fig 7 E-G).

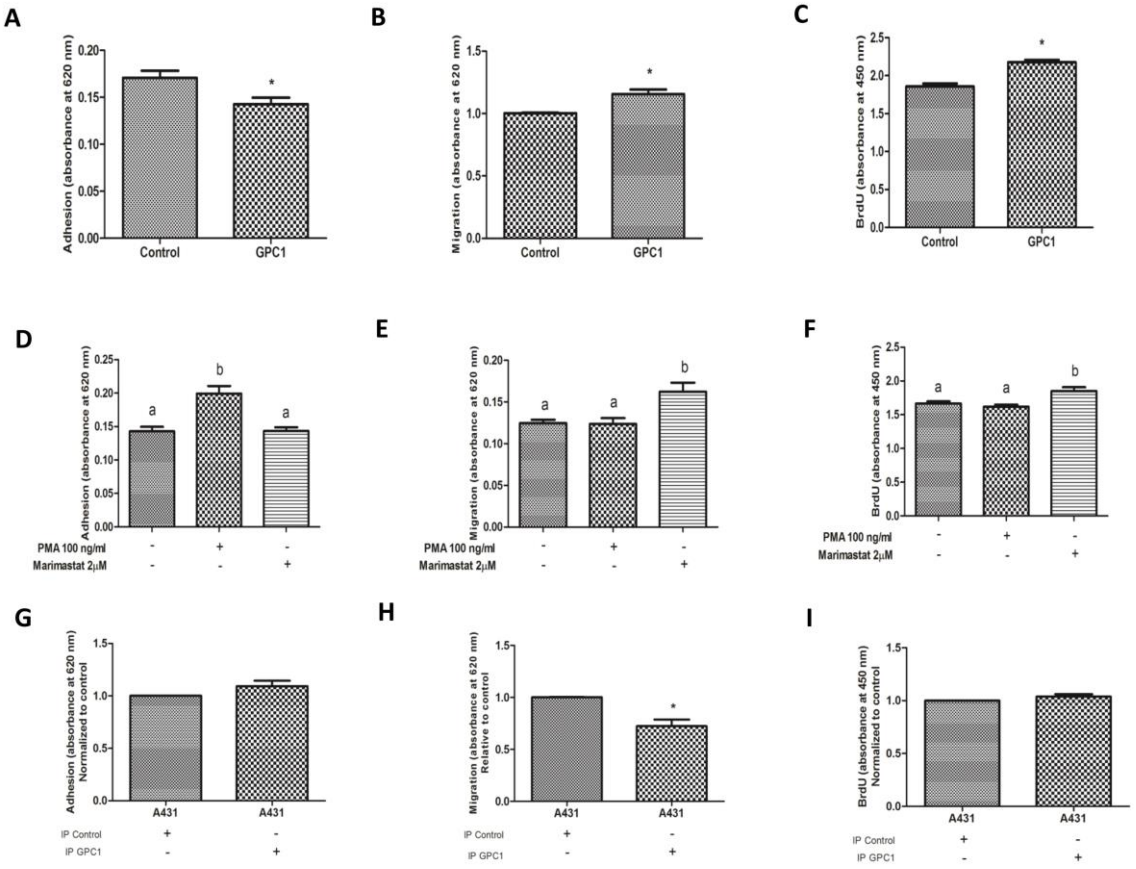

Figure 7 Functional assays of cells overexpressing GPC1 treated or not with PMA or marimastat. A) A431 cells overexpressing GPC1 have diminished cell adhesion to Matrigel compared to control (n = 3, Student's *t*-test, *p <0.05). B) A431 cells overexpressing GPC1 have increased migration compared to control. 1% FBS was added in serum-free media in the lower chamber (n = 2, Student's *t*-test, * p<0.05). C) A431 cells overexpressing GPC1 have increased proliferation compared to control (n = 2, Student's *t*-test, * p<0.05). D) Adhesion assays in A431 cells stably expressing GPC1 treated with PMA (100 ng/ml) or Marimastat (2µM) or both for 1 h (n = 3, ANOVA, followed by Tukey test, different letters indicate statistically difference at p <0.05). E) Adhesion assays in A431 cells stably expressing GPC1 were treated with PMA (100 ng/ml) or Marimastat (2µM) or both for 1 h and seeded in serum-free media in the upper chamber of 96-well transwell plates. 1% FBS was added in serum free media in the lower chamber (n = 2, ANOVA, followed by Tukey test, different letters indicate statistically difference at p <0.05). F) Proliferation assay was performed in stably expressing GPC1 A431 cells by measuring BrdU incorporation into DNA after incubation with PMA (100 ng/ml) or Marimastat (2µM) in serum-free media for 16 h (n = 2, ANOVA, followed by Tukey test, different letters indicate statistically difference at p <0.05). G) A431 cells were seeded in Matrigel coated 96-well plates treated with control or GPC1 complex in serum free media (10% v/v) for 1h (n=2). H) A431 cells were seeded in serum-free media in the upper chamber of 96-well transwell plates. Immunoprecipitation from control and GPC1 complex were added in serum-free media (10% v/v) in the lower chamber (n = 2, Student's *t*-test, p<0.05). I) Proliferation assay was performed in A431 cells by measuring BrdU incorporation into DNA after incubation with control or GPC1 complex in serum-free media for 16 h (n = 2).

The N-terminal fragment generated at the cleavage site was also cloned and expressed in A431 cells to test whether it could promote functional responses. First, we observed that the GPC1 fragment is expressed in the same molecular weight (~40kDa) as the shed form verified in the medium of scrambled SCC-9 and A431 cells (Figure 8A). Although no significant difference was observed in adhesion, the expression of GPC1 fragment decreased the ability of A431 cells to migrate (Figure 8B and 8C). Interestingly, GPC1 fragment also promoted an increased A431 cell proliferation (Figure 8D).

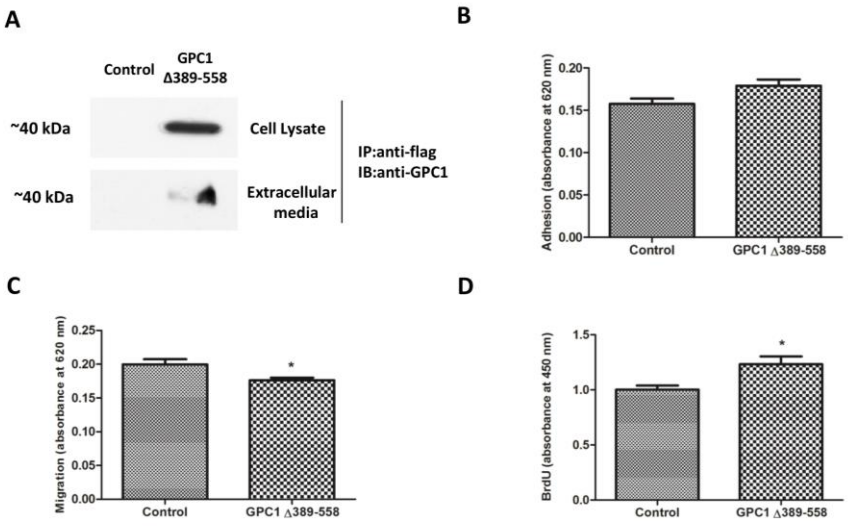

Figure 8 GPC1 fragment generated at ADAM17 cleavage site has functional properties. A) The expression of GPC1 fragment was confirmed by immunoblotting anti-GPC1 (IB) in cell lysate and extracellular media of A431 cells transient transfected with control (empty vector) or pcDNA 3.1/ GPC1 Δ389-558 (fragment) after anti-flag immunoprecipitation (IP). B) No difference in adhesion to Matrigel was observed in A431 cells overexpressing GPC1 fragment (n=3, Student's *t*-test, $p>0.05$). C) A431 cells overexpressing GPC1 fragment have diminished migration (n=2, Student's *t*-test, * $p<0.05$), but (D) increased proliferation (n=2, Student's *t*-test, * $p<0.05$).

DISCUSSION

Although many evidences showed that metalloproteases contribute to tumor progression, clinical trials still show lack of success with non-specific metalloprotease inhibitors (35). These unanticipated functions might be explained by the incomplete knowledge of individual degradomes as well as the role of targeted cleavage substrates in a cellular context. Therefore, the characterization of protease substrates is important not only to understand the function of proteases in specific biology conditions but also to predict the effect of the inhibition of a metalloprotease in tumor development (36).

ADAM17 was shown to be overexpressed in many types of cancer, including brain, breast, colon, gastric, kidney, liver, lung, ovary, pancreas and prostate (revised in Murphy, 2008 (19)). Although many studies demonstrated the involvement of ADAM17 in tumorigenic processes, using mainly knockdown (6,8,37,38) or overexpression (16,39) model system, the repertoire of its substrates in each type of cancer is far to be elucidated

In this study we presented two mass spectrometry-based strategies applied to SCC-9 cell line that were able to reveal in a large scale known ADAM17 substrates and new candidate substrates. Different proteomic methods have been undertaken to search for novel ADAM17 substrates. Using DIGE (Difference Gel Electrophoresis), Bech-Serra et al. (40) showed that two cell-cell adhesion molecules such as activated leukocyte cell adhesion molecule (ALCAM) and desmoglein 2 (Dsg-2) are cleaved by ADAM17 and ADAM10. Another proteomic screening was performed by Esselens et al. (36) using SILAC (Stable Isotope Labeling by Amino acids in Cell culture) in control and shRNA for ADAM10 and ADAM17 MCF7 cells, C4.4A, a GPI-anchored protein, was identified as a new substrate for ADAM17 and ADAM10. Recently, the secretome analysis of mouse embryonic fibroblasts (mEFs) wild-type and knockout for ADAM17 revealed other proteins candidate substrates modulated by ADAM17 (32). Even though the use of secretome to find ADAM17 substrates was previously demonstrated, in this study we presented complementary approaches that were able to reduce the complexity of the sample and reveal known and novel substrates for ADAM17. By biotinylating cell surface proteins, biotinylated peptides/proteins are released in the extracellular milieu and captured by streptavidin affinity purification. With that, we obtained many extracellular matrix and plasma membrane proteins exclusively identified in the SCC-9 cells incubated with ADAM17. As an additional approach, we performed a label-free mass spectrometry analysis of the secretome of control and knockdown for ADAM17 in SCC-9 cells. By these approaches, glypican-1 (GPC1) was found in increased

abundance in SCC-9 cells incubated with recombinant ADAM17 as well as in shRNA control compared to shRNA-ADAM17 SCC-9 cells.

Glypican-1 is a proteoglycan attached to the cell surface via a glycosyl-phosphatidylinositol (GPI) anchor (41,42) and preferentially localizes to cholesterol-rich lipid rafts and caveolae in the membrane. It was demonstrated that GPC1 is overexpressed in human astrocytoma and oligodendroglioma samples (43), human pancreatic cancer (44) and breast cancer (45). Although phospholipase D and C as well as notum were shown to be capable of releasing GPI-linked proteins from the cell surface (46-48), there was still no evidence in GPC1 ADAM17-mediated cleavage.

We first showed that GPC1 shedding is metalloprotease-dependent and induced by PMA and EGF (Figure 3). Second, we demonstrated that ADAM17 is the main responsible for GPC1 shedding (Figure 4 and 5). Moreover, the enhanced GPC1 release was confirmed to not be related to mRNA expression levels of GPC1, despite the knockdown of ADAM17 (Figure 4E). Some studies reported that ADAM17 responds rapidly to stimulation with PMA whereas ADAM10 does not, in short-term assays such as two hours or less (49-51). In addition, Le Gall et al. (52) demonstrated that only ADAM17, but not ADAM10, was able to respond to the physiological signaling pathways stimulated by EGF. Besides characterizing GPC1 induced shedding by PMA and EGF, we showed that the cleavage is mainly accomplished by ADAM17, since GPC1 abundance in the secretome of SCC-9 and A431 cells were decreased by targeting knockdown of ADAM17, in a condition of unaltered mRNA expression levels of ADAM10 (Figure 4F). In addition, the cleavage site of GPC1 was also determined in this study, in which valine and glutamine was found in P1' and P2', respectively, which is also in accordance with ADAM17 cleavage site specificity reported by Tucher et al. (53), where valine was the main residue found in P1'.

