



Faculdade de Medicina de São José do Rio Preto
Programa de Pós-graduação em Ciências da
Saúde

Tiago Henrique

Abordagem Computacional para Identificação de
Marcadores Moleculares e de seus Ligantes com
Potencial Aplicação no Tratamento do Carcinoma
Epidermoide de Cabeça e Pescoço

São José do Rio Preto
2016

Tiago Henrique

Abordagem computacional para identificação de
marcadores moleculares e de seus ligantes com
potencial aplicação no tratamento do carcinoma
epidermoide de cabeça e pescoço

Tese apresentada à Faculdade de
Medicina de São José do Rio Preto
para obtenção do Título de Doutor
no Curso de Pós-graduação em
Ciências da Saúde, Eixo Temático:
Ciências Correlatas

Orientadora: Profa. Dra. Eloiza Helena Tajara

São José do Rio Preto
2016

Henrique, Tiago

Abordagem computacional para identificação de marcadores moleculares e de seus ligantes com potencial aplicação no tratamento do carcinoma epidermoide de cabeça e pescoço / Tiago Henrique
São José do Rio Preto, 2016
80p.

Tese (Doutorado) - Faculdade de Medicina de São José do Rio Preto - FAMERP
Eixo Temático: Ciências Correlatas

Orientadora: Profa. Dra. Eloiza Helena Tajara

1. Câncer de cabeça e pescoço, 3. Revisão sistemática da literatura, 3. Marcadores Moleculares, 4. Bioinformática, 5. Desenvolvimento de drogas.

Tiago Henrique

Abordagem computacional para identificação de
marcadores moleculares e de seus ligantes com
potencial aplicação no tratamento do carcinoma
epidermoide de cabeça e pescoço

BANCA EXAMINADORA
TESE PARA OBTENÇÃO DO GRAU DE
DOUTOR

Presidente e Orientador: Profa. Dra. Eloiza Helena Tajara

2 Examinador: Nelson José Freitas da Silveira

3 Examinador: Marinonio Cornelio

4 Examinador: Erika Cristina Pavarino

5 Examinador: Mauricio Lacerda Nogueira

Suplentes: Sidney Pinheiro Junio

Lilian Castiglioni

São José do Rio Preto, 16/05/2016

Sumário

1 INTRODUÇÃO.....	1
1.1. CARCINOMA EPIDERMÓIDE DE CABEÇA E PESCOÇO: FATORES DE RISCO, PROGNÓSTICO E TRATAMENTO	1
1.2. INFLAMAÇÃO ASSOCIADA A TUMOR.....	3
1.3. ALVOS MOLECULARES.....	5
1.4. MINERAÇÃO DA LITERATURA PARA DETECÇÃO DE ALVOS MOLECULARES... 	9
2 OBJETIVOS.....	12
3 MATERIAIS E MÉTODOS	13
3.1. IDENTIFICAÇÃO E SELEÇÃO DE MARCADORES	13
3.1.1. OBTENÇÃO DOS DADOS E MINERAÇÃO DA LITERATURA.....	13
3.1.2. BANCO DE DADOS E INTERFACE WEB	17
3.1.3. ONTOLOGIAS E ANÁLISE DE VIAS METABÓLICAS	18
3.1.4. ESTRUTURA TRIDIMENSIONAL	19
3.1.5. SELEÇÃO DE LIGANTES PARA PROTEÍNAS DE INTERESSE	19
3.1.6. DOCKING MOLECULAR.....	22
3.1.7. CULTURA DE CÉLULAS.....	22
3.1.8. TRATAMENTO COM PIPLARTINA.....	24
3.1.9. ENSAIO DE VIABILIDADE CELULAR	24
3.1.10. ENSAIO DE MIGRAÇÃO	25
3.1.11. EXTRAÇÃO DE RNA DE CULTURA CELULAR.....	26
3.1.12. SÍNTESE DE DNA COMPLEMENTAR (CDNA).....	27
3.1.13. PCR EM TEMPO REAL	27

4 RESULTADOS	30
4.1. IDENTIFICAÇÃO E SELEÇÃO DOS MARCADORES.....	30
4.2 DOCKING MOLECULAR.....	33
4.3. ENSAIOS FUNCIONAIS.....	35
4.3.1. VIABILIDADE CELULAR	35
4.3.2. MIGRAÇÃO E INVASÃO	35
4.4.RNA DA LINHAGEM CELULAR.....	39
4.5 EXPRESSÃO GÊNICA	39
5 DISCUSSÃO.....	42
6. CONCLUSÕES	53
8. REFERÊNCIAS.....	54
9. TRABALHO PUBLICADO.....	63
10. OUTROS TRABALHOS PUBLICADOS	64

Dedico este trabalho

Aos meus pais, **Pedro** (*in memoriam*) e **Irma**,
À minha amada esposa, **Letícia**
com muito amor e carinho
Ao meu irmão, **Mateus**
E aos meus sobrinhos **Miguel e Gabriel**

Agradecimentos

A realização do presente trabalho somente foi possível devido à contribuição de algumas pessoas. A todos eles, meus mais sinceros agradecimentos.

À minha orientadora Profa. Dra. Eloiza Helena Tajara da Silva, por ter confiado em meu trabalho. Por estar sempre tão preocupada com a qualidade do trabalho que desenvolvemos e com nossa formação científica, por saber sempre posso contar e por ser um exemplo de pessoa.

Aos amigos do Laboratório de Marcadores Moleculares e Bioinformática Médica, da Faculdade de Medicina de São José do Rio Preto - FAMERP, pelo apoio e convivências sempre tão agradável. Agradeço à Bianca, Carol, Giovana, Ariane e Mayra.

À minha família pelo amor e compreensão. Aos meus pais Pedro (*in memoriam*) e Irma, que investiram muito em minha formação pessoal e profissional, muitas vezes deixando de realizar seus sonhos para realizar os meus. Obrigado. Tudo que sou hoje devo a vocês.

À Letícia, minha esposa e pessoa mais doce e carinhosa que conheço. Te amo !

Ao grupo GENCAPO

A FAPESP e CNPq pelo auxílio financeiro sem o qual esse trabalho não seria realizado.

A Deus, por tudo !

Muito obrigado, a todos !

Epígrafe

**"Por vezes sentimos que aquilo que fazemos não é senão uma gota de
água no mar. Mas o mar seria menor se lhe faltasse uma gota"**

MadreTeresa de Calcuta

Lista de Figuras

Figura 1. Fruto da pimenta <i>Piper longum</i> e estrutura química da piplartina	21
Figura 2. Redocking do complexo MAPK1 e seu inibidor	34
Figura 3. Tempos e concentrações escolhidos para o teste de citotoxicidade da piplartina	36
Figura 4. Fotomicrografias ensaios de migração	37
Figura 5. Fotomicrografias ensaios de invasão	38
Figura 6. Expressão gênica após tratamento com a piplartina	41
Figura 7. Redes regulatórias dos 1370 genes relacionados a CECF	45

Lista de Quadros e Tabelas

Tabela 1. Termos MeSH utilizados para a revisão sistemática da literatura	14
Tabela 2. Genes relacionados a processos inflamatórios	29
Tabela 3. Características físico químicas das drogas utilizadas para o <i>Docking</i> molecular	34
Tabela 4. Quantificação e análise do RNA extraído	40
Tabela 6. Genes diferencialmente expressos após o tratamento com a piplartina	40

Lista de Símbolos e Abreviaturas

CECP – Carcinoma epidermoide de cabeça e pescoço

AINES – Anti-inflamatórios não esteroides

CD20 – *Cluster differentiation 20*

cDNA – DNA complementar

PCR – Polimerase Chain Reaction

Da - Daltons

MeSH -

ln – Logaritmo natural

HTML - HyperText Markup Language

PHP - Hypertext Preprocessor

mM - Milimol

PBS - Phosphate Buffered Saline

DMEM - Dulbecco's Modified Eagle Medium

EDTA - Ethylenediamine tetraacetic acid

SFB – Soro Fetal Bovino

DMSO - Dimethyl sulfoxide

uL - Microlitro

CO₂ – Dióxido de Carbono

g – Força g

Pb – Pares de bases

Ct - Cycle threshold

RPM – Rotações por minuto

RESUMO

Introdução: A literatura científica sobre câncer tem crescido rapidamente nos últimos anos, o que torna difícil, se não impossível, a tarefa de recuperar e analisar manualmente as informações relevantes sobre os mecanismos que governam o processo neoplásico. Além disso, o câncer é uma doença complexa e sua heterogeneidade é particularmente evidente no câncer epidermoide de cabeça e pescoço (CECP), um dos tipos mais comuns de câncer em todo o mundo. **Objetivos:** Os objetivos do estudo foram (a) identificar genes/proteínas relacionados a CECP a partir de dados da literatura, (b) identificar ligantes que interajam eficiente e especificamente com alvos moleculares selecionados, (c) avaliar computacionalmente o complexo proteína/ligante e (d) avaliar a ação de ligantes na expressão gênica e no comportamento de células de carcinoma. **Métodos:** A busca de marcadores potenciais relacionados a CECP utilizou a mineração da literatura disponível publicamente, seguindo um fluxograma que incluiu seleção de artigos científicos no PubMed por termos MeSH, associação de artigos com genes/proteínas por meio do arquivo gene2pubmed, seleção de genes em bancos de dados externos e etapas de curação manual. As proteínas identificadas como sendo relacionadas a CECP que apresentaram envolvimento em processos inflamatórios foram submetidas a experimento de *docking* molecular para identificação de seus ligantes entre drogas disponíveis no mercado. Finalmente, o papel da piplartina, uma substância natural extraída da pimenta *Piper longum* com evidências de ação anti-inflamatória e antineoplásica, foi avaliado na proliferação, na migração e na expressão gênica de células neoplásicas. **Resultados:** Um total de 1370 genes relacionados a CECP foi identificado pela abordagem proposta, que mostrou especificidade de 74% e

sensibilidade de 87%. A diversidade dos dados permitiu obter associações potenciais ainda não exploradas, revelando, por exemplo, que a resposta ao estímulo esteróide hormonal está significativamente enriquecida com genes relacionados a CECP. Os resultados também mostraram que a piplartina reduz a viabilidade e a migração celular e modifica o padrão de expressão de um painel de genes que atuam em processos inflamatórios. **Conclusão:** A abordagem empregada permite a identificação de genes relacionados a CECP e revela novas associações que merecem ser estudadas. O composto piplartina selecionado para estudos *in vitro* interage com alvos moleculares de forma semelhante à de medicamentos anti-inflamatórios conhecidos e é capaz de diminuir a proliferação e a migração celular e de alterar a expressão de genes relacionados à CECP e a processos inflamatórios.

Palavras Chave: Câncer de cabeça e pescoço, Revisão sistemática da literatura, Marcadores Moleculares, Bioinformática, Desenvolvimento de drogas

ABSTRACT

Introduction: The total amount of scientific literature on cancer has grown rapidly in recent years. This makes it difficult, if not impossible, to manually retrieve relevant information on the mechanisms that govern the neoplastic process. Furthermore, cancer is a complex disease, and its Heterogeneity is particularly evident in head and neck squamous cell carcinoma (HNSCC); one of the most common types of cancer worldwide. **Objectives:** The present study aimed: a) to identify genes/proteins related to HNSCC; b) to identify ligands that specifically target molecular biomarkers of interest; c) to evaluate computationally protein-ligand complexes and d) to evaluate the effect of ligands on gene expression and on carcinoma cell behavior. **Methods:** The search for potential markers related to HNSCC was performed by literature mining, following a flow chart that included selection of scientific articles in PubMed by MeSH terms, association of articles with genes/proteins through the gene2pubmed file, selection of genes in external data bases and manual curation steps. In order to identify potential ligands, proteins related to HNSCC and involved in inflammatory processes were used to perform molecular docking assays with known anti-inflammatory drugs. Finally, the role of piperlongumine, a substance extracted from the *Piper longum* with anti-inflammatory and antineoplastic effects, on proliferation, migration and gene expression was investigated in neoplastic cells. **Results:** The curated gene-to-publication assignment yielded a total of 1,370 genes related to HNSCC, with specificity of 74% and sensibility of 87%. The diversity of results allowed identifying new and mostly unexplored gene associations, revealing, for example, that processes linked to response to steroid hormone stimulus are significantly enriched with genes related to HNSCC. The results

also showed that piplartine decreases viability and cell migration, and alters expression of genes involved in inflammatory responses. **Conclusion:** This approach allows the identification of genes related to HNSCC and is able to reveal new associations that deserve to be further studied. Piplartine, the compound selected for *in vitro* studies, interacts with molecular targets similar to known anti-inflammatory drugs, decreases proliferation and cell migration, and alter the expression of genes associated with HNSCC and inflammatory processes.

Key words: Head and Neck Cancer, Systematic review, Molecular Markers, Bioinformatic, Drug design

1 INTRODUÇÃO

1.1. CARCINOMA EPIDERMÓIDE DE CABEÇA E PESCOÇO: FATORES DE RISCO, PROGNÓSTICO E TRATAMENTO

O câncer de cabeça e pescoço compreende um grupo de neoplasias epiteliais que surgem nas cavidades nasal e oral, na laringe e na faringe, e a maioria é do tipo histológico epidermoide (1). É o sexto tipo mais frequente de câncer no mundo, sendo responsável por 6% de todos os tumores. São estimados aproximadamente 650.000 novos casos e 350.000 mortes por ano decorrentes destas neoplasias (2). No Brasil, segundo o Instituto Nacional do Câncer (INCA), foram esperados para o ano de 2014 cerca de 15 mil novos casos de câncer oral e 7 mil de laringe (3).

Entre os principais fatores de risco para o desenvolvimento do carcinoma epidermoide de cabeça e pescoço (CECP) está o consumo de tabaco e álcool. Nos países ocidentais, a exposição ao tabaco ocorre principalmente pelo fumo de cigarro, charuto e cachimbo. No sudeste da Ásia (incluindo a Índia), a população usa produtos derivados do tabaco, tais como *macheri*, *snuff*, *khaini*, *betel quid*, *gutkha* e *panmasala* (4). Existem mais de 60 carcinógenos conhecidos na fumaça do cigarro e pelo menos 16 em tabaco não queimado, entre eles as nitrosaminas e os hidrocarbonetos aromáticos policíclicos (5). Esses carcinógenos são conhecidos por causarem transversoão do nucleotídeo guanina em genes importantes envolvidos no desenvolvimento do câncer de cabeça e pescoço.

O risco conferido pelo álcool também é significativo e o consumo de 50 gramas por dia representa um aumento de aproximadamente três vezes o risco de câncer de

cavidade oral e faringe e cerca de duas vezes o risco de câncer de laringe. Juntos, tabaco e álcool são responsáveis por aproximadamente 51% dos tumores de cabeça e pescoço nos Estados Unidos, 84% na Europa e 83% na América Latina (6).

Outro fator etiológico importante para o câncer de faringe, e em escala menor para o câncer oral, é a infecção pelo papilomavírus humano (HPV). Os mecanismos moleculares da tumorigênese associada ao HPV é mediada principalmente por duas oncoproteínas E6 e E7, que inativam as proteínas supressoras de tumor p53 e pRb, respectivamente, ambas envolvidas no controle do ciclo celular. A proteína E6 liga-se à p53 e à proteína pró-apoptótica BAK levando à sua ubiquitinação e degradação. Da mesma maneira, a E7 liga-se à pRb, impedindo que ela forme um complexo com o fator de transcrição E2F. Como consequência, este fator fica disponível para promover a transcrição de genes relacionados com a progressão do ciclo celular ou seus inibidores, como o *CDKN2A* (7), um gene supressor de tumor cujo produto (p16-INK4) tem sido apontado como marcador de positividade para HPV (revisto por (8) e de prognóstico nos casos HPV positivos em câncer de orofaringe. Em células normais, a proteína p16-INK4 inibe a interação entre ciclina D e a quinase CDK4 e, portanto, a progressão do ciclo celular (9).

Considerando a infecção por HPV e os outros fatores de risco de CECP, é possível subdividir CECP em três grupos. Dois desses grupos envolvem os casos HPV positivos (principalmente de orofaringe com um prognóstico favorável) e os casos HPV negativos (com prognóstico menos favorável e um perfil molecular diferente (10). Os tumores HPV positivos possuem primariamente o gene *TP53* selvagem, enquanto os HPV negativos apresentam mutações neste gene e mostram instabilidade cromossômica elevada (10, 11), o que pode sustentar padrões metabólicos vantajosos, auxiliar no

escape a fatores inibitórios de sinais supressores (13) ou promover efeitos oncogênicos (14). Uma terceira classe de CECP compreende tumores nasofaríngeos nos quais fatores etiológicos distintos são conhecidos, incluindo a infecção pelo vírus Epstein-Barr (EBV) (15).

