UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

Instituto de Biociências
de Botucatu - Unesp

IDENTIFICAÇÃO DE POTENCIAIS BIOMARCADORES DE CAQUEXIA NO SECRETOMA DE CARCINOMAS DE PULMÃO DE CÉLULAS NÃO PEQUENAS

SARAH SANTILONI CURY

BOTUCATU – SP

2019

UNIVERSIDADE ESTADUAL PAULISTA
"Júlio de Mesquita Filho"
INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

IDENTIFICAÇÃO DE POTENCIAIS BIOMARCADORES DE CAQUEXIA NO SECRETOMA DE CARCINOMAS DE PULMÃO DE CÉLULAS NÃO PEQUENAS

Mestranda: Sarah Santiloni Cury

Orientador: Robson Francisco Carvalho

Dissertação apresentada ao Instituto de Biociências, Campus de Botucatu, UNESP, como parte dos pré-requisitos necessários para obtenção do título de Mestre no Programa de Pós-Graduação em Ciências Biológicas (Genética).

BOTUCATU – SP

2019

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

Instituto de Biociências de Botucatu - Unesp

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Cury, Sarah Santiloni.
Identificação de potenciais biomarcadores de caquexia no secretoma de carcinomas de pulmão de células não pequenas / Sarah Santiloni Cury. - Botucatu, 2019

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Robson Francisco Carvalho
Capes: 20205007

1. Interleucina-8. 2. Tomografia. 3. Linhagem celular. 4. Mioblastos. 5. Marcadores bioquímicos. 6. Músculos peitorais. 7. Caquexia. 8. Pulmões - Câncer.

Palavras-chave: C2C12; Interleucina-8; Muscularidade; Secretoma; Tomografia computadorizada.

Dedicatória

À minha mãe, *Valquiria*, que
foi meu alicerce e sempre será
o maior amor de minha vida.

Ao meu irmão, *Caio*, que
acreditou em meu potencial e
se orgulhou de cada conquista.

Eu amo vocês.

Agradecimentos

À minha mãe que sempre estimulou meu prazer por estudar, aprender, ler e me informar. Sempre procurou me apoiar, mesmo sabendo que isso causaria nela muitas preocupações. Obrigada por ser essa mãezona e me educar para ser íntegra e honesta. Inspiro-me em você.

Ao meu irmão, por todas as ajudas. Quando a matemática e a informática eram demais para meu entendimento, quando meu computador não rodava minha análise. E obrigada por se orgulhar de mim e me achar tão inteligente, sei que é porque você é irmão coruja, mas fico muito feliz quando te ouço falar isso.

A todos meus familiares, em especial, à minha avó *Valquiria Cury* por ser meu maior exemplo de profissional. Obrigada por tudo o que a senhora fez por nós e obrigada por nunca ter desistido da gente.

Aos meus grandes parceiros de laboratório *Geysson, Jakeline, Leonardo, Juarez, Bruno Fantinatti, Bruna, Jéssica e Tassiana* que sempre me ensinam muito sobre ciência. Em especial à professora *Maeli Dal-Pai* que é nosso alicerce e um grande exemplo de pesquisadora, professora, mãe e mulher.

Aos meus amigos do laboratório e da vida *Paula Freire, Bruno Duran, Grasieli de Oliveira e Rafaela Nunes.* Agradeço à vida por ter guiado o meu caminho para encontrar vocês, vocês são extremamente importantes em minha vida.

Aos alunos de Iniciação Científica, em especial, ao *Douglas* e *Diogo* que foram indispensáveis para a realização deste trabalho.

Aos membros banca professoras *Sandra Linde* e *Renée Amorim*. Muito obrigada por estarem presentes neste momento tão importante de minha vida e contribuírem com o conhecimento tão valioso de vocês.

Às minhas amigas *Samara, Tamara, Amanda, Beatriz, Mariane, Yasmin, Betina e Camila* por sempre estarem por perto em todos os momentos, pela amizade de tantos anos e por todo apoio. Tenho muita sorte por ter vocês.

Ao *Leandro Gomes dos Santos* por me ajudar a ver o lado positivo da vida, mesmo quando a distância dificultava.

Aos colaboradores *Diogo de Moraes, Paula Paccielli Freire, Grasieli de Oliveira, Douglas Venâncio Pereira Marques, Geysson Javier Fernandez, Maeli Dal-Pai-Silva, Érica Nishida Hasimoto, Patricia Pintor dos Reis, Silvia Regina Rogatto, Robson Francisco Carvalho* que tanto se empenharam no desenvolvimento e escrita do manuscrito gerado a partir desde estudo.

À todos que conheci no Departamento de Clínica Genética do Hospital de Vejle, Dinamarca. Em especial à *Luisa Matos do Canto* que me acolheu com os braços abertos para minha primeira experiência internacional tão longa.

À todos os funcionários e docentes do Departamento de Morfologia do Instituto de Biociências – UNESP – Botucatu. E aos amigos que fiz durante esses anos, desde a minha Iniciação Científica neste departamento, em especial aos alunos do LabMec e agregados.

Ao programa de Pós-Graduação Ciências Biológicas (Genética) pela infra-estrutura e auxílio concedido através da bolsa CAPES do programa.

A Fundação de Amparo à Pesquisa do Estado de São Paulo (Processo nº 2017/21223-9) pela bolsa e auxílio financeiro concedidos. E ao *Ministry of Higher Education and Science of Denmark* pelo auxílio financeiro durante o estágio no hospital de Vejle.

Agradecimentos especiais

À professora *Silvia Rogatto* que me ensinou muito mais do que os conhecimentos dos livros, me ensinou sobre o amor, sobre a generosidade e sobre amizade. Foi um processo muito enriquecedor, e hoje, posso afirmar que sou uma profissional bem melhor graças aos seus ensinamentos e exemplos de postura crítica na pesquisa. Mas ainda tenho muito a aprender com você. Não paramos por aqui.

E finalmente, agradeço imensamente ao meu orientador *Prof Robson Carvalho* por me presentear com este trabalho lindo, confiar em meu potencial mesmo quando eu mesma não acreditava, e estimular a paixão pela pesquisa e ensino. E me dar todo o suporte e condições para ser a minha melhor versão. Além de ser um profissional exemplar você é um ser humano incrível que tive a sorte de ter como orientador e amigo.

Epígrafe

Há na vida momentos privilegiados nos quais parece que o universo se ilumina, que nossa vida nos revela sua significação, que nós queremos o destino mesmo que nos coube, como se nós próprios o tivéssemos escolhido.

(Louis Lavelle)

Sumário

Resumo

A caquexia é uma síndrome metabólica complexa que frequentemente acomete pacientes portadores de neoplasia maligna em estádio clínico avançado. Caracteriza-se pela perda de massa muscular (com ou sem perda de tecido adiposo), a qual não pode ser completamente revertida por suporte nutricional. As vias moleculares responsáveis pela caquexia não estão completamente esclarecidas, entretanto, os avanços em estudos genômicos, transcriptômicos e proteômicos no câncer tem auxiliado na compreensão da importante relação entre o secretoma tumoral com alterações em órgãos e tecidos adjacentes ou distantes do tumor. Evidências têm demonstrado que componentes do secretoma do ambiente tumoral, incluindo citocinas pró-inflamatórias, possuem um papel fundamental no desenvolvimento de alterações metabólicas que resultam em sarcopenia (perda de função e massa muscular) em pacientes caquéticos. A perda de massa muscular é considerada um importante fator prognóstico de caquexia para pacientes com carcinomas de pulmão de células não pequenas (CPCNP) e, portanto, a avaliação de área muscular utilizando-se de tomografias computadorizadas (CTs) tem sido utilizada com grande eficácia para determinar a sobrevida e a presença de caquexia e sarcopenia nesses pacientes. Portanto, a hipótese desse trabalho é que a integração de dados clínicos e prognósticos, área dos músculos peitorais obtidas por CTs e perfil transcricional tumoral permitirá identificar potenciais biomarcadores de caquexia no secretoma de carcinomas de pulmão de células não pequenas. Para isso, 89 CTs de pacientes com CPCNP, disponíveis na plataforma TCIA (*The Cancer Imaging Archive*), foram analisadas para determinação da área dos músculos peitorais, a qual foi utilizada para seleção de pacientes com baixa ou alta muscularidade. A análise de expressão diferencial de genes do tumor desses mesmos pacientes foi realizada para selecionar transcritos com expressão aumentada no tumor de pacientes com baixa muscularidade. Destes, selecionamos os genes relacionados ao secretoma do tumor, a partir de análises *in silico* para predição de proteínas secretadas. Além disso, essas moléculas foram avaliadas quanto à sua classificação funcional e relação com dados clínicos dos pacientes. Nossos dados demonstram a relevância da citocina pró-inflamatória, *Inteleucina-8*, como potencial biomarcador de caquexia associado ao pior prognóstico de CPCNP. Esses dados são úteis para detecção precoce da caquexia e podem ter impacto na clínica de pacientes com CPCNP.

Palavras chave: Muscularidade, Secretoma, Interleucina-8, Tomografia computadorizada, C2C12

Abstract

Cachexia is a metabolic syndrome characterized by an ongoing loss of skeletal muscle mass (with or without loss of adipose tissue) that cannot be fully reversed by conventional nutritional support, most found in patients with advanced cancer. The molecular pathways of cancer cachexia are not completely known. The advances in genomic, transcriptomic and proteomic studies in cancer have helped in understanding the relationship of the tumor's secretome with changes in organs and tissues adjacent to or distant from the tumor. Evidences show that components of the tumor secretome, including pro-inflammatory cytokines, play a key role in metabolic alterations that result in sarcopenia (loss of muscle mass and function) of cachectic patients. Loss of skeletal muscle mass is important prognostic factor for cachexia in patients with non-small cell lung cancer (NSCLC), thus the evaluation of muscular area using computed tomography (CT) has been effective in determine survival and the presence of cachexia and sarcopenia in these patients. Therefore, our hypothesis is that the integration of clinical and prognostic data, pectoralis muscle area obtained by CTs, and tumor transcriptional profile will allow the identification of potential biomarkers of cachexia in the secretome of non-small cell lung carcinomas. To do this, 89 CTs from patients with NSCLC, available on the TCIA (The Cancer Imaging Archive) platform were used to measure the pectoralis muscle area, and to select patients with low or high muscularity. Differential gene expression analysis of tumor from these same patients was performed to select transcripts with increased expression in the tumor of low muscularity patients. Transcripts related to the tumor's secretome were selected through *in silico* prediction analyzes. In addition, these molecules were evaluated for their functional classification and relationship to the clinical data of each patient. Our results demonstrated the relevance of *Interleukin-8* as potential biomarkers of cachexia associated with poor prognosis of patients with NSCLC. These data are useful for early detection of cachexia, which may have clinical impact in the management of patients with NSCLC.

Keywords: Muscularity, Secretome, Interleukin-8, Computed tomography, C2C12

1. Introdução

1.1. Caquexia associada ao câncer

A caquexia é definida como uma síndrome metabólica complexa caracterizada pela perda de massa muscular (com ou sem perda de tecido adiposo), a qual não pode ser completamente revertida por suporte nutricional, acarretando em alterações funcionais nos processos fisiológicos, metabólicos e imunológicos que levam a desequilíbrio energético e proteico[1,2]. Essa síndrome metabólica é caracterizada pela presença de sarcopenia (perda de função e massa muscular) [3], anorexia, inflamação, escurecimento e/ou diminuição do tecido adiposo, resistência à insulina e a perda de proteínas musculares [1].

Dentre as principais condições crônicas relacionadas com o desenvolvimento de caquexia estão: AIDS, doença renal crônica, doença pulmonar obstrutiva crônica, queimaduras, falência de órgãos, trauma, sepse e câncer [4–6]. Entretanto, é evidente que a caquexia associada ao câncer apresenta uma importância significativa - mais de 50% dos pacientes com câncer desenvolvem caquexia [7]. A caquexia predomina em alguns tipos de câncer, sendo os cânceres de pâncreas, gastroesofágico, cabeça e pescoço e pulmão aqueles que apresentam maior prevalência de caquexia. Pacientes que desenvolvem tumores nesses locais, em média, perdem mais peso (revisado por [8]). Além disso, a caquexia pode impactar a sobrevida de pacientes com câncer em estádios avançados, os quais podem apresentar uma taxa de mortalidade de 20 a 80% [6]. Sendo assim, estado caquético no câncer representa prognóstico desfavorável e, pacientes acometidos por essa síndrome podem apresentar diminuição da resposta aos tratamentos rádio e quimioterápico [9,10]. Embora a caquexia não esteja relacionada com o tamanho e dimensão do tumor [11], há uma prevalência de pacientes com caquexia associada ao câncer de pulmão de células não pequenas; 61% desses pacientes, nos Estados Unidos, apresentam uma significativa perda de peso (revisado por [11]).

Mesmo apresentando tal relevância, a caquexia é pouco diagnosticada e raramente tratada, e seus critérios para definição ainda não estão completamente estabelecidos na prática médica [1,6]. Recente estudo demonstrou que diversos profissionais da área da saúde ainda possuem muita dificuldade para reconhecer a caquexia e conduzir os tratamentos adequados [12]. Nesse estudo, os profissionais da saúde foram questionados sobre qual é a porcentagem de perda de peso necessária para considerar o paciente caquético e iniciar os tratamentos; surpreendentemente 48% esperariam uma perda de peso maior que 15% [12], o que é preocupante, visto que uma perda de 5% do peso corporal em seis meses já é considerada indicativa da presença de caquexia [7].

Uma forma para definir melhor os parâmetros necessários para diagnóstico da doença é entender seus mecanismos moleculares. Entretanto, as vias moleculares responsáveis pela caquexia ainda não são totalmente conhecidas, embora muitos estudos já tenham demonstrado que citocinas pró-inflamatórias, como interleucinas (IL-1β, IL-6), fator de necrose tumoral (TNF-α) e interferon (INF-γ), apresentam uma grande importância para o desenvolvimento de alterações celulares e moleculares que resultam na perda de função muscular. As citocinas são responsáveis por induzirem a degradação de proteínas musculares específicas (revisado por [8]). Além disso, essas citocinas são secretadas pelo tumor, liberadas na circulação sanguínea e auxiliam na indução de anorexia, aumento de lipólise e da perda de massa muscular por degradação proteica [10,13].

A caquexia é rotineiramente associada com inflamação, resistência a insulina e atrofia muscular [3]. A atrofia muscular nessa condição é representada por perda de massa muscular decorrente de degradação de proteínas miofibrilares nas células musculares esqueléticas, resultando em fadiga e fraqueza muscular [14], portanto indivíduos caquéticos geralmente são sarcopênicos, mas indivíduos sarcopênicos não são considerados caquéticos. Por isso, existe um consenso para se definir e distinguir pacientes caquéticos e sarcopênicos [15].

Na caquexia, outros importantes processos da célula muscular encontram-se comprometidos e contribuem para o aumento da atrofia muscular, tais como: desregulação na síntese de novas proteínas, aumento de apoptose e menor capacidade de regeneração (para uma revisão, ver Argilés et al.,[16]). O processo de atrofia muscular é tão importante para o desenvolvimento da caquexia que, independente da perda de peso corporal, a diminuição da área muscular, por si só, representa um prognóstico desfavorável em pacientes com câncer [16–20], incluindo câncer de pulmão [21–23].

1.2. Secretoma

Recentes avanços na genômica, transcriptômica e proteômica tem auxiliado a esclarecer o efeito do tumor em tecidos e órgãos adjacentes ou distantes através da análise do seu secretoma [24]. Secretoma representa o conjunto de macromoléculas secretadas por células, as quais permitem comunicação celular [25,26]; este termo foi cunhado por Tjalsma e colaboradores para denotar todos os fatores secretados por uma célula e, também, os constituintes da via secretória [27]. A definição de secretoma, então, foi revisada para incluir apenas proteínas secretadas no espaço extracelular [24, 25].

O secretoma tumoral parece apresentar papel importante em aspectos conhecidos do câncer, como a proliferação excessiva, apoptose reduzida, angiogênese, alteração no metabolismo energético e desenvolvimento de resistência contra a terapia anticâncer [25]. O secretoma tumoral pode ser avaliado

através de meio condicionado de linhagens celulares, líquidos intersticiais do tecido/tumor e fluídos corporais (revisado por[28]). Muitas metodologias estão disponíveis para a detecção dos componentes do secretoma, entretanto o uso de ferramentas ômicas (microarranjos, sequênciamento e espectometria de massas) tem se destacado nos últimos anos (revisado por [29]). Análises proteômicas revelaram que o secretoma de pacientes com câncer de pulmão, mama e colorretal possuia biomarcadores relacionados ao aumento de sobrevida dos pacientes (revisado por [28]). Além disso, Papaleo, Gromova e Gromov discutem a relevância da análise integrada do secretoma com dados ômicos, incluindo perfil proteômico e mapeamento de microRNAs tumoral, para compreender de forma abrangente as alterações biomoleculares do câncer. Esses autores ainda sugerem que são necessárias estratégias para tornar os dados de secretoma tumoral já publicados mais acessíveis, com maior consistência em suas anotações, e também padronizar e harmonizar os protocolos de análise para uma pesquisa transparente e reproduzível [30]. A existência de bancos de dados públicos é uma forma de contribuir para esses aspectos. Para dados de secretoma, bancos de dados públicos contendo informações de proteínas, mRNAs, microRNAS presentes em vesículas extracelulares, exossomos, meio de cultura ou líquidos corporais podem ser fácilmente acessados.

Devido ao fato da caquexia ser uma síndrome metabólica que afeta diversos tecidos, as ferramentas ômicas apresentam grande potencial para compreensão das alterações que ocorrem em vias metabólicas do tumor e de órgãos e tecidos distantes, permitindo assim a identificação de potenciais novos tratamentos e biomarcadores para essa condição [24]. Considerando que ainda não há um tratamento efetivo para caquexia, e a viabilidade da utilização de ferramentas "ômicas" para identificação de alterações moleculares nessa condição, torna-se relevante avaliar como o perfil transcricional de genes codificadores de proteínas secretadas por tumores, especialmente no câncer de pulmão, está relacionado à caquexia.

1.2. Câncer de pulmão

O câncer de pulmão é o mais comum no mundo - mais de 2 milhões de novos casos foram diagnosticados em 2018 - e o que mais leva à óbito, sendo responsável por cerca de 1,7 milhões de mortes [31]. Segundo o INCA (Instituto Nacional de Câncer), no Brasil, estimam-se para o ano de 2018, 18.740 novos casos de câncer de traqueia, brônquios e pulmão entre homens, e 12.530 entre mulheres [32]. O câncer de pulmão é uma das principais causas de morte evitáveis, pois é geralmente diagnosticado em estádios mais avançados da doença [32]; a sobrevida média em um total de cinco anos varia entre 13 a 21%, em países desenvolvidos, e entre 7 e 10%, nos países em desenvolvimento [32].