The regulation of proteolysis on the substrate level has been emerging as an important mechanism for regulation of enzyme activity (54). Our study showed that the cleavage of GPC1 is regulated by its availability in the plasma membrane. Depletion of cholesterol enhanced GPC1 localization in the plasma membrane, resulting in a stronger PMA-induced GPC1 shedding, but did not change the total protein levels, neither the substrate (GPC1) nor the protease (ADAM17) (Figure 3D/E). Previously, studies have reported the involvement of lipid rafts in the regulation of ADAM17 and ADAM10 substrate shedding. Membrane cholesterol depletion increases the ADAM17-dependent shedding of TNF and its receptors TNFR1 and TNFR2 (55), ShhNp (lipidated and membrane-tethered Sonic hedgehog) (56) and IL-6R (57). Tellier et al. (55) showed a

regulation at a protease level, where the disruption of lipid rafts displaced the mature form of ADAM17 in the non-raft region of the membrane that contains the major part of ADAM17 substrates. Kojro et al. (58) also showed that methyl-cyclodextrin treatment had a stimulating effect on the cleavage of amyloid precursor protein (APP) presumably by increasing the alpha-secretase activity of ADAM10 and inhibiting endocytosis of APP. Besides, Dierker et al. showed that ShhNp (localized in lipid-rafts) shedding is increased by ADAM17 co-expression and cholesterol depletion of cells with methyl-β-cyclodextrin in a HS-dependent fashion (56). Altogether, these findings supported that the shedding of ADAM17 substrates can be modulated by altering their availability to the membrane.

The specificity in some glypican–protein interactions, especially growth factors such as FGF2 (43), VEGF (59), heparin-binding growth factors, heregulin and hepatocyte growth factor/scatter factor (44,45) was reported. These partners have linked GPC1 to tumorigenic process, including angiogenesis (60) and proliferation (44) in glioma and breast cancer, respectively. Therefore, to investigate the GPC1 partners and its signaling pathways in carcinoma, we characterized the interactome of soluble GPC1. The soluble GPC1 indeed showed known partners including growth factors, but also not previously reported partners, such as granulins (GRN), platelet-derived growth factor subunit A (PDGFA) and fibroblast growth factor-binding protein 1 (FGFBP1). The GPC1 complex had mainly proteins involved in cellular movement, adhesion and proliferation and the networ of GPC1 complex had a predicted activation of pathways such as ERK, PKCs, NFκB and focal adhesion kinase, which are all well established to be involved in tumorigenic process (Figure 6) (61-64).

We further demonstrated that the overexpressing of membrane-anchored GPC1 and soluble GPC1 might induce at least two distinct effects in carcinoma cells: the first condition, it decreased the adhesion, but favored migration and proliferation, whereas the soluble form decreased the cell migration (Figure 7). PMA treatment reversed cell adhesion solely, but not migration and adhesion. Marimastat treatment inhibited the cleavage favoring cell migration and proliferation. Supporting the evidences that the cleavage of GPC1 performs an opposite role than when it is anchored in cell-surface, Kleff et al. (44) showed that the release of GPI-anchor proteins by phosphoinositide-specific phospholipase-C (PI-PLC) abrogated pancreatic cancer cell mitogenic responses to fibroblast growth factor 2 (FGF2) and heparin-binding EGF-like growth factor (HB-EGF). Besides, Traiser et al. (48) demonstrated that notum, which also cleaves GPI anchor of cell surface proteins, acts as a negative regulator of growth factor responsiveness.

Using the information of GPC1 cleavage site analysis, we expressed the cleavage form of GPC1 and performed the same experiments of adhesion, migration and proliferation. Interestingly, the GPC1 fragment solely is capable to reduce cell migration and increase proliferation in cancer cells (Figure 8). Altogether, this study underscored molecules that are modulated by ADAM17 and uncovered glypican-1 as direct novel ADAM17 substrate. Further, it opened new avenues regarding the proteolysis-mediated function of GPC1.

FOOTNOTES

*This work was supported by FAPESP Grants 2009/54067-3, 2010/19278-0, 2011/22421-2, and 2009/53839-2 and CNPq Grants 470567/2009-0, 470549/2011-4, and 301702/2011-0.

[1]The abbreviations used are: ADAM17, a disintegrin and metalloproteinase 17; HNSCC, neck squamous cell carcinomas; GPC1, glypican-1; PMA, phorbol 12-myristate 13-acetate; EGF, epidermal growth factor; TACE, tumor necrosis factor-α converting enzyme; HB-EGF, heparin-binding epidermal growth; OSCC, oral squamous cell carcinoma; IP, co-immunoprecipitation; AP, alkaline phosphatase; PBS, phosphate-buffered saline; TTBS, Tris-Tween-buffered saline; MβCD, methyl-β-cyclodextrin; IPA, Ingenuity Pathways; DAVID, Database for Annotation, Visualization and Integrated Discovery; MAP, Molecular Activity Predictor; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; BrDu, Bromodeoxyuridine-labeling; mEFs, mouse embryonic fibroblasts; GPI, glycosyl-phosphatidylinositol; ANOVA, analysis of variance

The mass spectrometry proteomics data have been deposited to the PeptideAtlas repository with the data set identifier PASS00611.

REFERENCES

1. Weber, S., and Saftig, P. (2012) Ectodomain shedding and ADAMs in development. *Development* **139**, 3693-3709
2. Edwards, D. R., Handsley, M. M., and Pennington, C. J. (2008) The ADAM metalloproteinases. *Molecular aspects of medicine* **29**, 258-289
3. Black, R. A., Rauch, C. T., Kozlosky, C. J., Peschon, J. J., Slack, J. L., Wolfson, M. F., Castner, B. J., Stocking, K. L., Reddy, P., Srinivasan, S., Nelson, N., Boiani, N., Schooley, K. A., Gerhart, M., Davis, R., Fitzner, J. N., Johnson, R. S., Paxton, R. J., March, C. J., and Cerretti, D. P. (1997) A metalloproteinase disintegrin that releases tumour-necrosis factor-alpha from cells. *Nature* **385**, 729-733
4. Blobel, C. P. (2005) ADAMs: key components in EGFR signalling and development. *Nature reviews. Molecular cell biology* **6**, 32-43
5. Arribas, J., and Esselens, C. (2009) ADAM17 as a therapeutic target in multiple diseases. *Current pharmaceutical design* **15**, 2319-2335
6. Lin, P., Sun, X., Feng, T., Zou, H., Jiang, Y., Liu, Z., Zhao, D., and Yu, X. (2012) ADAM17 regulates prostate cancer cell proliferation through mediating cell cycle progression by EGFR/PI3K/AKT pathway. *Molecular and cellular biochemistry* **359**, 235-243
7. Zheng, X., Jiang, F., Katakowski, M., Zhang, Z. G., Lu, Q. E., and Chopp, M. (2009) ADAM17 promotes breast cancer cell malignant phenotype through EGFR-PI3K-AKT activation. *Cancer biology & therapy* **8**, 1045-1054
8. Franovic, A., Robert, I., Smith, K., Kurban, G., Pause, A., Gunaratnam, L., and Lee, S. (2006) Multiple acquired renal carcinoma tumor capabilities abolished upon silencing of ADAM17. *Cancer research* **66**, 8083-8090
9. Bahia, M. S., and Silakari, O. (2010) Tumor necrosis factor alpha converting enzyme: an encouraging target for various inflammatory disorders. *Chemical biology & drug design* **75**, 415-443
10. Horiuchi, K., Kimura, T., Miyamoto, T., Takaishi, H., Okada, Y., Toyama, Y., and Blobel, C. P. (2007) Cutting edge: TNF-alpha-converting enzyme (TACE/ADAM17) inactivation in mouse myeloid cells prevents lethality from endotoxin shock. *Journal of immunology* **179**, 2686-2689
11. Ruuls, S. R., Hoek, R. M., Ngo, V. N., McNeil, T., Lucian, L. A., Janatpour, M. J., Korner, H., Scheerens, H., Hessel, E. M., Cyster, J. G., McEvoy, L. M., and Sedgwick, J. D. (2001) Membrane-bound TNF supports secondary lymphoid organ structure but is subservient to secreted TNF in driving autoimmune inflammation. *Immunity* **15**, 533-543
12. Leemans, C. R., Braakhuis, B. J., and Brakenhoff, R. H. (2011) The molecular biology of head and neck cancer. *Nature reviews. Cancer* **11**, 9-22
13. Stokes, A., Joutsa, J., Ala-Aho, R., Pitchers, M., Pennington, C. J., Martin, C., Premachandra, D. J., Okada, Y., Peltonen, J., Grenman, R., James, H. A., Edwards, D. R., and Kahari, V. M. (2010) Expression profiles and clinical correlations of degradome components in the tumor microenvironment of head and neck squamous cell carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* **16**, 2022-2035
14. Kornfeld, J. W., Meder, S., Wohlberg, M., Friedrich, R. E., Rau, T., Riethdorf, L., Loning, T., Pantel, K., and Riethdorf, S. (2011) Overexpression of TACE and TIMP3 mRNA in head and neck cancer: association with tumour development and progression. *British journal of cancer* **104**, 138-145

15. Ge, L., Baskic, D., Basse, P., Vujanovic, L., Unlu, S., Yoneyama, T., Vujanovic, A., Han, J., Bankovic, D., Szczepanski, M. J., Hunt, J. L., Herberman, R. B., Gollin, S. M., Ferris, R. L., Whiteside, T. L., Myers, E. N., and Vujanovic, N. L. (2009) Sheddase activity of tumor necrosis factor-alpha converting enzyme is increased and prognostically valuable in head and neck cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **18**, 2913-2922
16. Simabuco, F. M., Kawahara, R., Yokoo, S., Granato, D. C., Miguel, L., Agostini, M., Aragao, A. Z., Domingues, R. R., Flores, I. L., Macedo, C. C., Della Coletta, R., Graner, E., and Paes Leme, A. F. (2014) ADAM17 mediates OSCC development in an orthotopic murine model. *Molecular cancer* **13**, 24
17. Huang, Y., Benaich, N., Tape, C., Kwok, H. F., and Murphy, G. (2014) Targeting the sheddase activity of ADAM17 by an anti-ADAM17 antibody D1(A12) inhibits head and neck squamous cell carcinoma cell proliferation and motility via blockage of bradykinin induced HERs transactivation. *International journal of biological sciences* **10**, 702-714
18. Ge, H., Liu, H., Fu, Z., and Sun, Z. (2012) Therapeutic and preventive effects of an epidermal growth factor receptor inhibitor on oral squamous cell carcinoma. *The Journal of international medical research* **40**, 455-466
19. Murphy, G. (2008) The ADAMs: signalling scissors in the tumour microenvironment. *Nature reviews. Cancer* **8**, 929-941
20. Trindade, D. M., Silva, J. C., Navarro, M. S., Torriani, I. C., and Kobarg, J. (2009) Low-resolution structural studies of human Stanniocalcin-1. *BMC structural biology* **9**, 57
21. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372
22. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10**, 1794-1805
23. Kim, J., Lin, J., Adam, R. M., Lamb, C., Shively, S. B., and Freeman, M. R. (2005) An oxidative stress mechanism mediates chelerythrine-induced heparin-binding EGF-like growth factor ectodomain shedding. *Journal of cellular biochemistry* **94**, 39-49
24. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57
25. Aragao, A. Z., Belloni, M., Simabuco, F. M., Zanetti, M. R., Yokoo, S., Domingues, R. R., Kawahara, R., Pauletti, B. A., Goncalves, A., Agostini, M., Graner, E., Coletta, R. D., Fox, J. W., and Paes Leme, A. F. (2012) Novel processed form of syndecan-1 shed from SCC-9 cells plays a role in cell migration. *PloS one* **7**, e43521
26. Milla, M. E., Leesnitzer, M. A., Moss, M. L., Clay, W. C., Carter, H. L., Miller, A. B., Su, J. L., Lambert, M. H., Willard, D. H., Sheeley, D. M., Kost, T. A., Burkhart, W., Moyer, M., Blackburn, R. K., Pahel, G. L., Mitchell, J. L., Hoffman, C. R., and Becherer, J. D. (1999) Specific sequence elements are required for the expression of functional tumor necrosis factor-alpha-converting enzyme (TACE). *The Journal of biological chemistry* **274**, 30563-30570
27. Maretzky, T., Schulte, M., Ludwig, A., Rose-John, S., Blobel, C., Hartmann, D., Altevogt, P., Saftig, P., and Reiss, K. (2005) L1 is sequentially processed by two differently activated metalloproteases and presenilin/gamma-secretase and regulates neural cell adhesion, cell migration, and neurite outgrowth. *Molecular and cellular biology* **25**, 9040-9053