Atualmente, o prognóstico para pacientes com CECP tem sido baseado em variáveis clinicopatológicas, que incluem tamanho do tumor, presença de metástases em linfonodos e grau de diferenciação da lesão (16).

O tratamento de CECP envolve decisões complexas que dependem de uma equipe multidisciplinar (cirurgiões de cabeça e pescoço, cirurgiões plásticos, oncologistas, radiologistas e dentistas). Alguns fatores, como sítio primário, estadiamento e preservação do órgão e da deglutição, são levados em conta no processo de escolha da melhor conduta que, em geral, envolve cirurgia e radioterapia (1).

1.2. INFLAMAÇÃO ASSOCIADA A TUMOR

Mantovani e colaboradores (2008) demonstraram que a inflamação causa alterações no microambiente tumoral e promove o desenvolvimento e a progressão do câncer. Alguns tipos celulares e mediadores inflamatórios específicos podem influenciar migração e invasão celular e o desenvolvimento de metástases (17).

A relação entre inflamação e câncer não é recente. Em 1863, Virchow defendeu a hipótese de que o câncer se desenvolve preferencialmente próximo a locais de inflamação crônica com base no fato de que lesões teciduais e a consequente inflamação podem aumentar a proliferação celular. (18). Estudos recentes têm demonstrado que a inflamação relacionada ao câncer deriva da interação entre as células de defesa do

organismo e as células tumorais, resultando em alterações sistêmicas que podem levar à supressão imune e à progressão maligna. Citocinas e outras proteínas inflamatórias sintetizadas e liberadas por células estromais, endoteliais e do sistema imune ou pelas próprias células neoplásicas são as principais responsáveis por coordenar a interação intercelular no microambiente tumoral (19).

A exposição constante a citocinas e a espécies reativas de oxigênio bem como a ativação de fatores de transcrição, como o fator nuclear kappa B (NF-KB) e o transdutor de sinal de transcrição 3 (STAT3), são considerados mecanismos moleculares críticos para o desenvolvimento e a progressão do câncer na presença de processos inflamatórios crônicos. Acredita-se que esses mecanismos levem à instabilidade genética e subseqüentes mutações em oncogenes ou em genes supressores de tumores (20).

Os processos inflamatórios envolvem ativação de uma grande gama de mediadores e vias metabólicas, o que torna difícil o desenvolvimento de tratamentos para esses processos. Atualmente, o tratamento para doenças inflamatórias na medicina ocidental utiliza anti-inflamatórios não esteroidais (AINES) e corticoesteroides. Os AINES são amplamente empregados no tratamento de doenças inflamatórias por apresentarem um mecanismo de ação bastante rápido e altas taxa curativas. No entanto, esses medicamentos apresentam efeitos colaterais diversos incluindo úlceras gastrintestinais (21-23) e disfunção plaquetária (24). Considerando as limitações impostas pelos efeitos colaterais decorrentes do uso dos atuais anti-inflamatórios, torna-se de extrema importância a identificação e a validação de substâncias que exerçam sua função anti-inflamatória ocasionando menos efeitos colaterais.

1.3. ALVOS MOLECULARES

Com o enorme progresso biotecnológico dos últimos anos e, conseqüentemente, com a identificação de muitas alterações moleculares em câncer, novas terapias têm sido desenvolvidas que utilizam alvos com função ou expressão alterada nas células neoplásicas. Um dos principais desafios dessas terapias é minimizar os seus efeitos colaterais, focando apenas o tecido tumoral e diminuindo o dano em tecidos normais. Muitas drogas desenvolvidas a partir de dados moleculares já estão na prática clínica. É o caso do mesilato de imatinib (Glivec) para leucemia mielóide crônica, uma neoplasia hematopoética que apresenta translocação dos genes *BCR* e *ABL* resultando em elevada atividade de uma quinase serina/treonina quimérica (25). Outros exemplos incluem o rituximab para linfomas positivos para o antígeno CD20 (26) e o trastuzumab para tumores de mama que exibem amplificação do receptor de membrana tirosina/quinase HER2/neu (27).

No câncer de cabeça e pescoço, um dos alvos moleculares é o receptor do fator de crescimento epidérmico (codificado pelo gene *EGFR*) e sua via metabólica, que apresentam um padrão de expressão elevada e estão relacionados com o processo neoplásico em si e com a diminuição da sobrevida. Além da mudança de expressão de *EGFR*, outras alterações genéticas e epigenéticas relacionadas a desenvolvimento, sobrevida e agressividade de CECP compreendem alvos potenciais de agentes terapêuticos (28).

O estudo destas alterações mostra que elas já estão presentes em hiperplasias benignas, como a inativação do gene *CDKN2*. De acordo com o modelo de progressão molecular de CECP, a progressão clínica a displasia, carcinoma *in situ* e, finalmente a

carcinoma invasivo, é acompanhada pelo acúmulo de alterações moleculares (29). As mutações no gene *TP53*, a amplificação de *CCND1*, a ativação de *EGFR*/inativação de *PTEN* e a deleção de diferentes segmentos do genoma são alguns exemplos de alterações relacionadas com a progressão de CECP, como estabelecido por Leemans e colaboradores (10). Tais alterações promovem o fenótipo neoplásico definido por Hanahan e Weinberg (2011, (30)).

Várias técnicas recentes para estudo do câncer têm permitido a análise em larga escala das características das células malignas. Entre elas, estão os arranjos de cDNA (31) e os métodos de sequenciamento de última geração (32), que são muito auxiliados na etapa de validação pela técnica da reação em cadeia da polimerase (PCR) em tempo real (RT-PCR) (33). Tais abordagens têm tornado possível, de maneira robusta e reprodutível, a identificação de ganhos e perdas de sequências e de alterações na expressão gênica nos tumores. Embora importantes como ferramentas diagnósticas e prognósticas, os seus resultados não oferecem uma fonte direta para o desenvolvimento de terapias contra o câncer, já que não são os genes e os seus transcritos, mas sim os produtos proteicos os alvos habituais de novas drogas

Por esse motivo, para o entendimento dos mecanismos envolvidos na tumorigênese e seu controle, é importante que, paralelamente aos dados derivados do genoma, do transcritoma e dos dados clínicos, sejam também obtidas informações sobre as diferenças proteicas entre a célula normal e a maligna.

O proteoma é o conjunto de todas as proteínas expressas por um organismo e é muito maior, mais complexo e mais dinâmico que o genoma. Por exemplo, são estimados cerca de 20.000 genes no genoma humano enquanto alguns dados recentes sugerem que mais de um milhão de diferentes espécies de proteínas (34) estão presentes

nas células e nos tecidos em condições e/ou momentos distintos.

Esta complexidade possui várias fontes. Por exemplo, alguns genes, por meio de *splicing* alternativo, codificam diferentes isoformas. Modificações pós-traducionais, tais como fosforilação, glicosilação, clivagem proteolítica e adição de oligossacarídeos ou lipídeos, podem alterar a estrutura e a função de uma mesma proteína e determinar mudanças em sua localização intracelular (35). Aumentando a complexidade, estão as associações de proteínas com outras moléculas, grupos prostéticos e a formação de complexos multiméricos.

A interação proteína-proteína é um processo importante na transmissão de sinais que controlam os processos celulares. Quando essas interações não ocorrem corretamente ou as proteínas envolvidas estão alteradas, os sinais enviados podem resultar em características do fenótipo neoplásico. Tais características foram bem definidas por Hanahan e Weinberg (2000; 2011) (30, 36) e incluem proliferação celular aumentada, insensibilidade a fatores supressores de crescimento, resistência a apoptose, angiogênese sustentada, reprogramação do metabolismo energético, evasão ao ataque imune e capacidade para invasão e metástase. Se os sinais desencadeados por interações proteicas anômalas puderem ser bloqueados, será possível deter o processo tumoral.

Muitas drogas utilizadas no tratamento do câncer são pequenas moléculas que apresentam estrutura geométrica similar ao sítio de ligação das proteínas alvo, local da molécula no qual são estabelecidas interações que garantem a afinidade pelo alvo de forma eficiente e específica (37). Como referido acima, um dos maiores desafios encontrados para o desenvolvimento de novas drogas é a inibição de proteínas alvo sem interferir nas funções normais da célula. As ferramentas de Bioinformática estão auxiliando este processo e se tornando parte importante das pesquisas acadêmicas e

industriais (38).

Diversas metodologias computacionais podem ser empregadas nos vários estágios do descobrimento de novas drogas (39). Dentre elas, podem ser citadas: o *Screening* Virtual e o *Docking* Molecular.

O *Screening* ou Escaneamento Virtual é utilizado nas etapas iniciais do desenvolvimento de fármacos e tem como objetivo diminuir o tamanho das bibliotecas de ligantes. Uma pré-filtragem das bibliotecas é realizada considerando alguns fatores físico-químicos, eliminando assim componentes que possivelmente não são adequados (39). Após esta pré-filtragem, os ligantes selecionados são utilizados para a realização do *Docking* ou Docagem Molecular contra as proteínas alvo, o que torna possível a predição de sua interação com o alvo (40). O processo de *Docking* envolve duas etapas básicas, a predição da conformação do ligante bem como sua posição e orientação nos sítios de ligação da proteína alvo e a avaliação da afinidade de ligação. Entretanto, a ligação de pequenas moléculas às proteínas alvo não é simples e pode sofrer influência de alguns fatores entrópicos e entálpicos, como mobilidade, cargas elétricas e ligações com moléculas de água, o que pode afetar a docagem (39).

Apesar de diferentes níveis de dificuldades, o desenho de novas drogas pode ser visto como um campo promissor, principalmente graças à grande quantidade de dados genômicos e proteômicos essenciais para a identificação de novos marcadores moleculares alvo (39).

1.4. MINERAÇÃO DA LITERATURA PARA DETECÇÃO DE ALVOS MOLECULARES

As tecnologias de última geração (genômica, transcritômica, proteômica e metabolômica) e ferramentas computacionais avançadas têm levado a um melhor entendimento do processo neoplásico bem como da identificação de biomarcadores potenciais para diagnóstico e prognóstico do câncer. Tais tecnologias resultaram em um grande número de dados em uma escala sem precedentes. Entretanto, estes dados estão dispersos em vários bancos de dados (41-44). Esta estrutura descentralizada dificulta a elaboração de conclusões e a formulação de novas hipóteses.

O banco de dados público PubMed, desenvolvido e mantido pela *US National Library of Medicine*, é uma das mais importantes ferramentas de busca de informação científica. Atualmente, o PubMed possui mais de 3 milhões de citações em câncer. Portanto, fica extremamente difícil extrair todos os dados relevantes, mesmo após a subdivisão da área de interesse ou a utilização de perguntas específicas. Além disso, as buscas na literatura são dificultadas pela complexidade da doença. O câncer e seus subtipos são provavelmente um conjunto de doenças, cada uma com diferentes características e questões a serem respondidas. Neste aspecto, o câncer de cabeça e pescoço é um dos melhores exemplos de grande heterogeneidade. De forma similar ao que ocorre em muitos tipos de neoplasmas, o desafio na busca de literatura em câncer de cabeça e pescoço é particularmente difícil em função de sua diversidade, que envolve tipo histológico, localização anatômica e fatores de risco primários.

As investigações sobre CECp enfatizam a importância da identificação de mecanismos e de alterações moleculares que ocorrem durante a transformação maligna. Novos dados nesta área podem auxiliar no entendimento da biologia do tumor e,

consequentemente, levar ao desenvolvimento de novas drogas. A mineração da literatura é um ponto de início fundamental neste processo de descoberta, mas o crescimento exponencial dos dados biológicos está muito além do limite de uma busca manual completa na maioria dos casos. Por outro lado, a mineração automática da literatura pode auxiliar a identificar biomarcadores relacionados com a doença e suas inter-relações, e extrair informações não claramente expressas com ferramentas capazes de eficientemente focar questões importantes e gerar hipóteses testáveis. Durante este processo, os artigos científicos são extraídos, as entidades biológicas são identificadas nos textos, e a informação é obtida.

Um dos desafios em abordagens automáticas é a identificação exata de genes, proteínas ou doenças considerando que, na literatura, podem ser utilizados diferentes nomes e símbolos, sinônimos ou mesmo nomenclatura não oficial (45). Um outro desafio é identificar descrições consistentes de produtos gênicos e suas características. Para superar estas limitações, as abordagens de mineração de literatura têm incorporado ferramentas para reconhecer palavras-chaves específicas e para capturar sentenças relevantes e ontologias. Por exemplo, as relações entre entidades podem ser extraídas investigando sua co-ocorrência no mesmo texto, título ou sentença, ou pelo método denominado Processamento de Linguagem Natural (NLP). Os métodos NLP têm como base a estrutura de sentenças e o modo como os dados biológicos são mencionados (46). Entretanto, esta abordagem tem vantagens e limitações porque pode levar a relações incorretas dependendo dos parâmetros utilizados (47).

Os vocabulários controlados do projeto *Gene Ontology* (GO) (48) permitem associar produtos gênicos a seus processos biológicos, componentes e funções celulares e moleculares (49). A identificação automática da associação literatura-GO é porém

menos acurada que os métodos de curação manual, tais como a utilização dos termos MeSH (*Medical Subject Headings*) (50) para indexação de artigos do PubMed, um processo realizado por pessoas treinadas que potencialmente gera poucas associações falso-positivas. Além disso, tais associações são ligadas a genes e doenças, facilitando a identificação de relações previamente não conhecidas entre entidades, tais como as de proteína-proteína, droga-efeito e proteína-doença (51, 52).

No presente trabalho, foi desenvolvida uma metodologia para mineração da literatura com o objetivo de identificar genes e produtos gênicos relacionados com vários aspectos de CECP. Em resumo, o método empregou duas etapas automáticas. Na primeira etapa, foram extraídos artigos da literatura utilizando termos de interesse e, na segunda etapa, os artigos foram associados com os genes estudados pelos autores. Finalmente, na terceira etapa, os dados foram manualmente curados para obtenção de informações consistentes relacionadas com os objetivos propostos.

2 OBJETIVOS

O presente projeto teve como objetivo geral identificar *in silico* marcadores moleculares de CECP e de ligantes com potencial para utilização em tratamentos de pacientes com esse tipo de tumor.

Seus objetivos específicos compreendem:

a. Identificar e selecionar alvos moleculares potenciais de carcinoma epidermoide de cabeça e pescoço a partir de dados da literatura científica, microarranjos e proteômica;

b. Identificar ligantes que interajam eficiente e especificamente com alvos moleculares selecionados;

c. Avaliar o complexo proteína-ligante selecionado por meio de energia livre obtida na ligação;

d. Avaliar a ação dos ligantes na expressão gênica e no comportamento de células de carcinoma.

3 MATERIAIS E MÉTODOS

3.1. IDENTIFICAÇÃO E SELEÇÃO DE MARCADORES

3.1.1. OBTENÇÃO DOS DADOS E MINERAÇÃO DA LITERATURA

Mineração da literatura. A extração de artigos sobre CECp foi realizada no banco de dados PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) combinando termos descritores *MeSH* (*Medical Subject Headings*) e operadores booleanos. Três buscas sem restrições foram realizadas em 29 de junho de 2015. A primeira busca incluiu o termo *MeSH* “*Neoplasms by site*” e o resultado obtido foi utilizado apenas para comparações de grupos (utilizando tabelas de contingência). A segunda busca, denominada “*NOT*”, incluiu o termo “*Head and Neck Neoplasms*” e excluiu termos que não obedeciam aos critérios para CECp nos sítios anatômicos considerados no presente estudo (gengiva, palato, língua, orofaringe, hipofaringe e laringe). Finalmente, a busca “*OR*” incluiu somente os termos de interesse da árvore “*Head and Neck Neoplasms*” (Tabela 1).

Associação artigo científico-gene. A associação artigo científico - gene foi obtida utilizando o *gene2pubmed* (53), um arquivo que contém os identificadores de genes (GeneID) e os respectivos identificadores dos artigos do PubMed (PMIDs). Para esta associação, foram considerados somente os genes humanos.