O câncer de pulmão apresenta-se em duas formas principais, que são classificados de acordo com sua característica histológica: câncer pulmonar de células pequenas (CPCP, da sigla, em inglês, SCLC, *Small-Cell Lung Cancer*), que representa cerca de 15% de todos os tipos de câncer de pulmão, e o câncer pulmonar de células não pequenas (CPCNP, da sigla, em inglês, NSCLC, *Non-Small-Cell Lung Cancer*), que prevalece em cerca de 85% dos casos [33]. Os CPCNP mais frequentes são três subtipos: adenocarcinoma, carcinoma de células escamosas (*Squamous-cell carcinoma*, em inglês) e carcinoma de grandes células (em inglês, *Large-cell lung cancer*) [33], sendo o adenocarcinoma e o carcinoma de células escamosas os mais comumente encontrados na população [33, 34]. O fumo ou exposição passiva ao tabaco estão associados à 90% dos casos de câncer de pulmão [35]. Fumar pode causar todos os tipos de câncer, porém, está fortemente relacionado com CPCP e carcinoma de células escamosas; e o adenocarcinoma é o tipo mais comum em pacientes que nunca fumaram [33]. Embora, existam avanços no diagnóstico precoce e tratamento padrão, o câncer de pulmão é frequentemente diagnosticado em um estádio avançado e tem prognóstico desfavorável [33].

Dentre as causas desse prognóstico desfavorável do câncer de pulmão destaca-se a caquexia, a qual acomete aproximadamente 60% dos pacientes e contribui para maior mortalidade da doença; estima-se que 20% das mortes são secundárias à caquexia [36–38]. Embora bem estabelecida uma estreita relação entre câncer de pulmão e a caquexia, as moléculas produzidas e liberadas pelo ambiente tumorais, potencialmente envolvidas no desenvolvimento de atrofia muscular na caquexia associada a esse tipo de câncer, ainda não são conhecidas. Dessa maneira, diversos estudos buscam entender como o secretoma do câncer altera o metabolismo de outros tecidos, incluindo a atrofia do tecido muscular esquelético (revisado por Twelkmeyer *et al*., 2017) [24]. Portanto, a análise do secretoma tumoral além de ser extremamente importante para a compreensão da biologia tumoral [28] pode, também, auxiliar na identificação de moléculas potencialmente responsáveis pela indução de atrofia muscular na caquexia associada ao câncer, incluindo alvos para o desenvolvimento de medicamentos, além da identificação de novos potenciais biomarcadores dessa condição.

Embora muitos estudos já tenham sido desenvolvidos para a análise do transcriptoma de pacientes com CPCNP [39,40], para nosso conhecimento, nenhum estudo explorou esses dados na busca de elementos do secretoma potencialmente associados à caquexia; em especial, pela dificuldade de obtenção de dados clínicos e moleculares de pacientes com caquexia (revisado por [41]). Entretanto, com o avanço da tecnologia, nossa compreensão da expressão gênica mudou drasticamente na última década, experimentos em larga escala - microarranjos e sequenciamento de próxima geração - geraram grandes quantidades de dados de expressão genômica que são depositados em arquivos

públicos (revisado por [42]). Desta forma, a reutilização de dados inseridos em bancos de dados públicos se tornou indispensável. Os dados públicos, geralmente, são combinados com novos dados gerados pelos pesquisadores, mas também podem ser re-analisados para responder questões diferentes daquelas colocadas nos estudos originais. A integração de dados ômicos disponíveis em bancos de dados é capaz de gerar *insights* biológicos, que, talvez, em estudos individuais não pudessem ser obtidos (revisado por [42]).

O *The Cancer Imaging Archive* (TCIA; http://www.cancerimagingarchive.net/) é um exemplo de plataforma de acesso público que possui coleções de imagens referentes à tipos ou ao local de câncer em comum [43]. Nessa plataforma podem ser depositadas ou acessadas imagens, além de outras informações relevantes para suporte dos dados de imagem, como dados genômicos e clínicos de cada paciente. A reutilização de dados da literatura, importante fonte de informações, incluindo dados do TCIA, TCGA (The Cancer Genome Atlas [44]) e GEO (Gene Expression Omnibus [45]) podem ser utilizados para integrar análises da musculatura dos pacientes através de imagens de tomografias computadorizadas (TCs) com perfil transcricional do tumor (dados de microarranjo ou seqüenciamento de RNA).

1.3. Tomografias computadorizadas para avaliação da área muscular de pacientes com câncer de pulmão

A avaliação da área do músculo não pertence ao repertório padrão na prática da oncologia médica, entretanto, dados recentes de TCs sugerem que essas medidas fornecem informações sobre o prognóstico de pacientes com câncer, e já tem se mostrado eficientes na avaliação clínica e na pesquisa [46], sendo considerada como o método padrão ouro de análise da composição corporal para diagnóstico de fenótipos anormais [3]. A análise tomográfica é realizada rotineiramente durante o diagnóstico e o estadiamento do câncer, e a introdução da avaliação da composição corporal por TC pode ser rápida, levando aproximadamente 20 minutos por exame, tornando-a extremamente viável [47].

A revisão sistemática realizada por Kazemi-Bajestani e seus colaboradores demonstrou o crescimento e importância do estudo das tomografias para correlacionar caquexia e sarcopenia com diferentes eventos relacionados ao câncer, tais como sobrevida, toxicidade da quimioterapia e complicações pós-cirúrgicas [48]. Essa revisão identificou uma importante relação entre a área muscular e a sobrevida dos pacientes, na qual a diminuição da área do músculo prediz a menor sobrevida, independente de outras variáveis tais como sexo, idade e estadiamento do câncer [17–19]

A atrofia muscular em pacientes com câncer de pulmão é característica importante de se avaliar, mesmo quando não existe alteração no peso, ou quando os pacientes são obesos [49]. Estudos recentes têm discutido a relevância da utilização do Índice de Massa Corpórea (IMC) e perda de peso corporal para classificação de indivíduos saudáveis e caquéticos [17,50], além da dificuldade de se avaliar a composição corpórea em pacientes obesos, a perda de peso em adultos pode ser corrigida pela retenção de líquidos [1]. A desnutrição também foi associada com fenótipos anormais de composição corpórea em um grupo de 725 pacientes oncológicos, apesar do alto nível de sobrepeso e obesidade registrado nessa grande coorte [47]. Embora esse estudo tenha revelado altas taxas de perda de massa muscular detectadas pela análise de TC, a adiposidade excessiva pareceu mascarar essa perda de peso e desnutrição, sendo que uma grande parte dos pacientes considerados caquéticos mantiveram um peso corporal global estável, mas tinham alterações na massa muscular e gordurosa que só eram detectáveis com a análise de TC. Ainda, 38% dos casos de caquexia nesta coorte foram diagnosticados usando apenas a tomografia computadorizada e não sendo identificáveis usando o IMC. Bhuachalla e seus colaboradores sugerem que menos importância deveria ser dada ao IMC, porque, como mostrado, pacientes com um IMC saudável, com excesso de peso ou obesos podem apresentar níveis severos de desnutrição e caquexia [47]. Por isso, estudos indicam uma maior eficiência das TCs para análise de área muscular na avaliação de caquexia em pacientes com câncer [51–54].

Em câncer de pulmão, os estudos envolvendo a avaliação da área muscular através de imagens de TCs tem sido relevantes para associar este parâmetro a dados clínicos, como a resposta ao tratamento, sobrevida e marcadores de inflamação (concentração de albumina sérica e de proteína C-reativa) [revisado por [55]]. Portanto, é evidente que a análise do músculo esquelético por TC apresenta forte relação com a caquexia associada ao câncer de pulmão. De fato, revisão sistemática revelou o impacto da utilização de TCs para acessar a sarcopenia em pacientes com câncer de pulmão e para determinar a relação da perda de massa muscular com morbidez dos pacientes com diferentes subtipos de câncer de pulmão. Além disso, essa revisão ressaltou que a perda de massa e função muscular (sarcopenia) avaliada por imagens pode anteceder a caquexia clinicamente evidente, sublinhando a importância de avaliar a sarcopenia, ao invés da perda de peso sozinha [55].

Esta revisão previamente citada traz dados de artigos publicados até 2013, por isso buscamos artigos mais atuais (até 2018) que avaliaram a muscularidade (medida muscular normalizada) de pacientes com câncer de pulmão utilizando imagens de TCs. A **tabela 1** mostra que diversos autores utilizaram essa estratégia para associar a área de músculos com características clínicas dos pacientes com câncer de pulmão. Muitos desses estudos avaliaram a área do mesmo conjunto total de músculos

na altura da terceira vértebra lombar (L3), e alguns sugerem uso de outros músculos e normalizações que não incluem a altura dos pacientes, a qual é a mais frequentemente utilizada. Porém, isso demonstra que não possuímos uma metodologia uniforme para detecção da sarcopenia. Diferentes estratégias tem sido utilizadas para normalizar a musculatura e identificar pacientes com menor massa muscular – uma vez que nem sempre a medida da altura é aferida de forma correta ou relatada corretamente pelo paciente (revisado por [56]). Portanto, uma medida da muscularidade, independente da altura, pode trazer, de forma rápida, dados relevantes sobre a composição corpórea de pacientes oncológicos, por isso medidas ósseas podem funcionar como alternativas para normalização.

É importante destacar que de forma geral, mesmo utilizando diferentes estratégias, estes estudos conseguiram associar a perda de massa muscular com um prognóstico desfavorável. De fato, para o carcinoma de pulmão de células não pequenas, a área dos músculos peitorais é um importante biomarcador para sarcopenia e caquexia, pois está relacionada com a predição de sobrevida desses pacientes [51,57]. Além disso, Stene et al., 2015 [20] mostraram que nos pacientes com CPCNP que receberam tratamento quimioterápico, a atrofia muscular está relacionada com menor sobrevida, mesmo quando eles não se enquadravam nos parâmetros para identificação de sarcopenia utilizados na análise, como o aumento da proteína C-reativa e diminuição de albumina sérica. Esses dados demonstram uma forte relação entre a diminuição da área do músculo e um prognóstico desfavorável para os pacientes com CPCNP. Entretanto, para nosso conhecimento, ainda não existem na literatura estudos que correlacionem área de músculo de pacientes com CPCNP e o perfil transcricional/secretoma do tumor.

Tabela 1. Análise de tomografias computadorizadas utilizando diferentes parâmetros para avaliação de muscularidade em pacientes com câncer de pulmão.

Autores	Tipo histológico	Músculo avaliado	Normalização	N	Resultado Principal
Jafri et al, 2015	CPCNP	Músculos na altura da L3	Altura (m^2)	112	Índice de caquexia combinando TC, nível de albumina sérica, razão neutrófilo/linfócito.
Stene et al, 2014	CPCNP	Músculos na altura da L3	Altura (m^2)	35	Mudanças na massa muscular é fator prognóstico.
Suziki et al, 2016	CPCNP	Músculos na altura da L3	Altura (m^2)	90	Sarcopenia associada à pior prognóstico.
Recio-Boiles et al, 2018	CPCNP	Músculos na altura da L1	Área da superfície corporal	37	Músculos da região da L1 são confiáveis para avaliar sarcopenia.
Kim et al, 2016	CPCP	Músculos na altura da L3, L1 e PMA	Altura (m^2)	90	L1 correlaciona fortemente com L3.
Kinsey, 2016	CPCNP	PMA	None	252	PMA associado com menor sobrevida.
Kim et al, 2015	CPCP	Músculos na altura da L3	Altura (m^2)	149	Sarcopenia prediz pior prognóstico.
Go et al, 2016	CPCP	PMA	Altura (m^2)	117	Pacientes sarcopênicos e com alta razão neutrófilo/linfócito apresentam pior prognóstico.
Sjøblom et al, 2016	CPCNP	Músculos na altura da L3	Altura (m^2)	734	Radiodensidade muscular associada à pior prognóstico.
Kim et al, 2016	CPCP	Músculos na altura da L3	Altura (m^2)	186	Significante relação entre massa muscular determinada por TC e proteína C reativa.
Kim et al, 2017	CPCP	Músculos na altura da L3	Altura (m^2)	127	Sarcopenia e/ou adipopenia prediz prognóstico em homens.
Wysham et al, 2016	CPCNP avançado	Músculos na altura da T4	Altura (m^2)	86	Sarcopenia torácica não se correlacionou com tolerância à atividade física e sobrevida.
Nattenmüller et al, 2017	CPCP e CPCNP	Músculos na altura da L2 e L3	Altura (m^2)	200	Sarcopenia associada ao tratamento quimioterápico e baixa sobrevida
Shoji et al, 2017	CPCNP	Músculos na altura da L3	Altura (m^2)	147	Sarcopenia avaliada antes da cirurgia se relaciona com parâmetros imuno-nutricionais.
Kim et al, 2018	CPCP	Músculos na altura da L3	Altura (m^2)	127	Avaliação da sarcopenia e adipopenia podem ser utilizadas para predizer prognóstico em homens com CPCP.
Recio-Boiles et al, 2018	CPCNP	Músculos na altura da L1	Altura (m^2)	73	Uso da avaliação dos músculos da L1 são confiáveis para avaliar sarcopenia.

CPCP: câncer de pulmão de células pequenas; CPCNP: câncer de pulmão de células não pequenas; L1: área de todos os músculos da região da primeira vértebra lombar; L2: área de todos os músculos da região da segunda vértebra lombar; L3: área de todos os músculos da região da terceira vértebra lombar; T4: área de todos os músculos da região da quarta vértebra torácica; PMA: área dos músculos peitorais; TC: tomografia computadorizada; N: número da pacientes.

2. Hipótese

A hipótese desse trabalho é que a integração de dados clínicos e prognósticos, da área muscular, e do perfil transcricional tumoral permitirá identificar potenciais biomarcadores de caquexia no secretoma do carcinoma de pulmão de células não pequenas. Nosso estudo também poderá servir de base para a identificação de componentes do secretoma que poderão ser utilizados como biomarcadores ou para o desenvolvimento de futuras estratégias terapêuticas visando à minimização da perda de massa muscular, aumentando assim a sobrevida e a melhora da qualidade de vida dos pacientes com caquexia associada ao câncer de pulmão.

3. Objetivos

3.1. Objetivos Gerais

i. Estabelecer um novo e aplicável método para determinação da muscularidade utilizando somente dados obtidos a partir de TCs de pacientes com CPCNP;

ii. Identificar potenciais biomarcadores de caquexia no secretoma do carcinoma de pulmão de células não pequenas em pacientes com baixa muscularidade.

3.2. Objetivos Específicos

a. Avaliar a área dos músculos peitorais através de TCs dos pacientes com CPCNP;

b. Identificar os pacientes com baixa muscularidade e alta muscularidade, a partir dos dados de TCs;

c. Comparar o perfil transcriptômico de pacientes com baixa muscularidade e alta muscularidade;

d. Comparar os achados moleculares com informações clínicas e histopatológicas;

e. Identificar a ontologia funcional dos transcritos com expressão alterada no carcinoma de pacientes com baixa muscularidade;

f. Determinar transcritos que codificam proteínas potencialmente secretadas pelo carcinoma de pulmão;

g. Validar os potenciais biomarcadores através da integração de dados de expressão gênica e dados de sobrevida dos pacientes.

4. Capítulo 1

Manuscrito submetido ao *Journal of Cachexia, Sarcopenia and Muscle* (JCSM-D-19-00053).

Tumor transcriptome in patients with lower pectoralis muscle area reveals IL-8 as a prognostic biomarker in non-small cell lung cancer

Sarah Santiloni Cury[1], Diogo de Moraes[1], Paula Paccielli Freire [1], Grasieli de Oliveira[1], Douglas Venâncio Pereira Marques[1], Geysson Javier Fernandez[1], Maeli Dal-Pai-Silva[1], Érica Nishida Hasimoto[2], Patricia Pintor dos Reis[2,3], Silvia Regina Rogatto[4], Robson Francisco Carvalho*[1].

1 São Paulo State University (UNESP), Institute of Biosciences, Department of Morphology, Botucatu São Paulo, Brazil.

2 São Paulo State University (UNESP), Department of Surgery and Orthopedics, Faculty of Medicine, Botucatu, São Paulo, Brazil.

3 São Paulo State University (UNESP), Experimental Research Unity, Faculty of Medicine, Botucatu, São Paulo, Brazil.

4 Department of Clinical Genetics, Vejle Hospital, Institute of Regional Health Research, University of Southern Denmark, Denmark.

* Corresponding author
Address: Prof. Dr. Antonio Celso Wagner Zanin, 250, Botucatu, SP, Brazil
Email address: robson.carvalho@unesp.br
Telephone: +55 14 3880 0473

Abstract

Background

Cachexia is a syndrome characterized by an ongoing loss of skeletal muscle mass associated with poor patient prognosis in non-small cell lung cancer (NSCLC). Up to 60% of NSCLC patients develop significant loss of muscle mass. However, prognostic biomarkers of cachexia in NSCLC are unknown. Here, we analyzed and integrated computed tomography (CT) images and tumor transcriptome data to identify novel potentially secreted cachexia biomarkers in NSCLC patients with low muscularity.

Methods

We integrated radiomics features (pectoralis muscle, sternum, and T10 vertebra) from CT of 89 NSCLC patients, which allowed us to identify an index for screening muscularity. Next, a tumor transcriptomic-based secretome analysis from these patients (discovery set) was evaluated to identify potential cachexia biomarkers in patients with low muscularity. The prognostic value of these biomarkers for predicting recurrence and survival outcome were confirmed using expression data from eight lung cancer datasets (validation set). Finally, C2C12 myoblasts differentiated into myotubes were used to evaluate the ability of the selected biomarker, IL-8, in inducing muscle cell atrophy.

Results

We identified 75 over-expressed transcripts in patients with low muscularity, which included several pro-inflammatory cytokines, such as *IL6, CSF3,* and *IL8*. In addition, we identified *NCAM1*, *CNTN1*, *SCG2*, *CADM1, IL8, NPTX1*, and *APOD* as novel potential cachexia biomarkers in the tumor secretome. These biomarkers were capable of distinguishing worse and better prognosis (recurrence

and survival) in NSCLC patients. *IL8* was confirmed as a predictor of worse prognosis in all validation sets. *In vitro* assays revealed that IL-8 promoted C2C12 myotube atrophy of myotubes differentiated from C2C12 myoblasts.