28. Rosso, O., Piazza, T., Bongarzone, I., Rossello, A., Mezzanzanica, D., Canevari, S., Orengo, A. M., Puppo, A., Ferrini, S., and Fabbi, M. (2007) The ALCAM shedding by the metalloprotease ADAM17/TACE is involved in motility of ovarian carcinoma cells. *Molecular cancer research : MCR* **5**, 1246-1253
29. Stoeck, A., Keller, S., Riedle, S., Sanderson, M. P., Runz, S., Le Naour, F., Gutwein, P., Ludwig, A., Rubinstein, E., and Altevogt, P. (2006) A role for exosomes in the constitutive and stimulus-induced ectodomain cleavage of L1 and CD44. *The Biochemical journal* **393**, 609-618
30. Pruessmeyer, J., Martin, C., Hess, F. M., Schwarz, N., Schmidt, S., Kogel, T., Hoettecke, N., Schmidt, B., Sechi, A., Uhlig, S., and Ludwig, A. (2010) A disintegrin and metalloproteinase 17 (ADAM17) mediates inflammation-induced shedding of syndecan-1 and -4 by lung epithelial cells. *The Journal of biological chemistry* **285**, 555-564
31. Budagian, V., Bulanova, E., Orinska, Z., Duitman, E., Brandt, K., Ludwig, A., Hartmann, D., Lemke, G., Saftig, P., and Bulfone-Paus, S. (2005) Soluble Axl is generated by ADAM10-dependent cleavage and associates with Gas6 in mouse serum. *Molecular and cellular biology* **25**, 9324-9339
32. Kawahara, R., Lima, R. N., Domingues, R. R., Pauletti, B. A., Meirelles, G. V., Assis, M., Figueira, A. C., and Paes Leme, A. F. (2014) Deciphering the role of the ADAM17-dependent secretome in cell signaling. *Journal of proteome research* **13**, 2080-2093
33. Cheng, F., Mani, K., van den Born, J., Ding, K., Belting, M., and Fransson, L. A. (2002) Nitric oxide-dependent processing of heparan sulfate in recycling S-nitrosylated glypican-1 takes place in caveolin-1-containing endosomes. *The Journal of biological chemistry* **277**, 44431-44439
34. Gutierrez, J., and Brandan, E. (2010) A novel mechanism of sequestering fibroblast growth factor 2 by glypican in lipid rafts, allowing skeletal muscle differentiation. *Molecular and cellular biology* **30**, 1634-1649
35. Saftig, P., and Reiss, K. (2011) The "A Disintegrin And Metalloproteases" ADAM10 and ADAM17: novel drug targets with therapeutic potential? *European journal of cell biology* **90**, 527-535
36. Esselens, C. W., Malapeira, J., Colome, N., Moss, M., Canals, F., and Arribas, J. (2008) Metastasis-associated C4.4A, a GPI-anchored protein cleaved by ADAM10 and ADAM17. *Biological chemistry* **389**, 1075-1084
37. Chen, X., Chen, L., Chen, J., Hu, W., Gao, H., Xie, B., Wang, X., Yin, Z., Li, S., and Wang, X. (2013) ADAM17 promotes U87 glioblastoma stem cell migration and invasion. *Brain research* **1538**, 151-158
38. Baumgart, A., Seidl, S., Vlachou, P., Michel, L., Mitova, N., Schatz, N., Specht, K., Koch, I., Schuster, T., Grundler, R., Kremer, M., Fend, F., Siveke, J. T., Peschel, C., Duyster, J., and Dechow, T. (2010) ADAM17 regulates epidermal growth factor receptor expression through the activation of Notch1 in non-small cell lung cancer. *Cancer research* **70**, 5368-5378
39. Xiao, L. J., Lin, P., Lin, F., Liu, X., Qin, W., Zou, H. F., Guo, L., Liu, W., Wang, S. J., and Yu, X. G. (2012) ADAM17 targets MMP-2 and MMP-9 via EGFR-MEK-ERK pathway activation to promote prostate cancer cell invasion. *International journal of oncology* **40**, 1714-1724
40. Bech-Serra, J. J., Santiago-Josefat, B., Esselens, C., Saftig, P., Baselga, J., Arribas, J., and Canals, F. (2006) Proteomic identification of desmoglein 2 and activated leukocyte cell adhesion molecule as substrates of ADAM17 and ADAM10 by difference gel electrophoresis. *Molecular and cellular biology* **26**, 5086-5095

41. Filmus, J., Shi, W., Wong, Z. M., and Wong, M. J. (1995) Identification of a new membrane-bound heparan sulphate proteoglycan. *The Biochemical journal* **311 (Pt 2)**, 561-565
42. Filmus, J., Capurro, M., and Rast, J. (2008) Glypicans. *Genome biology* **9**, 224
43. Su, G., Meyer, K., Nandini, C. D., Qiao, D., Salamat, S., and Friedl, A. (2006) Glypican-1 is frequently overexpressed in human gliomas and enhances FGF-2 signaling in glioma cells. *The American journal of pathology* **168**, 2014-2026
44. Kleeff, J., Ishiwata, T., Kumbasar, A., Friess, H., Buchler, M. W., Lander, A. D., and Korc, M. (1998) The cell-surface heparan sulfate proteoglycan glypican-1 regulates growth factor action in pancreatic carcinoma cells and is overexpressed in human pancreatic cancer. *The Journal of clinical investigation* **102**, 1662-1673
45. Matsuda, K., Maruyama, H., Guo, F., Kleeff, J., Itakura, J., Matsumoto, Y., Lander, A. D., and Korc, M. (2001) Glypican-1 is overexpressed in human breast cancer and modulates the mitogenic effects of multiple heparin-binding growth factors in breast cancer cells. *Cancer research* **61**, 5562-5569
46. Brunner, G., Metz, C. N., Nguyen, H., Gabrilove, J., Patel, S. R., Davitz, M. A., Rifkin, D. B., and Wilson, E. L. (1994) An endogenous glycosylphosphatidylinositol-specific phospholipase D releases basic fibroblast growth factor-heparan sulfate proteoglycan complexes from human bone marrow cultures. *Blood* **83**, 2115-2125
47. Kleeff, J., Wildi, S., Kumbasar, A., Friess, H., Lander, A. D., and Korc, M. (1999) Stable transfection of a glypican-1 antisense construct decreases tumorigenicity in PANC-1 pancreatic carcinoma cells. *Pancreas* **19**, 281-288
48. Traister, A., Shi, W., and Filmus, J. (2008) Mammalian Notum induces the release of glypicans and other GPI-anchored proteins from the cell surface. *The Biochemical journal* **410**, 503-511
49. Horiuchi, K., Le Gall, S., Schulte, M., Yamaguchi, T., Reiss, K., Murphy, G., Toyama, Y., Hartmann, D., Saftig, P., and Blobel, C. P. (2007) Substrate selectivity of epidermal growth factor-receptor ligand sheddases and their regulation by phorbol esters and calcium influx. *Molecular biology of the cell* **18**, 176-188
50. Le Gall, S. M., Bobe, P., Reiss, K., Horiuchi, K., Niu, X. D., Lundell, D., Gibb, D. R., Conrad, D., Saftig, P., and Blobel, C. P. (2009) ADAMs 10 and 17 represent differentially regulated components of a general shedding machinery for membrane proteins such as transforming growth factor alpha, L-selectin, and tumor necrosis factor alpha. *Molecular biology of the cell* **20**, 1785-1794
51. Sahin, U., Weskamp, G., Kelly, K., Zhou, H. M., Higashiyama, S., Peschon, J., Hartmann, D., Saftig, P., and Blobel, C. P. (2004) Distinct roles for ADAM10 and ADAM17 in ectodomain shedding of six EGFR ligands. *The Journal of cell biology* **164**, 769-779
52. Le Gall, S. M., Maretzky, T., Issuree, P. D., Niu, X. D., Reiss, K., Saftig, P., Khokha, R., Lundell, D., and Blobel, C. P. (2010) ADAM17 is regulated by a rapid and reversible mechanism that controls access to its catalytic site. *Journal of cell science* **123**, 3913-3922
53. Tucher, J., Linke, D., Koudelka, T., Cassidy, L., Tredup, C., Wichert, R., Pietrzik, C., Becker-Pauly, C., and Tholey, A. (2014) LC-MS based cleavage site profiling of the proteases ADAM10 and ADAM17 using proteome-derived peptide libraries. *Journal of proteome research* **13**, 2205-2214
54. Dang, M., Armbruster, N., Miller, M. A., Cermeno, E., Hartmann, M., Bell, G. W., Root, D. E., Lauffenburger, D. A., Lodish, H. F., and Herrlich, A. (2013) Regulated ADAM17-dependent EGF family ligand release by substrate-selecting signaling pathways. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9776-9781

55. Tellier, E., Canault, M., Rebsomen, L., Bonardo, B., Juhan-Vague, I., Nalbone, G., and Peiretti, F. (2006) The shedding activity of ADAM17 is sequestered in lipid rafts. *Experimental cell research* **312**, 3969-3980
56. Dierker, T., Dreier, R., Petersen, A., Bordych, C., and Grobe, K. (2009) Heparan sulfate-modulated, metalloprotease-mediated sonic hedgehog release from producing cells. *The Journal of biological chemistry* **284**, 8013-8022
57. Matthews, V., Schuster, B., Schutze, S., Bussmeyer, I., Ludwig, A., Hundhausen, C., Sadowski, T., Saftig, P., Hartmann, D., Kallen, K. J., and Rose-John, S. (2003) Cellular cholesterol depletion triggers shedding of the human interleukin-6 receptor by ADAM10 and ADAM17 (TACE). *The Journal of biological chemistry* **278**, 38829-38839
58. Kojro, E., Gimpl, G., Lammich, S., Marz, W., and Fahrenholz, F. (2001) Low cholesterol stimulates the nonamyloidogenic pathway by its effect on the alpha -secretase ADAM 10. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5815-5820
59. Gengrinovitch, S., Berman, B., David, G., Witte, L., Neufeld, G., and Ron, D. (1999) Glypican-1 is a VEGF165 binding proteoglycan that acts as an extracellular chaperone for VEGF165. *The Journal of biological chemistry* **274**, 10816-10822
60. Qiao, D., Meyer, K., Mundhenke, C., Drew, S. A., and Friedl, A. (2003) Heparan sulfate proteoglycans as regulators of fibroblast growth factor-2 signaling in brain endothelial cells. Specific role for glypican-1 in glioma angiogenesis. *The Journal of biological chemistry* **278**, 16045-16053
61. Teicher, B. A. (2006) Protein kinase C as a therapeutic target. *Clinical cancer research : an official journal of the American Association for Cancer Research* **12**, 5336-5345
62. Dolcet, X., Llobet, D., Pallares, J., and Matias-Guiu, X. (2005) NF-kB in development and progression of human cancer. *Virchows Archiv : an international journal of pathology* **446**, 475-482
63. Kohno, M., and Pouyssegur, J. (2006) Targeting the ERK signaling pathway in cancer therapy. *Annals of medicine* **38**, 200-211
64. McLean, G. W., Carragher, N. O., Avizienyte, E., Evans, J., Brunton, V. G., and Frame, M. C. (2005) The role of focal-adhesion kinase in cancer - a new therapeutic opportunity. *Nature reviews. Cancer* **5**, 505-515

3. DISCUSSÃO

3.1 Proteômica baseada em espectrometria de massas no estudo de candidados à biomarcadores em câncer oral

A proteômica baseada em espectrometria de massas progrediu enormemente nos últimos cinco anos, possibilitando a identificação e quantificação de proteínas bem como modificações pós-traducionais em larga escala em diferentes condições e sistemas biológicos. As últimas gerações de espectrômetros de massas sequencial combinam alta resolução e exatidão de massas com alta velocidade de aquisição de íons. Juntamente com o desenvolvimento de sistemas para análise de dados proteômicos, a espectrometria de massas tornou-se método de escolha para estudo de proteínas bem como para formulação de hipóteses diante de uma pergunta biológica (Mann e Kelleher, 2008).

O método científico consiste na observação de um fenômeno seguido de formulação, teste e modificação de uma hipótese. Por meio da proteômica baseada em descoberta, hipóteses são geradas a partir de um conjunto de informações sobre proteínas que podem estar diferencialmente reguladas em uma determinada condição em estudo. Essas proteínas direcionadas à hipótese podem ser testadas e quantificadas também pela espectrometria de massas, através da proteômica baseada em alvos. Assim, a espectrometria de massas tem sido uma poderosa ferramenta de aplicação no método científico, tanto para gerar hipóteses quanto para confirmá-las (Picotti, Bodenmiller *et al.*, 2013).

Os temas desenvolvidos nesse trabalho envolvem principalmente o estudo do subproteoma do secretoma de linhagens celulares, seja na busca à candidados marcadores em câncer oral ou no estudo do degradoma da metaloprotease ADAM17. Em ambos os temas, utilizou-se a espectrometria de massas para gerar hipóteses (*discovery-based proteomics*) e/ou testar hipóteses (*targeted proteomics*).

No contexto de formulação de hipótese por proteômica baseada em descoberta, o estudo apresentado no capítulo 1, intitulado *"Integrative Analysis to Select Cancer Candidate Biomarkers to Targeted Validation",* trouxe contribuições no sentido de apresentar métodos estatísticos de seleção de atributos (Beta-binomial, NSC e SVM-FRE), os quais em conjunto com diferentes estratégias de visualização dos dados em diagrama de Venn, agrupamento e redes de interações, foram capazes de revelar e priorizar um conjunto de proteínas diferencialmente

expressas no secretoma de seis linhagens celulares, subdivididas em três classes (melanoma, carcinoma e normal).