Tabela 1. Termos utilizados para a revisão sistemática da literatura.

Tipo de busca	Termos MeSH	Tipo de Busca	Termos MeSH	
A	Head and Neck Neoplasms	Metastasis	Lymphatic Metastasis	
	Ear Neoplasms		Neoplasm Invasiveness	
	Esophageal Neoplasms		Neoplasm Metastasis	
	Facial Neoplasms		Neoplasm Micrometastasis	
	Leukoplakia, oral		Neoplasm Seeding	
	Lip Neoplasms		Neoplastic Cells, Circulation	
	Nasopharyngeal Neoplasms		Treatment Outcome	
	Nose Neoplasms		Prognosis/Outcome	Disease Progression
	Parathyroid Neoplasms			Disease-free Survival
	Salivary Gland Neoplasms			Fatal Outcome
	Thyroid Neoplasms			Follow-up Studies
Tracheal Neoplasms	Neoplasm Grading			
B	Gingival Neoplasms	Neoplasm Staging		
	Palatal Neoplasms	Nomogram		
	Tongue Neoplasms	Prognosis		
	Oropharyngeal Neoplasms	Survival Rate		
	Hypopharyngeal Neoplasms	Treatment Failure		
C	Laryngeal Neoplasms	Treatment Outcome		
	Neoplasms by site			

Curação manual de dados. Considerando que a estratégia de busca automatizada pode ter perdido artigos e genes relevantes, foram também analisados genes de outros dois bancos de dados (*HNOCDB/Head and Neck and Oral Cancer Database* <http://gyanxet.com/hno.html> e *OrCGDB/Oral Cancer Gene Database* <http://www.actrec.gov.in/OCDB/index.htm>), ambos relacionados com dados sobre câncer de cabeça e pescoço. Após curação manual (54, 55) e confirmação da relação com o CECP, os genes foram adicionados à lista obtida no PubMed. Genes sem evidências de envolvimento com os subsítios anatômicos de interesse foram excluídos.

Os artigos que mencionavam genes identificados pelas buscas “NOT” ou “OR” foram lidos por dois investigadores independentes para confirmação de sua associação com carcinomas orais, de laringe e de faringe. Os PMIDs e os genes correspondentes foram excluídos quando nenhum envolvimento era detectado ou quando o estudo analisava múltiplos tipos de tumores.

Análises estatísticas dos dados de mineração de literatura. Para o estabelecimento da associação gene - CECP, foram construídas tabelas de contingência utilizando os PMIDs curados e o grupo de PMIDs remanescentes da busca “*Neoplasms by site*”. O teste exato de Fisher também foi realizado para avaliar esta associação e valores de $p < 0.05$ foram considerados estatisticamente significantes. A análise foi realizada com o *software* SAS 9.3 (*SAS Institute Inc. Cary, NC, USA*) para Windows. Para ranqueamento dos genes, do mais relevante para o menos relevante definidos pelo número de publicações obtido pelas buscas “NOT” / “OR” e pelo arquivo de associação gene2pubmed, um teste hipergeométrico (56) foi empregado usando a aproximação de Stirling's (57). O método calcula a probabilidade de k, ou seja, a probabilidade de

publicações relevantes para o gene A ser casual, sendo S o escore para o gene A, m as publicações no arquivo gene2pubmed, n o número de publicações recuperadas pelas buscas e presentes no arquivo gene2pubmed, j o número de publicações que envolvem o gene A, e k o número de publicações relevantes que envolvem o gene A. A fórmula (ln = logaritmo natural) é (Equações 1 e 2):

$$1 \quad S_A = \ln(f(m, n, j, k))$$

$$2 \quad f(m, n, j, k) = \sum_{i=k}^{\min(n,j)} \frac{\binom{m-j}{n-i} \binom{j}{i}}{\binom{m}{n}}$$

A precisão (especificidade) e o *recall* (sensibilidade) foram calculados, respectivamente, como a proporção de genes relevantes da pesquisa e como a proporção de genes relevantes que foram recuperados [Precisão = genes recuperados e relevantes/total de genes recuperados; *Recall* = genes recuperados e relevantes/total de genes relevantes na coleção]. Para superar a dificuldade de prever o número total de genes no PubMed que são relevantes para a pesquisa, foram utilizados dados dos bancos HNOCDDB e OrGDB.

Em função da importância da identificação de uma assinatura prognóstica para o câncer de cabeça e pescoço assim como de marcadores associados com progressão da doença, termos *MeSH* e *non-MeSH* relacionados com metástase e com prognóstico e desfecho foram pesquisados nos títulos e nos resumos do conjunto de PMIDs curados manualmente (Tabela 1).

3.1.2. BANCO DE DADOS E INTERFACE WEB

Para integrar biomarcadores potenciais envolvidos na tumorigênese de cabeça e pescoço com dados disponíveis na literatura, foi construído um banco de dados relacional *MYSQL* e um sistema implementado em um servidor *Apache* usando o sistema operacional *Linux*. A plataforma Web foi desenvolvida utilizando as linguagens de programação JavaScript, PHP e também HTML para o *front-end* e o *back-end* suportado pelas linguagens de programação PHP e PERL. A plataforma possibilita ao usuário buscar e baixar informações relacionadas aos genes e proteínas envolvidos em câncer de cabeça e pescoço.

A página de entrada apresenta os objetivos do banco de dados e fornece ferramentas para pesquisa de genes relacionados a CECP, seus padrões de expressão e localização cromossômica. Dados de fontes externas foram incluídos no banco de dados para facilitar o acesso a uma maior quantidade de informações a respeito de um gene ou proteína em particular. Por exemplo, os genes selecionados pelos usuários são ligados a PMIDs, vias metabólicas (58, 59), ontologias associadas (48), mutações somáticas (60), doenças genéticas (61) e dados de microarranjos. Os dados de microarranjos foram obtidos a partir dos repositórios GEO (41) e ONCOMINE (42) (números de acesso GEO GSE9844, GSE6631, GSE1722, GSE13601, GSE3524, GSE2379, GSE25099 e dados Ginos ONCOMINE Head-Neck) (62-69) e podem ajudar o usuário a identificar genes com padrão semelhante de expressão. Os dados de proteínas incluem informações sobre drogas relacionadas e interações moleculares (43, 44, 70-77).

O banco de dados pode ser pesquisado gratuitamente no endereço <http://www.gencapo.famerp.br/hndb> com três formas para entrada de dados. As entradas

permitidas para buscas são o símbolo oficial do gene, o número de identificação e os *aliases* do *National Center for Biotechnology Information/NCBI* ou *Ensembl Project* (78, 79), assim como o número de acesso da proteína (*UniProt ID*) do *Universal Protein Resource/UniProt* (79). Digitando uma dessas entradas na caixa de busca, os usuários serão informados se o gene já foi relacionado com CECP. Os dados de todos os genes relacionados com CECP também podem ser recuperados de uma só vez.

Os dados extraídos pelas buscas podem ser exportados para tabelas PDF ou para arquivos TXT. Os resultados para um gene particular são exibidos em uma nova página que fornece o nome oficial e o número de identificação do gene, *aliases*, localização cromossômica, padrão de expressão gênica gerado por estudos de microarranjos em tecido tumoral e artigos que suportam o envolvimento do gene com o CECP. O resultado inclui também ontologias, vias metabólicas, mutações somáticas em câncer, interações entre proteínas e drogas para os alvos de interesse.

3.1.3. ONTOLOGIAS E ANÁLISE DE VIAS METABÓLICAS

O conjunto de genes curados relacionados com CECP foram importados para o DAVID (80, 81), um banco de dados para anotação, visualização e descoberta integrada (80), e suas ontologias e vias metabólicas foram anotadas utilizando como referência o genoma humano completo. O teste exato de Fisher unicaudal foi utilizado para enriquecimento da análise dos genes e o teste de Bonferroni para obtenção dos valores de p corrigidos (valores de $p < 0.05$ foram considerados significantes). O *software Ingenuity Pathway Analysis (IPA)* (Quiagen, Redwood City, CA, USA) foi empregado para identificar as vias metabólicas canônicas mais representadas no conjunto de genes

relacionados com CECP.

Uma análise similar pelas ferramentas DAVID e pelo programa *Ingenuity* foi empregada para genes ativados por ligante selecionado para proteínas de interesse, conforme descrição a seguir.

3.1.4. ESTRUTURA TRIDIMENSIONAL

A partir dos resultados obtidos pela mineração da literatura relacionada a CECP, foram selecionadas algumas proteínas que atuam em processos inflamatório. As estruturas dessas proteínas juntamente com informações sobre as abordagens empregadas para resolução das estruturas foram obtidas a partir do banco de dados *Protein Data Bank PDB*.

3.1.5. SELEÇÃO DE LIGANTES PARA PROTEÍNAS DE INTERESSE

A seleção de compostos é uma etapa fundamental na identificação de um ligante adequado para uma molécula de interesse. No presente estudo, esta seleção foi realizada no *ZINC*, um banco de dados de domínio público que contém a maior coleção de compostos químicos disponíveis comercialmente para rastreamento virtual de estruturas.

Para se tornar um fármaco, um composto precisa apresentar uma série de características físico-químicas como pré-requisitos. A regra dos cinco de Lipinski é o conjunto de pré-requisitos mais utilizados e leva em consideração o coeficiente de partição óleo-água, expresso pelo log P, a massa molecular e o número de doadores e

aceptores de hidrogênio. Um fármaco que satisfaça esses requisitos apresenta coeficiente de partição log P igual ou maior que cinco, massa molecular menor ou igual a 500 Da, mais de 5 doadores de hidrogênio e menos de 10 receptores de hidrogênio.

Em função de estudos prévios do grupo e da disponibilidade do extrato puro, foi estudado em detalhes o potencial como ligante da piperlongumina, também conhecida como piplartina, um composto extraído da pimenta *Piper longum* (Figura 1 A e B) originária do sul da Índia e do sudeste da Ásia. Para avaliar a ação anti-inflamatória da piplartina, foram também selecionados medicamentos anti-inflamatórios já conhecidos com estruturas depositadas no *ZINC*.

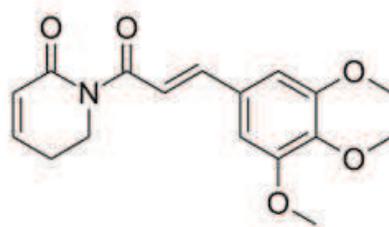
**A****B**

Figura 1. A. Fruto da pimenta *Piper Longum*. B. Estrutura química da pipartina

3.1.6. DOCKING MOLECULAR

Para os experimentos de *docking* molecular, foi utilizado o programa *AutoDock Vina 1.1.2* para linux (82). As estruturas tridimensionais das proteínas selecionadas foram preparadas com o programa *autodock tools 1.4.5*. As estruturas tiveram todos os átomos de hidrogênios adicionados, as moléculas de água e ligantes foram retirados, as cargas parciais de *Gasteiger* foram adicionadas e as coordenadas para criação do *grid* para o *docking* foram determinadas conforme o sítio ativo das proteínas alvo.

3.1.7. CULTURA DE CÉLULAS

Para os experimentos *in vitro*, a linhagem celular HEP-2 (descrita originalmente como procedentes de carcinoma epidermoide de laringe humana) (*ATCC Rockville, Maryland, USA*) foi cultivada em frascos de 75cm³ contendo 8mL de meio *Dulbecco's Modified Eagle Medium – DMEM (Cultilab, Campinas, SP)* suplementado com 10% de soro fetal bovino (SFB, Cultilab), aminoácidos não essenciais 10mM (*Gibco, Carlsbad, CA, USA*), L-glutamina 2mM(*Gibco, Carlsbad, CA, USA*), piruvato de sódio 1mM (*Sigma-Aldrich, St. Louis, MO, USA*) e solução de antibiótico/antimicótico 0,1% (*Sigma-Aldrich, St. Louis, MO, USA*). As células foram incubadas a 37°C, em atmosfera úmida com 5% de dióxido de carbono (CO²). O crescimento e a morfologia celular foram monitorados a cada 24h em microscópio invertido e o meio de cultura do frasco foi trocado a cada dois dias. Ao atingirem uma confluência de aproximadamente 70%, o meio de cultura foi aspirado e as células foram lavadas com 5mL de solução tampão fosfato-salina sem cálcio e magnésio pH 7,2 (PBS). Esse meio de lavagem foi aspirado

e as células foram tripsinizadas com a adição de 2mL de solução de tripsina 0,05% - EDTA 0,02% e incubadas a 37°C por 2 ou 3 min para que a tripsina pudesse agir e induzir o descolamento das células. Em seguida, a tripsina foi inativada com a adição de 3mL de meio DMEM com 10% de soro fetal bovino. O material foi aspirado e dividido em dois tubos de ensaio, que foram centrifugados a 1500rpm por 5 min. O *pellet* com células foi ressuspensionado em 1mL de meio e recultivado. Periodicamente, um dos *pellets* era ressuspensionado em SFB e solução de dimetilsulfóxido (DMSO) (Sigma-Aldrich) 9:1, e congelado para manutenção do estoque de linhagens.

Para avaliação do número de células a serem semeadas na fase de recultivo, foi utilizado o contador automático *Countess® Cell Counter* (Invitrogen, Forster City, CA, USA). Uma alíquota de 5uL de meio contendo as células tripsinizadas foi adicionada a um tubo com 5uL de azul de tripano (*Sigma-Aldrich, St. Louis, MO, USA*) e os 10uL totais foram aplicados em lâmina própria para contagem. O aparelho utiliza um algoritmo para análise de imagens e para calcular o número de células vivas e mortas. As primeiras brilham devido à ausência do corante em seu interior, pois a membrana celular íntegra impede sua entrada. As células mortas, ao contrário, ficam opacas por causa da presença do azul de tripano em seu citoplasma, uma vez que a membrana celular não é mais seletiva. O número de células viáveis/mL identificadas pelo aparelho foi utilizado para calcular o volume de meio com células a ser semeado. Para obtenção de réplicas, as células foram semeadas com densidade de 1×10^6 e cultivadas a 37°C em atmosfera de 5% de CO² por 24h.

Após sua adesão à garrafa, o meio de cultura foi trocado por meio sem soro para sincronização do ciclo celular. As células foram mantidas nesse meio por 24 horas, tempo suficiente para que ingressassem na fase G0. Após esse período, o meio foi

substituído por meio completo e as células foram mantidas em cultivo por 72h, quando foi realizado o bloqueio da cultura com solução de tripsina/EDTA. Uma das réplicas foi encaminhada para análise de viabilidade celular e a outra continuou em cultivo para análises posteriores. Todos os procedimentos de cultivo celular foram realizados em fluxo laminar bidirecional, específico para esse fim, para manutenção da esterilidade dos materiais bem como para evitar a contaminação das culturas.

3.1.8. TRATAMENTO COM PIPLARTINA

A piplartina foi diluída em DMSO nas seguintes concentrações: 5mM, 10mM e 20mM. Para avaliar a ação da piplartina na linhagem celular HEP-2, as células foram tripsinizadas, recultivadas com densidade de 5×10^3 em placas de 96 poços, tratadas com a piplartina ou com o veículo utilizado para diluição da piplartina (DMSO) por um período de 24, 48 e 72h. Estas placas foram utilizadas para ensaios de viabilidade celular.

3.1.9. ENSAIO DE VIABILIDADE CELULAR

Para o ensaio de viabilidade celular, foi utilizado o *The CellTiter 96® Aqueous One Solution Cell Proliferation Assay- MTS* (Promega, Madison, WI, USA), um método colorimétrico que determina o número de células viáveis. Neste ensaio, o sal de tetrazolium (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) é biorreduzido pelas células vivas com formação do composto solúvel formazan, que produz mudança de coloração do meio. A quantidade

de formazan é mensurada pela absorvância de 490 nm e diretamente proporcional ao número de células vivas na cultura.

3.1.10. ENSAIO DE MIGRAÇÃO

O ensaio de migração foi originalmente introduzido por Boyden (e por isso, frequentemente chamado de ensaio de câmara de Boyden) para analisar a resposta quimiostática de leucócitos (83). O princípio do ensaio é baseado na separação por membrana porosa (através da qual as células migram) de duas câmaras contendo meio de cultura. O tamanho da célula é determinante para a escolha do diâmetro do poro, pois é necessário que o poro possibilite a transmigração ativa das células. As membranas disponíveis apresentam poros com diâmetro que variam entre 3 a 12µm. Geralmente as células são cultivadas em meio apropriado na parte superior e migram na direção vertical através dos poros da membrana para o compartimento inferior contendo meio com atrativo ou simplesmente contendo soro.