Conclusions

Tumors from low muscularity patients presented a set of upregulated genes encoding for secreted proteins, including pro-inflammatory cytokines that predict worse overall survival in NSCLC. Among these up-regulated genes, *IL8* expression in NSCLC tissues was also associated with worse prognosis and the recombinant IL-8 was capable of triggering atrophy in C2C12 myotubes.

Keywords: Secretome, Computed tomography, Interleukin-8, Tumor-derived factor, C2C12 cells, Cachexia

Introduction

Lung cancer is the most prevalent cancer type worldwide and responsible for an estimated 1.8 million deaths, each year [80]. Most patients (~ 85%) develop non-small cell lung cancer (NSCLC) [81], which is frequently diagnosed in an advanced stage, and consequently has an unfavorable prognosis [33]. Cancer cachexia is a syndrome that affects a considerable proportion of NSCLC patients [49]. It is characterized by an ongoing loss of skeletal muscle mass (with or without loss of fat mass) that cannot be fully reversed by conventional nutritional support and is associated with significant functional impairments [82].

The loss of skeletal muscle mass in cancer cachexia may lead to substantial weight loss and decreased body mass index (BMI), which are associated with worse outcome in NSCLC patients [36,83,84]. Studies using computed tomography (CT) images have revealed occult muscle depletion in NSCLC patients, regardless of overall body weight [49,85]. In addition, the detection of muscle depletion or low muscle mass by CT images have been associated with shorter time to tumor progression, increased risk of chemotherapy toxicity, and shorter survival in NSCLC patients [20,49,85–89]. Skeletal muscle depletion detected by CT images in these patients also negatively affects their functional status and quality of life [90,91]. To our knowledge, tumor-secreted factors with prognostic value associated with low muscle mass as detected by CT in NSCLC are unknown.

Several studies have highlighted that macromolecules secreted from cancer cells and cells within the tumor microenvironment (secretome), including many pro-inflammatory cytokines, act systemically leading to muscle wasting in cancer cachexia [8,10,24]. However, the secretome complexity and differences found in distinct lung cancer and cells lines [28,92,93] illustrates the need to apply global approaches, in order to identify tumor-specific secreted molecules associated with skeletal muscle depletion. Moreover, previous "omics" studies of cancer secretome in cachexia have

focused on the analysis of cachectic conditioned media of single cancer cells lines to identify mediators of the syndrome [94–96]. However, *in vitro* systems ignore the contributions of the host–tumor microenvironment and the tumor heterogeneity as well as provide no insight into the disease progression [93]. These findings emphasize the importance of cancer cachexia studies in exploring the tumor secretome. Thus, we hypothesized that a tumor transcriptome-based secretome analysis in NSCLC patients with low muscularity is a strategy capable to identify prognostic biomarkers and mediators of cancer-associated weight loss.

Herein, we analyzed a cohort of NSCLC patients with CT images, clinical findings and tumor expression microarrays data from a previous study that decoded tumor radiomics features associated with gene expression levels [97]. For these patients, we compared the pectoralis muscle area (PMA) with muscle normalizations based on different radiomics features to select an approach for screening muscularity. Next, we identified genes predicted to be secreted in patients with low muscularity and assessed their prognostic value as tumor markers of recurrence-free survival and overall survival. Finally, we demonstrated the potential of IL-8 as a putative secreted marker capable of inducing atrophy in C2C12 myotubes.

Material and Methods

The workflow of the integrative analyses of CT images and tumor transcriptome used to identify potentially secreted cachexia mediators and biomarkers in NSCLC patients with low-muscularity is depicted in **Figure 1**.

Datasets

CT images and clinical data were downloaded from *The Cancer Imaging Archive* (TCIA, http://cancerimagingarchive.net/) database [43]. The dataset (NSCLC-Radiomics-Genomics collection) [98] contains information from 89 NSCLC adult patients treated at MAASTRO Clinic, The Netherlands, as previously published [58]. TCIA data are anonymized and the institutional ethical review board approval is not needed [99]. CT images were taken on diagnosis and the patients were treated with surgical procedure. Clinical data (age, gender, diagnosis, tumor stage), CT images and tumor microarrays data are available for all 89 patients. The NSCLC-Radiomics-Genomics microarrays data is available on Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo; microarrays dataset GSE58661) [97].

CT Imaging Analyses

The CT collection "NSCLC-Radiomics-Genomics" on TCIA database present CT images with radiomics features that can be used as noninvasive prognostic or predictive biomarkers [97]. This collection is also the most appropriate due the homogeneity of the CT images. The pectoralis muscle was analyzed on a single axial slice of the image. This region was selected by a single trained physician (ENH) who identified the aortic arch and then selected the first image just above the arch. The cross-sectional area (cm^2) of bilateral major and minor pectoralis muscles was measured by two independent examiners, using Slice-O-Matic software (v.5.0; Tomovision, Montreal, Quebec, Canada). Muscles were manually traced using Region of Interest (ROI) tool by summing the

appropriate pixels determined by CT Hounsfield unit (HU) for skeletal muscle (range −29 HU to 150 HU). The borders of the pectoralis muscles were corrected manually when necessary, as previously described [51,57,100]. The pectoralis muscle area (PMA) was calculated by adding up the four muscles area. To test the reproducibility of this analysis, an interobserver coefficient of variation was determined by comparing the results of the analyses conducted by the two observers. The mean of this coefficient of variation was 8.1%.

We also compared the PMA with muscle normalizations based on different radiomics features, as previously described [101–103], to test different approaches for screening muscularity in NSCLC patients. For this purpose, the pectoralis muscle area was also normalized by the following sternum measurements: 1) manubrium length; 2) sternum body length; 3) total manubrium and sternum body lengths; 4) distance between the beginning of manubrium and the end of sternum body measured in 90º (not considering the xiphoid process) (**Supplementary figure 1a**). Different T10 vertebrae measurements were also tested for muscle normalizations: 1) horizontal length of T10 body; 2) vertical length of T10 body; 3) distance between T10 body and spinous process; 4) distance between transverse processes; 5) distance between pedicles; and 6) T10 body area. We also analyzed the body cross-section anteroposterior diameter (APD) at the tenth thoracic vertebra (T10) level to normalize the muscle area (**Supplementary Figure 1b**). The bone images were selected in the cross-section where the bones appeared in higher extent and dimension. The measurements were performed at the tenth thoracic vertebra (T10), which is a common region for all patients in this CT collection. Skeletal muscle index (or muscularity) was defined as the PMA divided by each bone or body measure (mentioned above) squared (cm^2/cm^2). The measurements generated were z-score normalized and submitted to a non-hierarchical k-means clustering analysis using Bioconductor Package Complex Heatmap (v 3.5) in RStudio software (http://www.rstudio.org/).

Gene expression analysis

Tumor gene expression analysis was performed by comparing low and high muscularity patients using the GEO2R tool (http://www.ncbi.nlm.nih.gov/geo/geo2r/) [45]. The adjusted p values (adj. p) were applied using Benjamini and Hochberg false discovery rate (FDR) method by default. The cut-off criteria to define differential expression were adj. $p<0.05$ and $|\text{Fold Change (FC)}|>1.5$.

Gene ontology enrichment analysis

Gene ontology (GO) functional enrichment analysis was performed to identify the overrepresented GO categories of differential expressed genes using Gene Ontology Consortium database (http://geneontology.org/) [104]. The GO categories with p-value and FDR <0.05 were considered significant.

Protein-protein interactions (PPI) networks

PPI networks were generated using the STRING tool [105,106] (http:// string-db.org/). We considered experiments, database, co-expression, neighborhood, and co-occurrence as active interaction sources. The minimum required interaction score was 0.700 (high confidence), and the disconnected nodes in the network were hidden for display simplifications. The PPI enrichment p-value indicates the statistical significance provided by STRING.

In Silico Identification of Secreted Proteins

The over-expressed genes in the tumor of patients with low muscularity were filtered for genes encoding secreted proteins or proteins presented in microvesicles based on a pipeline of seven tools: SignalP 4.1 [65], SecretomeP 2.0 [66], ExoCarta [70], TargetP 1.1 [67], Human Cancer Secretome (HCS) [71], Vesiclepedia [68], and Evpedia [69]. Firstly, we accessed the UniProtKB database to obtain amino acid sequences of proteins in FASTA format [64]. These data were used in the prediction servers SignalP, TargetP, and SecretomeP at CBS portal (http://www.cbs.dtu.dk/services/). SignalP 4.1 server was used to identify classical secretory proteins (presenting signal peptide and D-value > 0.45). Proteins without signal peptide were evaluated in the SecretomeP 2.0 server to determine non-classical secreted proteins, using the cut-off for a neural network (NN) score > 0.6. TargetP 1.1 server was used to selectively collect proteins involved in secretory pathways and exclude mitochondrial proteins [67]. These potential secreted proteins were also investigated in lung cancer using the tools ExoCarta, HCS, Vesiclepedia, and Evpedia. Finally, the Plasma Proteome Database was consulted to identify human plasma proteins and their isoforms potentially encoded by the over-expressed genes from low muscularity patients [72]. The tumor over-expressed genes, detected by all eight prediction tools, were next used to assess their prognostic performance in predicting overall survival and time to recurrence in multiple NSCLC independent datasets (validation set).

Prognostic performance of secretory genes in predicting NSCLC outcome

SurvExpress [107] database (http://bioinformatica.mty.itesm.mx/SurvExpress) was used to assess the effect the differentially expressed genes on survival (datasets: Lung Meta-base, TCGA-LUAD and LUSC [108], GSE30219 [75], GSE31210 [76,77], and the Director's Challenge Consortium NCI [78]) and time to recurrence (dataset: GSE8894 [74]) of NSCLC patients. This tool allowed us to assess the expression of secretory genes and their association with the survival or time to recurrence

by Cox Proportional Hazard regression according to the risk groups estimated by an optimization algorithm. The prognostic value of the secretory genes in predicting survival were further determined in 1053 NSCLC patients using Kaplan-Meier Plotter – KM plotter [73]. Here, gene expression was specifically associated with survival and time to recurrence (worse prognosis) due to the lack of other clinical characteristics available in the databases.

Functional assay using the C2C12 cell culture

C2C12 mouse myoblasts (ATCC® CRL-1772™) were cultured in Dulbecco's modified Eagle's medium (DMEM, Thermo Fisher Scientific, USA) with 1% Penicillin–Streptomycin (Thermo Fisher Scientific, USA) and 10% fetal bovine serum (FBS, Thermo Fisher Scientific, USA) at 37 °C and 5% CO2 for growth and expansion. After reaching a confluence of 80-90%, the myoblasts were induced to differentiate in DMEM serum-free supplemented with 1% Penicillin–Streptomycin for 5 days. Recombinant IL-8 (10, 100 or 1000 ng/ml; Abcam, USA) was added to a new differentiation medium for 24 hours. All experiments were conducted using three independent replicates per group. Control myotubes (Ctrl) received a water solution containing bovine serum albumin 0.1%, the same solution used to dilute IL-8.

Immunofluorescence assay

C2C12 myotubes cultured in 6-well plates were fixed in 4% paraformaldehyde for 15 min, washed with phosphate buffered saline (PBS) and 0.1% TritonX-100 (Sigma, USA), and blocked with 3% bovine serum albumin (BSA), 1% glycine, 8% fetal bovine serum in PBS and 0.1% TritonX-100 for 1h at room temperature. Subsequently, the cells were incubated with primary (Myh2, 1:600 dilution, M7523, Sigma, USA) antibody overnight at 4°C. After washing, the cells were incubated with secondary antibody at the same concentration of the primary antibody (1h at room temperature), and counter stained with 4′,6-diamidino-2-phenylindole - DAPI (ProLong Gold Antifade Mountant

with DAPI, Thermo Fisher Scientific, USA). All images were acquired using scanning confocal microscope Fluoview FV10i (Olympus, Japan). The myotube diameter was measured by ImageJ software.

Statistical analysis

For the statistical analyses not previously described, we used the *GraphPad Prism*® (*GraphPad Software*, USA). Student's t-test or Mann-Whitney U-test was applied for independent samples with normal distributed or non-parametric data, respectively. The comparison of the effect of three different IL-8 concentrations on C2C12 myotubes with the respective controls was performed using one-way ANOVA followed by Tukey test. Data are expressed as mean ± standard deviation (SD).

Results

Study population

CT images, clinical, and microarrays data of 89 NSCLC patients, with average age of 65.2 ± 8.7 years, were included in this study. The most common NSCLC histological type was adenocarcinoma (47.2%), and 20.2% of the patients were diagnosed with advanced stage cancers (stages III or IV). Adenocarcinoma was prevalent in women while squamous cell carcinoma was more frequent in men. The muscle measurements revealed differences between sexes, with men and women presenting PMA of 42.5 ± 9.3 and 27 ± 6.0 cm^2, respectively. Based on this finding, sex-specific categorical variable was taken into consideration for further analyses. The **Table 1** summarizes the clinical, histopathological and muscle measurements in this cohort of NSCLS patients.

PMA distinguishes NSCLC patients with low and high muscularity

Considering that CTs from NSCLC patients have information that goes beyond the tumor, we integrated different radiomics features to determine an approach to be used for screening muscularity. The non-hierarchical, unsupervised clustering analysis of the PMA and its normalization by 11 CTs features (z-score normalized) revealed a similar pattern of patients' distribution according to all muscularity indexes. This analysis resulted in a cluster composed by 34 patients with low muscularity (**Figure 2a**). Therefore, a descending PMA order using a sex-specific categorical variable was performed. Subsequently, we segregated into terciles to generate two groups of study based on the patients' muscularity. The low muscularity group includes patients within the third tercile, while the high muscularity group includes patients within the first and second terciles, regardless of the patient gender (**Figure 2a**). The mean PMA differed significantly between the high and low muscularity groups considering all patients or comparing male and female patients (**Table 1**). We further compared high and low muscularity patients with other clinical variables using patient demographic information (**Figure 2b**). The comparison between these groups (high and low muscularity) revealed that squamous cell carcinomas were more frequently detected in patients with low muscularity (**Figure 2b**).

Patients with low muscularity upregulate tumor genes previously associated with cachexia

Considering that mediators released from cancer cells and cells within the tumor microenvironment have been associated with cachexia in lung cancers, we hypothesized that the identification of tumor deregulated genes in NSCLC patients with low muscularity could reveal potential factors associated with cachexia. Thus, an analysis using differential gene expression between patients with low and high muscularity revealed 105 genes exclusively deregulated (adj. p-value ≤ 0.05 and fold change ≥ 1.5) in patients with low muscularity, of which 75 and 30 were over- or down-expressed, respectively (**Supplementary Table 1**). Gene ontology and KEGG pathway analyses of the over-expressed transcripts highlighted cytokine activity and cytokine-receptor

interaction activity as the most enriched categories in low muscularity patients. (**Figure 3a**). PPI analysis identified the interactions among these proteins (**Figure 3b**), including the pro-inflammatory cytokines IL-6, IL-8, and CSF3, which have been previously implicated in the development of cancer cachexia [109–112].

Secretome-related genes with prognostic value in NSCLC

We then investigated whether these 75 up-regulated transcripts in the tumors from low muscularity patients are translated into secreted proteins. The intersection of the secretome databases CBS Servers, Vesiclepedia, Human Cancer Secretome Database, and Plasma Proteome Database showed seven overlapping proteins: IL-8, SCG2, NCAM1, CNTN1, CADM1, NPTX1, and APOD (**Figure 4a**). The microvesicle databases revealed that the predicted proteins in Evpedia (LPL, APOD and COL14A1) were also identified in the Vesiclepedia dataset. However, the Exocarta did not show any of these proteins in lung cancer samples, possibly due to the limited number of exosomes studies in lung cancers deposited in this database.

The prognostic value related to worse prognosis of *IL8, SCG2, NCAM1, CNTN1, CADM1, NPTX1,* and *APOD* transcripts were evaluated in seven lung cancer transcriptome datasets (validation set). Notably, these biomarkers were capable to distinguish worse and better prognosis (recurrence and survival) in seven NSCLC cohorts from the SurvExpress database (**Figure 4c**). Interestingly, only *IL8* was found with increased expression in high-risk group in all NSCLC validation set (**Figure 4d** and **Supplementary Figure 3**).

High IL8 expression in tumor tissues is associated with poor prognosis in NSCLC

All seven potential biomarkers were individually analyzed in the KM plotter server using gene expression and survival data of lung cancer patients available on the database (N = 1053), and *IL8* proved to be a strong predictor of poor survival (**Figure 5a**). Moreover, as *IL6* is a key regulator of

muscle mass during cachexia [113] and has been associated with worse prognosis in lung cancer patients [114,115], we compared the prognostic value of *IL8* with *IL6* using KM plotter server. Notably, both *IL8* and *IL6* tumor transcripts presented similar prognostic values (*IL8*: HR = 1.28, 95% CI = 1.12-1.45; *IL6*: HR = 1.32, 95% CI = 1.16-1.5). These results demonstrate the upregulation of *IL8* as a new biomarker associated with poor prognosis in lung cancer patients.

The 75 over-expressed transcripts were carefully evaluated in patients with low muscularity in KM plotter to detect additional potential cachexia biomarkers associated with poor prognosis in lung cancer patients. Nine genes (*IL6*, *IL8*, *IL1R2*, *CEMIP*, *CLEC4E*, *FCGR3B*, *HAL*, *MAP2K6*, *and KIF1A*) were validated as over-expressed in patients with worse overall survival (**Supplementary Figure 3**). Importantly, *IL6*, *IL8*, *IL1R2*, *CEMIP*, *FCGR3B*, *and KIF1A* are predicted as potentially secreted protein in at least two secretome databases (**Supplementary Table 2**). Collectively, these results emphasize that *IL8* is highly expressed in tumors from NSCLC patients with low muscularity and is associated with poor prognosis in this cancer type.

IL-8 treatment induces in vitro myotube atrophy

The ability of IL-8 in inducing muscle atrophy was evaluated by treating C2C12 myotubes with different concentrations of this cytokine (10, 100, and 1000 ng/mL). The myotubes treated with 100 ng/mL presented significant decrease in diameter compared to the control group after 24 hours (**Supplementary Figure 5**). The C2C12 myotubes treated with 100 ng/mL of IL-8 for 24 hours were evaluated by Myh2 immunostaining, which confirmed the significant decrease in myotubes diameter (**Figure 6a, b**). Myotubes treated with IL-8 also presented a higher number of myotubes with < 10 µm of diameter compared to the control group. Conversely, a higher number of controls myotubes with > 35 µm of diameter compared to those myotubes treated with IL-8 was observed (**Figure 6c**).