Até então, para fins de análise de proteínas diferencialmente expressas entre condições, os estudos proteômicos têm se baseado principalmente em estatísticas univariadas (ANOVA, test T, teste de Fisher etc) (Liu, Braakman *et al.*, 2012; Granato, Zanetti *et al.*, 2014; Simabuco, Kawahara *et al.*, 2014; Van Der Post e Hansson, 2014). Embora amplamente aplicados, esses testes são limitados à comparação multi-classes e na determinação de um conjunto ótimo de proteínas capazes de discriminar diferentes amostras com acurácia e sensibilidade satisfatória. A escolha dos métodos de seleção de atributos, Beta-binomial, NSC e SVM-FRE, se baseou em publicações recentes, nas quais foi comprovado melhor desempenho desses métodos na análise de dados proteômicos que utilizam contagem espectral como valor quantitativo e na comparação de múltiplas condições, possibilitando resultados com menor índice de falsos positivos (Christin *et al*., 2013; Pham *et al*., 2010). Além disso, NSC e SVM-FRE possibilitam uma etapa de validação juntamente com *double cross-validation* para definição de um conjunto ótimo de proteínas diferencialmente expressas capazes de distinguir dados provenientes de diferentes classes, com uma baixa taxa de erro.

Após aplicação desses métodos nos dados do secretoma, obteve-se um painel de 137 e 271 proteínas diferencialmente expressas nas classes carcinoma e melanoma, respectivamente. A hipótese gerada a partir desses dados foi que essas proteínas participam de processos biológicos e vias particulares dentro de cada classe e são candidatas à biomarcadores de cada condição: carcinoma, melanoma ou não cancerosa. As análises bioinformáticas de redes de interação proteína-proteína e enriquecimento de vias indicaram possíveis funções associadas aos marcadores de cada classe, por exemplo, ciclo celular e adesão celular em melanoma e coagulação e sistema complemento em carcinoma.

A validação da hipótese de que esses candidatos à biomarcadores se encontram com expressão aumentada e possuem um papel em câncer foi realizada utilizando técnicas quantitativas e funcionais complementares. Por imunohistoquímica, validou-se que a expressão de tenascin-C e GDF15 é significantemente maior em tecido de pacientes com melanoma comparado ao normal e a expressão de C3 e CFB é significantemente maior em tecidos de pacientes com carcinoma oral de células escamosas comparado ao tecido adjacente normal. Por meio da estratégia de proteômica baseada em alvos (*targeted proteomics)*, mostrou-se também que C3 e CFB possuem expressão aumentada em saliva de pacientes com carcinoma oral de células escamosas. Por fim, observou-se que CFB possui um papel funcional na migração de células tumorigênicas e quimiotaxia de macrófagos.

A metodologia de proteômica baseada em alvos proposta no capítulo 1 utilizou uma análise por monitoramento de reações paralelas (PRM) em instrumento do tipo LTQ Orbitrap Velos. Nesse tipo de análise, o precursor é selecionado (em uma faixa de massa de 15 Da) no modo de aquisição por *single ion monitoring* (SIM). Cada precursor referente ao peptídeo de interesse com *m/z* pré-determinada é monitorado em um evento durante a corrida cromatográfica e adquirido em alta resolução no analisador Orbitrap. Em tempos de retenções específicos e dentro de uma janela de 4 minutos, conhecida a partir de uma análise prévia por DDA, a *m/z* referente a cada peptídeo alvo é programada para ser selecionada e fragmentada no ion trap. Essa abordagem possui alta seletividade, uma vez que tanto a informação em alta resolução do precursor e quanto a confirmação deste pelos íons fragmentos são obtidos. Apesar dessa metodologia ter sido apropriada para esse trabalho, quando se tem um grande número de peptídeos de interesse a serem monitorados, esse tipo de análise perde em sensibilidade devido ao elevado tempo de ciclo (*cycle time*, tempo em que o espectrômetro leva para realizar todos os eventos: monitorar cada peptídeo e determinar os fragmentos) (Majovsky, Naumann *et al.*, 2014).

Como mostrado na introdução, o "padrão-ouro" que tem sido utilizado para proteômica baseada em alvos é o monitoramento seletivo de reações (SRM) realizado em triplo-quadrupolos e mais recentemente por PRM em quadrupolo-orbitrap (Peterson, Russell *et al.*, 2012). Nos triplo-quadrupolos, os dois quadrupolos atuam como filtros, nos quais somente precursor de *m/z* específico e seu respectivo fragmento será estável durante a trajetória e, portanto, detectado. A velocidade com que os quadrupolos monitoram cada transição é muito maior que a velocidade com que analisadores de alta resolução operam. Dessa forma, o tempo de ciclo para monitorar vários peptídeos é menor, possibilitando análise de um maior número de transições sem perda de sensibilidade.

Porém, a análise em triplo-quadrupolos necessita de uma otimização dos peptídeos (proteotípicos) bem como dos fragmentos que melhor se comportam (em termos de intensidade, reprodutibilidade e linearidade) no equipamento e na matriz de estudo. Muitos estudos utilizam algoritmos de predição ou banco de dados de experimentos por proteômica *shotgun* (conduzidos por aquisição dependente dos dados, DDA) (Lange, Picotti *et al.*, 2008). Contudo, nem sempre os peptídeos identificados por DDA irão produzir a melhor resposta em SRM (Stergachis, Maclean *et al.*, 2011).

Dessa forma, no estudo apresentado no capítulo 2, decidiu-se realizar a metodologia por proteômica baseada em alvos em espectrômetro de massas do tipo triplo-quadrupolo a fim de validar em saliva de pacientes com CEC oral 14 candidatos a biomarcadores em 30 amostras de

saliva humana, sendo 22 de pacientes com CEC oral e 8 de indivíduos saudáveis. Esse projeto foi realizado no laboratório do pesquisador Dr. Michael MacCoss (University of Washington), um dos mais reconhecidos na área de *targeted proteomics* principalmente pelo desenvolvimento de métodos em espectrometria de massas e de um *software* aberto, chamado de Skyline, para análise de dados de SRM, PRM e DDA. A grande importância desse trabalho encontra-se no desenvolvimento de um método de alta sensibilidade, especificidade e reprodutibilidade para análise de proteínas em saliva humana de vários candidatos biomarcadores ao mesmo tempo. Em um contexto clínico isso se torna ainda mais relevante devido à praticidade e facilidade da coleta da saliva, que pode ser usada para *screening* e monitoramento de CEC oral.

As 14 proteínas foram selecionadas a partir do painel de candidatos biomarcadores à carcinoma revelados no capítulo 1. Usamos como critério de seleção a disponibilidade de proteínas recombinantes padrão dos respectivos candidatos, os quais foram utilizados para determinar empiricamente os peptídeos proteotípicos (Stergachis, Maclean *et al.*, 2011) e à localização celular, na qual priorizamos as proteínas extracelulares.

Dois métodos de normalização foram realizados para proteína CFB: com peptídeos sintéticos do CFB marcados isotopicamente (VSEADSSNADWVT[$^{13}C_6$]K e YGLVTYATYP[$^{13}C_6$]K) ou peptídeos padrões de tempos de retenção presente em uma mistura comercial (*global standards*). Inicialmente, para avaliar o método analítico, foram realizadas curvas de calibração com diluições conhecidas da proteína recombinante CFB de 0,1 a 100 fmol/µl, previamente digerida com tripsina e misturada com quantidades conhecidas dos padrões sintéticos dos peptídeos do CFB ou da mistura comercial. Tanto a normalização dos peptídeos VSEADSSNADWVTCK e YGLVTYATYPCK do CFB com o respectivo peptídeo isotopicamente marcado quanto a normalização desses com peptídeos da mistura comercial resultaram em uma curva linear com valores de R superiores a 0.99. O coeficiente de variação dos peptídeos padrões, tanto originado do CFB *heavy* quanto da mistura comercial, entre as triplicatas da curva e entre as corridas com as amostras de saliva foi inferior a 25%, indicando alta reprodutibilidade e estabilidade desses peptídeos. Além disso, diferença estatística foi observada entre intensidades dos dois peptídeos do CFB (VSEADSSNADWVTCK e YGLVTYATYPCK) normalizados tanto pelo peptídeo do CFB *heavy* quanto pelo padrão da mistura comercial na comparação entre saliva de pacientes com CEC oral e saudáveis. Esse resultado mostrou que para o propósito de comparação relativa entre proteínas de diferentes condições, a normalização por *global standards* é uma estratégia que pode ser utilizada sem perda de acurácia e reprodutibilidade. Tal fato torna-se

relevante no contexto de análise de vários peptídeos, não sendo necessária a síntese de todos os respectivos padrões marcados.

O trabalho apresentado no capítulo 2 forneceu também uma descrição detalhada de como usar o programa Skyline. O desenvolvimento desse programa é liderado pelo pesquisador Dr. Brendan MacLean do grupo do Dr. Michael MacCoss e é um software aberto com milhares de usuários em todo o mundo. Em todas as etapas do estudo este software foi extensivamente utilizado, pois permite com muita facilidade exportar métodos contendo as transições dos peptídeos de interesse bem como visualizar os resultados para refinamento e otimização do método (Maclean, Tomazela *et al.*, 2010). Além disso, o programa possui acesso à ferramentas externas, como Quasar e MSstats, os quais permitem análise estatística dos dados provenientes de SRM, muito úteis por exemplo para cálculo de limite de quantificação e detecção, reprodutibilidade e comparação das intensidades dos peptídeos avaliados entre as condições em estudo (Choi, Chang *et al.*, 2014).

Embora bastante seletivo e sensível, a baixa resolução inerente do quadrupolo possibilita a detecção de interferentes que podem gerar precursores e fragmentos com mesma *m/z* do peptídeo de interesse. Por esse motivo, muitas vezes, sem um padrão interno, a confiabilidade na identificação do peptídeo de interesse torna-se limitada. Por isso, o programa Skyline possui ferramentas que auxiliam na identificação do pico correto através da comparação com bibliotecas de espectros bem como predições de tempo de retenção. Essas ferramentas também foram detalhadas no capítulo 2.

Ao final desse estudo, foram apresentados cinco candidatos à biomarcadores (CFB, C3, C4B, SERPINA1 e LRG1) validados com maior expressão em pacientes com CEC oral comparados com pacientes saudáveis. Além disso, foi verificado que a alta expressão desses candidatos possui associação com maior risco de desenvolvimento de CEC oral. Interessante notar que quatro dessas cinco proteínas fazem parte do processo biológico coagulação e sistema complemento, o qual foi descrito no capítulo 1 como um dos processos exclusivamente enriquecido na classe de carcinoma.

Esse trabalho abriu novas perspectivas para futuros estudos que contemplem um maior número de amostras com histórico clínico, de forma a possibilitar análises de correlação da expressão dessas proteínas com prognóstico ou resposta ao tratamento. Isso é de fato um dos aspectos mais relevantes na busca de biomarcadores em câncer oral, uma vez que como descrito na introdução, atualmente o sistema TNM é ainda o principal marcador de prognóstico, mas ainda tem uso limitado devido ao curso clínico variado e à heterogeneidade do tumor.

Vale lembrar que o estudo de proteínas com expressão alterada em câncer é util não somente para descoberta de biomarcadores e alvos terapêuticos, mas também para entender

mecanismos moleculares em tumorigênese, progressão e metástase. Nesse contexto, ensaios funcionais envolvendo manipulações da expressão de genes *in vitro* utilizando, por exemplo, tratamento com inibidores, RNA antisense, RNA interferência, super-expressão de genes ou *in vivo* com modelos ortotópicos ou de camundongos *knockout* fazem parte das validações de candidatos à biomarcadores que podem ser úteis clinicamente (Guo, Zou *et al.*, 2013).

O capítulo 3 apresenta um estudo funcional de dois proteoglicanos de matriz extracelular encontrados com expressão alterada no secretoma de células de carcinoma (capítulo 1) e validados com maior expressão por RTqPCR em tecidos de pacientes com CEC oral. Agrina e perlecan foram estudados quanto à sua função em câncer oral, por meio de silenciamento transiente por siRNA seguido de ensaios funcionais *in vitro* relacionados à processos tumorigênicos como proliferação, adesão, migração e resistência à cisplatina.

Nesse trabalho verificamos que a expressão de agrina e perlecan depende do sítio de origem, sendo que em células de câncer oral de origem não metastática (SCC-9) a expressão de agrina é maior e em células de origem metastática (SCC-9 LN1) a expressão de perlecan é maior. Funcionalmente, o silenciamento da expressão de agrina e perlecan por siRNA resultou na diminuição da capacidade de células SCC-9 e/ou SCC-9 LN1 em aderir à proteínas de matriz extracelular, migrar e proliferar.

Verificamos também que a modificação de condroitin sulfato, presente tanto na agrina quanto no perlecan, é importante na adesão de células à matriz extracelular e na resistência à cisplatina. Tanto a remoção dessa modificação pela enzima condroitinase quanto o silenciamento da expressão da agrina resultaram no aumento da sensibilidade de células SCC-9, SCC-9 LN1 e A431 à cisplatina.