Para os ensaios de migração, um total de 5×10^4 /300µL de meio DMEM sem soro fetal bovino foram semeadas no compartimento superior do inserto e tratadas com piplartina conforme descrito anteriormente. No compartimento inferior, foram adicionados 500µL de meio DMEM suplementado com soro fetal bovino 10%. Em seguida, a placa foi incubada por um período de 12h a 37°C e 5% de CO₂, durante o qual as células, foram atraídas pelo soro e migraram em direção ao compartimento inferior da câmara. As células contidas no inserto foram fixadas com formaldeído 0,4% à temperatura ambiente por 20min. Aquelas presente na superfície superior da membrana (e que, portanto, não migraram) foram removidas suavemente com uma

haste de algodão, enquanto as células contidas no compartimento inferior foram coradas por 5min com o composto fluorescente 4',6-diamidino-2-fenilindol/DAPI (Uniscience, São Paulo, SP, Brasil) e submetidas a um passo de lavagem com PBS 1X. Cinco campos da superfície inferior da membrana, selecionados casualmente, foram fotografados em microscópio de fluorescência invertido utilizando o *software Axiovision Release 4.8* (ZEISS Oberkochen, Alemanha) e as células foram contadas com o *software ImageJ* (National Institutes of Health – NIH Bethesda, EUA).

3.1.11. EXTRAÇÃO DE RNA DE CULTURACELULAR

O RNA de células submetidas ao tratamento com a piplartina foi extraído com Trizol (*Life Technologies, Grand Island, NY, USA*) utilizando protocolo já estabelecido no laboratório. Em resumo, o meio de cultura foi retirado das garrafas e as células foram lavadas com 5mL de PBS. O PBS foi aspirado e foram adicionados 3 mL de Trizol. As amostras foram distribuídas em três tubos de 1,5mL e mantidas à temperatura ambiente por 5 min. O próximo passo foi a adição de 200uL de clorofórmio gelado para cada mL de amostra em Trizol. Os tubos foram homogeneizados por 1min e incubados por 3min à temperatura ambiente. As amostras foram centrifugadas a 14.000g por 20min a 4°C, para separar a fase aquosa que contém RNA, da interfase branca e leitosa que contém principalmente DNA, e da fase orgânica com proteínas. A fase aquosa com RNA foi transferida para um novo tubo com 400uL de isopropanol gelado e 10ug de glicogênio, e mantida por 15min à temperatura ambiente. Após essa etapa de precipitação, as amostras foram centrifugadas a 14.000g, por 15min a 4°C, e o sobrenadante foi descartado por inversão. O sedimento foi lavado com 1mL de etanol 75% gelado e

agitado até se soltar do tubo. A amostra foi centrifugada a 7.500g por 5min a 4°C, o sobrenadante foi descartado e o sedimento foi seco sobre papel absorvente à temperatura ambiente, sendo posteriormente ressuspensionado em 20-50µL de água livre de nucleases (dependendo do tamanho do sedimento). As amostras foram colocadas em banho seco a 500g por 10min a 57°C. O RNA foi quantificado e a razão de absorvância foi obtida em espectrofotômetro *NanoDrop ND 1000 (Thermo Scientific, Wilmington, Delaware, USA)*. A qualidade e a integridade do RNA foram avaliadas no equipamento *2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA)*.

3.1.12. SÍNTESE DE DNA COMPLEMENTAR (cDNA)

O cDNA foi sintetizado a partir de 1µg de RNA total, utilizando o *High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, Califórnia, USA)*. A qualidade do cDNA foi avaliada em reação de PCR de fragmento do gene *glyceraldehyde-3-phosphate dehydrogenase, GAPDH*, com 101 pb. As condições da reação foram cinco minutos a 94°C e 35 ciclos de 50 segundos a 94°C para desnaturação do DNA, 30 segundos para anelamento dos iniciadores a 60°C e 45 segundos a 72°C para extensão das cadeias, além de sete minutos a 72°C para extensão final.

3.1.13. PCR EM TEMPO REAL

A reação de PCR em tempo real foi realizada no equipamento ABI Prism 7500 Fast Sequence Detection System (Thermo Fisher, Waltham, MA, USA). As reações foram preparadas para um volume final de 10µL, seguindo protocolo da empresa. Para

análise da expressão por PCR em tempo real, foram selecionados genes relacionados com processos inflamatórios (*PCR array* 4418719, Thermo Fisher Scientific, Waltham, MA USA) (Tabela 2). As análises dos resultados das reações foram realizadas com o *software Data Assist v.3.01* (Thermo Fisher). Esta ferramenta utiliza o método comparativo de CT (ddCT) para quantificar a expressão relativa dos genes entre grande número de genes e de amostras. O CT (*cycle threshold*) é definido como o número de ciclos necessários para que o sinal fluorescente atinja o limiar de detecção e depende do número de DNAs moldes no início da PCR. Em função da natureza exponencial da PCR, o valor de *fold change* é calculado como $2^{-\Delta\Delta CT}$. Os valores foram transformados em log₂: valores abaixo de -1 significam expressão reduzida e valores acima de 1 representam expressão elevada.

Tabela 2. Genes relacionados a processos inflamatórios

ADRB1	CACNA2D1	HRH2	KLK1	PDE4D	MAPK8	TNF	HRH3
ADRB2	CACNB2	HTR3A	KLK2	PLA2G1B	PTAFR	TNFRSF1A	PLA2G2D
ALX12	CACNB4	ICAM1	KLKB1	PLA2G2A	PTGDR	TNFRSF1B	IL1RAPL2
ALOX5	CASP1	IL1R1	KNG1	PLA2G5	PTGER2	VCAM1	KLK14
ANXA1	CD40	IL2RA	LTA4H	PLCB2	PTGER3	ILIR2	PLCE1
ANXA3	CD40LG	IL2RB	LTC4S	PLCB3	PTGFR	PLA2G7	KLK15
ANXA5	CES1	IL2RG	MC2R	PLCB4	PTGIR	PLA2G10	LTB4R2
KLK3	LTB4R	IL13	NFKB1	PLCD1	PTGIS	PLA2G4C	
BDKRB1	MAPK14	ITGAL	NOS2	PLCG1	PTGS1	IL1RL1	
BDKRB2	NR3C1	ITGAM	PDE4A	PLCG2	PTGS2	HTR3B	
CACNA1C	HPGD	ITGB1	PDE4B	MAPK1	TBXA2R	TNFSF13B	
CACNA1D	HRH1	ITGB2	PDE4C	MAPK3	TBXAS1	CYSLTR1	

4 RESULTADOS

4.1. IDENTIFICAÇÃO E SELEÇÃO DOS MARCADORES

No total, as buscas “*Neoplasms by site*”, “*NOT*” e “*OR*” resultaram em 1.819.931 (no período entre 2015 e 1928), 38.862 e 41086 (entre 2015 e 1945) artigos, respectivamente que, após a associação gene2pubmed e exclusão de redundância, geraram 2480 PMIDs e 1611 genes.

A análise de tabelas de contingência comparando os dados obtidos para CECP e para todos os tipos de tumores (exceto CECP) teve como objetivo identificar marcadores específicos para o câncer de cabeça e pescoço. Apesar de alguns genes mostrarem nível de significância $p > 0.05$, nenhum deles foi específico para este tumor.

Após a curação manual dos 1611 genes não redundantes, 421 não relacionados com CECP foram excluídos e uma lista de 1190 genes foi obtida. Para esta lista, 180 entre 517 genes identificados pelos bancos de dados HNOCDDB e OrCGDB mas não detectados pela abordagem inicial foram acrescentados após reavaliação manual, resultando em 1370 genes no total. Cento e quarenta e oito dos 1370 genes foram associados a pelo menos cinco PMIDs, sendo *TP53* e *EGFR* os dois primeiros genes da lista. Considerando esses dados, a precisão de nosso método automatizado foi estimada em 74%, e a especificidade foi estimada em 87%. Os escores para estes genes foram confirmados pelo teste hipergeométrico e indicam que eles representam os mais extensivamente estudados e exibem resultados mais relevantes. Em relação aos 893 genes mencionados por somente um artigo, muitos deles ainda não foram completamente explorados como potenciais marcadores e merecem mais investigações.

Os 1370 genes relacionados com CECF mostraram uma distribuição heterogênea ao longo dos cromossomos e, como esperado, muitos deles estão localizados em regiões conhecidas como *hot spot* para CECF, tais como 11q13. Contudo, alguns outros foram localizados em regiões menos citadas. Aproximadamente 11% dos genes foram mapeados no cromossomo 1, 7% no cromossomo 11 e a mesma porcentagem no cromossomo 17, uma distribuição não correlacionada com o tamanho em MB de cada cromossomo.

Para avaliar o desempenho da abordagem *NOT/OR* utilizada para mineração da literatura, a lista de 1190 genes não redundantes foi comparada com os genes identificados pelos bancos de dados HNOADB e OrCGDB (atualmente congelados). Considerando os mesmos sítios anatômicos analisados pela abordagem *NOT/OR*, o HNOADB extraiu 133 genes relacionados com câncer oral, 14 de língua, 7 de hipofaringe, 3 de orofaringe e 60 com câncer de laringe. O banco de dados OrCGDB selecionou 374 genes envolvidos em câncer oral pesquisando resumos do PubMed e termos MeSH. Um total de 520 genes não redundantes foram identificados pelos dois bancos de dados. Destes, 180 não estavam presentes na lista de 1190 genes. Ao contrário da estratégia utilizada pelos autores de HNOADB e OrCGDB, o presente estudo realizou duas buscas com termos MeSH e foi mais estrigente por excluir artigos que também analisaram outros tipos de tumores.

Além disto, as análises dos 1190 genes realizadas pelo *Ingenuity Pathway Analysis (IPA)* mostrou que a via metabólica canônica principal foi a de Mecanismo Molecular do Câncer ($p=6.64^{-66}$, *overlap* 34.5%, 126/365), corroborando assim sua relevância no processo neoplásico. Diferentemente, esta via metabólica não estava associada aos genes dos bancos OrCGDB e HNOADB, que mostraram os menores

valores de p para as vias Sinalização do Receptor de Aril Hidrocarboneto ($p=1.42E^{-32}$, *overlap* 29.3% 41/140), Sinalização do Câncer de Bexiga ($p=4.44E^{-32}$, *overlap* 39.1% 34/87) e Fibrose Hepática / Ativação da Célula Estrelada Hepática ($p=5.49E^{-32}$, *overlap* 24.6% 45/183).

Para investigar o significado biológico dos genes relacionados a CECP, foi empregada a ferramenta DAVID, que permite a identificação de ontologias e vias metabólicas. O DAVID mapeou um total de 1329 identificadores a partir da lista de 1370 genes e os termos similares foram agrupados, removendo as redundâncias. Mais de quinhentos grupos de anotação foram obtidos, 86 deles enriquecidos com escores >5.0 e valor de p corrigido por Bonferroni < 0.05 . Os resultados mostraram maior representação de grupos relacionados a desenvolvimento e diferenciação, resposta a estímulo, transdução de sinais, apoptose, transcrição, adesão, migração e ciclo celular, que são processos biológicos relevantes para o câncer. Também mostraram processos relacionados a resposta a hormônios esteroides (escores=37.99 e valor de p corrigido por Bonferroni= $7.50E^{-31}$).

As cinco principais vias canônicas identificados pelo *Ingenuity Pathway Analysis* para o total de 1370 genes foram Mecanismos Moleculares de Câncer (Figura 2A), Sinalização de Metástase de Câncer Colorretal, Papel de Macrófagos, Fibroblastos e Células Endoteliais em Artrite Reumatóide, Sinalização em Adenocarcinoma Pancreático e Sinalização IL -8 (valores de $p=4.90E^{-71}$, $1.95E^{-58}$, $7.11E^{-56}$, $5.46E^{-53}$, $2.25E^{-48}$, respectivamente), assim validando fortemente a estratégia utilizada e a característica informativa do conjunto de genes. Da mesma forma que a ferramenta DAVIS, o *Ingenuity* mostrou o beta estradiol como um dos principais reguladores a montante ($p= 3.35E^{-163}$), próximo aos genes *TGFB1*, *TNF*, and *TP53*.

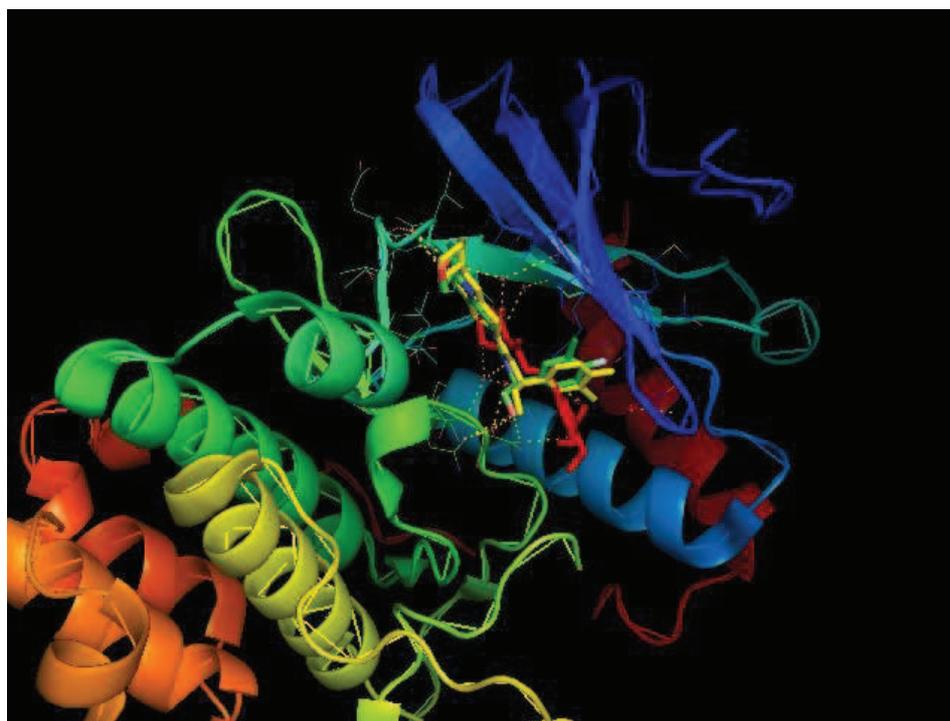
4.2 DOCKING MOLECULAR

As proteínas recuperadas por mineração da literatura relacionadas a CECP, que apresentaram envolvimento em processos inflamatórios e com estruturas tridimensionais resolvidas e depositadas no PDB (*Protein Data Bank*), foram submetidas a experimento de *docking* molecular. As drogas utilizadas como ligantes são drogas disponíveis no mercado que apresentam ação anti-inflamatória e características físico-químicas que obedecem às regras de Lipinski (84) (Tabela 3). Os resultados mostraram que a piplartina possui valores de energia livre comparáveis a drogas anti-inflamatórias conhecidas. Tais valores são menores ou iguais a -6.0 para 13 (92.85%) de 14 proteínas selecionadas, perdendo em eficiência somente para dexametasona e para piroxicam, que apresentaram energia livre menor ou igual a -6.0 para todas as proteínas (100%) utilizadas. As drogas diclofenaco e paracetamol foram as que apresentaram mais baixo desempenho, com energia livre menor ou igual a -6.0 somente para oito (57.14%) e uma (7.14%) das proteínas, respectivamente.

Os parâmetros utilizados para construção do *gridbox* foram validados por um processo de *redocking* (Figura 2). Para isso, foram empregadas estruturas de proteínas alvo depositadas no PDB na forma complexada proteína/molécula inibidora. A molécula inibidora foi retirada e o processo de *docking* foi refeito para verificar se a molécula inibidora/ligante voltaria para sua posição original. Foi observado que, nas estruturas originalmente resolvidas em complexo com a molécula inibidora e nas estruturas docadas com os parâmetros utilizados no presente trabalho, as drogas ficaram aproximadamente na posição original.

Tabela 3. Características físico químicas das drogas utilizadas para *docking* molecular.