Discussion

Using a tumor transcriptome-based secretome analysis in NSCLC patients with low muscularity we aimed to identify potential cancer biomarkers of prognostic value and mediators of cancer-associated weight loss. This strategy revealed increased expression levels of cachexia-related genes predicted to be secreted in NSCLC from patients with lower PMA. These genes were further associated with shorter recurrence-free survival and decreased overall survival in different validation sets of patients with NSCLC. Importantly, increased expression levels of *IL8* was detected in high-risk group in all NSCLC validation sets, and IL-8 was sufficient to trigger atrophy in C2C12 myotubes.

Muscle depletion or low muscle mass in NSCLC patients identified by CT images has been extensively associated with poor outcome [20,49,85–88,90,91]. Previous studies using the same methodology to ours - the objective assessment of the PMA on CT scans -reported lower PMA associated with worse overall survival in NSCLC patients or in cases with chronic obstructive pulmonary disease, despite normalization for BMI and performance status [89,100]. Teigen *et al.*, reported that the PMA divided by height (used to standardize for body size) is a powerful predictor of outcome after left ventricular assist device implantation [116]. Unfortunately, the height in our cohort of PMA CT-based analysis was not available. However, the high quality of these CT images previously allowed the identification of new tumor radiomics features with prognostic value in NSCLC patients [58]. Thus, we hypothesized that the comparison of the PMA with muscle normalizations based on different radiomics features aiming the standardization for body size could reveal new approaches for screening muscularity in NSCLC patients. Interestingly, PMA distinguished NSCLC patients with low and high muscularity similarly in all muscle normalizations tested. Considering that CTs images of lung cancer patients are preferentially performed in the thoracic region, our data additionally confirm that PMA is a feasible measurement easily applied to the clinical practice to distinguish NSCLC patients with different muscularity.

Although a large range of changes in body composition has been associated with tumor-derived factors, including many pro-inflammatory cytokines [8,10,24,117], only few NSCLC studies associated CT-derived body composition with systemic inflammatory response [54,118]. These studies showed that lower muscularity was associated with systemic inflammatory response (IL-6, C-reactive protein and albumin blood levels, and neutrophil-to-lymphocyte ratio). However, the specific tumor-derived factors that induce muscle loss in NSCLC patients are still unknown. Using the tumor transcriptome analysis of NSCLC patients with low muscularity, we found 105 deregulated genes, of which 75 were up-regulated and 30 down-regulated. The functional enrichment analysis revealed up-regulated genes related to cytokine activity (*CSF3, IL8, IL6, BMP6, SCG2, CCL8, BMP2*) and extracellular space (*CSF3, FLRT2, IL8, PLA2G3, IL6, ATP1B1, COL14A1, LPL, HBB, ADAMTS4*). These results suggest that tumor of patients with low muscularity possibly secrete cachexia associated-factors.

The *in-silico* analysis confirming that a set of over-expressed genes are translated into proteins presented in the plasma or in the secretome of NSCLC patients. Seven of these predicted proteins (NCAM1, CNTN1, SCG2, CADM1, IL-8, NPTX1, and APOD) were identified in five databases (SignalP 4.1, SecretomeP 2.0, Vesiclepedia, Human Cancer Secretome, and Plasma Proteome), giving support to their relevance in NSCLC. Although not all NSCLC patients with low muscularity are cachectic, the tumor gene expression profile identified molecules, such as *IL6* and *IL8*, consistently linked to inflammation and cancer cachexia pathogenesis [109,111,119–126]. The low muscle mass detected by CT images can occur in the absence of systemic inflammation in other malignancies such as colorectal cancer, but the proportion of patients with low muscularity is substantially greater in the presence of systemic inflammation [127]. In cases where the inflammation coexists with low muscle mass, the prognosis is especially poor [128]. Taken together, we identified a specific set of up-

regulated genes coding for secreted proteins that may constitute potential mediators of muscle loss in NSCLC.

Based on the fact that circulating levels of tumor derived factors were correlated with cachexia development and predicted outcome in cancer [109,111,119–126], we also investigated the predictive potential of seven transcripts (*NCAM1*, *CNTN1*, *SCG2*, *CADM1*, *IL8*, *NPTX1*, and *APOD*). All of them were associated with shorter overall survival and recurrence free survival for the predicted high-risk groups in the NSCLC validation set. However, only *IL8* was over-expressed in the high-risk group in all cohorts of our NSCLC validation set. We further confirmed that high *IL8* expression level in tumor tissue is a strong predictive biomarker significantly associated with worse survival (validation cohort of 1053 NSCLC patients). In agreement with our results, IL-8 expression in tumor tissues were recently associated with cachectic status and outcome in pancreatic cancer; cachectic patients with high IL-8 expression in tumor tissues had shorter overall survival or disease-free survival [111]. Importantly, these authors also showed that IL-8 expression level in tumor specimen paired with serum sample from the same patients were associated with tumor size.

We demonstrated that IL-8 can directly induce myotube atrophy, reinforcing its potential as a new mediator of cancer cachexia. Muscle wasting in cancer cachexia has been attributed to the combinatorial action of mediators from host and tumor microenvironment [8,10,24,117]. In addition, tumor expression and serum levels of IL-8 have been associated with muscle wasting in patients with different tumor types [109,111,119–123]. Yet, the potential direct effect of IL-8 in inducing muscle cell atrophy is still unknown. In this study, we provide evidence that IL-8 is a biomarker of worse prognosis that has the potential to define the cachectic state in NSCLC patients

The main strength of the present investigation is the identification of potential tumor-derived mediators of muscle wasting in patients with low muscularity, which have prognostic value in

NSCLC. However, our study is based on the reuse of transcriptomic and clinical data, which results in limitations that can be pointed out. Firstly, the validation of the findings at protein levels in NSCLC patients with low muscularity would be a strategy to define the cachexia blood biomarkers useful for clinical routine. Secondly, our survival analyses were restricted to the validation set; the survival information was not available in our discovery dataset. Finally, since the *IL8* gene is not present in the rodent genome, the atrophy phenotype observed in mice myotubes was likely induced by orthologue receptors to the human IL-8 [129]. In agreement with our study, Gerber et al., reported that IL-8 protein expression was significantly associated with tumor free body weight and skeletal muscle weight in a human pancreatic cancer xenograft mouse model [130]. Further studies are needed to elucidate the mechanisms of action of IL-8 in human muscle cells.

In conclusion, our study demonstrated that PMA is a clinical and practical method to distinguish NSCLC patients with different muscularity from routinely acquired CT images. Tumors from patients with low muscularity have a set of up-regulated genes coding for secreted proteins within the tumor microenvironment, including pro-inflammatory cytokines, which predict worse overall survival in NSCLC. Among the up-regulated genes, high *IL8* expression in tumor tissues is also associated with worse prognosis in NSCLC and recombinant IL-8 is capable to trigger atrophy in C2C12 myotubes.

Acknowledgements

The results shown here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/. We would like to thank Dr. Jean-Marie Delaissé and Jacob Bastholm Olesen, Vejle Hospital, University of Southern Denmark, for the contributions with the immunofluorescence assay. The authors of this manuscript certify that they comply with the ethical guidelines for authorship and publishing in the Journal of Cachexia, Sarcopenia and Muscle.

Funding

The following grant supported the development of this work: São Paulo Research Foundation – FAPESP (grant #17/21223-9).

Conflict of interests

Sarah Santiloni Cury, Diogo de Moraes, Paula Paccielli Freire, Grasieli de Oliveira, Douglas Venâncio Pereira Marques, Geysson Javier Fernandez, Maeli Dal-Pai-Silva, Érica Nishida Hasimoto, Patricia Pintor dos Reis, Silvia Regina Rogatto, Robson Francisco Carvalho declare that they have no conflict of interest.

Table 1 Clinical findings and skeletal muscle parameters of patients with NSCLC with low and high muscularity defined by pectoralis muscle cross-sectional area assessed by computed tomography.

Characteristics	*All*	*Men*	*Women*	*p value**
Number of patients	89	60	29	
Age	65.2 ± 8.7	66.9 ± 7.3	61.8 ± 10.3	0.011 [a]
Cancer Stage (%)				
Early Stages (I-II)	79.7	74.6	90	0.08 [b]
Advanced Stages (III-IV)	20.2	25.4	10	
Histological Type (%)				
Adenocarcinoma	47.2	37.3	66.6	0.02 [b]
Squamous Cell Carcinoma	40.4	49.2	23.3	
Other Subtypes	12.4	13.5	10.1	
PMA (cm^2)	37.3 ± 11.1	42.5 ± 9.3	27 ± 6.0	<0.001 [a]
HM (N)	59	40	19	
LM (N)	30	20	10	
LM PMA (cm^2)	28.6 ± 6.5 [#c]	32.3 ± 4.3 [#c]	21 ± 1.4 [#c]	<0.001 [a]
HM PMA (cm^2)	41.7 ± 10.2	47.5 ± 6.3	30 ± 5.2	<0.001 [a]

N: number of patients; PMA: Pectoralis Muscle Area; LM: low muscularity patients; HM: high-muscularity patients. The data represent mean ± standard deviation. a: Student's t-test; b: Chi-squared test; c: Mann-Whitney' U-test; * comparisons between men and women; # statistical difference between patients with low- and high-muscularity (P<0.001).

Figures

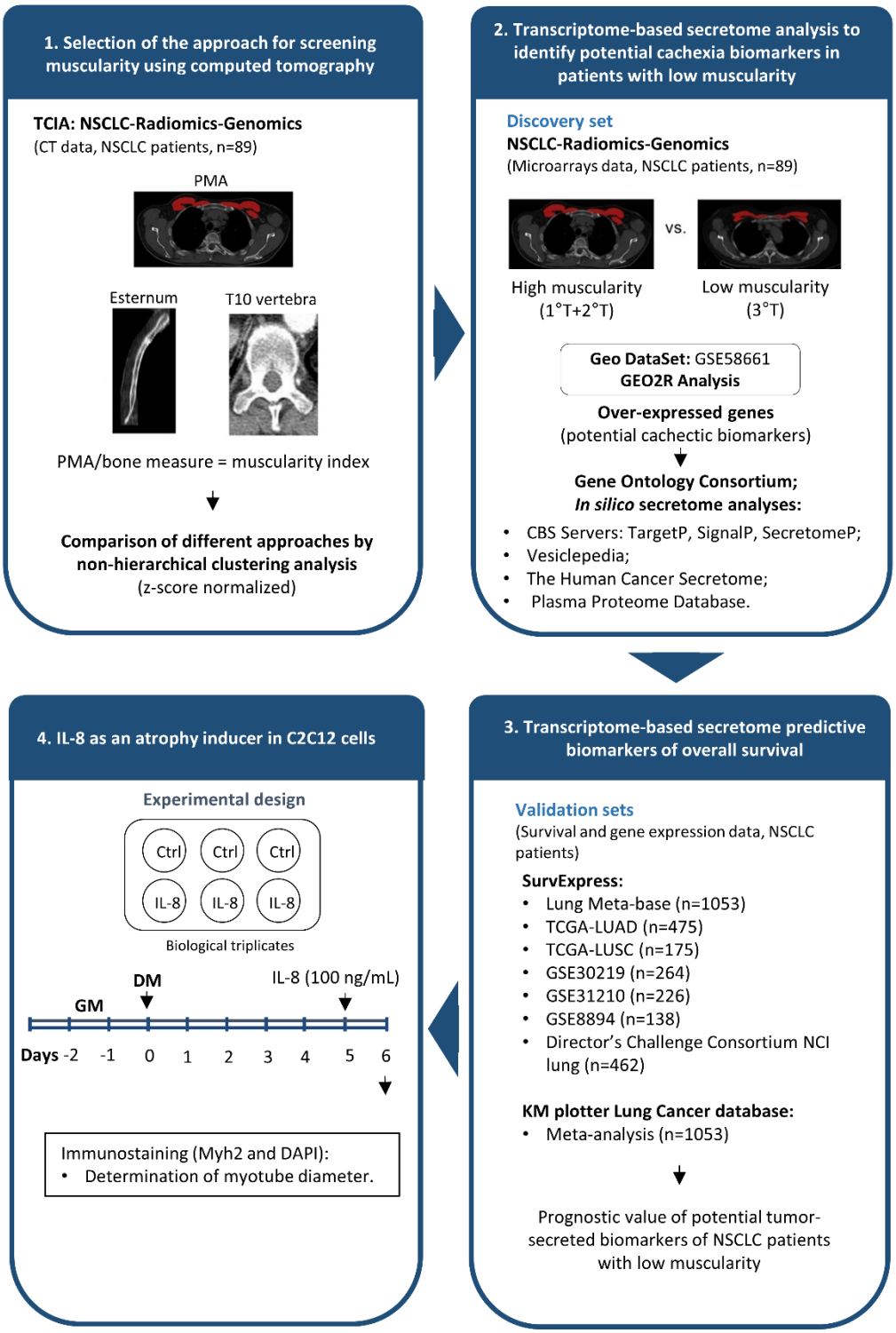

1. Selection of the approach for screening muscularity using computed tomography
TCIA: NSCLC-Radiomics-Genomics
(CT data, NSCLC patients, n=89)
PMA
Esternum
T10 vertebra
PMA/bone measure = muscularity index
Comparison of different approaches by non-hierarchical clustering analysis
(z-score normalized)
2. Transcriptome-based secretome analysis to identify potential cachexia biomarkers in patients with low muscularity
Discovery set
NSCLC-Radiomics-Genomics
(Microarrays data, NSCLC patients, n=89)
vs.
High muscularity
(1°T+2°T)
Low muscularity
(3°T)
Geo DataSet: GSE58661
GEO2R Analysis
Over-expressed genes
(potential cachectic biomarkers)
Gene Ontology Consortium;
In silico secretome analyses:
CBS Servers: TargetP, SignalP, SecretomeP;
Vesiclepedia;
The Human Cancer Secretome;
Plasma Proteome Database.
3. Transcriptome-based secretome predictive biomarkers of overall survival
Validation sets
(Survival and gene expression data, NSCLC patients)
SurvExpress:
Lung Meta-base (n=1053)
TCGA-LUAD (n=475)
TCGA-LUSC (n=175)
GSE30219 (n=264)
GSE31210 (n=226)
GSE8894 (n=138)
Director's Challenge Consortium NCI lung (n=462)
KM plotter Lung Cancer database:
Meta-analysis (n=1053)
Prognostic value of potential tumor-secreted biomarkers of NSCLC patients with low muscularity
4. IL-8 as an atrophy inducer in C2C12 cells
Experimental design
Ctrl
Ctrl
Ctrl
IL-8
IL-8
IL-8
Biological triplicates
DM
IL-8 (100 ng/mL)
GM
Days -2 -1 0 1 2 3 4 5 6
Immunostaining (Myh2 and DAPI):
Determination of myotube diameter.

Fig. 1 Workflow of the integrative analyses of CT images and tumor transcriptome data to identify secreted cachexia biomarkers in NSCLC patients with low muscularity. (1) We selected computed tomographies (CT) from 89 patients with non-small cell lung cancer (NSCLC) from "NSCLC-Radiomics-Genomics" collection [98], available on *The Cancer Imaging Archive* (TCIA, http://cancerimagingarchive.net/) database [43]. A total of 12 CTs features, including pectoralis muscle area (PMA), manubrium and sternum body lengths, six T10 different measures, and an anteroposterior length were used to determine an approach for screening muscularity. (2) This analysis revealed that PMA allows the identification of NSCLC patients with low (third tercile, 3^{rd}T) and high muscularity (first and second terciles, 1^{st}T + 2^{nd}T). These groups were compared by using a tumor transcriptomic-based secretome analysis (discovery set; microarray data; GSE58661) to identify potential cachexia biomarkers (over-expressed genes) in patients with low muscularity. Transcripts with increased expression were further analyzed to identify enriched terms by Gene Ontology Consortium and to predict potentially secreted proteins using secretome and microvesicle databases. (3) The performance of these transcripts as tumor biomarkers able to determine patients survival outcome were validated in multiple independent lung cancer validation sets available on SurvExpress [50] and KM plotter [73] databases. (4) C2C12 myoblast differentiated into myotubes were used to evaluate the ability of the selected biomarker (IL-8) in inducing atrophy. C2C12 mouse myoblasts were cultured in a growth medium (GM) for two days. Myoblasts with 80% to 90% of confluence were induced to differentiate in a differentiation medium (DM) for 5 days, when the cells were treated with recombinant IL-8 (100 ng/mL for 24h). Control myotubes (Ctrl) received a water solution containing bovine serum albumin 0.1%. Next, myotubes were fixed and immunostained for myosin heavy chain 2 (Myh2) and counter stained with DAPI. Finally, cellular atrophy parameters were evaluated in treated and control myotubes. n: number of patients

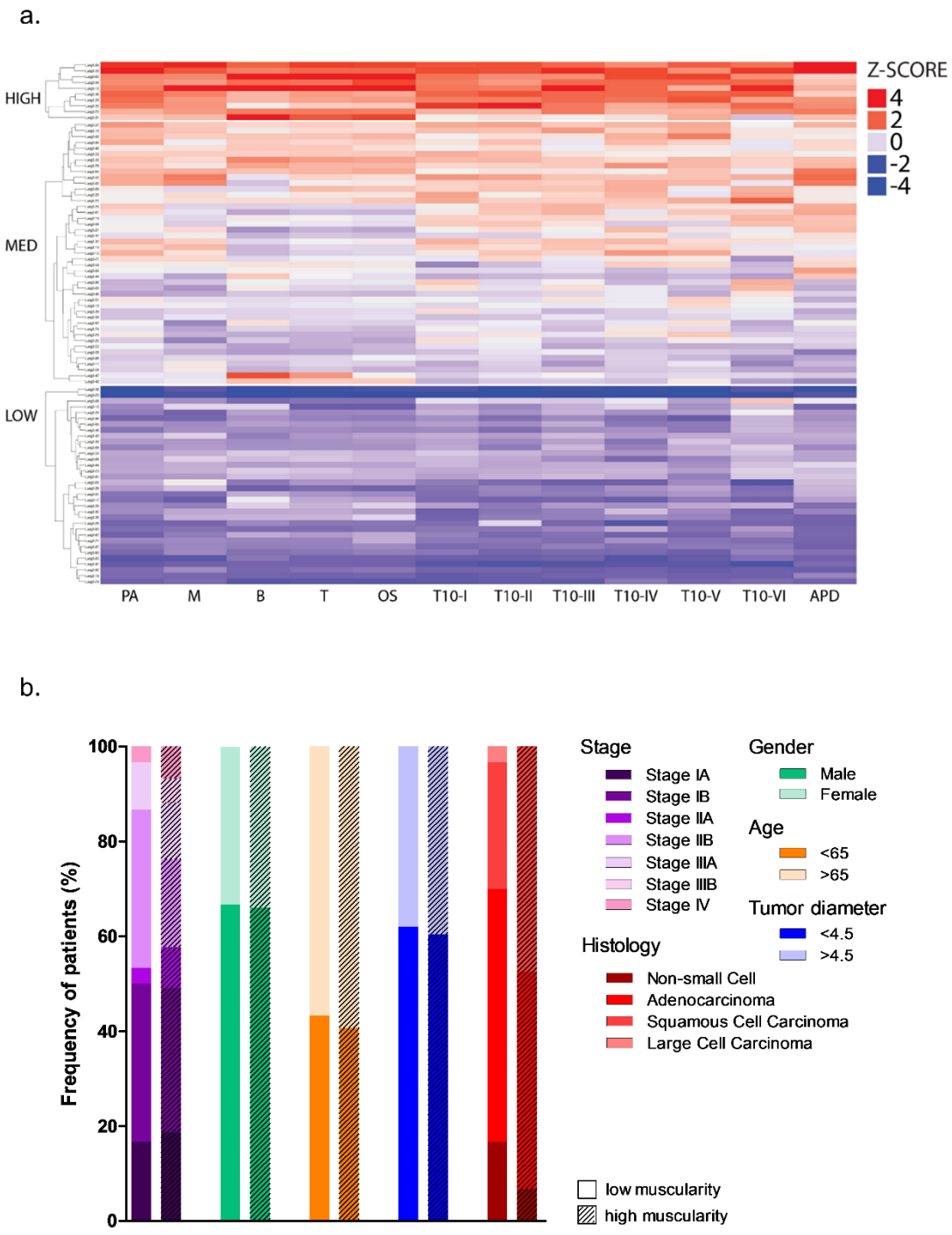


Fig. 2 Pectoralis muscle area as an approach for screening muscularity. (a) Heatmap showing patients stratification into high, medium, and low muscularity by a non-hierarchical k-means clustering analysis of the pectoralis muscle area (PA) and its normalization by eleven computed tomographies (CT) features that includes: manubrium and sternum body lengths, six T10 different

measures, and an anteroposterior length. Manubrium length (M); sternum Body length (B); T (M+B); total sternum length (OS); T10 body vertical length (T10-I); distance between T10 body and spinous process (T10-II); T10 body horizontal length (T10-III); distance between T10 pedicles (T10-IV); distance between T10 transverse processes (T10-V); T10 body area (T10-VI); anteroposterior distance (APD). (b) Bar graphs comparing the percentage of patients for clinical prognostic variables between high and low muscularity groups. These groups were generated based on the ordination of patients according to their PMA in descending order using a sex-specific categorical variable followed by segregation into terciles (high muscularity group: 1^{st} and 2^{nd} terciles; low muscularity group: 3^{rd} tercile).

a.

b.