Esse trabalho, por fim publicado na *Plos One* (v.9, n.12, 2014), trouxe importante contribuição no estudo da função de proteoglicanos em câncer oral, abrindo novos caminhos para estratégias terapêuticas inovadoras que visem tanto a regulação de níveis de expressão de agrina e perlecan no microambiente tumoral quanto de modificações pós-traducionais de condroitin sulfato.

3.2 Proteômica baseada em espectrometria de massas na identificação de novos alvos da ADAM17

A ADAM17 é uma metaloprotease de membrana envolvida em diversos processos tumorigênicos como adesão, migração, proliferação e invasão. A caracterização de substratos dessa

protease é importante para elucidar sua função biológica bem como para entender o papel da proteólise nos fenótipos observados, sobretudo em câncer (Arribas e Esselens, 2009).

Uma vez que a ADAM17 gera produtos de clivagem que atuam no meio extracelular, como ectodomínios de proteínas de membrana, fatores de crescimento, receptores de superfície e proteínas de matriz extracelular, o secretoma torna-se o modelo mais adequado para estudo do degradoma da ADAM17.

No capítulo 4 caracterizamos o secretoma de linhagens de fibroblastos embrionários provenientes de camundongos selvagem e *knockout* para ADAM17. Utilizando a proteômica quantitativa *label-free* e *SILAC*, proteínas diferencialmente expressas presentes nesses subproteomas foram determinadas e correlacionadas com vias de ERK e FAK, bem como fenótipos celulares como proliferação e migração, determinados pela presença ou ausência da ADAM17.

Novos alvos de matriz extracelular e de membrana para ADAM17 foram elucidados, como mimecan e perlecan, os quais foram validados pela incubação das recombinantes da protease e substrato *in vitro* e os sítios de clivagens determinados por marcação do N-terminal seguido de análise por espectrometria de massas. Além disso, confirmamos que ADAM17 não possui somente uma função de clivagem direta de substratos, como também participa modulando indiretamente a expressão dessas proteínas. Por meio de análise bioinformática de predição de sequência, verificamos que proteínas reguladas positivamente no secretoma de mEFs selvagem, isto é, na presença de ADAM17, possuem em comum o motivo de ligação ao fator de transcrição PPARγ. Como prova de conceito, confirmamos que na presença da ADAM17 esse fator de transcrição está mais ativo.

A grande contribuição desse trabalho, publicado na *Journal of Proteome Research* (v.13, n.4, 2014), foi que a ADAM17 está envolvida na regulação direta e indireta de substratos, não somente em nível de proteólise como também em nível de expressão gênica. Estudos futuros ainda são necessários para elucidar o papel biológico da clivagem do mimecan e perlecan, bem como os mecanismos envolvidos na regulação da expressão gênica de outras proteínas e ativação de PPARγ pela ADAM17.

Tendo em vista as várias evidências do envolvimento da ADAM17 em câncer oral, embora nenhum estudo tenha caracterizado o repertório de substratos nesse tipo de câncer, realizamos um trabalho de busca de novos alvos da ADAM17 em células de carcinoma oral de células escamosas isoladas de língua humana (SCC-9) mostrado no capítulo 5.

Utilizando duas estratégias de espectrometria de massas aplicada à análise de proteínas presentes no meio extracelular, glypican-1 foi revelado como candidato a novo alvo da ADAM17.

Glypican-1 é um proteoglicano ancorado à superfície da célula por meio de uma âncora de glicosilfosfatidilinositol (GPI). Diversas foram as evidências que suportaram nossa escolha para expandir o estudo do glypican-1 como novo alvo da ADAM17. (1) Glypican-1 foi encontrada exclusivamente no secretoma de células SCC-9 incubadas com ADAM17 recombinante e com abundância significantemente superior no secretoma de células SCC-9 controle em relação à silenciada para ADAM17. (2) Glypican-1 foi identificada como diferencialmente expressa no secretoma das mEFs *Wild-type* e *knockout* para ADAM17, mostrada no capítulo 4. (3) Glypican-1 faz parte da mesma classe de proteoglicanos de superfície de um alvo conhecido da ADAM17: os sindecanos (Pruessmeyer, Martin *et al.*, 2010).

Para avaliar se a presença de glypican-1 no meio extracelular era resultado de um evento de clivagem por metaloprotease, a sequência do glypican-1 foi clonada em fusão com a enzima fosfatase alcalina e transfectada em células de carcinoma de células escamosas da pele (A431). Esse ensaio é bem estabelecido e publicado em diversos trabalhos para estudo de alvos da ADAM17 (Sahin, Weskamp *et al.*, 2004; Le Gall, Maretzky *et al.*, 2010; Maretzky, Mcilwain *et al.*, 2013). Resumidamente, o ensaio consiste em super-expressar essa quimera (substrato + fosfatase alcalina) e tratar as células com conhecidos ativadores pleiotrópicos (PMA) ou fisiológicos (EGF) da ADAM17 ou inibidores de metaloproteases (marimastat). Se houver clivagem do substrato, a fosfatase alcalina fusionada ao substrato será liberada para o meio extracelular e a atividade dela poderá ser medida assim que for adicionado o respectivo substrato. A atividade da fosfatase alcalina sobre o substrato produz um composto de coloração amarela que é medido em absorbância de 405 nm. Verificamos que tanto o ativador pleiotrópico quanto fisiológico promovem aumento de liberação de glypican-1 para o meio extracelular, enquanto que a inibição de metaloproteases com marimastat reverte esse efeito.

Além disso, mostramos que a clivagem de GPC1 pode ser regulada pela modulação da sua disponibilidade na membrana. GPC1 foi descrita localizada preferencialmente em *lipids rafts*, os quais consistem em microdomínios de membrana ricos em colesterol (Gutierrez e Brandan, 2010). Primeiramente, verificamos que a depleção de colesterol por *methyl-β-ciclodextrin* (MβCD) promove o aumento da liberação de GPC1 para o meio extracelular por PMA. Levantamos então a hipótese de que esse aumento da clivagem de GPC1 pelo PMA após depleção de colesterol deve-se a maior disponibilidade do substrato na superfície da célula, onde ADAM17 se localiza e realiza a proteólise. De fato, observamos que o tratamento com MβCD promove aumento de GPC1 na superfície celular, detectado por *western blot* após enriquecimento de proteínas de membrana por biotinilação.

Também validamos por *western blot* que a clivagem de GPC1 é mediada pela ADAM17 em secretoma de células SCC-9 e A431 *scrambled* e *knockdown* para ADAM17 e o sítio de clivagem do GPC1 foi determinado por espectrometria de massas após incubação *in vitro* de GPC1 e ADAM17 recombinante.

A fim de estudar o papel do glypican-1 solúvel no meio extracelular, ensaios de imunoprecipitação de GPC1 recombinante seguido de análise por espectrometria de massas foram realizados. Por meio da determinação dos parceiros de interação do GPC1 presentes no meio extracelular, seguido de análises bioinformáticas de rede de interações e enriquecimento de processos biológicos, observamos que o complexo de proteínas que se liga ao GPC1 está envolvido principalmente em adesão, migração e proliferação. Esses eventos foram avaliados em ensaios funcionais com células A431 controle e super-expressando GPC1 (inteiro e fragmento referente à porção N-terminal clivada pela ADAM17), bem como em células submetidas a tratamento com PMA, marimastat ou com eluídos do complexo GPC1 obtidos da imunoprecipitação. Os resultados revelaram que na super-expressão de GPC1, isto é, provavelmente ancorado à membrana, há menor adesão e maior migração e proliferação. Quando a clivagem de GPC1 é estimulada por PMA, há aumento da adesão, porém nenhum efeito em migração e proliferação. Em contrapartida, quando a clivagem de GPC1 é inibida por marimastat, há aumento de migração e proliferação. Interessante notar que quando as células A431 são tratadas somente com o complexo de proteínas do GPC1 presente no meio extracelular ou quando super-expressamos somente a porção clivada do GPC1, há inibição da migração.

Por fim, o trabalho apresentado do capítulo 5 sugeriu que a clivagem de GPC1 é mediada por ADAM17 e que a função dessa proteína em diminuir adesão e aumentar proliferação e migração pode estar associada à sua forma ancorada à membrana. Dessa forma, a função de glypican-1 em câncer pode ser modulada por clivagem mediada pela ADAM17.

Apesar da ADAM17 ser descrita envolvida no aumento de adesão, proliferação e migração celular, esse trabalho trouxe uma nova compreensão em um contexto mais amplo, onde a função de proteases também é determinada por seus substratos. Uma vez que a expressão de substratos depende de cada tipo celular e condição biológica, metodologias que permitam o estudo do degradoma da ADAM17 em diferentes contextos são essenciais para entender sua função.

Cabe relembrar que estratégias utilizando inibidores de amplo espectro de metaloproteases bem como específicos para ADAM17 tem falhado clinicamente e em partes isso se deve à falta de conhecimento dos substratos e do efeito biológico resultante da clivagem. Por isso grande é a importância do estudo do degradoma de proteases e, como mostrado nos capítulos 4 e 5,

a proteômica baseada em espectrometria de massas, é sem dúvidas uma ferramenta muito robusta para esse fim.

Em resumo, a espectrometria de massas em conjunto com diversas técnicas complementares contribuiu na descoberta de novos alvos regulados direta ou indiretamente pela ADAM17, bem como na determinação de sítios de clivagem de substratos. Além disso, indicou-se vias de sinalização e processos biológicos que são alterados pela expressão diferencial de proteínas na presença da ADAM17 ou de proteínas presentes em um complexo de interação com um alvo (glypican-1) dessa protease.

4. CONSIDERAÇÕES FINAIS

A proteômica baseada em espectrometria de massas aplicada na análise de proteínas secretadas, seja na busca por candidatos biomarcadores ou na descoberta de novos alvos da ADAM17, trouxe novas compreensões na biologia do câncer, além de várias perspectivas para futuros estudos.

No capítulo 1, o *pipeline* proposto, contendo métodos de seleção de atributos, pela primeira vez utilizados em conjunto para revelar proteínas com diferença em abundância entre amostras, levantou um painel de candidatos à biomarcadores em carcinoma e melanoma. As análises bioinformáticas revelaram processos biológicos particularmente enriquecidos em cada condição, por exemplo, ciclo celular e adesão celular em melanoma e coagulação e sistema complemento em carcinoma. A validação de quatro proteínas (C3, CFB, C4B e SERPINA1) envolvidas no mesmo processo de coagulação e sistema complemento despertou novas perspectivas para futuros estudos dessas proteínas, sobretudo do contexto de biomarcadores de prognóstico, bem como desse processo biológico em câncer oral (Capítulo 2).

O estudo funcional de proteoglicanos de matriz extracelular desvendou *in vitro* o papel da agrina e perlecan em células de câncer oral na adesão, migração, proliferação e sensibilidade à cisplatina (Capítulo 3). Esses resultados deram início a uma nova etapa, na qual a função *in vivo* dessas proteínas bem como a utilidade delas como marcadores clínicos de prognóstico e resposta terapêutica serão explorados.

No contexto do microambiente tumoral, a ADAM17 exerce importante função pela clivagem de proteínas de membrana responsáveis por sinalizar de forma autócrina ou parácrina vários processos biológicos, entre eles proliferação, migração e adesão celular. A análise do secretoma de células na presença ou ausência da ADAM17 revelou um conjunto de proteínas relacionadas à essas conhecidas funções, mas que podem ser reguladas direta ou indiretamente pela ADAM17. Tanto o trabalho apresentado no capítulo 4 como no capítulo 5 resultaram em uma lista de candidatos à novos alvos para essa protease que também podem ser avaliados em estudos futuros. A validação direta de clivagem dos novos alvos mimecan, perlecan e glypican-1 pela ADAM17 ainda necessita de mais investigações quanto ao papel *in vivo* desses alvos como também dos seus produtos de clivagem gerados.

Por fim, dentro do método científico, a espectrometria de massas levantou diversas hipóteses na identificação de candidatos à biomarcadores em câncer e novos alvos da ADAM17. Por meio desses estudos, novas perguntas biológicas surgiram, permitindo que o ciclo dentro do

método científico recomece, na formulação, teste e confirmação de hipóteses, onde certamente a proteômica baseada em espectrometria de massas pode ser usada como ferramenta na busca pelo novo conhecimento.

5. REFERÊNCIAS

AEBERSOLD, R.; MANN, M. Mass spectrometry-based proteomics. **Nature,** v. 422, n. 6928, p. 198-207, Mar 13 2003.

AIKAWA, T. et al. Glypican-1 modulates the angiogenic and metastatic potential of human and mouse cancer cells. **J Clin Invest,** v. 118, n. 1, p. 89-99, Jan 2008.