Droga	Peso Molecular	LogP	Doador de H⁺	Aceptor de H⁺
Diclofenaco	295.145	4.57	1	3
Piroxicam	331.353	0.88	2	7
Paracetamol	151.165	0.68	2	3
Nimesulida	307.307	2.81	0	7
Naproxeno	229.255	3.38	0	3
Ketoprofen	253.277	3.59	0	3
Ketorolac	254.265	2.20	0	4
Dexametasona	392.467	2.06	3	5
Resveratrol	228.247	0.79	3	3
Piplartina	317.341	1.62	0	6

**Figura 2.** *Redocking* da molécula inibidora da proteína MAPK1. Verde: molécula inibidora cristalizada e resolvida em complexo com a MAPK1. Amarelo: molécula inibidora da MAPK1 redocada. Vermelho: piplartina docada com a MAPK1.

4.3. ENSAIOS FUNCIONAIS

4.3.1. VIABILIDADE CELULAR

Os resultados dos ensaios com MTS mostraram que o tratamento com piplartina reduziu significativamente (ANOVA $p < 0.05$) o número de células viáveis em relação aos controles não tratados. Todas as concentrações (5mM, 10mM e 20mM) e tempos (24, 48 e 72 horas) foram testados (Figura 3).

4.3.2. MIGRAÇÃO E INVASÃO

Para tratamento das células Hep-2 nos ensaios de migração e invasão, a piplartina foi utilizada na concentração de 20mM. Os resultados mostraram que as células tratadas com piplartina apresentaram taxa de migração significativamente menor (ANOVA $p < 0.001$) em relação às células em DMSO e às células controle (Figuras 4 e 5).

Os ensaios de invasão não apresentaram resultados significativamente diferentes entre as células tratadas e os controles (ANOVA $p < 0.1021$).

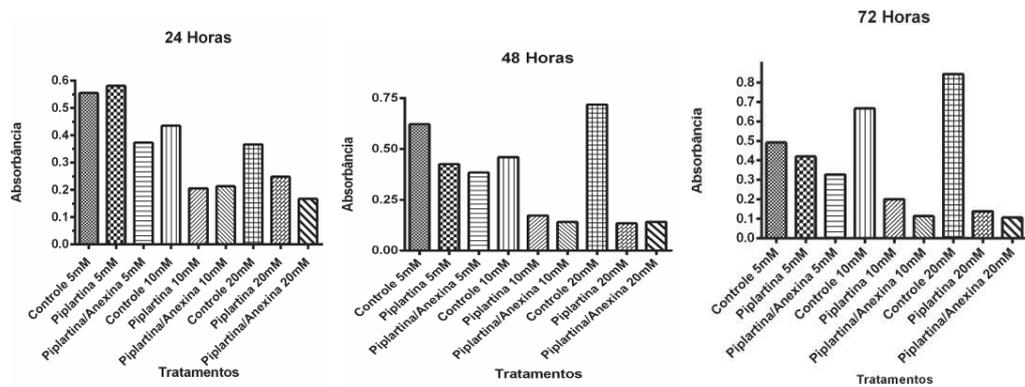


Figura 3. Gráficos representando os tempos de tratamento e as concentrações de pipartina utilizados para os testes de citotoxicidade

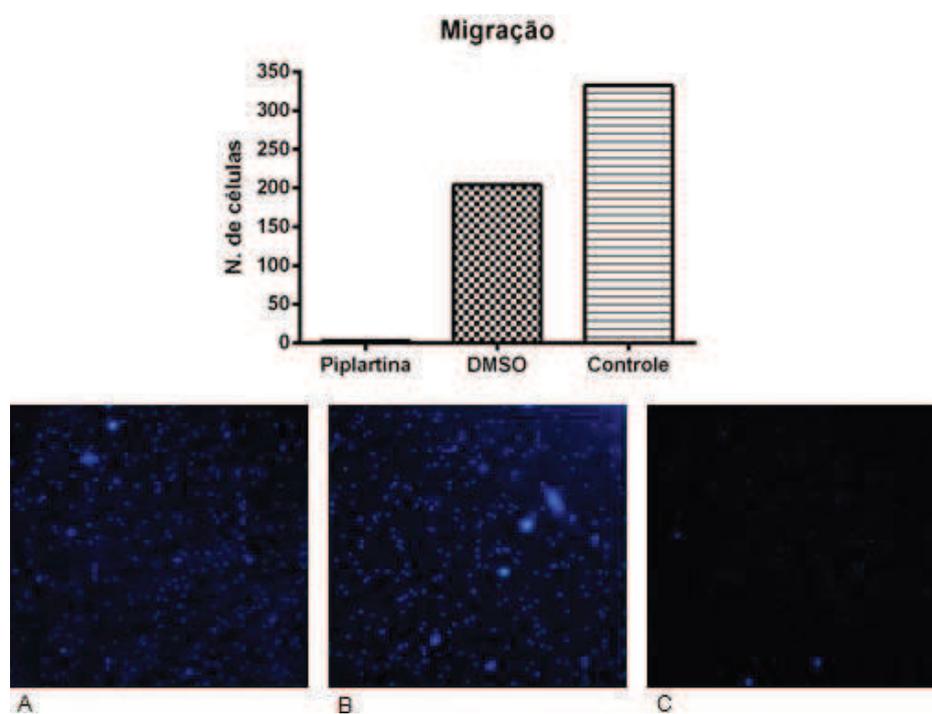


Figura 4. Gráfico e fotomicrografias do ensaio de migração. **A:** células controle sem tratamento, **B:** células em DMSO, **C:** células tratadas com a piplartina

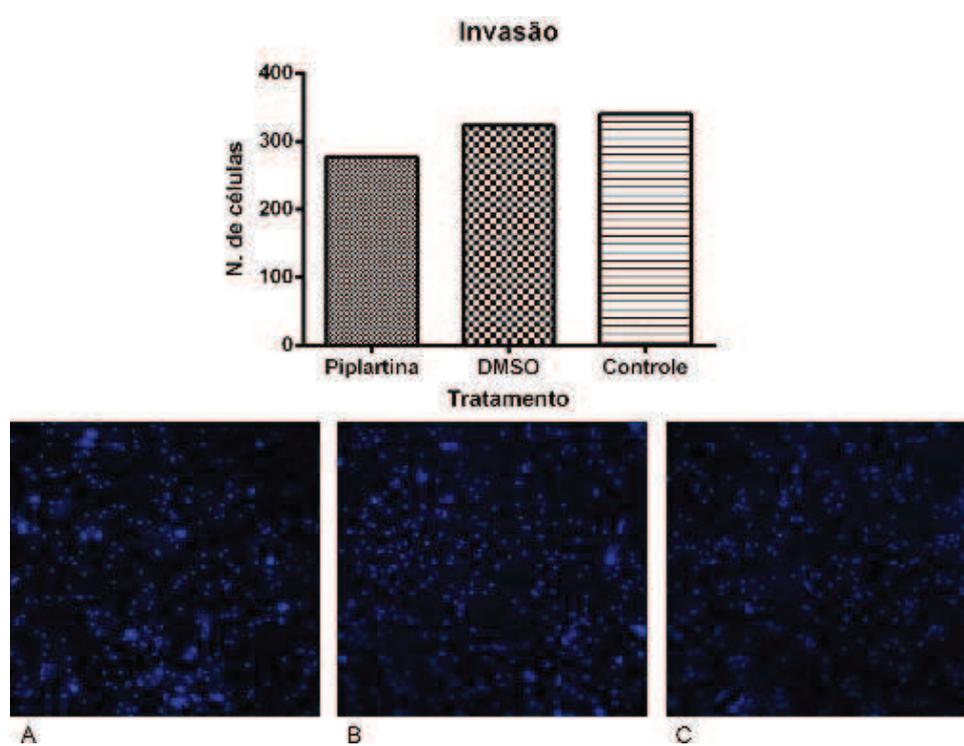


Figura 5. Gráfico e fotomicrografias do ensaio de invasão. **A:** células controle sem tratamento, **B:** células em DMSO, **C:** células tratadas com a piplartina

4.4. RNA DA LINHAGEM CELULAR

Os RNAs extraídos da linhagem celular Hep-2 tratada com piplartina, DMSO e controle mostraram quantidade e razão de absorbâncias satisfatórias (Tabela 5). A análise no equipamento *Bioanalyser* mostrou RIN (*RNA integrity number*) maior ou igual a 7 em todas as amostras.

4.5 EXPRESSÃO GÊNICA

A análise estatística dos resultados obtidos em experimentos de PCR em tempo real revelou que, dos 96 genes relacionados com processos inflamatórios utilizados no estudo, 43 apresentaram níveis de expressão alterados, sendo 25 genes com padrão de expressão reduzido e 18 genes com padrão de expressão aumentado (Tabela 6 e Figura 6). Dos 43 genes diferencialmente expressos, 11 (*CD40*, *IL2RA*, *ITGB1*, *MAPK1*, *MAPK8*, *MAPK14*, *NOS2*, *PLCE1*, *PLCG1*, *TNF*) foram identificados pelo presente estudo como sendo relacionados a CECP.

Como esperado para *PCR array* específico para processos inflamatórios, as duas principais vias canônicas identificados pelo programa *Ingenuity* para os 43 genes diferencialmente expressos foram Ação Antioxidante da Vitamina C e Sinalização de Eicosanoides (valores de $p=5.27E^{-15}$, $6.10E^{-15}$, respectivamente). Os cinco reguladores a montante com menores valores de p foram TNF, IL1B, IL4, lipopolissacarídeo e CD40 (valores de p de *overlap*= $6.09E^{-15}$ a $1.20E^{-10}$), e as principais doenças e funções biológicas foram Doença Cardiovascular, Resposta Inflamatória, Doença Imunológica, Doença Inflamatória e Doença Respiratória (valores de $p=7.43E^{-20}$ a $1.21E^{-05}$).

Tabela 4. Quantificação e análise da qualidade do RNA extraído

RNA	ng/uL	Razão 260/280	RIN
Hep-2 – piplartina	412	1.9	8.1
Hep-2 – DMSO	4.271	1.8	7.0
Hep-2 – controle	4.252	1.9	7.5

Tabela 5. Genes diferencialmente expressos após tratamento com o composto piplartina e suas funções. TNF=fator de necrose tumoral; IL=interleucina; FC=fator de crescimento

Genes com expressão reduzida	Função molecular*	Fold change (log2)	Genes com expressão elevada	Função molecular*	Fold Change (log2)
<i>LTB4R2</i>	receptor de leucotrieno	- 8,13	<i>TNF</i>	citocina	7,89
<i>LTB4R</i>	receptor de leucotrieno	- 2,21	<i>IL2RA</i>	receptor de interleucina	6,00
<i>PLCE1</i>	fosfolipase	- 2,15	<i>IL1R2</i>	receptor de interleucina	5,90
<i>PDE4D</i>	fosfodiesterase	- 2,13	<i>PLA2G2A</i>	fosfolipase	3,90
<i>TBXA2R</i>	receptor de tromboxane	- 2,12	<i>ITGAM</i>	ligadora de	3,91
<i>TNFRSF1A</i>	receptor de TNF	- 2,11	<i>CYSLTR1</i>	receptor de leucotrieno	2,98
<i>PTGER2</i>	receptor de	- 2,06	<i>HTR3A</i>	ligadora de serotina	2,97
<i>IL2RB</i>	receptor de interleucina	- 2,06	<i>ALOX12</i>	lipoxigenase de	2,94
<i>ITGB1</i>	integrina	- 1,17	<i>ADRB2</i>	receptor adrenérgico	2,91
<i>MAPK8</i>	quinase serina/treonina	- 1,16	<i>CD40</i>	receptor de TNF	2,87
<i>MAPK14</i>	quinase serina/treonina	- 1,15	<i>IL1RL1</i>	receptor de interleucina	2,87
<i>MAPK1</i>	quinase serina/treonina	- 1,15	<i>KLK2</i>	endopeptidase	2,87
<i>BDKRB1</i>	receptor de bradiquinina	- 1,13	<i>KLK3</i>	endopeptidase	2,87
<i>HRH1</i>	receptor de histamina	- 1,12	<i>KNG1</i>	inibidor de	2,87
<i>CACNB4</i>	canal de cálcio	- 1,11	<i>PLA2G5</i>	fosfolipase	2,87
<i>PLCB3</i>	fosfolipase	- 1,11	<i>NOS2</i>	sintase de óxido nítrico	2,72
<i>LTA4H</i>	hidrolase de leucotrieno	- 1,11	<i>CES1</i>	hidrolase	1,91
<i>A2M</i>	ligador de TNF, IL, FC	- 1,10	<i>KLK14</i>	endopeptidase	1,88
<i>PLCG2</i>	fosfolipase	- 1,10			
<i>NFKB1</i>	regulador de transcrição	- 1,09			
<i>NR3C1</i>	fator de transcrição	- 1,09			
<i>PDE4B</i>	fosfodiesterase	- 1,07			
<i>PLCG1</i>	fosfolipase	- 1,06			
<i>ADRB1</i>	receptor adrenérgico	- 1,05			
<i>CACNB2</i>	canal de cálcio	- 1,04			

*Segundo o Consórcio Gene Ontology (<http://geneontology.org/>).

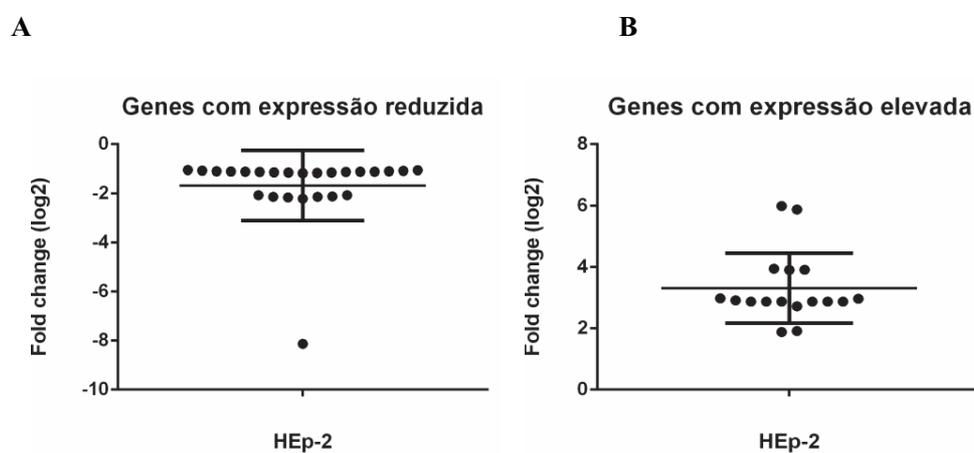


Figura 6. Gráficos dos dados de expressão gênica determinados por PCR em tempo real em células Hep-2. (A) Genes com expressão reduzida e (B) genes com expressão elevada após tratamento com a piplartina. A quantificação relativa foi calculada pelo método comparativo de CT. O experimento controle sem tratamento foi utilizado como calibrador para os cálculos realizados. Os valores foram transformados em \log_2 (eixo y): valores abaixo de -1 significam expressão reduzida e valores acima de 1 representam expressão elevada. As barras indicam a média \pm desvio padrão. O valor de *fold change* foi calculado como $2^{-\Delta\Delta CT}$.

5 DISCUSSÃO

A enorme quantidade de informações biológicas obtidas em estudos de genômica, transcritômica, proteômica e epigenômica disponíveis em bancos de dados públicos compreende uma fonte potencial para pesquisa e novas descobertas. Para tornar tais descobertas possíveis, contudo, é necessário o uso de ferramentas de bioinformática capazes de integrar dados provindos de diferentes plataformas. Atualmente, vários esforços estão sendo feitos para criar métodos destinados a processar e disponibilizar estes dados para a comunidade científica. Entretanto, tais iniciativas requerem grandes investimentos em recuperação, curação e processamento de informação (42, 85, 86).

No presente trabalho, foi realizada a mineração da literatura disponível publicamente para identificar marcadores moleculares relacionados a CECP. Para isto, foi utilizado um fluxograma que incluiu a busca de artigos científicos no PubMed por termos MeSH, a associação desses artigos com genes/proteínas por meio do arquivo gene2pubmed e etapas de curação. Os marcadores identificados e as anotações funcionais relacionadas a eles foram armazenados em um banco de dados público, que pode ser utilizado com fonte para identificação de novos alvos terapêuticos.