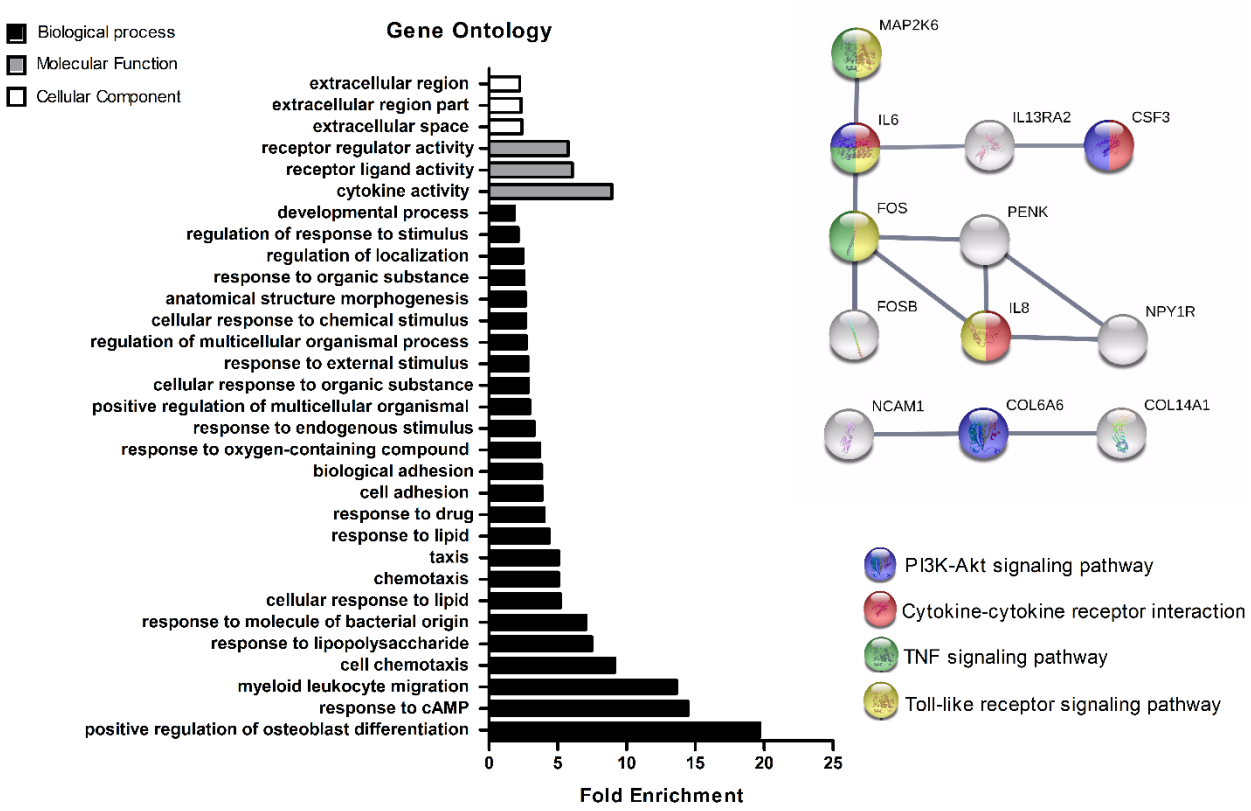


Fig. 3 Over-expressed genes in tumors from patients with low muscularity (a) Enriched terms in gene ontology analysis of the 75 transcripts up-regulated in patients with low muscularity. (b) Protein-protein interaction (PPI) network of 75 up-regulated transcripts in patients with low muscularity generated by STRING using a high confidence interaction score (0.700)

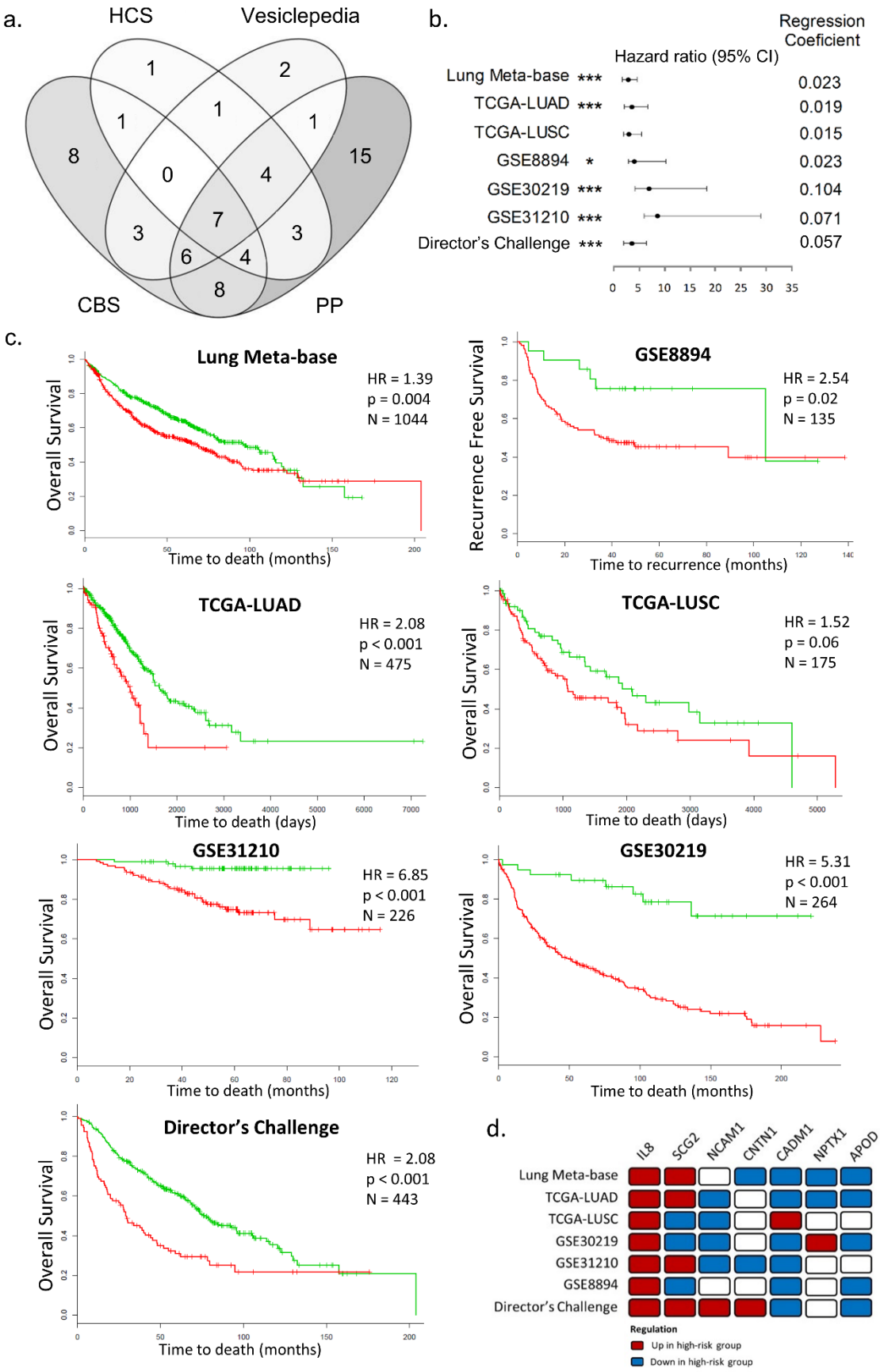

a.
HCS
Vesiclepedia
CBS
PP
1
2
1
1
1
8
0
4
15
7
3
3
6
4
8
b.
Regression Coeficient
Hazard ratio (95% CI)
Lung Meta-base *** 0.023
TCGA-LUAD *** 0.019
TCGA-LUSC 0.015
GSE8894 * 0.023
GSE30219 *** 0.104
GSE31210 *** 0.071
Director's Challenge *** 0.057
0 5 10 15 20 25 30 35
c.
Lung Meta-base
HR = 1.39
p = 0.004
N = 1044
Overall Survival
Time to death (months)
GSE8894
HR = 2.54
p = 0.02
N = 135
Recurrence Free Survival
Time to recurrence (months)
TCGA-LUAD
HR = 2.08
p < 0.001
N = 475
Overall Survival
Time to death (days)
TCGA-LUSC
HR = 1.52
p = 0.06
N = 175
Overall Survival
Time to death (days)
GSE31210
HR = 6.85
p < 0.001
N = 226
Overall Survival
Time to death (months)
GSE30219
HR = 5.31
p < 0.001
N = 264
Overall Survival
Time to death (months)
Director's Challenge
HR = 2.08
p < 0.001
N = 443
Overall Survival
Time to death (months)
d.
IL8
SCG2
NCAM1
CNTN1
CADM1
NPTX1
APOD
Lung Meta-base
TCGA-LUAD
TCGA-LUSC
GSE30219
GSE31210
GSE8894
Director's Challenge
Regulation
Up in high-risk group
Down in high-risk group

Fig. 4 Prognostic values of the potentially secreted proteins in tumors from patients with low muscularity. (a) The intersection of databases used for prediction of secreted proteins revealed seven overlapped proteins: IL8, SCG2, NCAM1, CNTN1, CADM1, NPTX1, and APOD. (b) Forest plots representing the set of potentially secreted biomarkers in each validation set. The horizontal axis represents confidence intervals estimated by using a Cox proportional hazards model. The asterisks represent the statistical significance in patient survival outcome (*** $p < 0.001$ and * $p < 0.05$, log-rank P-value). (c) Kaplan-Meier plots generated in SurvExpress [50]database for the NSCLC datasets (gene expression and survival or time to recurrence): Lung Meta-base, TCGA-LUAD, TCGA-LUSC, GSE30219, GSE31210, and Director's Challenge Consortium NCI. The Kaplan-Meier plot generated using the dataset GSE8894 was based on gene expression and time to recurrence data. (d) Direction of expression for the seven biomarkers in each validation set. N: number of patients; HR: adjusted hazard ratio; p: log-rank p-value determined from univariate Cox regression analyses (green curve: low-risk group; red curve: high-risk group); TCGA: The Cancer Genome Atlas; LUAD: Lung Adenocarcinoma; LUSC: Lung Squamous Cell Carcinoma

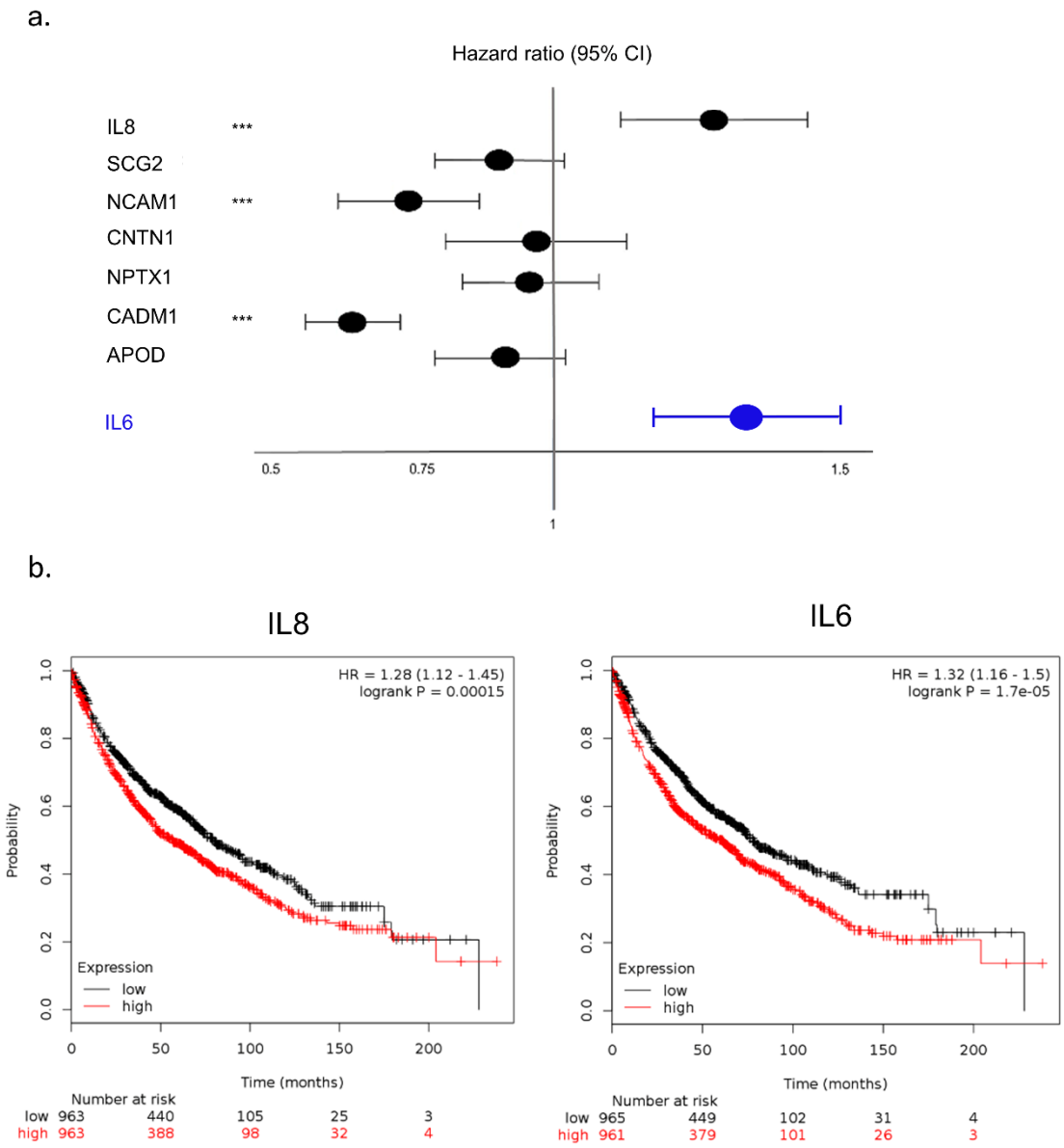


Fig. 5 IL8 is associated with poor prognosis in NSCLC. (a) Forest plot for each tumor biomarkers (*IL8*, *SCG2*, *NCAM1*, *CNTN1*, *CADM1*, *NPTX1*, and *APOD)* in NSCLC patients from the dataset available on KM plotter[73] database. The hazard ratio (HR) with 95% confidence intervals (CI) determined by Cox proportional hazards model are represented in the horizontal axis. *** represent the statistical significance in NSCLC patient survival outcome ($p < 0.001$; log-rank p-value). (b) Kaplan-Meier overall survival curves for *IL8* or *IL6* in NSCLC patients from the dataset available on KM plotter [[73]] database. The resulting p-values for the log-rank test are shown