ARAGAO, A. Z. et al. Novel processed form of syndecan-1 shed from SCC-9 cells plays a role in cell migration. **PLoS One,** v. 7, n. 8, p. e43521, 2012.

ARRIBAS, J.; ESSELENS, C. ADAM17 as a therapeutic target in multiple diseases. **Curr Pharm Des,** v. 15, n. 20, p. 2319-35, 2009.

BANTSCHEFF, M. et al. Quantitative mass spectrometry in proteomics: a critical review. **Anal Bioanal Chem,** v. 389, n. 4, p. 1017-31, Oct 2007.

BARASH, U. et al. Proteoglycans in health and disease: new concepts for heparanase function in tumor progression and metastasis. **FEBS J,** v. 277, n. 19, p. 3890-903, Oct 2010.

BATMUNKH, E. et al. Comparison of the expression of agrin, a basement membrane heparan sulfate proteoglycan, in cholangiocarcinoma and hepatocellular carcinoma. **Hum Pathol,** v. 38, n. 10, p. 1508-15, Oct 2007.

BAUVOIS, B. New facets of matrix metalloproteinases MMP-2 and MMP-9 as cell surface transducers: outside-in signaling and relationship to tumor progression. **Biochim Biophys Acta,** v. 1825, n. 1, p. 29-36, Jan 2012.

BENSIMON, A.; HECK, A. J.; AEBERSOLD, R. Mass spectrometry-based proteomics and network biology. **Annu Rev Biochem,** v. 81, p. 379-405, 2012.

BEREMAN, M. S. et al. The development of selected reaction monitoring methods for targeted proteomics via empirical refinement. **Proteomics,** v. 12, n. 8, p. 1134-41, Apr 2012.

BERGER, A.; SCHECHTER, I. Mapping the active site of papain with the aid of peptide substrates and inhibitors. **Philos Trans R Soc Lond B Biol Sci,** v. 257, n. 813, p. 249-64, Feb 12 1970.

BIGLER, L. R. et al. The potential use of saliva to detect recurrence of disease in women with breast carcinoma. **J Oral Pathol Med,** v. 31, n. 7, p. 421-31, Aug 2002.

BIOMARKERS DEFINITIONS WORKING, G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. **Clin Pharmacol Ther,** v. 69, n. 3, p. 89-95, Mar 2001.

BLACK, R. A.; WHITE, J. M. ADAMs: focus on the protease domain. **Curr Opin Cell Biol,** v. 10, n. 5, p. 654-9, Oct 1998.

CARR, S. A. et al. Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. **Mol Cell Proteomics,** v. 13, n. 3, p. 907-17, Mar 2014.

CHAKRABORTY, S. et al. An oncogenic role of Agrin in regulating focal adhesion integrity in hepatocellular carcinoma. **Nat Commun,** v. 6, p. 6184, 2015.

CHANG, C. Y. et al. Protein significance analysis in selected reaction monitoring (SRM) measurements. **Mol Cell Proteomics,** v. 11, n. 4, p. M111 014662, Apr 2012.

CHEN, X. et al. ADAM17 promotes U87 glioblastoma stem cell migration and invasion. **Brain Res,** v. 1538, p. 151-8, Nov 13 2013.

CHENG, Y. S.; REES, T.; WRIGHT, J. A review of research on salivary biomarkers for oral cancer detection. **Clin Transl Med,** v. 3, n. 1, p. 3, 2014.

CHOI, M. et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. **Bioinformatics,** v. 30, n. 17, p. 2524-6, Sep 1 2014.

CLAMP, M. et al. Distinguishing protein-coding and noncoding genes in the human genome. **Proc Natl Acad Sci U S A,** v. 104, n. 49, p. 19428-33, Dec 4 2007.

COHEN, I. R. et al. Abnormal expression of perlecan proteoglycan in metastatic melanomas. **Cancer Res,** v. 54, n. 22, p. 5771-4, Nov 15 1994.

COSTA, A. L. L.; DE ARAÚJO JÚNIOR, R. F.; RAMOS, C. C. F. Correlation between TNM classification and malignancy histological feature of oral squamous cell carcinoma. **Rev. Bras. Otorrinolaringol.,** v. 71, n. 2, 2005.

COX, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. **J Proteome Res,** v. 10, n. 4, p. 1794-805, Apr 1 2011.

DA SILVA, S. D. et al. Advances and applications of oral cancer basic research. **Oral Oncol,** v. 47, n. 9, p. 783-91, Sep 2011.

DASS, C. **Fundamentals of contemporary mass spectrometry**. Hoboken, New Jersey: John Wiley & Sons, Inc, 2007. ISBN 9780471682295.

DE JONG, E. P. et al. Quantitative proteomics reveals myosin and actin as promising saliva biomarkers for distinguishing pre-malignant and malignant oral lesions. **PLoS One,** v. 5, n. 6, p. e11148, 2010.

DEAN, R. A.; OVERALL, C. M. Proteomics discovery of metalloproteinase substrates in the cellular context by iTRAQ labeling reveals a diverse MMP-2 substrate degradome. **Mol Cell Proteomics,** v. 6, n. 4, p. 611-23, Apr 2007.

DECOCK, J. et al. Plasma MMP1 and MMP8 expression in breast cancer: protective role of MMP8 against lymph node metastasis. **BMC Cancer,** v. 8, p. 77, 2008.

DIONNE, K. R. et al. Potentially malignant disorders of the oral cavity: Current practice and future directions in the clinic and laboratory. **Int J Cancer**, Jan 31 2014.

DOMON, B.; AEBERSOLD, R. Mass spectrometry and protein analysis. **Science,** v. 312, n. 5771, p. 212-7, Apr 14 2006.

DOMON, B. A., R. Options and considerations when selecting a quantitative proteomics strategy. **Nat Biotechnol,** v. 28, n. 7, p. 710-21, Jul 2010.

DOUCET, A. et al. Metadegradomics: toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. **Mol Cell Proteomics,** v. 7, n. 10, p. 1925-51, Oct 2008.

DOUCET, A.; OVERALL, C. M. Protease proteomics: revealing protease in vivo functions using systems biology approaches. **Mol Aspects Med,** v. 29, n. 5, p. 339-58, Oct 2008.

DUFFY, M. J. et al. The ADAMs family of proteases: new biomarkers and therapeutic targets for cancer? **Clin Proteomics,** v. 8, n. 1, p. 9, 2011.

EDWARDS, I. J. Proteoglycans in prostate cancer. **Nat Rev Urol,** v. 9, n. 4, p. 196-206, Apr 2012.

ENG, J. K.; MCCORMACK, A. L.; YATES, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. **J Am Soc Mass Spectrom,** v. 5, n. 11, p. 976-89, Nov 1994.

FELLER, L.; LEMMER, J. Oral Squamous Cell Carcinoma: Epidemiology, Clinical Presentation and Treatment. **Journal of Cancer Therapy,** v. 3, n. 4, p. 263-268, 2012.

FIGEYS, D. **Industrial Proteomics: Applications for Biotechnology and Pharmaceuticals**. John Wiley & Sons, Inc, 2005. ISBN 0-471-45714-0.

FRANTZ, C.; STEWART, K. M.; WEAVER, V. M. The extracellular matrix at a glance. **J Cell Sci,** v. 123, n. Pt 24, p. 4195-200, Dec 15 2010.

FUZERY, A. K. et al. Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. **Clin Proteomics,** v. 10, n. 1, p. 13, 2013.

GALLIEN, S. et al. Selectivity of LC-MS/MS analysis: implication for proteomics experiments. **J Proteomics,** v. 81, p. 148-58, Apr 9 2013.

GIALELI, C.; THEOCHARIS, A. D.; KARAMANOS, N. K. Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting. **FEBS J,** v. 278, n. 1, p. 16-27, Jan 2011.

GILLETTE, M. A.; CARR, S. A. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. **Nat Methods,** v. 10, n. 1, p. 28-34, Jan 2013.

GLISH, G. L.; VACHET, R. W. The basics of mass spectrometry in the twenty-first century. **Nat Rev Drug Discov,** v. 2, n. 2, p. 140-50, Feb 2003.

GOOZ, M. ADAM-17: the enzyme that does it all. **Crit Rev Biochem Mol Biol,** v. 45, n. 2, p. 146-69, Apr 2010.

GRANATO, D. C. et al. Integrated proteomics identified up-regulated focal adhesion-mediated proteins in human squamous cell carcinoma in an orthotopic murine model. **PLoS One,** v. 9, n. 5, p. e98208, 2014.

GSTAIGER, M.; AEBERSOLD, R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. **Nat Rev Genet,** v. 10, n. 9, p. 617-27, Sep 2009.

GUO, S.; ZOU, J.; WANG, G. Advances in the proteomic discovery of novel therapeutic targets in cancer. **Drug Des Devel Ther,** v. 7, p. 1259-71, 2013.

GUTIERREZ, J.; BRANDAN, E. A novel mechanism of sequestering fibroblast growth factor 2 by glypican in lipid rafts, allowing skeletal muscle differentiation. **Mol Cell Biol,** v. 30, n. 7, p. 1634-49, Apr 2010.

GYGI, S. P. et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. **Nat Biotechnol**, v.17, n.10, p.994-999, 1999.

HAN, X.; ASLANIAN, A.; YATES, J. R., 3RD. Mass spectrometry for proteomics. **Curr Opin Chem Biol,** v. 12, n. 5, p. 483-90, Oct 2008.

HANAHAN, D.; WEINBERG, R. A. Hallmarks of cancer: the next generation. **Cell,** v. 144, n. 5, p. 646-74, Mar 4 2011.

HERZOG, C. et al. ADAM10 is the major sheddase responsible for the release of membrane-associated meprin A. **J Biol Chem,** v. 289, n. 19, p. 13308-22, May 9 2014.

HSU, C. W. et al. Secretome Profiling of Primary Cells Reveals That THBS2 Is a Salivary Biomarker of Oral Cavity Squamous Cell Carcinoma. **J Proteome Res**, Apr 16 2014.

HU, Q. et al. The Orbitrap: a new mass spectrometer. **J Mass Spectrom,** v. 40, n. 4, p. 430-43, Apr 2005.

HU, S. et al. Salivary proteomics for oral cancer biomarker discovery. **Clin Cancer Res,** v. 14, n. 19, p. 6246-52, Oct 1 2008.

HU, S. et al. Preclinical validation of salivary biomarkers for primary Sjogren's syndrome. **Arthritis Care Res (Hoboken),** v. 62, n. 11, p. 1633-8, Nov 2010.

HUANG, Y. et al. Targeting the sheddase activity of ADAM17 by an anti-ADAM17 antibody D1(A12) inhibits head and neck squamous cell carcinoma cell proliferation and motility via blockage of bradykinin induced HERs transactivation. **Int J Biol Sci,** v. 10, n. 7, p. 702-14, 2014.

IMPENS, F. et al. MS-driven protease substrate degradomics. **Proteomics,** v. 10, n. 6, p. 1284-96, Mar 2010.

INSTITUTO NACIONAL DE CÂNCER, I. **Estimativa 2014: Incidência de Câncer no Brasil**. Rio de Janeiro: INCA**:** 124 p. 2014.

IOZZO, R. V.; SANDERSON, R. D. Proteoglycans in cancer biology, tumour microenvironment and angiogenesis. **J Cell Mol Med,** v. 15, n. 5, p. 1013-31, May 2011.

IOZZO, R. V.; ZOELLER, J. J.; NYSTROM, A. Basement membrane proteoglycans: modulators Par Excellence of cancer growth and angiogenesis. **Mol Cells,** v. 27, n. 5, p. 503-13, May 31 2009.

JEDRYCHOWSKI, M. P. et al. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. **Mol Cell Proteomics,** v. 10, n. 12, p. M111 009910, Dec 2011.

KARAGIANNIS, G. S.; PAVLOU, M. P.; DIAMANDIS, E. P. Cancer secretomics reveal pathophysiological pathways in cancer molecular oncology. **Mol Oncol,** v. 4, n. 6, p. 496-510, Dec 2010.

KESHISHIAN, H. et al. Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. **Mol Cell Proteomics,** v. 8, n. 10, p. 2339-49, Oct 2009.

KESSENBROCK, K.; PLAKS, V.; WERB, Z. Matrix metalloproteinases: regulators of the tumor microenvironment. **Cell,** v. 141, n. 1, p. 52-67, Apr 2 2010.

KINTER, M. K. S., N. E. **Protein sequence and identification using tandem mass spectrometry**. New York: A John Wiley and Songs, Inc., Publication, 2000. ISBN 0-471-32249-0.

KLEIFELD, O. et al. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. **Nat Biotechnol,** v. 28, n. 3, p. 281-8, Mar 2010.

KOOMEN, J. M. et al. Proteomic contributions to personalized cancer care. **Mol Cell Proteomics,** v. 7, n. 10, p. 1780-94, Oct 2008.