A comparação dos resultados obtidos para CECP com aqueles obtidos para todos os tipos de tumores (exceto CECP) foi realizada por tabelas de contingência e os resultados mostraram que, apesar de alguns genes exibirem nível de significância $p > 0.05$, nenhum deles foi específico para CECP. Realmente, genes já estabelecidos para carcinoma de cabeça e pescoço listados por Leemans e colaboradores (2011) (10) (*CCND1*, *CDKN2A*, *EGFR*, *MET*, *PIK3CA*, *PTEN*, *SMAD4*, *TP53*) estão também

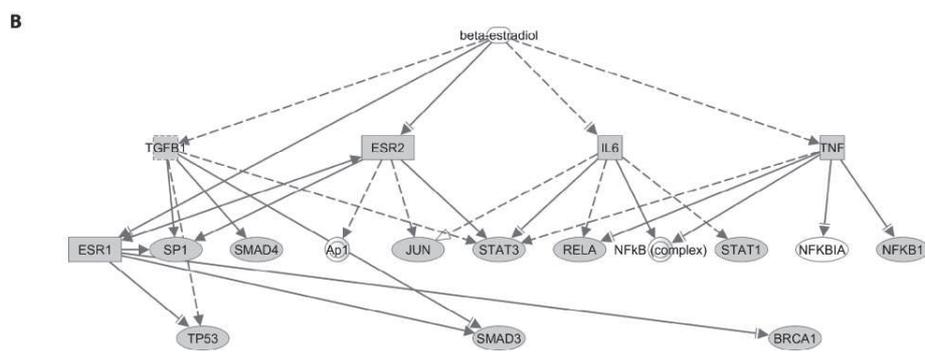
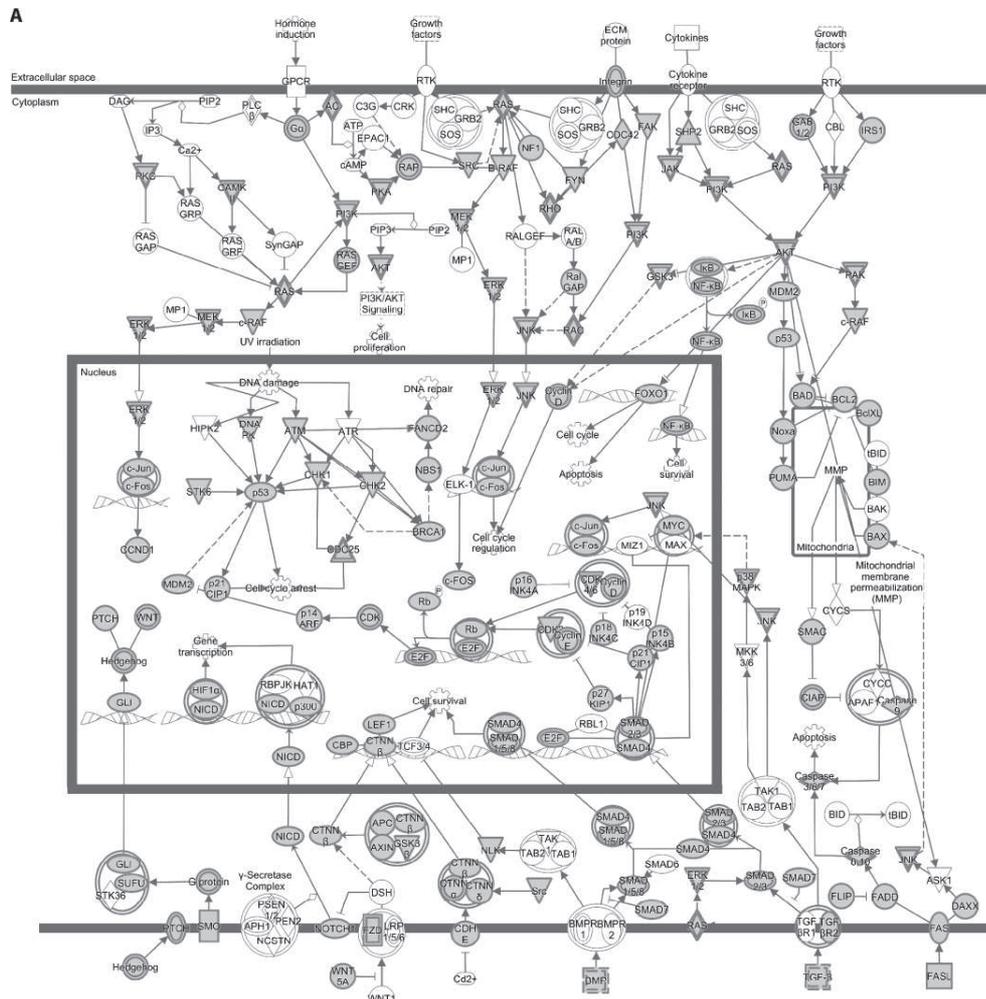
associados a outros tipos de tumores(87-94).

Pela abordagem proposta, foram obtidos inicialmente 1611 genes ligados a CECP que, após curação manual, resultaram em 1190 genes, aos quais foram adicionados 180 de 517 genes previamente selecionados por dois bancos de dados em câncer de cabeça e pescoço (HNOCDDB e OrCGDB). Considerando este resultado, a precisão (especificidade) da abordagem empregada foi estimada em 74% e a sensibilidade em 87%. Embora estes valores sejam satisfatórios, será necessário melhorar a performance da abordagem, dado que nem todos os genes extraídos pelo método foram relevantes após a curação manual. Além disso, vários genes relevantes foram perdidos, como verificado na análise dos dois bancos externos. Para minimizar a perda de informação, estão previstas atualizações da base de dados, nas quais serão incluídas metodologias de processamento de linguagens natural (PLN), assim como buscas em outras fontes, como *Google Scholar*, *Web of Science* e *Scopus*, além do PubMed.

Os resultados obtidos mostraram que a principal via metabólica congregando os 1370 marcadores potenciais relacionados a CECP é a de Mecanismos Moleculares do Câncer (Figura 7A). Também permitiram a identificação de associações ainda não exploradas. Por exemplo, a análise pelo DAVID revelou que processos relacionados a resposta a hormônios esteroides foram significativamente enriquecidos na lista de genes e o programa *Ingenuity* também mostrou que o beta-estradiol está entre os principais reguladores (Figura 7B).

Poucos estudos exploraram as vias metabólicas envolvidas na resposta a esteróides em CECP. Egloff e colaboradores (95) observaram que o estrogênio induz a ativação de membros da família da proteína quinase ativada por mitógeno (MAPK) em linhagens de células de carcinoma epidermoide. Os autores também relataram evidências de interação entre o receptor de estrogênio e o receptor do factor de crescimento epidérmico em CECP. Por sua vez, Brooks e colaboradores (96) verificaram que o aumento dos níveis do receptor β de estrogênio promove a expressão de *NOTCH1* e a diferenciação de células de CECP tanto *in vitro* como *in vivo*. Estes resultados demonstram que a integração de múltiplos tipos de dados expande muito as possibilidades de investigação, fornecendo associações potenciais para serem testadas.

Figura 7. Redes biológicas de acordo com o programa *Ingenuity* para 1370 genes relacionados a CECP. (A) Via canônica "Mecanismos Moleculares do Câncer". As vias de sinalização iniciadas por receptores são induzidas por hormônios, fatores de crescimento, proteínas da matriz extracelular ou citocinas, e levam à ativação de diferentes efetores citoplasmáticos e fatores de transcrição, que resulta em regulação do ciclo celular, reparo do DNA, apoptose, sobrevivência e proliferação celular. (B) Beta-estradiol como regulador a montante da tumorigênese de cabeça e pescoço. Rede enriquecida de fatores de transcrição / reguladores (*ESR1* e *2*, *JUN*, *TP53*, *NFKB1*, *RELA*, *STAT1* e *3*, *SP1*), transdutores de sinais (*Smad3* e *4*), proteína de reparo do DNA (*BRCA1*) e citocinas pró-inflamatórias (*IL6*, *TGFb1*, *TNF*) identificados pelo presente estudo. Setas sólidas = interações conhecidas; setas pontilhadas = interações indiretas; triângulo invertido = quinase; círculo duplo = complexo proteico; círculo, elipse, diamante ou quadrado = outros; formas cheias = genes identificados; formas vazias = genes que fazem parte da rede, mas não foram identificados pelo presente trabalho.



Além da busca de marcadores relacionados a CECP, o presente trabalho investigou ligantes para alvos moleculares selecionados. Para isto, o papel da piplartina, uma substância natural extraída da pimenta *Piper longum* com evidências de ação anti-inflamatória e antineoplásica, foi avaliado na proliferação, na migração e na expressão gênica de células neoplásicas. Os resultados mostraram que a piplartina reduz a viabilidade e a migração celular, sem afetar a capacidade de invasão das células. A piplartina também modificou o padrão de expressão de um painel de genes relacionados com processos inflamatórios. Entre os genes que apresentaram diminuição de expressão após o tratamento com o composto estão receptores ou enzimas da biossíntese de leucotrienos (*LTA4H*, *LTB4R*, *LTB4R2*), receptor de prostaglandina (*PTGER2*), interleucina (*IL2RB*), adrenérgico (*ADRB1*) e de membrana acoplados a proteína G (*BDKRB1*, *HRH1*, *TBXA2R*) ou envolvido em adesão celular (*ITGB1*), efetor de resposta inflamatória (*TNFRSF1A*), reguladores ou membros de vias de sinalização (*A2M*, *MAPK1*, *MAPK8*, *MAPK14*, *PDE4B*, *PDE4D*, *PLCB3*, *PLCE1*, *PLCG1*, *PLCG2*), reguladores de transcrição (*NFKB1*, *NR3C1*) e subunidades de canal de cálcio (*CACNB2*, *CACNB4*).

A redução de expressão desses genes confirma a atuação da piplartina como substância capaz de regular a síntese ou a atividade de importantes membros da reação inflamatória. Os leucotrienos (LTs) e as prostaglandinas (PGs) são dois bons exemplos. Estes compostos são representantes da classe de lipídeos eicosanóides que derivam do ácido araquidônico liberado da membrana pela fosfolipase A2, e posteriormente oxigenados pelas vias da lipooxigenase (LOX) e da ciclooxigenase (COX), respectivamente (97). Tanto os leucotrienos como as prostaglandinas atuam em homeostase e inflamação através de receptores acoplados a proteína G e podem ser

bloqueados por drogas anti-inflamatórias não-esteroidais (98). As prostaglandinas são sintetizadas pela maioria das células e atuam de forma autócrina e parácrina, enquanto os leucotrienos, tanto os cisteinil (cys-LTs) como o LTB₄, são gerados por células inflamatórias após o estímulo desencadeado por fatores exógenos ou eventos de fosforilação intracelular (99, 100).

A sinalização pró-inflamatória iniciada por LTB₄ através de seus receptores BLT1 e BLT2 (codificados pelos genes *LTB4R* e *LTB4R2*) tem sido associada a várias doenças, como asma, artrite reumatoide, aterosclerose, aneurisma aórtico abdominal, esclerose múltipla e câncer. A sinalização dos receptores cys-LT, por sua vez, tem emergido como um componente fundamental da inflamação vascular, com um importante papel na patogênese e na progressão de doenças cardiovasculares, como aterosclerose, infarto do miocárdio e estenose aórtica (revisto por (101)).

As prostaglandinas, em especial a PGE₂, também estão envolvidas na patofisiologia de doenças vasculares e neurodegenerativas. Através da ativação de seus receptores EP1 e EP3, a PGE₂ induz vasoconstrição, que é mediada pela via de cálcio intracelular ou pela redução da concentração de AMP cíclico (cAMP). Por outro lado, a estimulação dos receptores EP2 e EP4 aumenta os níveis de cAMP e provoca vasodilatação (revisto por (97)). Estudos *in vitro* têm demonstrado que a via de sinalização desencadeada por prostaglandinas e seu receptor EP2 em presença de lipopolissacarídeos está associada tanto a efeitos anti-inflamatórios, com redução da produção de interleucina, como a efeitos pró-inflamatórios, e que estas respostas dependem não somente do tipo de estímulo, mas também do tipo de receptor e da célula imune em que foram ativadas (revisto por (102)).

As prostaglandinas têm sido muito estudadas em câncer e altos níveis de PGE₂

já foram observados em vários subtipos desta doença. Como mediador de inflamação, a PEG2 está envolvida em crescimento e progressão tumoral e, juntamente com LTB₄, pode estimular vias de sinalização que envolvem fosfolipases (PLCs), fosfodiesterases (PDEs), nucleotídeos cíclicos, inositol trifosfato (IP₃), canais de cálcio, quinases fosfatidilinositol 3 (PI3K) e MAPK, resultando em diferentes eventos, como proliferação, angiogênese, migração, invasão e sobrevida (103, 104). Vários destes efetores tiveram sua expressão reduzida pela piplartina. Além destes, outras proteínas pró-inflamatórias foram igualmente inibidas, como aquelas que atuam como receptores e ligadores de interleucinas (codificadas pelos genes *A2M*, *IL2RB* e *TNFRSF1A*) e de fatores de transcrição (codificados pelos genes *NFKB1* e *NR3C1*), todos reconhecidamente promotores de inflamação. Em especial, os genes das fosfolipases C estiveram bastante representados entre aqueles inibidos pela piplartina. Estas enzimas fazem parte de uma família de seis isoformas (β , γ , δ , ϵ , ζ and η), que catalisam a hidrólise do fosfatidilinositol 4,5 bifosfato em dois mensageiros secundários, diacilglicerol e inositol 1,4,5-trifosfato. Recentemente, Cui e colaboradores (105) observaram que a elevação da expressão da PLCE1 em pacientes com carcinoma de esôfago está associada com metástase em linfonodos e com estadiamento da lesão. Confirmando seu papel em agressividade tumoral, um estudo realizado com linhagem celular de câncer de bexiga demonstrou que o silenciamento desta enzima levou à redução dos níveis de metaloproteinases e do regulador de apoptose Bcl-2, e conseqüentemente da capacidade de invasão das células (106).

Outra proteína que tem sido relacionada a doenças inflamatórias (107), mas também participa do processo e da progressão neoplásica (108), é a integrina beta-1, cuja expressão no nível de transcritos foi reduzida pela piplartina no presente estudo.

A explicação para a diminuição de expressão de vários fatores pró-inflamatórios pela piplartina é relativamente simples, considerando que este composto possui valores de energia livre comparáveis a algumas drogas anti-inflamatórias bastante utilizadas. No caso dos genes que mostraram expressão elevada nas mesmas condições, a explicação pode residir em atividades antagônicas pró- e anti-inflamatórias dos produtos gênicos, como já relatado pela literatura, ou em *feedback* negativo.

Os genes de duas fosfolipases (*PLA2G2A* e *PLA2G5*), do fator de necrose tumoral (*TNF*), receptores de interleucinas e citocinas (*CD40*, *IL1RL1*, *IL2RA* e *IL1R2*) e leucotrieno (*CYSLTR1*) tiveram sua expressão elevada após o tratamento. A fosfolipase A2 grupo IIA secretada (*PLA2G2A*) é uma enzima encontrada em processos inflamatórios (109, 110) e em doenças malignas (111, 112) associada com progressão da doença (113). Neste caso, a piplartina estaria favorecendo e não inibindo mecanismos favoráveis à inflamação. Entretanto, o papel desta fosfolipase em processos inflamatórios já foi questionado e levantada a hipótese de que sua expressão elevada poderia ter um aspecto protetor (114).

Como esperado para o efeito observado em presença da piplartina, a integrina alfa-M (codificada pelo gene *ITGAM*) regula negativamente a resposta imune desencadeada por receptores *Toll-like* (115) de modo tipo celular-específico (116). Confirmando estes dados, os estudos de Han e colaboradores (2010) *in-vitro* e *in-vivo* mostraram que a deficiência desta integrina resulta em maior produção de citocinas pró-inflamatórias e interferon do tipo I induzida por receptores *Toll-like* e que, portanto, em níveis elevados é capaz de inibir esta via de sinalização (115).