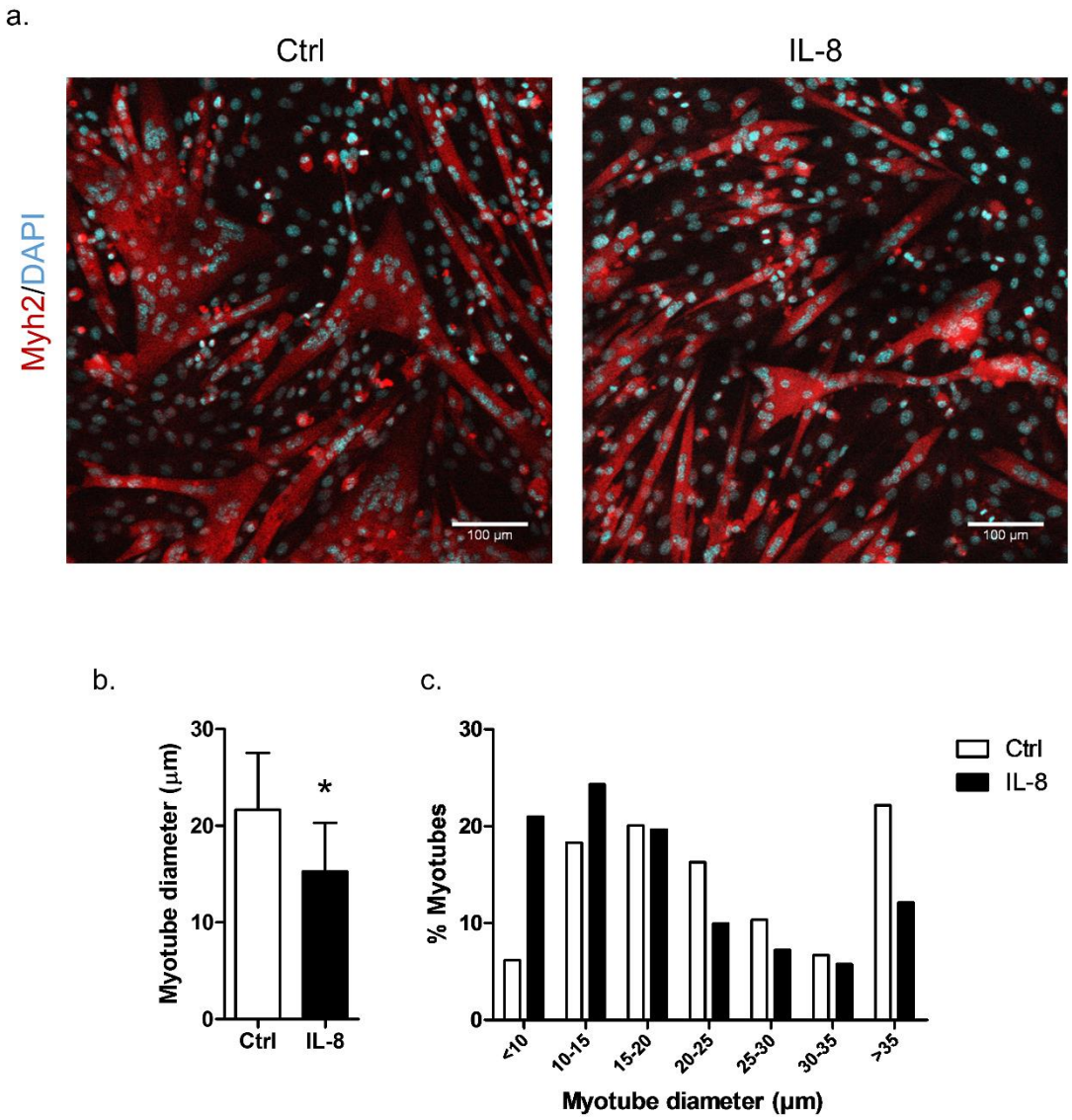
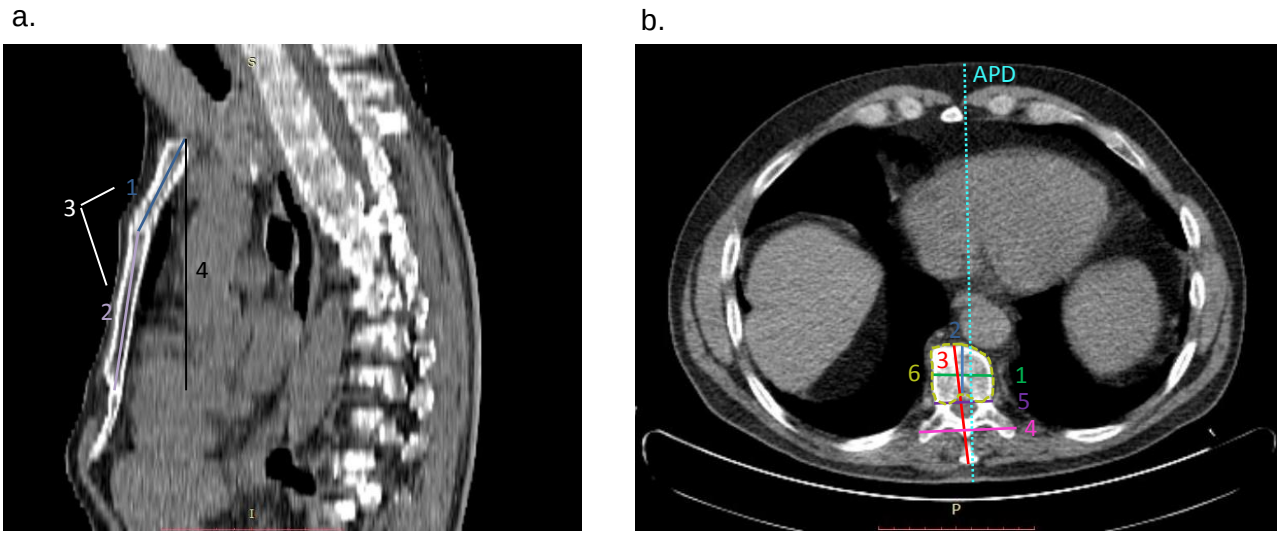


Fig. 6 IL-8 induces atrophy in C2C12 myotubes. (a) Immunofluorescence of C2C12 myotubes treated with recombinant IL-8 (100 ng/mL for 24 hours) immunostained for Myh2 (red) and the nuclei were counterstained with DAPI. (b) Myotube diameter (µm) quantification using ImageJ software. (c) Determination of the frequency of myotubes according to the diameter classes. The data represents the mean ± standard deviation from three independent experiments. The statistical significance was analyzed using the Student's T-test. * p value < 0.05. Ctrl: control myotubes; IL-8: myotubes treated with recombinant interleukin 8

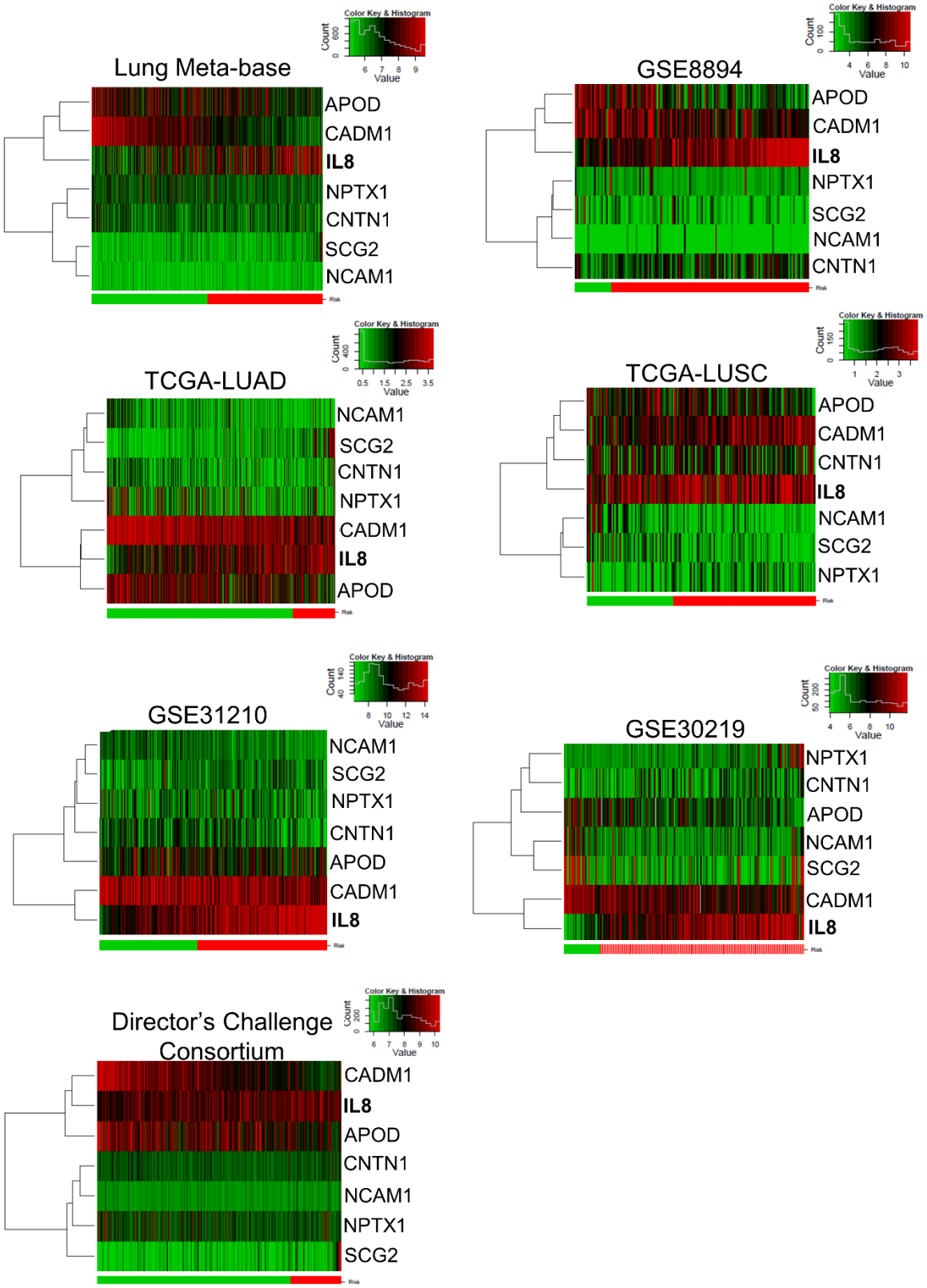
Supporting Material

Supplementary figure 1 Radiomics features (length or distance) used to test different pectoralis muscle area normalizations.

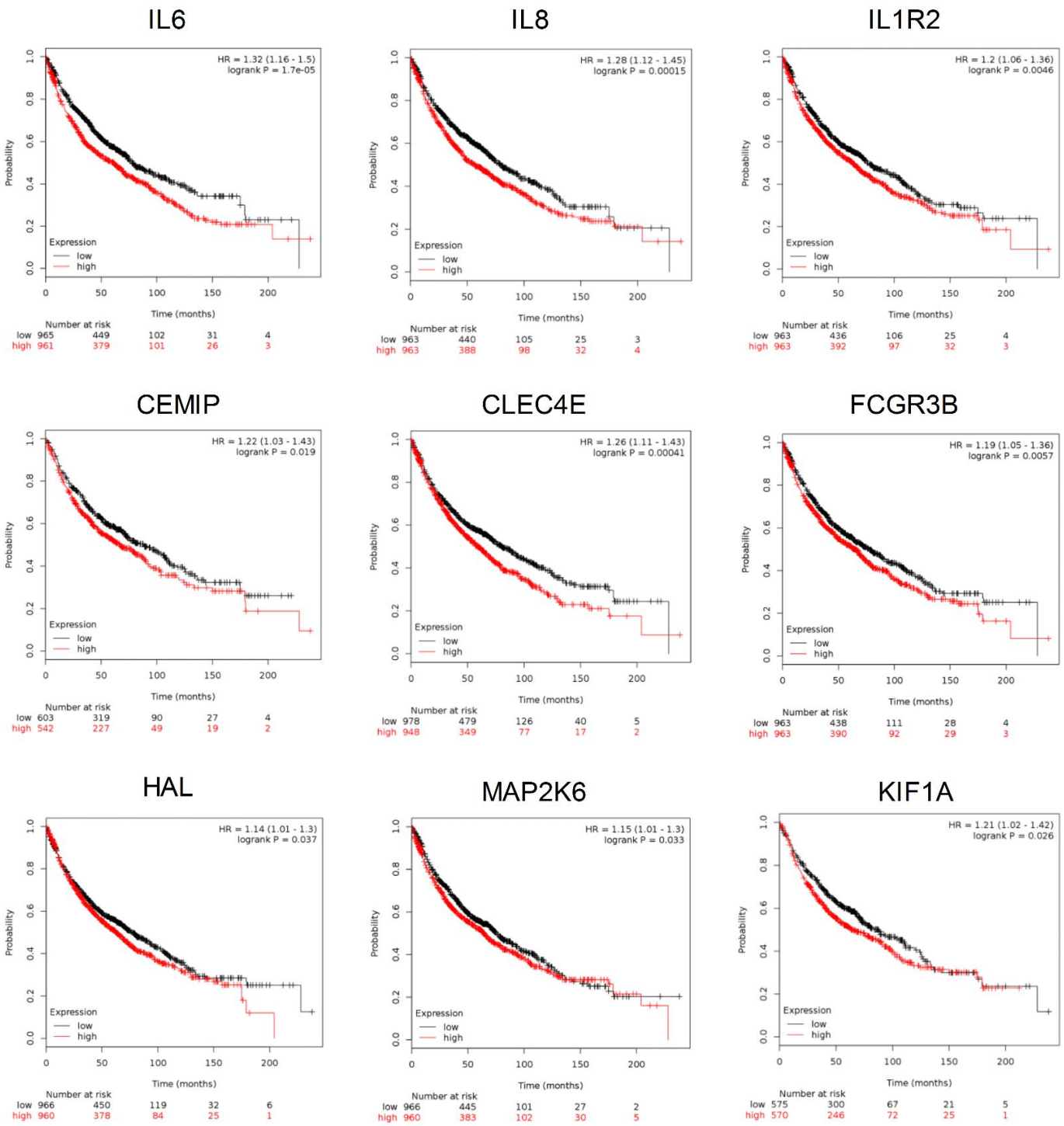


(a) 1: manubrium length; 2: sternum body length; 3: total of manubrium and sternum body lengths; 4: distance between the beginning of manubrium and the end of sternum body measured in 90º (without considering xiphoid process). (b) 1: horizontal length of T10 body; 2: vertical length of T10 body; 3: distance between T10 body and spinous process; 4: distance between transverse processes; 5: distance between pedicles; 6: T10 body area; APD: anteroposterior diameter. CT source The Cancer Imaging Archive (TCIA) (http://www.cancerimagingarchive.net/ [43]).

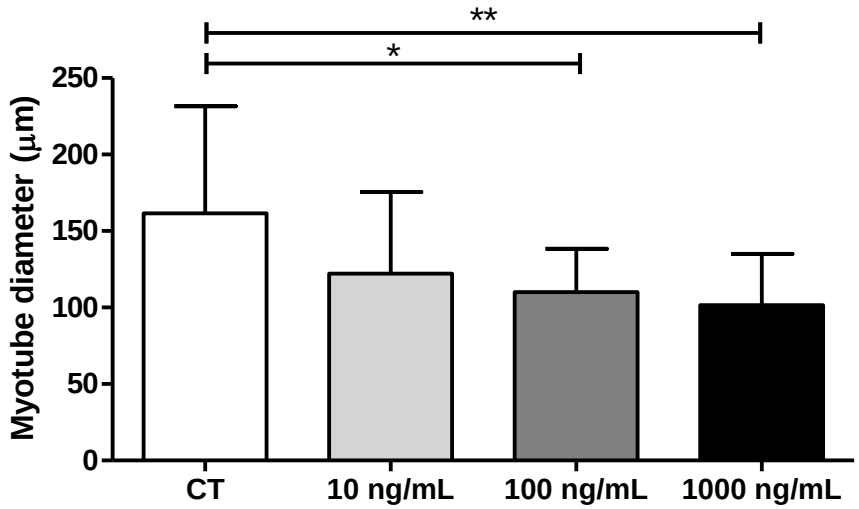
Supplementary figure 2 Heatmaps database showing the gene expression findings of seven potential biomarkers in low- and high-risk groups in NSCLC validation sets. (SurvExpress [107], http://bioinformatica.mty.itesm.mx:8080/Biomatec/SurvivaX.jsp)



Supplementary figure 3 Highly expressed mRNAs in NSCLC tissues from patients with low muscularity are associated with poor prognosis. Kaplan-Meier curves (KM plotter database [73], http://kmplot.com/analysis/index.php?p=service&cancer=lung) were generated in a cohort of 1053 NSCLC patients showing that the upregulated genes *IL6*, *IL8*, *IL1R2*, *CEMIP*, *CLEC4E*, *FCGR3B*, *HAL*, *MAP2K6*, and *KIF1A* are associated with worse survival.



Supplementary figure 4 IL-8 induces atrophy of C2C12 myotubes. Cells were treated with different concentrations (10, 100, 1000 ng/mL) of recombinant IL-8.



Myotube diameter of C2C12 cells exposed to recombinant IL-8. Experiments were performed in duplicate, and the data represent the mean ± standard deviation. Statistical analysis was performed using one-way analysis of variance test. * $p < 0.05$ and ** $p < 0.01$ compared with control group (CT).

Supplementary Table 1 mRNA differentially expressed in tumors of 30 NSCLC patients with low muscularity compared with the gene expression of 59 NSCLC patients with high muscularity.