KULASINGAM, V.; DIAMANDIS, E. P. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. **Nat Clin Pract Oncol,** v. 5, n. 10, p. 588-99, Oct 2008.

LANGE, V. et al. Selected reaction monitoring for quantitative proteomics: a tutorial. **Mol Syst Biol,** v. 4, p. 222, 2008.

LARBI, N. B.; JEFFERIES, C. 2D-DIGE: comparative proteomics of cellular signalling pathways. **Methods Mol Biol,** v. 517, p. 105-32, 2009.

LAU, C. et al. Role of pancreatic cancer-derived exosomes in salivary biomarker development. **J Biol Chem,** v. 288, n. 37, p. 26888-97, Sep 13 2013.

LE GALL, S. M. et al. ADAM17 is regulated by a rapid and reversible mechanism that controls access to its catalytic site. **J Cell Sci,** v. 123, n. Pt 22, p. 3913-22, Nov 15 2010.

LEE, K. H. et al. Role of combined modality treatment of buccal mucosa squamous cell carcinoma. **Aust Dent J,** v. 50, n. 2, p. 108-13, Jun 2005.

LIN, P. et al. ADAM17 regulates prostate cancer cell proliferation through mediating cell cycle progression by EGFR/PI3K/AKT pathway. **Mol Cell Biochem,** v. 359, n. 1-2, p. 235-43, Jan 2012.

LIU, N. Q. et al. Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. **J Mammary Gland Biol Neoplasia,** v. 17, n. 2, p. 155-64, Jun 2012.

MACLEAN, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. **Bioinformatics,** v. 26, n. 7, p. 966-8, Apr 1 2010.

MAJOVSKY, P. et al. Targeted Proteomics Analysis of Protein Degradation in Plant Signaling on an LTQ-Orbitrap Mass Spectrometer. **J Proteome Res,** v. 13, n. 10, p. 4246-58, Oct 3 2014.

MAKAROV, A. **Mass spectrometer**. PATENT, U. S. US5886346 A 1999.

MAKAWITA, S.; DIAMANDIS, E. P. The bottleneck in the cancer biomarker pipeline and protein quantification through mass spectrometry-based approaches: current strategies for candidate verification. **Clin Chem,** v. 56, n. 2, p. 212-22, Feb 2010.

MAKRIDAKIS, M.; VLAHOU, A. Secretome proteomics for discovery of cancer biomarkers. **J Proteomics,** v. 73, n. 12, p. 2291-305, Nov 10 2010.

MALLICK, P.; KUSTER, B. Proteomics: a pragmatic perspective. **Nat Biotechnol,** v. 28, n. 7, p. 695-709, Jul 2010.

MANN, M.; KELLEHER, N. L. Precision proteomics: the case for high resolution and high mass accuracy. **Proc Natl Acad Sci U S A,** v. 105, n. 47, p. 18132-8, Nov 25 2008.

MARETZKY, T. et al. iRhom2 controls the substrate selectivity of stimulated ADAM17-dependent ectodomain shedding. **Proc Natl Acad Sci U S A,** v. 110, n. 28, p. 11433-8, Jul 9 2013.

MARETZKY, T. et al. ADAM10 mediates E-cadherin shedding and regulates epithelial cell-cell adhesion, migration, and beta-catenin translocation. **Proc Natl Acad Sci U S A,** v. 102, n. 26, p. 9182-7, Jun 28 2005.

MARKOPOULOS, A. K. Current aspects on oral squamous cell carcinoma. **Open Dent J,** v. 6, p. 126-30, 2012.

MARTI-ALAMO, S. et al. Saliva as a diagnostic fluid. Literature review. **J Clin Exp Dent,** v. 4, n. 4, p. e237-43, Oct 2012.

MARTINS-DE-SOUZA, D. Is the word 'biomarker' being properly used by proteomics research in neuroscience? **Eur Arch Psychiatry Clin Neurosci,** v. 260, n. 7, p. 561-2, Oct 2010.

MARX, V. Targeted proteomics. **Nat Methods,** v. 10, n. 1, p. 19-22, Jan 2013.

MUCKE, T. et al. Recurrence interval affects survival after local relapse of oral cancer. **Oral Oncol,** v. 45, n. 8, p. 687-91, Aug 2009.

MURPHY, G. The ADAMs: signalling scissors in the tumour microenvironment. **Nat Rev Cancer,** v. 8, n. 12, p. 929-41, Dec 2008.

NAGARAJ, N. et al. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. **Mol Cell Proteomics,** v. 11, n. 3, p. M111 013722, Mar 2012.

NAGASE, H.; VISSE, R.; MURPHY, G. Structure and function of matrix metalloproteinases and TIMPs. **Cardiovasc Res,** v. 69, n. 3, p. 562-73, Feb 15 2006.

NEILSON, K. A. et al. Less label, more free: approaches in label-free quantitative mass spectrometry. **Proteomics,** v. 11, n. 4, p. 535-53, Feb 2011.

NOGUEIRA, F. C.; DOMONT, G. B. Survey of shotgun proteomics. **Methods Mol Biol,** v. 1156, p. 3-23, 2014.

NOGUTI, J. et al. Metastasis from oral cancer: an overview. **Cancer Genomics Proteomics,** v. 9, n. 5, p. 329-35, Sep-Oct 2012.

O'FARRELL, P. H. High resolution two-dimensional electrophoresis of proteins. **J Biol Chem,** v. 250, n. 10, p. 4007-21, May 25 1975.

OHTSU, H.; DEMPSEY, P. J.; EGUCHI, S. ADAMs as mediators of EGF receptor transactivation by G protein-coupled receptors. **Am J Physiol Cell Physiol,** v. 291, n. 1, p. C1-10, Jul 2006.

OLSEN, J. V. et al. Higher-energy C-trap dissociation for peptide modification analysis. **Nat Methods,** v. 4, n. 9, p. 709-12, Sep 2007.

ONG, S. E.; FOSTER, L. J.; MANN, M. Mass spectrometric-based approaches in quantitative proteomics. **Methods,** v. 29, n. 2, p. 124-30, Feb 2003.

ONG, S. E.; MANN, M. Mass spectrometry-based proteomics turns quantitative. **Nat Chem Biol,** v. 1, n. 5, p. 252-62, Oct 2005.

ONG, S. E.; MANN, M.. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). **Nat Protoc,** v. 1, n. 6, p. 2650-60, 2006.

PERKINS, D. N. et al. Probability-based protein identification by searching sequence databases using mass spectrometry data. **Electrophoresis,** v. 20, n. 18, p. 3551-67, Dec 1999.

PETERSON, A. C. et al. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. **Mol Cell Proteomics,** v. 11, n. 11, p. 1475-88, Nov 2012.

PICOTTI, P.; AEBERSOLD, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. **Nat Methods,** v. 9, n. 6, p. 555-66, Jun 2012.

PICOTTI, P.; BODENMILLER, B.; AEBERSOLD, R. Proteomics meets the scientific method. **Nat Methods,** v. 10, n. 1, p. 24-7, Jan 2013.

PICOTTI, P. et al. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. **Cell,** v. 138, n. 4, p. 795-806, Aug 21 2009.

PIETRAS, K.; OSTMAN, A. Hallmarks of cancer: interactions with the tumor stroma. **Exp Cell Res,** v. 316, n. 8, p. 1324-31, May 1 2010.

PRUDOVA, A. et al. Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. **Mol Cell Proteomics,** v. 9, n. 5, p. 894-911, May 2010.

PRUESSMEYER, J.; LUDWIG, A. The good, the bad and the ugly substrates for ADAM10 and ADAM17 in brain pathology, inflammation and cancer. **Semin Cell Dev Biol,** v. 20, n. 2, p. 164-74, Apr 2009.

PRUESSMEYER, J. et al. A disintegrin and metalloproteinase 17 (ADAM17) mediates inflammation-induced shedding of syndecan-1 and -4 by lung epithelial cells. **J Biol Chem,** v. 285, n. 1, p. 555-64, Jan 1 2010.

QIAO, D. et al. Heparan sulfate proteoglycans as regulators of fibroblast growth factor-2 signaling in brain endothelial cells. Specific role for glypican-1 in glioma angiogenesis. **J Biol Chem,** v. 278, n. 18, p. 16045-53, May 2 2003.

QUAIL, D. F.; JOYCE, J. A. Microenvironmental regulation of tumor progression and metastasis. **Nat Med,** v. 19, n. 11, p. 1423-37, Nov 2013.

RIFAI, N.; GILLETTE, M. A.; CARR, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. **Nat Biotechnol,** v. 24, n. 8, p. 971-83, Aug 2006.

ROEPSTORFF, P.; FOHLMAN, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. **Biomed Mass Spectrom,** v. 11, n. 11, p. 601, Nov 1984.

ROSS, P. L. et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. **Mol Cell Proteomics,** v. 3, n. 12, p. 1154-69, Dec 2004.

SADYGOV, R. G.; COCIORVA, D.; YATES, J. R., 3RD. Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. **Nat Methods,** v. 1, n. 3, p. 195-202, Dec 2004.

SAHIN, U. et al. Distinct roles for ADAM10 and ADAM17 in ectodomain shedding of six EGFR ligands. **J Cell Biol,** v. 164, n. 5, p. 769-79, Mar 1 2004.

SAVORE, C. et al. Perlecan knockdown in metastatic prostate cancer cells reduces heparin-binding growth factor responses in vitro and tumor growth in vivo. **Clin Exp Metastasis,** v. 22, n. 5, p. 377-90, 2005.

SCHELLER, J. et al. ADAM17: a molecular switch to control inflammation and tissue regeneration. **Trends Immunol,** v. 32, n. 8, p. 380-7, Aug 2011.

SCHIESS, R.; WOLLSCHEID, B.; AEBERSOLD, R. Targeted proteomic strategy for clinical biomarker discovery. **Mol Oncol,** v. 3, n. 1, p. 33-44, Feb 2009.

SCHILLING, O.; OVERALL, C. M. Proteomic discovery of protease substrates. **Curr Opin Chem Biol,** v. 11, n. 1, p. 36-45, Feb 2007.

SCHILLING, O.; OVERALL, C. M. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. **Nat Biotechnol,** v. 26, n. 6, p. 685-94, Jun 2008.

SCHULZE, W. X.; USADEL, B. Quantitation in mass-spectrometry-based proteomics. **Annu Rev Plant Biol,** v. 61, p. 491-516, 2010.

SHAH, F. D. et al. A review on salivary genomics and proteomics biomarkers in oral cancer. **Indian J Clin Biochem,** v. 26, n. 4, p. 326-34, Oct 2011.

SHAH, J. P.; GIL, Z. Current concepts in management of oral cancer--surgery. **Oral Oncol,** v. 45, n. 4-5, p. 394-401, Apr-May 2009.

SHAO, C.; ZHANG, Y.; SUN, W. Statistical characterization of HCD fragmentation patterns of tryptic peptides on an LTQ Orbitrap Velos mass spectrometer. **J Proteomics,** v. 109C, p. 26-37, Jun 27 2014.

SIMABUCO, F. M. et al. ADAM17 mediates OSCC development in an orthotopic murine model. **Mol Cancer,** v. 13, p. 24, 2014.

SOMORACZ, A. et al. Agrin immunohistochemistry facilitates the determination of primary versus metastatic origin of liver carcinomas. **Hum Pathol,** v. 41, n. 9, p. 1310-9, Sep 2010.

STARR, A. E. et al. Biochemical characterization and N-terminomics analysis of leukolysin, the membrane-type 6 matrix metalloprotease (MMP25): chemokine and vimentin cleavages enhance cell migration and macrophage phagocytic activities. **J Biol Chem,** v. 287, n. 16, p. 13382-95, Apr 13 2012.

STASTNA, M.; VAN EYK, J. E. Secreted proteins as a fundamental source for biomarker discovery. **Proteomics,** v. 12, n. 4-5, p. 722-35, Feb 2012.

STEEN, H.; MANN, M. The ABC's (and XYZ's) of peptide sequencing. **Nat Rev Mol Cell Biol,** v. 5, n. 9, p. 699-711, Sep 2004.

STERGACHIS, A. B. et al. Rapid empirical discovery of optimal peptides for targeted proteomics. **Nat Methods,** v. 8, n. 12, p. 1041-3, Dec 2011.

STERNLICHT, M. D.; WERB, Z. How matrix metalloproteinases regulate cell behavior. **Annu Rev Cell Dev Biol,** v. 17, p. 463-516, 2001.

STOKES, A. et al. Expression profiles and clinical correlations of degradome components in the tumor microenvironment of head and neck squamous cell carcinoma. **Clin Cancer Res,** v. 16, n. 7, p. 2022-35, Apr 1 2010.

STURGEON, C. et al. Taking a new biomarker into routine use--a perspective from the routine clinical biochemistry laboratory. **Proteomics Clin Appl,** v. 4, n. 12, p. 892-903, Dec 2010.