Um comportamento mostrando efeitos antagônicos dependendo do contexto é também ser esperado para algumas citocinas inflamatórias, como as interleucinas IL-1 e

IL-18, que exercem sua atividade sobre as células alvos ligando-se a receptores específicos. Em função da necessidade de controle rigoroso para que não ocorram danos teciduais e, conseqüentemente, desenvolvimento de doenças inflamatórias e autoimunes severas (117, 118), a família IL-1 possui tanto ativadores como supressores de inflamação. Por exemplo, o produto do gene *IL1R2* (induzido pela piplartina no presente estudo) funciona como um receptor *decoy*, neutralizando seus ligantes IL-1 β e IL-1 α e inibindo o sinal (117). O papel anti-inflamatório do *IL1R2* foi confirmado por estudos *in vivo* em modelo murino transgênico com expressão constitutiva de *IL1R2*, que foi incapaz de exibir inflamação crônica induzida (119). A transfecção de células com este gene também reduziu parâmetros clínicos e histológicos de artrite (120), diminuiu a capacidade de tecido endometrial de invadir, crescer, disseminar e estimular angiogênese (121) e atenuou rejeição a alotransplante cardíaco em ratos (122).

A piplartina também estimulou a expressão do fator de necrose tumoral (TNF), uma citocina presente na membrana plasmática predominantemente de macrófagos e linfócitos T. O TNF desempenha papel central na patogênese de algumas doenças inflamatórias e tem sido muito estudado por mediar processos biológicos importantes como sobrevivência, apoptose e comunicação entre células (123, 124). A desregulação de tais processos é uma característica de doenças inflamatórias e câncer.

A via de transdução de sinais do TNF é complexa. As respostas a este fator são desencadeadas pela ativação de um de seus receptores, TNFR1 e TNFR2 (codificados pelos genes *TNFRSF1A* e *TNFRSF1B*, respectivamente). Os domínios extracelulares destes receptores são homólogos e apresentam similar afinidade pelo TNF, mas as regiões citoplasmáticas são distintas e ativam eventos de sinalização com diferentes efeitos biológicos. Enquanto o TNFR1 contém o domínio morte, o TNFR2 perde este

domínio.

A ligação TNF-TNFR1 induz uma modificação conformacional do seu domínio citoplasmático, que recruta a proteína adaptadora TRADD (*TNFRSF1A-associated via death domain*) e a quinase de serina/treonina RIPK1 (*cell death protein RIP*) iniciando a formação do complexo I com ligação de TRAF2 (*TNF receptor-associated factor 2*) e das proteínas inibidoras de apoptose-1 e 2 (cIAP1 e 2). Nesta via, a proteína RIPK1 se comporta como um interruptor para sobrevivência ou morte celular dependendo de seu estado de ubiquitinação. Quando ubiquitinada, liga-se a outros membros da cascata, o que resulta na ativação da sinalização promovida pelo fator de transcrição NFκB, levando à síntese de fatores anti-apoptóticos como FLIP_L/CFLAR (*FLICE-like inhibitory protein*) e BCL-XL (*Bcl2-associated agonist of cell death*). Após sua desubiquitinação, ou previamente à sua ubiquitinação inicial, RIPK2 forma o complexo IIa com as proteínas TRADD, FADD (*FAS-associated death domain protein*), procaspase 8 e a proteína inibidora FLIP_L, ou forma o complexo IIb com as mesmas proteínas e com RIPK3 (mas não TRADD). Tanto o complexo IIa como o IIb resultam em sinais de sobrevivência e necrose. Para tanto, RIPK1 e 3 precisam ser inativados pela caspase 8, que é então liberada dos complexos e induz apoptose clássica. Ambas as vias desencadeadas por TNF-TNFR1 e TNF-TNFR2 através de TRAF2 podem levar à ativação de NFκB, mas a via TNFR2 promove sobrevivência e TNFR1-TRADD resulta em sobrevivência ou morte, como descrito acima, dependendo da sinalização a jusante e do contexto celular. Ainda não está claro como as vias TNFR1 e TNFR2 são reguladas, exceto pela expressão diferencial dos dois receptores na superfície das células (revisado por (125)).

O receptor TNFR1 mostrou níveis reduzidos após tratamento com a piplartina, enquanto o TNFR2 não exibiu mudança de perfil. Este dado sugere que a razão TNFR1/TNFR2 diminuiu em relação ao controle, o que pode alterar o destino da célula. Entretanto, uma conclusão definitiva sobre o resultado final não pode ser atingida, considerando que outros membros da via são importantes. Por exemplo, em células com atividade satisfatória do complexo I, a sinalização NF κ B induz a expressão de FLIP_L, cujos altos níveis inibem a caspase 8 e a apoptose (126, 127). Entretanto, é improvável que o programa anti-apoptótico de NF κ B foque apenas FLIP_L porque esta via controla a transcrição de múltiplos genes que atuam no bloqueio da morte celular (128).

Embora a literatura tenha amplamente revelado atuação direta de TNF no controle da via do NF κ B, alguns dados recentes mostram uma nova função deste fator, como regulador da geração de espécies reativas de oxigênio (ROS) e nitrogênio (RNS). Tanto ROS como RNS podem mudar o destino da célula de sobrevivência para morte. A produção de NOS, por exemplo, que se mostrou aumentada após tratamento com piplartina, leva à geração de óxido nítrico (NO) a partir de arginina. O NO tende a difundir-se pela membrana e favorecer a morte fagocítica de microorganismos no espaço extracelular, promovendo também inflamação. O NO intracelular interage com ROS e resulta na formação de altos níveis de RNS, que causam lesões no DNA e induzem apoptose ou necrose (revisado por (129)). Estes efeitos antagônicos evidenciam a complexidade do processo inflamatório e da via do TNF e mostram que o desenvolvimento de mecanismos eficientes de controle é ainda um grande desafio.

6. CONCLUSÕES

As principais conclusões obtidas no presente trabalho estão listadas a seguir:

1. A abordagem empregada no presente estudo para mineração da literatura apresenta bom desempenho na identificação e na recuperação de genes relacionados a CECP, com valores de sensibilidade e especificidade elevados.
2. Os marcadores relacionados a CECP identificados pela abordagem proposta não são específicos para este tumor e também estão associados a outros tipos de neoplasias.
3. Os marcadores relacionados a CECP estão ligados, entre outros, a processos envolvidos na resposta a hormônios esteroides.
4. A integração de múltiplos tipos de dados expande muito as possibilidades de investigação e revela associações potenciais para serem testadas.
5. A piplartina é um ligante que apresenta capacidade de interagir com alvos moleculares de forma semelhante à de medicamentos anti-inflamatórios utilizados atualmente na clínica médica.
6. A piplartina é capaz de diminuir a proliferação e a migração celular, mas não a invasividade de células de carcinoma. Também é capaz de alterar a expressão de genes relacionados a CECP e a processos inflamatórios.

8. REFERÊNCIAS

1. Argiris A, Karamouzis MV, Johnson JT, Heron DE, Myers E, Eibling D, et al. Long-term results of a phase III randomized trial of postoperative radiotherapy with or without carboplatin in patients with high-risk head and neck cancer. *The Laryngoscope*. 2008;118(3):444-9. Epub 2007/12/20.
2. IARC. Globocan. <http://globocan.iarc.fr/Default.aspx>; 2015; Available from: <http://globocan.iarc.fr/Default.aspx>.
3. Brasil. Estimativa 2014 - Incidência de Câncer no Brasil. Instituto Nacional do Câncer - INCA; 2014.
4. Paterson IC, Eveson JW, Prime SS. Molecular changes in oral cancer may reflect aetiology and ethnic origin. *European journal of cancer Part B, Oral oncology*. 1996;32B(3):150-3. Epub 1996/05/01.
5. Hecht SS. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nature reviews Cancer*. 2003;3(10):733-44. Epub 2003/10/23.
6. Hashibe M, Brennan P, Chuang SC, Boccia S, Castellsague X, Chen C, et al. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2009;18(2):541-50. Epub 2009/02/05.
7. zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. *Nature reviews Cancer*. 2002;2(5):342-50. Epub 2002/06/05.
8. Fotopoulos G, Pavlidis N. The role of human papilloma virus and p16 in occult primary of the head and neck: a comprehensive review of the literature. *Oral oncology*. 2015;51(2):119-23. Epub 2014/12/04.
9. Weinberger PM, Yu Z, Haffty BG, Kowalski D, Harigopal M, Sasaki C, et al. Prognostic significance of p16 protein levels in oropharyngeal squamous cell cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2004;10(17):5684-91. Epub 2004/09/10.
10. Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. *Nature reviews Cancer*. 2011;11(1):9-22. Epub 2010/12/17.
11. Jin C, Jin Y, Wennerberg J, Annertz K, Enoksson J, Mertens F. Cytogenetic abnormalities in 106 oral squamous cell carcinomas. *Cancer genetics and cytogenetics*. 2006;164(1):44-53. Epub 2005/12/21.
12. Wang C, Wang J, Ju Z, Zhai R, Zhou L, Li Q, et al. Reconstruction of metabolic network in the bovine mammary gland tissue. *Molecular biology reports*. 2012;39(7):7311-8. Epub 2012/02/14.
13. Wang D, Song H, Evans JA, Lang JC, Schuller DE, Weghorst CM. Mutation and downregulation of the transforming growth factor beta type II receptor gene in primary squamous cell carcinomas of the head and neck. *Carcinogenesis*. 1997;18(11):2285-90. Epub 1997/12/12.
14. Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer cell*. 2004;5(5):489-500. Epub 2004/05/18.
15. Jia WH, Qin HD. Non-viral environmental risk factors for nasopharyngeal

- carcinoma: a systematic review. *Seminars in cancer biology*. 2012;22(2):117-26. Epub 2012/02/09.
16. Martone T, Gillio-Tos A, De Marco L, Fiano V, Maule M, Cavalot A, et al. Association between hypermethylated tumor and paired surgical margins in head and neck squamous cell carcinomas. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2007;13(17):5089-94. Epub 2007/09/06.
 17. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. *Nature*. 2008;454(7203):436-44. Epub 2008/07/25.
 18. Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet*. 2001;357(9255):539-45. Epub 2001/03/07.
 19. Diakos CI, Charles KA, McMillan DC, Clarke SJ. Cancer-related inflammation and treatment effectiveness. *The Lancet Oncology*. 2014;15(11):e493-503. Epub 2014/10/05.
 20. Moore MM, Chua W, Charles KA, Clarke SJ. Inflammation and cancer: causes and consequences. *Clinical pharmacology and therapeutics*. 2010;87(4):504-8. Epub 2010/02/12.
 21. Gabriel SE, Jaakkimainen L, Bombardier C. Risk for serious gastrointestinal complications related to use of nonsteroidal anti-inflammatory drugs. A meta-analysis. *Annals of internal medicine*. 1991;115(10):787-96. Epub 1991/11/15.
 22. Allison MC, Howatson AG, Torrance CJ, Lee FD, Russell RI. Gastrointestinal damage associated with the use of nonsteroidal antiinflammatory drugs. *The New England journal of medicine*. 1992;327(11):749-54. Epub 1992/09/10.
 23. Hawkey CJ. Nonsteroidal anti-inflammatory drug gastropathy. *Gastroenterology*. 2000;119(2):521-35. Epub 2000/08/10.
 24. Aygun D, Kaplan S, Odaci E, Onger ME, Altunkaynak ME. Toxicity of non-steroidal anti-inflammatory drugs: a review of melatonin and diclofenac sodium association. *Histology and histopathology*. 2012;27(4):417-36. Epub 2012/03/01.
 25. O'Brien SG, Guilhot F, Larson RA, Gathmann I, Baccarani M, Cervantes F, et al. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *The New England journal of medicine*. 2003;348(11):994-1004. Epub 2003/03/15.
 26. Cartron G, Trappe RU, Solal-Celigny P, Hallek M. Interindividual variability of response to rituximab: from biological origins to individualized therapies. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2011;17(1):19-30. Epub 2011/01/07.
 27. Tsang RY, Finn RS. Beyond trastuzumab: novel therapeutic strategies in HER2-positive metastatic breast cancer. *British journal of cancer*. 2012;106(1):6-13. Epub 2012/01/05.
 28. Bebek G, Orloff M, Eng C. Microenvironmental genomic alterations reveal signaling networks for head and neck squamous cell carcinoma. *Journal of clinical bioinformatics*. 2011;1(1):21. Epub 2011/09/03.
 29. Califano J, van der Riet P, Westra W, Nawroz H, Clayman G, Piantadosi S, et al. Genetic progression model for head and neck cancer: implications for field cancerization. *Cancer research*. 1996;56(11):2488-92. Epub 1996/06/01.
 30. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-74. Epub 2011/03/08.
 31. DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, et al. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature genetics*.

- 1996;14(4):457-60. Epub 1996/12/01.
32. Morozova O, Hirst M, Marra MA. Applications of new sequencing technologies for transcriptome analysis. *Annual review of genomics and human genetics*. 2009;10:135-51. Epub 2009/09/01.
 33. Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, et al. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proceedings of the National Academy of Sciences of the United States of America*. 1996;93(12):6025-30. Epub 1996/06/11.
 34. Jensen ON. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Current opinion in chemical biology*. 2004;8(1):33-41. Epub 2004/03/24.
 35. Fels LM, Buschmann T, Meuer J, Reymond MA, Lamer S, Rocken C, et al. Proteome analysis for the identification of tumor-associated biomarkers in gastrointestinal cancer. *Dig Dis*. 2003;21(4):292-8. Epub 2004/01/31.
 36. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57-70. Epub 2000/01/27.
 37. Geromichalos GD. Importance of molecular computer modeling in anticancer drug development. *Journal of BUON : official journal of the Balkan Union of Oncology*. 2007;12 Suppl 1:S101-18. Epub 2007/11/06.
 38. Pitchai D, Manikkam R, V SL, Singaram R. Structure based virtual screening of anticancerous polyphenolic phytochemicals against G-protein coupled receptor and identification of potent antagonist ligand(s) through molecular docking. *Bioinformation*. 2011;6(6):226-8. Epub 2011/07/09.
 39. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. *Medicinal research reviews*. 2006;26(5):531-68. Epub 2006/06/08.
 40. Satyanarayanajois S, Villalba S, Jianchao L, Lin GM. Design, synthesis, and docking studies of peptidomimetics based on HER2-herceptin binding site with potential antiproliferative activity against breast cancer cell lines. *Chemical biology & drug design*. 2009;74(3):246-57. Epub 2009/08/26.
 41. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research*. 2002;30(1):207-10. Epub 2001/12/26.
 42. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004;6(1):1-6. Epub 2004/04/08.
 43. Uhlen M, Bjorling E, Agaton C, Szigyarto CA, Amini B, Andersen E, et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Molecular & cellular proteomics : MCP*. 2005;4(12):1920-32. Epub 2005/08/30.
 44. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research*. 2008;36(Database issue):D901-6. Epub 2007/12/01.
 45. Morgan AA, Lu Z, Wang X, Cohen AM, Fluck J, Ruch P, et al. Overview of BioCreative II gene normalization. *Genome biology*. 2008;9 Suppl 2:S3. Epub 2008/10/18.
 46. Fleuren WW, Alkema W. Application of text mining in the biomedical domain. *Methods*. 2015;74:97-106. Epub 2015/02/03.
 47. Jensen LJ, Saric J, Bork P. Literature mining for the biologist: from information