Gene Symbol	Description	NCBI RefSeq	p-value	Fold Change
FGL1	Fibrinogen like 1	NM_004467	0.0377	2.25
PENK	Proenkephalin	NM_001135690	0.0151	2.22
BMPER	BMP binding endothelial regulator	NM_133468	0.0002	2.16
NPTX1	Neuronal pentraxin 1	NM_002522	0.0198	2.09
NCAM1	Neural cell adhesion molecule 1	NM_000615	0.0001	2.02
IL13RA2	Interleukin 13 receptor subunit alpha 2	NM_000640	0.0059	2
KIF1A	Kinesin family member 1A	NM_004321	0.0344	1.99
BMP2	Bone morphogenetic protein 2	NM_001200	0.0053	1.93
CNTN1	Contactin 1	BX648591	0.0475	1.89
HBB	Hemoglobin subunit beta	NM_000518	0.0371	1.84
SLN	Sarcopilin	NM_003063	0.0035	1.83
INSM1	INSM Transcriptional Repressor 1	NM_002196	0.0180	1.82
CADM1	Cell Adhesion Molecule 1	NM_014333	0.0036	1.82
MTMR7	Myotubularin Related Protein 7	NM_004686	0.0116	1.81
LINC01554	Long Intergenic Non-Protein Coding RNA 1554	NR_026936	0.0001	1.8
FOSB	FosB Proto-Oncogene, AP-1 Transcription Factor Subunit	NM_006732	0.0210	1.8
RNF182	Ring Finger Protein 182	NM_001165032	0.0082	1.8
IL6	Interleukin 6	NM_000600	0.0088	1.79
FAM177A1	Family With Sequence Similarity 177 Member A1	NM_001079519	0.0054	1.78
FCGR3B	Fc Fragment Of IgG Receptor IIIb	NM_000570	0.0165	1.77
PRR15	Proline Rich 15	NM_175887	0.0351	1.77
APOD	Apolipoprotein D	NM_001647	0.0206	1.75
HAL	Histidine Ammonia-Lyase	NM_002108	0.0256	1.73
SLC22A20	Solute Carrier Family 22 Member 20	NR_033396	0.0060	1.72
UBE2QL1	Ubiquitin Conjugating Enzyme E2 Q Family Like 1	NM_001145161	0.0078	1.72
CEMIP	Cell Migration Inducing Hyaluronan Binding Protein	NM_018689	0.0092	1.7
RASD1	Ras Related Dexamethasone Induced 1	NM_016084	0.0198	1.69
RAB3IP	RAB3A Interacting Protein	NM_001024647	0.0097	1.67
TCHH	Trichohyalin	NM_007113	0.0428	1.67
COL14A1	Collagen Type XIV Alpha 1 Chain	NM_021110	0.0265	1.66
NR4A3	Nuclear Receptor Subfamily 4 Group A Member 3	NM_006981	0.0078	1.66
CSF3	Colony Stimulating Factor 3	NM_000759	0.0117	1.66
COL6A6	Collagen Type VI Alpha 6 Chain	NM_001102608	0.0359	1.65
FLRT2	Fibronectin Leucine Rich Transmembrane Protein 2	NM_013231	0.0091	1.65
ITIH5	Inter-Alpha-Trypsin Inhibitor Heavy Chain Family Member 5	NM_030569	0.0015	1.64
PCSK2	Proprotein Convertase Subtilisin/Kexin Type 2	NM_002594	0.0267	1.63
SCARA5	Scavenger Receptor Class A Member 5	NM_173833	0.0293	1.63
HS3ST2	Heparan Sulfate-Glucosamine 3-Sulfotransferase 2	NM_006043	0.0325	1.62
AKAP12	A-Kinase Anchoring Protein 12	NM_005100	0.0051	1.62
BCAT1	Branched Chain Amino Acid Transaminase 1	NM_005504	0.0144	1.61
IL8/ CXCL8	C-X-C Motif Chemokine Ligand 8	NM_000584	0.0377	1.6

NR4A1	Nuclear Receptor Subfamily 4 Group A Member 1	NM_002135	0.0059	1.59
LPL	Lipoprotein Lipase	NM_000237	0.0440	1.59
ATP6V1C2	ATPase H+ Transporting V1 Subunit C2	NM_001039362	0.0382	1.59
KCNK3	Potassium Two Pore Domain Channel Subfamily K Member 3	NM_002246	0.0316	1.59
SV2B	Synaptic Vesicle Glycoprotein 2B	NM_014848	0.0107	1.59
NPY1R	Neuropeptide Y Receptor Y1	NM_000909	0.0494	1.58
CLEC4E	C-Type Lectin Domain Family 4 Member E	NM_014358	0.0086	1.58
IL1R2	Interleukin 1 Receptor Type 2	NM_004633	0.0334	1.58
VPS37A	VPS37A, ESCRT-I Subunit	NM_152415	0.0042	1.57
CD300E	CD300e Molecule	NM_181449	0.0054	1.57
HIST2H2BF	Histone Cluster 2 H2B Family Member F	NM_001024599	0.0075	1.57
HYMAI	Hydatidiform Mole Associated And Imprinted (Non-Protein Coding)	NR_002768	0.0016	1.56
ADAMTS4	ADAM Metallopeptidase With Thrombospondin Type 1 Motif 4	NM_005099	0.0162	1.55
PLA2G3	Phospholipase A2 Group III	NM_015715	0.0259	1.55
MAP2K6	Mitogen-Activated Protein Kinase Kinase 6	AK225719	0.0035	1.54
BMP6	Bone Morphogenetic Protein 6	NM_001718	0.0114	1.54
RDH10	Retinol Dehydrogenase 10	NM_172037	0.0241	1.54
SCG2	Secretogranin II	NM_003469	0.0089	1.53
ATP1B1	ATPase Na+/K+ Transporting Subunit Beta 1	NM_001677	0.0100	1.53
SELE	Selectin E	NM_000450	0.0094	1.53
LYVE1	Lymphatic Vessel Endothelial Hyaluronan Receptor 1	NM_006691	0.0170	1.53
MAP2	Microtubule Associated Protein 2	NM_002374	0.0441	1.52
PGM5	Phosphoglucomutase 5	NM_021965	0.0458	1.52
NPNT	Nephronectin	NM_001184690	0.0154	1.52
PRKAA2	Protein Kinase AMP-Activated Catalytic Subunit Alpha 2	NM_006252	0.0462	1.52
CCL8	C-C Motif Chemokine Ligand 8	NM_005623	0.0419	1.52
PCK1	Phosphoenolpyruvate Carboxykinase 1	NM_002591	0.0300	1.51
RGS2	Regulator Of G Protein Signaling 2	NM_002923	0.0066	1.51
PDE4B	Phosphodiesterase 4B	NM_002600	0.0066	1.51
CCDC144A	Coiled-Coil Domain Containing 144A	NM_014695	0.0055	1.5
MMP19	Matrix Metallopeptidase 19	NM_002429	0.0013	1.5
FOS	Fos Proto-Oncogene, AP-1 Transcription Factor Subunit	NM_005252	0.0137	1.5
RSPO3	R-Spondin 3	NM_032784	0.0374	1.5
PLAGL1	PLAG1 Like Zinc Finger 1	NM_002656	0.0084	1.5
MIR205HG	MIR205 Host Gene (Non-Protein Coding)	NM_001104548	0.0242	-2.85
GPR87	G Protein-Coupled Receptor 87	NM_023915	0.0288	-2.24
COL17A1	Collagen Type XVII Alpha 1 Chain	NM_000494	0.0219	-2.23
PRSS21	Protease, Serine 21	NM_006799	0.0122	-2.03
CLDN1	Claudin 1	NM_021101	0.0452	-1.81
FAT2	FAT Atypical Cadherin 2	NM_001447	0.0458	-1.79
S100A14	S100 Calcium Binding Protein A14	NM_020672	0.0194	-1.78
EGFR	Epidermal growth factor receptor	NM_005228	0.0021	-1.75
HLA-DQB2	Major Histocompatibility Complex, Class II, DQ Beta 2	NM_001198858	0.0460	-1.75
HAS3	Hyaluronan Synthase 3	NM_005329	0.0467	-1.72
ATG9B	Autophagy Related 9B	NM_173681	0.0036	-1.71
PLAT	Plasminogen Activator, Tissue Type	NM_000930	0.0039	-1.7
FERMT1	Fermitin Family Member 1	NM_017671	0.0110	-1.7
CHP2	Calcineurin Like EF-Hand Protein 2	NM_022097	0.0211	-1.7
MMP28	Matrix Metallopeptidase 28	NM_001032278	0.0255	-1.62
NTF4	Neurotrophin 4	NM_006179	0.0165	-1.62
SH3RF2	SH3 Domain Containing Ring Finger 2	NM_152550	0.0217	-1.59
ALDH1L1	Aldehyde Dehydrogenase 1 Family Member L1	NM_012190	0.0362	-1.58
CALB2	Calbindin 2	NM_001740	0.0394	-1.58
DLK1	Delta Like Non-Canonical Notch Ligand	NM_003836	0.0373	-1.58
DDX43	DEAD-Box Helicase 43	NM_018665	0.0411	-1.58
GALNT6	Polypeptide N-Acetylgalactosaminyltransferase 6	NM_007210	0.0046	-1.56
NECTIN1	Nectin Cell Adhesion Molecule 1	NM_002855	0.0085	-1.56
OOEP	Oocyte Expressed Protein	NM_001080507	0.0238	-1.54
LAD1	Ladinin 1	NM_005558	0.0007	-1.54
KRT19	Keratin 19	NM_002276	0.0081	-1.53
MELTF	Melanotransferrin	NM_033316	0.0306	-1.52
ARNTL2	Aryl Hydrocarbon Receptor Nuclear Translocator Like 2	NM_020183	0.0382	-1.52
TUBA4B	Tubulin Alpha 4b	NR_003063	0.0144	-1.51
CX3CL1	C-X3-C Motif Chemokine Ligand 1	NM_002996	0.0246	-1.5

The transcriptomic data from “NSCLC-Radiomics-Genomics” TCIA collection were accessed using GEO (GSE58661).
RefSeq: NCBI reference sequence; TCIA: The Cancer Imaging Archive; GEO: Gene Expression Omnibus.

Supplementary Table 2 Potentially secreted proteins identified in secretome-related databases based on the over-expressed genes in NSCLC patients with low muscularity.

mRNA	*CBS Server*	*Lung Cancer Secretome*	*Lung Cancer Vesiclepedia*	*Plasma Proteome*	*Total*
IL8	+	+	+	+	4
APOD	+	+	+	+	4
CADM1	+	+	+	+	4
NPTX1	+	+	+	+	4
CNTN1	+	+	+	+	4
SCG2	+	+	+	+	4
NCAM1	+	+	+	+	4
BMP2	+	-	+	+	3
BMP6	+	-	+	+	3
ITIH5	+	-	+	+	3
LPL	+	-	+	+	3
MMP19	+	-	+	+	3
FAM177A1	+	-	+	+	3
AKAP12	-	+	+	+	3
BCAT1	-	+	+	+	3
COL14A1	-	+	+	+	3
HBB	-	+	+	+	3
IL1R2	+	+	-	+	3
IL6	+	+	-	+	3
LYVE1	+	+	-	+	3
PENK	+	+	-	+	3
BMPER	+	-	-	+	2
CCL8	+	-	-	+	2
CSF3	+	-	-	+	2
FCGR3B	+	-	-	+	2
FLRT2	+	-	-	+	2
SELE	+	-	-	+	2
FOSB	+	-	-	+	2
NR4A1	+	-	-	+	2
IL13RA2	+	-	+	-	2
NPNT	+	-	+	-	2
PRR15	+	-	+	-	2
CEMIP	+	+	-	-	2
ATP1B1	-	+	+	-	2
ADAMTS4	-	-	+	+	2
KIF1A	-	+	-	+	2
MAP2	-	+	-	+	2
PCSK2	-	+	-	+	2
RSPO3	+	-	-	+	2

CD300E	+	-	-	-	1
COL6A6	+	-	-	-	1
FGL1	+	-	-	-	1
PCSK2	+	-	-	-	1
CLEC4E	+	-	-	-	1
LINC01554	+	-	-	-	1
UBE2QL1	+	-	-	-	1
PLA2G3	-	+	-	-	1
NPY1R	-	-	+	-	1
RDH10	-	-	+	-	1
FGL1	-	-	-	+	1
FOS	-	-	-	+	1
HAL	-	-	-	+	1
INSM1	-	-	-	+	1
KCNK3	-	-	-	+	1
MAP2K6	-	-	-	+	1
NR4A3	-	-	-	+	1
PCK1	-	-	-	+	1
PDE4B	-	-	-	+	1
PGM5	-	-	-	+	1
PRKAA2	-	-	-	+	1
RGS2	-	-	-	+	1
TCHH	-	-	-	+	1
VPS37A	-	-	-	+	1

(+): indicates the presence of the protein in the database; (-): indicates the absence of the protein in the database. CBS users: TargetP 1.1 [67], SignalP 4.0 [65], SecretomeP 2.0 [66]; Lung Cancer Secretome: Human Cancer Secretome Database [71]; Vesiclepedia [68]; Plasma Proteome database [72].

5. Considerações Finais

A caquexia associada ao câncer é uma condição que implica em baixa qualidade de vida e prognóstico desfavorável aos pacientes com câncer de pulmão de células não pequenas. Portanto, a compreensão dos mecanismos moleculares que diferem nos tumores dos pacientes que desenvolvem esta doença torna-se indispensável para obtenção de novos biomarcadores e alvos terapêuticos.

Desta forma, concluímos que a análise da área dos músculos peitorais é uma metodologia eficiente para distinguir pacientes com diferentes muscularidades através de imagens de TCs. Tumores de pacientes com CPCNP e baixa muscularidade apresentam um conjunto de genes com expressão aumentada que codificam proteínas possivelmente secretadas, incluindo as citocinas pró-inflamatórias, as quais predizem pior sobrevida em pacientes com CPCNP. Dentre esses genes com aumento de expressão, a *IL8* está associada ao pior prognóstico de pacientes com CPCNP e, além disso, a IL-8 recombinante é capaz de induzir diretamente atrofia em miotubos C2C12.

6. *Referências Bibliográficas*

1. Evans, W. J. *et al.* Cachexia: A new definition. *Clin. Nutr.* **27,** 793–799 (2008).
2. Fearon, K. *et al.* Definition and classifi cation of cancer cachexia: an international consensus. *Lancet Oncol.* **12,** 489–495 (2011).
3. Cruz-Jentoft, A. *et al.* Sarcopenia : European consensus on definition and diagnosis Report of the European Working Group on Sarcopenia in Older People. *Age Ageing* **39,** 412–423 (2010).
4. Cohen, S., Nathan, J. A. & Goldberg, A. L. Muscle wasting in disease : molecular mechanisms and promising therapies. *Nat. Publ. Gr.* **14,** 58–74 (2015).
5. Busquets, S., Stemmler, B. & Lo, F. J. Cachexia and sarcopenia : mechanisms and potential targets for intervention. *Curr Opin Pharmacol.* **22,** 100–106 (2015).
6. Haehling, S. Von, Anker, M. S. & Anker, S. D. Prevalence and clinical impact of cachexia in chronic illness in Europe , USA , and Japan : facts and numbers update 2016. *J Cachexia Sarcopenia Muscle* **7,** 507–509 (2016).
7. Donohoe, C. L., Ryan, A. M. & Reynolds, J. V. Cancer Cachexia : Mechanisms and Clinical Implications. *Gastroenterol Res Pr.* **2011,** 601430 (2011).
8. Baracos, V. E., Martin, L., Korc, M., Guttridge, D. C. & Fearon, K. C. H. Cancer-associated cachexia. *Nat. Publ. Gr.* **4,** 1–18 (2018).
9. Tisdale, M. J. Cachexia in cancer patients. *Nat. Rev. cancer* **3,** 883–9 (2002).
10. Tsoli, M. & Robertson, G. Cancer cachexia : malignant inflammation, tumorkines, and metabolic mayhem. *Trends Endocrinol Metab* **24,** 174–183 (2013).
11. Fox, K. M., Brooks, J. M., Gandra, S. R., Markus, R. & Chiou, C. Estimation of Cachexia among Cancer Patients Based on Four Definitions. *J Oncol* **2009,** 693458 (2009).
12. Muscaritoli, M., Fanelli, F. R. & Molfino, A. Perspectives of health care professionals on cancer cachexia : results from three global surveys. *Ann Oncol* **27,** 2230–2236 (2016).
13. Petruzzelli, M. & Wagner, E. F. Mechanisms of metabolic dysfunction in cancer-associated cachexia. *Genes Dev* **30,** 489–501 (2016).
14. Cosper, P. & Leinwand, L. Myosin heavy chain is not selectively decreased in murine cancer cachexia. *Int J Cancer* **130,** 2722–2727 (2012).
15. Muscaritoli, M., Anker, S. & Argiles, J. Consensus definition of sarcopenia, cachexia and pre-cachexia: joint document ela- borated by Special Interest Groups (SIG) 'cachexia–anorexia in chronic wasting diseases' and 'nutrition in geriatrics'. *Clin Nutr* **29,** 154–9 (2010).
16. Argilés, J. M., Busquets, S., Stemmler, B. & López-soriano, F. J. Cancer cachexia : understanding the molecular basis. *Nat. Publ. Gr.* **14,** 754–762 (2014).
17. Martin, L. *et al.* Cancer Cachexia in the Age of Obesity : Skeletal Muscle Depletion Is a Powerful Prognostic Factor , Independent of Body Mass Index. **31,** (2016).
18. Miyamoto, Y., Baba, Y., Sakamoto, Y. & Ohuchi, M. Sarcopenia is a Negative Prognostic Factor After Curative Resection of Colorectal Cancer. *Ann Surg Oncol* **22,** 2663–2668 (2015).
19. van Vledder, M. G. Van *et al.* Body composition and outcome in patients undergoing resection of colorectal liver metastases. *Br J Surg.* **99,** 550–557 (2012).
20. Stene, G. B. *et al.* Changes in skeletal muscle mass during palliative chemotherapy in patients with advanced lung cancer. *Acta Oncol* **54,** 340–348 (2015).
21. Kim, E. Y., Kim, Y. S., Park, I. & Ahn, H. K. Prognostic Significance of CT-Determined Sarcopenia in Patients with Small-Cell Lung Cancer Characteristics of the Study Population. *J. Thorac. Oncol.* **10,** 1795–1799 (2015).
22. Kim, E. Y. *et al.* Prognostic significance of cachexia score assessed by CT in male patients with small cell lung cancer. *Eur. J. Cancer Care (Engl).* **27,** 1–6 (2017).
23. Nattenmüller, J. *et al.* Prognostic Impact of CT-Quantified Muscle and Fat Distribution before and after First-Line-Chemotherapy in Lung Cancer Patients. *PLoS One* **12,** e0169136 (2017).
24. Twelkmeyer, B., Tardif, N. & Rooyackers, O. Omics and cachexia. *Curr Opin Clin Nutr Metab Care* **20,** 181–185 (2017).
25. Patel, S., Ngounou Wetie, A., Darie, C. & Clarkson, B. Cancer secretomes and their place in supplementing other hallmarks of cancer. *Adv Exp Med Biol* **806,** 409–42 (2014).
26. Makridakis, M. & Vlahou, A. Secretome proteomics for discovery of cancer biomarkers. *Proteomics, J* **73,** 2291–305 (2010).
27. Tjalsma, H., Bolhuis, A., Jongbloed, J., Bron, S. & van Dijl, J. Signal peptide-dependent protein transport in Bacillus subtilis: a genome-based survey of the secretome. *Microbiol Mol Biol Rev* **64,** 515–547 (2000).
28. Schaaij-visser, T. B. M., Wit, M. De, Lam, S. W. & Jiménez, C. R. The cancer secretome, current status and