SYKA, J. E. et al. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. **Proc Natl Acad Sci U S A,** v. 101, n. 26, p. 9528-33, Jun 29 2004.

TAM, E. M. et al. Membrane protease proteomics: Isotope-coded affinity tag MS identification of undescribed MT1-matrix metalloproteinase substrates. **Proc Natl Acad Sci U S A,** v. 101, n. 18, p. 6917-22, May 4 2004.

TATRAI, P. et al. Agrin, a novel basement membrane component in human and rat liver, accumulates in cirrhosis and hepatocellular carcinoma. **Lab Invest,** v. 86, n. 11, p. 1149-60, Nov 2006.

THEOCHARIS, A. D. et al. Proteoglycans in health and disease: novel roles for proteoglycans in malignancy and their pharmacological targeting. **FEBS J,** v. 277, n. 19, p. 3904-23, Oct 2010.

TJALSMA, H. et al. Signal peptide-dependent protein transport in Bacillus subtilis: a genome-based survey of the secretome. **Microbiol Mol Biol Rev,** v. 64, n. 3, p. 515-47, Sep 2000.

TUCHER, J. et al. LC-MS based cleavage site profiling of the proteases ADAM10 and ADAM17 using proteome-derived peptide libraries. **J Proteome Res,** v. 13, n. 4, p. 2205-14, Apr 4 2014.

UNLU, M.; MORGAN, M. E.; MINDEN, J. S. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. **Electrophoresis,** v. 18, n. 11, p. 2071-7, Oct 1997.

VAN DEN BERG, B. H.; THOLEY, A. Mass spectrometry-based proteomics strategies for protease cleavage site identification. **Proteomics,** v. 12, n. 4-5, p. 516-29, Feb 2012.

VAN DER POST, S.; HANSSON, G. C. Membrane protein profiling of human colon reveals distinct regional differences. **Mol Cell Proteomics,** v. 13, n. 9, p. 2277-87, Sep 2014.

VESSECCHI, R. L., N. P.; GOZZO, F. C.; DORR, F. A.; MURGU, M. Nomenclaturas de espectrometria de massas em língua portuguesa. **Química Nova,** v. 34, n. 10, p. 1875-1887, 2011.

WALTHER, T. C.; MANN, M. Mass spectrometry-based proteomics in cell biology. **J Cell Biol,** v. 190, n. 4, p. 491-500, Aug 23 2010.

WANG, B. et al. The recurrence and survival of oral squamous cell carcinoma: a report of 275 cases. **Chin J Cancer,** v. 32, n. 11, p. 614-8, Nov 2013.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nat Rev Genet,** v. 10, n. 1, p. 57-63, Jan 2009.

WASHBURN, M. P. Utilisation of proteomics datasets generated via multidimensional protein identification technology (MudPIT). **Brief Funct Genomic Proteomic,** v. 3, n. 3, p. 280-6, Nov 2004.

WASINGER, V. C. et al. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. **Electrophoresis,** v. 16, n. 7, p. 1090-4, Jul 1995.

WEINBERG, R. A. **A biologia do câncer**. Artmed, 2008.

WELINDER, C. et al. Feasibility study on measuring selected proteins in malignant melanoma tissue by SRM quantification. **J Proteome Res,** v. 13, n. 3, p. 1315-26, Mar 7 2014.

WHITEAKER, J. R. et al. A targeted proteomics-based pipeline for verification of biomarkers in plasma. **Nat Biotechnol,** v. 29, n. 7, p. 625-34, Jul 2011.

WHITESIDE, T. L. The tumor microenvironment and its role in promoting tumor growth. **Oncogene,** v. 27, n. 45, p. 5904-12, Oct 6 2008.

WOLFSBERG, T. G. et al. ADAM, a widely distributed and developmentally regulated gene family encoding membrane proteins with a disintegrin and metalloprotease domain. **Dev Biol,** v. 169, n. 1, p. 378-83, May 1995.

XIAO, H. et al. Proteomic analysis of human saliva from lung cancer patients using two-dimensional difference gel electrophoresis and mass spectrometry. **Mol Cell Proteomics,** v. 11, n. 2, p. M111 012112, Feb 2012.

XUE, H.; LU, B.; LAI, M. The cancer secretome: a reservoir of biomarkers. **J Transl Med,** v. 6, p. 52, 2008.

YAKOB, M. et al. Salivary biomarkers for detection of oral squamous cell carcinoma - current state and recent advances. **Curr Oral Health Rep,** v. 1, n. 2, p. 133-141, Jun 1 2014.

YAN, W.; CHEN, S. S. Mass spectrometry-based quantitative proteomic profiling. **Brief Funct Genomic Proteomic,** v. 4, n. 1, p. 27-38, May 2005.

YATES, J. R., 3RD. Mass spectrometry and the age of the proteome. **J Mass Spectrom,** v. 33, n. 1, p. 1-19, Jan 1998.

YATES, J. R., 3RD. The revolution and evolution of shotgun proteomics for large-scale proteome analysis. **J Am Chem Soc,** v. 135, n. 5, p. 1629-40, Feb 6 2013.

ZHENG, X. et al. ADAM17 promotes glioma cell malignant phenotype. **Mol Carcinog,** v. 51, n. 2, p. 150-64, Feb 2012.

ZHENG, X. et al. ADAM17 promotes breast cancer cell malignant phenotype through EGFR-PI3K-AKT activation. **Cancer Biol Ther,** v. 8, n. 11, p. 1045-54, Jun 2009.

ZINI, A.; CZERNINSKI, R.; SGAN-COHEN, H. D. Oral cancer over four decades: epidemiology, trends, histology, and survival by anatomical sites. **J Oral Pathol Med,** v. 39, n. 4, p. 299-305, Apr 2010.

ZONG, F. et al. Specific syndecan-1 domains regulate mesenchymal tumor cell adhesion, motility and migration. **PLoS One,** v. 6, n. 6, p. e14816, 2011.

ZUBAREV, R. A. K., N.L.; MCLAFFERTY, F. W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. **J. Am. Chem. Soc,** v. 120, p. 3265–3266, 1998.

ANEXOS

DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha dissertação de Mestrado/tese de Doutorado intitulada PROTEÔMICA BASEADA EM ESPECTROMETRIA DE MASSAS NA DESCOBERTA DE CANDIDATOS À BIOMARCADORES EM CÂNCER E NA IDENTIFICAÇÃO DE NOVOS ALVOS DE ADAM17

( ) não se enquadra no § 4º do Artigo 1º da Informação CCPG 002/13, referente a bioética e biossegurança.

Tem autorização da(s) seguinte(s) Comissão(ões):

(x) CIBio – Comissão Interna de Biossegurança , projeto No. AFPL 2.2, Instituição: ABTLUS – Associação Brasileira de Tecnologia de Luz Síncrotron.

( ) CEUA – Comissão de Ética no Uso de Animais , projeto No. _____, Instituição: _____.

( ) CEP - Comissão de Ética em Pesquisa, protocolo No. _____, Instituição: _____

* *Caso a Comissão seja externa ao IB/UNICAMP, anexar o comprovante de autorização dada ao trabalho. Se a autorização não tiver sido dada diretamente ao trabalho de tese ou dissertação, deverá ser anexado também um comprovante do vínculo do trabalho do aluno com o que constar no documento de autorização apresentado.*

Rebeca Kawahara
Aluno: (nome completo)

Adriana Franco Paschoal
Orientador: (nome completo)

Para uso da Comissão ou Comitê pertinente:
(X) Deferido ( ) Indeferido

Carimbo e assinatura

Profa. Dra. HELOÍSA COUTINHO FRANCO DE OLIVEIRA
Presidente
CIBio/IB-UNICAMP

Para uso da Comissão ou Comitê pertinente:
( ) Deferido ( ) Indeferido

Carimbo e assinatura

Uso exclusivo da CIBio:

Número de projeto / processo: AFPL 2.2

Formulário de encaminhamento de projetos de pesquisa para análise pela CIBio - Comissão Interna de Biossegurança da ABTLuS – Associação Brasileira de Tecnologia de Luz Síncrotron

Título do projeto: Determinação dos sítios de clivagens de alvos de ADAM17 recombinante em cultura de células humanas e estudo de vias de sinalização de ADAM17 utilizando proteômica quantitativa.
Pesquisador responsável: Dra. Adriana Franco Paes Leme
Experimentador: Rebeca Kawahara

Nível do treinamento do experimentador: [ ]-Iniciação científica, [ ]-mestrado, [ ]-doutorado, [x]-doutorado direto, [ ]-pós-doutorado, [ ]-nível técnico, [ ]-outro, especifique: _____

Resumo do projeto:

Os objetivos desse estudo serão buscar novos alvos para ADAM-17 e determinar os sítios de clivagens específicos dessa proteinase utilizando abordagens de biologia molecular e proteômica/espectrometria de massas. Para isso, as seqüências de cDNA correspondentes aos domínios metaloproteinase, tipo-disintegrina e rico em cisteínas da ADAM-17 e tipo epidermal serão amplificadas pela reação em cadeia com polimerase e subclonadas nos vetores de expressão. Após expressão e purificação, a proteinase recombinante será incubada com cultura de células humanas não tumorigênicas (controle normal) e tumorigênicas, respectivamente HaCaT e SCC-9. As proteínas clivadas ou liberadas da superfície das células HaCaT e SCC-9 serão digeridas com tripsina e os sítios de clivagens serão identificados por espectrometria de massas. As estratégias utilizadas nesse estudo podem identificar novos alvos celulares e de matriz extracelular da ADAM-17 em células tumorigênicas e não tumorigênicas e elucidar os sítios de preferência de uma proteinase intrínsecos ao seu mecanismo de ação, como também é uma das estratégias para o desenho de substratos peptídicos e inibidores. Também será analisada a expressão de proteínas de membrana e fosforilação de seus domínios intracelulares em condição de super-expressão de ADAMs utilizando estratégia de proteômica quantitativa por marcação metabólica (SILAC: *stable isotope labeling with amino acid in cell culture*). A caracterização da expressão de proteínas de membrana, bem como a fosforilação de seus domínios intracelulares poderão elucidar eventos de sinalização e também de interações entre as células e o microambiente do tumor.

A CIBio analisou este projeto em reunião realizada no dia: 11.1.12
Parecer final: [X]-projeto aprovado, [ ]-projeto recusado, [ ]-projeto com deficiências, favor comentários anexo.

Presidente da CIBio – ABTLuS-LNBio
Jörg Kobarg

Membro da CIBio da ABTLuS-CTBE
Fabio Squina

Membro da CIBio – ABTLuS-LNBio
Prof. Dr. Celso Eduardo Benedetti

Membro da CIBio da ABTLuS-CTBE
Sindelia Freitas Azzoni

Membro da CIBio – ABTLuS-LNBio
Andrea Balan

Membro da CIBio da ABTLuS-CTBE
Roberto Ruller

Membro da CIBio – ABTLuS-SGT
Carolina Barbosa Marini

Campinas, 20 de Março de 2015

Para Profa. Dra. Helena Coutinho Franco de Oliveira
Presidente da CIBio/IB-UNICAMP

Assunto aprovação de projeto na CIBio – CNPEM

Prezada Profa. Dra. Helena Coutinho Franco de Oliveira

Declaramos para os devidos fins que Rebeca Kawahara, teve seu projeto analisado e aprovado pela CIBio do CNPEM em 11 de Janeiro de 2012, protocolo AFPL 2.2. O Titulo originalmente submetido foi "Determinação dos sitios de clivagens de alvos de ADAM 17 recombinante em cultura de células humanas e estudo de vias de sinalização de ADAM 17 utilizando proteômica quantitativa", sob orientação da pesquisadora Adriana Franco Paes Leme, PhD. Este projeto rendeu resultados que foram utilizados para o desenvolvimento da tese, intitulada "PROTEÔMICA BASEADA EM ESPECTROMETRIA DE MASSAS NA DESCOBERTA DE CANDIDATOS À BIOMARCADORES EM CÂNCER E NA IDENTIFICAÇÃO DE NOVOS ALVOS DE ADAM17". Outrossim, informamos que os trabalhos contidos na sua tese estão contemplados no projeto aprovado pela CIBio.

Atenciosamente,

Marcio Chaim Bajgelman, PhD
Presidente da CIBIO/ LNBio/CNPEM

Adriana Franco Paes Leme, PhD
Pesquisadora/orientadora – LNBio/CNPEM

Rebeca Kawahara
Doutoranda do programa de pós graduação em
Biologia Funcional e Molecular/IB/Unicamp

O LNBio integra o CNPEM, nova denominação da ABTLuS, Organização Social qualificada pelo
Ministério da Ciência, Tecnologia e Inovação (MCTI)
Campus: Rua Giuseppe Máximo Scolfaro, 10.000 - Polo II de Alta Tecnologia - Caixa Postal 6192 - 13083-970 - Campinas/SP
Fone: +55.19.3512.1010 | Fax: +55.19.3512.1006 | www.lnbio.org.br