- retrieval to biological discovery. *Nature reviews Genetics*. 2006;7(2):119-29. Epub 2006/01/19.
48. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*. 2000;25(1):25-9. Epub 2000/05/10.
49. Mao Y, Van Auken K, Li D, Arighi CN, McQuilton P, Hayman GT, et al. Overview of the gene ontology task at BioCreative IV. *Database : the journal of biological databases and curation*. 2014;2014. Epub 2014/08/27.
50. NCBI. MeSH. <http://www.ncbi.nlm.nih.gov/mesh>. 2015 [cited 2015 29/06]; Available from: <http://www.ncbi.nlm.nih.gov/mesh>.
51. Xiang Z, Qin T, Qin ZS, He Y. A genome-wide MeSH-based literature mining system predicts implicit gene-to-gene relationships and networks. *BMC systems biology*. 2013;7 Suppl 3:S9. Epub 2014/02/22.
52. Zhang Y, Sarkar IN, Chen ES. PubMedMiner: Mining and Visualizing MeSH-based Associations in PubMed. *AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium*. 2014;2014:1990-9. Epub 2014/01/01.
53. gene2pubmed. <ftp://ftp.ncbi.nih.gov/gene/DATA/gene2pubmed.gz> [database on the Internet]. 2015 [cited 29/06/2015]. Available from: <ftp://ftp.ncbi.nih.gov/gene/DATA/gene2pubmed.gz>.
54. Mitra S, Das S, Ghosal S, Chakrabarti J. HNOCDDB: a comprehensive database of genes and miRNAs relevant to head and neck and oral cancer. *Oral oncology*. 2012;48(2):117-9. Epub 2011/10/26.
55. Gadewal NS, Zingde SM. Database and interaction network of genes involved in oral cancer: Version II. *Bioinformatics*. 2011;6(4):169-70. Epub 2011/05/17.
56. Jourquin J, Duncan D, Shi Z, Zhang B. GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC genomics*. 2012;13 Suppl 8:S20. Epub 2013/01/11.
57. Mermin N. Stirling's formula! *American Journal of Physics*. 1984;52:362-5.
58. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 2000;28(1):27-30. Epub 1999/12/11.
59. Reactome [database on the Internet]. 2015 [cited 29/06/2015]. Available from: <http://www.reactome.org/pages/download-data/>.
60. Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Current protocols in human genetics / editorial board, Jonathan L Haines [et al]*. 2008;Chapter 10:Unit 10 1. Epub 2008/04/23.
61. Online Mendelian Inheritance in Man, OMIM [database on the Internet]. 2015 [cited 29/06/2015]. Available from: <http://omim.org/>.
62. Ye H, Yu T, Temam S, Ziober BL, Wang J, Schwartz JL, et al. Transcriptomic dissection of tongue squamous cell carcinoma. *BMC genomics*. 2008;9:69. Epub 2008/02/08.
63. Kuriakose MA, Chen WT, He ZM, Sikora AG, Zhang P, Zhang ZY, et al. Selection and validation of differentially expressed genes in head and neck cancer. *Cellular and molecular life sciences : CMLS*. 2004;61(11):1372-83. Epub 2004/06/02.
64. Schlingemann J, Habtemichael N, Ittrich C, Toedt G, Kramer H, Hambek M, et al. Patient-based cross-platform comparison of oligonucleotide microarray expression profiles. *Laboratory investigation; a journal of technical methods and pathology*. 2005;85(8):1024-39. Epub 2005/10/06.
65. Estilo CL, P Oc, Talbot S, Socci ND, Carlson DL, Ghossein R, et al. Oral tongue

- cancer gene expression profiling: Identification of novel potential prognosticators by oligonucleotide microarray analysis. *BMC cancer*. 2009;9:11. Epub 2009/01/14.
66. Toruner GA, Ulger C, Alkan M, Galante AT, Rinaggio J, Wilk R, et al. Association between gene expression profile and tumor invasion in oral squamous cell carcinoma. *Cancer genetics and cytogenetics*. 2004;154(1):27-35. Epub 2004/09/24.
67. Peng CH, Liao CT, Peng SC, Chen YJ, Cheng AJ, Juang JL, et al. A novel molecular signature identified by systems genetics approach predicts prognosis in oral squamous cell carcinoma. *PloS one*. 2011;6(8):e23452. Epub 2011/08/20.
68. Ginos MA, Page GP, Michalowicz BS, Patel KJ, Volker SE, Pambuccian SE, et al. Identification of a gene expression signature associated with recurrent disease in squamous cell carcinoma of the head and neck. *Cancer research*. 2004;64(1):55-63. Epub 2004/01/20.
69. Cromer A, Carles A, Millon R, Ganguli G, Chalmel F, Lemaire F, et al. Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis. *Oncogene*. 2004;23(14):2484-98. Epub 2003/12/17.
70. Kapushesky M, Adamusiak T, Burdett T, Culhane A, Farne A, Filippov A, et al. Gene Expression Atlas update--a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic acids research*. 2012;40(Database issue):D1077-81. Epub 2011/11/09.
71. Petryszak R, Burdett T, Fiorelli B, Fonseca NA, Gonzalez-Porta M, Hastings E, et al. Expression Atlas update--a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic acids research*. 2014;42(Database issue):D926-32. Epub 2013/12/07.
72. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic acids research*. 2000;28(1):235-42. Epub 1999/12/11.
73. Prasad SM, Massad MG, Chedrawy EG, Snow NJ, Yeh JT, Lele H, et al. Weathering the storm: how can thoracic surgery training programs meet the new challenges in the era of less-invasive technologies? *The Journal of thoracic and cardiovascular surgery*. 2009;137(6):1317-25, discussion 26. Epub 2009/05/26.
74. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic acids research*. 2009;37(Database issue):D412-6. Epub 2008/10/23.
75. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic acids research*. 2014;42(Database issue):D1091-7. Epub 2013/11/10.
76. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic acids research*. 2011;39(Database issue):D1035-41. Epub 2010/11/10.
77. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research*. 2006;34(Database issue):D668-72. Epub 2005/12/31.
78. Ensembl Project. <http://www.ensembl.org/> [database on the Internet]. 2015 [cited 29/06/2015]. Available from: <http://www.ensembl.org/>.
79. UniProt: a hub for protein information. *Nucleic acids research*. 2015;43(Database issue):D204-12. Epub 2014/10/29.
80. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*. 2009;4(1):44-

57. Epub 2009/01/10.
81. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*. 2009;37(1):1-13. Epub 2008/11/27.
82. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*. 2010;31(2):455-61. Epub 2009/06/06.
83. Boyden S. The chemotactic effect of mixtures of antibody and antigen on polymorphonuclear leucocytes. *The Journal of experimental medicine*. 1962;115:453-66. Epub 1962/03/01.
84. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*. 2001;46(1-3):3-26. Epub 2001/03/22.
85. McCall MN, Jaffee HA, Zelisko SJ, Sinha N, Hooiveld G, Irizarry RA, et al. The Gene Expression Barcode 3.0: improved data processing and mining tools. *Nucleic acids research*. 2014;42(Database issue):D938-43. Epub 2013/11/26.
86. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery*. 2012;2(5):401-4. Epub 2012/05/17.
87. Musgrove EA, Caldon CE, Barraclough J, Stone A, Sutherland RL. Cyclin D as a therapeutic target in cancer. *Nature reviews Cancer*. 2011;11(8):558-72. Epub 2011/07/08.
88. Witkiewicz AK, Knudsen KE, Dicker AP, Knudsen ES. The meaning of p16(ink4a) expression in tumors: functional significance, clinical associations and future developments. *Cell Cycle*. 2011;10(15):2497-503. Epub 2011/07/22.
89. Sharma SV, Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. *Nature reviews Cancer*. 2007;7(3):169-81. Epub 2007/02/24.
90. Gherardi E, Birchmeier W, Birchmeier C, Vande Woude G. Targeting MET in cancer: rationale and progress. *Nature reviews Cancer*. 2012;12(2):89-103. Epub 2012/01/25.
91. Bader AG, Kang S, Zhao L, Vogt PK. Oncogenic PI3K deregulates transcription and translation. *Nature reviews Cancer*. 2005;5(12):921-9. Epub 2005/12/13.
92. Hollander MC, Blumenthal GM, Dennis PA. PTEN loss in the continuum of common cancers, rare syndromes and mouse models. *Nature reviews Cancer*. 2011;11(4):289-301. Epub 2011/03/25.
93. Kin C, Kidess E, Poultsides GA, Visser BC, Jeffrey SS. Colorectal cancer diagnostics: biomarkers, cell-free DNA, circulating tumor cells and defining heterogeneous populations by single-cell analysis. *Expert review of molecular diagnostics*. 2013;13(6):581-99. Epub 2013/07/31.
94. Robles AI, Harris CC. Clinical outcomes and correlates of TP53 mutations and cancer. *Cold Spring Harbor perspectives in biology*. 2010;2(3):a001016. Epub 2010/03/20.
95. Egloff AM, Rothstein ME, Seethala R, Siegfried JM, Grandis JR, Stabile LP. Cross-talk between estrogen receptor and epidermal growth factor receptor in head and neck squamous cell carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2009;15(21):6529-40. Epub 2009/10/15.
96. Brooks YS, Ostano P, Jo SH, Dai J, Getsios S, Dziunycz P, et al. Multifactorial

- ERbeta and NOTCH1 control of squamous differentiation and cancer. *The Journal of clinical investigation*. 2014;124(5):2260-76. Epub 2014/04/20.
97. Gomez I, Foudi N, Longrois D, Norel X. The role of prostaglandin E2 in human vascular inflammation. *Prostaglandins, leukotrienes, and essential fatty acids*. 2013;89(2-3):55-63. Epub 2013/06/13.
98. Funk CD. Prostaglandins and leukotrienes: advances in eicosanoid biology. *Science*. 2001;294(5548):1871-5. Epub 2001/12/01.
99. Kanaoka Y, Boyce JA. Cysteinyl leukotrienes and their receptors; emerging concepts. *Allergy, asthma & immunology research*. 2014;6(4):288-95. Epub 2014/07/06.
100. Horn T, Adel S, Schumann R, Sur S, Kakularam KR, Polamarasetty A, et al. Evolutionary aspects of lipoxygenases and genetic diversity of human leukotriene signaling. *Progress in lipid research*. 2015;57:13-39. Epub 2014/12/02.
101. Back M, Powell WS, Dahlen SE, Drazen JM, Evans JF, Serhan CN, et al. Update on leukotriene, lipoxin and oxoecosanoid receptors: IUPHAR Review 7. *British journal of pharmacology*. 2014;171(15):3551-74. Epub 2014/03/05.
102. Johansson JU, Pradhan S, Lokteva LA, Woodling NS, Ko N, Brown HD, et al. Suppression of inflammation with conditional deletion of the prostaglandin E2 EP2 receptor in macrophages and brain microglia. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2013;33(40):16016-32. Epub 2013/10/04.
103. Wang D, Dubois RN. Eicosanoids and cancer. *Nature reviews Cancer*. 2010;10(3):181-93. Epub 2010/02/20.
104. Rundhaug JE, Simper MS, Surh I, Fischer SM. The role of the EP receptors for prostaglandin E2 in skin and skin cancer. *Cancer metastasis reviews*. 2011;30(3-4):465-80. Epub 2011/10/21.
105. Cui XB, Pang XL, Li S, Jin J, Hu JM, Yang L, et al. Elevated expression patterns and tight correlation of the PLCE1 and NF-kappaB signaling in Kazakh patients with esophageal carcinoma. *Med Oncol*. 2014;31(1):791. Epub 2013/12/07.
106. Ou L, Guo Y, Luo C, Wu X, Zhao Y, Cai X. RNA interference suppressing PLCE1 gene expression decreases invasive power of human bladder cancer T24 cell line. *Cancer genetics and cytogenetics*. 2010;200(2):110-9. Epub 2010/07/14.
107. Keely S, Glover LE, MacManus CF, Campbell EL, Scully MM, Furuta GT, et al. Selective induction of integrin beta1 by hypoxia-inducible factor: implications for wound healing. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2009;23(5):1338-46. Epub 2008/12/24.
108. Chen MB, Lamar JM, Li R, Hynes RO, Kamm RD. Elucidation of the roles of tumor integrin ss1 in the extravasation stage of the metastasis cascade. *Cancer research*. 2016. Epub 2016/03/19.
109. Triggiani M, Granata F, Giannattasio G, Marone G. Secretory phospholipases A2 in inflammatory and allergic diseases: not just enzymes. *The Journal of allergy and clinical immunology*. 2005;116(5):1000-6. Epub 2005/11/09.
110. Boudreau LH, Duchez AC, Cloutier N, Soulet D, Martin N, Bollinger J, et al. Platelets release mitochondria serving as substrate for bactericidal group IIA-secreted phospholipase A2 to promote inflammation. *Blood*. 2014;124(14):2173-83. Epub 2014/08/02.
111. Kupert E, Anderson M, Liu Y, Succop P, Levin L, Wang J, et al. Plasma secretory phospholipase A2-IIa as a potential biomarker for lung cancer in patients with solitary pulmonary nodules. *BMC cancer*. 2011;11:513. Epub 2011/12/14.

112. Menschikowski M, Hagelgans A, Schuler U, Froeschke S, Rosner A, Siegert G. Plasma levels of phospholipase A2-IIA in patients with different types of malignancies: prognosis and association with inflammatory and coagulation biomarkers. *Pathology oncology research : POR*. 2013;19(4):839-46. Epub 2013/06/01.
113. Oleksowicz L, Liu Y, Bracken RB, Gaitonde K, Burke B, Succop P, et al. Secretory phospholipase A2-IIa is a target gene of the HER/HER2-elicited pathway and a potential plasma biomarker for poor prognosis of prostate cancer. *The Prostate*. 2012;72(10):1140-9. Epub 2011/12/01.
114. Menschikowski M, Hagelgans A, Siegert G. Secretory phospholipase A2 of group IIA: is it an offensive or a defensive player during atherosclerosis and other inflammatory diseases? *Prostaglandins & other lipid mediators*. 2006;79(1-2):1-33. Epub 2006/03/07.
115. Han C, Jin J, Xu S, Liu H, Li N, Cao X. Integrin CD11b negatively regulates TLR-triggered inflammatory responses by activating Syk and promoting degradation of MyD88 and TRIF via Cbl-b. *Nature immunology*. 2010;11(8):734-42. Epub 2010/07/20.
116. Ling GS, Bennett J, Woollard KJ, Szajna M, Fossati-Jimack L, Taylor PR, et al. Integrin CD11b positively regulates TLR4-induced signalling pathways in dendritic cells but not in macrophages. *Nature communications*. 2014;5:3039. Epub 2014/01/16.
117. Dinarello CA. Immunological and inflammatory functions of the interleukin-1 family. *Annual review of immunology*. 2009;27:519-50. Epub 2009/03/24.
118. Boraschi D, Dinarello CA. IL-18 in autoimmunity: review. *European cytokine network*. 2006;17(4):224-52. Epub 2007/03/14.
119. Rauschmayr T, Groves RW, Kupper TS. Keratinocyte expression of the type 2 interleukin 1 receptor mediates local and specific inhibition of interleukin 1-mediated inflammation. *Proceedings of the National Academy of Sciences of the United States of America*. 1997;94(11):5814-9. Epub 1997/05/27.
120. Bessis N, Guery L, Mantovani A, Vecchi A, Sims JE, Fradelizi D, et al. The type II decoy receptor of IL-1 inhibits murine collagen-induced arthritis. *European journal of immunology*. 2000;30(3):867-75. Epub 2000/03/31.
121. Khoufache K, Bondza PK, Harir N, Daris M, Leboeuf M, Mailloux J, et al. Soluble human IL-1 receptor type 2 inhibits ectopic endometrial tissue implantation and growth: identification of a novel potential target for endometriosis treatment. *The American journal of pathology*. 2012;181(4):1197-205. Epub 2012/07/31.
122. Simeoni E, Dudler J, Fleury S, Li J, Pagnotta M, Pascual M, et al. Gene transfer of a soluble IL-1 type 2 receptor-Ig fusion protein improves cardiac allograft survival in rats. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery*. 2007;31(2):222-8. Epub 2006/12/22.
123. McInnes IB, Buckley CD, Isaacs JD. Cytokines in rheumatoid arthritis - shaping the immunological landscape. *Nature reviews Rheumatology*. 2016;12(1):63-8. Epub 2015/12/15.
124. Wcislo-Dziadecka D, Zbiciak-Nylec M, Brzezinska-Wcislo L, Mazurek U. TNF-alpha in a molecularly targeted therapy of psoriasis and psoriatic arthritis. *Postgraduate medical journal*. 2016;92(1085):172-8. Epub 2016/01/01.
125. Brenner D, Blaser H, Mak TW. Regulation of tumour necrosis factor signalling: live or let die. *Nature reviews Immunology*. 2015;15(6):362-74. Epub 2015/05/27.
126. Micheau O, Lens S, Gaide O, Alevizopoulos K, Tschopp J. NF-kappaB signals induce the expression of c-FLIP. *Molecular and cellular biology*. 2001;21(16):5299-305. Epub 2001/07/21.

127. Golks A, Brenner D, Fritsch C, Krammer PH, Lavrik IN. c-FLIPR, a new regulator of death receptor-induced apoptosis. *The Journal of biological chemistry*. 2005;280(15):14507-13. Epub 2005/02/11.
128. Karin M, Lin A. NF-kappaB at the crossroads of life and death. *Nature immunology*. 2002;3(3):221-7. Epub 2002/03/05.
129. Blaser H, Dostert C, Mak TW, Brenner D. TNF and ROS Crosstalk in Inflammation. *Trends in cell biology*. 2016;26(4):249-61. Epub 2016/01/23.

9. TRABALHO PUBLICADO

10. OUTROS TRABALHOS PUBLICADOS