opportunities in the lung , breast and colorectal cancer context. *BBA - Proteins Proteomics* **1834,** 2242–2258 (2013).
29. Mukherjee, P. & Mani, S. Methodologies to decipher the cell secretome. *Biochim Biophys Acta* **1834,** 2226–32 (2013).
30. Papaleo, E., Gromova, I. & Gromov, P. Gaining insights into cancer biology through exploration of the cancer secretome using proteomic and bioinformatic tools. *Expert Rev. Proteomics* **14,** 1021–1035 (2017).
31. Bray, F. *et al.* Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Cli* **68,** 394–424 (2018).
32. Instituto Nacional de Cancer José Alencar Gomes da Silva. *INCA - Instituto Nacional de Câncer - Estimativa 2016*. *Ministério da Saúde Instituto Nacional de Cancer José Alencar Gomes da Silva* (2016). doi:978-85-7318-283-5
33. Mitchell, G. & Mitchell, C. Lung cancer. *Aust Fam Physician* **33,** 321–325 (2004).
34. Wahbah, M., Boroumand, N., Castro, C., El-zeky, F. & Eltorky, M. Changing trends in the distribution of the histologic types of lung cancer : a review of 4439 cases. *Ann Diagn Pathol* **11,** 89–96 (2007).
35. Hecht, S. S. Tobacco Smoke Carcinogens and Lung Cancer. *J Natl Cancer Inst* **91,** 1194–1210 (1999).
36. Unit, L., Hospital, R. M. & Road, D. Do patients with weight loss have a worse outcome when undergoing chemotherapy for lung cancers ? *Br J Cancer* **90,** 1905–1911 (2004).
37. Del Ferraro, C., Grant, M., Koczywas, M. & Dorr-Uyemura, L. Management of Anorexia-Cachexia in Late Stage Lung Cancer Patients. *J Hosp Palliat Nurs* **14,** (2012).
38. Miller, A. *et al.* Blockade of the IL-6 trans-signalling / STAT3 axis suppresses cachexia in Kras-induced lung adenocarcinoma. *Oncogene* **36,** 3059–3066 (2016).
39. Gomez, D. *et al.* Integrative proteomic and transcriptomic analysis provides evidence for TrkB (NTRK2) as a therapeutic target in combination with tyrosine kinase inhibitors for non-small cell lung cancer. *Oncotarget* **9,** 14268–14284 (2018).
40. Mitchell, K., Zingone, A., Toulabi, L., Boeckelman, J. & Ryan, B. Comparative Transcriptome Profiling Reveals Coding and Noncoding RNA Differences in NSCLC from African Americans and European Americans. *Clin Cancer Res* **23,** 7412–7425 (2017).
41. Penna, F., Busquets, S. & Argilés, J. M. Experimental cancer cachexia : Evolving strategies for getting closer to the human scenario. *Semin. Cell Dev. Biol.* **54,** 20–27 (2016).
42. Rung, J. & Brazma, A. Reuse of public genome-wide gene expression data. *Nat. Rev. Genet.* **14,** 89–99 (2013).
43. Clark, K. *et al.* The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J Digit Imaging* **26,** 1045–1057 (2013).
44. Tomczak, K., Czerwińska, P. & Wiznerowicz, M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* **19,** A68–77 (2015).
45. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* **41,** D991–5 (2013).
46. Hubbard, J. M., Cohen, H. J. & Muss, H. B. Incorporating Biomarkers Into Cancer and Aging Research. *J Clin Oncol* **32,** 2611–2616 (2014).
47. Bhuachalla, É. B. N. *et al.* Computed tomography diagnosed cachexia and sarcopenia in 725 oncology patients : is nutritional screening capturing hidden malnutrition? *J Cachexia Sarcopenia Muscle* 295–305 (2018). doi:10.1002/jcsm.12258
48. Mohammad, S., Kazemi-bajestani, R., Mazurak, V. C. & Baracos, V. Computed tomography-defined muscle and fat wasting are associated with cancer clinical outcomes. *Semin. Cell Dev. Biol.* **54,** 2–10 (2016).
49. Baracos, V. E., Reiman, T., Mourtzakis, M., Gioulbasanis, I. & Antoun, S. Body composition in patients with non-small cell lung cancer: a contemporary view of cancer cachexia with the use of computed tomography image analysis. *Am J Clin Nutr* **91,** 1133–1137 (2010).
50. Roeland, E. J. *et al.* Weight loss versus muscle loss : re-evaluating inclusion criteria for future cancer cachexia interventional trials. *Support. Care Cancer* **25,** 365–369 (2017).
51. Kim, E. Y., Kim, Y. S., Park, I. & Ahn, H. K. Evaluation of sarcopenia in small-cell lung cancer patients by routine chest CT. *Support. Care Cancer* **24,** 4721–4726 (2016).
52. Tamandl, D. *et al.* Markers of sarcopenia quantified by computed tomography predict adverse long-term outcome in patients with resected oesophageal or gastro-oesophageal junction cancer. *Eur Radiol* **26,** 1359–1367 (2016).
53. Kim, E. Y., Kim, Y. S., Seo, J., Park, I. & Ahn, H. K. The Relationship between Sarcopenia and Systemic Inflammatory Response for Cancer Cachexia in Small Cell Lung Cancer. *PLoS One* **11,** e0161125 (2016).
54. Jafri, R., Previgliano, C., Khandelwal, K. & Shi, R. Cachexia Index in Advanced Non-Small-Cell Lung Cancer Patients. *Clin Med Insights Oncol* **9,** 87–93 (2015).
55. Collins, J., Noble, S., Chester, J., Coles, B. & Byrne, A. The assessment and impact of sarcopenia in lung cancer : a systematic literature review. *BMJ Open* **4,** e003697 (2014).

56. Müller, M. J., Geisler, C., Pourhassan, M. & Glüer, C. Assessment and definition of lean body mass deficiency in the elderly. *Eur J Clin Nutr* **68,** 1220–1227 (2014).
57. Kinsey, C. M. & Cole, B. F. Lower Pectoralis Muscle Area is Associated with a Worse Overall Survival in Non-Small Cell Lung Cancer. *Cancer Epidemiol Biomarkers Prev* **26,** 38–43 (2017).
58. Aerts, H. J. W. L. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun* **3,** 5:4006 (2014).
59. Albertina, B. *et al.* Radiology Data from The Cancer Genome Atlas Lung Adenocarcinoma [TCGA-LUAD] collection. *The Cancer Imaging Archive* http://doi.org/10.7937/K9/TCIA.2016.JGNIHEP5 (2016). doi:10.7937
60. Kirk, S. *et al.* Radiology Data from The Cancer Genome Atlas Lung Squamous Cell Carcinoma [TCGA-LUSC] collection. *The Cancer Imaging Archive* http://doi.org/10.7937/K9/TCIA.2016.TYGKKFMQ (2016). doi:10.7937/K9/TCIA.2016.TYGKKFMQ
61. Colaprico, A., Silva, T., Olsen, C., Garofano, L. & Cava, C. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44,** e71 (2015).
62. Robinson, M., McCarthy, D. & Smyth, G. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
63. Cheung, K. *et al.* Extending gene ontology in the context of extracellular RNA and vesicle communication. *J. Biomed. Semantics* **7,** 19 (2016).
64. Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M. & Bairoch, A. UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. *Methods Mol Biol* **1374,** 23–54 (2016).
65. Petersen, T., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* **29,** 785–786 (2011).
66. Bendtsen, J., Jensen, L., Blom, N., Von Heijne, G. & Brunak, S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel* **17,** 349–356 (2004).
67. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* **300,** 1005–1016 (2000).
68. Kalra, H. *et al.* Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation. *PLoS Biol* **10,** e1001450 (2012).
69. Kim, D., Lee, J., Kim, S., Choi, D. & Yoon, Y. EVpedia: a community web portal for extracellular vesicles research. *Bioinformatics* **31,** 933–939 (2015).
70. Mathivanan, S., Fahner, C., Reid, G. & Simpson, R. ExoCarta 2012: database of exosomal proteins, RNA and lipids. *Nucleic Acids Res* **40,** 1241–1244 (2012).
71. Feizi, A., Banaei-esfahani, A. & Nielsen, J. Database tool HCSD : the human cancer secretome database. *Database (Oxford)* bav051 (2015). doi:10.1093/database/bav051
72. Nanjappa, V., Thomas, J., Marimuthu, A. & Muthusamy, B. Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Res* **42,** 959–965 (2014).
73. Surowiak, P. & Budczies, J. Online Survival Analysis Software to Assess the Prognostic Value of Biomarkers Using Transcriptomic Data in Non-Small-Cell Lung Cancer. *PLoS One* **8,** e82241 (2013).
74. Lee, E., Son, D., Kim, S. & Lee, J. Prediction of recurrence-free survival in postoperative non-small cell lung cancer patients by using an integrated model of clinical information and gene expression. *Clin Cancer Res* **14,** 7397–404 (2008).
75. Rousseaux, S., Debernardi, A., Jacquiau, B. & Vitte, A. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Transl Med* **5,** 186ra66 (2013).
76. Yamauchi, M., Yamaguchi, R., Nakata, A. & Kohno, T. Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. *PLoS One* **7,** e43923 (2012).
77. Okayama, H., Kohno, T., Ishii, Y. & Shimada, Y. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* **72,** 100–111 (2012).
78. Shedden, K., Taylor, J. & Enkemann, S. Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* **14,** 822–7 (2008).
79. Yaffe, D. & Saxel, O. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature* **270,** 725–7
80. Ferlay, J. *et al.* Cancer incidence and mortality worldwide : Sources , methods and major patterns in GLOBOCAN 2012. **386,** (2015).
81. Herbst, R. S., Morgensztern, D. & Boshoff, C. The biology and management of non-small cell lung cancer. *Nat. Publ. Gr.* **553,** 446–454 (2018).
82. Fearon, K. *et al.* Definition and classification of cancer cachexia: An international consensus. *Lancet Oncol.* **12,** 489–495 (2011).

83. Mytelka, D. S., Li, L. & Benoit, K. Post-diagnosis weight loss as a prognostic factor in non-small cell lung cancer. *J. Cachexia. Sarcopenia Muscle* **9,** 86–92 (2018).
84. Dahlberg, S. E. *et al.* Body mass index and its association with clinical outcomes for advanced non-small-cell lung cancer patients enrolled on eastern cooperative oncology group clinical trials. *J. Thorac. Oncol.* **8,** 1121–1127 (2013).
85. Martin, L. *et al.* Cancer cachexia in the age of obesity: Skeletal muscle depletion is a powerful prognostic factor, independent of body mass index. *J. Clin. Oncol.* **31,** 1539–1547 (2013).
86. Prado, C. M. M. *et al.* Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts : a population-based study. 629–635 doi:10.1016/S1470-2045(08)70153-0
87. Management, S., Care, S. & Assessment, G. Nutritional Status , Body Surface , and Low Lean Body Mass / Body Mass Index Are Related to Dose Reduction and Severe Gastrointestinal Toxicity Induced by Afatinib in Patients With Non-Small Cell Lung Cancer. 967–974 (2015).
88. Fløtten, Ø., Hjermstad, M. J., Aass, N. & Jordhøy, M. Low muscle mass is associated with chemotherapy-induced haematological toxicity in advanced non-small cell lung cancer. 1–7 (2015). doi:10.1016/j.lungcan.2015.07.001
89. Kinsey, C. M. *et al.* Lower Pectoralis Muscle Area Is Associated with a Worse Overall Survival in Non – Small. *Cancer Epidemiol. Biomarkers Prev.* **26,** 38–44 (2017).
90. Bye, A. *et al.* Muscle mass and association to quality of life in non-small cell lung cancer patients. *J. Cachexia. Sarcopenia Muscle* **8,** 759–767 (2017).
91. Kilgour, R. D. *et al.* Cancer-related fatigue: The impact of skeletal muscle mass and strength in patients with advanced cancer. *J. Cachexia. Sarcopenia Muscle* **1,** 177–185 (2010).
92. Hsiao, Y., Chu, L., Chen, J., Yeh, T. & Yu, J. Proteomic profiling of the cancer cell secretome: informing clinical research. *Expert Rev Proteomics* **14,** 737–756 (2017).
93. Pavlou, M. P. & Diamandis, E. P. The cancer cell secretome: A good source for discovering biomarkers? *J. Proteomics* **73,** 1896–1906 (2010).
94. Fukawa, T. *et al.* Excessive fatty acid oxidation induces muscle atrophy in cancer cachexia. *Nat. Med.* **22,** 666–671 (2016).
95. Schäfer, M. *et al.* Ataxin-10 is part of a cachexokine cocktail triggering cardiac metabolic dysfunction in cancer cachexia. *Mol. Metab.* **5,** 67–78 (2016).
96. McLean, J. B., Moylan, J. S., Horrell, E. M. W. & Andrade, F. H. Proteomic analysis of media from lung cancer cells reveals role of 14-3-3 proteins in cachexia. *Front. Physiol.* **6,** 1–8 (2015).
97. Aerts, H. J. W. L. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5,** (2014).
98. Aerts, Hugo J. W. L., Rios Velazquez, Emmanuel, Leijenaar, Ralph T. H., Parmar, Chintan, Grossmann, Patrick, Carvalho, Sara, … Lambin, P. Data From NSCLC-Radiomics-Genomics.The Cancer Imaging Archive. (2015). doi:http://doi.org/10.7937/K9/TCIA.2015.L4FRET6Z
99. Freymann, J., Kirby, J., Perry, J., Clunie, D. & Jaffe, C. Image data sharing for biomedical research--meeting HIPAA requirements for De-identification. *J Digit Imaging* **25,** 14–24 (2012).
100. McDonald, M. *et al.* Quantitative computed tomography measures of pectoralis muscle area and disease severity in chronic obstructive pulmonary disease. A cross-sectional study. *Ann Am Thorac Soc* **11,** 326–334 (2014).
101. Nilufer, G. & Ayse, Y. Estimation of stature and sex from sternal lengths : an autopsy study. 89–96 (2015). doi:10.1007/s12565-014-0235-0
102. Mccarthy, I. D., Mcgregor, A. H. & Coombs, R. R. H. Geometrical dimensions of the lower lumbar vertebrae – analysis of data from digitised CT images. 242–248 (2000).
103. Yaguchi, Y., Kumata, Y., Horikawa, M. & Kiyokawa, T. Clinical Significance of Area of Psoas Major Muscle on Computed Tomography after Gastrectomy in Gastric Cancer Patients. 145–149 (2017). doi:10.1159/000480520
104. Blake, J. *et al.* Gene Ontology Consortium: going forward. *Nucleic Acids Res* **43,** 1049–1056 (2015).
105. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45,** D362–D368 (2017).
106. Snel, B., Lehmann, G., Bork, P. & Huynen, M. A. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* **28,** 3442–4 (2000).
107. Aguirre-gamboa, R. *et al.* SurvExpress : An Online Biomarker Validation Tool and Database for Cancer Gene Expression Data Using Survival Analysis. *PLoS One* **8,** e74250 (2013).
108. Tomczak, K., Czerwińska, P. & Wiznerowicz, M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. (2014). doi:10.5114/wo.2014.47136
109. Penafuerte, C. A. *et al.* Identification of neutrophil-derived proteases and angiotensin II as biomarkers of cancer cachexia. *Br. J. Cancer* **114,** 680–687 (2016).

110. Kuroda, K. *et al.* Interleukin 6 is associated with cachexia in patients with prostate cancer. *Urology* **69,** 113–117 (2007).
111. Hou, Y.-C. *et al.* Elevated Serum Interleukin-8 Level Correlates with Cancer-Related Cachexia and Sarcopenia: An Indicator for Pancreatic Cancer Outcomes. *J. Clin. Med.* **7,** 502 (2018).
112. Kandarian, S. C. *et al.* Tumour-derived leukaemia inhibitory factor is a major driver of cancer cachexia and morbidity in C26 tumour-bearing mice. *J. Cachexia. Sarcopenia Muscle* **9,** 1109–1120 (2018).
113. Carson, J. A. & Baltgalvis, K. A. Interleukin 6 as a key regulator of muscle mass during cachexia. *Exerc. Sport Sci. Rev.* **38,** 168–76 (2010).
114. Silva, E. M. *et al.* High systemic IL-6 is associated with worse prognosis in patients with non-small cell lung cancer. *PLoS One* **12,** 1–12 (2017).
115. Pettersen, K. *et al.* Cancer cachexia associates with a systemic autophagy-inducing activity mimicked by cancer cell-derived IL-6 trans-signaling. *Sci. Rep.* **7,** 2046 (2017).
116. Teigen, L. *et al.* Preoperative Pectoralis Muscle Quantity and Attenuation by Computed Tomography Are Novel and Powerful Predictors of Mortality After Left Ventricular Assist Device Implantation. *Circ Hear. Fail* **10(9),** pii: e004069 (2017).
117. Fearon, K. C. H., Glass, D. J. & Guttridge, D. C. Cancer cachexia: mediators, signaling, and metabolic pathways. *Cell Metab.* **16,** 153–66 (2012).
118. Srdic, D. *et al.* Cancer cachexia, sarcopenia and biochemical markers in patients with advanced non-small cell lung cancer—chemotherapy toxicity and prognostic value. *Support. Care Cancer* **24,** 4495–4502 (2016).
119. Pfitzenmaier, J. *et al.* Elevation of cytokine levels in cachectic patients with prostate carcinoma. *Cancer* **97,** 1211–1216 (2003).
120. Tazaki, E. *et al.* Serum cytokine profiles in patients with prostate carcinoma. *Exp. Ther. Med.* **2,** 887–891 (2011).
121. Alves, M. J. *et al.* Adipose tissue fibrosis in human cancer cachexia: The role of TGFβ pathway. *BMC Cancer* **17,** 1–12 (2017).
122. Zhang, D., Song, B., Wang, S., Zheng, H. & Wang, X. Association of interleukin-8 with cachexia from patients with low-third gastric cancer. *Comp. Funct. Genomics* **2009,** 1–6 (2009).
123. Lerner, L. *et al.* Plasma growth differentiation factor 15 is associated with weight loss and mortality in cancer patients. *J. Cachexia. Sarcopenia Muscle* **6,** 317–324 (2015).
124. Kuroda, K. *et al.* Interleukin 6 Is Associated with Cachexia in Patients with Prostate Cancer. *Urology* **69,** 113–117 (2007).
125. Richey, L. M. *et al.* Defining cancer cachexia in head and neck squamous cell carcinoma. *Clin. Cancer Res.* **13,** 6561–6567 (2007).
126. Krzystek-Korpacka, M. *et al.* Impact of weight loss on circulating IL-1, IL-6, IL-8, TNF-α, VEGF-A, VEGF–C and midkine in gastroesophageal cancer patients. *Clin. Biochem.* **40,** 1353–1360 (2007).
127. Dolan, R. D. *et al.* The relationship between computed tomography-derived body composition, systemic inflammatory response, and survival in patients undergoing surgery for colorectal cancer. *J. Cachexia. Sarcopenia Muscle* (2018). doi:10.1002/jcsm.12357
128. van Dijk, D. P. J. *et al.* Host phenotype is associated with reduced survival independent of tumour biology in patients with colorectal liver metastases. *J. Cachexia. Sarcopenia Muscle* (2018). doi:10.1002/jcsm.12358
129. Dunstan, C. A. N. *et al.* Identification of two rat genes orthologous to the human interleukin-8 receptors. *J. Biol. Chem.* **271,** 32770–32776 (1996).
130. Gerber, M. H. *et al.* Local and Systemic Cytokine Profiling for Pancreatic Ductal Adenocarcinoma to Study Cancer Cachexia in an Era of Precision Medicine. *Int. J. Mol. Sci.* **19,** 3836 (2018).